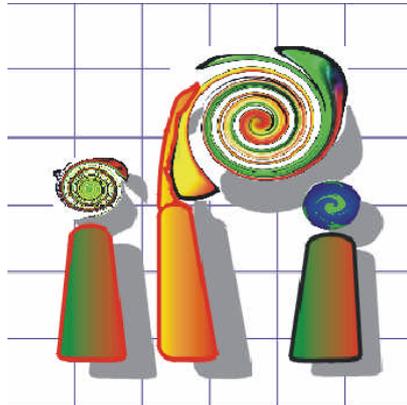


Institut für Photogrammetrie und Ingenieurvermessungen



UNIVERSITÄT HANNOVER

Diplomarbeit

**Modellierung der geometrischen Genauigkeit als
Qualitätsaspekt von Geodaten**

Zohar Kara

Hannover, im Februar 2001

1. Einführung:

1.1 Einleitung und Zielstellung

Der Einsatz von Geoinformationssystemen (im folgenden als GIS bezeichnet) zur digitalen Bearbeitung räumlicher Fragestellungen nimmt mit dem Einsatz von EDV zu und ist zu einem alltäglichen Vorgang geworden.

Die Gewinnung von Basisdaten für GIS hat sich mit der Entwicklung der vermessungstechnischen und der photogrammetrischen Geräte stark verbessert. Es werden heutzutage mit der tachymetrischen Geländeaufnahme oder mit Hilfe des Global Positioning Systems (GPS) große Mengen von Daten gewonnen, die direkt in einem GIS integriert werden können. Sowohl geometrische, als auch thematische Daten können durch Methoden der Fernerkundung gewonnen werden. Grundlage hierfür sind Satelliten- oder Luftbilder.

Der Erfassungsprozess von Objekten der realen Welt bis zur Verwaltung der Daten ist mit gewissen Unsicherheiten aufgrund systematischer, grober und zufälliger Fehler behaftet. Da ein GIS nicht nur zur Speicherung, sondern zur Weiterverarbeitung (Analyse) der Geodaten (Daten eines Geoinformationssystem) dient, werden sich solche Unsicherheiten fortpflanzen und führen dazu, dass die vom GIS ermittelten Folgedaten auch mit Unsicherheiten behaftet sind. Somit sind sowohl Genauigkeitsaussagen in GIS einzuführen, als auch deren Fortpflanzungsverhalten bei Datenanalysen zu berücksichtigen. Solche Aussagen sollten als Entscheidungskriterien zur Datenqualität dienen, denn: „ *Ist ein GIS nicht in der Lage, Genauigkeitsaussagen zu erfassen und zu verarbeiten, so kann es auch keine sicheren und vertrauenswürdigen Ergebnisse liefern*“ [Bill & P.K., 1998a, S.247].

In letzter Zeit wird viel über Datenqualität diskutiert. Es gibt eine ganze Reihe von internationalen Standards, die sich mit Fragen der Qualitätssicherung beschäftigen, z.B. DIN bzw. ISO 9000 – 9004 [DIN, 1995]. Trotz solcher Standards haben sämtliche GIS-Vorhaben gegen das Prinzip der Qualitätssicherung verstoßen, da erst mit zunehmender Verfügbarkeit der Daten über deren Qualitätssicherung diskutiert wird und nicht vor deren Entstehung.

Hinsichtlich der Qualität werden folgende Aspekte (Kriterien) genannt: Herkunft der Daten, Positions- und Attributgenauigkeit, Konsistenz, Vollständigkeit und Aktualität. Diese Kriterien u.a.m. werden in Kapitel drei weiter beschrieben.

Ein GIS bildet die reale Welt in einem Schichtenmodell verschiedener Datenebenen ab. Die raumbezogenen GIS- Daten beschreiben die Objekte über ihre Position (Geometrie), ihre Attribute (Sachdaten) und ihre räumliche Verknüpfung (Topologie). Die Erfahrungen in den letzten Jahren haben gezeigt, daß die beiden Aspekte Geometrie und Thematik sehr schwer voneinander zu trennen sind. Obwohl die Genauigkeitsfrage der Datenqualität von den beiden Aspekten stark abhängig ist, soll diese Arbeit keinen speziellen Beitrag zur Thematik liefern. Der Schwerpunkt dieser Arbeit ist die Beschreibung und Erläuterung der in der Literatur vorhandenen Modelle zur Untersuchung von geometrischer Genauigkeit und ihre Integration in GIS- Funktionen.

1.2 Aufbau der Arbeit

Die vorliegende Arbeit beschreibt einige in der Literatur vorhandene Modelle zur Beurteilung der geometrischen Genauigkeit als Qualitätsmerkmale in einem GIS, wie das sogenannte Minimum- Maximum- Modell, Fehlerband und das stochastische Modell.

In Kapitel 2 werden einige Definitionen im GIS- Bereich dargestellt, die vor allem für diese Arbeit wichtig sind, wie z. B. Datenerfassungsmethoden , Vektor- und Rasterdatenstrukturen, hybrides GIS. Weiterhin werden in Kapitel 3 Begriffe wie Qualität, Qualitätsmerkmale, Qualitätskriterien und Qualitätsanforderungen bezüglich der Geodaten näher diskutiert. Außerdem wird eine allgemeine Beschreibung zu der thematischen und geometrischen Genauigkeit (Unsicherheiten) gegeben.

In Kapitel 4 werden die theoretischen Grundlagen der Statistik erläutert, die zur Anwendung des gewählten stochastischen Modells dienen. Der Schwerpunkt liegt auf dem Ansatz von Wahrscheinlichkeitstheorie. Dabei werden einige Begriffe wie Zufallsexperimente, zufällige Variable, bedingte Wahrscheinlichkeit, Dichte- und Normalverteilungsfunktion näher erläutert.

In Kapitel 5 werden drei verschiedene Modelle zur Modellierung und Beurteilung der geometrischen Genauigkeit, wie in der Literatur vorhanden, kurz beschrieben und anhand von Beispielen diskutiert. Anschließend wird in Kapitel 6 durch Anwendung des stochastischen Modells in hybriden GIS geprüft, ob die Frage nach einer Identität zweier Objekten A und B, die in zwei verschiedenen Datensätze mit unterschiedlicher Herkunft und Genauigkeit gegeben sind, lösbar ist. Dieser Sachverhalt wird ausführlich an einem Beispiel beschrieben werden. Anschließend werden die Ergebnisse zusammengestellt und diskutiert und die Problematik solcher Anwendungen gezeigt.

Die Arbeit schließt mit einer Zusammenfassung ab.

2. Geodaten

Unter dem Begriff „Geodaten“ versteht man die Daten eines Geoinformationssystems (GIS). Die Vorsilbe „Geo“ leitet sich aus der griechischen Bezeichnung für „Erde“ ab, und ist charakteristisch für den deutschen Sprachraum. Im Englischen sagt man „Geographical Data“. Die Datenelemente, die auf die Erdoberfläche oder einen Teil der Erdkruste bzw. auch der Erdatmosphäre bezogen sind, und in einem GIS verwaltet werden, werden als Geodaten bezeichnet. Man unterscheidet bei den Geodaten zwischen geometrischen und thematischen Daten, die in unterschiedlichen Datenstrukturen in einem GIS vorliegen können. In diesem Kapitel werden einige Begriffe der GIS-Welt definiert.

2.1 Definitionen von Geoinformationssystemen

Mittlerweile sind Geographische Informationssysteme (GIS) weit verbreitet, die über die Möglichkeit verfügen, Sach- und Geometriedaten in ihren komplexen, logisch-inhaltlichen und räumlichen Zusammenhängen zu erfassen, zu verwalten und außerdem über räumliche Analysemöglichkeiten neue Informationen zu generieren. Für den Begriff GIS gibt es in der Literatur zahlreiche Definitionen [www1, GIS] und [Lothar, 1999b], die sehr allgemein gehalten und sich in der Regel auf die Forderung der Möglichkeit zur Erfassung, Verwaltung, Analyse und Ausgabe raumbezogener Daten beschränken, beispielweise:

1. *„ Ein Geoinformationssystem dient der Erfassung, Speicherung, Analyse und Darstellung aller Daten, die einen Teil der Erdoberfläche und der darauf befindlichen technischen und administrativen Einrichtungen sowie geowissenschaftliche, ökonomische und ökologische Gegebenheiten beschreiben.“* [Bartelme, 1995, S.12]
2. *„ Ein Geoinformationssystem ist ein rechnergestütztes System, das aus Hardware, Software, Daten und den Anwendungen besteht. Mit ihm können raumbezogene Daten erfaßt und redigiert, gespeichert und reorganisiert, modelliert und analysiert sowie alphanumerisch und graphisch präsentiert werden.“* [Bill & D.F., 1994, S. 5]

3. *„Ein Geoinformationssystem ist ein Informationssystem, das alle raumbezogenen Daten der Atmosphäre, der Erdoberfläche und der Lithosphäre enthält und eine systematische Erfassung, Aktualisierung, Verarbeitung und Umsetzung dieser Daten auf der Grundlage eines einheitlichen räumlichen Bezugssystemes gestattet.“* [Göpfert, 1991, S.3]
4. *„Sie strecken sich auf Daten aus allen Geo- Bereiche (Geodaten), die zur digitalen Auswertung und zur analogen Wiedergabe in Karten aller Maßstäbe geeignet“* oder *„Nach dem heutigen Sprachgebrauch bezieht sich die Bezeichnung GIS jedoch auf den Fall, daß unter Einsatz digitaler Technologien objektorientierte Modelle der Umwelt erzeugt und daraus Geoinformationen abgeleitet und dargestellt werden“* [Hake & D.G., 1994, S. 22 und 498f].

Diese vier Aspekte: Erfassung, Speicherung (Verwaltung), Analyse und Darstellung (Präsentation) der Daten sollten im einzelnen beachtet werden, um ein passendes Modell zur Beschreibung der Genauigkeit definieren zu können. Bei der Verwaltung und Analyse kommen insbesondere EDV- technische Aspekte zum Tragen z. B. Datenstrukturierung, Programmierung u.a.m.. Zu den Einflußgrößen aus der Präsentation zählen die gerätespezifischen Aspekte wie Auflösung und Präzision des Ausgabegerätes. Der größte Einfluß auf die Genauigkeit geht auf die Erfassungskomponenten zurück. Diese ergeben die Basis an Genauigkeit, die den anderen Komponenten zugrunde liegt. Deswegen wird diese Komponente im folgenden weiter beschrieben werden.

2.2 Erfassung von Daten

Die Daten bilden den Kern eines GIS, sie werden im Computer von zwei unterschiedlichen Datentypen genutzt: Rasterdaten und Vektordaten, um die Objekte der realen Welt in einem GIS verwalten zu können.

Die Erfassung von raumbezogenen Daten und ihre Nachbearbeitung, bis diese digital vorliegen, ist die entscheidende Basis für Anwendung und Erfolg eines GIS. Bei den Daten unterscheidet man zwischen geometrischen und thematischen Hauptklassen, so müssen die Erfassungsmethoden entsprechend gewählt werden, um die entsprechenden Objektinformationen für eine GIS- Anwendung zu

erhalten, z. B. werden tachymetrische Aufnahmen für Lageinformationen angewendet, während u.a. Klassifizierungen Objektattribute liefern. Dabei sollte die Datenerfassung zum einen so genau und vollständig wie notwendig, zum anderen so wirtschaftlich wie möglich erfolgen. Bartelme unterscheidet die Geodaten in vier Varianten [Bartelme,1995]:

- Rohdaten: Direkt registrierte, über Meßgeräte, Sensoren oder auf photogrammetrische Wege erfaßte Daten.
- Interpretierte Daten: Aufgrund eines Hintergrundwissens klassifizierte, mit einer Bedeutung belegte und ergänzte Daten.
- Symbolisierte Daten: Aufgrund kartographischer Konventionen erzeugte Symbole, Signaturen und Schriften.
- Strukturierte Daten: Aufgrund eines Hintergrundwissens und einer Anwendungsanforderung zu vielfältigen Strukturen zusammengefaßte Daten.

Es wird bezüglich der Herkunft der Daten zwischen zwei Erfassungsmethoden unterschieden, nämlich originäre und sekundäre Erfassungsmethoden [Bill & D.F, 1994], die im folgenden beschrieben werden.

2.2.1 Originäre Erfassungsmethoden

Originäre bzw. primäre Erfassungsmethoden (Abb. 2-1) sind solche Methoden, wie Vermessung, Photogrammetrie und Fernerkundung, bei denen die Primärdaten durch Messungen am Objekt oder dessen Abbild erfaßt werden.

Bei der Vermessung werden geometrische Beschreibungen von Liegenschaft- und topographischen Objekten geliefert. Bei Methoden und Anwendungen unterscheidet man zwischen drei Hauptbereichen der Vermessung: Erdmessung, Landesvermessung und Detailmessung. [Bill & D.F., 1994]. Während der Zusammenhang zwischen den ersten zwei Bereichen mit GIS darin besteht, daß die Erdmessung das fundamentale Bezugssystem für GIS liefert, und die GIS-Anwendung auf einem Landeskoordinatensystem, wie Gauß- Krüger-Koordinaten, basiert, sind die Methoden der Detailvermessung die direkten Datenlieferanten für ein GIS.

Unter Detailvermessung versteht man die Geländeaufnahme (Feldmessung), mit vermessungstechnischen Geräten, zur Lagebestimmung oder zur kombinierten

Lage- und Höhenbestimmung. Es wird beispielsweise mit dem Verfahren der Tachymetrie ein Objekt der realen Welt durch kombinierte Winkel- und Streckenmessung punktweise erfaßt und mit dem anschließenden Berechnungsverfahren die dreidimensionalen Koordinaten abgeleitet. Wenn die Geländeaufnahme mit einem registrierenden elektronischen Tachymeter durchgeführt wird, welche heute sehr entwickelt sind, können die gewonnenen Daten direkt in einem GIS übernommen werden.

Zur Gewinnung von Daten ist noch der Ansatz des GPS zu erwähnen, das in letzter Zeit vermehrt zum Einsatz kommt, weil u.a. die Erfassung von thematischen und geometrischen Daten mit GPS- Vermessung schnell vorgenommen werden kann und ausreichend genau ist.

Die Photogrammetrie und Fernerkundung als flächenhafte Aufnahmemethoden sind von besonderer Bedeutung für GIS. Bei der Photogrammetrie als indirekte Meßmethode, werden die geometrischen Daten nicht durch direkte Messungen am Objekt, sondern anhand photographischer Abbildungen des Objektes gewonnen. Mit verschiedenen photogrammetrischen Luftbildinterpretationsgeräten lassen sich für ein GIS beschreibende Daten wie Landnutzung etc. aus dem Luftbild entnehmen. Die Ergebnisse der Interpretationen werden entweder direkt im Bild vermerkt oder auf eine parallel mitgeführte Karte übertragen und dann im GIS digitalisiert. Weiterführendes im nächsten Abschnitt dieses Kapitels.

Die Fernerkundung liefert multispektrale digitale Daten, die zur Interpretation und Klassifizierung genutzt werden, und im GIS als Attribute zu den raumbezogenen Daten eingesetzt werden können.

Es kommen noch andere Erfassungsmethoden aus verschiedenen Fachdisziplinen zur Erhebung von thematischen Informationen mit Raumbezug vor. Diese Methoden zählen auch zu den primären Erfassungsmethoden, wie Erhebungen durch Befragungen oder Messungen von Temperatur, Luftdruck etc.. Solche Methoden werden in dieser Arbeit nicht behandelt.

2.2.2 Sekundäre Erfassungsmethoden

Die sekundären Erfassungsmethoden (Abb. 2-1) spielen eine große Rolle im GIS. Die primären Daten, die auf Kartenwerken, Luft- und Satellitenbildern vorliegen, werden mit den verschiedenen Digitalisierungsmethoden (manuell oder

automatisch) erfasst. Diese erfaßten Daten (sekundär) zählen zum größeren Teil der Daten, die in einem GIS eingespeist sind.

Bei der Digitalisierung unterscheidet man zwischen manueller und automatischer Digitalisierung. Die am meisten verbreitete Methode zur Erfassung von Vektordaten aus analogen Vorlagen ist die manuelle Digitalisierung. Dabei werden mit Hilfe eines Digitalisiertisches (Digitizer) die Objekte punktweise oder in benutzerdefinierbaren, kontinuierlichen Weg- oder Zeitintervallen erfaßt. Die am Digitizer registrierten lokalen Tischkoordinaten werden in das Koordinatensystem der Vorlage umgerechnet. Bei der automatischen Digitalisierung werden die analogen Vorlagen durch Abtasten (Scannen) in Rasterform gewandelt, und nach der sogenannten Raster-Vektor-Konvertierung wird die Objektgeometrie abgeleitet. Mehr dazu in Kapitel 6.2.

Primäre Erfassungsmethoden

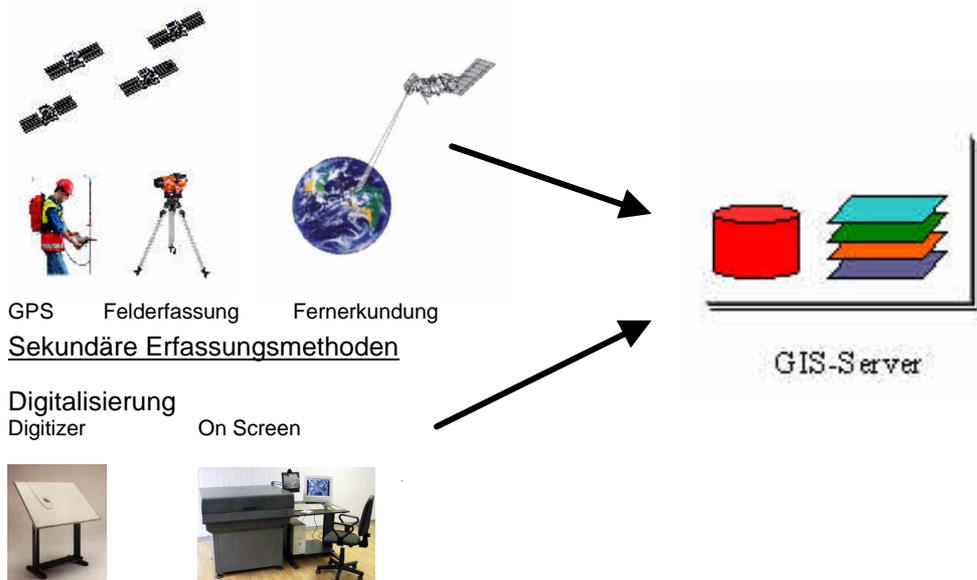


Abb. 2-1: Primäre und sekundäre Erfassungsmethoden

Man bezeichnet ein GIS als vektororientiert, wenn seine Fachdaten sich aus der Kombination von Vektor-, Grafik- und zugehörigen Sachdaten zusammensetzen. Demgegenüber ergibt sich ein rein rasterorientiertes GIS, wenn seine Fachdaten ausschließlich aus Raster-, Grafik- und entsprechenden Sachdaten bestehen. Ein hybrides GIS ist dann die Vereinigung von den beiden obengenannten Systemen. Im folgendem werden die beiden Datenarten beschrieben.

2.3 Datenstrukturen

Bevor die Datenstrukturen, innerhalb der verschiedenen Datenmodelle, kurz beschrieben werden, wird eine schematische Darstellung der drei Datenmodelle dargestellt: Raster-, Vektor- und hybrides Datenmodell. Der Abbildung 2-2 kann man folgendes entnehmen:

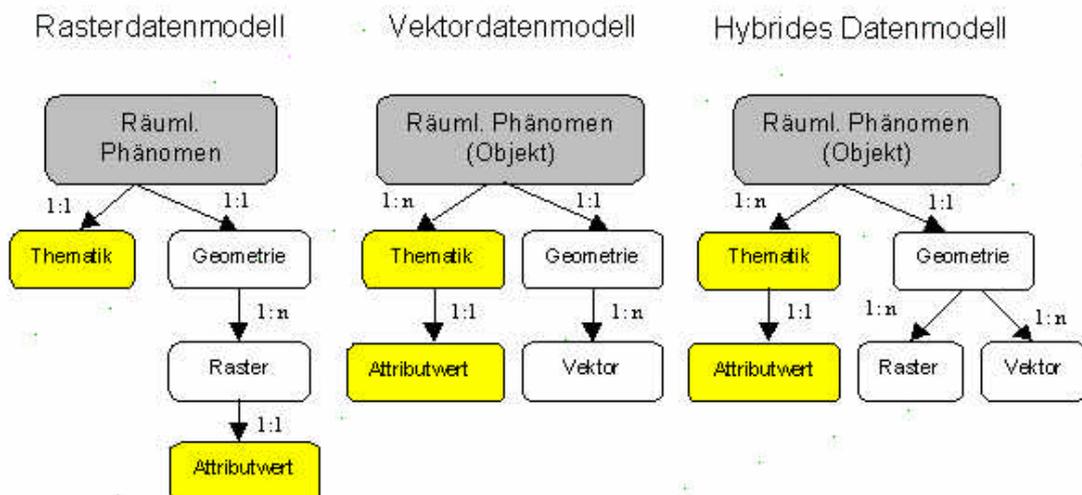


Abb. 2-2: Gegenüberstellung des Raster-, Vektor- und hybriden Datenmodells nach [Glemser & U.K.,1999d]

Ein hybrides Datenmodell ergibt sich aus der Vereinigung der beiden Modelle für Raster- und Vektordaten. Schwierigkeiten treten bei der Vereinigung aufgrund der folgenden Unterschiede auf:

1. Das Rastermodell verbindet Geometrie und Thematik durch die Zuweisung genau eines Attributwertes zu jeder Rasterzelle 1:1-Relation. Zelle und Wert definieren dabei eine untrennbare Einheit. Im Gegensatz dazu sind im Vektormodell Thematik und Geometrie über den Begriff des Objektes gekoppelt. Eine direkte Verbindung wie im Rastermodell besteht nicht mehr.
2. Im Rastermodell kann im allgemeinen nur ein Attributwert jeder Rasterzelle beigefügt werden. 1:1-Relation, dagegen erlaubt dem Vektormodell die Beschreibung beliebig vieler Thematiken zu jedem Objekt 1:n-Relation [Glemser & U.K., 1999d].

Die in gelb hervorgehobenen Kästen (Thematik und Attributwerte) werden, wie schon erwähnt, nicht behandelt. Nachfolgend werden die Datenstrukturen erläutert.

2.3.1 Vektordaten

Das Grundelement für Vektordaten ist der Punkt. Mit Hilfe von Punkten lassen sich höhere Strukturen (Linien, Flächen und Körper) definieren.

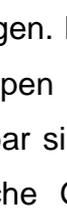
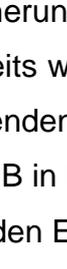
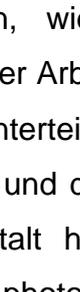
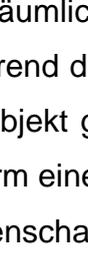
	Punkt	Linie	Fläche
Analytische Objekte			
Regellose Objekte			

Abb2-3: Vektordaten

Außerdem lassen sich aus den Koordinaten der Punkte sämtliche geometrischen Aussagen ableiten, wie Abstand zwischen Figuren oder die Länge von Verbindungen. In der Arbeit von Bethge [Bethge, 1997] werden die Vektordaten in zwei Gruppen unterteilt, die eine Gruppe beinhaltet Objekte, die analytisch beschreibbar sind, und die andere Gruppe beinhaltet Objekte, die eine regellose geometrische Gestalt haben. Die Gewinnung von Vektordaten erfolgt durch geodätische oder photogrammetrische Aufnahme- und Berechnungsverfahren, Digitalisierung von analogen Karteninformationen sowie Konstruktionen am graphischen Arbeitsplatz [Scheu, 1996].

Die Speicherung räumlicher Objekte mittels Vektordaten gibt die Möglichkeit, die Daten bereits während der Erfassung zu strukturieren, indem bestimmte zu einem entsprechenden Objekt gehörende Punkte in geeigneter Weise zusammengefaßt werden, z. B in Form einer Punktfolge für linienförmige Objekte.

