

Michael Wurm

# Verknüpfung von Fernerkundungsdaten und Survey-Daten (SOEP und BASE-II) in städtischen Räumen für sozialwissenschaftliche Analysen

DISSERTATION  
zur Erlangung des akademischen Grades  
Doktor der Naturwissenschaften



Betreuer:

Dipl.-Forstw. Prof. Dr. Mathias Schardt  
Institut für Fernerkundung und Photogrammetrie, Technische Universität Graz

Dipl.-Volksw. Prof. Dr. Gert G. Wagner  
Institut für Volkswirtschaftslehre und Wirtschaftsrecht, Technische Universität Berlin  
Deutsches Institut für Wirtschaftsforschung Berlin

Graz, Jänner 2013





**[...] bei gleicher Umgebung lebt doch jeder in einer andern Welt.**

*Arthur Schopenhauer*



# Eidesstattliche Erklärung

Ich erkläre an Eides statt, dass ich die vorliegende Arbeit selbstständig verfasst, andere als die angegebenen Quellen/Hilfsmittel nicht benutzt und die den benutzten Quellen wörtlich und inhaltlich entnommenen Stellen als solche kenntlich gemacht habe.

Graz, \_\_\_\_\_

# Statutory Declaration

I declare that I have authored this thesis independently, that I have not used other than the declared sources/resources and that I have explicitly marked all material which has been quoted either literally or by content from the used sources.

Graz, \_\_\_\_\_



# Danksagung

An dieser Stelle möchte ich mich bei allen, die zum Gelingen dieser Arbeit beigetragen haben, sehr herzlich bedanken.

In erster Linie danke ich Prof. Mathias Schardt, der die Betreuung der Arbeit übernommen hat. Durch seine fachlichen Beiträge und kritischen Fragen behielt ich den Fokus der Arbeit stets im Blickfeld und bekam eine andere Sicht auf die Dinge. Darüber hinaus bedanke ich mich bei ihm für die entgegenkommende Art der Betreuung, auch über große Distanzen hinweg. Des Weiteren möchte ich dem Zweitbetreuer der Arbeit, Prof. Gert G. Wagner vom Deutschen Institut für Wirtschaftsforschung Berlin (DIW) für sein uneingeschränktes Entgegenkommen, die einmalige Kooperation und die Unterstützung danken.

Diese Arbeit wurde am Deutschen Fernerkundungsdatenzentrum (DFD) im Deutschen Zentrum für Luft- und Raumfahrt e.V. (DLR) in Oberpfaffenhofen, in Kooperation mit dem Lehrstuhl für Fernerkundung der Universität Würzburg und am DIW Berlin erstellt. Ich möchte mich bei allen Mitarbeitern in diesen Einrichtungen für die herzliche Arbeitsatmosphäre und für die vielen Gespräche fachlicher und nicht-fachlicher Natur bedanken. Sowohl auf fachlicher als auch auf menschlicher Ebene konnte ich immer auf die Unterstützung von Dr. Hannes Taubenböck zählen. Sein großes Engagement und die Diskussionen, auch fernab der Fernerkundung trugen erheblich zum Gelingen dieser Arbeit bei.

Bei Dr. Jan Goebel vom DIW Berlin bedanke ich mich besonders herzlich für die vielen Gespräche über das SOEP, statistische Datenauswertung,  $R$ , die Hilfestellung bei der Datenverknüpfung und vor allem für seine Geduld. Prof. Stefan Dech, Andreas Müller, Dr. Thomas Esch und Achim Roth danke ich für die mir gebotenen Entfaltungsmöglichkeiten.

Dr.-Ing. Andreas Schmitt hat mich in statistischen und mathematischen Fragestellungen stets großartig unterstützt und von Dr. Martin Wegmann bekam ich viele wertvolle Tipps. Martin Habermeyer danke ich für die kritischen Kommentare. Ihm, Andreas Felbier, Christian Geiß und Igor Klein danke ich für Diskussionen über die Fernerkundung und darüber hinaus.

Für die kostenlose Bereitstellung von Daten bedanke ich mich bei Rolf Annecke vom Referat für Gesundheit und Umwelt der Stadt München, Mathias Boes von European Space Imaging, Thomas Schneider und Jörn Welsch von der Senatsverwaltung für Stadtentwicklung der Stadt Berlin, Frank Lehmann, Corinna Brüßhaber und Anna-Maria Trosset vom DLR Berlin-Adlershof und Udo Lapschies vom Umwelt- und Verbraucherschutzamt der Stadt Köln.

Mein persönlicher Dank richtet sich an meine Mutter Maria Wurm für ihre stets vorbehaltlose Unterstützung durch das Studium und die Promotion und vor allem an Anne-Sophie Hérisset für den immerwährenden Rückhalt und ihre unendliche Geduld.



# Zusammenfassung

Zwischen dem Menschen und seiner Umwelt besteht eine wechselseitige Beziehung. Durch sein Handeln verändert der Mensch seine Umwelt und er wird durch die Umgebung in seinem Verhalten im Raum beeinflusst. Deshalb stellt der Raum für die Erforschung von sozialen Fragestellungen eine bedeutende Rolle dar. Für die physische Erfassung des Raumes und seiner Merkmale hat sich die Erdbeobachtung als effizientes Werkzeug etabliert.

*Vor diesem Hintergrund ist das übergeordnete Ziel dieser Dissertation die Ableitung räumlicher Merkmale städtischer Räume aus Fernerkundungsdaten für die Einbindung in sozialwissenschaftliche Studien am Beispiel der beiden Städte Berlin und München.*

Für die Erarbeitung dieses Ziels wird in der vorliegenden Arbeit ein konzeptioneller Rahmen erstellt, der die Verknüpfung von Erdbeobachtungsdaten und sozialwissenschaftlichen Daten der sozioökonomischen Langzeitstudien SOEP und BASE-II beschreibt. Die Adressdaten der Teilnehmer können zum ersten Mal in der Geschichte der Studien räumlich verarbeitet werden, ohne dabei den Datenschutz zu verletzen.

Die physische Charakterisierung des Raumes erfolgt auf Basis hoch aufgelöster, optischer Satellitenbilddaten für die Erfassung der Landbedeckung und digitaler Oberflächenmodelle (DOM) für die Ableitung von Einzelgebäuden, welche die Bebauungsstruktur beschreiben. Dabei werden in einem objektorientierten Bildanalyseverfahren die Informationen aus dem Satellitenbild und dem DOM komplementär verwendet. Für die Einzelgebäude wird darüber hinaus auch noch ein Verfahren zur Abschätzung der Geschosshöhe entwickelt und einzelne Gebäudetypen werden auf Basis von 2D und 3D Formmerkmalen klassifiziert. Die aus den Fernerkundungsdaten abgeleiteten physischen Merkmale werden auf der räumlich übergeordneten Ebene des Baublocks zu Nachbarschaftsmerkmalen zusammengefasst und stellen die räumlichen Kontextmerkmale für die sozialwissenschaftliche Analyse dar.

Diese Nachbarschaftsmerkmale werden mit den Daten sozialwissenschaftlicher Surveys verknüpft, wodurch im erstellten Datensatz sowohl die sozioökonomischen Merkmale als auch die räumlichen Kontextmerkmale aus den Fernerkundungsdaten zur Verfügung stehen. Durch die Analyse beider Informationen werden die unterschiedlichen räumlichen Bezüge der Surveydaten und der räumlichen Kontextmerkmale untersucht. Konkret wird anhand einer empirischen Untersuchung zur innerstädtischen, räumlichen Segregation des Haushaltseinkommens sowohl die Möglichkeit zur technischen Integrierung der räumlichen Merkmale evaluiert, als auch quantitative Beziehungen zwischen sozioökonomischen und räumlichen Merkmalen untersucht. Sie zeigen, dass der Raum in Kombination mit Informationen über Personen und Haushaltsdaten aus den Surveys einen relevanten, zusätzlichen Erklärungsgehalt in statistischen, sozialwissenschaftlichen Modellen aufweist.



# Abstract

Humans change and shape the landscape. There exists a reciprocal relationship, whereby the natural environment is affected by human activities, while the physical landscape affects the behavior of people. This provides context for social scientific research. For the physical detection of space and its characteristics, earth observation can be considered as an established tool.

*Against this background, the major goal in this dissertation is the derivation of spatial features in urban areas derived by remote sensing data for the integration into socio-economic studies by the example of the two cities, Berlin and Munich.*

To reach this goal, a conceptual framework is developed in the context of this thesis. This framework describes the linking of remote sensing data and social scientific data from the socio-economical surveys SOEP and BASE-II. The address information of the participants of these studies can be processed spatially for the first time in the history of these studies, without impacting individual data protection.

The physical characterization of space is achieved with high resolution satellite images and digital surface models. These data sets are used to derive land cover classifications and individual building footprints, including their height to describe the development structure. An object-based image analysis procedure is then applied to extract the complementary information from both data sets. For the individual buildings another procedure is applied to estimate the number of building levels and several building types are classified by 2D and 3D features. All the derived physical features from the remote sensing data are summarized on the spatial level of the building block and represent the contextual features for the social scientific analysis.

Those contextual features are linked with the data of the socio-economic surveys. Hence, both information from the contextual features and socio-economic features from the surveys are accessible for the analysis. By analyzing both information types, the various spatial references are investigated. Specifically, an empirical study is performed to investigate the intra-urban spatial segregation of household incomes. Additionally, potential for the technical integration of spatial features, as well as the quantitative relationships between socio-economic and spatial features, are investigated. They show that space, in combination with information about individuals and households from the surveys, shows additional, relevant explanatory information in statistical social scientific models.



# Inhaltsverzeichnis

Danksagung .....	vii
Zusammenfassung.....	ix
Abstract.....	xi
Abbildungsverzeichnis .....	xvii
Tabellenverzeichnis.....	xx
Abkürzungsverzeichnis .....	xxiii
<b>1. Problemstellung und Zielsetzung .....</b>	<b>1</b>
1.1. Einleitung .....	1
1.2. Herausforderungen der Arbeit .....	5
1.3. Zielsetzung.....	8
1.4. Gliederung der Arbeit .....	9
<b>2. Grundlagen.....</b>	<b>11</b>
2.1. Eigenschaften von Fernerkundungsdaten.....	11
2.2. Stand der Forschung.....	13
2.2.1. Landbedeckungsklassifikation von optischen Fernerkundungsdaten in städtischen Räumen.....	13
2.2.2. Ableitung von Gebäuden aus digitalen Oberflächenmodellen.....	16
2.2.3. Ableitung von Gebäuden aus der Kombination von digitalen Oberflächenmodellen und Spektraldaten.....	17
2.2.4. Ableitung von semantischen Bebauungsklassen.....	18
2.2.4.1. Gebäudeebene .....	19
2.2.4.2. Strukturebene .....	20
2.2.5. Erdbeobachtungsdaten für die sozialwissenschaftliche Forschung .....	24
2.3. Sozialwissenschaftliche Surveydaten: SOEP und BASE-II.....	27
2.3.1. Die SOEP-Studie .....	27
2.3.2. Der räumliche Bezug der SOEP-Daten .....	30
2.3.2.1. Landkreise.....	32
2.3.2.2. Postleitzahlenbezirke.....	33
2.3.2.3. Mikrogeographische Nachbarschaften.....	33
2.3.2.4. Andere Langzeitstudien.....	34
2.3.2.5. SOEP-Studien und Fernerkundung.....	34

2.3.3. SOEP-Variablen mit indirektem Raumbezug .....	35
2.3.4. BASE-II .....	37
2.4. Innovationen der Dissertation aus der Synthese des Forschungsstandes.....	38
2.4.1. Identifizierte Forschungslücken aus dem Forschungsstand.....	38
2.4.2. Innovation der vorliegenden Arbeit.....	40
2.5. Übersicht der Arbeitsschritte.....	41
2.6. Zusammenfassung.....	42
<b>3. Konzeptioneller Rahmen und Geodaten .....</b>	<b>43</b>
3.1. Konzeptioneller Rahmen .....	43
3.1.1. Technische Verknüpfung von Geodaten und SOEP/BASE-II-Daten.....	43
3.1.2. Die räumliche Bezugsebene für die Datenverknüpfung .....	45
3.2. Untersuchungsgebiete .....	47
3.3. Geodatengrundlagen .....	49
3.3.1. Fernerkundungsdaten .....	49
3.3.1.1. Satellitenbilddaten.....	50
3.3.1.2. Digitale Oberflächenmodelle .....	51
3.3.2. Weitere Geodaten.....	55
3.3.2.1. Digitale Geländemodelle .....	55
3.3.2.2. Baublöcke/Stadtstrukturtypenklassifikation.....	55
3.3.2.3. Digitale Orthofotos .....	56
3.3.2.4. Digitale Gebäudemodelle aus dem Kataster.....	56
3.4. Validierung der Klassifikationsergebnisse .....	57
3.5. Zusammenfassung.....	61
<b>4. Erfassung der Gebäudestruktur und der Landbedeckung in den</b>	
<b>Testgebieten .....</b>	<b>63</b>
4.1. Datenvorverarbeitung .....	64
4.2. Ableitung der Gebäudestruktur.....	67
4.2.1. Normalisierung der digitalen Oberflächenmodelle .....	67
4.2.2. Eliminierung der Vegetation .....	70
4.2.3. Abgrenzung der Einzelgebäude .....	72
4.2.4. Ergebnisse der Gebäudeabgrenzung.....	76
4.2.5. Ableitung der Gebäudehöhen .....	81
4.2.6. Abschätzung der Geschosshöhe.....	82
4.2.7. Zusammenfassung für die Gebäudestruktur .....	87
4.3. Ableitung der urbanen Landbedeckung.....	88
4.3.1. Objektbasierte Bildklassifikation .....	88
4.3.2. Klassifikationsergebnisse .....	91

4.3.3. Genauigkeitsabschätzung.....	93
4.4. Zusammenfassung zur Erfassung der Gebäudestruktur und der Landbedeckung...	95
<b>5. Gebäudeklassifikation und Aggregation von Kontextmerkmalen .....</b>	<b>97</b>
5.1. Klassifikation der dominierenden Gebäudeklasse .....	97
5.1.1. Formmerkmale für die Beschreibung der Gebäudephysiognomie.....	100
5.1.1.1. Einfache Formmerkmale .....	100
5.1.1.2. Kreisapproximation .....	101
5.1.1.3. Quadratapproximation.....	102
5.1.1.4. Rechteckapproximation.....	103
5.1.1.5. Würfelapproximation .....	104
5.1.1.6. Weitere Merkmale.....	105
5.1.2. Häufigkeitsverteilung der Formmerkmale .....	105
5.1.3. Semantische Klassifikation von Gebäuden auf Basis von Formmerkmalen und linearer Diskriminanzfunktion .....	110
5.2. Bebauungs- und Geschossflächendichte.....	123
5.3. Vegetations- und Versiegelungsanteil.....	124
5.4. Lage.....	125
5.5. Zusammenfassung zur Gebäudeklassifikation und Aggregation zu Kontextmerkmalen.....	126
<b>6. Sozialwissenschaftliche Analysepotenziale der Kontextmerkmale   und der SOEP/BASE-II-Daten .....</b>	<b>129</b>
6.1. Verknüpfung der Kontextmerkmale und der Surveydaten.....	130
6.1.1. Stichprobenumfang .....	130
6.1.2. Die Merkmale im verknüpften Datensatz .....	131
6.2. Der räumliche Bezug der Surveydaten und der räumlichen Kontextmerkmale .....	135
6.2.1. Lage.....	135
6.2.1.1. Personenebene .....	136
6.2.1.2. Haushaltsebene .....	137
6.2.1.3. Interpretation des emotionalen Raumbezuges für München.....	138
6.2.2. Grünanteil.....	141
6.2.3. Art der Bebauung.....	143
6.2.4. Zusammenfassung zum räumlichen Bezug .....	145
6.3. Quantitative empirische Analsepotenziale.....	147
6.3.1. Die lineare Regression.....	148
6.3.2. Exemplarische Analyse zum Haushaltseinkommen .....	150
6.3.2.1. Hintergrund .....	150
6.3.2.2. Datenaufbereitung .....	151
6.3.2.3. Modellschätzungen.....	155

6.3.2.4. Ergebnisse.....	158
6.3.2.5. Diskussion .....	159
<b>7. Zusammenfassung und Ausblick .....</b>	<b>161</b>
7.1. Zusammenfassende Betrachtung .....	161
7.2. Zukünftiger Forschungsbedarf.....	165
<b>Anhang A: Fernerkundungsdaten .....</b>	<b>169</b>
<b>Anhang B: Boxplots zur Geschosshöhenabschätzung für Berlin und München .....</b>	<b>173</b>
<b>Anhang C: Addendum zu Kapitel 6.....</b>	<b>174</b>
<b>Anhang D: Tabellen .....</b>	<b>176</b>
<b>Literaturverzeichnis .....</b>	<b>186</b>

# Abbildungsverzeichnis

## Kapitel 2

Abbildung 1: Bilddaten unterschiedlicher fernerkundlicher Aufnahmesysteme. ....	13
Abbildung 2: Rasterbild und erzeugte Objekte durch Multiresolutionsegmentierung.....	14
Abbildung 3: Unterschiedliche Bebauungsstrukturen am Beispiel Münchens. ....	18
Abbildung 4: Sonderstichproben und Teilnehmerentwicklung der ersten 27 Befragungswellen im SOEP. ....	30
Abbildung 5: Anzahl der SOEP-Haushalte pro Landkreis.....	31
Abbildung 6: Auszug aus dem Haushaltsfragebogen 2009 zur subjektiven Einschätzung gegenüber der Lage der Wohngegend des Haushaltes.....	35
Abbildung 7: Auszug aus dem Haushaltsfragebogen 2009 zum Wohnumfeld (WUM1).....	36
Abbildung 8: Auszug aus dem Haushaltsfragebogen 2009 zum Wohnumfeld (WUM2).....	36
Abbildung 9: Auszug aus dem Haushaltsfragebogen 2009 zum Wohnumfeld (WUM3).....	36
Abbildung 10: Auszug aus dem Haushaltsfragebogen 2009 zur subjektiven Einschätzung gegenüber Umwelteinflüssen. ....	37
Abbildung 11: Graphische Übersicht über die Arbeitsschritte. ....	41

## Kapitel 3

Abbildung 12: Schema der Verknüpfung von Geodaten und SOEP/BASE-II-Daten. ....	44
Abbildung 13: Maßstabsabhängige Strukturen im urbanen Raum.....	45
Abbildung 14: Das räumliche Konzept zur Integration von Nachbarschaftsmerkmalen in die Surveydaten. ....	47
Abbildung 15: Lage der Untersuchungsgebiete in Deutschland.....	49
Abbildung 16: Räumliche Abdeckung der Fernerkundungsdatengrundlagen und ihre beispielhafte Darstellung für die beiden Untersuchungsgebiete München und Berlin.....	54
Abbildung 17: Ausschnitt aus der Stadtstrukturtypenklassifikation für München.....	56
Abbildung 18: Erstellung des Trainings- und Validierungsdatensatzes auf Gebäudeebene anhand der Stadtstrukturtypenklassifizierung. ....	60

## Kapitel 4

Abbildung 19: Darstellung des ehemaligen Berliner Flughafen Tempelhof in der Quickbird-Aufnahme in einer Falschfarbeninfrarotdarstellung und im digitalen Oberflächenmodell.....	64
Abbildung 20: Lage von Objekten in den Satellitenbilddaten und den digitalen Oberflächenmodellen ....	65
Abbildung 21: Ikonos-Aufnahme für München vor und nach der Datenvorverarbeitung. ....	66
Abbildung 22: Vergleich zwischen Höheninformation im DGM und im DOM. ....	68
Abbildung 23: Normalisierung von digitalen Oberflächenmodellen. ....	68
Abbildung 24: Normalisiertes digitales Oberflächenmodell (nDOM).....	69
Abbildung 25: Auswirkungen hoher Vegetation in nächster Nachbarschaft zu Häusern auf die Berechnung des kleinsten und höchsten Grauwertes innerhalb eines Filterfensters.....	70

