Datenunsicherheit in hybriden Geoinformationssystemen

Michael Glemser[1] und Ulrike Klein[1]

Abstract

An essential requirement for a comprehensive use of hybrid data is the consideration and processing of its uncertainty. Erroneous interpretations of analyses can be avoided if uncertainty is integrated as a mandatory component, stored and considered in all operations. In this contribution, a probabilistic approach is presented for modelling geometrical and thematic uncertainty. The used hybrid data model is extended to manage data uncertainty.

1 Einführung

Digitale raumbezogene Datenbestände, wie sie in zukünftigen interoperablen geowissenschaftlichen Informationssystemen verwaltet und analysiert werden sollen, beinhalten alle einen gewissen Grad an Unsicherheit in ihren Daten. Der Ursprung dieser Unsicherheit liegt zum einen in der Erfassung selbst als auch in der anwendungsspezifischen Modellbildung begründet, die eine mehr oder weniger exakte Abstraktion der Phänomene der realen Welt bewirkt. Bislang blieb diese Eigenschaft im wesentlichen unberücksichtigt. Im Zuge der Nutzung dieser Daten innerhalb von interoperablen Anwendungen, in denen komplexe Analyseprozesse mit einer Vielzahl an Daten unterschiedlicher Herkunft, Thematik und Qualität gemeinsam analysiert werden, kommt der Datenunsicherheit eine gestiegene Bedeutung zu. Zur Berücksichtigung dieser Belange bedarf es der Modellierung der Unsicherheit, der Integration dieser Modelle in die einzelnen Systeme und der Erweiterung der darin verfügbaren Funktionalität, so daß auch am Ende von Analyseprozessen eine Beurteilung hinsichtlich der Unsicherheit des Ergebnisses möglich wird. Die Berücksichtigung ist als Erweiterung der bisherigen Arbeitsweise zu sehen, die ohne zusätzliche Nutzerinteraktionen automatisch ablaufen soll, um damit die Akzeptanz zu gewährleisten.

[1] Universität Stuttgart, Institut für Photogrammetrie, Geschwister-Scholl-Straße 24, D-70174 Stuttgart, email: {Michael.Glemser|Ulrike.Klein}@ifp.uni-stuttgart.de,
Internet: http://www.uni-stuttgart.de

2 Geometrische und thematische Unsicherheit

Zur Strukturierung von räumlichen Daten stützen sich heutige GIS auf das Objektmodell. Die Grundidee besteht darin, aus allen relevanten räumlichen Phänomenen Objekte zu bilden. Ein einzelnes Objekt setzt sich im wesentlichen aus den beiden Komponenten Geometrie und Thematik zusammen (Bill/Fritsch 1991). Zusätzlich kann es mit einer graphischen Beschreibung verbunden werden, die die äußere Darstellung mit Hilfe von graphischen Attributen (Farben, Strichstärken, Symbolen, ...) regelt. Gleichartige Objekte werden zumeist zu abstrakten Objektklassen zusammengefügt. Der Raumbezug als charakteristische Eigenschaft von räumlichen Objekten ist in der Geometriekomponente enthalten, während alle beschreibenden Informationen in Form von thematischen Attributen zur Thematikkomponente zählen. Thematische Attribute besitzen entweder diskrete oder kontinuierliche Werte. Entsprechend können sie der nominalen oder ordinalen Skala (bei diskreten Werten) bzw. der Intervall- oder Ratio-Skala (bei kontinuierlichen Werten) zugeordnet werden.

Die Untersuchung und Beschreibung der in den Daten enthaltenen Unsicherheit wird im folgenden getrennt nach den beiden Komponenten durchgeführt. Das notwendige Zusammenfügen der dabei entwickelten Konzepte erfolgt dann im nachfolgenden Kapitel durch die Definition eines Unsicherheitsmodells für hybride Daten.

2.1 Geometrische Unsicherheit

In GIS kommen zwei grundlegend unterschiedliche Geometriearten vor, die gegensätzliche Eigenschaften aufweisen: Raster- bzw. Vektorgeometrie. Die Rastergeometrie unterteilt den Raum vollständig in gleichförmige Zellen. Eine oder mehrere zumeist zusammenhängende Zellen definieren dann die Geometrie eines Objektes. Prinzipiell sind mehrere Zellformen (z.B. Quadrat, Dreieck, ...) möglich, wobei in GIS die quadratische Form üblich ist. Die Unterteilung erfolgt i.d.R. strikt nach den Koordinatenachsen und wird durch Definition von Zellgröße (Auflösung) und Ursprung des Rasters festgelegt. Das zugehörige geometrische Modell wird als Enumerationsverfahren bezeichnet. Im Gegensatz dazu setzt die Vektorgeometrie eine Reihe von geometrischen Grundprimitiven (Punkt, Linie, Fläche) ein, um die Ausdehnung des Objektes zu beschreiben. Mehrere aus dem CAD-Bereich stammende Modelle sind hierbei möglich, wobei sich die meisten Systemrealisierungen auf die Randdarstellung stützen. Darin repräsentiert die Geometrie den Rand des Objektes (z.B. Umringslinien bei einem flächenhaften Objekt). Die Form eines Objektes kann dadurch beliebig variabel gehalten werden. Die Lage ist durch individuell zu bestimmende Koordinaten der Primitive festlegt.