Die folgenden Eigenschaften gelten für Vektordaten [Bill & D.F., 1994]

- Punkt und Linie als graphische Grundstrukturen, Fläche als geschlossener Linienzug
- Daten nach Objektlinien geordnet, dadurch linienhafte Betrachtungsweise
- logische Datenstrukturierung und Objektbezug leicht möglich

- punktuelle Datenerfassung durch den Einsatz von bewährten Methoden, jedoch hohe Erfassungszeiten
- geringe Datenmengen

2.3.2 Rasterdaten

Rasterdarstellung bezieht sich auf Flächen statt auf Punkte bzw. Linien. Das Grundelement ist das Pixel (Rasterzelle, Rastermasche), welches zeilen- und spaltenweise in einer Matrix angeordnet ist.

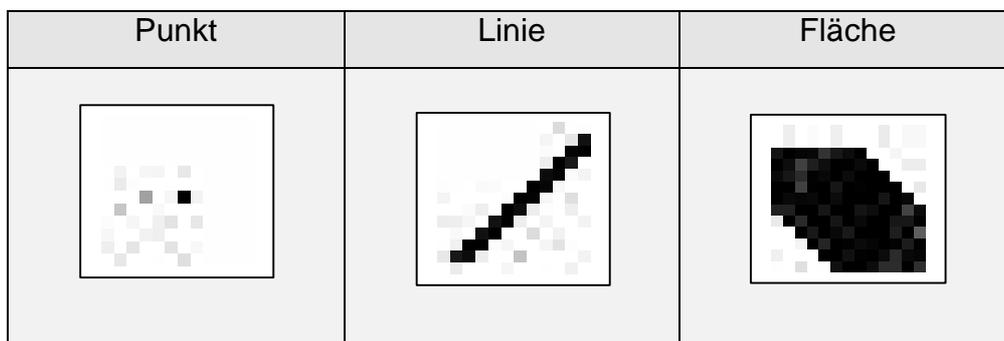


Abb.2-4: Rasterdaten

Rasterdaten erhalten Werte über Eigenschaften der Pixel wie Grau- oder Farbwerte. Dies sind nicht nur visuell erfassbare Größen, sondern sie stehen für eine inhaltliche Aussage. Die Gewinnung von Rasterdaten erfolgt durch Scannen von analogen Vorlagen (Luftbild, Orthophoto, Karten u.ä.) oder durch Aufnahme der Erdoberfläche mit Sensoren oder Kameras, die von verschiedenen Satelliten oder Flugzeugen getragen werden. Optische Sensoren liefern z. B. Objektinformationen, wie Grauton oder Farbe. Durch Klassifizierungsalgorithmen in Rasterbildern lassen sich Farbwerte einer bestimmten Objektklasse zuordnen. Pixel mit ähnlicher multispektraler Information werden so zu einer Gruppe zusammengefaßt, wie Wälder, Gewässer, Siedlungen, Straßen, Wiesen, etc., wobei die Bildinformation sich nicht auf das sichtbare Licht beschränken muß, sondern auch Information aller anderen Wellenlänge enthalten kann (z.B. Infrarot, Radar, usw.). Eine detaillierte Darstellung der Methoden der Klassifizierung wird

hier nicht beabsichtigt. Für eine Einführung in dieses umfangreiche Gebiet sei auf [Haberäcker, 1991], [Albertz, 1991] und [www2,Tutorial] verwiesen.

Die Rasterdaten haben u.a. folgende Eigenschaften:

- Pixel als graphische Grundstruktur, die gleich groß sind und einer regelmäßigen Anordnung unterliegen
- flächenhafte Betrachtungsweise, dadurch Vorzüge in diesem Bereich
- Ordnung nur nach der Position der Pixel
- einfache Datenerfassung, kurze Erfassungszeiten
- evtl. große Datenmengen, dadurch hoher Rechenaufwand

Die beiden Typen basieren auf einem unterschiedlichen Datenmodell, das entsprechend Vektormodell oder Rastermodell genannt werden kann. Angesichts der Vor- und Nachteile, die beide Strategien aufweisen, kann man weder der einen noch der anderen Methode grundsätzlich den Vorzug geben; jede kann in dem ihr gemäßen Umfeld optimal operieren, und aus diesem Grunde betonte Barteme [in: Bartelme, 1995, S. 100], daß es kein entweder- oder gibt, sondern vielmehr ein Nebeneinander, wo die beiden Strategien einander ergänzen. Diese Beschreibung führt zum Einsatz eines hybriden Modells, in dem die Vorteile beider Modelle kombiniert werden können. Dabei liegt die Stärke von Rastermodellen in ihrem flächenhaften Aspekt und in der Einfachheit ihrer Geometrie, und im Gegensatz dazu zeichnen sich Vektormodelle durch die hohe Genauigkeit aus, die sie ermöglichen, sowie durch eine beliebig steigerungsfähige Komplexität in den Beziehungen der einzelnen Elemente [Bartelme, 95]. Im folgenden wird das hybride GIS beschrieben.

2.3.3 Hybrides GIS

In einem hybriden Datenmodell können die Vorteile beider Modelle kombiniert werden, wie es oben erwähnt wurde. Die gemeinsame Verarbeitung von Vektor- und Rasterdaten in einem System führte zu dem allgemeinen Begriff des hybriden GIS. Unter diesem Begriff lassen sich mehrere Strategien einordnen, die zu drei Hauptkategorien zusammengefaßt werden können [Bartelme, 1995] und [Glemser & U.K., 1999a aus: (Ehlers et al.,1989)]:

1. Überlagerung: Es handelt sich dabei um ein Hinterlegen von Vektordaten mit Rasterbildern, allerdings bleibt die Verwaltung jedoch getrennt (siehe Abb.6-1 in Kapitel 6). Diese Strategie kommt in einem umfassenden GIS vor, weil sowohl Themen, die sich besser für eine Modellierung nach dem Vektorprinzip eignen, als auch solche, die mit dem Rasterprinzip besser modelliert werden können, vorhanden sein können. Eine Anwendungsmöglichkeit ergibt sich durch Hinterlegung von Straßennetz mit Satellitenbildern. Dieses Prinzip der gleichzeitigen Verwendung von Vektor- und Rasterbildern kann im Detail noch verschiedenen Ausprägungen haben [in: Bartelme, 1995, S.113]:
 - simultane Bereitstellung von Vektor- und Rasterbildern am Bildschirm
 - deckungsgleiches Übereinanderblenden
 - teilweise, dem Bedarf angepaßte Rasterisierung von Vektordaten oder umgekehrt

Der zweite Fall ist in dieser Arbeit von Bedeutung, weil das deckungsgleiche Übereinanderblenden eine Variante darstellt, in dem man das Vektorbild über das Rasterbild geschoben werden kann, ohne dieses zu verdecken (Kapitel 6).

2. Integration: Es lässt sich die eine systemfremde Datenart durch Konversion in die eigene Datenart umwandeln und so kann eine gemeinsame Verwaltung von Vektor- und Rasterdaten durchgeführt werden. Somit können beispielsweise Vektordaten zum Zwecke der Analyse in einem Raster- GIS über den Weg einer Vektor-Raster-Konversion in Rasterdaten konvertiert werden.
3. Verzahnung: Diese Stufe der Integration wird erreicht, wenn das im System angewandte Datenmodell selbst hybrid ist. Dazu sind die beiden bisher getrennten Datenmodelle für Vektor- und Rasterdaten zu vereinigen. Bartelme

[in: Bartelme, 1995] betonte, daß ein solches Modell im groben eine Rasterstrategie verfolgt, im Detail hingegen hat es Vektoraspekte. Bartelme gibt folgendes Beispiel zur Entstehung eines hybriden Modells dieser Art: Eine Teilung eines Interessengebiets in gleich große Rasterquadrate und Aufschneiden aller Elemente entlang der Rastermaschen.

Nur wenn die dritte Integrationsstufe in einem System realisiert wurde, kann die Integration als vollständig bezeichnet werden. Die darin enthaltenen Funktionen arbeiten unabhängig von der Datenart und erlauben somit die Ausnutzung der Vorteile beider Datenmodelle.

In dieser Arbeit ist die erste und die zweite Strategie von Interesse. Bei der ersten Stufe werden die untersuchten Vektordaten (linienförmige Objekte) über Satellitenbilder geschoben, und bei der zweiten Stufe werden die Vektordaten im Interesse der Anwendung in Rasterdaten umgewandelt.

3. Qualität von Geodaten

Das Einsatzspektrum der Geodaten hängt wesentlich von ihrer Qualität ab, die durch geeignete Maßnahmen bei der Datenerfassung und Datenverarbeitung zu sichern ist.

Der Begriff der Qualität von bestimmten Dingen hat in einer umgangssprachlichen Verwendung keinen festen Inhalt. In DIN (Deutsches Institut für Normen), EN (European de Normalisation) und ISO (International Standardization Organization) 8402 (1995) wird die Qualität so definiert [www3, DIN-CEN-ISO] und [DIN, 1995]:

„Qualität ist die Gesamtheit von Merkmalen einer Einheit, bezüglich ihrer Eignung festgelegte und vorausgesetzte Erfordernisse zu erfüllen.“

Sie bezieht sich auf objektiv beschreibbare oder meßbare Eigenschaften eines Produktes oder einer Dienstleistung, sie bedeutet dabei nicht die genaueste und perfektteste Lösung zu erzielen, sondern die geforderten Merkmale einer Einheit so genau wie nötig zu erfüllen.

In den Normen DIN ISO 9000 bis 9004 wird festgelegt, wie Qualität gesichert und festgelegt werden kann. Dabei ist ISO 9000, Teil 3 für Software und ISO 9004, Teil 2 für Dienstleistung zuständig. Diese Normen können keine Qualitätsstandards geben oder ein Qualitätsniveau definieren, daß für technische Bewertungen anwendbar ist, aber sie bilden die nötige Basis für den Aufbau von Qualitätsmanagementssystemen (im Folgenden QMS), das im nächsten Abschnitt kurz erläutert wird.

3.1 Qualitätsmanagementssystemen

Qualitätsmanagement (QM) im Sinne der genannten Normen setzt mindestens folgende Maßnahmen voraus nach [Caspary, 1992b und Grünreich, 1996]:

1. Die Merkmale des Produktes sind vom Auftraggeber genau vorzugeben, im Idealfall wirken dabei Auftraggeber , Käufer und Nutzer zusammen.
2. In den Herstellungsprozeß werden streng festgelegte und formalisierte Qualitätskontrollen integriert. Die Ergebnisse der Kontrollen werden in einem Protokoll festgehalten.
3. Auftraggeber, Käufer und Kunde erstellen einen umfassenden Katalog der Erfordernisse , denen das Produkt mit seinen Merkmalen entsprechen soll.
4. Der Hersteller unterwirft sich einer Überprüfung hinsichtlich der Maßnahmen und Erfolge des Qualitätsmanagements durch eine zur Zertifizierung autorisierte Institution.
5. Der Grad der Übereinstimmung von Merkmalen und Erfordernissen wird vom Käufer, im Ideal Fall gemeinsam mit dem Hersteller, ermittelt.

QM umfaßt alle Tätigkeiten der Gesamtführungsaufgaben, welche die Qualitätspolitik, Ziele und Verantwortungen festlegen, sowie diese durch Mittel der Qualitätsplanung, -sicherung und -verbesserung im Rahmen des QMS verwirklichen. Dabei bezeichnet Qualitätssicherung alle geplanten und systematischen Fähigkeiten, die notwendig sind, um ein angemessenes Vertrauen zu schaffen, so daß ein Produkt oder eine Dienstleistung die gegebenen Qualitätsanforderungen erfüllt wird [Caspary, 1993] und [Bill, 1997].

Es hat sich gezeigt, daß durch die Einführung von QMS Kosten gespart werden, die für den Aufbau einer GIS- Datenbank ca. 80% der Gesamtkosten eines GIS ausmachen, und mögliche Image- und Marktverluste für Anbieter bei der Auslieferung fehlerhafte Produkte vermieden [Lothar, 1999a].

Wenn das Konzept des QMS auf GIS in Zusammenhang mit der massenhaften Sammlung und Verwaltung von Geodaten angewandt werden soll, fällt auf, daß die vorausgesetzte Hersteller- Benutzer - Beziehung in aller Regel nicht gegeben ist. In folgende werden die Qualitätsmerkmale bezüglich der Geodaten beschrieben, und wird ein Vergleich zwischen die verschiedenen Autoren diskutiert.

3.2 Überblick über die Qualitätsmerkmale der Geodaten

Es ist für GIS- Daten erforderlich, objektive Kriterien zur Beurteilung ihrer Qualität anzugeben, d.h. die Merkmale (Eigenschaften, Beziehungen) von GIS-Objekten sind zusätzlich mit Qualitätsmerkmalen zu versehen, die deren Qualität beschreiben. So kann z.B. für ein Flurstück die Qualität seiner Flächenangabe durch die Angabe der Standardabweichung beschrieben werden. Mittels Vergleich der Standardabweichung mit den vorgegebenen Fehlergrenzen lässt sich dann prüfen, ob die geforderte Qualität erreicht wurde. Die Qualitätsmerkmale sollen dem Benutzer von GIS- Daten hinreichende Auskunft darüber geben, in wieweit die Daten für seine Aufgaben überhaupt geeignet sind [Lother, 1999a].

In der Literatur werden meist fünf bis sechs Aspekten als Qualitätsmerkmale angesprochen, nämlich *Herkunft der Daten*, *Positionsgenauigkeit*, *Attributgenauigkeit*, *Logische Konsistenz*, *Vollständigkeit* und *Aktualität* und darüber sind sich alle Verfasser einig [Caspar, 1993]. Caspar teilt diese wichtigsten Qualitätsmerkmalen in zwei Gruppen (Caspar, 92a). Siehe folgende Tabelle:

Qualitätsmerkmale	
Geometriedaten	Themadaten
Genauigkeit von Position und Geometrie	Attributauswahl, Objektartenreichtum
Fortführung, Aktualität	Richtigkeit der Attributzuweisung und –
Vollständigkeit	Vollständigkeit der Attribute
Auflösung, Generalisierungsgrad	Aktualität

Tab. 3-1: Qualitätsmerkmale nach [Caspar, 1992a]

Ein Teil der o.a. Merkmale kann aus dem Bereich des Vermessungswesen, andere können aus der Photogrammetrie, der Fernerkundung, der Kartographie, der Geographie und der Geostatistik übernommen werden.

Nach [Bartelme, 1995] werden die beiden Aspekte Positions- und Attributgenauigkeit unter dem Qualitätsmerkmal „Genauigkeit“ zusammengefaßt, die dann *Präzision*, *Auflösung* und *Zuverlässigkeit* beinhaltet (siehe unten). Es werden zusätzlich zwei Aspekte eingeführt und zwar die *Zugreifbarkeit* und die *Identifizierbarkeit*. Im folgenden werden die obengenannten Begriffe definiert, allerdings wird die *Genauigkeit* in späteren Abschnitten dieses Kapitels weiter behandelt.

- **Präzision:** Sie entspricht der Wahrscheinlichkeit, einen richtigen Wert (‘wahren Wert’) zuweisen zu können. Sie kann auch das Maß dafür sein, wie weit der ‘gemessene’ Wert von der besten Schätzung des wahren Wertes (Mittelwertes) abweicht. Für geometrische Aussagen ist sie gleichbedeutend mit der Angabe des im Vermessungswesen gebräuchlichen mittleren Punktlagefehlers bzw. mit der Angabe von Konfidenzintervallen.
- **Auflösung:** Sie gibt an, wie genau eine Entität überhaupt repräsentiert werden kann. Beispiel hierfür sei die Auflösung eines Rasterbilds in Pixel, deren Abmessungen einem genäherten Quadrat von 10 m x 10 m entspricht. In gewissen Sinn stellt die Auflösung die (theoretisch) maximal erreichbare Präzision dar.
- **Zuverlässigkeit:** Sie stellt ein Maß für die Wahrscheinlichkeit dar, daß Fehler in einem Modell entdeckt werden können. Wenn etwa das Modell keine unabhängigen Kontrollmessungen erlaubt, ist diese Wahrscheinlichkeit gleich Null.
- **Zugreifbarkeit:** Dieses Kriterium gibt an, wie einfach oder schwierig sich der Zugriff auf das jeweilige Objekt gestaltet. Dabei geht es nicht um die Zugriffsberechtigung rechtlicher oder administrativer Art, sondern um die Möglichkeit der technischen Realisierung.
- **Identifizierbarkeit:** Sie gibt an, wie einfach oder schwierig es ist, ein Objekt von anderen Vertretern derselben Kategorie zu unterscheiden. Klar definierte Geometrien sind leicht unterscheidbar und werden daher von dieser Warte aus

als qualitativ hochwertig eingestuft. Schwammige Definition hingegen lassen diese Eigenschaft vermissen.

Nach [Lother, 1999a] werden die Merkmale in zwei Gruppen unterteilt:

Qualitätsmerkmale	
Pragmatische Qualitätsmerkmale	Deskriptive Qualitätsmerkmale
Herkunft	Metrische Genauigkeit
Aktualität	Semantische Richtigkeit
Datenbeschaffenheit	Logische Konsistenz

Tab. 3-2: Qualitätsmerkmale nach [Lother, 1999]

Die pragmatischen Qualitätsmerkmale geben allgemeine anwendungsorientierte Hinweise über einen Datenbestand. Sie beinhalten i.d.R. die Ausschlußkriterien für seine Verwendung und lassen sich meist schnell beurteilen. Diese Qualitätsmerkmale können meist mittels freiem Text oder Verweis beschrieben werden und gelten entweder für den gesamten Datensatz, einen Themenbereich oder ein Gebietsfragment und beinhalten damit globale Aussagen über die Qualität der Daten. Die Datenbeschaffenheit kann u.a. Hinweise über die erforderlichen Verarbeitungsmethoden, die Datenvolumina, die nötigen Schnittstellen sowie den Erfassungsmaßstab und die Datendichte oder die Auflösung geben. Die deskriptiven Qualitätsmerkmale haben meist lokalen Charakter und geben Auskunft über die Qualität einzelner Merkmale, Objekte oder Gruppen von Objekten. Sie müssen durch geeignete Tests bestimmt werden. Hierfür sind Prüfmittel und Kenngrößen erforderlich, deren Einsatz oft sehr aufwendig werden kann.

In der Diplomarbeit von Jens Schumann [Schumann, 2000] wurde ein Qualitätsmodell nach WILTSCHEK dargestellt, in dem von sieben

Qualitätsmerkmale ausgegangen wird, die in die Gruppen *Zuverlässigkeit*, *Integrität* und *Genauigkeit* eingeteilt sind.

Qualitätsmerkmale		
Zuverlässigkeit	Integrität	Genauigkeit
Verfügbarkeit	Vollständigkeit	Metrische Genauigkeit
Aktualität	Konsistenz	Semantische Genauigkeit
	Korrektheit	

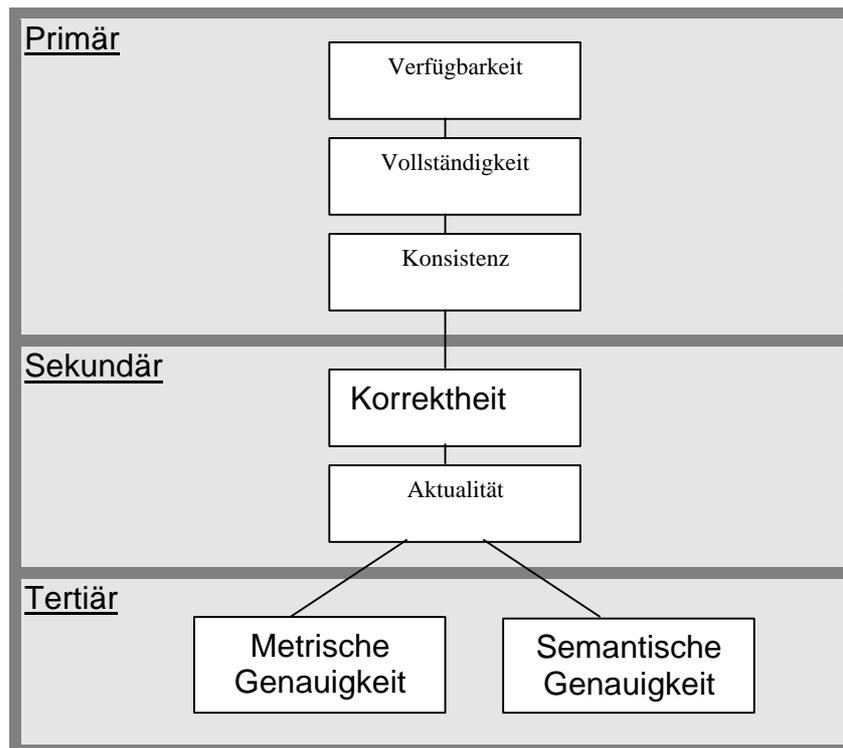
Tab. 3-3: Einordnung der Qualitätsmerkmale nach WILTSCHKO (2000)

Außer dieser Einteilung werden diese sieben Eigenschaften darüber hinaus hierarchisch in

- a) Primärqualitätsmerkmale
- b) Sekundärqualitätsmerkmale
- c) Tertiärqualitätsmerkmale

gegliedert, wodurch die Bedeutung der einzelnen Merkmale zum Ausdruck gebracht wird.

Bevor eine Information nicht verfügbar, vollständig und konsistent ist, sind Gedanken darüber irrelevant, ob diese Information korrekt oder aktuell ist. Ebenso ist es zweckmäßig die Information auf Korrektheit und Aktualität zu prüfen, bevor eine Genauigkeitsbetrachtung durchgeführt wird. Aus diesen Überlegungen ergibt sich die in der folgenden Abbildung 3-4 dargestellte hierarchische Gliederung.



Tab. 3-4: Hierarchische Gliederung der Qualitätsmerkmale nach WILTSCHKO
übernommen aus [Schumann, 2000]

Im folgenden werden die nach Caspary [Caspary, 1993] oben erwähnten Merkmalen kurz erläutert :

- **Herkunft der Daten:** Es wird verlangt, daß für den gesamten Datensatz die wichtigen Angaben vorhanden sein sollen, die Hinweise auf die folgenden Fragen geben: z. B. wann wurden die Daten erfaßt, von welcher Person bzw. Organisation, in welcher Umgebung geschah dies und welche modellierenden Annahmen wurden zugrunde gelegt (zu erwähnen sind u.a. Abstraktions-, Interpolations- und Generalisierungsaspekte oder auch Transformationsparameter). Hierzu gehören u.a. die Angaben zum Kartenmaßstab sowie über die Zeitpunkte der Datensammlung, der Aufbereitung und der Veröffentlichung der Daten.
- **Genauigkeit:** Die Genauigkeit bezieht sich sowohl auf Position und Geometrie als auch auf Thematik (Attributgenauigkeit). Bei der Geometrie kann die Genauigkeit als Übereinstimmung mit einem wahren Wert, mit einer Regel oder mit einem Modell aufgefaßt werden. Im Zusammenhang mit einem GIS-

Datenmodell wird die Genauigkeit als Grad der Übereinstimmung mit der Wirklichkeit definiert. Die Positionsgenauigkeit (Lagegenauigkeit) wird meistens nach den Regeln der Fehlerlehre bestimmt und gibt den Zusammenhang zwischen dem gemessenen und dem wahren bzw. plausiblesten Wert an. Außerdem ist es möglich die Genauigkeit mit einer ordinalen Skala wie „gut“, „sehr gut“ usw. sowie mit einer nominalen Skala wie „genau“ und „ungenau“ festzulegen. Die Attributgenauigkeit bezieht sich zunächst auf einzelne Objekte. Dabei werden je nach Skala metrische Werte bei kontinuierlich meßbaren Attributen (z. B. Höhe eines Gebäudes) oder die Angaben „richtig“ oder „falsch“ bei der Zuordnung von Objekten- oder Attributsschlüssel (z. B. Nutzungsart eines Gebäudes) verwendet. Außerdem sind noch die Klassifizierungs- und Abgrenzungsgenauigkeit zu beachten (siehe 3.3). Im Gegensatz dazu wird die Genauigkeit nach Bartelme [Bartelme, 1995] als ein Merkmal definiert, daß die Begriffe Präzision, Auflösung und Zuverlässigkeit umfasst (siehe oben)

- **Logische Konsistenz:** Darunter wird allgemein ein widerspruchsfreier Zustand eines Datenbestands verstanden. Sie bezieht sich meistens auf die Richtigkeit der Geometrie und Topologie des Datenmodells (räumliche Konsistenz), sowie auf die Richtigkeit (Plausibilität) der Referenzen in der Datenbank. Bezüglich der Konsistenz des Datenmodells wird auch zwischen räumlicher und semantischer Konsistenz differenziert. Dabei bedeutet räumliche Konsistenz, daß bestimmte topologische Regeln eingehalten werden, z. B. müssen alle Kanten an Knoten beginnen und enden. Die semantische Konsistenz bezieht sich auf die Übereinstimmung der im Datenmodell gespeicherten Attribute mit einem Objektartenkatalog oder mit erwarteten Beziehungen zwischen Attributen. Auftretende Inkonsistenzen beziehen sich überwiegend auf einzelne Objekte oder Beziehungen zwischen benachbarten Objekten, seltener auf den gesamten Datenbestand.
- **Vollständigkeit:** Vollständigkeit bezieht sich auf das gesamte Datenmodell. Sie ist insgesamt abhängig von der semantischen, räumlichen und zeitlichen Auflösung der Datenmodellierung. Die Vollständigkeit ist zunächst eine Funktion der semantischen und geometrischen Modellauflösung, die im konzeptionellen Modell der Wirklichkeit festgelegt wird. Demnach kann die

Forderung nach vollständiger Erfassung aller Objekte einer bestimmten Objektart nicht absolut aufgefaßt werden, sondern es ist notwendig, eine bestimmte Mindestgröße des Objekts der Wirklichkeit festzustellen. Die Vollständigkeit eines Modells wird dann gegen diesen Schwellwert gemessen. Sind andere als die bei der Entwicklung des konzeptionellen Modells berücksichtigten Anwendungen eines Datenmodells beabsichtigt, wird man die Auswirkung der Erfassungsgeneralisierung gegen das entsprechende konzeptionelle Modell ermittelt. Da es aus wirtschaftlichen Gründen nicht möglich sein wird, ein Datenmodell kontinuierlich zu aktualisieren, ist Vollständigkeit auch abhängig von der Organisation der Erfassung von Veränderungen in der Realität, d.h. abhängig vom Aktualisierungsturnus.

