

6 Geomarketing-Prozess I: Modellaufbau, GIS- und Geostatistik-Methoden

Im Folgenden wird ein Modell aufgestellt, das die Grundlage für die durchzuführenden Analysen für eine Filialnetzoptimierung unter besonderer Berücksichtigung des Wettbewerbs bildet. Die einzelnen Bausteine des Modells werden erörtert, insbesondere die Methodenwahl aus dem GIS und Data Mining Bereich wird ausführlich dargelegt. Eine Strukturierung der Methoden, die sich für Geomarketing anbieten, wird versucht. Durch das iterative Vorgehen bei der Durchführung von Analysen werden Methoden in Hinblick auf mögliche Zusatzkenntnisse für die Filialnetzoptimierung und Wettbewerberanalyse gegenüber anderen Methoden getestet.

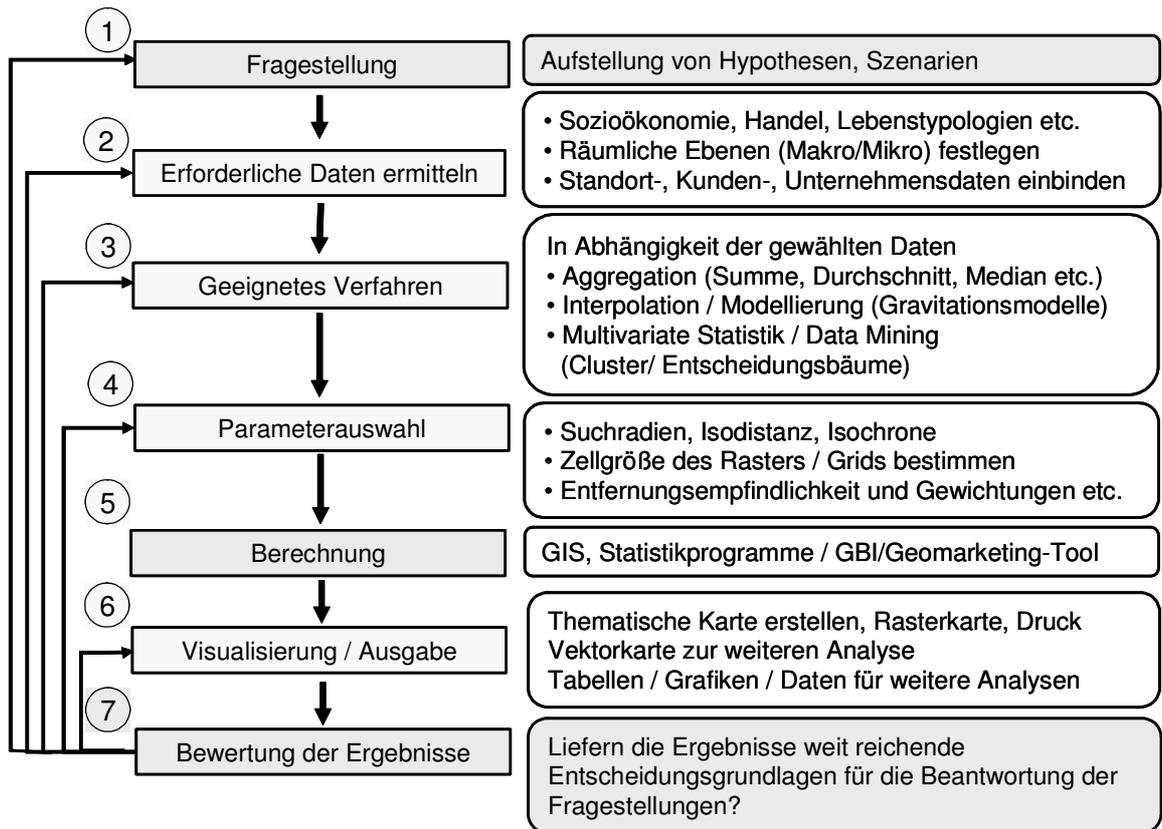
6.1 Methodisches Vorgehen für eine Modellentwicklung

Die Idee der räumlichen Marktbearbeitung mit GIS, das Karten, Daten und Methoden zusammenführt, galt in den 90er Jahren noch als Innovation (vgl. FRÜHLING & STEINGRUBE 1996, S. 197). Heute stellt Geomarketing eine solche Innovation dar, in der die Marktbearbeitung mit einer sequentiellen Kombination von einzelnen methodischen Bausteinen aus GIS, Statistik und *Spatial Data Mining* vorgenommen wird. Die einzelnen Methoden, die im Geomarketing verwendet werden, sind die kleinsten elementarsten Bausteine einer komplexen Analyse. Das Ergebnis einer Geomarketinganalyse dient als Grundlage für geplante Maßnahmen der Marktbearbeitung.

Die folgende Abbildung (*Abb. 6-1*) zeigt ein iteratives Vorgehen bei einer solchen Modellbildung. Es werden sieben Punkte von der Aufstellung von Fragestellungen aus den Geschäftsprozessen bis zur Bewertung der Analyseergebnisse aufgezeigt und erläutert. Der Inhalt der Fragestellungen und das Ableiten von Hypothesen (*Punkt 1*) ergibt sich z. B. aus den veränderten Kundenbedürfnissen und aus dem zunehmenden Wettbewerbsdruck (siehe dazu *Kapitel 2*). Die qualifizierte Datenermittlung, die Qualitätsprüfung der Daten und der Aufbau eines adäquaten Datenmanagements (*Punkt 2*) sind Voraussetzung für ein hochwertiges Analyseergebnis. Nur wenn die Ausgangsdaten entsprechend detailliert und verlässlich sind, sind auch die Ergebnisse von Rechenoperationen, die darauf zugreifen, qualitativ hochwertig.

Geeignete Verfahren, Ermittlung von Berechnungsmethoden für die aufgestellten Fragestellungen (*Punkt 3*), die Parameterauswahl (*Punkt 4*) zu treffen, die Berechnung durchzuführen (*Punkt 5*) und die Visualisierung der Ergebnisse (*Punkt 6*) werden Gegenstand dieses Kapitels sein. Am Ende des Prozesses steht die Bewertung der

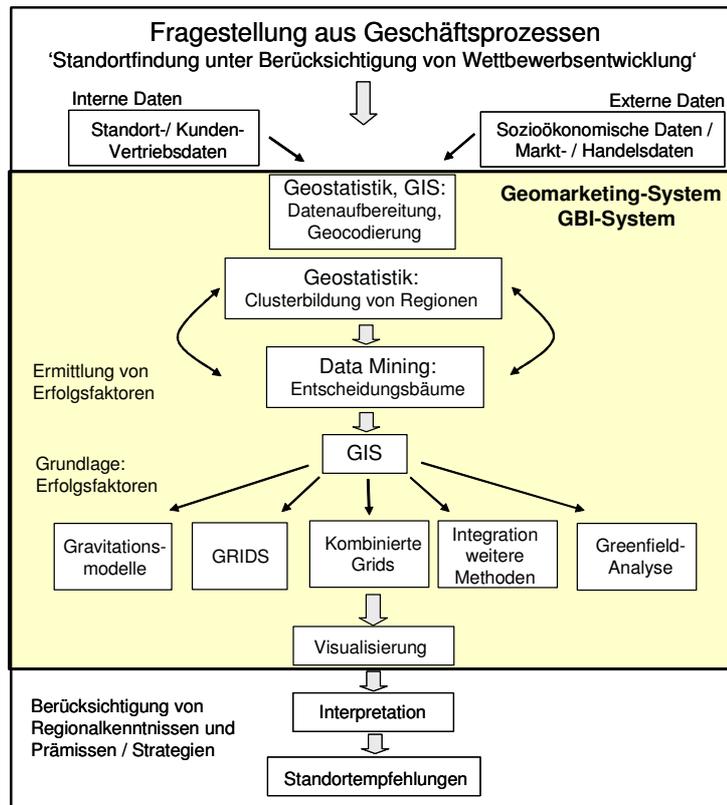
Ergebnisse (*Punkt 7*), aus denen sich möglicherweise wieder weitere Fragestellungen ergeben, die den Prozess erneut anstoßen. Im Falle, dass die Fragestellung nicht hinreichend beantwortet ist, wird erneut ein Einstiegspunkt gewählt (*Punkt 2,3,4 oder Punkt 6*), erforderliche werden Daten überprüft, ergänzt und die Berechnungsmethoden oder Parameterauswahl verändert. Insofern handelt es sich um ein iteratives Vorgehen, das im ‘*Geo Business Intelligence*’ Prozess zu weiteren Erkenntnissen führt.