Die geometrische Unsicherheit besitzt für die beiden Geometriearten ganz unterschiedliche Bedeutung. Im Falle der Rastergeometrie liegen Form und Lage der Zellen per Definition fest. Es ist also von einer sicheren Geometrie auszugehen. Unsicherheit in den Daten besteht hier lediglich in der zugeordneten Thematik (z.B.

aufgrund von Mischnutzungen). Einen erheblichen Anteil an geometrischer Unsicherheit kann dagegen im Falle der Vektorgeometrie in den Daten vorliegen. Er wird in erster Linie durch den Erfassungsschritt hervorgerufen. Die Ursachen dafür liegen sowohl in den Meßverfahren, die nur eine bestimmte Punktmeßgenauigkeit gewährleisten, als auch in der Modellierung der Objekte selbst. Die zweite Ursache wird offensichtlich, indem man die Repräsentation eines Objektes in GIS mit seinem Original in der Natur vergleicht. Im GIS-Modell ist die Beschreibung in diskreter Form gefordert, die ein Objekt scharf von seinen Nachbarobjekten abgrenzt. Eine scharfe Grenze wird aber besonders für natürliche Objekte nicht anzutreffen sein. Vielmehr besitzen Objekte häufig unscharfe Grenzen, die ausgedehnten Übergangszonen entsprechen, in denen Lage und Form nur unsicher festzustellen sind. In vielen Fällen wird sich diese Ursache in dominanter Weise gegenüber anderen Ursachen äußern.

Um die geometrische Unsicherheit in GIS mitzuverarbeiten, wird eine Beschreibung durch ein Modell benötigt. Bei der Messung werden Punktkoordinaten beobachtet, deren Unsicherheiten sich mit den Methoden der mathematischen Statistik ausdrücken lassen. Jede Koordinate ist dabei als kontinuierliche Zufallsvariable aufzufassen, die durch eine Verteilung charakterisiert werden kann (z.B. Bill/Korduan 1998). Man geht i.d.R. vom Vorliegen einer Normalverteilung aus. Damit werden alle Koordinaten (x, y, z) durch

- Mittelwerte (μ_x, μ_y, μ_z) und
- Varianzen ($\sigma_x^2, \sigma_y^2, \sigma_z^2$)

beschrieben. Unter Einbeziehung des Modellierungseinflusses (unscharfe Grenze) ist der Modellansatz dahingehend zu erweitern, daß alle zum Rand gehörenden Punkte stochastische Größen darstellen, unabhängig davon, ob deren Koordinaten beobachtet (bei Anfangs- oder Endpunkten von Linien) oder durch Definition (bei beliebigen Zwischenpunkten) festgelegt sind. Die Variation wirkt jeweils senkrecht zur Richtung der Linie. Gemessene Punkte stellen in diesem Sinne keine herausragenden Punkte mehr dar, sondern alle Punkte werden in gleicher Weise behandelt. Damit wird der allgemeine Fall abgedeckt, daß die Unsicherheit des Objektes sowohl von Unschärfe im Modell als auch von der Beobachtungsgenauigkeit bestimmt wird.

Für jeden Punkt der Objektgrenze sind somit Mittelwert und Varianz anzugeben. Mittelwerte sind durch die Objektgeometrie festgelegt, während die Varianz bestimmt werden muß. Da eine individuelle Angabe für jeden Punkt (bei unendlich vielen Randpunkten) unmöglich ist, sollten gemeinsame Werte für Liniensegmente oder für das ganze Objekt gefunden werden.

Als Alternative zu Mittelwert und Varianz bietet sich die Möglichkeit, auf Wahrscheinlichkeitswerte überzugehen (Kraus/Haussteiner 1993). Als Aussage ist die Zugehörigkeit eines beliebigen Punktes zum Objekt zu bewerten, d.h. es wird ausgedrückt, mit welcher Wahrscheinlichkeit man sich an einer beliebigen Stelle innerhalb des Objektes befindet. Die Berechnung der Wahrscheinlichkeiten stützt sich direkt auf Mittelwert und Varianz. Genaue Formeln für unterschiedliche Objekttypen (punkt-, linien-, flächenförmiges Objekt) können aus Glemser und Klein (1999) entnommen werden. Bewertet man alle möglichen Stellen, so entsteht eine räumliche Wahrscheinlichkeitsdichte. Sie entspricht einer räumlich stetigen Funktion, die zur besseren Verarbeitung in eine diskrete Form überzuführen ist. Dafür eignet sich ein Raster, das sich dann als Wahrscheinlichkeitsraster bezeichnen läßt. Die Rastergröße r sollte mindestens so klein sein, daß eine Rekonstruktion der Originalfunktion möglich bleibt. Sie ist damit abhängig von der Standardabweichung σ der Punkte und kann zu

$$r = 0.25\ \sigma$$

abgeschätzt werden. Ein wesentlicher Vorteil des Übergangs auf Wahrscheinlichkeiten äußert sich in der räumlichen Verteilung der Unsicherheit, die eine besonders anschauliche Darstellung ermöglicht. Dazu sind die Wahrscheinlichkeiten lediglich in Grauwerte zu transformieren. Abbildung 1 zeigt Beispiele zur Darstellung der Wahrscheinlichkeitsraster für verschiedene Objekttypen.