Abbildung 26: Ableitung der Vegetationsmaske auf Basis der Reflexion im nahen Infrarot und Überlagerung mit dem nDOM.....	72
Abbildung 27: Vergleich der Blockgrenzen zwischen ATKIS Basis-DLM und Baublöcken für einen Ausschnitt im Untersuchungsgebiet München. ....	73
Abbildung 28: Ablauf der Abgrenzung der Einzelgebäude. ....	73
Abbildung 29: Berechnung des trennenden Schwellwertes auf Basis des besten Kontrastes. ....	74
Abbildung 30: Anwendung des in Tabelle 8 berechneten Schwellwertes für die Abgrenzung von Gebäuden.....	76
Abbildung 31: Ergebnisse der Gebäudeabgrenzung für München und Berlin.....	77
Abbildung 32: Vergleich der abgegrenzten Gebäudegrundflächen mit dem amtlichen Gebäudemodell für München.....	79
Abbildung 33: Vergleich der abgegrenzten Gebäudegrundflächen mit dem amtlichen Gebäudemodell für Berlin.....	79
Abbildung 34: Gebäudegrundflächen und Gebäudehöhen für die Stadtzentren von München und Berlin. ....	81
Abbildung 35: Gebäudetypenabhängige Darstellung der Gebäudehöhen pro Geschosszahl für München und Berlin. ....	84
Abbildung 36: Darstellung der Gebäudehöhen pro Geschosszahl für alle Gebäude.....	86
Abbildung 37: Zusammenhang der Parameter für die Multiresolutionsegmentierung in eCognition. ....	89
Abbildung 38: Die einzelnen Ablaufschritte zur Klassifikation der Landbedeckung.....	90
Abbildung 39: Ergebnis der Landbedeckungsklassifikation für Berlin. ....	91

## Kapitel 5

Abbildung 40: Ergebnis der Landbedeckungsklassifikation für München. ....	92
Abbildung 41: Vergleich der fünf unterschiedlichen Gebäudetypen zwischen Schrägluftbilddaufnahme und 3D-Gebäudemodell. ....	98
Abbildung 42: Hierarchischer Ablauf für die Klassifikation der Gebäude. ....	99
Abbildung 43: Häufigkeitsverteilung der Formmerkmale für die 4 Gebäudeklassen für das Untersuchungsgebiet Berlin. ....	106
Abbildung 44: Häufigkeitsverteilung der Formmerkmale für die 4 Gebäudeklassen für das Untersuchungsgebiet München.....	108
Abbildung 45: Untersuchung der Gruppenunterschiede und Klassifikation neuer Objekte.....	111
Abbildung 46: Geometrische Darstellung der Diskriminanzachse an einem Zwei-Gruppen-Fall.....	117
Abbildung 47: Geometrische Abgrenzung zwischen den Gebäudeklassen auf Basis der 3 linearen Diskriminanzfunktionen (LD1, LD2 und LD3) für Berlin.....	118
Abbildung 48: Geometrische Abgrenzung zwischen den Gebäudeklassen auf Basis der 3 Diskriminanzfunktionen (LD1, LD2 und LD3) für München.....	119
Abbildung 49: Ergebnis der Klassifikation der Bebauungstypen für Berlin. ....	121
Abbildung 50: Ergebnis der Klassifikation der Bebauungstypen für München. ....	122
Abbildung 51: Bebauungsdichte am Beispiel Berlins. ....	123
Abbildung 52: Geschossflächendichte am Beispiel Berlins.....	123
Abbildung 53: Vegetationsanteil am Beispiel Berlins.....	124
Abbildung 54: Versiegelungsanteil am Beispiel Berlins.....	124

## **Kapitel 6**

Abbildung 55: Distanz der Baublöcke zum Stadtzentrum am Beispiel Berlins.....	125
Abbildung 56: Anzahl der teilnehmenden Haushalte über den Untersuchungszeitraum für die einzelnen Stichproben. ....	131
Abbildung 57: Vergleich zwischen gemessener Distanz der Wohnung zum Stadtzentrum und subjektiver Einschätzung durch die Surveyteilnehmer (SOEP und BASE-II).....	138
Abbildung 58: Mediandistanz der Haushalte im Großstadtzentrum für das Untersuchungsgebiet München.....	139
Abbildung 59: Stadtentwicklungsplan von München aus dem Jahr 1963.....	140
Abbildung 60: Vergleich zwischen Grünanteil in der Nachbarschaft basierend auf Klassifikationsergebnissen aus Erdbeobachtungsdaten und subjektiver Einschätzung. ....	142
Abbildung 61: Frage zum monatlichen Haushaltseinkommen. ....	151

## **Anhang A**

Anhang 1: Ikonos Falschfarbeninfrarotdarstellung für München.....	169
Anhang 2: Digitales Oberflächenmodell für München.....	170
Anhang 3: Quickbird Falschfarbeninfrarotdarstellung für Berlin.....	171
Anhang 4: Digitales Oberflächenmodell für Berlin. ....	172

## **Anhang B**

Anhang 5: Boxplots zur Geschosshaltenabschätzung für Berlin und München.....	173
--	-----

## **Anhang C**

Anhang 6: Vergleich zwischen Grünanteil in der Nachbarschaft basierend auf Klassifikationsergebnissen aus Erdbeobachtungsdaten und subjektiver Einschätzung getrennt nach München und Berlin.....	174
---	-----

# Tabellenverzeichnis

## Kapitel 2

Tabelle 1: Übersicht über Verfahren zur Ableitung von Bebauungsklassen.....	23
Tabelle 2: Zentrale Themen des SOEP.....	28
Tabelle 3: Stichproben im SOEP.....	29

## Kapitel 3

Tabelle 4: Eigenschaften von Ikonos und Quickbird.....	51
Tabelle 5: Eigenschaften der Fernerkundungsdatengrundlagen für die Untersuchungsgebiete Berlin und München.....	53
Tabelle 6: Übersicht über die verwendeten Geodaten.....	57
Tabelle 7: Konfusionsmatrix.....	58

## Kapitel 4

Tabelle 8: Berechnete Variablenwerte und bester Schwellwert für einen Beispielbaublock.....	75
Tabelle 9: Flächenbasierte Genauigkeitsabschätzung der Gebäudemaske für München.....	80
Tabelle 10: Flächenbasierte Genauigkeitsabschätzung der Gebäudemaske für Berlin.....	80
Tabelle 11: Genauigkeitsabschätzung für die Klassifikation von Berlin.....	94
Tabelle 12: Genauigkeitsabschätzung für die Klassifikation von München.....	94

## Kapitel 5

Tabelle 13: Korrelationsmatrix der Formmerkmale.....	112
Tabelle 14: Überprüfung der Diskriminanzfunktionen auf Basis der Referenzdaten für Berlin.....	114
Tabelle 15: Überprüfung der Diskriminanzfunktionen auf Basis der Referenzdaten für München.....	114
Tabelle 16: Mittlere standardisierte Diskriminanzkoeffizienten für Berlin und München.....	116
Tabelle 17: Genauigkeitsabschätzung für die Klassifikation auf Basis von Trainingssamples für Berlin.....	118
Tabelle 18: Genauigkeitsabschätzung für die Klassifikation auf Basis von Trainingssamples für München.....	119
Tabelle 19: Genauigkeitsabschätzung für die Übertragung der Trainingssamples von München auf Berlin.....	120
Tabelle 20: Genauigkeitsabschätzung für die Übertragung der Trainingssamples von Berlin auf München.....	120
Tabelle 21: Räumliche Merkmale für die Einbindung in SOEP/BASE-II.....	127

## Kapitel 6

Tabelle 22: Auszug aus den verknüpften Einzeldatensätzen: SOEP/BASE-II und die fernerkundungsbasierten räumlichen Merkmale und die Distanz der Baublöcke, in denen die Surveyteilnehmer wohnen.....	134
Tabelle 23: Vergleich zwischen klassifizierten Bebauungsstrukturen und Wohnumfeld aus den Surveys für Berlin und München. ....	144
Tabelle 24: Anzahl der beobachteten Personen und Haushalte für jedes Beobachtungsjahr nach Auswahl auf ein einzelnes Beobachtungsjahr für jeden Haushalt für Berlin und München.....	152
Tabelle 25: Anpassung des Haushaltsäquivalenzeinkommens mittels VPI.....	152
Tabelle 26: Beschreibung der Surveyvariablen und Nachbarschaftsmerkmale für die empirische Modellschätzung des Nettoäquivalenzeinkommens. ....	154
Tabelle 27: Regressionsschätzungen zum Nettoäquivalenzeinkommen für sechs verschiedene Modelle mit und ohne kleinräumige Nachbarschaftsmerkmale.....	156
Tabelle 28: Regressionsschätzungen zum Nettoäquivalenzeinkommen für drei verschiedene Modelle nur auf Basis der kleinräumigen Nachbarschaftsmerkmale (ohne Dummyvariable für die Stadt). ....	157
Tabelle 29: Regressionsschätzungen zum Nettoäquivalenzeinkommen für drei verschiedene Modelle nur auf Basis der kleinräumigen Nachbarschaftsmerkmale.....	157

## Anhang C

Anhang 7: Deskriptive Statistiken der verwendeten Variablen. ....	175
---	-----

## Anhang D

Anhang 8: Regressionsschätzung zum Nettoäquivalenzeinkommen nur auf Basis der Surveyinformationen ohne Wohnumfeldmerkmale.....	176
Anhang 9: Regressionsschätzung zum Nettoäquivalenzeinkommen nur auf Basis der Surveyinformationen mit Wohnumfeldmerkmalen. ....	177
Anhang 10: Regressionsschätzung zum Nettoäquivalenzeinkommen auf Basis der Surveyinformationen und Nachbarschaftsmerkmale aus Fernerkundungsdaten (nur 4er Nachbarschaft).....	178
Anhang 11: Regressionsschätzung zum Nettoäquivalenzeinkommen auf Basis der Surveyinformationen und Nachbarschaftsmerkmale aus Fernerkundungsdaten (nur Baublock). ....	179
Anhang 12: Regressionsschätzung zum Nettoäquivalenzeinkommen auf Basis der Surveyinformationen und Nachbarschaftsmerkmale aus Fernerkundungsdaten (Baublock und 4er Nachbarschaft). ....	180
Anhang 13: Regressionsschätzung zum Nettoäquivalenzeinkommen auf Basis der Surveyinformationen und Nachbarschaftsmerkmale aus Fernerkundungsdaten (Baublock und 4er Nachbarschaft, ohne Survey-Wohnumfeldmerkmale).....	181
Anhang 14: Regressionsschätzung zum Nettoäquivalenzeinkommen nur auf Basis der Nachbarschaftsmerkmale aus Fernerkundungsdaten (Baublock, ohne Stadt). ....	182
Anhang 15: Regressionsschätzung zum Nettoäquivalenzeinkommen nur auf Basis der Nachbarschaftsmerkmale aus Fernerkundungsdaten (4er Nachbarschaft, ohne Stadt).....	182
Anhang 16: Regressionsschätzung zum Nettoäquivalenzeinkommen nur auf Basis der Nachbarschaftsmerkmale aus Fernerkundungsdaten (Baublock und 4er Nachbarschaft, ohne Stadt). ....	183

Anhang 17: Regressionsschätzung zum Nettoäquivalenzeinkommen nur auf Basis der Nachbarschaftsmerkmale aus Fernerkundungsdaten (Baublock, mit Stadt) .....	184
Anhang 18: Regressionsschätzung zum Nettoäquivalenzeinkommen nur auf Basis der Nachbarschaftsmerkmale aus Fernerkundungsdaten (4er Nachbarschaft, mit Stadt).....	184
Anhang 19: Regressionsschätzung zum Nettoäquivalenzeinkommen nur auf Basis der Nachbarschaftsmerkmale aus Fernerkundungsdaten (Baublock und 4er Nachbarschaft, mit Stadt) .....	185

# Abkürzungsverzeichnis

<b>AdV</b>	Arbeitsgemeinschaft der Vermessungsverwaltungen der Länder der Bundesrepublik Deutschland
<b>ATCOR</b>	Atmospheric and Topographic Correction
<b>ATKIS</b>	Amtliches Topographisches-Kartographisches Informationssystem
<b>AVHRR</b>	Advanced very high resolution radiometer
<b>BauNVO</b>	Baunutzungsverordnung
<b>BASE-II</b>	Berliner Altersstudie II
<b>Basis-DLM</b>	Basis-Digitales Landschaftsmodell
<b>BHPS</b>	British household panel study
<b>BKG</b>	Bundesamt für Kartographie und Geoinformation
<b>BMZ</b>	Baummassenzahl
<b>BRD</b>	Bundesrepublik Deutschland
<b>CCD</b>	Charge-coupled device
<b>CIR</b>	Color infrared (Falschfarbeninfrarot)
<b>CNEF</b>	Cross national equivalent file
<b>CNL</b>	Cognition Network Language
<b>CORINE</b>	Coordinated Information on the Environment
<b>DDR</b>	Deutsche Demokratische Republik
<b>DGM</b>	Digitales Geländemodell
<b>DIW</b>	Deutsches Institut für Wirtschaftsforschung
<b>DLR</b>	Deutsches Zentrum für Luft- und Raumfahrt e.V.
<b>DMSP-OLS</b>	Defense Meteorological Satellite Program Operational Linescan System
<b>DOM</b>	Digitales Oberflächenmodell
<b>DTK</b>	Digitale Topographische Karte
<b>GIS</b>	Geographisches Informationssystem
<b>GFD</b>	Geschoßflächendichte
<b>GFZ</b>	Geschoßflächenzahl
<b>GRZ</b>	Grundflächenzahl
<b>GPS</b>	Globales Positionierungssystem
<b>GLCM</b>	Gray-level Co-occurrence matrix
<b>HID</b>	Haushaltsidentifikationsnummer
<b>HILDA</b>	Household, income and labour dynamics in Australia Survey
<b>HRSC-AX</b>	High resolution stereo camera – airborne extended
<b>INS</b>	Inertiales Navigationssystem
<b>IRS</b>	Indian remote sensing satellite

<b>LiDAR</b>	Light detection and ranging
<b>LoD</b>	Level of detail
<b>MAUP</b>	Modifiable areal unit problem
<b>nDOM</b>	normalisiertes Digitales Oberflächenmodell
<b>NDVI</b>	Normalized Difference Vegetation Index
<b>OBIA</b>	Object-based image analysis
<b>OECD</b>	Organisation for Economic Co-operation and Development (Organisation für wirtschaftliche Entwicklung und Zusammenarbeit)
<b>OLS</b>	Ordinary least squares; Methode der kleinsten Quadrate
<b>PID</b>	Personenidentifikationsnummer
<b>PSID</b>	Panel study of income dynamics
<b>Radar</b>	Radio detection and ranging
<b>RatSWD</b>	Rat für Sozial- und Wirtschaftsdaten
<b>RMSE</b>	Root mean squared error
<b>SAR</b>	Synthetic aperture radar
<b>SHP</b>	Schweizer Haushaltspanel
<b>SLID</b>	Survey of Labor and Income Dynamics
<b>SOEP</b>	Längsschnittstudie Sozio-oekonomisches Panel
<b>SPOT</b>	Système probatoire d'observation de la terre bzw. Satellite pour l'observation de la terre
<b>SVM</b>	Support vector machines
<b>TM</b>	Thematic mapper
<b>VPI</b>	Verbraucherpreisindex





# Kapitel 1

## Problemstellung und Zielsetzung

*“We shape our buildings; thereafter they shape us.”*

Winston Churchill

### 1.1. Einleitung

Der Mensch verändert und prägt den Raum. Zwischen ihm und seiner Umwelt besteht eine wechselseitige Beziehung. Durch die Aktivitäten des Menschen wird der Naturraum verändert und die Ausstattung der Umgebung wirkt sich auf die Verhaltensweisen des Menschen aus. Für die wissenschaftliche Untersuchung dieser Zusammenhänge bedarf es sowohl räumlicher als auch sozialer Datengrundlagen. Vor allem die raumbezogenen Informationen haben in der empirischen Sozialforschung in den letzten Jahrzehnten einen erheblichen Bedeutungszuwachs erfahren da durch sie die Beziehungen zwischen Mensch und Raum untersucht werden können. Dies ist unter anderem auf technische Entwicklungen in den geographischen Informationstechnologien wie den geographischen Informationssystemen (GIS), globalen Positionierungssystemen (GPS) und satellitengestützter Erdbeobachtung zurückzuführen. Goodchild (2007) schreibt in diesem Zusammenhang von einem *spatial turn* für die Sozialwissenschaften. Der Rat für Sozial- und Wirtschaftsdaten (RatSWD) sieht die Georeferenzierung von sozialwissenschaftlichen Daten als eine der großen positiven Herausforderungen der letzten Jahre (RatSWD, 2010).

Die Beziehung zwischen Mensch und Raum ist in den Sozialwissenschaften grundsätzlich seit vielen Jahrzehnten bekannt (Werlen, 2000). Vor allem die Arbeiten der *Chicagoer Schule* am Beginn des vorigen Jahrhunderts (Burgess, 1925, 1929; Hoyt, 1939; Harris und Ullman, 1945) trugen zur Etablierung der räumlichen Sozialforschung bei, auch wenn viele Theorien aus heutiger Sicht bereits überholt sind. Es kann aber festgestellt werden, dass der Bedarf und die Relevanz raumbezogener Stadt- und Sozialforschung in ihrer Vielschichtigkeit bereits früh erkannt wurden. Zudem bringen viele sozialwissenschaftliche Theorien individuelles oder haushaltsbezogenes Verhalten in einen räumlichen Kontext, in welchen die Person oder der Haushalt eingebettet ist.

*Für die Erforschung dieser Fragestellungen ist demnach die Betrachtung des räumlichen Umfeldes von Menschen von erheblicher Bedeutung.*

Eine praktische Umsetzung dieser Erkenntnis in der Forschung erfolgte allerdings für viele Jahrzehnte nur sehr schleppend und einige Autoren vertreten die Ansicht, dass im Laufe des 20. Jahrhunderts verhältnismäßig wenig raumbezogene Arbeiten in den Sozialwissenschaften durchgeführt wurden (z.B. Tickamyer, 2000; Goodchild und Janelle, 2004a; Boyle et al., 2004). Ein gestiegenes Interesse an raumbezogenen Fragestellungen in der sozialwissenschaftlichen Forschung konnte erst gegen Ende des 20. Jahrhunderts festgestellt werden, was in der Literatur durch Arbeiten mit einem expliziten Fokus auf die Möglichkeiten der Verknüpfung von sozialwissenschaftlichen Daten mit räumlichen Daten sichtbar ist (z.B. Goodchild und Janelle, 2004b; Owen, Green und Elias, 2009; Hintze und Lakes, 2009; Lakes, 2010). Vor allem der steigende Bestand an räumlichen Daten – etwa 80 % aller Daten weisen einen räumlichen Bezug auf (Lakes, 2010) – bietet viele Möglichkeiten, um soziale Strukturen in ihrem räumlichen Umfeld zu untersuchen.