- **Aktualität:** Sie bezieht sich auf die Aktualität der verwendeten Datenquellen, auf Zeitangaben zur Erfassung und Speicherung des Datenmodells sowie geschätzte Veränderungsraten und auf die letzte und nächste Fortführung.

Nach der allgemeinen Beschreibung der Qualitätsmerkmale von Geodaten, werden in den folgenden Abschnitten die beiden Aspekte thematische bzw. geometrische Genauigkeit (Unsicherheit) in GIS kurz erläutert.

3.3 Thematische und geometrische Genauigkeit

Wie es schon in Kapitel 2 in Abbildung 2-2 dargestellt ist, beinhalten alle drei verschiedenen Modelle, das Rastermodell, Vektormodell und das hybride Modell jeweils geometrische und thematische Anteile.

Im Rastermodell werden die Objekte über eine Flächendarstellung beschrieben, die ein geometrisches und thematisches Attribut als Information tragen (1:1-Relation). Die Geometrie wird durch das Enumerationsverfahren bestimmt, dabei wird ein Objekt durch gleichmäßige Rasterzellen (Pixel) beschrieben, d.h. über den Darstellungsraum wird ein festes Raugitter oder Raster gelegt. Die Pixel werden über ihre Zeile bzw. Spalte innerhalb dieses Rasters angeordnet. Im Vektormodell setzt sich die Geometrie, wie schon erwähnt, auf der Basis geometrischer Primitiver (Punkt, Linie) zusammen. Dabei kommt das Prinzip der Randbeschreibung zum Einsatz, in dem z. B. Flächen durch ihre Umringslinien dargestellt werden können. Einem Objekt können in diesem Modell n Attribute

zugeordnet werden, deswegen wird bei der Thematik von 1:n-Relation gesprochen.

3.3.1 Thematische Genauigkeit

Die Thematik wird durch Attribute beschrieben, die sowohl diskrete als auch kontinuierliche Werte annehmen können. Eine der wichtigsten Quellen zur Erhaltung von thematischen Daten sind die multispektralen Fernerkundungsaufnahmen (siehe Kapitel 2.2). Die Auswertung von Fernerkundungsdaten beinhaltet die Definition von Klassen, z. B. Landnutzungsklassen, und die Bestimmung der Zugehörigkeit der flächenhaft erfaßten Ausschnitte der Erdoberfläche zu dieser Klasse. Die thematische Genauigkeit ist von der Aufnahme und der Klassifizierung abhängig. Durch Klassifizierungsverfahren werden dann Maßzahlen für die Zugehörigkeit der betrachteten Bildelemente zu den einzelnen Klassen berechnet. Bei solchen Klassifizierungsmethoden wird mit Hilfe von Entscheidungskriterien jedem Bildelement eine einzige Klasse zugewiesen, aber man kann den Grad der Zugehörigkeit nicht feststellen. So ist es z. B. nicht möglich, zu beurteilen, ob eine starke Zugehörigkeit (z. B. 90% Wald und 10% Wiese) oder eine schwache Zugehörigkeit (z. B. 51% Wald und 49% Wiese) vorliegt. Diese Vorgehensweise führt zu ungenauen Grenzen zwischen den beiden Klassen (in diesem Fall Wald und Wiese). Zur Beurteilung der Qualität der Klassifizierung werden die Ergebnisse der Klassifizierung mit Gebieten bekannter Thematik verglichen. Die Behandlung der thematischen Genauigkeit (Unsicherheit) und ihre Beurteilung wird in dieser Arbeit nicht behandelt und es sei auf die folgende Literatur verwiesen [Glemser & Klein, 1999a], [Glemser & Klein, 1999d], [Fritsch et al., 1998], [Caspary, 1992a], und [Goodchild, 1995].

3.3.2 Geometrische Genauigkeit

Um die Unsicherheit in der Geometrie eines Objektes veranschaulichen zu können, betrachtet man folgendes Beispiel: Läßt man ein bestimmtes Objekt (Naturgebiet) durch tachymetrische Aufnahme oder durch Digitalisierung aus einem Luftbild am PC mehrmals unabhängig voneinander erfassen, wie in der

Abbildung 3-1 dargestellt ist, und überlagert man anschließend alle Erfassungen, so zeigen sich deutliche Unterschiede zueinander. Durch diese Überlagerung entsteht ein Bereich, in dem die Geometrie streut. Dieser Bereich repräsentiert die Variation und damit die Unsicherheit in der Geometrie des Gebietes. In Abbildung 3-1a sieht man zwei Erfassungen: ATKIS Daten (grün) und manuell digitalisierte Daten (blau). Wenn diese Erfassungen mehrmals unabhängig von einander durchgeführt werden, entsteht ein Bereich, in dem alle möglichen Realisierungen sich befinden können und damit die wahre Geometrie. Dieser Bereich ist in der Abbildung 3-1b dargestellt.

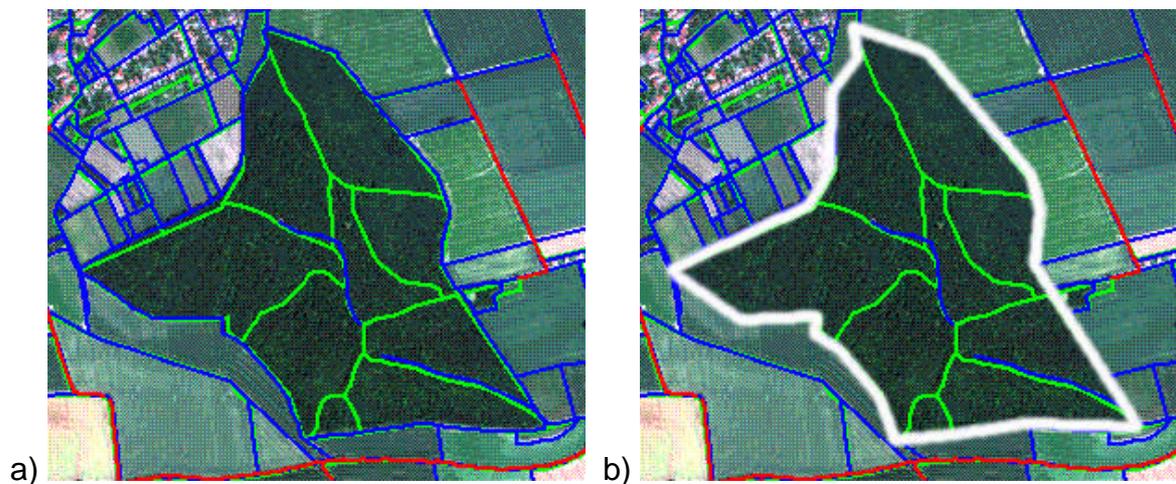


Abb. 3-1: verschiedene Erfassungen der Grenze eines Gebietes

a) zwei Datensätze b) der Bereich von allen möglichen Realisierungen

Es stellt sich die Frage, welche Ursachen können auf die Größe der Variation Einfluß nehmen. Zum größten Teil der Ursachen der geometrischen Unsicherheit trägt der Erfassungsvorgang bei (siehe Abschnitt 2.2). Der Erfassungsvorgang von realer Welt bis in ein GIS- Modell, läßt sich in drei verschiedene Schritte zerlegen nach [Glemser, 1999d]: zunächst wird das zu erfassende Objekt der realen Welt räumlich von Nachbarobjekten abgegrenzt. Im nächsten Schritt ist der gedachte Verlauf der Grenzlinie durch die Bestimmung markanter Punkte in diskreter Form festzulegen, und im dritten und letzten Schritt werden dann die gefundenen Punkte digitalisiert, falls diese nicht schon in digitaler Form vorliegen.

Bei der Digitalisierung wird die Genauigkeit von den eingesetzten Geräten und Meßmethoden beeinflusst, sowie von dem Betrachter selbst. Die Auswirkungen sind eher zufälliger Natur. Im Gegensatz dazu haben die Fehler bei der

Diskretisierung systematischen Charakter. Je nach Art des Objektes kommen die einzelnen Einflüsse bei den Erfassungen in unterschiedlicher Stärke zum Tragen. Bei der Diskretisierung lassen sich künstliche Objekte (Bauwerke oder Flurstücke) besser von anderen Objekten unterscheiden, als dieses für natürliche Objekte der Fall ist. Caspary [in: Caspary, 1992b] verdeutlicht dieses Problem anhand folgendem Beispiel: Die Grenze zwischen Nutzungsarten Wiese und Wald kann in der Natur eine breite Übergangszone haben, und die Festlegung dieser Grenze kann von verschiedenen Beobachtern unterschiedlich interpretiert werden. Glemser [in: Glemser & U.K., 1999b] bezeichnet diese breite Übergangszone als einen fließenden Übergang, der eine Zone definiert, innerhalb der die Geometrie variieren kann. Die Breite der Übergangszone ist ein Maß für die geometrische Genauigkeit des erfaßten Objektes und damit seiner Koordinaten. Eine relativ genaue Geometrie liegt bei einer schmalen Übergangszone vor. Analog ist die Geometrie für eine breite Übergangszone ungenau. Die Fehler bei der Digitalisierung dürften bei gleichen Geräten, Methoden und Bearbeitern nahe zu gleich groß ausfallen.

Um die Geometrie in GIS mitzuverarbeiten, wird eine Beschreibung durch ein Modell benötigt. Im Kapitel 5 werden die verschiedenen Modelle kurz beschrieben und die Anwendung des stochastischen Modells wird anschließend im Kapitel 6 detailliert behandelt. Davor aber werden im nächsten Kapitel Grundlagen, Definitionen und Begriffe der Statistik kurz erläutert, die zum Verständnis dieser Kapitel beitragen.

4. Theoretische Grundlagen der Statistik

In diesem Kapitel werden die wichtigsten Begriffe der Statistik, wie Zufallsexperimente (Ereignisse), Zufallsvariable, Wahrscheinlichkeiten und Normalverteilung u.v.a.m. definiert. Dazu muß zunächst die Schreibweise der später benutzten Buchstaben (Variable) I vereinbart werden.

Der Inhalt dieses Kapitels orientiert sich an den Lehrbüchern der Statistik und Ausgleichsrechnung, [Kreyszig, 1973],[Höpcke, 1980] [Pelzer, 1985] und [Bosch, 1998].

4.1 Beschreibende Statistik

In der beschreibenden Statistik sollen Untersuchungsergebnisse (Stichproben), wie Genauigkeit der erfaßten Geodaten übersichtlich dargestellt werden. Aus den Stichproben werden eine oder mehrere Kenngrößen berechnet wie Erwartungswert, Varianz und Standardabweichung usw., die über das vorliegende Stichprobenmaterial möglichst viel Informationen liefern sollen. Mit diesen Kenngrößen können jedoch ohne weitere Untersuchungen nur Aussagen über die Ausgangsstichprobe gemacht werden. Solche Aussagen dürfen nicht ohne weiteres auf größere Grundgesamtheiten übertragen werden, so müssen solche Kenngrößen z. B. mit anderen Stichproben verglichen werden, damit man bessere und kräftigere Aussagen bekommen kann.

4.2 Schreibweisen in dieser Diplomarbeit

In der Sprache der Ereignisse werden die Operationen der Mengenlehre benutzt. Die Ergebnisse eines Zufallsexperiments werden mit kleinen Buchstaben bezeichnet, z. B. mit a , b , c oder ω . Die Zufallsvariable wird mit X und ihre Realisierung mit x bezeichnet. Die Ergebnismenge Ω besteht aus allen möglichen Ergebnissen des Zufallsexperiments. Experimente werden mit großen Buchstaben A , B oder C bezeichnet. Nachfolgend sind die wichtigsten Mengenoperationen, die in dieser Arbeit verwendet werden, und ihre Regeln dargestellt:

Durchschnitt:

Das Ereignis $A \cap B = A * B$ („A und B“, der Durchschnitt von A und B) tritt genau dann ein, wenn sowohl A als auch B, also beide gleichzeitig eintreten. Es gilt $A \cap B = \{\omega \mid \omega \in A \text{ und } \omega \in B\}$.

Vereinigung:

Das Ereignis $A \cup B$ („A oder B“ oder die Vereinigung von A und B) tritt genau dann ein, wenn mindestens eines der beiden Ereignisse A und B eintritt. Es ist $A \cup B = \{\omega \mid \omega \in A \text{ oder } \omega \in B\} = \{\omega \mid \omega \text{ liegt in mindestens einer der beiden Mengen A und B}\}$.

Das Ereignis \bar{A} („A nicht“, das Komplement bzw. Komplementärereignis von A) tritt genau dann ein, wenn A nicht eintritt. Dabei gilt $\bar{A} = \{\omega \mid \omega \in \Omega \text{ und } \omega \notin A\}$.

4.3 Definitionen

Im folgenden werden die wichtigen Begriffe bezüglich der Anwendung des stochastischen Modells kurz erläutert.

4.3.1 Zufallsexperiment

Darunter versteht man ein Experiment, bei dem vor jeder Durchführung nicht mit absoluter Sicherheit vorausgesagt werden kann, welche der möglichen Ergebnisse bei der bevorstehenden Versuchsdurchführung tatsächlich eintreten werden. [Als Beispiel kann der Fall der Bestimmung der Ausdehnung eines linienhaften Objektes (Straße, Uferlinie usw.) genommen werden].

Relative Häufigkeiten von Ereignissen:

Das gleiche Zufallsexperiment wird n-mal durchgeführt. Bei jedem Einzelversuch soll festgestellt werden, ob das Ereignis A eintreten wird.

Die relative Häufigkeit $r_n(A) = \frac{h_n(A)}{n}$ stellt den relativen Anteil der Versuche dar,

bei denen das Ereignis A eingetreten ist.

Die relative Häufigkeit besitzt folgende Eigenschaften:

$0 \leq r_n(A) \leq 1$ für jedes Ereignis A (**Nichtnegativität**);

$r_n(\Omega) = 1$ (**Normierung**);

$r_n(A \cup B) = r_n(A) + r_n(B)$, falls $A \cap B = \emptyset$

(Additivität bei unvereinbaren (disjunkten) Ereignissen);

$r_n\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} r_n(A_i)$ für paarweise unvereinbare Ereignisse mit

$A_i \cap A_k = \emptyset$ für $i \neq k$ (**s- Additivität**)

Die relative Häufigkeit einer Vereinigung paarweise disjunkter Ereignisse ist gleich der Summe der relativen Häufigkeit.

4.3.2 Wahrscheinlichkeit von Ereignissen:

Es gibt verschiedene Begriffe der Wahrscheinlichkeit, wie klassische Wahrscheinlichkeit (zur Untersuchung der Chancen bei Glücksspielen) oder die sog. geometrische Wahrscheinlichkeit (Bestimmung der Wahrscheinlichkeit mit Hilfe der Längen- und Flächenberechnung) und die Wahrscheinlichkeiten, die C.F. Gauß bei stetigen Merkmalen bei der Fehler- und Ausgleichsrechnung berechnet hat. Diese Methode ist auch unter dem Namen Gauß'sche Glockenkurve bekannt. Diese Wahrscheinlichkeit wird in dieser Arbeit ihre Anwendung finden.

Interpretation einer Wahrscheinlichkeit

Das Ereignis A besitze bei einem Einzelexperiment die Wahrscheinlichkeit $p = P(A)$. Das Zufallsexperiment werde n - mal unabhängig unter den gleichen Bedingungen durchgeführt. Dann tritt für große n im allgemeinen das Ereignis A ungefähr $n \cdot p$ - mal ein. Damit liegt die relative Häufigkeit des Ereignisses A in der Nähe von p . Für große n gilt also der Regel die Näherung $r_n(A) \approx P(A)$.

Aus diesem Grund kann eine unbekannte Wahrscheinlichkeit p durch die relative Häufigkeit in einer genügend großen unabhängigen Versuchsserie geschätzt werden.

Die oben erwähnten Wahrscheinlichkeiten dürfen als Maß für die Chance des Eintretens des Ereignisses A nur dann benutzt werden, wenn das Zufallsexperiment noch nicht begonnen (stattgefunden) hat oder wenn über das laufende bzw. beendete Zufallsexperiment keinerlei Information vorliegen. Sind jedoch bereits Teilinformationen über den Ausgang bekannt, so weiß man, daß gewisse Ereignisse nicht eingetreten sein können. Durch eine solche Information ändert sich die Ergebnismenge und eventuell auch manche Wahrscheinlichkeit. Die vom Informationsstand abhängigen Wahrscheinlichkeiten heißen bedingte Wahrscheinlichkeiten. Diese Wahrscheinlichkeiten können von Interesse sein, wenn man die Wahrscheinlichkeit eines Ereignisses A unter der Bedingung, daß ein bestimmtes Ereignis B eintritt, berechnen möchte. Bedingte Wahrscheinlichkeiten sind Schätzwerte für die entsprechenden bedingten relativen Häufigkeiten, die folgendermaßen definiert sind:

$$\frac{h_n(A \cap B)}{h_n(B)} = \frac{h_n(A \cap B)/n}{h_n(B)/n} = \frac{r_n(A \cap B)}{r_n(B)} = r_n(A|B) \text{ mit:}$$

$h_n(A)$ und $h_n(B)$: die absolute Häufigkeit der Ereignisse A und B und $h_n(B) > 0$.

Definition der bedingten Wahrscheinlichkeit:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \text{ mit } P(B) > 0 \text{ heißt die bedingte Wahrscheinlichkeit des}$$

Ereignisses A unter der Bedingung B. Es ist die Wahrscheinlichkeit dafür, daß A eintritt unter der Bedingung, daß B eintritt (eingetreten ist).

Das ist die gesuchte Wahrscheinlichkeit, mit der zu entscheiden ist, ob Objekt X aus dem Modell A mit Objekt Y aus dem Modell B identisch sind, unter der Bedingung, daß sie sich schneiden (Mehr dazu in Kapitel 5). Im folgenden werden die Eigenschaften der bedingten Wahrscheinlichkeit zusammengestellt:

Das Ereignis B mit $P(B) > 0$ wird festgehalten. Die durch

$$P_B(A) = P(A|B) = \frac{P(A \cap B)}{P(B)}$$

definierte Ereignisfunktion P_B ist dann eine Wahrscheinlichkeit auf den Teilereignissen von B mit den Eigenschaften :

$$0 \leq P_B(A) \leq 1 \text{ für jedes Ereignis A} \quad (\text{Nichtnegativität});$$

$$P_B(B) = 1 \quad (\text{Normierung});$$

$$P_B(A \cup C) = P_B(A) + P_B(C), \text{ falls } A \cap C = \emptyset \quad (\text{Additivität});$$

$$P_B\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P_B(A_i) \text{ für paarweise unvereinbare Ereignisse}$$

$$A_i \cap A_k = \emptyset \text{ für } i \neq k \text{ } (\sigma\text{-Additivität}).$$

Es ist noch wichtig den Multiplikationssatz bei bedingten Wahrscheinlichkeiten zu erwähnen: $P(A \cap B) = P(A|B) \cdot P(B) = P(B|A) \cdot P(A)$ für $P(A), P(B) > 0$.

4.3.3 Unabhängige Ereignisse

Die (stochastische) Unabhängigkeit von zwei oder mehreren Ereignissen kann folgendermaßen definiert werden:

Das Ereignis A heißt unabhängig vom Ereignis B mit $0 < P(B) < 1$, wenn das Eintreten des Ereignisses B die Chance für das Eintreten des Ereignisses A nicht ändert. Dann besitzt das Ereignis A immer die gleiche Chance, unabhängig davon, ob B eintritt oder nicht. Dies ist genau der Fall, wenn gilt $P(A|B) = P(A|\bar{B})$.

Falls diese Bedingung nicht erfüllt ist, heißt das Ereignis A vom Ereignis B abhängig.

Falls A vom B unabhängig ist, ist zur Berechnung der Wahrscheinlichkeit des Ereignisses A die Information, daß B eingetreten ist, wertlos. Sie ändert die Wahrscheinlichkeit nicht. Dann gilt für die entsprechenden bedingten relativen Häufigkeiten die Näherung $r_n(A|B) \approx r_n(A|\bar{B}) \approx r_n(A)$.

Aus dem Produktsatz für bedingte Wahrscheinlichkeit ergibt sich unmittelbar der

Produktsatz für unabhängige Ereignisse:

Das Ereignis A ist vom Ereignis B mit $P(B) > 0$ genau dann unabhängig, wenn gilt $P(A \cap B) = P(A) \cdot P(B)$.

Dieser spezielle Produktsatz wird häufig zur Definition unabhängiger Ereignisse benutzt. Die Einführung über bedingte Wahrscheinlichkeit ist jedoch etwas anschaulicher. Aus der Produktdarstellung erhält man sofort die Symmetrie der Unabhängigkeit. Wenn A von B unabhängig ist, dann ist auch B von A unabhängig.

4.3.4 Unabhängige Zufallsexperimente

In diesem Abschnitt sollen mehrere Zufallsexperimente untersucht werden, die von einander unabhängig durchgeführt werden, als Beispiel betrachtet man zwei Zufallsexperimenten mit den Ergebnismengen Ω_1 und Ω_2 und die Wahrscheinlichkeiten P_1 und P_2 . Falls beide Zufallsexperimente gleichzeitig oder nacheinander durchgeführt werden, sind die Versuchsergebnisse dieses zweistufigen Zufallsexperiments geordnete Paare (ω_1, ω_2) mit $\omega_1 \in \Omega_1$ und $\omega_2 \in \Omega_2$. Oft interessiert man sich nur dafür, ob beim ersten Experiment ein Ereignis $A_1 \subset \Omega_1$ und beim zweiten ein Ereignis $A_2 \subset \Omega_2$ eintritt. Dieses Gesamtereignis bezeichnet man mit (A_1, A_2) . Es stellt den Durchschnitt der beiden Ereignisse (A_1, Ω_2) und (Ω_1, A_2) dar, also $(A_1, A_2) = (A_1, \Omega_2) \cap (\Omega_1, A_2)$.