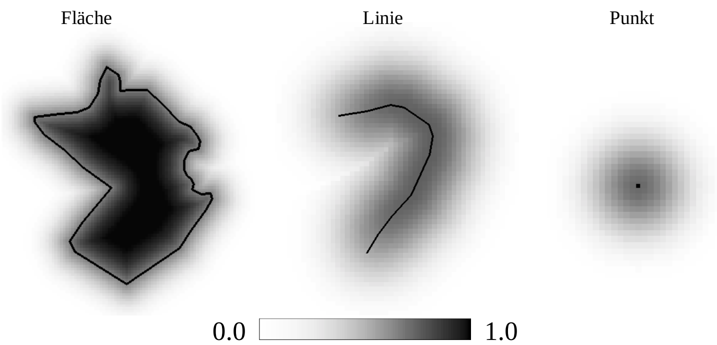


Abbildung 1
Beispiele zur Darstellung der Wahrscheinlichkeiten

Grundsätzlich ist die Rastergeometrie als frei von geometrischer Unsicherheit zu betrachten. Ein Sonderfall tritt ein, wenn die geometrische Lage und Form des Rasters aus einer geometrischen Entzerrung hervorgegangen sind, wie sie im Falle von Fernerkundungsdaten häufig als Vorverarbeitungsschritt (Richards 1993) durchgeführt wird. Eine Rasterzelle weist dabei meist zu einem bestimmten Betrag Restklaffungen zur Wirklichkeit auf, d.h. die zugewiesene Thematik bezieht sich nicht auf die aktuelle sondern auf eine zufällige Position der Zelle. Die Restklaffungen sind also als geometrische Unsicherheit zu werten. Jedoch können die Wirkungen der Restklaffungen auch als thematische Effekte interpretiert werden. Dazu ist die Rastergeometrie wiederum als fest anzusehen, wogegen die Thematik an der nun festen Position eine Abweichung zur Wirklichkeit aufweisen kann. In diesem Sinn entsprechen geometrische Restklaffungen thematischen Abweichungen zur Wirklichkeit und können als Teil der thematischen Unsicherheit angesehen werden.

2.2 Thematische Unsicherheit

Attribute sind aufgrund der Ausprägung ihrer Wertebereiche zu unterteilen in Attribute, die diskrete Werte annehmen können, und in Attribute, die mit kontinuierlichen Werten beschrieben werden. Entsprechend unterschiedlich ist die Vorgehensweise zur Modellierung der Unsicherheit der Attribute mit Wahrscheinlichkeiten.

(a) (b)

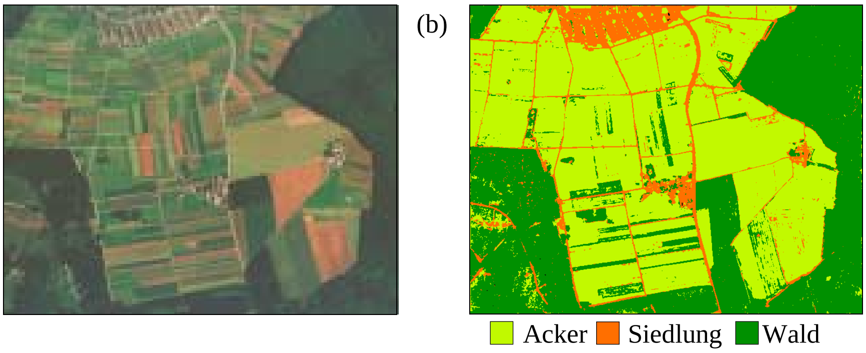


Abbildung 2
Testgebiet (a) und Ergebnis der Klassifizierung (b)

Im Bereich der Rasterdaten ist die Fernerkundung ein wichtiger Lieferant von Daten, deren Attribute mit diskreten Werten beschrieben werden. Die derzeitige Zunahme der satellitengestützten Fernerkundungssensoren führt dazu, daß in vermehrtem Maße Anwendungsmethoden für den Umweltschutz, die öffentliche Verwaltung und den privaten Sektor entwickelt werden. In diesem Zusammenhang stellt natürlich auch die Beurteilung der Datenqualität einen wichtigen Aspekt dar. Die Auswertung von Fernerkundungsdaten beinhaltet in einem ersten Schritt die Definition von Klassen, z.B. von Landnutzungsklassen. Dabei besteht das Problem, aus der Vielzahl an möglichen Objektarten eine begrenzte Anzahl von Klassen festzulegen. Im zweiten Schritt ist die Zugehörigkeit der flächenhaft erfaßten Ausschnitte der Erdoberfläche zu den Klassen zu bestimmen. Durch Klassifizierungsverfahren werden Maßzahlen für die Zugehörigkeiten der einzelnen Bildelemente zu den definierten Klassen berechnet; im Falle der Maximum-Likelihood-Methode (Richards 1993) handelt es sich bei diesen Maßzahlen um Wahrscheinlichkeiten. Herkömmlicherweise wird mit Hilfe eines Entscheidungskriteriums jedem Bildelement eine einzige Klasse zugewiesen und ausschließlich dies als Ergebnis der Klassifizierung festgehalten (Abbildung 2).

Im Hinblick auf die Beschreibung der thematischen Unsicherheit ist diese Vorgehensweise mit einem Informationsverlust verbunden, da im nachhinein der Grad der Zugehörigkeit, der als Unsicherheit der zugewiesenen Objektklasse anzusehen ist, nicht mehr beurteilt werden kann. Es bietet sich daher alternativ an, die Wahrscheinlichkeiten pro Bildelement als Information zur Genauigkeitsbeschreibung in einem Vektor abzuspeichern. Prinzipiell liegen so viele Wahrscheinlichkeiten pro Bildelement vor, wie Klassen definiert werden. Untersuchungen haben jedoch gezeigt, daß lediglich die größten und zweitgrößten Wahrscheinlichkeiten signifikante Werte annehmen (Fritsch et al. 1998). Zur Veranschaulichung sind in Abbildung 3 die jeweils pro Bildelement größten und zweitgrößten Wahrscheinlichkeiten, die aus der Maximum-Likelihood-Klassifizierung des in Abbildung 2 dargestellten Testgebietes resultieren, durch Grauwerte dargestellt. Die dunklen Bereiche zeigen eine hohe Wahrscheinlichkeit an, während die hellen Regionen einen hohen Grad an Unsicherheit aufweisen.

