

**Zur Berücksichtigung der geometrischen Objektunsicherheit
in der Geoinformatik**

Von der Fakultät Bauingenieur- und Vermessungswesen
der Universität Stuttgart
zur Erlangung des akademischen Grades
Doktor-Ingenieur (Dr.-Ing.) genehmigte Abhandlung

Vorgelegt von
Dipl.-Ing. Michael Glemser
aus Löffingen

Institut für Photogrammetrie der Universität Stuttgart

2000

Hauptberichter: Prof. Dr.-Ing. habil. D. Fritsch
Mitberichter: Prof. Dr.-Ing. R. Bill
Prof. Dr.-Ing. habil. Dr. tech. h.c. mult. Dr.-Ing. E.h. Erik W. Grafarend

Tag der mündlichen Prüfung: 30. Januar 2001

Zusammenfassung

Geo-Daten sind stets zu einem gewissen Grad mit Unsicherheit behaftet. Dennoch ermöglichen es die heute verfügbaren Geo-Informationssysteme nicht, die Unsicherheit in den verschiedenen Verarbeitungsschritten einer Anwendung zu berücksichtigen. Eine solche Situation ist aber nur dann zu tolerieren, wenn ausschließlich einfache Operationen innerhalb einer abgeschlossenen Anwendung benutzt werden. Nur in diesem Fall reicht das Fachwissen eines erfahrenen Experten aus, um die Qualität zuverlässig prüfen zu können. Moderne Anwendungen greifen dagegen immer häufiger auf externe Datenbestände zurück und setzen vermehrt komplexe Operatoren zur Problemlösung ein, die eine einfache Prüfung nicht mehr zulassen. Aus diesem Grund ist eine vollständige Integration der Unsicherheit in den Arbeitsablauf einer GIS-Umgebung zu realisieren.

Die Arbeit hat zum Ziel, Modelle und Methoden zur Berücksichtigung der Unsicherheit zu entwickeln. Im Besonderen konzentrieren sich die Untersuchungen auf die Behandlung der Geometrie als wichtigste Komponente von räumlichen Daten. Die Phänomene der realen Welt werden dabei durch eine Objektsichtweise abgebildet. Ausgangspunkt für die Entwicklung von Modellen bildet die Diskussion von Ursachen, die zum Entstehen der Unsicherheit beitragen. Wesentliche Einflüsse sind bereits innerhalb der Erfassung im Rahmen der Modellbildung, Diskretisierung und Messung aufzufinden. Die Gesamtheit der Einflüsse ist schließlich durch ein geeignetes Unsicherheitsmodell zu beschreiben. Aufgrund der vielfältigen Anwendungen kann aber kein universelles Unsicherheitsmodell definiert werden, das allen Anforderungen der Praxis gerecht wird. Vielmehr sind hierfür mehrere Alternativen aufzuzeigen. Aufbauend auf bekannten theoretischen Grundlagen werden drei verschiedene Modelle, das Stochastische Modell, das Minimum-Maximum-Modell und das Fuzzy-Modell entwickelt und in das bestehende Objektmodell integriert.

Für Anwendungen ist es wichtig, dass die Datenunsicherheit in die verschiedenen Komponenten eines GIS einfließt. Die Erfassungskomponente erhält die zusätzliche Aufgabe, die Werte der definierten Unsicherheitsmaße festzulegen. Verschiedene Verfahren werden vorgestellt, die beispielsweise über Mehrfacherfassungen oder durch Vergleiche mit übergeordneten Referenzdatensätzen die gesuchten Werte ermitteln. Neben der üblichen Visualisierung dieser Werte in Form von darstellbaren Attributen zu einem Objekt interessiert bei der Präsentation im Besonderen die räumliche Auswirkung der Unsicherheit. Als mögliche Techniken können Isolinien- oder Matrixdarstellungen eingesetzt werden. Weiterhin sind zur Verwaltung der Unsicherheitsmaße geeignete Speicherkonzepte zu entwickeln. Dafür bietet sich eine geometrische oder eine attributive Vorgehensweise an. Die Analyse als wichtigste Komponenten eines GIS hat im Zusammenhang mit Datenunsicherheit eine Übertragung der Unsicherheitsmaße auf das Ergebnis zu lösen. Im Besonderen widmen sich die Ausführungen den geometrischen Analysemethoden wie Längen- und Flächenberechnung, Bestimmung der kürzesten Distanz und der geometrischen Verschneidung. Für alle beschriebenen Unsicherheitsmodelle werden die notwendigen Übertragungsmethoden abgeleitet und diskutiert.

Die durchgeführten Entwicklungen erfolgen unabhängig von bestehenden GIS-Produkten. Durch die Arbeit steht jedoch ein vollständiges Konzept bereit, das sich unmittelbar zur Umsetzung in verschiedenen Systemen eignet. In diesem Zusammenhang sind die momentan in der Entwicklung befindlichen Geo-Daten-Standards von Bedeutung. Eine Überprüfung ergibt, dass sich die getroffenen Festlegungen zu Modellen und Methoden problemlos in die Standardisierung einbetten lassen, so dass Integrationen auch von dieser Seite keine Probleme im Wege stehen werden.

Summary

Spatial data are always uncertain to some degree. Nevertheless, existing geographical information systems do not allow the consideration of uncertainty in any of the processing steps of an application. However, such a situation can only be tolerated if simple operations are used exclusively in a closed application environment. Only in this case the know-how of an experienced expert is sufficient in order to be able to test the quality reliably. On the other hand, modern applications have the opportunity to access external databases and utilise complex operators in an increased manner. A simple examination is no longer possible. Instead, a complete integration of uncertainty has to be developed and implemented into the work flow of GIS processing.

The major objective of this research is to develop data models and methods to enable the processing of uncertainty in a GIS. The investigations in particular focus on the processing of geometry as the most important component of spatial data. The phenomena of the real world are represented by objects. The discussion of causes which add for uncertainty forms the starting point for the development of models. Essential influences can already be found in the data collection step caused by abstraction of the real world, discretisation and measurement of data. All the influences have to be covered and described by a suitable uncertainty model. However, due to varied applications, no universal uncertainty model can be defined which meets the diverse requirements of practical work. Therefore, several alternatives are required. Basing on fundamental approaches, three different models are developed and integrated in the well-known object model: the Stochastic Model, the Minimum Maximum Model and the Fuzzy Model.

It is important for applications that data uncertainty is integrated into the different components of a GIS. Data collection receives the additional task to determine the values of the defined uncertainty measures. Different procedures are presented for this purpose, e.g. using multiple repeated collection of geometry or comparison with accurate reference data. Beside the ordinary visualisation of the uncertainty values as special displayable attributes it is interesting to see the spatial effect of uncertainty. Isolines and matrix representations can be used as possible techniques. The existing management concepts have to be extended in order to include the additional information. A geometrical or an attributive procedure provides a solution for this. The analysis step as the most important component in a GIS has to solve a transfer of uncertainty to the result. In particular, this step concentrates on geometrical analysis methods, i.e. length and area calculation, determination of the shortest distance and the geometrical overlay of objects. The necessary propagation methods are derived and discussed for all described uncertainty models.

The development is carried out without regard to special GIS products and therefore can be seen as independent. However, the research defines a complete concept which is suitable for the transfer to any system. In this connection the Geo data standards which are currently under development play an important role. User acceptance will profit from compatibility with these standards. A verification shows that the given definitions of models and methods can be embedded into standardisation without any problems, so that there should be no serious impediments to the practical implementation of the concepts.

Inhaltsverzeichnis

1	Einführung	7
1.1	MOTIVATION.....	7
1.2	STAND DER FORSCHUNG.....	9
1.2.1	<i>Datenmodellierung in GIS</i>	9
1.2.2	<i>Unsicherheit von räumlichen Daten</i>	10
1.2.3	<i>Diskussion</i>	13
1.3	ZIELE DER ARBEIT.....	13
1.4	AUFBAU DER ARBEIT.....	14
2	Grundlagen zur Modellierung und Integration der Unsicherheit	16
2.1	GEOMETRIE VON OBJEKTEN.....	16
2.2	UNSICHERHEITSBEGRIFFE.....	18
2.3	URSACHEN DER GEOMETRISCHEN UNSICHERHEIT.....	20
2.3.1	<i>Messung</i>	23
2.3.2	<i>Diskretisierung</i>	24
2.3.3	<i>Modellbildung</i>	27
2.3.4	<i>Diskussion</i>	31
2.4	UNSICHERHEITSMODELLE.....	31
2.4.1	<i>Stochastisches Modell</i>	33
2.4.2	<i>Minimum-Maximum-Modell</i>	41
2.4.3	<i>Fuzzy-Modell</i>	45
2.4.4	<i>Diskussion</i>	50
2.5	INTEGRATION DER UNSICHERHEIT IN DIE DATENMODELLIERUNG.....	53
3	Anwendung der Unsicherheit in den GIS-Komponenten	57
3.1	GIS-KOMPONENTEN.....	57
3.2	ERFASSUNG DER UNSICHERHEIT.....	58
3.2.1	<i>Bestimmung der mittleren Geometrie</i>	58
3.2.2	<i>Modellparameter und Variationsmaß</i>	63
3.2.3	<i>Diskussion</i>	70
3.3	PRÄSENTATION DER UNSICHERHEIT.....	71
3.3.1	<i>Unsicherheitsmaße als darstellbares Attribut</i>	71
3.3.2	<i>Unsicherheit in der räumlichen Ausdehnung</i>	73
3.3.3	<i>Diskussion</i>	78
3.4	VERWALTUNG DER UNSICHERHEIT.....	78
3.4.1	<i>Geometrische Verwaltung</i>	79
3.4.2	<i>Attributverwaltung</i>	79
3.4.3	<i>Diskussion</i>	80
3.5	INTEGRATION DER UNSICHERHEIT IN DIE ANALYSE.....	80
3.5.1	<i>Bestimmung der Linienlänge</i>	81
3.5.2	<i>Bestimmung des Flächeninhalts</i>	84
3.5.3	<i>Kürzeste Distanz zwischen zwei Objekten</i>	86
3.5.4	<i>Verschneidung von Objekten</i>	91
3.5.5	<i>Diskussion</i>	99
4	Diskussion der Ergebnisse	101
4.1	EIGENSCHAFTEN DER UNSICHERHEITSMODELLE.....	101
4.2	ANMERKUNGEN ZUR PRAKTISCHEN UMSETZUNG.....	102
5	Schlussbetrachtungen	105
5.1	ZUSAMMENFASSUNG.....	105
5.2	AUSBlick.....	106

6	Literatur	107
A	Ergänzungen zum Stand der Forschung.....	114
A.1	FUNDAMENTALE ANSÄTZE ZUR DATENMODELLIERUNG	114
A.1.1	<i>Objektmodell</i>	114
A.1.2	<i>Feldmodell</i>	115
A.2	ANSÄTZE ZUR MODELLIERUNG DER UNSICHERHEIT VON FELDERN	115
B	Grundlagen der Unsicherheitsmodelle.....	118
B.1	GRUNDLAGEN DES STOCHASTISCHEN MODELLS	118
B.2	ABLEITUNG DER WAHRSCHEINLICHKEIT FÜR DIE LINIENZUGEHÖRIGKEIT BEI STOCHASTISCHER BREITE DES PHÄNOMENS.....	123
B.3	GRUNDLAGEN DES MINIMUM-MAXIMUM-MODELLS	124
B.4	GRUNDLAGEN DES FUZZY-MODELLS	126
B.5	EINIGE WICHTIGE SPRACHKONSTRUKTE DER „UNIFIED MODELING LANGUAGE“ (UML)	130
C	Ergänzungen zu den Anwendungen der Unsicherheit.....	131
C.1	ÜBERGANGSMODELLE ZWISCHEN ZWEI LINIEN	131
C.2	GRUNDLAGEN ZUR GENERIERUNG VON SIMULATIONSDATENSÄTZEN AM BEISPIEL EINER LINIE	134
Dank	137
Lebenslauf	138

1 Einführung

1.1 Motivation

Geo-Informationssysteme (GIS) haben in den vergangenen Jahren eine rasante Verbreitung in unterschiedlichsten Anwendungsfeldern erfahren. Heute finden sich neben den traditionellen Einsatzgebieten wie Liegenschaftskataster, Topographie oder Versorgungsnetzwerke (Strom, Wasser, Gas usw.) ganz neue Bereiche wie z.B. Geo-Marketing, Telekommunikation, Verkehrsnavigation oder Umweltmanagement. Besonders das Entstehen von neuen Anwendungen wirkt innovativ auf weitere Entwicklungen und fördert damit eine fortschreitende Ausbreitung, so dass auch zukünftig mit steigenden Nutzerzahlen und neuen Anwendungen zu rechnen ist.

Generell eignen sich nahezu alle Aufgaben mit räumlichen Inhalten zur effizienten Bearbeitung in einem GIS. Dabei können bestehende analoge Arbeitsweisen automatisiert und beschleunigt werden. Aber nicht nur beim Schritt vom analogen zum digitalen Arbeiten, sondern insbesondere mit den Möglichkeiten zur Lösung von Problemstellungen, die seither als zu aufwendig und komplex erschienen, eröffnet sich die ganze Leistungsfähigkeit von GIS. Hierzu bedarf es in vielen Fällen eines interdisziplinären Ansatzes, bei dem Fachdaten und Analysemethoden aus unterschiedlichen Quellen zusammenzuführen sind (z.B. in Umweltinformationssystemen). Eine solche Anwendung steht und fällt mit der Verfügbarkeit an Daten. Während in den Anfängen der GIS-Technologie ein Nutzer die benötigten Daten stets selbst erfassen musste, hat sich die Situation in heutiger Zeit grundlegend verändert. Es entstehen immer mehr Datenbestände, auf die zugegriffen werden kann. Mit Hilfe dieser externen Daten kann ein Anwender die eigene Datenerfassung auf ein Minimum reduzieren. Es ist durchaus keine Vision, dass zukünftig zahlreiche Anwendungen vollkommen auf eigene Erfassungen verzichten und ihre Analysen gänzlich mit Fremddaten bestreiten werden. Bedenkt man, dass die Datenerfassung als der kostenintensivste Verarbeitungsschritt innerhalb eines GIS-Projekts gilt, so lassen sich durch den Bezug von Fremddaten erhebliche Einsparungen erzielen, wodurch die Einführung von GIS besonders für kleinere Aufgaben (z.B. im Geo-Marketing-Bereich) erst interessant und lohnend wird.

Unterstützung erfährt diese Entwicklung durch die modernen Möglichkeiten des Informationsaustausches über Computernetzwerke, die einen schnellen und direkten Zugriff auf weltweit verteilte Datenbestände erlauben. Auch hier ist der heutige Trend zur Globalisierung zu erkennen. Innerhalb von wenigen Augenblicken kann ein Anwender sich mit unterschiedlichsten Daten versorgen, die dann sofort bereitstehen, um in Analysen miteinander kombiniert und verarbeitet zu werden. Die notwendigen technischen Voraussetzungen dafür sind momentan im Entstehen begriffen. Dazu zählt insbesondere die Definition eines strukturierten Austausches der Daten, wie er in Normungsgremien (DIN/CEN/ISO, 1998) oder von privatwirtschaftlichen Organisationen (OGC, 2000) erarbeitet wird.

Diese optimistischen Aussichten auf eine uneingeschränkte Anwendbarkeit von GIS werden jedoch durch einige Risiken getrübt. Ein wichtiges Problem liegt in der fehlerhaften Nutzung von Daten. Die Möglichkeit, Datenbestände beliebiger Art in GIS zusammenzuführen, kann zu Situationen führen, in denen Daten für Zwecke eingesetzt werden, für die sie ungeeignet sind. Beispielsweise lassen sich durch Verknüpfung von kleinmaßstäbigen Landnutzungsdaten mit großmaßstäbigen Katasterinformationen Aussagen zur parzellenscharfen Flächennutzung ableiten, die in den Daten gar nicht enthalten sind. Solche Kombinationen von unterschiedlichsten Daten können falsche oder widersprüchliche Ergebnisse liefern, die in vielen Fällen zwangsläufig zu Fehlentscheidungen führen. Prinzipiell ist sogar jedes beliebige Analyseergebnis zu erzielen, wenn nur die entsprechenden Daten dafür herangezogen werden.

Probleme entstehen nicht nur durch gestiegene Datenverfügbarkeit und freien Datenaustausch. Auch in Anwendungen mit selbsterzeugten Daten sind Datenüberprüfungen unerlässlich. Zwar werden die Daten normalerweise so erfasst, wie sie für den vorbestimmten Zweck benötigt werden, doch lassen sich mit heutigen GIS äußerst komplexe Analysen durchführen, bei denen die Prüfung des Endproduktes auch für Experten nicht mehr ohne weiteres möglich ist. Insofern besteht bei der Verarbeitung kein größerer Unterschied, ob eigene oder fremdbezogene Daten eingesetzt werden.

Die Ursache für eine eingeschränkte Anwendbarkeit ist in der Qualität der Daten zu suchen. Der Begriff „Qualität“ findet im allgemeinen Sprachgebrauch recht häufig Verwendung. Er dient dabei der Bewertung von Waren oder Leistungen. Dies lässt sich auch aus der nachfolgenden Definition (Brockhaus, 1998) ableiten, die hier speziell auf „Daten als Ware“ bezogen ist. Danach versteht man unter Qualität:

„die Beschaffenheit von Daten nach ihren Unterscheidungsmerkmalen gegenüber anderen Daten, nach ihren Vorzügen oder Mängeln“.

In ähnlicher Weise wird der Begriff in einem Entwurf zu einem Qualitätsstandard für Geo-Daten definiert (DIN/CEN/ISO, 1998): Qualität ist

„die Gesamtheit der Eigenschaften eines Datensatzes, die zu seiner Fähigkeit beitragen, expliziten oder impliziten Anforderungen zu genügen“.

Jeder Datenbestand besitzt demnach ganz spezifische Qualitätseigenschaften, die seinen Anwendungsbereich festlegen. Zur Beschreibung der Qualität sind individuelle Angaben zu einer Reihe von Qualitätsmerkmalen notwendig. Mögliche Parameter können in folgende drei Gruppen untergliedert werden (Aronoff, 1991):

- Lokale Parameter: geometrische, thematische und zeitliche Unsicherheit,
- Globale Parameter: Herkunft, Vollständigkeit, Aktualität, logische Konsistenz, Maßstab und Auflösung,
- Nutzungsparameter: Format, Kosten und Zugriff.

Lokale Parameter verdeutlichen die Existenz von Schwankungen der Qualität bis in einzelne Teile der Daten hinein. Beispielsweise können geometrische Elemente mit ganz unterschiedlicher geometrischer Unsicherheit behaftet sein. Eine Straße besitzt z.B. in einem mittelmaßstäbigen topographischen Datenbestand eine Lagegenauigkeit von mehreren Metern, während für das korrespondierende Objekt im Katasterbereich Dezimeter-Genauigkeit erreicht wird. Die Werte der globalen Parameter beziehen sich dagegen auf den gesamten Datenbestand, ebenso wie die Nutzungsparameter, die sich den zum ordnungsgemäßen Einsatz der Daten notwendigen Aufwendungen widmen.

Um das diskutierte Risiko einer fehlerhaften Nutzung und daraus resultierender Fehlentscheidungen zu mindern, sind aus Anwendersicht zwei grundlegende Problembereiche angemessen zu berücksichtigen. Dazu zählen:

1. Dateneignung (Angebotsaspekt):

Der Anwender hat zu beurteilen, ob ein angebotener Datensatz für die vorgesehenen Zwecke einzusetzen ist. Stehen mehrere Datensätze mit entsprechendem Inhalt zur Verfügung, dann sind die Daten zu ermitteln, mit denen das zuverlässigste Ergebnis erreicht werden kann.

2. Ergebnisprüfung (Nutzungsaspekt):

Alle aus Analyseaufgaben erzielten Ergebnisdaten weisen wiederum bestimmte Qualitätseigenschaften auf, die einerseits von den Eigenschaften der Eingangsdaten und andererseits von den angewendeten Analyseschritten abhängen. Ziel muss es sein, zusätzlich zum Ergebnis immer auch eine qualitative Beschreibung zu erzeugen. Für diesen Zweck sind die Qualitätseigenschaften in allen Analyseschritten zu berücksichtigen und auf das Ergebnis zu übertragen.

Die Angebotsproblematik ist dadurch zu lösen, dass ein Datenanbieter jedem Datensatz eine Qualitätsbeschreibung in Form der erwähnten Parameter anfügt. Gleichfalls hat der Nutzer seine Anforderungen an die gewünschten Daten in ähnlicher Weise zu formulieren. Aus dem Vergleich von Beschreibung und Anforderung kann direkt auf die Eignung geschlossen werden. Für den Vergleich sind insbesondere die Nutzungs- und die globalen Parameter von Bedeutung. Sie können als Metadaten, d.h. Daten über Daten, verstanden werden, da der gesamte Datenbestand charakterisiert wird. Eine formalisierte Beschreibung erleichtert die programmgestützte Durchführung des Vergleichs, so dass eine automatische Suche nach optimalen Datensätzen möglich wird. Obwohl bezüglich der Dateneignung und der Auswahl mit Hilfe von Metadaten noch einiges an Entwicklungsarbeit notwendig ist (z.B. Entwicklung geeigneter Prüf- und Suchalgorithmen), scheint das Problem aufgrund bereits bestehender Ansätze (DIN/CEN/ISO, 1998) ausreichend gelöst zu sein. Daraus haben sich Forderungen an Datenanbieter und GIS-Systementwickler ergeben, entsprechend standardisierte Qualitätsbeschreibungen bzw. Programmrealisierungen bereitzustellen und zu implementieren.

Sehr viel schwieriger zu lösen ist dagegen die Nutzungsproblematik, die eine tiefgehende Integration der Qualität sowohl auf Datenseite als auch auf Systemseite notwendig macht. Die eigentlichen Daten sind um Werte der lokalen Qualitätsparameter zu ergänzen. Beispielsweise ist der Geometrie eines Objektes eine Beschreibung der geometrischen Unsicherheit beizufügen. Dies erfordert die Definition von Methoden zur detaillierten Bestimmung der Parameter. Beim Austausch der Daten müssen diese Werte zusätzlich übertragen werden, damit sie in der neuen Anwendung weiterhin zur Verfügung stehen. Für GI-Systeme bedeutet Integration in erster Linie eine Erweiterung der vorhandenen Datenmodelle, um die zusätzliche Information qualifiziert verwalten zu können. Zugriff und Verarbeitung setzen Datenstrukturen voraus, in die Qualitätsparameter eingebettet sind. Eine wichtige Aufgabe besteht in der Übertragung der Qualität auf Analyseergebnisse. Bisher eingesetzte Daumenregeln, wie die Bestimmung der Endqualität durch die „schlechtesten“ Eingabedaten, besitzen aufgrund von unterschiedlichen Gewichtungen von Eingabewerten keine zuverlässige

Aussagekraft. Auch Beurteilungen durch das Fachwissen des Anwenders kann nur bedingt Vertrauen geschenkt werden, da die Beeinflussung der Qualität über zahlreiche Operationen hinweg kaum abzuschätzen ist. Deshalb bedarf es einer Reihe methodischer Entwicklungen, die eine weitgehend automatisierte Übertragung erlauben, ohne dass der Nutzer größere Eingriffe vorzunehmen hat. Dazu sind sowohl bestehende Funktionalitäten zu erweitern, als auch neue Methoden zu erstellen. Insbesondere den notwendigen Systemveränderungen in Modellen, Strukturen und Methoden widmet sich der Hauptteil dieser Arbeit.

1.2 Stand der Forschung

In diesem Abschnitt erfolgt eine Sammlung und Beschreibung bisheriger Arbeiten zum Thema Datenunsicherheit, um einen Eindruck über den momentanen Stand der Forschung zu vermitteln. Zwei Schwerpunkte werden gebildet. Zum einen sind die Erkenntnisse über mögliche Sichtweisen auf die reale Welt und die sich daraus ableitbaren Datenmodelle von Interesse. Zum anderen soll ausführlich auf Arbeiten zur Modellierung der geometrischen Datenunsicherheit eingegangen werden.

1.2.1 Datenmodellierung in GIS

Unter räumlicher Datenmodellierung versteht man einen Prozess, der das Definieren und Organisieren von Daten über die reale Welt umfasst (Peuquet, 1984). Der Gesamtprozess lässt sich in mehrere Schritte unterteilen, die in ein Schichtenmodell (Bill, 1999a) eingeordnet werden können. Zunächst erfolgt eine anwendungsbezogene Auswahl der Inhalte der realen Welt (thematisches Modell). Die getroffenen Festlegungen werden anschließend in Form eines Datenmodells (konzeptionelles Modell) beschrieben und strukturiert, das dann auf ein Datenhaltungskonzept (logisches Modell) abzubilden ist. Am Ende steht die technische Implementation in einem bestimmten Dateiformat (physikalisches Modell). Das Datenmodell als wichtigste Komponente spiegelt die anwendungsspezifische Sichtweise auf die reale Welt wider und kann deshalb auch unabhängig von den Daten als Raummodell (Kemp, 1992) bezeichnet werden. Grundsätzlich sind die vorhandenen räumlichen Eigenschaften in diskrete und kontinuierliche Phänomene zu unterteilen. Diskrete Phänomene (Gebäude, Straßen, Verwaltungseinheiten, Gesteinskörper, etc.) werden als Objekte, kontinuierliche Phänomene (Temperatur, Schwere, Höhe über N.N., etc.) als Felder modelliert. Felder und Objekte repräsentieren fundamental verschiedene Abstraktionsformen der Realität (Goodchild, 1989). Eine allgemeine Beschreibung der beiden Sichtweisen ist in Anhang A.1 zusammengetragen.

In verfügbaren GIS-Produkten ist häufiger die Objektsichtweise vorzufinden, so dass eine Vielzahl von Anwendungen auf diese Weise realisiert werden. In Couclelis (1996) wird sogar davon gesprochen, dass die heutige GIS-Technik die eigentlich zu modellierenden räumlichen Phänomene „objektifiziert“ hat. Frank (1996) versucht zu ergründen, warum die Objektsicht eine solch dominante Rolle spielt. Wichtige Argumente sind dabei die Bedeutung von Eigentum und die damit verbundenen rechtlichen Verhältnisse. Nur für Objekte können Eigentumsansprüche formuliert und entsprechend Rechtsfragen geklärt werden. Ein anderes Argument zielt auf die häufig durchzuführenden Aufgaben des Messens, Zählens und Beschreibens, die mit Objekten leicht zu realisieren sind. Des Weiteren sind Felder konzeptionell leicht in Objekte umzuwandeln, da ein Feld zur digitalen Verarbeitung stets zu diskretisieren ist (z.B. in Form eines gleichmäßigen Rasters). Die diskrete Form kann direkt als Einzelobjekte (jede Rasterzelle entspricht einem Objekt) interpretiert werden (Goodchild et al., 1992a, Bonham-Carter, 1994).

Als wichtiges Merkmal von Objekten gilt die scharfe Abgrenzung voneinander. Scharfe Grenzen sind aber in der Realität nur in Ausnahmefällen anzutreffen (Couclelis, 1992). Es handelt sich dabei fast ausschließlich um von Menschen geschaffene (künstliche) Phänomene (z.B. Ingenieurbauten, Verwaltungseinheiten). Im Gegensatz dazu stehen natürliche Phänomene, die nicht eindeutig zu definierende Grenzen besitzen und somit nur als „schlechte“ Objekte (Couclelis, 1996) zu bezeichnen sind. Hierbei kommt die fraktale Eigenschaft von natürlichen Objekten zum Tragen (Burrough, 1986), so dass beliebig verfeinerte Beobachtungen durch Vergrößerung des Maßstabes trotzdem keine scharfe Abgrenzung liefern (Couclelis, 1992). Es ist stets eine Abstraktion und Generalisierung der Realität auszuführen, um eine scharfe Grenze zu approximieren. Dadurch wird die Objektgeometrie zu einem gewissen Grad unsicher (Burrough, 1996). Bei Objekten können neben der Unschärfe in den Grenzen die Attribute inhomogen sein oder eine kontinuierliche Variation aufweisen. Dies führt wiederum zu Unsicherheiten, diesmal aber in den Attributen der Objekte. Bei Feldern können innerhalb der kontinuierlich angenommenen Phänomene Diskontinuitäten (im Sinne von sprunghaften Änderungen) vorkommen, die sich nur unzureichend modellieren lassen und Unsicherheiten in den Werten der Feldvariablen bewirken. Zusammenfassend ist festzustellen, dass sich die beiden fundamentalen Sichtweisen (Objekte und Felder) nur mit Einschränkungen zur Beschreibung des Detailreichtums der realen Welt eignen.

Unabhängig von den Datenmodellen kann eine alternative Einteilung räumlicher Daten gefunden werden, die auf die unterschiedliche Bedeutung der Komponenten räumlicher Information bei der Erfassung eingeht (Sinton, 1978). Als Komponenten können Position, Thematik und Zeit unterschieden werden. Da nur eine der drei Komponenten gemessen werden kann, ist bei der Erfassung eine andere zu fixieren und die verbleibende zu kontrollieren. In den meisten Anwendungen wird die Zeit als feste Komponente angenommen, so dass sich zwei zu unterscheidende Fälle ergeben. Der erste Fall kontrolliert die Thematik und beobachtet die Position (zumeist für Objekte zutreffend), während im zweiten Fall die Position kontrolliert und die Thematik gemessen (z.B. bei Rasterzellen) wird. Daraus leiten sich direkt die beiden fundamentalen Datenarten in GIS ab: Vektordaten sind der ersten, Rasterdaten der zweiten Kategorie zuzuordnen. Eine Komponente zu kontrollieren, bedeutet dabei, die Variation in vernachlässigbaren Grenzen zu halten. Chrisman (1991) stellt heraus, dass der Unterschied zwischen Kontrollieren und Messen zum Aufbau eines Unsicherheitsmodells von besonderer Bedeutung ist. Die kontrollierte Komponente kann aber nicht generell als völlig frei von Unsicherheiten angenommen werden. Sie wirkt sich zu einem gewissen Grad auf die Unsicherheit der gemessenen Komponente aus. Im Falle von Vektordaten wird die Bestimmung der Position von der räumlichen Variation der Attribute beeinflusst. Danach kann die Unschärfe in der thematischen Beschreibung als Ausdruck der Unsicherheit in der Ausdehnung des Objektes aufgefasst werden (Molenaar, 1996a).

1.2.2 Unsicherheit von räumlichen Daten

Im Mittelpunkt der bisherigen Arbeiten steht die Suche nach einem geeigneten Unsicherheitsmodell. Grundsätzlich können die Arbeiten gemäß der angewendeten Sichtweise auf die reale Welt in Objekt- und Feldansätze unterteilt werden. Bei den Objektansätzen steht die Beschreibung der geometrischen Unsicherheit im Vordergrund, während die Feldansätze sich auf Aspekte der thematischen Unsicherheit konzentrieren. Die unterschiedliche Gewichtung korrespondiert zur beschriebenen möglichen Charakterisierung der Daten aufgrund der Bedeutung der Komponenten (Sinton, 1978, Chrisman, 1991), wonach Objekte hauptsächlich von geometrischer Unsicherheit, kontinuierliche Felder dagegen von thematischer Unsicherheit geprägt werden. Nachfolgend wird vertiefend auf die Ansätze zur Objektsicht eingegangen, während die bestehenden Ansätze zur Feldsicht im Anhang A.2 erläutert werden.

Epsilon-Band

Mit Hilfe des Epsilon-Bandes soll bei der Digitalisierung von Linien aus bestehenden Karten der unbekannten Ort der wahren Linie eingegrenzt werden. Dazu ist ein Puffer mit der Weite e um jede digitalisierte Linie zu legen (Abb. 1.1). Es entsteht ein Band mit der Breite $2e$ um die Linie. Der Parameter e wird auch als Epsilon-Toleranz bezeichnet. Ursprünglich nutzte Perkal (1966) den Ansatz zum Zwecke der kartographischen Generalisierung. Chrisman (1982) und Blakemore (1984) verwenden dann das Epsilon-Band zur Beschreibung der Unsicherheit der digitalisierten Linie. Es handelt sich um eine diskrete Modellierung, da keine Aussagen über mögliche Verteilungen innerhalb des Bandes getroffen werden.

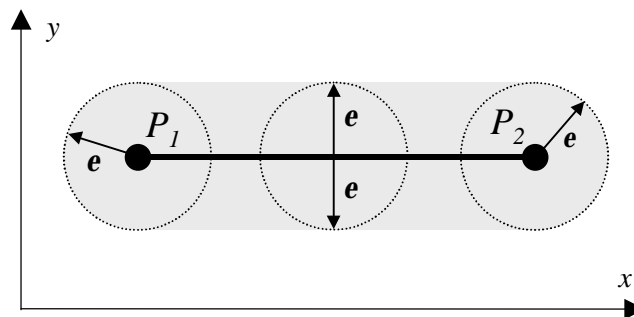


Abb. 1.1: Epsilon-Band um eine Linie von P_1 nach P_2 .

Blakemore (1984) nutzt das Modell im Zusammenhang mit Punkt-im-Polygon-Tests, bei denen über die Zugehörigkeit eines Punktes zu einem flächenhaften Objekt zu entscheiden ist. An die Stelle der eindeutigen Ja-Nein-Antwort tritt dabei eine diskrete Bewertung der Zugehörigkeit, bei der fünf verschiedene Klassen unterschieden werden können: sicher innerhalb, sicher außerhalb, möglicherweise innerhalb, möglicherweise außerhalb und mehrdeutig. Die Epsilon-Toleranz bildet bei Edwards und Lowell (1996) die Grundlage, um eine veränderte Darstellung der Daten zu erzeugen, die eine angepasste Modellbeschreibung zum Ziel hat. Mit Hilfe von morphologischen Operatoren (Serra, 1982) werden zunächst alle Objekte um die Epsilon-Toleranz geschrumpft (Erosion), so dass sie sich auf sichere Kernbereiche (Kernzonen) reduzieren. Anschließend erfolgt eine Aufblähung der Kerne durch Pufferbildung um den Faktor e in Umkehrung des ersten Schrittes (Dilatation). Es entstehen Bänder um die Kerne, die die Unsicherheit im Grenzverlauf ver-

deutlichen. Des Weiteren ergeben sich aber auch Löcher zwischen den Objekten, in denen die Zuweisung gänzlich zweifelhaft ist. Die so generierte Darstellung wird als „Fuzzy“-Karte bezeichnet. Einige weitere Anwendungen, wie z.B. bei der Bestimmung und Bewertung von topologischen Relationen (Clementini und Di Felice, 1996), bauen ebenfalls auf Epsilon-Bänder als Basisansatz auf.

Kovarianzansatz (stochastischer Ansatz)

Grundlage für den Kovarianzansatz bildet die Betrachtung der erfassten Punkte als stochastische Größen. Die Unsicherheit eines einzelnen Punktes lässt sich durch die Angabe einer Kovarianzmatrix D der Koordinaten (x,y) :

$$D \begin{pmatrix} x \\ y \end{pmatrix} = \begin{bmatrix} \mathbf{s}_x^2 & \mathbf{s}_{xy} \\ \mathbf{s}_{yx} & \mathbf{s}_y^2 \end{bmatrix} \quad (1-1)$$

veranschaulichen. Auf der Hauptdiagonalen stehen die Varianzen der Koordinaten, während die Nebendiagonalen vorhandene Korrelationen beschreiben. Vereinfachend wird dabei angenommen, dass nur zufällige Einflüsse wirken. Die Geodäsie stellt das Wissen und die Methoden für diese Modellbeschreibung bereit. Werden Auswertungen, wie z.B. Schnittberechnungen oder Bestimmung des Flächeninhalts mit Hilfe der Punktdaten ausgeführt, so ermöglicht das allgemeine Fehlerfortpflanzungsgesetz die Schätzung der Kovarianzmatrix des Ergebnisses (z.B. Wolf, 1975, Koch, 1980). Anwendung und Auswirkung des Gesetzes mit räumlichen Daten werden ausführlich in Drummond (1995) diskutiert. Aufgrund der sich auf Punkten aufbauenden Konstruktion von Linien bzw. Flächen aus Linien vergrößert sich die Dimension der Kovarianzmatrix für höherdimensionale Primitive in Abhängigkeit von der Punktzahl.

Der Kovarianzansatz wird z.B. bei der Flächenverschneidung eingesetzt, bei der die Unsicherheit auf Linienschnittpunkte fortzupflanzen ist (Kraus und Kager, 1994, Kraus und Ludwig, 1998, Bill und Korduan, 1998). Die bestehenden Arbeiten nutzen jeweils die Kovarianzmatrix der Linienpunkte, um über den funktionalen Zusammenhang von Schnitt- und Linienpunkten die Kovarianzmatrix der Schnittpunkte zu schätzen. Eine bedeutende Rolle spielen mögliche Korrelationen zwischen den Punkten, die vor allem bei adjazenten Punkten anzunehmen sind. Kraus und Kager (1998) beschreiben den Einfluss von Korrelationen anhand der Genauigkeit des Flächeninhalts. Ein Ansatz zur Bewertung von topologischen Relationen zwischen zwei flächenhaften Objekten wird in Winter (1996) entwickelt. Das Zutreffen einer Relation (z. B. Berühren oder Überschneiden von Objekten) ist abhängig von den Varianzen der Punkte der Objektgrenze. Das Verfahren berechnet zunächst den kleinsten und den größten Abstand zwischen den Objekten und bewertet anschließend die beiden Werte unter Berücksichtigung der Genauigkeit mit Hilfe von statistischen Testmethoden. Als Ergebnis wird für jede mögliche Relation die Wahrscheinlichkeit ihres Zutreffens erhalten.

Fehlerband

Die Werte der Kovarianzmatrix für Koordinaten eines Punktes definieren eine zweidimensionale Wahrscheinlichkeitsdichte, aus der sich zur visuellen Darstellung der Punktgenauigkeit eine Fehlerellipse ableiten lässt (Helmert, 1907). Form, Richtung und Größe der Ellipse geben wichtige Informationen zur Güte der Punktbestimmung. Für die Fläche der Ellipse gilt, dass sie mit einer Wahrscheinlichkeit von 39% die wahre Punktlage abdeckt. Eine ähnliche Darstellung aber mit höherer Wahrscheinlichkeit ist durch die Fußpunktskurve gegeben (z.B. Höpcke, 1980). Eine beliebige Wahrscheinlichkeit der Punktlage kann durch Verwendung von Konfidenzellipsen erzielt werden (Koch, 1980).

Für höherdimensionale Primitive wechselt die Darstellung von der Ellipsenform zu einem Band. Es wird als Fehlerband bezeichnet und soll die Streuung der Geometrie verdeutlichen. Die Bestimmung der Form des Fehlerbandes ist Inhalt verschiedener Arbeiten (Zhang und Tulip, 1990, Caspary und Scheuring, 1992, Shi, 1994). Ansatzpunkt ist stets die Ableitung der Kovarianzmatrix eines beliebigen Linienpunktes aus den Kovarianzmatrizen der Linienendpunkte. Die Ausdehnung des Bandes ergibt sich aus der senkrecht zur Linienrichtung wirkenden Randverteilung in jedem Linienpunkt unter Annahme einer bestimmten Wahrscheinlichkeit p (Abb. 1.2a).

Es kann festgestellt werden, dass die Standardabweichung der Randverteilung in der Mitte einer Linie am geringsten ist, so dass das Fehlerband die bekannte „Knochen“-Form annimmt (Abb. 1.2b). Eine analytische Definition der Randverteilung für beliebige Richtungen eines Linienstückes ist in Shi (1994) zu finden.

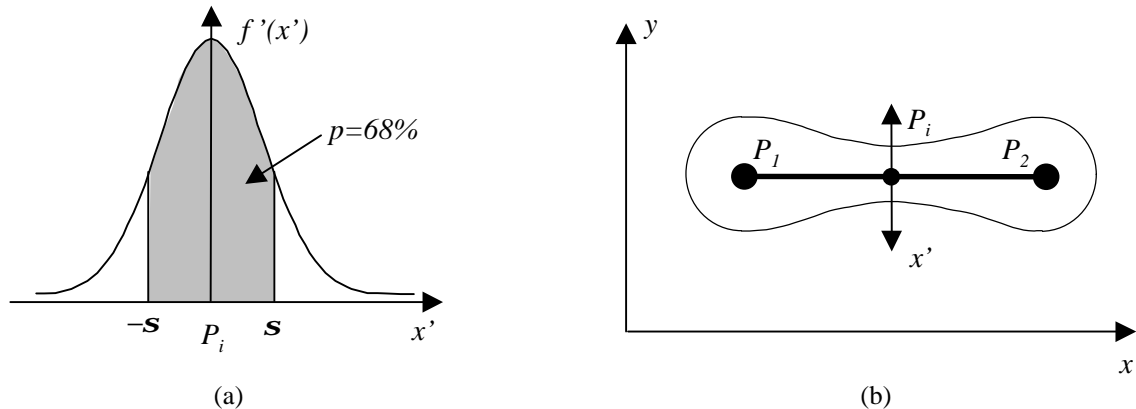


Abb. 1.2: Randverteilung senkrecht zur Linie in P_i (a), Fehlerband um die Linie von P_1 nach P_2 (b).

Die Komplexität des Problems steigt zunehmend, wenn Linien zu Linienzügen und weiter zu geschlossenen Randbeschreibungen für Flächen kombiniert werden. Eine allgemeine analytische Beschreibung wird zu komplex, so dass Approximationen unumgänglich sind. Ein vereinfachter Ansatz behandelt zunächst alle geraden Linienstücke getrennt und vereinigt dann deren Fehlerbänder mengentheoretisch (Shi, 1994). Einen guten Eindruck über die zu erwartende Form geben Simulationen (Casparly und Scheuring, 1992). Sie zeigen, dass die Größe des Brechungswinkel zwischen den angrenzenden Linienstücken ein wichtiger Formparameter darstellt.

Die Integration von Fehlerbändern in Analysemethoden kann für verschiedene Zwecke nützlich sein. Zhang und Tulip (1990) nutzen eine Approximation des Fehlerbandes, um identische Linien innerhalb von Datensätzen aufzufinden und einander zuzuordnen. Dieses Verfahren kann insbesondere zum Bereinigen von Splitterpolygonen nützlich sein, die häufig bei der Flächenverschneidung von unsicheren Daten entstehen.

Generell ist festzuhalten, dass Epsilon-Band und Fehlerband unterschiedliche Ziele verfolgen. Das Fehlerband visualisiert die Streuung in den Linienpunkten, während das Epsilon-Band die Lage der wahren Linie einzugrenzen versucht. Eine dem Epsilon-Band entsprechende Darstellung auf der Basis von stochastischen Punkten kann über statistische Testmethoden unter Vorgabe eines bestimmten Signifikanzniveaus gewonnen werden (Shi, 1994). Es definiert eine Art Vertrauensintervall um die Linie und ist aus diesem Grund zur besseren Unterscheidung als Konfidenzband zu bezeichnen.

Wahrscheinlichkeiten

Die stochastische Modellierung, wie sie im Falle der Kovarianzmatrizen oder der Fehlerbänder eingesetzt wird, bietet die Möglichkeit auf Wahrscheinlichkeiten überzugehen. In diesem Fall bewerten Wahrscheinlichkeiten die Zugehörigkeit eines beliebigen Punktes P zum Objekt. Entsprechende Ableitungen existieren für flächenhafte Objekte (Kraus und Haussteiner, 1993, Shi, 1994). Der zu bewertende Punkt wird auf die nächstgelegene Grenzlinie projiziert und für den gefundenen Lotfußpunkt die zugehörige Randverteilung geschätzt. Über den Abstand des Punktes zur Linie ergibt sich die gesuchte Wahrscheinlichkeit. Sie sagt aus, wie sicher ein Punkt innerhalb des Objektes liegt. Shi (1994) nutzt diesen Ansatz, um eine Zusammenführung von geometrischer und thematischer Unsicherheit zu erreichen. Dies gelingt allerdings nur für den Fall von diskreten Attributwerten (z.B. Landnutzungsklassen).

Anfelderungsmethode

Zu den vorgestellten Bandmodellen, die sich zur Beschreibung der Unsicherheit stark an den geometrischen Grundprimitiven (vor allem Punkte und Linien) orientieren, weist eine objektbezogene Betrachtungsweise einen alternativen Weg. Ein Objekt ist aufgrund der Geometrie in seiner Ausdehnung unsicher. Zumeist besitzt es einen sicheren Kern und einen unsicheren Randbereich. In der Anfelderungsmethode (Molenaar, 1996b) wird der unsichere Randbereich durch kleine Randflächen (fuzzy fields) repräsentiert, die alle eine unterschiedliche Zugehörigkeit zum Objekt aufweisen. Abhängig vom gewünschten Sicherheitsniveau werden einzelne Randflächen zur Objektausdehnung hinzugenommen oder fallen weg. Entsprechend variabel kann das Objekt dargestellt werden.

1.2.3 Diskussion

Die Integration der Unsicherheit in die Geoinformatik zählt seit einigen Jahren zu den ausgewiesenen Forschungsthemen mit vordringlicher Bedeutung im Bereich der räumlichen Datenverarbeitung (z.B. NCGIA, 1989, Goodchild, 1992, Caspary, 1992a). Trotz der Vielzahl an diskutierten Arbeiten hat bislang kein nennenswerter Einzug der Datenunsicherheit in die Praxis stattgefunden. Das liegt einerseits an den GIS-Nutzern selbst, denen bisher das notwendige Bewusstsein noch fehlt, dass generell alle Daten zu einem gewissen Grad unsicher sind. Gestützt wird diese Haltung immer noch von der exakten und scharfen Darstellung der Daten in den Systemen, die einen sicheren Datenbestand vortäuschen. Andererseits ist die Situation den bestehenden Ansätze anzulasten. Sie sind häufig in Konzeption und Anwendung zu speziell und damit kaum allgemein einzusetzen. So erlaubt es z.B. der Kovarianzansatz nur, die Unsicherheit von künstlichen Objekten angemessen zu repräsentieren, da er nicht in der Lage ist, die Einflüsse aus der Natur der Phänomene zu fassen. Um bestehende Ansätze zu nutzen, wird zudem umfangreiches Spezialwissen vorausgesetzt, das ein gewöhnlicher Anwender meist nicht vorweisen kann. Dies führt zwangsläufig zu erheblichen Akzeptanzproblemen, zumal die Behandlung der Unsicherheit zusätzlichen Aufwand verlangt. Grundlegend fehlt es an brauchbaren Definitionen, Konzepten und Modellen (Caspary, 1992b), die eine durchgängige Integration der Unsicherheit in alle vier Komponenten (Erfassung, Verwaltung, Analyse und Präsentation) eines GIS ermöglichen. Insbesondere ist die Berücksichtigung und Fortpflanzung der Unsicherheit in gängigen Methoden des Analysebereichs bisher nur an wenigen Fällen erprobt, so dass dem Anwender der Nutzen nur unzureichend aufgezeigt wird.

Die technischen Voraussetzungen dürften geschaffen sein, jedoch stellt sich grundsätzlich das Problem, die anfänglichen Entwicklungen aufzugreifen, zu vervollständigen und in GIS-Anwendungen einfließen zu lassen. Auf diesem Weg bedarf es der Klärung folgender offener Fragen, die bislang nicht ausreichend beantwortet sind (Goodchild, 1992):

- Wie soll eine genaue Repräsentation der räumlichen Phänomene aussehen?
- Wie soll die Unsicherheit digital repräsentiert werden?
- Wie kann Unsicherheit auf Ergebnisse fortgepflanzt werden?

Besonderen Wert ist auf die Definition eines Unsicherheitsmodells zu legen, das sich in die bestehende Datenmodellierung und Datenstrukturierung einzufügen hat. Bislang existiert keine zufriedenstellende Lösung der Unsicherheitsmodellierung für das Objektmodell. Die vorhandenen Ansätze stützen sich in ihrer Beschreibung auf die geometrischen Basisprimitive (Punkt und Linie) ohne einen Objektzusammenhang zu formulieren. Dies gilt insbesondere für die Bandansätze (Epsilon- und Fehlerband), die auf die Beschreibung der Unsicherheit von Linien ausgerichtet sind. Im Wesentlichen trifft dies auch für die Modellierung mit Wahrscheinlichkeiten zu, die auf Basis von Punktgenauigkeiten abgeleitet werden. Bei diesem Ansatz kommt hinzu, dass Wahrscheinlichkeiten nicht direkt zu erfassen sind, sondern sich nur über Verteilungen und deren Parameter berechnen lassen, so dass kein eigenständiges Modell vorliegt. Als Ausnahme ist die Anfängerermittlungsmethode zu sehen, die einen solchen objektweisen Zusammenhang beinhaltet, jedoch auf den Spezialfall von flächenhaften Objekten beschränkt bleibt. Aus diesen Gründen widmet sich diese Arbeit der Formulierung einer durchgängigen Konzeption zur Berücksichtigung der Objektunsicherheit.

1.3 Ziele der Arbeit

Räumliche Daten werden von heutigen am Markt verfügbaren GIS-Produkten als exakt angenommen und entsprechend modelliert, verarbeitet und präsentiert. Alle geometrischen Elemente werden koordinatenscharf dargestellt. Ein Zoom erweckt den Eindruck eines maßstabslosen GIS mit absolut genauen Daten. Zur Verwaltung setzen die Systeme Datenmodelle ein, die ausschließlich auf die Modellierung von exakten Daten ausgerichtet sind. Da Datenmodelle die Basis für sämtliche Methodenentwicklungen bilden, sind alle verfügbaren GIS-Methoden nur mit solchen Daten sinnvoll zu nutzen. Dagegen führt die Benutzung von unsicheren Daten mit diesen Methoden in vielen Fällen zu problematischen Ergebnissen, die eine Nachbehandlung erfordern (z.B. Beseitigung von Splitterpolygonen bei der Flächenverschneidung (Zhang und Tulip, 1990, Pullar, 1991)). Grundsätzlich müssen Ergebnisse aus Analysen vom Nutzer aber zwangsläufig als zweifelsfrei und korrekt angesehen werden, d.h. der Nutzer hat die Ergebnisse stets so zu interpretieren, als ob sie die tatsächlichen Verhältnisse in der Wirklichkeit direkt wiedergeben.

Es existiert jedoch eine Vielzahl an Ursachen, warum eine verlustfreie Abbildung der realen Welt in ein GIS-Modell in keinem Fall erreicht werden kann. Bereits ein einfacher Vergleich eines erfassten Objektes (z.B. Waldobjekt) mit seinem Erscheinungsbild in der Wirklichkeit offenbart, dass jeder Datensatz nur eine Approximation der realen Situation darstellt. Die Approximation berücksichtigt dabei anwendungsbezogene Aspekte, die je nach Kontext zu unterschiedlichen Repräsentationen des Phänomens führen. Zusätzlich kann die eigentliche Erfassung nur mit limitierter Genauigkeit

erfolgen, da eingesetzte Instrumente und Verfahren nicht absolut fehlerfrei arbeiten. Aus diesem Grund ist die Annahme von exakten Daten nicht zu rechtfertigen und deshalb zu verwerfen. Alle Daten sind demnach stets zu einem gewissen Grad mit Unsicherheit behaftet. Anstatt jedoch wie bislang die Unsicherheit zu vernachlässigen, ist als wichtigstes Ziel eine Integration der Unsicherheit herbeizuführen. Dadurch lässt sich insbesondere eine verbesserte Beschreibung der realen Welt erreichen. Beispielsweise können so auch unscharfe Objekte (insbesondere natürliche Objekte) angemessen modelliert werden. Durch Berücksichtigung der Unsicherheit innerhalb von Analysen eröffnet sich die Möglichkeit, Aussagen zur Zuverlässigkeit von Ergebnissen zu treffen. Dies führt zu einer verbesserten Interpretation und Beurteilung durch den Nutzer. Vergleiche zwischen widersprüchlichen Ergebnissen, die sich auf verschiedenen Wegen (mit unterschiedlichen Methoden und/oder mit unterschiedlichen Daten) erzielen lassen, werden möglich. Es reduziert sich das Risiko, auf der Basis von unsicheren Ergebnissen (die in heutigen Systemen als exakt gelten) falsche Entscheidungen zu treffen. Allgemein kann durch die Integration der Unsicherheit die Zuverlässigkeit des Werkzeuges „GIS“ gesteigert und so die Nutzerakzeptanz weiter verbessert werden.

Im Einzelnen werden im Rahmen der Arbeit die beiden folgenden Ziele angestrebt:

- Definition eines Datenmodells für geometrisch unsichere Objekte und
- Integration der Datenunsicherheit in die Komponenten eines GIS.

Ein Datenmodell für geometrisch unsichere Daten benötigt als Grundlage die Definition eines Unsicherheitsmodells, welches die Unsicherheit als Eigenschaft der eigentlichen Daten beschreibt. Dabei müssen alle Objekttypen, d.h. punkt-, linien- und flächenförmige Objekte in gleichem Maße beachtet werden. Das Unsicherheitsmodell ist als Erweiterung in das bestehende Objektmodell einzubetten. Diese Modellerweiterung soll in Form eines allgemeinen und offenen Konzeptes erfolgen und so Unabhängigkeit von Anwendungen und Systemen bieten. Offenheit ist insbesondere auch im Hinblick auf eine spätere Erweiterung um Aspekte der thematischen Unsicherheit zu fordern. Aufgrund der vielen verschiedenen Anwendungsfelder von GIS kann aber kein universelles Unsicherheitsmodell gefunden werden, das alle Anforderungen erfüllt. Vielmehr sind mehrere Modelle mit unterschiedlichen Eigenschaften zu formulieren und zu untersuchen.

Die geometrische Unsicherheit soll in allen Komponenten eines GIS (Erfassung, Verwaltung, Analyse und Präsentation) Beachtung finden. Dazu müssen die bestehenden Datenstrukturen und Methoden den hinzukommenden Inhalten angepasst werden. Für die Erfassung entsteht zusätzlich die Aufgabe, Werte der Unsicherheitsmaße zu bestimmen. Zur Verwaltung der Werte wird die Konzeption von Speicherstrukturen notwendig. Von besonderem Interesse ist die Fortpflanzung der Unsicherheit auf Analyseergebnisse bei der Verwendung bestehender Methoden (z.B. bei der Flächenberechnung). Entsprechende Funktionen, die diese Aufgabe übernehmen, sind zu definieren und als Ergänzungen den Methoden beizufügen. Dadurch soll am Ende nicht nur das Ergebnis, sondern auch die damit verbundene Unsicherheit bestimmt werden, um zusätzlich eine Prüfung der Zuverlässigkeit zu ermöglichen. In der Präsentationskomponente werden kombinierte Darstellungen von Daten und ihre Unsicherheit erforderlich. Dies kann z.B. in der Form geschehen, dass nur Bereiche angezeigt werden, die mit einer vorab gewählten Sicherheit die gesuchten Eigenschaften besitzen.

Um die Diskussionen und die Entwicklungen auf grundlegende und entscheidende Aspekte zu konzentrieren, sind an verschiedenen Stellen Einschränkungen im Umfang und Inhalt der Arbeit unumgänglich. Der Schwerpunkt der Arbeit liegt auf der Behandlung der geometrischen Unsicherheit. Thematische Aspekte werden nur in dem Maße einbezogen, wie sie sich räumlich auswirken. Hierunter lässt sich z.B. die Unschärfe in der Objektdefinition einordnen. Einerseits kann Unschärfe als Unsicherheit in der thematischen Zuweisung zu einer bestimmten Objektart angesehen werden. Interpretiert man andererseits Unschärfe als Problem der Festlegung der Objektausdehnung, dann wirkt sich ihr Einfluss räumlich aus und ist aus diesem Grund der geometrischen Unsicherheit zuzurechnen.

Eine weitere Einschränkung ergibt sich in Bezug auf die Datenmodellierung. Im Rahmen der Arbeit werden nur räumliche Phänomene betrachtet, die sich in Form von Objekten repräsentieren lassen. Als Datenart werden vornehmlich Vektordaten benutzt. Zwar kommen Rasterdaten im Speziellen bei der Darstellung und Verwaltung der Unsicherheit zum Einsatz, jedoch geht die Arbeit nur am Rande auf Aspekte der Unsicherheit im Zusammenhang mit rasterförmigen Objekten ein. Zusätzlich bleiben die Betrachtungen auf eine zweidimensionale Geometrie beschränkt.

1.4 Aufbau der Arbeit

In der Einführung wird zunächst auf die Motivation (Kapitel 1.1) eingegangen, warum die Entwicklung von Lösungsansätzen zur Integration der geometrischen Unsicherheit ein für die zukünftige Nutzung von GIS entscheidendes Kriterium darstellt. Daran schließen sich die Beschreibung und Diskussion des momentanen Standes der Forschung in diesem Themenumfeld an (Kapitel 1.2). Im Mittelpunkt stehen die existierenden Datenmodelle und deren Verbindungen und

die Einflüsse auf die Unsicherheit von Geo-Daten, sowie bereits existierenden Ansätze zur Modellierung und Verarbeitung der Datenunsicherheit. Den Abschluss des Kapitels bilden die Definition der Zielsetzung und die Beschreibung der daraus resultierenden Anforderungen an eine Lösung (Kapitel 1.3).

Kapitel 2 widmet sich den Grundlagen zur Modellierung und Integration der Unsicherheit. Es wird zunächst auf die verschiedenen Arten der geometrischen Objektbeschreibung eingegangen (Kapitel 2.1). Danach folgen einige Erläuterungen zu den gebräuchlichen Unsicherheitsbegriffen, um eine einheitliche Verwendung zu gewährleisten (Kapitel 2.2). Es schließen sich allgemeine Betrachtungen zu den Ursachen der Datenunsicherheit an (Kapitel 2.3). Die Erfassung als ein erster Verarbeitungsschritt beinhaltet einige wichtige Ursachen der Unsicherheit, die in diesem Kapitel näher untersucht werden. Die Integration der Unsicherheit benötigt ein Unsicherheitsmodell als Basisdefinition. Ausgehend von den bestehenden Ansätzen lassen sich drei unterschiedliche Modelle - Stochastisches Modell, Minimum-Maximum-Modell und Fuzzy-Modell - ableiten, die in Kapitel 2.4 vorgestellt werden. Kapitel 2.5 erläutert, wie die bestehenden Datenmodelle um die entwickelten Unsicherheitsmodelle zu erweitern sind.

Kapitel 3 geht vertiefend auf die Integration der Datenunsicherheit in die Anwendungsumgebung aus der Sichtweise der vier Komponenten eines GIS (Erfassung, Verwaltung, Analyse und Präsentation) ein. Zunächst werden die Aufgabengebiete der einzelnen Komponenten abgegrenzt und ihre Bedeutung im Hinblick auf die Datenunsicherheit beleuchtet (Kapitel 3.1). Die Erfassung als erste Komponente in der Projektbearbeitung erhält zusätzlich die Aufgabe, die Datenunsicherheit festzulegen. Einige Verfahren werden vorgestellt und diskutiert (Kapitel 3.2). Ein Blick auf die Möglichkeiten zur Visualisierung der Unsicherheit im Rahmen der Präsentationskomponente schließt sich an (Kapitel 3.3). Es werden mehrere Darstellungsarten entwickelt, die alle eine übersichtliche und verständliche Veranschaulichung der Unsicherheit als ergänzende Information zum Ziel haben. Innerhalb der Verwaltungskomponente ist die Frage zu behandeln, wie eine angemessene Speicherung der zusätzlichen Informationen erfolgen kann (Kapitel 3.4). Kapitel 3.5 widmet sich der Berücksichtigung der Unsicherheit in der Analyse. Dabei wird die Fortpflanzung der Unsicherheit über verschiedene geometrische Methoden (Linien- und Flächenberechnung, kürzeste Distanz und Verschneidung) untersucht.

Kapitel 4 enthält eine Diskussion der erzielten Ergebnisse. Insbesondere wird vertiefend auf die besonderen Eigenschaften der Unsicherheitsmodelle eingegangen, um auf deren Basis Empfehlungen zu treffen, für welchen Anwendungszweck das jeweilige Modell am besten geeignet ist (Kapitel 4.1). Es schließen sich einige Anmerkungen zur praktischen Umsetzung und den sich dabei ergebenden Problembereichen an (Kapitel 4.2). Einen wichtigen Aspekt bilden die aktuellen Standardisierungsbestrebungen im Zusammenhang mit Datenqualität. Die entwickelten Unsicherheitsmodelle werden auf Konformität mit den vorliegenden Entwürfen überprüft.

Die Zusammenfassung der wesentlichen Inhalte und der gefundenen Erkenntnisse bildet den Abschluss der Arbeit (Kapitel 5.1). Sie wird durch einen Ausblick auf darauf aufbauende zukünftige Aufgabenstellungen vervollständigt (Kapitel 5.2).

2 Grundlagen zur Modellierung und Integration der Unsicherheit

2.1 Geometrie von Objekten

Das Objektmodell bildet die Basis für die weiteren Betrachtungen im Rahmen dieser Arbeit. Es definiert ein Objekt als universelle Geo-Einheit, die sich aus den Komponenten Geometrie und Thematik aufbaut (Abb. 2.1). Von besonderem Interesse ist die Geometriekomponente. Die Geometrie hat die Aufgabe, Lage und Form des Objektes festzulegen. Die Lageinformation umfasst sowohl die Beschreibung der Position als auch der Orientierung innerhalb eines definierten Bezugssystems. Dazu werden in den meisten Fällen Koordinaten verwendet, die einen primären Raumbezug definieren. Beispiele sind geographische Koordinaten (Länge und Breite) oder kartesische Koordinaten, die einer bestimmten Abbildung des Bezugsellipsoids auf die Ebene zugehören. Alternative Möglichkeiten sind durch Kennzahlen, wie z.B. durch Postleitzahlen, Telefonvorwahlen usw. gegeben. Es handelt sich dabei um einen sekundären Raumbezug, der im Vergleich zu Koordinaten die Lage in schwächerer Form festlegt. Lageangaben helfen insbesondere beim Identifizieren und Auffinden von gesuchten Objekten. Die Formbeschreibung gibt dagegen konkret die räumliche Ausdehnung der mit dem Objekt verbundenen Thematik wieder. Im Allgemeinen können dabei beliebige Formen gebildet werden. Zusätzlich finden sich in der Geometriekomponente häufig Informationen zur Topologie wieder, die Inzidenz- und Adjazenzbeziehungen der Grundprimitive (z.B. Kanten-Knoten-Inzidenz) abbildet. Ihr Vorkommen erleichtert im weiteren Verlauf der Verarbeitungen die Analyse von Nachbarschaftsbeziehungen zwischen Objekten (z.B. Suche nach angrenzenden Objekten) in erheblichem Maße.

Als geometrische Grundprimitive werden Punkte, Linien und Flächen für Vektordaten und Rasterzellen für Rasterdaten in verschiedenen Variationen verwendet (Abb. 2.1). Beispielsweise können Linien als Geraden, Linienzüge, Kreisbögen, Polynome, Splines etc. auftreten. Die Primitive können ihrem Verlauf entsprechend in diskrete und kontinuierliche Formen unterschieden werden. Eine diskrete Form (Abb. 2.2b) liegt vor, wenn der Verlauf durch diskrete Einzelpunkte beschrieben wird, die durch Geraden miteinander verbunden sind. Die Einzelpunkte definieren Unstetigkeitsstellen im Verlauf. Dagegen hat man eine kontinuierliche Form (Abb. 2.2a) vorliegen, wenn die Verbindungen zwischen den Punkten durch beliebige Funktionen ausgefüllt sind, die an den Zwischenpunkten keine Unstetigkeiten aufweisen. In den meisten GIS-Produkten werden zwar kontinuierliche Verläufe, wie Kreisbögen, Splines und Polynome als Grundprimitive angeboten, doch werden diese nur in Ausnahmefällen innerhalb von Analysen auch als solche verwendet. Statt dessen werden kontinuierliche Verläufe zu Analyse Zwecken durch eine diskrete Form angenähert. Kontinuierliche Formen werden vor allem deshalb nicht verwendet, weil sie die Funktionsweise der Methoden in erheblichem Maße erschweren.

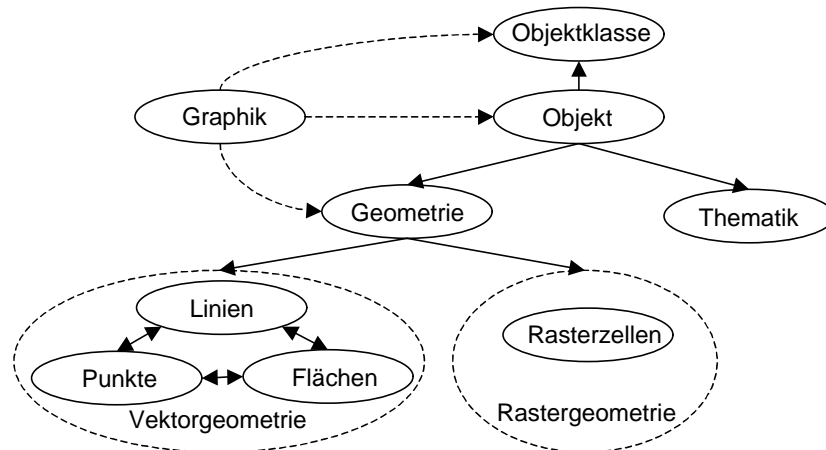


Abb. 2.1: Vektor- und Rastergeometrie und graphische Beschreibung im Objektmodell.

Im Rahmen der Arbeit wird ebenfalls das Vorliegen eines diskreten Verlaufs gefordert, obwohl die meisten in der realen Welt vorkommende Objekte eher einem kontinuierlichen Verlauf folgen. Die Festlegung der Forderung ist hier aber nicht ausschließlich durch die Komplexität in den Verarbeitungsmethoden begründet, sondern hängt ursächlich mit der Unsicherheit der Geometrie zusammen und ist damit fundamental mit der Themenstellung der Arbeit verknüpft. Wie in den weiteren Ausführungen erläutert wird, ist jede Geometrie zu einem bestimmten Grad mit Unsicherheit behaftet. Dadurch ist die geometrische Repräsentation eines Objektes, ob kontinuierlich oder diskret, als eine von vielen mögli-

chen Formbeschreibungen zu werten. Die kontinuierliche Form mag vielleicht exakter sein, ob sie jedoch genauer als eine korrespondierende diskrete Form ist, wird von einer Reihe von Ursachen, die zur gesamten Unsicherheit beitragen, bestimmt. Im Allgemeinen ist daher eine diskretisierte der kontinuierlichen Form gleichzusetzen, wenn auf eine angepasste Diskretisierung geachtet wird. Das ist dann der Fall, wenn die Formabweichung unbedeutend gegenüber der Unsicherheit in der Geometrie ausfällt (Abb. 2.2). Insofern ist die Forderung eines diskreten Verlaufs nicht mit einer Einschränkung der Anwendbarkeit der entwickelten Konzepte gleichzusetzen.

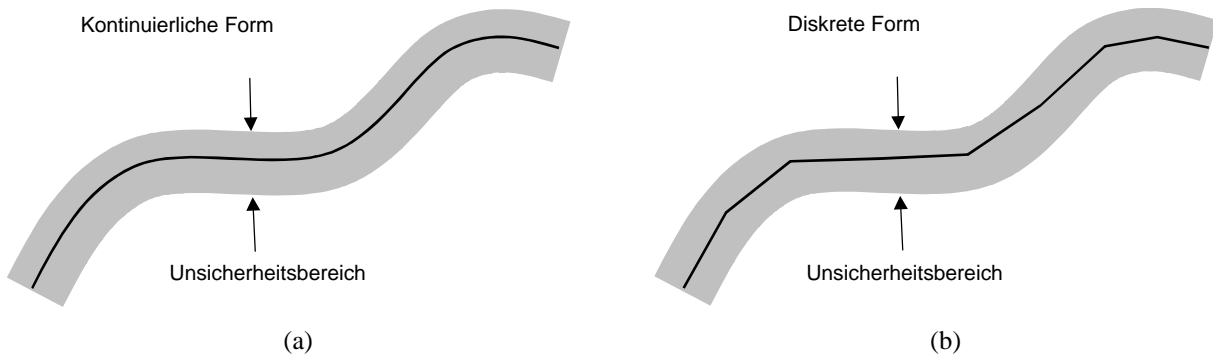


Abb. 2.2: Kontinuierliche (a) und diskrete (b) geometrische Formen im Zusammenhang mit dem zugehörigen Unsicherheitsbereich.

Letztendlich setzt sich jede Objektgeometrie aus beliebigen Kombinationen der Grundprimitive zusammen. Welcher Bauplan dabei gilt, ist durch ein geometrisches Modell festzulegen. Der CAD-Bereich kennt eine ganze Reihe solcher Modelle, die im Wesentlichen auch auf GIS-Anwendungen zu übertragen sind. Bill (1999a) gibt eine detaillierte Aufstellung verschiedener Modelle wieder. In bestehenden GIS-Produkten finden vor allem zwei Modelle Anwendung:

- die Randdarstellung (Boundary Representation) und
- das Enumerationsverfahren (Spatial Enumeration).

Bei der Randdarstellung (Abb. 2.3a) wird ein höherdimensionales Element stets durch seinen Rand (z.B. eine Fläche durch ihre Umringslinien) repräsentiert. Zur konkreten Beschreibung sind beliebige geometrische Primitive einzusetzen. Zum Einsatz kommen Punkte (für punkt- und linienförmige Objekte), Linien (für flächenförmige Objekte) und Flächen (für volumenförmige Objekte), so dass das Modell eine Vektorgeometrie bildet. In das Modell lassen sich in einfacher Weise topologische Beziehungen zwischen den Primitiven einbauen. Diese können getrennt von der metrischen Information strukturiert werden, wodurch topologische Konsistenzbedingungen unabhängig von der Metrik zu prüfen sind. Das Enumerationsverfahren (Abb. 2.3b) benutzt eine gleichmäßige Unterteilung des Raumes in Raumzellen von einheitlicher Größe und Form. Das Objekt ergibt sich dann aus einer Auflistung der zum Objekt gehörenden Zellen. Rasterzellen aus den bekannten Rasterdatenstrukturen korrespondieren genau zu diesem Modell, weshalb bei dessen Verwendung eine Rastergeometrie entsteht.

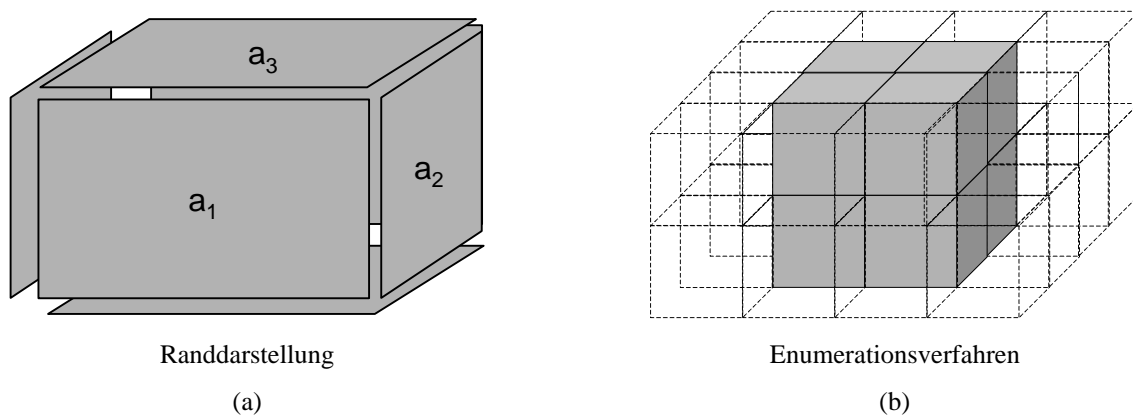


Abb. 2.3: Beispiel für die Randdarstellung (a) und das Enumerationsverfahren (b) für einen 3-dimensionalen Körper (nach Bill (1999a)).

Stellt das implementierte Objektmodell dem Nutzer beide Geometriearten in einer Systemumgebung zur Verfügung, so spricht man von einer Integration zu einem hybriden Datenmodell (z.B. Fritsch, 1988a). Damit lassen sich die einzelnen

Vorteile, die bei der Nutzung der beiden Arten entstehen, gezielt kombinieren. Es bestehen jedoch gravierende Unterschiede in der Art und Weise, wie eine Integration durchgeführt werden kann (Ehlers et al., 1989, Yang, 1992).

Zusätzlich werden Geometriedaten durch die graphische Beschreibung der Primitive ergänzt. Dazu können jedem Primitive graphische Attribute zugeordnet werden, die die visuelle Darstellung regeln. Zu den wichtigsten Attributen zählen Farbe, Linienstärke, Linienart, Flächenfüllung, Orientierung, Muster, Größenverhältnisse und Symbole (Buttenfield und Mackaness, 1991). Sie werden als Graphikdaten bezeichnet. Systemimplementationen siedeln die graphische Beschreibung auf unterschiedlichen Ebenen des Datenmodells an. Im allgemeinsten Ansatz werden die graphischen Attribute den Objektklassen zugewiesen, so dass alle Objekte einer Klasse das gleiche Erscheinungsbild aufweisen. Eine Erweiterung dieser Lösung erlaubt dennoch eine Variation zwischen den Objekten, indem die Werte eines thematischen Attributes die Ausprägung eines graphischen Attributes bestimmen (z.B. die Farbe von Straßenobjekten variiert mit dem thematischen Attribut Straßentyp). In der nächsten Ebene werden die graphischen Attribute jedem einzelnen Objekt zugewiesen und so eine individuellere Ausgestaltung der Objekte ermöglicht. Es entsteht der Nachteil, dass der Speicheraufwand steigt. Eine gänzlich freie Gestaltungsmöglichkeit bietet die letzte Ebene, die eine individuelle Attributierung einzelner geometrischer Primitive zulässt, wodurch aber ein noch größerer Aufwand entsteht. Eine solche Detaillierung dürfte aber nur in speziellen Anwendungen benötigt werden.

2.2 Unsicherheitsbegriffe

Im Zusammenhang mit der Unsicherheit von Daten finden eine Reihe von Begriffen Verwendung. „Ein Objekt ist unsicher“, „eine Grenze ist unscharf“, „ein Attribut ist ungenau“ oder „eine Koordinate ist fehlerhaft“. Die Aufzählung umfasst verschiedene Aussagen, die alle in qualitativer Weise ein Element charakterisieren. Die Liste der dazu benutzten Prädikate (hier: unsicher, unscharf, ungenau, fehlerhaft) ist keinesfalls als erschöpfend anzusehen. Man denke allein an die Gegenteile der erwähnten Begriffe (z.B. genau und ungenau), die ebenfalls mögliche Aussagen ergeben. Es ist leicht vorzustellen, dass bei der Menge an ähnlichen Begriffen deren Gebrauch häufig wechselweise in gleichen Zusammenhängen erfolgt. Z.B. stellt sich die Frage: Ist die Geometrie eines Objektes ungenau oder ist sie als unsicher zu bezeichnen? Jeder Benutzer verbindet mit einem bestimmten Begriff eine eigene Bedeutung, wodurch sich zwangsläufig Missverständnisse ergeben, wenn zwei Benutzer einen Begriff in unterschiedlicher Weise verwenden. Aus diesem Grund soll hier eine Definition der Begriffe erfolgen, die innerhalb der Arbeit eingesetzt werden. Die Definition hat die Aufgabe für einen einheitlichen Gebrauch zu sorgen und so das Verständnis und die Interpretation zu erleichtern. Die Zusammenhänge der Begriffe untereinander sind in Abb. 2.4 mit Hilfe der UML-Notation (OMG, 2000) veranschaulicht. Die Darstellung zeigt aber nur Beziehungen auf der Ebene der Begriffe. Sie gibt keine Auskunft darüber, wie die einzelnen Komponenten sich gegenseitig beeinflussen. Erläuterungen zur UML-Notation finden sich in Anhang B.5.

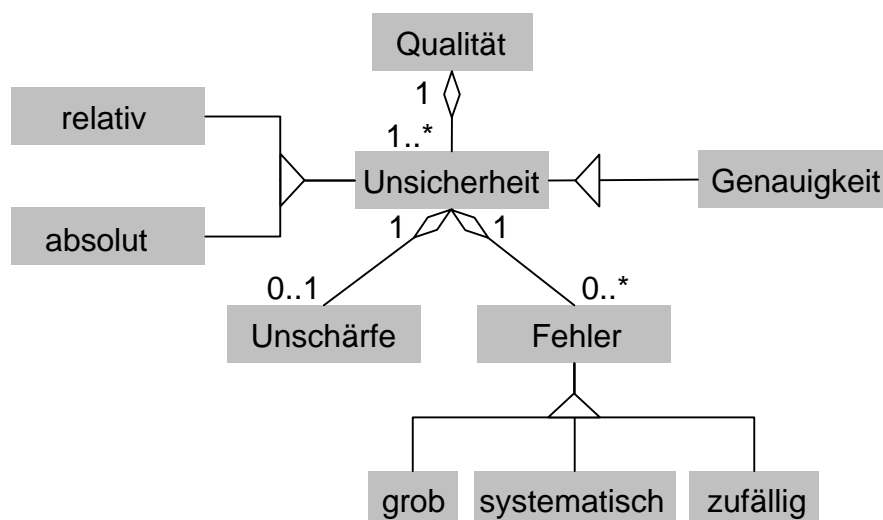


Abb. 2.4: Zusammenhänge der Unsicherheitsbegriffe.

Qualität

Die Qualität ergibt sich durch die Beschreibung von Eigenschaften einer Einheit (d.h. eines Elementes z.B. eines Datensatzes) im Hinblick auf die Fähigkeit, bestimmten Anforderungen zu genügen. Eine Definition und ausführliche Dis-

kussion wurde bereits in Kapitel 1.1 gegeben, so dass an dieser Stelle lediglich festgehalten werden soll, dass Qualität eine Vielzahl an Teilaspekten umfasst, von denen Datenunsicherheit mit seinen Komponenten geometrische und thematische Unsicherheit nur einen, wenn auch wichtigen Aspekt darstellt.

Unsicherheit

Unsicherheit definiert sich als Abweichung eines erfassten Elementes von der Wirklichkeit. Sie ist daher zu den lokalen Qualitätsmerkmalen zu zählen. Gleichzeitig dient Unsicherheit auch als Überbegriff und besitzt weitere Spezifizierungen (z.B. Genauigkeit). Als Elemente können wiederum beliebige Inhalte aufgefasst werden (z.B. Objekte, Attribute, Geometrie, Koordinaten, Aussagen, Beziehungen). Ein Element ist sicher, wenn es der Wirklichkeit exakt entspricht. Es kann als wahres Element bezeichnet werden. In der Regel weichen Elemente aufgrund verschiedenster Ursachen von der Wirklichkeit ab. Es wird unsicher. Als Unsicherheitsmaß ist der Wahrheitsgehalt des Elementes zu bewerten (z.B. durch Wahrscheinlichkeiten) oder es ist die Nähe zur Wirklichkeit und damit die Größe der Abweichung (z.B. durch Varianzen) festzustellen. Das Maß repräsentiert den Grad der Unsicherheit. Ist die Wirklichkeit in Form eines wahren Wertes bekannt, geben die Abweichungen Aufschluss über die absolute Unsicherheit. Sie werden z.B. durch Vergleiche mit übergeordnet genauen Referenzen ermittelt. Es kann aber vorkommen, dass kein wahrer Wert zu finden ist, da ein solcher in der Wirklichkeit nicht definiert werden kann oder als unbekannt gilt. In diesem Fall lässt sich nur eine relative Unsicherheit bestimmen, die sich auf die Abweichung zu einem mittleren Wert bezieht, der als Approximation an die Stelle des wahren Wertes tritt.

Beispiele:

- Positionen sind unsicher: Standpunkte, Grenzpunkte, Linienpunkte.
- Diskrete Attribute sind unsicher: Landnutzung ist Wald, Bodentyp ist sandig.
- Topologische Beziehungen sind unsicher: Objekt A liegt vollständig innerhalb des Objektes B.

Unschärfe

Als Unschärfe wird die Abweichung eines Modells von der Wirklichkeit bezeichnet. Dabei ist ein thematisches Modell gemeint, das sich aus der anwendungsspezifischen Abstraktion der Wirklichkeit zum Zwecke der Vereinfachung und Reduzierung auf wichtige Aspekte ergibt. Unschärfe wirkt als wichtiger Bestandteil auf die Unsicherheit von Elementen ein. Insbesondere ist die Unsicherheit von natürlichen Objekten davon betroffen, da diese oftmals nur unzureichend abstrahiert werden können und so größere Differenzen zwischen Modell und Wirklichkeit auftreten. Aber auch Aussagen können unscharf getroffen sein. Hier steht das vereinfachte linguistische Konzept der komplexeren Wirklichkeit gegenüber.

Beispiele:

- Natürliche Grenzen sind unscharf: Eine Grenze zwischen den Landnutzungsarten Wald und Wiese.
- Aussagen zur Nachbarschaft oder Güte sind unscharf: Zwei Objekte sind benachbart. Ein Bodentyp ist gut für Getreideanbau geeignet.

Fehler

Unter Fehlern sind Abweichungen erfasster Elemente vom Modell zu verstehen. Zusammen mit der Unschärfe bilden sie den zweiten wichtigen Bestandteil der Unsicherheit. Ausgangspunkt für Abweichungen ist das thematische Modell, nicht die Wirklichkeit selbst. Daher beziehen sich Fehler nur auf Unterschiede zum Modell. Drei Arten von Fehlern lassen sich unterscheiden: grobe Fehler, systematische Fehler und zufällige Fehler. Als grobe Fehler werden falsch erfasste Elemente bezeichnet, die aufgrund unzureichender Kontrolle auftreten. Systematische Fehler definieren Abweichungen, die stets in gleicher Weise wirken. So z.B. wenn ein Messgerät stets einen zu großen Messwert liefert. Zufällige Fehler entsprechen statistischen Abweichungen, die einem Zufallsprozess entstammen. Ohne besondere Aufwendungen werden alle drei Fehlerarten in gleicher Weise auftreten. Sowohl grobe als auch systematische Fehler können durch Kontrollen und spezielle Verfahren eliminiert werden, so dass im Idealfall nur zufällige Fehler auf die Unsicherheit einwirken.

Beispiele:

- Ein Messwert ist fehlerhaft:
 - ◆ Grob: Für ein Attribut Wasserstand wurde ein negativer Wert eingetragen.
 - ◆ Systematisch: Für das Attribut Landnutzung werden Streuobstwiesen stets zu Wäldern klassifiziert.
 - ◆ Zufällig: Die Position eines Punktes variiert um seinen wahren Wert.

Genauigkeit

Genauigkeit ist in ähnlicher Weise wie Unsicherheit zu benutzen. Entscheidender Unterschied ist die Einschränkung der Genauigkeit als Unsicherheitsbegriff für kontinuierliche Werte, während Unsicherheit allgemeiner zu benutzen ist. Damit ist Genauigkeit als Spezialisierung von Unsicherheit anzusehen und kann als Abweichung eines kontinuierlichen Wertes von seinem wahren Wert definiert werden. Die begriffliche Spezialisierung in absolute und relative Genauigkeit trifft ebenso zu, wie die Zuordnung der Begriffe Unschärfe und Fehler als Bestandteile der Genauigkeit.

Beispiele:

- Kontinuierliche Attribute sind ungenau: Wasserstände, Schadstoffbelastungen der Luft.
- Geometrische Primitive sind ungenau: ein Punkt, eine Linie, eine Fläche.
- Koordinaten sind ungenau: X-, Y-, Z-Koordinaten.

Für diskrete Werte ist demnach nur der Begriff Unsicherheit zulässig (ein diskretes Attribut ist nicht ungenau, sondern unsicher), während für kontinuierliche Werte beide Begriffe in äquivalenter Weise zu benutzen sind. Die Geometrie als das zentrale Element in der vorliegenden Arbeit weist in zweierlei Hinsicht kontinuierliche Eigenschaften auf. Als Erstes werden für Punktkoordinaten, die den geometrischen Primitiven zugrunde liegen, kontinuierliche Werte bestimmt. Als Zweites weisen die Primitive zumeist einen kontinuierlichen Verlauf auf. Dies gilt im Besonderen auch dann, wenn die Formbeschreibung diskret ist (z.B. im Falle einer Polygonlinie). Somit ist die Geometrie, also auch die geometrischen Primitive, durch den Begriff Genauigkeit in ihrer Unsicherheit zu charakterisieren.

2.3 Ursachen der geometrischen Unsicherheit

Die Ursachen für das Entstehen der geometrischen Unsicherheit sind vielschichtig. Generell können Gründe in allen vier Komponenten eines GIS gefunden werden. Eine dominante Rolle fällt aber den Einflüssen bei der Erfassung zu. Sie bilden eine Basisunsicherheit, die in anderen Komponenten durch weitere Einflüsse verstärkt werden kann. Aus diesem Grund konzentriert sich die Untersuchung auf die Ursachen im Erfassungsschritt. Ursachen innerhalb anderer Komponenten werden z.B. in Burrough (1986) angesprochen.

Ein möglicher Erklärungsansatz für die geometrische Unsicherheit (Molenaar, 1993) sieht die Gründe in der häufig durchzuführenden Zuweisung eines Elementes x zu einer Menge M : $x \in M$. Drei unterschiedliche Zusammenhänge können die Zuweisung unsicher gestalten. Als Erstes kann das Element x unsicher sein, da es nur mit beschränkter Genauigkeit zu erfassen ist. In der zweiten Möglichkeit gilt ähnliches für die Menge M . Sie ist dann unsicher, wenn M nur unscharf definiert werden kann. Beim dritten Zusammenhang sind sowohl M als auch x sicher, aber dennoch kann die Zuweisung nicht eindeutig erfolgen, da ein Nutzer sich nicht sicher für die Zuweisung entscheiden kann. Zuweisungen kommen in GIS in Form von Objektklassenzuweisungen (x = Objekt, M = Klasse), Attributwertzuweisungen (x = Attributwert, M = Attributdomäne) und Geometriezuweisungen (x = beliebiger Punkt, M = Objektausdehnung) vor (Molenaar, 1996b). Die Geometriezuweisung, ausschlaggebend für die geometrische Unsicherheit, definiert die Zugehörigkeit beliebiger Punkte aufgrund ihrer Lage zum Objekt. Alle nur unsicher zuzuweisenden Punkte veranschaulichen die Unsicherheit in der Geometrie. Gründe für eine unsichere Zuweisung leiten sich direkt aus den oben angeführten Zusammenhängen ab. Die Messung als eine der Ursachen (unsichere Elemente x) wird nur für ausgewählte Punkte durchgeführt (z.B. Polygonpunkte der Objektgrenze). Sie wirkt sich jedoch indirekt auch auf alle anderen Punkte aus, wenn der Punkt ein höherdimensionales geometrisches Primitiv (Linie oder Fläche) festlegt. Die Unschärfe in der Definition der Objektausdehnung (unsichere Menge M) hat dagegen auf beliebige Punkte in gleichem Maße Einfluss. Ist eine Objektausdehnung nur unscharf festzulegen, so muss die Geometrie entsprechend als unsicher bewertet werden. Genau so ergibt sich auf der Basis einer scharfen Definition eine sichere Geometrie. Dass sowohl Punktlage als auch Mengendefinition sicher sind, die Zuweisung aber trotzdem unsicher ist, trifft im Falle der Geometrie nur für Randpunkte zu. Für diese ist es zweifelhaft, ob sie zum Inneren oder Äußeren des Objektes zählen sollen. Dieser Sonderfall lässt sich jedoch durch eine festgeschriebene Regelung in einfacher Weise beseitigen.