Durch diesen steigenden Bestand an räumlichen Daten ist auch die Vielfalt der räumlichen Datenquellen für die Sozialforschung gestiegen. Diese unterscheiden sich sowohl inhaltlich als auch sehr stark hinsichtlich ihrer räumlichen Bezugsebene. Geosachdaten aus der öffentlichen Verwaltung sind an die Anforderungen der Verwaltung angepasst und orientieren sich meist an statistisch-administrativen Raumeinheiten (z.B. Bundesländer, Raumordnungsregionen, Landkreise oder Gemeinden). In der kommunalen Verwaltung existieren kleinräumigere Daten auf darunter liegenden räumlichen Ebenen (z.B. Stadtbezirke, Stadtbezirksteile oder Stadtviertel). Nichtamtliche Geodaten und Geodaten von kommerziellen Anbietern bereiten ihre Daten wiederum in anderen räumlichen Einheiten auf (z.B. Postleitzahlenbezirke, Mikrozellen, Haushalte). Für bestimmte sozialwissenschaftliche Fragestellungen zum räumlichen Kontext von Haushalten sind diese Geodaten allerdings unzureichend, da sie entweder nicht auf einer ausreichend differenzierten räumlichen Bezugsebene vorliegen, nicht die geforderte Aktualität aufweisen oder nicht für das gesamte Bundesgebiet in der gleichen Qualität vorliegen. Je kleinräumiger die Anforderungen an die Geodatengrundlage sind, desto heterogener sind auch Geodaten. Kleinräumige, flächendeckende Daten über die Stadtstruktur liegen in Deutschland nur für einige Städte vor. Viele dieser Daten sind nicht mehr aktuell oder sind mit den anderen Daten hinsichtlich ihrer Taxonomie oder Geometrie nicht identisch. Damit die gleichen räumlichen Informationen für alle Städte erhoben werden können, bedarf es einer einheitlichen Datenbasis.

Eine solche Datenbasis kann mithilfe von Erdbeobachtungsdaten hergestellt werden. Mit Erdbeobachtungsdaten können detaillierte Informationen flächendeckend und unabhängig von administrativen Grenzen erhoben und rasch aktualisiert werden. Sie tragen damit einen wichtigen Baustein zur Beobachtung und Bewertung von urbanen Räumen bei (Dech, 1997). Mit aktuellen Methoden ist eine flächenhafte Auswertung von großmaßstäbigen Daten für viele

urbane Anwendungsgebiete möglich (z.B. Taubenböck und Dech, 2010). Diese detaillierten Daten sind notwendig, um die kleinräumige Struktur in den Städten zu erfassen und ihre einzelnen Elemente wie Gebäude, Straßen, Freiflächen oder Vegetation abzuleiten. Anhand von Fernerkundungsdaten vergleichbarer Sensoren kann eine weitgehend ähnliche Datenbasis geschaffen werden, welche mit gleichen Auswertelgorithmen in unterschiedlichen geographischen Regionen verarbeitet werden und vergleichbare Ergebnisse erzielen kann.

*Räumlich hoch aufgelöste Fernerkundungsdaten können demnach für die Erfassung des räumlichen Umfeldes von Menschen in Städten verwendet werden, um die Erforschung sozialwissenschaftlicher Fragestellungen zu unterstützen.*

Eine Quelle für sozialwissenschaftliche Informationen stellen Personenbefragungen dar. Im Rahmen der sozialwissenschaftlichen Surveys *SOEP* und *BASE-II* werden Personen in Privathaushalten über verschiedene sozialwissenschaftliche Themen befragt. Räumliche Auswertungen dieser Daten wurden bisher aufgrund eines rigiden Datenschutzkonzeptes ausschließlich auf statistisch-administrativen Raumeinheiten durchgeführt. Deshalb stellt die geographische Verortung der Wohnadressen der Umfrageteilnehmer eine bedeutende Innovation in der internationalen sozialwissenschaftlichen Forschung dar<sup>1</sup>.

*Durch die exakte Position der Personen im Raum ist es möglich, aus Fernerkundungsdaten abgeleitete Informationen mit den verorteten Haushaltsadressen zu verknüpfen und als räumliche Kontextinformation für die Bearbeitung sozialwissenschaftlicher Themen zu verwenden.*

In der vorliegenden Arbeit werden Möglichkeiten der Einbindung von hoch aufgelösten Fernerkundungsdaten in urbanen Räumen und sozialwissenschaftlichen Daten exemplarisch demonstriert. Zu diesem Zweck werden Fernerkundungsdaten mit einem der meistgenutzten sozialwissenschaftlichen Surveys, der *Längsschnittstudie das sozio-oekonomische Panel (SOEP)*, verknüpft (Wagner, Frick und Schupp, 2007). Zusätzlich werden auch Umfragedaten der *Berliner Altersstudie II (BASE-II)* herangezogen, da diese weitgehend dieselben Variablen enthält wie das *SOEP* und dadurch die Anzahl an teilnehmenden Personen deutlich erhöht wird (zu *BASE-II* vgl. Bertram et al., 2012).

Diese Haushaltsinterviews beinhalten viele Fragestellungen, die die unterschiedlichsten Themen der Umfrageteilnehmer berühren. Dabei werden auch Merkmale über die Qualität des Wohnumfeldes und der Lage der Wohnung innerhalb des Stadtgebietes abgefragt. Sie erlauben aber nur eine indirekte Beschreibung des räumlichen Umfeldes der Umfrageteilnehmer, da sie nur subjektive Einschätzungen der Personen darstellen (vgl. Kapitel 2.3.2). Diese *räumlichen*

---

<sup>1</sup> Die Surveydaten unterliegen einem strengen Datenschutzkonzept. Im Rahmen des Konzeptes wird zu keinem Zeitpunkt der Datenauswertung die tatsächliche geographische Position eines Haushaltes dem Datennutzer preisgegeben. Eine detaillierte Erläuterung des Datenschutzkonzeptes findet sich in Abschnitt 3.1.1.

Angaben stellen nur subjektive Einschätzungen der Personen dar. Darüber hinaus ist der Raumbezug nicht exakt definiert oder nicht quantifizierbar: er kann sich auf eine punktuelle Information beziehen, auf eine relative Distanz zu einer bestimmten geographischen Position innerhalb des Stadtgebietes oder auf eine nicht näher definierte Nachbarschaft der Haushalte. Die Merkmale, welche im Rahmen der SOEP/BASE-II-Studien über die physische Ausstattung und die Qualität des Wohnumfeldes erfragt werden, beziehen sich auf die *Lage des Wohngebietes innerhalb der Stadt, die Ausstattung mit Grünflächen und die Art und Qualität der Bebauungsstruktur*. Erst durch die geographische Verortung der Haushaltsadressen und durch die Einbindung von räumlichen Kontextinformationen aus Fernerkundungsdaten können exakte Raumbezüge definiert werden. Von erheblicher Bedeutung bei der Einbindung flächenhafter Kontextinformationen ist, dass die Informationen aus der Fernerkundung *flächendeckend* für das gesamte Gebiet vorliegen.

*Nach der Verschmelzung der Surveydaten mit den Fernerkundungsdaten können für jeden Haushalt nicht nur die punktuellen, subjektiven räumlichen Angaben aus den Surveys verwendet werden, sondern auch räumliche Informationen in frei definierbaren Raumeinheiten und Nachbarschaften.*

Erdbeobachtungsdaten können flächendeckende Informationen über ganze Stadtregionen erfassen und vor allem unabhängig von der administrativen Raumeinheit aufgenommen werden. Sie können für verschiedene Gebiete Daten in der gleichen Qualität darstellen und der betrachtete Raumausschnitt kann je nach Fragestellung frei definiert werden: vom Einzelgebäude bis zur größeren Nachbarschaft aus mehreren Straßenabschnitten. Diese Informationen aus Erdbeobachtungsdaten stellen räumliche Kontextinformationen über die physische Ausstattung des räumlichen Wohnumfeldes dar. Allerdings ist die Erfassung von Merkmalen für sozialwissenschaftliche Studien aus Fernerkundungsdaten auch limitiert. Durch die Betrachtung von oben können Aussagen über die physische Ausstattung des Raumes getroffen werden. Informationen über den Zustand oder die Qualität von Wohnungen können damit jedoch nicht erhoben werden.

*Die vorliegende Arbeit soll die Möglichkeiten für die Verwendung von Informationen über das räumliche Umfeld von Personen im Sinne von räumlichen Kontextmerkmalen auf Basis von großmaßstäbigen Fernerkundungsdaten im Rahmen der SOEP/BASE-II-Studien erweitern.*

Dabei wird eine technische und inhaltliche Verknüpfung von Kontextinformationen aus Fernerkundungsdaten mit sozialwissenschaftlichen Daten exemplarisch an zwei Untersuchungsgebieten durchgeführt. Die Anwendung der verwendeten Auswerteverfahren der Fernerkundungsdaten wird stets vor dem Hintergrund einer möglichen Übertragbarkeit der Ansätze auf weitere Daten und Untersuchungsgebiete vorgenommen.

Die Möglichkeiten der Auswertung sozialwissenschaftlicher Daten sind sehr vielfältig und stellen nicht den Fokus der vorliegenden Arbeit dar. Für die empirische Bearbeitung sozialwissenschaftlicher Themen existiert eine Vielzahl an quantitativen Verfahren, deren

Anwendung und Erläuterung den Rahmen dieser Arbeit weit übersteigen würden. Für die Darstellung des interdisziplinären Mehrwertes und der Anwendungsmöglichkeit der verknüpften Daten werden in einer vereinfachten empirischen Analyse die statistischen Zusammenhänge zwischen räumlicher Segregation und dem Haushaltseinkommen untersucht. Dabei werden verschiedene statistische Modelle aufgestellt und durch Hinzu- bzw. Wegnahme von einzelnen erklärenden Variablen die Auswirkungen auf die Gesamterklärung des Modells untersucht. Als erklärende Variablen werden dabei die Merkmale aus den Surveydaten verwendet und die räumlichen Kontextinformationen aus den Fernerkundungsdaten. Es kann auf diese Weise untersucht werden, ob es einen Zusammenhang zwischen der physischen Ausstattung einer Nachbarschaft und dem Einkommen gibt.

## 1.2. Herausforderungen der Arbeit

Durch die Unterschiedlichkeit der im Rahmen der Arbeit behandelten Forschungsdisziplinen und Daten ergeben sich für die Verknüpfung der Daten folgende allgemeine Herausforderungen, die sich auch im Aufbau der Arbeit widerspiegeln:

- **Definition des räumlichen, zeitlichen und taxonomischen Maßstabes für die Untersuchung.**

Der Maßstab ist bei der Bearbeitung räumlicher Fragestellungen von erheblicher Bedeutung. In der sozialwissenschaftlichen Forschung hat die Relevanz des räumlichen Maßstabes für die Erarbeitung von räumlichen Fragestellungen erst mit der Verarbeitung von räumlichen Daten an Bedeutung gewonnen (Gibson, Ostrom und Ahn, 2000). In der Erdbeobachtung bestimmt die geometrische Auflösung des Fernerkundungsbildes den räumlichen Maßstab, wobei das Spektrum von globalen Daten bis zu Daten mit einer Auflösung unter einem Meter reicht.

Für die Fernerkundung ist demnach die Auflösung des Bildes von Bedeutung, in der Sozialwissenschaft ist die Erklärungsebene relevant, auf welcher die Fragestellung formuliert ist. Somit ist es erheblich, ob sich die Variablen, welche zur Erklärung eines Verhaltens verwendet werden, auf derselben Bezugseinheit befinden. In sozialen Daten ist die kleinste Einheit das Individuum. Personen können in Haushalten zusammengefasst werden und Haushalte können in Nachbarschaften zusammengefasst werden und diese können wiederum zu noch größeren Einheiten aggregiert werden. In Fernerkundungsdaten stellt das Bildelement (Pixel) die kleinste Einheit dar. Mehrere Bildelemente können zu Objekten zusammengefasst werden und diese wiederum in Regionen aggregiert werden. Die Auswahl des entsprechenden räumlichen Maßstabes bzw. des Aggregationsniveaus und der Untersuchungseinheit ist also bedeutend für die gemeinschaftliche Untersuchung.

Neben dem räumlichen Maßstab ist der taxonomische Maßstab für die Verknüpfung wichtig. Die Wahl der zu verwendenden Fernerkundungsdatengrundlage richtet sich nach der Bestimmung der abzuleitenden Variablen für die Einbindung in die sozialwissenschaftliche Untersuchung. Die abgeleiteten Klassen aus dem Fernerkundungsbild müssen mit dem Aggregationsniveau der sozialen Daten vereinbar sein.

Erdbeobachtungsdaten stellen den Zustand auf der Erdoberfläche zum Zeitpunkt der Aufnahme dar. Die Aktualität der Daten richtet sich aber auch nach der Dynamik der Veränderungen auf der Erdoberfläche. In statischen Regionen mit geringen Veränderungen können Aufnahmen über einen längeren Zeitraum weniger an Aktualität einbüßen als in hochdynamischen Regionen (z.B. schnellwachsende Megastädte). Umfragedaten können sich dagegen auf einen exakt definierten Zeitpunkt oder Zeitraum beziehen, oder gar keinen konkreten zeitlichen Bezug aufweisen. Die Auswahl der beiden Datengrundlagen richtet sich demnach auch nach der Aktualität.

- **Erhebung von Kontextmerkmalen auf Basis von geometrisch hoch aufgelösten 2D und 3D Fernerkundungsdaten.**

Für die Erhebung von räumlichen Kontextmerkmalen in städtischen Räumen sind großmaßstäbige Fernerkundungsdaten erforderlich, da sie die kleinräumigen Strukturen von Städten abbilden können. Mit der Auflösung der Bilddaten steigt allerdings auch die Komplexität des Bildinhaltes. Eine Herausforderung stellt demnach die Auswertung der detaillierten Fernerkundungsdaten dar. In der computergestützten Bildanalyse können bestimmte Verfahrensschritte so konzipiert werden, dass sie auf mehreren Datensätzen wiederholt angewendet werden können. Diesem Vorteil steht allerdings auch der Nachteil gegenüber, dass häufig geringere Genauigkeiten als bei manuellen Bildauswertungen erzielt werden.

Die erhobenen räumlichen Kontextmerkmale aus Fernerkundungsdaten beziehen sich auf die physische Ausstattung des städtischen Gebietes. Zur physischen Ausstattung wird die Landbedeckung gezählt, welche aus optischen, hoch aufgelösten Satellitenbilddaten abgeleitet wird. Neben der Landbedeckung wird aber auch die Gebäudestruktur der Städte zur physischen Ausstattung gezählt. Die Gebäude unterscheiden sich voneinander durch ihre unterschiedliche Größe, Form, Anordnung oder Dichte. Die Ableitung der Gebäudestrukturen wird auf Basis von digitalen Oberflächenmodellen durchgeführt, welche die dreidimensionale Information über die Gebäudestruktur enthalten. Durch die gemeinschaftliche Verwendung von multisensoralen und multidimensionalen Fernerkundungsdaten bestehen Herausforderungen durch die unterschiedliche Geometrie, den unterschiedlichen Aufnahmezeitpunkt, den unterschiedlichen Bildinhalt und die unterschiedliche Bodenauflösung der Bilddaten, welche bei der Auswertung der Bilddaten berücksichtigt werden müssen.

- **Verknüpfung der Kontextmerkmale mit den Umfragedaten aus SOEP und BASE-II.**

**Berücksichtigung des Datenschutzes:** Die räumliche Auswertung von personenbezogenen Daten stößt vor allem in Deutschland häufig auf Kritik von Datenschützern. Eine Vereinigung von räumlichen und sozialwissenschaftlichen Daten stellt daher in Bezug auf den Schutz persönlicher Daten eine große Herausforderung dar (Blumberg und Jacobson, 1997; Rindfuss und Stern, 1998). Die Herausforderungen bei der Verknüpfung von Fernerkundungsdaten mit Surveydaten der SOEP und BASE-II-Studie beziehen sich einerseits auf den Schutz der persönlichen Daten der Studienteilnehmer und andererseits auf die Festlegung der räumlichen Bezugsebene, auf welcher die Verknüpfung stattfindet. Die Wahrung des Datenschutzes ist für das Fortbestehen einer Studie von erheblicher Bedeutung, da die Studie vom Vertrauen der Teilnehmer lebt. Damit die räumlichen Kontextmerkmale mit den Surveydaten verknüpft werden können, müssen entsprechende Datenschutzvorkehrungen getroffen werden. Dabei ist es wichtig, dass die Daten nicht deanonymisiert werden. Damit dies nicht geschieht, müssen entsprechende Vorkehrungen getroffen werden. Der Datenschutz wird im Rahmen der Arbeit im Sinne eines am DIW Berlin entwickelten Datenschutzverfahrens respektiert. Dieses Verfahren beinhaltet eine gesonderte technische Infrastruktur sowie rechtliche Grundlagen. Weitere Erläuterungen zum Thema Datenschutz finden sich in Abschnitt 3.1.1.

**Schaffung von räumlichen Bezugsebenen:** Wie oben erwähnt, stellt die Definition der räumlichen Bezugsebene die zweite Herausforderung bei der Verknüpfung von Kontextmerkmalen dar. Fernerkundungsbilder bestehen aus einer rasterförmigen Aneinanderreihung von einzelnen Bildelementen (Pixeln). In hoch aufgelösten Bilddaten ist die Anzahl dieser Bildpunkte pro Raumeinheit sehr groß, weshalb ein Objekt in der Natur (z.B. ein Haus) durch mehrere Bildpunkte dargestellt wird. Die Adressen der SOEP/BASE-II-Teilnehmer wurden in einem geographischen Koordinatensystem verortet und können durch einen eindeutigen Punkt in diesem Koordinatensystem lokalisiert werden. Bei einer räumlichen Überlagerung der Bilddaten und der verorteten Haushaltsadressen entspricht ein Adresspunkt einem Bildpunkt des Rasterdatensatzes mit der gleichen Position im Koordinatensystem. Ein einzelner Bildpunkt kann aber nicht die gesamte Information einer Nachbarschaft beinhalten. Deshalb ist es notwendig, den Raumausschnitt für die Nachbarschaft eines verorteten SOEP/BASE-II-Haushaltes zu definieren und abzugrenzen. Für die Verknüpfung der Informationen über die physische Ausstattung aus Fernerkundungsdaten mit den Surveydaten wurde eine räumliche Bezugsebene definiert, welche zwischen dem einzelnen Bildpunkt und einer administrativen Raumeinheit liegt. Sie stellt die aggregierte Information aus den einzelnen Strukturelementen auf der räumlichen Ebene des Baublocks dar. Auf die Beschreibung dieses räumlichen Konzeptes wird in Abschnitt 3.1.2 eingegangen.

### 1.3. Zielsetzung

Entsprechend der in Abschnitt 1.2 definierten Herausforderungen werden im Rahmen der Arbeit räumliche Kontextmerkmale flächenhaft aus Fernerkundungsdaten abgeleitet und mit punktueller, teilweise subjektiver Information aus der SOEP-Studie verknüpft und untersucht. Diese Arbeit stellt eine interdisziplinäre Arbeit dar, da sie sich im Schnittpunkt der Forschungsdisziplinen *urbane Fernerkundung* und *empirische Sozialwissenschaften* bewegt. Die Arbeit beschreibt Wege, um Informationen aus Erdbeobachtungsdaten zu gewinnen welche eine neue Informationsquelle für sozialwissenschaftliche Disziplinen darstellt. Aufgrund der Interdisziplinarität der Themenstellung ist es unumgänglich, dass bestimmte Bereiche der Arbeit einem Leserkreis geläufiger sind, als einem anderen Leserkreis. Das Ziel der Arbeit kann wie folgt zusammengefasst werden:

*Die empirische Untersuchung der technischen Einbindung von flächenhaften, kleinräumigen Kontextmerkmalen aus Erdbeobachtungsdaten in die SOEP/BASE-II-Studien am Beispiel von zwei Testgebieten.*

Durch die Aggregation der räumlichen Kontextmerkmale auf kleinräumige Nachbarschaften ist es erstmals möglich, trotz strikter Datenschutzerfordernungen, SOEP-Daten mit zusätzlichen, raumbezogenen Kontextmerkmalen aus Fernerkundungsdaten zu verknüpfen und zu untersuchen, ohne dabei die persönlichen Rechte von Einzelpersonen zu verletzen.