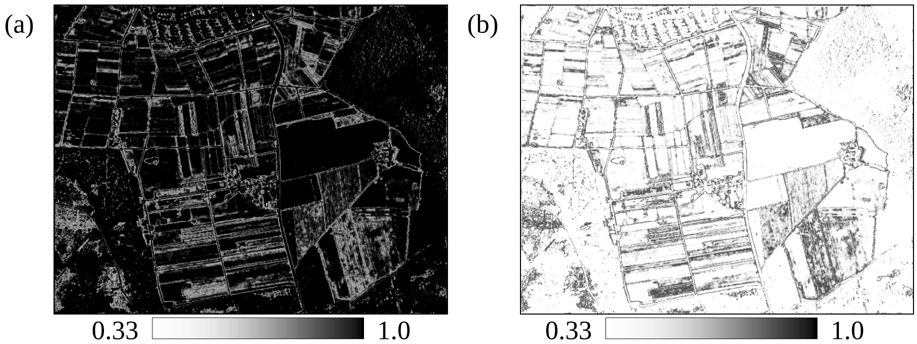


Abbildung 3
Darstellung der größten (a) und zweitgrößten (b) Wahrscheinlichkeiten

Im Bereich der thematischen Unsicherheit ist es wichtig, die Genauigkeitsbetrachtungen auf kontinuierliche Attributwerte auszudehnen, um auch Daten wie z.B. Klima- oder Höhendaten qualitativ beschreiben zu können. Diese im Rasterformat vorliegenden Daten werden in der Regel aus Messungen in unregelmäßig verteilten Punkten abgeleitet. Die Genauigkeitsbeschreibung dieser als Zufallsgrößen betrachteten Meßwerte erfolgt üblicherweise mit deren Varianzen bzw. Standardabweichungen. Da diese Informationen jedoch nur spärlich zur Verfügung stehen, werden daher ersatzweise einfache Modelle zur Beschreibung der Genauigkeit der abgeleiteten Rasterdaten aufgestellt. Im Falle eines digitalen Höhenmodells könnte dies etwa so aussehen, daß ein einziges Genauigkeitsmaß für alle Höhen verwendet wird oder individuell für jede Rasterzelle ein Maß in Abhängigkeit der Landnutzung oder der Neigung angegeben wird. Prinzipiell sind die Genauigkeiten der originären Meßwerte mit den Methoden der mathematischen Statistik auf die Rasterdaten zu übertragen, da die Genauigkeit innerhalb des Datensatzes in Abhängigkeit von unterschiedlichen Faktoren lokal stark schwanken kann. Durch die Ersatzmodelle wird diese Variabilität hingegen nicht immer ausreichend beschrieben. Im Gegensatz zu den Attributen, die diskrete Werte annehmen, wird die Genauigkeitsbeschreibung der kontinuierlichen Attributwerte durch die zugehörige Varianz bzw. Standardabweichung pro Bildelement verwaltet.

Im Bereich der Vektordaten erfolgt die Datenerfassung, indem zu vorgegebenen diskreten oder kontinuierlichen Attributwerten die Ausdehnung dieser Attribute erfaßt wird. Aufgrund dieser Vorgehensweise wird zur Vereinfachung vielfach angenommen, daß die als fest betrachteten Attributwerte keine Unsicherheiten beinhalten. Diese Einschränkung ist jedoch in der Wirklichkeit nicht immer gerechtfertigt, da die erfaßten Objekte Inhomogenitäten bezüglich der Thematik aufweisen können oder die Attributzuweisung unsicher ist. Zur Modellierung dieser thematischen Unsicherheit sind die prozentualen Anteile der jeweiligen Attribute anzugeben.

3 Modellierung der Unsicherheit

Ausgangspunkt der Integration der Unsicherheit innerhalb dieser Arbeit bildet ein hybrides Datenmodell, das sich für die räumliche Modellierung auf das Objektmodell stützt. Hybrid wird das Datenmodell durch die Erweiterung der Geometriekomponente, sowohl eine Raster- wie auch eine Vektordarstellung zuzulassen. Die Thematikkomponente folgt einer üblichen Unterteilung der Attribute gemäß ihrer Skalenzugehörigkeit. Abbildung 4 zeigt die graphische Repräsentation des Modells in UML-Notation (Quatrani 1998).

Das hybride Datenmodell ist nun um die Beschreibung der Unsicherheit zu erweitern. Sie ist an beiden Komponenten in gleicher Weise zu berücksichtigen, da sowohl Thematik als auch die Geometrie eines Objektes unsicher sein können. Als Unsicherheitsmaße werden, wie bereits erläutert, Wahrscheinlichkeiten und Standardabweichungen eingesetzt. Während die Geometriekomponente beide Maße in äquivalenter Weise nutzt, sie also als Alternativen zu verstehen sind, setzt die Thematik beide unabhängig voneinander jeweils zur Beschreibung der Unsicherheit von diskreten (Wahrscheinlichkeiten) und kontinuierlichen Attributen (Standardabweichungen) ein. Zu beachten ist die Festlegung, daß die geometrische Unsicherheit nur im Falle einer Vektorgeometrie existiert. Die Rastergeometrie wird dagegen als sicher betrachtet, während die gesamte Unsicherheit in diesem Fall die Thematikkomponente trägt. Besitzt ein Objekt mehrere unsichere Attribute, so ist für jedes eine eigene Beschreibung notwendig. Ein Objekt könnte demnach z.B. ein Wahrscheinlichkeitsraster für die geometrische Unsicherheit und mehrere verschiedene Unsicherheitsmaße (Standardabweichungen und Wahrscheinlichkeiten) für die thematische Unsicherheit seiner Attribute umfassen. Abbildung 5 zeigt einen Ausschnitt des hybriden Modells mit den entsprechenden Erweiterungen zur Unsicherheit.