Quelle: eigene Darstellung

Abb. 6-1: Geo Business Intelligence: Iteratives Vorgehen bei der Modellbildung

Für die Standort-, Filialnetzplanung und -optimierung wird das hier dargestellte allgemeine Modell ‘*Geo Business Intelligence*’ (*Abb. 6-1*). in den methodischen Punkten für ein Modell *Standortempfehlung* präzisiert (*Abb. 6-2*) und im Weiteren in Anwendungsfällen erprobt. Das Modell lässt sich auf das Thema Kundenanalyse und weitere Themenfelder des Geomarketings übertragen. Es wird auf die Themengebiete immer wieder Bezug genommen, um dies aufzuzeigen.



Quelle: eigene Darstellung

Abb. 6-2: Prozess Geo Business Intelligence: Modell Standortempfehlung (vereinfacht)

Voran geht der Anwendung der Methoden eine Aufstellung und Strukturierung von Geomarketing-Verfahren aus GIS, Statistik und Data Mining, die die Vielfältigkeit der möglichen Einzelmethoden verdeutlicht. Einige der Methoden sind Grundlage von komplexen hier angewendeten Methoden und werden von daher beschrieben. Aufgrund der Vielzahl möglicher und sinnvoller Kombinationen können diese allerdings nicht erschöpfend herausgearbeitet werden. Gleichwohl gilt es zu beachten, dass sich kein Modell neutral bezüglich der erzielten Erkenntnisse verhält: Jedes Modell stellt eine Abstraktion der Realität dar und beeinflusst die Ergebnisse der Analyse (MARTIN 1995, S. 62).

Die Darstellung (Abb. 6-2) zeigt anhand einer Auswahl von Methoden ein mögliches Vorgehen zur Analyse von Standortplanung bzw. deutschlandweiter Filialnetzplanung. Die Methoden, die hier aus Geostatistik, GIS und Data Mining zusammengefasst sind, sind Bestandteile eines Geomarketing-Systems. Bisher allerdings gibt es ein solches System mit diesem Umfang an Funktionen nicht. Das im Zusammenhang mit dieser Arbeit konzipierte und eingesetzte *GBI-Tool* ist mit dem GIS-System *Filialinfo* ein weiterer Schritt zur Schaffung eines solchen umfassenden Geomarketing-Systems.

6.2 Methoden räumlicher Analysen

6.2.1 Systematisierung

Einteilungen von Methoden und Techniken räumlicher Analysen werden aus unterschiedlichen methodisch-fachlichen Richtungen definiert, so dass teilweise idente Techniken mit unterschiedlichen Namen belegt sind. Eine klare allgemeingültige Ordnung existiert demnach bisher nicht. Hier eine Systematik aufgezeigt, die sich um die Pragmatik der hier dargelegten Anwendungen auf die Analyse räumlicher Probleme bezieht.

Geographische Informationssysteme und statistische Verfahren sind im Geomarketing unentbehrlich. Speziell auf dem Gebiet des Geomarketings setzen die weitaus meisten Fragestellungen die Verarbeitung extrem große Datenmengen voraus (MAGUIRE 1995, S. 173), wie es auch im Data Mining der Fall ist. Daten müssen geocodiert und transformiert werden, um sie anderen Daten anzupassen und zu integrieren. Sie werden anschließend nach räumlichen Mustern analysiert. Dieser Prozess stellt hohe Anforderungen an die verwendete Software, insbesondere wenn die Analysen zeitnah berechnet werden sollen.

OPENSHAW (1995, S. 151) vertritt die Ansicht, dass raumbezogene Analysen auf Methoden ausgerichtet werden sollten, die zur Darstellung und Identifizierung geographischer Muster auf verschiedenen räumlichen Maßstabsebenen geeignet sind. In dieser Hinsicht ist wiederum die Stärke Geographischer Informationssysteme von großem Nutzen, Datenmodelle konstruieren zu können, die einen schnellen Wechsel zwischen verschiedenen Aggregationsebenen¹ ermöglichen. GIS-Methoden werden benötigt, um die integrierten Geodaten so zu verarbeiten, dass Geographische Informationssysteme zu anwendungsorientierten Instrumenten der unternehmerischen Entscheidungsunterstützung werden (EBENDA S. 159). Daher stellt GIS einen der vielversprechendsten Anwendungsbereiche für raumbezogene Analysemethoden dar. Die gezielte Anwendung von GIS-Methoden auf interne und externe Daten eröffnet oftmals eine neue Sicht auf Unternehmensaspekte. Von daher wird dem Gebiet GIS ein besonders ausführlicher Abschnitt gewidmet.

In der folgenden Abbildung (Abb. 6-3) werden die für Geomarketing relevanten Methoden, die in GIS abgebildet sind, dargestellt. Es wird versucht, sie in eine logische

¹ Dies wird hier als Funktionen 'Spatial drill-down' und 'Spatial drill-up' bezeichnet (siehe Kapitel 6.2.5).

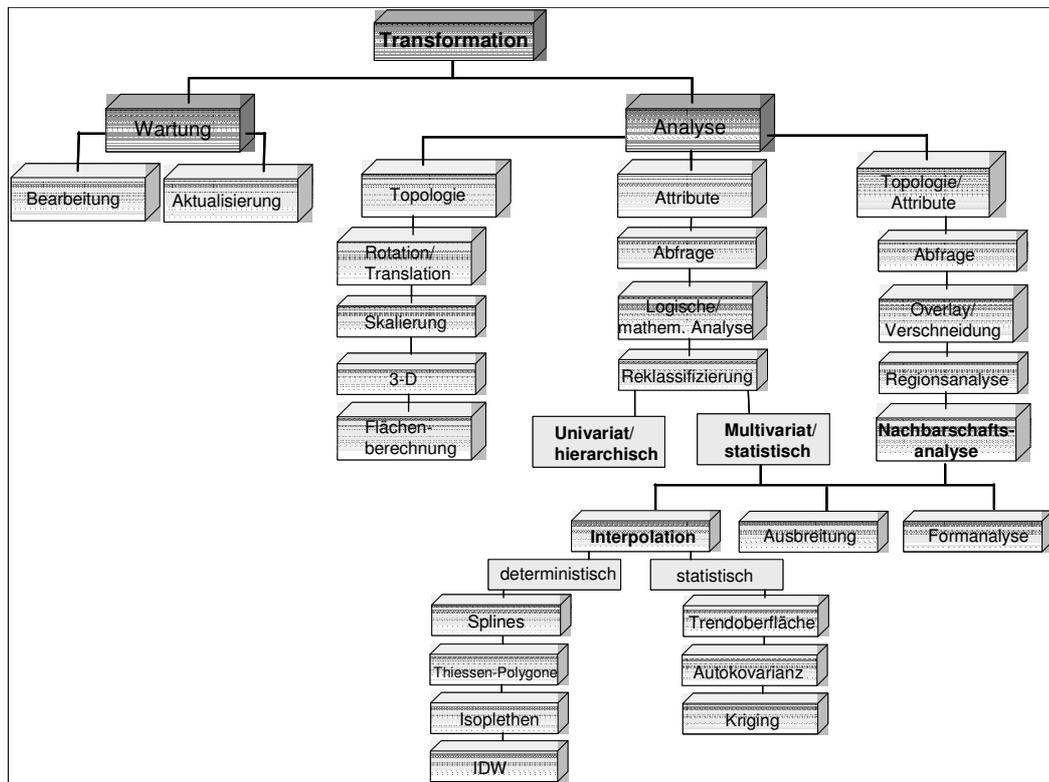
Ordnung mit ihren jeweiligen Abhängigkeiten abzubilden. Einige komplexere Verfahren, die hier von Relevanz sind, werden in den weiteren Kapiteln erläutert.