Aus den im vorigen Abschnitt genannten Herausforderungen lassen sich die folgenden, konkreten Teilziele ableiten:

- Erstellung eines allgemeinen Konzeptes zur Verknüpfung von flächenhaften Erdbeobachtungsdaten mit punktuellen, sozialwissenschaftlichen Daten
- Entwicklung und Anwendung eines übertragbaren Klassifikationsablaufs zur Ableitung der städtischen Landbedeckung auf Basis von geometrisch hoch aufgelösten Fernerkundungsdaten
- Ableitung der dreidimensionalen Gebäudestruktur auf Basis von digitalen Oberflächenmodellen
- Untersuchung von geometrischen Merkmalen auf ihre Eignung zur Klassifikation unterschiedlicher Gebäudetypen
- Aggregation der Klassifikationsergebnisse zu räumlichen Kontextmerkmalen
- Validierung der aus der Fernerkundung abgeleiteten Information
- Demonstration von Potenzialen und Limitierungen der Einbindung von Fernerkundungsdaten in sozialwissenschaftliche Daten auf Basis von ausgewählten Beispielen.

## 1.4. Gliederung der Arbeit

Die Arbeit ist in mehrere Kapitel gegliedert. Im ersten Kapitel wurde die Motivation der Arbeit beschrieben und warum die Einbindung von Fernerkundungsdaten in sozialwissenschaftliche Daten eine aktuelle Herausforderung an die Forschung darstellt. Des Weiteren werden die daraus abgeleiteten Aufgabenstellungen und die Ziele definiert.

Im zweiten Kapitel werden die Grundlagen dargelegt, welche zum Verständnis der vorliegenden Arbeit notwendig sind. Da die Arbeit in einem interdisziplinären Thema eingebettet ist, wird in diesem Kapitel sowohl auf Grundlagen der Fernerkundung als auch auf grundlegende Informationen über die verwendeten sozialwissenschaftlichen Surveys eingegangen. In den einzelnen Unterabschnitten wird auch auf den aktuellen Stand der Forschung der verwendeten Fernerkundungsverfahren und der Verknüpfung mit sozialwissenschaftlichen Daten verwiesen.

Im dritten Kapitel wird der konzeptionelle Rahmen erläutert, nach welchem die Verknüpfung der aus den Fernerkundungsdaten abgeleiteten Informationen mit den SOEP/BASE-II-Studien durchgeführt wird. Darüber hinaus werden die verwendeten Datengrundlagen und ihre Eigenschaften beschrieben sowie die Untersuchungsgebiete charakterisiert und deren Auswahl begründet. Im letzten Teil des Kapitels wird auf das verwendete Konzept zur empirischen Validierung der einzelnen Klassifikationsergebnisse verwiesen.

Das vierte Kapitel beschreibt die einzelnen Schritte zur Ableitung der urbanen Landbedeckung und der 3D-Bebauungsstruktur. Dabei wird im ersten Schritt auf die Extraktion von Gebäudegrundrissen und der Gebäudehöhe aus den digitalen Oberflächenmodellen eingegangen und in einem weiteren Schritt ein empirisches Verfahren zur Abschätzung der Geschoszahl für jedes Gebäude beschrieben. Des Weiteren wird basierend auf der abgeleiteten Bebauungsstruktur und den Satellitenbilddaten die Klassifikation der urbanen Landbedeckung dargestellt. Die Ergebnisse aus diesen Klassifikationsschritten stellen die räumlichen Kontextmerkmale dar. Sie repräsentieren die Merkmale, welche mit den sozialwissenschaftlichen Surveydaten verknüpft werden.

Das fünfte Kapitel widmet sich der Klassifikation dieser räumlichen Kontextmerkmale wie die Bebauungsstruktur und die physische Ausstattung der Nachbarschaft. Dabei werden zunächst auf Basis der 3D-Bebauungsstruktur verschiedene geometrische Merkmale für die Klassifikation von unterschiedlichen Gebäudeklassen untersucht und die Gebäude klassifiziert. Im letzten Teil des Kapitels werden die Informationen aus den einzelnen Klassifikationsergebnissen auf der Ebene des Baublocks zusammengefasst und die räumlichen Kontextmerkmale für die Verknüpfung mit den SOEP/BASE-II-Daten beschrieben.

Im sechsten Kapitel wird die Verknüpfung der abgeleiteten räumlichen Kontextmerkmale aus den Fernerkundungsdaten mit den Surveydaten beschrieben und sozialwissenschaftliche Analysepotenziale beleuchtet. In einem weiteren Abschnitt des Kapitels wird der räumliche Bezug der Surveydaten und der abgeleiteten räumlichen Kontextmerkmale aus den Fernerkundungsdaten gegenübergestellt. Den letzten Teil des Kapitels stellt eine empirische Analyse der erstellten Daten dar, um Möglichkeiten und Limitierungen der Verknüpfung aufzuzeigen.

Im abschließenden siebenten Kapitel werden die Ergebnisse der Untersuchungen zusammengefasst und ein Ausblick auf mögliche weiterführende Entwicklungsschritte gegeben.

# Kapitel 2

## Grundlagen

*„Integrating social science and remote sensing will require the fusion not only of data, but also of quite different scientific traditions.“*

Rindfuss und Stern (1998)

In diesem Kapitel werden die dem Verständnis der vorliegenden Arbeit zugrundeliegenden Rahmenbedingungen erläutert und auf den aktuellen Stand der Forschung Bezug genommen. Dabei werden zunächst kurz die Eigenschaften von Fernerkundungsdaten zusammengefasst. Der zweite Abschnitt verweist auf relevante Arbeiten, die den aktuellen Stand der Forschung in der urbanen Fernerkundung darstellen. Des Weiteren wird auf Forschungsarbeiten eingegangen, die den Forschungsstand in der räumlichen Auswertung von sozialwissenschaftlichen Daten und in der Kombination mit Fernerkundungsdaten darstellen. Im letzten Abschnitt werden die sozialwissenschaftlichen Surveydaten SOEP/BASE-II ausführlich vorgestellt, mit welchen die Fernerkundungsdaten verknüpft werden.

### 2.1. Eigenschaften von Fernerkundungsdaten

Auf eine ausführliche Beschreibung der Grundlagen der Fernerkundung wird in der Arbeit verzichtet, da sie im Fachgebiet der Fernerkundung vorausgesetzt werden können. Eine detaillierte Beschreibung von gängigen Aufnahme- und Auswerteverfahren findet sich in Fernerkundungslehrbüchern, auf die an dieser Stelle verwiesen wird. Sie stellen auch die Quellen für die folgenden Ausführungen dar (Hildebrandt, 1996; Mather, 1999; Albers, 2001; Lillesand, Kiefer und Chipman, 2004; Kraus, 2004; De Jong und van de Meer, 2004; Richards und Jia, 2006). Die Themenstellung der Arbeit bewegt sich im Schnittpunkt zwischen zwei unterschiedlichen Forschungsrichtungen, weshalb im aktuellen Abschnitt für den in der Fernerkundung nicht erfahrenen Leser die für die Arbeit relevanten Grundlagen der Fernerkundung zusammengefasst werden. Im darauf folgenden Abschnitt 2.2 wird auf den aktuellen Stand der Forschung in der für die vorliegende Arbeit relevanten Fernerkundungsarbeiten eingegangen.

Fernerkundung bedeutet die Messung von Informationen über den physikalischen Zustand der Erdoberfläche, ohne dass der Ort des Objektes, für welches Informationen gewonnen werden, weder von einer Person noch einem Messgerät aufgesucht werden muss. Die Messung der Informationen beruht auf der Strahlung und stellt somit eine berührungslose Beobachtung dar. Fernerkundungsverfahren werden seit Mitte des 19. Jahrhunderts für die Erdbeobachtung verwendet: während zunächst fotografische Aufnahmen aus Heißluftballonen gemacht wurden (Tournachon, 1864), wurden vor allem durch technische Weiterentwicklungen auf dem Gebiet der Luftfahrt und der Fotografie flugzeuggetragene Aufnahmesysteme zu Aufklärungszwecken in den beiden Weltkriegen verwendet bis sich die Fernerkundung als eigene Forschungsdisziplin etablieren konnte. Seit den 1970er Jahren wurden flugzeuggetragene Aufnahmesysteme zunehmend durch verbesserte Satellitensysteme ergänzt.

Neben den unterschiedlichen Aufnahmeplattformen (z.B. flugzeuggetragen oder satellitengestützt), können Fernerkundungssensoren nach ihren Aufnahmesystemen unterschieden werden. Die klassische Aufnahme stellt das Luftbild dar. Dabei fotografiert eine in einem Flugzeug montierte Kamera den überflogenen Ausschnitt auf der Erdoberfläche. Diese Art der Aufnahme wird *passives Aufnahmesystem* genannt, da für die Fotografie die natürliche Strahlung der Sonne benötigt wird. Ähnliche Aufnahmesysteme können auch auf Satelliten angebracht sein, welche die Erde umkreisen. Sie können durch die erhöhte Aufnahmeposition bis zu mehrere hundert Kilometer große Aufnahmen machen. Demgegenüber stehen mit *aktiven Aufnahmesystemen* jene, welche die Strahlung aktiv aussenden und empfangen können. In SAR-Systemen (Synthetic Aperture Radar) wird Mikrowellenstrahlung zwischen 1 und 100 cm Wellenlänge ausgesendet, weshalb unabhängig von natürlichen Strahlungsverhältnissen auch bei Nacht oder Bewölkung aufgezeichnet werden kann (Albertz, 2001). Seit Ende der 1970er Jahre wurden als weitere aktive Verfahren vermehrt *LiDAR*-gestützte Aufnahmeverfahren (Light detection and ranging) für die Erfassung von Gelände- und Objekthöhen verwendet. Die diesem Verfahren zugrundeliegende Technologie verwendet Laserimpulse, welche auf die Erdoberfläche ausgesendet werden und über die Laufzeit des Impulses die Distanz messen. Laserscanbefliegungen werden neben der Erstellung digitaler Bodenmodelle für die Herstellung von geometrisch hochaufgelösten *digitalen Oberflächenmodellen* (DOM) verwendet, welche in unterschiedlichen Anwendungsgebieten zum Einsatz kommen. Vor allem in forstlichen Fragestellungen (z.B. Nilsson, 1996; Goodwin, Coops und Culvenor, 2006) und urbanen Anwendungen (z.B. Murakami et al., 1999; Yoon und Shan, 2002; Zhou et al., 2004; Filin, 2004) stellt die Auswertung der Höheninformation einen bedeutenden Mehrwert dar.

Die Aufnahmen der Fernerkundungssensoren werden in der Regel als digitales Fernerkundungsbild gespeichert. Digitale Rasterbilddaten sind durch die rasterförmige Anordnung von Bildpunkten in Zeilen und Spalten dargestellt. Die Bildinformation ist somit als Zahlenwert der einzelnen Bildelemente abgelegt und kann mit Hilfe von Computerprogrammen verarbeitet werden. Die Kantenlänge der Bildelemente (Pixel: *picture*

*element*) wird in der Erdbeobachtung als Angabe für die *geometrische Auflösung* verwendet. Sie ist neben den Objekteigenschaften und dem Kontrast maßgeblich für die Erkennbarkeit von Objekten verantwortlich. In untenstehender Abbildung 1 sind Beispiele von Aufnahmen gängiger Fernerkundungssysteme dargestellt. Sie zeigt links eine Satellitenbilddaufnahme mit einer geometrischen Auflösung von einem Meter, in welcher selbst kleinräumige Strukturen innerhalb des Stadtgebietes erfasst werden können. In der Mitte ist für den gleichen Ausschnitt ein digitales Oberflächenmodell dargestellt. Die Helligkeit der Pixel repräsentiert die Höhe des Punktes in der Natur. Rechts ist eine Aufnahme eines aktiven SAR-Systems abgebildet. Durch die besondere, schräge Aufnahmegeometrie scheinen die erhöhten Objekte in der Abbildung zu „kippen“.



**Abbildung 1:** Bilddaten unterschiedlicher fernerkundlicher Aufnahmesysteme.

(Quelle: Taubenböck und Dech, 2010)

## 2.2. Stand der Forschung

In diesem Teil werden relevante Vorarbeiten und aktuelle Arbeiten in der urbanen Fernerkundung beschrieben, die den Stand der Forschung darstellen. Dieser Abschnitt gliedert sich in Unterabschnitte zu den Themenbereichen der Landbedeckungsklassifikation von hoch aufgelösten Bilddaten, der Ableitung von Gebäuden, der Ableitung von semantischen Bebauungsklassen aus Fernerkundungsdaten und der Einbindung von Erdbeobachtungsdaten in sozialwissenschaftliche Analysen.

### 2.2.1. Landbedeckungsklassifikation von optischen Fernerkundungsdaten in städtischen Räumen

Die Anzahl von Fernerkundungsanwendungen im städtischen Raum ist in den vergangenen beiden Jahrzehnten stark gestiegen. Beigetragen zu diesem Anstieg hat die breite Verfügbarkeit von geometrisch sehr hoch aufgelösten Satellitenbilddaten, welche mit der Entwicklung von

hoch auflösenden Sensoren einhergeht (Maktav, Erbek und Jürgens, 2005). Neueste Satellitensysteme wie Ikonos, SPOT-5, Quickbird, World View I und II sind in der Lage, die kleinräumigen urbanen Strukturen mit einem sehr hohen räumlichen Detaillierungsgrad von bis zu einem halben Meter abzubilden. Die Anwendung von Fernerkundung in städtischen Räumen ist nicht zuletzt durch die weltweit steigende Urbanisierung ein hochaktuelles Thema und einige Fachbücher setzen sich explizit mit dieser Thematik auseinander (Donnay, Barnsley und Longley, 2001; Mesev, 2003; Weng, 2006; Netzband, 2007; Gamba und Herold, 2009; Rashed und Jürgens, 2010; Bhatta, 2010; Soergel, 2010; Taubenböck und Dech, 2010; Yang, 2011).

Die computergestützte Auswertung von fernerkundlichen Bilddaten kann anhand von vielen unterschiedlichen Methoden erfolgen. Im Zusammenhang mit der gestiegenen geometrischen Auflösung von Satellitenbildern für die Auswertung in städtischen Räumen kann jedoch ein allgemeiner Trend von der pixelbasierten zur objektbasierten Klassifikation festgestellt werden (Blaschke, 2010). In den geometrisch hoch aufgelösten Bilddaten wird ein Objekt in der Natur durch mehrere Pixel repräsentiert, wodurch der Bildinhalt und auch die Auswertung komplexer geworden sind. Ein einzelnes Pixel stellt nicht mehr die erforderliche semantische Information dar, da es nur mehr einen Teil eines Objektes abbildet (Baatz und Schäpe, 2000). Neben der hohen Auflösung sind auch die spektrale Ähnlichkeit von urbanen Landbedeckungsarten in multispektralen Aufnahmen und die hohe spektrale Heterogenität innerhalb eines Objekts dafür verantwortlich, dass vor allem in städtischen Gebieten die Genauigkeiten klassischer, pixelbasierter Methoden für diese Daten nicht mehr ausreichend sind (Schiewe, 2002). Ein Beispiel für einen segmentierten Bildausschnitt mit den gebildeten Objekten ist in Abbildung 2 dargestellt. Die einzelnen Pixelgruppen im Luftbild weisen weitgehend ähnliche spektrale Eigenschaften auf, aber die dargestellten Strukturen im Bild können durch eine isolierte Betrachtung der einzelnen Pixel nicht ausreichend beschrieben werden. Durch die Gruppierung der Einzelpixel zu mehr oder weniger homogenen Objekten, können semantische Klassen gebildet werden, die über ihre spektrale Zusammensetzung, ihre Form und ihre Nachbarschaftsbeziehungen beschrieben werden können.



**Abbildung 2:** Rasterbild (links) und erzeugte Objekte durch Multiresolutionsegmentierung (rechts).

(Quelle: Esch, 2006)

In der aktuellen Literatur finden sich dennoch verschiedene Ansätze, wie pixelbasierte Klassifikationen mit hoch aufgelösten Daten in städtischen Räumen verbessert werden können. Van de Voorde, De Genst und Canters (2007) beschreiben verschiedene Postklassifikationsmethoden um das Ergebnis zu verbessern. Sie wenden einen Strukturfilter auf das Klassifikationsergebnis an, um die hohe Variabilität der Landbedeckungsklassen für benachbarte Pixel zu reduzieren. Der Filter sucht zunächst Gruppen benachbarter Pixel mit der gleichen Landbedeckungsklasse und weist jene Gruppen, welche eine vordefinierte Fläche unterschreiten, der benachbarten Gruppe mit der größten Fläche zu. Cablk und Minor (2003) verwenden neben Hauptkomponentenanalysen auch räumliche morphologische Operatoren, um unterschiedliche Siedlungsflächen in Städten auf Basis einer Ikonos-Aufnahme zu klassifizieren. Ausschließlich mit mathematischen Operatoren auf Basis von panchromatischen Quickbird-Daten klassifizieren Tuia et al. (2009) städtische Landnutzungstypen mit Support-vector-machines (SVM). Sie verwenden ein automatisches Verfahren für die Auswahl der besten Merkmale um unterschiedliche Bebauungstypen zu klassifizieren. In anderen Arbeiten werden Texturmaße verwendet, um zu den spektralen Merkmalen der Pixel zusätzliche Merkmale zu berechnen (z.B. Puissant, Hirsch und Weber, 2005). Die Autoren verwenden ausschließlich den panchromatischen Kanal für die Berechnung von Texturmaßen anhand der Grey level co-occurrence matrix (GLCM). Sie beschreibt die Grauwertbeziehungen eines Pixels mit seiner Nachbarschaft. Durch die zusätzliche Verwendung dieser Merkmale konnten sie die Genauigkeit der Klassifikation von Landbedeckungs- und -nutzungsklassen in städtischen Räumen erhöhen. Ebenfalls basierend auf texturaler Information in panchromatischen Bilddaten klassifizieren Pacifici, Chini und Emery (2009) WorldView-1 und Quickbird-Daten. Die Methode beschreibt die systematische Analyse von über 191 verschiedenen Texturmerkmalen, aus welchen die wichtigsten für die Klassifikation ausgewählt werden. Sie klassifizieren unterschiedliche versiegelte Flächen wie Straßen oder Parkplätze oder verschiedene Bebauungstypen, -größen und -höhen.

Alle oben beschriebenen Klassifikationsverfahren haben gemeinsam, dass sie Strategien gesucht haben, um das Einzelpixel nicht mehr isoliert, sondern im Kontext seiner Nachbarschaft zu betrachten. Genau diese Betrachtung einer Gruppe an ähnlichen Bildelementen hat die objektbasierte Bildanalyse (OBIA: object-based image analysis) zum Ziel. Das Konzept von OBIA (Blaschke, Lang und Hay, 2008) ist darauf ausgerichtet, nicht mehr das einzelne Pixel, sondern eine Gruppierung von Pixeln als die relevante Größe zu betrachten. Diese Gruppierung wird als Objekt bezeichnet. Durch die Betrachtung des Objektes gewinnen zusätzliche Merkmale wie die Größe, die Form oder die räumliche Beziehung zu anderen Objekten für die Klassifikation eine wichtige Bedeutung. Diese form- und kontextbasierten Eigenschaften der Objekte können bei der Klassifikation bestimmter Bildinhalte einen hohen Automatisierungsgrad erzielen (Thomas, Hendrix und Congalton, 2003). Das Konzept von OBIA erlaubt also neben der Verwendung von spektralen, texturalen, räumlichen, hierarchischen und topologischen Eigenschaften (Lang, 2008) auch die Vereinigung von

Informationen unterschiedlicher Quellen. Die Fusion an Bildinformation aus unterschiedlichen Quellen stellt zwar kein Alleinstellungsmerkmal in der objektbasierten Bildanalyse dar, die Verknüpfung der unterschiedlichen Informationsebenen in den Objekten kann mittlerweile aber als etabliertes Auswertekonzept gesehen werden. Bedeutend ist, dass durch die Erstellung von Objekten die Weiterverarbeitung auf einer semantischen Ebene durchgeführt werden kann.