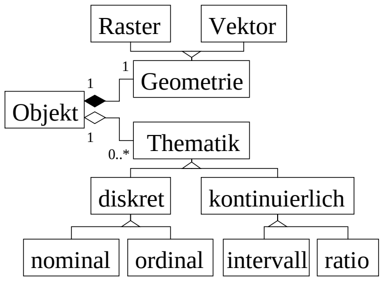


Abbildung 4
Hybrides Datenmodell

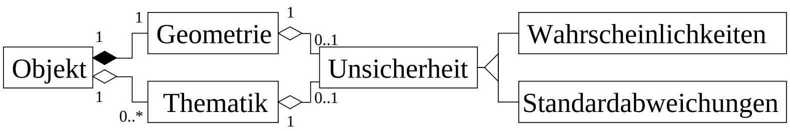


Abbildung 5
Unsicherheitsmodell

4 Objektbildung

Die Grundlage zur Realisierung des definierten hybriden Datenmodells ist die Definition bzw. Bildung von Objekten aus räumlichen Phänomenen. Im Bereich der Vektordaten entsteht ein Objekt dadurch, daß seine Thematik festgelegt und die zugehörige Geometrie erfaßt wird. Im Bereich der Rasterdaten ist gemäß der Ausprägung des Wertebereichs der Attribute zu unterscheiden. Nehmen die Attribute kontinuierliche Werte an, so besitzt jede einzelne Rasterzelle eine individuelle Thematik. Folglich bildet jede Rasterzelle ein Objekt, dessen Geometrie durch die Form der Rasterzelle festgelegt wird. Für weiterführende Analysen können mit Hilfe von selektiven Abfragen (Kapitel 5.1.1) Objekte mit gleichen Eigenschaften ausgewählt und zu neuen Objekten zusammengefaßt werden. Im Falle diskreter Attributwerte ist es notwendig, eine Objektbildung durchzuführen, d.h. benachbarte Rasterzellen sind zu einem Objekt zusammenzufassen, wenn sie dieselbe Thematik aufweisen.

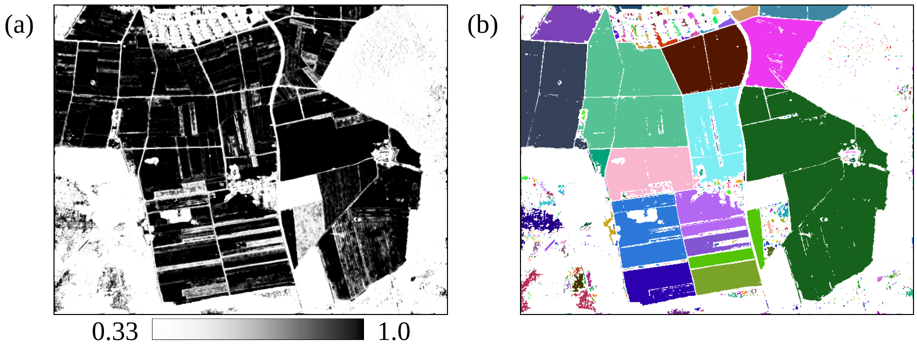


Abbildung 6
Darstellung der Wahrscheinlichkeiten (a) und des Ergebnisses der Objektbildung (b) für die Landnutzungsklasse Acker

Im Zusammenhang mit der Modellierung der thematischen Unsicherheit ist es wichtig, die Wahrscheinlichkeiten der ursprünglichen Bildelemente auf das neu gebildete Objekt zu übertragen bzw. in die Objektbildung zu integrieren, um ihre Berücksichtigung in weiterführenden Analysen zu gewährleisten. Dies soll an einem konkreten Beispiel aus dem Bereich der Fernerkundung verdeutlicht werden. Ausgangsdaten sind das Ergebnis der Klassifizierung für die Landnutzungsklasse Acker und die dazugehörigen Wahrscheinlichkeiten (Abbildung 6 (a)). Benachbarte Bildelemente werden zu einem Objekt zusammengefaßt, wenn ihre Wahrscheinlichkeiten ein vordefiniertes Einheitlichkeitskriterium erfüllen, z.B. die Wahrscheinlichkeiten größer als ein gegebener Grenzwert sind. Für dieses recht einfache Kriterium ist das Ergebnis der Objektbildung in Abbildung 6 (b) dargestellt.

5 Analysen

Nach der Beschreibung der Modellierung der Unsicherheit und ihrer Integration in das hybride Datenmodell wird auf den Einfluß der Unsicherheit auf verschiedene Funktionalitäten eines GIS eingegangen.

5.1 Thematische Analysen

Selektive Abfragen im Datenbestand nach räumlichen, zeitlichen und sachbezogenen Aspekten werden den sog. Mengenoperationen zugeordnet. Anhand von zwei ausgewählten Beispielen aus diesem Bereich werden die Grundlagen zur Berücksichtigung der Unsicherheit erläutert.