Der diskutierte Ansatz begründet die Existenz der Unsicherheit und deren Ursachen zwar eingängig, bleibt aber zu allgemein. Eine detailliertere Beschreibung soll insbesondere die verschiedenen Quellen der geometrischen Unsicherheit aufdecken und gezielt den Eigenschaften der Ursachen nachgehen. Dazu sind zunächst im Einzelnen die bei der Datenerfassung auszuführenden Schritte zu untersuchen. Die Datengewinnung kann formal durch das in Abb. 2.5 gezeigte Schema repräsentiert werden (DIN/CEN/ISO, 1998, Davids et al., 1996). Ausgangspunkt bildet die reale Welt mit den zu erfassenden räumlichen Eigenschaften. Im ersten Schritt, der Modellbildung, ist daraus ein Anwendungsmodell zu erstellen, das die Eigenschaften in anwendungsspezifischer Sichtweise wiedergibt. Das Anwendungsmodell greift die für den bestimmten Zweck wichtigen Aspekte heraus und vernachlässigt dafür andere unwichtige Informationen und Details. Zum Beispiel wird ein Objekt Straße jeweils unterschiedliche Inhalte aufweisen, wenn es innerhalb einer Straßenplanung, im Liegenschaftskataster oder für topographische Aufgaben genutzt werden soll. Eine anwendungsunabhängige Abbildung scheint, abgesehen von der Machbarkeit, nicht sinnvoll, da für jegliche Nutzung immer erst relevante von irrelevanten Informationen zu trennen sind. Welche Inhalte für eine Anwendung von Bedeutung sind, wird in einer Modellierungsvorschrift festgehalten. Ein Beispiel für eine Modellierungsvorschrift ist im Objektartenkatalog des Amtlichen Topographisch-Kartographischen Informationssystems (ATKIS, 2000) gegeben. Die Vorschrift dient als Vorlage zur konkreten Umsetzung des ersten Schrittes. Da die Detailfülle der realen Welt in keinem Fall exakt übertragen werden kann, hat der Nutzer eine Generalisierung (thematischer Art) durchzuführen. Insofern entsteht immer ein mehr oder weniger abstraktes Abbild, ein Modell, der Realität. Das Anwendungsmodell kann deshalb auch als abstrakte Sicht der realen Welt bezeichnet werden. Durch die notwendige Abstraktion werden bereits in diesem ersten Schritt Differenzen (nämlich zwischen Modell und Wirklichkeit), die zur geometrischen Unsicherheit beitragen, erzeugt. Die Modellbildung entspricht aber nur einem gedanklichen Schritt, der kein physikalisches Ergebnis liefert. Das Anwendungsmodell existiert lediglich als virtuelles Gedankenbild, wodurch sich Betrachtungen zur Unsicherheit erschweren. Eine von anderen Schritten unabhängige Untersuchung der Einflüsse ist nicht möglich.

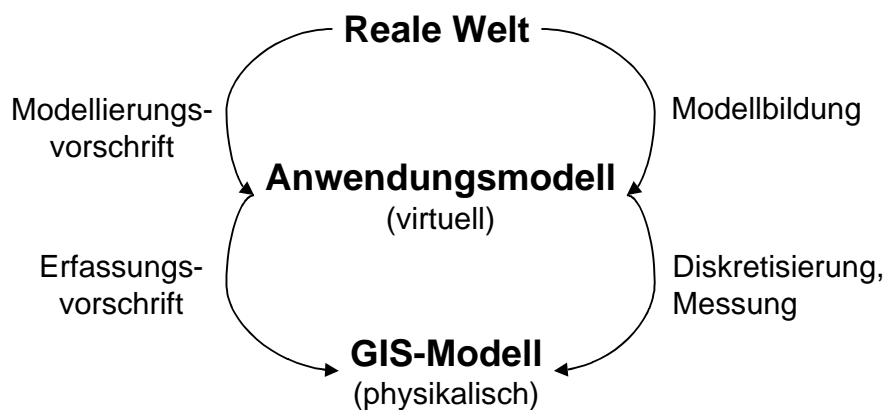


Abb. 2.5: Formales Schema zur Datengewinnung.

Der zweite Schritt baut auf dem gewonnenen Anwendungsmodell auf und hat die Aufgabe, die ausgewählten Informationen in computerlesbare Daten umzuwandeln. Es handelt sich dabei um den eigentlichen Erfassungsvorgang, bei dem durch Diskretisierung und Messung die Geometrie konkret festgelegt wird. Eine Diskretisierung ist nur dann notwendig, wenn die zu beschreibende Form einen kontinuierlichen Verlauf aufweist. Zur exakten Nachbildung einer kontinuierlichen Form wären unendlich viele Punkte zu erfassen. Die notwendige Reduktion auf endlich viele Punkte entspricht einem weiteren Generalisierungsschritt, der im Gegensatz zur Modellbildung geometrischer Art ist. Bei bereits diskreten Formen entfällt dieser Zwischenschritt. Letztendlich werden die Koordinaten von allen Punkten, die an der Geometriebeschreibung beteiligt sind, gemessen bzw. digitalisiert. Am Ende entsteht ein physikalischer Datensatz, der als GIS-Modell der Anwendung die Grundlage für alle weiteren Analysen bildet. Die gezielte Steuerung des Vorgangs übernimmt eine Erfassungsvorschrift, die dem Bearbeiter als Hilfe zur Interpretation von Einzelsituationen dient. Darin sind Festlegungen, wie z.B. die Mindestgröße eines Objektes oder wann ein Objekt linienhaft oder flächenhaft (z.B. bei Gewässern) zu erfassen ist, abgelegt.

Vollführt man die beschriebenen Schritte bei der Objekterfassung (Abb. 2.6), so wandelt sich das beobachtete Phänomen der realen Welt zunächst durch Abgrenzen von Nachbarobjekten in ein abstraktes, virtuelles Objekt, das eine kontinuierliche geometrische Form aufweist. Anschließend werden bei der Diskretisierung markante Formpunkte ausgewählt. Als Ergebnis erhält man ein diskretes Objekt, das aus den Punkten und ihren zugehörigen Verbindungen besteht. Den Abschluss bildet das Messen der Punktpositionen. Es entsteht das bekannte Objekt im Rechner.

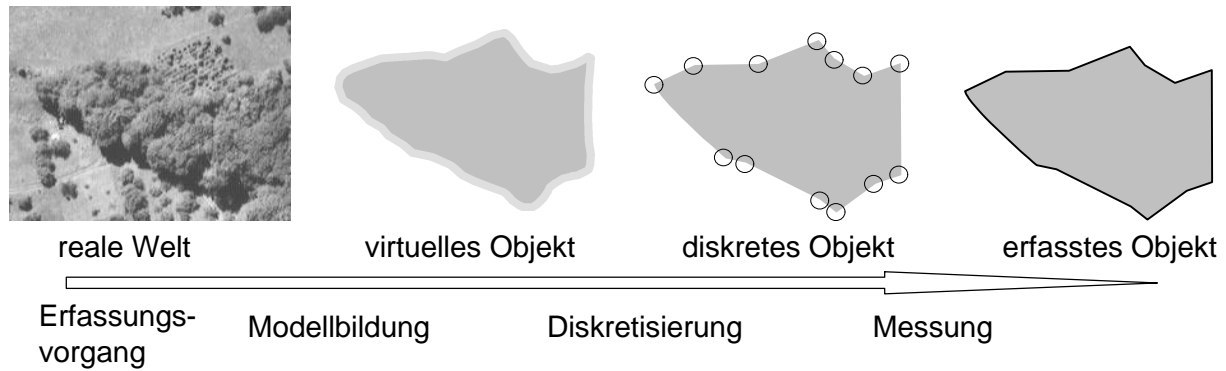


Abb. 2.6: Der Weg von der realen Welt zum digitalen Objekt: die Einzelschritte des Erfassungsvorgangs.

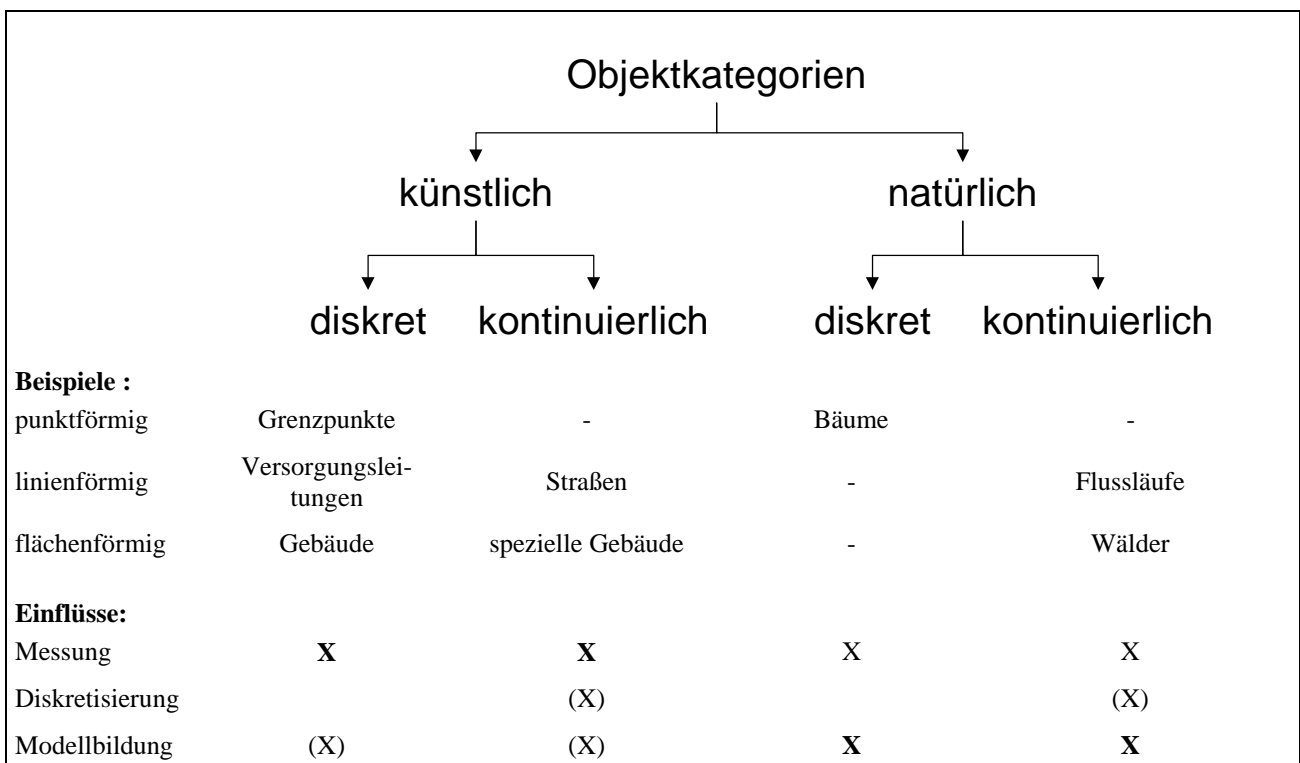


Abb. 2.7: Kategorisierung von GIS-Objekten nach Einflüssen auf die Datenunsicherheit mit Bewertung der Einflüsse auf die Kategorien (X = vorhanden, **X** = dominant, (X) = einzuschränken).

In allen beschriebenen Teilschritten sind Ursachen zu finden, die zur Entstehung der geometrischen Unsicherheit beitragen. Sie spielen jedoch abhängig von der Kategorie des zu erfassenden Objektes (Abb. 2.7) eine unterschiedliche Rolle. Zunächst ergeben sich Unterschiede zwischen natürlichen und künstlichen Objekten. Von natürlichen Objekten spricht man, wenn Existenz und Aussehen maßgeblich von der Natur beeinflusst werden. Künstliche Objekte sind dagegen durch Menschenhand geschaffen. Die Objektabgrenzung gestaltet sich in der Regel nur für natürliche Objekte schwierig, so dass hier mit Einflüssen aus der Definition des Anwendungsmodells zu rechnen ist. Künstliche Objekte sind meist aufgrund von Planungsprozessen entstanden und weisen daher eine sichere Abgrenzung zur Nachbarschaft auf. Wenn nach dem Abgrenzen die Formbeschreibung eines Objektes bereits diskret ist, dann entfällt die aufwendige Suche nach markanten Punkten. Die Form kann sofort gemessen werden, wodurch lediglich Einflüsse aus der Messung wirken. Generell ist die Messung von markanten Punkten bei allen Objekten notwendig, so dass zumindest diese Ursache die Unsicherheit beeinflusst.

Im Folgenden wird konkret auf die einzelnen Ursachen eingegangen und jeweils eine Abschätzung zur Größenordnung der Einflüsse gegeben. Als Grundlage dienen Auswertungen eines Erfassungstests, der speziell zu diesem Zweck durchgeführt wurde. Die Aufgabe bestand darin, verschiedene Testobjekte zu erfassen. Um statistische Aussagen treffen zu können, wurde die Erfassung von zehn unterschiedlichen Operateuren ausgeführt, wobei in allen Fällen identi-

sche Arbeitsumgebungen vorlagen. Als Gerät kam ein Digitalisierertisch zum Einsatz, an dem eine manuelle Erfassung der Testobjekte erfolgte. Um die unterschiedlichen Einflüsse untersuchen zu können, wurden die Testobjekte gemäß den diskutierten Objektkategorien (Abb. 2.7) ausgewählt, so dass der jeweilige Einfluss dominant in Erscheinung trat. Zur Beobachtung der Messeinflüsse wurden Gebäudegrundrisse, zur Beobachtung der Diskretisierungseinflüsse synthetisch erzeugte Kreise und zur Beobachtung der Modellbildungseinflüsse Waldgebiete ausgesucht. Einschränkend ist zu bemerken, dass die angegebenen Werte lediglich die Größenverhältnisse verdeutlichen sollen, jedoch nicht als absolute Maße allgemein zu interpretieren sind. Viel wichtiger sind die aus den Testergebnissen abgeleiteten Eigenschaften der Einflüsse. Als eine Art Kategorisierung wird dabei eine Einteilung in zufällige, systematische und grobe Einflüsse vorgenommen, die analog zu den Fehlerarten in der Geodäsie (z.B. Höpcke, 1980) zu verstehen ist.

2.3.1 Messung

Die Einflüsse aus der Messung werden durch die eingesetzten Geräte und den erfassenden Operateur bestimmt. Die Geräte besitzen stets eine begrenzte Auflösung, die den minimal noch zu messenden Koordinatenunterschied festlegt. Ein Digitalisierertablett z.B. weist durchschnittlich eine Auflösung von 0.025 mm ($1000 \text{ Linien} / \text{inch}$) auf. Die Messgenauigkeit von Koordinaten reicht aber bei weitem nicht an die Auflösung heran. Verantwortlich dafür sind die internen Komponenten des Gerätes, die in ihrer Gesamtheit das Ergebnis beeinflussen. So können Schwankungen in den ausgegebenen Koordinatenwerten auftreten, obwohl die Messposition unverändert bleibt. Für die genannte Auflösung ist nach Herstellerangaben des Digitalisierertablets eine Messgenauigkeit von $s = 0.15 \text{ mm}$ zu erwarten. Eine weitere Rolle spielt der Operateur selbst. Die Fähigkeit mehrmals hintereinander den gleichen Punkt in identischer Weise einzustellen, ist ebenfalls begrenzt. Hier stößt man bei manueller Messung an die Grenzen der Augenauflösung und der menschlichen Motorik. Auch können sich Faktoren wie Zeitdruck oder Arbeitsdauer bemerkbar machen.

Zur Beobachtung der Einflüsse aus der Messung eignen sich diskrete, künstliche Objekte, da auf diese sich keine weiteren Einflüsse auswirken. Als Testobjekte wurden daher eine Reihe von Gebäudegrundrissen (Abb. 2.8a) ausgewählt, die aus einer analogen Karte zu digitalisieren waren. Konkret zu messen waren jeweils die Eckpunkte der einzelnen Gebäude, aus denen schließlich ein geschlossenes Polygon erzeugt wurde. Die eingesetzte Digitalisiereinheit besaß die oben beschriebenen Kennwerte. Eine Überlagerung der zehn Erfassungen ist exemplarisch für ein Gebäude in Abb. 2.8b zu sehen.

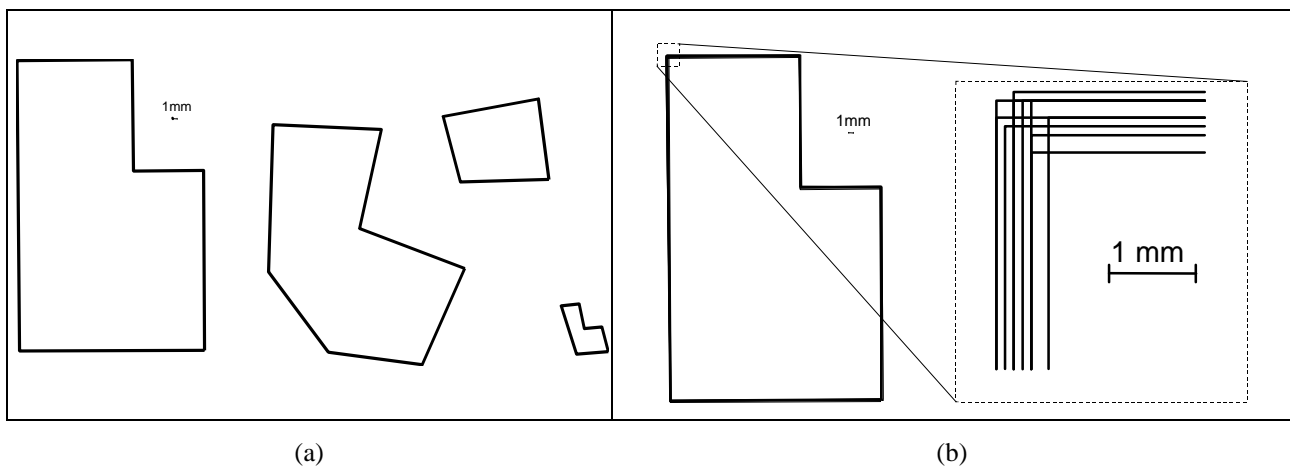


Abb. 2.8: Testaufbau: Gebäudegrundrisse (a) und Beispiel der unterschiedlichen Erfassungen (b) mit vergrößertem Ausschnitt eines Eckpunktes.

Die Auswertung der Messdaten erfolgt durch Mittelbildung der Punktkoordinaten und Berechnung der empirischen Standardabweichungen (s_x , s_y). Einen Gesamtwert für jeden Punkt kann durch den Punktfehler nach Helmert s_p ausgedrückt werden. Die für den Testaufbau erzielten Werte sind in Tab. 2.1 zusammengefasst. Sie beziehen sich auf das Tischkoordinatensystem und sind daher unabhängig vom Maßstab der Vorlage. Für die beiden Koordinatenrichtungen x und y ergeben sich nur geringfügige Unterschiede in den Standardabweichungen $s_x = 0.17 \text{ mm}$ bzw. $s_y = 0.15 \text{ mm}$. Der Punktfehler berechnet sich zu $s_p = 0.22 \text{ mm}$. Der durchschnittliche Wert mit $\bar{s} = 0.16 \text{ mm}$ liegt im Bereich der vom Hersteller angegebenen Messgenauigkeit. Übertragen auf eine Katasterkarte im Maßstab 1:2.500 ist mit einer Koordinatengenauigkeit von 40 cm in der Natur zu rechnen. Aus den Testergebnissen lässt sich einerseits ableiten, dass der Einfluss des Operateurs auf die Messgenauigkeit hier eine untergeordnete Rolle spielt. Die Grenzbereiche der menschlichen Fähigkeiten werden im Test nicht erreicht. Schwankungen aufgrund der Gerätekom-

ponenten bilden den Hauptanteil. Andererseits bestätigt das Ergebnis die gute Arbeit der zehn Operateure. Es wurde getestet, ob die Verteilung der Koordinaten der erwarteten Normalverteilung folgt. Aufgrund des geringen Stichprobenumfangs kam ein Kolmogoroff-Smirnow-Test (z.B. Kreyszig, 1965) zum Einsatz. Hierbei wird das Maximum a der Abweichungen zwischen der empirischen Verteilungsfunktion $\tilde{F}(x)$, die sich auf Basis der Stichprobe berechnen lässt, und der angenommenen Verteilungsfunktion $F(x)$ (hier: Normalverteilung) als Testgröße verwendet. Das Testergebnis (Tab. 2.1) zeigt, dass die Normalverteilung stets mit einem Signifikanzniveau von $\alpha = 10\%$ angenommen werden kann. Damit sind die Koordinaten als normalverteilte Zufallsvariablen ($x \sim N(\bar{x}, \mathbf{s}_x^2)$, $y \sim N(\bar{y}, \mathbf{s}_y^2)$) anzusehen, die dem Zufallsereignis „Messung“ entstammen.

Tab. 2.1: Standardabweichungen der Punktkoordinaten und Testergebnisse des Kolmogoroff-Smirnov-Test zur Anpassung an die Normalverteilung.

Objekt	Standardabweichungen		Test für $\alpha = 10\%$	
	\mathbf{s}_x [mm]	\mathbf{s}_y [mm]	Maximum a	Grenzwert c
Gebäude 1	0.17	0.13	0.08	0.11
Gebäude 2	0.17	0.14	0.09	0.10
Gebäude 3	0.17	0.14	0.12	0.14
Gebäude 4	0.16	0.15	0.09	0.11

In Bolstad et al. (1990) wurde die Messgenauigkeit deutlich besser zu $\bar{s} = 0.05 \text{ mm}$ ermittelt. Begründen lässt sich das Ergebnis mit der idealen Erfassungsumgebung (z.B. Kontrolle der Lufttemperatur, Messung nur idealer Punkte), die speziell für diesen Zweck hergestellt wurde. Dieses Ergebnis sollte aber eher als untere Grenze angesehen werden, da kontrollierte Verhältnisse in normalen Arbeitsabläufen nur unzureichend einzuhalten sind. Maffini et al. (1989) diskutieren andere Faktoren, die wichtige Einflussgrößen darstellen können. Dazu zählt der Zeitdruck, dem der Operateur bei der Bearbeitung ausgesetzt wird, aber auch der Maßstab der Vorlage. Die Untersuchungen zeigen, dass mit knapper werdender Zeit die Unsicherheit steigt. Demnach wird das Anzielen der Punkte bei Zeitmangel mit geringerer Exaktheit ausgeführt. Bei einem Maßstab der Vorlage von 1:50.000 und ausreichender Digitalisierzeit liegen 90% der Abweichungen vom Mittel unter $\pm 20 \text{ m}$. Dies entspricht einem maßstabsfreien Punktfehler von $\mathbf{s}_p = 0.24 \text{ mm}$, wodurch ein vergleichbares Ergebnis zu dem in dieser Arbeit durchgeführten Test erzielt wurde. Die absolute Genauigkeit verschlechtert sich aber deutlich mit steigender Maßstabszahl. So liegen bei einem Maßstab von 1:250.000 nur 33% der Abweichungen vom Mittel unter $\pm 20 \text{ m}$. Der daraus resultierende maßstabsfreie mittlere Punktfehler verändert sich mit $\mathbf{s}_p = 0.20 \text{ mm}$ aber nur unwesentlich.

Geometriedaten werden sehr häufig aus bestehenden analogen Karten erfasst. Da es sich hierbei um eine sekundäre Erfassungsmethode handelt, ist die Erfassungsgenauigkeit für die Vorlage (Einflüsse der primären Erfassung) den Einflüssen aus der Messung hinzuzurechnen. Hinzu kommt die Zeichengenauigkeit, für die im Allgemeinen ein Wert von $\mathbf{s}_z = 0.1 \text{ mm}$ anzunehmen ist. Aufgrund möglicher Veränderungen der Vorlage gegenüber der Wirklichkeit können weitere Größen wirken. Dazu gehören insbesondere Generalisierungseffekte und der Papierverzug. Eine sichere Abschätzung der Gesamtgenauigkeit lässt sich daher nur über einen Vergleich mit einer Referenz erreichen. Man erhält eine absolute Genauigkeit im Gegensatz zu der im Test ermittelten Wiederholgenauigkeit, die als relative Genauigkeit aufzufassen ist.

Zusammenfassend ist festzuhalten, dass die Einflüsse aus der Messung auf alle gemessenen Punkte in gleicher Weise einwirken. Im Allgemeinen handelt es sich dabei um zufällige Einflüsse, die in ihrer Gesamtheit durch verbesserte Messverfahren zwar verringert werden können, jedoch niemals gänzlich zu beseitigen sind. Da die Messung einen stets notwendigen Schritt innerhalb der Erfassung darstellt, sind alle Geometriedaten zumindest mit diesem Anteil an geometrischer Unsicherheit behaftet.

2.3.2 Diskretisierung

Eine Diskretisierung der Geometrie ist nur dann notwendig, wenn eine kontinuierliche Form beschrieben werden soll. Dazu sind markante Punkte, die die Form repräsentieren, auszuwählen. Dementsprechend spielt die Wahl der Punkte eine wichtige Rolle. Bei gleichmäßiger Krümmung (z.B. für einen Kreis, Abb. 2.9) sind die auszuwählenden Punkte ebenfalls gleichmäßig zu verteilen. Eine ungleichmäßige Verteilung der Punkte bewirkt eine verschlechterte Approximation der kontinuierlichen Form und ergibt eine größere geometrische Unsicherheit. Variiert die Krümmung innerhalb

des Objektes, dann ist die Auswahl entsprechend anzupassen. Welche Krümmung vorliegt, hängt von der individuellen Objektform ab, die damit einen zweiten Einflussfaktor bildet. Die Größenordnung der Abweichung zwischen diskreter und kontinuierlicher Beschreibung wird abschließend durch die Anzahl der gewählten Punkte bestimmt. Mit einer unendlich großen Anzahl lässt sich die Form theoretisch exakt beschreiben. In der Praxis sind nur begrenzte Mengen an Punkten möglich, so dass bei notwendiger Diskretisierung immer Unsicherheit in der Geometriebeschreibung entsteht.

Beispielsweise ist ein Kreis durch eine gleichmäßige Verteilung der Punkte auf seinem Rand zu diskretisieren. Betrachtet man einen Randausschnitt näher, so kann festgestellt werden, dass die Abweichung zwischen Approximation und wahrer Form immer nur in eine Richtung (in Richtung des Kreismittelpunktes) wirkt (Abb. 2.9). Aus diesem Grund muss die Diskretisierung als systematischer Einfluss eingestuft werden. Die Normalverteilung ist zur Beschreibung auszuschließen. Die mögliche Form der Wahrscheinlichkeitsdichte wird in Caspary (1992b) angedeutet. Die maximale Abweichung d_{max} tritt im Falle eines Kreises immer in der Mitte zwischen zwei Punkten auf (Abb. 2.9) und kann wie folgt abgeschätzt werden:

$$d_{max} = (1 - \cos \frac{\mathbf{a}}{2}) r = r - \sqrt{r^2 - (\frac{d}{2})^2}, \tag{2-1}$$

mit \mathbf{a} als Sektorenwinkel zum Digitalisierintervall d und r als Radius des Kreises. Das Digitalisierintervall entspricht dem Abstand von zwei aufeinanderfolgenden Punkten. Aus (2-1) resultiert der Zusammenhang:

$$d = 2 \sqrt{2rd_{max} - d_{max}^2} \tag{2-2}$$

dafür.

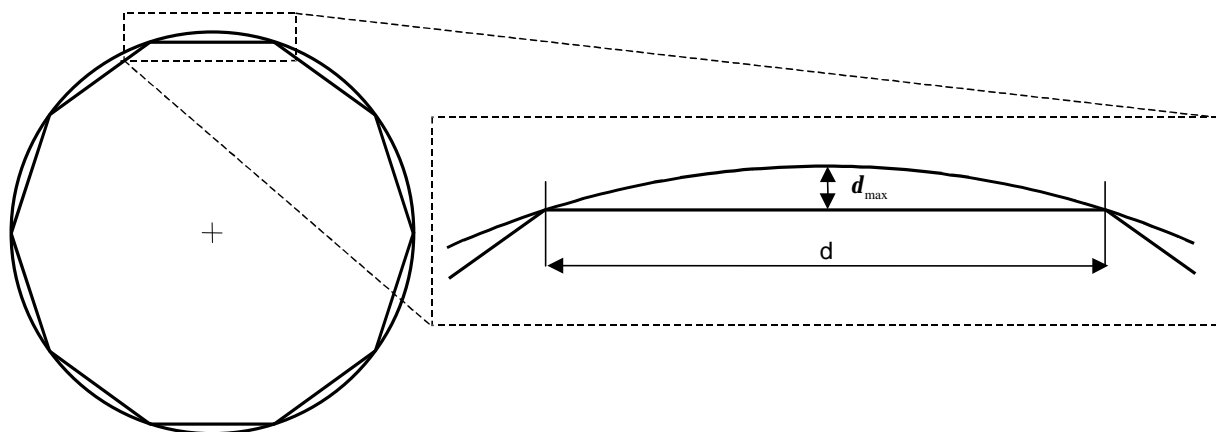


Abb. 2.9: Beispiel für eine gleichmäßige Diskretisierung eines Kreises mit $n = 10$ Punkten und zugehörigem vergrößerten Ausschnitt mit Punktabstand d und maximaler Abweichung d_{max} vom wahren Kreis.

Systematische Einflüsse sind nach Möglichkeit zu eliminieren oder in geeigneter Weise einzuschränken. Die Größenordnung des Diskretisierungseinflusses lässt sich über das Digitalisierintervall d steuern. Es ist so zu wählen, dass die maximale Abweichung deutlich kleiner als die Digitalisiergenauigkeit \bar{s} ausfällt, d.h. die Messung den dominanten Einfluss auf die geometrische Unsicherheit darstellt. In diesem Fall kann der Diskretisierungseinfluss vernachlässigt werden. Für beliebige kontinuierliche Formen liegen eine Reihe von Arbeiten vor, die sich mit der Bestimmung des optimalen Digitalisierintervalls, insbesondere im Zusammenhang mit Punkterfassungen für Höhenmodelle, auseinandersetzen (Fritsch, 1984, Fritsch, 1988b, Fritsch, 1991).

Das festgelegte Digitalisierintervall kann z.B. bei der Messung im „Stream“-Modus als Steuerparameter Verwendung finden. In diesem Modus wird eine neue Punktmessung automatisch ausgelöst, wenn von der Digitalisierlupe eine bestimmte Strecke zurückgelegt wurde. Die Strecke entspricht dem Digitalisierintervall und ist vom Nutzer vorab festzulegen. Der Operateur hat lediglich dem Linienverlauf zu folgen. Auslösung und Messung erfolgen automatisch. Dieses Verfahren zählt zu den gängigen Digitalisierungsmethoden für kontinuierliche Formen, wie z.B. Höhenlinien.

Zur konkreten Untersuchung des Diskretisierungseinflusses auf die geometrische Unsicherheit wurde ebenfalls ein Erfassungstest ausgeführt. Hierbei waren Kreise zu digitalisieren, die verschiedene Durchmesser aufweisen (Abb. 2.10). Die Festlegung, welche und wieviel Punkte zu messen waren, wurde den Operateuren freigestellt. Es galt lediglich, die Form möglichst gut zu approximieren. Dadurch entstand eine erhebliche Variation in der für notwendig gehaltenen Punktzahl (Abb. 2.11). Für den Operateur scheint demnach schwierig intuitiv festzulegen, wieviel Punkte notwendig

sind. Ein Vergleich von Mittelwert und aus (2-1) und (2-2) über $n_{\min} = 2pr/d_{\max}$ berechneter Mindestanzahl für die maximale Abweichung $d_{\max} = \bar{s} = 0.15 \text{ mm}$ zeigt, dass der Mittelwert in allen Fällen größer ist. In der Regel werden also mehr Punkte als unbedingt notwendig erfasst. Nur in Ausnahmefällen sind es zu wenige.

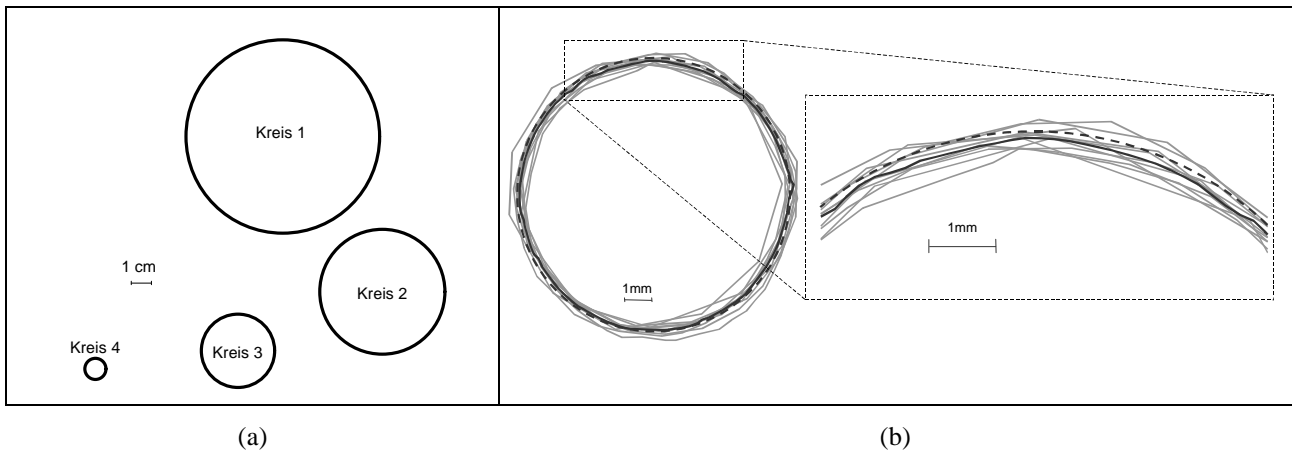


Abb. 2.10: Testaufbau: Kreise mit verschiedenen Radien (a) und ein Beispiel der unterschiedlichen Erfassungen (grau) mit Darstellung des wahren Kreises (gestrichelt) und der mittleren Geometrie (durchgezogen) (b).

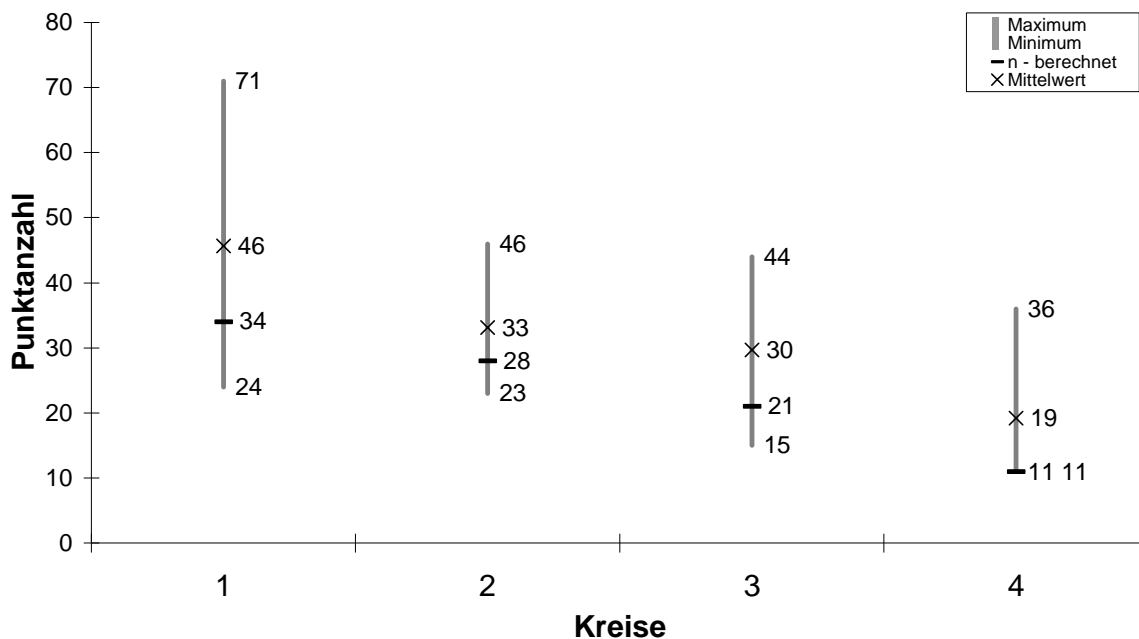


Abb. 2.11: Ergebnisse der Testdatensätze: Spannweite der Punktzahl der Realisierungen, mittlere Punktzahl und aufgrund der Vermeidung des Diskretisiereinflusses ermittelte minimal notwendige Punktzahl.

Zusammenfassend ist festzuhalten, dass sich die Einflüsse aus der Diskretisierung nur auf kontinuierliche Formen auswirken. Sie beeinflussen die Erfassung in systematischer Weise. Aus Nutzersicht ist daher eine Verringerung bzw. Elimination wünschenswert. Entscheidende Größen sind Objektform, Anzahl und Auswahl der Punkte. Zur Überprüfung eignet sich lediglich die Anzahl der Punkte. Wird als Ergebnis der Überprüfung festgestellt, dass zu wenige Punkte erfasst wurden, dann ist die Erfassung entsprechend durch weitere Punkte zu ergänzen. Zu viele Punkte erlauben dagegen eine anschließende Glättung der Form. Dadurch lässt sich eine Optimierung der Punktzahl nach dem Gesichtspunkt der Diskretisierung erreichen. Letztendlich bestimmt die Menge der gewählten Punkte die Güte der Approximation und damit die Größe des Diskretisiereinflusses.

2.3.3 Modellbildung

Die Modellbildung ist ein Schritt, der für alle Objekte notwendig ist. Aber nur im Falle von natürlichen Objekten entstehen Probleme, die zu Differenzen zwischen Modell und Wirklichkeit führen. Bei künstlichen (von Menschenhand geschaffenen) Objekten kann davon ausgegangen werden, dass sie bereits auf der Basis eines Modells entstanden sind, so dass eine neuerliche Modellbildung erheblich erleichtert und im Idealfall zu exakten Ergebnissen führen wird. Bei natürlichen Objekten ist dagegen stets ein neues Modell zu generieren. Im Wesentlichen besteht für den Anwender die Aufgabe, einzelne Objekte zu identifizieren und gegeneinander räumlich abzugrenzen. Identifizierung und Abgrenzung sind Teile eines Interpretationsprozesses, der gemäß der anwendungsbezogenen Modellierungsvorschrift abläuft. Bei der Identifizierung wird die Existenz eines Objektes festgestellt, um anschließend dessen räumliche Ausdehnung zu bestimmen. Dazu sind die Grenzen zu den Umgebungsobjekten zu definieren.

Auf die Unsicherheit in der Bestimmung der Ausdehnung wirken verschiedene Einflüsse (Abb. 2.12). Ein möglicher Beitrag kann der Modellierungsvorschrift (Abb. 2.12a) entstammen. Sie weist im Normalfall einen bestimmten Detaillierungsgrad auf, der den Maßstabsbereich der Anwendung widerspiegelt. Jedoch hält die Natur eine nahezu unendlich große Vielzahl an möglichen Fällen bereit, die unmöglich in einer Vorschrift abzudecken sind. Im Regelfall sind nur die wichtigsten Unterscheidungsmerkmale darin zu finden. Alle weiteren Entscheidungen bleiben dem Operateur überlassen. Dementsprechend können verschiedene Sichtweisen und damit, übertragen auf ein individuelles Objekt, Variationen in der Ausdehnung entstehen. Z.B. ergeben sich bei der Abgrenzung von Waldobjekten Schwierigkeiten durch einzelne Bäume oder Aufforstungsflächen, für die im Speziellen zu entscheiden ist, ob diese noch zum Wald oder bereits zur angrenzenden Wiese zählen sollen. Die Wirkung ist als grober Einfluss einzuschätzen, da die entsprechenden Regionen entweder ganz oder gar nicht hinzugehören. Einen weiteren Beitrag liefert die verwendete Datenquelle (Abb. 2.12b) selbst. Sie kann unter Umständen nicht ausreichend sein, um sich sicher für einen bestimmten Verlauf der Abgrenzung entscheiden zu können. Dies trifft z.B. dann zu, wenn Verdeckungen auftreten. Luftbilder (Orthophotos) als mögliche Datenquelle enthalten Verdeckungen in Form von Schattenbereichen und aufgrund perspektivischer Umklappereffekte. Im Gegensatz zum Detaillierungsgrad handelt es sich dabei um einen zufälligen Einfluss, da jede beliebige Abgrenzung im verdeckten Bereich möglich ist. Schließlich ist im Übergang (Abb. 2.12c) zwischen Objekten ein weiterer Beitrag begründet. Die Modellierung verlangt die Festlegung eines unendlich kleinen Übergangs zwischen zwei benachbarten Objekten und beschreibt diesen durch einen linienhaften Grenzverlauf (Grenzlinie). Natürliche Objekte weisen in der realen Welt aber einen mehr oder weniger breiten Übergang auf. Dieser definiert eine Zone, in der es schwierig ist festzulegen, wo das Objekt endet und das Nachbarobjekt beginnt. Aus diesem Grund werden mehrere Erfassungen räumlich voneinander abweichen, da jeder Operateur eine individuelle Grenze definiert. Entsprechend der Zone variiert die Ausdehnung des Objekts. In diesem Fall spricht man von einer unscharfen Grenze, im Gegensatz zu einer scharfen Abgrenzung, wie sie bei vielen künstlichen Objekten anzutreffen ist (z.B. Gebäuden, Straßen). Die Ursache für das Entstehen des Einflusses ist alleinig in der per Definition geforderten scharfen Abgrenzung begründet, so dass man diesen als Definitionsunschärfe bezeichnen kann. Würde das Datenmodell die Beschreibung von unscharfen Grenzen vorsehen, wäre diese Quelle der geometrischen Unsicherheit eliminiert.



Modellierungsvorschrift
z.B. Einzelbäume,
junger Wald

(a)

Datenquelle
z.B. Schatten

(b)

Übergangszone
z.B. Wald gegen Feldlage

(c)

Abb. 2.12: Beispiele für verschiedene Ursachen der Unsicherheit aus der Modellbildung: Modellierungsvorschrift (a), Datenquelle (b) und Übergangszone (Definitionsunschärfe) (c).

Die Übergangszone kann z.B. für Wald eine Breite von mehreren Metern, im Falle von geologischen Schichten oder Bodenarten sogar von mehreren hundert Metern besitzen. Damit wird die Definitionsunschärfe, verglichen mit Messung und Diskretisierung, zum dominierenden Faktor auf die geometrische Unsicherheit. In der Art handelt es sich um einen zufälligen Einfluss, da innerhalb der Zone prinzipiell jede Abgrenzung möglich ist. Die Größenordnung äußert sich objektindividuell, d.h. für Objekte gleicher Objektart können ganz unterschiedlich breite Übergangszonen auftreten. Gleiches gilt bereits für die Grenzen eines Objektes, in denen ebenso beträchtliche Unterschiede möglich sind.

Die Beobachtung der genannten Einflüsse innerhalb eines Untersuchungstests erweist sich als schwierig, da immer alle diskutierten Größen aus Modellbildung, Diskretisierung und Messung gemeinsam auftreten. Trotzdem lassen sich einige Erkenntnisse mit Hilfe einer weiteren Testreihe gewinnen. Aus den gebildeten Objektkategorien kann entnommen werden, dass sich punktförmige, natürliche Objekte besonders für diese Untersuchung eignen, da ihre Unsicherheit frei von Diskretisierungseinflüssen ist. Dazu sollten die Operateure einen durch einen unregelmäßigen Grauwertverlauf unscharf definierten Punkt erfassen. Die statistische Auswertung der Messungen ergibt deutlich erhöhte empirische Standardabweichungen für die Punktkoordinaten von $s_x = 0.34 \text{ mm}$ bzw. $s_y = 0.21 \text{ mm}$ gegenüber dem vorher beobachteten durchschnittlichen Wert von $\bar{s} = 0.16 \text{ mm}$, als nur die Messung als Einflussgröße wirkte. Bereits für dieses Beispiel gilt, dass die Definitionsunschärfe die dominante Rolle spielt. Die Messung bleibt in ihrer Größenordnung dagegen unbedeutend.

In einer weiteren Untersuchung galt es zwei Waldgebiete zu digitalisieren (Abb. 2.13 und Abb. 2.14). Datenquellen waren Darstellungen der Objekte in Luftbildern. Welche Abgrenzung zu den umliegenden Wiesen gefunden wurde, lag alleinig an der individuellen Interpretation der Operateure. Ein Vergleich der verschiedenen Realisierungen zeigt, dass alle möglichen Einflussgrößen der Modellbildung auf die Unsicherheit in den Daten anzutreffen sind. Da für natürliche Objekte keine scharfe Referenz existiert, startet die Auswertung mit einer Bestimmung der mittleren Geometrie eines jeden Objektes. Auf das hierzu verwendete Verfahren wird in Kapitel 3.2.1 näher eingegangen. Die senkrecht zur Linierrichtung gemessenen Abweichungen der Erfassungen zur mittleren Geometrie werden anschließend zur Bestimmung von Standardabweichungen für die Linienposition s_l genutzt. Die in Abb. 2.13 und Abb. 2.14 enthaltenen Beispiele zeigen den objektindividuellen Charakter der Modellbildungseinflüsse. Beispiel 1 (Abb. 2.13) mit $s_l = 0.50 \text{ mm}$ besitzt gegenüber Beispiel 2 (Abb. 2.14) mit $s_l = 0.20 \text{ mm}$ einen mehr als doppelt so großen Wert für die Genauigkeit der erfassten Begrenzungslinien. Dies kann dadurch begründet werden, dass das zweite Objekt sich deutlich von seiner Umgebung abgrenzt. Es besitzt nur einen geringen Übergangsbereich zu seinen Nachbarn. Weiterhin weist es im Gegensatz zu Objekt 1 nur geringe Verdeckungen durch Schattenwurf auf. Objekt 2 entspricht demnach besser den bisherigen Modellvorstellungen eines scharfen GIS-Objektes.

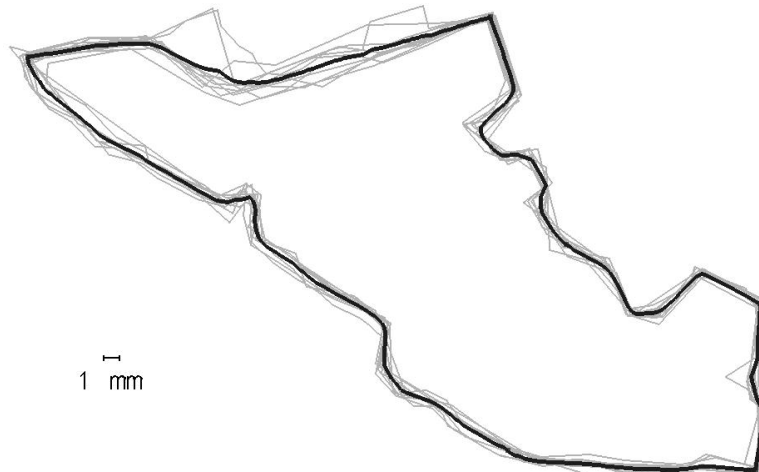
Zusätzlich erschweren Interpretationsprobleme durch Jungwald und angrenzende Baumreihen die Erfassung des Objektes 1. Einige Operateure schlossen diese Bereiche mit ein, während andere diese den Nachbarn hinzurechneten oder als eigenständige Objekte betrachteten. Besonders das Jungwaldgebiet verdeutlicht die systematische Wirkung des Einflusses auf die mittlere Geometrie des Objektes. Sie verläuft durch die kritische Region, obwohl kein Operateur einen solchen Verlauf erfassen würde. Um einen solchen offensichtlichen Fehler auszuschließen, sollten die einzelnen Realisierungen vorab überprüft werden. Eine Möglichkeit besteht in Ersetzung der Mittelbildung durch den Median. Damit können auch grob fehlerhaft wirkende Effekte beseitigt werden.

Betrachtet man Objekt 1 in seinen lokalen Verhältnissen etwas genauer, dann lassen sich deutliche Unterschiede in der Standardabweichung der Linienpositionen feststellen. Neben einigen Bereichen mit geringer Unsicherheit existieren andere Teile mit übermäßig großen Abweichungen vom Mittel. Der für das Objekt berechnete Wert der Standardabweichung repräsentiert demnach nur einen globalen Wert, der kaum auf die individuellen Eigenheiten eingeht. Eine bessere Beschreibung gelingt daher, wenn das Objekt in einzelne Liniensegmente zerlegt wird, die jeweils eine weitgehend homogene Unsicherheit beinhalten.

Zusammenfassend ist festzustellen, dass mehrere Einflüsse aus der Modellbildung auf die geometrische Unsicherheit einwirken. Aufzählen lassen sich dazu die Probleme mit der Detaillierung der Modellierungsvorschrift, mit der Datenquelle und mit der Definitionsunschärfe. Während natürliche Objekte meist von allen Einflüssen betroffen sind, fällt bei künstlichen Objekten die Definitionsunschärfe weg. Die Einflüsse äußern sich sowohl in systematischer als auch in zufälliger Weise. Sie besitzen häufig objektindividuellen Charakter und können zudem innerhalb der Objekte stark variieren, so dass zur Bestimmung der Unsicherheit eine Unterteilung der Objektgeometrie in kleinere Einheiten mit homogenen Eigenschaften sinnvoll ist. Im Vergleich mit den Größenwerten der Einflüsse aus Diskretisierung und Messung kann die Wirkung der Modellbildung den dominierenden Faktor darstellen. In der Regel bilden jedoch alle drei Einflüsse gemeinsam den gesamten Wert der Unsicherheit für ein Objekt.

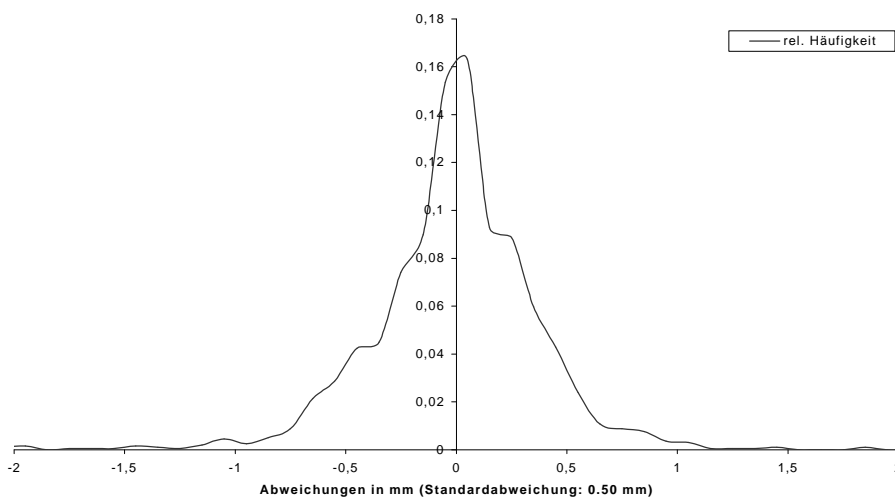


(a)



(b)

Test: Relative Häufigkeit der Abweichungen von der mittleren Geometrie (Wald 1)

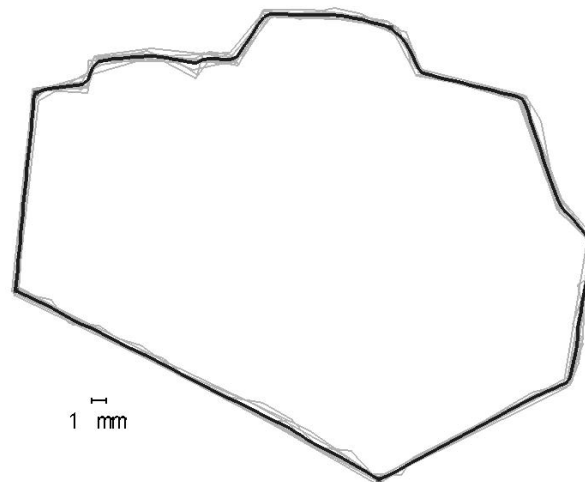


(c)

Abb. 2.13: Beispiel 1 für einen großen Einfluss der Modellbildung auf die Unsicherheit bei natürlichen Objekten: Luftaufnahme (a), die Ergebnisgeometrien von 10 unabhängigen Erfassungen (grau) mit zusätzlicher Darstellung der mittleren Geometrie (schwarz) (b) und Häufigkeitsverteilungen der Abweichungen der Erfassungen von der Mittelgeometrie (c).

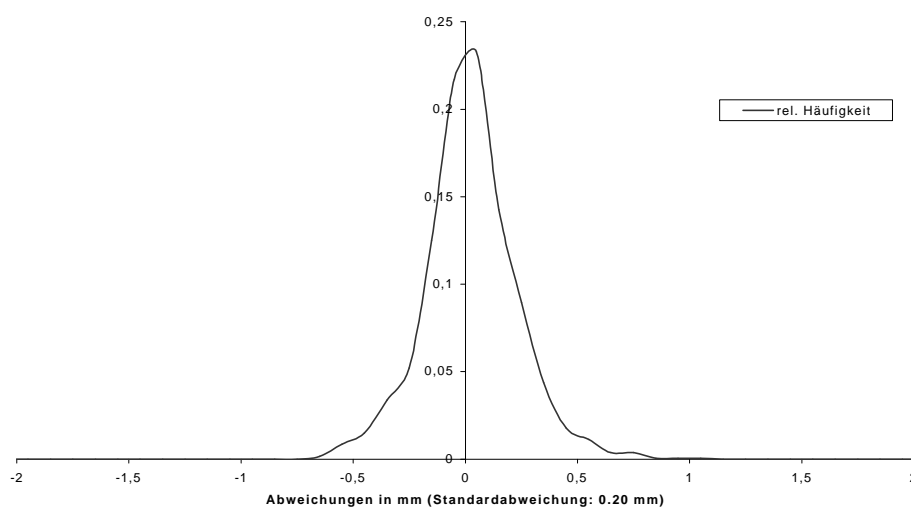


(a)



(b)

Test: Relative Häufigkeiten der Abweichungen von der mittleren Geometrie (Wald 3)



(c)

Abb. 2.14: Beispiel 2 für einen kleinen Einfluss der Modellbildung auf die Unsicherheit bei natürlichen Objekten: Luftaufnahme (a), die Ergebnisgeometrien von 10 unabhängigen Erfassungen (grau) mit zusätzlicher Darstellung der mittleren Geometrie (schwarz) (b) und Häufigkeitsverteilungen der Abweichungen der Erfassungen von der Mittelgeometrie (c).

2.3.4 Diskussion

Aus den Untersuchungen zeigt sich, dass alle drei genannten Ursachen einen Beitrag zur Unsicherheit der Geometrie leisten. Die Diskretisierung nimmt eine Sonderstellung ein, da deren Wirkung durch eine geeignete Punktauswahl und eine an die Krümmung der Objektgeometrie angepasste Punktzahl vernachlässigbar gering gegenüber den beiden anderen Einflüssen gehalten werden kann. Ebenso ist eine Erhöhung der Messgenauigkeit möglich, jedoch lässt sich der Einfluss nicht gänzlich eliminieren.

Am wichtigsten erscheint die unterschiedliche Wirkung der Einflüsse auf die beiden Objektkategorien der natürlichen und der künstlichen Objekte. Die geometrische Unsicherheit für künstliche Objekte hängt im Wesentlichen von der Messung ab. Direkt davon betroffen sind lediglich die gemessenen Punkte, wobei sich die Wirkung indirekt auf die Unsicherheit höherdimensionaler Formen (Linien und Flächen), d.h. auf die weiteren Geometriepunkte fortpflanzt. Aus der Fortpflanzung der Genauigkeit von Linienanfangs- und Linienendpunkt auf einen beliebigen Linienpunkt ergibt sich z.B. die bekannte Steigerung der Genauigkeit zur Linienmitte hin (Caspary und Scheuring, 1992). Somit genügt für künstliche Objekte in vielen Fällen die Beschreibung der Genauigkeit der Punktkoordinaten aller Messpunkte zur Veranschaulichung der Unsicherheit in der Geometrie.

Im Falle natürlicher Objekte verlieren die Messpunkte ihre entscheidende Bedeutung für die Unsicherheit. Sie legen zwar als markante Punkte weiterhin die Form des Objektes fest, doch bestimmen alle anderen Geometriepunkte die Objektausdehnung in gleicher Weise mit. Die Begründung dafür findet sich im Schritt der Modellbildung, der eine Abstraktion der als zufällig aufzufassenden Objektausdehnung (z.B. wegen der natürlichen Unschärfe) bewirkt. Alle Geometriepunkte sind in gleicher Weise von diesem Schritt betroffen, so dass die geometrische Unsicherheit eine individuelle Eigenschaft eines jeden Punktes (nicht nur eines Messpunktes) darstellt. Wie die Untersuchungen ergaben, spielt der Einfluss der Modellbildung auf die Unsicherheit natürlicher Objekte die dominante Rolle. Zwar sind Messung und Diskretisierung stets mit enthalten, doch äußern sich diese in den meisten Fällen unbedeutend gegenüber der Modellbildung. Deshalb hat die Beschreibung der Unsicherheit individuell für jeden Geometriepunkt zu erfolgen. Da sich für Linien und Flächengrenzen unendlich viele Geometriepunkte unterscheiden lassen, ist eine Zusammenfassung in einzelne Segmente mit gleichen Eigenschaften sinnvoll. Segmente besitzen den Vorteil, dass der objektspezifische Charakter der Unsicherheit bewahrt bleibt.

Die beiden beschriebenen Teilschritte zur Erfassung von Objekten lassen sich innerhalb eines realen Verfahrens nicht unmittelbar voneinander trennen. Das ist daran abzulesen, dass das Anwendungsmodell nur eine virtuelle gedankliche Ausbildung der Objekte darstellt, die erst im zweiten Teilschritt zu einem physikalischen GIS-Modell führt. Dennoch besteht eine klare Trennung, die im individuellen Fall sinnvoll ausgenutzt werden kann. Z.B. braucht ein Objekt (z.B. Fluss) nicht mit hohem Aufwand durch ein genaues Verfahren (z.B. terrestrische Vermessung) gemessen werden, wenn es nur unscharf zu definieren ist. In diesem Fall entscheidet die Modellbildung, welcher Messaufwand sinnvoll ist.

2.4 Unsicherheitsmodelle

Die Berücksichtigung der Unsicherheit bei der Verarbeitung von Daten erfordert die Definition eines Unsicherheitsmodells. Ein solches Modell liefert die Grundlage zur Repräsentation der Unsicherheit innerhalb des Systems und ist daher als wichtige Komponente in das Datenmodell zu integrieren. Als wesentlicher Bestandteil werden darin Maße definiert, die die Unsicherheit quantitativ oder qualitativ beschreiben. Die Maße sind zumeist in ein theoretisches Umfeld eingebettet, mit dessen Hilfe Verarbeitungsformeln abgeleitet werden können. Die Formeln finden z.B. dann Verwendung, wenn beliebige Kombinationen von unsicheren Daten durchgeführt werden und daraufhin eine Aussage zur Unsicherheit des Ergebnisses herzustellen ist oder bei Bewertungen von Schlussfolgerungen (wenn..., dann...) in regelbasierten Systemen, die zur Entscheidungsfindung eingesetzt werden. Die Maße sind als zusätzliche Angaben in die Datenmodellierung einzubringen. Sie werden dadurch selbst Bestandteil der Daten, die durch diese erweiterte Beschreibung höher in ihrer Qualität einzustufen sind, d.h. die Nutzung eines Unsicherheitsmodells bewirkt eine Steigerung der Qualität der Daten.

Es sind eine ganze Reihe von Unsicherheitsmaßen bekannt, die jeweils in unterschiedlichen Zusammenhängen angewendet werden. Die Einordnung der Maße kann aufgrund der Eigenschaften der zu bewertenden Inhalte geschehen. Eine Gruppe hat diskrete Ereignisse (z.B. Landnutzung des Objektes A ist Acker) oder Aussagen (z.B. Objekt A ist länglich) zum Inhalt. Dazu lassen sich die folgenden Maße zuordnen (Bhatnagar und Kanal, 1986, Witke, 1987, Vosselman, 1996): Wahrscheinlichkeiten, Glaubwürdigkeits- und Plausibilitätsmaße gemäß der Dempster-Shafer-Theorie (Shafer, 1976), Sicherheitsfaktoren, wie sie im bekannten System MYCIN (Buchanan und Shortliffe, 1984) verwendet werden und Möglichkeitswerte der Unschärfen-Mengen-Theorie (Zadeh, 1965). Sie können als Aussagemäße bezeich-

net werden. Die Aufzählung ist keineswegs erschöpfend, sondern es sind hier lediglich die Bekanntesten genannt. Ein Beispiel zur Bewertung eines Ereignisses mit Hilfe von Wahrscheinlichkeiten könnte damit folgendermaßen aussehen:

Die Landnutzung des Objektes A ist Acker mit einer Wahrscheinlichkeit von $p = 80\%$.

Eine zweite Gruppe bilden Ereignisse, die in ihrer Ausprägung kontinuierlich sind, d.h. im Sinne der Unsicherheit einer Variation unterliegen. Für sie eignen sich alle Beschreibungsarten, die in der Lage sind, die Variation in Größe und/oder Form zu erfassen. Dazu dienen im Wesentlichen die Maße der mathematischen Statistik: Variationsbreite (Spannweite), Mittelwert und Varianz zur Charakterisierung der zugehörigen Verteilung und andere (z.B. Sachs, 1978). Sie lassen sich als Variationsmaße oder Streuungsmaße bezeichnen. Auch hierzu ein Beispiel mit Hilfe von Mittelwert und Varianz:

*Die Länge einer Straße beträgt 200 m (Mittelwert)
mit einer Standardabweichung von $s = 10$ m.*

Im Beispiel wird implizit das Vorliegen einer bestimmten Verteilung angenommen. Da auf kontinuierliche Messwerte in aller Regel nur zufällige Einflüsse wirken, ist in den meisten Fällen von der Normalverteilung auszugehen.

Die Aussagemäße finden im Besonderen in Problemlösungs- und Entscheidungsfindungssystemen ihre Anwendung (Kanal und Lemmer, 1986). Solche Systeme sind gezielt auf die Verarbeitung von unsicherem Wissen ausgerichtet. Die Basis bilden Aussagen, die mit Hilfe eines Unsicherheitsmaßes auf ihre Vertrauenswürdigkeit hin bewertet werden. Verschiedene Regeln ermöglichen die Verknüpfung der Aussagen mit dem Ziel, neue Aussagen abzuleiten oder bestehende zu bestätigen oder zu verwerfen. Dabei dürfen auch die Regeln selbst unsicher sein, solange sich deren Grad der Unsicherheit bewerten lässt. Am Ende erhält der Nutzer eine Reihe von untereinander konsistenten Aussagen mit zugehörigen Unsicherheiten, die den möglichen Sachverhalt der Wirklichkeit repräsentieren und die er als Entscheidungshilfe heranziehen kann.

Da geometrische Primitive zu den kontinuierlichen Ereignissen zählen, kommen zunächst nur Variationsmaße zur Beschreibung der geometrischen Unsicherheit in Betracht. Sie bilden die Basis für die Definition möglicher Modelle. Aussagemäße lassen sich dagegen nicht ohne weiteres zur Repräsentation benutzen. Jedoch sind beliebige Aussagen auf der Basis der geometrischen Verhältnisse abzuleiten, die dann durch Aussagemäße in ihrer Unsicherheit zu beschreiben sind. Beispielsweise ist vielfach die Frage interessant, ob man sich an einer beliebigen Position im Raum innerhalb eines bestimmten Objektes befindet. Die Bewertung dieser Aussage (man befindet sich innerhalb des Objektes) hat durch die Ableitung eines Aussagemäßes aus der Unsicherheit der geometrischen Primitive zu erfolgen. Da es sich hierbei um eine häufig benötigte Aussage handelt, soll diese als geometrische Grundaussage besondere Berücksichtigung finden. Insbesondere soll deren Bewertung in die nachfolgend beschriebenen Modelle integriert werden.

Die Definition eines Unsicherheitsmodells für die Geometrie hat zwei wichtige Teilaspekte zu umfassen:

1. Beschreibung der Variation der geometrischen Primitive mit Hilfe eines Variationsmaßes
2. Bewertung von Aussagen, die sich aus der Geometrie heraus ableiten mit Hilfe eines Aussagemäßes, das auf der Beschreibung der Variation aufbaut.

Im Folgenden werden drei unterschiedliche Modelle definiert. Das Stochastische Modell stützt sich auf die Mittel der mathematischen Statistik und verwendet Varianz und Mittelwert als Maße zur Beschreibung der Variation. Als zweite Komponente bedient es sich der Wahrscheinlichkeitstheorie zur Bewertung von Aussagen. Minimale und maximale Ausdehnung als Ausdruck der Variationsbreite bilden die Grundlage für das Minimum-Maximum-Modell. Die Unsicherheit von Aussagen wird darin durch qualitative Möglichkeitsmaße repräsentiert. Das Unschärfe-Mengen-Modell oder Fuzzy-Modell verwendet die geometrische Grundaussage (Position innerhalb des Objektes) direkt als Basis und unterscheidet sich daher schon in der Vorgehensweise von den anderen beiden genannten Ansätzen. Die Bewertung wird durch ein quantitatives Möglichkeitsmaß realisiert, das sich mit Hilfe einer Zugehörigkeitsfunktion bestimmen lässt. Theoretische Grundlage dafür bildet die Unschärfe-Mengen-Theorie (Fuzzy-Subset-Theorie) nach Zadeh (1965). Die Parameter der Zugehörigkeitsfunktion können als Variationsmaße aufgefasst werden. Sie spielen in diesem Ansatz jedoch nur eine untergeordnete Rolle. Die Definitionen der Modelle werden nachfolgend im Einzelnen erläutert.

In den Erläuterungen wird häufig der Begriff der Realisierung einer Objektgeometrie verwendet. Darunter ist genau eine mögliche Erfassung der Geometrie zu verstehen, auf die alle bereits diskutierten Ursachen, die zum Entstehen der Unsicherheit beitragen, einwirken. Im statistischen Sinne entspricht die Realisierung einem Zufallsexperiment eines zugeordneten Versuchs. Die Sammlung aller Realisierungen bildet die Grundgesamtheit, die im Falle der Geometrie unendlich viele Elemente enthält. In praktischen Anwendungen sind jedoch nur endlich viele Erfassungen zu erzeugen, so dass immer nur eine Stichprobe mit endlich vielen Realisierungen vorliegt.

2.4.1 Stochastisches Modell

Grundlage des Stochastischen Modells bilden statistische Kenngrößen, die die Aufgabe besitzen, die Variation der Geometrie zu beschreiben. Die geometrischen Primitive sind als Zufallsvariablen aufzufassen. Zugeordnetes Zufallsexperiment ist die jeweilige Erfassungsmethode, die als Ergebnis einzelne Realisierungen liefert. Zur Beschreibung der Variation werden als Maßzahlen insbesondere die ersten beiden Momente der zugehörigen Verteilungsfunktion genutzt. Das erste Moment definiert die mittlere Objektgeometrie (kurz: Mittel), die die mittlere geometrische Lage des Objektes veranschaulicht. Das zweite zentrale Moment bezeichnet die Varianz der Lage. Einige allgemeine Grundlagen zur Stochastik können dem Anhang B.1 entnommen werden.

Um eine Beschreibung der Variation zu realisieren, betrachtet das Modell die geometrischen Primitive Punkt, Linie und Fläche als stochastische Variablen. Da es sich dabei um komplexe Gebilde handelt, können diese nicht direkt und in einfacher Weise durch die genannten Momente beschrieben werden. Vielmehr ist es Aufgabe der Modelldefinition, eine einfache und einheitliche Vorgehensweise für alle Primitive zu entwickeln. Grundsätzlich bieten sich zwei Beschreibungsmöglichkeiten dafür: ein punktwiser Ansatz, der auf die gemessenen Koordinaten basiert, und ein linienwiser Ansatz, bei dem die Objektlinien im Mittelpunkt stehen. Beide Ansätze werden nachfolgend näher beschrieben und diskutiert.

Punktweiser Ansatz zur geometrischen Unsicherheit

Aufgrund der inneren topologischen Zusammenhänge der Grundprimitive liegt es nahe, die Beschreibung den Punktkoordinaten (x, y) anzuhängen. Sie bilden die unterste geometrische Ebene und finden sich als Grundbestandteile in allen Primitiven wieder, die hierarchisch auf diese aufbauen (Punkt->Linie->Fläche). In solch einem punktweisen Ansatz besteht nun die Möglichkeit, alle Koordinaten als stetige Zufallsvariablen zu betrachten und diese individuell durch ihre Momente zu beschreiben (Abb. 2.15). Für einen Punkt ergeben sich die folgenden Ausdrücke:

$$\text{Mittelwert : } \mathbf{m}_P = \begin{pmatrix} \mathbf{m}_x \\ \mathbf{m}_y \end{pmatrix} \tag{2-3}$$

und

$$\text{Dispersionsmatrix: } \mathbf{D}(P) = \begin{pmatrix} \mathbf{s}_x^2 & \mathbf{s}_{xy} \\ \mathbf{s}_{yx} & \mathbf{s}_y^2 \end{pmatrix}. \tag{2-4}$$

An die Stelle der einfachen Varianz tritt die Dispersionsmatrix \mathbf{D} , die die Varianzen und Kovarianzen der individuellen Zufallsvariablen (hier: Koordinaten) beinhaltet.

Darauf aufbauend leitet sich die Variation einer Linie über die Koordinaten des zugehörigen Anfangs- und Endpunktes ab und kann durch Angabe der Variation in einem beliebigen Linienpunkt P_i repräsentiert werden (Abb. 2.15b). Grundlage dafür bildet der folgende funktionale Zusammenhang f zwischen Linienpunkt und Anfangs- und Endpunkt (P_A, P_B) der Linie (Zhang und Tulip, 1990):

$$f : \begin{pmatrix} x_{P_i} \\ y_{P_i} \end{pmatrix} = \begin{pmatrix} 1-\mathbf{a}_i & 0 & \mathbf{a}_i & 0 \\ 0 & 1-\mathbf{a}_i & 0 & \mathbf{a}_i \end{pmatrix} \begin{pmatrix} x_{P_A} \\ y_{P_A} \\ x_{P_B} \\ y_{P_B} \end{pmatrix} \text{ mit } 0 \leq \mathbf{a}_i \leq 1. \tag{2-5}$$

Während sich die Mittelwerte direkt aus der Funktion ergeben, ist zur Bestimmung der Dispersionsmatrix eine Varianzfortpflanzung durchzuführen. Sie liefert das folgende Ergebnis für die Dispersionsmatrix \mathbf{D} eines beliebigen Linienpunktes P_i :

$$\mathbf{D}(P_i) = \begin{pmatrix} (1-\mathbf{a}_i)^2 \mathbf{s}_{x_A}^2 + \mathbf{a}_i^2 \mathbf{s}_{x_B}^2 & (1-\mathbf{a}_i)^2 \mathbf{s}_{x_A y_A} + \mathbf{a}_i^2 \mathbf{s}_{x_B y_B} \\ (1-\mathbf{a}_i)^2 \mathbf{s}_{y_A x_A} + \mathbf{a}_i^2 \mathbf{s}_{y_B x_B} & (1-\mathbf{a}_i)^2 \mathbf{s}_{y_A}^2 + \mathbf{a}_i^2 \mathbf{s}_{y_B}^2 \end{pmatrix}. \tag{2-6}$$

Für die Ableitung gilt die Annahme, dass Anfangs- und Endpunkt voneinander unabhängig erfasst werden, d.h. es bestehen keine Korrelationen zwischen beiden Punkten. Nimmt man zusätzlich an, dass die Varianzen der einzelnen

Punktkoordinaten identisch und die Kovarianzen gleich Null sind, dann vereinfacht sich die Dispersionsmatrix weiter zu:

$$D(P_i) = \begin{pmatrix} (1-2a_i + 2a_i^2)s^2 & 0 \\ 0 & (1-2a_i + 2a_i^2)s^2 \end{pmatrix} \tag{2-7}$$

mit

$$s_{x_{P_A}}^2 = s_{x_{P_B}}^2 = s_{y_{P_A}}^2 = s_{y_{P_B}}^2 = s^2, s_{x_A y_A} = s_{x_B y_B} = 0. \tag{2-8}$$

Betrachtet man die Festlegung der Laufvariablen zu $0 \leq a_i \leq 1$ genauer, dann wird offenkundig, dass die Varianzen der Linienpunkte mit wachsender Entfernung von Anfangs- und Endpunkt abnehmen, wobei das Minimum in der Linienmitte aufzufinden ist. Es entsteht die bekannte Knochenform (Abb. 2.15b), d.h. eine zur Linienmitte eingedrückte Form eines auf dieser Basis berechneten Fehlerbandes (Zhang und Tulip, 1990, Caspary und Scheuring, 1992).

Über die geometrische Beschreibung einer Fläche durch eine geschlossene Folge an Randlinien ergibt sich, dass Flächen nur am Rand unsicher sind. Insofern lässt sich die Variation einer Fläche direkt aus den einzelnen Variationen der umschreibenden Linien ableiten (Abb. 2.15c). Somit gilt auch hier, dass die Variation einer Fläche durch Angabe der Variation eines beliebigen Randlinienpunktes P_i repräsentiert wird. Individuell ist stets zu klären, zu welcher Randlinie der Punkt hinzugehört. Für das Beispiel einer Dreiecksfläche kann die resultierende Dispersion formal folgendermaßen dargestellt werden:

$$D(P_i)_D = \begin{cases} D(P_i)_{AB}, & \forall P_i \in \overline{AB} \\ D(P_i)_{BC}, & \forall P_i \in \overline{BC} \\ D(P_i)_{CA}, & \forall P_i \in \overline{CA}. \end{cases} \tag{2-9}$$

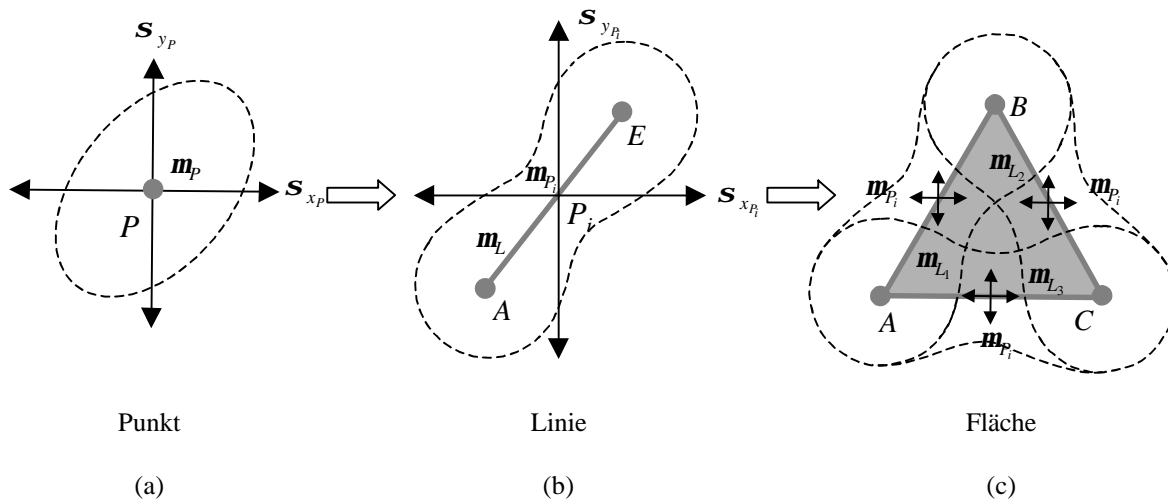


Abb. 2.15: Variation der geometrischen Primitive im punktwisen Ansatz für Punkt (a), Linie (b) und Fläche (c).

Der punktwise Ansatz eignet sich gut für die Beschreibung von Punktprimitive. Er weist jedoch sowohl konzeptionell als auch methodisch Nachteile auf, wenn Linien oder Flächen behandelt werden sollen. Dies wird am Beispiel einer Linie weiter ausgeführt. Der Ansatz geht davon aus, dass Ursachen für Variationen allein in den Koordinaten von Anfangs- und Endpunktes enthalten sind. Die Variation eines beliebigen Linienpunktes hängt demnach ausschließlich von seiner Lage gegenüber den beiden Linienbegrenzern und deren Variationen ab. Eigene Ursachen finden keine Berücksichtigung. Nur ein bestimmter Teil an möglichen Objekten erfüllt diese Anforderung. Sie gehören der Objektkategorie der künstlichen, diskreten Objekte an. Dazu zählen z.B. Flurstücksgrenzen oder Gebäudekanten, die durch diesen Ansatz in ausreichender Weise zu beschreiben sind. Alle anderen Objekte, insbesondere natürliche Objekte, deren Linienpunkte stets individuellen Ursachen ausgesetzt sind, lassen sich nicht zufriedenstellend im Ansatz unterbringen. Insbesondere bleibt in diesen Fällen die Abnahme der Variation in der Linienmitte unbegründet. Im Normalfall enthält die Dispersionsmatrix individuelle Werte und ist meist vollbesetzt, so dass ein erheblicher Speicheraufwand entsteht, da mindestens die Matrizen der Anfangs- und Endpunkte explizit abzuspeichern sind. Geeignete Kompressionsverfahren werden dafür benötigt. Wird z.B. die Fortpflanzung der Unsicherheit auf das Analyseergebnis notwendig, so sind vorab aufwendige Berechnungen durchzuführen, um die unterschiedlichen Werte für die einzelnen Linienpunkte zu erzeugen.

Hier stellt sich zusätzlich die Frage, wieviel Punkte zu berechnen und anschließend zu verarbeiten sind. Theoretisch entspricht die Linie einem Zufallsvektor unendlicher Dimension mit zugehöriger multivariater Verteilung, die in geeigneter Weise zu diskretisieren ist. Zudem repräsentiert die Dispersionsmatrix die Unsicherheit in Richtung der Koordinatenachsen. Gesucht ist jedoch eine Darstellung der Unsicherheit in jedem Punkt bezüglich der Linie, bei der die Komponenten die Variation senkrecht zur Linie und in Richtung der Linie veranschaulichen. Der Anteil in Richtung der Linie kann für einen allgemeinen Linienpunkt vernachlässigt werden, da keine nach außen ersichtliche Einwirkung auf die Geometrie des Objektes davon ausgeht. Es verbleibt die Komponente senkrecht zur Linienrichtung, die über die Dispersionsmatrix eines Linienpunktes individuell abzuleiten ist.

Wie an den diskutierten Problempunkten zu sehen ist, eignet sich der punktweise Ansatz nur mit erheblichen Einschränkungen für eine allgemeine Modelldefinition. Aus diesem Grund soll auf eine linienbasierte Modelldefinition übergegangen werden, die die Nachteile weitgehend überwindet. Daher folgt zunächst die Beschreibung für eine Linie, aus der sich dann Definitionen für Punkt und Fläche heraus ableiten.

Linienweiser Ansatz zur geometrischen Unsicherheit

Die mittlere Liniengeometrie setzt sich weiterhin aus den einzelnen mittleren Positionen aller zugehörigen Linienpunkte zusammen, während an die Stelle einer vollbesetzten Dispersionsmatrix eine einzige Varianz für die Linie tritt (Abb. 2.16b). Sie repräsentiert die Komponente der Unsicherheit senkrecht zur Linienrichtung in jedem Punkt und gilt einheitlich für die gesamte Ausdehnung der Linie. Ausnahmen bilden Anfangs- und Endpunkt der Linie. Dort beschreibt die Varianz auch die Unsicherheit in alle beliebigen Richtungen. Damit genügen für jede Linie folgende Angaben:

- mittlere Liniengeometrie $\mathbf{m}_L = (\bigcup_i \mathbf{m}_{L_i})$ und deren Varianz \mathbf{s}_L^2 .

Eine Linie besitzt damit stets eine konstante Varianz. Ändert sich die Varianz, so sind Liniensegmente zu bilden, die als eigenständige Linien zu behandeln sind. Damit lassen sich auch komplexere Situationen abbilden, in denen sich die Variation innerhalb des Linienverlaufes stark ändert.

Ein Punkt kann als Linie der Länge Null angesehen werden, d.h. Anfangs- und Endpunkt der Linie fallen zusammen. Somit können die Festlegungen, die zu den Anfangs- und Endpunkten einer Linie getroffen wurden, auch von Einzelpunkten übernommen werden (Abb. 2.16a). Damit sind für Punkte anzugeben:

- mittlere Punktgeometrie $\mathbf{m}_p = \{ \mathbf{m}_x, \mathbf{m}_y \}$ und deren Varianz \mathbf{s}_p^2 .

Die Varianz gilt für alle Koordinaten bzw. wirkt in alle beliebigen Richtungen r in gleicher Weise:

$$\mathbf{s}_x^2 = \mathbf{s}_y^2 = \mathbf{s}_r^2 = \mathbf{s}_p^2. \tag{2-10}$$

Dies führt dazu, dass die häufig im Entstehungsprozess (z.B. Netzausgleichung, Homogenisierung) erzeugte vollbesetzte Kovarianzmatrix der Koordinaten durch die vorgenommene Generalisierung verloren geht. Jedoch wird dadurch erst die in diesem Ansatz verfolgte universelle Anwendbarkeit ermöglicht.

Im Allgemeinen baut sich die Geometrie einer Fläche aus einer gewissen Anzahl an Linien L_i auf. Sie sind zur Beschreibung der Unsicherheit getrennt voneinander zu betrachten (Abb. 2.16c). Für eine Fläche sind daher anzugeben:

- mittlere Flächengeometrie $\mathbf{m}_F = (\bigcup_i \mathbf{m}_{L_i})$ und deren Varianz $\mathbf{s}_F^2 = (\bigcup_i \mathbf{s}_{L_i}^2)$.

Die Flächenvarianz besteht somit aus der Vereinigungsmenge der einzelnen Linienvarianzen der den Rand bildenden Linien.

Ein formaler Ausdruck für die Variation V kann folgendermaßen gegeben werden:

$$\text{Variation } V : V_{P,L,F} = \{ \mathbf{m}_{P,L,F}, \mathbf{s}_{P,L,F}^2 \}. \tag{2-11}$$

Die jeweils gültigen Größen sind gemäß den gegebenen Definitionen einzusetzen. Zur Repräsentation der Variation ist ein Variationsmaß abzuleiten, das einen Vergleich unterschiedlicher Objekte ermöglichen soll. Als Variationsmaß eignet sich die Standardabweichung \mathbf{S} , die einen direkten metrischen Vergleich möglich macht.

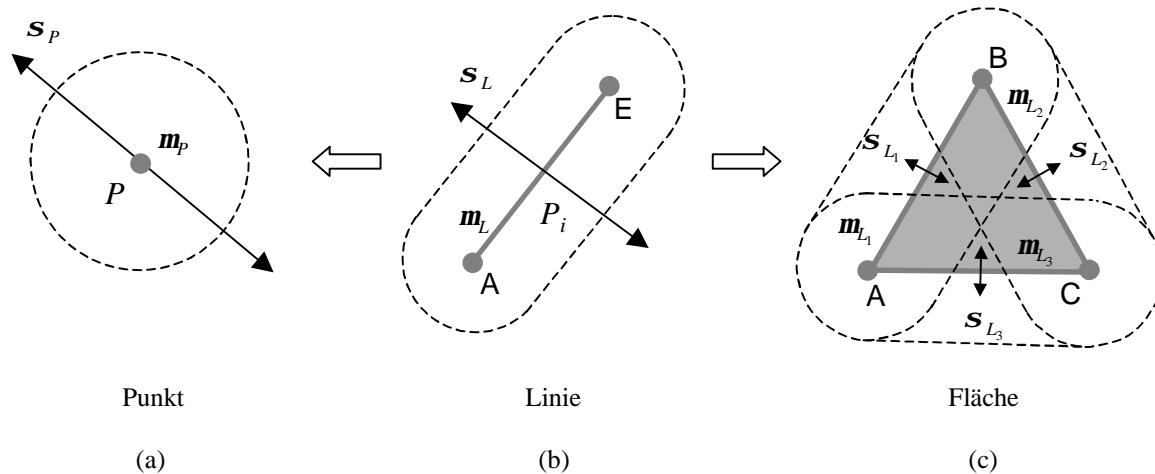


Abb. 2.16: Variation der geometrischen Primitive im linienbasierten Ansatz für Punkt (a), Linie (b) und Fläche (c).

Wie bereits erwähnt, benötigt man zusätzlich eine Annahme zur Verteilung der Variablen. Aus den durchgeführten Tests kann beobachtet werden, dass sich die Ursachen der Unsicherheit in ihrer Gesamtheit annähernd normalverteilt verhalten. Die Annahme der Normalverteilung ist zudem gängige Praxis in der Geodäsie, als einem der wichtigsten Bereiche, der sich mit der Erfassung von geometrischen Formen beschäftigt. Dabei wird davon ausgegangen, dass die Summe der wirkenden Einflüsse nur kleine, zufällige Abweichungen vom wahren Wert entstehen lässt, systematische Einwirkungen dagegen in nur unbedeutendem Umfang vorhanden sind. Unter dieser Voraussetzung kann eine Herleitung der Normalverteilung als Verteilung dieser zufälligen Einflüsse erfolgen (Koch, 1980, Höpcke, 1980). Von den hier behandelten Ursachen wirken nur Diskretisierungseinflüsse in systematischer Weise. Wird ihre Größenordnung beispielsweise durch eine angepasste Punktzahl gering gehalten, so kann begründeter Weise von einer Normalverteilung des Gesamteinflusses ausgegangen werden. Eine wichtige Argumentationshilfe liefert weiterhin der zentrale Grenzwertsatz (Cramer, 1946, Wilks, 1962). Er besagt, dass die Verteilung der Summe voneinander unabhängiger Zufallsvariablen mit beliebigen Verteilungen asymptotisch gegen eine Normalverteilung strebt, wenn deren Anzahl gegen unendlich geht. Wie in Kapitel 2.3 erläutert, trägt eine Reihe unterschiedlicher Einflüsse zur Entstehung der Unsicherheit bei, so dass die Annahme der Normalverteilung aufgrund des zentralen Grenzwertsatzes berechtigt erscheint. Grundsätzlich fordert das Modell von sich aus keine konkrete Festlegung einer bestimmten Verteilung, da die Definitionen davon unabhängig formuliert sind. Jedoch wird aus Gründen eines besseren Verständnisses allen in der Arbeit enthaltenen Beispielanwendungen die Annahme der Normalverteilung zu Grunde gelegt.