Die Grundlagen für die objektbasierte Bildanalyse wurden bereits vor Jahrzehnten von Kettig und Landgrebe (1976) erarbeitet und Grundlagen auf dem Gebiet der Bildsegmentierung zur Herstellung von Bildobjekten von Haralick und Shapiro (1985) geschaffen. Allerdings zeigte ihre Anwendung auf den zu dieser Zeit zur Verfügung stehenden, geringer aufgelösten Satellitenbilddaten keine nennenswerten Verbesserungen und war der Komplexität des Bildinhaltes von Luftbildern nicht gewachsen (Müller, 2001) bzw. durch die damals zur Verfügung stehende Rechenleistung limitiert (Neubert, 2005). Zum Zeitpunkt der Jahrtausendwende wurden klassische, pixelbasierte Klassifizierungsverfahren vermehrt in Frage gestellt (Blaschke und Strobl, 2001) und das Interesse an der objektbasierten Bildanalyse hat seither stark zugenommen (Blaschke, 2010). Seit dieser Zeit sind auch erstmals kommerzielle Satellitenbilddaten mit einer geometrischen Auflösung von einem Meter verfügbar (GeoEye, 2009).

Bildsegmentierungsalgorithmen zielen auf die Gruppierung und Verschmelzung benachbarter Objekte basierend auf ihrer spektralen oder räumlichen Homogenität ab oder sie zerschneiden bestehende Objekte aufgrund ihrer Heterogenität (Schiewe, 2002). Für die Erzeugung dieser nicht überlappenden Bildobjekte wurde eine ganze Reihe an unterschiedlichen Segmentierungsalgorithmen und Anwendungen entwickelt. Neben verschiedenen kantenbasierten und regionenbasierten Segmentierungsverfahren stehen auch viele hybride Verfahren zur Verfügung (vgl. Neubert, 2005: 25ff.). Durch die Einbindung in das kommerzielle Softwarepaket eCognition© (Trimble, 2010a) findet das Verfahren der *Multiresolutionsegmentierung* (Baatz und Schäpe, 2000) eine breite Anwendung. Es wird im Rahmen dieser Arbeit zur Klassifikation der städtischen Landbedeckung verwendet und in Kapitel 4 beschrieben.

### **2.2.2. Ableitung von Gebäuden aus digitalen Oberflächenmodellen**

Neben der Klassifikation der städtischen Landbedeckung ist auch die Ableitung der dreidimensionalen Bebauungsstruktur auf Basis von digitalen Oberflächenmodellen (DOM) ein Ziel der Arbeit. Digitale Oberflächenmodelle werden vor allem für Anwendungen in städtischen Räumen und für die Ableitung von Vegetationsparametern verwendet. Besonders seit Ende der 1990er Jahre wurden DOMs aus aktiven, flugzeuggetragenen Aufnahmesystemen hergestellt. Neben diesen Laserscannern werden heute auch photogrammetrische Verfahren aus Daten von digitalen Luftbildkameras für die Herstellung von digitalen

Oberflächenmodellen verwendet. Durch die Höheninformation in DOMs können erhabene Objekte auf der Erdoberfläche, wie zum Beispiel Gebäude oder hohe Vegetation, aufgrund ihres relativen Höhenunterschiedes zu benachbarten Objekten unterschieden werden. In der Literatur finden sich verschiedene Ansätze zur Ableitung von Gebäuden aus digitalen Oberflächenmodellen. Brenner (2000) beschreibt verschiedene Segmentierungsverfahren wie die morphologische Segmentierung, die Steigungssegmentierung, die Ebenensegmentierung, die krümmungsbasierte Segmentierung, die Richtungssegmentierung und die Segmentierung basierend auf Höhenlinien für die Abgrenzung von Einzelgebäuden in DOMs. Ein weiterer Ansatz für die Einzelhaussegmentierung wurde von Miliaris und Kokkas (2007) beschrieben. Die Autoren entwickelten einen geomorphometrischen Segmentierungsansatz für die Anwendung auf einem normalisierten digitalen Oberflächenmodell (nDOM). Ein Nachteil der Methode ist, dass für die Anwendung ein großes Maß an Nutzerinteraktion notwendig ist, um relevante Parameter zu bestimmen.

Neben der Ableitung von Gebäudegrundflächen beschäftigen sich einige Arbeiten mit der Gebäuderekonstruktion aus digitalen Oberflächenmodellen. Sirmacek et al. (2012) vergleichen verschiedene Arten von digitalen Oberflächenmodellen für die realistische Beschreibung von komplexen Gebäudeformen, -höhen und Dachformen. Weitere Alternativen zu den genannten Daten und Methoden finden sich in der Literatur in der Verwendung von SAR-Interferometrie zur Erkennung von Gebäudestrukturen (z.B. Gamba, Houshmand und Saccani, 2000).

### **2.2.3. Ableitung von Gebäuden aus der Kombination von digitalen Oberflächenmodellen und Spektraldaten**

Es existieren in der Literatur viele Arbeiten, die zusätzliche Informationsquellen wie Spektraldaten für die Ableitung der Gebäudestruktur verwenden. Diese Art der Bild- und Informationsfusionierung von spektraler Information und Höheninformation wird verwendet, um Gebäudestrukturen von anderen Objekten im Bild zu trennen. Eine Übersicht über den aktuellen Status und Trends über Bild- und Informationsfusionsverfahren bieten sowohl Gamba, Dell'Acqua und Dasarathy (2005) als auch Zhang (2010). Exemplarisch zeigen die im Folgenden angeführten Arbeiten verschiedene Verfahren. Haala und Brenner (1999) verwenden ein ISODATA-Clusteringverfahren, um verschiedene Landbedeckungsklassen mit einem Falschfarbinfrarotbild und einem nDOM zu klassifizieren. Chen et al. (2009) verwenden ein hierarchisches, objektbasiertes Verfahren, um neun städtische Landbedeckungsarten auf Basis eines nDOMs und einer Quickbird-Aufnahme zu klassifizieren. Sowohl Oczipka (2007), als auch Trosset, Bucher und Lehmann (2009) und Brüßhaber, Trosset und Bucher (2010) verwenden objektbasierte Methoden auf digitalen Oberflächenmodellen und spektralen Daten aus flugzeuggetragenen HRSC bzw. UltraCam Befliegungen für die Klassifikation von städtischen Landbedeckungsklassen und Vegetations- bzw. Gebäudehöhen.

### 2.2.4. Ableitung von semantischen Bebauungsklassen

Zusätzlich zur reinen Ableitung von Einzelgebäuden für die Bestimmung von physischen Kenngrößen wie der Höhe oder der Fläche der Gebäude, sind für bestimmte Anwendungen semantische Informationen über die Gebäudetypologie relevant. Geiß et al. (2011) verwenden beispielsweise unterschiedliche Gebäudetypologien, um ihnen gebäudespezifische Kennwerte zuzuweisen und darauf eine Abschätzung des Wärmebedarfs von Städten abzuschätzen. Im Kontext der vorliegenden Arbeit wird eine flächendeckende Typologisierung der Gebäude benötigt, da sie gemeinsam mit quantifizierbaren Gebäudemerkmalen, wie die Größe, weitere Aussagen über die Beschaffenheit der lokalen Bebauungsstruktur zulassen. Beispiele für unterschiedliche Bebauungsklassen sind in Abbildung 3 für München dargestellt. Von links oben bis rechts unten lassen sich fünf unterschiedliche semantische Bebauungsklassen in unterschiedlichen Ausprägungen unterscheiden: a) lockere Bauformen wie Einzel- und Doppelhausbebauung (auch Schrebergartensiedlungen), b) Zeilenbebauung, c) mehrgeschossige Wohnbauformen (hauptsächlich Blockbebauung und Blockrandbebauung), d) Hochhausbebauung und e) Bebauung, welche aufgrund ihrer Physiognomie offensichtlich keiner Wohnnutzung zugeführt wird (industriell bzw. gewerblich genutzte Bebauung).



**Abbildung 3:** Unterschiedliche Bebauungsstrukturen am Beispiel Münchens.

(© Bundesamt für Kartographie und Geodäsie)

Diese fünf Bebauungsklassen umfassen thematisch und flächenmäßig einen Großteil der Fläche. Am Beispiel für München werden durch diese Klassen etwa 80 % der bebauten Fläche abgedeckt, die restlichen Klassen sind Mischklassen, Sonderformen und Gebäude, die sich auf Basis ihrer Nutzung von anderen Gebäudeklassen unterscheiden. Bei einer flächendeckenden Klassifikation dieser Gebäudeklassen wird also ein Großteil der Gebäude klassifiziert. Die semantische Klassifikation dieser fünf Bebauungsklassen wird in Abschnitt 5.1 beschrieben.

Bei der Betrachtung der Arbeiten, welche sich mit der Klassifikation von Bebauungsklassen auseinandersetzen, ist die räumliche Abgrenzung, auf welcher die Klassifikation durchgeführt wird, entscheidend. Einige Arbeiten verfolgen die semantische Klassifikation von Bebauungsklassen auf der Gebäudeebene und andere Arbeiten auf einer übergeordneten Ebene, der Strukturebene. Bezeichnend bei der Betrachtung ist die Vielfalt der Taxonomien der semantischen Bebauungsklassen, da sie sehr stark zwischen unterschiedlichen physischen Bebauungsformen und Nutzungsarten variieren. Im folgenden Abschnitt wird auf Arbeiten auf der Gebäudeebene eingegangen, wobei in diesem Zusammenhang auch auf Arbeiten verwiesen wird, welche andere Daten als Fernerkundungsdaten verwenden. Im darauffolgenden Abschnitt wird auf die Arbeiten der Strukturebene eingegangen und im letzten Abschnitt werden die unterschiedlichen Anwendungen und Verfahren zusammenfassend betrachtet.

#### **2.2.4.1. Gebäudeebene**

Auf der Ebene der Einzelgebäude klassifizieren Henn et al. (2012) sieben verschiedene Gebäudeklassen anhand von *support vector machines (SVM)*. Die Autoren verwenden ein digitales Gebäudemodell mit einem *Level-of-Detail 1 (LoD1)*, welches Einzelgebäude durch einen generalisierten Gebäudegrundriss mit einem mittleren Höhenwert, vertikalen Wänden und flachen Dächern repräsentiert. Für das überwachte Klassifikationsverfahren wurden über 1200 Gebäude für das Training der SVM verwendet. Für die Klassifikation verwenden die Autoren sowohl Formmerkmale wie Länge, Breite und Höhe, aber auch Informationen über die Gebäudenutzung und Informationen über die Gebäude in der direkten Nachbarschaft wie die Anzahl der Gebäude pro Baublock oder die Anzahl der direkten Nachbarn. Aus zusätzlichen Datenquellen werden auch Distanzen zu diversen öffentlichen und kulturellen Einrichtungen für die Klassifikation verwendet. Die Autoren erzielen mit dieser Methode Genauigkeiten von über 90 %.

Aus der Digitalen Topographischen Karte 1:25000 (DTK-25), Baublockgrenzen und Nutzungsinformationen aus ATKIS (Amtliches Topographisches-Kartographisches Informationssystem) klassifizieren Meinel et al. (2008) sowohl Bebauungsklassen auf der Gebäudeebene als auch auf der Strukturebene. In diesem Verfahren werden Einzelgebäude aus der DTK-25 extrahiert und eine Vielzahl an gebäudebezogenen Formmerkmalen wie Größe und Form, als auch Informationen über die Nachbarschaftsbeziehung berechnet und in einem

hierarchischen, regelbezogenen Verfahren in 8 Gebäudetypen klassifiziert. Die klassifizierten Einzelgebäude werden in einem weiteren Schritt ebenfalls mit einem regelbezogenen Verfahren auf der räumlichen Ebene der Baublöcke aggregiert. Diese Aggregation zu neun *Siedlungsstrukturtypen* erfolgt nach dem Prinzip der Flächendominanz der klassifizierten Einzelgebäude und Nutzungsinformationen aus den ATKIS-Baublöcken.

Einen Sonderfall der Arbeiten zur Gebäudetypisierung in dieser Auflistung stellt der Ansatz von Aubrecht et al. (2009) dar, da sowohl eine Gebäudetypisierung als auch eine Verknüpfung mit sozioökonomischen Daten durchgeführt wird. Die Autoren beschreiben zunächst die Ableitung von Einzelgebäuden aus einem digitalen Oberflächenmodell und eine anschließende Erstellung eines Gebäudeteilmodells auf Basis von geokodierten Adressdaten. Die Typisierung der Einzelgebäude erfolgt dabei nicht nach ihrer Form und Größe, sondern nach ihrer Nutzung. Dafür verwenden die Autoren sowohl Flächenwidmungsinformationen als auch punktuelle Firmendaten. Das finale Produkt, ein funktionales 3D-Gebäudemodell, wird für die detaillierte räumliche Disaggregation von Bevölkerungsdaten verwendet.

### 2.2.4.2. Strukturebene

Arbeiten zur direkten Klassifikation von urbanen Struktureinheiten auf einer räumlich übergeordneten Ebene, wie der Ebene des Baublocks, wurden in vielen deutschen Städten bis zum Ende des 20. Jahrhunderts durchgeführt. Das Ziel dieser Arbeiten war die Klassifikation von städtischen Strukturtypen oder städtischen Biotoptypen für eine nachhaltige, ökologische Stadtplanung. Die Vorgaben zu diesen Arbeiten gehen auf die Arbeitsgruppe „Methodik der Biotopkartierung im besiedelten Bereich“ zurück, welche die Klassifizierungen auf Basis von visuellen Luftbildinterpretationen, Kartierungen und Geländebegehungen empfiehlt (Schulte et al., 1986; Schulte, Sukopp und Werner, 1993). Nach diesem Vorgehen wurden bis zum Jahr 1993 in 160 deutschen Mittel- und Großstädten Biotoptypenkartierungen durchgeführt, aber durch den entstehenden Zeit- und Kostenaufwand bei der visuellen Auswertung dieser Datengrundlagen nicht regelmäßig aktualisiert. Durch den Zugang zu Satellitenbilddaten entfielen aufwändige Luftbildbefliegungen. Erste Arbeiten zur Ableitung relevanter Parameter für die Klassifikation städtischer Strukturtypen mit Unterstützung von Satellitenbilddaten wurden von Breuste, Wächter und Bauer (2001) zur Erhebung des Grünanteiles und von Socher (1999) zur Abschätzung des Versiegelungsgrades mittels SPOT 4 und IRS-1-C Daten gemacht.

Zur weitgehend automatisierten, computergestützten Klassifikation von städtischen Struktur- und Nutzungstypen auf Basis von Fernerkundungsdaten finden sich in der Literatur Ansätze mit unterschiedlichen Datengrundlagen, Methoden und Maßstäben. Eine frühe Arbeit von Møller-Jensen (1990) beschreibt die Klassifikation von sieben Strukturtypen und drei Landbedeckungsklassen in einer Landsat-Aufnahme. Der Autor verwendet Segmentierungsverfahren für die Abgrenzung von homogenen Raumeinheiten und ein

hierarchisches, wissenbasiertes Verfahren für die Klassifikation auf Basis von Texturmerkmalen, spektralen Merkmalen und Kontextmerkmalen. Die Grundidee von späteren Klassifikationsverfahren ist bereits in dieser Arbeit zu finden, allerdings ist die geometrische Auflösung der Landsat-Daten für eine belastbare Abgrenzung und Klassifikation von kleinräumigen Strukturen nicht ausreichend.

In den letzten Jahren wurden für die Klassifikation von Bebauungsklassen auf der Strukturebene vermehrt hoch aufgelöste Bilddaten unter Hinzunahme von digitalen Oberflächenmodellen und Vektordaten verwendet. Ein Verfahren zur Beschreibung von Landnutzungsveränderungen auf Basis von Landschaftsstrukturmaßen und CIR-Luftbildern aus drei verschiedenen Zeitpunkten beschreiben Herold, Scean und Clarke (2002). Die Autoren quantifizieren die Landnutzungsveränderungen von drei Stadtstrukturtypen anhand von sieben Landschaftsstrukturmaßen: Fraktale Dimension, Anteil der Klassenfläche an der gesamten Fläche, *patch density*, *patch size standard deviation*, *edge density*, *area weighted mean patch fractal dimension* und der *contagion index*. Das Konzept der Landschaftsstrukturmaße kommt ursprünglich aus der Landschaftsökologie und beinhaltet eine quantitative Beschreibung von räumlichen Einheiten in unterschiedlichen räumlichen Hierarchien (O'Neill et al., 1988; McGarigal et al., 2011). In einer Weiterentwicklung des beschriebenen Ansatzes wird die Methode auch auf Ikonos Satellitenbilddaten angewendet und die Strukturtypen nicht nur quantitativ beschrieben, sondern auch klassifiziert. Die Autoren Herold, Liu und Clarke (2003) beschreiben ein objektbasiertes Klassifikationsverfahren zur Ableitung der Landbedeckung und eine Analyse von Landschaftsstrukturmaßen und Texturmaßen basierend auf der Grey level co-occurrence matrix (GLCM) für die Charakterisierung und überwachte Klassifikation von neun Strukturtypen auf Basis von *Fisher's Linear Discriminant*. Eine ausführliche Beschreibung dieses Ansatzes und weitere Entwicklungen finden sich bei Herold (2004).

Auf Basis von CIR-Orthofotos klassifizieren Banzhaf und Höfer (2008) verschiedene Bebauungsklassen. Die Autoren verwenden ein objektbasiertes Klassifikationsverfahren für die Ableitung verschiedener Landbedeckungsklassen (Wald, Gras-/Rasenflächen, Gebäude, versiegelte Flächen, offener Boden, Schatten). Die abgeleiteten Klassen werden auf der räumlichen Ebene des Baublocks gemeinsam mit Nutzungsinformationen des Amtlichen Topographisch-Kartographischen Informationssystems (ATKIS) auf Basis eines wissenbasierten Verfahrens zu fünfzehn verschiedenen Stadtstrukturtypen zusammengefasst. Die Merkmale müssen allerdings für jede Klassifikation an die Datengrundlage angepasst werden.

Bochow (2009) beschreibt die Klassifikation von städtischen Biotoptypen, welche neben verschiedenen Flächennutzungstypen auch sechs verschiedene Bebauungsklassen beinhaltet. In diesem Verfahren werden hyperspektrale Flugzeugscannerdaten und digitale Oberflächenmodelle verwendet, um mittels eines objektbasierten Klassifikationsverfahrens 67 verschiedene Oberflächenarten zu klassifizieren. Für die Klassifikation der Bebauungsklassen

verwendet der Autor eine sehr große Anzahl an gebäudebezogenen Merkmalen, klassenbezogenen Merkmalen und Nachbarschaftsmerkmale innerhalb des zu klassifizierenden Biotoptyps. Anhand von zwei unterschiedlichen Klassifikationsverfahren, nämlich einem *Maximum-Likelihood*-Klassifikator und einem wissensbasierten Klassifikationsmodell basierend auf *Fuzzy-Logik*, vergleicht der Autor das Automatisierungspotenzial von Biotoptypenklassifikationen.

Neben den oben genannten Verfahren zur Klassifikation von städtischen Strukturtypen oder Gebäudetypen, beschreiben Heiden et al. (2012) ein Verfahren zur Ableitung von sieben räumlichen Indikatoren für die Charakterisierung von neun verschiedenen Stadtstrukturtypen. Die Autoren verwenden flugzeuggetragene Hyperspektraldaten und ein digitales Oberflächenmodell, um zunächst die Oberflächenmaterialien mit einem mittleren Fehler kleiner als 8 % zu klassifizieren. Diese 38 Oberflächenmaterialien bilden gemeinsam mit der Höheninformation die Grundlage, um die räumlichen Indikatoren zu klassifizieren, anhand welcher die Strukturtypen charakterisiert werden.

Eine Übersicht über die in diesem Zusammenhang beschriebenen Arbeiten zur Ableitung von Bebauungsklassen findet sich in Tabelle 1. Sie zeigt die unterschiedlichen Daten, Methoden und semantischen Bebauungsklassen sowohl auf der Gebäudeebene als auch auf der Strukturebene.