GIS Analyse-Begriff	Einzelne Methoden, Verfahren
Generalisierung	Vereinfachung, Zusammenfassen, Selektieren, Betonung, Verdrängung
GIS-Abfragen	Raumbezogene Abfragen (Innerhalb, Außerhalb, Nähe, Berührend, Überschneidung, Schnittmenge)
	Attributbezogene Abfragen (SQL)
GIS-Analyse	Geometrische Berechnungen (Länge, Flächeninhalt etc.)
	Räumliche Zuordnung
	Nachbarschaftsanalysen
	Einzugsgebietsberechnung anhand Kundenreichweiten, Einzugsgebietsberechnung anhand der Anzahl der Ausprägung eines Parameters (z.B. Einwohneranzahl)
	Netzwerke, Routingoptimierung, Navigation, Allokations-, Distributionsanalyse
	Routing, Isodistanzen, Isochronen
	Verschneidungen
	Pufferzonen, Korridore
	Interpolationen
	Thiessen, Voronoi-Polygonisierung
	Kriging
Point Pattern Analysis	
Visualisierung	Choroplethenkarte, Säulen-, Balkendiagrammkarten, Kreissektorenkarten, 3D-Darstellung (Größe, Dichte, Farbe, Form, Muster, Ausrichtung)
Gravitationsmodelle	‘Gravitation Retail Trade Law’ (Reilly, 1931), Huff Modell (1963), Multiplicative Competitive Interaction Model MCI (Nakanishi & Cooper, 1974), Competing Destinations Model CDM (Fortheringham 1983, 1989) und weitere
GRID Analyse	Berechnung von Daten anhand von unterschiedlichen Zellgrößen z.B. 100 x100 Meter. Umlegung der Daten auf Zellen / GRIDS - anteilmäßig, Schwerpunkt, Zentroid, Summe, Durchschnitt, Median, Modus, Gewichtung (siehe auch Interpolationsmethoden)
Greenfield-Analyse GRID-basiert	Ausgehend von Erfolgsfaktorermittlung (z.B. durch Entscheidungsbäume) Einsatz der Erfolgsfaktoren in GRID-Analysen zur Auffindung gleich strukturierter Gebiete

Quelle: eigene Zusammenstellung (erweitert nach SCHÜSSLER 2000, S. 56ff)

Abb. 6-3: GIS-Methoden für den Einsatz im Geomarketing

Eine Ordnung der vielfältigen Analysemöglichkeiten der GIS-Operatoren in eine eindeutige Struktur ist nicht durchgehend definiert. Als mögliche Einordnung dient die folgende Grafik, wobei allerdings in der Praxis die ‘*Transformationsprozesse*’ ineinander übergehen.



Quelle: nach BURROUGH 1986, BLASCHKE 1997

Abb. 6-4: Annäherung Transformationsoperationen in GIS

Entscheidend ist, dass alle Ansätze der Systematisierung von räumlichen Analysen das Bestreben haben, den Aufbau eines einheitlichen und konsistenten Zugangs zur Bearbeitung mit Geoinformationen darzulegen. Die in der folgenden Abbildung (Abb. 6-5) dargestellte weitere Systematisierung versucht nicht eine Bestehende in Frage zu stellen, sondern zeigt, dass sich durch die Anwendung von Mind Mapping, die im Rahmen des Geomarketings notwendigen Kombinationen von GIS, uni- und multivariater Statistik und Data Mining ideal abbilden lässt.

Das Mind Mapping bietet die Möglichkeit einer detaillierteren Darstellung, die auch die Verknüpfungen der einzelnen Methoden für die Anwendungsfälle darstellt. Die Methode eignet sich hervorragend dadurch, Prozesse, die sich nicht in einer einfachen hierarchischen Form abbilden lassen, zu entwickeln. Es wird deutlich, dass eine stetige Weiterentwicklung und Vertiefung in einem der einzelnen Äste möglich ist, und dass dieses auch durchaus anwendungsfallbedingt begründet werden kann.

Muster, Zusammenhänge und Veränderungen untersucht wird, nimmt dieses visuelle Kommunikationsprinzip eine zentrale Stellung ein.

Geographische Aspekte können insbesondere durch die Verwendung von Karten dargestellt werden. Der Begriff Visualisierung beinhaltet alle Verfahren und Maßnahmen, digitale Daten in einer Form darstellen, die zur Vermittlung ihres Inhaltes dem menschlichen Auge und der Denkweise gerecht wird. Da die Hälfte aller Neuronen des menschlichen Gehirns mit dem Sehen assoziiert sind, können Menschen graphische Zusammenhänge weit schneller aufnehmen als z. B. schriftliche oder akustische Informationen (vgl. BARTELME 1995, S. 5).

Hinsichtlich der Betonung und Hervorhebung der darzustellenden Phänomene gibt es in Geographischen Informationssystemen einige Besonderheiten im Vergleich zu traditionellen kartographischen Arbeitsmethoden². Der bedeutendste Vorteil von GIS gegenüber der manuellen Kartographie liegt in der Flexibilität: Kartenlayer können ein- oder ausgeblendet werden. Farben, Symbol-, Linien- oder Flächendarstellungen können für die gesamte Karte oder auch für einen Kartenseriendruck schnell geändert werden. Entwurfsregeln verlieren daher ihre Wichtigkeit und der Prozess der Kartenherstellung und -verbesserung tritt in den Vordergrund (vgl. POIKER 1997, S. 99). Einen weiteren Vorteil bietet die Möglichkeit auch maßstabsunabhängig arbeiten zu können.