Für die Wahrscheinlichkeit P des zweistufigen Gesamtexperimentes gilt

$$P(A_1, \Omega_2) = P_1(A_1) \text{ und } P(\Omega_1, A_2) = P_2(A_2).$$

Die beiden Ereignisse (A_1, Ω_2) und (Ω_1, A_2) sind im Gesamtexperiment (stochastisch) unabhängig, wenn die Produktdarstellung

$$P((A_1, \Omega_2) \cap (\Omega_1, A_2)) = P(A_1, \Omega_2) \cdot P(\Omega_1, A_2) = P_1(A_1) \cdot P_2(A_2)$$

gilt. Diese Eigenschaft ergibt die Definition der

Unabhängigkeit zweier Zufallsexperimente:

Zwei Zufallsexperimente mit den Ergebnismengen Ω_1, Ω_2 und den Wahrscheinlichkeiten P_1, P_2 heißen unabhängig, wenn für alle Ereignisse $A_1 \subset \Omega_1$ und $A_2 \subset \Omega_2$ bezüglich der Wahrscheinlichkeit P des Gesamtexperimentes die Produktdarstellung $P(A_1, A_2) = P_1(A_1) \cdot P_2(A_2)$ gilt, wenn also die Ereignisse

(A_1, Ω_2) und (Ω_1, A_2) unabhängig sind. Darunter versteht man allgemein die Unabhängigkeit der beiden Ereignisse A_1 und A_2 .

Entsprechend gelten die oben genannte Produktdarstellung auch für unabhängige mehrstufige Zufallsexperimente: $P(A_1, A_2, \dots, A_n) = P_1(A_1) \cdot P_2(A_2) \cdot \dots \cdot P_n(A_n)$.

Wird ein Zufallsexperiment häufiger unter denselben Bedingungen unabhängig durchgeführt und zu einem Gesamtexperiment zusammengeführt, gilt auch in diesem Fall die gleiche Produktdarstellung von oben.

4.3.5 Zufallsvariable eines Experiments

Eine Zufallsvariable X ist das durch eine Zahl ausgedrückte Ergebnis eines Experiments, dessen Realisierung x innerhalb eines Intervalls liegt. Je nach Wertebereich unterscheidet man zwischen diskreter und stetiger Zufallsvariable.

Eine Zufallsvariable X , deren Wertebereich endlich oder höchstens abzählbar unendlich ist, heißt diskret. In dieser Arbeit beschäftigt man sich grundsätzlich mit linienhaften Objekten wie Straßenlänge, Abständen zwischen Objekten usw., deswegen wird die diskrete Zufallsvariable nicht in diesem Abschnitt behandelt.

Stetige Zufallsvariable

Im Gegensatz zu der diskreten Zufallsvariablen ist der Wertebereich bei einer stetigen Zufallsvariable überabzählbar unendlich. Er enthält ganze Intervalle der reellen Zahlengeraden. Dabei besitzt jede einzelne Realisierung einer stetigen Zufallsvariablen die Wahrscheinlichkeit Null $P(X = x) = 0$ für jedes $x \in \mathfrak{R}$.

Bei der Untersuchung von stetigen Zufallsvariablen sind einige wichtige Begriffe zu beachten, die im folgenden beschrieben werden:

⇒ **Dichtefunktion einer stetigen Zufallsvariablen:**

Eine über ganze \mathfrak{R} integrierbare Funktion $f(x)$ heißt Dichte (Dichtefunktion oder Wahrscheinlichkeitsdichte), wenn sie folgende Bedingungen erfüllt:

a) $f(x) \geq 0$ für alle $x \in \mathfrak{R}$;

b) $\int_{-\infty}^{+\infty} f(x)dx = 1$.

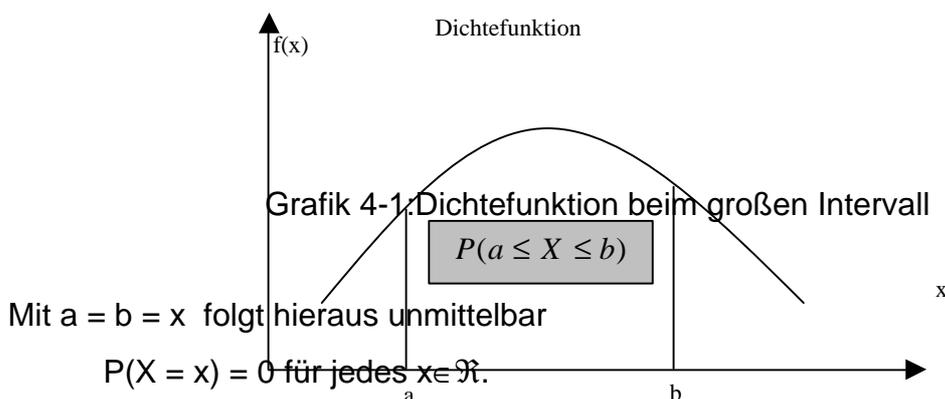
Der Graf der Dichte verläuft oberhalb oder auf der x- Achse (reelle Zahlengerade).

Die Fläche zwischen der Dichte $f(x)$ und der x- Achse hat den Inhalt Eins.

Eine Zufallsvariable X heißt stetig mit der Dichte $f(x)$, wenn für alle $a < b$ gilt

$$P(a \leq X \leq b) = P(\{\omega \in \Omega \mid a \leq X(\omega) \leq b\}) = \int_a^b f(x)dx .$$

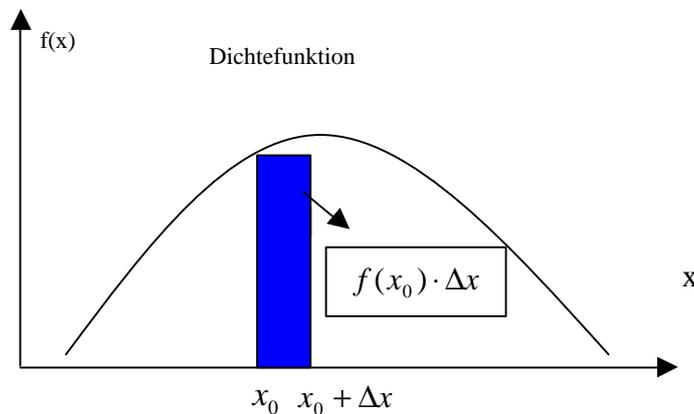
Die Wahrscheinlichkeit dafür, daß die Realisierung der Zufallsvariablen X im Intervall $[a,b]$ liegt, ist gleich dem Inhalt der Fläche, den die Dichte mit der x- Achse über diesem Intervall einschließt.



Wegen der Integraldarstellung beschreibt die Dichte $f(x)$ global (über Intervalle) Wahrscheinlichkeiten einer Zufallsvariablen. Sie kann auch Wahrscheinlichkeiten lokal beschreiben. Ist $f(x)$ an der Stelle x_0 stetig, so gilt für kleine Δx die Näherung

$$P(x_0 \leq X \leq x_0 + \Delta x) \approx f(x_0) \cdot \Delta x .$$

Die Näherung wird umso besser, je kleiner die Länge Δx des Intervalls ist.



Grafik 4-2: Dichtefunktion beim kleinen Intervall

⇒ Die Verteilungsfunktion einer stetigen Zufallsvariablen

Die stetige Zufallsvariable X besitze die Dichte $f(x)$, dann heißt die durch

$$F(x) = P(X \leq x) = P(\{\omega \in \Omega \mid X(\omega) \leq x\}) = \int_{-\infty}^x f(u) du ; x \in \mathfrak{R}$$

definierte Funktion $F(x)$ die Verteilungsfunktion der stetigen Zufallsvariablen X . Dabei stellt der Funktionswert $F(x)$ den Inhalt der Fläche zwischen der Dichte $f(x)$ und der x -Achse links von der Stelle x dar. Diese Funktion hat folgende Eigenschaften:

- F ist monoton nicht fallend, d.h. aus $x_1 < x_2$ folgt $F(x_1) \leq F(x_2)$.
- F ist an jeder Stelle stetig.
- Es gilt $\lim_{x \rightarrow -\infty} F(x) = F(-\infty) = 0$ und $\lim_{x \rightarrow +\infty} F(x) = F(+\infty) = 1$.
- Falls die Dichte $f(x)$ an der Stelle x_0 stetig ist, so ist die Verteilungsfunktion an dieser Stelle differenzierbar mit

$$F'(x_0) = \lim_{h \rightarrow 0} \frac{F(x_0 + h) - F(x_0)}{h} = f(x_0), \text{ falls } x_0 \text{ Stetigkeitspunkt der Dichte } f$$

ist.

Aus der Verteilungsfunktion F erhält man unmittelbar die Wahrscheinlichkeiten

$$P(a \leq X \leq b) = P(a < X < b) = \int_a^b f(x) dx = F(b) - F(a), a \leq b;$$

$$P(X \geq x) = P(X > x) = 1 - P(X \leq x) = 1 - F(x), x \in \mathfrak{R}.$$

Die Wahrscheinlichkeitsverteilung einer stetigen Zufallsvariablen X heißt stetige Verteilung. Durch die Angabe der Dichte bzw. der Verteilungsfunktion ist die Wahrscheinlichkeitsverteilung eindeutig bestimmt.

In der Geodäsie folgt der größte Teil der auftretenden Zufallsvariablen mehr oder weniger streng der Normalverteilung oder kann aus normalverteilten Größen abgeleitet werden. Bevor die Beschreibung der Normalverteilung durchgeführt wird, sollen noch die wichtigen Parameter dieser Verteilung definiert werden.

4.3.6 Normalverteilung und ihre Parameter

1. Der Modalwert:

Jeder Wert $x_{Mod} \in \mathfrak{X}$, für den die Dichte $f(x)$ maximal ist, heißt Modalwert oder Modus der stetigen Zufallsvariablen X . Falls die Dichte an der Stelle des Modalwerts stetig ist, besitzen Intervalle mit gleicher, aber sehr kleiner Länge in der Nähe des Modalwerts die größte Wahrscheinlichkeit. Bei kleinem Δx ist die Wahrscheinlichkeit $P(x \leq X \leq x + \Delta x)$ für $x = x_{Mod}$ am größten, falls $f(x)$ an der Stelle x_{Mod} stetig ist.

2. Der Erwartungswert:

Die Zufallsvariable X besitze die Dichte $f(x)$. Dann heißt

$$m = E(X) = \int_{-\infty}^{+\infty} x \cdot f(x) \cdot dx$$

der Erwartungswert von X .

Läßt man das entsprechende Zufallsexperiment sehr oft unabhängig durchführen, wobei die Realisierungen der Zufallsvariablen X zu einer Stichprobe mit dem Umfang n zusammengefaßt werden. Dann gilt für das arithmetische Mittel

$\bar{x} = \frac{1}{n} \cdot \sum_{i=1}^n x_i$ der Stichprobe für große n die Näherung: $\bar{x} \approx E(X) = m$. Von dieser

Näherung ist der wahre Wert einer Zufallsvariable zu unterscheiden. Einerseits sind bei geodätischen Problemstellungen häufig Zufallsvariablen anzutreffen, für die ein wahrer Wert nicht existiert. Auf der anderen Seite gibt es Zufallsvariablen, für die man sich sehr wohl einen wahren Wert vorstellen kann (z. B. gemessene Winkel oder Strecken). Dieser wahre Wert muß dann aber keineswegs immer mit

dem Erwartungswert zusammenfallen, der als Mittel aus unendlich vielen Messungen unter gleichen Bedingungen aufgefaßt werden kann.

3. Varianz und Standardabweichung (Streuungsparameter):

Die stetige Zufallsvariable X besitze die Dichte $f(x)$ und den Erwartungswert $m = E(X)$. Dann heißt im Falle der Existenz

$$\text{Var}(X) = s^2 = E((X - m)^2) = \int_{-\infty}^{+\infty} (x - m)^2 f(x) \cdot dx$$

die Varianz und die positive Quadratwurzel $s = +\sqrt{s^2}$ die Standardabweichung der stetigen Zufallsvariablen X .

Normalverteilung:

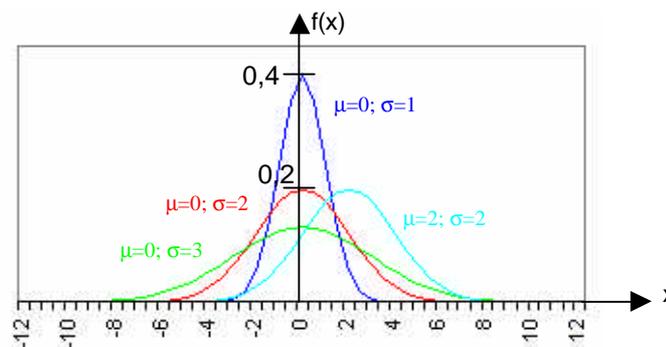
Eine Normalverteilung entsteht durch additives Zusammenwirken vieler unabhängiger Zufallsgrößen, von denen keine dominierend sein darf. Die Dichtefunktion $f(x)$

$$f(x) = \frac{1}{s\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2s^2}}$$

und die Integration der Dichtefunktion führt zur Verteilungsfunktion $F(x)$

$$F(x) = \frac{1}{s\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{(x-m)^2}{2s^2}} \cdot dx$$

einer Zufallsvariablen sind durch die Angabe von zwei Parametern, nämlich des Erwartungswertes μ und der Varianz σ^2 kurz $N(\mu, \sigma^2)$, vollständig bestimmt. Die Dichtefunktion $f(x)$ ist symmetrisch zur Achse $x=\mu$. Sie besitzt an der Stelle μ das Maximum (Modalwert). An den Stellen $\mu \pm \sigma$ hat $f(x)$ Wendepunkte. Bei einer Normalverteilung ist μ gleichzeitig der Erwartungswert und der Modalwert. In der nachfolgenden Abbildung sind verschiedene Dichten graphisch dargestellt.



Grafik 4-3: Dichten von Normalverteilungen

Jede Normalverteilung kann durch die Standardisierung (siehe oben) auf die Standard-Normalverteilung zurückgeführt werden. Die Standardisierung der Zufallsvariable X ergibt die $N(0;1)$ - verteilte Zufallsvariable

$$Z = \frac{x - \mathbf{m}}{\mathbf{s}},$$

und daher sieht die Wahrscheinlichkeitsdichte folgendermaßen aus:

$$f(z) = \frac{1}{\sqrt{2\mathbf{p}}} \cdot e^{-\frac{z^2}{2}}$$

Mit Integralrechnung kann man beweisen, daß die sog. Gauß'sche Glockenkurve $\phi(z)$ tatsächlich Dichte einer stetigen Zufallsvariablen Z ist.

Für die Dichte $f(z)$ läßt sich keine Stammfunktion in geschlossener Form angeben. Aus diesem Grund müssen die Werte der Verteilungsfunktion

$$F(z) = P(Z \leq z) = \frac{1}{\sqrt{2\mathbf{p}}} \int_{-\infty}^z e^{-\frac{u^2}{2}} \cdot du$$

mit Hilfe numerischer Methoden berechnet werden.

Wegen der Symmetrie der Dichte $f(z)$ zur y-Achse ist $F(0) = 0,5$. Aus symmetrischen Gründen genügt es, die Verteilungsfunktion $F(z)$ für nichtnegative Werte zu tabellieren und es gilt

$$F(-z) = 1 - F(z) \text{ für jedes } z.$$

Für die sogenannte α -Quantile z_a mit $F(z_a) = a$ gilt wegen Symmetrie zum Nullpunkt

$$z_{1-a} = -z_a \text{ für } 0 < a < 1.$$

Daher genügt die Vertafelung der rechtsseitigen Quantile (Tabellen finden sich in den statistischen Bücher, z.B. [Pelzer, 1985])

Allgemein gilt

$$P(a \leq Z \leq b) = F(b) - F(a)$$

und speziell

$$P(-z \leq Z \leq z) = F(z) - F(-z) = F(z) - [1 - F(z)] = 2 \cdot F(z) - 1.$$

Setzt man $Z = \frac{x - \mathbf{m}}{\mathbf{s}}$ so folgt:

$$P(-1 \leq \frac{x - \mathbf{m}}{\mathbf{s}} \leq 1) = P(-\mathbf{s} \leq x - \mathbf{m} \leq \mathbf{s}) = P(|x - \mathbf{m}| \leq \mathbf{s})$$

Für beliebige $k > 0$ erhält man: $P(|x - \mathbf{m}| \leq k\mathbf{s})$.

Im folgenden sind einige Werte für k zusammengestellt:

$k \cdot \mathbf{s}$	$P(x - \mathbf{m} \leq \mathbf{s})$	$P(x - \mathbf{m} \geq \mathbf{s})$
$1 \cdot \sigma$	0.683	0.317
$2 \cdot \sigma$	0.954	0.046
$3 \cdot \sigma$	0.997	0.003

Tabelle. 4-1: Wahrscheinlichkeit für Zufallsvariable

Diese Tabelle zeigt die Tatsache, daß bei normalverteilten Meßgrößen mehr als 2/3 aller zufälligen Abweichungen ($x - \mu = \varepsilon$) kleiner als die Standardabweichung σ sind und größere Abweichungen als 3σ äußerst selten vorkommen. Anders gesagt liegt z. B. zwischen $\mu - 2\sigma$ und $\mu + 2\sigma$ ungefähr 95.4 % der Wahrscheinlichkeitsmasse der Zufallsvariablen X , und zwischen $\mu - 3\sigma$ und $\mu + 3\sigma$ sogar 99.7 %.

Mit Hilfe der statistischen Auswertungen der Datensätze zwischen zwei Datenmodellen, wird versucht, Aussagen über die Wahrscheinlichkeit zu treffen, ob ein Objekt A des einen Datensatzes dem vermutlichen gleichen Objekt A' des anderen Datensatzes zugeordnet sein kann. Hierzu werden die Dichtefunktionen der Objekte A und A' (Zufallsvariable) durch ihre Varianzen und Erwartungswerte berechnet, die dann zur Verschneidung gebracht werden, und daraus kann man die Wahrscheinlichkeit für die Verschneidung der Objekte berechnen. Diese Wahrscheinlichkeit soll dazu beitragen, daß man die Entscheidung treffen kann, ob die Objekte A und A' identisch sind. Diese Berechnung erfolgt unter der Bedingung (bedingte Wahrscheinlichkeit), daß beide Dichtefunktionen der Objekte sich schneiden. Deswegen sollen die Ober- und Untergrenzen des Intervalls der Dichtefunktionen von $f(a)$ und $f(a')$ der beiden Linien und damit der Abstand berechnet bzw. bekannt sein.

5. Modellierung der geometrischen Genauigkeit

Bei der geometrischen Erfassung der Objekte der realen Welt zu deren Integration in einem GIS entstehen Unsicherheiten in der Geometrie eines Objektes. Diese Unsicherheiten berühren hauptsächlich auf den Erfassungsmethoden und damit verbundenen Fehlern (siehe Kapitel 2). Diese geometrische Unsicherheit wird im folgenden durch den Begriff geometrische Genauigkeit eines Objektes ausgedrückt. Die geometrische Genauigkeit ist eine der wichtigsten Qualitätskriterien eines GIS und dafür sollten allgemein anwendbare Genauigkeitsmaße vorhanden sein, um die Güte der Repräsentation von Objekten der realen Welt zu quantifizieren. Es wird sich zeigen, daß sich kein einheitliches Genauigkeitsmaß definieren lässt. Daher ist es notwendig, daß man für die verschiedenen Objekttypen unterschiedliche Maße definiert. Unter geometrischer Genauigkeit eines Modells bzw. einer Repräsentation ist die Güte der Übereinstimmung mit der Wirklichkeit zu verstehen. Dafür sollte ein Vergleich der zu prüfenden Daten mit richtigen (fehlerfreien) Daten (Ist-Soll- Vergleich) oder mit Daten höherer Genauigkeit durchgeführt werden. Der Vergleich kann stichprobenhaft geschehen und erfordert den Einsatz statistischer Methoden zur Auswertung. Damit kann man die geometrischen Abweichungen bzw. Fehler aufdecken. Diese Fehler sollten im Interesse der GIS-Nutzer bei Datenanalysen veranschaulicht werden, um die richtigen Schlüsse über die Qualität der Daten zu ziehen. Die Genauigkeitsmaße werden zunächst für punkt- und linienförmige Objekte untersucht und für die abgeleiteten ebenen Figuren berechnet. Die geometrische Genauigkeit der Punkte kann mit den Methoden der Fehlerlehre und Statistik abgeschätzt werden, sofern die Punkte in der Örtlichkeit eindeutig identifizierbar sind. Zur Bestimmung dieser Genauigkeit wird meistens der mittlere Fehler (auch als theoretische Streuung bezeichnet) angewendet. Sind die Koordinaten eines Punktes mit den dazu gehörenden mittleren Fehlern s_x und s_y angegeben, berechnet sich der mittlere Fehler des Punktes folgendermaßen aus:

$$s_p = \pm \sqrt{s_x^2 + s_y^2}$$

Er ist von der zufälligen Lage des Koordinatensystems unabhängig und somit als eindeutiges Genauigkeitsmaß für die Punktbestimmung geeignet [Reißmann,

1980]. Der Helmert'sche Punktfehler, die mittlere Fehlerellipse und die Konfidenzellipse zu einem angegebenen Wahrscheinlichkeitsniveau sind durchaus geeignete Maße zur Charakterisierung der Genauigkeit einzelner Punkte. Sie stehen jedoch nur dann zur Verfügung, wenn die Koordinaten durch Ausgleichung redundanter Messungen gewonnen wurden [Caspary & R.S., 1992c]. Der Einfluß des Punktfehlers auf andere geometrische Größen, wie z. B. die Länge einer Linie oder den Flächeninhalt eines geschlossenen Polygons, kann mit Hilfe des Fehlerfortpflanzungsgesetzes (kurz FFG) bestimmt werden. Zur Behandlung der Genauigkeit der Punkte und Anwendung von FFG sei auf die folgende Literatur [Bethge, 1997], [Shi & W.L., 2000] und [Höpcke, 1980] verwiesen. Bei den von Punkten abgeleiteten Objekten können solche Ansätze dazu beitragen, einige Verfahren zu entwickeln, die die geometrische Genauigkeit von Geraden und Polygonen (auch geschlossene Polygone) besser veranschaulichen. In diesem Kapitel werden die folgenden drei Ansätze

1. das ϵ - und Fehlerband,
2. das Minimum- Maximum- Modell und
3. das stochastische Modell

mit ihren Grundideen präsentiert. Bei diesen Ansätzen soll untersucht werden, wie die geometrische Genauigkeit der Objekte in einem GIS modelliert und visualisiert werden kann, so daß die Ergebnisse mit einer Bewertung der Qualität versehen und mit anderen, unabhängig voneinander erzielten Ergebnissen verglichen werden können. Als Beispiele zur Anwendung solcher Modelle zur Untersuchung der geometrischen Genauigkeit der Objekte werden hier zwei in GIS häufig genutzte Analysefunktionen betrachtet, nämlich der „Punkt-im-Polygon-Test“ und die sogenannte Flächenverschneidung. Der bekannte Punkt-im-Polygon-Test soll im GIS-Bereich feststellen, ob ein Punkt innerhalb, außerhalb oder auf dem Rand eines gegebenen beliebig geformten Polygons liegt. Bei den existierenden Algorithmen, die dieses Problem behandelt haben [Bill, 1996], ergibt sich als Lösung eine eindeutige Antwort, nämlich innerhalb oder außerhalb, wobei der Rand meist zum Inneren gezählt wird. Die Verschneidung ist allgemein eine der wichtigsten GIS-Analysefunktionen, die als Grundlage für Entscheidungen dienen können. Bei der geometrischen Verschneidung sollen die geometrischen Formen (Punkte, Linien, Flächen) der Objekte miteinander verschnitten werden, um zum

einen aus den sich überlappenden Teilen neue Objekte zu bilden, und zum anderen die Gleichheit eines Objektes aus verschiedenen Datensätzen zu überprüfen. Die Verschneidung entspricht der logischen UND-Operation der Mengentheorie [Bill, 1996], [Bill & P.K., 1998a] und [Glemser & C.H., 1999c]. Mit den obengenannten Modellen sollen im folgenden die beiden Sachverhalte näher betrachtet werden.