**Tabelle 1:** Übersicht über Verfahren zur Ableitung von Bebauungsklassen.

Autoren	Daten	Methode	Merkmale	Bezeichnung	Kl.	Klassenamen	Ebene
							G S
<b>Henn et al. (2012)</b>	Gebäudemodell	SVM	Form Nutzung Nachbarschaft	<i>Building types</i>	7	Einzelhaus, Doppelhaus, Reihenhaushaus, Gründerzeitbebauung, Villa, Zeilenbebauung, sehr groß dimensionierte Gebäude	x
<b>Meinel et al. (2008)</b>	DTK-25 ATKIS	regelbezogen hierarchisch	Form Nutzung Nachbarschaft	Gebäudetypen Stadtstrukturtypen	8/9	Mehrfamilienhaus (MFH) in geschl. Blockbauweise, freist. MFH, MFH in offener Blockstruktur, tradit. Zeilenbebauung, industrielles MFH in Zeile, Hochhaus, Ein- und Zweifamilienreihenhäuser, Ein- und Zweifamilienhaus, Reihenhaushaus, dörflich traditionelles Haus, Geschl. Blockbebauung, off. Blockbebauung, Zeilenbebauung (tradit.) Zeilenbebauung (industr.), Hochhausbebauung, Ein- und Zweifamilienhausbebauung, Reihenhaushausbebauung, dörflich traditionelle Bauweise, Industrie/Gewerbe, besondere funktionale Prägung	X X
<b>Aubrecht (2009)</b>	Digitales Oberflächenmodell Adressdaten	diverse	Nutzung	Nutzungstypen	11	Wohngebäude, gemischte Nutzung, Kleingartenhaus, Garage/Nebengebäude, Dienstleistung, öffentliche Nutzung, Landwirtschaft, Freizeit, Gewerbe, Erzeugung, Handel, unbekannte Nutzung	X
<b>Møller-Jensen (1990)</b>	Landsat	Segmentierung hierarchisch wissensbasiert	Textur Spektral Kontext	<i>Urban land-use</i>	7	dicht bebaute Gebiete, dicht bebaute Gebiete mit einem hohen Anteil an Slum und Wohnbebauung von niedrigem Status, mitteldichte Bebauung von Einzelgebäuden, mitteldichte Bebauung von Einzelgebäuden mit einem hohen Anteil an Slum und Wohnbebauung von niedrigem Status, Einzelgebäude geringer Dichte, Bebauung geringer Dichte, öffentliche Gebäude	X
<b>Herold et al. (2003)</b>	Ikonos Digitales Oberflächenmodell Baublockgrenzen	Fisher's Linear Discriminant	Landscape metrics	<i>Land-use class</i>	9	Ein- und Zweifamilienhausbebauung geringer Dichte, mittlerer Dichte, hoher Dichte, MFH, Industrie und Gewerbe, öffentliche Einrichtungen, Freizeitanlagen, Agrarland, Wälder und Feuchtgebiete	X
<b>Banzhaf und Höfer (2008)</b>	CIR Orthofotos ATKIS	wissensbasiert	Form Kontext Spektral Textur	Stadtstrukturtypen	15	Durchgehende Blockbebauung, nicht durchgehende Blockbebauung, durchgehende Blockrandbebauung, nicht durchgehende Blockrandbebauung, MFH, Gebäude nach 1960, Ein- und Zweifamilienhäuser, Industrie und Gewerbe, öffentliche Gebäude, Einkaufszentren, Verkehrsflächen, Parks und Grünflächen, Kleingärten, Friedhöfe, Freizeitanlagen	X
<b>Bochow (2009)</b>	Hyperspektral Digitales Oberflächenmodell	überwacht wissensbasiert	Gebäude Nachbarschaft	Biotoptypen	6	Einzel- und Reihenhaushausbebauung, Block-, Blockrand- und Zeilenbebauung, Großform- und Hochhausbebauung, öffentl. Einrichtungen, sonstige Bebauung, Kerngebiet (City)	X

Kl.: Klassen, Ebene G: Gebäudeebene, Ebene S: Strukturebene

### 2.2.5. Erdbeobachtungsdaten für die sozialwissenschaftliche Forschung

Während die Verwendung von Erdbeobachtungsdaten in vielen naturwissenschaftlichen und vor allem umweltwissenschaftlichen Forschungsdisziplinen bereits zu den Standardverfahren gezählt werden kann, stellt ihre Verwendung in sozialwissenschaftlichen Studien ein relativ junges Betätigungsfeld dar. Es existieren zwar viele Anwendungen, welche Erdbeobachtungsinformationen in einen sozialen Bezug stellen, aber nur wenige Arbeiten, welche sie für die sozialwissenschaftliche Forschung verwenden. Vor allem die Unterschiede in den Forschungstraditionen, in den Datengrundlagen, Fragestellungen und Analysemethoden verzögerten die interdisziplinäre Verwendung der Daten. Darüberhinaus ist die Sozialwissenschaft stärker darauf fokussiert zu erfahren, *warum* Dinge passieren und nicht *wo* sie passieren. Weitere Gründe für die geringere Akzeptanz der Fernerkundung in der Vergangenheit sind auch, dass die traditionellen, für Sozialwissenschaftler relevanten Variablen „[...] nicht direkt aus der Luft gemessen werden können sondern meist nur indirekt abgeleitet werden können [...]“ (Turner, 1998).

Wie oben erwähnt, finden sich in der Literatur sehr wohl einige Arbeiten, welche Fernerkundungsdaten für die Generierung sozialrelevanten Wissens verwenden (z.B. in der Meteorologie, Katastrophenhilfe oder Erntevorhersagen), die Beispiele für die *sozialwissenschaftliche Forschung* sind allerdings deutlich geringer. Durch die Interdisziplinarität der Arbeiten findet sich die Literatur auch selten in reinen Fernerkundungszeitschriften oder Sozialwissenschaftlichen Zeitschriften, sondern eben in interdisziplinären Zeitschriften (Hall, 2010).

Aus der Literatur wird untenstehend auf einige Beispiele verwiesen, welche durch ihre inhaltliche Fragestellung, ihre Methode oder ihre verwendeten Daten mit der vorliegenden Arbeit in einen Bezug gebracht werden können. Es existieren diesbezüglich einige Arbeiten mit demographischem Bezug, da der Zusammenhang zwischen den verwendeten Daten und sozialwissenschaftlichem Mehrwert klar ersichtlich ist. In diesem Zusammenhang sind vor allem Arbeiten zur Abschätzung von Bevölkerungsgrößen und -verteilungen auf Basis von Satellitenbilddaten zu nennen. Es können zwar keine Menschen direkt aus Satellitenbilddaten gezählt werden, allerdings ist eine indirekte Ableitung der Bevölkerung über die Kartierung von Stadtflächen oder die Identifikation von individuellen Wohneinheiten möglich (Jensen und Cowen, 1999). Um aus der abgeleiteten Bebauungsstruktur auf Bevölkerungsgrößen zu schließen, sind ungefähre Kenngrößen der Bevölkerung notwendig und einige Annahmen zu treffen, wie zum Beispiel, dass sämtliche Wohneinheiten auch tatsächlich bewohnt sind. Für Gebiete, welche durch starke Abwanderung gekennzeichnet sind, müssen die Annahmen/Hypothesen entsprechend angepasst werden. In der Literatur werden Arbeiten zu *bottom-up*- und *top-down*-Bevölkerungsverteilungen und Auswertungen über den statistischen Zusammenhang zwischen Bildmerkmalen und Bevölkerungszahlen beschrieben.

Der *bottom-up* Ansatz benötigt *in-situ* Kenngrößen für ausgewählte Bereiche des Untersuchungsgebietes worüber die Bevölkerungszahlen hochgerechnet werden können. Jensen und Cowen (1999) erzielen mit dieser Methode auf Basis von Luftbildern eine hohe Übereinstimmung mit Referenzdaten ( $r^2=0,81$ ).

Liu, Clarke und Herold (2006) untersuchen den Zusammenhang zwischen verschiedenen Texturmaßen und der Bevölkerungsdichte, argumentieren aber, dass der statistische Zusammenhang nicht stark genug ist, um genaue Abschätzungen über die ansässige Bevölkerung zu machen. Allerdings eignet sich die Methode, um Registerzählungen räumlich zu verfeinern. Aminipouri, Sliuzas und Kuffer (2009) verwenden geometrisch hoch aufgelöste Luftbilddaten für die Zählung von Einzelgebäuden in einem Slumgebiet in Daressalam, Tansania und schätzen die ansässige Bevölkerung mit Genauigkeiten zwischen 68 und 82 %.

In *top-down*-Ansätzen werden mittels räumlicher Disaggregation bekannte Bevölkerungsgrößen auf Siedlungsstrukturen verteilt (Steinnocher et al., 2005). Die Ansätze dieses Verfahrens, im englischen Sprachraum auch als *dasymeric mapping* bezeichnet, gehen in das frühe 20. Jahrhundert zurück (Wright, 1936). Auf diese Weise ist es möglich, die Verteilung einer Bevölkerung im Raum, beispielsweise auf nationaler Ebene oder innerhalb einer Stadt, abzuschätzen. Taubenböck, Roth und Dech (2007) verwenden sowohl *top-down*- als auch *bottom-up*-Abschätzungen am Beispiel der Megastadt Istanbul. Die Vergleiche für beide Verfahren zeigen Genauigkeiten in ähnlichen Dimensionen. Weitere Ausführungen über Fernerkundung und Bevölkerungsabschätzungen finden sich bei Baudot (2000) und eine Aufarbeitung von Literatur zu Methoden des *dasymeric mapping* im GIS-Umfeld findet sich bei Reibel (2007).

Den statistischen Zusammenhang zwischen Bildinformation und Bevölkerungsdichte untersucht Chen (2002) mit Landsat-TM Daten für New South Wales, Australien. Sutton et al. (1997) untersuchen anhand von Nachtaufnahmen des Defense Meteorological Satellite Program Operational Linescan System (DMSP-OLS) den statistischen Zusammenhang zwischen Bevölkerungszahlen und Helligkeit mit einer geometrischen Auflösung von 1x1 km für die Vereinigten Staaten von Amerika. Sutton et al. (2001) schätzen mit den gleichen Datengrundlagen für knapp 23000 Städte weltweit die Bevölkerungszahlen ab. Eine detaillierte Beschreibung über die Nutzung von DMSP-OLS Daten als Hilfsvariable vor allem für ökonomische Untersuchungen findet sich bei Chen und Nordhaus (2010) und für die Schätzung des Bruttoinlandsproduktes bei Henderson, Storeygard und Weil (2012).

Den gleichen Datensatz verwenden auch Vollmer, Michel und Randler (2012) für die Bestimmung eines empirisch ermittelten Zusammenhangs zwischen menschlichen Schlaf- und Wachtypen und nächtlichem Licht. In einer Befragung von 1507 Erwachsenen wurde ein Bestimmtheitsmaß von 0,201 ermittelt.

Neben der Erhebung von Bevölkerungsgrößen und ihrer Verteilung im Raum haben sich einige Autoren mit der Erhebung physischer Merkmale von städtischen Gebieten aus

Fernerkundungsdaten beschäftigt, welche beispielsweise als Indikatoren über die Lebensqualität verwendet werden können. Einige Arbeiten versuchen einen Zusammenhang zwischen der Ausstattung einer Stadt mit Grünflächen bzw. dem Anteil an versiegelten Flächen und sozioökonomischen Merkmalen herzustellen. Lo und Faber (1997) klassifizieren in diesem Zusammenhang Vegetationsflächen aus Landsat-TM Daten und finden einen positiven, statistischen Zusammenhang zwischen Grünanteil und Einkommen und dem Wert der Immobilie. Umgekehrt fanden sie einen negativen Zusammenhang mit der Bevölkerungsdichte. Einen ähnlichen Ansatz verwenden Jensen et al. (2004) um den positiven Zusammenhang zwischen der *urban leaf area*, dem Einkommen und dem Wert der Immobilie zu untersuchen. Im Gegensatz zu dieser Arbeit hat Emmanuel (1997) den Zusammenhang zwischen Grünanteil und sozialer Benachteiligung festgestellt. Es kann aus der Betrachtung dieser Beispiele festgehalten werden, dass die Berücksichtigung von kontextueller Information aus Fernerkundungsdaten für die sozialwissenschaftliche Erklärung von bestimmten Eigenschaften noch nicht ausreichend untersucht ist. Die Unterschiedlichkeit der in der Literatur gefundenen Erkenntnisse unterstreicht aber die Wichtigkeit für weitere Forschungsarbeiten in diesem Zusammenhang.

Andere Autoren beschreiben Verfahren, in denen die physische Ausstattung der Stadt mittels Erdbeobachtungsdaten erfasst wird um dadurch indirekt Rückschlüsse auf sozioökonomische Merkmale zu ziehen. Avelar, Zah und Tavares-Corre (2009) und Tapiador et al. (2011) klassifizieren Ikonos-Satellitenbilder im ariden Lima, Peru um bestimmte Oberflächenmaterialien, Bebauungstypen und private Schwimmbäder zu klassifizieren und mit sozioökonomischen Daten zu vergleichen. Anhand des Vorhandenseins bzw. Nichtvorhandenseins bestimmter physischer Merkmale wie beispielsweise Vegetation kann auf einen sozioökonomischen Status geschlossen werden. Die Autoren beschreiben einen hohen statistischen Zusammenhang ( $R^2=0,70$ ) im Vergleich mit Referenzdaten. In weiteren Arbeiten wird ein Zusammenhang zwischen der Gebäudeform und -anordnung und dem sozioökonomischen Status hergestellt. Diese Ansätze sind vor allem in Schwellen- und Entwicklungsländern anwendbar, da hier besonders durch die Form und Anordnung von Gebäuden Rückschlüsse auf den Status der Bewohner gemacht werden kann. Slums zeichnen sich beispielsweise durch ihre besonders hohe Bebauungsdichte und kleine Gebäude aus. Weitere Arbeiten, welche sich vor allem mit der Ausweisung von Slumgebieten mit Methoden der Fernerkundung beschäftigen, sind zu finden in: Sliuzas und Kuffer (2008), Sliuzas, Kerle und Kuffer (2008), Taubenböck et al. (2009) und Baud et al. (2010).

In der Stadtforschung eignen sich kleinmaßstäbige Fernerkundungsdaten vor allem auch für die Kartierung und Abgrenzung von städtischen Gebieten. Durch charakteristische Signaturen der von Menschen gebauten Städte können diese in Satellitenbildern identifiziert werden. Dieser Umstand kommt insbesondere bei der Abgrenzung von schnell wachsenden Megastädten zur Anwendung, da hier die Fernerkundung aktuelle Daten bereitstellen kann (z.B. Griffiths et al., 2010). Seit den 1970er Jahren finden sich in der Literatur auch vermehrt

Arbeiten, welche auf Basis von Satellitenbilddaten (z.B. AVHRR oder Landsat) den *urban heat island effect* messen (z.B. Roth, Oke und Emery (1989) und somit relevante Daten für die Stadtklimatologie zur Verfügung stellen können. Neben der Nutzung von Fernerkundungsdaten für die Erstellung von Variablen in sozialwissenschaftlichen Analysen werden umgekehrt sozialwissenschaftliche Daten verwendet, um die Qualität von Fernerkundungsprodukten zu verbessern (z.B. Harris und Ventura, 1995).

Übersichten über interdisziplinäre Arbeiten zwischen Erdbeobachtung und Sozialwissenschaften finden sich auch bei Rindfuss und Stern (1998), Liverman und Cuesta (2008), bei De Sherbinin et al. (2006) und Hall (2010). Die letzte Autorin erwähnt auch, dass es zwar vielfältige Anwendungsfelder zwischen den beiden Forschungsdisziplinen gibt, aber noch kein ausgeprägter Reifegrad entwickelt wurde. Darüberhinaus besteht die heutige Herausforderung weniger in technischen Problemen, sondern in einem umfassenden Verständnis von Erdbeobachtung mit ihren Chancen und Limitierungen als räumliche Informationsquelle für die Sozialwissenschaften.

### **2.3. Sozialwissenschaftliche Surveydaten: SOEP und BASE-II**

Die sozialwissenschaftlichen Daten, welche in dieser Arbeit mit Erdbeobachtungsdaten verknüpft werden, entstammen der deutschlandweiten Längsschnittstudie *das sozioökonomische Panel*. Die Studie stellt eine der weltweit am meisten genutzten Längsschnittdaten dar und ist vor allem in sozial- und wirtschaftswissenschaftlichen Untersuchungen, sowie psychologischen und medizinischen Untersuchungen sehr bekannt und entsprechend verbreitet. Da die Nutzung dieser Daten für Wissenschaftler aus den Bereichen der Fernerkundung weniger bekannt ist, wird in diesem Abschnitt auf die relevanten Informationen über die Surveys SOEP und BASE-II eingegangen. Zunächst wird der Aufbau des SOEP vorgestellt und danach das traditionelle Konzept des Raumbezuges und auf raumbezogene Forschungen mit dem Datensatz eingegangen. Daraufaufgehend werden die in der Arbeit verwendeten Merkmale beschrieben. Da die Struktur der Surveydaten von BASE-II gleich mit der des SOEP ist, wird nur das SOEP ausführlich beschrieben. Die Beschreibung von BASE-II erfolgt ergänzend und entsprechend knapp (vgl. Abschnitt 2.3.4).

#### **2.3.1. Die SOEP-Studie**

Verantwortlich für die Planung, Durchführung und Aufbereitung der Längsschnittstudie „*das sozio-ökonomische Panel*“ ist die gleichnamige Forschungseinheit des Deutschen Instituts für Wirtschaftsforschung Berlin (DIW Berlin). Die Studie basiert auf einer seit 1984 laufenden, repräsentativen Wiederholungsbefragung von etwa 23000 Personen in über 12000

Privathaushalten in ganz Deutschland (Wagner et al., 2008). Im Rahmen dieses *Surveys* werden Daten für die sozial- und wirtschaftswissenschaftliche Grundlagenforschung erhoben. Die Teilnehmer werden Jahr für Jahr zu ihren objektiven und subjektiven Lebensbedingungen befragt. Es werden repräsentative Mikrodaten auf Personen-, Haushalts- und Familienebene erhoben. Die Teilnehmer werden zu objektiven Indikatoren wie beispielsweise dem Einkommen und subjektiven Indikatoren wie die persönliche Zufriedenheit befragt, wodurch ein Bild über den Wandel der Gesellschaft dargestellt werden kann.

Die Themengebiete decken unterschiedliche Bereiche in den einzelnen Lebensabschnitten der Teilnehmer ab. Dies beinhaltet Themen wie die Haushaltszusammensetzung, die Erwerbs- und Familienbiographie, die Erwerbsbeteiligung und die berufliche Mobilität, individuelle Einkommensverläufe, die Gesundheit und Lebenszufriedenheit. Die zentralen, übergeordneten Fragestellungen des SOEP sind in Tabelle 2 dargestellt.

**Tabelle 2:** Zentrale Themen des SOEP. (Quelle: Wagner et al., 2008).

---

<b>Zentrale Themen des SOEP</b>	
Demographie und Wohnsituation	Vorschul- und Schulbildung, berufliche Bildung und Weiterbildung
Arbeitsmarkt und Berufsmobilität	Einkommen, Vermögen und soziale Sicherheit
Sorgen und Zufriedenheit	Gesundheit
Haushaltsproduktion	Grundlegende Orientierungen (Vorlieben, Werte usw.)

---

Somit gibt der Datensatz Auskunft über viele objektive Lebensbedingungen, Persönlichkeitsmerkmale, Wertvorstellungen, Risikoeinstellungen und über Abhängigkeiten zwischen diesen Bereichen und deren Veränderungen. Die Daten beinhalten über vier Millionen Beobachtungen und über 37000 Variablen (Goebel et al., 2008). Anhand des SOEP können eine Vielzahl an Theorien aus den Sozial- und Wirtschaftswissenschaften und der Psychologie getestet werden (Wagner et al., 2008: 303).