Aussagemaß

Die Stochastik zeichnet sich durch eine Verbindung der Statistik mit der Wahrscheinlichkeitstheorie aus (Anhang B.1). Diese Verbindung ermöglicht die Nutzung von Wahrscheinlichkeiten als alternatives Maß zur Beschreibung der Unsicherheit. Insbesondere eignen sich Wahrscheinlichkeiten als Aussagemaß und in diesem Zusammenhang für die Bewertung der geometrischen Grundaussage. Entscheidend dafür ist der minimale Abstand einer beliebigen Position im Raum zur Objektgeometrie. Der minimale Abstand wird bei Linien und Flächen durch den senkrechten Abstand zur nächstgelegenen Objektklinie festgelegt, während bei Punkten die einfache lineare Distanz zu bestimmen ist. Der Abstand D berechnet sich aus:

$$D = \sqrt{(x_{pos} - x_{\bar{P}})^2 + (y_{pos} - y_{\bar{P}})^2}, \quad (2-12)$$

mit \bar{P} als Lotfußpunkt auf der Objektgeometrie und pos als beliebige Position im Raum. Durch den funktionalen Zusammenhang des Abstandes zur variablen Geometrie pflanzt sich die Zufallsabhängigkeit vom geometrischen Primitiv auf die abgeleitete Größe fort, so dass der Abstand gleichfalls als stochastische Variable zu behandeln ist. Die Variation der Geometrie wird dabei vollständig auf den Abstand übertragen. Es gilt:

$$\mathbf{s}_D^2 = \mathbf{s}_{P,L,F}^2. \quad (2-13)$$

Die Position im Raum wird als bekannt und damit fehlerfrei angenommen. Ableitungen der Varianz des Abstandes mit fehlerbehafteten Positionen sind in Kraus und Haussteiner (1993) gegeben.

Wahrscheinlichkeit für die Zugehörigkeit zu einer Fläche

Für die möglichen Objekttypen (Punkt, Linie, Fläche) ergeben sich unterschiedliche Formeln zur Berechnung der Wahrscheinlichkeiten. Für eine Fläche ist zu unterscheiden, ob sich die ausgewählte Position innerhalb oder außerhalb der mittleren Objektgeometrie befindet (Abb. 2.17a). Dies kann durch einen einfachen Punkt-im-Polygon-Test (z.B. mit Hilfe des Theorems von Jordan) festgestellt werden. Je nach Ergebnis dieser Untersuchung erhält das Abstandsmaß ein Vorzeichen nach folgender Konvention:

$$\begin{aligned}
 \text{Position innerhalb: } & d > 0 \\
 \text{Position außerhalb: } & d < 0 \\
 \text{Position auf Rand: } & d = 0.
 \end{aligned}
 \tag{2-14}$$

Die Lage einer Position auf dem Rand ergibt direkt den Abstand Null. Als zu bewertende Aussage ist gefragt, ob sich die beliebige Position innerhalb der Fläche befindet, d.h. der Abstand D_F ein positives Vorzeichen aufweist (Abb. 2.17b):

$$A_F = \text{Position innerhalb Fläche} = D_F \geq 0.
 \tag{2-15}$$

Die Randlinie gehört dabei zu der Fläche, die sie begrenzt. Es ergibt sich (Abb. 2.17c):

$$p(A_F) = p(D_F \geq 0) = 1 - F_D(0),
 \tag{2-16}$$

mit F_D als Verteilungsfunktion des Abstandes. (2-16) liefert eine Wahrscheinlichkeit von $p=0.5$ für Positionen exakt auf dem Rand. Wandert die Position ins Innere der Fläche, erhält man größere Werte (innerhalb: $p > 0.5$). Entsprechend nehmen die Werte ab, wenn man sich vom Objekt entfernt (außerhalb: $p < 0.5$). Theoretisch sind die Werte $p = 1$ bzw. $p = 0$ nur an unendlich entfernten Positionen innerhalb bzw. außerhalb des Objektes anzutreffen.

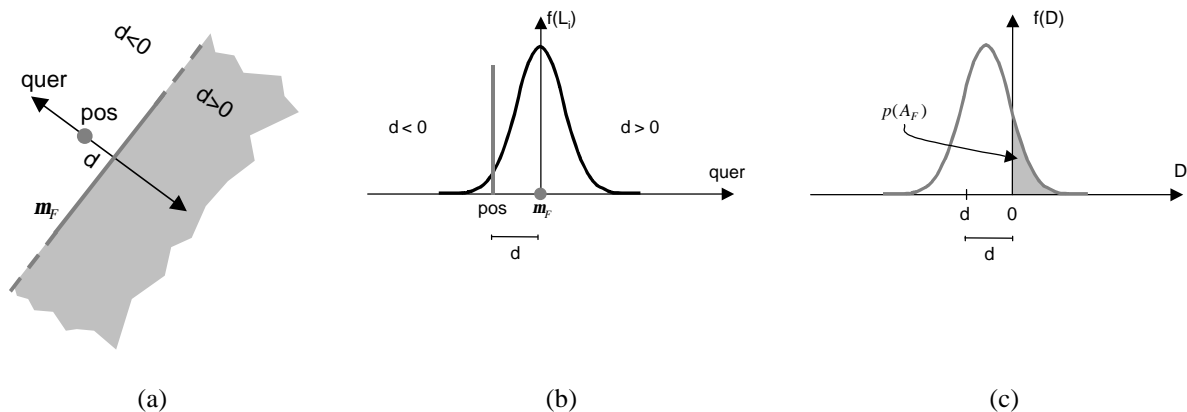


Abb. 2.17: Bestimmung der Wahrscheinlichkeit für die Flächenzugehörigkeit einer beliebigen Position mit Darstellung der Lagesituation (a), einem Querschnitt durch die Dichtefunktionen der Grenzlinie L_i (b) und durch die Dichtefunktion der Distanz D mit Angabe der gesuchten Wahrscheinlichkeit $p(A_F)$ (c).

Wahrscheinlichkeit für die Zugehörigkeit zu einer Linie

Die Anwendung der Bewertungsformel auf Linien führt zum Problem, dass Linien nur in eine Richtung (nämlich in Richtung der Linie) eine Ausdehnung besitzen. Die Aussage, ob eine beliebige Position innerhalb einer Linie liegt, ist folglich nicht sinnvoll zu formulieren. Beachtet man aber, dass sich hinter den meisten linienhaft abgebildeten Objekten zwar langgestreckte, aber dennoch ausgedehnte Phänomene der realen Welt stehen (z.B. Straßen, Leitungen, Schienenstrecken, Flussläufe), dann braucht der für Flächen gefundene Berechnungsansatz nur eine entsprechende Anpassung erfahren. Dazu ist für jede Linie die mittlere Ausdehnung b abzuschätzen. Sie definiert ein Band, das sich um die Linie legt und durch eine Pufferbildung mit Distanz $b/2$ erzeugt werden kann (Abb. 2.18a). Die beiden Ränder des Bandes (\mathbf{a}, \mathbf{b}) begrenzen die flächenhafte Ausdehnung der Linie und können mit den Randlinien einer Fläche verglichen werden. Ihre Varianzen \mathbf{s}_a^2 und \mathbf{s}_b^2 entsprechen der Linienvarianz \mathbf{s}_L^2 :

$$s_a^2 = s_b^2 = s_L^2. \quad (2-17)$$

Dabei wird standardmäßig die Breite b als eine fehlerfrei vorliegende Konstante angenommen. Anhang B.2 zeigt die entsprechenden Ableitungen, wenn die Breite b ebenfalls als eine stochastische Größe behandelt wird. Zu beachten ist, dass die beiden Ränder vollständig miteinander korreliert sind (Korrelationskoeffizient $r = 1$).

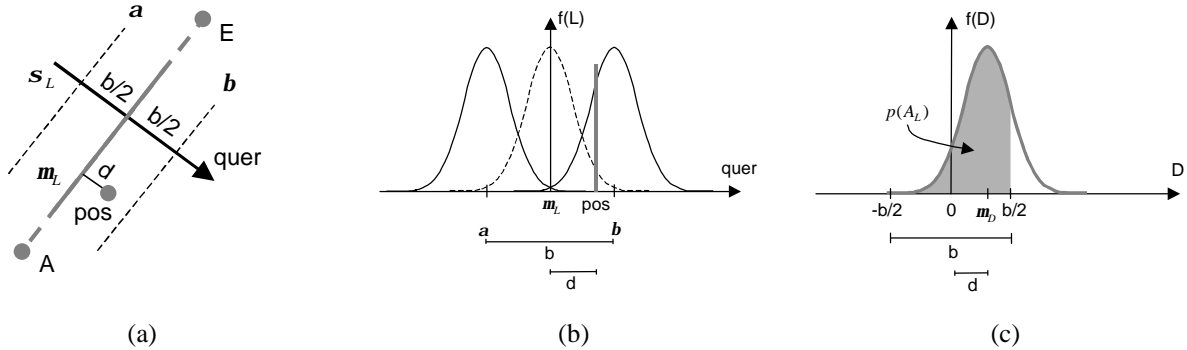


Abb. 2.18: Bestimmung der Wahrscheinlichkeit für die Linienzugehörigkeit einer beliebigen Position mit Darstellung der Lagesituation (a), einem Querschnitt durch die Dichtefunktionen der Zufallsvariablen L , a und b (Linie und Ränder) (b) und einen Querschnitt durch die Dichtefunktion der Distanz D mit Angabe der gesuchten Wahrscheinlichkeit $p(A_L)$ (c).

Eine beliebige Position pos gehört genau dann zur Linie, wenn sie sich innerhalb des Bandes befindet. Auf die beiden Ränder übertragen, müssen demnach die folgenden Aussagen gleichzeitig erfüllt sein (Abb. 2.18b):

$$A_L = A_{L_a} \wedge A_{L_b}, \quad \text{mit} \quad \begin{array}{l} A_{L_a} = \text{Position rechts vom linken Rand} = pos > a \\ A_{L_b} = \text{Position links vom rechten Rand} = pos < b. \end{array} \quad (2-18)$$

Aufgrund der Korrelation der Ränder ergibt sich die Bewertung des gleichzeitigen Eintreffens der beiden Ereignisse über den allgemeinen Multiplikationssatz und lautet:

$$p(A_L) = p(A_{L_a} \wedge A_{L_b}) = p(A_{L_a}) \cdot p(A_{L_b} | A_{L_a}). \quad (2-19)$$

Die Wahrscheinlichkeiten für die beiden Ränder und für die Bedingung berechnen sich aus:

$$p(A_{L_a}) = p(a \leq pos) = p(L \leq pos + b/2), \quad (2-20)$$

$$p(A_{L_b}) = p(b \geq pos) = p(L \geq pos - b/2), \quad (2-21)$$

$$p(A_{L_b} | A_{L_a}) = \frac{p(pos \leq b \leq pos + b)}{p(b \leq pos + b)} = \frac{p(pos - b/2 \leq L \leq pos + b/2)}{p(L \leq pos + b/2)}. \quad (2-22)$$

Aufgrund der vollständigen Korrelation der Ränder und der konstanten Breite gilt zudem:

$$p(b \leq pos + b) = p(a \leq pos). \quad (2-23)$$

Daraus resultiert die gesuchte Bewertung der Linienzugehörigkeit als:

$$p(A_L) = p(pos \leq b \leq pos + b) = p(pos - b/2 \leq L \leq pos + b/2). \quad (2-24)$$

Die Wahrscheinlichkeit wird demnach über ein kleines, um die beliebige Position zu bildendes Intervall bestimmt. Das Intervall bildet einen Ausschnitt der Gesamtfläche unter der Dichtefunktion f_L . Die Größe des Intervalls orientiert sich an der Breite b . Geht die Breite gegen Null, dann geht auch die Wahrscheinlichkeit gegen Null. Dies trifft z.B. für alle künstlichen Linien (z.B. Flurstücksgrenzen) zu. Um auch in diesen Fällen eine Wahrscheinlichkeit zu erhalten, wird für das Intervall eine Mindestgröße b_{min} (Mindestbreite) festgelegt. Liegt der wahre Breitenwert unter dieser Schranke, wird die Mindestbreite zur Berechnung verwendet. Für (2-24) ergibt sich damit:

$$p(A_L) = p(pos - b_{min}/2 \leq L \leq pos + b_{min}/2) \quad \text{für} \quad b \leq b_{min}. \quad (2-25)$$

Anstelle der bisher verwendeten absoluten Positionen kann die senkrechte Distanz D_L zur Linie als stochastische Variable eingesetzt werden (Abb. 2.18c). Für die Distanz ergeben sich hier stets positive Werte. (2-24) verändert sich zu:

$$p(A_L) = p(-b/2 \leq D_L \leq b/2) \quad (2-26)$$

oder in Funktionswerten der Verteilungsfunktion F_D ausgedrückt:

$$p(A_L) = F_D(b/2) - F_D(-b/2). \quad (2-27)$$

Wird die Breite b kleiner als die festgelegte Mindestbreite b_{min} , so wird die Mindestbreite in die Berechnung folgendermaßen eingesetzt:

$$p(A_L) = p(-b_{min}/2 \leq D_L \leq b_{min}/2) \quad \text{für } b \leq b_{min}. \quad (2-28)$$

Diese Festlegung definiert eine Minimalwahrscheinlichkeit für Positionen auf der mittleren Objektgeometrie ($d_L = 0$), die unabhängig von der Unsicherheit nicht unterschritten wird. Beispielsweise ergibt sich für $b_{min} = s_L/4$ eine Minimalwahrscheinlichkeit p_{min} von:

$$p_{min}(d_L = 0) = 0.10 \quad \text{für } b \leq b_{min} \text{ und } b_{min} = \frac{s_L}{4}. \quad (2-29)$$

Da ein linienhaftes Objekt im Allgemeinen eine endliche Länge aufweist, sind die Regionen um die Linienendpunkte (A und E) gesondert zu betrachten (Abb. 2.19a). Zusätzlich zur Querrichtung ist zu fordern, dass die Linie auch in Richtung von A nach B so weit ausgedehnt sein muss, damit die zu bewertende Position als Teil der Linie aufgefasst werden kann (Abb. 2.19b). Dies kann in folgender Aussage formalisiert werden:

$$A_{L_{AE}} = \begin{cases} \text{rechts von } A = A \leq pos \\ \text{links von } E = E \geq pos. \end{cases} \quad (2-30)$$

Die Bewertung dieser Aussage kann analog zur Flächenbewertung erfolgen. Es ist somit der Abstand D_{AE} zum jeweiligen Endpunkt in Laufrichtung der Linie zu ermitteln (Abb. 2.19c):

$$p(A_{L_{AE}}) = p(D_{AE} \geq 0). \quad (2-31)$$

Das Vorzeichen ist identisch zum flächenhaften Fall zu setzen, wobei innerhalb sich hier auf die senkrechte Projektion der Position auf die Linie bezieht. Die gesuchte Gesamtwahrscheinlichkeit ergibt sich dann durch Verknüpfung der beiden Einzelwahrscheinlichkeiten für die Zugehörigkeiten quer und längs der Linienrichtung:

$$p(A_{L_{gesamt}}) = p(A_{L_{AE}} \wedge A_L) = p(A_{L_{AE}}) \cdot p(A_L). \quad (2-32)$$

Dies soll anhand eines Beispiels verdeutlicht werden. Für eine Straße mit abgeschätzter Breite $b = 8 \text{ m}$ und einer Standardabweichung $s_L = 3 \text{ m}$ werden für zwei ausgewählte Positionen (pos_1, pos_2), die jeweils einen Abstand von $d = 5 \text{ m}$ zur Straßenmitte aufweisen, die Wahrscheinlichkeiten für die Zugehörigkeit bestimmt. Die erste Position wird so gewählt, dass Anfangs- und Endpunkt keinen Einfluss ausüben ($p(A_{L_{AE}}) = 1$). Die Wahrscheinlichkeit ergibt sich zu:

$$p(pos_1) = p(-4.0 \text{ m} \leq D_L \leq 4.0 \text{ m}) = 0.370. \quad (2-33)$$

Wählt man im zweiten Fall die Position so, dass sie im Abstand von $d_{AE} = 1 \text{ m}$ zu einem Endpunkt liegt, dann berechnet sich für die Wahrscheinlichkeit zu:

$$p(pos_2) = p(-4.0 \text{ m} \leq D_L \leq 4.0 \text{ m}) \cdot p(D_{AE} \geq 0 \text{ m}) = 0.370 \cdot 0.629 = 0.233. \quad (2-34)$$

Die geringe Entfernung zum Endpunkt bewirkt eine deutliche Verkleinerung der Wahrscheinlichkeit.

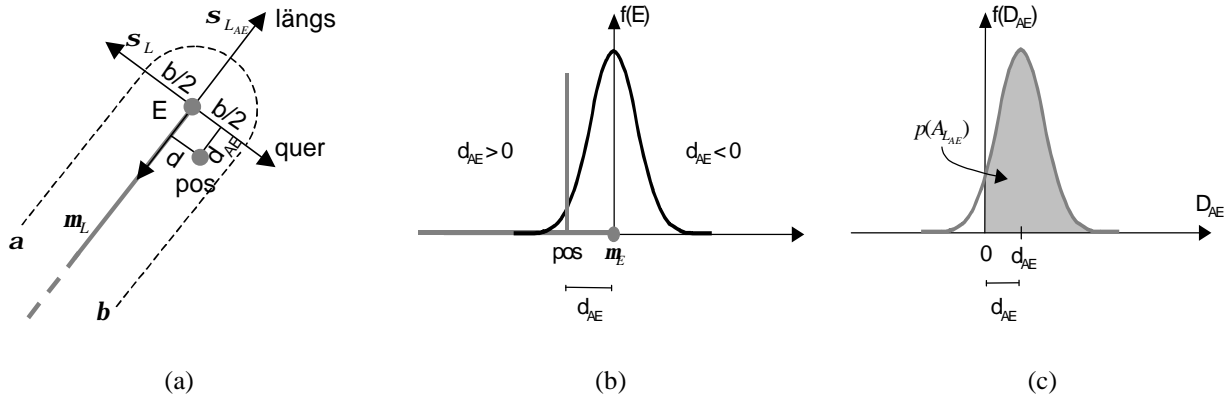


Abb. 2.19: Sonderfall der Linienzugehörigkeit für Positionen im Bereich von Anfangs- bzw. Endpunkt mit Darstellung der Lagesituation (a), einem Längsschnitt durch die Dichtefunktion der Zufallsvariable L (b) und durch die Dichtefunktion der Distanz zum Anfangs- bzw. Endpunkt mit Angabe der gesuchten Wahrscheinlichkeit $p(A_{L_{AE}})$ (c).

Wahrscheinlichkeit für die Zugehörigkeit zu einem Punkt

Bleibt zuletzt noch die Bewertung der Zugehörigkeit zu punktförmigen Objekten. Theoretisch besitzt ein Punkt keinerlei räumliche Ausdehnung. Doch gilt im Allgemeinen, dass punktförmig modellierte Phänomene in der Realität durchaus ausgedehnt sein können. Die Vorgehensweise zur Berechnung der Wahrscheinlichkeit kann daher analog zur Lösung bei linienförmigen Objekten erfolgen. Zu jedem Punkt ist die Breite b des Phänomens abzuschätzen. Während die Bewertung der Basisaussage im Linienfall im Allgemeinen nur von der Variation quer zur Linienrichtung beeinflusst wird, sind bei Punktobjekten zwei zueinander senkrechte Richtungen unabhängig voneinander zu betrachten. Prinzipiell kann man die Wahl der Richtungen beliebig gestalten. Zur Formalisierung soll jedoch eine bestimmte Anordnung vorgegeben werden. Die Richtung vom Punkt zur betrachteten Position wird als Querrichtung und die Richtung senkrecht dazu als Längsrichtung festgelegt (Abb. 2.20a). Die Aussagen für die Zugehörigkeit können dann für beide Richtungen jeweils identisch zum allgemeinen Linienfall ((2-24) bzw. (2-26)) formuliert werden. Sie lauten:

$$A_p = A_{p_{quer}} \wedge A_{p_{längs}} \quad \text{mit} \quad A_{p_{quer}} = A_{p_{längs}} = A_L. \quad (2-35)$$

Die einzelnen Aussagen sind unabhängig voneinander und müssen zur Bewertung der Gesamtaussage gleichzeitig auftreten. Damit ergibt sich für die gesuchte Wahrscheinlichkeit:

$$p(A_p) = p(A_{p_{quer}}) \cdot p(A_{p_{längs}}). \quad (2-36)$$

Zu beachten ist, dass die jeweils einzusetzende Distanz der Position vom Punktobjekt von der Richtung abhängt. Während für die Querrichtung unterschiedliche Werte (D_{quer}) vorkommen (Abb. 2.20b), erhält man für die Distanz in Längsrichtung ($D_{längs}$) stets den Wert Null. Die Distanz ist aber keine konstante Größe, sondern muss als Zufallsvariable behandelt werden (Abb. 2.20c). Damit ergeben sich die folgenden Formeln für die beiden Richtungen:

$$p(A_{p_{quer}}) = p(-b/2 \leq D_{quer} \leq b/2), \quad (2-37)$$

$$p(A_{p_{längs}}) = p(-b/2 \leq D_{längs} \leq b/2), \quad \text{mit} \quad d_{längs} = 0. \quad (2-38)$$

Damit berechnet sich die gesuchte Wahrscheinlichkeit einer beliebigen Position für die Zugehörigkeit zu einem Punkt zu:

$$p(A_p) = p(A_{p_{quer}}) \cdot p(A_{p_{längs}}) = p(-b/2 \leq D_{quer} \leq b/2) \cdot p(-b/2 \leq D_{längs} \leq b/2). \quad (2-39)$$

Bei kleinen Breiten b ist analog zur linienförmigen Lösung eine Mindestbreite b_{min} festzulegen.

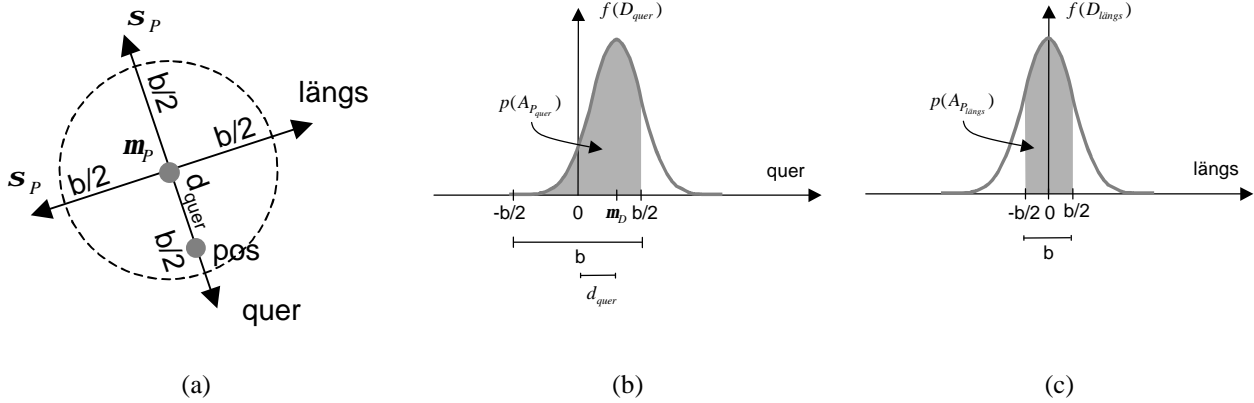


Abb. 2.20: Bestimmung der Punktzugehörigkeit für beliebige Positionen mit Darstellung der Lagesituation (a), einem Querschnitt durch die Dichtefunktion der Distanz D zur Position (b) und einen Querschnitt durch die Dichtefunktion in Längsrichtung (c) mit Angabe der gesuchten Wahrscheinlichkeiten $p(A_{p_{quer}})$ und $p(A_{p_{längs}})$.

Zur Verdeutlichung wird für eine Position die Zugehörigkeit zu zwei Objekten unterschiedlicher Ausdehnung bestimmt. Als Objekte werden ein Strommast mit einer räumlichen Ausdehnung von $b_1 = 1\text{ m}$ und ein Grenzpunkt mit der Breite $b_2 = 0\text{ m}$ ausgewählt. Die beiden Objekte besitzen jeweils eine Standardabweichung von $s_p = 0.5\text{ m}$. Für die Mindestbreite gilt der eingeführte Schrankenwert von $b_{min} = s_p / 4$. Dieser Wert findet beim Grenzpunkt Verwendung. Die betrachtete Position befindet sich im Abstand von $d=0.5\text{ m}$ von den Punkten. Die Wahrscheinlichkeiten für die Zugehörigkeit der Position errechnen sich in den beiden Fällen zu:

$$\begin{aligned} \text{Strommast : } p(\text{pos}) &= p(-0.5\text{ m} \leq D_{quer} \leq 0.5\text{ m}) \cdot p(-0.5\text{ m} \leq D_{längs} \leq 0.5\text{ m}) \\ &= 0.477 \cdot 0.683 = 0.326 \end{aligned}$$

$$\begin{aligned} \text{Grenzpunkt : } p(\text{pos}) &= p(-0.0625\text{ m} \leq D_{quer} \leq 0.0625\text{ m}) \cdot p(-0.0625\text{ m} \leq D_{längs} \leq 0.0625\text{ m}) \\ &= 0.06 \cdot 0.10 = 0.006. \end{aligned}$$

Aufgrund der größeren Ausdehnung erhält die Position bezüglich des Strommasts eine bedeutend höhere Wahrscheinlichkeit gegenüber der Position im Falle des Grenzpunktes.

2.4.2 Minimum-Maximum-Modell

Grundlage für die Modelldefinition bildet das Eingrenzen der Variation durch die Angabe eines Wertebereichs für die Objektgeometrie. Es wird davon ausgegangen, dass nur in einem eingeschränkten räumlichen Bereich geometrische Primitive eines Objektes aufzufinden sind. Diesen Bereich gilt es einzugrenzen. Dies soll in Form eines Intervalls erfolgen, in das alle möglichen Realisierungen der Objektgeometrie einzuordnen sind, d.h. außerhalb der Intervallgrenzen existieren keine weiteren Realisierungen. Als obere Intervallgrenze wird der maximal ausgedehnte räumliche Bereich definiert, in dem Realisierungen vorkommen können. Er soll als Maximum-Geometrie (kurz: Maximum) bezeichnet werden. Das Maximum stellt eine Art umschließende Geometrie dar, die alle Realisierungen des Objektes beinhaltet. Die untere Intervallgrenze beschreibt den Bereich minimaler Ausdehnung eines Objekts, der in allen Realisierungen als Teilmenge enthalten ist. Sie definiert die Minimum-Geometrie (kurz: Minimum). Für das Minimum gilt, dass die umschlossene Ausdehnung als sicher zum Objekt gehörend angenommen wird, während alle Bereiche außerhalb des Maximums sicher nicht zum Objekt zählen. Mit x als beliebige Realisierung, die der Grundgesamtheit M entstammt, hat eine formale Definition von Minimum und Maximum folgendes Aussehen:

$$\begin{aligned} \text{Maximum-Geometrie : } x &\subseteq \text{MAX}, \forall x \in M \\ \text{Minimum-Geometrie : } \text{MIN} &\subseteq x, \forall x \in M. \end{aligned} \tag{2-40}$$

Auf die Verhältnisse der realen Welt übertragen, können Maximum und Minimum als maximale und minimale Ausdehnung des Phänomens interpretiert werden.

Eine allgemeine Beschreibung der geometrischen Variation im Rahmen des Modells kann durch Angabe des Intervalls:

$$\text{Variation} : V = [MIN, MAX] = \{x / x \in M, MIN \leq x \leq MAX\} \quad (2-41)$$

erfolgen. Die Variation entspricht damit einer Intervallzahl, auf die sich die Beziehungen der Intervallarithmetik anwenden lassen (Grosche et al., 1995). Weiterführende theoretische Grundlagen des Minimum-Maximum-Modells können im Anhang B.3 eingesehen werden.

Zusätzlich zum Minimum und Maximum wird eine mittlere Ausdehnung des Objektes formuliert, die die mittlere Geometrie des Objektes (kurz: Mittel) veranschaulicht. Sie stellt eine ausgezeichnete Realisierung dar, die in der Gesamtbetrachtung aller Realisierungen eine mittlere Lage und eine mittlere Ausdehnung besitzt. Methoden zur Schätzung des Mittels sind bereitzustellen (Kapitel 3.2.1).

Die unsichere Geometrie eines Objektes im Rahmen des Modells ergibt sich schließlich in Form des folgenden 3-Tupels:

$$\text{unsichere Geometrie} = \{ \text{Minimum}, \text{Mittel}, \text{Maximum} \}. \quad (2-42)$$

Für ein einzelnes Objekt existieren stets drei geometrische Repräsentationen (Einzelgeometrien), die entsprechend ihrer Definition unterschiedliche Bedeutung besitzen. Die geometrische Unsicherheit ist in ausreichender Weise im Tupel integriert, da Minimum und Maximum einen Zusammenhang zur Intervalldefinition der geometrischen Variation herstellen. Aufgrund der verschiedenen räumlichen Ausdehnungen kann zwischen den drei Repräsentationen folgende Beziehung hergestellt werden:

$$\text{Minimum} \subseteq \text{Mittel} \subseteq \text{Maximum}, \quad (2-43)$$

d.h. das 3-Tupel beschreibt eine der Größe nach geordnete Abfolge. Maximum und Minimum sind nur in dem Fall identisch, wenn keine Unsicherheit besteht und die Geometrie als sicher aufzufassen ist. In Kombination mit der bekannten Abfolge

$$\text{Punkt} \subset \text{Linie} \subset \text{Fläche} \quad (2-44)$$

wird der Typ eingeschränkt, den eine Einzelgeometrie annehmen kann. Ist z.B. das Mittel eine Linie, so kann das Minimum nur ebenfalls eine Linie oder ein Punkt sein.

Zum Zwecke einer geordneten Verarbeitung der unsicheren Geometrie wird als Vorschrift eingeführt, dass jedes Objekt durch ein vollständiges 3-Tupel zu beschreiben ist. Alle drei Einzelgeometrien sind demnach durch Primitive zu repräsentieren. Besonders punkt- und linienförmige Objekte sind von dieser Regel betroffen. Das Problem besteht darin, dass im Allgemeinen keine Geometrie gefunden werden kann, die als Teilmenge in allen punkt- bzw. linienförmigen Realisierungen enthalten ist. Es existieren somit keine Minima, die der gegebenen Definition genügen. Um dennoch die Vorschrift zu erfüllen, müssen künstliche Repräsentanten festgelegt werden. Prinzipiell sind beliebige Festlegungen denkbar, solange die geordnete Abfolge der Einzelgeometrien erfüllt bleibt. Als Standard wird festgelegt, dass für linien- und punktförmige Objekte das Minimum dem Mittel identisch sein soll:

$$\text{Standard für Linie und Punkt} : \text{Minimum} \equiv \text{Mittel}. \quad (2-45)$$

In diesem Fall sind alle Punkte der mittleren Geometrie als sicher zum Objekt gehörend anzusehen. Auch für Flächen ist es denkbar, dass kein Minimum und/oder kein Mittel gefunden werden kann (Kapitel 3.5.4). Auch hier ist jeweils ein geometrischer Repräsentant zu definieren. In solch einem Fall kann standardmäßig der Flächenschwerpunkt des Maximums als Repräsentant benutzt werden. Abb. 2.21 illustriert einige Beispiele für unterschiedliche Objekttypen.

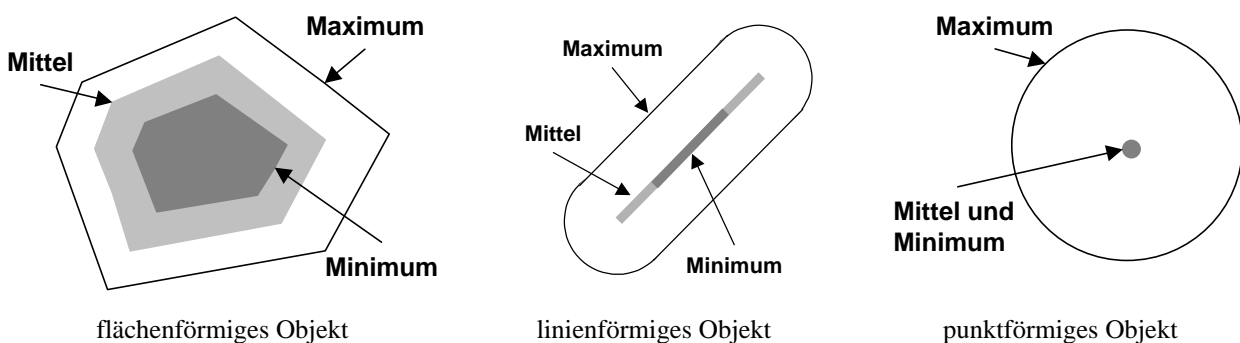


Abb. 2.21: Beispiele für die Definition von Minimum, Mittel und Maximum bei verschiedenen Objekttypen.

Im Fall der Linie weicht das Minimum von der Standardfestlegung ab, indem eine verkürzte mittlere Linie Verwendung findet. Dies ist mit der Definition konform, da das Minimum eine kleinere Ausdehnung als das Mittel aufweist.

Die Variationsbreite R als beschreibendes Maß für die Variation wird innerhalb des Ansatzes durch eine Region identifiziert, die sich zwischen Minimum und Maximum befindet. Sie lässt sich in einfacher Weise durch eine mengentheoretische Differenzbildung erzeugen, indem Maximum und Minimum jeweils als Punktmengen aufgefasst werden. Die Variationsbreite ergibt sich dann zu:

$$\text{Variationsbreite : } R = \text{MAX} - \text{MIN}. \tag{2-46}$$

Die Variationsbreite besitzt ebenfalls eine räumliche Ausdehnung, die durch geometrische Primitive beschrieben werden kann. Dieser Umstand erschwert einen Vergleich der Variation verschiedener Objekte erheblich. Deshalb soll zusätzlich ein numerisches Maß festgelegt werden, das den Vergleich vereinfacht. Ein solches Maß ist in einfacher Weise über die Abstände der Einzelgeometrien abzuleiten. Allgemein wird der Abstand als kürzeste Distanz eines Randpunktes der einen Geometrie vom Rand der anderen Geometrie definiert. Bei punkt- oder linienförmigen Geometrien tritt an die Stelle des Randes jeweils die Geometriebeschreibung des Objektes. Ausgehend vom Mittel sind die Abstände zum Minimum und zum Maximum zu bestimmen. Als Ergebnisse erhält man Intervalle für die Extremwerte der Distanzen:

$$\begin{aligned} \text{Abstände Mittel - Minimum : } d_{\text{Mittel,MIN}} &= [\min d_{\text{Mittel,MIN}}, \max d_{\text{Mittel,MIN}}] \\ \text{Abstände Mittel - Maximum : } d_{\text{Mittel,MAX}} &= [\min d_{\text{Mittel,MAX}}, \max d_{\text{Mittel,MAX}}]. \end{aligned} \tag{2-47}$$

Es ist sinnvoll, die Intervallangaben auf einen einzigen Wert zu reduzieren. Da das Mittel stets zwischen den Einzelgeometrien liegt, kann von annähernd gleichen Intervallgrenzen ausgegangen werden. Um die Variation in jedem Fall einzugrenzen, sind die oberen Grenzen der Intervalle als ausgezeichnete Distanzen wiederzuverwenden. In Anlehnung an die Epsilon-Band-Methode können diese als Epsilon-Distanzen (e_{\min}, e_{\max}) bezeichnet werden. Sie ergeben sich folgendermaßen:

$$\begin{aligned} \text{Größte Distanz Mittel - Minimum : } e_{\text{MIN}} &= \max d_{\text{Mittel,MIN}} \\ \text{Größte Distanz Mittel - Maximum : } e_{\text{MAX}} &= \max d_{\text{Mittel,MAX}}. \end{aligned} \tag{2-48}$$

Ein geeignetes Maß für die Variation definiert sich letztlich durch das Maximum der beiden Epsilonwerte zu:

$$\text{Variationsmaß : } e = \max(e_{\text{MIN}}, e_{\text{MAX}}). \tag{2-49}$$

Es entspricht dem maximalen Distanzwert, der zwischen Mittel und Minimum oder Maximum auftreten kann. In Abb. 2.22 ist die Bestimmung des Variationsmaßes beispielhaft für die unterschiedlichen Objekttypen angedeutet.

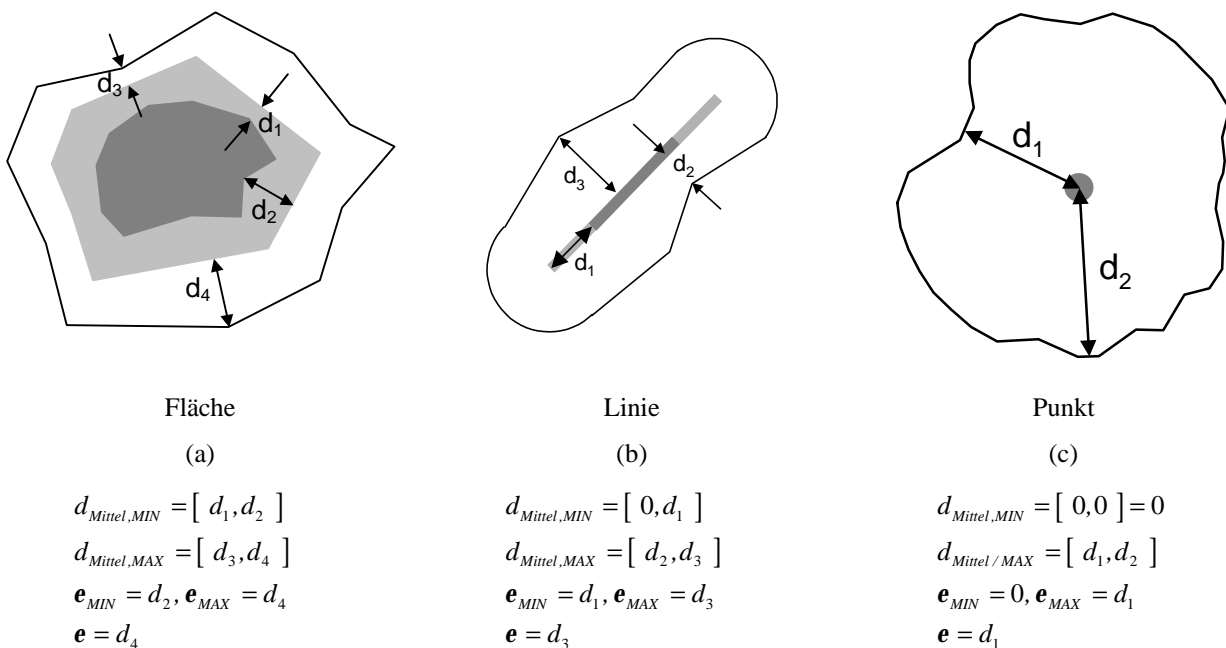


Abb. 2.22: Bestimmung des Variationsmaßes e aus den Abständen zwischen Minimum und Maximum für die Objekttypen Fläche (a), Linie (b) und Punkt (c).

Ist der Epsilonwert als übergeordnetes Fachwissen a-priori bekannt, so genügt es, die mittlere Geometrie zu bestimmen, um automatisch durch zwei Pufferbildungen mit Radius ϵ (Puffer nach außen) und Radius $-\epsilon$ (Puffer nach innen) das Maximum und das Minimum erzeugen zu können. Diese Möglichkeit ist im Besonderen für bereits bestehende Datensätze interessant, die in der Regel nur über eine Geometrie pro Objekt verfügen. Diese besondere Art der Einzelgeometrie soll als traditionelle Geometrie (kurz: Normalgeometrie) bezeichnet werden. Fasst man die Normalgeometrie als Schätzung der mittleren Geometrie auf, so lassen sich mit dem bekannten Wert ϵ Minimum und Maximum erzeugen.

Die unsichere Geometrie nimmt genau drei ausgezeichnete Zustände (Minimum, Mittel und Maximum) ein. Aufgrund dieser diskreten Anzahl kann die Bewertung von Aussagen nur in qualitativer Weise erfolgen. Entsprechend der drei Zustände lassen sich vier Werte zuweisen und so ein diskretes Möglichkeitsmaß m definieren:

$$m \in \{ \text{sicher}, \text{möglicherweise}, \text{möglicherweise nicht}, \text{sicher nicht} \}.$$

Die Werte sind in Form natürlichsprachlicher Ausdrücke formuliert, so dass sie sich direkt den Aussagen anfügen lassen (z.B. Ein Industriegebiet grenzt *möglicherweise* an ein Biotop). Kodierungen in numerischer Form sind denkbar (z.B. sicher = 3, *möglicherweise* = 2, *möglicherweise nicht* = 1, sicher nicht = 0).

Ist nun die geometrische Basisaussage (Position innerhalb Objekt) zu bewerten, dann ergibt sich durch Anwendung des Möglichkeitsmaßes eine Erweiterung der Aussagemenge E zu:

$$E = \{ \text{sicher innerhalb}, \text{möglicherweise innerhalb}, \text{möglicherweise nicht innerhalb}, \text{sicher nicht innerhalb} \}.$$

Zur Zuweisung eines Möglichkeitswertes ist nun nicht nur die Position gegenüber der Normalgeometrie zu überprüfen, sondern zusätzlich auch gegenüber Maximum und Minimum. Anschließend kann der in Abb. 2.23 aufgestellte Bewertungsbaum genutzt werden, um die letztendlich gültige Aussage zu finden. Das Schema zeigt, dass stets nur zwei Prüfungsschritte notwendig sind. Die Überprüfung gegenüber dem Mittel entscheidet, ob nachfolgend Maximum oder Minimum relevant werden. Einige Beispiel der Bewertung innerhalb eines flächenhaften Objektes zeigt Abb. 2.24.

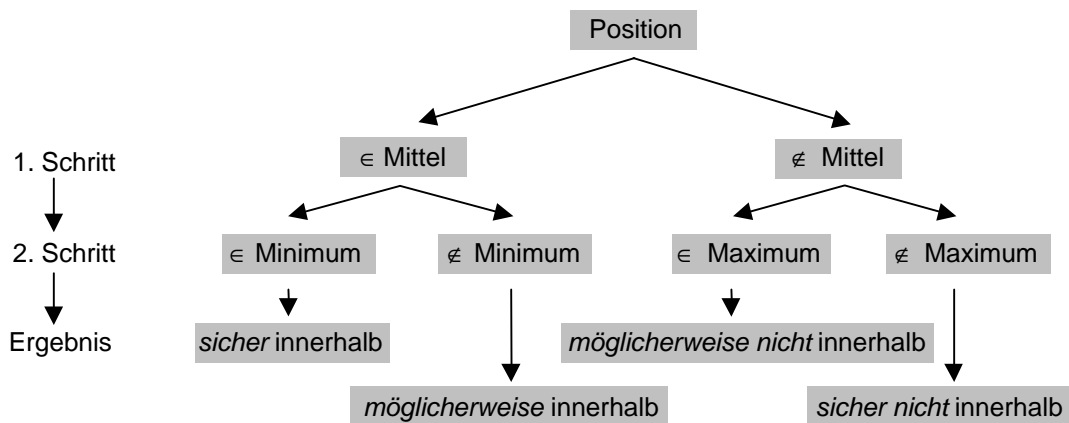


Abb. 2.23: Bewertungsbaum für die geometrische Grundaussage „Position innerhalb Objekt“.

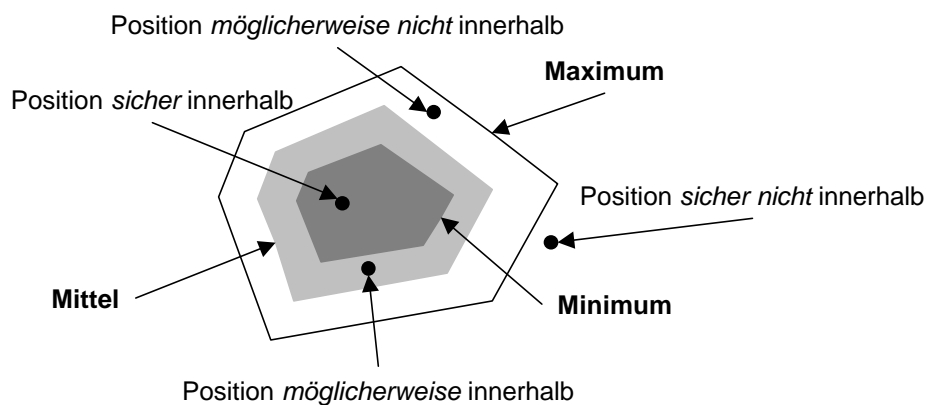


Abb. 2.24: Bewertung der geometrischen Grundaussage am Beispiel eines flächenhaften Objekts.

2.4.3 Fuzzy-Modell

Räumliche Phänomene sind in aller Regel individuelle, komplexe Gebilde. Sie enthalten interne Inhomogenitäten und kleinräumige Variationen, so dass vollständige Beschreibungen durch räumliche und nicht-räumliche Eigenschaften nur angenähert gelingen. Die Phänomene wirken unscharf, was zur Unsicherheit auf Datenebene führt. Ziel ist es, eine verbesserte Objektrepräsentation zu erzeugen, die die Unschärfe der Phänomene mit enthält. Eine solche Betrachtungsweise bildet den Ansatzpunkt für die Anwendung der unscharfen Mengentheorie (Fuzzy-Subset-Theorie) (Zadeh, 1965). Die Theorie soll hier zur Beschreibung der geometrischen Unsicherheit eingesetzt werden. Dabei gilt die Annahme, dass die Unsicherheit hauptsächlich von der Unschärfe in der räumlichen Ausdehnung der Phänomene verursacht wird. Jedoch ist unabhängig von möglichen Ursachen ein allgemeingültiges Modell zu definieren, das auf alle Objekte gleichermaßen angewendet werden kann. Basisbausteine der unscharfen Mengentheorie sind dem Anhang B.4 zu entnehmen.

Grundlegendes Konzept des Fuzzy-Modells ist die Formulierung der Objektgeometrie als unscharfe Punktmenge. Punkte selbst sind scharfe Elemente des Raumes und können durch Positionen $(x, y) \in \mathfrak{R}^2$ charakterisiert werden. Die unscharfe Objektgeometrie im Fuzzy-Modell (Fuzzy-Geometrie) lässt sich folgendermaßen beschreiben:

$$\text{Fuzzy-Geometrie: } \underline{G} = \left\{ \left((x, y) / \underline{m}_{\underline{G}}(x, y) \right) \right\}, \quad (x, y) \in \mathfrak{R}^2, \quad \underline{m}_{\underline{G}}(x, y) \in [0, 1], \quad (2-50)$$

mit $\underline{m}_{\underline{G}}(x, y)$ als Zugehörigkeitsfunktion, die jedem Punkt einen Zugehörigkeitswert zuweist. Er gibt an, zu welchem Grad der Punkt zur Objektgeometrie zu zählen ist. Mögliche Zugehörigkeitswerte entstammen dem Intervall $[0, 1]$, wobei ein Wert nahe Eins eine hohe und ein Wert nahe Null eine niedrige Zugehörigkeit bedeutet. Interpretieren lässt er sich als Grad der Übereinstimmung zwischen den Eigenschaften an der Position und des beobachteten Objektes. Die Zuweisung des Wertes erfolgt über eine festzulegende Zugehörigkeitsfunktion, deren Form von der enthaltenen Variation abhängt. Im Allgemeinen sind für alle Punkte im Raum individuelle Zugehörigkeitswerte möglich.

Es existieren zwei unterschiedliche Ansätze, eine Zugehörigkeitsfunktion festzulegen. Liegt zu jedem Objekt eine große Anzahl an Referenzdaten (z.B. mehrere Realisierungen) vor, dann kann darauf aufbauend die Festlegung der Funktion erfolgen. Als Alternative bietet sich die Möglichkeit, die Zugehörigkeitsfunktion vom Anwender interaktiv festlegen zu lassen. Hierbei wird vorausgesetzt, dass der Anwender über ausreichend Expertenwissen über die Unschärfe der Phänomene verfügt. Der erste Ansatz hat den Vorteil, individuelle Lösungen zu erzeugen, ist aber deutlich aufwendiger zu handhaben als der Zweite. Der Zweite ist weniger flexibel und weniger exakt, dafür aber schneller auszuführen. Die Lösung, die in dieser Arbeit entwickelt wird, stellt eine einfache Kombination von beiden Ansätzen dar. Basis bildet eine Standardfunktion, deren Parameter eine variable Ausgestaltung ermöglichen. Die Parameter werden mit Hilfe von erfassten Referenzdaten objektindividuell geschätzt. Als Standardfunktionen kommen eine Vielzahl an Funktionen in Betracht. Es soll hier aber von einer konstanten Änderung der Zugehörigkeit ausgegangen werden, so dass sich eine lineare Standardfunktion ergibt.

Allgemein hängt die Zugehörigkeitsfunktion von den beiden Koordinaten (x, y) ab. Zur Vereinfachung der Definition ist eine Verringerung der Dimensionalität sinnvoll. Dies kann dadurch erreicht werden, dass an die Stelle der Koordinaten der Abstand $d(x, y)$ zur mittleren Geometrie (kurz: Mittel) tritt. Der Abstand definiert sich als minimale euklidische Distanz des Punktes zum Lotfußpunkt auf der mittleren Geometrie. Dadurch entsteht eine feste Bindung zwischen der Zugehörigkeitsfunktion und der mittleren Geometrie. Zur Bestimmung des Mittels sind entsprechende Methoden bereitzustellen (Kapitel 3.2.1).

Modelldefinition für eine Fläche

Die drei Geometrietypen (Fläche, Linie, Punkt) erfordern geringfügig unterschiedliche Definitionen der Standardfunktionen. Zuerst erfolgt die Festlegung für eine Fläche. Allgemein lässt sich eine unscharfe Fläche folgendermaßen beschreiben:

$$\text{unscharfe Flächengeometrie: } \underline{F} = \left\{ \left((x, y) / \underline{m}_{\underline{F}}(d(x, y)) \right) \right\}, \quad (2-51)$$

mit (x, y) als beliebiger Punkt und $d=d(x, y)$ als Abstand zur mittleren Geometrie \bar{F} . Gemäß den Einflüssen wirkt sich die geometrische Variation nur in den Randbereichen einer Fläche aus, so dass dort die lineare Änderung der Zugehörigkeit zu modellieren ist. Es kann die folgende lineare Flächenzugehörigkeitsfunktion formuliert werden:

$$m_f(d) = \begin{cases} 0 & \text{für } d < -a, \\ \frac{d+a}{2a} & \text{für } -a \leq d \leq a, \ a \neq 0, \\ 1 & \text{für } a < d \text{ und } d = a = 0, \end{cases} \quad (2-52)$$

mit a als zu wählender Parameter.

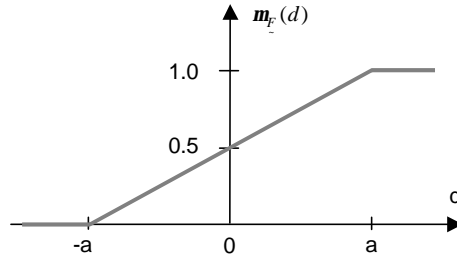


Abb. 2.25: Lineare Zugehörigkeitsfunktion für eine Fläche.

Abb. 2.25 zeigt die graphische Darstellung der Funktion. Das Vorzeichen des Abstandes d wird innerhalb der Fläche positiv und außerhalb negativ gewählt. Für den Abstand $d=0$ ergibt mit Ausnahme von $a=0$ immer die konstante Zugehörigkeit $m_f(0)=0.5$, so dass die mittlere Randlinie der Fläche eine Isolinie gleicher Zugehörigkeit bedeutet. Die mittlere Geometrie erhält man durch Bildung eines Alpha-Schnitts (Anhang B.4) mit $\alpha = 0.5$:

$$\text{mittlere Flächengeometrie : } \bar{F} = F_{\alpha=0.5} \quad (2-53)$$

Sie stellt im Sinne der Modelltheorie eine scharfe Menge dar.

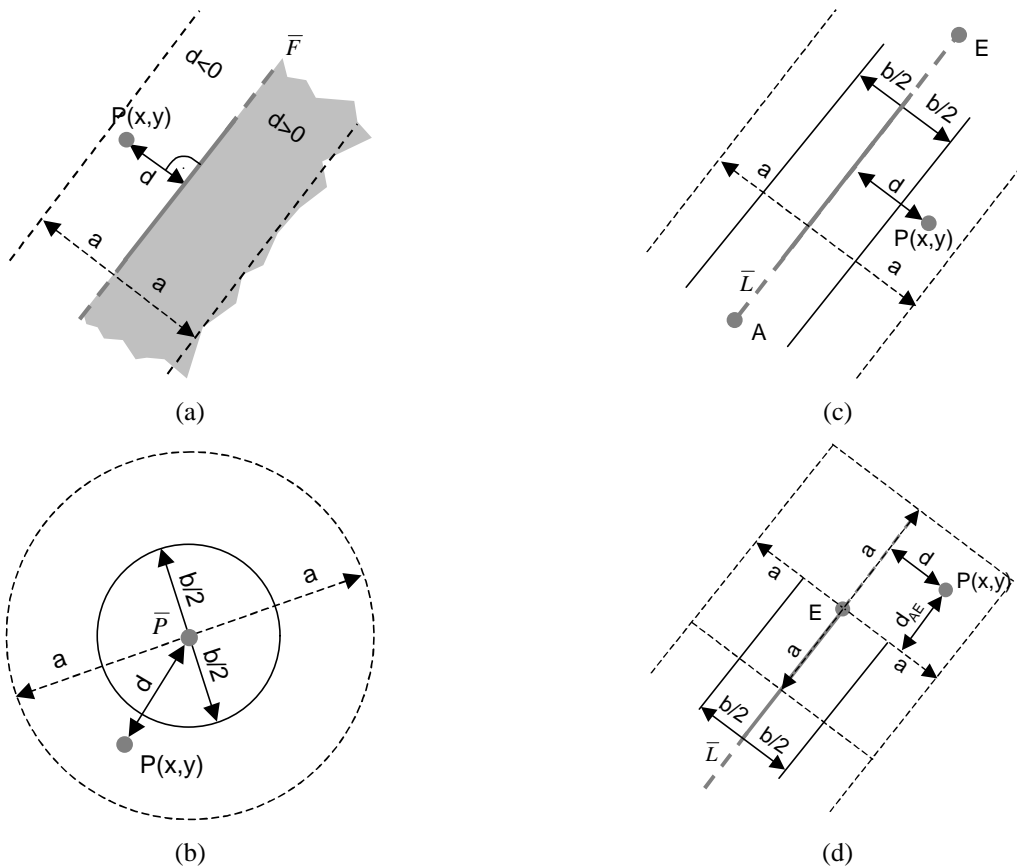


Abb. 2.26: Darstellung der Lagesituation für die Bestimmung der Zugehörigkeit eines beliebigen Punktes zur unscharfen Geometrie im Falle einer Fläche (a), eines Punktes (b) und einer Linie (c), (d). Für die Linie wird zusätzlich unterschieden, ob der Punkt innerhalb der Linie (c) oder im Bereich eines Randpunktes (d) liegt.

Der Parameter a dient der individuellen Ausgestaltung der Standardfunktion, die dadurch flexibel an die jeweilige Situation eines Objektes angepasst werden kann. Er verdeutlicht einerseits den maximalen Abstand eines noch zum Objekt zu zählenden Punktes von der mittleren Geometrie und andererseits den minimalen Abstand der Punkte, die zum Kern der unscharfen Menge gehören (Abb. 2.26a). Im Extremfall $a=0$ liefert die Funktion eine scharfe Zugehörigkeit. In diesem Fall wird die Geometrie zu einer gewöhnlichen scharfen Menge. Aufgrund dieser Eigenschaft eignet sich der Parameter als anschauliches Maß, um die Unschärfe der Flächengeometrie wiederzugeben. Unterschiede in der Größe der Variation innerhalb eines Objektes können dadurch ausgedrückt werden, dass man die Randlinien in einzelne Segmente aufteilt und jedem Segment einen eigenen Parameterwert zuweist.

Die Berechnung der Zugehörigkeit soll durch ein einfaches Beispiel verdeutlicht werden: Ein Punkt befindet sich im Abstand von $d = -5 m$ außerhalb der mittleren Objektgeometrie (z.B. für ein Waldobjekt). Die maximale Entfernung, in der Punkte noch zum Objekt zählen, beträgt $a=15 m$. Die Zugehörigkeit des Punktes zum Objekt bestimmt sich dann zu:

$$m_{\underline{z}}(d = -5 m) = \frac{-5 m + 15 m}{30 m} = 0.33. \tag{2-54}$$

Modelldefinition für eine Linie

Betrachtet man eine Linie als eine unscharfe Menge, so kann ihre Definition folgendermaßen gegeben werden:

$$\text{unscharfe Liniengeometrie: } \underline{L} = \left\{ (x, y) \mid m_{\underline{z}}(d(x, y)) \right\}. \tag{2-55}$$

Die Zugehörigkeit ist wiederum vom Abstand des Punktes zur mittleren Geometrie abhängig. Geht man davon aus, dass die durch Linien modellierten Phänomene trotzdem eine Ausdehnung quer zur Linienrichtung besitzen, dann ist zu deren Berücksichtigung zusätzlich die Breite b des Phänomens abzuschätzen. Die Breite soll als feste Größe angesehen werden, was bedeutet, dass eine sichere Minimalbreite einzusetzen ist. Variiert die Breite beispielsweise im Intervall $[5 m, 10 m]$, so ist der untere Schrankenwert (hier: $5 m$) als Minimalbreite auszuwählen. Der obere Schrankenwert findet im Parameter a seine Berücksichtigung. Grundsätzlich wird die Kenntnis der Objektbreite vom Fuzzy-Modell nicht zwingend vorausgesetzt. Ist kein Wert bekannt oder erscheint die Bestimmung zu aufwendig, sollte näherungsweise von der Breite $b=0$ ausgegangen werden. Die Linienzugehörigkeitsfunktion besitzt folgendes Aussehen:

$$m_{\underline{z}}(d) = \begin{cases} 0 & \text{für } a < d, \\ \frac{a-d}{a-b/2} & \text{für } \frac{b}{2} \leq d \leq a, a \neq 0, b \neq 0, a \neq \frac{b}{2}, \\ 1 & d < \frac{b}{2} \text{ und } d = a = b = 0 \text{ und } a = d = \frac{b}{2}, \end{cases} \tag{2-56}$$

mit a und b als zu wählende Parameter der Standardfunktion.

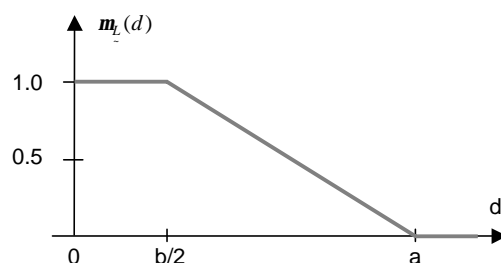


Abb. 2.27: Zugehörigkeitsfunktion für eine Linie

Eine graphische Visualisierung der Funktion ist in Abb. 2.27 zu sehen. Punkte, die direkt auf der mittleren Linie liegen ($d=0$), besitzen immer die volle Zugehörigkeit und gehören damit zum Kern der unscharfen Geometrie (Abb. 2.26c). Als weitere Verallgemeinerung der Standardfunktion wäre es denkbar, den maximalen Funktionswert der Zugehörigkeitsfunktion variabel zu halten. Dies könnte für solche Phänomene sinnvoll sein, für die nicht sicher festzulegen ist, ob die Eigenschaften überhaupt auf die Existenz eines Objektes hindeuten.

Die Modelldefinition soll wiederum durch ein Beispiel verdeutlicht werden: Für einen Punkt mit Entfernung $d=5\text{ m}$ von der mittleren Linie (z.B. Straße) ist die Zugehörigkeit zum linienförmigen Objekt zu berechnen. Die unscharfe Linie hat eine maximale Ausdehnung von $a=13\text{ m}$ und eine Minimalbreite von $b=8\text{ m}$. Die Zugehörigkeitsfunktion ergibt den folgenden Wert:

$$\mathbf{m}_L(d=5\text{ m}) = \frac{13\text{ m} - 5\text{ m}}{13\text{ m} - 4\text{ m}} = 0.89. \quad (2-57)$$

Wurde bisher nur die Zugehörigkeit quer zur Linienrichtung beurteilt, so ist an den Rändern, definiert durch Start- und Endpunkt der Linie, zusätzlich die Längsrichtung zu beachten. Die Längszugehörigkeit kann in diesen Bereichen analog zum flächenhaften Fall bestimmt werden. Insbesondere lässt sich direkt die Flächenzugehörigkeitsfunktion nutzen. Es gilt somit:

$$\mathbf{m}_{L_{\text{längs}}}(d_{AE}) = \mathbf{m}_F(d_{AE}). \quad (2-58)$$

Die zur Bestimmung notwendige Distanz d_{AE} definiert sich als Entfernung des Lotfußpunktes zum Randpunkt entlang der mittleren Linie, wobei der Lotfußpunkt das Ergebnis der senkrechten Projektion des betreffenden Punktes auf die mittlere Linie darstellt (Abb. 2.26d). Dazu ist die mittlere Linie über den Randpunkt fortgesetzt zu denken. Die Distanz ist positiv, wenn der Lotfußpunkt zwischen den Randpunkten liegt und ansonsten negativ. Da für die Längszugehörigkeit die Objektbreite keine Rolle spielt, ist der für die Flächenzugehörigkeitsfunktion notwendige Parameter a' wie folgt zu berechnen:

$$a' = a - \frac{b}{2}. \quad (2-59)$$

Längs- und Querszugehörigkeit sind in den Randbereichen zu kombinieren. Es handelt sich dabei um eine Konjunktion, so dass der Minimumoperator anzuwenden ist. Die Gesamtzugehörigkeit sieht schließlich wie folgt aus:

$$\mathbf{m}_{L_{AE}}(d_{AE}, d) = \min(\mathbf{m}_F(d_{AE}), \mathbf{m}_L(d)). \quad (2-60)$$

Die Zugehörigkeit der Randpunkte (Start- und Endpunkt der Linie) hat den konstanten Wert $\mathbf{m}_{L_{AE}} = 0.5$ und nimmt linear zur Mitte zwischen beiden Randpunkten zu. Der Kern der unscharfen Linie ist kürzer als die mittlere Linie, da Bereiche um die Ränder eine geringere Zugehörigkeit aufweisen. Die Erweiterung wird ebenfalls am bereits bekannten Beispiel verdeutlicht: Der Punkt, dessen Zugehörigkeit gesucht ist, befindet sich im Abstand $d=5\text{ m}$ von der mittleren Linie ($a=13\text{ m}$, $b=8\text{ m}$) und liegt jedoch jetzt im Bereich eines der Randpunkte. Der Abstand vom Randpunkt beträgt $d_{AE} = -5\text{ m}$. Damit ergibt sich die Zugehörigkeit zu:

$$\mathbf{m}_{L_{AE}}(d_{AE}, d) = \min(\mathbf{m}_F(-5\text{ m}) = 0.31, \mathbf{m}_L(5\text{ m}) = 0.89) = 0.31. \quad (2-61)$$

Die Gesamtzugehörigkeit wird in diesem Fall von der Flächenzugehörigkeit dominiert.

Modelldefinition für einen Punkt

Zuletzt erfolgt die Definition für ein punktförmiges Objekt. Ein Punkt als unscharfe Menge hat folgendes Aussehen:

$$\text{unscharfe Punktgeometrie: } P = \left\{ (x, y) \mid \mathbf{m}_P(d(x, y)) \right\}. \quad (2-62)$$

Die Zugehörigkeit orientiert sich an der Distanz d eines beliebigen Punktes zum mittleren Punktobjekt \bar{P} . Wie schon im Falle einer Linie wird allgemein angenommen, dass das Objekt eine gewisse Ausdehnung in der Realität aufweist, die in Form der Breite b abzuschätzen ist. Die bei der Linie bezüglich der Breite gegebenen Erläuterungen besitzen auch hier ihre Gültigkeit. Damit kann die Punktzugehörigkeitsfunktion entsprechend der Linienzugehörigkeitsfunktion definiert werden. Es gilt:

$$\mathbf{m}_P(d) = \mathbf{m}_L(d). \quad (2-63)$$

Der mittlere Objektzentrum hat unabhängig von der festgelegten Breite immer den Zugehörigkeitswert $\mathbf{m}_P = 1$ und zählt damit zum Kern des Objektes. Mit wachsender Entfernung nimmt die Zugehörigkeit der Punkte in linearer Weise ab (Abb. 2.26b).

Allgemein wird die Variation durch die Zugehörigkeitsfunktion repräsentiert. Aus ihr ist die Definition eines Variationsmaßes abzuleiten. Auf den ersten Blick scheinen die Parameter der Standardfunktionen (a, b) dafür geeignet. Ein Vergleich der Variation unterschiedlicher Geometrietypen ist jedoch schwierig, da bei Punkten und Linien beide Parameter benötigt werden, bei Flächen hingegen ein Parameter ausreicht. Ein besseres Maß stellt der Fuzzy-Index (Anhang B.4) dar. Er veranschaulicht den Grad der Unschärfe des Objektes, indem die Distanz der unscharfen Menge zur nächstgelegenen scharfen Menge \underline{G} bestimmt wird. Die nächstgelegene scharfe Menge (Abb. 2.28) gilt als Alpha-Schnitt mit $\alpha = 0.5$:

$$\text{nächstgelegene scharfe Menge : } \underline{G} = G_{\alpha=0.5}. \tag{2-64}$$

Es wird der Hamming-Fuzzy-Index (Anhang B.4) genutzt, der auf der Hamming-Distanz (lineare Distanz) basiert:

$$\text{Fuzzy-Index : } \mathbf{u}_G = \mathbf{u}(\underline{G}) = \int_{-\infty}^{\infty} |m_G(x) - m_{\underline{G}}(x)| dx. \tag{2-65}$$

In die Indexberechnung gehen die Zugehörigkeitsfunktionen von unscharfer und nächstliegend scharfer Menge ein. (2-65) verzichtet auf die übliche Normierung des Wertes, um einen Vergleich zwischen unterschiedlichen Geometrietypen zu ermöglichen. Der Fuzzy-Index nimmt damit Werte im Intervall $[0, \infty)$ an. Ein hoher Wert deutet auf große Unschärfe, ein niedriger Wert auf kleine Unschärfe hin. Für eine Fläche ist die nächstgelegene scharfe Menge durch die mittlere Geometrie gegeben (Abb. 2.28a). Der daraus abzuleitende Fuzzy-Index einer Fläche lautet:

$$\text{Flächenindex : } \mathbf{u}_F = \frac{a}{2}. \tag{2-66}$$

Aufgrund der identischen Zugehörigkeitsfunktionen für unscharfe Punkte und Linien (Abb. 2.28b) sind Linien- und Punktindex in gleicher Weise nach folgender Formel zu berechnen:

$$\text{Punkt- /Linienindex : } \mathbf{u}_P = \mathbf{u}_L = \frac{1}{4} \left(a - \frac{b}{2} \right). \tag{2-67}$$

Für Breiten $b=0$ m sind Flächen-, Linien- und Punktindex äquivalent.

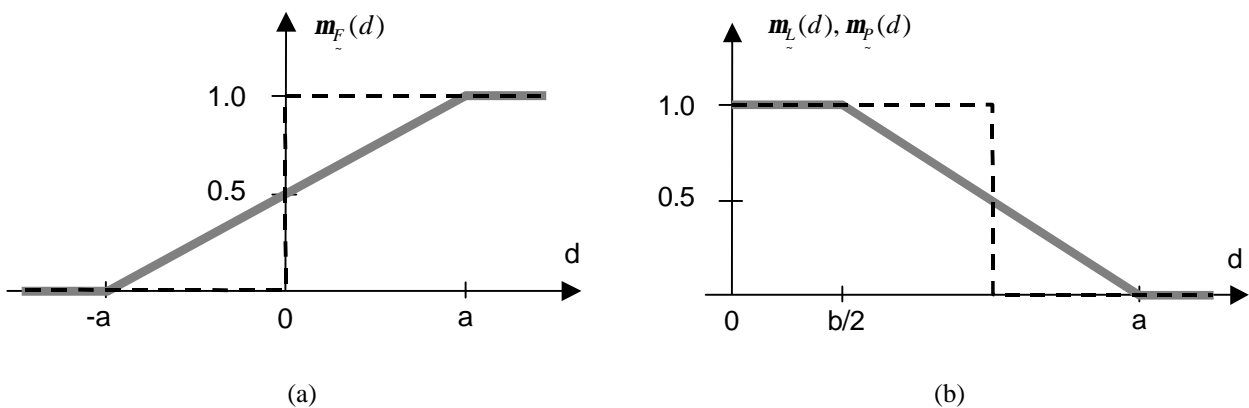


Abb. 2.28: Vergleich der Zugehörigkeitsfunktionen der unscharfen Mengen Fläche (a), Linie und Punkt (b) mit ihren nächstgelegenen scharfen Mengen.

Um abschließend eine formale Beschreibung der geometrischen Variation zu geben, bieten sich zwei Möglichkeiten. Die Erste benutzt direkt die Zugehörigkeitsfunktion, die vollständig die Variation repräsentiert. Die Beschreibung besitzt folgendes Aussehen:

$$\text{Variation : } V_{P,L,F} = \{ m_{P,L,F} \}. \tag{2-68}$$

Durch Verwendung des Fuzzy-Index ergibt sich eine alternative Darstellung. Um den räumlichen Bezug herzustellen, muss hierfür zusätzlich die mittlere Geometrie \bar{G} integriert werden, so dass sich die Variation folgendermaßen beschreiben lässt:

$$\text{Variation : } V_{P,L,F} = \{ \bar{G}, \mathbf{u}_{P,L,F} \}. \tag{2-69}$$

Im Fuzzy-Modell werden Aussagen durch quantitative Möglichkeitswerte bewertet. Möglichkeitswerte können als Zugehörigkeiten interpretiert werden, insbesondere dann, wenn man eine Aussage als eine Zuweisung zu einem unscharfen Zustand ansieht. Der Zustand definiert eine unscharfe Menge, und die Aussage ist die Zugehörigkeit eines Elementes zu dieser Menge. Damit wird der Grad beurteilt, inwieweit für das Element der Zustand zutrifft. Beispielsweise ist die Aussage „Objekt A ist lang“ zu bewerten. Die Zustandsbeschreibung „lang“ stellt einen unscharfen Begriff dar und kann durch eine unscharfe Menge repräsentiert werden. Die Zugehörigkeit des bestimmten Objektes A zur unscharfen Menge „lang“ liefert die gesuchte Möglichkeit der Aussage. Dazu ist die Zugehörigkeitsfunktion auszuwerten.

Die Bewertung der geometrischen Grundaussage „Position (x,y) innerhalb der Objektgeometrie“ fällt im vorliegenden Fuzzy-Modell besonders leicht, da die Lösung implizit im Ansatz enthalten ist. Noch deutlicher wird der Umstand, wenn man die Grundaussagen in die gleichbedeutende Aussage „Punkt $(x,y) \in Geometrie$ “ umformuliert. Die Geometrie ist im Fuzzy-Modell bereits als unscharfe Punktmenge G beschrieben, deren Elemente $P(x,y)$ ein bestimmter Zugehörigkeitswert beigelegt ist. Der gesuchte Möglichkeitswert p der Grundaussage entspricht damit dem Zugehörigkeitswert des betreffenden Punktes zur Geometrie. Formal kann dieser Zusammenhang folgendermaßen beschrieben werden:

$$\text{Möglichkeitsmaß} : p((x,y) \text{ innerhalb Geometrie}) = p(P(x,y) \in G) = m_G((x,y)). \quad (2-70)$$

Die allgemeine Bewertungsformel ist durch Einsetzen der Zugehörigkeitsfunktionen entsprechend dem vorliegenden Geometrietyp zu verfeinern.

2.4.4 Diskussion

In den letzten Abschnitten wurden drei unterschiedliche Modelle definiert, die alle in der Lage sind, die geometrische Unsicherheit zu beschreiben. Sie weisen ganz verschiedene Eigenschaften auf, die im Folgenden näher untersucht werden sollen. Zunächst wird dabei jedes Modell für sich betrachtet, um schließlich die markanten Eigenschaften der Modelle zusammenfassend gegenüberzustellen. Am Ende steht die Bereitstellung von Transformationen, die es ermöglichen, einen Wechsel der Modelle vollziehen zu können.

Das Stochastische Modell lässt deutliche Ähnlichkeiten zu einer Reihe von bestehenden Ansätzen zur Unsicherheitsbeschreibung erkennen, ohne dass konkret ein einzelner Ansatz als Basis Verwendung findet. Vielmehr kann die Modelldefinition als eine Vereinigung und Erweiterung mehrere Ansätze angesehen werden. Aufgrund der theoretischen Vorgehensweise kann ein Zusammenhang zum Kovarianzansatz hergestellt werden. Die darin enthaltene punktweise Modellierung wird jedoch durch eine linienweise Betrachtung ersetzt. Der wichtigste Vorteil des Modells liegt in einer vereinfachten Unsicherheitsbeschreibung. Man braucht lediglich die Varianz quer zur Linienrichtung angeben und nur sie wird auch für Weiterverarbeitungszwecke benötigt. Zusätzlich gilt für alle Linienpunkte der gleiche Wert der Varianz. Der vereinfachten Beschreibung erwachsen aber auch verschiedene Nachteile. So ist in manchen Fällen eine Linie in mehrere Segmente zu unterteilen. Konkret ist dies immer dann notwendig, wenn sich die Varianz im Verlauf der Linie ändert. Daraufhin ist jedes Segment unabhängig als eigene Linie mit eigener Varianz weiterzubehandeln. Die Behandlung von diskreten, künstlichen Objekten erweist sich als weiterer Nachteil. Sie sind im linienweisen Ansatz nur genähert zu beschreiben, während der punktweise Ansatz auf solche Objekte besonders gut anzuwenden ist. Die Varianzen differieren aber nur geringfügig (maximales Verhältnis der Standardabweichungen: $s_{L_{Linie}} : s_{L_{Punkt}} = \sqrt{2} : 1$), wobei die Varianz im linienweisen Ansatz stets etwas größer ausfällt, so dass sich der Nachteil unkritisch gegenüber einer weiteren Verarbeitung verhält. Problematisch ist weiterhin die Behandlung von Punktobjekten. Deren Unsicherheit wird anstatt der sonst üblichen Komponenten (s_x^2, s_y^2) in Richtung der Koordinatenachsen lediglich durch eine einzige Varianz repräsentiert. Eine Bestimmung eines einzelnen Wertes (s_p^2) unter angemessener Berücksichtigung der beiden Komponenten kann z.B. über den bekannten Punktfehler nach Helmert (z.B. Höpcke, 1980) erfolgen.

In der Modellbildung ist auch der diskutierte Wahrscheinlichkeitsansatz wiederzufinden. Neben der Anpassung auf die linienweise Betrachtung wird der Ansatz vor allem im Umfang erweitert. Konzentrierte man sich seither auf Flächen als Objekttyp, so ist die Zugehörigkeit jetzt auch für Linien und Punkte zu bewerten. Die Modelldefinition selbst beinhaltet bisher keine Festlegungen zur graphischen Visualisierung der Unsicherheit, so dass Bezüge zu den graphischen Ansätzen nur schwer herzustellen sind. Auf solche Festlegungen wird bewusst verzichtet, um unabhängig von möglichen Darstellungsformen zu bleiben. Dadurch wird eine Offenheit hergestellt, die die Anwendung einer ganzen Reihe an graphischen Präsentationen ermöglicht. Konkret werden die Darstellungsaspekte erst in Kapitel 3.3 behandelt. Dennoch kann auch ohne weitergehende Betrachtung festgehalten werden, dass Fehlerbänder eine mögliche graphische Präsentationsform darstellen.

Im Vergleich zu den bereits bekannten Ansätzen weist das Minimum-Maximum-Modell nur zur Epsilon-Band-Methode eine gewisse Verbindung auf. Beide beschreiben die Unsicherheit in geometrischer Form und setzen geometrische Primitive ein, um die Wirkung der Unsicherheit zu verdeutlichen. Für Linien ist das Epsilon-Band identisch mit dem Maximum, wenn für dessen Bildung das Variationsmaß e benutzt wird. Während sich das Epsilon-Band im Wesentlichen nur auf Linien bezieht und dabei nur die maximale Auswirkung beschreibt, wird durch das Minimum-Maximum-Modell eine verallgemeinerte Beschreibung gegeben. Hinzu kommen die weiteren Objekttypen (Punkt, Fläche), sowie die Beschreibung des Minimums als sichere Ausdehnung eines Objektes. Zusätzlich ist das Minimum-Maximum-Modell als flexibler einzustufen, da im Gegensatz zum Epsilon-Band keine expliziten Konstruktionsregeln vorgeschrieben werden. Die Bewertung der geometrischen Basisaussage wird in der im Minimum-Maximum-Modell eingesetzten Form bereits in Blakemore (1984) formuliert. Dort ist ein zusätzlicher Wert für den Fall eingeführt, dass die Position genau auf dem Rand der mittleren Geometrie zu liegen kommt. Die Aussage wird dann mit „nicht eindeutig - innerhalb/außerhalb“ bewertet. In logischer Konsequenz hätten ebenfalls eigene Aussagen für die Fälle, dass die Position sich auf dem Rand des Maximums oder des Minimums befindet, gebildet werden müssen. Verglichen mit der Anzahl möglicher Positionen kommt exakten Randlagen nur eine unbedeutende Rolle zu. Es wird hier auf eine Sonderbehandlung verzichtet, da eine spezielle Aufweitung des Aussageraumes nicht zu rechtfertigen ist. Stattdessen ist der Rand stets zum Inneren hinzuzählen (z.B. die Randlinie des Maximums gehört zur Maximum-Fläche selbst).

Das Fuzzy-Modell stellt eine ganz neue Möglichkeit der Beschreibung der geometrischen Unsicherheit dar, die bislang in dieser Weise noch nicht formuliert wurde. Das Unschärfe Mengenmodell kommt jedoch im Attributbereich bereits zur Anwendung. Betrachtet man die Geometrie als ein die räumliche Ausdehnung beschreibendes Attribut, dann lassen sich einzelne Querbezüge zu diesem Bereich herstellen. Im Falle von Attributen steht die begriffliche Unschärfe im Mittelpunkt (z.B. Temperatur ist hoch), während im geometrischen Fall die räumliche Unschärfe zu behandeln ist. Aufgrund der räumlichen Unschärfe entsteht Unsicherheit erst dadurch, dass üblicherweise eine scharfe Objektdefinition für nur unsharp zu beschreibende Phänomene zu geben ist. Wären unscharfe Objektdefinitionen von vornherein möglich, dann ließe sich die Unsicherheit um diesen Aspekt reduzieren. Im Allgemeinen sind alle natürlichen Objekte von der Problematik der Unschärfe betroffen, während künstliche Objekte zumeist einer scharfen Definition unterliegen. Wegen der Fixierung auf den Einfluss der räumlichen Unschärfe kann die Unsicherheit künstlicher Objekte nur aufgrund einer generalisierten Betrachtungsweise angemessen durch das Fuzzy-Modell repräsentiert werden. Erfassungseinflüsse sind dazu ebenfalls als eine Art Unschärfe (z.B. Unschärfe der Beobachtungsmethodik, Auflösung) zu charakterisieren, obwohl es sich streng gesehen um Messunsicherheiten handelt.

Zum Vergleich der einzelnen Modelle erfolgt eine Gegenüberstellung einiger markanter Eigenschaften. Die wichtigsten Modellfestlegungen sind in Tab. 2.2 zusammengefasst. Den Modellen liegen unterschiedliche Theorien zu Grunde, so dass die Repräsentation der Variation jeweils individuelle Formen annimmt. Ein Verknüpfungspunkt ist durch die mittlere Geometrie gegeben, die in allen drei Modellen eine wichtige Komponente darstellt. Einer der Gründe für ihre herausragende Stellung ist darin zu sehen, dass in den bereits bestehenden Datensätzen zumeist nur eine geometrische Repräsentation gespeichert ist, die zwangsläufig näherungsweise als mittlere Geometrie zu interpretieren ist. Darauf aufbauend ermöglichen alle Modelle, den Daten im nachhinein Unsicherheitsbeschreibungen hinzuzufügen, indem übergeordnetes Expertenwissen über die Unsicherheit integriert wird. Gemäß der Bedeutung der mittleren Geometrie sollte auf die Schätzung und die Definition eines Schätzverfahrens besonderes Augenmerk gelegt werden (Kapitel 3.2.1).

Tab. 2.2: Gegenüberstellung der wichtigsten Definitionen der Modelle

Eigenschaften	Stochastisches Modell	Minimum-Maximum-Modell	Fuzzy-Modell
Basismodell:	Stochastik: Mathematische Statistik und Wahrscheinlichkeitstheorie	Intervallarithmetik	Unschärfe Mengentheorie
Variation:	$V_{p,L,F} = \{ m_{p,L,F}, s_{p,L,F}^2 \}$	$V_{p,L,F} = [MIN, MAX]$	$V_{p,L,F} = \{ m_{p,L,F} \}$
Variationsmaß:	s Standardabweichung	e Maximaler Abstand zwischen Mittel und Maximum / Minimum	u Fuzzy-Index, Abstand zur nächstliegenden scharfen Menge
Aussagemäß:	p Wahrscheinlichkeit	m qualitatives Möglichkeitsmaß	p quantitatives Möglichkeitsmaß

Während die Variation im stochastischen Modell theoretisch eine unendlich große räumliche Ausdehnung des Objektes erzeugt, ist die Wirkung der Unsicherheit im Minimum-Maximum-Modell und im Fuzzy-Modell stets begrenzt. Die Maximum-Geometrie lässt sich daher mit dem Träger der unscharfen Menge $\text{supp}(A)$ (Anhang B.4), die Minimum-Geometrie mit dem Kern (Alpha-Schnitt $A_{\alpha=1}$) (Anhang B.4) vergleichen. Ein unmittelbarer Vergleich der Größenordnung der drei Maße ist aufgrund der verschiedenartigen Definition nicht möglich. Jedoch besitzen alle Werte eine metrische Einheit. Weiterhin können der Epsilon-Wert und die Standardabweichung als Abstände einer ausgezeichneten Geometrie zum Mittel interpretiert werden. In die Berechnung des Fuzzy-Indexes gehen die Parameter der Zugehörigkeitsfunktion ein, die ebenfalls Entfernungen zum Mittel angeben.

Im mathematischen Sinne handelt es sich bei Linien um unendlich dünne, bei Punkten um unendlich kleine Objekte, die in mindestens einer Richtung keine räumliche Ausdehnung besitzen. Sowohl das stochastische Modell als auch das Fuzzy-Modell sehen die Möglichkeit vor, solchen Objekten trotzdem einen Breitenwert zuzuweisen. Während die Zuweisung im Fuzzy-Modell optional geschieht, ist die Angabe der Breite im stochastischen Modell für die Bewertung der geometrischen Grundaussage zwingend notwendig. Für Phänomene, die in der realen Welt keine Ausdehnung besitzen (z.B. Verwaltungsgrenzen), muss sogar eine Mindestbreite eingesetzt werden. Im Minimum-Maximum-Modell spielt die Breite keine Rolle und bleibt unberücksichtigt.

Ein Blick auf die Aussagemasse zeigt, dass das Minimum-Maximum-Modell nur ein qualitatives Maß bereithält, während stochastisches Modell und Fuzzy-Modell quantitative Werte dafür vorsehen. Von quantitativen Werten kann ohne weiteres auf qualitative Werte übergegangen werden, jedoch nicht umgekehrt, so dass quantitative Werte als hochwertiger einzuschätzen sind. Wahrscheinlichkeiten zeichnen sich dadurch aus, dass eine Interpretation der Werte in Form von relativen Häufigkeiten einen einfacheren und allgemeinen Zugang zur Unsicherheitsbeschreibung ermöglicht. Eine ähnlich einfache Interpretation kann für Möglichkeitswerte nicht gegeben werden. Sie veranschaulichen im Allgemeinen die Zugehörigkeit zum entsprechend unscharfen Konzept. Rückt man von der traditionellen Interpretation der Wahrscheinlichkeiten ab und sieht diese in einem verallgemeinerten Zusammenhang als Maße der Glaubwürdigkeit einer Aussage an, dann können Aussagen in vergleichbarer Weise durch beide Ansätze bewertet werden (Cheeseman, 1986). In diesem Zusammenhang lassen sich Möglichkeitswerte mit bedingten Wahrscheinlichkeiten gleichsetzen. Der Wahrscheinlichkeitsansatz bietet dann zusätzlich den Vorteil, dass sich Abhängigkeiten von Einzelaussagen in beliebigem Grade berücksichtigen lassen, während Möglichkeitswerte den schlechtesten Fall einer vollen Abhängigkeit annehmen. Im Allgemeinen kann aber nicht davon ausgegangen werden, dass ein hoher Möglichkeitswert auch eine hohe Wahrscheinlichkeit bedeutet (Bhatnagar und Kanal, 1986). Es trifft sicherlich zu, dass unscharfe Konzepte (z.B. zwei Waldobjekte sind benachbart) angepasster mit Fuzzy-Methoden zu bearbeiten sind (Zadeh, 1986), dagegen Wahrscheinlichkeiten Vorteile im Bereich von Messunsicherheiten (z.B. Länge einer Straße) besitzen.