5.1 Das ϵ - und Fehlerband

Dieses Verfahren war ein erster Schritt zur Entwicklung eines Konzeptes zur Charakterisierung und Visualisierung der Genauigkeit bei linienförmigen Objekten [Casparly, 1992a]. In diesem Konzept wird ein Vertrauensbereich definiert, innerhalb dessen die wahre Geometrie der Linie liegen soll. Dieser Bereich wird ϵ -Band (Chrisman, 1982) genannt. Bei [Casparly & R.S., 1992c] wird die Konstruktion eines solchen Bandes durch zwei Parallelen mit der Länge l im Abstand ϵ und durch zwei Halbkreise an den Linienenden mit dem Radius ϵ gebildet (siehe Abbildung 5-1).

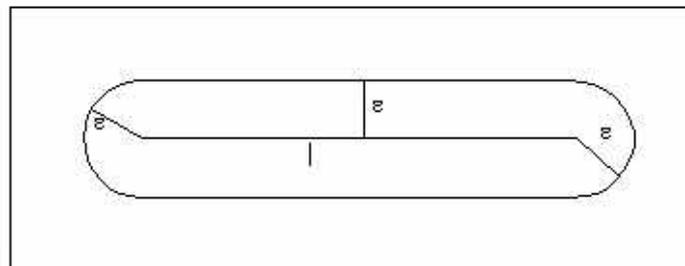


Abb. 5-1: Konstruktion des ϵ - Bandes

Dabei entspricht der Wert ϵ der Standardabweichung der Linienpunkte. Berechnet man die Fläche dieses Bandes, die zur Bewertung der Güte der Erfassung einer Strecke dienen kann, bekommt man folgendes Ergebnis

$$A_1 = 2 \cdot l \cdot \epsilon + \pi \cdot \epsilon^2 .$$

Dieses Ergebnis wird später mit den Ergebnissen der weiterentwickelten Fehlerbänder verglichen. Die Darstellung durch das ϵ - Band ist eine Vereinfachung der richtigen statistischen Definition, weil die Punkte, die in der

Mitte der Linie liegen, kleinere mittlere Fehler als die Linienendpunkte haben. Die Berechnung der mittleren Fehler unter der Bedingung, daß die Koordinatenfehler der Endpunkte unabhängig und gleich sind, d.h. $s = s_x = s_y$ ergibt sich $e = s \cdot \sqrt{2}$. Mit dieser Beschreibung erhält man das sogenannte Fehlerband. Dabei sieht das Fehlerband nur an den Punktenden ähnlich wie das ε - Band aus. Zur Linienmitte $l_i = l/2$ verjüngt es sich mit dem Faktor $1/\sqrt{2}$ und gewinnt eine Knochenform (siehe Abbildung 5-2).

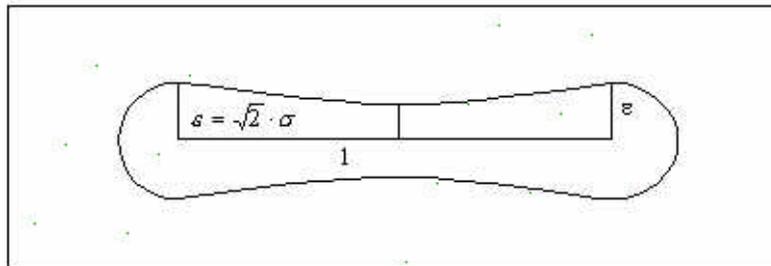


Abb. 5-2: Fehlerband nach [Caspary & R.S., 1992c]

Die Fläche des Fehlerbands in Vergleich zum ε - Band kann folgendermaßen berechnet werden:

$$A_2 = p \cdot e^2 + 2 \cdot l \cdot e \cdot \int_0^1 \sqrt{1 - 2x - 2x^2} \cdot dx \approx p \cdot e^2 + 1,62 \cdot l \cdot e^2$$

$$A_2 = A_1 - 0,38 \cdot l \cdot e \approx 0,8 \cdot A_1$$

Diese Fläche ist also mit dem Faktor 0,8 kleiner als die des ε - Bands. Aufgrund dieser Tatsache wird in [Caspary & R.S., 1992c] vorgeschlagen, daß das Fehlerband das konventionelle ε - Band ersetzen sollte. Da die analytische Beschreibung dieses Bandes sehr kompliziert ist, kann man für praktische Anwendungen das Fehlerband approximieren, so daß das Band von vier Geraden wie in Abbildung 5-3 gebildet wird. Die Fläche beträgt:

$$A_3 = 1,71 \cdot l \cdot e + p \cdot e^2$$

$$A_3 = A_1 - 0,29 \cdot l \cdot e$$

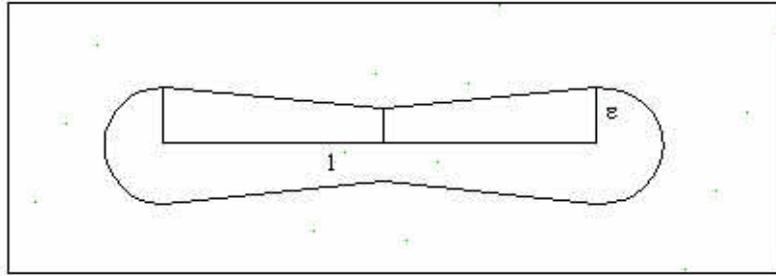


Abb. 5-3: Approximation des Fehlerbands

In der Arbeit [Shi & W.L., 2000] wird das Fehlerband-Modell auch bei dem Fall, daß die Linienendpunkten voneinander abhängig sind und verschiedene Varianzen haben, weiter entwickelt. Diese Entwicklung führt zur Definition des sogenannten „Generic error band, G-Band“, das in der folgenden Abbildung 5-4 dargestellt ist.

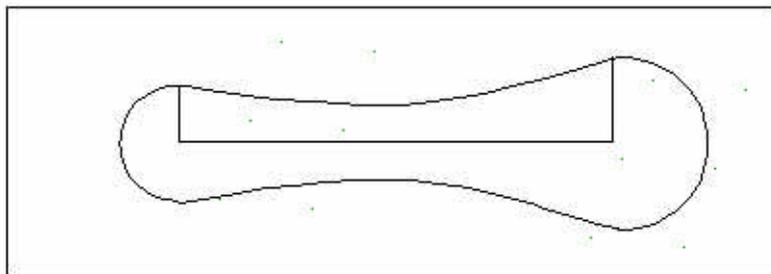


Abb. 5-4: Ein Beispiel zum G-Band

Dieses G-Band wird in dieser Arbeit nicht weiter behandelt.

Der Vertrauensbereich für die Lage einer Linie als Zufallsobjekt hat jedoch eine etwas andere Form, wie es durch die Untersuchung mittels Simulation gezeigt wurde. So wird das Fehlerband aus den Punkten der Ebene konstruiert, die mindestens mit einer bestimmten Häufigkeit von den digitalisierten Strecken geschnitten werden. Die Fläche dieses Fehlerband beträgt:

$$A_4 = 1.6 \cdot l \cdot e$$

Mehr zu dieser Untersuchung findet sich in der Arbeit von [Caspary & R.S., 1992c]. Für die Berechnung der Breite des Fehlerbands von Polygonen können alle über die Genauigkeit von Punkten und Linien angestellten Betrachtungen übertragen werden. Die Berechnung, wie z. B. mit der Fehlerfortpflanzung führt in diesem Fall problemlos zu Ergebnissen, die allerdings den Nachteil haben, daß

sie für graphische Darstellung wenig geeignet sind [Bill & P.K., 1998a]. Insofern kann die Fläche des Fehlerbandes als anschauliches Fehlermaß für Polygone verwendet werden. Dieses Band ist auch geeignet, eine Entscheidung über das Bestehen identischer Linien bei einer Überlagerung von verschiedenen Daten von unterschiedlicher Herkunft und Genauigkeit zu treffen. Die folgende Abbildung 5-5 zeigt ein geschlossenes Polygon mit ε -Band und die Fälle, die sich bei dem Punkt-im-Polygon-Test ergeben können.

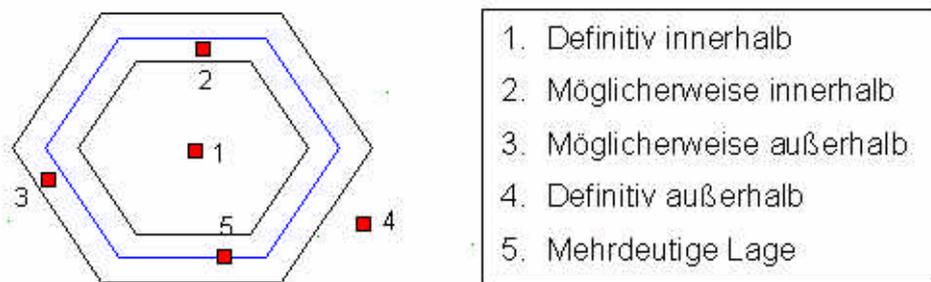


Abb.5-5: ε -Band und der „Punkt-im-Polygon-Test“

Durch diese Beschreibung der geometrischen Genauigkeit geht die Eindeutigkeit verloren; dabei erweitert sich die Lösung und liefert zusätzliche Aussagen: möglicherweise innerhalb bzw. außerhalb. Der fünfte Fall mit der mehrdeutigen Lage kann zum Objektrand zählen (also innerhalb). Diese Bewertung kann mittels Berechnung der Wahrscheinlichkeiten der Punktabstände zu den Polygonseiten erfolgen, wobei die Ergebnisse als Wahrscheinlichkeitsaussagen gegeben werden können, z. B. im Falle, daß der Punkt genau auf eine Polygonseite fällt, beträgt die Wahrscheinlichkeit 50%, mit Vergrößerung des Abstandes des beliebigen Punktes von der Polygonseite nimmt die Wahrscheinlichkeit zu. Mehr zu dieser Untersuchung findet sich bei [Kraus & H. K., 1993a], [Kraus & K.H., 1993b] und [Kraus & M.L., 1998].

Bei der Flächenverschneidung in diesem Modell wird die geometrische Genauigkeit mit der Anlehnung an die Berechnung des Fehlerbands durchgeführt. Man geht von dem einfachsten Fall der Verschneidung, nämlich der Berechnung

eines Schnittpunkts **S** von zwei Polygonseiten, wie es folgende Abbildung 5-6 darstellt:

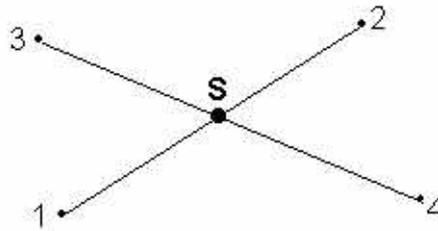


Abb.5-6: Schnitt zweier Polygonseiten

Die Berechnung der Koordinaten des Schnittpunktes und ihre Genauigkeit sollte geläufig sein. Auf die Formel dieser Berechnung wird das allgemeine FFG angewandt werden. Das FFG hat die allgemeine Form: $\Sigma_{SS} = A \cdot \Sigma_{PP} \cdot A^T$. Dabei enthält unter anderem die Varianz-Kovarianzmatrix (im folgenden VK-Matrix) Σ_{SS} die Varianzen $s_{x_s}^2$ und $s_{y_s}^2$ der gesuchten Koordinaten des Schnittpunktes. Die Funktional-Matrix A (oder auch als Jacobi-Matrix bezeichnet) enthält alle partiellen Differentialquotienten der entsprechenden Gleichung. Die VK-Matrix Σ_{PP} enthält die Genauigkeitsangaben über die Koordinaten der vier Ausgangspunkte $P_1 - P_4$. Bei der Verschneidung von zwei Polygonmengen werden mehr Punkte beteiligt und dementsprechend sollen mehrere Schnittpunkte und ihre Genauigkeit berechnet werden. Dabei sollte man sich Gedanken machen über die VK-Matrix und die daraus resultierenden Probleme, wie Speicherplatz, Rechenzeit und Modellierungsform der geometrischen Genauigkeit. Die Besetzungsstruktur und der Besetzungsgrad der VK-Matrix hat eine große Bedeutung bei der Berechnung und Modellierung der Genauigkeit des Schnittpunktes. In der Literatur wird von verschiedenen Bezeichnungen und Modellen gesprochen. Z. B. werden in der Arbeit von Bill und Korduan [in: Bill & P.K., 1998a] drei verschiedene Bezeichnungen der VK-Matrix gegeben:

1. „BEST CASE“: Eine Diagonalmatrix, d.h. es gibt keine Korrelation im Punkt und zwischen den Punkten einer Polygonmenge und keine Korrelation zwischen den beteiligten Polygonmengen. Hier wird zwischen zwei Modellen unterschieden: in dem einen haben alle Punkte die gleiche Varianz, d.h.

Skalarfaktor und Einheitsmatrix. Dieses Modell wählt man, wenn keinerlei Vorinformation zum Entstehungsprozeß der Geodaten vorliegt. In dem anderen Modell haben die beteiligten Polygonmengen unterschiedliche Varianzen und jede Koordinate hat innerhalb der Polygonmenge eine unterschiedliche Varianz.

2. „AVERAGE CASE“: dieser durchschnittliche Fall entspricht der Annahme eines Modells mit zufälligen und systematischen relativen Fehleranteilen. Es bestehen Korrelationen zwischen den y- und x-Koordinaten von Punkten, die mittels Kanten miteinander verbunden sind. Hier können gleiche Kovarianzfunktionen für die beiden Punktsätze angenommen werden.
3. „WORST CASE“: Die VK-Matrix ist in diesem Fall nahezu vollbesetzt (nur die Varianzen in den Punkten sind Null) und ergibt sich, wenn die Polygonmenge Ergebnis einer Netzausgleichung oder Homogenisierung ist.

Je nach Wahl der VK-Matrix bei der Durchführung der Verschneidung kann die Bildung des Fehlerbands der beteiligten Polygonlinien verschiedene Formen annehmen, wie es oben geschildert wurde. Die Genauigkeit der neu entstandenen Objekte (Objektpunkte) ist von den Ausgangspunkten abhängig, und wird durch das FFG oder einen Ausgleichungsprozess berechnet. Informationen über die Güte der Genauigkeit der Ergebnisse entnimmt man aus der Ergebnismatrix. Die VK-Matrix der Koordinatenunbekannten ist alleiniger Träger der Genauigkeitsinformation der neu entstandenen Punkte. Der Antwort auf die Frage, ob eine Verschneidung stattgefunden hat oder nicht, kann mit Hilfe eines Signifikanztests oder durch Berechnung von Konfidenzbereich gegeben werden. Der Konfidenzbereich ist ein Bereich, in dem die Wahre Punktlage mit einer vorzugebenden Wahrscheinlichkeit $1-\alpha$ sich befindet, diese Betrachtungsweise führt zur Anwendung der Fehler- oder Konfidenzellipse als Genauigkeitsmaß, wobei die gerechneten Standardabweichungen und Korrelationen der Punktkoordinaten für die Berechnung der große bzw. kleine Halbachse der Ellipse und ihren Richtungswinkel genutzt wird. Mit der Irrtumswahrscheinlichkeit α und das Quantil der c^2 -Verteilung mit bestimmten Freiheitsgraden kann man die Wahrscheinlichkeit der Schnittpunkte aus den Tabellen der c^2 -Verteilung ergreifen. (Tabellen finden sich in vielen Lehrbücher

der Statistik z.B. in [Pelzer, 1985]). Diese Thema wird hier nicht weiter behandelt und sei auf die folgenden Literatur [Pelzer, 1985] und [Höpcke, 1980] verwiesen. Die Modellierung kann wiederum durch Fehlerbänder (z. B. mit den berechneten Fehlerellipsen oder Konfidenzellipsen) durchgeführt werden. Die folgende Abbildung 5-7 ist ein mögliches Beispiel zu solchen Bändern.

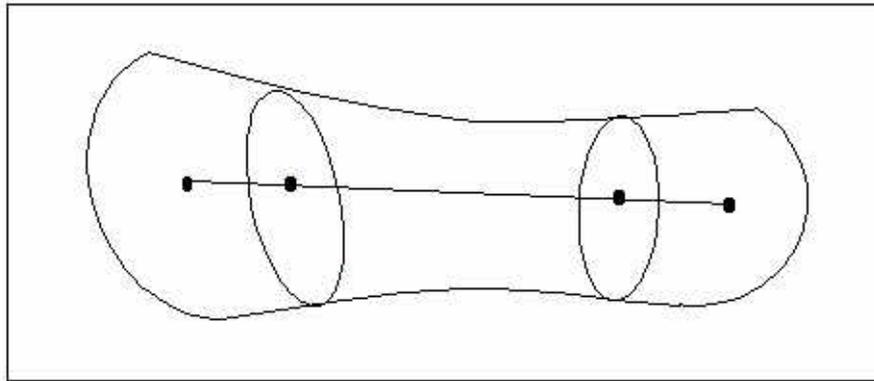


Abb. 5-7: Fehlerband aus Fehlerellipsen nach [Shi & W.S.,2000]

Nach dieser kurzen Erläuterung kann man feststellen, daß die Anwendung des ε -bands eine vereinfachte Form des Fehlerbands ist, was aber eine günstige Lösung bietet und deswegen bevorzugt wird. Die Berechnung der Breite des Bandes, um die Genauigkeit der Linie zu visualisieren, hängt von den Genauigkeiten der gemessenen vier Variablen ab (x , y - Koordinaten der beiden Punkte oder x , y - Koordinaten eines Punktes und Strecke und Richtung zum anderen Punkt). Diese Situation trifft dann zu, wenn die Linie eine diskrete Form besitzt, wie es bei der Trennungslinie zwischen zwei Grundstücken der Fall ist. Bei allen anderen Objekten sollte die Linie als Gesamtheit aller Linienpunkte betrachtet werden. Da auf alle Linienpunkte Einflüsse der Definition und der Diskretisierung in gleicher Weise wirken, sollten keine Unterschiede in der Behandlung einzelner Linienpunkte auftreten. Die Gleichbehandlung aller Linienpunkte erfordert eine Anpassung des oben beschriebenen Ansatzes. Dieser Sachverhalt wird im folgenden Abschnitt diskutiert, und die Erweiterung des ε -Bands führt zur Definition des Minimum- Maximum- Modells.

5.2 Das Minimum- Maximum- Modell

Die folgende Beschreibung bezieht sich auf die Arbeiten von [Glemser, 1996], [Schneider, 1999], [Krauß, 1999] und [Glemser et al., 1999e] an dem Institut für Photogrammetrie der Universität Stuttgart. In diesem Modell werden zur Objektgeometrie zwei zusätzliche Geometrien erzeugt, nämlich eine minimale und eine maximale Geometrie. Die maximale Geometrie (kurz Maximum) veranschaulicht die mögliche maximale Ausdehnung der Objektgeometrie, die ähnlich bestimmt wird, wie bei dem Ansatz des oben beschriebenen ε -Bandes. Außerhalb dieses Maximums existieren keine weitere Realisierungen der Objektgeometrie. Die minimale Geometrie (kurz Minimum) existiert für flächenhafte Objekte und sie veranschaulicht die minimale Ausdehnung der Objektgeometrie. Die mittlere Geometrie kann über eine Schätzung einer mittleren Ausdehnung des Objektes formuliert werden [Glemser et al., 1999e]. Es wird auch festgelegt, daß bei punkt- und linienförmigen Objekten die minimale Geometrie und die mittlere Geometrie identisch sind. Die folgende Abbildung 5-8 veranschaulicht diese Beschreibung an einem flächenhaften Objekt.

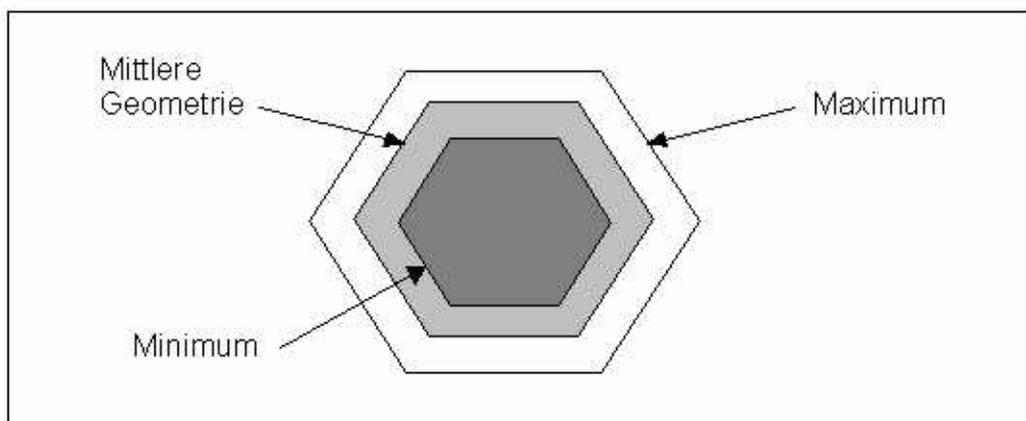


Abb. 5-8: Darstellung des Minimums, des Mittels und des Maximums eines Flächenobjektes nach [Glemser et al., 1999e]

Die Bildung der Geometrien für die einzelnen Elemente kann in der Tat nicht direkt erfaßt werden, sondern es erfolgt auf der Grundlage des ε -Bands von Chrismann, Die Ausdehnungen der Elemente werden als Toleranzbänder bezeichnet und ihre Breite durch den Wert ε definiert. Bei den punkt- und linienförmigen Objekten entspricht das Toleranzband automatisch der Maximum-Geometrie und die Minimum-Geometrie entspricht den vorhandenen Objektgeometrien. Bei

flächenhaften Objekten entsteht die Maximum-Geometrie durch eine zusätzliche Vereinigung des Bandes mit der Normalgeometrie, während die Minimum-Geometrie durch Differenzbildung von Mittelgeometrie und Band zu erzeugen ist, wie es in der folgenden Abbildung 5-9 dargestellt ist.

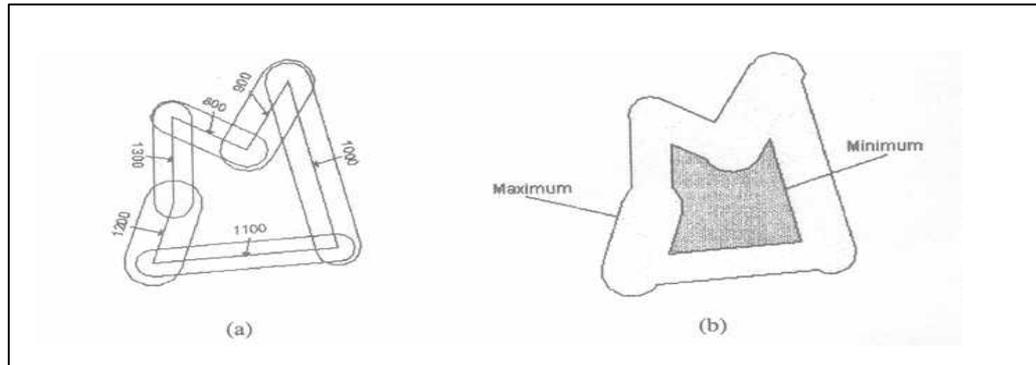


Abb. 5-9: a) Toleranzbänder um die Randlinien, b) Minimum- und Maximum-Geometrie

Die Modellierung der Breite der Toleranzbänder für das Minimum und das Maximum eines einzelnen Linienelements ist in einer praktikablen Erfassungsweise durch Angabe verschiedener ε -Werte denkbar. Diese werden mit ε -min und ε -max bezeichnet. Die Angabe solcher Werte bedarf der Durchführung von Tests, innerhalb derer durch Auswertung von Mehrfacherfassungen Wissen über die Variation der Realisierung erarbeitet wird. Die Tests sollten beispielhafte Objekte aller relevanten Objektarten umfassen. Dieses Modell hat den Nachteil, daß eine Angabe zur Verteilung der Realisierungen innerhalb des Toleranzbands fehlt, was dazu führt, daß alle möglichen Realisierungen eines Objektes als gleich wahrscheinlich betrachtet werden. Diese Annahme widerspricht dem Fall, daß die Realisierungen eines Objektes der Normalverteilung folgen. Die Integration dieses Modells kann in graphischer Form in einem GIS durchgeführt werden, wenn die drei getrennten Geometrien in verschiedenen Zeichenebenen (Layer) gespeichert werden. So wird für den Nutzer die Möglichkeit gegeben, mit der Normalgeometrie nach wie vor zu arbeiten, und bei der Frage nach der Genauigkeit zeigt das System zusätzlich Minimum und Maximum der Objekte (einzeln oder gemeinsam) an.