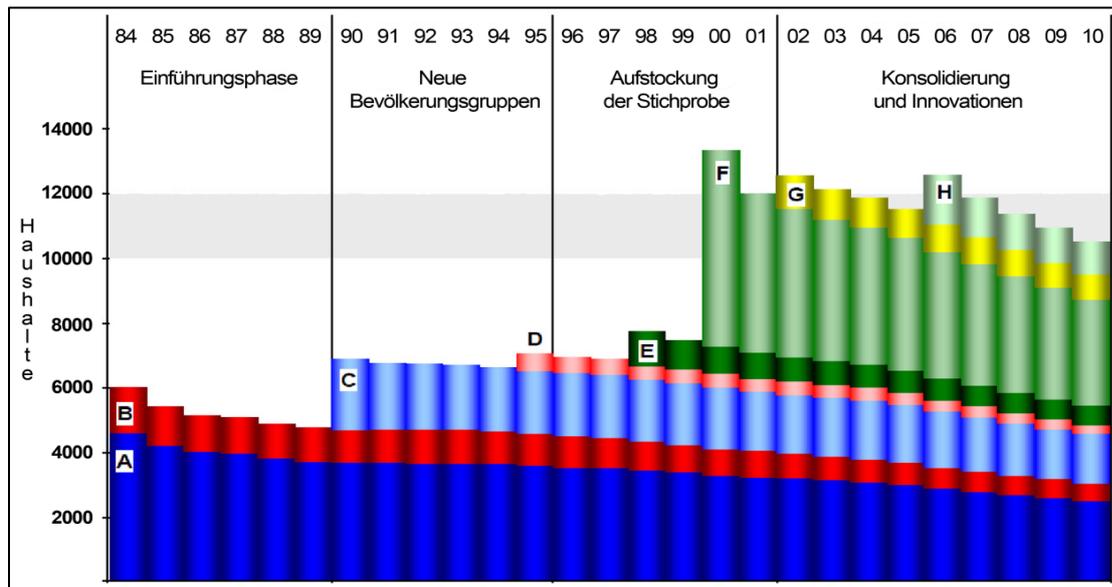
Die erste Befragungswelle wurde 1984 in der Bundesrepublik Deutschland (BRD) an etwa 12000 Personen in beinahe 6000 Haushalten in zwei Stichproben durchgeführt. Hauptstichprobe A repräsentiert die Bevölkerung der Privathaushalte; die Sonderstichprobe B stellt einen Querschnitt über die damaligen Gastarbeiter dar (Wagner, Frick und Schupp, 2007). Im Juni 1990, also noch vor der deutschen Wiedervereinigung, wurde die Befragung auf die damalige Deutsche Demokratische Republik (DDR) ausgeweitet. Im Laufe der Zeit wurden auch weitere Sonderstichproben (z.B. Migranten, Hocheinkommen) gezogen (vgl. Tabelle 3). Die große Immigrationswelle zwischen 1985 und dem Beginn der 1990er Jahre wurde in einer weiteren Sonderstichprobe D abgedeckt. In den beiden Auffrischungstichproben E und H wurden zusätzliche Haushalte gezogen, um für eine Stabilisierung der Teilnehmerzahl zu sorgen. Durch den Längsschnittcharakter der Studie wird die Anzahl der teilnehmenden

Haushalte über die Jahre immer kleiner (vgl. Abbildung 4). Die Ursachen dafür können zum Beispiel Umzug, Krankheit oder verlorenes Interesse an der weiteren Teilnahme der Studie sein. Besonders interessant für die Längsschnittuntersuchungen sind die Teilnehmer, welche über den gesamten Zeitraum der Untersuchung teilgenommen haben. Je länger die Studie andauert, desto kleiner wird auch die Zahl der ununterbrochenen Teilnehmer. Wenn sich die ausgeschiedenen Teilnehmer durch bestimmte Merkmale von den verbleibenden Teilnehmern unterscheiden, wird in diesem Zusammenhang von *Panelmortalität* gesprochen. Sie kann zu systematischen Verzerrungen der Daten führen und eine Schwierigkeit bei der Interpretation der Daten darstellen.

Im Jahr 2006 wurde neuerlich eine zufällige Auffrischungstichprobe über die Gesamtbevölkerung von rund 1500 Haushalten gezogen. Neue Personen gelangen in das SOEP durch Ziehung neuer Stichproben, durch Geburt und Umzug in einen bestehenden SOEP-Haushalt, das Erreichen des 18. Lebensjahres oder durch Teilung eines bestehenden SOEP-Haushaltes in einen neuen SOEP-Haushalt.

**Tabelle 3:** Stichproben im SOEP. (Quelle: Wagner et al., 2008).

Stichprobe	Bezeichnung	Jahr	Anzahl Haushalte	Beschreibung (HH=Haushalt)
A	Westdeutsche	1984	4528	HH-Vorstand ist D. oder and. National. als Sample B
B	Ausländer	1984	1393	HH-Vorstand ist Türke, Ital., Span., Gr. oder Jugosl.
C	Ostdeutsche	1990	2179	HH-Vorstand war DDR Bürger
D	Immigranten	1994/95	522	Mind. ein HH-mitglied migrierte nach D nach 1984
E	Auffrischung	1998	1067	Zufälliges Sample über die Gesamtbevölkerung
F	Innovation	2000	6052	Zufälliges Sample über die Gesamtbevölkerung
G	Hocheinkommen	2002	1224	Monatl. HH-Nettoeinkommen > 7500DM (4500€)
H	Auffrischung	2006	1506	Zufälliges Sample über die Gesamtbevölkerung



**Abbildung 4:** Sonderstichproben und Teilnehmerentwicklung der ersten 27 Befragungswellen im SOEP.  
(Verändert nach: Wagner, Frick und Schupp, 2007)

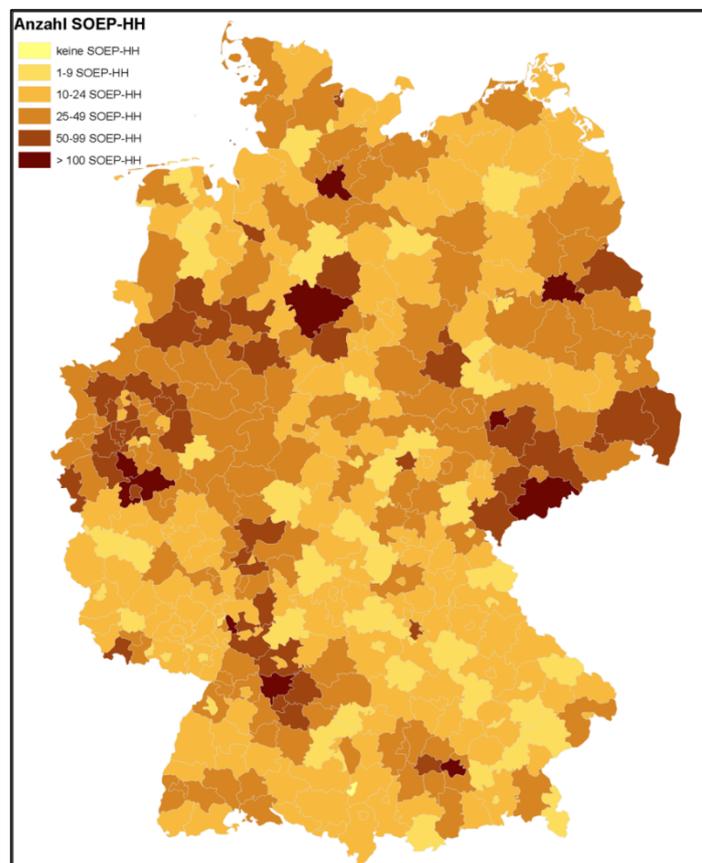
Neben dem SOEP gibt es in Deutschland eine Reihe an Längsschnittstudien wie z.B. das IAB-Betriebspanel (Bellmann, Kohaut und Lahner, 2002), die Sächsische Längsschnittstudie (Förster und Berth, 2011) oder die Berliner Altersstudie BASE (Mayer und Baltes, 1996). Im internationalen Umfeld vergleichbare Studien sind z.B. HILDA - Household, income and labour dynamics in Australia Survey (Melbourne Institute, 2011), *BHPS* - British household panel study (BHPS, 2011), *SLID* - Survey of Labor and Income Dynamics in Canada (Statistics Canada, 2011), *SHP* - Schweizer Haushaltspanel (FORS, 2011) oder *PSID* - Panel study of income dynamics (PSID, 2011) in den Vereinigten Staaten von Amerika. Die Daten aus den Studien dieser Länder können anhand des übergreifenden Datensatzes Cross national equivalent file (CNEF, 2011) gemeinsam ausgewertet werden (Haisken-DeNew, 2001).

### 2.3.2. Der räumliche Bezug der SOEP-Daten

Die Erhebung der SOEP-Daten findet in der Wohnung der Teilnehmer statt. Das bedeutet, dass der befragte Haushalt eindeutig über die Wohnadresse für die Befragungen in den folgenden Jahren wieder aufgesucht werden kann. Eine Verortung der Haushalte in den Daten, welche für wissenschaftliche Zwecke der Forschergemeinschaft zu Verfügung gestellt werden, ist allerdings nicht möglich, da die Daten faktisch anonymisiert sind. Eine Regionalisierung der Daten ist im Standarddatensatz nur nach den Bundesländern möglich (Spieß, 2005). Weitere Angaben über den Wohnort des SOEP-Haushaltes beziehen sich nur mehr auf die Gemeindegröße, den Gemeindetyp nach Boustedt (Boustedt, 1953) und BIK (BIK, 2001), die

Entfernung zur nächsten Großstadt, die Siedlungsform, oder die Wohngegend (z.B. Lärmbelästigung, Luftverschmutzung, Mangel an Grünflächen, Ausländeranteil).

Unterhalb der Bundesländer ist eine räumliche Analyse der SOEP-Daten nur im Rahmen von speziellen Datenschutzbedingungen auf der Ebene der 96 Raumordnungsregionen (BBSR, 2011a), der über 400 Landkreise oder auf der Ebene der Postleitzahlenbezirke möglich. Die Aufbereitung der SOEP-Daten mit diesen Regionalinformationen ermöglicht die Verknüpfung von SOEP-Informationen mit weiteren statistischen Informationen, welche auf der gleichen räumlichen Ebene verfügbar sind. Problematisch ist die Auswertung dieser Daten aber durch die starke Streuung der teilnehmenden Haushalte in den Raumeinheiten, welche zwischen 7 und 521 Haushalte in den Raumordnungsregionen schwankt. Veränderungen im Gebietsstand der Landkreise erschweren eine raumbezogene Analyse über die Zeit (Goebel, 2007). Seit 1993 hat die Anzahl der Landkreise in Deutschland von 543 auf zunächst 439 (bis Ende 2006) und aktuell auf 413 abgenommen (BBSR, 2011b). Die Anzahl der in der SOEP-Stichprobe gezogenen Haushalte pro Landkreis hängt von der Gesamtbevölkerung des jeweiligen Landkreises ab. Somit ergibt sich ein stark differenziertes Bild der SOEP-Teilnehmer auf der Ebene der Landkreise mit Häufungen in den Großstädten und Ballungsräumen (vgl. Abbildung 5).



**Abbildung 5:** Anzahl der SOEP-Haushalte pro Landkreis.  
(Gebietsstand der Kreisgrenzen: 31.12.2000; Datengrundlage: SOEP 2009)

Die Reduzierung des räumlichen Bezuges der SOEP-Daten auf unterschiedliche administrative Raumeinheiten und der restriktive Datenschutz haben dazu geführt, dass die für die wissenschaftliche Forschung zur Verfügung gestellten Daten auch nur auf diesen räumlichen Bezugsebenen ausgewertet werden konnten. Kleinräumige Auswertungen und Verknüpfungen der SOEP-Daten mit zusätzlichen Informationen wurden auf der Ebene der Landkreise und auf der Ebene der Postleitzahlen gemacht. Die in den folgenden Unterabschnitten dargestellten Forschungsarbeiten stellen den Stand der Forschung zu räumlichen Auswertungen der SOEP-Studie auf unterschiedlichen Raumeinheiten dar.

### 2.3.2.1. Landkreise

Auf der Ebene der Landkreise untersuchen Rehdanz und Maddison (2009) mit einem hedonischen Ansatz die Annehmlichkeiten des lokalen Klimas für die Bewohner. Die Autoren verknüpfen Temperatur- und Niederschlagsdaten mit den Landkreisen und vergleichen sie mit dem Einkommen und den Werten der Häuser und Wohnungen aus den SOEP-Daten. Sie finden eine statistische Signifikanz der mittleren Temperaturen und Niederschläge im Jänner und im Juli auf diese ökonomischen Eigenschaften. Baumgartner (2004) untersucht die Effekte der Schulqualität auf ihre Auswirkung auf das spätere Einkommen. Als Hilfsvariable für die Schulqualität wird die Größe der Schulklassen verwendet, welche für die Landkreise aus statistischen Daten generiert wurden. Der Autor findet keine relevanten Effekte der Klassengröße auf das spätere Einkommen. Hank und Kreyenfeld (2004) untersuchen den Einfluss des Zugangs zu Kinderbetreuungseinrichtungen auf die Fruchtbarkeit westdeutscher Frauen. Sie verknüpfen hierzu ebenfalls statistische Daten der einzelnen Landkreise für die multivariate Analyse mit dem Ergebnis, dass es keinen statistisch signifikanten Zusammenhang gibt.

Eine Arbeit zur Untersuchung der Luftqualität auf Mietpreise und die Lebenszufriedenheit von SOEP-Teilnehmern beschreibt Luechinger (2009). Der Autor interpoliert punktuelle Daten aus den Jahren 1985, 1990 und 1995 über die Schwefeldioxidbelastung von 553 Meßstationen auf die Landkreise und verknüpft sie auf dieser räumlichen Ebene mit den Daten aus dem SOEP. Die in dieser Arbeit aufgestellten empirischen Modelle werden mit zusätzlichen Variablen kontrolliert und bestätigen nicht nur den negativen Zusammenhang zwischen der Schwefeldioxidkonzentration und der Lebenszufriedenheit, sondern auch einen negativen Effekt auf Mietpreise. Den Einfluss vom Wohnort auf die Lebensdauer und die Lebenszufriedenheit von SOEP-Teilnehmern untersuchen Gerstorff et al. (2010). Die Autoren analysieren in diesem Zusammenhang statistische Daten auf der Ebene der Landkreise (Wohlstand, Arbeitslosenrate und Zugang zu Gesundheitseinrichtungen) mit SOEP-Information auf Personenebene für über 3400 Teilnehmer. Die Ergebnisse zeigen einen Zusammenhang zwischen diesen regionalen Indikatoren auf der Landkreisebene und der

Lebenszufriedenheit im letzten Lebensabschnitt von Personen. Sie weisen allerdings auch darauf hin, dass eine detailliertere Erfassung der geographischen Distanz für den Zugang zu den Gesundheitseinrichtungen vorteilhaft wäre.

### **2.3.2.2. Postleitzahlenbezirke**

Auf einer räumlich feiner gegliederten Ebene untersuchen Spiess und Wrohlich (2008) die Auswirkungen der Distanz zwischen dem Wohnort und der nächsten Hochschuleinrichtung (Universitäten und Fachhochschulen) auf die Entscheidung, an einer Universität zu studieren. Sie berechnen die Distanzen zwischen den Postleitzahlenbezirken der SOEP-Teilnehmer, zu dem Jahr, in welchem sie die Hochschulreife erworben haben und den Postleitzahlenbezirken der nächsten Hochschule. Die Autoren finden einen signifikant positiven Effekt der Distanz für die Entscheidung zur Hochschule: pro Kilometer Nähe steigt die Wahrscheinlichkeit an einer Hochschule zu studieren um 0,2-0,3 Prozentpunkte. Eine weitere Arbeit auf dieser räumlichen Ebene untersucht inwiefern ethnische Nachbarschaften wirtschaftlich, sozial und/oder kulturell isolierte Räume darstellen (Drever, 2004). Die Autorin verknüpft SOEP-Daten mit Daten über die Konzentration an Minderheiten und Einkommen auf der Ebene der Postleitzahlen. Die Ergebnisse zeigen, dass trotzdem die Qualität der Nachbarschaften für Minderheiten in ethnischen Nachbarschaften geringer ist, sie aber nicht mehr kulturell isoliert sind als andere Minderheiten außerhalb ethnischer Nachbarschaften.

### **2.3.2.3. Mikrogeographische Nachbarschaften**

Nachbarschaftseffekte auf einer räumlich höher aufgelösten Ebene untersuchen Dittmann und Goebel (2010). Im SOEP-Datensatz werden standardmäßig Informationen über das nächste Wohnumfeld der Teilnehmer erhoben. Diese Informationen beinhalten den Gebäudetyp (Einzel-/Doppelhausbebauung, mehrgeschossige Wohnbebauung, Hochhausbebauung, landwirtschaftliches Wohngebäude; Privathaushalt oder Wohnheim) und das Wohnumfeld (Altbauwohngebiet, Neubauwohngebiet, Mischgebiet, Geschäftszentrum, Industriegebiet). Eine detailliertere Beschreibung über die Nachbarschaft ist im SOEP nicht möglich. Das räumliche Umfeld der Nachbarschaft ist durch den Interviewer bzw. den Interviewten festgelegt und kann nicht nachträglich verändert werden. Vor diesem Hintergrund haben die Autoren kommerzielle, mikrogeographische Daten über die direkte Nachbarschaft der SOEP-Teilnehmer (microm©) verknüpft (Goebel et al., 2007). Die verknüpfte Information beschreibt den sozioökonomischen Status der Nachbarschaft, welche mit Informationen über die Lebenszufriedenheit der SOEP-Teilnehmer in Bezug gebracht wird. Neben dem sozioökonomischen Status finden die Autoren allerdings auch starke Effekte des sozialen Netzes auf die Lebenszufriedenheit.

### 2.3.2.4. Andere Langzeitstudien

In vergleichbaren Langzeitstudien wie dem SOEP (vgl. Studien auf S. 30) werden räumliche Analysen ebenfalls auf unterschiedlichen administrativ-statistischen Raumeinheiten bis auf die kleinste verfügbare Ebene der Nachbarschaften, der *census collection districts* oder *statistical local areas* (ABS, 2005) durchgeführt (z.B. Sucoff und Upchurch, 1998; Harris, 1999). In einer aktuellen Arbeit (Ambrey und Fleming, 2011) werden für die australische Langzeitstudie HILDA die räumlichen Distanzen zu naturräumlichen Annehmlichkeiten (Nationalparks oder Küsten) als zusätzliche erklärende Variable für die Lebenszufriedenheit verwendet. Die Autoren berechnen dafür die Distanzen zwischen den Schwerpunkten der *statistical local areas* und den naturräumlichen Merkmalen und finden einen positiven, signifikanten Zusammenhang mit der individuellen Lebenszufriedenheit der Umfrageteilnehmer.

Anhand der raumbezogenen Arbeiten von Daten aus Längsschnittstudien ist erkennbar, dass sehr wohl ein Bewusstsein dafür besteht, das räumliche Umfeld von Umfrageteilnehmern in die Erklärung sozialwissenschaftlicher Hypothesen einzubeziehen. Allerdings sind viele Fragestellungen datengetrieben und durch die Zugänglichkeit und Verwendbarkeit von zusätzlichen Daten limitiert (siehe auch: Gerstorf et al., 2010). Eine ausführliche Betrachtung von Möglichkeiten zur Einbindung von geographischen Variablen und geographischen Angaben über die Lage der Untersuchungsteilnehmer ist bei Boyle et al. (2004) beschrieben. Der Bericht spiegelt das gewachsene Interesse der sozialwissenschaftlichen Forschungsgemeinschaft an kleinräumiger Information über die nächste Nachbarschaft von Befragungsteilnehmern wider. Darüber hinaus werden die vergleichbaren Längsschnittstudien im internationalen Umfeld auf die Möglichkeit untersucht, kleinräumige Untersuchungen durchzuführen. Das Ergebnis zeigt, dass bei allen Studien eine direkte Verortung der Teilnehmer nicht möglich ist, sondern höchstens unter strengen Datenschutzvorkehrungen die Verwendung von kleinräumigen statistischen Einheiten wie Postleitzahlenbezirken oder Zensusdistrikte möglich ist. Weitere, vor allem technische Arbeiten zur Verknüpfung von sozialwissenschaftlichen Umfragedaten mit Geodaten sind bei Downey (2006) und Reibel (2007) beschrieben.