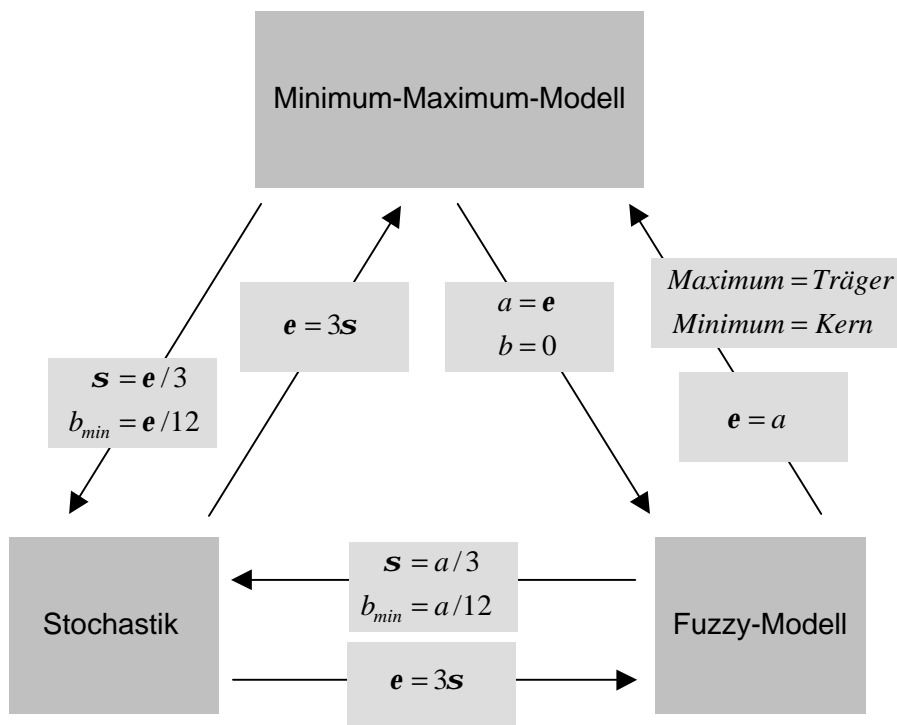


Abb. 2.29: Transformationen zwischen den Parametern der Unsicherheitsmodelle.

Der weitergehende Vergleich der Modelle ist dann erst sinnvoll anzustellen, wenn die Anwendungsaspekte der Modelle untersucht sind. Dies wird im Kapitel 3 erfolgen, so dass dort eine abschließende Bewertung der Modelle auch hinsichtlich ihrer Eignung für bestimmte Anwendungsfälle getroffen wird. Daher sei hier auf die Diskussion in Kapitel 4.1 verwiesen.

Die Diskussion zeigt, dass zwischen den Modellen erhebliche Unterschiede in der Interpretation der Variationsbeschreibung und der Maße bestehen. Trotzdem kann es für bestimmte Anwendungszwecke sinnvoll sein, einen formalen Wechsel des Unsicherheitsmodells durchführen zu können. Wird beispielsweise bei der Analyse ein unscharfes Anwendungsmodell (z.B. Suche alle „sehr weit entfernten“ Objekte) mit Hilfe der unscharfen Mengentheorie realisiert, dann ist es vorteilhaft, die Unsicherheit ebenfalls im Fuzzy-Modell zu beschreiben und so eine einfache Berücksichtigung zu ermöglichen. Durch einen solchen Wechsel lassen sich die Vorteile der einzelnen Modelle optimal ausnutzen, da je nach Anwendungszweck auf das beste Unsicherheitsmodell zurückgegriffen werden kann. Es entsteht eine Art hybride Unsicherheit, die eine gewisse Modellunabhängigkeit erlaubt. Einschränkend ist zu bemerken, dass stets den Sinn anwendungsabhängig zu prüfen ist.

Transformationen der Modellparameter bilden die Grundlage für den Wechsel. Sie sind für alle Kombinationen zu definieren. Eine wichtige Rolle spielt dabei die mittlere Geometrie, die in allen drei Modellen enthalten ist und somit unverändert übertragen werden kann. Sie dient als Bezugspunkt für die meisten Modellparameter, so dass sich eine Transformation entscheidend vereinfacht. Abb. 2.29 veranschaulicht die gefundenen Definitionen.

In erster Linie fließen die diskutierten Ähnlichkeiten der Modelle ein. Auf diesem Weg kann man durch die Bildung des Trägers und des Kerns vom Fuzzy-Modell zum Minimum-Maximum-Modell gelangen. Als Alternative bietet sich an, das Varianzmaß e dem Parameter a der linearen Zugehörigkeitsfunktion gleichzusetzen. Beide repräsentieren ein Maß für die maximale Ausdehnung des unsicheren Objektes. Aus e sind anschließend Minimum- und Maximum-Geometrie zu generieren. Der umgekehrte Weg ist durch die gleichen Beziehungen definiert. Da im Minimum-Maximum-Modell keine Breitenangaben Berücksichtigung finden, wird standardmäßig die Breite zu $b=0$ gesetzt. Vom stochastischen Modell zum Minimum-Maximum-Modell gelangt man, indem man die theoretisch unendliche Ausdehnung begrenzt. Allgemein werden mögliche Beziehungen zwischen Standardabweichung und Variationsbreite in Sachs (1978) diskutiert. Bekannt und häufig von der Praxis angewendet wird eine $3s$ -Toleranzgrenze, die aussagt, dass mit einer Wahrscheinlichkeit von 99,7% alle Realisierungen innerhalb einer Entfernung von $3s$ um das Mittel liegen. Dabei hat man von einem unendlich großen Stichprobenumfang auszugehen. Sowohl Hin- als auch Rücktransformation ist auf der Basis eines solchen Grenzwertes abgeleitet. Zusätzlich benötigt ein Wechsel zum stochastischen Modell die Angabe einer minimalen Breite, die ebenfalls auf den Grenzwert bezogen werden kann. In beiden Modellen ist die Breite b integriert, so dass im Normalfall dieser Parameter ohne Änderung übertragen werden kann. Da im Fuzzy-Modell die Angabe von b nur eine optionale Möglichkeit darstellt, ist eine Minimumbreite festzulegen. Auch hier ergibt sich aus der Verbindung der Teilschritte eine Lösung.

Vergleicht man die Ergebnisse aufgrund von hintereinander ausgeführten Hin- und Rücktransformationen, so kann festgestellt werden, dass nur Wechsel, die vom Maximum-Minimum-Modell ausgehen, verlustfrei ablaufen. Das Ergebnis bei den anderen Modellen hängt unmittelbar mit der Behandlung der Breite zusammen. Da das Minimum-Maximum-Modell keine Breitenwerte verarbeiten kann, entsteht immer ein Verlust, wenn zu diesem Modell gewechselt wird. Im Fuzzy-Modell wird die Breite nur als optionaler Parameter benutzt, so dass ein Verlust nicht zwangsläufig stattfinden muss. Zwischen Fuzzy-Modell und stochastischem Modell kann normalerweise verlustfrei gewechselt werden, da beide Breitenangaben integrieren. Zusammenfassend ist festzustellen, dass die Schwierigkeit des Wechsels in der Reihenfolge vom Minimum-Maximum-Modell über das Fuzzy-Modell zum stochastischen Modell zunimmt.

2.5 Integration der Unsicherheit in die Datenmodellierung

Nach der Definition der verschiedenen Unsicherheitsmodelle gilt es nun, die Unsicherheitsbeschreibung als einen neuen Bestandteil in die Datenmodellierung einzufügen. Das so erweiterte Datenmodell dient als Grundlage für Anwendungsaufgaben, die im nächsten Kapitel im Mittelpunkt stehen. Die Modellbeschreibung erfolgt mit Hilfe der „Unified Modeling Language“ (UML) (OMG, 2000). Es handelt sich dabei um eine objektorientierte Modellierungssprache, mit der Modelldefinitionen in graphischer Weise repräsentiert werden können. Objekt- und Klassenbildung, Vererbung, Aggregation, Einkapselung sind einige der objektorientierten Eigenschaften (Atkinson et al., 1989), die in UML anzutreffen sind. Die Notation ist implementationsunabhängig, so dass der Umsetzung unterschiedliche Wege offen stehen. Insbesondere zieht die objektorientierte Modellierung nicht notwendigerweise eine Implementation in einer objektorientierten Programmiersprache nach sich (Fritsch und Anders, 1996). Vielmehr können Umsetzungen in beliebigen Sprachen ausgeführt werden. Einige kommerzielle Programmpakete bieten dem Anwender bei der Modellbildung und der Umsetzung Unterstützung (z.B. Rational, 1998). Die wichtigsten Sprachkonstrukte werden im Anhang B.5 kurz erläutert.

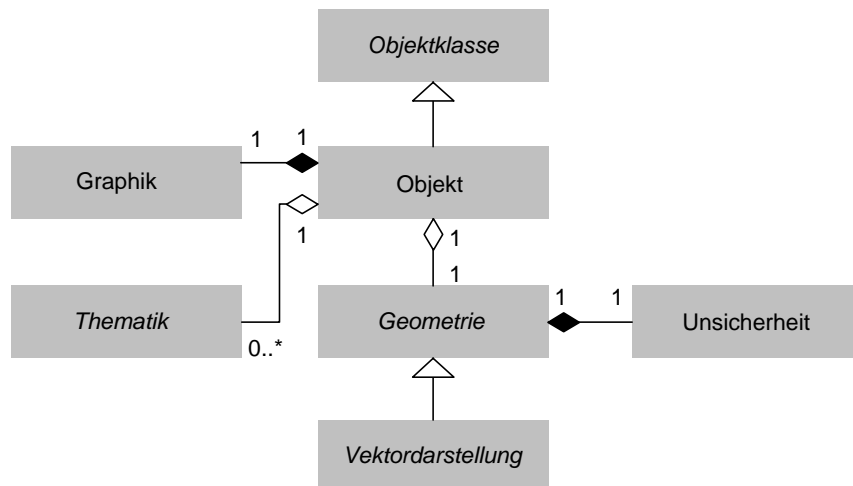


Abb. 2.30: Objektdatenmodell in UML-Notation.

Das erweiterte Datenmodell besitzt einen solchen Umfang, dass keine geschlossene Darstellung mehr möglich ist. Daher wird das Modell in kleinere Pakete aufgeteilt, die als einzelne Module aufzufassen sind. Die vorliegende Modellierung konzentriert sich im Schwerpunkt auf die Darstellung der Beziehungen zwischen den Objekten, so dass an dieser Stelle auf die Beschreibung von Methoden und Attributen verzichtet werden kann, wodurch sich die Übersichtlichkeit erheblich steigern lässt. Das erste Paket (Abb. 2.30) zeigt die allgemeine Erweiterung des Objektmodells um die Aspekte der geometrischen Unsicherheit. Ein Objekt gilt danach als Spezialisierung einer Objektklasse und besitzt eine Geometrie in Vektordarstellung und optional eine Reihe an Attributen (Thematik). Weiterhin gehört eine graphische Beschreibung dazu. Die Unsicherheitsbeschreibung wird als Bestandteil der Geometrie eingebunden, wobei die Geometrie über genau eine Unsicherheitsbeschreibung verfügt. Durch den Zusammenhang zwischen Objekt und Geometrie kann die Unsicherheit indirekt auch als Bestandteil des Objektes angesehen werden. Das Paket ist gegenüber eventuellen Erweiterungen in Richtung auf ein hybrides Datenmodell offen gestaltet, indem die Vektordarstellung noch um eine Rasterdarstellung ergänzt werden kann (Fritsch et al., 1998).

Die Vektordarstellung ist in einem weiteren Paket zusammengefasst (Abb. 2.31). Hinter der Vektordarstellung verbergen sich die verschiedenen Objekttypen als punktförmiges, linienförmiges und flächenförmiges Objekt. Die Objekttypen bauen sich aus den topologischen Primitiven Knoten, Kante und Masche auf, die ihrerseits aufgrund der verwendeten Randdarstellung aus den geometrischen Primitiven Punkt und Linie bestehen. Eine Besonderheit bilden topologische Beziehungen (z.B. innerhalb (inside), enthalten (contains), überlagern (intersect) usw. (Egenhofer und Herring, 1990)), die aufgrund ihrer Bedeutung für Analysezwecke explizit in die Modellierung aufgenommen wurden. Sie sind auf der Ebene der Objekttypen angesiedelt und in Form einer Assoziationsklasse formuliert, die auf die Beziehungen zwischen den Objekttypen einwirkt. Die Vektordarstellung umfasst bisher nur eine 2D-Darstellung, doch ist eine Erweiterung zur dritten Dimension konzeptionell berücksichtigt. Im Wesentlichen muss dazu eine weiterer Objekttyp, das Volumenobjekt, hinzugefügt werden.

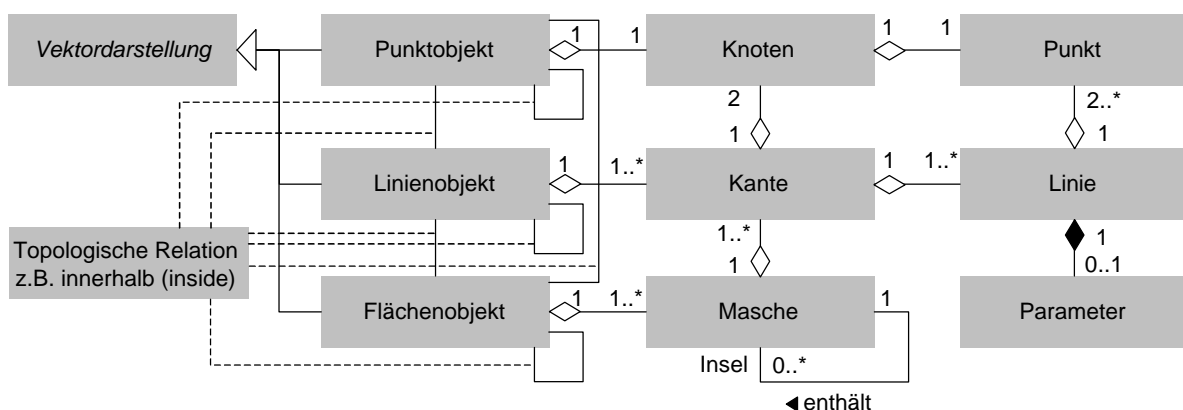


Abb. 2.31: Vektordarstellung in UML-Notation.

Die Unsicherheit als neuer Bestandteil im Datenmodell bildet ein eigenes abgeschlossenes Paket (Abb. 2.32). Dadurch wird deutlich, dass die Integration der Unsicherheit in kompakter, modularer Form erfolgt, die keine Änderung der inneren Struktur nach sich zieht, sondern das bestehende Modell an einer einzigen spezifizierten Stelle ergänzt. Kernstück der Unsicherheit bildet das Unsicherheitsmodell, für das als Spezialisierung eines der drei definierten Modelle (Stochastisches Modell, Minimum-Maximum-Modell, Fuzzy-Modell) in Frage kommt. Der herausragenden Stellung der mittleren Geometrie wird dadurch Rechnung getragen, dass sie einen weiteren wichtigen Bestandteil liefert. Entsprechend der eigentlichen Geometrie ist die mittlere Geometrie ebenfalls in einer Vektordarstellung zu geben. Die mittlere Geometrie kann mit der Objektgeometrie identisch sein. Da aber auch Unterschiede möglich sind, müssen beide Geometrien gesondert modelliert werden. Zuletzt ist noch eine graphische Darstellung vorzusehen, die die Unsicherheit in visueller Form nach außen präsentiert. Hier sind verschiedene Möglichkeiten denkbar, die im Detail in Kapitel 3.3 besprochen werden und daher unspezifiziert bleiben. Zwar sind nur die drei definierten Modelle genannt, doch grundsätzlich besteht die Möglichkeit, die Modellierung um beliebig andere Unsicherheitsmodelle zu ergänzen.

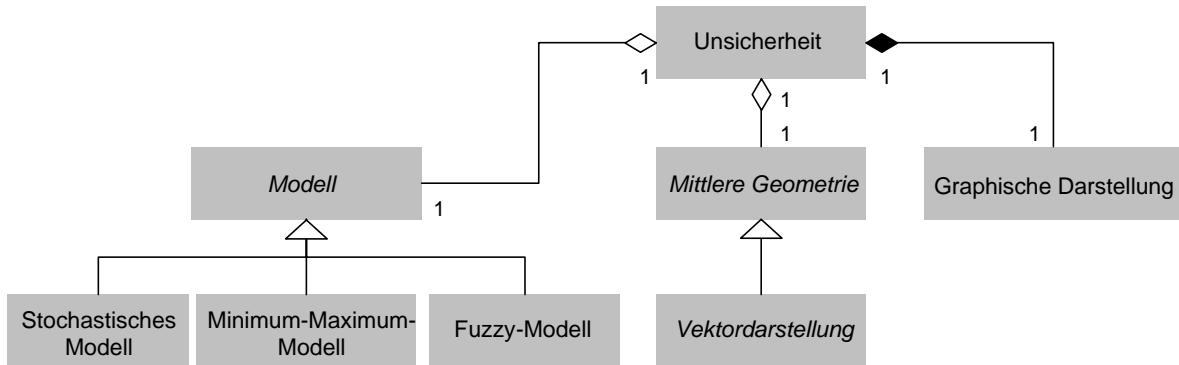


Abb. 2.32: Unsicherheitsmodell in UML-Notation.

Aussagemaß und Variationsmaß helfen, die Unsicherheit zu charakterisieren und bilden daher feste Bestandteile eines jeden Unsicherheitsmodells. Im stochastischen Modell (Abb. 2.33) werden dafür Wahrscheinlichkeiten bzw. eine Standardabweichung eingesetzt. Die Standardabweichung als Variationsmaß kann direkt aus den anderen Modellangaben ermittelt werden, so dass sie als abgeleitetes Objekt anzusehen ist. Die wesentliche Unsicherheitsbeschreibung wird durch mindestens einen Parametersatz gegeben. Mehrere Parametersätze sind deshalb vorzusehen, da sich die Unsicherheit innerhalb der Objektgeometrie ändern kann und so eine Unterteilung in Segmente notwendig wird. Jeder Parametersatz besteht aus einer individuellen Verteilungsfunktion, der Objektbreite und einer aus der Verteilungsfunktion und der Breite abzuleitenden Mindestbreite. Wie diskutiert, werden die Breitenangaben im Zusammenhang mit der Bestimmung von Wahrscheinlichkeiten von Aussagen, insbesondere der geometrischen Grundaussage benötigt. Als Verteilungsfunktion wird hier die zentrale Normalverteilung ($\sim N(0, \mathbf{s}^2)$) angenommen, die als einziger Parameter die zugehörige Varianz besitzt. Der Weg steht offen, weitere Verteilungen in das Modell einzubringen.

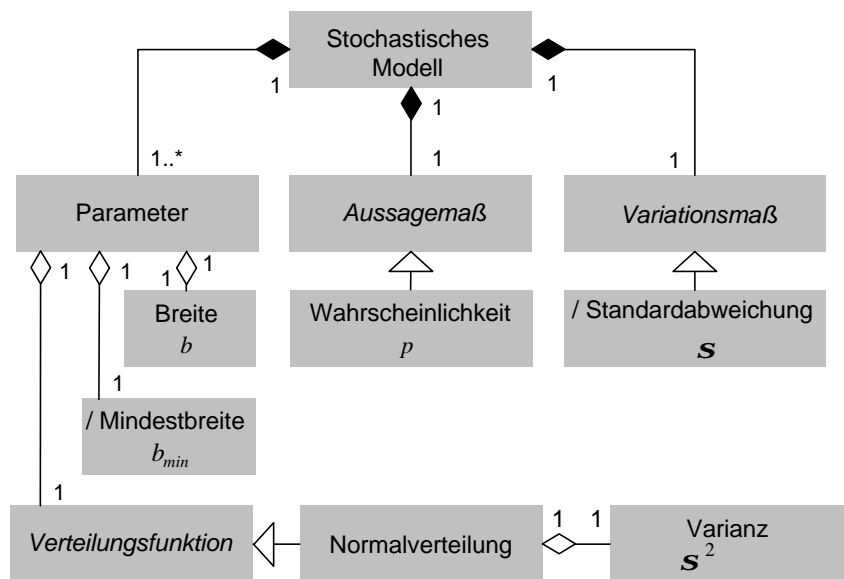


Abb. 2.33: Stochastisches Modell in UML-Notation.

Neben dem Aussagemaß, gegeben durch einen Möglichkeitswert, und dem Variationsmaß, das sich in Form eines Epsilonwertes ebenfalls aus anderen Angaben vollständig ableiten lässt, verfügt das Minimum-Maximum-Modell (Abb. 2.34) über zwei zusätzliche geometrische Beschreibungen, dem Minimum und dem Maximum. Sie sind analog zur eigentlichen Objektgeometrie in Vektordarstellung modelliert. Optional kann auch in diesem Modell die Variation durch Parametersätze mit Epsilonwerten beschrieben werden. Jedoch bleibt deren Angabe freigestellt, da bereits durch Minimum und Maximum eine vollständige Beschreibung existiert. Mehrere Parametersätze ermöglichen wiederum die Aufteilung in verschiedene Segmente. Prinzipiell dürfen auch hier beliebige Erweiterungen eingebracht werden, doch ist das Modell so allgemein und umfassend formuliert, dass keine offensichtlichen Stellen dafür zu identifizieren sind.

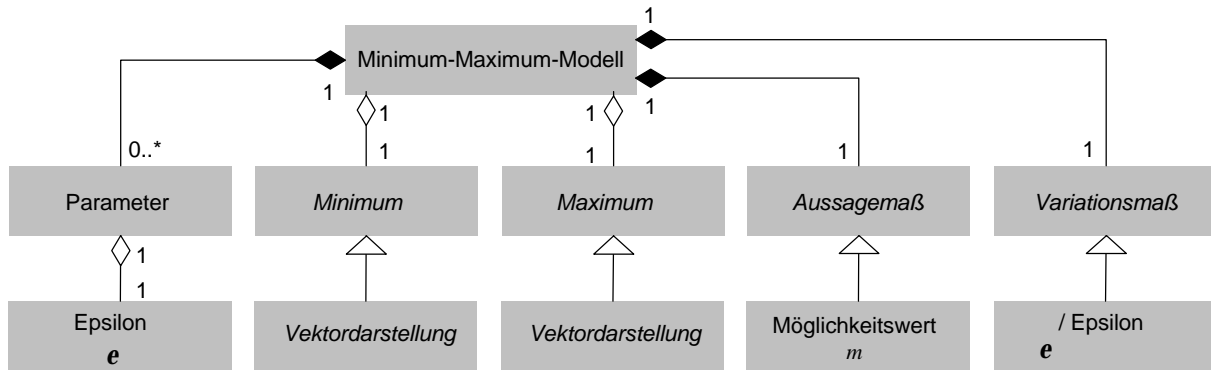


Abb. 2.34: Minimum-Maximum-Modell in UML-Notation.

Wie auch in den beiden anderen Modellen bilden Aussagemaß und Variationsmaß Bestandteile des Fuzzy-Modells (Abb. 2.35). Sie werden zum einen durch einen Möglichkeitswert und zum anderen durch den Fuzzy-Index repräsentiert. Weiterhin gehört mindestens eine Zugehörigkeitsfunktion zum Modell. Auch hier ermöglicht die Angabe mehrerer Funktionen die Unterteilung in Segmente. Die vorliegende Modelldefinition beschränkt sich bisher auf einen linearen Funktionsansatz als einzusetzende Zugehörigkeitsfunktion. Dazu gehören die beiden Parameter Ausdehnung und Breite, wobei die Breite optional anzugeben ist. Zu den zukünftigen Erweiterungen gehören insbesondere die Hinzunahme weiterer Funktionstypen der Zugehörigkeitsfunktion. Kaufmann (1975) gibt eine Übersicht, welche vielfältigen Möglichkeiten sich hier bieten.

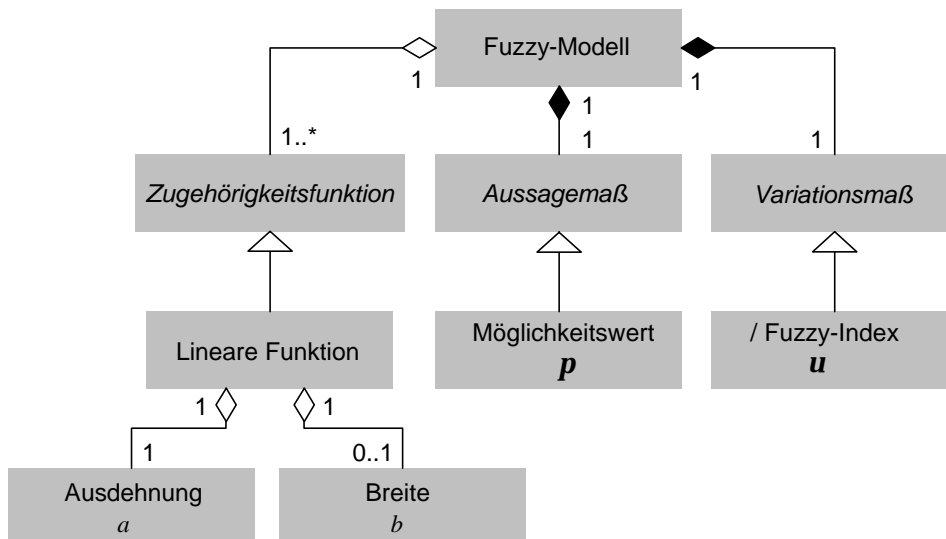


Abb. 2.35: Fuzzy-Modell in UML-Notation.

Um eine Darstellung des Gesamtmodells zu erhalten, sind die einzelnen Pakete an den entsprechenden Stellen zusammenzufügen. Durch die Modularität der Anbindung erfährt das bisherige Datenmodell eine einfache Erweiterung um die Unsicherheitsbeschreibung. Dies hat den Vorteil, dass problemlos auf bestehenden Lösungen aufgesetzt werden kann. Die darin vorhandenen Funktionalitäten lassen sich auch weiterhin nutzen, wobei aber die Berücksichtigung der Unsicherheit im Normalfall eine Anpassung erfordert. Konkret bewirkt die gefundene Anbindung, dass zu jedem spezialisierten Objekttyp genau eine Unsicherheitsbeschreibung hinzuzufügen ist. Welches Aussehen diese hat, entscheidet sich im Wesentlichen durch das benutzte Unsicherheitsmodell. Alle im Kapitel 3 diskutierten Anwendungsaufgaben stützen sich auf das hier vorgestellte Modell, das damit die Grundlage für die Berücksichtigung der Unsicherheit bildet.

3 Anwendung der Unsicherheit in den GIS-Komponenten

3.1 GIS-Komponenten

Ein Geo-Informationssystem ist ein auf einen speziellen Anwendungsbereich fixiertes gewöhnliches Informationssystem, das den üblichen physischen Aufbau mit Hardware, Software, Daten und Anwendern aufweist. Die Spezialisierung steckt in der Vorsilbe „Geo“, die verdeutlicht, dass der Raumbezug als verbindendes Element in allen darin behandelten Daten vorhanden ist. Das System hat zur Verarbeitung dieser Geo-Daten im Rahmen einer bestimmten Anwendung verschiedene Aufgaben wahrzunehmen. Die Aufgaben können in vier verschiedene Komponenten aufgeteilt werden. Am Anfang jeder Anwendung steht die Erfassung der Daten, die in einer Verwaltungskomponente, z.B. einer Datenbank, gemäß dem Datenmodell strukturiert und gespeichert werden. Die Mächtigkeit eines Systems offenbart sich in den verfügbaren Analysen. Ohne Analysefunktionen läge ein reines Auskunftssystem vor, das nur in der Lage ist, die gespeicherten Daten wieder auszugeben. Der eigentliche Gewinn liegt aber in der gemeinsamen Verarbeitung aller Daten, durch die neue Sichtweisen erzeugt, unbekannte Zusammenhänge aufgedeckt und damit Grundlagen für Entscheidungsprozesse geschaffen werden können. Grundsätzlich kann ein GIS als Werkzeug angesehen werden, das dem Menschen als Informationsquelle dienen soll. In dieser Funktion spielt die Präsentation der Inhalte und der Analyseergebnisse eine wichtige Rolle. Durch die fortschreitende Automatisierung gewinnt die Bildschirmpräsentation gegenüber der Ausgabe in Papierform immer mehr an Bedeutung. Dabei gelten ganz verschiedene Ansprüche, so dass beide Ausgabeformen als eigenständige Bereiche anzusehen sind. Zusammengefasst bilden die vier Komponenten das sogenannte EVAP-Modell. An diesem Modell orientieren sich häufig die in einem GIS vorzufindenden Softwaremodule.

Die vorliegende Arbeit hat die Aufgabe übernommen, die Geo-Daten um den Aspekt der geometrischen Unsicherheit zu ergänzen. Der neue Inhalt erfordert eine Anpassung in den Aufgabenbereichen eines GIS, so dass eine angemessene Berücksichtigung erfolgen kann. Alle vier Komponenten sind in gleichem Maße davon betroffen, da die Unsicherheit als zusätzlicher Bestandteil genauso wie die Daten erfasst, verwaltet, analysiert und präsentiert werden muss. Bei der Erfassung der Unsicherheit sind konkret die Parameterwerte zur Beschreibung der Variation für jedes individuelle Objekt entsprechend dem gewählten Unsicherheitsmodell festzulegen. Verschiedene Vorgehensweisen sind möglich. Die mittlere Geometrie nimmt eine besondere Stellung innerhalb der Modelldefinitionen ein. Ein Verfahren ist zu entwickeln, das eine Bestimmung der mittleren Geometrie für alle drei Objekttypen (Punkt, Linie, Fläche) ermöglicht. Die erzeugten Werte sind gemeinsam mit den eigentlichen Objektdaten abzuspeichern. Es können verschiedene Ansätze gefunden werden, die eine Erweiterung des Verwaltungskonzeptes bewirken.

Aufwendigste Anpassungsarbeiten werden in der Analysekomponente notwendig. Von Interesse ist dabei, wie sich die Unsicherheit in den Eingangsdaten auf das Ergebnis einer Analyse überträgt. Dieser Vorgang soll in Verallgemeinerung zur Fehlerfortpflanzung als Fortpflanzung der Unsicherheit bezeichnet werden. Prinzipiell sind alle Verarbeitungsfunktionen davon betroffen und bedürfen einer entsprechenden Erweiterung der Funktionalität. Die Einbindung sollte so realisiert werden, dass ein automatischer Prozess entsteht, der keine Nutzereingaben erforderlich macht. Damit kann die Fortpflanzung verdeckt im Hintergrund während der eigentlichen Analyse ablaufen. Anschließend steht dem Nutzer neben dem Analyseergebnis auch dessen Unsicherheit zur weiteren Behandlung zur Verfügung. Zur Präsentation gehört sowohl die alphanumerische Ausgabe der Unsicherheitswerte als Reaktion auf eine Nutzeranfrage als auch die graphische Visualisierung der Unsicherheit der Objektgeometrie. Hier ist die Problematik zu lösen, dass zu diesem Zweck die Unsicherheit zusätzlich zur eigentlichen Objektinformation (geometrische und thematische Inhalte) so dargestellt werden muss, dass alle Inhalte ansprechend zur Geltung kommen. Überdeckungen oder Überfrachtungen sind zu vermeiden.

Die folgenden Abschnitte widmen sich den diskutierten Erweiterungen der Komponenten. Am Beginn stehen die Betrachtungen zur Erfassung der Unsicherheit (Kapitel 3.2). Normalerweise wäre die Präsentation der Unsicherheit gemäß der sonst üblichen zeitlichen Abfolge der Bearbeitungsschritte am Ende zu behandeln. In Abweichung von dieser Reihenfolge wird die Präsentation bewusst an die zweite Stelle (Kapitel 3.3) vorgezogen, um dadurch Darstellungsmöglichkeiten der Unsicherheit zur Verfügung zu stellen, von denen die anderen Komponenten durch eine anschaulichere Erläuterung profitieren. Die Betrachtungen zur Verwaltung der Unsicherheit folgen im Anschluss daran (Kapitel 3.4). Abschließend widmet sich ein wichtiger Abschnitt der Berücksichtigung der Unsicherheit in der Analysekomponente (Kapitel 3.5). Die Erläuterungen konzentrieren sich insbesondere auf geometrische Funktionen, d.h. solche Operationen an denen die Objektgeometrie maßgeblich beteiligt ist. In allen Abschnitten werden alle drei gegebenen Unsicherheitsmodelle in gleichem Maße berücksichtigt und jeweils auf deren Besonderheiten eingegangen.

3.2 Erfassung der Unsicherheit

Die Beschreibung der Variation setzt sich allgemein aus mehreren Teilen zusammen. Einheitlich bauen alle Unsicherheitsmodelle auf der mittleren Geometrie als Basis auf, so dass deren Bestimmung eine besondere Bedeutung zukommt. Zusammen mit der mittleren Geometrie verwenden die einzelnen Modelle eine Reihe individueller Modellparameter als weitere Beschreibungshilfen für die Unsicherheit. Während diese Modellparameter die lokale Variation in detaillierter Form charakterisieren, geben die Variationsmaße einen übergeordneten Eindruck über die globale Objektunsicherheit. Mittlere Geometrie, Modellparameter und Variationsmaß bilden zusammen die Unsicherheitsbeschreibung, die es im Erfassungsschritt zu bestimmen gilt.

3.2.1 Bestimmung der mittleren Geometrie

Die mittlere Geometrie bildet den geometrischen Bezugspunkt für die weitere Beschreibung der Unsicherheit durch die Modellparameter. Die bisherige Praxis bei der Erfassung der Objektgeometrie verfährt so, dass nur eine einmalige Beobachtung durchgeführt wird. Als Ergebnis erhält man eine einzige räumliche Repräsentation des Objektes, die im System verwaltet und in Analysen verwendet wird. Sie ist als einfache Schätzung der mittleren Geometrie anzusehen. Eine mehrfache Erfassung führt zu einer besseren Bestimmung der mittleren Geometrie. Voraussetzung hierfür ist die unabhängige Beobachtung der einzelnen Realisierungen. Dies gilt z.B. dann, wenn die Erfassungen von verschiedenen Operateuren durchgeführt werden.

Liegt eine Mehrfacherfassung vor, so ist die mittlere Geometrie durch Mittelbildung sämtlicher Realisierungen zu schätzen. Wie Abb. 3.1 zeigt, handelt es sich aufgrund der möglichen Formenvielfalt der Objekte um ein komplexes Problem, für das bisher noch keine allgemeine Lösungsweise existiert, die auf beliebige Eingabedaten angewendet werden kann.

Spezielle Lösungen sind sowohl im Vektor- als auch im Rasterbereich zu finden. Edwards (1994) berichtet über ein Vektorverfahren, das zunächst alle Realisierungen in eine parametrisierte Darstellung überführt. Anschließend werden die korrespondierenden Parameter gemittelt und eine mittlere Darstellung erzeugt. Theoretisch können beliebig viele Realisierungen in die Berechnung einbezogen werden. Probleme bereiten allerdings stark unterschiedliche Linienlängen der Realisierungen, die zu Verzerrungen in der Mittelbildung führen (Keul, 1998). Verfahren im Rasterbereich verwenden Distanztransformationen, um eine Skelettierung der Fläche, die sich zwischen der Geometrie zweier Realisierungen befindet, zu erhalten. Es sind verschiedene Möglichkeiten bekannt, wie eine solche Skelettierung ausgeführt werden kann (z.B. Kreifelts und Wötzel, 1977, Serra, 1982, Cramer, 1993, Winter, 1996). Neben der Abhängigkeit von der Auflösung weisen diese Verfahren den großen Nachteil auf, dass sich gleichzeitig nur zwei Realisierungen mitteln lassen. Für jede weitere Realisierung ist das komplette Verfahren zu wiederholen.

Aufgrund der Schwächen der bestehenden Verfahren wird in dieser Arbeit ein Lösungsansatz eingeführt werden, der die Bestimmung der mittleren Geometrie für beliebige Eingabedaten ermöglicht. Er lässt sich daher als allgemeine geometrische Mittelbildung bezeichnen. Das Verfahren ist in vier Einzelschritte (Mulfacherfassung, Formbeschreibung, Segmentierung und Mittelbildung) unterteilt, die nacheinander zu durchlaufen sind. Die einzelnen Schritte werden nachfolgend näher erläutert.

1. Mulfacherfassung

Die Objektgeometrie ist mehrfach (mindestens zweifach) unabhängig voneinander zu erfassen. Es entsteht eine gewisse Anzahl an Realisierungen, die zu mitteln sind. Je höher die Anzahl ist, desto besser fällt die Bestimmung der mittleren Geometrie aus. Aus diesem Grund sollte das angewendete Verfahren mit beliebig vielen Realisierungen umgehen können.

2. Generierung einer vereinfachten Formbeschreibung

Der zweite Schritt stellt für jede Realisierung eine vereinfachte Formbeschreibung her. Sie dient dem Zweck, die Realisierungen über identische Punkte zusammenzuführen. Dazu wird die Geometrie auf eine möglichst geringe Anzahl an markanten Punkten reduziert. Ein Punkt gilt dann als markant, wenn er zur globalen Formbeschreibung des Objektes beiträgt. Die globale Formbeschreibung entspricht einer generalisierten Darstellung des Objektes, die sich aufgrund einer Glättung der Geometrie ergibt. Der bekannte Douglas-Peucker-Algorithmus (Douglas und Peucker, 1973) liefert als globales Glättungsverfahren gute Ergebnisse. Eine ausführliche Beschreibung des Algorithmus kann z.B. in Bill (1999b) gefunden werden. In diesem Verfahren bleiben nur solche Punkte erhalten, die aufgrund ihrer Lage im Objekt eine ausgezeichnete Stelle einnehmen. Gesteuert wird das Verfahren durch Angabe ei-

nes Schrankenwertes. Er verdeutlicht den minimal erforderlichen Abstand eines markanten Punktes von der Sehne seiner zwei Nachbarn. Im Normalfall ist der Schrankenwert situationsabhängig vom Nutzer zu wählen. Eine Automatisierung des Verfahrens erfordert aber die Formulierung einer Funktion, mit deren Hilfe sich ein angemessener Schrankenwert berechnen lässt.

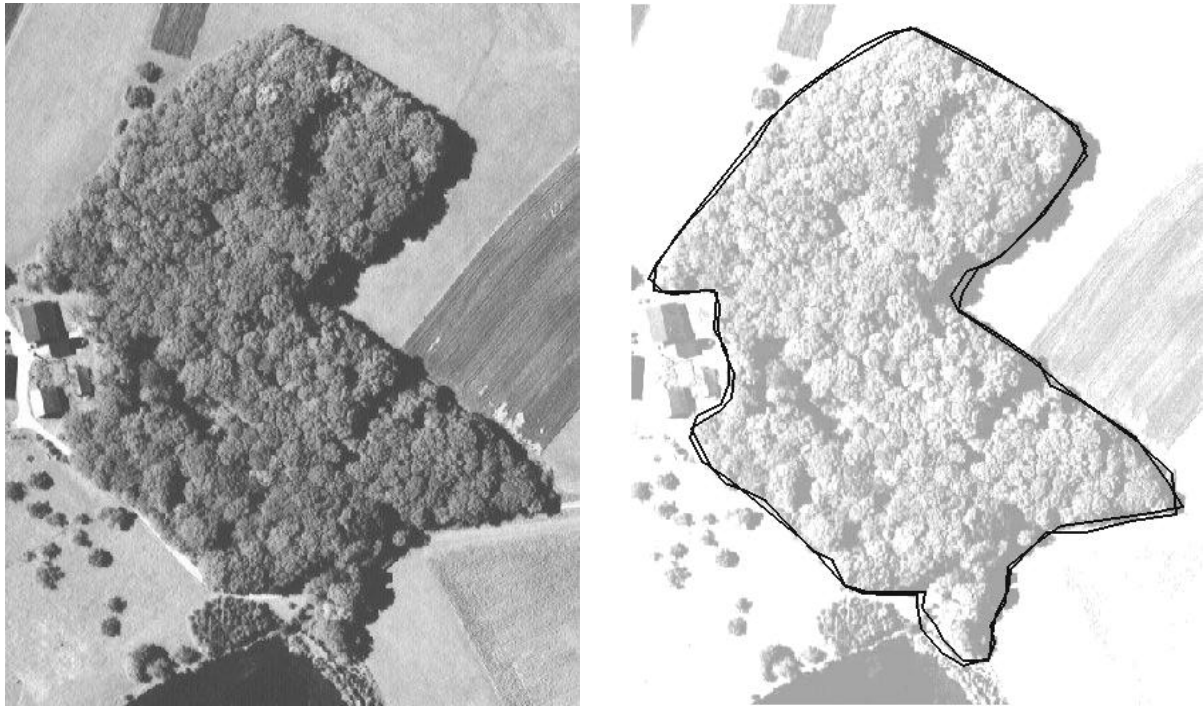


Abb. 3.1: Beispiel für ein Waldobjekt (Orthophoto) und dessen zweifache unabhängige Erfassung der Objektgeometrie.

Weiterhin bleibt die Anwendung des Douglas-Peucker-Algorithmus auf offene Polygonlinien beschränkt, da eine Pfeilhöhe nur dann definiert ist, wenn Anfangs- und Endpunkt verschieden sind. Geschlossene Linienzüge, wie sie Flächen bilden, müssen vorab in mindestens zwei Einzelsegmente zerteilt werden, die dann als getrennte Glättungsaufgaben zu bearbeiten sind. Problematisch ist dabei, dass die beiden Trennpunkte nicht willkürlich gewählt werden dürfen, weil sie später als markante Punkte erhalten bleiben.

Daher sind der Glättung zwei Vorverarbeitungsschritte voranzustellen. Benötigt werden zum einen ein Schrankenwert und zum anderen die Bestimmung von zwei Trennpunkten für geschlossene Linienzüge. Ansatzpunkt dafür bildet die Berechnung der mittleren Kantenlängen und der mittleren Winkeländerung der Kanten. Die Formeln lauten folgendermaßen:

$$\text{Mittlere Kantenlänge : } \mathbf{m} = \frac{1}{m} \sum_{i=1}^m l_i, \quad (3-1)$$

$$\text{Mittlere Winkeländerung : } \mathbf{m}_w = \frac{1}{n} \sum_{i=1}^n |w_i - \mathbf{p}| = \frac{1}{n} \sum_{i=1}^n w'_i, \quad (3-2)$$

mit l_i als Kantenlänge für m Kanten, w_i als Brechungswinkel und w'_i als Betrag der Winkeländerung im Punkt P_i für n Punkte (Abb. 3.2a). Mit den berechneten Mittelwerten lässt sich eine spezielle mittlere Situation analysieren. Es wird angenommen, dass sich ein markanter Punkt (P_i) mindestens in mittlerer Entfernung \mathbf{m} vom Vorgänger (P_{i-1}) und Nachfolger (P_{i+1}) befindet, wobei im Punkt die Änderung in Kantenrichtung der mittleren Winkeländerung \mathbf{m}_w entspricht (Abb. 3.2b). Der zu dieser Situationsbeschreibung gehörende Schrankenwert d lautet:

$$\text{Douglas-Peucker-Schrankenwert : } d = \mathbf{m} \cdot \sin(\mathbf{m}_w / 2). \quad (3-3)$$

Die Berechnung der Mittelwerte und des Schrankenwertes ist unabhängig für jede Realisierung durchzuführen, um die typischen Eigenheiten einer Erfassung (und eines Erfassers) zu berücksichtigen.

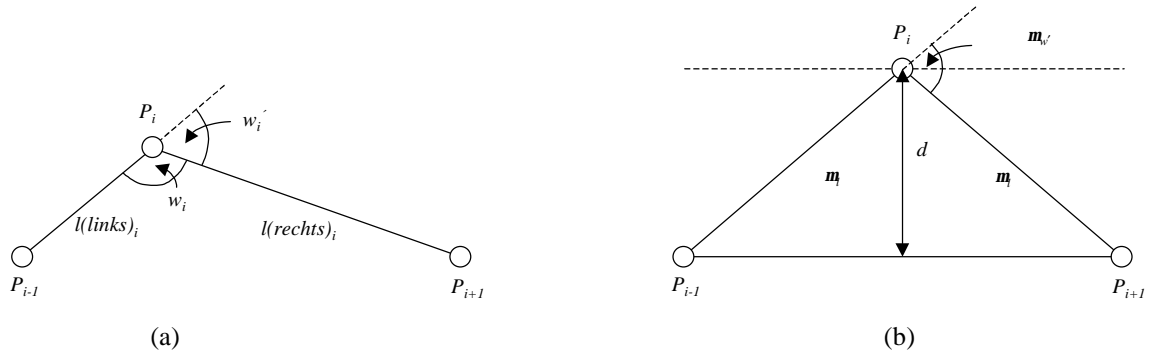


Abb. 3.2: Mittelung von Kantenlängen und Richtungsänderungen (a) zur Bestimmung des Schrankenwerts (b) für eine automatische Glättung der Geometrie.

Die bereits berechneten Kantenlängen und Winkeländerungen können ebenfalls für die Bestimmung der Trennpunkte im Falle von geschlossenen Linienzügen genutzt werden. Gesucht sind zwei markante Punkte, die eine herausragende Bedeutung in der Formbeschreibung besitzen, da die Trennpunkte während der anschließenden Glättung erhalten bleiben. Die Bewertung der Punkte erfolgt aufgrund der folgenden beiden Kriterien:

- Längenkriterium: Längen der inzidierenden Kanten größer als Mittelwert:

$$\begin{aligned} \text{Länge der linken Kante : } l(\text{links})_i &> \mathbf{m}, \\ \text{Länge der rechten Kante : } l(\text{rechts})_i &> \mathbf{m}, \end{aligned} \quad (3-4)$$

- Winkelkriterium: Winkeländerungen größer als Mittelwert:

$$\text{Winkeländerung : } w'_i > \mathbf{m}_w. \quad (3-5)$$

Beim Längenkriterium (3-4) wird davon ausgegangen, dass ein Punkt dann eine größere Bedeutung besitzt, wenn der Erfasser den Punkt nach einem überdurchschnittlichen Zwischenraum gesetzt hat. Gleichfalls von besonderem Interesse ist ein Punkt, in dem die Linie eine große Winkeländerung erfährt, womit sich das Winkelkriterium (3-5) begründen lässt.

In Abb. 3.3 sind für eine Realisierung des Waldbeispiels (Abb. 3.1) die Elemente hervorgehoben, die das Längenkriterium (Abb. 3.3a) bzw. Winkelkriterium (Abb. 3.3b) erfüllen. Beide Kriterien können in einer Bewertungsfunktion zusammengefügt werden. Sie lautet:

$$I(P_i) = p_w \frac{w'_i - \mathbf{m}_w}{\max(w'_i) - \mathbf{m}_w} + \frac{1}{2} p_l \frac{l(\text{links})_i - \mathbf{m}}{\max(l_i) - \mathbf{m}} + \frac{1}{2} p_r \frac{l(\text{rechts})_i - \mathbf{m}}{\max(l_i) - \mathbf{m}}. \quad (3-6)$$

(3-6) sieht eine unterschiedliche Berücksichtigung des Längen- und des Winkelkriterium durch entsprechende Wahl von Gewichten (Längengewicht p_l bzw. Winkelgewicht p_w) vor (gleichgewichtete Standardeinstellung: $p_l = 1$ und $p_w = 1$).

Die beiden größten Werte ergeben die beiden gesuchten Trennpunkte (Abb. 3.3c), an denen die geschlossene Form in zwei unabhängig voneinander zu behandelnde offene Linienzüge aufgetrennt werden kann. Bereits das Ergebnis der Bewertungsfunktion könnte genutzt werden, um eine Auswahl von markanten Punkten zu treffen, die dann eine Formbeschreibung des Objektes definieren. Um jedoch besser auf die globalen geometrischen Verhältnisse einzugehen, schließt sich eine Glättung mit Hilfe des Douglas-Peucker-Algorithmus an. Alle nach dieser Glättung erhalten gebliebenen Punkte sind als markante Punkte aufzufassen, die zusammen die gesuchte genäherte Form des Objektes bilden. Die auf diese Weise für das Beispiel ermittelten markanten Punkte sind in Abb. 3.3d zu sehen.

3. Segmentierung

Die markanten Punkte legen eine Approximationsgeometrie für jede Realisierung fest. Mit ihrer Hilfe kann die Gesamtgeometrie in einzelne Segmente aufgeteilt werden, die kleinere Einheiten für das Mittelungsproblem bilden (Abb. 3.4a). Wichtig ist, dass in allen Realisierungen eine identische Segmentierung vorliegt, d.h. Start- und Endpunkte in der Wirklichkeit dieselben Punkte bezeichnen. Dazu wird als beste Approximation die Realisierung ausgewählt, die die geringste Anzahl an markanten Punkten besitzt. Für sie kann angenommen werden, dass ihre Punkte auch in den anderen Realisierungen enthalten sind. Um den konkreten Zusammenhang zu klären, ist eine

eindeutige Zuordnung der markanten Punkte über alle Realisierungen durchzuführen. Dazu wird jeder markante Punkt der besten Approximation in alle anderen Realisierungen übertragen, um dort jeweils einen korrespondierenden Punkt festzulegen (Abb. 3.4b). Zunächst erfolgt eine Suche nach markanten Punkten innerhalb eines Fangkreises (z.B. mit Radius $r = \min(l_i)$), von denen dann der nächstgelegene Punkt zuzuordnen ist. Liegt kein markanter Punkt im Fangkreis, so soll der Punkt korrespondieren, der den absolut geringsten Abstand zum übertragenen Punkt aufweist.

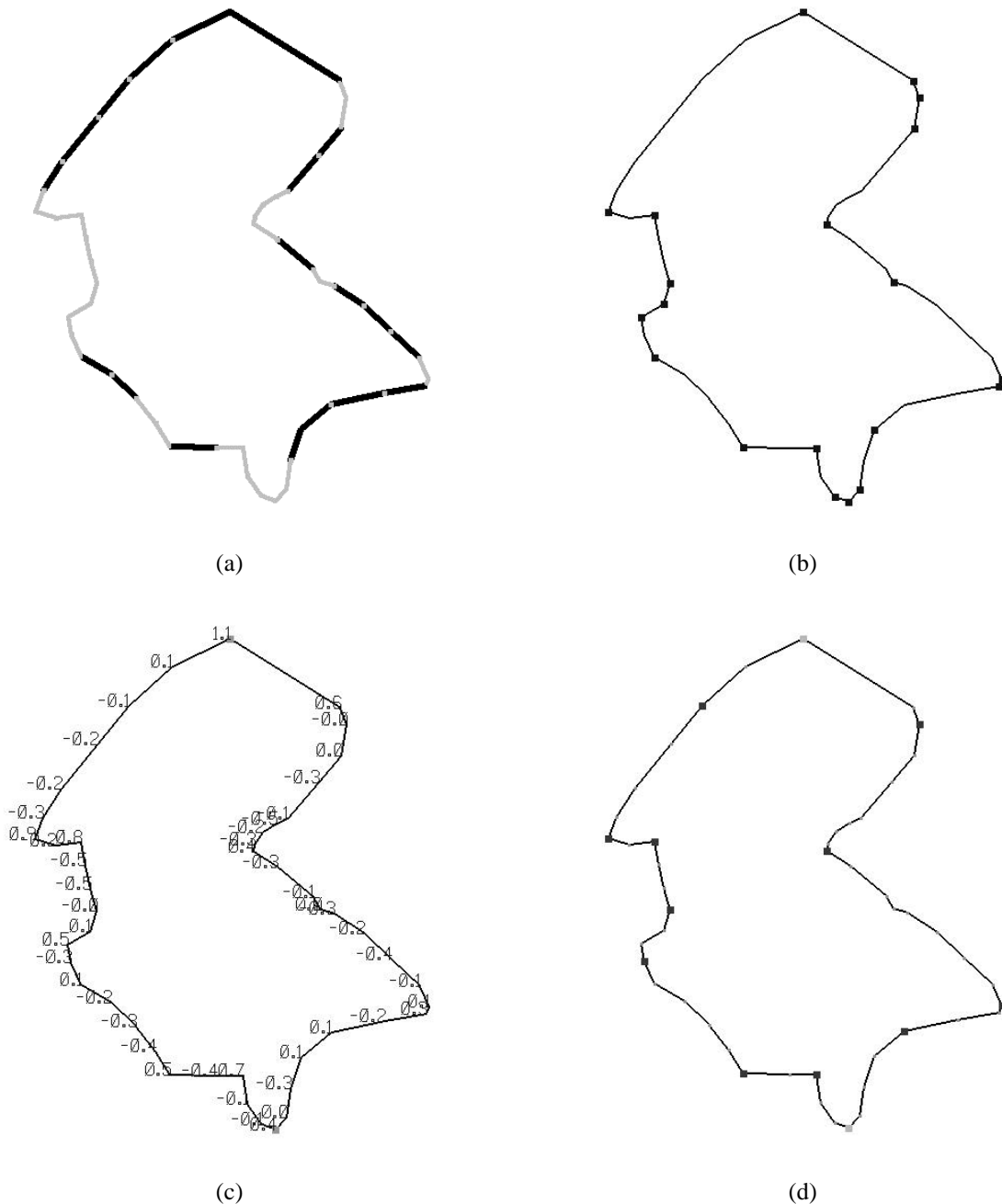


Abb. 3.3: Bestimmung der markanten Punkte einer flächenförmigen Realisierung: überdurchschnittliche Kantenlängen (dicke, dunkle Kanten) (a), überdurchschnittliche Winkeländerungen (markierte Punkte) (b), Werte der Bewertungsfunktion (I -Werte) mit Punkten der Maximalwerte (markiert durch kleine Pfeile) (c) und endgültige markante Punkte (dunkel markierte Punkte) (d).

Wenn alle Zuordnungen gefunden sind, kann eine verbesserte Approximationsgeometrie geschätzt werden, indem alle korrespondierenden Punkte gemittelt werden. Dadurch ergeben sich die endgültigen Positionen der markanten Punkte in der Approximationsgeometrie, die gleichzeitig das Grundgerüst der mittleren Geometrie bildet. Aus der Zuordnung der Punkte kann direkt auf die Zuordnung von Teilabschnitten jeder Realisierung zu den Segmenten geschlossen werden. Eine solche Zuordnung bildet die Grundlage für den nachfolgenden Mittelungsprozess.

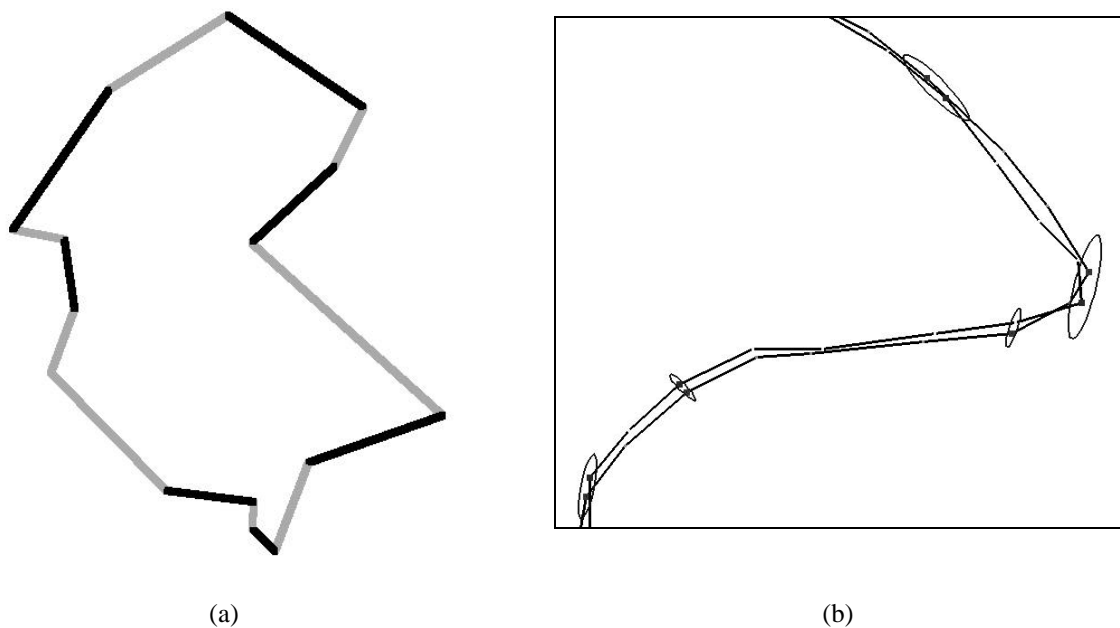


Abb. 3.4: Auftrennung des Gesamtproblems in einzelne Segmente (a) und Suche nach korrespondierenden markanten Punkte in den beiden dargestellten Realisierungen (b). Korrespondierende Punkte sind in einer Ellipse eingefangen.

4. Mittelbildung

Im letzten Schritt erfolgt die eigentliche Mittelbildung. Sie wird segmentweise ausgeführt, d.h. alle korrespondierenden Teilabschnitte (Segmente) aus den verschiedenen Realisierungen bilden zusammen eine Auswerteeinheit. Start- und Endpunkt für jedes Segment sind die jeweils gemittelten markanten Punkte, die bereits ihre endgültigen Positionen besitzen. Die geradlinige Verbindung von Start- und Endpunkt definiert eine Näherung an die gesuchte mittlere Segmentlinie, die mit Hilfe der vorliegenden Realisierungen zu verbessern ist. Dazu führt man in diskreten, gleichmäßigen Abständen Zwischenpunkte ein. Die Güte der Schätzung der mittleren Geometrie hängt unmittelbar von der Dichte der Zwischenpunkte ab, so dass möglichst kleine Abstände gewählt werden sollten (z.B. im Abstand $d = \min l / 2$). Für jeden Zwischenpunkt sind anschließend die senkrechten Abstände zwischen der Näherung an das Mittel und allen Realisierungen zu berechnen und zu mitteln (Abb. 3.5). Der durch die gemittelten Abstände definierte mittlere Zwischenpunkt, legt einen neuen Punkt der endgültigen Geometrie fest.

Durch Anwendung des Verfahrens auf alle Segmente ergibt sich eine verbesserte Schätzung der mittleren Geometrie (Abb. 3.6). Die mittlere Geometrie weist zumeist eine hohe Punktdichte auf, die durch eine abschließende Glättung sinnvoll reduziert werden kann. Der dazu notwendige Schrankenwert ist so zu wählen, dass die geometrischen Änderungen durch den Wegfall von Zwischenpunkten im Verhältnis zur enthaltenen Unsicherheit klein ausfallen. Aus diesem Grund erfolgt die konkrete Festlegung des Schrankenwertes erst nach der Bestimmung der Unsicherheitsparameter (Kapitel 3.2.2).

Die vorgestellte allgemeine Mittelbildung weist einige typische Eigenschaften auf, in denen sich die Unterschiede zu bestehenden Ansätzen offenbaren. Das Verfahren kann sowohl auf Linien als auch auf Flächen angewendet werden. Es entstehen keine geometrischen Verzerrung, die z.B. bei einer parametrischen Lösung auftreten können (Keul, 1998). Es können beliebig viele Realisierungen einbezogen werden. Die Qualität der Lösung lässt sich durch die Punktdichte steuern, wobei eine anschließende Glättung eine qualifizierte Reduktion ermöglicht. Mit der Mittelbildung wird gleichzeitig eine Approximationsgeometrie erzeugt, die eine generalisierte Form des Objektes darstellt. Während des Verfahrens erfolgt eine Zuordnung der Realisierungen zueinander und eine Unterteilung in korrespondierende Segmente, die eine Aufspaltung und Parallelisierung erlaubt. Ein entscheidender Vorteil liegt in der Möglichkeit, den Ansatz auch zur Bestimmung der Unsicherheitswerte anzuwenden. Durch einfache Erweiterungen lassen sich Modellparameter und Variationsmaße zusammen mit der mittleren Geometrie in einem Arbeitsschritt schätzen.

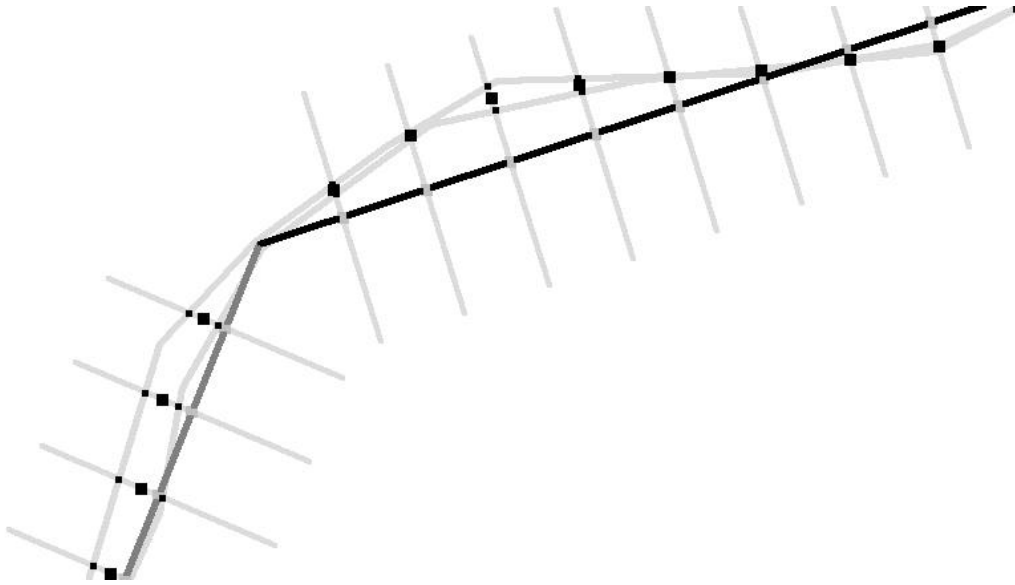


Abb. 3.5: Beispiel für die Mittelbildung zwischen zwei Realisierungen in zwei Segmenten (dunkle, gerade Linien) mit den Zwischenpunkten (helle Quadrate), den Schnittpunkten der Senkrechten mit den Realisierungen (kleine Punkte) und den gemittelten Zwischenpunkten (dunkle Quadrate).

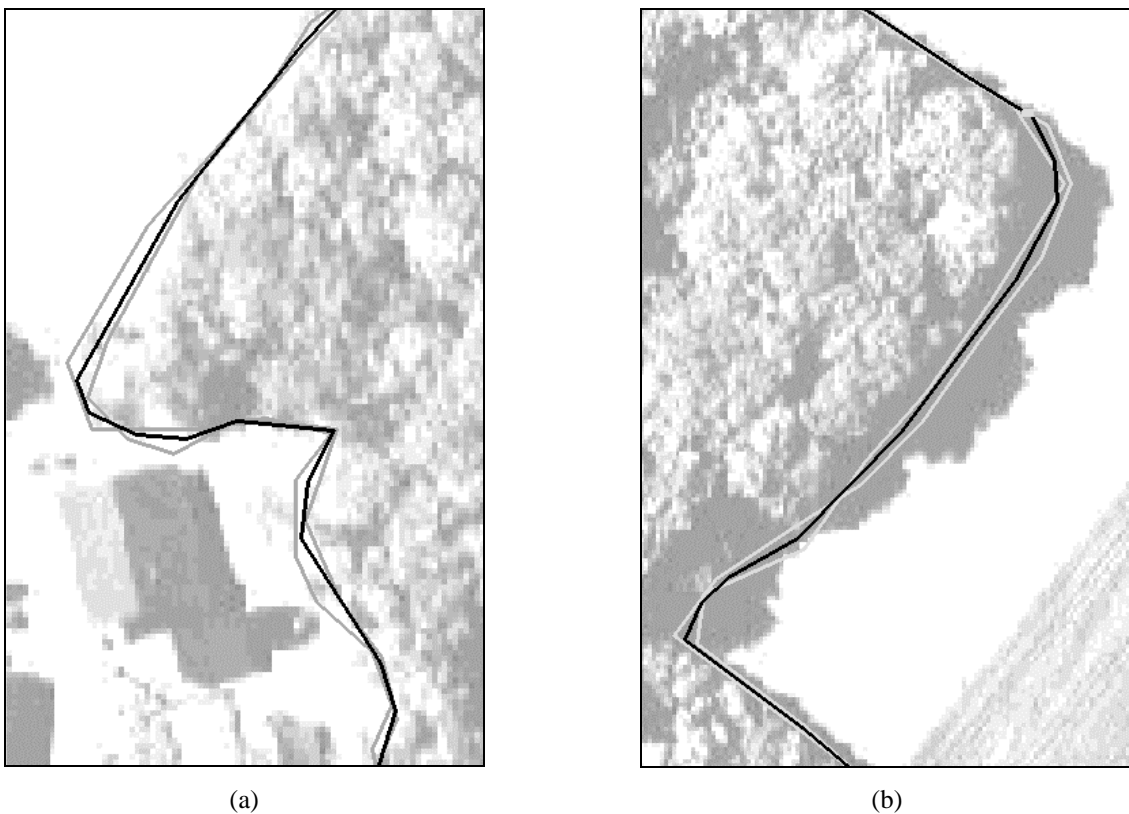


Abb. 3.6: Zwei Beispiele für die endgültige Form der mittleren Geometrie (dunkle Linie) generiert aus zwei Realisierungen (helle Linien).

3.2.2 Modellparameter und Variationsmaß

Zur detaillierten Beschreibung der Unsicherheit sind die Werte der Modellparameter zum angewendeten Unsicherheitsmodell gesucht. Aus ihnen lassen sich dann die zugehörigen Variationsmaße ableiten. Tab. 3.1 fasst die Parameter zusammen.

Tab. 3.1: Übersicht über Modellparameter und Variationsmaße der einzelnen Unsicherheitsmodelle.

	Modellparameter	Variationsmaß
Stochastisches Modell	Varianz s^2 , Breite b	Standardabweichung s
Minimum-Maximum-Modell	Minimum- und Maximum-Geometrie oder Epsilon-Distanzen e (lokal)	Epsilon-Distanz e (global)
Fuzzy-Modell	Lineare Parameter a, b	Fuzzy-Index u

Zur Bestimmung der Unsicherheitswerte bieten sich prinzipiell drei verschiedene Vorgehensweisen: das Schätzen aufgrund von Fachwissen, das Messen auf Basis von Referenzdaten und das Messen durch Mehrfacherfassungen. Alle drei Methoden werden nachfolgend im Einzelnen vorgestellt.

Schätzen der Unsicherheit aufgrund von Fachwissen

Beim Schätzen wird vorausgesetzt, dass die wirkenden Einflüsse in ihren Größenverhältnissen und ihren gegenseitigen Abhängigkeiten bekannt sind. Dies kann z.B. dann zutreffen, wenn an Beispieldatensätzen bereits Untersuchungen durchgeführt wurden und sich die Ergebnisse auf die vorliegende Situation übertragen lassen. Dazu können exemplarisch einige Objekte einer Objektart ausgewählt und deren Unsicherheit untersucht werden. Das Ergebnis ist dann auf alle anderen Objekte der gleichen Objektart zu übertragen. Je nach Detaillierungsgrad der Untersuchung sind individuelle Anpassungen möglich. Aber auch Expertenwissen kann auf diese Weise genutzt werden. Ein erfahrener Erfasser z.B. ist häufig in der Lage, die vorliegende Unsicherheit in angemessener Weise abzuschätzen. Für den Fall, dass detaillierte Kenntnis der einzelnen wirkenden Komponenten verfügbar sind, kann die Schätzung der einzelnen Modellparameter folgendermaßen durchgeführt werden:

$$\begin{aligned}
 \text{Stochastisches Modell : } s^2 &= s_1^2 + s_2^2 + \dots + s_n^2 + s_{12} + s_{23} + (\text{weitere Korrelationen}), \\
 \text{Min- / Max-Modell : } e &= e_1 + e_2 + \dots + e_n, \\
 \text{Fuzzy-Modell : } a &= a_1 + a_2 + \dots + a_n.
 \end{aligned}
 \tag{3-7}$$

Nur im Stochastischen Modell können Abhängigkeiten der Komponenten über Korrelationen angemessen berücksichtigt werden. Das Minimum-Maximum-Modell geht dagegen vom ungünstigsten Fall aus und betrachtet alle Einflüsse als unabhängig. Das Fuzzy-Modell verfügt über keine adäquate allgemeine Vereinigungsmöglichkeit auf Basis der Parameter. In Anlehnung an das Minimum-Maximum-Modell wird daher beim Fuzzy-Modell eine einfache Addition als Lösung eingeführt.

Messen der Unsicherheit auf Basis von Referenzdaten

Beim Messen der Einflüsse werden die Unsicherheitswerte individuell aus den vorliegenden Daten bestimmt. Eine Möglichkeit besteht darin, durch Vergleich mit einem externen Datensatz (Referenzdatensatz) auf die Unsicherheit zu schließen. Dadurch wird die äußere Unsicherheit charakterisiert. Um mit diesem Verfahren vertrauenswürdige Werte zu erhalten, muss der Referenzdatensatz einige Voraussetzungen erfüllen. Beide Datensätze, d.h. der zu bewertende Datensatz und der Referenzdatensatz, müssen unabhängig voneinander erfasst sein und den gleichen räumlichen und thematischen Ausschnitt abdecken. Wichtig ist, dass die Referenzdaten eine übergeordnete Qualität aufweisen, damit ihre eigene Unsicherheit vernachlässigt werden kann. Die Referenzdaten können damit als eine exakte Abbildung der Realität betrachtet werden. Eine weitere Rolle spielen mögliche Unterschiede im konzeptionellen Modell, die sich nachteilig auf die Auswertung auswirken können.

Die Vorgehensweise und die Probleme bei Verwendung eines Referenzdatensatzes werden anhand eines Beispiels verdeutlicht (Abb. 3.7). Aufgabe ist es, die Unsicherheit eines Straßendatensatzes innerhalb eines Testgebietes zu bewer-

ten. Bei den Straßendaten handelt es sich um einen Auszug aus dem Digitalen Landschaftsmodell 25/1 (DLM 25/1) des Amtlichen Topographisch-Kartographischen Informationssystems (ATKIS). Sie sind im Maßstabsbereich um 1:25.000 angesiedelt und repräsentieren die Straßen linienförmig durch eine Mittelachsendarstellung. Als Quelle für Referenzdaten eignet sich das Liegenschaftskataster, das sich ebenfalls dem Straßenraum zuwendet, ihn aber in anderer Weise durch flächenhafte Objekte der Art Flurstück (in Verkehrswegen) beschreibt. Solche Daten sind der Automatisierten Liegenschaftskarte (ALK) zu entnehmen, die ein Teilvorhaben im Verfahren des Aufbaus eines Automatisierten Liegenschaftskatasters bildet.



Abb. 3.7: Überlagerung der Beispieldatensätze für die Ableitung der Unsicherheit von ATKIS-Straßen (dunkel) aus ALK-Flurstücken (hell). Hinterlegt ist das zugehörige Orthophoto.

Da Katasterdaten dem großmaßstäbigen Bereich zuzuordnen sind (Maßstabsbereich 1:500 – 1:2.500), ist die Forderung nach übergeordneter Qualität erfüllt. Es ist eine mindestens zehnfach bessere geometrische Genauigkeit zu erwarten. Während ATKIS-Daten durch Digitalisierung von Orthophotos abgeleitet werden, erfolgt die Erfassung von ALK-Daten überwiegend durch terrestrische Vermessungen, so dass beide Datensätze bezüglich ihrer Herkunft als unabhängig anzusehen sind. Aufgrund der unterschiedlichen thematischen Ausrichtungen bestehen jedoch Unterschiede in den verwendeten konzeptionellen Modellen. Abb. 3.7 bringt die Verschiedenartigkeit der geometrischen Repräsentation der Straßen zum Ausdruck. In ALK sind Straßen in einer flächenhaften Randdarstellung gegeben, während in ATKIS diese linienhaft durch ihre Mittelachse repräsentiert werden. Ein weiterer Unterschied besteht in der anwendungsspezifischen Betrachtungsweise des Phänomens Straße. Im Kataster werden die Eigentumsverhältnisse abgebildet, während in ATKIS die Topographie und damit die Landnutzung im Vordergrund stehen. Dadurch kann es zu erheblichen Unterschieden in der geometrischen Formbeschreibung kommen, z.B. wenn ein Flurstück zusätzliche Bereiche umfasst, die nicht mehr zum Straßenraum zählen, aber eigentumsrechtlich eine Einheit bilden. Probleme bereitet auch die logische Zuordnung der Objekte. Nur in Ausnahmefällen wird ein bestimmtes Straßensegment in ATKIS mit genau einem Flurstück übereinstimmen.

Um trotz der Unterschiede einen Vergleich zu ermöglichen, ist der Referenzdatensatz geeignet aufzubereiten. Dazu sind zunächst die Flurstücke auszuschließen, die aufgrund der anwendungsspezifischen Sichtweise keine Weiterbehandlung

erlauben. Für die verbleibenden Flurstücke kann dann eine Mittelachsentransformation durchgeführt werden, die aus der Randdarstellung eine Mittelachsendarstellung erzeugt (Abb. 3.8).

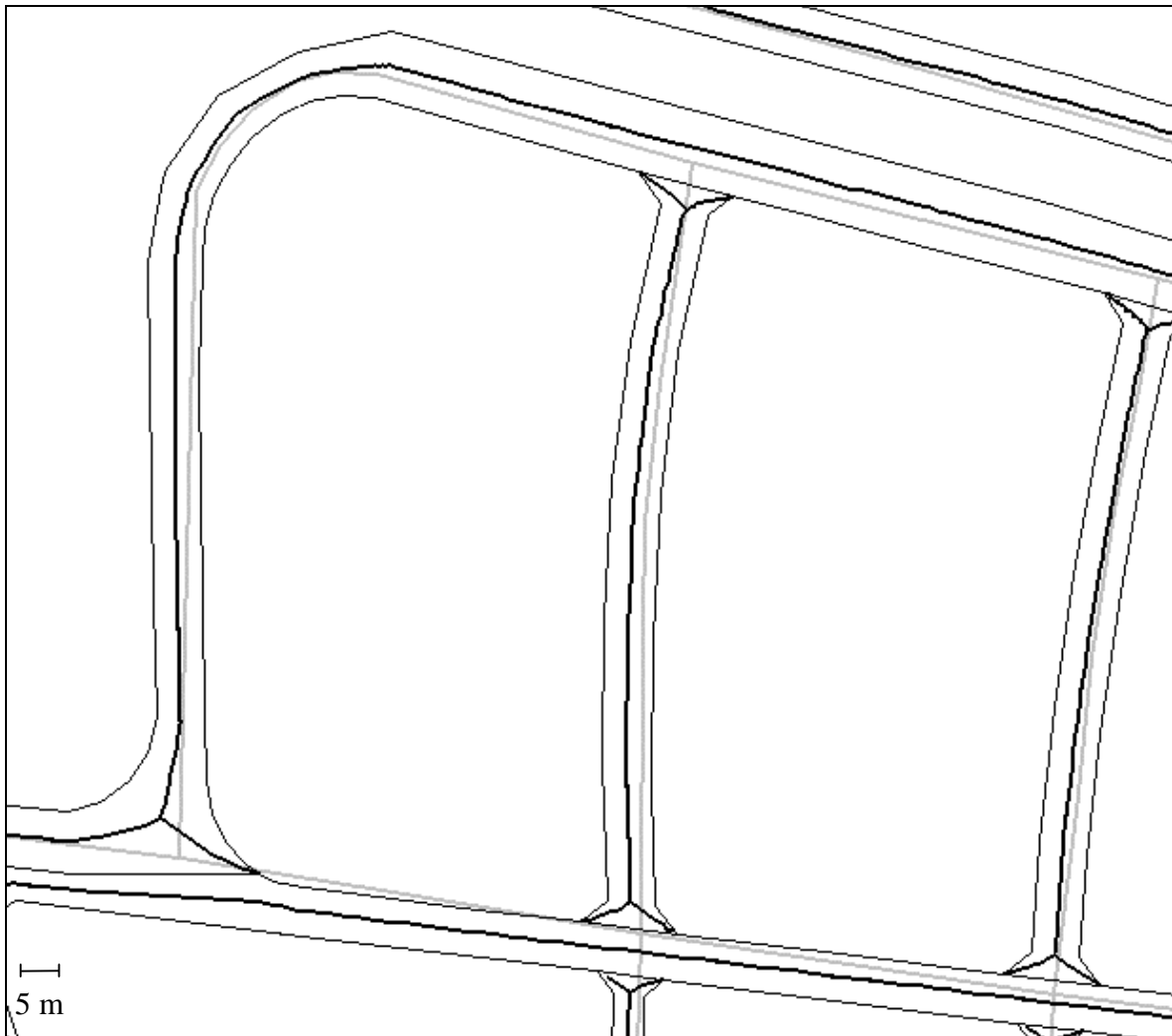


Abb. 3.8: Ausschnitt aus dem Beispieldatenbestand mit Darstellung der berechneten Mittelachsen (dicke, schwarze Linien), der Flurstücke (dünne, schwarze Linien), überlagert mit den ATKIS-Straßen (helle Linien).

Die Transformation entspricht einer Skelettierung der Fläche. Existierende Verfahren lösen die Aufgabe häufig im Rasterbereich und nutzen dabei die Vorteile der einfachen Rasterdatenstruktur. Das im Testbeispiel angewendete Verfahren (Cramer, 1993) wandelt zunächst die Fläche in Rasterdaten um und führt nacheinander die Schritte Distanztransformation, topologische Skelettierung, Linienverfolgung und Linienglättung aus. Die Distanztransformation bestimmt den Abstand jeder Rasterzelle (Pixel) zum Rand der Fläche in Abhängigkeit von der eingesetzten Metrik. Das Ergebnis wird zur topologischen Skelettierung eingesetzt, um ein Minimalgerüst (Skelett) zu erzeugen, das die ursprüngliche Fläche auf die Breite einer Rasterzelle reduziert. Das Skelett entspricht bereits der gesuchten Mittelachsendarstellung, allerdings in Form einer Rasterdarstellung. Eine Besonderheit der topologischen Skelettierung liegt in der gleichzeitigen Klassifizierung der zum Skelett gehörenden Rasterzellen nach ihrer topologischen Bedeutung in Linienanfangs- bzw. Linienendpixel, Knoten- und Linienpixel. Dadurch wird die nachfolgende Raster-Vektor-Transformation vorbereitet. Innerhalb der Linienverfolgung werden ausgehend von einem Linienanfangspixel solange benachbarte Linienpixel aneinandergereiht, bis man an einem End- oder Knotenpixel ankommt. Die Prozedur wiederholt sich bis alle Linienpixel durchlaufen sind. Alle aneinandergereihten Linienpixel werden anschließend in Linienpunkte umgewandelt, so dass sich letztendlich eine Vektorrepräsentation des Skeletts ergibt. Abschließend kann darauf eine Glättung ausgeführt werden, mit dem Ziel, das stufige Erscheinungsbild zu beseitigen. Im letzten Teil der Aufbereitung sind die berechneten Mittelachsen den korrespondierenden Straßen eindeutig zuzuordnen, d.h. es ist eine 1:1-Relation herzustellen. Es gibt eine ganze Reihe möglicher Problemefälle, die einer automatischen Lösung der Zuordnung entgegenstehen (Walter, 1997). Im für das Testbeispiel durchgeführten manuellen Verfahren werden die Mittelachsen der Flurstücke zunächst in zu den Straßenachsen korrespondierende Einheiten geometrisch zusammengefügt oder geteilt und anschließend eine logische Verknüpfung zwischen den korrespondierenden Objekten der Datensätze aufgebaut.

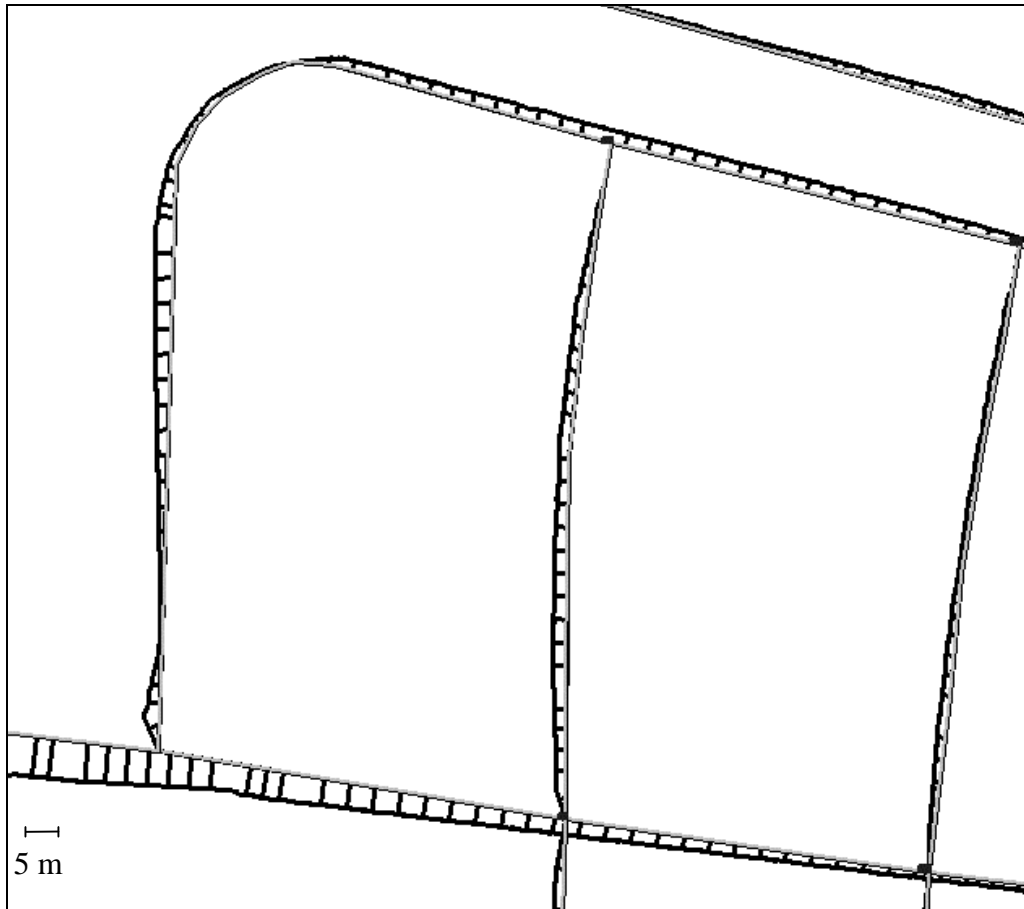


Abb. 3.9: Abweichungen der ATKIS-Straßen (hell) von der Mittelachse (schwarz) an diskreten Stellen.

Nach Aufbereitung der Katasterdaten kann die eigentliche Schätzung der Unsicherheit erfolgen. Dazu sind die geometrischen Abweichungen der beiden Datensätze zu bestimmen (Abb. 3.9). Die gesuchten Abweichungen definieren sich als senkrechte Abstände des Straßenobjekts zur berechneten Mittelachse des korrespondierenden Flurstücks. Aufgrund des kontinuierlichen Verlaufs der zu vergleichenden Linien, lassen sich Abweichungen an unendlich vielen Stellen bestimmen. Um aber die Unabhängigkeit der Werte zu gewährleisten, wird der Vergleich nur an einigen wenigen ausgewählten Stellen pro Straßenobjekt durchgeführt. Durch die hergestellte Zuordnung korrespondierender Elemente kann der Vergleich vollautomatisch, d.h. ohne Eingriff eines Operateurs erfolgen. Aus den Abweichungen leiten sich die gesuchten Modellparameter folgendermaßen ab:

$$\begin{array}{ll}
 \text{Stochastisches Modell :} & \text{Varianz } \mathbf{s}^2 = \sum_{i=1}^n \frac{v_i^2}{n}, \\
 \text{Minimum - Maximum - Modell :} & \text{Epsilon } \mathbf{e} = \max |v_i|, \\
 \text{Fuzzy - Modell :} & \text{Parameter } \mathbf{a} = \max |v_i| + \frac{b}{2}.
 \end{array} \quad (3-8)$$

mit n_i als einzelne Abweichung zwischen Vergleichs- und Referenzdatensatz bei insgesamt n Beobachtungen. Minimum-Maximum-Modell und Fuzzy-Modell basieren im Wesentlichen auf dem Extremwert der Abweichungen. Daher ergibt sich bereits bei geringem Stichprobenumfang (z.B. bei Beobachtung von nur einer oder zwei Stellen) eine gute Beschreibung der Unsicherheitssituation (Sachs, 1978). Das Stochastische Modell erfordert für eine zuverlässige Bestimmung deutlich mehr Beobachtungen (im Idealfall sogar $n \rightarrow \infty$), da anstelle der Extremwerte eine innere Verteilung zu charakterisieren ist.

Die Untersuchung für das Testgebiet ergibt die in der Tab. 3.2 zusammengestellten Unsicherheitsmaße, die aus den Abweichungen aller Objekte abgeleitet wurden. Die Stichprobe umfasst $n = 2348$ Beobachtungen. Das Ergebnis zeigt, dass die in der ATKIS-Modellierungsvorschrift (ATKIS, 2000) verankerte Forderung nach einer Lagegenauigkeit bei Straßen von 3 m mit $s_1 = 2.2 \text{ m}$ im Allgemeinen gut erreicht wird. Die maximale Abweichung liegt bei 11.0 m . Die Verteilung der ermittelten Abweichungen kann Abb. 3.10 entnommen werden. Daraus lässt sich feststellen, dass

99.73% aller Abweichungen unter 9.6 m liegen. Individuelle Werte für einzelne Objekte sind im Kapitel 3.3 in Abb. 3.13 zu sehen.

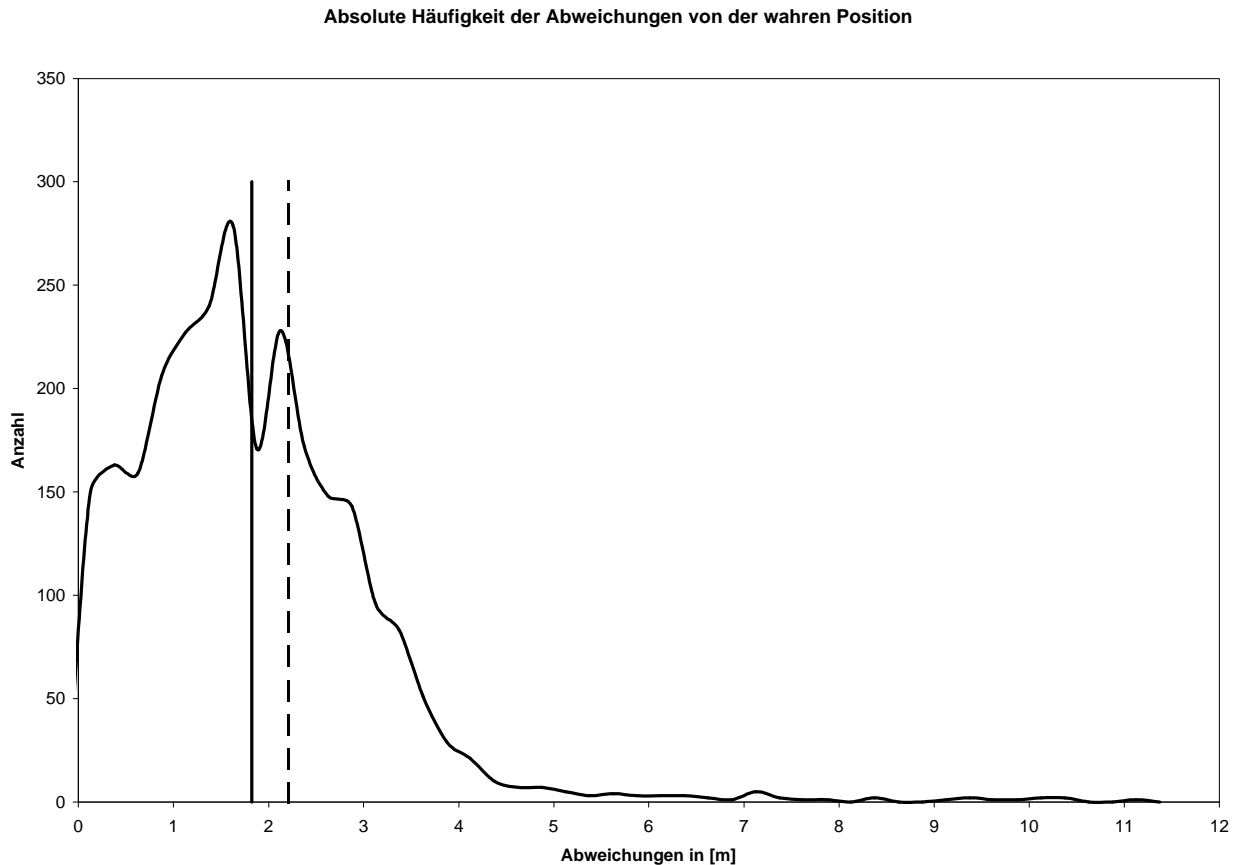


Abb. 3.10: Häufigkeitsverteilung der Abweichungen der ATKIS-Mittelachse von der geschätzten Mittelachse aus ALK mit Mittelwert (durchgezogene Linie) und Standardabweichung (gestrichelte Linie).