5.1.1 Relationale Operatoren

Mittels Selektion können aus einer Tabelle die Zeilen entnommen werden, die eine bestimmte Bedingung erfüllen, oder es können im Bereich der Rasterdaten die Rasterelemente ausgewählt werden, deren Attributwerte in einem vorgegebenen Intervall liegen. Selektionen werden im Zusammenhang mit digitalen Höhenmodellen in vielfältigen Anwendungen der Geowissenschaften als Analysemethode eingesetzt. Gilt es etwa im Bereich der Hydrologie die Gebietsgrenzen von Überschwemmungsgebieten zu ermitteln, so ist zu untersuchen, ob die Rasterzellen unter- oder oberhalb einer vorgegebenen Höhenschranke liegen oder nicht. Dabei ist es natürlich wichtig, das Ergebnis um eine Genauigkeitsbeschreibung zu ergänzen. Dazu ist jede Höhe, analog zu Kapitel 2.1, als kontinuierliche Zufallsvariable aufzufassen. Geht man auch hier von der Normalverteilung aus, so ist diese Zufallsvariable ausreichend durch Mittelwert und Varianz beschrieben. Dem Mittelwert entspricht die Höhe der jeweils betrachteten Rasterzelle. Zur Bewertung der Selektion ist die Wahrscheinlichkeit zu berechnen, mit der die Zufallsgröße „Höhe“ einen Wert annimmt, der kleiner oder größer als die vorgegebene Höhenschranke ist.

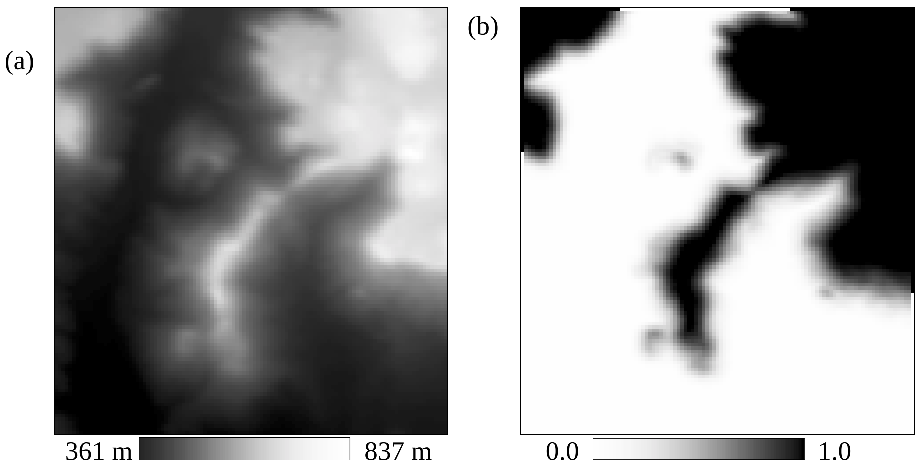


Abbildung 7
Darstellung des Höhenmodells (a) und der
Wahrscheinlichkeiten des selektierten Bereiches (> 600 m) (b)

In Abbildung 7 ist in (a) ein 5.5 km $\times$ 6.0 km großes Höhenmodell (Rasterweite 50m) dargestellt. Mittels Selektion wird unter Berücksichtigung der Genauigkeiten der Bereich ausgewählt, der über 600 m liegt. Aufgrund der Unsicherheit des Höhenmodells können die Rasterzellen, deren Höhenwerte ungefähr 600 m betragen, nicht mit einer Wahrscheinlichkeit von 100% zugewiesen werden. Daraus resultiert in der Darstellung der Wahrscheinlichkeiten (b) die unscharfe Begrenzungslinie des selektierten Bereiches.

5.1.2 Funktionale Ableitung neuer Informationen

In der vorangegangenen Selektionsmethode werden nach einem vorgegebenen Kriterium Objekte aus einem Datenbestand ausgewählt. Eine mathematische Weiterverarbeitung der Daten erfolgt jedoch nicht. Soll etwa aus dem Höhenmodell für jede Rasterzelle die Neigung berechnet werden, um beispielsweise erosionsgefährdete Gebiete detektieren zu können, so werden durch eine funktionale Verknüpfung der Attribute neue Informationen abgeleitet. Ausgehend von der Genauigkeit des Höhenmodells und dem funktionalen Zusammenhang zwischen den Höhen und den Neigungen ist mit den Methoden der mathematischen Statistik die Genauigkeit des Neigungsmodells zu bestimmen. Die Standardabweichungen der Neigungen sind für das in Abbildung 7 dargestellte Höhenmodell in Abbildung 8 graphisch veranschaulicht.

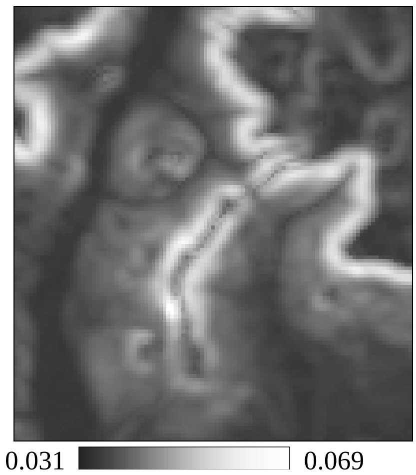


Abbildung 8
Standardabweichungen des Neigungsmodells

5.2 Geometrische Analysen

Geometrische Analysen bewirken eine Veränderung der Geometrie von Objekten. Zum einen kann sich die Änderung auf ein bereits bestehendes Objekt beziehen (z.B. Verkleinern oder Vergrößern eines Objektes durch Pufferbildungen), oder aber es wird ein komplett neues Objekt als Ergebnis der Analyse erzeugt (z.B. bei Verschneidungen). Aus der Vielzahl solcher Analysemöglichkeiten sollen im folgenden zwei wichtige Methoden ausgewählt werden, die als Basisoperationen grundlegende Bedeutung besitzen. Zum einen werden Raster-Vektor-Konversionen behandelt, denen besonders in einer hybriden Systemumgebung eine zentrale Rolle zukommt. Zum anderen sollen Verschneidungen betrachtet werden, die als typische Operationen in einem GIS häufig zum Einsatz kommen. Für beide Methoden werden die Auswirkungen der Berücksichtigung der Unsicherheit während der Verarbeitung erläutert.