Diese Vorgehensweise wird im folgenden an den obengenannten Analysefunktionen praktiziert werden. Die Bildung des Minimums und des

Maximums um die mittlere Geometrie eines geschlossenen Polygons hat die gleiche Auswirkung auf die Lösung des „Punkt-im-Polygon-Tests“.

Es ergeben sich die folgenden Fälle, die in Abbildung 5-10 dargestellt sind.

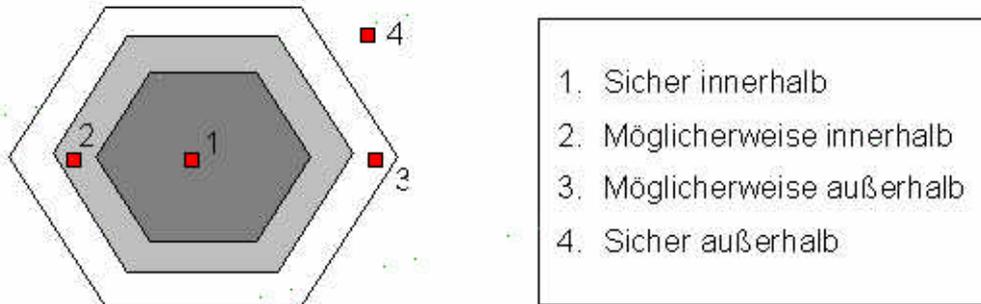


Abb. 5-10: Die Fälle bei der Bewertung des Punkt-im-Polygon-Tests

Diese Bewertung hat im Vergleich zu der oben beschriebenen Bewertung bei dem Fehlerband (siehe Abb. 5-5) einen kleinen Unterschied, und zwar wird der 5. Fall „mehrdeutige Lage“ nicht behandelt. Demzufolge geht man davon aus, daß die Punkte am Objektrand zum Inneren gehören. Für die verbleibenden Fälle bleibt die Entscheidung identisch, lediglich die Wortwahl ist eine andere (sicher = definitiv).

Die Untersuchung der geometrischen Verschneidung von zwei Objekten A und B wird in diesem Modell in zwei Schritten durchgeführt. Im ersten Schritt soll festgestellt werden, ob ein Schnitt vorliegt (Schnittprüfung) und im zweiten Schritt werden dann die neuen Objekte gebildet, falls ein Schnitt überhaupt existiert. Die Untersuchung beginnt mit der ebenenweisen Verschneidung der drei Geometrien der beteiligten Objekte A und B, d.h. Minimum A mit Minimum B ($A_{Min} \cap B_{Min}$), Mittel A mit Mittel B ($A_{Mittel} \cap B_{Mittel}$) und Maximum A mit Maximum B ($A_{Max} \cap B_{Max}$). Erwartungsgemäß müssen bei der Verschneidung nicht alle Schnittbildungen eine Ergebnismenge erzielen. Je nach nichtleerer Ergebnismenge kann festgestellt werden, wie sicher ein Schnitt vorliegt.

In der folgenden Tabelle 5-1 werden die verschiedenen Bewertungsmöglichkeiten schematisch dargestellt.

$(A_{Min} \cap B_{Min})$	$(A_{Mittel} \cap B_{Mittel})$	$(A_{Max} \cap B_{Max})$	Bewertung
--------------------------	--------------------------------	--------------------------	-----------

$E=O$	$E=O$	$E=O$	Objekte schneiden sich sicher nicht
$E=O$	$E=O$	$E=\emptyset$	Objekte schneiden sich möglicherweise nicht
$E=O$	$E=\emptyset$	$E=\emptyset$	Objekte schneiden sich möglicherweise
$E=\emptyset$	$E=\emptyset$	$E=\emptyset$	Objekte schneiden sich sicher

Tab.5-1: Bewertung des Schnitts ($E=O$: leere Ergebnismenge, $E=\emptyset$: nichtleere Ergebnismenge) nach Glemser [in: Glemser et al., 1999e]

Mit solchen Ergebnissen der Bewertungen des Schnitts kann sich der Nutzer entscheiden, wie sicher seine Schnittobjekte sein sollen. Bei einem sicheren Schnitt sollen alle Objektschnitte geprüft werden. Diese Entscheidung erfolgt automatisch und dadurch reduziert sich der Arbeits- und Zeitaufwand [Schneider, 1999].

5.3 Das stochastische Modell

Bei diesem Modell werden die einzelnen Linienpunkte als eine kontinuierliche stochastische Variable betrachtet. Die Verteilung der Realisierungen einer Linie wird hier miteinbezogen, was in den anderen Modelle nicht der Fall ist. Zur Beschreibung der Genauigkeit der Objekte stützt sich dieser Ansatz auf die Angabe der Parameter Mittelwert und Varianz der statistischen Verteilung der zu beschreibenden Größe. Diese Größe wird dabei als Zufallsvariablen angesehen und ihre Variation wirkt sich senkrecht zur eigentlichen Richtung der Linie aus. Die Gesamtheit aller Mittelwerte ergibt die mittlere Geometrie des Objektes. Theoretisch wären unterschiedliche Varianzen für jeden Linienpunkt denkbar (siehe Abschnitt 5.1), doch soll entsprechend der Gleichbehandlung und im Hinblick auf eine praktikable Anwendbarkeit nur eine Varianz für eine Linie gelten (entspricht dem „BEST CASE“). Im Normalfall besteht die Geometrie eines Objektes aus mehreren einzelnen Linien. Dabei gehören die meisten Endpunkte zwei Linien an. Durch sie entstehen Korrelationen der Linien im Objekt und damit der Linienpunkte untereinander. Diese Korrelationen bleiben hier unbeachtet, so daß die Varianzen aller Linienpunkte voneinander unabhängig sind. Durch das

stochastische Modell lassen sich nur zufällige Einflüsse beschreiben. Wie es aber schon im Kapitel 2 erwähnt wurde, wirken andere Einflüsse auf die Erfassung der Daten ein, die nicht ausschließlich zufälliger Natur sind. Um die geometrische Genauigkeit dennoch modellieren zu können, wird angenommen, daß die Einflüsse, die nicht zufällig sind, unbedeutend oder als annähernd zufällig zu betrachten sind. Als zugehörige Verteilung wird die Normalverteilung der Linie angenommen. Mit der Funktion werden für Linienpunkte Wahrscheinlichkeiten berechnet. Bildet man Isolinien für Punkte mit gleicher Wahrscheinlichkeit, so entstehen Bänder um das Objekt, die die Vertrauensbereiche der Geometrie analog zu den Toleranzbändern darstellen. Die geometrischen Genauigkeiten im stochastischen Modell werden durch Wahrscheinlichkeiten ausgedrückt. Sie sind quantitative Werte und folglich ein räumlich kontinuierliches Phänomen. Die Darstellung solcher Phänomene mit Vektordaten ist wenig geeignet, deswegen zeigt sich die Anwendung des Rastermodelles als besser geeignet. Der Übergang in den Rasterbereich ermöglicht eine andere Art der Darstellung [Kraus & K.H.,1993a]. Dazu werden die kontinuierlichen Werte durch Rasterelemente diskretisiert und in Grauwerte transformiert. Die Größe des dazu einzusetzenden Diskretisierungsintervalls Δ ist von der Varianz σ^2 der Liniengeometrie abhängig. Die Festlegung zu $\Delta=\sigma/4$ gewährleistet eine ausreichende Genauigkeit für eine eventuelle Rekonstruktion der Funktion [Glemser & U.K.,1999d]. Mit Hilfe des Intervalls Δ wird über das Objekt ein gleichmäßiges Raster gelegt, dessen einzelne Pixel den Punkten entsprechen, für die die Wahrscheinlichkeiten berechnet werden müssen. Die Berechnung der Wahrscheinlichkeit erfolgt mit der Dichtefunktion der Normalverteilung. Die einzelnen Grauschattierungen visualisieren dabei die Zu- bzw. Abnahme der Wahrscheinlichkeit, mit der ein beliebiger Punkt zum Objekt gehört.

In der Abbildung 5-11 werden die Wahrscheinlichkeitsmatrizen für Fläche und Linie dargestellt.

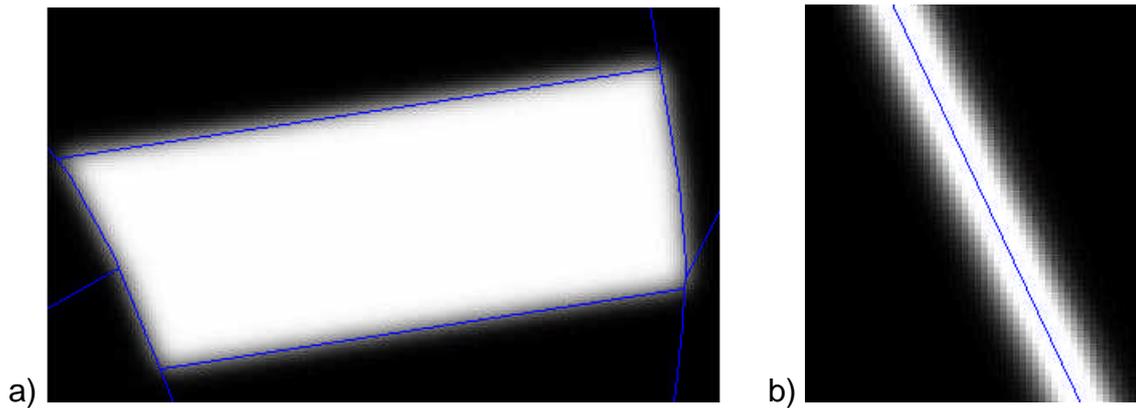


Abb. 5-11: Darstellung der Wahrscheinlichkeit a) Fläche, b) Linie

In den obengenannten zwei Modellen wurden die beiden Aufgaben im Vektorbereich behandelt. Im Gegensatz dazu werden die Aufgaben in diesem Modell im Rasterbereich durchgeführt. Die Rasterstruktur hat den Vorteil, daß die Geometrie der Rasterzellen (Pixel) in einer regelmäßigen Anordnung bereits vorliegt.

Die Entscheidung bei dem Punkt-im-Polygon-Test, ob ein Punkt zum Objekt gehört oder nicht, passiert mit Hilfe der berechneten Wahrscheinlichkeitsmatrizen. Die Berechnung der Wahrscheinlichkeit $p(x, y)$, mit der ein beliebiger Punkt $P(x, y)$ zum Polygon gehört, ergibt sich in Abhängigkeit vom Abstand des Punktes, wie in dieser Formel dargestellt:

$$p(x, y) = p(d) = \int_{-\infty}^d f(x) dx = F(d)$$

$$f(x) = \frac{1}{s\sqrt{2\pi}} \cdot e^{-\frac{1}{2}\left(\frac{x-m}{s}\right)^2}$$

Dabei ist $f(x)$ die Dichtefunktion der Normalverteilung und $F(d)$ deren Verteilungsfunktion. Für den Abstand d zur mittleren Grenzlinie (mittlere Geometrie) gilt, daß er innerhalb des Objektes positiv und außerhalb negativ sein soll (Abb.5-12).

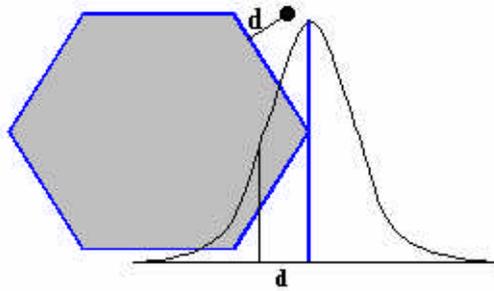


Abb. 5-12: Berechnung der Wahrscheinlichkeit für Punkt-im-Polygon-Test
 In dieser Abbildung bezeichnet die blaue Grenzlinie die mittlere Geometrie des Polygons. Die Wahrscheinlichkeit nimmt von Außen nach Innen zu, und beträgt 50%, wenn der Punkt genau auf der mittleren Linie liegt. Bei Linien- und Punktobjekte benötigt man zusätzlich eine Annahme über die Breite b des Objektes in der Wirklichkeit. Die entsprechende Wahrscheinlichkeit $p(x, y)$ eines linienförmige Objektes ergibt sich aus

$$p(x, y) = p(d, b) \cdot p(-d, b) \text{ mit}$$

$$p(d, b) = \int_{-\infty}^{\frac{b}{2}+d} f(x) dx$$

mit die Breite b des Objektes und Abstand d zur mittleren Linie (siehe Abb. 5-13).

Bei Punktobjekten berechnet sich die gesuchte Wahrscheinlichkeit $p(x, y)$ aus

$$p(x, y) = p(d, b)^2 \cdot p(-d, b)^2$$

Wobei die Wahrscheinlichkeiten $p(d, b)$ und $p(-d, b)$ wie oben berechnet werden.

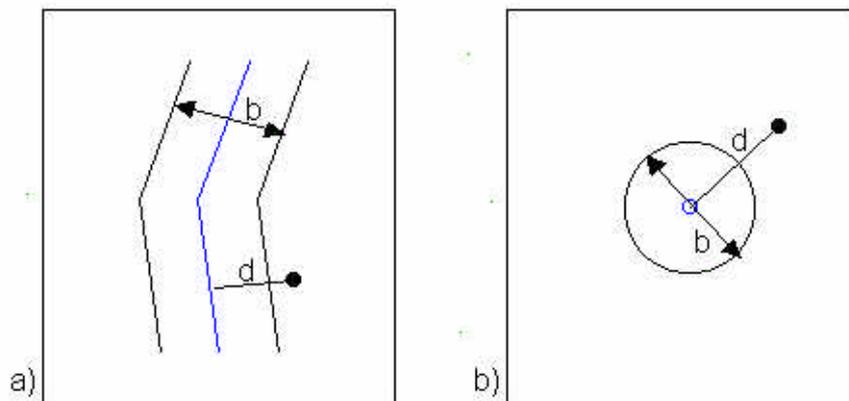


Abb. 5-13: Berechnung der Wahrscheinlichkeiten für a) Linie und b) Punkt

Die Lösung des Punkt-im-Polygon-Tests ist deshalb mit der Wahrscheinlichkeitsaussage machbar. Mehr zur Berechnung und Visualisierung solche Wahrscheinlichkeiten findet sich in der Arbeit von Kraus [in: Kraus & K.H., 1993b].

Bei der Verschneidung sollten die beteiligten Objekte in der Rasterstruktur vorliegen. Sind sie aber in Vektorstruktur vorhanden, müssen sie zuerst in Rasterstruktur umgewandelt werden.

Die Verschneidung im Rasterbereich besteht aus einer logischen Verknüpfung von überlagerten Zellen. Als Ergebnis der Verschneidung entstehen neue Objekte, bei denen die Eigenschaften der beteiligten Objekte übertragen werden sollen (Fortpflanzung). Es ist dabei von Interesse, daß die geometrische Genauigkeit der neuen Objekten bestimmbar und in einer Grafikform darstellbar ist. Die Verschneidung kann im Zusammenhang mit der logischen UND-Operation und der Wahrscheinlichkeitstheorie durch die folgende Formel:

$$P(A \cap B) = P(A) \cdot P(B)$$

definiert werden. Mit $P(A)$ bzw. $P(B)$ werden die Wahrscheinlichkeiten, daß eine Rasterzelle die jeweiligen Attribute aufweist, bezeichnet. Das geschieht unter der Voraussetzung, daß die beiden Thematiken unabhängig voneinander sind (siehe Kapitel 4). Durch diese Methode können auch für die neu entstandenen Objekte Wahrscheinlichkeitsmatrizen berechnet werden, und mit dieser Art der Modellierung ist der Nutzer in der Lage wichtige Schlüsse über die Genauigkeit zu entnehmen. Im nächsten Kapitel wird die Anwendung des stochastischen Modells an zwei Fällen untersucht und die Vorgehensweise dieses Modells detailliert erklärt werden.

6. Anwendung des stochastischen Modells

Der Kern dieses Kapitels ist die Anwendung des stochastischen Modells, um die Frage beantworten zu können, ob vermutlich gleiche Objekte in verschiedenen Datensätzen identisch sind. Dabei konzentriert sich die Anwendung auf die linienhaften Objekte.

Hier wird die Anwendung des stochastischen Modells in GIS aus folgenden Gründen näher erläutert:

- Die zwei Methoden, die in dem vorigen Kapitel beschrieben wurden, werden bei der Behandlung der Genauigkeit der Objekte im GIS mit Hilfe von Vektordaten durchgeführt, wie es mit dem Fehlerband und mit dem Minimum-Maximum-Modell gezeigt wurde. Diese beide Methoden können mit strengen und mühsamen Berechnungen eine Aussage zu der Wahrscheinlichkeit geben, ob zwei Objekte sich schneiden, oder ob sie identisch sind. Die Visualisierung dieser Wahrscheinlichkeiten mit Fehler- oder Konfidenzellipsen ist nicht unbedingt die ideale Möglichkeit [Bill & P. K., 1998b]. Für die Wahrscheinlichkeitsdarstellung sind die Rasterdaten besser geeignet, was heute aufgrund der Entwicklung der Softwarepakete zur digitalen Bildverarbeitung mit einem hybriden GIS möglich geworden ist (siehe Kapitel 3). Bei diesen Anwendungen werden die Vektordaten in Rasterdaten konvertiert und mit der entsprechenden Software können die gewünschten Arbeitsschritte durchgeführt werden. Mehr dazu im Abschnitt 6.3.

Bevor das Untersuchungsschema detailliert dargestellt wird, werden im nächsten Abschnitt die Datensätze, die in dieser Arbeit zur Verfügung standen, kurz beschrieben.

6.1 Beschreibung der Datensätze

Die drei zur Verfügung stehenden Datensätze sind die folgenden (siehe Abb. 6-1):

- a) ATKIS- Daten (grün)
- b) automatisch digitalisierte Daten (rot)
- c) manuell digitalisierte Daten (blau)

Die Luftbilder, von denen die automatisch- und manuell digitalisierten Daten gewonnen worden, sind im Sommer 1997 von der DPA (Digitale Photogrammetric Assemble) in vier Kanälen: blau, grün, rot und infrarot mit 1m Auflösung aufgenommen worden, mehr dazu in [Fritsch, 1997]. Die folgende Abbildung 6-1 stellt die oben genannten drei Datensätze in einem Bildausschnitt dar.

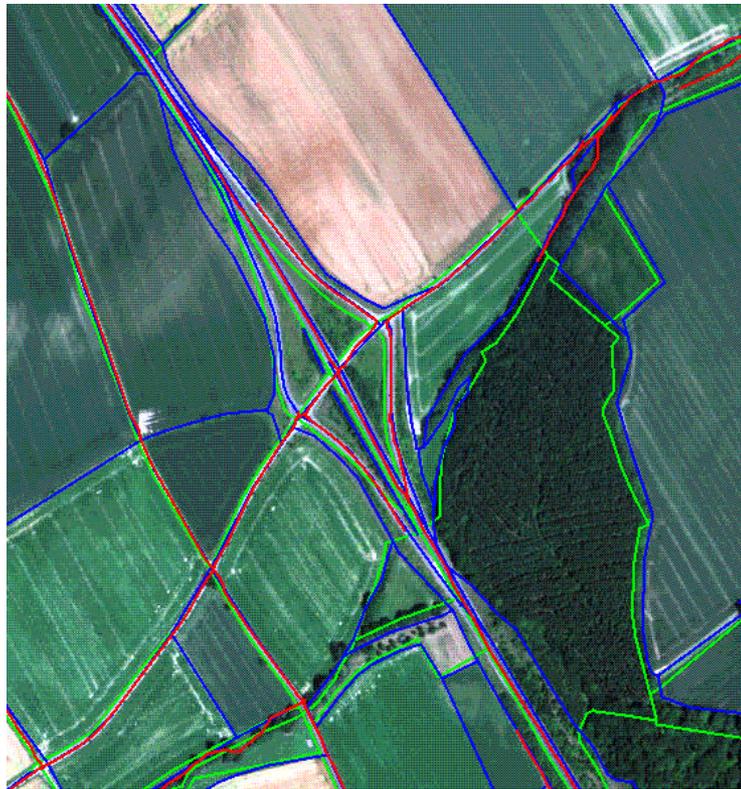


Abb. 6-1: Ausschnitt eines Luftbildes mit drei Vektordatensätzen

Diese Daten sind von drei verschiedenen Institutionen unabhängig voneinander erfaßt, wobei für die Digitalisierung der Datensätze b) und c) die gleichen Luftbilder zur Verfügung standen.

6.1.1 ATKIS- Daten

Der Aufbau des **Amtlichen Topographischen Kartographischen Informationssystems (ATKIS)** wurde im Oktober 1986 nach dem Beschluß der Arbeitsgemeinschaft der Vermessungsverwaltungen der Länder der Bundesrepublik Deutschland (AdV) begonnen, und kam in den meisten Bundesländern zum Abschluß mit der Ersterfassung der Daten für ATKIS, DLM 25/1 1996. [Kresse, 1997]

In ATKIS sollen die topographischen Informationen als digitale Daten vorgehalten werden und damit in zeitgemäßer Form die analogen Kartenwerke ablösen bzw. ergänzen [Rossipal, 1998]. ATKIS ist auch als Basisinformationssystem zu verstehen, d.h. die Anwender von raumbezogenen Daten sollen ATKIS-Daten als Basisdaten für ihre Anwendungen nutzen und falls erforderlich mit eigenen Daten anreichern [Walter, 1997].

Die Hauptziele von ATKIS als raumbezogenem Basisinformationssystem sind nach (AdV 1988) bei [Walter, 1997] und [Rossipal, 1998] folgendermaßen definiert:

- Den Benutzern soll eine einfache und zuverlässige Beschaffung digitaler topographischer Daten ermöglicht werden, mit denen sie ihre Fachdaten verknüpfen können.
- Es soll sichergestellt werden, daß authentische und aktuelle topographische Informationen eingesetzt werden.
- Es soll die Wirtschaftlichkeit der Informationsgewinnung, -verarbeitung und -ausgabe gesteigert werden.

ATKIS enthält generell zwei verschiedene Datenmodelle:

- Das **D**igitale **L**andschafts**m**odell (DLM) und
- Das **D**igitale **K**artographische **M**odell (DKM).

Im DLM wird die dreidimensionale Struktur der Erdoberfläche durch digitale topographische Daten beschrieben. Dabei wird die Landschaft in topographische Objekte gegliedert, die mit Namen, Lage und Eigenschaften versehen werden. In dieser Arbeit ist insbesondere das DLM 25 von Bedeutung, in dem die Genauigkeit der Straßen- und Gewässernetze mit einer Genauigkeit von $\pm 3\text{m}$ und die übrigen Objekte mit ± 10 bis 30m erfaßt sind. Ein reduziertes DLM 25/1 wurde

von den Landesvermessungsbehörden aufgebaut, in dem die Informationen in den Bereichen Verkehrsnetz, Gewässernetz und Vegetation geometrisch-topologisch vollständig beschrieben werden; der Objektbereich Siedlung wird dagegen reduziert und z.B. ohne Gebäude erfaßt [Hake & D.G., 1994]. Die reduzierte Form (DLM 25/1) ist für das gesamte Gebiet der Bundesrepublik Deutschland bereits vorhanden. Zum Vergleich der Inhalte der TK 25, des DLM und des DLM 25/1 kann man die Tabelle aus der Quelle [Heipke,1998] entnehmen. Kurze Beschreibungen von ATKIS, DLM, DKM usw. findet man in folgender Literatur: [Hake & D.G., 1994], [Rossipal, 1998] und [Walter, 1997].

6.1.2 Automatisch digitalisierte Daten

Allgemein werden bei der automatischen Digitalisierung die Kartenvorlagen gescannt und in ein digitales Bild umgewandelt. Anschließend wird das Rasterbild mit Verfahren der Raster-Vektor-Konvertierung in eine Datei von unstrukturierten Vektordaten konvertiert. Durch entsprechende Mustererkennungssoftware erfolgt dann eine automatische Strukturierung der Geometrie. Dies heißt, daß den Linien Bedeutungen zugewiesen werden. Nach Ablauf der Zuordnungen kontrolliert der Operateur die Ergebnisse und korrigiert gegebenenfalls interaktiv den Datenbestand [Rossipal, 1998].