Tab. 3.2: Unsicherheitsmaße für die ATKIS-Straßenobjekte des Testgebietes.

	Modellparameter	Variationsmaß
Stochastisches Modell	Varianz $s_L^2 = 4.84 \text{ m}^2$ Breite $b = 6.0 \text{ m}$	Standardabweichung $s = 2.2 \text{ m}$
Minimum-Maximum-Modell	Epsilon $e = 11.0 \text{ m}$	Epsilon $e = 11.0 \text{ m}$
Fuzzy-Modell	Lineare Parameter $a = 14.0 \text{ m}$ $b = 6.0 \text{ m}$	Fuzzy-Index $u = 5.5 \text{ m}$

Der für die Modellbeschreibung von Linien wichtige Parameter Breite des Objektes b wird von ATKIS als beschreibendes Attribut zu jedem Objekt geführt. Ändert sich der Wert im Verlauf eines Objektes, dann schreibt die ATKIS-Erfassungsvorschrift eine Objektteilung vor, solange bis in den entstehenden Objektteilen eine annähernd konstante Breite vorliegt. Den Objektteilen werden dann einzelne individuelle Breitenwerte zugeordnet. Im Falle des Testdatensatzes liegen keine Breitenwerte zu den Objekten vor, so dass alternativ eine Schätzung mit Hilfe der ALK-Daten erfolgt. Ausgehend von der geschätzten ALK-Mittelachse ist für jede Straße der senkrechte Abstand zum in den Originaldaten gegebenen Rand zu bestimmen (Abb. 3.11). Der doppelte Wert ergibt die gesuchte Breite. Es genügt, die Bestimmung an wenigen Stellen pro Objektteil durchzuführen, da sich die Breite nur geringfügig ändert. Auch hier kann

das Verfahren vollautomatisch ablaufen, ohne dass manuelle Eingaben notwendig werden. Einen Eindruck über die ermittelten Breiten liefert die Abb. 3.20 in Kapitel 3.3.2.

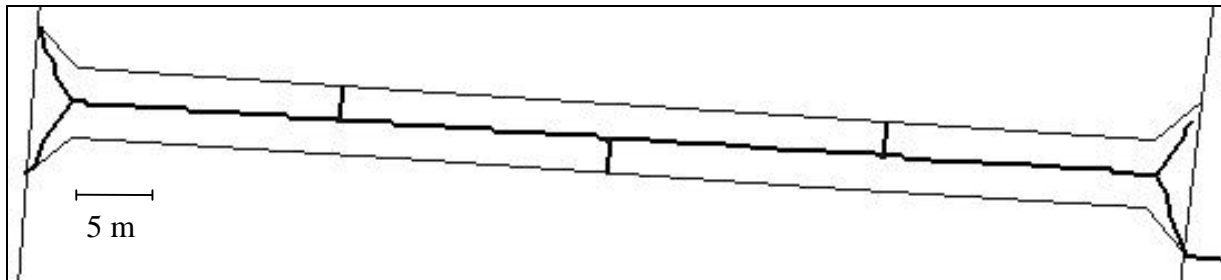


Abb. 3.11: Bestimmung der Breite für ein ausgewähltes ALK-Straßenobjekt (helle, dünne Linie) über die geschätzte Mittelachse (schwarze, dicke Linie) an drei ausgewählten Stellen.

Messen der Unsicherheit durch Mehrfacherfassungen

Die dritte Möglichkeit stützt sich auf die Auswertung von Mehrfacherfassungen. Dabei wird davon ausgegangen, dass sämtliche Einflüsse in zufälliger Weise auf jede Erfassung einwirken. Mit steigender Anzahl an Realisierungen zeigt sich die daraus resultierende Variation in ihrer Gesamtheit und mit ihr die enthaltene Unsicherheit. Die aus Mehrfacherfassungen abgeleiteten Unsicherheitsmaße charakterisieren die innere Unsicherheit, da die Werte ausschließlich aus eigenerfassten Daten abgeleitet werden. Im Gegensatz dazu legen die durch Messung über Referenzdaten bestimmten Maße die äußere Unsicherheit fest.

Die Bestimmung der Unsicherheitsmaße aus Mehrfacherfassungen stützt sich auf eine vorab erfolgte Mittelbildung (Kapitel 3.2.1). Betrachtet man die Objektgeometrie als Zufallsvariable und ihre Erfassungen als Realisierungen der Zufallsvariablen, so kann die mittlere Geometrie als Schätzung des Erwartungswertes der Zufallsgröße verwendet werden. Die Unsicherheitsmaße lassen sich aus den Abweichungen der einzelnen Realisierungen zum Mittel ableiten. Aus diesem Grund ist es sinnvoll den im Kapitel 3.2.1 gegebenen Ablauf zur Mittelbildung um einen weiteren Schritt, die Schätzung der Unsicherheit, zu ergänzen. Die Berechnung der Unsicherheit erfolgt segmentweise, da von konstanten Unsicherheitseigenschaften innerhalb eines Segmentes ausgegangen werden kann. Als Ergebnis entsteht ein unabhängiger Datensatz an Unsicherheitsmaßen für jedes Segment.

Analog zur Mittelbildung werden an diskreten Stellen (z.B. im Abstand $d = \min l$) Zwischenpunkte in die mittlere Geometrie eingefügt, an denen die senkrechten Abstände zu den Realisierungen gemessen werden (Abb. 3.12, für das Waldobjekt in Abb. 3.1). Sämtliche so ermittelte Abstandswerte werden zusammengenommen und gehen in die Bestimmung der Parameter ein. Ihre Bestimmung selbst ist wiederum abhängig vom gewählten Unsicherheitsmodell und kann wie folgt durchgeführt werden:

$$\begin{array}{ll}
 \text{Stochastisches Modell :} & \text{Varianz } \mathbf{s}^2 = \frac{1}{n-1} \sum_{i=1}^n v_i'^2 \\
 \text{Minimum - Maximum - Modell :} & \text{Epsilon } \mathbf{e} = \max |v_i'| \\
 \text{Fuzzy - Modell :} & \text{Parameter } \mathbf{a} = \max |v_i'| + \frac{b}{2}
 \end{array} \quad (3-9)$$

mit v_i' als einzelne Abweichungen vom Mittel bei insgesamt n Beobachtungen. Analog zum Vergleich mit Referenzdaten basieren Minimum-Maximum-Modell und Fuzzy-Modell im Wesentlichen auf Extremwerte der Abweichungen. Es ergibt sich wiederum bereits bei geringem Stichprobenumfang eine gute Beschreibung der Unsicherheitssituation. Das Stochastische Modell erfordert für eine zuverlässige Bestimmung deutlich mehr Beobachtungen.

Zusätzlich zum Epsilon-Parameter verfügt das Minimum-Maximum-Modell mit der direkten Angabe von Minimum und Maximum über eine alternative Repräsentationsform. Allgemein lässt sich das Maximum auch durch eine Einhüllende um alle Realisierungen festlegen. Das Minimum gleicht bei Punkten und Linien dem Mittel, während bei Flächen eine Art innere Hülle zu beschreiben ist. Eine Schätzung über das Verfahren der Mittelbildung ist möglich, indem auf der Senkrechten zur Mittellinie nach links und nach rechts jeweils die Punkte der maximal entfernten Realisierungen extrahiert werden. Bei Linien legen alle gefundenen Extrempunkte gemeinsam das Maximum fest. Bei Flächen sind die Punkte gemäß den Richtungen zu trennen. Nach außen gehören die Punkte zum Maximum, nach innen zum Minimum.

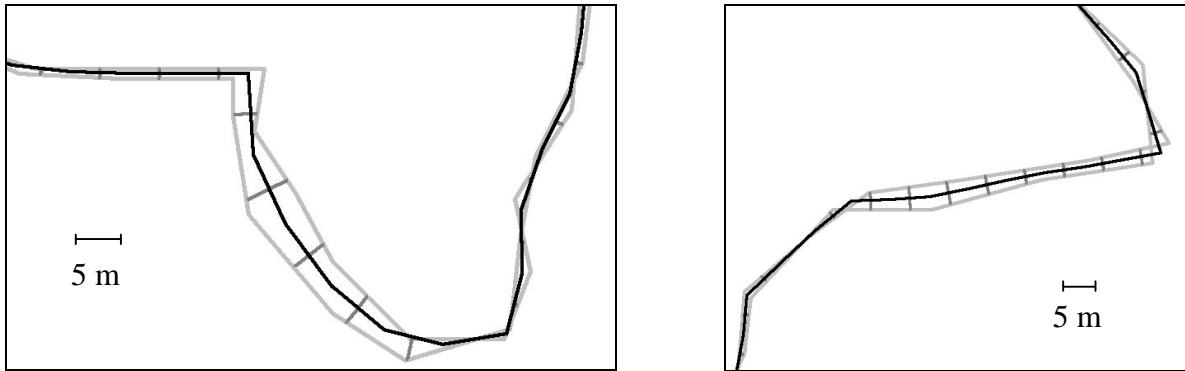


Abb. 3.12: Zwei Beispiele zur Messung der Abweichungen (dunkel) von der mittleren Geometrie (schwarz) zu zwei Realisierungen (hell). Die gezeigten Ausschnitte gehören zum Waldobjekt (Abb. 3.1).

Mit Hilfe der gemessenen Modellparameter gelingt auch die bereits in Kapitel 3.2.1 diskutierte qualitative Glättung der mittleren Geometrie. Die Glättung lässt sich an die konkret vorliegende Unsicherheit so anpassen, dass die geglättete Geometrie nur geringfügig im Rahmen des Variationsbereiches verändert wird und dadurch keine zusätzlichen Auswirkungen auf die Unsicherheit entstehen. Dazu ist der zu wählende Schrankenwert d entsprechend einzustellen. Das Maximum definiert den Bereich, in dem alle möglichen Realisierungen enthalten sind. Interpretiert man die mittlere Geometrie als eine spezielle Realisierung, dann dürfen sich die Änderungen durch die Glättung nur innerhalb des Maximums abspielen. Daher sollte der Schrankenwert maximal zu $d = e$ gewählt werden. Um aber den Einfluss der Glättung auf die geometrische Unsicherheit vernachlässigbar klein zu halten, empfiehlt es sich jedoch, den Wert weiter zu verringern. Ein praktischer Wert kann mit $d = s$ festgelegt werden. Die geglättete Geometrie bleibt damit in einen Bereich, in dem ebenfalls die Mehrheit an Realisierungen (s -Umgebung enthält 68% der Realisierungen) anzutreffen ist.

3.2.3 Diskussion

Obwohl das Schätzen der Unsicherheitsmaße keiner mathematisch strengen Ableitung entspricht, besitzt diese Vorgehensweise in der Praxis eine hohe Bedeutung. Bei den Anwendern liegen oftmals vielfältige und detaillierte Kenntnisse über die Unsicherheitssituation vor, ohne dass diese explizit notiert sind. Durch das Schätzen können solche wertvollen Erfahrungswerte integriert werden. Zunächst lassen sich die Angaben zum Aufbau von Metadaten nutzen, um anschließend in die Bestimmung der Unsicherheitswerte einzufließen. Hingegen wird eine objektspezifische Abschätzung durch dieses Verfahren nur in Ausnahmefällen zu realisieren sein. Sinnvolle Unterscheidungen können häufig nur zwischen Objektarten oder größeren räumlichen Bereichen getroffen werden.

Die Erfassung der Unsicherheitsmaße über den Vergleich mit Referenzdaten liefert das beste Ergebnis. Es werden individuelle Maße für jedes Objekt unter Berücksichtigung aller wirkenden Einflüsse erzeugt. Voraussetzung ist jedoch, dass alle diskutierten Anforderungen an die Referenzdaten erfüllt sind. Unter dieser Bedingung dürfte es allerdings schwer fallen, einen geeigneten Datensatz zu finden, der sich ohne großen Aufwand für einen Vergleich aufbereiten lässt. Das gezeigte Beispiel (ALK als Referenz für ATKIS) gibt einen guten Einblick in die zu beachtenden Schwierigkeiten. Grundsätzlich wirft sich beim Vorliegen eines idealen Vergleichsdatensatzes aber auch die Frage auf, weshalb der zu vergleichende Datensatz überhaupt erzeugt wurde, wenn eine qualitativ bessere Alternative mit gleichen Eigenschaften bereits existiert. Ein realistischer Fall ist die Anwendung des Verfahrens zu Testzwecken, um Erfahrungswerte zu sammeln. Dafür werden lediglich innerhalb von Testgebieten Referenzdaten benötigt, die häufig einfacher zu finden sind. Von den erzielten Erkenntnissen kann anschließend analog zum Schätzverfahren auf das Verhalten aller Objekte geschlossen werden.

Mehrfacherfassungen erlauben ebenfalls eine individuelle Ableitung der Unsicherheitswerte unabhängig für jedes Objekt und bestimmen gleichzeitig eine Schätzung der mittleren Geometrie. Vorteil ist, dass die Werte direkt bei der Dateneingabe ermittelt werden und direkt nach der Erfassung bereitstehen. Über Mehrfacherfassungen erhält man eine Aussage zur inneren Unsicherheit der Objekte. Im Gegensatz zur äußeren Unsicherheit über Referenzdaten werden dabei nicht alle Einflüsse berücksichtigt. Doch spiegeln sich die für Analysezwecke wichtigen relativen Verhältnisse wider. Ein möglicher Hinderungsgrund für den praktischen Einsatz des Verfahrens ist durch den dadurch erheblich steigenden Aufwand an manueller Erfassung gegeben. Jedes Objekt muss mindestens doppelt erfasst werden. Da es sich bei der Erfassung um den kostenintensivsten und aufwendigsten Arbeitsschritt innerhalb jeder Projektbearbeitung han-

delt, sind Erweiterungen der Arbeiten unweigerlich mit Akzeptanzproblemen verbunden. Doch sollte der Gewinn an Informationen den Aufwand rechtfertigen.

3.3 Präsentation der Unsicherheit

Die Präsentationskomponente übernimmt die Aufgabe, die in den Objekten enthaltene Unsicherheit zu visualisieren. In erster Linie dient sie dazu, den Nutzer in anschaulicher und leicht verständlicher Weise über die vorliegenden Größenordnungen der Unsicherheit zu informieren. Graphische Darstellungen sind ein geeignetes Mittel für diesen Zweck. Sie helfen, insbesondere komplexe Sachverhalte verständlich und schnell erfassbar darzulegen.

Die Präsentation der Unsicherheit hat sich zwei unterschiedlichen Aspekten zu widmen. Zum einen kann von Interesse sein, die objektindividuellen Werte der Unsicherheitsparameter, die üblicherweise als Zahlen gespeichert sind, darzustellen. Die Unsicherheitsparameter wirken dabei wie gewöhnliche darstellbare Attribute. Eine solche Präsentation verfolgt das Ziel, den Vergleich der Parameterwerte zwischen den Objekten eines Datensatzes zu ermöglichen. So lassen sich in einfacher Form Qualitätsunterschiede aufzeigen und eventuell vorhandene Defizite aufdecken. Zum anderen kann die Auswirkung der Unsicherheit auf die Ausdehnung des Objektes von Bedeutung sein. Es handelt sich dabei um eine räumliche Betrachtungsweise, die die enthaltene geometrische Variation aufzeigen will. Dies gelingt dadurch, dass die übliche Objektdarstellung um die Visualisierung des Variationsbereiches ergänzt wird. Eine solche Darstellung unterstützt die Interpretation der Daten, indem direkt sichere von unsicheren Bereichen unterschieden werden können. Welche graphischen Möglichkeiten sich für beide Aspekte bieten, wird nachfolgend mit Hilfe von Beispieldatensätzen erläutert.

3.3.1 Unsicherheitsmaße als darstellbares Attribut

Im Normalfall wird zu einem Objekt eine große Anzahl an beschreibenden Informationen in Form von Attributen abgelegt. Besitzt ein Attribut innerhalb einer Anwendung eine herausragende Bedeutung (z.B. Straßentyp für Straßen, Art der Landnutzung für topographische Objekte, Einwohnerzahl für Siedlungen, ...), so ist es sinnvoll, den individuellen Attributwert in die graphische Darstellung einzubringen. Es wird zum darstellbaren Attribut. Dies erfolgt dadurch, dass die Werte die graphische Ausprägung der Objekte steuern. Allgemein wird die Ausprägung durch verschiedene visuelle Variablen festgelegt. Dazu zählen Variationsmöglichkeiten in Farbe, Stärke, Art, Form, Orientierung und Symbolik. Bei einem darstellbaren Attribut wirken die Werte direkt auf die Ausgestaltung einer oder mehrerer visueller Variablen ein, so dass ein bestimmter Attributwert eine bestimmte graphische Ausprägung erzeugt.

Unsicherheitsmaße lassen sich ebenfalls als eine Art beschreibender Attribute auffassen. Dies ermöglicht einen Vergleich der Unsicherheit zwischen verschiedenen Objekten oder ganz generell die Prüfung der im Datensatz vorhandenen Unsicherheitssituation. Der definierte Modellierungsansatz ordnet die Unsicherheitsmaße den Objektlinien zu, so dass nur solche visuelle Variablen geeignet sind, die insbesondere auf die Darstellung von Linien einwirken. Dazu gehören Farbe, Linienstärke und Linienart. Dem kontinuierlichen und nach oben offenen Wertebereich der Unsicherheit kommt die Linienstärke am nächsten. Die Linienstärke kann ebenfalls unendlich viele Werte annehmen und so die Unsicherheitswerte direkt abbilden. Abb. 3.13a zeigt ein Beispiel für eine solche Umsetzung für die Epsilon-Werte der ATKIS-Straßenobjekte (Kapitel 3.2.2). Die Zunahme der Unsicherheit wird hier durch Zunahme der Linienstärke angezeigt. Dünne Linien weisen auf sichere, dicke Linien auf unsichere Geometrien hin. Eine Umkehrung dieser Abbildungsregel ist genauso denkbar. Eine Darstellung mit Hilfe von Farbe gelingt durch Variation der Sättigung eines Farbwertes entsprechend der Unsicherheitswerte (HSV-Farbmodell: Hue = Farbton, Saturation = Sättigung, Value = Helligkeit). Einschränkend können hierbei lediglich 256 Werte unterschieden werden. Es muss deshalb eine Klassenbildung der Unsicherheitswerte durchgeführt werden. Oftmals genügt aber bereits eine sehr viel geringere Anzahl an Farbwerten. Soll beispielsweise die Einhaltung von vorgegebenen Unsicherheitsgrenzwerten überprüft werden, so lassen sich dafür die drei Farben einer Verkehrsampel, rot, gelb und grün einsetzen. Ausgenutzt wird dabei die Vertrautheit der Farben und die intuitive Verknüpfung mit ihrer Bedeutung, die vielen Nutzern aus dem Straßenverkehr geläufig sind. Im Zusammenhang mit der Unsicherheit gilt folgende Interpretation:

- Rot: Unsicherheitsgrenzwert überschritten,
- Gelb: Unsicherheit liegt nahe am Grenzwert, Unsicherheitsgrenzwert knapp eingehalten,
- Grün: Unsicherheitsgrenzwert eingehalten.



Abb. 3.13: Visualisierungen der objektindividuellen Unsicherheitswerte am Beispiel von ATKIS-Straßenobjekten: als darstellbares Attribut mit Hilfe der Linienstärke (dünne Linie = geringe Unsicherheit, dicke Linie = große Unsicherheit) (a) und Visualisierung der Einhaltung der geforderten Lageunsicherheit durch Grauwerte (schwarz: nicht eingehalten (entspricht rot), dunkel: kritischer Bereich (entspricht gelb), hell: eingehalten (entspricht grün)) (b).

Abb. 3.13b zeigt eine Anwendung dieser Visualisierungsmöglichkeit ebenfalls am Beispiel der ATKIS-Straßendaten, wobei die Farben aus Druckgründen durch Grauwerte ersetzt sind. Die ATKIS-Modellierungsvorschrift gibt für Straßen als wichtige Objektklasse einen Grenzwert für die geometrische Unsicherheit in Form einer Standardabweichung mit $s_L = 3\text{ m}$ vor. Für die Darstellung im Beispiel gelten die folgenden Klassengrenzen:

- Rot: $s_L > 3.0\text{ m}$
- Gelb: $2.5\text{ m} \leq s_L \leq 3.0\text{ m}$
- Grün: $s_L < 2.5\text{ m}$

Die Prüfung zeigt, dass nur 3 von 59 Straßen im Untersuchungsgebiet die Forderung nicht einhalten. Zwei davon überschreiten den Grenzwert nur geringfügig ($s_L < 3.5\text{ m}$) und nur ein Objekt besitzt eine nicht zu tolerierende Unsicherheit ($s_L = 6.8\text{ m}$). Die weitere Untersuchung dieses Objektes offenbart, dass die ALK-Daten dort grob von den ATKIS-Daten abweichen, so dass eine Änderung der realen Situation zu vermuten ist (Abb. 3.14). Welche der beiden Datensätze die aktuelle Situation widerspiegelt, kann nicht ohne weiteres abgeleitet werden. Aufgrund seiner permanenten Fortführung sollte im Normalfall jedoch der ALK-Datenbestand aktueller sein.

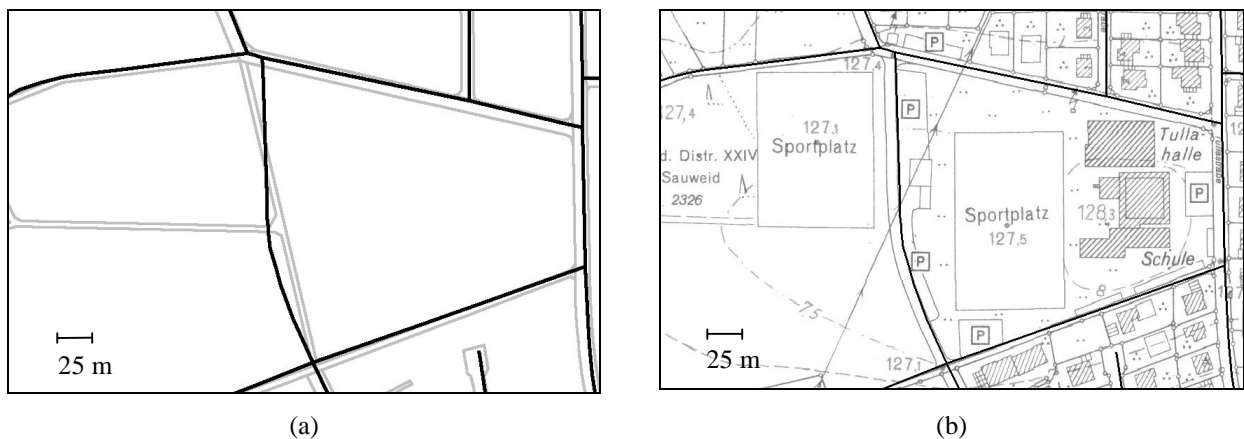


Abb. 3.14: Beispiel für ein Straßenobjekt (schwarz), das deutlich die in ATKIS geforderte Lageunsicherheit überschreitet. Als visuelle Vergleichsinformationen dienen ALK-Daten (hell) (a) bzw. die Katasterkarte (b).

3.3.2 Unsicherheit in der räumlichen Ausdehnung

Der zweite bereits angedeutete Aspekt beschäftigt sich mit der räumlichen Auswirkung der Unsicherheit. Während Attributwerte die Größenordnung der Unsicherheit im Objekt verdeutlichen, lassen sich die Werte nicht unmittelbar nutzen, um die räumliche Ausdehnung des Variationsbereiches darzustellen. Für diesen Zweck sind die Ergebnisse der geometrischen Basisaussage (Position befindet sich innerhalb des Objektes) besser geeignet. Sie liefern für jede beliebige Position im Raum eine Bewertung, mit welcher Sicherheit sich die Stelle innerhalb des Objektes befindet. Jedes mögliche Aussagemass definiert dabei eine zweidimensionale Bewertungsfunktion für jedes Objekt. Um die gesuchte räumliche Auswirkung der Unsicherheit aufzuzeigen, gilt es, die Bewertungsfunktion in geeigneter Weise darzustellen. Dazu gibt es verschiedene Möglichkeiten, die nachfolgend im Einzelnen erläutert werden. Die Reihenfolge orientiert sich an der aufsteigenden räumlichen Dimension der Darstellungen (von eindimensional bis dreidimensional).

On-line-Auskunft (eindimensional)

Die einfachste Art der Darstellung ist die einzelne Ausgabe von Unsicherheitswerten zur geometrischen Basisaussage. Der Nutzer hat dabei die Position, für die eine Bestimmung erfolgen soll, und das betreffende Objekt selbst auszuwählen. Die Festlegung der Position kann beispielsweise interaktiv dadurch erfolgen, indem der Nutzer mit dem Cursor die entsprechende Position anfährt. Anschließend erfolgt die Bewertung anhand der diskutierten Formeln. Auf dieses Vorgehen aufbauend lässt sich eine On-line-Auskunft realisieren, bei der der Cursor beliebig durch den Raum bewegt werden darf und die Bewertung dann stets in dynamischer Weise fortgeführt wird. Ein Beispiel für eine solche On-line-Auskunft zeigt Abb. 3.15. Darin werden alle drei definierten Aussagemasse (Wahrscheinlichkeit p , Möglichkeitswert m , Möglichkeitswert \mathbf{p}) permanent für die aktuelle Cursorposition berechnet.

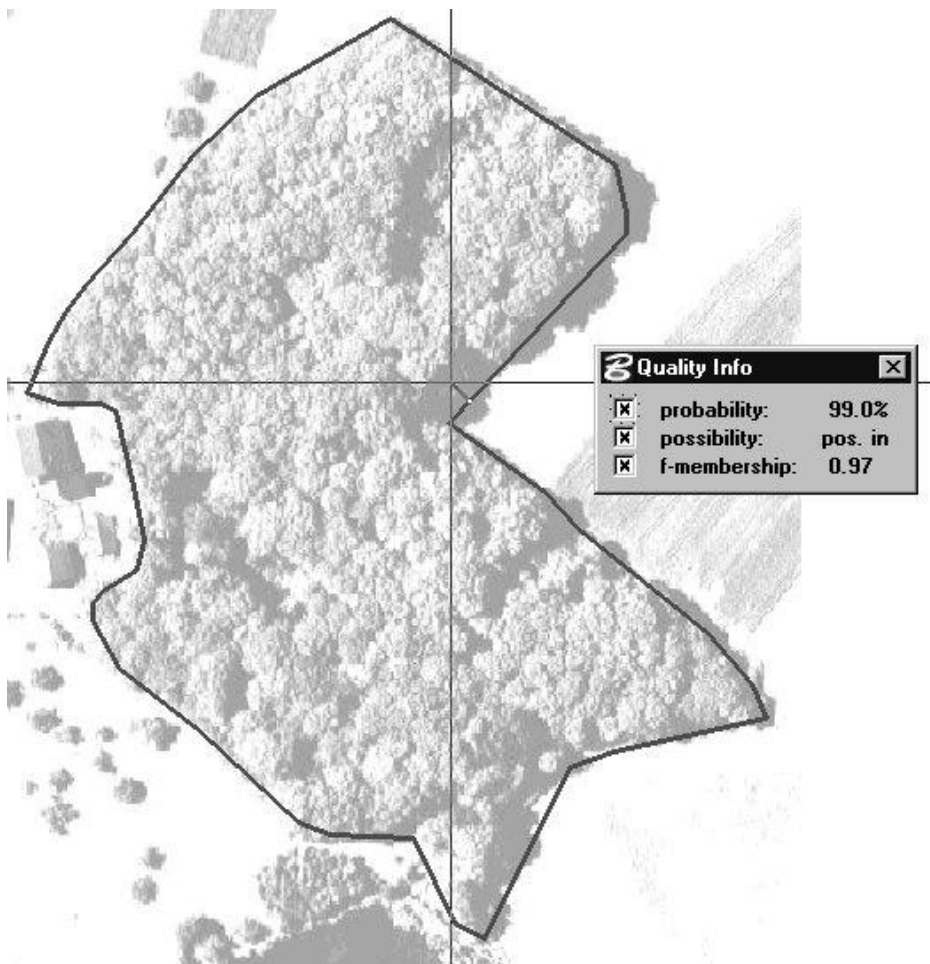


Abb. 3.15: Beispiel für eine On-line-Auskunft über die zugehörigen Werte der Aussagemasse für die sich im Fadenkreuz befindende Position. Zusätzlich eingeblendet ist die Verbindung von der Position zum Objekt, über die die Distanz berechnet wird. Als Hintergrund dient ein Orthophoto.

Entscheidend für die Akzeptanz einer solchen Auskunftsmöglichkeit ist ein schnelles Antwortverhalten des Systems. Der Nutzer erwartet eine Reaktion in Echtzeit, auch dann, wenn die Position dynamisch verändert wird. Nur so lässt sich der Nachteil der Darstellung eines Einzelwertes ausgleichen und die Unsicherheit sinnvoll evaluieren.

Als kritisch im Hinblick auf die Antwortzeit erweist sich die programminterne Bestimmung der Distanz von der betrachteten Position zur Objektgeometrie. Nur im Falle gleicher Variationswerte für alle Linien im Objekt ist die gesuchte Distanz gleich der kürzesten Distanz zum Objekt (Abb. 3.16a). Bei unterschiedlicher Variation der Linien entscheidet die Größe des Unsicherheitswertes in Kombination mit der Objektform zu welcher Linie die Distanz zu ermitteln ist (Abb. 3.16b). Aus der einfachen Bestimmung der kürzesten Distanz wird ein komplexes, fallspezifisches Verfahren, das im Allgemeinen alle Linien in die Untersuchung einschließt.

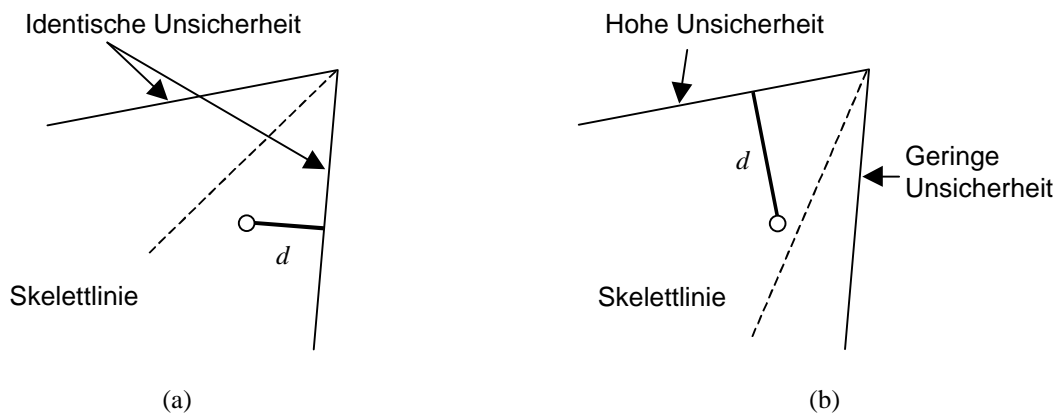


Abb. 3.16: Beispiele zur Problematik der Distanzbestimmung mit kürzester Distanz bei identischer Unsicherheit (a) und fallspezifischer Distanz bei unterschiedlichen Unsicherheiten der Linien (b).

Ein akzeptables Echtzeitverhalten kann erreicht werden, wenn das Problem der Liniensuche vorab für alle Positionen gelöst wird. Dazu ist in einem Vorverarbeitungsschritt für jede Objektlinie ein räumlicher Bereich zu bestimmen, in dem die Linie für die Distanzberechnung verantwortlich ist. Es ergibt sich eine Zerlegung des Objektes in einzelne Segmente, wobei jede Objektlinie genau einem Segment zugehört. Abb. 3.17 vergleicht die Form der Segmente im Inneren eines Beispielobjektes für gleiche Variation (Abb. 3.17a) und für unterschiedliche Variation der Objektlinien (Abb. 3.17b). Es treten Unterschiede in Größe, Form und Nachbarschaften der Segmente hervor, die verdeutlichen, welche komplexe Problematik das Liniensuchproblem beinhaltet. Aus den Darstellungen ist zu erkennen, dass die Randlinien der Segmente das Skelett des Objektes definieren. Somit kann die Segmentbestimmung durch eine Skelettierung gelöst werden.

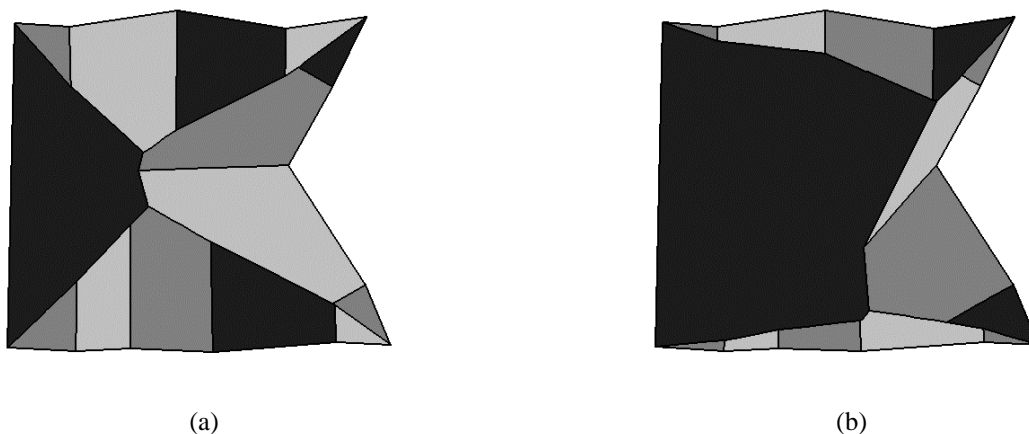


Abb. 3.17: Segmentierung des Objektinneren für identische (a) und unterschiedliche Unsicherheiten der Linien (b).

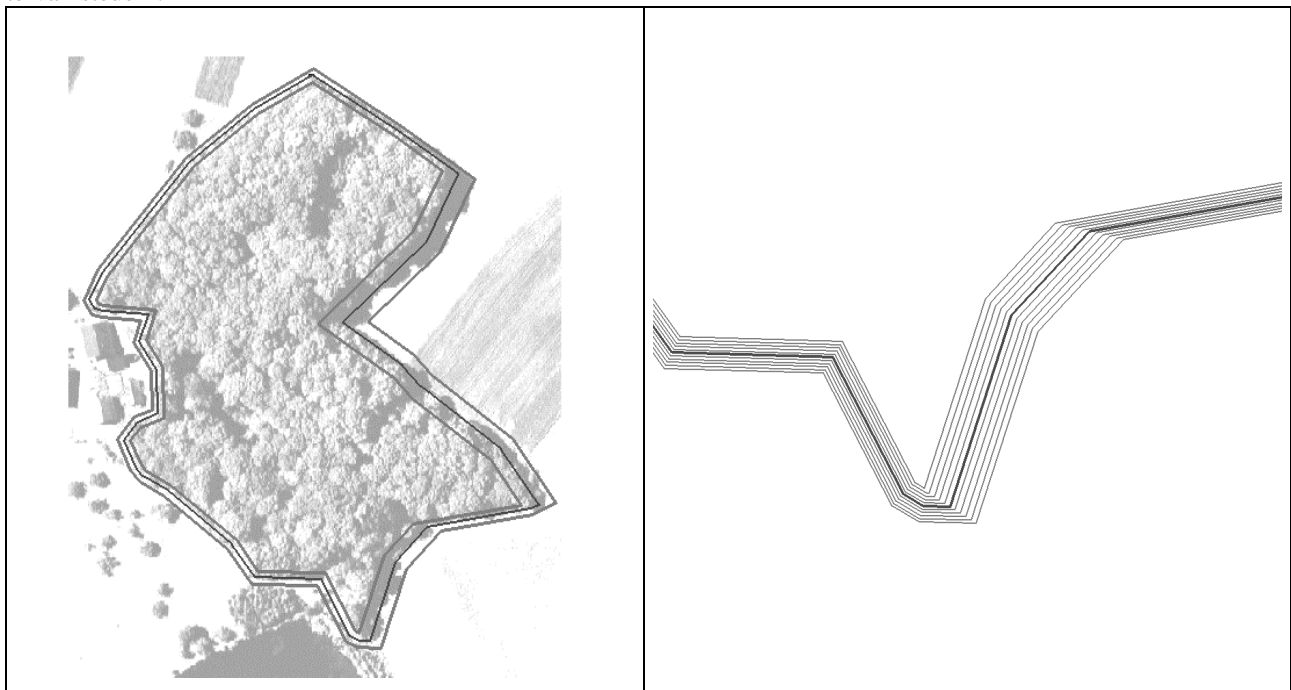
Eine anschauliche Erläuterung der Skelettierung vergleicht das Verfahren mit einer Überschwemmung, bei der sich Wasserfronten gleichzeitig von allen Objektlinien ausbreiten. Die Stellen, an denen Wasserfronten zusammentreffen, bilden die Skelettlinien. Es sind einige Algorithmen bekannt (z.B. Lee, 1982, Yang und Gold, 1996), die eine Skelettierung durchführen. Dabei wird grundsätzlich davon ausgegangen, dass sich die Wasserfronten mit gleicher Geschwin-

digkeit bewegen. Im Zusammenhang mit der Unsicherheit entspricht dies dem Fall gleicher Variationen in den Linien. Unterschiedliche Variation bedingen dagegen unterschiedliche Ausbreitungsgeschwindigkeiten. Dabei gilt folgender Zusammenhang: je kleiner die Unsicherheit ist, desto langsamer bewegt sich das Wasser und umgekehrt. Die Lösung einer solchen Aufgabe erfordert eine modifizierte Skelettbildung. Die Unterschiede in den Linien werden dabei in Form von Gewichten berücksichtigt, wobei die Gewichte umgekehrt proportional zu den Unsicherheitswerten anzusetzen sind. Die Gewichte bewirken, dass eine zwischen zwei Objektlinien ursprünglich als Winkelhalbierende definierte Skelettlinie sich in Richtung der sicheren Linie verändert (Abb. 3.16b). Das Ergebnis dieser modifizierten Skelettbestimmung ist in Abb. 3.17b zu sehen, das über geometrische Schnitte der gewichteten Winkelhalbierenden erzeugt wurde. Die Vorgehensweise kann auch auf das Äußere einer Fläche angewendet werden. Entsprechende allgemeine Ableitungen werden ausführlicher in Anhang C.1 behandelt.

Auf der Basis des Skeletts können die gesuchten Segmente durch eine Polygonbildung abgeleitet werden. Die Segmentierung ist für jedes Objekt nur einmal auszuführen. Sie kann anschließend für die Unsicherheitsbestimmung beliebiger Positionen genutzt werden. Dazu ist zuerst festzustellen, in welchem Segment sich die aktuelle Position befindet, um danach die kürzeste Distanz zur zugehörigen Objektlinie zu berechnen. Anschließend können die diskutierten Bewertungsformel angewendet werden, die das gesuchte Ergebnis liefern.

Isoliniendarstellung (zweidimensional)

Die Bewertungsfunktion der Objektzugehörigkeit beschreibt die Unsicherheit in Abhängigkeit von der Position im Raum. Wie andere thematische Attribute kann die Funktion ebenfalls als ein räumliches Phänomen aufgefasst und in ähnlicher Weise behandelt werden. Für eine Betrachtungsweise eignet sich in diesem Fall das Feldmodell besser als das Objektmodell (Kapitel 1.2.1 und Anhang A.1), da ein einzelnes, kontinuierliches Phänomen beobachtet wird, das zudem stark variiert. Die Repräsentation von Feldern kann sowohl in Vektor- wie auch in Rasterform erfolgen. Ein Vertreter der Vektorform sind Isoliniendarstellungen. Dabei werden Linien gleicher Funktionswerte (hier: Isolinien gleicher Unsicherheit) gebildet und durch Vektorelemente repräsentiert. Die Informationsdichte lässt sich durch das Isolinienintervall steuern.



(a)

(b)

Abb. 3.18: Isoliniendarstellung für ein Waldobjekt im Minimum-Maximum-Modell (a) und für einen Ausschnitt im Fuzzy-Modell (b) mit unterschiedlichen Unsicherheiten innerhalb der Objekte.

Isolinien eignen sich insbesondere für diskrete Funktionswerte, da sich genau eine Isolinie pro Wert ergibt. Probleme bereiten dagegen kontinuierliche Werte, für die theoretisch unendlich viele Isolinien zu bilden sind, um eine genaue Abbildung des Phänomens zu erreichen. Die notwendige Einschränkung auf eine endliche Anzahl erfordert die zusätzliche Festlegung einer Interpolationsmethode (z.B. lineare Interpolation), damit auch Werte zwischen den Isolinien zu bestimmen sind. Von den Aussagemäßen weist nur das Möglichkeitsmaß m einen diskreten Wertebereich auf, so dass

sich die Isoliniendarstellung besonders für die Visualisierung der Unsicherheit im Minimum-Maximum-Modell eignet. Zwar definieren die Möglichkeitswerte im Allgemeinen jeweils einen flächenförmigen Bereich (nämlich entsprechend der Lage im Minimum, Mittel oder Maximum) und damit eine Isofläche, doch können die Ränder der Flächen, die den Sprungstellen der Bewertungsfunktion entsprechen, als die gesuchten Isolinien aufgefasst werden. Es ergibt sich eine Banddarstellung um die mittlere Geometrie des Objektes (Abb. 3.18a). Unterschiedliche Bandbreiten spiegeln unterschiedliche Unsicherheiten in der räumlichen Ausdehnung wider. Aber auch für den stochastischen Ansatz und das Fuzzy-Modell kann mit Hilfe der Isoliniendarstellung eine aussagekräftige Visualisierung der Unsicherheit erzeugt werden (Abb. 3.18b). Entscheidend ist jedoch, dass das Isolinenintervall in angemessener Weise gewählt wird. In Anhang C.1 werden Möglichkeiten diskutiert, wie Isolinien bei unterschiedlichen Unsicherheitsmaßen der adjazenten Objektlinien zu bilden sind.

Matrixdarstellung (zweidimensional)

Eine alternative Möglichkeit bietet sich durch die Repräsentation der Unsicherheit in Rasterform. Dazu ist für jedes Objekt eine Rastermatrix zu definieren, die eine Diskretisierung der zugehörigen räumlich kontinuierlichen Bewertungsfunktion erzeugt. Zur Festlegung der Rastermatrix werden die Position, die Ausdehnung und die Zellgröße (Auflösung) benötigt. Position und Ausdehnung sind so zu wählen, dass die gesamte Objektfläche einschließlich des Unsicherheitsbereiches überdeckt wird. Um eine anschauliche Darstellung zu erzielen, hat sich die Zellgröße an der zugehörigen Unsicherheit zu orientieren. Grundsätzlich gilt dabei, je kleiner die Unsicherheit ist, desto kleiner muss auch die Zellgröße ausfallen. Große Unsicherheiten erlauben dagegen eine gröbere Darstellung, so dass bereits visuell die enthaltene Größenordnung sichtbar wird. Besondere Anforderung an die Zellgröße stellt das Ziel, eine Rekonstruktion der ursprünglichen Bewertungsfunktion aus der Rastermatrix zu ermöglichen. Dazu ist das Abtasttheorem (Stearns, 1979) zu beachten, das für die Zellgröße Δr fordert:

$$\Delta r < \frac{1}{2f_{max}}, \quad (3-10)$$

mit f_{max} als maximale, innerhalb der Bewertungsfunktion enthaltene Frequenz (Nyquistfrequenz). Für eine Standardnormalverteilung ist die maximale Frequenz mit Hilfe einer Fourier-Transformation (Brigham, 1987) als $f_{max} \approx 0.977$ zu bestimmen, woraus sich die folgende Festlegung für die optimale Zellgröße ableiten lässt:

$$\Delta r_{stochastisch} < 0.5 \mathbf{s}, \quad (3-11)$$

mit Standardabweichung \mathbf{s} als Maß der geometrischen Unsicherheit im Stochastischen Modell. Durch die linearen Ansätze der Bewertungsfunktionen in den anderen Modellen genügen dort die Angabe von zwei Stellen, um eine Rekonstruktion zu ermöglichen. Die zugehörigen Zellgrößen lauten:

$$\begin{aligned} \Delta r_{min/max} &< 0.5 \mathbf{e} \\ \Delta r_{fuzzy} &< 0.5 (a - b / 2), \end{aligned} \quad (3-12)$$

mit den Parametern Epsilon \mathbf{e} im Minimum-Maximum-Modell bzw. a und b im Fuzzy-Modell. Die auf der Basis des Abtasttheorems abgeleiteten Zellgrößen definieren obere Grenzwerte, die nicht überschritten werden dürfen. Um eine besonders anschauliche Visualisierung zu erreichen, sollten die letztendlich gewählten Werte aber deutlich unter den ermittelten Grenzen liegen. Unter Beachtung der möglichen Übergänge zwischen den Unsicherheitsparametern kann folgende Empfehlung für die praktische Festlegung der Zellgröße gegeben werden:

$$\begin{aligned} \Delta r_{stochastisch} &= 0.25 \mathbf{s} \\ \Delta r_{min/max} &= 0.1 \mathbf{e} \\ \Delta r_{fuzzy} &= 0.1 (a - b / 2). \end{aligned} \quad (3-13)$$

Alle Matrixdarstellungen im Rahmen der Arbeit stützen sich auf diese Festlegung.

Jede Rasterzelle umfasst einen flächenförmigen Bereich, für den eine konstante Zugehörigkeit zum Objekt angenommen wird. Als Position für die Wertermittlung ist der Mittelpunkt der Zelle heranzuziehen. Die einzelnen Werte ergeben sich dann durch Anwendung der Formeln zur Bestimmung der Aussagemaße der geometrischen Grundaussage (Kapitel 2.4). Für Darstellungszwecke sind die originären Matrixwerte in Grau- oder Farbwerte umzuwandeln. Beispielsweise lässt sich der für die graphische Darstellung der Objektgeometrie genutzte Farbton verwenden, um die Unsicherheit durch die Variation in der Sättigung darzustellen. Das Ergebnis eines solchen Verfahrens ist für ein linienförmiges Objekt in Abb. 3.19 zu sehen. Ein intensiver Farbton weist in diesem Fall (Stochastisches Modell) auf eine hohe, ein schwacher Farbton auf eine niedrige Wahrscheinlichkeit hin. Das Beispiel verdeutlicht zusätzlich, welche Differenzen entstehen, wenn die Unsicherheiten in Längsrichtung der Linie vernachlässigt (Abb. 3.19a) oder mit be-

rücksichtigt (Abb. 3.19b) werden. Durch Mitberücksichtigung fallen die Wahrscheinlichkeiten in den Regionen um die Endpunkte kleiner aus, so dass die Darstellung dort entsprechend heller wird.

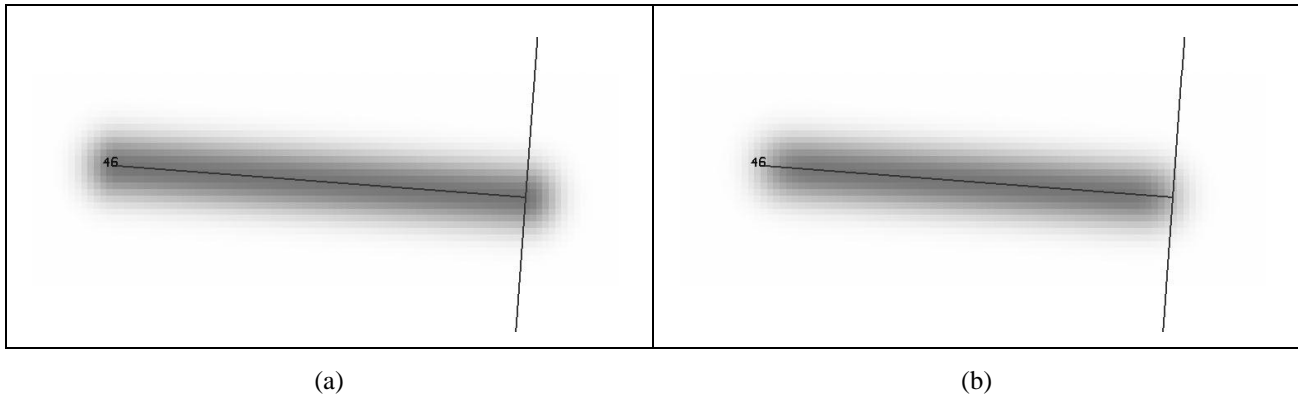


Abb. 3.19: Matrixdarstellung der Unsicherheit eines Straßenobjektes im Stochastischen Modell, ohne (a) und mit Berücksichtigung des Einflusses längs der Linienrichtung (b).

Ein zweites Beispiel (Abb. 3.20) zeigt den Einfluss der Objektbreite auf die Berechnung der Wahrscheinlichkeiten. Im ersten Fall ist die Breite bekannt und wird entsprechend berücksichtigt (Abb. 3.20a), während im zweiten Fall die Breite als unbekannt gilt und somit die Standardbreite einzusetzen ist (Abb. 3.20b). Die sehr blasse Darstellung im zweiten Fall verdeutlicht, dass eine Linie mit geringer Breite eine große Unsicherheit im Bezug auf die räumliche Ausdehnung beinhaltet.

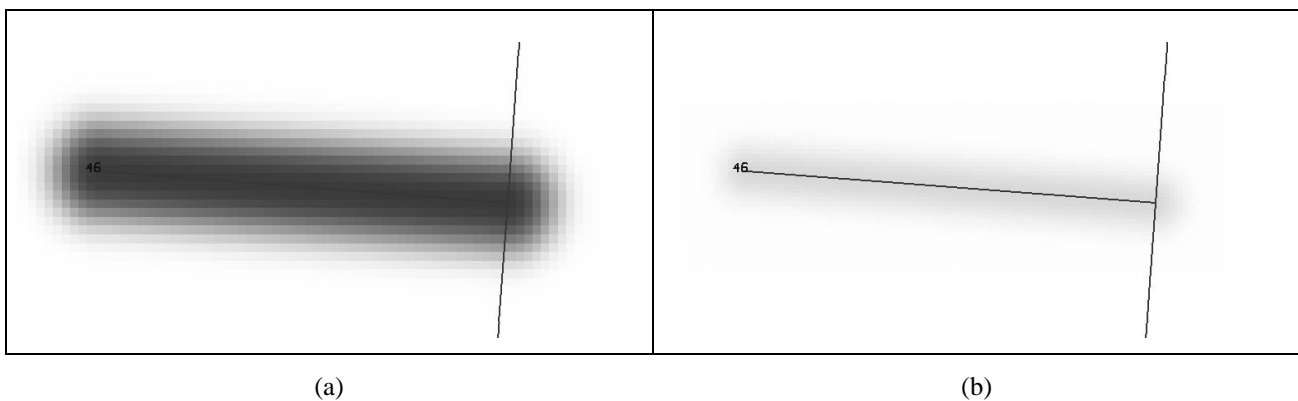


Abb. 3.20: Matrixdarstellung der Unsicherheit eines Straßenobjektes im Stochastischen Modell mit (a) und ohne Angabe der Breite (b).

Funktionsdarstellung (dreidimensional)

Neben der Interpretation als ein thematisches Attribut im Feldmodell kann die Unsicherheit auch als gleichberechtigte dritte Koordinate angesehen werden. Das führt zu einer dreidimensionalen Betrachtungsweise der Bewertungsfunktion und damit zu einer plastischen Visualisierung der Unsicherheit. Abb. 3.21 zeigt anhand eines Beispiels eine Perspektivansicht der Bewertungsfunktion eines linienförmigen Objektes. Mit zunehmender Höhe steigt die Zugehörigkeit zum Objekt. Zusätzlich ist die Darstellung mit einer zweidimensionalen Matrixdarstellung der Unsicherheit überlagert. Die Anwendung dieser Darstellungsmöglichkeit setzt voraus, dass das eingesetzte System in der Lage ist, die dritte Dimension zu berücksichtigen.

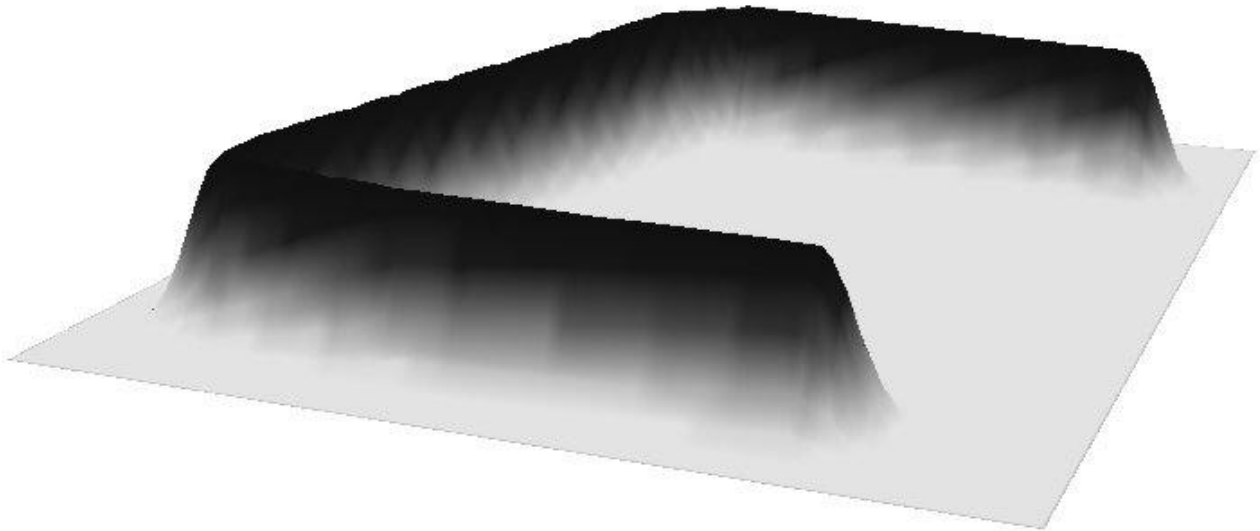


Abb. 3.21: 3-D Perspektivansicht der Unsicherheit eines linienförmigen Objektes mit Überlagerung der zugehörigen Matrixdarstellung.

3.3.3 Diskussion

Die Beispiele zeigen, dass eine anschauliche Darstellung der Unsicherheitssituation in vielfältiger Weise unter Nutzung der graphischen Möglichkeiten moderner Systeme gelingt. Sie decken aber auch einige Problembereiche auf. Die Unsicherheit hat aus Nutzersicht den Stellenwert einer Zusatzinformation zu den Objekten. Sie entspricht damit einer Art zweiten Thematik, die der ersten Thematik, der eigentlichen Objektinformation, überlagert wird. Die eigene graphische Ausprägung der Objektdaten führt zu gegenseitigen Überdeckungen der beiden Thematiken (Unsicherheit überdeckt Objektgeometrie und umgekehrt). Zusätzlich leidet die Übersichtlichkeit aufgrund der gestiegenen Menge an Informationen, die gemeinsam zu sehen sind. Daher ist die Darstellung der Unsicherheit nur als eine kurzzeitige Hilfe anzusehen, die bei Bedarf, z.B. zu Überprüfungszwecken, eingeblendet wird. Bei der gewöhnlichen Projektarbeit sollte darauf verzichtet werden, damit die eigentliche Thematik im Vordergrund bleibt. Die Interpretation der Unsicherheitsdarstellungen kann erleichtert werden, indem man die graphische Ausprägung an die Darstellung der Objekte anpasst. Beispielsweise gelingt dies durch Nutzung des gleichen Farbtons, so dass eine direkte Zuordnung gegeben ist.

Aufgrund der objektweisen Modellierung erfolgt die Darstellung der Unsicherheit ebenfalls unabhängig für jedes Objekt. Da sich die Unsicherheitsbereiche benachbarter Objekte häufig überlappen, entstehen Überlagerungen und Verdeckungen, wenn die Unsicherheitssituation gleichzeitig für mehrere Objekte betrachtet werden soll. Besonders die flächenhafte Matrixdarstellung ist davon betroffen. Als Lösungsansatz kann festgelegt werden, dass in den Überlappungsbereichen immer nur der höchste aller Unsicherheitswerte dargestellt wird. Auf diese Weise wird ein Gesamtbild der Unsicherheit über alle Objekte erzeugt, das keine Verdeckungen mehr enthält, jedoch nicht mehr für Einzelobjekte interpretiert werden darf (Abb. 3.30b).

3.4 Verwaltung der Unsicherheit

Um die geometrische Unsicherheit in die Verarbeitungsprozesse eines GIS zu integrieren, müssen die Unsicherheitsmaße zusammen mit den Objektdaten gespeichert und im System verfügbar gehalten werden. Dazu ist ein Verwaltungskonzept zu entwickeln, das die Umsetzung eines der Datenmodelle (Kapitel 2.5) realisiert. Generell eröffnen sich zwei Ansatzpunkte dazu. Begründet durch die Wirkung der Unsicherheit auf die räumliche Ausdehnung kann einerseits eine geometrische Modellierung eingesetzt werden. Andererseits entsprechen die Unsicherheitsmaße beschreibenden Informationen zum Objekt, so dass eine Behandlung als Attribute ebenfalls in Frage kommt. Beide Möglichkeiten werden nachfolgend erläutert.

3.4.1 Geometrische Verwaltung

Basis einer geometrischen Verwaltung bildet die Auswirkung der geometrischen Unsicherheit auf die räumliche Ausdehnung eines Objektes. Ziel ist es dabei, die Variation mit Hilfe von geometrischen Elementen auszudrücken. Die bisher einzige Objektgeometrie wird dadurch um weitere Geometriebeschreibungen ergänzt. Hinzu kommt die mittlere Geometrie (häufig identisch zur Objektgeometrie) und verschiedene Variationsgeometrien. Analog zu den vorgestellten Darstellungskonzepten lassen sich Isolinien oder Rastermatrizen als Variationsgeometrien einsetzen (Kapitel 3.3.2). Das Minimum-Maximum-Modell benötigt zwei Isolinien, das Minimum und das Maximum. Im Stochastischen Modell und im Fuzzy-Modell sind allgemein unendliche viele Isolinien möglich, so dass zur Verwaltung eine begrenzte Anzahl auszuwählen ist. Als Alternative kann eine Rastermatrix pro Objekt mit den Werten der entsprechenden Aussagemäßen aufgebaut werden. Isolinien besitzen den Vorteil, dass sie passend zur Objektgeometrie Vektordaten ergeben, während Rastermatrizen als Rasterdaten eine hybride Systemumgebung erfordern.

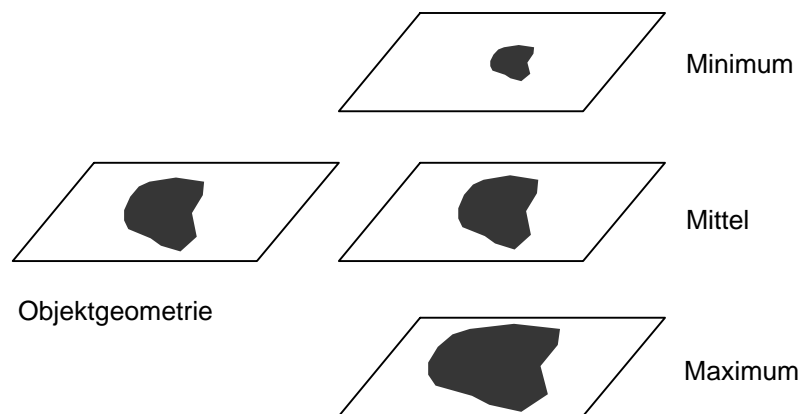


Abb. 3.22: Ebenenverwaltung der verschiedenen Geometriebeschreibungen.

Die verschiedenen Geometrien (Objektgeometrie, mittlere Geometrie, Variationsgeometrien) sind deutlich voneinander zu trennen, um mögliche Überschneidungen und Verwechslungen zu vermeiden. Dies kann beispielsweise dadurch erreicht werden, dass die zugehörigen geometrischen Primitive auf unterschiedliche Ebenen (Layer) aufgeteilt werden (Abb. 3.22). Die Möglichkeit der Ebenenverwaltungen wurde in den Anfangszeiten der Systementwicklungen benutzt, um eine thematische Modellierung zu realisieren und ist daher noch in vielen Systemen verfügbar. Wichtig ist, dass über die Ebenen hinweg der Objektzusammenhang in allen Geometrien gewahrt bleibt. Dazu sind alle zusammengehörenden Elemente über Objektverweise (z.B. die eindeutige Objekt-ID) zu verbinden.

3.4.2 Attributverwaltung

Die Unsicherheitsmaße können als zusätzliche beschreibende Informationen zum Objekt betrachtet werden und lassen sich deshalb wie gewöhnliche Attribute repräsentieren (Drummond und Ramlal, 1992). Sie besitzen einfache numerische Werte, die der Ratioskala zugehören und damit durch den Datentyp „float“ zu beschreiben sind. Wie für andere Attribute bietet die Tabellenform ein geeignetes Verwaltungsschema (z.B. innerhalb einer relationalen Datenbank). Da jede Objektlinie einen Satz an eigenen Parametern besitzt, ist pro Linie genau eine Zeile vorzusehen. Ein Beispiel für eine Umsetzung in einem kommerziellen GIS ist Abb. 3.23 zu entnehmen.

Shape	LID	OID	Varianz	Epsilon	Fuzzy-a	Erweite
PolyLine	1	1001	4.82	3.30	5.80	3.00
PolyLine	2	1001	6.90	4.70	9.50	4.00
PolyLine	3	1001	4.69	5.90	5.67	3.00
PolyLine	4	1001	1.20	3.30	6.20	4.50
PolyLine	1	1002	1.15	2.10	8.54	7.25
PolyLine	2	1002	0.01	0.40	9.10	9.00
PolyLine	3	1002	9.93	9.30	12.45	8.00
PolyLine	4	1002	5.67	6.50	12.19	8.75

Abb. 3.23: Verwaltung der Unsicherheitsmaße in Tabellenform mit zusammengesetztem Zugriffsschlüssel bestehend aus Identifikationsnummern der Linien (LID) und der Objekte (OID).

3.4.3 Diskussion

Ein wichtiger Vorteil der geometrischen Verwaltung besteht in der Möglichkeit, Unsicherheitsdarstellungen schnell zu erzeugen. Die dazu notwendigen Elemente (z.B. für eine Isoliniendarstellung) sind integraler Bestandteil und brauchen nicht zusätzlich konstruiert werden. Die Darstellungsaufgabe reduziert sich auf ein einfaches Ein- und Ausblenden der betreffenden Ebenen. Eine getrennte Verwaltung von Objektgeometrie und mittlerer Geometrie lässt sich ohne weiteren Aufwand realisieren. Besonders geeignet ist dieser Ansatz in Verbindung mit dem Minimum-Maximum-Modell, weil darin eine feste Anzahl an zu verwaltenden Geometrien (nämlich genau vier: Objektgeometrie, Mittel, Minimum und Maximum) vorliegt. Dagegen muss für das Stochastische Modell und für das Fuzzy-Modell im Allgemeinen eine Auswahl aus den theoretisch unendlich vielen Geometrien getroffen werden. Insofern kann die geometrische Verwaltung hier nur eine Näherungslösung für die Beschreibung leisten. Aufgrund des geometrischen Ansatzes sind die Unsicherheitsmaße nur implizit in den Variationsgeometrien enthalten. Werden direkt die Maße benötigt, wie z.B. bei einigen Verfahren zur Fortpflanzung der Unsicherheit innerhalb von Analysen, so müssen die Werte aus den vorhandenen Geometrien abgeleitet werden. Dazu sind Prozesse (z.B. Rekonstruktionsprozess der Bewertungsfunktion) notwendig, die teilweise aufwendig und schwierig zu realisieren sind.

Aufgrund der einfachen Datentypen, mit denen die Unsicherheitsmaße beschrieben werden können, lässt sich die Attributverwaltung in einfacher Weise umsetzen. Ein zugehöriges Schema (z.B. Tabellenform) steht normalerweise in jeder gängigen Datenbank dafür zur Verfügung. Probleme können dadurch entstehen, dass zu jeder Linie ein Satz an Attributen abgelegt werden muss. Üblicherweise werden in der Datenmodellierung die Attribute aber direkt dem Objekt zugewiesen, wobei ein Attribut maximal einen Wert tragen darf. Liegen mehrere verschiedene Werte vor, wird zumeist eine Objektteilung erforderlich. Da ein Objekt im Allgemeinen aus mehreren Linien besteht, sind die Unsicherheitsattribute gleich mehrmals pro Objekt zu verwalten, jedoch ohne dass eine Objektteilung erfolgt. Dies ist nur dann möglich, wenn das eingesetzte System so offen gestaltet ist, dass die Modellierung erweitert werden kann. In diesem Fall ist das Problem durch Einfügen einer Zwischentabelle zu lösen, die gleichzeitig den Linien- und den Objektbezug beinhaltet (Abb. 3.23). Im Gegensatz zur geometrischen Verwaltung sind im attributiven Ansatz alle Unsicherheitsdarstellungen stets neu zu erzeugen. Auch wenn es sich dabei um einen automatischen Prozess handelt, so kann sich der erforderliche Rechenaufwand dann besonders negativ auswirken, wenn solche Darstellungen häufig benötigt werden. Weiterhin ist innerhalb der Attributverwaltung keine zusätzliche geometrische Beschreibung vorgesehen, so dass angenommen wird, dass Objektgeometrie und mittlere Geometrie sich entsprechen. Abweichende Repräsentationen beider Geometrien können nicht ohne weiteres realisiert werden.

Neue Möglichkeiten der Verwaltung eröffnen objektorientierte Datenbanken (z.B. Objectivity/DB, (Objectivity, 2000)), auf denen das entwickelte Datenmodell (Kapitel 2.5) direkt implementiert werden kann. Die konkrete interne Ausführung der Verwaltung wird von der Datenbank geregelt und bewusst vor dem Nutzer verborgen gehalten. Dennoch wird dem Nutzer auch hier die Entscheidung nicht abgenommen, ob die Verwaltung attributiv oder geometrisch erfolgen soll, denn beide Varianten sind im Datenmodell offengehalten. Von der Entscheidung hängen im Weiteren die auf den Objekten zu implementierenden Methoden ab.

3.5 Integration der Unsicherheit in die Analyse

Aufgabe der Analysekomponente ist es, die gesammelten Daten so zu verarbeiten, dass bisher nur implizit in den Daten enthaltene Zusammenhänge aufgedeckt werden können, woraus sich neue Informationen und Interpretationen ableiten lassen. Aufgrund der Unsicherheit in den Eingangsdaten werden stets auch das Ergebnis einer Analyse und die daraus zu schließenden Erkenntnisse zu einem gewissen Grad unsicher sein. Die Übertragung der Unsicherheit kann als Fortpflanzung der Unsicherheit bezeichnet werden. Neben der Größe der Unsicherheit in den Eingangsdaten hängt die Fortpflanzung entscheidend von der eingesetzten Methode ab. Während bei einfachen Methoden noch eine näherungsweise Abschätzung der Ergebnisunsicherheit auf der Basis von Fachwissen gelingt, erfordern die heutzutage verfügbaren komplexen Methoden eine automatische Bestimmung innerhalb des eigentlichen Verarbeitungsprozesses. Folglich sind sämtliche Methoden um eine Fortpflanzungskomponente zu erweitern. Grundlage dafür bilden die Unsicherheitsmaße der Eingangsdaten. Sie sind in der Fortpflanzung so zu verarbeiten, dass am Ende wiederum Werte der Unsicherheitsmaße für das Ergebnis erzeugt werden.

Die Arbeit konzentriert sich ausschließlich auf geometrische Analysemethoden, d.h. solche Methoden, in denen die Verarbeitung der Geometrie im Vordergrund steht und die daher in besonderem Maße von der geometrischen Unsicherheit beeinflusst werden. Dazu zählen eine Reihe von Messfunktionen, wie die Längenberechnung einer Linie, die Bestimmung des Flächeninhalts oder die Distanzmessung zwischen zwei Objekten. Als wichtigste Methode im geometrischen Bereich gilt die Verschneidung, durch die sich räumlich überlagernde Bereiche von Eingangsobjekten extrahiert werden, um daraus neue Objekte zu bilden, die die vereinigten Eigenschaften besitzen. Für die genannten Analysen

werden nachfolgend Verfahren entwickelt, die eine Fortpflanzung der Unsicherheit ermöglichen. Aufgrund der verschiedenen Ansätze unterscheiden sich die Vorgehensweisen je nach eingesetztem Unsicherheitsmodell, so dass für alle drei Modelle jeweils ein gesondertes Verfahren bereitgestellt wird.

3.5.1 Bestimmung der Linienlänge

Die Länge gehört zu den beschreibenden Attributen eines geometrischen Elementes. Während Punkte aufgrund ihrer Definition die feste Länge Null besitzen, ist die Länge von Linien beliebig variabel, ebenso wie die Länge der Umringslinie bei Flächen. Allgemein ergibt sich die Länge l für eine einzelne Linie aus den Koordinaten von Anfangs- und Endpunkt (A, E) zu:

$$l = \sqrt{(x_E - x_A)^2 + (y_E - y_A)^2}, \quad (3-14)$$

für den Fall einer geradlinigen Verbindung. Rotiert man die Abszissenachse des Koordinatensystems in der Weise, dass sie parallel zur Linie verläuft, dann kann die Längenberechnung als ein eindimensionales Problem gelöst werden, wodurch sich (3-14) zu

$$l = l_E - l_A \quad (3-15)$$

vereinfacht, mit l_A und l_E als transformierte Abszissenwerte von Anfangs- bzw. Endpunkt. Setzt sich ein geometrisches Element aus n Liniestücken zusammen, so ist deren Gesamtlänge l_{gesamt} als Summe der Einzellängen l_i zu bestimmen:

$$l_{gesamt} = \sum_{i=1}^n l_i = \sum_{i=1}^n (l_A - l_E)_i. \quad (3-16)$$

Die geometrische Unsicherheit findet sich in den Koordinatenwerten des Anfangs- und Endpunktes wieder, so dass auch die berechnete Länge als unsicher aufzufassen ist. Die in dieser Arbeit bevorzugte linienbezogene Betrachtung der Unsicherheit beschreibt direkt die Unsicherheit der Abszissenwerte (l_A, l_E) im transformierten Koordinatensystem, während ein punktbezogener Ansatz sich auf die Originalkoordinaten (x_A, x_E, y_A, y_E) bezieht. Daher genügt es von der vereinfachten Längenfunktion auszugehen. Im Folgenden werden die zur Fortpflanzung der Unsicherheit notwendigen Verfahren für die einzelnen Unsicherheitsmodelle erläutert.

Stochastisches Modell

Basis im Stochastischen Modell bildet die allgemeine Varianzfortpflanzung (Anhang B.1), die den funktionalen Zusammenhang zwischen Eingangs- und Ergebnisgrößen benutzt, um eine Schätzung der Varianz des Ergebnisses zu erzielen. Angewendet auf die Längenberechnung ergibt sich für die gesuchte Varianz der Linienlänge:

$$\begin{aligned} \mathbf{s}_l^2 &= \mathbf{D}(\mathbf{l}) = \begin{bmatrix} -1 & 1 \end{bmatrix} \begin{bmatrix} \mathbf{s}_{l_A}^2 & \mathbf{s}_{l_A l_E} \\ \mathbf{s}_{l_E l_A} & \mathbf{s}_{l_E}^2 \end{bmatrix} \begin{bmatrix} -1 \\ 1 \end{bmatrix} \\ &= \mathbf{s}_{l_A}^2 + \mathbf{s}_{l_E}^2 - 2\mathbf{s}_{l_A l_E}. \end{aligned} \quad (3-17)$$

Im linienbezogenen Ansatz ist für die Unsicherheit der Abszissenwerte die Varianz $\mathbf{s}_{l_0}^2$ zu setzen, die für alle Linienelemente in gleicher Weise Gültigkeit besitzt:

$$\mathbf{s}_{l_A}^2 = \mathbf{s}_{l_E}^2 = \mathbf{s}_{l_0}^2. \quad (3-18)$$

Unter der zusätzlichen Annahme der Unabhängigkeit von Anfangs- und Endpunkt ergibt sich die vereinfachte Form für die Varianz der Linie zu:

$$\mathbf{s}_l^2 = 2\mathbf{s}_{l_0}^2. \quad (3-19)$$

Im Falle einer aus n Stücken zusammengesetzten Linie ist im Besonderen auf die Korrelation der einzelnen Abszissenwerte zu achten. Sie entsteht dadurch, dass der Endpunkt eines Linienelementes gleichzeitig auch immer Anfangspunkt des nächsten Stückes ist. Die Größe der Korrelation \mathbf{r} hängt vom Brechungswinkel \mathbf{a} zwischen den beiden betreffenden Linienelementen ab. Der Wert berechnet sich allgemein zu:

$$\mathbf{r} = -\cos \mathbf{a}. \quad (3-20)$$

Für die Kovarianz der beiden Abszissenwerte l_{E_1} und l_{A_2} zu einem Linienzwischenpunkt gilt dann:

$$\mathbf{s}_{l_{E_1}l_{A_2}} = \mathbf{r}_{l_{E_1}l_{A_2}} \cdot \mathbf{s}_{l_{E_1}} \cdot \mathbf{s}_{l_{A_2}} = -\mathbf{s}_{l_{E_1}} \cdot \mathbf{s}_{l_{A_2}} \cdot \cos \mathbf{a}_{E_1, A_2}. \quad (3-21)$$

Daraus ergibt sich für die Varianz der Linie folgender, auszugsweise dargestellter Zusammenhang:

$$\mathbf{s}_{l_{gesamt}}^2 = \begin{bmatrix} -1 \\ 1 \\ -1 \\ 1 \\ \vdots \end{bmatrix}^T \begin{bmatrix} \mathbf{s}_{l_{A_1}}^2 & \dots & \dots & \dots & \dots \\ \dots & \mathbf{s}_{l_{E_1}}^2 & -\cos \mathbf{a}_{E_1, A_2} \cdot \mathbf{s}_{l_{E_1}} \cdot \mathbf{s}_{l_{A_2}} & \dots & \dots \\ \dots & -\cos \mathbf{a}_{A_2, E_1} \cdot \mathbf{s}_{l_{A_2}} \cdot \mathbf{s}_{l_{E_1}} & \mathbf{s}_{l_{A_2}}^2 & \dots & \dots \\ \dots & \dots & \dots & \mathbf{s}_{l_{E_2}}^2 & \dots \\ \dots & \dots & \dots & \dots & \ddots \end{bmatrix} \begin{bmatrix} -1 \\ 1 \\ -1 \\ 1 \\ \vdots \end{bmatrix}. \quad (3-22)$$

Der vollständige Zusammenhang in Summennotation liefert folgenden Ausdruck:

$$\mathbf{s}_{l_{gesamt}}^2 = \sum_{i=1}^n \mathbf{s}_{l_{A_i}}^2 + \sum_{i=1}^n \mathbf{s}_{l_{E_i}}^2 - 2 \sum_{i=1}^n \mathbf{s}_{l_{A_i}l_{E_i}} + 2 \sum_{i=1}^{n-1} \cos \mathbf{a}_{E_i, A_{i+1}} \mathbf{s}_{l_{E_i}} \mathbf{s}_{l_{A_{i+1}}}. \quad (3-23)$$

Wiederum können einige Vereinfachungen eingeführt werden. Einheitliche Varianzen und Unabhängigkeit der Abszissenwerte innerhalb eines Linienstücks führen zur folgenden Form:

$$\mathbf{s}_{l_{gesamt}}^2 = 2 \sum_{i=1}^n \mathbf{s}_{L_0}^2 + 2 \sum_{i=1}^{n-1} \cos \mathbf{a}_{L_i, L_{i+1}} \mathbf{s}_{L_0} \mathbf{s}_{L_{0i+1}}. \quad (3-24)$$

Bei gleicher Varianz für alle Linienstücke ergibt sich dann:

$$\mathbf{s}_{l_{gesamt}}^2 = 2 \mathbf{s}_{L_0}^2 \left(n + \sum_{i=1}^{n-1} \cos \mathbf{a}_{L_i, L_{i+1}} \right) \quad (3-25)$$

Zur Verdeutlichung der Größenverhältnisse werden zwei Beispiele ausgewählt. Das Erste enthält einen langgestreckten Linienzug bestehend aus n Linienstücken. Die Brechungswinkel an den Zwischenpunkten besitzen den konstanten Wert $\mathbf{a} = 200 \text{ gon}$. Für gleiche Varianz aller Linienstücke und unter Vernachlässigung der Korrelationen innerhalb eines Linienstückes ergibt sich für die gesamte Varianz der Linie:

$$\mathbf{s}_{l_{gesamt}}^2 = 2 \mathbf{s}_{L_0}^2. \quad (3-26)$$

Das zweite Beispiel besteht aus einem geschlossenen Polygonzug, der ein Rechteck bildet, d.h. $n=4$ Punkte bzw. Linienstücke enthält, und als Brechungswinkel nur rechte Winkel besitzt. Bei gleichen Annahmen wie im ersten Beispiel erhält man für die Varianz der Linie folgenden Wert:

$$\mathbf{s}_{l_{gesamt}}^2 = 2n \mathbf{s}_{L_0}^2. \quad (3-27)$$

Allgemein kann festgehalten werden, dass die Gesamtvarianz linear mit der Anzahl der Linienstücke zunimmt. Dabei gilt, je gestreckter der Zug ist, desto geringer fällt die Zunahme aus, bis hin zu einer konstanten Varianz der Linie im Falle einer vollständig gestreckten Linie, bei der die Anzahl der Zwischenpunkte keine Rolle mehr spielt.

Minimum-Maximum-Modell

Als Unsicherheitsmaße im Minimum-Maximum-Modell sind die minimale und die maximale Länge der Linie gesucht. Sie können folgendermaßen festgelegt werden:

$$\begin{aligned} l_{min} &= l(MIN) \\ l_{max} &= l(MAX). \end{aligned} \quad (3-28)$$

Die Definition besagt, dass sich die minimale Länge als die Länge des Minimums und die maximale Länge als die Länge des Maximums ergeben. Somit ist die eigentliche Längenbestimmung einfach um die Messung der Längen der weiteren Geometrien zu ergänzen. Erfolgt die Unsicherheitsbeschreibung mit Hilfe von Epsilon-Werten, dann lässt sich die Länge des Minimums durch

$$l_{min} = l - \mathbf{e}_{l_A} - \mathbf{e}_{l_E} \quad (3-29)$$

ausdrücken, mit l als Länge des Mittels und \mathbf{e}_{l_A} bzw. \mathbf{e}_{l_E} als Epsilon-Werte der Abszissen des Linienanfangs- bzw. Linienendes. Unter Annahme gleicher Epsilon-Werte ($\mathbf{e}_{l_A} = \mathbf{e}_{l_E} = \mathbf{e}_L$) vereinfacht sich (3-29) zu:

$$l_{min} = l - 2 \mathbf{e}_L. \quad (3-30)$$

Schwieriger gestaltet sich die Berechnung der Länge des Maximums. Während das Minimum einer Linie ebenfalls eine Linie ergibt, fällt das Maximum dagegen immer flächenförmig aus, so dass die Länge nicht unmittelbar zu messen ist. Die bandförmige Struktur des Maximums legt eine Abschätzung auf der Basis des Umfangs U nahe. Näherungsweise entspricht der Umfang U_{max} der doppelten maximalen Linienlänge:

$$U_{max} \approx 2 l_{max}. \quad (3-31)$$

Dieser Ansatz lässt sich verbessern, indem man die Endbereiche der Linie gesondert behandelt. Dort schließen Halbkreise die Linie ab. Ihr Radius ist gleich dem zugehörigen Epsilon-Wert. Da die Kreisbögen aufgrund ihrer Form mit zu großen Werten zur Bestimmung des Umfangs beitragen, sind diese durch einfache Epsilon-Werte zu ersetzen. Für die maximale Linienlänge ergibt sich dann folgender Wert:

$$l_{max} = \frac{U}{2} + (1-p/2) \mathbf{e}_{l_A} + (1-p/2) \mathbf{e}_{l_E}. \quad (3-32)$$

In Anlehnung an die Bestimmung der minimalen Länge lässt sich eine alternative Variante definieren. Hierbei ergibt sich die maximale Länge durch einfache Addition der Epsilon-Werte zur mittleren Länge:

$$l_{max} = l + \mathbf{e}_{l_A} + \mathbf{e}_{l_E}. \quad (3-33)$$

Unter Annahme gleicher Werte vereinfacht sich die Bestimmung weiter zu:

$$l_{max} = l + 2 \mathbf{e}_L. \quad (3-34)$$

Anstelle von minimaler und maximaler Länge kann ebenfalls ein Epsilon-Wert als Unsicherheitsmaß angegeben werden. Er ergibt sich aus der halben Variationsbreite R , die sich allgemein aus der Differenz von Maximum und Minimum ableitet. Der Epsilon-Wert für eine Linie berechnet sich dann zu:

$$\mathbf{e}_l = R/2 = \frac{l_{max} - l_{min}}{2}. \quad (3-35)$$

Setzt man minimale und maximale Länge in vereinfachter Form darin ein, ergibt sich folgendes Ergebnis:

$$\mathbf{e}_l = 2 \mathbf{e}_L. \quad (3-36)$$

Hierbei erfolgt die Bestimmung der Längenunsicherheit unabhängig von der Form und der Anzahl der Linienzwischenpunkte. Es wirken lediglich die Epsilon-Werte des Linienanfangs- und Linienendpunktes auf die Unsicherheit ein.

Fuzzy-Modell

Zur Behandlung der Unsicherheit von Längen im Fuzzy-Modell ist die Länge als unscharfe Menge aufzufassen, die durch die Angabe einer Zugehörigkeitsfunktion charakterisiert wird. Allgemein besteht die unscharfe Längenmenge \underline{l} aus Elementpaaren mit individuellem Längen- und entsprechendem Zugehörigkeitswert:

$$\underline{l} = \{ (l_i / \mathbf{m}_i) \}. \quad (3-37)$$

Zur Bestimmung des Zugehörigkeitswertes muss die Linie, entlang der die Länge gemessen wurde, bekannt sein. Dann entscheiden die Zugehörigkeiten der Linienpunkte P_j zur unscharfen Linie \underline{l} über den Zugehörigkeitswert der Länge. Es gilt folgender Zusammenhang:

$$\mathbf{m}_i = \min(\mathbf{m}_i(P_j)) \quad (3-38)$$

Analog zur Linienunsicherheit soll auch hier eine lineare Zugehörigkeitsfunktion bestimmt werden (Abb. 3.24). Sie weist folgende Form auf:

$$\mathbf{m}_l(l) = \begin{cases} 1 & \text{für } 0 < l < b, \\ \frac{a-l}{a-b} & \text{für } b \leq l \leq a \text{ und } a \neq b, \\ 0 & \text{für } l > a \text{ und } a = b. \end{cases} \quad (3-39)$$

Um die Parameter a und b festzulegen, genügt im einfachsten Fall die Angabe von zwei Längen mit unterschiedlichen Zugehörigkeiten. Mit mittlerer Linienlänge und mit den im Minimum-Maximum-Modell behandelten minimalen und maximalen Längen lassen sich für alle Linien immer drei Werte ermitteln, deren Zugehörigkeiten bekannt sind. Sie besitzen stets die folgenden konstanten Werte:

$$\begin{aligned} (l_{\min} / \mathbf{m}_l = 1.0), \\ (l / \mathbf{m}_l = 0.5), \\ (l_{\max} / \mathbf{m}_l = 0.0). \end{aligned} \quad (3-40)$$

Daraus ergeben sich die gesuchten Parameter zu:

$$\begin{aligned} a &= l_{\max} = 2l - l_{\min}, \\ b &= l_{\min}. \end{aligned} \quad (3-41)$$

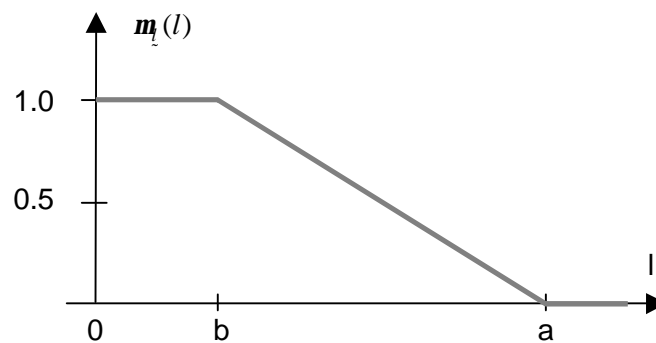


Abb. 3.24: Zugehörigkeitsfunktion der Linienlängen.

Der Fuzzy-Index \mathbf{n}_l als Maß für die Unschärfe der Menge beträgt:

$$\mathbf{n}_l = \frac{a-b}{4} = \frac{l-l_{\min}}{2}. \quad (3-42)$$

Solange die Längen unverändert bleiben, ist die Längenunsicherheit von der Form und der Anzahl an Zwischenpunkten unabhängig.