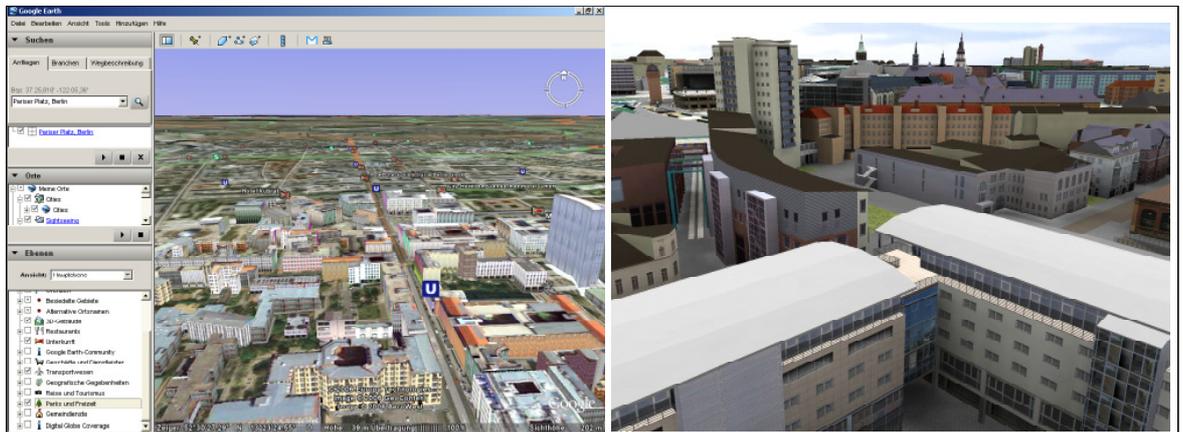
Unterschiedliche Kartenformen können miteinander kombiniert werden, indem mehrere Themen übereinandergelegt werden. Dabei gilt es zu beachten, dass mit steigendem Informationsgehalt einer Karte die Lesbarkeit und Interpretierbarkeit zurückgeht. Ein weiterer wichtiger Aspekt ist die Abbildung der angemessenen Klasseneinteilung für die Erstellung von thematischen Karten, d. h. die Zuordnung von Objekten gemäß ihres Wertes zu einer Kategorie. Geographische Informationssysteme stellen neben der Möglichkeit, die Klasseneinteilung individuell zu bearbeiten, einige standardmäßige Klassifikationsroutinen zur Verfügung. Weitere Besonderheiten werden durch Visualisierungsmethoden, wie 3D Darstellungen, zum Ausdruck gebracht. Anwendungsgebiete sind der Telekommunikationsbereich, Architektur, Städtebau, Immobilienwirtschaft (HOLWEG D., FRAUNHOFER IGD 2004). Weiterhin ist die 3D Visualisierung aussagekräftig, wenn die Bedeutung eines Gebietes anhand von einer Überhöhung prägnant abgebildet werden soll, wie bei der Darstellung des *'Attraktivitäts-Einflußgebietes'* eines Standortes nach dem Gravitationsmodell

2 Zum Thema Kartographie und kartographische Gestaltung siehe auch INTERNATIONALE KARTOGRAPHISCHE VEREINIGUNG 1973, S. 22 und HAKE (1982).

(Kapitel 8.4 und 8.5). Insgesamt wird bei allen durchgeführten Analysen deutlich, dass die Visualisierung der Ergebnisse entscheidend für die Bewertung der Verfahren in der Praxis ist.

Zum Thema Visualisierung und Kommunikation gibt es eigene Wissenschaftszweige und Werke, wie die vom Fraunhofer Institut: ENCARNAÇÃO & ARTS (2006) ENCARNAÇÃO (2004), ENCARNAÇÃO & KIRK (2005), ENCARNAÇÃO ET AL (1997). ENCARNAÇÃO hat bereits 1995 das Thema GIS als Management-Werkzeug der Zukunft entdeckt und die Bedeutung der Visualisierung für das Management hervorgehoben (ENCARNAÇÃO 1995).

Weiterführende wissenschaftliche Arbeiten und Darstellungen in Bezug auf Visualisierung und Data Mining finden sich bei FAYYAD ET AL (HRSG.) (2002). Für weitere aktuelle Forschungen spielt auch das Hasso-Plattner Institut eine zunehmend wichtige Rolle: DÖLLNER, J.; H. BUCHHOLZ, H.; NIENHAUS, M. & F. KIRSCH (2005, S. 42-51); DÖLLNER, J. & H. BUCHHOLZ (2005, S. 173-181); DÖLLNER, J. (2005, S. 325-344); DÖLLNER, J. & K. BAUMANN (2005, S. 217-230); WOOD, J.; KIRSCHENBAUER S.; DÖLLNER J.; LOPES A. & L. BODUM (2005). Die Forschungsarbeiten beschäftigen sich mit computergraphischen und multimedialen Systemen, z. B. intensiv mit der 3D Thematik und Stadtmodellen. Diese Themenfelder werden auch in Bezug auf Geomarketing-Fragestellungen an Bedeutung gewinnen, wie es im Kapitel Schlussfolgerungen und Ausblick nochmals dargelegt wird.



Quelle: eigene Zusammenstellung, Daten und Modelle: Döllner, Buchholz (Hasso-Plattner-Institut)

Abb. 6-6: Stadtmodelle (ALK Berlin) mit Texturen, Beleuchtung, weiteren Informationsschichten (ÖV, Hotels) in Google Earth visualisiert

Die Einbettung von einerseits Stadtmodellen (Abb. 6-6) und andererseits Informationsschichten wie Postfilialen, Wettbewerbern, Frequenzpunkten etc. in einen 'Google Earth Client' bietet neue Möglichkeiten der Visualisierung und damit der

Informationserfassung und Entscheidungsunterstützung. Die Einfärbung von Gebäuden bis hin zu einzelnen Stockwerken (Abb. 6-7) anhand von Kriterien (Umsatz, Energieverbrauch) ergibt eine weitergehende Möglichkeit der Erfassung von maßgeblichen Kennzeichen und Erkenntnissen im direkten Raumzusammenhang.³



Quelle: eigene Zusammenstellung, Modell: Döllner, Buchholz (Hasso-Plattner-Institut / 3D Geo)

Abb. 6-7: Gebäude eingefärbt nach Variablen, z. B. Umsatz in Google Earth visualisiert

6.2.3 Routing: Routenoptimierung, Isodistanz und Isochrone

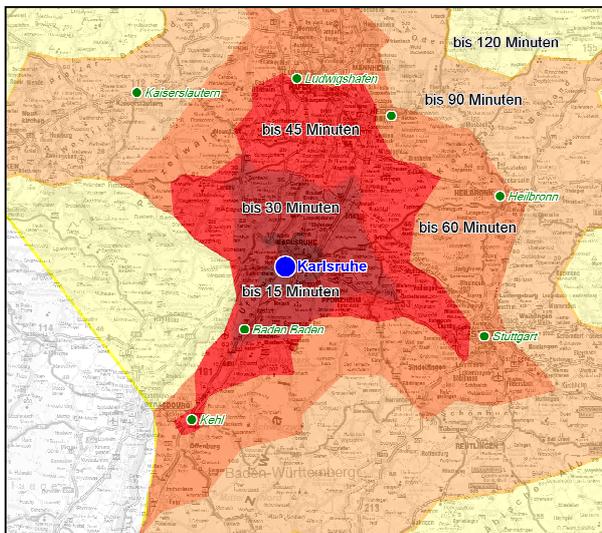
Routenoptimierung, Allokations- und Distributionsanalyse sind als verschiedene Ansätze von Netzwerk-Analysen bei Geographischen Informationssystemen bekannt und ebenfalls elementar für das Geomarketing. Netzwerke sind Systeme, die durch miteinander verbundene linienförmige Objekte gebildet werden (vgl. ARONOFF 1989, S. 225f). Diese werden für den Transport von Personen, Gütern, Dienstleistungen und Daten genutzt. Im Bereich Geomarketing werden Verkehrswege und Logistikströme als Netzwerke dargestellt.

Für alle Ansätze gilt die gleiche Voraussetzung hinsichtlich der Struktur der verwendeten linienhaften Daten. Die Struktur der Netzwerkdaten ist als Modell mit Knoten und Kanten realisiert. Als fachliche Parameter können die Streckenabschnitte Informationen zu Einbahnstraßen, Geschwindigkeits- und Gewichtsbegrenzungen bei Brücken, Baustellen, Verkehrsbelastung, Ampeln, Buslinien etc. beinhalten. Auf diese

³ Diese Darstellungsmöglichkeiten sind im Forschungsstadium (Hasso-Plattner-Institut, Potsdam).