5.2.1 Konversionen

In einer hybriden Systemumgebung werden Konversionen eingesetzt, um innerhalb der Geometriekomponente von einer Darstellungsart in die andere zu wechseln. Während manche hybride Ansätze durch Neuentwicklung von Datenstrukturen und Methoden versuchen, Konversionen zu vermeiden, weil sie zu einem gewissen Grade arbeits- und damit zeitaufwendig sind (Jung et al. 1998), werden hier die Konversionen gezielt zum Vorteil der Datenverarbeitung eingesetzt. Ein wesentlicher Vorteil ist demnach, daß die bereits vorhandenen Methoden eines bestehenden Systems, die i.d.R. mit einer bestimmten Datenart arbeiten, weiterhin genutzt werden können, da keine Änderungen in der Strukturierung der Geometriedaten zu berücksichtigen sind. Liegen z.B. zwei miteinander zu verarbeitende Datensätze in unterschiedlichen Datenarten vor, so ist einer der beiden zu konvertieren, und die bestehenden Methoden können ohne Probleme angewendet werden. Die vorgeschaltete Konversion bewirkt, daß sämtliche Methoden mit beliebigen Datenarten zu nutzen sind. Dadurch dürfen die Methoden selbst als hybride Methoden bezeichnet werden. Beachtet man zusätzlich, daß manche Operationen mit einer bestimmten Datenart optimal zu lösen sind, so kann gezielt dieser Vorteil ausgenutzt werden. Z.B. ist die Verschneidung sehr einfach mit Rasterdaten auszuführen, da sie dort nur einer einfachen Überlagerung der Zellen entspricht, während im Vektorbereich geometrische Schnitte aufwendig zu berechnen sind.

Im Normalfall ist es für den Nutzer eines hybriden Systems nicht wichtig zu wissen, mit welcher Datenart eine bestimmte Methode konkret arbeitet. Für ihn ist nur die Verfügbarkeit einer Methode von Interesse. Somit sollte eine eventuell notwendige Konversion als autonomer Hintergrundprozeß ablaufen, der keine Interaktion seitens des Nutzers verlangt. Da der Prozeß im Verborgenen abläuft, verliert der Nutzer auch die Kontrolle darüber. Beachtet man, daß in hybriden Anwendungen häufig Konversionen eingesetzt werden, dann ist zusätzlich vom System zu fordern, daß eine verlustfreie Konversion durchgeführt wird. Als verlustfrei kann die Konversion bezeichnet werden, wenn sich die Objektgeometrie während einer Reihe von aufeinanderfolgenden Hin- und Rückkonversionen nicht verändert hat. Bedingt durch Diskretisierung (von Vektor nach Raster) bzw. Glättung (von Raster nach Vektor) und der dazu notwendigen Festlegung der Parameter Rastergröße bzw. Glättungsfaktor wird eine kleine Abweichung nicht zu vermeiden sein. Da die Daten aber zu einem gewissen Grad mit Unsicherheit behaftet sind, kann eine Abweichung akzeptiert werden, sofern sie sich auch im Falle von unendlich vielen Konversionen deutlich geringer als die Unsicherheit äußert. Damit ist klar, daß die Steuerung der Konversion durch die Unsicherheit in den Daten geregelt werden muß. Die Rastergröße (Auflösung) bei der Vektor-Raster-Konversion sollte klein gegenüber der Vektorgenauigkeit sein. Liegt ein Wahrscheinlichkeitsraster als Beschreibung vor, dann kann direkt die darin verwirklichte Rastergröße übernommen werden. Im Falle von Standardabweichungen gilt die entsprechende Formel zur Festlegung der Rastergröße für das Wahrscheinlichkeitsraster (Kapitel 2.1). Der Glättungsfaktor bei der Raster-Vektor-Konversion sollte so klein sein, daß lediglich das treppenförmige Erscheinungsbild der resultierenden Vektorelemente beseitigt wird. Die Wahl erfolgt in Abhängigkeit von der aktuellen Rastergröße der Daten. Dies ist zulässig, da die mit den Daten verbundene thematische Unsicherheit deutlich größer ausfällt, so daß ein kleiner Verschiebungsbetrag die Eigenschaften des Objektes unbedeutend verändert.

Unter Beachtung der Festlegungen für die Konversionsparameter können Konversionen unbedenklich und in beliebiger Häufigkeit ohne Erhöhung der Unsicherheit angewendet werden. Dadurch bleibt die mit den Daten registrierte Unsicherheit in ihrer Größe erhalten.