Die in dieser Arbeit verwendeten automatisch digitalisierten Daten (im folgenden AdD) sind durch automatische Extraktion von Straßennetzen aus den obengenannten Luftbildern erzeugt worden. Zu der Extraktion des Straßennetzes sei auf folgende Quellen verwiesen. [Wiedemann, 1999a], [Wiedemann & S.H., 1999] und [Straub & C.W., 2000]

Die Genauigkeit der linienhaften Objekte wird mit 2m angenommen.

6.1.3 Manuell digitalisierte Daten

Die manuelle Digitalisierung kann in zwei Methoden durchgeführt werden, nämlich

- a) mit Hilfe eines Digitalisiertisches oder
- b) der Bildschirmdigitalisierung

Bei der ersten Methode werden von einer analogen, graphischen Vorlage, mit Hilfe eines Digitalisiertisches, diskrete Punkte direkt erfaßt. Bei dieser

Vorgehensweise werden die Objekte direkt gebildet und mit Attributen versehen, so daß kaum interaktive Nachbearbeitungen erforderlich sind. Diese Methode, die Zeit- und Kostenintensiv ist, wird insbesondere bei der Erfassung von komplexen Karteninhalten mit unregelmäßiger Geometrie, vielen Symbolen und unterschiedlichen Objektarten angewandt.

Bei der Bildschirmdigitalisierung werden die analogen Vorlagen mit einem Rasterscanner gescannt oder die Daten liegen schon in Rasterformat vor, wie es bei den genutzten Bildern der Fall ist. Anschließend werden die Daten am Bildschirm dargestellt. Die folgende Vektorisierung funktioniert im Prinzip wie oben beschrieben, mit dem Unterschied, daß sie am Bildschirm und nicht mit Papiervorlagen durchgeführt wird. Dabei können alle gängigen Methoden der Bildschirmdarstellung (z.B. Zoomen, Darstellen eines beliebigen Ausschnittes) genutzt werden.

Die angewandten manuell digitalisierten Daten (im folgenden MdD) sind am Institut für Photogrammetrie und Ingenieurvermessungen an der Universität Hannover aus digitalen Bildern erstellt worden. Die Genauigkeit der linienhaften Objekte beträgt 1m. Diese Daten werden als Referenzdaten angenommen.

6.2 Schematische Beschreibung der Anwendung

6.2.1 Überlegung zur angewendeten GIS-Software

Es wurde schon erwähnt, daß die Anwendung des stochastischen Modells am besten in einem hybriden System durchgeführt werden kann, weil die geprüften Daten im Vektorformat vorliegen und mit Luftbildern hinterlegt werden können (siehe 2.3). Außerdem ist die Modellierung der Genauigkeit in einem solchen Modell besser durchführbar. Deswegen benötigt man für diese Anwendung ein Softwarepaket, in dem mindestens folgende Funktionen vorhanden sein müssen:

1. Überlagerungsmöglichkeit von Vektor- und Rasterdaten
2. Vektor-Raster-Konvertierung und Umkehrung
3. Berechnung von großen Filtermatrizen (Binomialfilter)
4. mathematische und logische Operationen (Multiplikation zweier Bildern, UND-Verknüpfung, etc.)
5. Berechnung von Querprofilen

6. Elementare Funktionen der digitalen Bildverarbeitung

In dieser Arbeit fand die Anwendung mit dem Softwarepaket ERDAS (IMAGINE 8.3) statt. In diesem Paket kann nur mit ganz bestimmten Vektordaten gearbeitet werden, die im Arc/Info Format vorliegen [www2, Tutorial] und [www4, ERDAS].

6.2.2 Vorgehensweise der Arbeit

In diesem Abschnitt wird die Vorgehensweise der praktischen Arbeit bei der Anwendung des stochastischen Modells im hybriden GIS beschrieben, wie sie in der Abbildung 6.2 schrittweise (a bis c) dargestellt ist. Dabei werden einige wichtige Funktionen kurz erläutert.

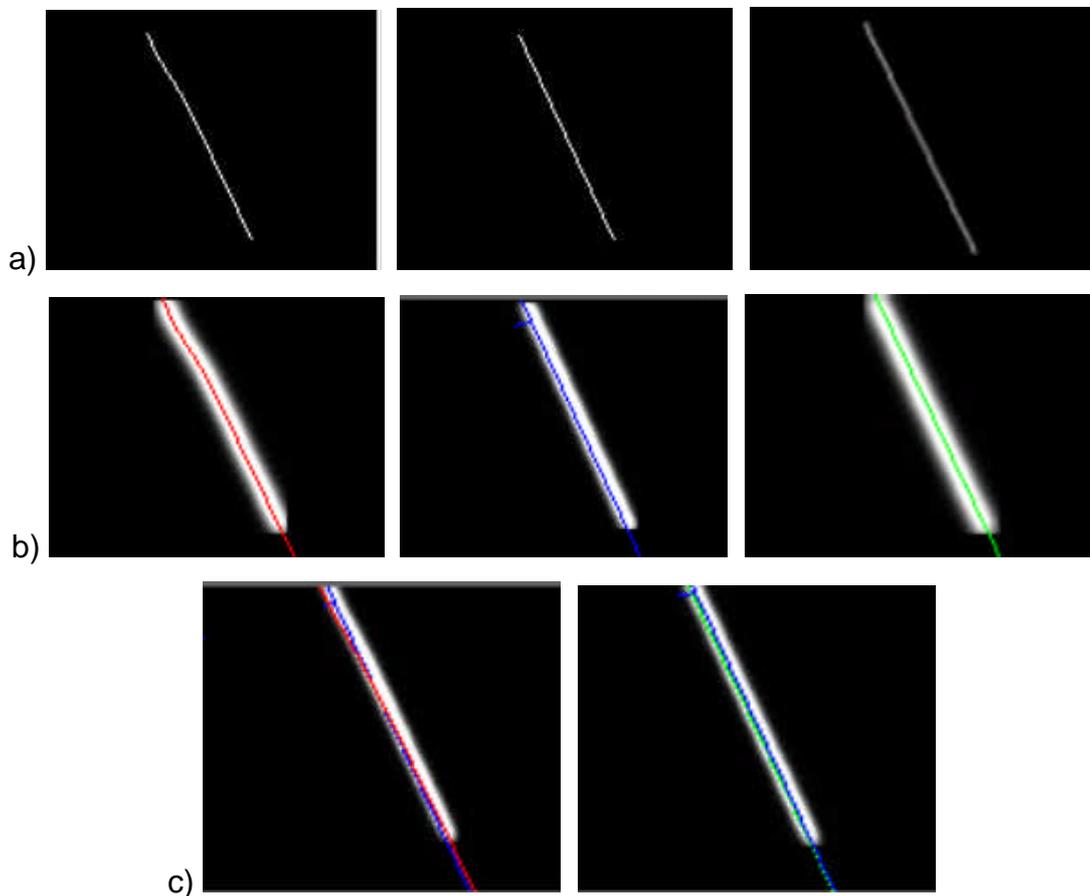


Abb. 6-2: Vorgehensweise der Anwendung, a) Vektorisierung, b) Glätten und
c) Multiplikation zweier Bildern

Wie in Abschnitt 6.1 kurz beschrieben wurde, liegen die drei Daten in Vektorformat vor. Um eine spätere Anwendung in Rasterformat durchzuführen, sollten diese

Daten umgewandelt werden. Dafür wird die Vektor-Raster-Konvertierung (im folgenden Rasterisierung) angewandt (Schritt a). Es ist von Interesse, daß die Rasterisierung objektweise oder abschnittsweise (stückweise) durchgeführt wird, so daß der spätere Vergleich objektweise gemacht werden kann. Außerdem werden der Arbeits- und Zeitaufwand im Vergleich zu einer vollständigen Rasterisierung geringer sein. Nachfolgend findet sich eine kurze Beschreibung der Vektor-Raster-Konvertierung (Schritt a).

Vektor-Raster-Konvertierung

Vektor-Raster-Konvertierung findet bekanntlich permanent Anwendung bei der Arbeit an einem Computerbildschirm, dabei wird der Bildschirm als eine Matrix angesehen, auf dem graphische Darstellungen von Vektorgrafik bearbeitet werden können. Die Umwandlung kann im einfachsten Fall durchgeführt werden, wenn eine Rastermatrix über das Gebiet, das die Vektorgrafik beinhaltet, gelegt wird, und danach jedes Pixel, das von einem Vektor geschnitten wird, einen Grau- oder Farbwert erhält. Wie es im folgenden zu sehen sein wird, erhalten die Objektpixel den maximalen Grauwert (255=weiß), während alle anderen Pixel, die zu keinem Objekt gehören, den Grauwert null erhalten (0=schwarz). Die Qualität der Umwandlung hängt von der Pixelgröße (Auflösung) der Rastermatrix ab. Die Entscheidung, auf welcher Pixelgröße die Umwandlung basieren soll, ist nach der Theorie aus der Signalverarbeitung zu treffen. Diese Theorie ist als NYQUIST- oder Abtasttheorem bekannt. Das Abtasttheorem besagt, daß sich eine ursprüngliche Funktion (Signal) ohne Informationsverlust wieder rekonstruieren läßt, wenn das Abtastintervall höchstens halb so groß ist, wie das kleinste noch aufzulösende Objekt (halbe Wellenlänge des Signals – Nyquistfrequenz).

Raster-Vektor-Konvertierung

Es gibt eine Vielzahl von Methoden zur Raster-Vektor-Konvertierung. Dabei besteht die einfachste Methode darin, die Randelemente des im Rasterformat vorliegenden Objektes mittels Vektoren zu beschreiben. Bei linienhaften Objekten wird die Konvertierung über Skelettbildung durchgeführt. Im Gegensatz dazu wird

bei flächenhaften Objekten ein Umringspolygon bestimmt [Bill, 1996] und [Hake & D.G., 1994].

Nachdem die beteiligten Objekte im Rasterformat vorliegen, sollte ihre Genauigkeit in einer Art und Weise modelliert werden, daß die Variation der Zufallsvariable der Normalverteilung (Gauß'sche Glockenkurve) entspricht (Schritt b). Eine gute Näherung für die Gauß'sche Glocken-Funktion auf einem diskreten Definitionsbereich stellen die Binomialfilter dar. Die Binomialkoeffizienten ermittelt man mit dem Pascal'schen Dreieck (eindimensionale Binomialverteilung) :

			1		
		1		1	
		1	2	1	
	1	3	3	1	
	1	4	6	4	1

Abb. 6-2 : Pascal'sche Dreieck

Zweidimensionale Binomialmasken erhält man durch Faltung zweier eindimensionaler Binomialmasken, z. B. eine 3x3- Operatormaske (Kernel) erfolgt durch:

$$h = \frac{1}{4} \begin{bmatrix} 1 & 2 & 1 \end{bmatrix} \otimes \frac{1}{4} \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} = \frac{1}{16} \begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix} \quad [\text{Kopp, 1997}] \text{ und } [\text{Heipke, 2000}].$$

Die Binomialfilter sind Glättungs- oder Tiefpaßfilter und haben in dieser Arbeit die Aufgabe, die Wahrscheinlichkeitsdichtefunktion des linienhaften Objektes zu bilden. Die Berechnung der Größe der Binomialfiltermatrix (Kernel) ist von der Varianz, dem Erwartungswert der Normalverteilung, einer Zufallsvariable und von der Abtastrate abhängig. Ein linienhaftes Objekt aus dem manuell digitalisierten Datensatz ist mit der Varianz $s_{MDD}^2 = 1m^2$ und Erwartungswert $\mu=0$ normal verteilt - der Erwartungswert entspricht der Linie selbst (mittlere Geometrie) -, nach der Berechnung der dazugehörenden Wahrscheinlichkeitsdichtefunktion dieser Linie ergibt sich ein Intervall zwischen $-4.08m$ und $+4.08m$, in dem sich die Realisierungen der Linie mit einer Wahrscheinlichkeit von 99.9% verteilen. Um das Abtasttheorem zu wahren muß das Objekt mit einer Pixelgröße von $\Delta=\sigma/2$

rasterisiert werden, für bessere Darstellung wäre eine kleinere Pixelgröße von $\Delta=\sigma/4$ denkbar (siehe [Glemser, 1999a]). In dieser Arbeit wurde die Rasterisierung mit einer Pixelgröße von $\Delta=\sigma/2=0.5\text{m}$ durchgeführt. Das führt dazu, daß die Breite des oben gerechneten Intervalls 18 Pixel entspricht. Da die eindimensionale Binomialmaske ungerade sein soll, nimmt man die nächst größere Binomialmaske. Infolgedessen ergibt sich der Kernel des entsprechenden Binomialfilters in diesem Fall von 19x19 Pixel. Bei den AdD hat die berechnete Wahrscheinlichkeitsdichtefunktion der Linien mit der Varianz $s_{AdD}^2 = (2m)^2 = 4m^2$ und dem Erwartungswert $\mu=0$ (siehe oben) einen Intervall mit der Breite $[-7.8\text{m}, +7.8\text{m}] \cong [32 \text{ Pixel}]$, so daß sich die Kernelgröße von 33x33 Pixel ergibt. Analog für ATKIS Daten mit der Varianz von $s_{ATKIS}^2 = (3m)^2$ ergibt sich eine Kernelgröße von 47x47 Pixel. Nach diesem Schritt liegen die Wahrscheinlichkeiten der linienhaften Objekte der drei Daten in Rasterformat vor (siehe Abb. 6-2b). Die Wahrscheinlichkeiten P_{MdD} , P_{AdD} und P_{ATKIS} können aus einem Querprofil berechnet werden, das senkrecht zu der Linie verlaufen sollte, wie es in folgender Abbildung zu sehen ist (die Linieninformationen sind nach wie vor in Vektordaten enthalten):

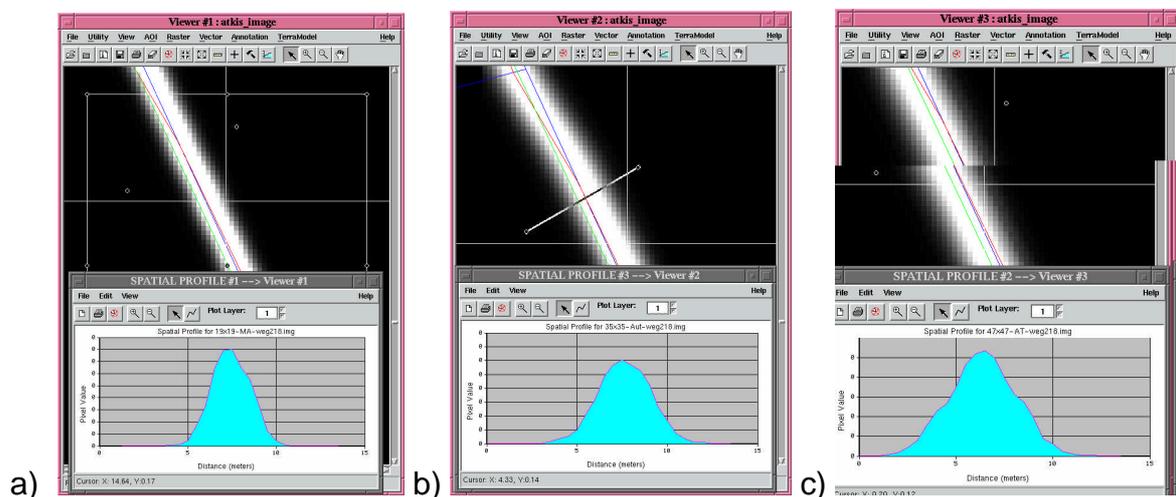


Abb. 6-3: Darstellung von Wahrscheinlichkeiten aus den Querprofilen für

a) MdD b) AdD und c) ATKIS- Daten

Die Darstellung von den Grauwerten der Querprofile haben die erwartete Form der Gauß'schen Glockenkurve. Je schlechter die Genauigkeit eines Objektes ist, desto breiter wird der Binomialfilter, wie es in der Abbildung 6-3 zu sehen ist.

Im nächsten Schritt (Schritt c) werden zwei Rasterbilder, die mit dem Binomialfilter geglättet wurden, miteinander multipliziert (siehe Abb. 6-2c), dabei werden ein Objekt A aus den MdD mit dem vermutlich gleichen Objekt A' aus den AdD bzw. A'' aus den ATKIS Daten verschnitten.

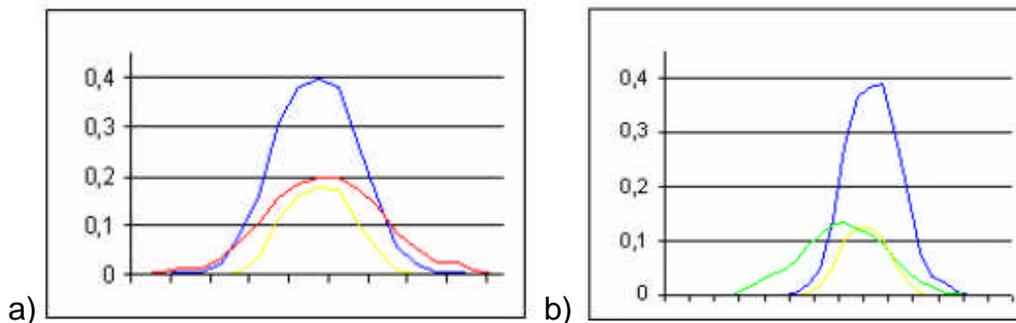


Abb. 6-4 : Querprofil der Verschnidung von : a) MdD mit AdD und
b) MdD mit ATKIS

Wie es in Kapitel 4 erwähnt wurde, entspricht diese Verschnidung der logischen UND-Operation, und es gilt nach dem Produktsatz der Wahrscheinlichkeit für unabhängige Ereignisse:

$$P(A \cap A') = P(A) \cdot P(A')$$

und analog

$$P(A \cap A'') = P(A) \cdot P(A'').$$

Das Ergebnis der Multiplikation ist ein Rasterbild, in dem nur diejenigen Grauwerte ungleich null existieren, an denen sich in beiden Bildern Objekte befinden. An den anderen Stellen löschen sich die Objekte aus (Multiplikation mit null ergibt immer null). Aus diesem neu entstandenen Bild sollten die Wahrscheinlichkeiten berechnet werden. Dafür sollen einige Querprofile aufgenommen und berechnet werden. Es wurde festgestellt, daß die berechneten Wahrscheinlichkeiten eines Schnitts P_{Schnitt} an verschiedenen Stellen unterschiedlich sind. Der Grund könnte darin liegen, daß die Abstände zwischen den beteiligten Linienobjekten unterschiedlich verlaufen. Um diese Annahme überprüfen zu können, legt man die ursprünglichen Vektordaten der beteiligten Linienobjekten über das Rasterbild.

Nun können Querprofile an verschiedenen Stellen mit verschiedenen Abständen d berechnet werden. Die Abstände können mit der Funktion „measure“ gemessen werden oder aus den Pixelkoordinaten berechnet werden. Die Annahme, daß zwischen den beteiligten Linienobjekte verschiedene Abstände vorliegen, wurde als richtig bestätigt. Ein Beispiel kann man aus der Abbildung 6-4 entnehmen: Mit einer Wahrscheinlichkeit von 18% ergibt sich ein Schnittpunkt ($d=0$) zwischen MdD und AdD (Abb. 6-4a) und bei der Verschneidung zwischen MdD und ATKIS-Daten mit dem Abstand zwischen den beiden Linienobjekten von 2 Pixel ($d=1m$) ergibt sich eine Wahrscheinlichkeit von 12%. Die Berechnung der Wahrscheinlichkeiten eines Schnitts in Abhängigkeit vom Abstand kann durch FFG (a-priori) oder durch Simulationen gemacht werden, wie es bei der Untersuchung von Flächenverschneidung mit Hilfe von Monte Carlo Simulation in der Arbeit von Glemser und Henneberg gezeigt wurde [Glemser & C.H., 1999c]. Die Varianz des Abstandes s_d ergibt sich aus der Liniengenauigkeit der beiden beteiligten Objekte, hier wird die Varianz für die zwei Fälle berechnet:

1. Fall für Verschneidung zwischen MdD und AdD:

$$s_{d1}^2 = s_{MdD}^2 + s_{AdD}^2 \text{ und}$$

2. Fall für Verschneidung zwischen MdD und ATKIS:

$$s_{d2}^2 = s_{MdD}^2 + s_{ATKIS}^2$$

Die Wahrscheinlichkeit, mit der die Zufallsvariable D die Werte d_1 und d_2 annimmt, wird über die entsprechenden Wahrscheinlichkeitsdichtefunktionen mit ihren Varianzen s_{d1}^2 und s_{d2}^2 und den Erwartungswerten $m_1 = m_2 = 0$ berechnet:

$$f(d_i) = \frac{1}{s_{di} \sqrt{2p}} e^{-\frac{(d_i - m)^2}{2s_{di}^2}}, \text{ mit } i=1,2.$$

Die Wahrscheinlichkeitsdichtefunktionen einer Zufallsvariablen ist um den Erwartungswert m symmetrisch (siehe Kapitel 4), dabei sollten die Abstände zwischen den Objekten nur als positiv betrachtet werden. Aus der folgenden Tabelle 6.1 kann man feststellen, daß die höchste Wahrscheinlichkeit eines Schnitts $P_{Schnitt}$ erwartungsgemäß vorhanden ist, wenn der Abstand d zwischen den beteiligten Objekten gleich null ist (Schnittpunkt der Linienobjekte). Sie beträgt bei der Verschneidung zwischen MdD und AdD 18%, - dieser Wert entspricht dem

berechneten Wert aus dem Querprofil (siehe oben) -, und bei der Verschneidung zwischen MdD und ATKIS 13%.

Abstand [in σ]	-6	-4	-3	-2	-1	0	1	2	3	4	6
Wahrsch. eines Schnitts bei s_{d1}	0	0,04	0,07	0,12	0,16	0,18	0,16	0,12	0,07	0,04	0
Wahrsch. eines Schnitts bei s_{d2}	0,02	0,06	0,08	0,10	0,12	0,13	0,12	0,10	0,08	0,06	0,02

Tab. 6-1: Wahrscheinlichkeiten eines Schnitts bei d1 und d2

Diese Abhängigkeit der Verschneidung vom Abstand führt dazu, daß man bei der Verschneidung nur die Objekten überprüfen soll, die sich innerhalb des in der Tabelle gegebenen Abstände befinden. So ist z.B. zu empfehlen, daß bei der Untersuchung der Verschneidung zwischen MdD und AdD nur die Objekte zu betrachten sind, deren Abstände kleiner als 4m sind. (Analog bei MdD und ATKIS, $d < 6m$). Infolgedessen gewinnt man eine wichtige Information, nämlich ob eine Schnittmenge überhaupt existiert. Das Ergebnis der Verschneidung ist eine Schnittmenge, in der die Wahrscheinlichkeiten durch die Grauwerte der beiden beteiligten Objekten berechnet werden können. Ist eine Schnittmenge vorhanden, kann man überprüfen, ob ein Punkt zu dieser Schnittmenge gehört. Das heißt die Untersuchung der Punktzugehörigkeit zur Schnittmenge erfolgt unter der Bedingung, daß ein Schnitt vorhanden sein sollte. Die Berechnung dafür ist in der folgenden Formel gegeben:

$$P'_{ZS} = \frac{P_A \cdot P_{A'}}{P_{Schnitt}}$$

und analog

$$P''_{ZS} = \frac{P_A \cdot P_{A''}}{P_{Schnitt}},$$

wobei die P_A , $P_{A'}$ und $P_{A''}$ aus den Grauwerte der Objekte, die mit dem Binomialfilter erzeugt worden sind, berechnet werden können. P'_{ZS} und P''_{ZS} sind die Wahrscheinlichkeiten, mit denen einen Punkt zur Schnittmenge der Verschneidung MdD mit AdD bzw. MdD mit ATKIS gehört. Diese sogenannte

bedingte Wahrscheinlichkeitsberechnung ist aus Anwendersicht erwünscht, da die Flächenverschneidung nur unter der Annahme eines vorhandenen Schnitts sinnvoll auszuführen ist. Im nächsten Abschnitt wird die praktische Anwendung des stochastischen Modells zur Feststellung einer Identität von Objekten durchgeführt.