Weise können verkehrsbeeinträchtigende Elemente in Zeiteinheiten umgerechnet werden. Letzendlich beeinflusst die Modellbildung die Analyseergebnisse stark, die je nach Anwendungsfall variieren kann. Beispielsweise kann sich ein Modell zur Abfrage nach verkehrsbeeinträchtigenden Elementen auf übergeordnete Straßennetze beziehen oder aber das Nebenstraßennetz einbeziehen. Daraus ergeben sich unterschiedliche Ergebnisse.

Die Routenoptimierung ermittelt die kürzeste Strecke zwischen zwei Punkten oder die schnellste Verbindung auf der Basis der aggregierten Fahrtzeiten der Streckenabschnitte. Allokations- und Distributionsanalysen sind geeignet; für ein komplexeres System von Quell- und Zielobjekten Optimierungsaufgaben zu erfüllen. Typische Anwendungsfälle sind z. B. Schulbus- oder Behindertentransporte, die viele Punkte in einer optimalen Reihenfolge ansteuern, um einen Zielpunkt zu erreichen oder umgekehrt. Anwendungen im Logistikbereich wie Briefzustellungsoptimierung, Außendienstmanagement bis hin zu Entsorgungstransporten und Baufahrzeugmanagement benötigen derartige Algorithmen.



Quelle: eigene Darstellung

Abb. 6-8: Isodistanzen anhand Tele Atlas Straßennetz, Routingalgorithmus⁴

Anhand von im GIS implementierten Routingalgorithmen und einem routingfähigen Straßennetz⁵ lassen sich auch Einzugsgebiete auf Grundlage des Straßennetzes entweder nach Distanz (Isodistanz) oder nach Zeithorizonten (Isochrone) für verschiedene Klassen wie PKW, LKW, Rad, Fußgänger generieren. Wenn es

um mikroräumliche Analysen geht, ist die Betrachtung von Einzugsgebieten anhand der Isodistanz oder Isochrone eine Möglichkeit, genauer die realen Gegebenheiten zu

⁴ Die Routingalgorithmen sind von der IVU Traffic Technologies AG erstmals in Zusammenarbeit mit dem Konrad-Zuse-Zentrum für Informationstechnologie Berlin (ZIB) erstellt und stetig weiterentwickelt worden. Sie sind in vielfältigen Verkehrs- und Logistikprojekten im Einsatz.

⁵ Zu routingfähigen Straßennetzen zählen zur Zeit die Netze der Firmen Tele Atlas und Navteq. In der Regel werden diese zweimal (Tele Atlas) bzw. viermal (Navteq) jährlich aktualisiert bereitgestellt. Die Deutsche Post World Net verwendet das Tele Atlas Straßennetz.

berücksichtigen. Es werden natürliche Barrieren, wie größere Flüsse oder Eisenbahntrassen, automatisch berücksichtigt. Im Geomarketing wird das unmittelbare Einflussgebiet hinsichtlich Wettbewerber, Bevölkerung, Kaufkraft und Frequenzen untersucht. Von daher ist eine genaue Definition des zu betrachtenden Gebietes für die Qualität des Ergebnisses entscheidend. Dies ist auch bei der Netzplanung von elementarer Bedeutung, da Einzugsbereiche einer Filiale, die anhand einfacher Radien gebildet werden, Flussläufe nicht beachten: Links- und rechtsseitig vom Rhein gelegene Filialen im Gebiet Köln und Bonn sind hierfür ein gutes Beispiel. Nach dem Isochronenansatz ist sofort ersichtlich, dass hier wesentlich längere Reisezeiten zur Erreichung der Filiale in Kauf genommen werden müssen, als es nach dem Radienansatz berechnet wird. Dieses Kriterium wird oftmals mit dem Ziel einer Vereinfachung der Modelle unterschätzt.

6.2.4 Abfragen, Verknüpfung, Verschneidung, Nachbarschaftsanalysen

Bei räumlichen Abfragen werden Kriterien verwendet, um aus den vorhandenen Daten gezielt diejenigen herauszufiltern, die entweder direkt von inhaltlicher Relevanz sind oder anschließend für weitere Analysen benötigt werden. Es wird zwischen räumlichen und attributbezogenen Abfragen differenziert. Räumliche Abfragen beziehen sich auf geometrische Objekte und ihre Attribute. Diese Funktionen sind Nachbarschaftsoperationen, die sich aus der räumlichen Nähe mindestens zweier Objekte ableiten lassen.

Es gibt Operatoren wie *'Innerhalb'* und *'Außerhalb'* bzw. *'Innerhalb invers'*, wobei alle geometrischen Objekte ausgewählt werden, die sich innerhalb oder außerhalb eines anderen Objektes befinden. Das Ausgangsobjekt muss eine Fläche sein. Beispielsweise können alle Filialen ausgewählt werden, die sich innerhalb eines Einzugsgebietes oder Kundenstandorte, die sich außerhalb des fußläufigen Einzugsgebietes einer Filiale befinden.

Zu weiteren Operatoren zählen *'Innerhalb einer Distanz'*, *'Nächstes Objekt'*, *'Nächstes Objekt invers'*, *'Nächste N Objekte'*, *'Nächste Objekte innerhalb eines Schwellenwertes'*. Attributbezogene Abfragen können über Filterungen oder auch Verknüpfungen über Spaltenwerte oder über die Verknüpfung mit einer dritten Tabelle (Referenztable) erfolgen.

Über die Vielfältigkeit der möglichen Verknüpfungsfunktionen, die in einem GIS⁶ abgebildet werden können, gibt folgende Abbildung (Abb. 6-9) Aufschluss.

anhand von Spalten		 alle	
anhand einer Tabelle		 alle	
innerhalb von Gebiet	prozentuale Überlappung	 Fläche zu Fläche	 Linie zu Fläche
	größte Überlappung	 Fläche zu Fläche	 Linie zu Fläche
	Schwerpunkt	 Fläche zu Fläche	 Punkt zu Fläche
	jede Überlappung	 Fläche zu Fläche	 Linie zu Fläche
innerhalb von Distanz	Luftlinie, Isodistanz, Isochrone	 Punkt zu Punkt	
nächstgelegenes Objekt	Luftlinie, Isodistanz, Isochrone	 Punkt zu Punkt	
nächste N Objekte	Luftlinie, Isodistanz, Isochrone	 Punkt zu Punkt	
nächste Objekte innerhalb Schwellwert	Luftlinie, Isodistanz, Isochrone	 Punkt zu Punkt	

Quelle: eigene Zusammenstellung aus dem GIS *Filialinfo* 2006

Abb. 6-9: Mögliche Verknüpfungs-, Verschneidungs-, Distanzanalysen

Zu den Verschneidungsoperationen zählt das Teilen, die Überlappung, das Berühren und das Ausschneiden ('clip', 'erase', 'split', 'identity', 'union', 'intersect') (LAURINI; THOMPSON 1992, S. 330). Verschneidungen werden dann eingesetzt, wenn arithmetische, logische oder statistische Operationen auf der Basis verschiedener räumlicher Bezugsebenen stattfinden sollen. Es können Flächen, Linien mit Linien oder Flächen mit Linien miteinander verschnitten werden. Wichtig ist, dass ein GIS dabei auch ermöglicht, die Attributdaten entsprechend anteilmäßig weiterzugeben. Verschneidungsoperationen stellen wichtige Basisfunktionen für die Feststellung von

⁶ Alle aufgeführten Verknüpfungsmöglichkeiten sind im GIS *Filialinfo* implementiert. *Filialinfo* bildet mit dem neu konzipierten *GBI-Tool* gemeinsam die Systemgrundlage der hier im Weiteren durchgeführten Analysen.