3.5.2 Bestimmung des Flächeninhalts

Der Flächeninhalt definiert ein weiteres geometrisches Attribut, das speziell nur für flächenhaft ausgedehnte Elemente zu bestimmen ist. Bei geometrischen Grundformen (Dreieck, Quadrat, Rechteck usw.) kann der Wert häufig durch einfache Formeln über Längen- und Höhenmaße von Grundseiten berechnet werden. Für beliebige Polygone kommt die bekannte Gaußsche Dreiecksformel zum Einsatz, die sich auf die Koordinaten der Eckpunkte stützt. Sie lautet für n Eckpunkte:

$$f = \frac{1}{2} \sum_{i=1}^n x_i (y_{i+1} - y_{i-1}) = \frac{1}{2} \sum_{i=1}^n y_i (x_{i-1} - x_{i+1}) \quad \text{mit } x_{n+1} = x_1 \text{ und } y_{n+1} = y_1. \quad (3-43)$$

(3-43) korrespondiert zum punktweisen Ansatz der Unsicherheit (Kapitel 2.4), bei dem Punkte die Basiselemente bilden. Sie legt nahe, dass alleinig die Unsicherheit der Eckpunkte auf die Unsicherheit des Flächeninhalts f wirkt. Der linienhafte Ansatz schreibt dagegen vor, dass jeder beliebige Linienpunkt unabhängig variieren kann. Für eine Realisierung bedeutet dies, dass Linienpunkte gleichhäufig links und rechts zur mittleren Linie auftreten werden. Aufgrund der

Zufälligkeit der Punktlagen gleicht sich die entstehende Zu- bzw. Abnahme des Flächeninhalts aber bereits innerhalb jeder Realisierung aus, so dass keine zusätzliche Variation entsteht (Abb. 3.25). Somit kann die linienhafte Betrachtung in diesem Fall auf einen punktweisen Ansatz zurückgeführt werden. Es folgen wiederum die Ableitungen zur Fortpflanzung der Unsicherheit entsprechend den verschiedenen Unsicherheitsmodellen.

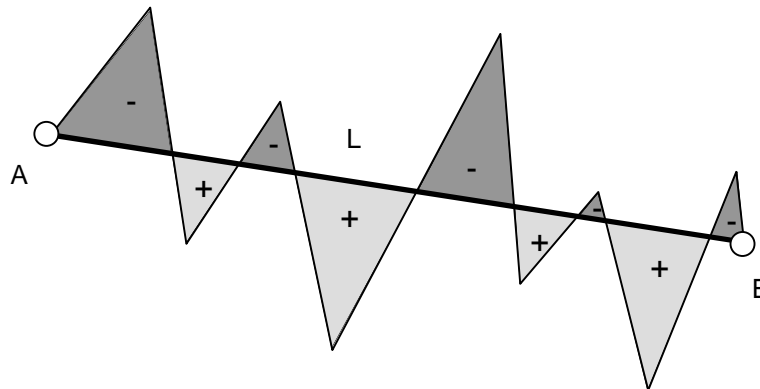


Abb. 3.25: Zu- und Abnahme des Flächeninhalts einer Realisierung der Linie durch zufällige Variation der Linienpunkte (aufgezeigt an einer endlichen Anzahl).

Stochastisches Modell

Die Unsicherheit des Flächeninhalts, deren Ursache sich in der Unsicherheit der Eckpunkte verbirgt, lässt sich über eine Varianzfortpflanzung der Flächenformel ermitteln. Dafür gilt ansatzweise folgender Zusammenhang:

$$\mathbf{s}_f^2 = \mathbf{a} \mathbf{D}(\mathbf{x}) \mathbf{a}^T = \frac{1}{4} \begin{bmatrix} \dots & \Delta y_i & \Delta x_i & \dots \end{bmatrix} \begin{bmatrix} \ddots & & & \\ & \mathbf{s}_{x_i}^2 & \mathbf{s}_{x_i y_i} & \\ & \mathbf{s}_{y_i x_i} & \mathbf{s}_{y_i}^2 & \\ & & & \ddots \end{bmatrix} \begin{bmatrix} \vdots \\ \Delta y_i \\ \Delta x_i \\ \vdots \end{bmatrix}, \quad (3-44)$$

mit

$$\begin{aligned} \Delta x_i &= x_{i-1} - x_{i+1}, \\ \Delta y_i &= y_{i+1} - y_{i-1}. \end{aligned} \quad (3-45)$$

Dabei wird ein linearer Ansatz verfolgt (Grafarend und Schaffrin, 1993), bei dem zweite und höhere Ableitungen vernachlässigt werden. Im Allgemeinen kann die Kovarianzmatrix $\mathbf{D}(\mathbf{x})$ der Punktkoordinaten vollständig besetzt sein (auch mit Kovarianzen zwischen den Punkten). Geht man dagegen von unabhängigen Punkten aus ($\mathbf{s}_{x_i x_j} = \mathbf{s}_{x_i y_j} = \mathbf{s}_{y_i x_j} = \mathbf{s}_{y_i y_j} = 0$), so vereinfacht sich $\mathbf{D}(\mathbf{x})$ zu einer schwachbesetzten Blockmatrix und die Flächenvarianz berechnet sich wie folgt:

$$\mathbf{s}_f^2 = \frac{1}{4} \sum_{i=1}^n \Delta y_i^2 \mathbf{s}_{x_i}^2 + 2 \Delta x_i \Delta y_i \mathbf{s}_{x_i y_i} + \Delta x_i^2 \mathbf{s}_{y_i}^2. \quad (3-46)$$

Bei gleichen, unkorrelierten Unsicherheitswerten für die Koordinaten ($\mathbf{s}_0^2 = \mathbf{s}_{x_i}^2 = \mathbf{s}_{y_i}^2$ und $\mathbf{s}_{x_i y_i} = 0$) ergibt sich:

$$\mathbf{s}_f^2 = \frac{1}{4} \mathbf{s}_0^2 \sum_{i=1}^n \Delta y_i^2 + \Delta x_i^2. \quad (3-47)$$

Da der Linienansatz keine einzelnen Unsicherheitswerte für die Punktkoordinaten bereithält, müssen diese aus den dort verfügbaren Linienvarianzen abgeleitet werden. Eine einfache Abschätzung für einen beliebigen Eckpunkt (x_i, y_i) liefert den folgenden Zusammenhang:

$$\mathbf{s}_{x_i}^2 = \mathbf{s}_{y_i}^2 = \max(\mathbf{s}_{L_{links_i}}^2, \mathbf{s}_{L_{rechts_i}}^2), \quad (3-48)$$

mit den Varianzen der inzidierenden Linien ($s_{L_{links_i}}^2, s_{L_{rechts_i}}^2$). Das Einsetzen des Maximums bewirkt, dass der Wert für die Varianz der Fläche etwas größer ausfällt, als er mit der strengen Lösung zu erzielen wäre.

Minimum-Maximum-Modell

Als Unsicherheitsmaße des Minimum-Maximum-Modells sind der minimale und der maximale Flächeninhalt gesucht. Die beiden Werte können mit Hilfe von Minimum und Maximum abgeleitet werden. Dabei gilt:

$$\begin{aligned} f_{min} &= f(MIN), \\ f_{max} &= f(MAX). \end{aligned} \quad (3-49)$$

Das bedeutet, dass sich die minimale bzw. maximale Fläche aus der Flächenbestimmung des Minimums bzw. Maximums ergibt. Als alternatives Unsicherheitsmaß kann ebenfalls ein Epsilon-Wert festgelegt werden. Er definiert sich zu:

$$e_f = R/2 = \frac{f_{max} - f_{min}}{2}, \quad (3-50)$$

mit Variationsbreite R , die dem Flächeninhalt des Bandes um die mittlere Objektgeometrie entspricht. Eine direkte Ableitung der Unsicherheitsmaße auf der Basis der Epsilon-Werte der Linien hängt stark von der Flächenform ab und kann deshalb nur in Ausnahmefällen (z.B. für Grundformen wie Rechtecke) formuliert werden.

Fuzzy-Modell

Im Rahmen der Unsicherheitsbetrachtungen im Fuzzy-Modell ist der Flächeninhalt ebenfalls als eine unscharfe Menge \underline{f} aufzufassen. Ihre Elemente bestehen jeweils aus einem beliebigen Flächenwert mit Angabe des entsprechenden Zugehörigkeitswertes. Formell ergibt sich die unscharfe Menge zu:

$$\underline{f} = \left\{ (f_i / m_f) \right\}, \quad (3-51)$$

mit f_i als Flächeninhalte beliebiger Realisierungen der Fläche F . Die Bestimmung der Zugehörigkeitswerte orientiert sich an den Zugehörigkeitswerten der Punkte P_j der betrachteten Realisierung zur unscharfen Flächenmenge \underline{F} . Für die Festlegung ist das Minimum der Flächenzugehörigkeiten der Punkte entscheidend. Es gilt:

$$m_f = \min(m_f(P_j)). \quad (3-52)$$

Die gegebenen Definitionen gleichen den Festlegungen zur Unsicherheit der Linienlänge. Aus diesem Grund gelten auch hier die dort getroffenen Ableitungen zur Bestimmung der Unsicherheitsparameter a und b , so dass auf eine ausführliche Beschreibung verzichtet werden kann. Die Linienwerte sind lediglich durch die entsprechenden Flächenwerte zu ersetzen.

3.5.3 Kürzeste Distanz zwischen zwei Objekten

Zwischen zwei beliebigen Punkten (A, B) in der Ebene lässt sich eine lineare Distanz wie folgt bestimmen:

$$d = \sqrt{(x_B - x_A)^2 + (y_B - y_A)^2}. \quad (3-53)$$

Im Folgenden wird dabei stets angenommen, dass die Punktkoordinaten einer ebenen Kartenprojektion (z.B. UTM- oder Gauß-Krüger-Projektion) zugehören. Entstammen die beiden Punkte unterschiedlichen Objekten, dann bildet die Distanz ein Maß für den Abstand der Objekte zueinander. Entsprechend der Anzahl der Punkte können im Allgemeinen unendlich viele Distanzen zwischen zwei Objekten ermittelt werden. Von besonderem Interesse ist vor allem die kürzeste Distanz. Dazu genügt es, die Betrachtungen auf die Randpunkte zu beschränken. Für jeden Randpunkt ist die Strecke zu ermitteln, über die man am schnellsten zum Rand des anderen Objektes gelangen kann. Die absolut kürzeste Strecke über alle Randpunkte definiert die gesuchte kürzeste Distanz (Abb. 3.26a). Bei sich überlappenden Objekten erhält man stets den Wert Null, da diese sich an mindestens einem Punkt berühren. Um für diesen Fall aber eine Fortsetzung des Abstandsmaßes zu erreichen, kann eine negative Distanz eingeführt werden. Es gilt allgemein:

$$\begin{aligned}
 d > 0: & \quad A \in \text{Objekt 1} \setminus \text{Objekt 2} \quad \wedge \quad B \in \text{Objekt 2} \setminus \text{Objekt 1}, \\
 d \leq 0: & \quad A \in \text{Objekt 1} \cap \text{Objekt 2} \quad \vee \quad B \in \text{Objekt 1} \cap \text{Objekt 2}.
 \end{aligned}
 \tag{3-54}$$

Danach ergibt sich ein negativer Wert, wenn mindestens einer der beteiligten Punkte zu beiden Objekten gehört. Als Gegenstück zur kürzesten Distanz ist der Grad der Überlappung gefragt. Wiederum sind von den Randpunkten die kürzesten Strecken zum jeweils anderen Rand zu bestimmen. Diesmal gilt es, für jeden Randpunkt auf schnellstmöglichem Wege das andere Objekt zu verlassen, wobei im Streckenverlauf nur Punkte aus dem Überlappungsbereich passiert werden dürfen (Abb. 3.26b). Insofern handelt es sich dabei ebenfalls um eine Bestimmung der kürzesten Distanz. Jedoch wird im Gegensatz zu disjunkten Objekten die betragsmäßig längste von allen diesen Strecken gesucht. Ihr Distanzwert repräsentiert den Grad der Überlappung. Man beachte, dass die kürzeste Distanz in diesem Fall immer negativ ausfällt.

Berechnungsgrundlage für die Distanz bilden die zugehörigen Punktpositionen. Da diese zur unsicheren Geometrie gehören, werden auch die Distanzwerte mit Unsicherheit behaftet sein. Als Folge davon wird auch der Grad der Überlappung unsicher. Daher wird im Folgenden zur Fortpflanzung der Unsicherheit auf die Distanz zusätzlich bewertet, wie sicher sich die Objekte überlappen.

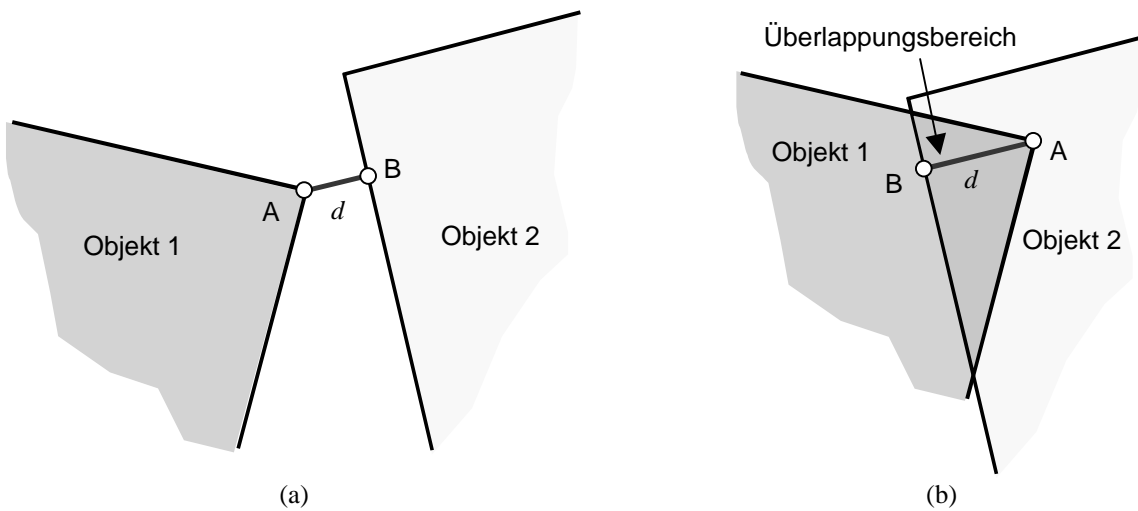


Abb. 3.26: Kürzeste Distanz bei disjunkten (a) und sich überlappenden (b) Objekten.

Stochastisches Modell

Wendet man die Varianzfortpflanzung auf die Distanzformel an, so ergibt sich für die Varianz der Distanz ansatzweise der folgende Zusammenhang:

$$\mathbf{s}_d^2 = \mathbf{a} \mathbf{D}(\mathbf{x}) \mathbf{a}^T = \frac{1}{d^2} \begin{bmatrix} -\Delta x & -\Delta y & \Delta x & \Delta y \end{bmatrix} \begin{bmatrix} \mathbf{s}_{x_A}^2 & \mathbf{s}_{x_A y_A} & \mathbf{s}_{x_A x_B} & \mathbf{s}_{x_A y_B} \\ \mathbf{s}_{y_A x_A} & \mathbf{s}_{y_A}^2 & \mathbf{s}_{y_A x_B} & \mathbf{s}_{y_A y_B} \\ \mathbf{s}_{x_B x_A} & \mathbf{s}_{x_B y_A} & \mathbf{s}_{x_B}^2 & \mathbf{s}_{x_B y_B} \\ \mathbf{s}_{y_B x_A} & \mathbf{s}_{y_B y_A} & \mathbf{s}_{y_B x_B} & \mathbf{s}_{y_B}^2 \end{bmatrix} \begin{bmatrix} -\Delta x \\ -\Delta y \\ \Delta x \\ \Delta y \end{bmatrix}, \tag{3-55}$$

mit

$$\begin{aligned}
 \Delta x &= x_B - x_A, \\
 \Delta y &= y_B - y_A.
 \end{aligned}
 \tag{3-56}$$

Im Allgemeinen kann $\mathbf{D}(\mathbf{x})$ vollbesetzt sein. Unter der Annahme von Unabhängigkeit der Punkte ($\mathbf{s}_{x_A x_B} = \mathbf{s}_{x_A y_B} = \mathbf{s}_{y_A x_B} = \mathbf{s}_{y_A y_B} = 0$ und $\mathbf{s}_{x_B x_A} = \mathbf{s}_{y_B x_A} = \mathbf{s}_{x_B y_A} = \mathbf{s}_{y_B y_A} = 0$) kann die Berechnung mit

$$\mathbf{s}_d^2 = \frac{1}{d^2} \left[\Delta x^2 (\mathbf{s}_{x_A}^2 + \mathbf{s}_{x_B}^2) + 2\Delta x \Delta y (\mathbf{s}_{x_A y_A} + \mathbf{s}_{x_B y_B}) + \Delta y^2 (\mathbf{s}_{y_A}^2 + \mathbf{s}_{y_B}^2) \right] \tag{3-57}$$

erfolgen. (3-57) vereinfacht sich weiter zu:

$$\mathbf{s}_d^2 = 2 \mathbf{s}_0^2, \quad (3-58)$$

wenn zusätzlich gleichgenaue, unkorrelierte Koordinaten vorliegen ($\mathbf{s}_0 = \mathbf{s}_{x_A} = \mathbf{s}_{x_B} = \mathbf{s}_{y_A} = \mathbf{s}_{y_B}$ und $\mathbf{s}_{x_A y_A} = \mathbf{s}_{x_B y_B} = 0$). Die Werte des linienweisen Ansatzes der Unsicherheitsmodellierung, in dem statt der Varianzen der Koordinaten für jeden Randpunkt die Varianz senkrecht zur Linienrichtung bekannt ist, lassen sich ebenfalls nutzen, indem die Messstrecke als Linienelement aufgefasst wird. Dann können die für die Längenbestimmung getroffenen Festlegungen zur Unsicherheit auf die Distanzmessung übertragen werden (Kapitel 3.5.1). Insbesondere gilt für unabhängige Punkte:

$$\mathbf{s}_d^2 = \mathbf{s}_{l_A}^2 + \mathbf{s}_{l_B}^2, \quad (3-59)$$

mit $\mathbf{s}_{l_A}^2$ und $\mathbf{s}_{l_B}^2$ als Varianzen in Richtung der Messstrecke, die den Varianzen der Abszissenwerte entsprechen. Die Richtungen der Linienvariation unterscheiden sich im Allgemeinen voneinander und von der Richtung der Messstrecke. Aus diesem Grund sind die Abszissenvarianzen als Komponenten der zugehörigen Linienvarianz zu bestimmen. Die Komponentenbestimmung hängt von der Richtungsdifferenz \mathbf{a} zwischen Variationsrichtung und Streckenrichtung ab und kann wie folgt beschrieben werden:

$$\begin{aligned} \mathbf{s}_{l_A}^2 &= \cos \mathbf{a}_A \mathbf{s}_{L_A}^2 & \text{mit } \mathbf{a}_A \in [0, \mathbf{p} / 2], \\ \mathbf{s}_{l_B}^2 &= \cos \mathbf{a}_B \mathbf{s}_{L_B}^2 & \text{mit } \mathbf{a}_B \in [0, \mathbf{p} / 2], \end{aligned} \quad (3-60)$$

mit $\mathbf{s}_{L_A}^2$ und $\mathbf{s}_{L_B}^2$ als geltende Linienvarianzen in den Punkten A und B . Daraus lässt sich für die Varianz der Distanz ableiten:

$$\mathbf{s}_d^2 = \cos \mathbf{a}_A \mathbf{s}_{L_A}^2 + \cos \mathbf{a}_B \mathbf{s}_{L_B}^2. \quad (3-61)$$

Beispielsweise ergibt sich für den Fall, dass die Richtungen übereinstimmen und die Linien gleiche Varianzen besitzen ($\mathbf{s}_{L_0} = \mathbf{s}_{L_A} = \mathbf{s}_{L_B}$):

$$\mathbf{s}_d^2 = 2 \mathbf{s}_{L_0}^2. \quad (3-62)$$

Die Objekte überlappen sich, wenn die Distanz einen negativen Wert annimmt. Um diese Aussage zu bewerten, ist festzustellen, mit welcher Wahrscheinlichkeit die Zufallsvariable „Distanz“ einen Wert kleiner als Null annehmen wird. Für die Wahrscheinlichkeit ergibt sich:

$$p(d < 0) = \int_{-\infty}^0 f(t) dt, \quad (3-63)$$

mit $f(t)$ als Dichtefunktion der Distanz.

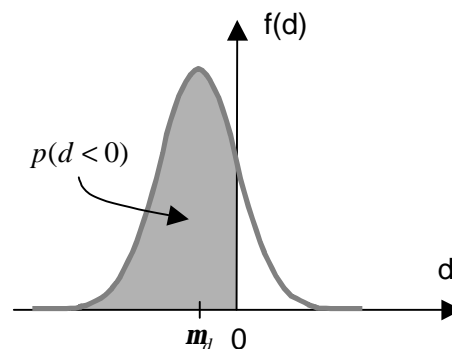


Abb. 3.27: Bestimmung der Wahrscheinlichkeit für die Überlappung zweier Objekte anhand der kürzesten Distanz unter Annahme einer Normalverteilung.

Tab. 3.3 stellt einige Beispiele zusammen, bei denen die Distanz in Abhängigkeit ihrer Varianzen gewählt werden. Hierzu gilt die Annahme der Normalverteilung.

Tab. 3.3: Wahrscheinlichkeiten der Überlappung für ausgewählte Distanzen der Objekte.

Distanz d	$-3s_d$	$-2s_d$	$-1s_d$	0	$1s_d$	$2s_d$	$3s_d$
Wahrscheinlichkeit $p(d < 0)$	0.001	0.023	0.159	0.5	0.841	0.977	0.999

Minimum-Maximum-Modell

Die Variation der kürzesten Distanz zweier Objekte lässt sich im Minimum-Maximum-Modell durch die beiden Werte der minimalen bzw. der maximalen kürzesten Distanz beschreiben. Sie ergeben sich als:

$$\begin{aligned} d_{min} &= d(MIN_1, MIN_2) \\ d_{max} &= d(MAX_1, MAX_2) \end{aligned} \tag{3-64}$$

aus den kürzesten Distanzen der Minima bzw. der Maxima der beiden Objekte. Eine alternative Vorgehensweise bietet die Verwendung von Epsilon-Werten. Wie bereits im Stochastischen Modell erläutert, sind dabei die Richtungen der Linienvariation im Bezug auf die Messstrecke zu beachten. Auch hier sind entsprechende Komponenten der originalen Epsilon-Werte zur Beschreibung der Variation in Richtung der Messstrecke abzuleiten. Sie lauten:

$$\begin{aligned} \mathbf{e}_{l_A} &= \cos \mathbf{a}_A \mathbf{e}_{L_A} \quad \text{mit } \mathbf{a}_A \in [0, \mathbf{p} / 2], \\ \mathbf{e}_{l_B} &= \cos \mathbf{a}_B \mathbf{e}_{L_B} \quad \text{mit } \mathbf{a}_B \in [0, \mathbf{p} / 2], \end{aligned} \tag{3-65}$$

mit \mathbf{e}_{L_A} und \mathbf{e}_{L_B} für die Epsilon-Werte der Linien. Daraus ergeben sich für die minimale bzw. maximale kürzeste Distanz:

$$\begin{aligned} d_{min} &= d - \cos \mathbf{a}_A \mathbf{e}_{L_A} - \cos \mathbf{a}_B \mathbf{e}_{L_B}, \\ d_{max} &= d + \cos \mathbf{a}_A \mathbf{e}_{L_A} + \cos \mathbf{a}_B \mathbf{e}_{L_B} \end{aligned} \tag{3-66}$$

und für den Epsilon-Wert als Unsicherheitsparameter der kürzesten Distanz:

$$\mathbf{e}_d = R / 2 = \frac{d_{max} - d_{min}}{2} = \cos \mathbf{a}_A \mathbf{e}_{L_A} + \cos \mathbf{a}_B \mathbf{e}_{L_B}. \tag{3-67}$$

mit R als Variationsbreite. Die Schätzung auf der Basis von Epsilon-Werten kann sich in manchen Fällen von der Messung über Minima und Maxima unterscheiden. Dies hängt damit zusammen, dass die Epsilon-Werte an der Stelle der kürzesten Distanz d der mittleren Geometrie bestimmt werden, während Maxima und Minima an anderen Stellen ihre kürzeste Distanz aufweisen können. Das trifft insbesondere dann zu, wenn im Objekt große Unterschiede in der geometrischen Unsicherheit auftreten. Aus diesem Grund ist die Bestimmung der Unsicherheit durch Messen nach Möglichkeit vorzuziehen.

Wie sicher sich die Objekte überlappen, wird durch das zum Minimum-Maximum-Modell gehörende Möglichkeitsmaß m ausgedrückt. Sein Wert ermittelt sich aus der Betrachtung aller Geometrien der Objekte. Dazu ist festzustellen, welche der gleichartigen Geometrien sich überlappen. Trifft dies beispielsweise für die beiden Minimum-Geometrien zu, dann ist die Überlappung als sicher anzusehen. Stellt man dagegen nur eine Überlappung der Maxima fest, dann ergibt sich die Bewertung $m = \text{möglicherweise nicht}$. Alle Entscheidungskriterien und daraus resultierenden Werte sind in Tab. 3.4 zusammengestellt.

Tab. 3.4: Bewertung der Überlappung zweier Objekte durch Untersuchung der Objektgeometrien.

Zustand	Möglichkeitswert m
$MIN_1 \cap MIN_2 \neq \emptyset$	Sicher
$MITTEL_1 \cap MITTEL_2 \neq \emptyset$	Möglicherweise
$MAX_1 \cap MAX_2 \neq \emptyset$	Möglicherweise nicht
$MAX_1 \cap MAX_2 = \emptyset$	Sicher nicht

Fuzzy-Modell

Alle möglichen Distanzen zwischen beliebigen Objektpunkten und deren Zugehörigkeitswerte bilden die Elemente der unscharfen Menge \underline{d} , durch die im Fuzzy-Modell die kürzeste Distanz zu beschreiben ist. Für sie kann geschrieben werden:

$$\underline{d} = \{ (d_i / \mathbf{m}_i) \}, \quad (3-68)$$

mit d_i als eine Distanz zwischen beliebigen Punkten (A_i, B_i) und \mathbf{m}_i als deren Grad der Zugehörigkeit zur Menge. Die Bestimmung des Wertes stützt sich auf die Zugehörigkeiten der Punkte zum jeweiligen Objekt und formuliert sich folgendermaßen:

$$\mathbf{m}_i = \min(\mathbf{m}_1(A_i), \mathbf{m}_2(B_i)). \quad (3-69)$$

Zur Abschätzung der Funktion in linearer Weise sind Werte der beiden Parameter a und b zu bestimmen. Sie ergeben sich wie gewohnt als:

$$\begin{aligned} a &= d_{\max}, \\ b &= d_{\min}. \end{aligned} \quad (3-70)$$

Im Gegensatz zu den Funktionen für Linienlänge und Flächeninhalt (3-39) nimmt die Zugehörigkeit im Falle der kürzesten Distanz mit kleiner werdenden Distanzwerten ab. Daher ist eine neue Funktion zu definieren (Abb. 3.28). Für sie gilt:

$$\mathbf{m}_i(d) = \begin{cases} 0 & \text{für } d < a, \\ \frac{d-a}{b-a} & \text{für } a \leq d \leq b \text{ und } a \neq b, \\ 1 & \text{für } d > b \text{ und } a = b. \end{cases} \quad (3-71)$$

Der Fuzzy-Index bleibt unverändert bestehen.

Die Bewertung der Überlappung erfolgt im Fuzzy-Modell durch die Angabe des Möglichkeitsmaßes \mathbf{p} . Dabei ist festzustellen, welche maximale Zugehörigkeit Distanzen kleiner als Null besitzen. Im Allgemeinen genügt es aufgrund der ansteigenden Funktionswerte die Zugehörigkeitsfunktion an der Stelle $d=0$ zu untersuchen. Unter der Annahme, dass an dieser Stelle mehrere Funktionswerte möglich sind, lautet die Bestimmung dann:

$$\mathbf{p}_{\text{Überlappung}} = \max(\mathbf{m}_i(0)). \quad (3-72)$$

Beispielsweise ergibt sich für Objekte, deren mittlere Geometrie einander berühren, im Normalfall der Möglichkeitswert von $\mathbf{p} = 0.5$.

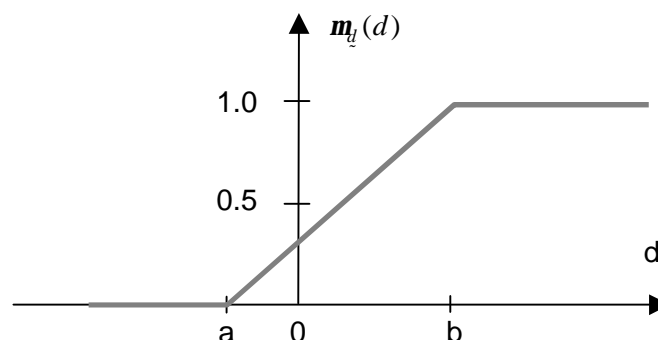


Abb. 3.28: Zugehörigkeitsfunktion der kürzesten Distanz.

3.5.4 Verschneidung von Objekten

Die Verschneidung zählt zu den wichtigsten geometrischen Operationen eines GIS. Sie stellt einen unverzichtbaren Bestandteil für jedes System dar und gilt als eine Minimalanforderung an die Analysemöglichkeiten eines GIS (Bill, 1999b). Die Aufgabe besteht darin, die geometrischen Primitive (Flächen, Linien, Punkte) zweier sich überlappender Objektmengen miteinander zu verschneiden, um aus den entstehenden Teilen neue Objekte zu bilden. Auf diese Weise lassen sich die unterschiedlichen Themenbereiche der Objekte miteinander vereinigen und damit neue bisher nur implizit enthaltene Informationen gewinnen. Im Einzelnen sind zunächst alle Objektkanten auf gegenseitige Schnitte zu prüfen und gegebenenfalls Schnittpunkte zu berechnen. Jeder Schnittpunkt ergibt einen neuen Knoten und teilt damit die ursprüngliche Kante in zwei Teile. Anschließend können aus den Teilstücken neue geometrische Primitive gebildet werden (z.B. durch Polygonbildung). Sie definieren die Geometrie der gesuchten Ergebnisobjekte. Wichtig ist, dass zu jedem neuen Primitiv festgehalten wird, welchem Eingangsobjekt es ursprünglich angehörte. Die Information benötigt man, um im abschließenden Schritt eine Übertragung der thematischen Informationen (Attribute) auf die neuen Objekte zu realisieren.

Betrachtet man Objekte als Punktmengen, so entstehen durch die Verschneidung zweier Objekte A und B im Allgemeinen die folgenden Teilmengen (Abb. 3.29):

$$A \setminus B, A \cap B, B \setminus A. \quad (3-73)$$

Der Nutzer hat die Wahl, welche Teilmengen die Ergebnisobjekte bilden sollen. Entscheidet er sich beispielsweise für alle Mengen, so gewinnt man eine Vereinigung der Eingangsobjekte ($A \cup B$). In den meisten Fällen sind Nutzer jedoch an der Schnittmenge ($A \cap B$) interessiert, so dass die Verschneidung häufig mit einer logischen UND-Verknüpfung von Punktmengen gleichgesetzt werden kann. Die weiteren Erläuterungen konzentrieren sich im Besonderen auf diese Art der Verschneidung.

Die Verschneidung erfolgt in traditioneller Weise auf Basis der gespeicherten Objektgeometrie, die im Allgemeinen der mittleren Geometrie entspricht. Insofern werden nur dann Schnittberechnungen ausgeführt, wenn sich die Geometrien zweier Objekte auch wirklich überlappen (Abb. 3.29a). Es entstehen dagegen keine neuen Objekte bei disjunkten Geometrien (Abb. 3.29b). Trotzdem kann es vorkommen, dass aufgrund der geometrischen Unsicherheit eine Überlappung möglich ist, obwohl sich die Geometrien im traditionellen Sinne nicht schneiden. Wie sicher eine Überlappung vorliegt, kann mit Hilfe der kürzesten Distanz und deren Unsicherheit festgestellt werden (Kapitel 3.5.3). Dem Nutzer bleibt hierbei die Entscheidung überlassen, ab welchem Sicherheitsniveau er eine mögliche Überlappung akzeptiert oder verwirft. Für diesen Zweck ist ein anwendungsabhängiger Schrankenwert in Form eines Unsicherheitswertes (z.B. Wahrscheinlichkeit $p_{\text{Schranke}} = 0.05$, Möglichkeitswert $m_{\text{Schranke}} = \text{möglicherweise nicht}$, Möglichkeitswert $p_{\text{Schranke}} = 0.1$) festzulegen. Er definiert das Mindestmaß an Überlappung, die zwei Objekte aufweisen müssen, damit eine Verschneidung ausgeführt wird. Bei sicherheitsrelevanten Analyseaufgaben, bei denen z.B. das Überlappen von zwei Themen unbedingt auszuschließen ist, wird der Schrankenwert äußerst niedrig anzusetzen sein, so dass mit großer Sicherheit alle möglichen Fälle berücksichtigt werden.

Das Zulassen und Bearbeiten der Fälle, in denen der Schnitt der Objektgeometrien kein Ergebnis mehr liefert, erfordert eine Erweiterung des traditionellen Verschneidungsansatzes. Die Lösung hängt entscheidend vom gewählten Unsicherheitsmodell ab, so dass die nachfolgenden Untersuchungen getrennt nach den verschiedenen Modelle erfolgen. Dabei ist zu allen akzeptierten Verschneidungsfällen jeweils die Geometrie der Ergebnisobjekte zu erzeugen und die geometrische Unsicherheit fortzupflanzen.

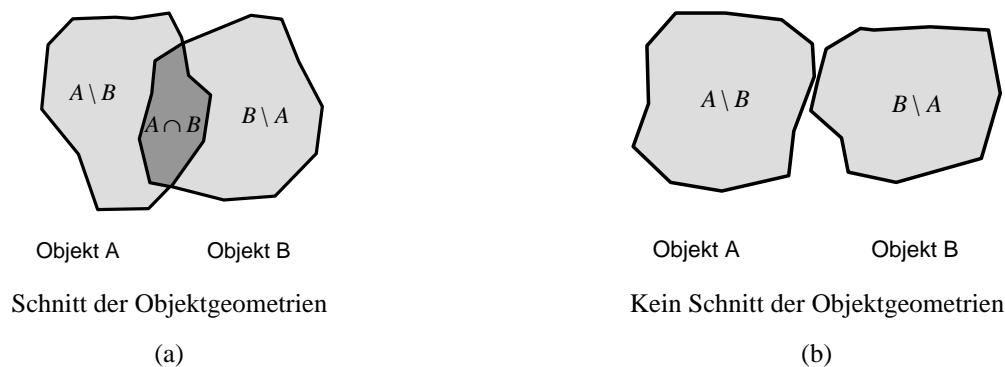


Abb. 3.29: Zwei Fälle des traditionellen Verschneidungsansatzes – Schnitt existiert (a), Schnitt existiert nicht (b).

Stochastisches Modell

Die Schnittbildung bewirkt bei der punktwweisen Betrachtung eine Übertragung der geometrischen Unsicherheit von den Linienendpunkten auf den Schnittpunkt. In diesem Fall ist auf der Basis der Schnittgleichung (Schnitt zweier Geraden) die Varianzfortpflanzung mit dem Ziel abzuleiten, die Kovarianzmatrizen der Schnittpunkte zu schätzen. Die dazugehörigen Ausführungen können in Bill und Korduan (1998) gefunden werden. Die Linien selbst werden durch die Schnittbildung nur geteilt, bleiben aber ansonsten unverändert. Insbesondere entstehen keine Auswirkungen auf die geometrische Variation in beliebigen Linienpunkten. Daher entspricht die Varianzfortpflanzung in der linienweisen Betrachtung einer einfachen Übertragung der Unsicherheitsmaße auf die generierten Teillinien. Bei der Erzeugung eines neuen Objektes übernehmen die zugehörigen Objektlinien die Linienvarianzen der Ursprungslinien.

Die einfache Übertragung der Unsicherheitsmaße ist auf den traditionellen Verschneidungsansatz begrenzt, da nur dann eine Schnittbildung möglich ist. In allen anderen Fällen, in denen keine Schnittgeometrie im traditionellen Sinne existiert, benötigt man einen erweiterten Lösungsansatz. Zwei unterschiedliche Vorgehensweisen werden im Folgenden vorgestellt.

Verschneidung der Punkt mengen

Fasst man die Objekte als Punkt mengen auf, so kann die Verschneidung als eine logische Verknüpfung der Mengen beschrieben werden. Die Schnittmenge wird durch Anwendung einer logischen UND-Operation erzeugt, wobei für jeden Punkt im Einzelnen festzustellen ist, ob er sowohl in der einen als auch in der anderen Eingangsmenge vorkommt. Es muss daher geklärt werden, wann ein beliebiger Punkt zu einem Objekt gehört. Das ist immer dann der Fall, wenn der Punkt sich innerhalb der Objektgeometrie befindet, d.h. die geometrische Basisaussage (Position liegt innerhalb des Objektes) erfüllt ist. Aufgrund der geometrischen Unsicherheit entstehen Fälle, bei denen die Basisaussage nur teilweise zutrifft. Bekanntermaßen lässt sich der Grad des Zutreffens mit Hilfe von Wahrscheinlichkeiten ausdrücken (Kapitel 2.4). Grundsätzlich gilt dann, dass ein Punkt zu einem bestimmten Objekt zu zählen ist, wenn die zugehörige Wahrscheinlichkeit einen Wert $p > 0$ besitzt. Der Umfang der aus dieser Festlegung resultierenden Punktmenge ist bedeutend ausgedehnter als die Punktmenge, die durch die mittlere Geometrie definiert wird, da sämtliche Punkte des Variationsbereiches hinzukommen.

Das entstehende Schnittobjekt umfasst schließlich alle Punkte, die jeweils in beiden Eingangsmengen enthalten sind. Diese Punkte variieren wiederum im Grad der Zugehörigkeit, verursacht durch die jeweiligen Zugehörigkeiten zu den Eingangsmengen, die sich dann über die angewendete logische Operation auf das Ergebnis fortpflanzen. Die Beschreibung der Fortpflanzung hat folglich auf Basis von Wahrscheinlichkeiten zu erfolgen. Grundlage dafür stellt die Wahrscheinlichkeitstheorie bereit (Anhang B.1). Für die logische UND-Operation zweier Punkt mengen ($C = A \cap B$) definiert sich die Fortpflanzung für einen beliebigen Punkt P zu:

$$\begin{aligned} p(P \in C) &= p((P \in A) \wedge (P \in B)) = p(P \in A) \cdot p(P \in B) \\ &= p(a \wedge b) = p(a) \cdot p(b). \end{aligned} \quad (3-74)$$

Hierbei gilt die Annahme, dass die Zugehörigkeiten zu den Eingangsobjekten unabhängig voneinander zu bestimmen sind. Die Formel setzt voraus, dass die Objekte sich sicher überlappen und damit ein Schnittobjekt sicher existiert. Wie den Ausführungen zur kürzesten Distanz zu entnehmen ist (Kapitel 3.5.3), trifft diese Vorbedingung nicht in allen Fällen zu. Die Überlappung kann ebenfalls unsicher sein. Wird zusätzlich die Überlappungsunsicherheit beachtet, dann lässt sich die Fortpflanzung als totale Wahrscheinlichkeit folgendermaßen beschreiben:

$$p(P \in C) = p(c) = p(o) \cdot p(a \wedge b / o) + p(\neg o) \cdot p(a \wedge b / \neg o) \quad (3-75)$$

mit o als Aussage zur Überlappung ($o = \text{„Objekte überlappen sich“}$). Wenn ein Punkt zum Objekt A und zum Objekt B gehört, dann kann daraus geschlossen werden, dass sich die Objekte überlappen. Daher schließen sich die Ereignisse $a \wedge b$ und $\neg o$ gegenseitig aus, so dass deren bedingte Wahrscheinlichkeit immer zu $p(a \wedge b / \neg o) = 0$ wird. Sinnvollerweise wird eine Verschneidung dazu ausgeführt, um ein Schnittobjekt zu erhalten. Ein solches Objekt existiert dann, wenn die Überlappung der Eingangsobjekte aus Anwendersicht groß genug ausfällt. Ein Akzeptieren der Überlappung hat zur Folge, dass die Existenz des erzeugten Objektes im Weiteren als sicher anzunehmen ist. Insofern ist anstelle der allgemeinen Situation ($a \wedge b$), die Punktzugehörigkeit unter der Bedingung der Überlappung zu beurteilen ($a \wedge b / o$). Die Fortpflanzung der Unsicherheit erhält dadurch folgendes verändertes Aussehen:

$$p((P \in C) / o) = p(a \wedge b / o) = p(c / o) = \frac{p(a \wedge b)}{p(o)} = \frac{p(a) \cdot p(b)}{p(o)} \quad (3-76)$$

Die bedingten Punktwahrscheinlichkeiten fallen meist größer als die allgemeinen Wahrscheinlichkeiten aus.

Jedes Objekt besteht im Allgemeinen aus einer unendlich großen Anzahl an Punkten, für die theoretisch alle diskutierten Untersuchungen auszuführen sind. Aus praktischen Gründen hat deshalb eine Reduzierung zu erfolgen. Eine Möglichkeit bietet sich durch Diskretisierung des Raumes in Form eines Raster. Jede Rasterzelle repräsentiert dabei genau einen Punkt, für den die Zugehörigkeit festzustellen ist.

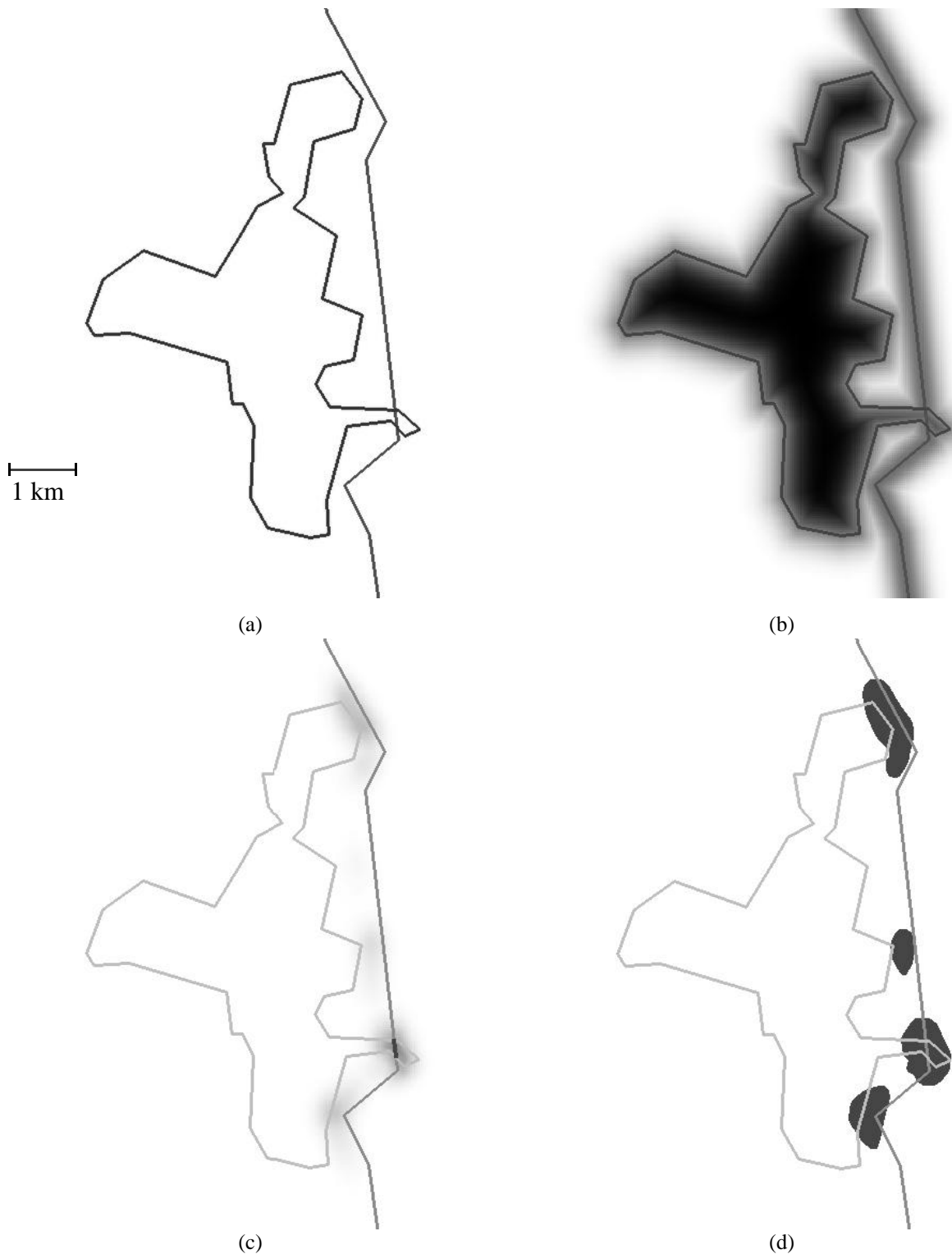


Abb. 3.30: Beispiel für die Fortpflanzung der Unsicherheit bei der Verschneidung im Stochastischen Modell. Dargestellt sind zwei zu verschneidende Objekte (a) mit ihrer Unsicherheit in Matrixdarstellung (b). Es ergibt sich ein linienförmiges Ergebnisobjekt, dessen fortgepflanzte Unsicherheit ebenfalls in einer Matrixdarstellung hinterlegt ist (c). Durch Schwellwertbildung (hier: $p_s = 0.1$) kann eine neue Objektgeometrie abgeleitet werden (d), die die Objektausdehnung mit einer größeren Sicherheit repräsentiert.

Das Vorgehen kann als eine Vektor-Raster-Transformation bezeichnet werden, mit der Besonderheit, dass die Rasterzellen Wahrscheinlichkeiten gemäß ihrer Objektzugehörigkeit enthalten. Das Verfahren gleicht damit der Erzeugung einer Matrixdarstellung zum Zwecke der Präsentation der Unsicherheit (Kapitel 3.3.2). Notwendige Festlegungen insbesondere zur Auflösung des Rasters können direkt von dort übernommen werden. Für jedes Objekt ist ein getrenntes Raster zu erzeugen. Aufgabe der Unsicherheitsfortpflanzung ist es, die Wahrscheinlichkeiten des Ergebnisrasters zu berechnen. Dazu sind zunächst die Raster der Eingangsobjekte einander zu überlagern. Die überlagerte Fläche begrenzt grob die Ausdehnung des Ergebnisobjektes. Dessen Auflösung kann entweder aus den Werten der Eingangsraster abgeleitet oder unabhängig davon festgelegt werden. Für jede Zelle ist dann die definierte bedingte Wahrscheinlichkeit zu berechnen, wobei als Eingangswerte die Wahrscheinlichkeiten der jeweils nächstgelegenen Zelle zu beiden Eingangsobjekten entnommen werden müssen.

Zusätzlich zur Fortpflanzung wird weiterhin eine Repräsentation der Objektgeometrie gesucht, ohne die kein räumliches Objekt bestehen darf. Die Ableitung kann auf Basis des fortgepflanzten Ergebnisrasters erfolgen, indem ein Schwellwert für die Wahrscheinlichkeiten festgelegt wird. Alle Rasterzellen, deren Werte größer als der Schwellwert ausfallen, werden daraufhin der Objektgeometrie zugerechnet. Es entsteht eine Rasterdarstellung der gesuchten Geometrie, die durch eine geeignete Raster-Vektor-Transformation (z.B. Lichtner, 1981, Cramer, 1993, Fritsch et al., 1998) in eine Vektordarstellung überführt werden kann. Beispielsweise erzeugt der Schwellwert $p_s = 0$ die maximale Geometrie des Objektes. Durch die anwendungsbezogene Wahl des Schwellwertes werden sich die entstehende Objektgeometrie und die mittlere Geometrie normalerweise voneinander unterscheiden. Der Ansatz bietet aber keine Möglichkeit, eine Schätzung der mittleren Geometrie zu erreichen. Aus diesem Grund kann nach Anwendung des Verfahrens nicht mehr auf eine Parameterdarstellung übergegangen werden, auch wenn auf einem noch zu entwickelnden Wege Varianzen aus den Wahrscheinlichkeiten abzuleiten sind, da weiterhin die mittlere Geometrie dazu fehlt.

Abb. 3.32 zeigt ein Anwendungsbeispiel, in dem eine linienhafte Überlandstromleitung mit einem flächenhaften Stadtgebiet verschnitten werden soll. Ziel ist es festzustellen, welche Stadtbereiche von der Leitung überquert werden. Entnommen sind die Objekte dem Datenbestand „Digital Chart of the World“ (DCW, 2000), der eine weltweite topographische Datenbasis für den kleinmaßstäbigen Bereich (Maßstabsbereich ca. 1:1.000.000) bereitstellt. Vereinfachend wird angenommen, dass die Objekte unabhängig voneinander erfasst wurden. Die Abschätzung der Unsicherheitsmaße erzeugt unterschiedliche Werte (Standardabweichungen der Linien für die Leitung: $s_L = 200\text{ m}$ mit Breite $b = 100\text{ m}$ bzw. für das Stadtgebiet: $s_L = 250\text{ m}$), die gemeinsam in einer Matrixdarstellung veranschaulicht werden. Die kürzeste Distanz beträgt $d = -85\text{ m}$, so dass die Wahrscheinlichkeit der Überlappung mit $p = 0.63$ zu berechnen ist. Die Verschneidung auf traditionellem Weg liefert erwartungsgemäß ein kleines Schnittobjekt, in dem Bereich, wo sich die Eingangsgeometrien überlappen. Die Verschneidung der Punktmengen hat dagegen ein Raster zum Ergebnis, das die Wahrscheinlichkeiten der Punktzugehörigkeiten zum Schnittobjekt enthält und ebenfalls durch eine Matrixdarstellung visualisiert werden kann. Das Raster bildet die Basis für die Ableitung der Objektgeometrie, für die hier ein Schwellwert von $p_s = 0.1$ gewählt wurde.

Verschneidungssimulation

Eine andere Möglichkeit bietet die Anwendung von Simulationen. Sie verfolgen allgemein das Ziel, Aussagen über das Verhalten von Systemen zu gewinnen (Piehler und Zschiesche, 1976). Notwendig werden Simulationen beispielsweise dann, wenn das direkte Beobachten der Systeme nicht möglich ist oder zu komplex erscheint. In diesen Fällen wird die Originalsituation durch ein Ersatzsystem in Form eines mathematischen Modells beschrieben, das dann zu Analyse-zwecken benutzt werden kann. Mit den so gewonnenen Ergebnissen lassen sich Rückschlüsse auf das eigentliche System herstellen. Eine spezielle Simulationsart findet sich in der Monte-Carlo-Methode (z.B. Sobol, 1985). Sie eignet sich im Besonderen zur Untersuchung von zufallsabhängigen Systemen. Basis bildet die Generierung einer großen Menge an Eingangswerten, die als Stichprobe dient. Im Allgemeinen handelt es sich dabei bereits um eine Ansammlung von Zufallswerten, die Realisierungen einer oder mehrerer Zufallsvariablen entsprechen. Das formulierte Modell, das zusätzlich von weiteren zufälligen Faktoren beeinflusst werden kann, wird nacheinander auf jede der Realisierungen angewendet, so dass für das Ergebnis ebenfalls eine Stichprobe entsteht. Zur abschließenden Analyse des Ergebnisses lassen sich statistische Methoden einsetzen. Die Güte des Verfahrens hängt einerseits von der Qualität der Modelldefinition, andererseits vom Umfang der Stichprobe ab. Je mehr Zufallswerte prozessiert werden, desto zuverlässiger fällt das Ergebnis aus. Daher ist die Anwendung des Verfahrens stets mit einem großen Aufwand (insbesondere Rechenaufwand) verbunden.

Im Falle der Verschneidung kann die Monte-Carlo-Methode dazu eingesetzt werden, die Fortpflanzung der Unsicherheit zu untersuchen. Die Zufallsabhängigkeit entsteht dabei durch die Variation in der Ausdehnung der zu verschneidenden Objekte. Diese Variation gilt es durch ein geeignetes Simulationsmodell nachzubilden. Grundlagen für das entwickelte Vorgehen sind im Anhang C.2 am Beispiel einer Linie erläutert. Der Modellansatz stützt sich auf die gespeicherten Unsicherheitsmaße der Objekte. Mit deren Hilfe werden zu jedem Eingangsobjekt eine Vielzahl an Reali-

sierungen der Objektgeometrie erzeugt, die dann nacheinander in traditioneller Weise miteinander verschritten werden. Dabei kann es vorkommen, dass manche Realisierungen kein Schnittergebnis liefern, da keine Überlappung existiert. Daraus lässt sich als ein Ergebnis der Simulation die relative Häufigkeit des Auftretens eines Schnittergebnisses als Maß für die Überlappung ableiten. Die Schnittergebnisse können wiederum als Realisierungen der Objektgeometrie des Schnittobjektes aufgefasst werden. In gleicher Weise wie im Erfassungsschritt (Kapitel 3.2) sind die Realisierungen auszuwerten. Als Ergebnis werden die mittlere Geometrie und die Varianzen der Linien als die bekannten Unsicherheitsparameter gewonnen.

Einige praktische Erkenntnisse lassen sich über ein einfaches Beispiel gewinnen. Aufgabe ist es, zwei rechteckige Objekte miteinander zu verschneiden. Für alle Linien sollen identische Linienvarianzen ($s_L = s_0$) gelten. Es werden vier verschiedene Fälle konstruiert und ausgewertet (Abb. 3.31). Ausgehend vom ersten Fall, bei dem eine vollständige Überlagerung ($d_1 = -8s_0$) vorliegt, rücken die beiden Objekte in den weiteren Fällen immer weiter voneinander ab, so dass ihre Entfernung zunimmt ($d_2 = -4s_0, d_3 = 0, d_4 = 2s_0$). Die aus der kürzesten Distanz zu ermittelnden Wahrscheinlichkeiten der Überlappung betragen $p_1 = 1.0, p_2 = 1.0, p_3 = 0.5$ und $p_4 = 0.24$. Nach Durchführung der Simulation mit der Anzahl von $n = 1000$ Realisierungen für beide Objekte ergeben sich die in Abb. 3.31 gezeigten mittleren Geometrien der Schnittergebnisse. In allen Fällen weichen die mittleren Geometrien von den auf traditionelle Weise zu ermittelten Schnittgeometrien ab. Bei disjunkten Objektgeometrien (Abb. 3.31d) fällt der Vergleich ganz aus, da kein traditioneller Schnitt mehr möglich ist. Ursache der Abweichungen liegt offensichtlich in der Ähnlichkeit von Objektkanten. Die maximale Differenz tritt bei identischem Verlauf auf (z.B. Abb. 3.31a), verringert sich mit zunehmenden Auseinanderrücken der Kanten und verschwindet bei unterschiedlichen Kanten ganz (Abb. 3.31b). Eine theoretische Ableitung des zu erwartenden Verschiebungsbetrages zwischen Simulationsergebnis und traditionellem Schnitt gelingt ansatzweise über die Formulierung einer mathematischen Verschneidungsfunktion und der daraus resultierenden Schätzung des Erwartungswertes (Glemser 1996, Glemser und Henneberg, 1999). Ein ähnliches Verhalten zeigen die Varianzen der Linien (Abb. 3.31), die sich zunehmend von den einfach übertragenen Werten ($s_L = s_0$) unterscheiden, wenn Linien sich im Verlauf annähern. Die einfache Übertragung bestätigt sich nur für unterschiedliche Linien bei einer deutlichen Überlappung der Objekte (Abb. 3.31b). Für den Flächeninhalt eines Schnittobjektes kann festgestellt werden, dass bei deutlicher Überlappung ($p = 1.0$) der gleiche oder ein kleinerer Wert als das traditionelle Ergebnis erzielt wird (Abb. 3.31a,b). Mit abnehmendem Grad der Überlappung ($p < 1.0$) wandelt sich das Bild dahingehend, dass auch größere Werte möglich werden (Abb. 3.31c,d).

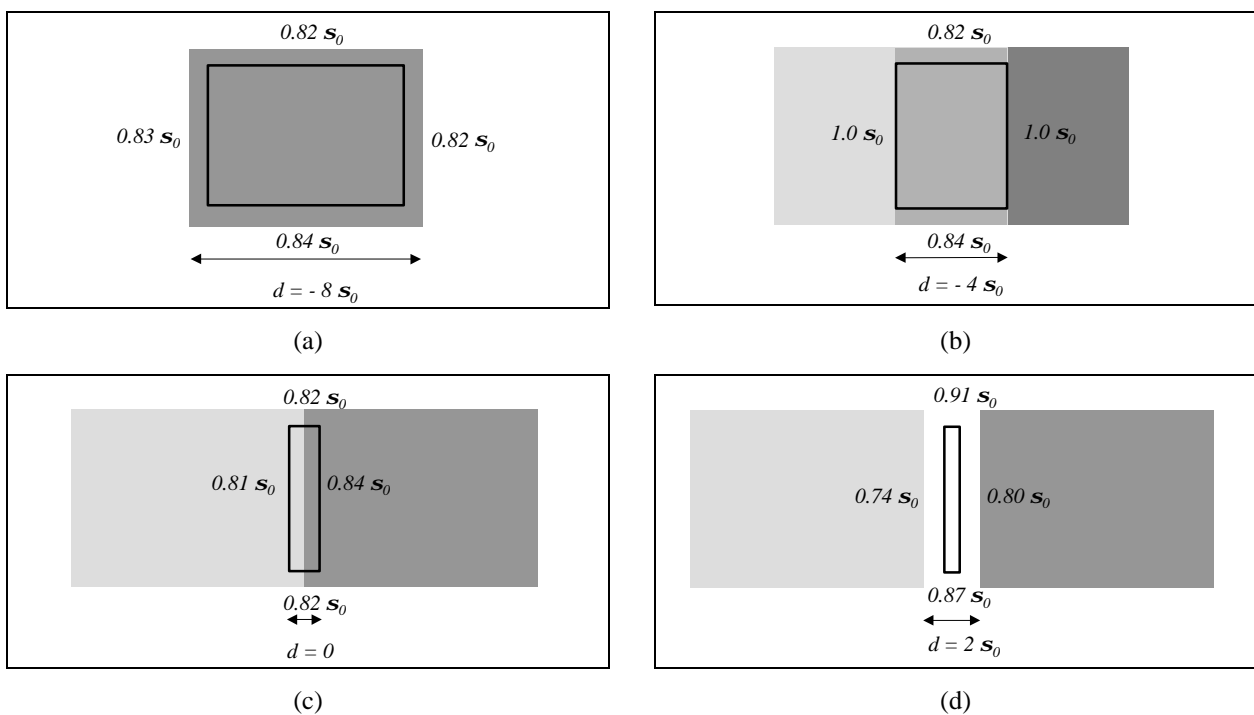


Abb. 3.31: Schnittgeometrie aus Simulationen (schwarze Linien) für vier verschiedene Abstände d_1 (a), d_2 (b), d_3 (c) und d_4 (d) zweier einfacher Beispielobjekte mit Angabe der empirischen Standardabweichungen der Linien des Schnittobjekts.

Kritischer Faktor bei der Anwendung von Simulationen ist der Aufwand, den es zum Erreichen eines ausreichend großen Stichprobenumfangs zu betreiben gilt. Die Anzahl der erzielten Schnittrealisierungen hängt proportional vom Grad der Überlappung ab. In manchen Fällen wird daher eine Erhöhung des Stichprobenumfangs notwendig sein, um eine zuverlässige Bestimmung der Unsicherheitsparameter zu erzielen.

Minimum-Maximum-Modell

Analog zum Ansatz im Stochastischen Modell kann auch hier eine einfache Übertragung der Unsicherheitsparameter (hier: Epsilon-Werte) auf die durch die traditionelle Schnittbildung geteilten und neu formierten Objektlinien definiert werden. Jedoch bleibt die Anwendung wiederum auf die Fälle beschränkt, in denen die traditionelle Verschneidung tatsächlich ein Schnittobjekt erzeugt. Eine allgemeine Fortpflanzung, die eine derartige Einschränkung überwindet, lässt sich auf Basis der minimalen und maximalen Geometrie aufbauen. Dazu sind als Erweiterung der bisherigen Vorgehensweise alle jeweils korrespondierenden Geometrien der Objekte miteinander zu verschneiden. In diesem Fall kann das übliche Verschneidungsverfahren unverändert eingesetzt werden. Im Einzelnen sind für die Verschneidung der beiden Objekte A und B die folgenden Schnittbildungen auszuführen:

$$\begin{aligned} MIN_C &= MIN_A \cap MIN_B, \\ MITTEL_C &= MITTEL_A \cap MITTEL_B, \\ MAX_C &= MAX_A \cap MAX_B. \end{aligned} \quad (3-77)$$

Auf diese Weise wird theoretisch eine vollständige geometrische Beschreibung mit Minimum, Mittel und Maximum für das resultierende Schnittobjekt C definiert. Welche der drei Schnittbildungen in der praktischen Anwendung ein Ergebnis liefert, hängt vom Grad der Überlappung ab. Wenn eine Überlappung existiert, kann mindestens das Maximum erzeugt werden. Fällt eine der Schnittbildungen aus (z.B. die Minima schneiden sich nicht), so ist aus Gründen der Vollständigkeit ein geometrischer Repräsentant anstelle der fehlenden Schnittgeometrie zu generieren, für den aber weiterhin die in der Modelldefinition getroffenen Festlegungen gelten (Kapitel 2.4). Für diesen Zweck eignet sich das Zentroid (Mittelpunkt) der nächstgelegenen noch gebildeten Schnittgeometrie. Somit kann das Minimum als Zentroid des Mittels und das Mittel als Zentroid des Maximums festgelegt werden.

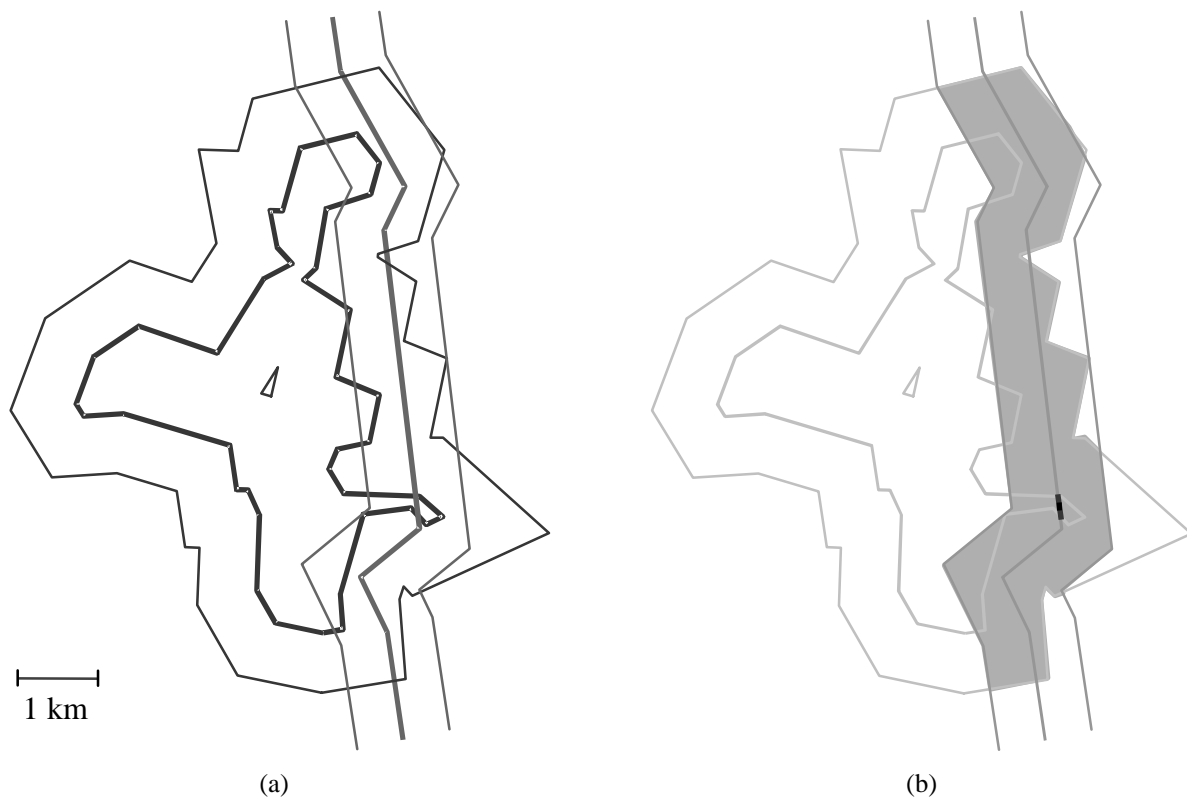


Abb. 3.32: Beispiel der Fortpflanzung der Unsicherheit bei der Verschneidung im Minimum-Maximum-Modell. In (a) sind zwei Objekte jeweils mit Objektgeometrie (dicke Linien) und Maximum bzw. Minimum (dünne Linien) dargestellt. Das Verschneidungsergebnis in (b) zeigt die Objektgeometrie (kleine, dunkle Linie), die dem traditionellen Ergebnis entspricht mit Maximum (flächig gefüllt) und Minimum (schwarzer Punkt).

Als Anwendungsbeispiel dient das bereits bekannte Verschneidungsproblem einer Stromleitung mit einem Stadtgebiet (Abb. 3.30). Zu beiden Objekten wurde eine Minimum- bzw. eine Maximum-Geometrie auf Basis von abgeschätzten Epsilon-Werten (Stromleitung: $e_L = 600\text{ m}$, Stadtgebiet: $e_L = 750\text{ m}$) erzeugt (Abb. 3.32a). Die Schnittbildung aller Geometrien liefert das in Abb. 3.32b gezeigte Ergebnis. Interessant erscheint, dass sich Geometrien unterschiedlichen Typs ergeben: Ein flächenförmiges Maximum, ein linienförmiges Mittel und ein punktförmiges Minimum als Repräsentant für den nicht existierenden Schnitt der beiden Minima.

Fuzzy-Modell

Wie bereits im Minimum-Maximum-Modell und im stochastischen Ansatz stellt die einfache Übertragung der Unsicherheitsparameter eine Möglichkeit der Unsicherheitsfortpflanzung dar. Dabei sind die Unsicherheitsparameter (hier: a, b) in gewohnter Weise auf die geteilten Linien zu übertragen. Doch bleibt die Anwendung weiterhin an die Existenz einer traditionellen Schnittgeometrie gebunden. Aus diesem Grund ist auch für das Fuzzy-Modell nach einer alternativen Vorgehensweise zu suchen, die allgemein verwendet werden kann. Die Modelldefinition, in der ein Objekt als unscharfe Menge an Punkten angesehen wird, legt die Formulierung eines Ansatzes über Punktmengen nahe. Bekanntlich umfasst ein Objekt alle Punkte P_i , die eine Zugehörigkeit von $m(P_i) > 0$ aufweisen. Darauf aufbauend lässt sich das Schnittobjekt C folgendermaßen formulieren:

$$C = \{ (P_i / m_c) \} = A \cap B. \quad (3-78)$$

Als Elemente kommen sämtliche Punkte der Eingangsobjekte A und B in Frage. Die unscharfe Verschneidung kann immer noch als eine logische UND-Operation beschrieben werden, bei der zusätzlich die Unsicherheit in Form der Punktzugehörigkeiten auf das Ergebnis fortzupflanzen ist. Die gesuchte Zugehörigkeit eines beliebigen Punktes P zum Schnittobjekt berechnet sich nach:

$$m_c(P) = \min(m_a(P), m_b(P)). \quad (3-79)$$

Die im Allgemeinen unendliche Anzahl an möglichen Objektpunkten lässt sich auch hier sinnvoll durch die Definition eines Rasters reduzieren. Für die Bestimmung der Rastergröße gelten die Festlegungen, die zur Erzeugung einer Matrixdarstellung getroffen wurden (Kapitel 3.3.2). Jedes Objekt hat ein eigenes Raster zu erhalten. Zur Lösung der Verschneidung sind die Raster der Eingangsobjekte einander zu überlagern und für jeden resultierenden Rasterpunkt die Fortpflanzung der Zugehörigkeit zu berechnen, so dass als Ergebnis wiederum ein Raster für das Schnittobjekt entsteht. Eine Repräsentation der Objektgeometrie kann durch Defuzzifizierung mit Hilfe eines Schwellwertes a_s (Alpha-Schnitt) gefunden werden. Die dadurch entstehende scharfe Punktmenge lässt sich über eine Raster-Vektor-Transformation zurück in eine Vektordarstellung bringen.

Wie bereits im Stochastischen Modell diskutiert, besteht auch hier keine praktikable Möglichkeit, die mittlere Geometrie zu bestimmen. Dennoch kann ein Übergang von Punktzugehörigkeiten auf eine Parameterform aufgebaut werden. Dazu ist die mittlere Geometrie als bisheriger Bezug der Zugehörigkeitsfunktionen durch die maximale Geometrie (Geometrie mit Schwellwert $a_s = 0$) auszutauschen. Als Auswirkung davon ist die lineare Zugehörigkeitsfunktion anzupassen (Abb. 3.33).

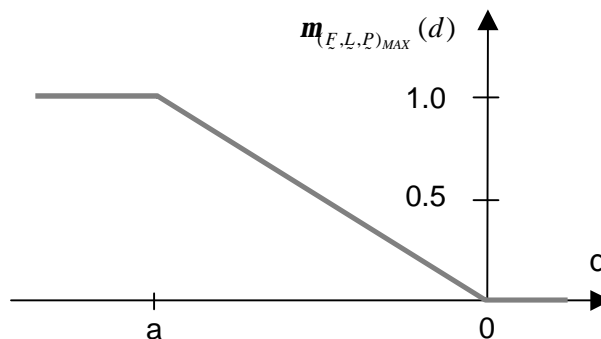


Abb. 3.33: Zugehörigkeitsfunktion zum Maximum als Bezugsgeometrie.

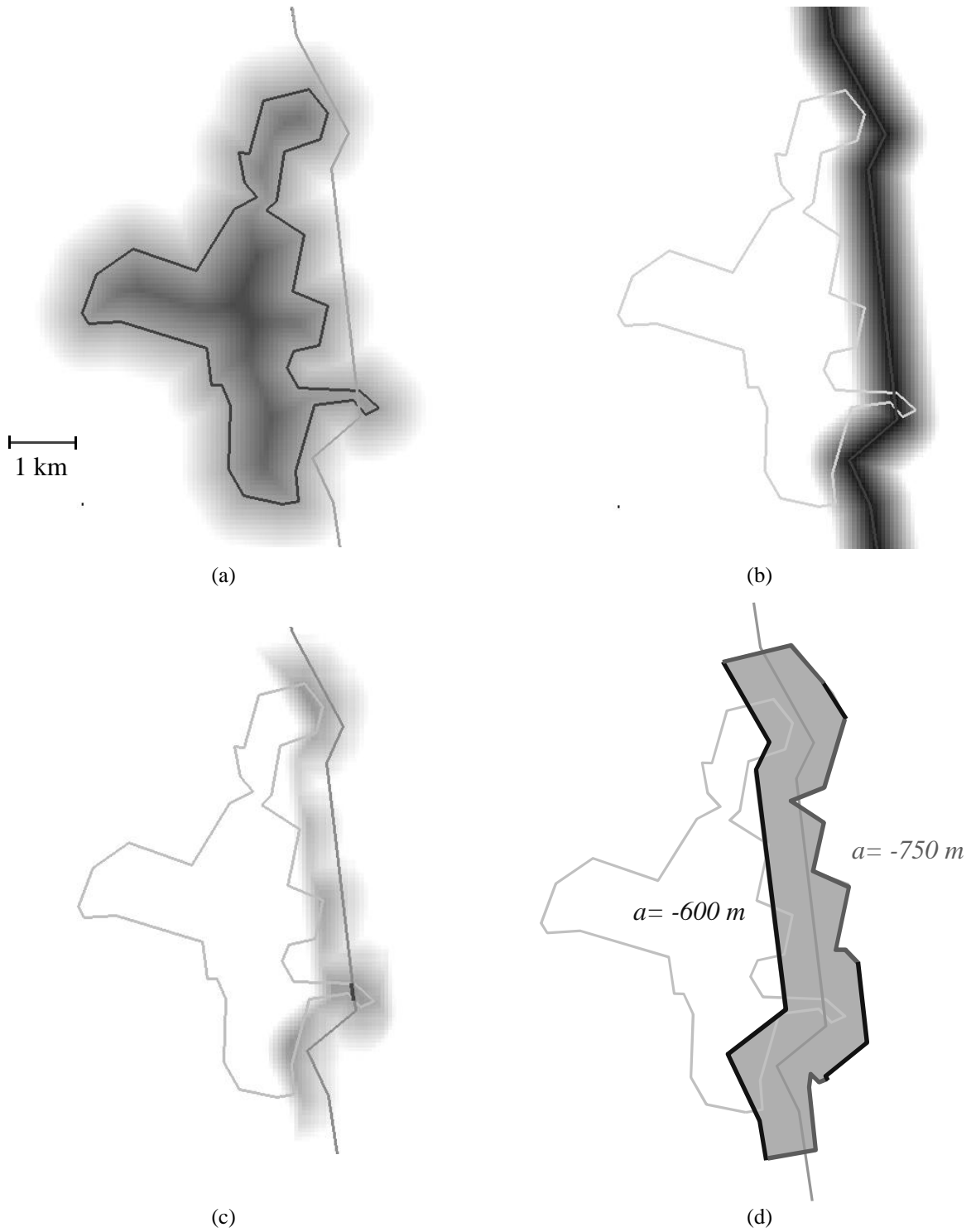


Abb. 3.34: Beispiel der Fortpflanzung der Unsicherheit bei der Verschneidung im Fuzzy-Modell. Dargestellt sind zwei Objekte mit jeweils einer Matrixdarstellung ihrer Unsicherheit (a und b), sowie die Unsicherheit des Schittergebnisses (c) und die segmentweise Abschätzung der Unsicherheitsparameter für das Maximum als Bezugsgeometrie (d).

In neuer Form lautet sie:

$$\mathbf{m}_{(E, L, P)_{MAX}}(d) = \begin{cases} 1 & \text{für } d \leq a, \\ \frac{d}{a} & \text{für } a < d \leq 0 \text{ und } a \neq 0, \\ 0 & \text{für } d > a, \end{cases} \tag{3-80}$$

mit a als Parameter, der in gewohnter Weise zu allen Objektklinien abzulegen ist. Aufgrund der veränderten Bezugsgeometrie (Maximum) gilt die neue Zugehörigkeitsfunktion für alle Objekttypen in gleichem Maße. Zu beachten ist, dass sowohl der Parameter a als auch die individuellen Punktdistanzen d nur Werte kleiner als Null annehmen. Zur Bestimmung des Parameters a genügen zwei Punkte (P_1, P_2) mit unterschiedlichen Zugehörigkeitswerten. Zusätzlich müssen die Punkte in Richtung der Variation (senkrecht zur Linienrichtung) angeordnet sein und möglichst nahe zur betreffenden Objektklinie liegen. Wenn beispielsweise der erste Punkt direkt aus dem Rand des Maximums gewählt wird, dann bestimmt sich der gesuchte Parameter nach:

$$a = -\frac{\Delta}{\mathbf{m}_{\zeta}(P_2)} \quad (3-81)$$

mit Δ als Abstand der beiden Punktpositionen P_1 und P_2 und $\mathbf{m}_{\zeta}(P_2)$ als Zugehörigkeit für den zweiten Punkt.

Die Anwendung der Verschneidung im Fuzzy-Modell soll ebenfalls am Beispiel von Stromleitung und Stadtgebiet gezeigt werden. Grundlage bilden die abgeschätzten Werte für die Parameter der Eingangsobjekte (Stromleitung: $a = 600 \text{ m}$, $b = 50 \text{ m}$ und Stadtgebiet: $a = 750 \text{ m}$), aus denen jeweils eine Rasterdarstellung abgeleitet wurde (Abb. 3.34a und Abb. 3.34b). Die Fortpflanzung der Zugehörigkeiten erzeugt wiederum ein Raster als Repräsentation der unscharfen Ergebnismenge (Abb. 3.34c). Über Bildung des Maximums lassen sich Parameter zu den einzelnen Objektklinien bestimmen, die zur neuen Zugehörigkeitsfunktion gehören (Abb. 3.34d).

3.5.5 Diskussion

Die Fortpflanzung der Unsicherheit auf das Analyseergebnis ist für die diskutierten geometrischen Aufgaben grundsätzlich mit allen Unsicherheitsmodellen möglich. Der Ansatz im Stochastischen Modell stützt sich zumeist auf eine Varianzfortpflanzung, zu der die Analyseaufgabe in Form eines mathematisch funktionalen Modells vorzuliegen hat. Bei der strengen Lösung werden Informationen zu den Abhängigkeiten der Eingangsdaten benötigt, die in der Fortpflanzung zu berücksichtigen sind. In vielen Fällen können Vereinfachungen formuliert werden (z.B. Unabhängigkeit, gleiche Unsicherheit), die den praktischen Gebrauch erleichtern. Probleme bereiten komplexe Analysen, die durch keine einfachen mathematischen Zusammenhänge beschrieben werden können. In diesem Fall sind alternative Ansätze (z.B. Simulationen) anzuwenden. Die Fortpflanzung im Minimum-Maximum-Modell lässt sich in einfacher Weise stets durch Analyse des Minimums und des Maximums realisieren. Dazu ist die zu untersuchende Analyseverfahren zusätzlich mit Minimum und Maximum durchzuführen. Die Ergebnisse verdeutlichen dann die maximal mögliche Variation. Da so immer nur eine Bereichsangabe entsteht, sind keine quantitativen Aussagen daraus abzuleiten. Ansatz für die Fortpflanzung im Fuzzy-Modell bildet die Betrachtung der Lösung als eigene unscharfe Menge, zu der es die Zugehörigkeitsfunktion zu bestimmen gilt. Bei Verwendung einer linearen Funktion genügt dazu die Kenntnis zweier Elemente der neuen unscharfen Menge, die im Allgemeinen einfach aufzufinden sind. Für komplexere Funktionen werden entsprechend mehr Elemente benötigt, was unter gewissen Umständen problematisch sein kann.

Die Verschneidung als Vertreter einer aufwendigen Analyseverfahren bereitet bei der Behandlung der Fortpflanzung allen Unsicherheitsmodellen konzeptionelle Schwierigkeiten. Ein bedeutendes Problem liegt in der Bestimmung der mittleren Geometrie, die in den Modelldefinitionen eine wichtige Rolle spielt. Sie entsteht in Anlehnung an die übliche Vorgehensweise bei der Verschneidung aus dem Schnitt der mittleren Geometrien der Eingangsobjekte. Nicht in allen Fällen schneiden sich die mittleren Geometrien, obwohl ein Schnittobjekt aufgrund des Überlappungsgrades existieren kann. Im Stochastischen Modell wird aber die mittlere Geometrie für alle Schnittobjekte benötigt, da sich die Varianzen als zentrale Momente darauf beziehen. Simulationen bieten eine Lösungsmöglichkeit. Sie gestalten sich jedoch aufwendig, insbesondere wenn eine geringe Überlappung vorliegt, da immer für eine ausreichend große Stichprobe zu sorgen ist. Ansonsten leidet die Zuverlässigkeit der Schätzung. Aus diesem Grund werden Simulationen nur in ausgewählten Situationen für den praktischen Gebrauch einzusetzen sein. Als Alternative ist es denkbar, dass sich aufbauend auf den Erkenntnissen der durchgeführten Simulationstests eine allgemeine funktionale Ableitung entwickeln lässt, die dann einen universellen Einsatz erlaubt. Das Minimum-Maximum-Modell verzichtet dagegen ganz darauf, ein Bestimmungsverfahren bereitzustellen. Existiert der Schnitt der Geometrie nicht, dann wird ersatzweise ein Repräsentant (z.B. Zentroid) eingesetzt, der das fehlende Schnittergebnis ersetzt. Einzige Funktion des Repräsentanten ist es, die Vollständigkeit der Modellkonzeption zu erhalten. Prinzipiell könnte das Modell aber auch ohne eine solche Lösung auskommen. Dagegen besitzt die mittlere Geometrie im Fuzzy-Modell den Status einer Bezugsgeometrie, die den Ursprung für die Zugehörigkeitsfunktion markiert. Dennoch wird auch hier auf eine Schätzung verzichtet und statt dessen die mittlere Geometrie durch eine beliebige andere Bezugsgeometrie ersetzt (z.B. Maximum). Dabei muss beachtet werden, dass die Definition der Zugehörigkeitsfunktion anzupassen ist.

Obwohl Objektgeometrie und mittlere Geometrie in den Unsicherheitsmodellen getrennt voneinander behandelt werden, sind in der praktischen Anwendung häufig beide einander gleichgesetzt und damit identisch. Daher gelten die diskutierten Probleme bei der Bestimmung der mittleren Geometrie in gleichem Maße auch für die Objektgeometrie, sofern an einer gemeinsamen Form festgehalten werden soll. Das vorgestellte Datenmodell schreibt aber keine Verknüpfung vor, so dass die Objektgeometrie eine beliebige Ausgestaltung annehmen kann, die noch eine gültige Repräsentation des Objektes darstellt. Diese Verallgemeinerung eröffnet die Möglichkeit, die Objektgeometrie in einfacher und flexibler Weise unabhängig von der mittleren Geometrie zu bestimmen. Beispielsweise können dazu an bestimmte Sicherheitsniveaus angepasste Geometrien (z.B. das Maximum im Minimum-Maximum-Modell) erzeugt werden, die dann bereits in der Objektdarstellung einen Eindruck über die enthaltene Unsicherheit vermitteln (Kapitel 3.3.2).

Das Stochastische Modell bietet durch die Verschneidung von Punktmengen eine interessante Variante zur Varianzfortpflanzung. Hierbei bestehen Ähnlichkeiten zum Ansatz im Fuzzy-Modell, das sich ebenfalls auf Punktmengen stützt. Im Unterschied dazu betrachtet das Stochastische Modell aber die Mengen als scharf und bewertet dann einzelne Teilmengen im Hinblick auf ein bestimmtes Ereignis. Auch unterscheiden sich die Bewertungsfunktionen beider Ansätze in der Weise, dass die Zugehörigkeiten von Punkten sich an den schwächsten Eingangswerten orientieren (Minimum-Operator), während zur Bestimmung der Wahrscheinlichkeiten alle Eingangswerte einbezogen werden. Insofern ist ein Vergleich oder eine direkte gegenseitige Transformation der Ergebnisse nicht ohne weiteres möglich. Zwar erzeugt die Verschneidung der Punktmengen anschauliche Visualisierungen der enthaltenen Unsicherheit, aber ohne zusätzliche Schätzung der mittleren Geometrie können die Ergebnisse nicht uneingeschränkt für weitere Analysen (z.B. eine anschließende Flächenbestimmung) genutzt werden.

Die aufgrund der Verschneidungsmethode naheliegende einfache Übertragung der Unsicherheitsparameter auf die Schnittgeometrie hängt vom Schnitt der mittleren Geometrien der Eingangsobjekte ab und kann deswegen nur bedingt eingesetzt werden. Aber auch in den Fällen, in denen ein Schnitt existiert, ist eine einfache Übertragung in Frage zu stellen. Betrachtet man die Ergebnisse des Simulationstests genauer (Abb. 3.31), so stellt man fest, dass nur bei ausreichenden Lagedifferenzen in den Eingangsgeometrien eine einfache Übertragung bestätigt werden kann. Aus diesem Grund wird dieses Verfahren nur in Ausnahmefällen anzuwenden sein.

Die vorgestellten Ansätze sind alle so definiert, dass gemäß der gestellten Forderung die Fortpflanzung als ein automatischer Prozess im Hintergrund der Analysemethoden ablaufen kann, ohne dass spezielle Nutzereingaben notwendig werden. Wird die Fortpflanzung als Bestandteil der Methode implementiert, dann steht mit dem Ergebnis gleichzeitig auch dessen Unsicherheit für weitere Beurteilungen bereit. In einem Gesamtsystem darf die Fortpflanzung aber nicht auf einzelne geometrische Analyseaufgaben beschränkt bleiben, sondern ist in allen Verarbeitungen zu berücksichtigen. Im Zusammenhang mit der Geometrie ist beispielsweise eine Ausdehnung auf topologische Operatoren von Interesse, deren Ergebnis maßgeblich von der geometrischen Unsicherheit beeinflusst wird. Hierzu ist auf einige Arbeiten zu verweisen, die sich diesem speziellen Thema widmen (Winter, 1996, Clementini und Di Felice, 1996, Glemser et al., 1999). Im Weiteren bedarf es grundsätzlich der Entwicklung von Ansätzen, die eine Fortpflanzung aller Komponenten der Datenunsicherheit über beliebig komplexe Analysen mit geometrischen, topologischen und attributiven Fragestellungen ermöglichen.

4 Diskussion der Ergebnisse

4.1 Eigenschaften der Unsicherheitsmodelle

Die Erläuterungen in Kapitel 3 haben gezeigt, dass die Anwendungsaufgaben im Bereich der geometrischen Unsicherheit prinzipiell mit allen drei definierten Unsicherheitsmodellen zu lösen sind. Dennoch bestehen Unterschiede, die dazu führen, dass sich ein bestimmtes Modell in manchen Fällen besser eignet als die anderen und deshalb vorzuziehen ist. Bedenkt man die Vielzahl an möglichen Anwendungen mit ihren verschiedenen Ausrichtungen, dann wird klar, dass nicht generell ein Modell als das beste gelten kann. Vielmehr bedarf es für jeden Einzelfall einer genauen Prüfung der Modelle gemäß den gestellten Anforderungen, um letztendlich die richtige Wahl zu treffen. Als Entscheidungshilfe für den Nutzer werden im Folgenden für jedes Modell einige charakteristische Merkmale diskutiert, anhand derer eine anwendungsspezifische Bewertung der Modelle durchgeführt werden kann.

Das Minimum-Maximum-Modell zeichnet sich durch eine einfache Handhabung aus. Es erfordert den geringsten Integrationsaufwand, da sich die vorhandenen Bestandteile der GIS-Komponenten weitgehend auch für die Bearbeitung der Unsicherheit nutzen lassen können. Z.B. können die bestehenden Analysemethoden auch für die Fortpflanzung der Unsicherheit eingesetzt werden. Aus diesem Grund eignet sich dieses Modell im Besonderen für eine Erweiterung von bestehenden Systemen, für die es grundsätzlich möglich sein müsste, die Modelldefinitionen umzusetzen und die erläuterte Verarbeitung der Unsicherheit in den einzelnen GIS-Komponenten auszuführen. Als Systemtyp genügt ein Vektor-GIS. Die vorliegende qualitative Unsicherheitsbeschreibung schränkt dagegen den Einsatz auf Anwendungen mit geringen Anforderungen an die Behandlung der Unsicherheit ein. Dazu zählen beispielsweise Anwendungen, die nur in Ausnahmefällen Angaben zur Unsicherheit benötigen. In diesen Fällen veranschaulichen Variationsbreite oder Minimum und Maximum die enthaltene Unsicherheit in ausreichender Weise. Sind speziell kritische Analysen von Interesse, d.h. Analysen, bei denen für das Ergebnis höchste Sicherheitsansprüche gelten, so kann ebenfalls auf das Minimum-Maximum-Modell zurückgegriffen werden. Dies ist beispielsweise dann der Fall, wenn das gleichzeitige Vorkommen zweier thematischer Eigenschaften (z.B. Altlast und Wasserschutzgebiet) sicher auszuschließen ist. Es werden darin stets die maximalen Einflüsse berücksichtigt, so dass Minimum und Maximum immer die größtmögliche Variation beschreiben.