6.2.3 Durchführung der Anwendung und die Ergebnisse

Die praktische Anwendung wird in zwei Fällen (Ausschnitte) durchgeführt, dabei sollten verschiedene Abstände betrachtet werden. Wie es in der Abbildung 6-5 zu sehen ist, beinhaltet das Gebiet eine Kreuzung und die damit verbundene Straße (Abb.6-5a). Dabei handelt es sich um zwei Ausschnitte, bei der einen befinden sich relativ kleine Abstände zwischen den Datensätzen (siehe Abb.6-5b), bei dem anderen Ausschnitt sind an manchen Stellen große Abstände zu erkennen (siehe Abb.6-5c).

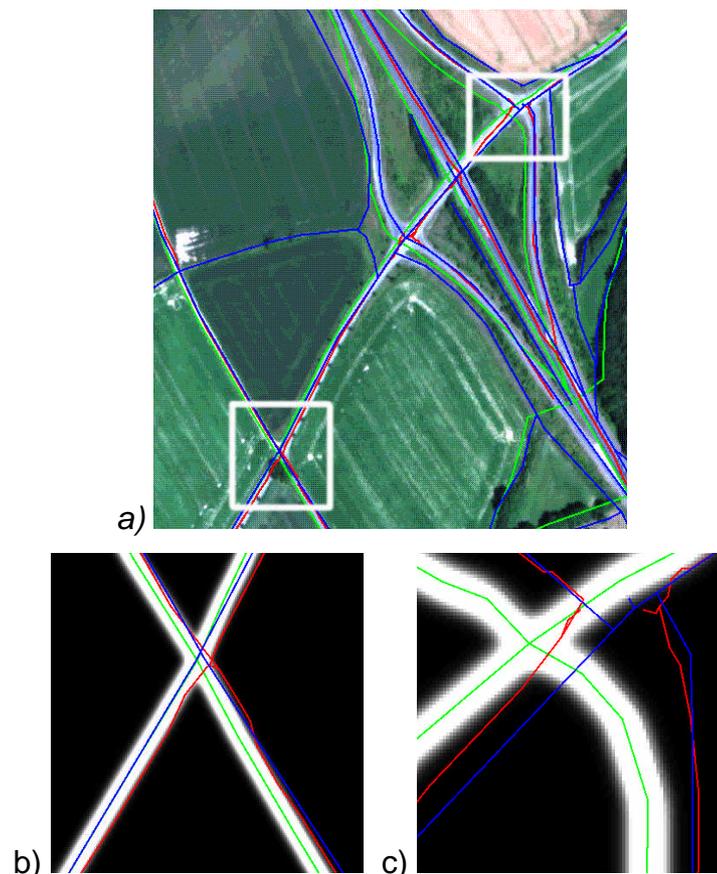


Abb. 6-5: Abstände zwischen den Objekte bei a) Ausschnitt einer Kreuzung
b) Straße und c) Kreuzungsteil

Die Querprofile werden in Abständen von $d=1\text{m}$ jeweils zwischen dem MdD und AdD bzw. ATKIS entlang der Linienobjekte aufgenommen. Die Ergebnisse, wie sie nach dem oberen Schema berechnet worden sind, werden in der folgenden Tabelle zusammengestellt:

Abstand in [m]	0	1	2	3	4	5
Wahrscheinlichkeit $P_A \cdot P_{A'}$	0,079	0,071	0,044	0,018	0,006	
P_{Schnitt} für MdD und AdD	0,178	0,162	0,119	0,072	0,036	
$P'_{\text{ZS}} = \frac{P_A \cdot P_{A'}}{P_{\text{Schnitt}}}$	0,444	0,438	0,370	0,248	0,167	
Abstand in [m]	0	1	2	3	4	5
Wahrscheinlichkeit $P_A \cdot P_{A''}$	0,053	0,048	0,038	0,017	0,008	0,003
P_{Schnitt} für MdD und ATKIS	0,126	0,120	0,103	0,080	0,056	0,036
$P''_{\text{ZS}} = \frac{P_A \cdot P_{A''}}{P_{\text{Schnitt}}}$	0,421	0,400	0,369	0,212	0,143	0,083

Tab. 6-2: Bedingte Wahrscheinlichkeiten der Zugehörigkeit zum Schnitt

Nun liegen die Wahrscheinlichkeiten der Punktzugehörigkeit P'_{ZS} und P''_{ZS} zu einem Schnitt vor. Da jeder Punkt der Schnittmenge zu den beiden Linienobjekten gehört, sollten seine Wahrscheinlichkeiten über die Identität der beiden beteiligten Linienobjekten festgestellt werden. Je größer die Schnittmenge ist, desto höher wird die Wahrscheinlichkeit, daß die beiden Linienobjekte identisch sind. Es war an manchen Stellen bei der Verschneidung kaum noch möglich einen Schnitt zu berechnen, demzufolge kann man keine Identität der Objekte feststellen. Dieser Fall trat mit der Verschneidung zwischen MdD und ATKIS-Daten häufiger im Vergleich zu der Verschneidung mit AdD auf. Ein Grund dafür könnte darin liegen, daß die Topologie in den ATKIS-Daten nicht mit der von MdD und AdD übereinstimmen, so daß die Abstände zwischen den Objekten (Abb. 6-5c) groß

sind. Diese Problematik beschränkt die Anwendung des stochastischen Modells zur Feststellung einer Identität zwischen den beteiligten Objekten. Die Behandlung der Genauigkeit von Objekten in Abhängigkeit von der Topologie ist noch Stand der Forschung, und ist in zahlreichen Arbeiten in der Literatur zu finden, hier sei auf folgenden Arbeiten verwiesen [Winter, 1996], [Walter, 1997], [Krauß, 1999] und [Gerke, 2000]. Anhand der Ergebnisse in der Tabelle 6-2 kann man dem Anwender die Entscheidung überlassen, ab welcher Wahrscheinlichkeitsgrenze eine Identität vorhanden sein könnte, z.B. könnte bei der Verschneidung zwischen MdD und AdD ab der Wahrscheinlichkeit von 25% eine Identität annehmbar sein.

7. Zusammenfassung

In dieser Arbeit wurden im Rahmen einer Literaturrecherche verschiedene Qualitätsaspekte von Daten in GIS dargestellt.

Nach einer kurze Einleitung in das Thema wurden in Kapitel 2 die Methoden zur Erfassung von Daten beschrieben, dabei war festzustellen, daß die Photogrammetrie und Fernerkundung eine große Menge von Daten liefern. Aus den Qualitätsaspekten der Geodaten (oder auch als Qualitätsmerkmale bezeichnet), die im Kapitel 3 detailliert beschrieben wurden, ist die geometrische Genauigkeit als das wichtigste und am meisten behandelte Qualitätsmerkmal ausgewählt, und wird in den beiden Kapiteln 5 und 6 näher betrachtet. Es werden in der Literatur verschieden Modelle zur Modellierung der geometrischen Genauigkeit vorgestellt, von denen die drei Modelle in Kapitel 5 beschrieben worden sind. Aus diesen drei Modelle wird das stochastische Modell ausgewählt, um die Frage nach der Identität von Objekten aus verschiedenen Datensätzen und unterschiedlichen Genauigkeiten zu behandeln. Die Entwicklung von GIS-Software macht diese Anwendung möglich, so daß einige wichtige statistische Berechnungen mit Hilfe von Rasterdaten und Grauwerte darstellbar sind. Die Berechnung und Visualisierung von Wahrscheinlichkeiten ist z.B. mit Hilfe von Operationen der digitalen Bildverarbeitung relativ leicht durchzuführen, wie z.B. bei dem Ansatz von Tiefpaßfiltern zur Darstellung von Wahrscheinlichkeiten der Dichtefunktion angezeigt wurde. Ein Voraussetzung für die Funktionalität dieses Modells ist damit verbunden, daß die Standardabweichungen der Objekte als wichtige Information vorhanden sein müssen. Hier wird vorgeschlagen, daß diese Informationen schon mit der Integration der Daten in einem GIS angegeben werden sollten. Die Entscheidung nach einer Identität von Objekten ist von der Verschneidung der beteiligten Objekte abhängig. In Kapitel 6 wurde von dem Autor dieser Arbeit versucht, diesen Sachverhalt mit Hilfe der Anwendung des stochastischen Modells anhand der Verschneidung von Linienobjekten zu bearbeiten. Es war festzustellen, daß die Verschneidung erwartungsgemäß von dem Abstand zwischen den Objekten abhängig ist. Deswegen ist dem Anwender zu empfehlen, bevor er die Verschneidung durchführt, beteiligte Objekte

übereinander zu legen und sich die Abstände zwischen ihnen anzuschauen. Man kann feststellen, daß die Topologie der Daten einen großen Einfluß auf die Abstände zwischen den Objekten hat, so daß man sich weitere Gedanken machen sollte, ob es noch sinnvoll ist, diese Anwendung durchzuführen. Meiner Meinung nach sollte die Topologie der beteiligten Objekte zu erst untersucht und anschließend das stochastische Modell angewendet werden.

Das stochastische Modell, wie es in dieser Arbeit beschrieben wurde, ist noch entwicklungsfähig und kann als eine Analysefunktion zur Behandlung der geometrische Genauigkeit in GIS eingesetzt werden, so kann z.B. der Verlauf der Anwendung automatisiert werden.

8. Literaturverzeichnis

[Albertz, 1991] Albertz, Jörg (1991):
"Grundlagen der Interpretation von Luft- und Satellitenbildern", Eine Einführung in die Fernerkundung, Berlin, 203 Seiten.

[Bartelme, 1995] Bartelme, Norbert (1995):
"Geoinformatik", Modelle, Strukturen, Funktionen, Springer-Verlag Berlin, Graz, 414 Seiten.

[Bethge, 1997] Bethge, Frank (1997):
Dissertation "Genauigkeit geometrischer Größen aus Vektordaten", Deutsche Geodätische Kommission (DGK), Münschen, Technische Universität Dresden, 92 Seiten.

[Bill, 1996] Bill, Ralf (1996):
"Grundlagen der Geo-Informationssysteme", Analysen, Anwendungen und neue Entwicklungen, Band 2, Wichmann Verlag, Heidelberg, 463 Seiten.

[Bill, 1997] Bill, Ralf (1997):
"Metainformationen als Basis interdisziplinärer GIS-Anwendungen", in: Zeitschrift für Photogrammetrie, Fernerkundung und Geoinformation, 5/97, Seite: 283-290.

[Bill & D. F., 1994] Bill, Ralf & Fritsch, Dieter (1994):
"Grundlagen der Geoinformationssystem", Hardware, Software und Daten, Band 1, Wichmann Verlag, Heidelberg, 415 Seiten.

[Bill & P.K., 1998a] Bill, Ralf & Korduan, Peter (1998a):
"Flächenverschneidung in GIS - Stochastische Modellierung und Effizienzbetrachtungen, Teil 1", in: Zeitschrift für Vermessungswesen (ZfV), Heft 8/98, Seite: 247-253.

[Bill & P.K., 1998b] Bill, Ralf & Korduan, Peter (1998b):

"Flächenverschneidung in GIS - Stochastische Modellierung und Effizienzbetrachtungen, Teil 2", in: Zeitschrift für Vermessungswesen (ZfV), Heft 10/98, Seite: 333-338.

[Bosch, 1998] Bosch, Karl (1998):
"Statistik-Taschenbuch", Dritte Auflage, R. Oldenbourg Verlag, München, 814 Seiten.

[Carver & C.B, 1992] Carver, Steve & Brunson, Chris (1992):
"Accuracy and Sampling Strategy in 2D and 3D GIS Databases", Egis 92, Third European Conference and Exhibition on Geographical Information System, Munich-Germany, March 23-26, 1992, Seite: 244-253.

[Caspary, 1992a] Caspary, Willi (1992a):
Genauigkeit als Qualitätsmerkmal digitaler Datenbestände, Deutsche Verein für Vermessungswesen (DVW) e.V., Heft 4/92, Seite: 157-166.

[Caspary, 1992b] Caspary, Willi (1992b):
"Qualitätsmerkmale von Geo-daten", in: Zeitschrift für Vermessungswesen (ZfV), Heft 7/92, Seite: 360-367.

[Caspary, 1993] Caspary, Willi (1993):
"Qualitätsaspekte bei Geoinformationssystemen", in: Zeitschrift für Vermessungswesen (ZfV), Heft 8,9/93, Seite: 444-450.

[Caspary & R.S., 1992c] Caspary, Wilhelm & Scheuring, Robert (1992):
"Error-Band as measures of geometrical accuracy", Proceeding of EGIS'92, Utrecht, Vol. 1, Seite: 226-233.

[DIN, 1995] Deutsche Institut Für Normung (DIN) (1995):
"Qualitätsmanagement- Begriffe", DIN EN ISO 8402, Berlin.

[Fritsch et al., 1998] Fritsch, Dieter & Glemser, Michael & Klein, Ulrike & Sester, Monika & Strunz, Günter (1998):

"Zur Integration von Unsicherheit bei Vektor- und Rasterdaten", in: Geo-Informationssystem (GIS), Heft 4/98, Seite: 26-35.

[Gerke, 2000] Gerke, Markus (2000):

Diplomarbeit, "Topologische und geometrische Analyse zum Vergleich ungenauen Flächen", Institut für Photogrammetrie und Ingenieurvermessungen, Universität Hannover.

[Glemser, 1996] Glemser, Michael (1996):

"Integration geometrischer Datenqualität in GIS-Funktionen", in: Heft5, Datenqualität und Metainformation in GIS - Tagungsband zum Workshop vom 7./8. Oktober 1996, Institut für Geodäsie und Geoinformatik / Universität Rostock, Seite: 69-84.

[Glemser & C.H., 1999c] Glemser, Michael & Henneberg, Christian (1999c):

"Untersuchung zu den Fortpflanzungseigenschaften der geometrischen Genauigkeit bei der Flächenverschneidung", in: M. Sester & F. Krumm, Hrsg., GIS-Forschung im Studiengang Geodäsie und Geoinformatik, Schriftenreihe der Institute des Fachbereichs Vermessungswesen, Universität Stuttgart, Seite: 44-56.

[Glemser & U.K., 1999a] Glemser, Michael & Klein, Ulrike (1999a):

"Datenunsicherheit in hybriden Geoinformationssystem", in: C. Rautenstrauch, & M. Schenk, Hrsg., Umweltinformatik zwischen Theorie und Industrieanwendung, 13. Internationales Symposium "Informatik für den Umweltschutz" der Gesellschaft für Informatik (GI), Magdeburg 1999, Seite: 457-471.

[Glemser & U.K., 1999b] Glemser, Michael & Klein, Ulrike (1999b):

"Datenqualität in hybriden Geoinformationssystemen", in: Mitteilung des Bundesamtes für Kartographie und Geodäsie, Arbeitsgruppe Automation in der Kartographie Tagung 1998, Band 2, Seite: 17-24.

[Glemser & U.K.,1999d] Glemser, Michael & Klein, Ulrike (1999d):
"Hybride Modellierung und Analyse von unsicheren Daten", in: GIS-Forschung im Studiengang Geodäsie und Geoinformatik, Schriftenreihe der Institute des Fachbereichs Vermessungswesen, Universität Stuttgart, Seite: 27-43.

[Glemser et al., 1999e] Glemser, Michael & Krauß, Sandra & Schneider, Christian (1999e):
"Das Minimum-Maximum-Modell - Eine qualitative Beschreibung der geometrischen Genauigkeit und ihre Integration in GIS", in: M. Sester & F. Krumm, Hrsg., GIS-Forschung im Studiengang Geodäsie und Geoinformatik, Schriftenreihe der Institute des Fachbereichs Vermessungswesen, Universität Stuttgart, Seite: 57-69.

[Goodchild, 1995] Goodchild, Michael F. (1995):
"Attribute accuracy", in: Elements of Spatial Data Quality, Chapter 4, Seite: 60-79.

[Göpfert, 1991] Göpfert, Wolfgang (1991):
"Raumbezogene Informationssysteme", Grundlagen der integrierten Verarbeitung von Punkt-, Vektor-, und Rasterdaten, Anwendungen in kartographie, Fernerkundung und Umweltplanung, 2. Auflage, Wichmann Verlag, Karlsruhe, 318 Seiten.

[Grünreich, 1996] Grünreich, Dietmar (1996):
Arbeitspapier, "GIS- Topographie / Hydrographie", Institut für Kartographie, Universität Hannover, Seite: 1-8.

[Haberäcker, 1991] Haberäcker, Peter (1991):
"Digitale Bildverarbeitung", Grundlagen und Anwendung, 4. Auflage, Carl Hanser Verlag, München, 404 Seiten.

[Hake & D.G., 1994] Hake, Günter & Grünreich, Dietmar (1994):
"Kartographie", 7. Auflage, Walter de Gruyter, Berlin , New York, 599 Seiten.

[Heipke, 1998] Heipke, Cristian (1998):

"Stand und Perspektiven der digitalen Photogrammetrie zur Erfassung und Aktualisierung von Geodaten", in: VGI, München, Band 1/98, Zusammengestellt nach Vortragsunterlagen von G. Brandstätter, TU Graz, Seite: 17-22.

[Heipke, 2000] Heipke, Christian (2000):
Vorlesungsskript 3. Semester, "Photogrammetrie I", Digitale Bildverarbeitung, Hannover, Institut für Photogrammetrie und Ingenieurvermessungen Universität Hannover.

[Höpcke, 1980] Höpcke, Walter (1980):
"Fehlerlehre und Ausgleichung", Walter de Gruyter- Berlin. New York, 227 Seiten.

[Kopp, 1997] Kopp, Herbert (1997):
"Bildverarbeitung interaktiv", eine Einführung mit multimedialem Lernsystem auf CD-ROM, B. G. Teubner, Stuttgart, 215 Seiten.

[Kraus & H.K., 1993a] Kraus, Karl & Kager, H. (1993a):
"Genauigkeiten abgeleiteter Daten in einem Informationssystem", in: Mitteilung der Geodätischen Institut TU Graz, Folge 78, Seite: 95-101.

[Kraus & K.H., 1993b] Kraus, Karl & Haussteiner, Karl (1993b):
"Visualisierung der Genauigkeit geometrischer Daten", in: Geo-Informationssystem (GIS), Heft 3/93, Seite 7-12.

[Kraus & M.L., 1998] Kraus, Karl & Ludwig, M. (1998):
"Genauigkeit der Verschneidung geometrischer Geodaten", in: Zeitschrift für Vermessungswesen (ZfV), Heft 3/98, Seite: 81-87.

[Krauß, 1999] Krauß, Sandra (1999):
Diplomarbeit, "Qualitative Beschreibung von unsicheren topologischen Relationen innerhalb des Minimum-/Maximum-Modells", Stuttgart, Universität Stuttgart, 70 Seiten.

[Kresse, 1997] Kresse, Wolfgang (1997):
"Visuelle Überlagerung von ALK-GIAP-Vektordaten und stereoluftbildpaar zur Verbesserung der geometrischen Genauigkeit und Fortführung der GIS-Daten", in: Heft5, Datenqualität und Metainformation in GIS - Tagungsband zum Workshop vom 7./8. Oktober 1996, Institut für Geodäsie und Geoinformatik / Universität Rostock, Seite: 111-115.

[Kreyszig, 1973] Kreyszig, Erwin (1973):
"Statistischen Methoden und ihre Anwendung", 4.Auflage, Vandenhoeck & Ruprecht in Göttingen.

[Lother, 1999a] Lother, Georg (1999a):
"Qualitätssicherung für GIS-Daten am Beispiel des Geoinformationssystems der Bayerischen Staatsforstverwaltung", in: Zeitschrift für Vermessungswesen (ZfV), Heft 9/99, Seite: 281-289.

[Lother, 1999b] Lother, Georg (1999b):
"Geo-Informationssysteme- Begriffe und Kategorien", in: Geo-Informationssystem (GIS), Heft 6/99, Seite: 20-27.

[Pelzer, 1985] Pelzer, Heinz (1985):
"Geodätische Netze in der Landes- und Ingenieurvermessung II", Wittwer Verlag.

[Reißmann, 1980] Reißmann, Günter (1980):
"Die Ausgleichsrechnung", Grundlagen und Anwendung in der Geodäsie, 5. Auflage, VEB Verlag für Bauwesen, Berlin.

[Roschlaub, 1996] Roschlaub, Robert (1996):
"Geometrische Datenqualität und Klassifikation von Geodaten", in: Heft5, Datenqualität und Metainformation in GIS - Tagungsband zum Workshop vom 7./8. Oktober 1996, Rostock, Institut für Geodäsie und Geoinformatik / Universität Rostock, Seite: 85-94.

[Rossipal, 1998] Rossipal, Silke (1998):
Dissertation, "Untersuchungen zur geometrischen Datenkonsistenz von ATKIS-Bayern", Deutsche Geodätische Kommission (DGK), Münschen, Universität Stuttgart, 98 Seiten.

[S.A & S.Q., 1992] S.A. John & S.Q. Lewis (1992):
"Compiling a GIS Databas the quastion of Data quality", proceedings of EGIS'92, Munich-Germany, March, 23-26, 1992, Seite: 521-531.

[Scheu, 1996] Scheu, Martin (1996):
Dissertation, "Erfassung und Rekonstruktion raumbezogener Objekte in Vektorform", Deutsche Geodätische Kommission (DGK), Münschen, Technische Universität Berlin, 136 Seiten.

[Schneider, 1999] Schneider, Christian (1999):
Diplomarbeit, "Integration der geometrischen Genauigkeit in MapInfo nach dem Minimum/Maximum-Model", Stuttgart, Universität Stuttgart, 61 Seiten.

[Schumann, 2000] Schumann, Jens (2000):
Diplomarbeit, "Einsatz von standardisierten Analysenverfahren zur Beurteilung der Informationsqualität geodätischer Messungen", Stuttgart, Universität Stuttgart, 68 Seiten.

[Shi & W.L., 2000] Shi, Wenzhong & Liu, Wenbao (2000):
"A stochastic process-based model for the positional error of line segments in GIS", in: International Journal of Geografical Informaion Sience, Vol. 14/1, Seite: 51-66.

[Straub & C.W., 2000] Straub, Bernd - M. & Wiedemann, Christian (2000):
"Towards the Update of Geo Data by Automatic objkt Extraction", EOS/SPIE Symposium on Remote Sensing, Barcelona, Spain, September, Seite: 25-29.

[Walter, 1997]

Walter, Volker (1997):

Dissertation, "Zuordnung von raumbezogenen Daten am Beispiel der Datenmodelle ATKIS und GDF", Deutsche Geodätische Kommission (DGK), Münschen, Universität Stuttgart, 130 Seiten.

[Wiedemann, 1999]

Wiedemann, Christian (1999):

"Completion of automatically extracted road networks based on the function of roads", in: International Archives of Photogrammetry and Remote Sensing, 32(3-2w5), Seite: 209-215.

[Wiedemann & S.H.,1999] Wiedemann, Christian & Hinz, Stefan (1999):

"Automatic Extraction and Evaluation of Road Networks from Satellite Imagery", in: International Archives of Photogrammetry and Remote Sensing, 32(3-2w5), Seite: 95-100.

[Winter, 1996]

Winter, Stephan (1996):

Dissertation, "Unsichere topologische Beziehungen zwischen ungenauen Flächen", Deutsche Geodätische Kommission (DGK), Münschen, Universität Bonn, 66 Seiten.

Internet Adressen (Stand Februar 2001)

[www1, GIS]

Dr. Thomas Ott

Geographische Informationssystemen, Ein Überblick.

GIS-Begriffe und Bedeutung.

<http://www.oaqis.com/gis3.htm>

[www2, Tutorial]

Blaschke, T. und Lang, S.

Tutorial: Einführung in Fernerkundung und digitale Bildverarbeitung.

<http://www.sbg.ac.at/geo/student/fernerkundung/>

[www3, DIN-CEN-ISO]

Helmholz, P., Richter, S. und Leitloff, J.

Normung von Geodaten.

http://www.igg.tu-berlin.de/~helmholz/vw/DIN_Cen_ISO/Vortrag.htm#einleitung

[www4, ERDAS]

<http://www.erdas.com>