‘über-’ und ‘unter’ versorgten Gebieten und für die Berechnung des Anteils, der in diesen Gebieten erreichbaren Bevölkerung, dar.

Nachbarschaftsanalysen (*‘adjacency analysis’, ‘neighbourhood analysis’*) untersuchen Objekte auf ihre nachbarschaftlichen Beziehungen zu anderen Objekten. Die Bestimmung der Zielobjekte, die Angabe des Nachbarschaftsraumes und die Spezifikation der Funktion, die auf die Objekte des Nachbarschaftsraumes angewendet wird, sind die drei benötigten Parameter für diese Analysen (vgl. ARONOFF 1989, S. 220 ff.). Eine Systematik zu topologischen Verhältnissen von Objekten wird ausführlich bei LAURINI; THOMPSON (1992, S. 175ff) beschrieben.

Die eigentlichen Stärken von Abfragen liegen in ihrer Kombinationsfähigkeit. Vielfältige Kombinationen sind in sequentieller Abfolge möglich. Dabei sind Zusammenstellungen wie auch Abfolgen von räumlichen sowie attributbezogenen Abfragen möglich. Die daraus generierten Ergebnisse, in Form einer Auswahl von Standorten, Gebieten etc. können als Grundlage für die geplanten Maßnahmen der Marktbearbeitung dienen.

6.2.5 Spatial drill-down - Spatial drill-up

Geometrische Objekte können gemäß ihrer räumlichen Position anderen geometrischen Objekten zugeordnet werden. Punkte, Linien oder Flächen können übergeordneten Flächen zugewiesen werden. Dabei werden die Attribute der zugewiesenen Objekte nach verschiedenen Methoden (z. B. Aggregation oder Mittelwertbildung) auf das Zielobjekt übertragen. Passend sind hier die Begriffe aus dem Data Mining *‘drill-down’* und *‘drill-up’*⁷. Entsprechend der Betrachtung auf räumlicher Ebene werden hier die Begriffe *‘spatial drill-down’* und *‘spatial drill-up’* als Bezeichnungen für das über die Karte gesteuerte *‘aggregieren’, ‘disaggregieren’* und visualisieren der Kennziffern und Daten der unterschiedlichsten aufeinander aufbauenden räumlichen Hierarchiestufen eingeführt.

Ähnlich der reinen Abfragefunktion *‘Innerhalb’* wird für eine definierte Menge an Punkten festgelegt, dass sie einer ebenfalls definierten Menge an Flächen zugewiesen wird. So lassen sich punktförmig vorliegende Daten den flächenhaften Daten (wie etwa

⁷ Im Data Mining bedeutet *‘drill-down’* das Betrachten von Daten auf einer übergeordneten Aggregationsstufe bis zu weiteren darunterliegenden Stufen, z. B. Anzeigen von Kennziffern übergeordneter Vertriebsgebiete und den jeweils einzelnen Kennziffern der zugeordneten Teilvertriebsgebiete.

Vertriebsgebiete) samt ihrer Attributdaten zuordnen und dort weiterverarbeiten oder mit anderen Daten auf Basis dieser Aggregationsstufe kombinieren. Es lassen sich die Flächenschwerpunkte (Zentroide) von Flächen ermitteln. Auch linienförmige und flächenhafte Strukturen können durch die oben genannte Methode Flächenobjekten zugewiesen werden (MAGUIRE 1995, S. 190).

6.3 Geostatistik

Die Besonderheit der Geostatistik ist, dass sie sich mit raumbezogenen Daten, d. h. Messungen oder Beobachtungen, denen Koordinaten zugeordnet sind, beschäftigt. In der Literatur wird die räumliche Statistik erst seit den 80er Jahren als eigenständiges Gebiet in Monographien behandelt: RIPLEY (1981, 1988), HAINING (1990), CRESSIE (1993), BAILEY & GATRELL (1995).

Räumliche Analyse ist neben der Visualisierung das wesentliche Element von GIS. Dennoch gibt es viele Verfahren der Geostatistik, die bisher in keinem GIS integriert sind⁸. Es gibt Überschneidungsgebiete von GIS und Geostatistik, aber keine Deckungsgleichheit. Die Verfahren der Geostatistik, von GIS und des Data Minings unterstützen sich gegenseitig (FOTHERINGHAM ET AL 2000, S. 61-64). Auch in dem mit diesem Vorhaben konzipierten und entwickelten *GBI-Tool* sind spezielle Methoden der räumlichen Statistik implementiert worden.

6.3.1 Analyse von raumbezogenen Daten und Attributdaten

Eine Besonderheit für die Analyse stellen raumbezogene Daten dar. Nach CRESSIE (1993) werden drei Typen von raumbezogenen Daten unterschieden: Räumlich kontinuierliche Daten (z. B. Niederschlagsmengen), Punktprozessdaten (Pflanzenstandorte, Krankheitsvorkommen, aber auch Einzelhandelsstandorte) und Regionaldaten (z. B. Anzahl der Bevölkerung, Einzelhändler in einem Gebiet). Wobei die Punktprozessdaten zu den 'Lokalen Daten', die Regionaldaten und die räumlich kontinuierlichen Daten zu den 'Attributdaten' gezählt werden (Abb. 6-10).

⁸ Einige Methoden wie die Point-Pattern Analyse oder Gravitationsmodelle sind in speziellen GIS-Modulen implementiert. Dazu zählen z. B. GIS-Extensions der Firma ESRI, wie ArcGIS® Spatial Analyst oder Filialinfo® der IVU, CommonGIS® des Fraunhofer Instituts IAIS und CrimeStat®, ein Geostatistik-Programm für die Analyse von Verbrechensorten vom National Institute of Justice, Washington DC & Ned Levine & Associates, Houston.

Die Ziele, die mit der Analyse raumbezogener Daten verfolgt werden, sind allerdings unabhängig von der Datenstruktur (vgl. DRESSMANN 2004, S. 71):

- Aufbereitung und Visualisierung der Daten für die Erkennung und Beschreibung von Strukturen und Mustern in den Daten.
- Identifizierung von Abweichungen von regelmäßigen Strukturen und Bewertung der Systematik.
- Berechnung von Prognosewerten, dort wo keine Messwerte zur Verfügung stehen.
- Anpassung statistischer Modelle an die Daten unter Berücksichtigung von Kovariablen.
-

DATENSTRUKTUR	UNIVARIATE Dimension	MULTIVARIATE Dimension
LOKALE DATEN (Punktprozessdaten)	Nächste-Nachbar-Methode (Nearest Neighbour-Methods), K-Funktion	Bivariate K-Funktionen, Raum-Zeit-Interaktionen
LOKALE DATEN und ATTRIBUTDATEN	Punktgedichteschätzung (Point Pattern Analysis), Kernel Density Estimation, Kernel Regression, Bayesian Smoothing	Angepasste Punktgedichteschätzungen (Adaptive Kernel Density Estimation), Bayesian Smoothing
ATTRIBUTDATEN (Regionaldaten, Räumlich-kontinuierliche Daten)	Räumliche Autokorrelation, Variogramme	Multivariate räumliche Korrelation
	Trendoberflächenanalysen, Kriging	Räumliche Regression, Co-Kriging, Raumzeitliche Modelle
		Räumlich lineare Modellierung (Spatial general linear modelling), Canonical Correlation, Multidimensional Scaling
LOKALE DATEN und ATTRIBUTDATEN		Hierarchische Cluster Analyse, Two-Step Cluster Analyse
		Entscheidungsbäume (Answer Trees), CHAID
		Gravitationsmodelle

Quelle: Zusammenstellung in Anlehnung an Bailey (1998)

Abb. 6-10: Komplexe räumlich-statistische Analyseverfahren

Die Abbildung (Abb. 6-10) dient als Überblick komplexer räumlich-statistischer Analysemethoden, die in Verbindung mit GIS, Geomarketing bzw. 'Geo Business Intelligence' von Bedeutung sein können.