Stellt die Anwendung hohe Anforderungen an die Integration der Unsicherheit, dann ist entweder auf das Stochastische Modell oder das Fuzzy-Modell zurückzugreifen. Die darin verfügbare quantitative Beschreibung ermöglicht eine vielfältige und detaillierte Behandlung der Unsicherheit. Es können Aussagen zu beliebigen Sicherheitsniveaus abgeleitet und dadurch die Weiterverarbeitung für beliebige Zwecke offengehalten werden. Die Nutzung der Matrixdarstellung als ansprechende Visualisierungsmöglichkeit und zur vereinfachten Analyse erfordert jedoch eine hybride GIS-Umgebung, in der sowohl Vektor- als auch Rasterdaten gemeinsam darzustellen sind. Von den drei Modellen besitzt das Stochastische Modell die in der Praxis am besten akzeptierte theoretische Basis. Insofern liegen hierfür reichhaltige Erfahrungswerte vor, die einen Einsatz des Modells erleichtern. Das Modell ist besonderes günstig einzusetzen, wenn die zu analysierenden Zusammenhänge in Form von mathematischen Funktionen beschrieben werden können. Dann ist die Übertragung der Unsicherheit durch die bekannte Varianzfortpflanzung zu lösen. Dagegen bereiten komplexe Analysen Schwierigkeiten, die sich häufig nur aufwendig durch den Einsatz von Simulationen beheben lassen. In allen Verarbeitungsschritten ist stets auf mögliche Korrelationen der einfließenden Daten zu achten, die das Ergebnis der Unsicherheitsbetrachtungen stark beeinflussen können (z.B. bei der Berechnung der kürzesten Distanz zwischen zwei identischen Kanten, die vollständig miteinander korreliert sind). Insofern sind die Korrelationswerte zusätzlich mitzuverwalten. In den beiden anderen Modellen existieren keine allgemeingültigen Konzepte, um Korrelationen angemessen einzubringen. Die Berechnungen ermitteln dort immer die maximalen Auswirkungen. Für eine entsprechende Berücksichtigung sind individuelle Bedingungen einzuführen. Problematisch ist weiterhin die übergeordnete Bedeutung der mittleren Geometrie, die als Bezugsgeometrie für die Beschreibung der Verteilung einen unverzichtbaren Bestandteil der Modelldefinition im Stochastischen Modell darstellt. Sie ist für jedes Ergebnis, das eine neue oder veränderte Geometrie erzeugt, zu schätzen, auch wenn dies nur unter erheblichem Aufwand erreicht werden kann (z.B. bei der Verschneidung). In den beiden anderen Modellen ist diese herausragende Stellung der mittleren Geometrie nicht anzutreffen. Zwar verwendet das Fuzzy-Modell die mittlere Geometrie ebenfalls als Bezugsgeometrie, doch kann diese theoretisch durch eine beliebige andere ersetzt werden. Im Minimum-Maximum-Modell ist es möglich, ganz ohne mittlere Geometrie auszukommen.

Wichtiges Merkmal des Fuzzy-Modell ist seine Flexibilität, sich an Besonderheiten anzupassen. Dazu zählt in erster Linie die Möglichkeit, die Zugehörigkeitsfunktion so zu definieren, dass sie optimal die gestellte Anforderung erfüllt. Dabei sind beliebige Formen erlaubt (nicht nur lineare Funktionen), die sich auf beliebige Bezugsgeometrien als Grundlage stützen dürfen. Die flexible Verwendung bringt häufig das Problem mit sich, dass erhebliches Fachwissen benötigt wird, um eine adäquate Bearbeitung zu gewährleisten. Insofern ist das Fuzzy-Modell mit seinen Möglichkeiten

nur von Experten voll auszuschöpfen. Gemäß seiner theoretischen Basis eignet sich das Fuzzy-Modell insbesondere für die Anwendung innerhalb unscharfer Analyseaufgaben (z.B. Suche nach „naheliegenden“ oder „weitentfernten“ Objekten). Dazu ist die unscharfe Aussage ebenfalls durch eine Zugehörigkeitsfunktion zu beschreiben und gemeinsam mit der Zugehörigkeitsfunktion der unsicheren Daten zu verarbeiten. Auch hierfür ist Expertenwissen unentbehrlich.

Für Anwendungen, die ganz unterschiedliche Anforderungen in sich vereinen, ist eine hybride Nutzung der Unsicherheitsmodelle denkbar. Dazu sind alle drei Modelldefinitionen zu integrieren, ohne dass aber sämtliche Methoden für alle Modelle zu realisieren sind. Vielmehr lassen sich gezielt die Stärken ausnutzen, indem die jeweilige Verarbeitung in dem Modell ausgeführt wird, in dem eine Lösung am besten gelingt. Ist z.B. eine kritische Analyse durchzuführen, so kann diese mit Hilfe des Minimum-Maximum-Modells bearbeitet werden, während beim Vorliegen einer mathematischen Funktion das Stochastische Modell zu nutzen ist. Den theoretischen Rahmen für die dazu häufig notwendigen Modellwechsel stellen die in Kapitel 2.4.4 definierten Transformationen bereit. Jedoch decken die behandelten Anwendungsbeispiele einige Hindernisse auf, die eine beliebige hybride Nutzung einschränken. Der Übergang zum Minimum-Maximum-Modell ist in den meisten Fällen problemlos möglich, da dieses Modell in gewisser Weise als eine Teilmenge in den anderen Modellen wiedergefunden werden kann. Schwieriger gestaltet sich der Wechsel zum Stochastischen Modell, da stets die mittlere Geometrie benötigt wird, die aber in den beiden anderen Modellen nicht zwingend vorhanden sein muss. Auch kann die Bestimmung der zugrundeliegenden Verteilung Probleme bereiten, insbesondere im Falle des Minimum-Maximum-Modells, das nur eine qualitative Beschreibung bereithält. Gleiches gilt für den Übergang zum Fuzzy-Modell, bei dem die Form der Zugehörigkeitsfunktion festzulegen ist. Während beim Minimum-Maximum-Modell nur eine lineare Form möglich ist, kann für das Stochastische Modell eine angepasste Form aus der Dichtefunktion hergeleitet werden.

Zusammenfassend kann als Nutzerempfehlung festgehalten werden, dass das Minimum-Maximum-Modell dann anzuwenden ist, wenn geringe Anforderungen an die Behandlung der Unsicherheit gestellt werden oder wenn eine einfach zu realisierende Lösung gefragt ist. Bei hohen Anforderungen kommen sowohl Fuzzy- als auch Stochastisches Modell in Frage, wobei man sich für das Stochastische Modell entscheiden sollte, wenn noch nicht endgültig geklärt werden kann, welche zukünftigen Aufgaben zu bearbeiten sind oder wenn ganz unterschiedliche Anforderungen auftreten. Ist dagegen bereits klar, dass Flexibilität eine wichtige Rolle spielt, so ist das Fuzzy-Modell vorzuziehen.

4.2 Anmerkungen zur praktischen Umsetzung

Durch die Integration der Unsicherheit entstehen einige Neuerungen in der Behandlung der Geometrie, die sich nicht immer problemlos umsetzen lassen. Statt genau einer Geometrie pro Objekt sind gleichzeitig mehrere Beschreibungen zu ermöglichen. Beispielsweise ist im Stochastischen Modell zwischen Objektgeometrie und mittlerer Geometrie zu unterscheiden. Die Objektgeometrie behält zwar weiterhin ihre Bedeutung als räumlicher Repräsentant des Objektes, aber die mittlere Geometrie ist zusätzlich als Bezugsgeometrie für die Unsicherheitsmodellierung einzuführen. Im Unterschied zur früheren Behandlung weist die Objektgeometrie aber keine unveränderliche Form mehr auf, sondern kann verschiedene Ausprägungen annehmen, die sich in Abhängigkeit vom gewählten Sicherheitsniveau ergeben. Um die Präsentation der Daten aber nicht zu überladen und die Übersichtlichkeit der Darstellung zu erhalten, ist darauf zu achten, dass die Unsicherheit nur auf Anfrage als Hintergrundthema dargestellt wird. Nicht zu vermeiden ist eine Vergrößerung des Aufwandes durch die individuelle Erfassung der Unsicherheit, wodurch Probleme bei der Nutzerakzeptanz entstehen können. In Einzelfällen schaffen automatische Vergleichsverfahren mit anderen Datensätzen Abhilfe. Ansonsten gilt es, den zusätzlichen Aufwand durch effektive Unterstützungsfunktionen möglichst gering zu halten. Die Fortpflanzung der Unsicherheit bei der Datenanalyse sollte möglichst ohne Nutzerinteraktion auskommen. Die benutzten Methoden sind derart zu erweitern, dass die Fortpflanzung stets als automatischer Hintergrundprozess abläuft, dessen Ergebnis gleichzeitig mit den Analyseberechnungen bereitsteht und bei Bedarf abgerufen werden kann.

Für die Nutzerakzeptanz ist die Konformität der Entwicklungen mit den aktuellen Standardisierungsbestrebungen förderlich. Auf europäischer Ebene wird momentan eine Norm zum übergeordneten Thema „Datenqualität“ erarbeitet, die zukünftig das standardisierte Beschreiben von Qualitätsinformationen regeln soll (DIN/CEN/ISO, 1998). Aufgabe der Normierung ist es, einen Rahmen vorzugeben, wie einerseits ein Datenerzeuger Qualitätsinformationen zu Datensätzen anzugeben hat und wie andererseits Anforderungen an benötigte Daten von Seiten eines Datennutzers zu formulieren sind. Die getroffenen Festlegungen können mit allen Ebenen von Geodatensätzen, vom ganzen Datensatz über bestimmte Untergliederungen bis zu individuellen Datenelementen, verknüpft werden. Nicht enthalten sind Techniken, wie Qualität gemessen, abgeschätzt, verwaltet oder kontrolliert werden kann. Auch werden keine Aussagen zur Qualität von einzelnen Produkten getroffen.

Die Norm sieht vor, dass sich die anzugebenden Qualitätsinformationen aus den Qualitätselementen Herkunft, Verwendung, Qualitätsparameter und Homogenität zusammensetzen (Abb. 4.1). Mindestens ein Qualitätselement soll zwingend zu jedem Datensatz anzugeben sein. Für diesen Zweck eignet sich besonders die Herkunft, die Angaben zur Hersteller-

organisation, zu den Datenquellen und zum Verarbeitungsprozess umfasst. In der Verwendung können bekannte erfolgreiche Anwendungsgebiete der Daten aufgezählt werden. Kernstück bilden die Qualitätsparameter, die eine detaillierte Beschreibung der konkreten Qualitätssituation liefern. Die Homogenität macht Angaben zur einheitlichen Gültigkeit der Qualitätsparameter über den gesamten Umfang der beschriebenen Ebene.

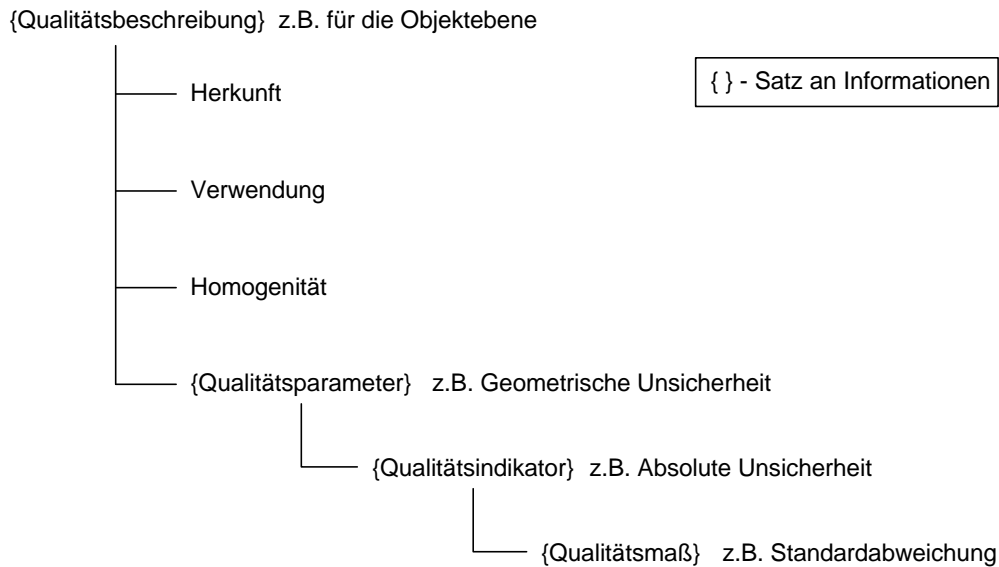


Abb. 4.1: Qualitätsschema des europäischen Normentwurfes zur Datenqualität.

Qualitätsparameter sind Träger der eigentlichen Qualitätsinformationen. Sie umfassen alle messbaren Aspekte der Güte von Datensätzen. Beispielsweise kann die geometrische Unsicherheit als ein solcher Aspekt angesehen werden. Die Parameter dokumentieren die Güte im Vergleich zur konzeptuellen Realität, die man als idealen Datensatz ansehen kann, der exakt den geforderten Spezifikationen entspricht. Die Norm gibt konkret eine Reihe von Qualitätsparametern vor, von denen erwartet wird, dass sie den jetzigen und auch zukünftigen Anforderungen gerecht werden. Dazu gehören Positionsgenauigkeit, semantische Genauigkeit, zeitliche Genauigkeit, Vollständigkeit und logische Konsistenz. Die Aufzählung entspricht einer Teilmenge der in Kapitel 1.1 diskutierten Qualitätsmerkmale. Eine anwendungsabhängige Erweiterung um zusätzliche Parameter ist vorgesehen und problemlos möglich. Die in der Arbeit behandelte geometrische Unsicherheit lässt sich mit der Positionsgenauigkeit gleichsetzen. Jeder Parameter wird durch Qualitätsindikatoren und diese ihrerseits durch Qualitätsmaße beschrieben. Mögliche Indikatoren für die Positionsgenauigkeit sind z.B. die relative horizontale Lagegenauigkeit oder die absolute Lagegenauigkeit in Bezug auf ein Referenzsystem. Qualitätsmaße enthalten das Ergebnis der Qualitätsmessung, die zur Prüfung des Datenbestandes ausgeführt wird. Standardabweichungen oder Fehlerellipsen sind Beispiele zur Positionsgenauigkeit.

Stochastisches Modell	Minimum-Maximum-Modell	Fuzzy-Modell
Qualitätsindikator: Absolute Unsicherheit	Qualitätsindikator: Absolute Unsicherheit	Qualitätsindikator: Absolute Unsicherheit
Maß: Standardabweichung Wertetyp: Real	Maß: Minimum Wertetyp: geometrisches Primitiv	Maß: Zugehörigkeitsfunktion Wertetyp: Funktion
Maß: Breite Wertetyp: Real	Maß: Maximum Wertetyp: geometrisches Primitiv	Maß: Parameter a Wertetyp: Real
Maß: mittlere Geometrie Wertetyp: geometrisches Primitiv	Maß: Mittel Wertetyp: geometrisches Primitiv	Maß: Parameter b Wertetyp: Real
	Maß: Epsilon Wertetyp: Real	Maß: Bezugsgeometrie Wertetyp: geometrisches Primitiv
		Maß: Fuzzy-Index Wertetyp: Real

Abb. 4.2: Einbettung der entwickelten Modelldefinitionen in den Rahmen des europäischen Normentwurfes.

Die Norm schreibt nicht vor, welche bestimmten Indikatoren oder Maße zu einem Parameter anzugeben sind. Die Auswahl bleibt vollständig dem Nutzer überlassen. Aus diesem Grund können die in der Arbeit entwickelten Unsicherheitsmodelle unproblematisch in die Beschreibung eingefügt werden. Als Indikator findet die allgemeine Bezeichnung „absolute geometrische Unsicherheit“ Verwendung, um so eine häufig implizierte Verbindung von Genauigkeit mit dem Stochastischen Modell zu vermeiden. Zu allen Modellen sind jeweils mehrere Qualitätsmaße zu definieren, die die notwendigen Unsicherheitsparameter abdecken. Maße und ihre Werte können im Einzelnen Abb. 4.2 entnommen werden. Beim Betrachten der Wertetypen fällt auf, dass neben einfachen Wertetypen auch abstrakte Datentypen auftreten. Dazu zählen insbesondere Funktionsbeschreibungen für die Zugehörigkeitsfunktionen im Fuzzy-Modell und geometrische Angaben, wie mittlere Geometrie, Minimum, Maximum oder sonstige Bezugsgeometrien in allen Modellen. Zur konformen Verwendung dieser Typen sind weitere Normen einzubeziehen.

Als Alternative zu den Normungsbestrebungen sind im GIS-Bereich die Arbeiten des Open GIS Consortiums (OGC, 2000) als wichtige Entwicklungstendenz zukünftiger Systemumgebungen zu beachten. Bei dieser Organisation handelt es sich um eine nicht-kommerzielle Vereinigung von freiwilligen Mitgliedern, die sich aus GIS-Herstellern, Anwendern und Forschungseinrichtungen zusammensetzt. Ziel des Zusammenschlusses ist es, gemeinsame Spezifikationen zu verabschieden, die eine offene Gestaltung der GIS-Umgebungen ermöglichen. Dahinter verbirgt sich die grundlegende Idee, zukünftige Aufgaben durch eine interoperable Geodatenverarbeitung zu lösen. Eine interoperable Anwendung äußert sich darin, dass mehrere spezialisierte Systeme zu einer integrierten Problemlösung beitragen, wobei die Einzelsysteme netzwerkweit verteilt sein können. Auf diese Weise wird insbesondere die Bearbeitung komplexer Aufgaben ermöglicht. In einem solchen Ansatz kommt der Datenqualität eine gewichtige Rolle zu, da unterschiedliche Daten in vielen verschiedenen Systemen weiterverarbeitet werden, über die dem Anwender im Allgemeinen kein Detailwissen vorliegt. Daher wendet sich ein spezieller Abschnitt der Spezifikation zum abstrakten Datenmodell dem Thema Datenqualität zu (The OpenGIS Abstract Specification Model – Topic 9: Quality, (OGC, 2000)). Darin werden Qualitätsbeschreibungen der Ebene der Metadaten zugeordnet, so dass ein Nutzer in die Lage versetzt wird, die angebotenen Daten auf ihre Eignung gegenüber den Anwendungsbedürfnissen zu prüfen. Es gilt das Prinzip der Angabe eines Qualitätsmaßes anstelle der Angabe des möglichen Eignungsbereiches der Daten. Die Spezifikation konzentriert sich im Wesentlichen auf den Aspekt der Positionsgenauigkeit von Punkten und definiert eine Reihe möglicher Qualitätsmaße. Es werden ausschließlich Stochastische Maße diskutiert, wie z.B. Standardabweichungen für einzelne Koordinaten, Kovarianzmatrix der Koordinaten, Punktfehler nach Helmert. Als feste Annahme wird stets eine Normalverteilung zugrunde gelegt. Zu kritisieren ist, dass die Spezifikation keine Ansätze zur Behandlung der Unsicherheit auf Datenebene enthält, obwohl die detaillierte Kenntnis der Unsicherheit in interoperablen Anwendungen eine wichtige Rolle spielt. Zudem bedeutet die Beschreibung der geometrischen Unsicherheit durch Punktgenauigkeiten mit Stochastischen Maßen eine erhebliche Einschränkung, die dem Gedanken der Offenheit entgegensteht und zwangsläufig zu Schwierigkeiten mit Anwendungen führt, für die andere Maße sinnvoller einzusetzen sind. Die Definition eines grundlegenden Rahmens, in den die jetzigen Spezifikationen als wichtige Bestandteile einfließen, könnte diese Probleme beseitigen. Innerhalb einer solchen Lösung ließen sich dann auch die in dieser Arbeit entwickelten Modelldefinitionen und Maße integrieren.

5 Schlussbetrachtungen

5.1 Zusammenfassung

Moderne GIS-Anwendungen profitieren von der Situation, dass Geoinformationen heutzutage zu allen möglichen Themengebieten verfügbar sind. Der Aufwand für eine eigene Datenerfassung lässt sich so durch das Nutzen externer Datenbanken auf ein Minimum reduzieren. Zudem bieten die heutigen Systeme eine Vielzahl an Funktionen, die eine Durchführung von komplexen Analyseaufgaben weitgehend problemlos ermöglichen. In diesem Zusammenhang kommt der Berücksichtigung der Datenqualität eine gestiegene Bedeutung zu. Durch das ausschließliche Arbeiten mit eigenerfassten Daten und den einfach gestellten Aufgaben gelang es einem Anwender bisher, die Qualität der Ergebnisse aufgrund seines Fachwissens selbständig zu bewerten und zu kontrollieren. Die Nutzung von Fremddaten und die Komplexität der Anwendungen erfordern die Entwicklung und den Einsatz neuer Methoden, die diese Aufgabe übernehmen und eine Ableitung der Qualitätsbeschreibung auf automatisiertem Wege erzeugen.

Die vorliegende Arbeit konzentriert sich auf die geometrische Unsicherheit als wichtige Komponente der Datenqualität von räumlichen Daten. Ziel ist es, Konzepte und Methoden für die Integration der geometrischen Unsicherheit in die Systemumgebung zu entwickeln. Im Einzelnen benötigt man dafür die Definition eines Datenmodells, das um die Beschreibung der Unsicherheit erweitert ist. Des Weiteren sind in allen Komponenten eines GIS, von der Erfassung, über die Verwaltung und Analyse bis zur Präsentation, Funktionen zu entwickeln, die eine gemeinsame Verarbeitung von Daten und Unsicherheit ermöglichen.

Ausgangspunkt für die Entwicklung von Konzepten bilden Untersuchungen zur Ursache der geometrischen Unsicherheit. Wesentliche Einflüsse entstehen innerhalb der Datenerfassung bei der Modellbildung, der Diskretisierung und der Messung. Die Größenordnung hängt entscheidend von der jeweiligen Objektkategorie ab. Die Unsicherheit von künstlichen Objekten wird beispielsweise maßgeblich von Messeinflüssen bestimmt, während bei natürlichen Objekten die Modellbildung die dominante Rolle spielt. Allgemein ist davon auszugehen, dass jeder Punkt der Geometrie zu einem gewissen Grad eine unabhängige Variation besitzt. Anstatt einer individuellen Behandlung von unendlich vielen Punkten, wird jedoch eine zusammenfassende linienweise Beschreibung vorgezogen. Aus den existierenden Ansätzen können drei Unsicherheitsmodelle abgeleitet werden, die auf unterschiedliche Weise in der Lage sind, die in einer Linie enthaltene geometrische Variation zu beschreiben. Im Einzelnen werden das Stochastische Modell, das Minimum-Maximum-Modell und das Fuzzy-Modell definiert. Wie aus den Bezeichnungen ersichtlich ist, basieren die Modelle auf ganz unterschiedlichen mathematischen Grundlagen. Charakteristisch für das Minimum-Maximum-Modell ist die qualitative Art und Weise der Beschreibung der Unsicherheit, während sowohl das Stochastische Modell als auch das Fuzzy-Modell eine quantitative Abstufung ermöglichen.

Zur Berücksichtigung der Unsicherheit bedarf es einer Erweiterung des bestehenden Objektmodells zu einem Datenmodell für geometrisch unsichere Objekte. Dazu sind die einzelnen Unsicherheitsmodelle als zusätzliche Bestandteile der Objektgeometrie anzufügen. Für die meisten Anwendungen reicht die Erweiterung um ein bestimmtes Modell aus. Für komplexe Aufgaben, die eine Kombination der einzelnen Merkmale erfordern, müssen im Rahmen einer hybriden Nutzung alle drei Modelle eingebracht werden.

In gleichem Maße wie die Daten selbst ist auch deren Unsicherheit in den verschiedenen Komponenten eines GIS zu behandeln. Dazu werden neue Ansätze und zusätzliche Funktionen zur Erfassung, Verwaltung, Analyse und Präsentation der Unsicherheit benötigt. Bei der Erfassung gilt es, die Unsicherheitsparameter individuell für jede Objektgeometrie zu bestimmen. Möglichkeiten bieten sich durch den Vergleich mit einer externen Referenz, durch Auswertung von Mehrfacherfassungen oder durch Festlegung der Parameter über Fachwissen. Alle drei Verfahren weisen gewisse Vor- und Nachteile auf, so dass im praktischen Gebrauch eine Kombination aller anzutreffen sein wird. Zur Präsentation können sowohl vektor- als auch rasterbasierte Methoden zum Einsatz kommen. Isolinien- und Matrixdarstellungen geben einen Einblick in die vorliegende Unsicherheit. Probleme bereiten die durch die Unsicherheit hinzukommenden zusätzlichen Informationen, die die eigentliche Thematik teilweise überlagern und so die Lesbarkeit und Übersichtlichkeit der Präsentation stören können. Aus diesem Grund sollte auf eine dauerhafte Anzeige verzichtet und die Unsicherheit nur zum Zwecke der Qualitätsprüfung dargestellt werden. Zur Verwaltung der Unsicherheitsparameter können im Wesentlichen die bereits bestehenden Konzepte eingesetzt werden. Es bieten sich die beiden Möglichkeiten, die Verwaltung auf geometrischem oder auf attributivem Wege zu realisieren. Während die geometrische Lösung im Speziellen für das Minimum-Maximum-Modell sinnvoll eingesetzt werden kann, ist für die beiden anderen Modelle der Attributansatz zu favorisieren. In der Analysekomponente ist die Fortpflanzung der Unsicherheit auf das Ergebnis zu lösen. Aufgrund der besonderen Beeinflussung werden ausschließlich geometrische Operatoren betrachtet. Dazu zählen Längen-, Flächen-, und Distanzbestimmung, sowie die bekannte Verschneidung. Für alle drei Unsicherheitsmodelle können Fortpflanzungsansätze formuliert werden, die eine Übertragung der Unsicherheit auf das Ergebnis im Rahmen des glei-

chen Modells erlauben. Es treten jedoch deutliche Unterschiede hinsichtlich des notwendigen Aufwandes zwischen den Modellen hervor, die den Anwendungszweck beeinflussen.

Aus den Erfahrungen der Integration in die GIS-Komponenten können Merkmale für den praktischen Einsatz der einzelnen Unsicherheitsmodelle abgeleitet werden. Auf dieser Basis lassen sich einige allgemeine Empfehlungen für die Wahl eines geeigneten Modells festlegen. Für Anwendungen, bei denen die Betrachtung der Unsicherheit nur selten benötigt wird, ist das Minimum-Maximum-Modell einzusetzen. Werden dagegen hohe Anforderungen gestellt, so sind das Stochastische Modell oder das Fuzzy-Modell vorzuziehen, wobei sich das Fuzzy-Modell insbesondere durch seine Flexibilität auszeichnet. Für Anwendungen, deren Ausrichtung noch unklar ist, empfiehlt sich das Stochastische Modell, da es den besten Ausgangspunkt bildet, um falls notwendig auf andere Modelle überzugehen. Theoretisch ist eine hybride Verwendung aller drei Modelle denkbar, wodurch sich eine optimale Nutzung der Stärken ergibt. Doch ist ein Modellwechsel teilweise nur unter Einschränkungen und mit Informationsverlust möglich.

Einen positiven Einfluss auf die Nutzerakzeptanz hat die Einhaltung von Standards. Momentan befindet sich ein europäischer Standard in der Entwicklung, der die Qualitätsbeschreibung von räumlichen Daten auf allen Ebenen von den Metadaten bis zu einzelnen Datenelementen regelt. Die Festlegungen definieren einen Rahmen, in den jeder Nutzer seine eigenen Modelldefinitionen einzufügen hat. In dieser Weise lassen sich auch die vorgestellten Definitionen ohne jeglichen Aufwand einbetten, wodurch die Unsicherheitsmodelle als konform mit dem Standard zu bezeichnen sind.

Es bleibt zu hoffen, dass die entwickelten Konzepte und Methoden den Eingang in die Praxis finden werden. Alle dazu notwendigen Voraussetzungen sind durch die vorliegende Arbeit soweit geschaffen.

5.2 Ausblick

Für eine durchgängige Berücksichtigung der Unsicherheit ist es notwendig, die Fortpflanzung der Unsicherheit nicht nur auf einige ausgewählte Operatoren zu beschränken. Grundsätzlich ist zu fordern, dass die Unsicherheit nach jedem Verarbeitungsschritt eingesehen und überprüft werden kann. Dazu sind Erweiterungen für alle verfügbaren Operatoren zu entwickeln. Den behandelten geometrischen Funktionen stehen topologische Operatoren am nächsten, die die Nachbarschaftsbeziehungen von Objekten auf der Basis der geometrischen Situation ermitteln. Anstelle einer eindeutigen Beziehung ergibt sich aufgrund der Unsicherheit in der Geometrie eine Reihe möglicher Zustände, für die jeweils durch Fortpflanzung eine Beschreibung der Unsicherheit abzuleiten ist. Sowohl für das Stochastische Modell (Winter, 1996) als auch für das Minimum-Maximum-Modell (Clementini und Di Felice, 1996, Krauß, 1998, Glemser et al., 1999) liegen bereits Ansätze dazu vor, während für das Fuzzy-Modell eine entsprechende Lösung noch aussteht.

Ein räumliches Objekt besteht nicht nur aus einer geometrischen Komponente, sondern wird maßgeblich auch von einem thematischen Anteil geprägt. Für eine gemeinsame Verarbeitung beider Teile in integrierten Fragestellungen ist es daher notwendig, die Unsicherheitsbeschreibung auch auf thematische Daten auszudehnen. Die Ansätze zum Feldmodell zeigen mögliche Richtungen an. Für die Verarbeitung wäre es von Vorteil, wenn beide Komponenten auf gleichen Unsicherheitsmodellen basieren würden. Daher ist für die vorgestellten Modelle im Einzelnen zu prüfen, inwieweit sie sich auch auf thematische Daten anwenden lassen und welche Erweiterungen oder Änderungen für ihren Einsatz notwendig sind.

Bisher unbeachtet blieben die Auswirkungen von unsicheren Analysemodellen auf die Datenverarbeitung. Sie bilden eine zusätzlich zu den Daten zu berücksichtigende Quelle an Unsicherheit. Insbesondere bei komplexen Analysen können die Verarbeitungsvorschriften häufig nicht als fest angesehen werden, sondern lassen eine gewisse Variation zu. Ursache dafür können beispielsweise unvollständiges Wissen oder aber nur unscharf zu formulierende Anweisungen sein. Sind z.B. alle Objekte in der Nähe zu einem anderen Objekt gefragt, so eröffnet der unscharfe Begriff „Nähe“ einen Spielraum, in dem sich die Lösung bewegen kann. Andere Beispiele sind Simulationen, deren Modelle die Realität meist nur genähert abbilden. Diese Quelle der Unsicherheit ist in gleichem Maße zu beschreiben und in der Verarbeitung der Daten zu berücksichtigen. Nur so lässt sich gewährleisten, dass das erzielte Ergebnis qualifiziert geprüft und darauf basierende Fehlentscheidungen vermieden werden können.

6 Literatur

- Altman, D. (1994): Fuzzy Set Theoretic Approaches for Handling Imprecision in Spatial Analysis. *Int. J. Geographic Information Systems*, Vol. 8, No. 3, 271-289.
- Aronoff, S. (1991): *Geographic Information Systems: A Management Perspective*. WDL Publication, Ottawa.
- Atkinson, M., Bancilhon, F., De Witt, D., Dittrich, K., Maier, D., Zdonik, S., (1989): The Object-Oriented Database System Manifesto. In: *Proceedings 1st DOOD*, 40-57.
- ATKIS (2000): Amtliches Topographisch-Kartographisches Informationssystem, ATKIS-Objektartenkatalog, <http://www.atkis.de>.
- Bezdek, J. C., Ehrlich, R., Full, W. (1984): FCM: The Fuzzy-C-Means Clustering Algorithm. *Computer and Geoscience*, Vol. 10, 191-203.
- Bhatnagar, R. K., Kanal, L. N. (1986): Handling Uncertain Information: A Review of Numeric and Non-Numeric Methods. In: Kanal und Lemmer (1986), *Uncertainty in Artificial Intelligence*. Elsevier Science Publisher, 3-26, Amsterdam.
- Bill, R. (1999a): *Grundlagen der Geo-Informationssysteme, Band 1. Hardware, Software und Daten*. Wichmann Verlag, Heidelberg.
- Bill, R. (1999b): *Grundlagen der Geo-Informationssysteme, Band 2. Analyse, Anwendungen und neue Entwicklungen*. Wichmann Verlag, Heidelberg.
- Bill, R., Korduan, P. (1998): Flächenverschneidung in GIS - Stochastische Modellierung und Effizienzbetrachtung. *Zeitschrift für Vermessungswesen*, Vol. 123, Teil 1 in No. 8, 247-253, Teil 2 in No. 10, 333-338.
- Blakemore, M. (1984): Generalisation and Error in Spatial Data Bases. *Cartographica*, Vol. 21, 131-139.
- Blaschke, T. (1997): Map Algebra und Fuzzy Logic in Behörden? Potential und Akzeptanz von GIS-Analysen bei Einbeziehung von räumlicher Unschärfe. *GIS*, Vol. 10, No. 6, 3-12.
- Bolstad, P. V., Gessler, P., Lillesand, T. M. (1990): Positional Uncertainty in Manually Digitized Map Data. *Int. J. Geographic Information Systems*, Vol. 4, No. 4, 399-412.
- Bonham-Carter, G. F. (1994): *Geographic Information Systems for Geoscientists: Modelling with GIS*. Pergamon, Elsevier Science Publications, Kidlington.
- Booch, G. (1991): *Object Oriented Analysis and Design with Applications*. The Benjamin Cummings Series in Object-Oriented Software Engineering, Redwood City.
- Brigham, E. O. (1987): *FFT: Schnelle Fourier-Transformation*. Oldenbourg Verlag, München.
- Brockhaus (1998): *Brockhaus – die Enzyklopädie*. Band 18, 20. Auflage, Mannheim.
- Buchanan, B. G., Shortliffe, E. H. (Eds.) (1984): *Rule Based Expert Systems. The MYCIN Experiments of Stanford Heuristic Programming Project*. Addison-Wesley, Reading.
- Burrough, P. A. (1986): *Principles of Geographical Information Systems for Land Resource Assessment*. Oxford Science Publications, Clarendon Press, Oxford.
- Burrough, P. A. (1989): Fuzzy Mathematical Methods for Soil Survey and Land Evaluation. *Journal of Soil Science*, No. 40, 477-492.
- Burrough, P. A. (1996): Natural Objects with Indeterminate Boundaries. In: Burrough und Frank (1996), *Geographic Objects with Indeterminate Boundaries*. ESF-GISDATA, Vol. 2, 3-28, Taylor & Francis, London.

- Burrough, P. A., Heuvelink, G. B. M. (1992): The Sensitivity of Boolean and Continuous (fuzzy) Logical Modelling to Uncertainty Data. In: Proceedings EGIS '92, München, 1032-1039, EGIS Foundation, Utrecht.
- Burrough, P. A., Frank, A. U. (1995): Concepts and Paradigms in Spatial Information: Are Current Geographical Information Systems Truly Generic? *Int. J. Geographical Information Systems*, Vol. 9, No. 2, 101-116.
- Burrough, P. A., Frank, A. U. (Eds.) (1996): *Geographic Objects with Indeterminate Boundaries*. ESF-GISDATA, Vol. 2, Taylor & Francis, London.
- Buttenfield, B. P., Mackaness W. A. (1991): Visualisation. In: Maguire et al. (1991), *Geographical Information Systems, Principles and Applications*, 427-443, Longman, New York.
- Caspary, W. (1992a): Qualitätsmerkmale von Geo-Daten. *Zeitschrift für Vermessungswesen*, Vol. 117, No. 7, 360-367.
- Caspary, W. (1992b): Genauigkeit als Qualitätsmerkmal digitaler Datenbestände. In: Grünreich, D., Buziek, G. (Hrsg.) (1992), *Gewinnung von Basisdaten für Geo-Informationssysteme*. DVW-Schriftenreihe, Heft 4, 157-166.
- Caspary, W., Scheuring, R. (1992): Error-Bands as Measures of Geometrical Accuracy. In: Proceedings EGIS '92, München, Vol. 1, 226-233, EGIS Foundation, Utrecht.
- Cheeseman, P. (1986): Probabilistic versus Fuzzy Reasoning. In: Kanal und Lemmer (1986), *Uncertainty in Artificial Intelligence*. Elsevier Science Publisher, 85-102, Amsterdam.
- Chrisman, N. R. (1982): A Theory of Cartographic Error and its Measurement in Digital Databases. In: Proceedings Auto-Carto 5, 159-168, Crystal City.
- Chrisman, N. R. (1991): The Error Component in Spatial Data. In: Maguire et al. (1991), *Geographical Information Systems, Principles and Applications*, 165-174, Longman, New York.
- Clementini, E., Di Felice, P. (1996): An Algebraic Model for Spatial Objects with Indeterminate Boundaries. In: Burrough und Frank (1996), *Geographic Objects with Indeterminate Boundaries*. ESF-GISDATA, Vol. 2, 155-169, Taylor & Francis, London.
- Congalton, R. G., Mead, R. A. (1983): A Qualitative Method to Test for Consistency and Correctness in Photointerpretation. *Photogrammetric Engineering & Remote Sensing*, Vol. 49, No. 1, 69-74.
- Couclelis, H. (1992): People Manipulate Objects (but Cultivate Fields): Beyond the Raster-Vector Debate in GIS. In: Frank et al. (1992), *Theories and Methods of Spatio-Temporal Reasoning in Geographic Space*. Lecture Notes in Computer Science, No. 639, 65-77, Springer, Berlin.
- Couclelis, H. (1996): Towards an Operational Typology of Geographic Entities with ill-defined Boundaries. In: Burrough und Frank (1996), *Geographic Objects with Indeterminate Boundaries*. ESF-GISDATA, Vol. 2, 45-55, Taylor & Francis, London.
- Cramer, H. (1946): *Mathematical Methods of Statistics*. Princeton University Press, Princeton.
- Cramer, M. (1993): *Implementation von Raster-Vektor-Konvertierungsbausteinen als Basis für eine GIS-Teachware*. Diplomarbeit am Institut für Photogrammetrie, Universität Stuttgart (unveröffentlicht).
- David, B., Van Den Herrewegen, M., Salge, F. (1996): Conceptual Models for Geometry and Quality of Geographical Information. In: Burrough und Frank (1996), *Geographic Objects with Indeterminate Boundaries*. ESF-GISDATA, Vol. 2, 193-206, Taylor & Francis, London.
- DCW (2000): *Digital Chart of the World Server*, Penn State University Libraries, <http://www.maproom.psu.edu/dcw>.
- DIN/CEN/ISO (1998): *Geoinformation – Datenbeschreibung – Qualität*. Vornorm E DIN EN 12656, CEN/TC 287 „Geoinformation“, ISO/TC 211 „Geoinformation/Geomatik“, Beuth Verlag, Berlin.
- Douglas, D. H., Peucker, T. K. (1973): Algorithmus for the Reduction of the Number of Points Required to Represent a Digitized Line or its Caricature. *The Canadian Cartographer*, Vol. 10., 112-122.

- Drummond, J. E. (1995): Elements of Spatial Data Quality. In: Guptill und Morrison (1995), Elements of Spatial Data Quality, 31- 58, Pergamon, Oxford.
- Drummond, J. E., Ramlal, B. (1992): A Prototype Uncertainty Sub-System Implemented in ITCs ILWIS PC-Based GIS, and Tested in a Dutch Land Reallotment Project. In: Proceedings EGIS '92, München, 234-243, EGIS Foundation, Utrecht.
- Dutton, G. (Hrsg.) (1978): Harvard Papers on Geographic Information System, Vol. 6, Laboratory for Computer Graphics and Spatial Analysis, Addison-Wesley, Cambridge.
- Edwards, G. (1994): Characterizing and Maintaining Polygons with Fuzzy Boundaries in Geographic Information Systems. In: Waugh, T. C., Healey, R. G. (Eds.) (1994), Proceedings of the 6th Spatial Data Handling, Vol. 1, 223-239, Taylor & Francis, London.
- Edwards, G., Lowell, K. E. (1996): Modelling Uncertainty in Photointerpreted Boundaries. Photogrammetric Engineering & Remote Sensing, Vol. 62, No. 4, 337-391.
- Egenhofer, M. J., Herring, J. R. (1990): A Mathematical Framework for the Definition of Topological Relationships. In: Proceedings of 4th International Symposium on Spatial Data Handling, International Geographical Union, 803-818, Zürich.
- Ehlers, M., Edwards, G., Bedard, Y. (1989): Integration of Remote Sensing with Geographic Information Systems: A Necessary Evolution. Photogrammetric Engineering & Remote Sensing, Vol. 55, No. 11, 1619-1627.
- Fisher, P. F., Pathirana, S. (1990): The Evaluation of Fuzzy Membership of Land Cover Classes in the Suburban Zone. Remote Sensing and Environment, Vol. 34, 121-132.
- Frank, A. U. (1996): The Prevalence of Objects with Sharp Boundaries in GIS. In: Burrough und Frank (1996), Geographic Objects with Indeterminate Boundaries. ESF-GISDATA, Vol. 2, 29-40, Taylor & Francis, London.
- Frank, A. U., Campari, I., Formentini, U. (Eds.) (1992): Theories and Methods of Spatio-Temporal Reasoning in Geographic Space. Lecture Notes in Computer Science, No. 639, Springer, Berlin.
- Fritsch, D. (1984): Proposal for the Determination of Least Sampling Interval for DEM Data Acquisition. Tech. Paper, Workshop Digital Elevation Models, Edmonton.
- Fritsch, D. (1988a): Hybride graphische Systeme – Eine neue Generation von raumbezogenen Informationssystemen. GIS, Vol. 1, No. 1, 12-20.
- Fritsch, D. (1988b): Some Experience with the Determination of the Optimum Sampling Density. In: Int. Archives of Photogrammetry and Remote Sensing, Vol. 27, Part B2, 493-504, Kyoto.
- Fritsch, D. (1991): Raumbezogene Informationssysteme und digitale Geländemodelle. Deutsche Geodätische Kommission, Reihe C, Heft Nr. 369, München.
- Fritsch, D., Anders, K.-H. (1996): Objektorientierte Konzepte in Geo-Informationssystemen. GIS, Vol. 9, No. 2, 2-14.
- Fritsch, D., Glemser, M., Klein, U., Strunz, G. (1998): Zur Integration von Unsicherheit bei Vektor- und Rasterdaten. GIS, Vol. 11, No. 4, 26-35.
- Glemser, M. (1996): Integration geometrischer Datenqualität in GIS-Funktionen. In: Proceedings Workshop Datenqualität und Metainformation in Geo-Informationssystemen, Universität Rostock, 69-84, Rostock.
- Glemser, M., Henneberg, C. (1999): Untersuchungen zu den Fortpflanzungseigenschaften der geometrischen Genauigkeit bei der Flächenverschnidung. In: Sester, M., Krumm, F. (Hrsg.) (1999): GIS-Forschung im Studiengang Geodäsie und Geoinformatik, Schriftenreihe der Institute des Fachbereichs Vermessungswesen, Universität Stuttgart, Report Nr. 1999.1, 44-56, Stuttgart.
- Glemser, M., Krauß, S., Schneider, C. (1999): Das Minimum-Maximum-Modell – Eine qualitative Beschreibung der geometrischen Genauigkeit und ihre Integration in GIS. In: Sester, M., Krumm, F. (Hrsg.) (1999): GIS-

Forschung im Studiengang Geodäsie und Geoinformatik, Schriftenreihe der Institute des Fachbereichs Vermessungswesen, Universität Stuttgart, Report Nr. 1999.1, 57-69, Stuttgart.

- Goodchild, M. F. (1989): Error in Objects and Fields. In: Goodchild und Gopal (1989), Accuracy of Spatial Databases, 107-113, Taylor & Francis, London.
- Goodchild, M.F. (1992): Geographical Information Science. *Int. J. Geographical Information Systems*, Vol. 6, No. 1, 31-45.
- Goodchild, M. F. (1995): Attribute Accuracy. In: Guptill und Morrison (1995), Elements of Spatial Data Quality, 59-79, Pergamon, Oxford.
- Goodchild, M. F., Dubuc, O. (1987): A Model of Error for Choropleth Maps, with Applications to Geographic Information Systems. In: Proceedings of Auto-Carto 8, 165-174.
- Goodchild, M. F., Wang, F. (1989): Modelling Errors for Remotely Sensed Data into GIS. In: Proceedings Auto-Carto 9, 530-537.
- Goodchild, M. F., Gopal, S. (Eds.) (1989). The Accuracy of Spatial Databases. Taylor & Francis, London.
- Goodchild, M. F., Haining, R. P., Wise S. M. et al., (1992a): Integrating GIS and Spatial Analysis: Problems and Possibilities. *Int. J. Geographical Information Systems*, Vol.6, No. 5, 407-423.
- Goodchild, M. F., Sun, G., Yang, S. (1992b): Development and Test of an Error Model for Categorical Data. *Int. J. Geographical Information Systems*, Vol. 6, No. 2, 87-104.
- Graf, U., Barner, M. (1978): Darstellende Geometrie. Uni-Taschenbücher 817, Quelle & Meyer, Heidelberg.
- Grafarend, E. W., Schaffrin, B. (1993): Ausgleichsrechnung in linearen Modellen. BI-Wissenschaftsverlag.
- Grosche, G., Ziegler, V., Ziegler, D., Zeidler, E. (1995): Teubner Taschenbuch der Mathematik, Teil II. B. G. Teubner Verlag, Stuttgart und Leipzig.
- Guptil, S. C., Morrison, J. L. (Eds.) (1995): Elements of Spatial Data Quality. Pergamon, Oxford.
- Hadzilacos, T. (1996): On Layer-Based Systems for Undetermined Boundaries. In: Burrough und Frank (1996), Geographic Objects with Indeterminate Boundaries. ESF-GISDATA, Vol. 2, 237-255, Taylor & Francis, London.
- Haralick, R. M. (1980): A Spatial Data Structure for Geographic Information Systems. In: Freeman, H., Pieroni, G.G. (1980), Map Data Processing. Academic Press, 63-99, New York.
- Helmert, F. R. (1907): Die Ausgleichsrechnung nach der Methode der kleinsten Quadrate. Teubner, Leipzig und Berlin.
- Heuvelink, G. B. M. (1992): An Iterativ Method for Multidimensional Simulation with Nearest Neighbor Models. Proceedings of the 2nd CODATA Conference on Geomathematics and Geostatistics, Nancy, 51-57.
- Heuvelink, G. B. M., Burrough, P. A. (1993): Error Propagation in Cartographic Modelling Using Boolean Logic and Continuous Classification. *Int. J. Geographic Information Systems*, Vol. 7, No. 3, 231-246.
- Heuvelink G. B. M., Burrough, P. A., Stein A. (1989): Propagation of Errors in Spatial Modelling with GIS. *Int. J. Geographic Information Systems*, Vol. 3, No. 4, 303-322.
- Höpcke, W. (1980): Fehlerlehre und Ausgleichsrechnung. De Gruyter, Berlin und New York.
- Hunter, G. J., Goodchild, M. F. (1997): Modeling the Uncertainty of Slope and Aspect Estimates Derived from Spatial Databases. *Geographical Analysis*, Vol. 29, No. 1, 35-49.
- Jacobsen, I. (1992): Object-Oriented Software Engineering: A Use Case Driven Approach. Academic Press, Addison Wesley, Reading.

- Kanal, L. N., Lemmer, J. F. (Eds.) (1986): *Uncertainty in Artificial Intelligence*. Elsevier Science Publisher, Amsterdam.
- Kaufmann, A. (1975): *Introduction to the Theory of Fuzzy Subsets. Volume 1: Fundamental Theoretical Elements*. Academic Press, London.
- Kemp, K. (1992): *Environmental Modeling with GIS: A Strategie for Dealing with Spatial Continuity*. In: *Proceedings of the GIS/LIS '92 Annual Conference and Exposition, Vol. 1, 397-406, San Jose*.
- Keul, E. (1998): *Anwendung der parametrischen Distanzfunktion zur Bestimmung der Mittelachse eines Polygons*. Studienarbeit am Institut für Photogrammetrie, Universität Stuttgart (unveröffentlicht).
- Klein, U., Sester, M., Strunz, G. (1998): *Segmentation of Remotely Sensed Images Based on the Uncertainty of Multispectral Classifications*. In: *Int. Archives of Photogrammetry and Remote Sensing, Vol. 32, Part 4, ISPRS Commission IV Symposium, GIS- Between Vision and Applications, 299-305, Stuttgart*.
- Koch, K. R. (1980): *Parameterschätzung und Hypothesentests in linearen Modellen*. Ferd. Dümmlers Verlag, Bonn.
- Kraus, K., Haussteiner, K. (1993): *Visualisierung der Genauigkeit geometrischer Daten*. GIS, Vol. 6, Heft 3, 7-12.
- Kraus, K., Karger, H. (1994): *Accuracy of Derived Data in a Geographic Information System*. *Comput., Environ. and Urban Systems, Vol. 18, No. 2, 87-94*.
- Kraus, K., Ludwig, M. (1998): *Genauigkeit der Verschneidung geometrischer Geodaten*. *Zeitschrift für Vermessungswesen, Vol. 123, No.3, 81-87*.
- Krauß, S. (1998): *Qualitative Beschreibung von unsicheren topologischen Relationen innerhalb des Minimum-Maximum-Modells*. Diplomarbeit am Institut für Photogrammetrie, Universität Stuttgart (unveröffentlicht).
- Kreifelts, T., Woetzel, G. (1977): *Extraktion von Linienmustern aus Binärbildern*, *Mitteilungen der Gesellschaft für Mathematik und Datenverarbeitung, St. Augustin*.
- Kreyszig, E. (1965): *Statistische Methoden und ihre Anwendungen*. Vandenhoeck & Ruprecht, 422p., Göttingen.
- Lee, D. T. (1982): *Medial Axis Transformation of a Planar Shape*. *IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 4, No. 4, 363-369*.
- Lichtner, W. (1981): *Anwendungsmöglichkeiten der Rasterdatenverarbeitung in der Kartographie*. *Wissenschaftliche Arbeiten der Fachrichtung Vermessungswesen, Universität Hannover, Nr. 105, Hannover*.
- Lillesand, T. M., Kiefer, R. W. (1979): *Remote Sensing and Image Interpretation*. John Wiley and Sons, New York.
- Lowell, K. (1994): *An Uncertainty-Based Spatial Representation for Natural Resources Phenomena*. In: *Proceedings of 6. Spatial Data Handling, Vol. 2, 933-944, Taylor & Francis, London*.
- Lowell, K. (1995): *Manipulating Uncertainty-Based Digital Maps to Perform like Polygon-Based Thematic Maps*. In: *Proceedings of ICA-Conference, Vol. 1, 168-175, Barcelona*.
- Maffini G., Arno, M., Bitterlich, W. (1989): *Observations and Comments on the Generation and Treatment of Error in Digital GIS Data*. In: *Goodchild und Gopal (1989), Accuracy of Spatial Databases, 55-67, Taylor & Francis, London*.
- Maguire, D. J., Goodchild, M. F., Rhind, D. W. (1991): *Geographical Information Systems, Principles and Applications*. Longman, Harlow.
- Mandl, P. (1994): *Räumliche Entscheidungsunterstützung mit GIS: Nutzwertanalyse und Fuzzy-Entscheidungsmodellierung*. In: *Dollinger, F., Strobl, J. (Hrsg.), Angewandte Geographische Wissensverarbeitung VI, Salzburger Geographische Materialien, No. 21, 463-473, Salzburg*.
- Molenaar, M. (1989): *Single Valued Vector Maps - A Concept in Geographic Information Systems*. GIS, Vol. 2, Heft 1, 18-26.

- Molenaar, M. (1993): Object Hierarchies and Uncertainty in GIS or Why is Standardisation so Difficult? GIS, Vol. 6, No. 4, 22-28.
- Molenaar, M. (1996a): The Extensional Uncertainty of Spatial Objects. In: Proceedings Spatial Data Handling, 9B.1-9B.13, Delft.
- Molenaar, M. (1996b): A Syntactic Approach for Handling the Semantics of Fuzzy Spatial Objects. In: Burrough und Frank (1996), Geographic Objects with Indeterminate Boundaries. ESF-GISDATA, Vol. 2, 207-224, Taylor & Francis, London.
- Molenaar, M., Fritsch, D. (1991): Combined Data Structures for Vector and Raster Representations in Geographic Information Systems. GIS, Vol. 4, No. 3, 26-32.
- NCGIA (1989): The Research Plan of the National Center for Geographic Information and Analysis. Int. J. Geographic Information Systems, Vol. 3, No. 2, 117-136.
- Objectivity (2000): Objectivity/DB, <http://www.objectivity.com>.
- OGC (2000): The OpenGIS Guide. Open GIS Consortium (OGC), <http://www.opengis.org/techno/guide.htm>.
- OMG (2000): UML Resource Page, <http://www.omg.org/uml>.
- Perkal, J. (1966): On the Length of Empirical Curves. Discussion Paper No. 10, Michigan Inter-University Community of Mathematical Geographers, Ann Arbor.
- Peuquet, D. J., (1984): A Conceptual Framework and Comparison of Spatial Data Models. Cartographica, Vol. 21, No. 14, 66-113.
- Piehler, J. Zschiesche, H.-U. (1976): Simulationsmethoden. Mathematik für Ingenieure, Band 20. Teubner Verlag, Leipzig.
- Pullar, D. (1991): Spatial Overlay with Inexact Numerical Data. Proceedings ACSM/ASPRS, Auto-Carto 10, Vol.6, 313-329, Baltimore.
- Quatrani, T. (1998): Visual Modeling with Rational Rose and UML. Addison Wesley, Longman, Reading.
- Rational (2000): UML Notation Guide, Version 1.3, <http://www.rational.com/uml/resources/documentation>.
- Richards, J. A. (1986): Remote Sensing Digital Image Analysis: An Introduction. Springer-Verlag, Berlin und Heidelberg.
- Ripley, B. D. (1981): Spatial Statistics. John Wiley & Sons, Inc., New York.
- Rosenfield, G. H., Fitzpatrick-Lins, K. (1986): A Coefficient of Agreement as a Measure of Thematic Classification Accuracy. Photogrammetric Engineering & Remote Sensing, Vol. 48, No. 5, 793-801.
- Rumbaugh, J., Blaha, M., Premerlani, W., Eddy, F., Lorenson, W. (1991): Object Oriented Modelling and Design, Prentice Hall, Englewood Cliffs.
- Sachs, L. (1978): Angewandte Statistik. Statistische Methoden und ihre Anwendungen. Springer-Verlag, Berlin, Heidelberg, New York.
- Serra, J. (1982): Image Analysis and Mathematical Morphology: Academic Press, 610 p., London, New York.
- Shafer, G. (1976): A Mathematical Theory of Evidence. Princeton University Press, Princeton.
- Shi, W. (1994): Modelling Positional and Thematic Uncertainties in Geographic Information Systems. ITC Publication, No. 22, Enschede.
- Sinton, D. (1978): The Inherent Structure of Information as a Constraint to Analysis: Mapped Thematic Data as a Case Study. In: Dutton (1978), Harvard Papers on Geographic Information Systems, Addison-Wesley, Cambridge.

- Sobol, I. M. (1985): Die Monte-Carlo-Methode. Deutsche Taschenbücher, Band 41, Verlag Harri Deutsch, Thun und Frankfurt am Main.
- Stearns, S. D. (1979): Digitale Verarbeitung analoger Signale. Oldenbourg Verlag, München.
- Story, M., Congalton, R. G. (1986): Accuracy Assessment: a User's Perspective. Photogrammetric Engineering & Remote Sensing, Vol. 52, No. 3, 397-399.
- Veregin, H. (1989): Error Modeling for the Map Overlay Operation. In: Goodchild und Gopal (1989), Accuracy of Spatial Databases, 3-18, Taylor & Francis, London.
- Veregin, H. (1995): Developing and Testing of an Error Propagation Model for GIS Overlay Operations. Int. J. Geographic Information Systems, Vol. 9, No. 6, 595-619.
- Vosselman, G. (1996): Uncertainty in GIS Supported Road Extraction. In: Int. Archives of Photogrammetry and Remote Sensing, Vol. 31, Part B3, 909-917, Wien.
- Walter, V. (1997): Zuordnung von raumbezogenen Daten – am Beispiel ATKIS und GDF. Deutsche Geodätische Kommission, Reihe C, Heft Nr. 480, München.
- Wang, F. (1990): Improving Remote Sensing Image Analysis through Fuzzy Information Representation. Photogrammetric Engineering & Remote Sensing, Vol. 56, No. 8, 1163-1169.
- Wilks, S. S. (1962): Mathematical Statistics. J. Wiley, New York.
- Winter, S. (1996): Unsichere topologische Beziehungen zwischen ungenauen Flächen. Deutsche Geodätische Kommission, Reihe C, Heft Nr. 465, München.
- Witke, T. (1987): Vergleich verschiedener Maße zur Bewertung von unsicheren Aussage. Diplomarbeit am Institut für Photogrammetrie, Universität Stuttgart (unveröffentlicht).
- Wolf, H. (1975): Ausgleichsrechnung. Formeln zur praktischen Anwendung. Ferd. Dümmlers Verlag, Bonn.
- Yang, H. (1992): Zur Integration von Vektor- und Rasterdaten in Geo-Informationssystemen. Deutsche Geodätische Kommission, Reihe C, Heft Nr. 389, München.
- Yang, W., Gold, C. (1996): Managing Spatial Objects with the VMO-Tree. In: Proceedings Spatial Data Handling, 11B.15-11B.30, Delft.
- Zadeh, L. A. (1965): Fuzzy Sets. Information and Control, Vol. 8, 338-353.
- Zadeh, L. A. (1986): Is Probability Theory Sufficient for Dealing with Uncertainty in AI: A Negative View. In: Kanal und Lemmer (1986), Uncertainty in Artificial Intelligence. Elsevier Science Publisher, 103-116, Amsterdam.
- Zhang, G. Y, Tulip, J. (1990): An Algorithm for the Avoidance of Sliver Polygons and Clusters of Points in Spatial Overlay. In: Proceedings of 4th Spatial Data Handling, 141-150, Zürich.

A Ergänzungen zum Stand der Forschung

A.1 Fundamentale Ansätze zur Datenmodellierung

Objektmodell und Feldmodell definieren zwei fundamentale Datenmodelle in der Geoinformatik. Sie werden im folgenden Abschnitt näher erläutert.

A.1.1 Objektmodell

Ein Objekt existiert stets unabhängig von anderen Objekten und besitzt eine unveränderliche Identität. Es umfasst eine bestimmte räumliche Ausdehnung (Geometrie) und wird durch Attribute beschrieben (Thematik) (Abb. A.1).

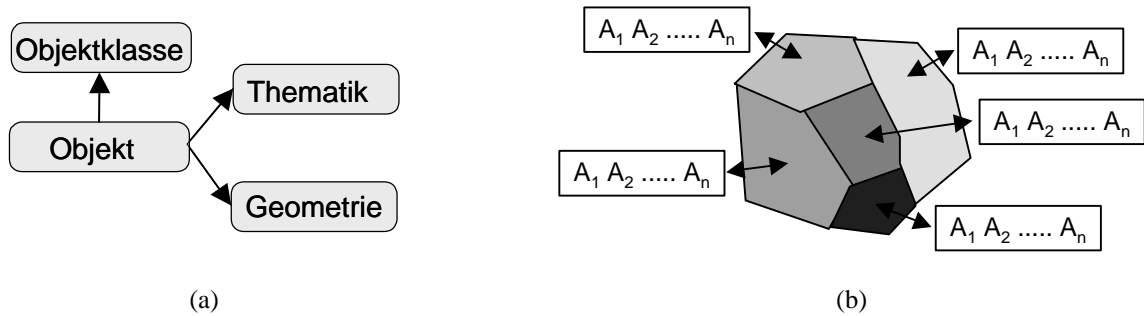


Abb. A.1: Objektmodell (a) mit Beispiel: Objekte mit geometrischer Repräsentation und Attributen (b).

Die räumliche Ausdehnung ergibt sich durch vorhandene Diskontinuitäten der Eigenschaften, so dass innerhalb des Objektes die beschreibenden Attribute als homogen gelten. Die Bildung einer Objektgrenze als geometrisch scharfe Abgrenzung gilt als ein Kennzeichen dieser Sichtweise (Hadzilacos, 1996). Objekte können einfach oder beliebig komplex aufgebaut sein. Gleichartige Objekte lassen sich klassenweise (in Objektklassen) zusammenfassen und darin prototypisch beschreiben (Abb. A.1a). Eine Repräsentation des Modells durch eine allgemeine formale Datenstruktur ist in Molenaar (1989) vorgestellt (Abb. A.2). Sie verwendet direkt die topologischen Beziehungen (vor allem Inzidenzbeziehungen, z.B. Linie hat Anfangs- und Endknoten) zwischen den geometrischen Primitiven (Punkt, Linie), um daraus Objekte verschiedenen Typs (punkt-, linien- und flächenförmige Objekte) aufzubauen. Aufgrund der aus Punkten und Linien bestehenden geometrischen Basis werden Daten, die in dieser Struktur vorliegen, als Vektordaten bezeichnet. Das Objektmodell lässt sich aber ebenso in Rasterform repräsentieren (Molenaar und Fritsch, 1991), so dass Objekte auch als Rasterdaten abgelegt sein können. Dadurch wird eine hybride Modellierung möglich, bei der die Repräsentation der Objektausdehnung alternativ in Raster- oder Vektorform erfolgen kann. Entsprechende hybride Strukturierungen (Haralick, 1980, Molenaar und Fritsch, 1991) setzen jeweils das Vorhandensein der Objektsichtweise voraus.

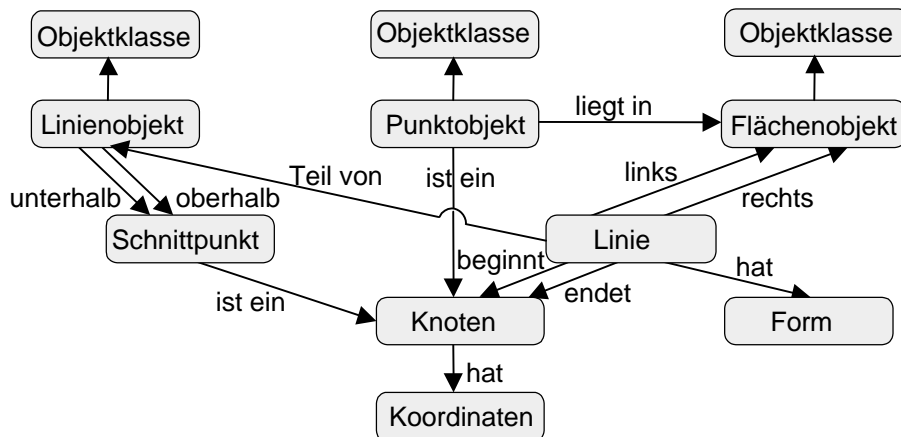


Abb. A.2: Formale Vektordatenstruktur (Molenaar, 1989).

A.1.2 Feldmodell

Felder repräsentieren genau eine kontinuierliche (zumeist physikalische) Eigenschaft des Raumes und damit genau eine räumliche Thematik. Die Kombination von mehreren Feldern, d.h. von verschiedenen Thematiken ist durch die identische Position festgelegt (Abb. A.3).

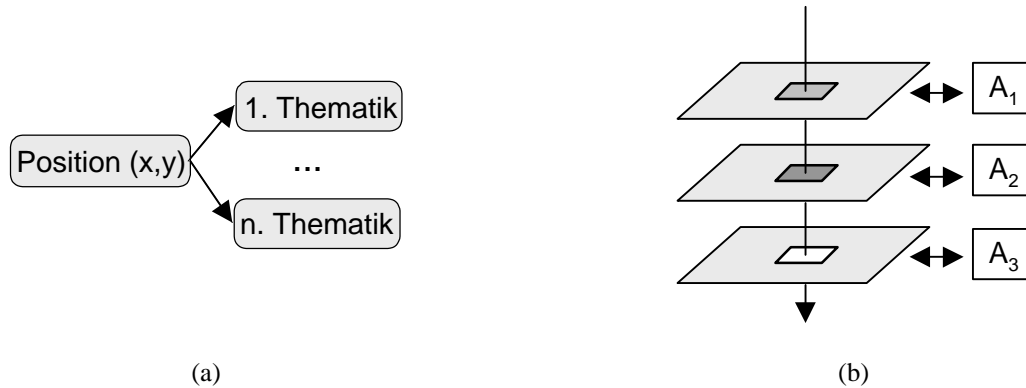


Abb. A.3: Feldmodell (a) mit Beispiel: diskretisierte Felder mit einzelnen Ebenen für jedes Attribut (b).

Allgemein gibt es hierzu zwei Gruppen von Modellen (Goodchild et al., 1992a, Kemp, 1992). In der ersten Gruppe wird die Eigenschaft präzise an einer bestimmten Position im Raum wiedergegeben (Punktraster, irreguläre Punktverteilungen, Konturlinien), während in der zweiten Gruppe (Polygone, Rasterzellen, irreguläre Dreiecke) explizite Zuweisungen des Wertes über geeignete Funktionen (z.B. Interpolationsfunktionen) notwendig werden. Ein Feld lässt sich auch als eine kontinuierlich verteilte Variable ansehen und entsprechend behandeln (Kemp, 1992). Es existieren zahlreiche Datenstrukturen, die die verschiedenen Feldmodelle repräsentieren. Am bekanntesten sind Rasterzellenstrukturen (Abb. A.3b), die die einfache Matrixform oder auch komplexere Ansätze (z.B. Lauflängenkodierung) umfassen, so dass Felder digital häufig als Rasterdaten anzutreffen sind. Felder können aber auch in Vektorform (z.B. durch Konturlinien) repräsentiert werden.

Als Kriterien, welches Konzept der realen Welt für welche Anwendungen am besten geeignet erscheint, sind die Natur der Phänomene (diskret oder kontinuierlich, homogen oder inhomogen, etc. (Couclelis, 1996)), der Anwendungsmaßstab (groß- oder kleinmaßstäbig (Couclelis, 1992)) und der Zweck (verschiedene Nutzergruppen mit verschiedenen Anforderungen (Burrough und Frank, 1995)) zu beurteilen. Eine Klassifikation der möglichen Anwendungen nach den Kriterien Exaktheit oder Komplexität der Phänomene, eingesetzte binäre oder mehrwertige Logik und deterministische oder stochastische Beschreibung der Realität ist in Burrough und Frank (1995) gegeben.

A.2 Ansätze zur Modellierung der Unsicherheit von Feldern

Während die bestehenden Ansätze zur Modellierung der Unsicherheit von Objekten ausführlich in Kapitel 1.2 erläutert werden, folgen hier einige Ausführungen zu den Ansätzen für Felder.

Durch die Feldsichtweise werden Phänomene beschrieben, die räumlich kontinuierliche Eigenschaften besitzen. Sie zeichnen sich häufig durch starke Variationen aus, so dass die Messung der Thematik im Vordergrund steht. Die Umsetzung erfolgt zumeist im Rasterdatenmodell, das die Geometrie durch eine strikte Definition vorgibt und nur die Thematik variabel hält. Daher spielt hier die Attributunsicherheit die entscheidende Rolle, während die geometrische Unsicherheit von untergeordneter Bedeutung ist. Die nachfolgend untersuchten Verfahren lassen sich grob nach den darin behandelten Attributtypen in quantitative (Varianzen) und qualitative Verfahren (Konfusionsmatrix, Wahrscheinlichkeitsvektor und Fuzzy-Ansatz) unterscheiden. Qualitative Attribute besitzen diskrete Werte, die der nominalen oder ordinalen Skala zuzuordnen sind. Quantitative Attribute sind dagegen kontinuierlich und daher der Intervall- oder Ratio-Skala zugehörig (z.B. Bonham-Carter, 1994).

Varianzen (stochastischer Ansatz)

Die Unsicherheit in quantitativen Attributen lässt sich auf der Basis der Stochastik behandeln (Goodchild, 1995). Analog zu den Koordinaten (siehe stochastischer Ansatz in der Geometrie), die ebenfalls quantitativen Charakter besitzen, können Varianzen zur Beschreibung eingesetzt werden. Ein Attribut (z.B. Temperatur) wird dabei als Zufallsvariable

betrachtet und durch Mittelwert und Varianz charakterisiert. Als zugehörige Verteilung kommt meist die Gaußsche Normalverteilung in Betracht. Da zu jedem Attributwert individuell eine eigene Varianz gehören kann, ist zur Verwaltung eine zusätzliche Datenebene pro Attribut notwendig. Individuelle Varianzen ergeben sich z.B., wenn aus Punktmessungen (z.B. in Form von Bodenprofilen) flächenhafte Daten interpoliert werden. Das bekannte Kriging-Verfahren (z.B. Ripley, 1981) ist in der Lage, Varianzen für jede Rasterzelle zu schätzen (Heuvelink et al., 1989, Heuvelink und Burrough, 1993). Das Verhältnis von Standardabweichung zu Attributwert ergibt ein relatives Fehlermaß, das zum Zwecke einer ansprechenden Darstellung der Unsicherheit genutzt werden kann (Heuvelink et al., 1989).

Die Fortpflanzung der Varianzen über verschiedene Verarbeitungsschritte erfordert die Kenntnis des funktionalen Zusammenhangs zwischen Eingabedaten und Ergebnis. Ist die entsprechende algebraische Funktion bekannt, bestimmt das allgemeine Fehlerfortpflanzungsgesetz die Berechnung der Ergebnisvarianz (Veregin, 1989, Goodchild, 1995). Heuvelink et al. (1989) erläutern die Fortpflanzung an Beispielen aus dem Bereich der Bodenkunde. Neben den Einflüssen aus den Eingabedaten finden zusätzlich Korrelationen zwischen den Daten in globaler Form und die Unsicherheit in den Modellparametern Berücksichtigung. Die Unsicherheit eines Analyseergebnisses wird maßgeblich durch die Wahl des Verarbeitungsmodells beeinflusst. In einer Untersuchung (Burrough und Heuvelink, 1992, Heuvelink und Burrough, 1993) werden die Auswirkungen von Logik-Modell und Fuzzy-Subset-Modell (Zadeh, 1965) einander gegenübergestellt. Simulationen übernehmen die Fortpflanzung der Unsicherheit, da in beiden Modellen logische Ausdrücke (z.B. UND oder ODER) verwendet werden. Aus den statistischen Auswertungen der Simulationsergebnisse ergibt sich, dass das Fuzzy-Subset-Modell verglichen mit dem logischen Modell stets eine geringere Varianz besitzt.

Konfusionsmatrix

Der Aufbau einer Konfusionsmatrix (Fehlermatrix) ist ein Verfahren, das in der Fernerkundung Anwendung findet (z.B. Lillesand und Kiefer, 1979). Es dient dort zur Kontrolle und Dokumentation der Güte von Klassifizierungen. Verallgemeinert eignet sich die Konfusionsmatrix zur Beschreibung der globalen Unsicherheit von qualitativen Attributen, d.h. es wird das Vorliegen von diskreten Werten vorausgesetzt. Zur Erstellung einer solchen Matrix wird ein Referenzdatensatz benötigt, der die Wirklichkeit übergeordnet genau repräsentiert. Aus dem Vergleich der Daten mit der Referenz werden absolute Häufigkeiten von Übereinstimmungen und Fehlzuzuweisungen klassenweise bestimmt und in der Matrix festgehalten (Tab. A.1). Auf der Hauptdiagonalen finden sich die Übereinstimmungen, während Nebendiagonalelemente die jeweilige Anzahl von Fehlzuzuweisungen enthalten, die wahren Fehlern (keine Residuen) entsprechen. Daraus können verschiedene andere Maße abgeleitet werden (z.B. Prozentsatz korrekt klassifizierter Zellen (PCC), Kappa-Koeffizient, Nutzer- und Produzentengenauigkeit; (Rosenfield und Fitzpatrick-Lins, 1986, Story und Congalton, 1986, Congalton und Mead, 1983)). Der PCC-Wert gilt als wichtiges Maß, das die relative Häufigkeit der Übereinstimmungen pro Klasse darstellt. Er lässt sich direkt als Wahrscheinlichkeit für die Richtigkeit des Attributwertes interpretieren. Jede Klasse besitzt genau einen Wert, d.h. eine räumliche Differenzierung ist nicht möglich.

Tab. A.1: Konfusionsmatrix nach Lillesand und Kiefer (1979).

Referenz- klasse	Anzahl der Pixel	Prozent- satz kor- rekt	Anzahl an Pixel in klassifizierter Klasse					
			1	2	3	4	5	6
1	5325	97	5165	0	42	44	53	21
2	328	66	0	216	0	108	4	0
3	4284	84	0	0	3599	16	482	187
4	945	42	12	92	228	397	132	84
5	2380	80	0	9	28	78	1904	361
6	3048	72	8	0	48	18	779	2195

Die klassenweisen Wahrscheinlichkeiten können für die Fortpflanzung der Unsicherheit innerhalb von Analysen eingesetzt werden. In Veregin (1989) werden für die logischen Operatoren UND und ODER die jeweiligen Formeln entwickelt und Abschätzungen für maximale und minimale Werte im allgemeinen Fall von n ($n > 1$) zu verarbeitenden Datensätzen gegeben. Korrelationen zwischen den Datensätzen äußern sich als bedingte Wahrscheinlichkeiten, die nicht direkt aus der Konfusionsmatrix abzuleiten sind, sich jedoch über einen erweiterten Vergleich der Datensätze mit der Referenz ermitteln lassen. Voraussetzung ist, dass die Referenz in allen Datensätzen an den gleichen Stellen vorliegt. Eine Erkenntnis der Untersuchungen besteht darin, dass nicht notwendigerweise der schlechteste Datensatz die Unsi-

cherheit bestimmt, sondern das Ergebnis stark vom angewendeten Operator abhängt. Im Falle von binominalen Daten (binäre Daten: nur Werte Ja/Nein, Wahr/Falsch oder 1/0 möglich) besteht die Möglichkeit, eine Fortpflanzung der kompletten Konfusionsmatrix zu formulieren (Veregin, 1995). Die einzelnen Konfusionsmatrizen werden dazu vereinigt, so dass eine zusammengesetzte Konfusionsmatrix als Basis für die Fortpflanzung entsteht. Für unterschiedliche Operatoren (z.B. logisches UND) sind für alle möglichen Kombinationen der beiden Datenebenen Verarbeitungsanweisungen in Form einer Wahrheitstabelle anzugeben, über die die Elemente der zusammengesetzten Konfusionsmatrix fortgepflanzt werden können.

Wahrscheinlichkeitsvektor

Ein zweites Verfahren aus dem Fernerkundungsbereich leitet einen Wahrscheinlichkeitsvektor individuell für jede Rasterzelle ab. Der Vektor enthält die Wahrscheinlichkeiten, mit der die Zelle aufgrund ihrer beobachteten Eigenschaften zu jeder einzelnen Klasse zugeordnet werden kann. Die Wahrscheinlichkeiten werden z.B. innerhalb der bekannten Maximum-Likelihood-Methode (z.B. Richards, 1986) geschätzt, wobei dort die höchste Wahrscheinlichkeit die endgültig dem Attribut zugewiesene Klasse definiert. Der Wahrscheinlichkeitsvektor repräsentiert eine Art innere Unsicherheit, da kein direkter Vergleich mit der Wirklichkeit erfolgt.

Wahrscheinlichkeitsvektoren für Rasterdaten und die bereits behandelten Wahrscheinlichkeiten für Vektordaten definieren eine gemeinsame Basis, auf der eine Intergration der Unsicherheit beider Datenarten möglich wird (Shi, 1994). Darauf aufbauend kann die Verarbeitung von hybriden Daten (z.B. die geometrische Überlagerung von Raster- und Vektordaten) unter Berücksichtigung der Unsicherheit realisiert werden (Fritsch et al., 1998). Der Wahrscheinlichkeitsvektor kann auch dazu genutzt werden, mit den klassifizierten Daten eine verbesserte Objektbildung durchzuführen (Klein et al., 1998). Die Objektbildung erfolgt direkt auf den aus dem Vektor abzuleitenden Wahrscheinlichkeitsmatrizen für jede Objektklasse. Im Vergleich zur bisherigen Lösung ergeben sich überlappende Objekte, wodurch die enthaltene Unsicherheit verdeutlicht wird.

Für komplexe Analysen, die sich nicht durch algebraische Zusammenhänge beschreiben lassen, müssen Simulationen für die Fortpflanzung der Unsicherheit sorgen. In Simulationen werden eine große Anzahl von möglichen Realisierungen der Daten erzeugt und mit ihnen die Analyse durchgeführt, um am Ende aus der Variation der Ergebnisse statistische Maße abzuleiten. Zur Erzeugung von Realisierungen ist die Definition eines Fehlermodells notwendig (Goodchild und Dubuc, 1987). Darin werden alle Einflüsse, die zur Unsicherheit der Daten beitragen, integriert. Goodchild und Wang (1989) beschreiben ein Fehlermodell in Form eines autoregressiven Prozesses, der als Grundlage die Wahrscheinlichkeitsvektoren der Zellen und einen globalen Autokorrelationskoeffizienten r beinhaltet, der die Abhängigkeiten zwischen Nachbarzellen angibt. Verschiedene Tests zeigen, dass sowohl Homogenität der Attributwerte als auch Übergangszonen zwischen verschiedenen Werten geeignet simuliert werden können (Goodchild et al., 1992b). Zur vereinfachten Beachtung der Autokorrelation wird vorgeschlagen, entsprechende Basiszufallsdatensätze für diskrete Korrelationswerte permanent im System bereitzuhalten.

Fuzzy-Ansatz

Der Fuzzy-Ansatz macht sich die Theorie der unscharfen Mengen (Fuzzy-Subset-Theorie), die von Zadeh (1965) eingeführt wurde, zunutze. Die Motivation für die Anwendung der Theorie besteht darin, dass häufig keine absolute Referenz für die Daten existiert, sondern die Klassen nur unscharf in der Wirklichkeit wiedergefunden werden können. Aus diesem Grund soll die Unschärfe in der Wirklichkeit durch eine unsichere Zuweisung zu einer Klasse repräsentiert werden. Als Maß für die Unsicherheit ist die Zugehörigkeit m zu einer jetzt unscharfen Klasse zu bestimmen. Üblicherweise erhält m einen Wert im Intervall zwischen 0 und 1. Die einzelnen Werte können analog zum Wahrscheinlichkeitsvektor in einem Zugehörigkeitsvektor abgelegt werden. Es existieren verschiedene Möglichkeiten, den Zugehörigkeitsvektor zu schätzen, die übersichtsweise in Lowell (1994) zusammengefasst sind. Ein interessantes Verfahren wendet sich der Datenerfassung zu. Dabei werden zunächst nur sichere Informationen (Kernbereiche von Objekten und klare Trennungslinien) erfasst. In einem zweiten Schritt schließt sich die Berechnung der Zugehörigkeitswerte (Fuzzifizierung) für alle Rasterzellen mit Hilfe einer festzulegenden Zugehörigkeitsfunktion an. Es entsteht eine Fuzzy-Oberflächen-Karte für jeden einzelnen Attributwert. Die Karte enthält Fuzzy-Regionen (Fisher und Pathirana, 1990), die die Verteilung der Konzentration eines Attributwertes über den Raum veranschaulichen. Eine andere Möglichkeit der Bestimmung bieten alternative Klassifikationsalgorithmen aus dem Fernerkundungsbereich, die sich die Fuzzy-Subset-Theorie zunutze machen, um Zugehörigkeiten von Pixeln zu verschiedenen Klassen explizit zuzulassen. Dazu zählen z.B. das Fuzzy-k-Means-Verfahren (Bezdek et al., 1984), ein unüberwachtes Clusterverfahren, und die überwachte Fuzzy-Klassifizierung (Wang, 1990).

Zur Fortpflanzung der Unsicherheit kommen Operatoren zum Einsatz, die speziell innerhalb der Fuzzy-Subset-Theorie definiert sind (Anhang B.4). Lowell (1995) untersucht dem logischen UND und logischen ODER entsprechende Operatoren (weiches UND bzw. ODER, hartes UND bzw. ODER). Es ergibt sich ein kontinuierliches Ergebnis, das durch Schrankenbildung mit einem bestimmten Zugehörigkeitswert (Defuzzifizierung) wieder in eine scharfe Darstellung umgewandelt werden kann.

In der gleichen Art und Weise können auch Entfernungs- und Richtungsbestimmungen zwischen Fuzzy-Regionen ausgedrückt werden (Altman, 1994). Zum Beispiel ergibt sich als Entfernung zwischen den Regionen A und B eine unscharfe Menge an verschiedenen Distanzen mit jeweiligen Zugehörigkeiten. Eine mögliche Defuzzifizierung sucht das Maximum der Zugehörigkeit und findet so die gesuchte geringste Entfernung zwischen den Regionen. Die unscharfe Distanzmenge lässt sich z.B. zur Bewertung eines unscharfen Prädikates (z.B. $benachbart(Region A, Region B)$) weiterverwenden.

B Grundlagen der Unsicherheitsmodelle

B.1 Grundlagen des Stochastischen Modells

Wichtigster Grundbaustein des Stochastischen Modells ist die beschreibende Statistik. Sie stellt Kenngrößen bereit, um die Variation in einer Beobachtungsfolge zu beschreiben. Eine Folge enthält eine Menge einzelner Werte (Realisierungen) einer Funktion X , die als Zufallsvariable oder stochastische Variable bezeichnet wird. Die Zufallsvariable veranschaulicht mögliche Ergebnisse eines Zufallsexperimentes und besitzt folgende Eigenschaften (Kreyszig, 1965):

1. Die Werte von X sind reelle Zahlen.
2. Für jede Zahl a und für jedes Intervall I auf einer Zahlengeraden existiert die Wahrscheinlichkeiten der Aussagen (Ereignisse) „ X besitzt den Wert a “ bzw. „ X liegt im Intervall I “ unter Beachtung der Axiome der Wahrscheinlichkeitstheorie.

Es können zwei Klassen an Variablen unterschieden werden: diskrete und stetige Variablen. Diskrete Variablen erlauben nur endlich viele oder abzählbar unendlich viele Werte mit positiver Wahrscheinlichkeit, während für stetige Variablen diese Einschränkung aufgehoben ist. Da die Geometrie von Objekten in Form von kontinuierlichen Werten beschrieben wird, konzentrieren sich die nachfolgenden Betrachtungen auf stetige Variablen.

Die zugehörige Grundgesamtheit (Menge aller möglichen Werte) besitzt eine charakteristische Verteilung, die sich durch Dichtefunktion $f(x)$ oder Verteilungsfunktion $F(x)$ vollständig beschreiben lässt. Die Verteilungsfunktion ist definiert als Wahrscheinlichkeit der Aussage „ $X \leq x$ “:

$$F(x) = p(X \leq x). \quad (\text{B-1})$$

Die zugehörige Dichte (auch Wahrscheinlichkeitsdichte genannt) ergibt sich dann als erste Ableitung der Verteilungsfunktion:

$$f(x) = F'(x). \quad (\text{B-2})$$

Zur weiteren Charakterisierung der Verteilung einer Zufallsvariablen dienen die Momente. Sie sind über die mathematische Erwartung definiert. Der Erwartungswert einer Funktion $g(x)$ ist folgendermaßen festgelegt:

$$E(g(x)) = \int_{-\infty}^{\infty} g(x) f(x) dx \quad (\text{B-3})$$

und benötigt die Angabe der Dichtefunktion $f(x)$. Der Wert lässt sich als die durchschnittliche Gewinnerwartung eines Glücksspielers pro Spiel interpretieren. Allgemein ergibt das k -te Moment einer Verteilung über $g(x) = X^k$ zu:

$$E(X^k) = \int_{-\infty}^{\infty} x^k f(x) dx. \quad (\text{B-4})$$

Das erste Moment definiert den Mittelwert m der Verteilung:

$$\text{Mittelwert : } \mathbf{m}_x = E(x) = \int_{-\infty}^{\infty} x f(x) dx. \quad (\text{B-5})$$

Vor Festlegung des zweiten Momentes wird speziell $g(x) = (X - \mathbf{m})^k$ gewählt. Die Reduktion der Zufallsvariablen auf den Mittelwert bewirkt, dass die Momente als zentrale Momente bezeichnet werden. Entsprechendes gilt für das zweite zentrale Moment, das die Varianz (Dispersion) festlegt:

$$\text{Varianz : } \mathbf{s}_x^2 = E((X - \mathbf{m})^2) = \int_{-\infty}^{\infty} (x - \mathbf{m})^2 f(x) dx. \quad (\text{B-6})$$

Weitere Momente, wie Schiefe und Exzess, werden in gleicher Weise abgeleitet. Anstelle der Varianz wird häufig die Standardabweichung benutzt, die sich als positive Quadratwurzel aus der Varianz berechnet. Auch für mehrere voneinander abhängige Zufallsvariablen können die Momente der nun mehrdimensionalen Verteilungen abgeleitet werden. Von besonderem Interesse sind dabei Kovarianzen:

$$\text{Kovarianz : } \mathbf{s}_{xy} = E((X - \mathbf{m}_x)(Y - \mathbf{m}_y)) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mathbf{m}_x)(y - \mathbf{m}_y) f(x, y) dx dy, \quad (\text{B-7})$$

über die sich der Grad der gegenseitigen Abhängigkeit zweier Zufallsvariablen in Form der Korrelation \mathbf{r} ausdrücken lässt. Sie wird festgelegt durch:

$$\text{Korrelation : } \mathbf{r} = \frac{\mathbf{s}_{xy}}{\mathbf{s}_x \mathbf{s}_y} \quad (\text{B-8})$$

Die Korrelation nimmt Werte im Intervall $I = [-1, 1]$ an. Eine vollständige Korrelation der beiden Zufallsvariablen liegt bei $|\mathbf{r}| = 1$ vor, während ein Korrelationswert von Null Unabhängigkeit anzeigt.