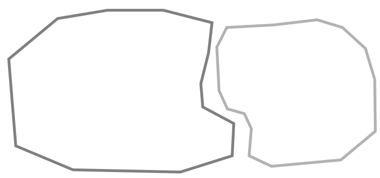
5.2.2 Verschneidung

Innerhalb von Verschneidungsanalysen sind die sich überlappenden Regionen der Objekte zweier Eingangsdatensätze gesucht. Aus den gefundenen Regionen werden dann neue Objekte gebildet, die nun die gesamten Eigenschaften der Eingangsobjekte erhalten. Wie bereits erwähnt, ist die Lösung des Problems besonders einfach mit Rasterdaten durchzuführen. Vektordatenbestände sind dazu in Rasterdaten zu konvertieren. Die Bildung neuer Objekte erfordert zusätzlich die Übertragung der Unsicherheit auf das Ergebnis. Die Übertragung entspricht einer Fortpflanzung der Unsicherheitseigenschaften unter Berücksichtigung der konkreten geometrischen Situation. Voraussetzung ist, daß die Unsicherheit als Wahrscheinlichkeitsraster vorliegt. Für Vektordaten ist dies durch die Anwendung der beschriebenen Formel einfach zu realisieren. Rasterdaten mit diskreten Attributen sind bereits mit Wahrscheinlichkeiten in ihrer Unsicherheit beschrieben. Dagegen erhält man bei kontinuierlichen Attributen Wahrscheinlichkeiten nur im Zusammenhang mit einer Selektion. Da die Verschneidung im Rasterbereich auch als logisches UND interpretiert werden kann, ergibt sich die Fortpflanzung der Unsicherheit gemäß der bekannten Formel der Wahrscheinlichkeitstheorie:

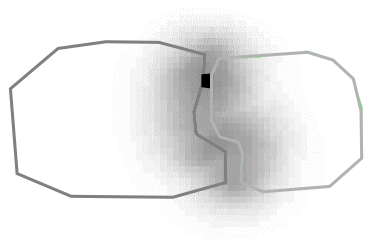
$$p(A_1 \cap A_2) = p(A_1) \cdot p(A_2).$$

A_1 und A_2 bezeichnen dabei die Eigenschaften der zu verschneidenden Rasterdatensätze. Als Einschränkung für die Formel gilt, daß die Eigenschaften voneinander unabhängig sein müssen. Ist dies nicht der Fall, dann ist der Grad der Abhängigkeit in Form von bedingten Wahrscheinlichkeiten zu berücksichtigen.

Abbildung 9 zeigt ein Beispiel der Verschneidung zweier Vektorobjekte. Das Beispiel ist von besonderem Interesse, da hier das traditionelle Ergebnis (ohne Unsicherheit) von der Lösung unter Berücksichtigung der Unsicherheit abweicht. Der traditionelle Ansatz würde in diesem Fall keinen Schnitt hervorbringen, da die dargestellten Geometrien disjunkt zueinander sind. Mit Berücksichtigung der Unsicherheit und unter der Annahme, daß sich die Wahrscheinlichkeitsraster der beiden Objekte in den Randbereichen überlappen, trifft dieses Ergebnis nicht mehr zu. Ein Schnitt existiert und somit ist ein geometrischer Repräsentant für das Schnittobjekt zu bilden. Zur Festlegung des Repräsentanten ist auf dem fortgepflanzten Wahrscheinlichkeitsraster des Ergebnisobjektes eine Schwellwertbildung durchzuführen, die eine gewisse räumliche Ausdehnung liefert. Im Beispiel ist der Schwellwert hoch angesetzt worden, so daß das Schnittobjekt eine punktförmige Ausdehnung zeigt. Zusätzlich wurde das zugehörige Wahrscheinlichkeitsraster überlagert, um die Unsicherheit mit darzustellen.



(a)



(b)

Abbildung 9
Verschneidung mit Fortpflanzung der Unsicherheit (b), bei der der traditionelle Ansatz kein Schnittobjekt liefert (a)

6 Zusammenfassung und Ausblick

Thema des Beitrags war die Erweiterung eines hybriden Datenmodells um die Modellierung der Unsicherheit. Damit ist die Grundlage zur Berücksichtigung der Unsicherheit innerhalb von interoperablen Anwendungen geschaffen. An einigen Beispielen wurde aufgezeigt, welche Änderungen sich dadurch für die Methodik ergeben und wie Analyseergebnisse hinsichtlich ihrer Unsicherheit zu interpretieren sind. Innerhalb einer Anwendung genügt es jedoch nicht, daß nur einzelne Methoden Unsicherheit berücksichtigen. Vielmehr müssen alle verfügbaren Funktionen mit entsprechenden Erweiterungen versehen werden.

Literaturverzeichnis

Bill, R., Fritsch, D. (1991): Grundlagen der Geo-Informationssysteme, Band 1: Hardware, Software und Daten, Karlsruhe

Bill, R., Korduan, P. (1998): Flächenverschneidung in GIS - Stochastische Modellierung und Effizienzbetrachtung, in: Zeitschrift für Vermessungswesen, 123, Teil 1 in No. 8, S. 247-253, Teil 2 in No. 10, S. 333-338

Fritsch, D., Glemser, M., Klein, U., Sester, M., Strunz, G. (1998): Zur Integration von Unsicherheit bei Vektor- und Rasterdaten, in: Geo-Informations-Systeme - Zeitschrift für raumbezogene Information und Entscheidungen, 11, Heft 4, S. 26-35

Glemser, M., Klein, U. (1999): Hybride Modellierung und Analyse von unsicheren Daten, Schriftenreihe der Institute des Fachbereichs Vermessungswesen, Universität Stuttgart, Report Nr. 1999.1

Jung, S., Voser, S.A., Ehlers, M. (1998): Hybrid Spatial Analysis Operations as a Foundation for Integrated GIS, in: IAPRS, 32, Part 4, „GIS – Between Visions and Applications“, Stuttgart

Kraus, K., Haussteiner, K. (1993): Visualisierung der Genauigkeit geometrischer Daten, in: GIS, 6, Heft 3, S. 7-12

Quatrani, T. (1998): Visual Modeling with Rational Rose and UML, Reading/MA

Richards, J.A. (1993): Remote Sensing Digital Image Analysis, Berlin et al.