Viele Untersuchungen zu diesen Verfahren werden im Bereich des Gesundheitswesens genutzt, insbesondere in der Epidemiologie. Ursächlich geht der Ursprung der Methoden auf Untersuchungen von Gesundheitsdaten durch SNOW (1848) zurück. Er untersuchte den Cholera-Ausbruch in London und identifizierte mittels der genannten Verfahren eine Wasserpumpe als Verursacher für die Verbreitung.

Analysen von Attributdaten erfolgen in einem Informationssystem, wenn integrierte unternehmensinterne und -externe Attribute mit Kennwerten charakterisiert und miteinander in Beziehung gesetzt werden. Beispielsweise werden Vertriebsserfolge in Zusammenhang mit sozioökonomischen Daten des Umfeldes der Filiale analysiert. Es werden Umsätze mit Kundenfrequenzen und mit weiteren unternehmensexternen Daten verknüpft, analysiert und kartographisch dargestellt. Entscheidend ist, dass nicht nur einfache statistische Verfahren bzw. univariate Methoden zur Datenanalyse in GIS angewendet werden, sondern auch die multivariaten statistischen Verfahren sowie Ansätze des Data Minings Berücksichtigung finden, die z. B. in der vorliegenden Untersuchung für die Ermittlung von Erfolgsfaktoren herangezogen werden.

Im Weiteren wird auf die Analyse von Attributdaten eingegangen, bis hin zur Clusteranalyse und zum Einsatz von Entscheidungsbäumen. Es geht darum, eine optimale Kombination dieser Methoden mit denen, die auf räumliche Daten ausgerichtet sind, zu finden.⁹

Zur Anwendung kommen verschiedene mathematisch-statistische Verfahren, wie sie seit den sechziger Jahren in der *'Quantitativen Geographie'* eingesetzt werden (vgl. BAHRENBERG, GIESE & NIPPER 1992, 1999). Diese Methoden können zunächst anhand der Zahl der betrachteten Variablen klassifiziert werden.

Die univariaten Methoden beschreiben die Ausprägungen von Werten eines einzigen Attributes¹⁰. Geht es darum, den Zusammenhang zwischen zwei Variablen zu messen (bivariate Methoden), werden einfache Korrelations- und Regressionsanalysen¹¹ eingesetzt. Multivariate Methoden stellen eine Erweiterung der bivariaten Analyse dar.

⁹ Auf die Methoden, die vor allem im GIS-Bereich Anwendung finden, die z. B. auf Interpolationen beruhen wie *Point Pattern Analysis* bzw. *Punkte-, Kerneldichteschätzungen* und auf *Gravitationsmodelle* wird an späterer Stelle ausführlich eingegangen (*Kapitel Fehler! Verweisquelle konnte nicht gefunden werden.*).

¹⁰ Dieses geschieht meist durch Häufigkeiten (Modus), Anteilswerte (Quartile), Verteilungen (Schiefe), Maße der Zentraltendenz (arithmetisches Mittel, Median) sowie Streuungsmaße (Abweichung der Variablenwerte vom Mittelwert).

¹¹ Die Korrelation misst die Stärke und die Richtung des Zusammenhanges der betrachteten Größen, während die Regression die funktionale Abhängigkeit der einen Variable von der anderen Variable ermittelt.

BAHRENBERG, GIESE & NIPPER (1992, S. 14) unterscheiden zwischen Verfahren, welche die Beziehungen zwischen den beobachteten und gemessenen Größen prüfen, und solchen Verfahren, die der Aufdeckung von Strukturen dienen, die hinter den Variablen stehen, sobald sie miteinander in Beziehung gesetzt werden.

Statistischer Oberbegriff	
Univariate Methoden	Maße der Zentraltendenz, Streuungsmaße
Bivariate Methoden	Korrelationsanalyse
	Regressionsanalyse
Multivariate Methoden	Multiple Korrelationsanalyse
	Multiple Regressionsanalyse
	Varianzanalyse
	Hauptkomponentenanalyse
	Faktorenanalyse
	Diskriminanzanalyse
Data Mining	Hierarchische Clusteranalyse
	Two-Step-Clusteranalyse
	Entscheidungsbäume (CHAID, C&RT, QUEST u.a.)
	Neuronale Netze

Quelle: erweitert nach Schüssler 2000, S. 68

Abb. 6-11: Mathematisch-statistische Methoden im Geomarketing

Zur multivariaten Analyse direkt gemessener Größen werden je nach vorliegenden Skalenniveaus der Attribute multiple Korrelations-, Regressions- und Varianzanalysen eingesetzt. Hauptkomponenten- und Faktorenanalysen stellen methodische Instrumente dar, um aus den beobachteten oder gemessenen Variablen komplexere, nicht direkt ermittelbare Größen abzuleiten (Dimensionsreduktion).

Für die Zwecke der Segmentation werden Verfahren der Clusteranalyse eingesetzt. Hierbei wird versucht, Gruppen von ähnlich strukturierten Fällen (hier: Raumeinheiten) zu finden. Raumeinheiten mit ähnlichen Ausprägungen von mehreren Attributen werden homogenen Gruppen zugeordnet, indem deren Ähnlichkeit durch die Lage der Raumeinheiten in einem von mehreren Variablen aufgespannten Koordinatensystem bestimmt wird (BACKHAUS 1987, FOTHERINGHAM U.A. 2000, S. 188ff). Anhand der Clusteranalyse können regelhafte Muster (z. B. Filialen mit verschiedenem sozioökonomischen Umfeld) ermittelt und zur Entscheidungsunterstützung hinsichtlich geplanter unternehmerischer Aktivitäten herangezogen werden (BATEY & BROWN 1995, S. 95). So können z. B. Standort- und Sortimentsoptimierung durchgeführt werden,

nachdem eine genauere Kenntnis über die Typisierung des voraussichtlichen potentiellen Kundenumfeldes erlangt wurde.

Als exploratives Verfahren werden bei der Clusteranalyse keine Gruppen vordefiniert. Die Zahl der Cluster und die Zugehörigkeit der Untersuchungsobjekte in eines der Cluster stellt das eigentliche Ergebnis der Analyse dar. Im Gegensatz dazu untersucht die Diskriminanzanalyse eine vorgegebene Gruppierung im Hinblick auf deren bestmögliche Trennung, beispielsweise bei der Prüfung der Kundeneigenschaften von Zeitungsabonnenten in Bezug auf Kündiger und Nicht-Kündiger.

Bei hinreichend großen Datensätzen und einer Vielzahl von unterschiedlich skalierten Variablen kommen die Verfahren des Data Mining zum Einsatz, die automatisch nach der optimalen Anzahl von Segmenten im Datenbestand suchen (*Two-Step-Cluster*), den Datenbestand bezüglich eines Zielkriteriums nach automatisch bestimmten Variablenwichtigkeiten aufsplitten (Entscheidungsbäume wie *CHAID*, *C&RT*) oder Klassifikationsmodelle mit großer Komplexität erzeugen (Neuronale Netze). Auf das Verfahren und die Begründung des Einsatzes der *Two-Step-Clusterung* und der Entscheidungsbäume wird in *Kapitel 7* detailliert eingegangen.