Voneinander abhängige Zufallsvariablen werden häufig in Vektoren (Zufallsvektoren) zusammengefasst. Während die zugehörigen Mittelwerte ebenfalls in einem Vektor Platz finden, sind Varianzen und Kovarianzen in Matrixform anzuordnen. Eine solche Matrix wird Kovarianz- oder Dispersionsmatrix genannt und hat folgendes Aussehen:

$$\text{Dispersionsmatrix : } \mathbf{D} \left(\begin{bmatrix} X_1 \\ X_2 \\ \dots \\ X_n \end{bmatrix} \right) = \begin{bmatrix} \mathbf{s}_{x_1}^2 & \mathbf{s}_{x_1 x_2} & \dots & \mathbf{s}_{x_1 x_n} \\ \mathbf{s}_{x_2 x_1} & \mathbf{s}_{x_2}^2 & \dots & \mathbf{s}_{x_2 x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{s}_{x_n x_2} & \mathbf{s}_{x_n x_2} & \dots & \mathbf{s}_{x_n}^2 \end{bmatrix}. \quad (\text{B-9})$$

Werden aus den Zufallsvariablen neue Werte rechnerisch abgeleitet, so sind diese wiederum als Zufallsvariablen zu betrachten. Zusätzlich muss eine Übertragung der Varianz realisiert werden. Die Übertragung ist unter dem Begriff der Varianzfortpflanzung (oder Fehlerfortpflanzung) bekannt. Für einen linearen Zusammenhang zwischen einem Eingangszufallsvektor \mathbf{x} und einem Ergebnisvektor \mathbf{y} der Form:

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{b} \quad (\text{B-10})$$

definiert sich die Varianzfortpflanzung zu (allgemeines Fehlerfortpflanzungsgesetz):

$$\mathbf{D}(\mathbf{y}) = \mathbf{D}(\mathbf{A}\mathbf{x} + \mathbf{b}) = \mathbf{A}\mathbf{D}(\mathbf{x})\mathbf{A}^T. \quad (\text{B-11})$$

Die Matrix \mathbf{A} wird als Designmatrix (oder Jakobimatrix) bezeichnet. Sie enthält die linearen Faktoren, die den funktionalen Zusammenhang repräsentieren. Bei allgemeinen nichtlinearen Zusammenhängen $f(x)$ ist zunächst eine Linearisierung an einer Näherungsstelle x_0 für die Erwartungswerte von x durchzuführen, woraufhin das Fortpflanzungsgesetz auch auf die Differenzen $(\Delta x, \Delta y)$ anzuwenden ist. An die Stelle der linearen Faktoren in der Designmatrix treten über partielle Ableitungen gewonnene linearisierte Koeffizienten $(\frac{\partial f}{\partial x})$.

Die Berechnung der Kenngrößen Mittelwert und Varianz erfolgt über die Auswertung der Realisierungen der Zufallsvariablen. Da stets nur eine Stichprobe mit endlich großer Anzahl ermittelt werden kann, ist nur eine Schätzung der Kenngrößen möglich. Eine Schätzung des Mittelwertes gelingt über die arithmetische Mittelbildung:

$$\hat{\mathbf{m}}_x = \frac{1}{n} \sum_{i=1}^n x_i \quad (\text{B-12})$$

Daraus aufbauend ergibt sich die Schätzung für die Varianz zu:

$$\hat{\mathbf{s}}_x^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \hat{\mathbf{m}}_x)^2. \quad (\text{B-13})$$

Als eine der wichtigsten Verteilungen gilt die Normalverteilung (Gauß-Verteilung). Viele Messgrößen, wie z.B. Koordinaten oder Strecken sind als normalverteilte Zufallsvariablen anzusehen. Gestärkt wird die Bedeutung der Normalverteilung durch die Eigenschaft, dass das Zusammentreffen einer großen Anzahl beliebig verteilter Einflüsse näherungsweise zu einer normalverteilten Größe führt (zentraler Grenzwertsatz) (Koch, 1980). Die Normalverteilung besitzt die folgende Wahrscheinlichkeitsdichte:

$$f(x) = \frac{1}{\mathbf{s}\sqrt{2\mathbf{p}}} e^{-\frac{1}{2}\left(\frac{x-\mathbf{m}}{\mathbf{s}}\right)^2} \quad \text{für } -\infty < x < \infty \text{ und } \mathbf{s} > 0. \quad (\text{B-14})$$

Als Parameter der Funktion finden sich Mittelwert \mathbf{m} und Varianz bzw. Standardabweichung \mathbf{s} wieder. Die Verteilungsfunktion ergibt sich durch Integration der Dichte in folgender Weise:

$$F(x) = \frac{1}{\mathbf{s}\sqrt{2\mathbf{p}}} \int_{-\infty}^x e^{-\frac{1}{2}\left(\frac{u-\mathbf{m}}{\mathbf{s}}\right)^2} du. \quad (\text{B-15})$$

Eine geschlossene Lösung des Integrals existiert nicht, so dass man sich einer Näherungslösung bedienen muss. Eine Möglichkeit bietet der Rückgriff auf tabellierte Werte, die vielfach für die Standardnormalverteilung $N(\mathbf{m}=0, \mathbf{s}^2=1)$ aufzufinden sind (z.B. Kreyszig, 1965). Zum Zwecke der Automation bedarf es numerischer Lösungsansätze. Eine polynomiale Approximation für die Standardnormalverteilung ist mit einer Genauigkeit von $1 \cdot 10^{-5}$ in Koch (1980) angegeben und lautet folgendermaßen:

$$F_{0,1}(x) = 1 - \frac{e^{-x^2/2}}{\sqrt{2\mathbf{p}}} (a_1 t + a_2 t^2 + a_3 t^3), \quad \text{für } x \geq 0 \quad (\text{B-16})$$

mit

$$\begin{aligned} t &= \frac{1}{1+px} \\ p &= 0.33267 \\ a_1 &= 0.4361836 \\ a_2 &= -0.1201676 \\ a_3 &= 0.9372980. \end{aligned} \quad (\text{B-17})$$

Der zu einem bestimmten Funktionswert \mathbf{a} gehörende Abzissenwert x_a wird als \mathbf{a} -Fraktile (oder Quantil) bezeichnet. Auch Fraktile können in Tafeln nachgesehen oder ebenfalls auf numerischem Wege berechnet werden. Eine Lösung mit einer Genauigkeit von $4.5 \cdot 10^{-4}$ findet sich in Koch (1980) und lautet:

$$x_a = t - \frac{c_0 + c_1 t + c_2 t^2}{1 + d_1 t + d_2 t^2 + d_3 t^3} \quad \text{für } 0.5 \leq \mathbf{a} \leq 1, \quad (\text{B-18})$$

mit

$$\begin{aligned} t &= \sqrt{\ln\left(\frac{1}{(1-\mathbf{a})^2}\right)} \\ c_0 &= 2.515517 & d_1 &= 1.432788 \\ c_1 &= 0.802853 & d_2 &= 0.189269 \\ c_2 &= 0.010328 & d_3 &= 0.001308. \end{aligned} \quad (\text{B-19})$$

Die enge Verzahnung der Stochastik mit der Wahrscheinlichkeitstheorie kann bereits an der Definition einer Zufallsvariablen abgelesen werden. Sie bildet einen wichtigen Bestandteil, mit dem es möglich wird, Zufallserscheinungen (Ereignisse oder Aussagen) zu bewerten. Die klassische Definition der mathematischen Wahrscheinlichkeit nach Laplace lautet:

Die Wahrscheinlichkeit $p(A)$ eines Ereignisses A bei einem Zufallsexperiment ist durch

$$p(A) = \frac{g}{m} \quad (\text{B-20})$$

mit

$$\begin{aligned} g &\text{ als Anzahl der günstigen Fälle, bei denen } A \text{ eintritt und} \\ m &\text{ als Anzahl aller gleichmöglichen Fälle} \end{aligned} \quad (\text{B-21})$$

gegeben.

Die Wahrscheinlichkeit kann gemäß der klassischen Definition als relative Häufigkeit des Auftretens eines bestimmten Ereignisses angesehen werden. Dabei wird vorausgesetzt, dass eine große Anzahl an Realisierungen zur Bestimmung vorliegt. Eine Idealisierung der zur relativen Häufigkeit zu beobachtenden Eigenschaften führt zur Formulierung von Axiomen der Wahrscheinlichkeit, die eine andere allgemeinere Definition des Begriffes erlauben. Folgende drei Axiome können aufgestellt werden:

$$\begin{aligned} \text{Nichtnegativität : } & 0 \leq p(A) \leq 1 \\ \text{Normierung : } & p(S) = 1 \\ \text{Additivität : } & p(A_1 \vee A_2) = p(A_1) + p(A_2), \end{aligned} \quad (\text{B-22})$$

mit A als eine beliebige, zu bewertende Aussage, S als sichere Aussage und A_1 und A_2 als zwei sich gegenseitig ausschließende Aussagen. Weitere Festlegungen kommen hinzu. Für ein komplementäres Ereignis \bar{A} gilt:

$$p(A) = 1 - p(\bar{A}). \quad (\text{B-23})$$

Aufgrund der Einschränkung der Additivität stellt sich die Frage, wie mit beliebigen Aussagen verfahren werden soll. Der allgemeine Additionssatz gibt darauf die Antwort. Er formuliert die Additivität zweier beliebiger Aussagen als:

$$p(A \vee B) = p(A) + p(B) - p(A \wedge B). \quad (\text{B-24})$$

Die Summe der Einzelwahrscheinlichkeiten wird um die Wahrscheinlichkeit des gleichzeitigen Eintreffens der beiden Aussagen A und B verringert. Sind diese voneinander unabhängig, dann lässt sich das gleichzeitige Eintreffen zu:

$$p(A \wedge B) = p(A) \cdot p(B) \quad (\text{B-25})$$

bestimmen. Im Allgemeinen kann nicht vom Vorliegen unabhängiger Aussagen ausgegangen werden. Daher interessiert zunächst, welche Einwirkungen das Eintreffen einer Aussage auf die Wahrscheinlichkeit einer anderen besitzt. Ein solcher Zusammenhang drückt sich dadurch aus, dass man das Auftreten einer Aussage bewertet, für die aber hypothetisch vorausgesetzt wird, dass eine andere Aussage bereits gültig ist. Es entsteht die bedingte Wahrscheinlichkeit einer Aussage B unter der Bedingung des Eintreffens der Aussage A . Sie ist wie folgt zu beschreiben:

$$\text{bedingte Wahrscheinlichkeit : } p(B/A) = \frac{p(A \wedge B)}{p(A)} \quad \text{mit } p(A) > 0. \quad (\text{B-26})$$

Sind beide Aussagen unabhängig, so gilt:

$$p(B/A) = p(B), \quad (\text{B-27})$$

und schließen sich die Aussagen gegenseitig aus ($p(A \wedge B) = 0$), so ergibt sich:

$$p(B/A) = 0. \quad (\text{B-28})$$

Die Umformulierung der bedingten Wahrscheinlichkeit (B-26) liefert den allgemeinen Multiplikationssatz, der die folgende Gleichung beschreibt:

$$\text{Multiplikation : } p(A \wedge B) = p(A) \cdot p(B/A) = p(B) \cdot p(A/B). \quad (\text{B-29})$$

Für eine Reihe paarweise sich gegenseitig ausschließenden Aussagen ($A_i \wedge A_j$, für $i \neq j$), die zusammen eine sichere Aussage bilden ($S = A_1 \vee A_2 \vee \dots \vee A_n$) ergibt sich die Wahrscheinlichkeit einer beliebigen anderen Aussage B , die in Kombination mit den anderen Aussagen zu sehen ist, zu:

$$\text{Totale Wahrscheinlichkeit: } p(B) = \sum_{i=1}^n p(A_i) \cdot p(B|A_i). \quad (\text{B-30})$$

Der so berechnete Wahrscheinlichkeitswert wird als totale Wahrscheinlichkeit bezeichnet. Er kann seinerseits genutzt werden, um die bedingten Wahrscheinlichkeiten des Eintreffens der Einzelereignisse A_i unter der Bedingung, dass B bereits eingetreten ist, zu bestimmen. Es ergibt sich das Theorem von Bayes:

$$\text{Bayes - Theorem: } p(A_i|B) = \frac{p(A_i) \cdot p(B|A_i)}{\sum_{j=1}^n p(A_j) \cdot p(B|A_j)}. \quad (\text{B-31})$$

Der bedingte Wahrscheinlichkeitswert $p(A_i|B)$ wird als a-posteriori-Wahrscheinlichkeit von A_i bezeichnet, während die Wahrscheinlichkeit $p(A_i)$ die a-priori-Wahrscheinlichkeit angibt. Das Bayes-Theorem eignet sich um festzustellen, wie sich die Wahrscheinlichkeit ändert (von a-priori nach a-posteriori), wenn neue Informationen (hier: B) hinzukommen.

Betreffen die formulierten Aussagen das Auftreten von bestimmten Werten einer Zufallsvariablen, so kann man zur Bestimmung der Wahrscheinlichkeit direkt auf die Verteilungsfunktion der Variablen zurückgreifen. Für die hier behandelten stetigen Zufallsvariablen sind die Aussagen so aufzustellen, dass alle möglichen Werte im Intervall $I = (-\infty, x]$ liegen. Eine gültige Aussage hat dann die Form:

$$A = X \leq x. \quad (\text{B-32})$$

Die Wahrscheinlichkeit p einer solchen Aussage ergibt sich durch den Funktionswert der Verteilungsfunktion F für die obere Intervallgrenze x :

$$p(A) = p(X \leq x) = F(x) \quad (\text{B-33})$$

oder mit Hilfe der zugehörigen Dichtefunktion der Verteilung $f(t)$ zu:

$$p(A) = \int_{-\infty}^x f(t) dt. \quad (\text{B-34})$$

Als Beispiel soll die Wahrscheinlichkeit bestimmt werden, mit der die gemessene Entfernung D kleiner als 100 m ist. Für die Größe D gilt die Annahme, sie sei normalverteilt, woraufhin die Bestimmung der Wahrscheinlichkeit folgendes Aussehen besitzt:

$$p(D \leq 100\text{m}) = \int_{-\infty}^{d=100} f(t) dt = \frac{1}{s_D \sqrt{2\pi}} \int_{-\infty}^{d=100} e^{-\frac{1}{2} \left(\frac{t-m_D}{s_D} \right)^2} dt = F_D(d). \quad (\text{B-35})$$

Wie bereits erwähnt, existiert keine geschlossene Lösung des Integrals, so dass auf die beschriebenen Näherungslösungen zurückgegriffen werden muss.

Nimmt die Zufallsvariable nur Werte in einem bestimmten Intervall $I =] a, b]$ an, dann ergibt sich die Wahrscheinlichkeit für das Auftreten solcher Werte durch:

$$p(a < X \leq b) = F(b) - F(a). \quad (\text{B-36})$$

Bei sehr kleinen Intervallen ($I =] a, a + dx]$) kann näherungsweise der Dichtefunktionswert $f(a)$ zusammen mit der Intervallgröße dx benutzt werden:

$$p(a < X < a + dx) = f(a) \cdot dx. \quad (\text{B-37})$$

Zu beachten ist jedoch, dass das Eintreffen genau eines exakten Wertes ($dx = 0$) im Falle von stetigen Zufallsvariablen immer die Wahrscheinlichkeit Null liefert ($p(X = x) = 0 !$).

Für manche Zwecke ist es sinnvoll abzuschätzen, mit welcher Wahrscheinlichkeit sich Realisierungen innerhalb eines bestimmten Toleranzbereiches um den Mittelwert befinden. Der Toleranzbereich wird häufig als ein Vielfaches n der Standardabweichung s gebildet und kann als Intervall T folgendermaßen formuliert werden:

$$T =] \mathbf{m} - n\mathbf{s}, \mathbf{m} + n\mathbf{s}]. \quad (\text{B-38})$$

Unter Annahme der Normalverteilung ergibt sich erst etwa mit einem Faktor $n=3$ eine so hohe Wahrscheinlichkeit $p(\mathbf{m} - 3\mathbf{s} < X \leq \mathbf{m} + 3\mathbf{s}) = 0.997$, dass in praktischen Anwendungen alle Realisierungen innerhalb des Intervalls liegen werden. Auf dieser Basis lässt sich die im Minimum-Maximum-Modell als Maßzahl eingesetzte Variationsbreite abschätzen. Dazu ist das Intervall T abzuschließen (T'), wodurch die Variationsbreite als Durchmesser des Intervalls mit $n=3$ lautet:

$$d(T') = | \mathbf{m} + n\mathbf{s} - (\mathbf{m} - n\mathbf{s}) | = 2n\mathbf{s} = 6\mathbf{s}. \quad (\text{B-39})$$

B.2 Ableitung der Wahrscheinlichkeit für die Linienzugehörigkeit bei stochastischer Breite des Phänomens

In ausgewählten Fällen kann es vorkommen, dass die Breite b , die die räumliche Ausdehnung eines linien- oder punkthaften Objektes abschätzt, nicht als Konstante behandelt werden kann, sondern ebenfalls als stochastische Variable anzusehen ist. Ein solcher Fall liegt beispielsweise dann vor, wenn die Breite über den Verlauf eines Objektes variiert (z.B. bei Wegen oder Flüssen). Anstelle des konstanten Wertes tritt dann eine mittlere Breite \mathbf{m}_b mit ihrer Varianz \mathbf{s}_b^2 . Die Breite bestimmt die Ränder des Bandes \mathbf{a} und \mathbf{b} um die Linie L . Die vollständige Dispersionsmatrix für die vier auftretenden Zufallsvariablen $\mathbf{a}, \mathbf{b}, L$ und b kann folgendermaßen angegeben werden:

$$D \begin{pmatrix} \mathbf{a} \\ \mathbf{b} \\ L \\ b \end{pmatrix} = \begin{bmatrix} \mathbf{s}_L^2 + \frac{\mathbf{s}_b^2}{4} & \mathbf{s}_L^2 - \frac{\mathbf{s}_b^2}{4} & \mathbf{s}_L^2 & -\frac{\mathbf{s}_b^2}{2} \\ \mathbf{s}_L^2 - \frac{\mathbf{s}_b^2}{4} & \mathbf{s}_L^2 + \frac{\mathbf{s}_b^2}{4} & \mathbf{s}_L^2 & \frac{\mathbf{s}_b^2}{2} \\ \mathbf{s}_L^2 & \mathbf{s}_L^2 & \mathbf{s}_L^2 & 0 \\ -\frac{\mathbf{s}_b^2}{2} & \frac{\mathbf{s}_b^2}{2} & 0 & \mathbf{s}_b^2 \end{bmatrix}. \quad (\text{B-40})$$

Durch die variable Entfernung der beiden Ränder ist explizit auszuschließen, dass diese Entfernung (also die Breite b) kleiner als Null wird. Dazu muss die folgende Bedingung eingehalten werden:

$$A_b = (b \geq 0). \quad (\text{B-41})$$

Zur Bestimmung der Zugehörigkeit einer beliebigen Position zur Linie ist folgende Gesamtaussage zu erfüllen:

$$A_L = (A_{L_a} \wedge A_{L_b}) / A_b. \quad (\text{B-42})$$

Die Bewertung dieser Aussage ergibt sich zu:

$$p(A_L) = \frac{p(A_{L_a}) \cdot p(A_{L_b} / A_{L_a})}{p(A_b)}. \quad (\text{B-43})$$

Für absolute Positionen resultiert:

$$p(A_L) = \frac{p(\mathbf{a} \leq \text{pos}) \cdot p(\text{pos} \leq \mathbf{b} \leq \text{pos} + b)}{p(\mathbf{b} \leq \text{pos} + b) \cdot p(b \geq 0)}. \quad (\text{B-44})$$

Da die Werte für $p(A_{L_a})$ und $p(A_{L_b})$ identisch sind, lässt sich (B-44) schreiben als:

$$p(A_L) = \frac{p(\text{pos} \leq \mathbf{b} \leq \text{pos} + b)}{p(b \geq 0)}. \quad (\text{B-45})$$

Geht man von absoluten Positionen zu Distanzwerten über, so hat (B-44) folgendes Aussehen:

$$p(A_L) = \frac{p(D-b/2 \leq 0) - p(D+b/2 \leq 0)}{p(b \geq 0)} = \frac{p(I_1 \leq 0) - p(I_2 \leq 0)}{p(b \geq 0)}. \quad (\text{B-46})$$

Die Bewertung hat über die beiden neu gebildeten Zufallsvariablen I_1 und I_2 zu erfolgen. Aus der gegebenen Dispersionsmatrix lassen sich die zugehörigen Varianzen zu:

$$s_{I_1}^2 = s_{I_2}^2 = s_L^2 + \frac{s_b^2}{4} \quad (\text{B-47})$$

ermitteln. Die Auswirkungen der Erweiterungen werden am bereits gegebenen Linienbeispiel (Straße mit $s_L = 3 \text{ m}$, $d = 5 \text{ m}$ und $b = 8 \text{ m}$) gezeigt (Kapitel 2.4). Dazu wird zusätzlich angenommen, dass die Breite b eine Standardabweichung von $s_b = 3 \text{ m}$ besitzt, d.h. es können sowohl breitere als auch schmalere Stellen vorkommen. Es ergeben sich die folgenden neuen Werte:

$$s_{I_1} = s_{I_2} = 3.35 \text{ m} \quad (\text{B-48})$$

$$\begin{aligned} p(D-b/2 \leq 0) &= p(I_1 \leq 0) = 0.382 \\ p(D+b/2 \leq 0) &= p(I_2 \leq 0) = 0.004 \\ p(A_b) &= p(b \geq 0) = 0.996 \end{aligned} \quad (\text{B-49})$$

$$p(A_L) = \frac{p(I_1 \leq 0) - p(I_2 \leq 0)}{p(b \geq 0)} = \frac{0.382 - 0.004}{0.996} = 0.380. \quad (\text{B-50})$$

Vergleicht man den berechneten Wert für die stochastische Breite mit der Lösung für eine konstante Breite ($p_{\text{const}}(A_L) = 0.370$), so stellt man einen leichten Anstieg der Wahrscheinlichkeit fest, d.h. es wird wahrscheinlicher, dass die Position zum Objekt gehört. Insgesamt vergrößern sich die Bereiche, innerhalb denen noch Positionen zur Linie zu zählen sind. Trotz der großen Standardabweichung ($s_L : s_b = 1 : 1$) ändert sich die Wahrscheinlichkeit im Beispiel aber nur geringfügig. Nur beträchtliche Variationen in der Breite wirken sich aus. Eine größere Rolle spielt dagegen die Breite selbst. Wird diese sehr viel kleiner angenommen, so ergibt sich auch im Quotienten ein kleinerer Wert (minimaler Wert: $p_{b=s}(b \geq 0) = 0.841$) und damit insgesamt eine größere Wahrscheinlichkeit (im Beispiel: $p_{b=s}(A_L) = 0.449$).

B.3 Grundlagen des Minimum-Maximum-Modells

Die Betrachtungen zur Intervallarithmetik bilden das theoretische Grundgerüst des Minimum-Maximum-Modells (Große et al., 1995). Die Verbindung zur geometrischen Unsicherheit wird über die Realisierungen eines Phänomens hergestellt, die einem bestimmten Wertebereich entstammen, der die Variation des Phänomens charakterisiert. Der Wertebereich definiert ein Intervall, in dem alle möglichen Realisierungen zu finden sind. Jeder einzelne Wert im Intervall stellt gleichzeitig eine mögliche Realisierung dar. Intervalle können im mathematischen Sinne durch Intervallzahlen repräsentiert werden. Voraussetzung dafür ist allerdings das Vorliegen eines beschränkten und abgeschlossenen Intervalls. Da räumliche Objekte im Normalfall eine endliche geometrische Ausdehnung besitzen, sollte diese Voraussetzung im hier vorliegenden Anwendungsfall stets erfüllt sein. Eine Intervallzahl A weist allgemein die folgende Form auf:

$$A = [a_1, a_2] = \{x / x \in R, a_1 \leq x \leq a_2\}, \quad (\text{B-51})$$

mit x als beliebiges Element der Menge R und den Intervallgrenzen a_1 und a_2 . Die Menge R bezeichnet die Universalmenge der möglichen Geometrien aller Objekte. Die Intervallgrenzen legen die obere (für a_2) bzw. die untere Grenze (für a_1) der Grundgesamtheit M des Phänomens fest. Diese sind entsprechend als Supremum bzw. Infimum des Phänomens zu bezeichnen. Durch die geforderte Abgeschlossenheit des Intervalls sind Supremum und Infimum immer auch Elemente der Grundgesamtheit selbst. In diesem Fall dürfen die Intervallgrenzen als Maximum bzw. Minimum bezeichnet werden. Sie sind formal definiert als:

$$\begin{aligned}
 \text{Minimum} : a_1 &= \min M = \inf_{x \in M} x, \quad \forall x \in M : a_1 \leq x, \quad a_1 \in M \\
 \text{Maximum} : a_2 &= \max M = \sup_{x \in M} x, \quad \forall x \in M : a_2 \geq x, \quad a_2 \in M.
 \end{aligned}
 \tag{B-52}$$

Die weitere Verarbeitung der Intervalle wird durch Rechenoperationen ermöglicht. Dazu müssen insbesondere die arithmetischen Grundoperationen zur Verfügung stehen. Sie sind für Intervallzahlen $A = [a_1, a_2]$ und $B = [b_1, b_2]$ in folgender Weise definiert:

$$\begin{aligned}
 A + B &= [a_1 + b_1, a_2 + b_2], \\
 A - B &= [a_1 - b_1, a_2 - b_2], \\
 A \cdot B &= [\min\{a_1 b_1, a_1 b_2, a_2 b_1, a_2 b_2\}, \max\{a_1 b_1, a_1 b_2, a_2 b_1, a_2 b_2\}], \\
 A : B &= A \cdot \left[\frac{1}{b_2}, \frac{1}{b_1} \right], \quad 0 \notin B.
 \end{aligned}
 \tag{B-53}$$

Zusätzlich kann auf einige spezielle Intervallfunktionen zurückgegriffen werden. Als Erstes ist der Abstand zweier Intervallzahlen festgelegt durch:

$$\text{Abstand} : q(A, B) = \max\{|a_1 - b_1|, |a_2 - b_2|\}.
 \tag{B-54}$$

Weiterhin kann der Betrag einer Intervallzahl ausgedrückt werden als:

$$\text{Betrag} : |A| = q(A, [0, 0]) = \max\{|a_1|, |a_2|\},
 \tag{B-55}$$

und der Durchmesser einer Intervallzahl wird definiert zu:

$$\text{Durchmesser} : d(A) = a_2 - a_1.
 \tag{B-56}$$

Die Variation eines Phänomens kann folglich durch eine Intervallzahl charakterisiert werden. Formal ergibt sich die Variation V dann zu:

$$V = \left[\min_{x \in M} x, \max_{x \in M} x \right] = [\min M, \max M].
 \tag{B-57}$$

Als Maßzahl zur Beschreibung der Unsicherheit kann der Durchmesser $d(V)$ fungieren. Er ist in diesem Zusammenhang auch als Spannweite oder Variationsbreite bekannt (Sachs, 1978).

Da stets nur eine Stichprobe aus der Grundgesamtheit beobachtet wird, können nur Schätzungen für die Variation und deren Durchmesser ermittelt werden. Eine Schätzung erfolgt so, dass der größte Wert der Stichprobe das Maximum, der kleinste Wert das Minimum und die Differenz beider den Durchmesser, d.h. die Variationsbreite bestimmen.

Die Variationsbreite kann genutzt werden, um die Standardabweichung der zugehörigen stochastischen Variablen abzuschätzen. Dazu müssen mehrere voneinander unabhängige Beobachtungen der Variationsbreite V vorliegen, so dass sich die gesuchte Standardabweichung folgendermaßen ableiten lässt:

$$\hat{s} = \sqrt{\frac{\sum_{i=1}^n V_i^2}{2n}}
 \tag{B-58}$$

für n beobachtete Variationsbreiten. Geht man von einer normalverteilten Zufallsvariablen aus und liegt eine umfangreiche Stichprobe der Schätzung der Variationsbreite zugrunde, so kann die Standardabweichung in vereinfachter Weise durch

$$\hat{s} \approx \frac{d(V)}{6}
 \tag{B-59}$$

angenähert werden.

B.4 Grundlagen des Fuzzy-Modells

Nicht immer lassen sich strikte und klare Definitionen von Begriffen oder Gegenständen geben. Ist auf dieser Basis zu entscheiden, ob ein bestimmtes Objekt durch eine solche Definition zu charakterisieren ist, so kann oftmals nicht eindeutig mit Ja oder Nein geantwortet werden. Definitionen sind häufig vage oder unscharf. Der Begriff eines „alten Menschen“ ist ein Beispiel dafür. Ab welchem Alter soll ein Mensch so bezeichnet werden? Im Bereich räumlicher Daten könnte beispielsweise zu klären sein, ob ein bestimmter Punkt noch zum Wald gehört oder nicht. Man neigt bei manchen Zuständen eher dazu den Begriff zu bejahen (z.B. bei einem 80-jährigen Menschen bzw. bei einem Punkt mitten im Wald) und bei manchen eher zu verneinen (z.B. bei einem 50-jährigen Menschen bzw. bei einem Punkt am Waldrand). Um nun auch solche unscharfen Definitionen in Verarbeitungsprozesse einbeziehen zu können, wurde von Zadeh (1965) ein Konzept entwickelt, das unscharfe Mengen als neue mathematische Objekte einführt. Das Konzept ist als unscharfe Mengentheorie (Fuzzy-Subset-Theorie) bekannt. Die im Umfeld der Theorie einzuordnenden Funktionalitäten werden häufig als Fuzzy-Methoden bezeichnet (Grosche et al., 1995). Aber nicht immer, wenn der Begriff „Fuzzy“ benutzt wird, kann stillschweigend die unscharfe Mengentheorie als Basis angenommen werden. Das in der Arbeit definierte Fuzzy-Modell legt jedoch diese Theorie zu Grunde.

Zu Beginn wird der klassische Mengenbegriff betrachtet. Eine Menge A besteht aus einer Reihe an Elementen x , die aufgrund bestimmter Eigenschaften der Menge zugerechnet werden. Ein Element ist dabei entweder ganz der Menge angehörig oder gar nicht. Man schreibt dafür:

$$x \in A \text{ oder } x \notin A. \tag{B-60}$$

Alle beliebigen Elemente entstammen einem definierten Grundbereich, der sich als Universalmenge E bezeichnen lässt. Ein Beispiel dafür wäre die Menge der reellen Zahlen \mathfrak{R} . Will man die Zugehörigkeit eines Elementes zu einer Menge durch eine charakteristische Funktion ausdrücken, dann hat die Funktion folgendes Aussehen:

$$\mathbf{m}_A(x) = \begin{cases} 1 & \forall x \in A, \\ 0 & \forall x \notin A. \end{cases} \tag{B-61}$$

Diese strikte Zuordnung der Elemente zu einer Menge wird im Konzept der unscharfen Menge aufgehoben. Alle Elemente können nun einen abgestuften Grad der Zugehörigkeit besitzen. Die gewöhnliche (oder scharfe) Menge wird um weitere Elemente mit teilweiser Zugehörigkeit erweitert. Dadurch entsteht eine unscharfe Menge. Jedem Element ist dabei sein Zugehörigkeitswert beizufügen. Eine unscharfe Menge \underline{A} kann wie folgt beschrieben werden:

$$\text{unscharfe Menge: } \underline{A} = \{ (x / \mathbf{m}_A(x)) \} \quad \forall x \in E \tag{B-62}$$

Die Menge besteht aus geordneten Paaren, die jeweils sowohl das Element als auch dessen Zugehörigkeit umfassen. Der Grad der Zugehörigkeit wird durch eine charakteristische Funktion ausgedrückt, die folgendermaßen definiert ist:

$$\text{Zugehörigkeitsgrad: } \mathbf{m}_A : x \rightarrow [0,1] \tag{B-63}$$

Die Zugehörigkeit zeigt an, inwieweit ein bestimmtes Element mit den definierten Eigenschaften der unscharfen Menge übereinstimmt. Ein hoher Wert weist auf eine starke bis volle Übereinstimmung hin, ein niedriger Wert auf eine geringe bis keine Übereinstimmung. Beispielsweise wird ein Ort inmitten eines Waldgebietes stärker mit der Modelldefinition von Wald zusammenpassen (daher hoher Zugehörigkeitswert) als ein Ort am Rand im Übergangsbereich zur angrenzenden Feldlage (daher niedriger Zugehörigkeitswert). Theoretisch sind zwar beliebige Intervallgrenzen für den Grad der Zugehörigkeit denkbar, solange ein beschränktes Intervall erhalten bleibt. Doch gehen die weiteren Betrachtungen von der Festlegung in (B-63) aus.

Es bestehen verschiedene Formen, wie eine unscharfe Menge beschrieben werden kann (Kaufmann, 1975). Zu den direkten Formen gehören:

- Elementaufzählung: $\underline{A} = \{ (x_1 / \mathbf{m}_A(x_1)), (x_2 / \mathbf{m}_A(x_2)), \dots \}, \tag{B-64}$

- Summenform: $\underline{A} = x_1 / \mathbf{m}_A(x_1) + x_2 / \mathbf{m}_A(x_2) + \dots = \sum_{i=1}^n x_i / \mathbf{m}_A(x_i), \tag{B-65}$

- Wertetabelle: $\underline{A} :$

x_1	x_2	...	x_n
$\mathbf{m}_A(x_1)$	$\mathbf{m}_A(x_2)$...	$\mathbf{m}_A(x_n)$

Diese Formen sind nur auf diskrete Grundbereiche sinnvoll anzuwenden. Eine andere Möglichkeit ergibt sich über die Angabe der Zugehörigkeitsfunktion, durch die der Grad der Zugehörigkeit für jedes beliebige Element des Grundbereiches berechnet werden kann. Eine einfache lineare Funktion hat beispielsweise folgende Beschreibung:

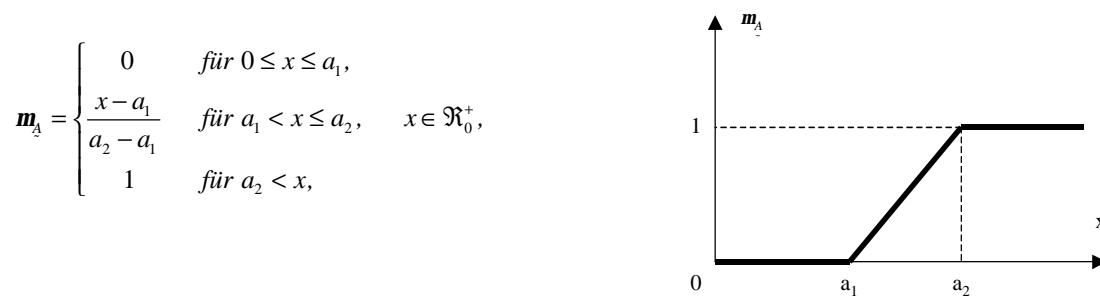


Abb. B.1: Einfache lineare Form einer Zugehörigkeitsfunktion.

Abb. B.1 verdeutlicht die Funktion. Sie eignet sich auch für stetige Grundbereiche. Die weiteren Ausführungen beziehen sich auf den stetigen Fall, besitzen aber zumeist auch Gültigkeit für diskrete Grundbereiche.

Die unscharfe Menge wird zum wesentlichen Teil durch die Zugehörigkeitsfunktion charakterisiert. Es stellt sich dann die Frage, in welcher Weise eine Zugehörigkeitsfunktion abgeleitet werden kann. Zwei Ansätze können unterschieden werden. Kennzeichen des ersten Ansatzes ist die Verwendung von Referenzdaten. Auf deren Basis wird eine möglichst angepasste Funktion bestimmt. Diese Vorgehensweise wird als SR-Modell (Similarity Relation Model) bezeichnet. Anwendung findet das SR-Modell beispielsweise bei der Klassifizierung mit der Fuzzy-k-Means-Methode (Bezdek et al., 1984). Die Festlegung im zweiten Ansatz erfolgt dagegen auf Basis von Expertenwissen und ist deshalb unabhängig vom Vorliegen von Daten. Man geht davon aus, dass eine gute Kenntnis der Situation besteht und im Voraus eine passende Funktion durch den Anwender festgelegt werden kann. Man nennt diese Art der Festlegung SI-Modell (Semantic Import Model). Burrough (1989) gibt ein Beispiel für die Anwendung dieses Modells.

Unschärfe Mengen können über verschiedene Kenngrößen charakterisiert werden. Dazu gehören:

- Träger: $\text{supp}(\underline{A}) = \{x \in E \mid \mathbf{m}_A(x) \neq 0\}$, (B-66)

- Höhe: $\text{hgt}(\underline{A}) = \sup(\mathbf{m}_A(x) \mid x \in E)$, (B-67)

- Beliebiger Alpha-Schnitt: $A_a = \{x \mid \mathbf{m}_A(x) \geq a\}$, (B-68)

- Kern: $A_{a=1} = \{x \mid \mathbf{m}_A(x) = 1\}$. (B-69)

Träger, Alpha-Schnitt und Kern sind scharfe (gewöhnliche) Mengen. Gilt für die Höhe $\text{hgt}(\underline{A}) = 1$, so spricht man von einer normalen, sonst von einer subnormalen unscharfen Menge. Eine unscharfe Menge heißt konvex, wenn alle ihre Alpha-Schnitte zusammenhängende Intervalle bilden. Ein spezielles Beispiel sind reelle unscharfe Zahlen. Sie besitzen zusätzlich einen Kern, der nur aus einem einzelnen Element besteht. Alpha-Schnitte lassen sich weiterhin nutzen um eine Zerlegung der unscharfen Menge zu definieren. Es gilt das folgende Zerlegungstheorem (Kaufmann, 1975):

$$\underline{A} = \max_{a_i}(\mathbf{a}_1 \cdot A_{a_1}, \mathbf{a}_2 \cdot A_{a_2}, \dots, \mathbf{a}_n \cdot A_{a_n}) \quad \text{für } 0 \leq \mathbf{a}_i \leq 1 \text{ und } i = 1, \dots, n \quad (\text{B-70})$$

Verknüpfungen von Mengen werden über mengenalgebraische Operationen durchgeführt. Sie sind für die Anwendung auf unscharfe Mengen zu erweitern. Während die bisher bekannten scharfen Operationen nur festlegen, welche Elemente in der verknüpften Endmenge enthalten sind, muss nun zusätzlich geregelt werden, welchen Zugehörigkeitsgrad diese erhalten sollen. Dafür ergeben sich folgende Festlegungen:

- Schnitt: $\underline{C} = \underline{A} \cap \underline{B}$, $\mathbf{m}_C(x) = \min(\mathbf{m}_A(x), \mathbf{m}_B(x))$, (B-71)

- Vereinigung: $\underline{C} = \underline{A} \cup \underline{B}$, $\mathbf{m}_C(x) = \max(\mathbf{m}_A(x), \mathbf{m}_B(x))$, (B-72)

- Komplement: $\underline{B} = \overline{\underline{A}}$, $\mathbf{m}_B(x) = 1 - \mathbf{m}_A(x)$, (B-73)

- Differenz: $\underline{C} = \underline{A} - \underline{B} = \underline{A} \cap \overline{\underline{B}}, \mathbf{m}_{\underline{C}}(x) = \min(\mathbf{m}_{\underline{A}}(x), 1 - \mathbf{m}_{\underline{B}}(x)).$ (B-74)

Der resultierende Zugehörigkeitswert wird bei diesen Operationen stets nur durch einen bestimmten Operanden (\underline{A} oder \underline{B}) festgelegt. Sie werden daher als nicht-interaktive Operationen bezeichnet (Grosche et al., 1995). Dagegen bilden interaktive Operationen eine zweite Gruppe an Verknüpfungsmöglichkeiten, bei denen beide Zugehörigkeiten jeweils ihren Teil zu jedem neuen Zugehörigkeitsgrad beitragen. Folgende Operationen lassen sich hinzuzählen:

- Beschränktes Produkt: $\underline{C} = \underline{A} \otimes \underline{B}, \mathbf{m}_{\underline{C}}(x) = \max(0, \mathbf{m}_{\underline{A}}(x) + \mathbf{m}_{\underline{B}}(x) - 1),$ (B-75)

- Beschränkte Summe: $\underline{C} = \underline{A} \oplus \underline{B}, \mathbf{m}_{\underline{C}}(x) = \min(1, \mathbf{m}_{\underline{A}}(x) + \mathbf{m}_{\underline{B}}(x)),$ (B-76)

- Algebraisches Produkt: $\underline{C} = \underline{A} \cdot \underline{B}, \mathbf{m}_{\underline{C}}(x) = \mathbf{m}_{\underline{A}}(x) \cdot \mathbf{m}_{\underline{B}}(x),$ (B-77)

- Algebraische Summe: $\underline{C} = \underline{A} + \underline{B}, \mathbf{m}_{\underline{C}}(x) = \mathbf{m}_{\underline{A}}(x) + \mathbf{m}_{\underline{B}}(x) - \mathbf{m}_{\underline{A}}(x) \cdot \mathbf{m}_{\underline{B}}(x),$ (B-78)

- Gamma-Operator: $\underline{C} = \underline{A} \mathbf{g} \underline{B}, \mathbf{m}_{\underline{C}}(x) = (\mathbf{m}_{\underline{A}}(x) + \mathbf{m}_{\underline{B}}(x) - \mathbf{m}_{\underline{A}}(x) \cdot \mathbf{m}_{\underline{B}}(x))^{\mathbf{g}} \cdot (\mathbf{m}_{\underline{A}}(x) \cdot \mathbf{m}_{\underline{B}}(x))^{1-\mathbf{g}}, \mathbf{g} \in [0,1].$ (B-79)

Die Anwendung der Produkt- bzw. Summenbildungen auf scharfe Mengen liefert das identische Ergebnis wie Schnitt bzw. Vereinigung dieser Mengen. Das algebraische Produkt hat im Allgemeinen die Eigenschaft, immer kleinere Werte als die Eingangswerte zu liefern. Der Operator wirkt abschwächend. Dagegen erhält man durch Anwendung der algebraischen Summe immer größere Werte, so dass der Operator als verstärkend zu charakterisieren ist. Ein Ausgleich zwischen beiden Tendenzen kann durch die Anwendung des Gamma-Operators erreicht werden (Bonham-Carter, 1994). Der Einfluss wird über die Wahl des Wertes $\mathbf{g} \in [0,1]$ gesteuert, wobei für $\mathbf{g} = 1$ die algebraische Summe und für $\mathbf{g} = 0$ das algebraische Produkt erhalten wird. Hinzuweisen ist noch auf zusätzliche Operationen, die arithmetische Berechnungen auf unscharfen Zahlen ausführen (Grosche et al., 1995).

Die Verschiedenheit zweier unscharfer Mengen kann über die Berechnung einer Distanz ausgedrückt werden. Dazu existieren die beiden folgenden Distanzmaße (Kaufmann, 1975):

- Generalisierte Hamming-Distanz (oder lineare Distanz):

$$d(\underline{A}, \underline{B}) = \int_{-\infty}^{\infty} |\mathbf{m}_{\underline{A}}(x) - \mathbf{m}_{\underline{B}}(x)| \cdot dx = \int_{-\infty}^{\infty} (\max(\mathbf{m}_{\underline{A}}(x), \mathbf{m}_{\underline{B}}(x)) - \min(\mathbf{m}_{\underline{A}}(x), \mathbf{m}_{\underline{B}}(x))) \cdot dx, \quad (\text{B-80})$$

- Euklidische Distanz (oder quadratische Distanz):

$$e(\underline{A}, \underline{B}) = \sqrt{\int_{-\infty}^{\infty} (\mathbf{m}_{\underline{A}}(x) - \mathbf{m}_{\underline{B}}(x))^2 \cdot dx}. \quad (\text{B-81})$$

Auf der Basis der Distanzmaße kann eine Bewertung des Unschärfegrades (Grad der Fuzziness) der gesamten Menge abgeleitet werden. Ein solcher Unschärfegrad kann als Fuzzy-Index der Menge bezeichnet werden. Zur dessen Berechnung ist in einem ersten Schritt die nächstliegende scharfe Menge zu definieren. Sie kann nach folgender Anweisung gefunden werden:

$$\text{nächstliegende scharfe Menge } \underline{A}: \mathbf{m}_{\underline{A}} = \begin{cases} 0 & \forall x \in E / \mathbf{m}_{\underline{A}}(x) < 0.5 \\ 1 & \forall x \in E / \mathbf{m}_{\underline{A}}(x) \geq 0.5. \end{cases} \quad (\text{B-82})$$

Im zweiten Schritt ergibt sich der Fuzzy-Index als Distanz der unscharfen Menge zur nächstliegenden scharfen Menge. Der Hamming-Fuzzy-Index (oder lineare Fuzzy-Index) lautet demnach:

$$\mathbf{u}(\underline{A}) = \frac{2}{b-a} \int_a^b |\mathbf{m}_{\underline{A}}(x) - \mathbf{m}_{\underline{A}}(x)| \cdot dx. \quad (\text{B-83})$$

Die Menge mit der konstanten Zugehörigkeitsfunktion von $\mathbf{m}_{\underline{A}}(x) = 0.5$ besitzt die maximale Unschärfe. Hier liegt die Zugehörigkeit der Elemente genau in der Mitte zwischen komplett zugehörig und nicht zugehörig. Der Fuzzy-Index eignet sich auch zum Vergleich der Unschärfe zweier beliebiger Mengen und kann als Maß für die Variation und somit zur Beschreibung der räumlichen Unsicherheit eingesetzt werden.

Eine alternative Festlegung des Unschärfemaßes kann über die Entropie erfolgen. Sie gilt als Maß für den Grad der Ordnung der Komponenten in einem System aufgrund vorliegender Zustandswahrscheinlichkeiten. Dazu wäre eine Ableitung von Wahrscheinlichkeiten für unscharfe Ereignisse durchzuführen. Da hier aber eine klare Trennung der unscharfen Mengentheorie und der Wahrscheinlichkeitstheorie erhalten bleiben soll, wird auf eine entsprechende Definition verzichtet.

In mathematischen Anwendungen spielen Graphen und Relationen eine wichtige Rolle. Für sie können korrespondierende Konzepte im Bereich der unscharfen Mengen gefunden werden. Ein unscharfer Graph ist demnach eine spezielle unscharfe Menge, die folgenden Festlegungen genügt:

$$\text{unscharfer Graph: } \underline{G} = \{ (x, y / \underline{m}_G(x, y)) \} \quad \forall (x, y) \in E_1 \times E_2. \quad (\text{B-84})$$

Der unscharfe Graph greift damit auf zwei unterschiedliche Grundbereiche E_1 und E_2 zurück, die beliebige Formen annehmen dürfen. Sind die Grundbereiche dagegen identisch und als Menge der reellen Zahlen definiert, dann handelt es sich um eine unscharfe Relation:

$$\text{unscharfe Relation: } \underline{R} = \{ (x, y / \underline{m}_R(x, y)) \} \quad \forall (x, y) \in \mathfrak{R}^2. \quad (\text{B-85})$$

Unschärfe Relationen können dazu dienen, bedingte unscharfe Mengen zu formulieren. Die Zugehörigkeit eines Elementes y hängt in diesem Fall von einem anderen Element x ab. Die Zugehörigkeitsfunktion kann dann folgendermaßen geschrieben werden:

$$\text{bedingte Zugehörigkeit: } \underline{m}_B(y // x) \quad x \in E_1, y \in E_2. \quad (\text{B-86})$$

Aus der Kenntnis der bedingten Zugehörigkeit lässt sich die Zugehörigkeitsfunktion der unscharfen Menge \underline{B} wie folgt ermitteln:

$$\underline{m}_B(y) = \max_{x \in E_1} (\min(\underline{m}_B(y // x), \underline{m}_A(x))) \quad x \in E_1, y \in E_2. \quad (\text{B-87})$$

Gedankengänge und Schlussfolgerungen zu formalisieren, ist die Aufgabe der Logik. Normalerweise zeichnet sie sich durch eine strikte und formale Vorgehensweise aus und steht damit im deutlichen Widerspruch zu den Konzepten der Unschärfe. Obwohl Fuzzy und Logik eigentlich unvereinbar sind, hat sich der Begriff „unscharfe Logik (Fuzzy Logik)“ durchgesetzt. Dies verdeutlicht, dass im Kontrast zur Booleschen Logik eine auf der unscharfen Mengentheorie basierende Logik aufgebaut werden kann. Ihre Existenz wird gestützt durch die Tatsache, dass sich die menschliche Denkweise keineswegs an logischen Konzepten orientiert, sondern einen mehr oder weniger unscharfen Mechanismus darstellt. Möchte man nun solche Mechanismen ebenso im Computer verfügbar machen, bedarf es eines erweiterten logischen Ansatzes, der Fuzzy Logik. Analog zu Booleschen Variablen, die als Werte Null (falsch) oder Eins (wahr) besitzen können (man spricht deshalb auch von einer zweiwertigen Logik), können hier unscharfe Variablen definiert werden. Sie entsprechen den Werten der Zugehörigkeitsfunktion für ein bestimmtes Element:

$$\underline{a} := \underline{m}_A(x) \quad (\text{B-88})$$

und erhalten damit Werte im Intervall von Null bis Eins. Für den Anwendungsbereich der Aussagenlogik ergibt sich damit ein gleitender Übergang zwischen den Zuständen falsch (Null) und wahr (Eins). Man spricht in diesem Zusammenhang auch von Möglichkeitswerten, im Vergleich zum sonst üblichen Wahrscheinlichkeitsmaß der Booleschen Logik. Die Kombination der Variablen gelingt in bekannter Weise durch logische Operatoren, deren Verhalten entsprechend der unscharfen Mengentheorie zu modifizieren ist. Die wichtigsten Operatoren lauten:

- Logisches UND (Konjunktion): $\underline{a} \wedge \underline{b} = \min(\underline{a}, \underline{b}), \quad (\text{B-89})$

- Logisches ODER (Disjunktion): $\underline{a} \vee \underline{b} = \max(\underline{a}, \underline{b}), \quad (\text{B-90})$

- Logisches NICHT (Negation): $\bar{\underline{a}} = 1 - \underline{a}. \quad (\text{B-91})$

Aus beliebig vielen Variablen können mit Hilfe der Operatoren unscharfe Funktionen definiert werden. Es gilt:

$$0 \leq \underline{f}(\underline{a}, \underline{b}, \dots) \leq 1. \quad (\text{B-92})$$

Dies ist immer dann erfüllt, wenn lediglich die drei oben aufgelisteten Operatoren eingesetzt werden.

Beispiele für die Anwendung der Fuzzy Logik mit raumbezogenen Daten finden sich vor allem in Verbindung mit Standortfragen. Dabei ist die Eignungsfähigkeit des Raumes gegenüber verschiedenen Kriterien abzuwägen. Beispielsweise soll bestimmt werden, welche Bereiche sich für den Bau eines Industriekomplexes eignen. Die Verwendung scharfer Kriterien (z.B. Mindestabstand von 1000 m zur bestehenden Wohnbebauung) und deren anschließenden logi-

schen Verknüpfung liefert eindeutige Ergebnisse (geeignet oder nicht). Im Gegensatz dazu können innerhalb der Fuzzy Logik die Kriterien unscharf formuliert, bewertet und verknüpft werden, so dass am Ende die Eignung der Bereiche individuell durch Möglichkeitswerte beschrieben wird. Anwendungsbeispiele dieser Art sind aus den Bereichen der Bodenkunde (Burrough, 1989) oder der Landschaftsplanung (Mandl, 1994, Blaschke, 1997) bekannt. Abschließend ist anzumerken, dass im genannten Anwendungsbereich das Verarbeitungsmodell, definiert durch die zu formulierenden Kriterien und deren logische Verknüpfungen, als unscharf und damit als unsicher einzustufen ist. Die notwendigen räumlichen Daten werden weiterhin als sicher angesehen. Nur in einem Fall wurden Untersuchungen durchgeführt, wie sich das Ergebnis verändert, wenn zusätzlich die Daten als unsicher gelten (Heuvelink und Burrough, 1993). Zur Beschreibung der Datenunsicherheit wird dort aber ein stochastisches Modell angesetzt.

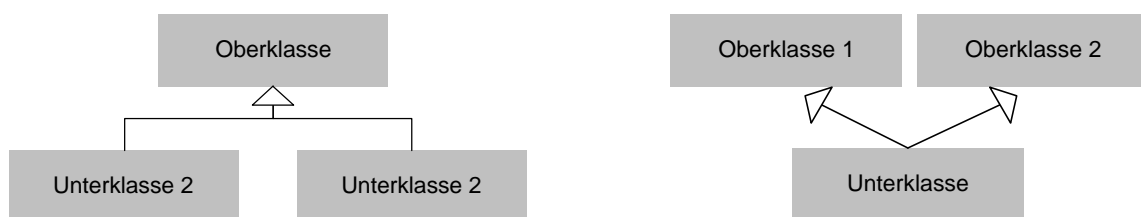
B.5 Einige wichtige Sprachkonstrukte der „Unified Modeling Language“ (UML)

Die „Unified Modeling Language“ (UML) erlaubt die objektorientierte Beschreibung von beliebig komplexen Modellen in graphischer Form. Dabei wird das Ziel verfolgt, durch eine abstrahierte visuelle Beschreibung ein Problem in seiner Komplexität aufzulösen und übersichtlich zu strukturieren, um letztendlich das Verständnis zu erleichtern. Die Notation soll in gleichem Maße die Analyse als auch das Design von Modellen unterstützen (Quatrani, 1998). UML steht vorläufig an der Spitze der Entwicklung verschiedener graphischer Modellierungssprachen (z.B. OMT (Rumbaugh et al., 1991), Booch (Booch, 1991), OOSE (Jacobson, 1992)), wobei die Definition versucht, die bekannten Konzepte zu integrieren, um so deren Vorteile auszunutzen und eine Vereinheitlichung herbeizuführen. Inzwischen ist die Notation ein Standard der unabhängigen „Object Management Group“ (OMG) (OMG, 2000). Die Sprache verfügt über eine Vielzahl an Konstrukten, von denen nachfolgend die wichtigsten herausgegriffen und kurz erläutert werden. Eine vollständige Beschreibung ist beispielsweise in Rational (2000) nachzulesen.

Zu den grundlegenden objektorientierten Eigenschaften zählt die Fähigkeit gleichartige Objekte zusammenzufassen und in Form von Klassen allgemeingültig zu beschreiben. Eine Klasse definiert Methoden, Attribute und Beziehungen zu anderen Objekten. Alle Objekte einer Klasse folgen dieser Definition. Abstrakte Klassen und abgeleitete Klassen sind Sonderformen. Eine Klasse ist dann abstrakt, wenn keine Objektinstanzen (reale Objekte) erzeugt werden können. Solche Klassen enthalten nur eine unvollständige Beschreibung des Sachverhalts, mit dem Ziel, die Inhalte auf Unterklassen zu vererben. Es liegt dann im Aufgabenbereich der Unterklassen, die Beschreibung zu konkretisieren und zu vervollständigen. Abgeleitete Klassen zeichnen sich dadurch aus, dass ihre Inhalte (Methoden und Attribute) vollständig aus anderen Klassen hergeleitet werden können. Der Zusammenhang ist der Klassendefinition beizufügen. Abgeleitete Klassen dienen allein der verbesserten Strukturierung verteilter Inhalte, um so für Verarbeitungszwecke Erleichterungen zu erbringen. In UML werden Klassen folgendermaßen dargestellt:



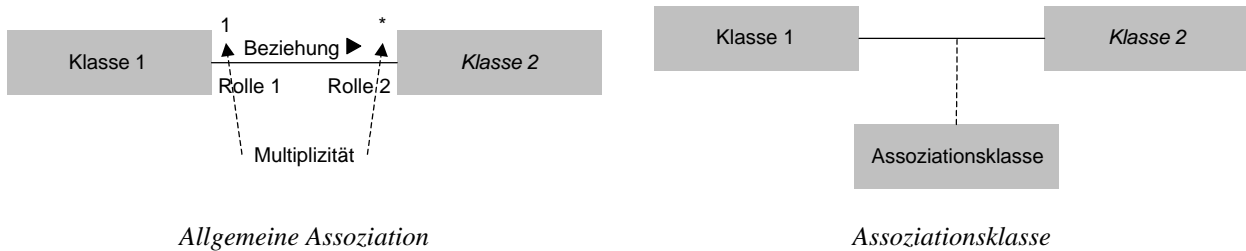
Zu den wichtigen objektorientierten Eigenschaften zählt weiterhin die Vererbung. Sie definiert die Übertragung von Methoden und Attributen von einer Oberklasse auf eine oder mehrere Unterklassen, die dann die übertragenen Eigenschaften wie eigene Inhalte benutzen können. In diesem Sinne stellt die Unterklasse eine Spezialisierung der Oberklasse dar und beschreibt eine „... ist eine ...“-Relation. Es ist möglich, dass eine Unterklasse von beliebig vielen Oberklassen erbt, genauso wie eine Oberklasse ihre Eigenschaften auf beliebig viele Unterklassen vererben kann. Diese spezielle Eigenschaft wird als Mehrfachvererbung bezeichnet. Die Vererbung hat in UML folgendes Aussehen:



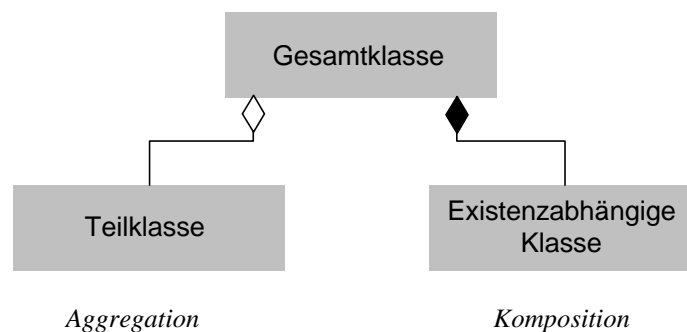
Vererbung

Zu beachten ist, dass die Vererbungsrichtung entgegen der Pfeilrichtung des Vererbungssymbols in UML von der Oberklasse zur Unterklasse wirkt, d.h. die Unterklasse erbt Methoden und Attribute von der Oberklasse.

Allgemeine Beziehungen (Relationen) zwischen Objekten werden als Assoziationen bezeichnet. In einem Modell können prinzipiell beliebig viele Beziehungen auftreten. Entsprechend vielgestaltig sind ihre möglichen Erscheinungsformen. Richtungen der Beziehungen (unidirektional, bidirektional) können ebenso von Bedeutung sein, wie die Anzahl (Multiplizität) der an einer Beziehung beteiligten Objekte. Der Sinn der Beziehung wird durch die Angabe der Rolle, die ein Objekt in der Beziehung einnimmt, und durch den Beziehungsnamen verdeutlicht. Besonders komplexe Beziehungen lassen sich in Form einer eigenen Klasse (Assoziationsklasse) zusammenfassen, wodurch eine vereinfachte Strukturierung erreicht werden kann. Die UML-Darstellung sieht dafür wie folgt aus:



Zwei häufig verwendete und daher speziell ausgewiesene Assoziationen sind Aggregation und Komposition. Mit ihnen lässt sich ausdrücken, aus welchen Bestandteilen sich eine Klasse zusammensetzt. Mögliche Bestandteile bilden wiederum eine oder mehrere Klassen (Teilklassen), die in einer „... Teil von ...“-Beziehung zur zusammensetzenden Klasse (Gesamtklasse) stehen. Während bei Aggregationen die Bestandteile unabhängig von der aggregierten Klasse existieren, können bei Kompositionen die Bestandteile nur für die Lebenszeit des Gesamtobjekts bestehen. Mit dem Gesamtobjekt verschwinden auch die zugehörigen existenzabhängigen Teile. In UML werden die beiden Beziehungen wie folgt dargestellt:



C Ergänzungen zu den Anwendungen der Unsicherheit

C.1 Übergangsmodelle zwischen zwei Linien

Im Normalfall baut sich ein Objekt aus mehreren Linien zusammen. Die Linien können die gleiche oder aber unterschiedliche geometrische Unsicherheiten besitzen. Durch den in der Arbeit verfolgten linienweisen Ansatz zur Berücksichtigung der Unsicherheit erhält ein Punkt, in dem zwei Linien enden, eine doppelte Rolle zugewiesen. Der Punkt gehört beiden Linien in gleichem Maße an und trägt deshalb die Unsicherheitsmaße beider Linien, die zusammen die Unsicherheit des Punktes ausdrücken. Diese punktweise Unsicherheit kann beispielsweise durch eine Kovarianzmatrix (Kapitel 2.4.1 in (2-4)) beschrieben werden.

Die doppelte Rolle solcher Punkte verursacht Probleme bei der Berechnung des Aussagemaßes für die Zugehörigkeit einer beliebigen Position zum Objekt (Kapitel 2.4.1 – 2.4.3). Ist der Punkt an der Bestimmung der Distanz von Objekt zur untersuchenden Position beteiligt, so ist zu entscheiden, welche der Unsicherheiten der beiden inzidierenden Linien

zur Berechnung des Aussagemaßes verwendet werden soll. Die Positionen, bei denen dieses Problem auftritt, bilden einen zusammenhängenden Sektor (Abb. C.1). Für diesen Sektor gilt es eine Lösung festzulegen.

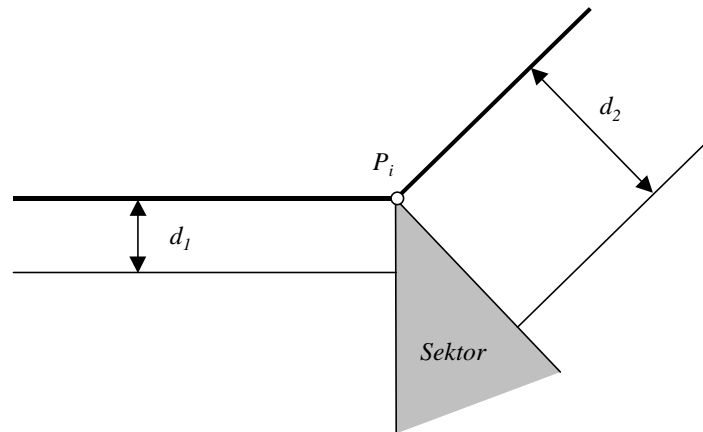


Abb. C.1: Problematischer Sektor für die Bestimmung des Aussagemaßes bei zusammengesetzten Linien mit unterschiedlichen Unsicherheiten.

Zwei Modelle können dazu definiert werden. Das erste Modell sieht einen rechnerischen Übergang zwischen den Unsicherheiten der beiden Linien vor. Ansatz ist eine lineare Veränderung der Unsicherheitsmaße in Abhängigkeit vom Sektorenwinkel. Die Vorgehensweise kann an Distanzen d_1 und d_2 von gleichwertigen Isolinien verdeutlicht werden (Abb. C.1). In den Distanzen spiegeln sich die Größenverhältnisse der Unsicherheit von beiden betroffenen Linien wider. Die Distanz vom Punkt P einer ausgewählten Isolinie ändert mit dem Teilsektorenwinkel α_i wie folgt:

$$d_i = d_1 + (d_2 - d_1) \cdot \frac{\alpha_i}{\alpha} \tag{C-1}$$

mit α als bekannter Winkel des Gesamtsektors (Abb. C.2).

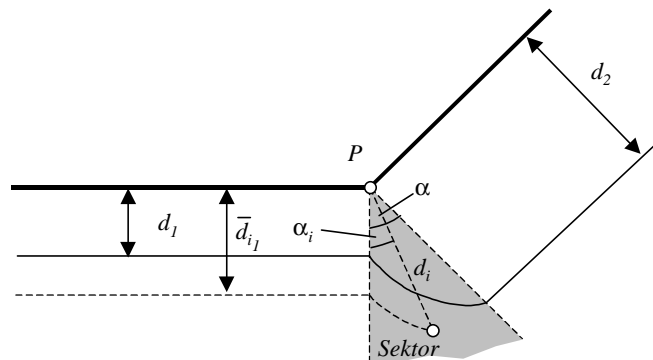


Abb. C.2: Lösung des Übergangsproblems im Spiralmodell.

Für eine beliebige Position mit Distanz \bar{d}_i vom Punkt P lässt sich aus (C-1) die zugehörige Distanz der Isolinie im Bezug auf die Unsicherheitsmaße der ersten Linie ausdrücken. Sie lautet:

$$\bar{d}_i = \bar{d}_i - (d_2 - d_1) \cdot \frac{\bar{\alpha}_i}{\alpha} \tag{C-2}$$

mit $\bar{\alpha}_i$ als Sektorenwinkel für die Richtung vom Punkt P zur untersuchenden Position. Auf den aus (C-2) berechneten Distanzwert können die Unsicherheitsmaße der ersten Linie angewendet und auf diese Weise das Aussagemaß bestimmt werden. Im Sektor nimmt eine Isolinie die Form einer archimedischen Spirale an (Abb. C.2), die in Polarkoordinaten (r, \mathbf{j}) die folgende Funktionsgleichung besitzt:

$$\mathbf{r} = \frac{d_2 - d_1}{\mathbf{a}} \mathbf{j} \quad \text{für } 0 \leq \mathbf{j} < \infty. \quad (\text{C-3})$$

Aus diesem Grund kann dieser Lösungsansatz auch als Spiralmodell bezeichnet werden. Die Kurve besitzt innerhalb des Sektors Ähnlichkeit mit einer Ellipse, die insbesondere dann als Isolinie entsteht, wenn das Problem geometrisch durch eine Abböschung (Graf und Barner, 1978) gelöst wird.

Das zweite Modell teilt den Sektor in zwei Bereiche, in denen jeweils nur die Unsicherheit einer Linie in die Berechnung einfließt. Die Teilung erfolgt gewichtet gemäß der Größe der Unsicherheit der Linien. Die entstehende Teilungslinie entspricht einer Skelettlinie (Kapitel 3.3.2) mit Ursprung in Punkt P (Abb. C.3). Zur Bestimmung der Teilungslinie ist von einem modifizierten Sektor auszugehen, der sich aus den über den Punkt hinaus verlängerten Linien ergibt. Im Unterschied zu anderen Bereichen wird die benötigte Distanz in diesem Modell auf Basis der verlängerten Linien bestimmt (Abb. C.3). Dadurch lässt sich erreichen, dass Positionen auf der Teilungslinie ein einheitliches Ausmaß erhalten, unabhängig davon, mit welchen der beiden beteiligten Unsicherheitswerten die Berechnung durchgeführt wird. Die Teilungslinie kann zusätzlich auf geometrische Weise gewonnen werden, indem man gleichwertige Isolinien der Punktzugehörigkeiten (Kapitel 3.3.2) miteinander schneidet. Alle Schnittpunkte definieren dann die gesuchte Teilungslinie. Diese Festlegung entspricht der Vorgehensweise bei einer Dachausmittlung (Graf und Barner, 1978), bei der ansteigende Flächen unterschiedlicher Neigungen (bedingt durch unterschiedliche Unsicherheiten) geschnitten und so Firste, Grate und Kehlen konstruiert werden. Aus diesem Grund lässt sich dieses Verfahren auch als Dachmodell bezeichnen. Eine andere Möglichkeit der Konstruktion der Teilungslinie bietet die rechnerische Bestimmung der Teilsektorenwinkel \mathbf{b}_1 und \mathbf{b}_2 . Sie können wiederum über die bekannten Distanzen d_1 und d_2 von gleichwertigen Isolinien der beiden Linien abgeleitet werden. Ansatz bildet das folgende Verhältnis:

$$\frac{\mathbf{b}_1}{\mathbf{b}_2} = \frac{d_2}{d_1}. \quad (\text{C-4})$$

Aus (C-4) ergeben sich die gesuchten Teilsektorenwinkel zu:

$$\begin{aligned} \mathbf{b}_1 &= \frac{d_2}{d_1 + d_2} \cdot \mathbf{b} \\ \mathbf{b}_2 &= \frac{d_1}{d_1 + d_2} \cdot \mathbf{b} \end{aligned} \quad (\text{C-5})$$

mit \mathbf{b} als bekannter Winkel des erweiterten Gesamtsektors.

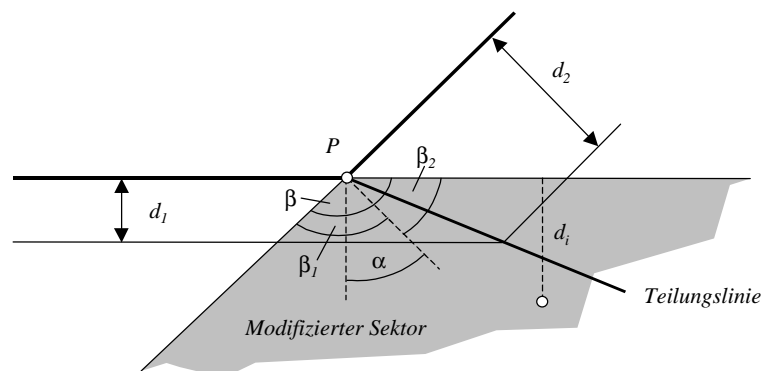


Abb. C.3: Bestimmung der Teilungslinie im Dachmodell.

Das Dachmodell hat den Vorteil, dass wegen der vorab zu bestimmenden Sektoraufteilung die Berechnung des Ausmaßes für Positionen innerhalb des Sektors vereinfacht wird. Weiterhin müssen zu Darstellungszwecken nur Geradenstücke konstruiert werden. Die entstehende Form jeder Isolinie behält die ursprüngliche Form der Linien bei. Dies kann aber insbesondere bei spitzen Winkeln der Linien zu ausladenden Figuren führen. Das Spiralmodell erzeugt dagegen einen kontinuierlichen Übergang zwischen den unterschiedlichen Unsicherheiten. Nur an den beiden Sektorgrenzen treten Unstetigkeitsstellen in der Form auf. Um diese zu beseitigen, bietet sich als Alternative die Möglichkeit an, Klothoiden als Übergangsbögen einzusetzen. Beispiele für die Anwendung der beiden Modelle sind in verschiedenen Abbildungen innerhalb der Arbeit zu finden. Das Spiralmodell wird z.B. in Abb. 3.30 benutzt, während die Anwendung des Dachmodells z.B. in Abb. 3.18 zu sehen ist.

C.2 Grundlagen zur Generierung von Simulationsdatensätzen am Beispiel einer Linie

Simulationen werden häufig dann eingesetzt, wenn das direkte Beobachten des Verhaltens eines Systems nicht möglich oder nur aufwendig zu realisieren ist. Deshalb wird im Computer mit Hilfe von Modellen ein Ersatzsystem erzeugt, an dem sich die Untersuchungen ausführen lassen. Als spezielles Verfahren kommt die Monte-Carlo-Methode zum Einsatz, wenn zufallsabhängige Systeme vorliegen. Dabei werden Eingangsdaten auf Basis der Eigenschaften der beteiligten Zufallsvariablen erzeugt. Sie entsprechen einer statistischen Stichprobe, die sich aus einer großen Anzahl an einzelnen Realisierungen der Zufallsvariablen zusammensetzt. Diese Daten werden unter Nutzung des Anwendungsmodells prozessiert und ergeben wiederum eine Stichprobe des Ergebnisses, die weitergehend zu analysieren ist.

Die Generierung einer solchen Stichprobe soll am Beispiel einer Objektlinie diskutiert werden. Wie alle GIS-Daten ist auch eine Linie zu einem bestimmten Teil geometrisch mit Unsicherheit behaftet. Die Größe ihrer Unsicherheit wird als bekannt vorausgesetzt. Sie soll im Stochastischen Modell in Form einer Linienvarianz \mathbf{s}_L vorliegen. Aufgabe ist es nun, Realisierungen der Linienpunkte zu generieren, die in ihrer Gesamtheit die Variation der Linie ergeben. Gemäß ihrer Definition wirkt die Variation in einem Linienpunkt stets senkrecht zur Linienrichtung in diesem Punkt. Daher wird für alle nachfolgenden Formeln ein lokales Koordinatensystem angenommen, dessen Abszissenachse in Richtung der Linie und die Ordinatenachse senkrecht dazu in Richtung der Variation definiert sind. Statt die im Allgemeinen unendliche Anzahl an Linienpunkten zu behandeln, erfolgt eine Beschränkung auf eine bestimmte endliche Anzahl an Stützpunkten (Abb. C.4).

Zur Generierung der Stichprobe bestehen verschiedene Möglichkeiten. Entscheidender Einflussfaktor bildet die Berücksichtigung der räumlichen Korrelation, die zwischen den Linienpunkten wirkt. Drei verschiedene Fälle können unterschieden werden (Hunter und Goodchild, 1997):

1. Fall: räumliche Unabhängigkeit der Linienpunkte ($r=0$)

Jeder Linienpunkt wird unabhängig von allen anderen betrachtet, d.h. der Korrelationskoeffizient hat den Wert $r=0$. Generell entsteht eine Realisierung, indem man zu den beobachteten Ordinatenwerten y' der Linienpunkte zugehörige Werte des Störungsfeldes e addiert:

$$y = y' + e. \quad (\text{C-6})$$

Im Fall der räumlichen Unabhängigkeit wird das Störungsfeld durch einen Vektor mit unabhängigen normalverteilten Zufallsvariablen $N(0, \mathbf{s}_L)$ definiert:

$$\mathbf{e} = N(0, \mathbf{s}_L). \quad (\text{C-7})$$

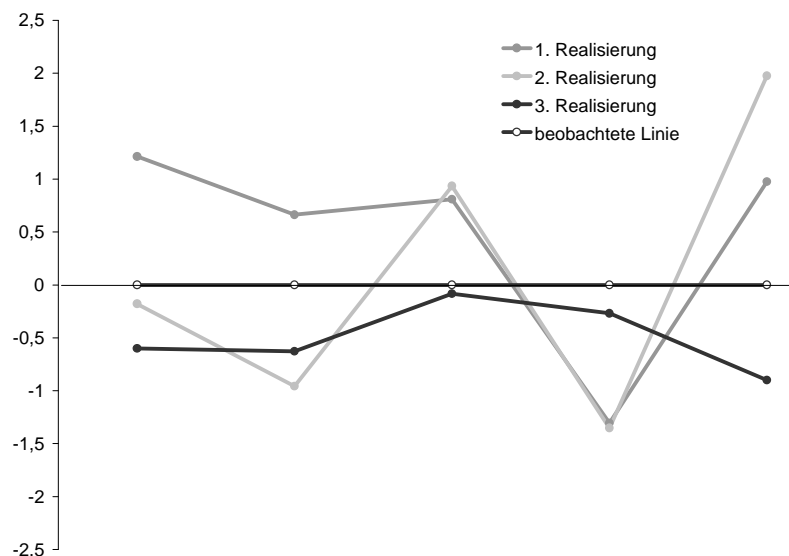


Abb. C.4: Drei Realisierungen einer Beispiellinie mit räumlicher Unabhängigkeit der Linienpunkte.

Das Störungsfeld ordnet jedem beobachteten Ordinatenwert unabhängig einen eigenen Zufallswert zu. Abb. C.4 zeigt drei auf der Basis von (C-6) und (C-7) erzeugte Realisierungen für eine Beispiellinie mit $\mathbf{s}_L = 1$. Die Ordinatenwerte der Linie wurden alle identisch zu Null beobachtet, so dass die beobachtete Beispiellinie einem Teil der Abzissenachse entspricht. Insgesamt besteht die Linie aus fünf Linienpunkten.

2. Fall: vollständige räumliche Abhängigkeit ($r=1$)

In diesem Fall sind die Linienpunkte vollständig funktional voneinander abhängig. Die Abhängigkeit wirkt wie ein systematischer Einfluss, d.h. innerhalb einer Realisierung ändern sich alle Linienpunkte immer in die gleiche Richtung und um den gleichen Betrag. Zur Erzeugung einer Realisierung wird daher nur ein einziger Zufallswert $N(0, \mathbf{s}_L)$ benötigt. Das Störungsfeld enthält innerhalb einer Realisierung nur konstante Werte:

$$\mathbf{e} = [N(0, \mathbf{s}_L) \ N(0, \mathbf{s}_L) \ \dots] = \mathbf{const}. \quad (\text{C-8})$$

Die Realisierungen für die Beispiellinie aus Fall 1 gleichen Parallelen zur beobachteten Linie, so dass an dieser Stelle auf eine Abbildung verzichtet werden kann.

3. Fall: teilweise räumliche Abhängigkeit ($0 < r < 1$)

Der Fall der teilweisen räumlichen Abhängigkeit siedelt sich genau zwischen den beiden vorher behandelten Extremen der Unabhängigkeit und der vollständigen Abhängigkeit an. Er bildet ein Szenario, das in der Realität am häufigsten anzutreffen sein wird. Als Ansatz wird hier das Störungsfeld als ein autoregressiver Prozess definiert (Goodchild et al., 1992b). Es gilt der folgende Zusammenhang:

$$\mathbf{e} = \mathbf{rW}\mathbf{e} + N(0, \mathbf{s}_L). \quad (\text{C-9})$$

Die Matrix \mathbf{W} enthält Gewichte, die die Beziehung zwischen den Linienpunkten beschreiben. Ein Element wird zu Eins gesetzt, wenn die entsprechenden Punkte benachbart sind. Ansonsten erhalten die Elemente den Wert Null. Der Parameter \mathbf{r} beschreibt den Grad der räumlichen Autokorrelation, d.h. den Grad der gegenseitigen Abhängigkeit zwischen benachbarten Linienpunkten. Er kann im Bereich $0 < \mathbf{r} < 0.5$ variieren. Ein Wert nahe 0.5 bedeutet eine hohe, ein Wert nahe Null eine geringe Autokorrelation. Um aus (C-9) das Störungsfeld zu berechnen, kann einerseits ein iterativer Lösungsansatz benutzt werden (Heuvelink, 1992). Andererseits ergibt eine Auflösung von (C-9) nach \mathbf{e} folgendes:

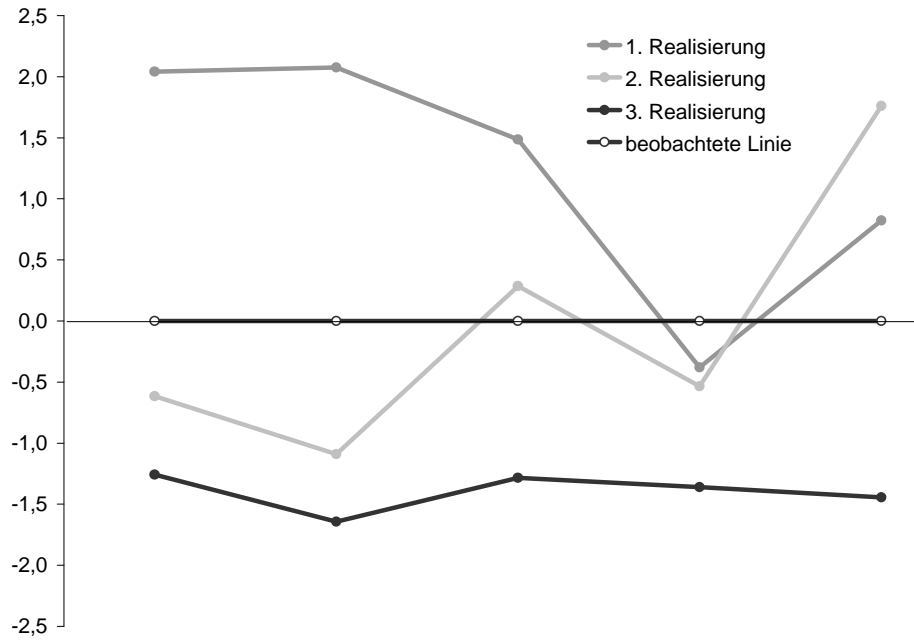
$$\mathbf{e} = (\mathbf{I} - \mathbf{rW})^{-1} \cdot N(0, \mathbf{s}_L). \quad (\text{C-10})$$

Der notwendige Berechnungsaufwand wird maßgeblich von der Inversion der Matrix $\mathbf{I} - \mathbf{rW}$ bestimmt. Er hängt entscheidend von der Dimension dieser Matrix ab. Die Dimension wird durch die Anzahl der beteiligten Linienpunkte festgelegt. Abb. C.5 zeigt jeweils drei Realisierungen der bekannten Beispiellinie mit den identischen Zufallswerten wie im Fall 1. Für Abb. C.5a gilt ein Autokorrelationskoeffizient von $\mathbf{r} = 0.4$ und für Abb. C.5b ein Wert von $\mathbf{r} = 0.49$.

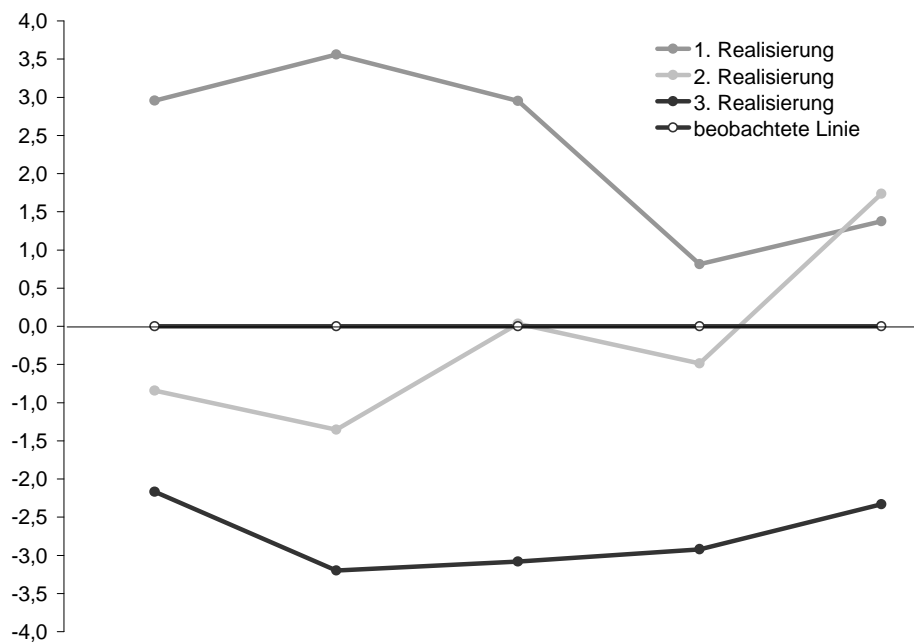
Die Beispiele zeigen, dass mit Zunahme der Autokorrelation eine Glättung der Linie erfolgt. Zur Generierung einer Realisierung ist der Autokorrelationskoeffizient in angemessener Weise vom Anwender festzulegen. Beispielsweise kann eine Schätzung über die Auswertung einer erfassten Stichprobe durchgeführt werden. Ein Ansatz dazu sind die Mittelung von lokalen Koeffizienten, die auf der Basis von drei benachbarten Linienpunkten berechnet werden können. Der Zusammenhang definiert sich in folgender Weise:

$$\mathbf{r}_i = \left| \frac{v_i}{v_{i+1} + v_{i-1}} \right|, \quad (\text{C-11})$$

mit den Werten v als Abweichungen der beobachteten Ordinatenwerte vom Mittel ($v_i = y_i - \mathbf{m}_y$). Generell ist festzustellen, dass für diskrete, künstliche Objekte (z.B. Straßen, Häuser) der Koeffizient stets nahe bei 0.5 liegen wird, da dort häufiger eine glatte Form auftritt. Bei natürlichen Objekten (z.B. Waldgrenzen) ergibt sich dagegen ein Koeffizient nahe Null. Die Größe des Wertes wird maßgeblich durch den vorliegenden Punktabstand beeinflusst. Geht dieser gegen Null, so ergibt sich automatisch ein hoher Wert. Um Simulationen schnell durchzuführen, ist es sinnvoll, den Autokorrelationskoeffizienten bei der Erfassung des Objektes mitzuschätzen und als zusätzlichen Parameter der Unsicherheitsbeschreibung beizufügen. Die bestehenden Unsicherheitsmodelle (Kapitel 2.5) sind dahingehend zu erweitern.



(a)



(b)

Abb. C.5: Zwei Beispiele für die Generierung von je drei Realisierungen mit unterschiedlichen Autokorrelationskoeffizienten mit $r = 0.4$ (a) und mit $r = 0.49$ (b).

Dank

Es ist mir ein besonderes Anliegen an dieser Stelle all denen zu danken, die zum Gelingen der Arbeit beigetragen haben. In erster Linie gilt der Dank dem Hauptreferenten Prof. Dr.-Ing. habil. Dieter Fritsch, der die Arbeit von Beginn an gefördert und sorgsam betreut hat. Weiter möchte ich Prof. Dr.-Ing. Ralf Bill und Prof. Dr. mult. Erik W. Grafarend für die Übernahme des Korreferats danken und für die auf diesem Wege erfahrene Unterstützung.

Im besonderen Maße hilfreich waren die zahlreichen intensiven und kritischen Diskussionen mit den Kolleginnen und Kollegen der GIS-Gruppe am Institut für Photogrammetrie (IfP) der Universität Stuttgart. Spezieller Dank gebührt Dr.-Ing. Monika Sester und Dr.-Ing. Ulrike Klein, die durch Korrekturlesen der Rohfassungen wertvolle Anregungen und Verbesserungsvorschläge gegeben haben. Positiv auf die Arbeit hat sich insbesondere auch das sehr freundschaftliche Arbeitsklima am IfP ausgewirkt. Dafür möchte ich allen jetzigen und ehemaligen Kolleginnen und Kollegen recht herzlich danken.

Des Weiteren sind die Beiträge der Studentinnen und Studenten zu würdigen, die sich intensiv mit Themen aus dem Umfeld der Arbeit im Rahmen von Diplom- und Studienarbeiten auseinandergesetzt haben. Dadurch erhielt ich die Gelegenheit, mich zusätzlichen Aspekten zu widmen. Meinen Dank möchte ich im Speziellen auch an Frau cand. geod. Sandra Waller richten, die als Hilfswissenschaftlerin über mehrere Jahre wertvolle Unterstützungstätigkeiten leistete.

Der größte Dank gebührt aber meiner Frau, die mich zu allen Zeiten unterstützte und geduldig in der Fertigstellung der Arbeit bestärkte. Ihr ist diese Arbeit gewidmet.

Abschließend möchte ich noch der Universität Stuttgart als meinem Arbeitgeber danken, die mir durch besondere vertragliche Regelungen die Vorbereitung der Promotion ermöglichte.

Stuttgart, im Mai 2000

Michael Glemser

Lebenslauf

Michael Glemser

Geboren am 19. Mai 1964 in Löffingen (Kreis Breisgau-Hochschwarzwald)

Verheiratet mit Imke Glemser seit 1994

Schulausbildung	1970 – 1974	Wilhelmsschule, Stuttgart- Wangen
	1974 – 1983	Wirtemberg-Gymnasium, Stuttgart-Untertürkheim
	1983	Abitur
Studium	1983 – 1989	Studium der Geodäsie an der Universität Stuttgart
	1989	Abschluss mit Diplom im April
Beruf	1987 – 1990	Wissenschaftliche Hilfskraft am Institut für Photogrammetrie der Universität Stuttgart
	1990	Softwareingenieur bei der Fa. Inpho GmbH, Stuttgart
	1990 – heute	Wissenschaftlicher Mitarbeiter am Institut für Photogrammetrie der Universität Stuttgart