6.3.2 Regionalisierung und Autokorrelation

Regionaldaten beziehen sich in der Regel auf administrative Einheiten oder aber auf künstlich geschaffene Regionen wie Marktzellen. Es handelt sich meist um Durchschnitts-, Summen-, Index- oder klassifizierte Werte des Gebietes, z. B. durchschnittliche Kaufkraft in einem Postleitzahlgebiet oder vorwiegender Soziotyp. Diese Daten erhalten bereits durch eine einfache Visualisierung eine Interpretierbarkeit und schaffen so eine Vergleichbarkeit zwischen Regionen. Verschiedene räumliche Statistiken können auf Regionaldaten angewandt werden (z. B. Autokorrelation *Morans I*), die aufzeigen, inwieweit benachbarte Gebiete räumlich-statistisch miteinander korrelieren. Regionaldaten werden zum Teil von öffentlichen Statistischen Ämtern bereit gestellt. In einer höheren Detailtiefe werden Daten über Marktforschungsinstitutionen oder private Datenbroker bezogen, oder es werden spezielle Erhebungen beauftragt.

Während bei geostatistischen Daten die Korrelation als Funktion der Distanz zweier Messpunkte analysiert wird, kommt für Regionaldaten ein Nachbarschaftsprinzip zum Tragen, welches untersucht, ob Regionen auf der Basis der Variablenwerte eines räumlichen Prozesses zueinander benachbart sind. Wenn die Werte eines räumlichen

Prozesses in benachbarten Regionen miteinander korrelieren, korreliert der Prozess mit sich selber: 'Räumliche Autokorrelation' (DRESSMANN 2004, S. 86). Die Anwendung der Autokorrelation¹² ist immer dann wichtig, wenn regionale Abhängigkeiten und Nachbarschaftseffekte festgestellt werden sollen. Die Messung der Autokorrelation erfolgt beispielsweise durch MORANS I oder GEARYS C¹³:

$$I = \frac{N \cdot \sum_{i=1}^N \sum_{j=1}^N w_{ij} (z_i - \bar{z})(z_j - \bar{z})}{\left(\sum_{i=1}^N (z_i - \bar{z})^2 \right) \cdot \left(\sum_{i=1}^N \sum_{j \neq i}^N w_{ij} \right)}$$

$$C = \frac{(N-1) \sum_{i=1}^N \sum_{j=1}^N w_{ij} (z_i - z_j)^2}{\left(\sum_{i=1}^N (z_i - \bar{z})^2 \right) \cdot \left(\sum_{i=1}^N \sum_{j \neq i}^N w_{ij} \right)}$$

N	Anzahl Fälle
Z	untersuchte räumliche Prozesse
Z _i	Variablenwert an einem bestimmten Ort i
Z _j	Variablenwert an einem bestimmten anderen Ort, wobei i ≠ j
W _{ij}	Gewichtungsfaktor, der von der räumlichen Beziehung zwischen dem Ort i und dem Ort j abhängig ist
W _{ij}	= 1 falls die Orte Z _i und Z _j benachbart sind
W _{ij}	= 0 falls keine Nachbarschaft vorliegt

Insbesondere Anwendungen und Besonderheiten von *Morans I* sind in Arbeiten von TIEFELSDORF (1995, 1997, 1998, 1999 und 2000) untersucht und weiterentwickelt worden. Im Hinblick auf die Verknüpfung zum GIS-Bereich ist die Anwendung von *Morans I* auf die Weiterentwicklung auf die Thyssen-Polygon-Struktur (*Tesselation*) (BOOTS, TIEFELSDORF 2000) übertragen worden. Anwendungen der Autokorrelation nach *Morans I* für den Raum Berlin, der auch in der folgenden Untersuchung zu einem der Pilotgebiete gewählt wurde, finden sich bei BRAUN, TIEFELSDORF (1993, 1996, 1998). Hier werden die Wanderungsbewegungen innerhalb der Bezirke Berlins zwischen 1991-97 untersucht und durch die Autokorrelation nach *Morans I* wird u.a. nachgewiesen, dass erst seit 1995 wieder eine Stabilisierung einer geregelten zentrifugalen Migration eingetreten ist.

Allerdings ist der Test auf räumliche Autokorrelation nicht geeignet, die Lokalisation einzelner homogener Gebiete (Cluster) in der Karte zu identifizieren (LAWSON 2001),

¹² Räumliche Autokorrelation ist als Analysemethodik insbesondere in statistischen Paketen wie SPSS, SAS, S-Plus, CrimeStat integriert.

¹³ Morans I entspricht dem Konzept des gewöhnlichen Korrelationskoeffizienten, wobei der Wertebereich nicht exakt auf das Intervall (-1, 1) begrenzt ist. Keine Korrelation existiert zwischen benachbarten Regionen, wenn der Wert um 0 liegt. Je größer der Wert, umso stärker ist die Korrelation, um so mehr tendieren benachbarte Regionen dazu, ähnliche Werte bezüglich des räumlichen Prozesses aufzuweisen. Ein negativer Moran I bedeutet, dass benachbarte Regionen dazu neigen, gegenläufige Werte anzunehmen. Gearys C nimmt Werte von 0 bis +2 an. Wenn der C-Wert gegen 2 geht, existiert keine räumliche Autokorrelation. Geht C gegen 0, so liegt eine starke räumliche Autokorrelation vor (vgl. DRESSMANN 2004, S. 87; BRAUN & TIEFELSDORF 1998).

weshalb sie hier im weiteren nicht zur Anwendung kommt. Grundsätzlich geht es bei der Clusteranalyse primär nicht um die Berücksichtigung von regionalen Abhängigkeiten und Nachbarschaftseffekten auf Basis eines einzelnen räumlichen Prozesses, sondern um die wesentlich kompaktere Beurteilung der Ähnlichkeiten von Messpunkten¹⁴ auf Basis mehrerer individueller Klassifikationsmerkmale.

So soll es in der vorliegenden Untersuchung weniger um die Abbildung von Regionalstrukturen in homogenen Räumen gehen, als vielmehr um das Auffinden von vergleichbaren Standorten hinsichtlich verschiedener Merkmale. Dies kann losgelöst von Nachbarschaftseffekten geschehen, um so auch die Übertragbarkeit der Informationen einzig auf Basis der Klassifikationsmerkmale zu gewährleisten. Wie die Ergebnisse der *Two-Step-Clusteranalyse* im Folgenden zeigen (*siehe Kapitel 7.1: Abb. 8, Abb. 9, Abb. 10, Abb. 11, Abb. 12*), konnten durch die Anwendung dieses multivariaten statistischen Verfahrens auch ohne Berücksichtigung von Nachbarschaftseffekten durchaus homogene Großräume (hier mehrere PLZ-Gebiete) im nachbarschaftlichen Verbund ausgewiesen werden. So werden beispielsweise bestimmte benachbarte städtische PLZ-Gebiete zusammengefasst, die eine ähnliche Struktur hinsichtlich mehrerer Eigenschaftsmerkmale aufweisen und somit indirekt auf das Vorliegen von wenigstens einem räumlichen Prozess hindeuten.

¹⁴ Da der Schwerpunkt der Untersuchung auf dem städtischen Raum bzw. auf Regionen mit hoher Einwohnerdichte liegt, werden hier die kleinräumigeren PLZ-Gebiete den Gemeinden vorgezogen.