### 2.3.2.5. SOEP-Studien und Fernerkundung

Während alle oben angeführten Arbeiten die Auswertung der SOEP-Daten immer auf Basis des klassischen Raumbezuges im Sinne von administrativen, statistischen oder mikrogeographischen Einheiten beziehen, wurden einige wenige Forschungsarbeiten unter Berücksichtigung von Erdbeobachtungsdaten veröffentlicht. Voigtländer et al. (2011) verwenden flächendeckende Umweltbelastungsdaten aus Fernerkundung und Bodenmessstationen, um die individuelle Umweltbelastung von SOEP-Teilnehmern über die Zeit zu untersuchen. Die Umweltbelastungsdaten liegen dabei in einer geometrischen Auflösung von 45x45km mit stündlichen Messwerten vor. Die einzelnen

Umweltbelastungsmerkmale  $\text{NO}_2$ ,  $\text{PM}_{10}$  und  $\text{O}_3$  werden über die geographische Lage der SOEP-Haushalte mit den subjektiven Angaben über die Umweltbelastungen der SOEP-Teilnehmern verknüpft. Die Auswertungen zeigen aufgrund der geringen räumlichen Auflösung der Daten allerdings neben einem kleinen, positiven Zusammenhang zwischen der subjektiven Gesundheit und der individuellen  $\text{NO}_2$ -Belastung keine nennenswerten Ergebnisse.

Für die kleinräumige Untersuchung des Armutsrisikos in Ost- und Westdeutschland verwenden Goebel und Wurm (2010) Daten aus CORINE Landcover (Corine Land Cover, 2010). Dabei leiten die Autoren physische Grenzen aus den Landbedeckungsklassen für alle deutschen Groß- und Mittelstädte ab und berechnen die Lage der SOEP-Haushalte innerhalb des Stadtgebietes bzw. ihre Lage zum Stadtrand. Auf Basis dieser Untersuchung zeigen die Autoren, dass das Armutsrisiko generell in der Nähe der Stadtzentren sehr hoch ist und am Stadtrand bzw. unmittelbar dahinter am niedrigsten ist.

In der Arbeit von Wurm, Goebel und Wagner (2010) werden erstmals auch kleinräumige Kontextmerkmale für die Städte Köln und München exemplarisch in die SOEP-Studie aufgenommen. Die Autoren leiten Kontextmerkmale wie Versiegelungsanteile und Vegetationsanteile aus hoch aufgelösten Satellitenbilddaten ab und berechnen exemplarisch den Zusammenhang dieser Kontextmerkmale mit der Wohnungsmiete. Sie stellen einen positiven Zusammenhang zwischen dem Grünflächenanteil in der Nachbarschaft und dem Quadratmetermietpreis fest.

### 2.3.3. SOEP-Variablen mit indirektem Raumbezug

Wie bereits beschrieben, werden im Rahmen der SOEP-Befragungen unterschiedliche Informationen über die Nachbarschaft der Umfrageteilnehmer erhoben. Dazu zählen subjektive Informationen über die räumliche Lage des Haushaltes innerhalb des Stadtgebietes bzw. zum nächstgelegenen Großstadtzentrum (Abbildung 6).

**54. Wie weit ist es von hier bis ins Zentrum der nächsten Großstadt?**

Wohnung liegt im Großstadtzentrum .....	<input type="checkbox"/>
unter 10 km .....	<input type="checkbox"/>
10 bis unter 25 km .....	<input type="checkbox"/>
25 bis unter 40 km .....	<input type="checkbox"/>
40 bis unter 60 km .....	<input type="checkbox"/>
60 km und mehr .....	<input type="checkbox"/>

**Abbildung 6:** Auszug aus dem Haushaltsfragebogen 2009 zur subjektiven Einschätzung gegenüber der Lage der Wohngegend des Haushaltes.

Quelle: SOEP Haushaltsfragebogen 2009.

Andere Fragen betreffen die Einschätzung über das Wohnumfeld. Das Wohnumfeld (WUM) wird über drei Fragen an den Haushaltsvorstand beschrieben. Die Erhebung dieser Informationen findet bei der erstmaligen Teilnahme am SOEP statt und wird danach in einem Intervall von fünf Jahren erneut erfragt (Abbildungen 7-9).

**5. Wie würden Sie die Wohngegend hier beschreiben?**

- Ein reines Wohngebiet mit überwiegend Altbauten .....
- Ein reines Wohngebiet mit überwiegend Neubauten .....
- Ein Mischgebiet mit Wohnungen und Geschäften bzw. Gewerbebetrieben .....
- Ein Geschäftszentrum (Läden, Banken, Verwaltungen) mit wenigen Wohnungen .....
- Ein Gewerbe- bzw. Industriegebiet mit wenigen Wohnungen .....

**Abbildung 7:** Auszug aus dem Haushaltsfragebogen 2009 zum Wohnumfeld (WUM1).

Quelle: SOEP Haushaltsfragebogen 2009.

**6. Was ist das für ein Haustyp, in dem Sie hier wohnen?**

- Landwirtschaftliches Wohngebäude .....
- Freistehendes Ein- / Zweifamilienhaus .....
- Ein- / Zweifamilienhaus als Reihenhaus oder Doppelhaus .....
- Wohnhaus mit 3 bis 4 Wohnungen .....
- Wohnhaus mit 5 bis 8 Wohnungen .....
- Wohnhaus mit 9 und mehr Wohnungen  
(aber höchstens 8 Stockwerken; also kein Hochhaus) .....
- Hochhaus (9 und mehr Stockwerke) .....

**Abbildung 8:** Auszug aus dem Haushaltsfragebogen 2009 zum Wohnumfeld (WUM2).

Quelle: SOEP Haushaltsfragebogen 2009.

**7. Handelt es sich um ein Wohnheim, eine Pension oder eine ähnliche Unterkunft?**

- Nein .....       Ja, und zwar: – Schüler- / Jugendlichenwohnheim .....
- Studentenwohnheim .....
- Berufstätigenwohnheim .....
- Altenheim / Pflegeheim .....
- Altenwohnheim .....
- Hotel / Pension .....
- Sonstiges Heim / Unterkunft .....

und  
zwar:

**Abbildung 9:** Auszug aus dem Haushaltsfragebogen 2009 zum Wohnumfeld (WUM3).

Quelle: SOEP Haushaltsfragebogen 2009.

Neben den Informationen zum direkten Wohnumfeld, werden auch subjektive Einschätzungen der Haushaltsvorstände zur Wohngegend abgefragt. Eine direkte Frage zum Anteil an Grünflächen in der direkten Nachbarschaft ist im SOEP nicht enthalten, eine Frage bezieht sich aber auf den Einfluss von fehlenden Grünflächen in der Nachbarschaft.

**56. Wie sehr fühlen Sie sich hier in dieser Wohngegend durch folgende Umwelteinflüsse beeinträchtigt?**

	Gar nicht	Gering	Gerade erträglich	Stark	Sehr stark
Durch Lärmbelästigung .....	<input type="checkbox"/>				
Durch Luftverschmutzung .....	<input type="checkbox"/>				
Durch <u>fehlende</u> zugängliche Grünflächen .....	<input type="checkbox"/>				

**Abbildung 10:** Auszug aus dem Haushaltsfragebogen 2009 zur subjektiven Einschätzung gegenüber Umwelteinflüssen.

Quelle: SOEP Haushaltsfragebogen 2009.

Die hier dargestellten Fragen aus dem SOEP zeigen exemplarisch, wie indirekte Nachbarschaftsinformationen im SOEP erhoben werden. Diese Angaben stellen *punktueller*, oder für eine räumlich nicht näher definierte Nachbarschaft, *subjektive* Informationen der Haushaltsvorstände dar. Eine individuelle Abgrenzung des Raumes für weitere Untersuchungen ist anhand dieser Daten nicht möglich, da keine flächendeckende Informationen vorliegen. Für die kleinräumige geographische Beschreibung der Nachbarschaft sind also objektive und flächendeckende Datengrundlagen erforderlich.

### 2.3.4. BASE-II

Um die Fallzahlen der Haushalte für die vorliegende Arbeit zu erhöhen, werden auch Daten aus einer dem SOEP verwandten Studie einbezogen. Die *Berliner Altersstudie* (BASE) ist eine interdisziplinäre Studie in den Forschungsbereichen Innere Medizin und Geriatrie, Psychiatrie, Psychologie und Soziologie und Sozialpolitik (Lindenberger et al., 2010). Die Studie wurde zunächst als Querschnittstudie begonnen und über die Jahre als Längsschnittstudie weitergeführt (BASE, 2007). In der Studie wurde unter anderem untersucht, welches Risiko für Krankheit besteht und welche Faktoren für die Gesundheit im Alter verantwortlich sind. Die Risikofaktoren werden in der aktuellen Fortführung der Studie als BASE-II untersucht. Im Rahmen dieser Fortführung wurde für BASE-II eine weitere Stichprobe mit mehr als 2200 Teilnehmern verschiedenen Alters gezogen (BASE II, 2010). Neben der Untersuchungspopulation von Berlinern höheren Alters (1600 Personen zwischen 60-75 Jahren), wird eine Kontrollgruppe mit jüngeren Erwachsenen berücksichtigt (600 Personen zwischen 20-35 Jahren). Dadurch können die Entwicklungsverläufe im Alter anhand der jüngeren Vergleichsgruppe untersucht werden. Durch die Ziehungsmethode der Stichprobe ergibt sich

für das Untersuchungsgebiet Berlin eine Konzentration der Studienteilnehmer auf die räumliche Umgebung der Berliner Innenstadt (vgl. Bertram et al., 2012).

Die Studienteilnehmer wurden im Jahr 2009 so befragt, als ob sie Stichprobenmitglieder des SOEP wären (Bertram et al., 2012). Dadurch stehen die gleichen Informationen wie in den Original-SOEP-Daten zur Verfügung und somit kann für die vorliegende Arbeit für das Jahr 2009 die auswertbare Fallzahl deutlich erhöht werden (vgl. Abbildung 56, S. 131).

## **2.4. Innovationen der Dissertation aus der Synthese des Forschungsstandes**

In diesem Abschnitt werden die Forschungslücken aus dem aktuellen Forschungsstand der urbanen Fernerkundung, der Verknüpfung von Erdbeobachtungsdaten für sozialwissenschaftliche Studien und der räumlichen Analyse der SOEP/BASE-II-Daten formuliert und in einen Bezug zu den Aufgaben der vorliegenden Arbeit gebracht. Aus dieser Betrachtung werden die Innovationen der Arbeit abgeleitet und dargelegt.

### **2.4.1. Identifizierte Forschungslücken aus dem Forschungsstand**

Aus der Betrachtung der Grundlagen zur räumlichen Auswertung sozialwissenschaftlicher Daten geht hervor, dass es zwar viele Forschungsarbeiten gibt, die Informationen aus Fernerkundungsdaten in einen sozialen Bezug stellen, aber nur eine überschaubare Anzahl an qualitativ hochwertigen sozialwissenschaftlichen Arbeiten, die Fernerkundungsdaten für die Erklärung von räumlichen sozialwissenschaftlichen Hypothesen verwenden. In der Forschungsliteratur muß dabei unterschieden werden, ob es sich um fernerkundliche Arbeiten mit einem sozialwissenschaftlichen Bezug handelt oder um sozialwissenschaftliche Arbeiten mit einem Raumbezug bzw. einem expliziten Bezug zur Erdbeobachtung. Vor allem geographische Arbeiten beschränken sich meist auf die Erfassung von biophysikalischen Merkmalen aus Fernerkundungsdaten und die Herstellung eines einfachen, statistischen Zusammenhanges zu ökonomischen Werten. Daneben existieren viele Arbeiten, die Fernerkundungsdaten als Grundlage für die räumliche Disaggregation oder Abschätzung von Bevölkerungszahlen und -verteilungen verwenden. Diese Arbeiten werden in der Forschungsliteratur auch unter dem Thema der sozioökonomischen Anwendungen von Fernerkundungsdaten beschrieben, sie weisen allerdings keine sozialwissenschaftlich relevante Tiefe oder fundierte Fragestellung auf.

Demgegenüber steht die empirische Sozialforschung als Forschungsdisziplin der Soziologie. In diesem Forschungsfeld ist in den letzten Jahren ein Trend zu raumbezogenen

Arbeiten zu beobachten. Allerdings ist der räumliche Bezug der Daten sehr häufig auf eine traditionelle, administrative oder statistische, Raumeinheit reduziert, weshalb die Einbindung von flächendeckender Erdbeobachtungsinformation nicht möglich ist. Sozialwissenschaftliche Studien, welche auf Erdbeobachtungsdaten als räumliche Datengrundlage zurückgreifen, beziehen sich häufig auf Landbedeckungs-/nutzungsveränderungen und ihre Auswirkungen auf die örtliche Gesellschaft. Wissenschaftliche Arbeiten, welche physische Merkmale aus Erdbeobachtungsdaten für die kleinräumige Analyse in Städten verwenden, sind forschungstechnisches Neuland.

Dem Schutz der persönlichen Daten von Umfrageteilnehmern wird in der sozialwissenschaftlichen Forschung ein sehr hoher Stellenwert beigemessen, da er ein wesentliches Element für das Vertrauen der Umfrageteilnehmer darstellt. Der Datenschutz stellt im Zusammenhang mit einer räumlichen Auswertung ein bedeutendes Hindernis dar, da eine Deanonymisierung der Umfrageteilnehmer ausgeschlossen werden muss. Viele sozialwissenschaftliche Daten werden vor diesem Hintergrund nicht in einen räumlichen Bezug gebracht.

In der SOEP-spezifischen Forschung finden sich sehr wenige Arbeiten mit einem expliziten räumlichen Bezug. Die meisten Arbeiten beziehen sich auf die in den SOEP-Daten abgelegten räumlichen Informationen zu den Bundesländern, Landkreisen oder Postleitzahlenbezirken. Erst drei Arbeiten haben bisher die Verknüpfung von Fernerkundungsdaten mit SOEP-Daten exemplarisch dargestellt, es wurden jedoch keine tiefergehenden Analysen von kleinräumigen Kontextmerkmalen über die physische Ausstattung der Nachbarschaft und der lokalen Bebauungsstruktur in städtischen Räumen mit einer ausreichenden Fallzahl durchgeführt.

Die Arbeiten in der aktuellen Forschungsliteratur zur Ableitung von urbanen Strukturen aus Fernerkundungsdaten sind methodisch sehr breit gestreut. Viele Arbeiten, die sich mit der Ableitung der Bebauungsstruktur aus digitalen Oberflächenmodellen beschäftigen, greifen auf LiDAR-Punktwolken zurück. Es sind allerdings nur wenige robuste Verfahren bekannt, welche die Gebäudestruktur flächendeckend für ganze Städte direkt aus digitalen Oberflächenmodellen unterschiedlicher Herstellungsverfahren ableiten können. Aktuelle Arbeiten beschäftigen sich daher weniger mit der Gebäuderekonstruktion als mit der direkten Extraktion von Gebäudestrukturen. In diesem Zusammenhang kann hier, genau wie in der Literatur zur urbanen Landbedeckungsklassifikation, ein Trend zu objektbasierten Klassifikationsmethoden festgestellt werden.

Die semantische Klassifikation von Bebauungsstrukturen wird in der Forschung ebenfalls durch ein sehr breites methodisches Spektrum behandelt. Die Arbeiten in diesem Zusammenhang beziehen sich meist auf eine räumlich übergeordnete Strukturebene, und nur in einigen Fällen auf die Klassifikation von Einzelgebäuden. Die Ausgangsdatengrundlage ist in diesen Arbeiten sehr unterschiedlich, eine Gemeinsamkeit stellen die abgeleiteten Gebäudegrundflächen aus den Ausgangsdaten dar. Für die semantische Klassifikation von

Gebäudetypen werden in den bekannten Arbeiten sehr viele verschiedene Merkmale verwendet. Es werden neben Formmerkmalen wie die Größe auch Texturmerkmale, Nachbarschaftsmerkmale, Distanzmaße und zusätzliche Nutzungsinformationen aus weiteren Datenquellen verwendet. Es ist keine Systematik oder systematische Evaluierung der Relevanz von den unterschiedlichen Merkmalen bekannt.

### **2.4.2. Innovation der vorliegenden Arbeit**

Die Innovation der vorliegenden Arbeit liegt in der erstmaligen Verknüpfung von räumlichen Informationen aus Erdbeobachtungsdaten mit punktuellen sozialwissenschaftlichen Daten am Beispiel von SOEP/BASE-II, da abgesehen von wenigen Vorstudien, bislang noch keine Verknüpfung mit kleinräumigen, stadtstrukturellen Informationen aus geometrisch hoch aufgelösten Erdbeobachtungsdaten durchgeführt wurde. Für die Umsetzung dieser übergeordneten Innovation wurden weitere Teilinnovationen erarbeitet, wobei am Beginn die Erstellung eines methodischen Konzeptes zur Verknüpfung von flächenhaften Fernerkundungsdaten mit punktuellen, sozialwissenschaftlichen Daten steht. Der Raumbezug der SOEP/BASE-II-Daten ist bisher nur sehr vage definiert und nicht eindeutig abgrenzbar, weshalb die zusätzlichen Variablen aus den räumlichen Kontextmerkmalen die Möglichkeit bieten, die Nachbarschaft der Umfrageteilnehmer objektiv und vor allem flächendeckend zu beschreiben. Somit steht für die sozialwissenschaftliche Auswertung eine exakte räumliche Bezugsebene zur Verfügung, in welcher der räumliche Bezug für die zu erforschende Fragestellung angepasst werden kann. Die Überprüfung von sozialwissenschaftlichen Theorien ist nicht mehr Bestandteil der vorliegenden Arbeit, es werden aber Gemeinsamkeiten und Unterschiede zwischen den räumlichen SOEP/BASE-II-Merkmalen und den räumlichen Kontextmerkmalen im Rahmen von statistischen Modellschätzungen dargestellt und diskutiert.

Zwischen der Erarbeitung des methodischen Konzeptes und der sozialwissenschaftlichen Modellschätzung wurden weitere Teilinnovationen erarbeitet um den Ansprüchen einer räumlichen Übertragbarkeit und der Anwendbarkeit der Bildauswerteverfahren auf unterschiedlichen Datensätzen Rechnung zu tragen. In diesem Zusammenhang wurden *State of the Art* Verfahren der digitalen Bildauswertung modifiziert und getestet. Konkret handelt es sich dabei um die Ableitung von Gebäudestrukturen aus digitalen Oberflächenmodellen, die Erstellung eines Verfahrens für die Abschätzung der Geschosshöhe aus digitalen Oberflächenmodellen und um die Überprüfung der Relevanz von Formmerkmalen für die semantische Klassifikation von Gebäudeklassen.

## 2.5. Übersicht der Arbeitsschritte

Eine Übersicht über die Arbeitsschritte ist in Abbildung 11 dargestellt. Sie zeigt die Gliederung der Arbeit in die drei praktischen Bereiche der Arbeit, welche in den entsprechenden Kapiteln ausführlich beschrieben werden. Die Auswertung der Fernerkundungsdaten mit dem Ergebnis der Landbedeckungsklassifikation inklusive der Ableitung der Einzelgebäude wird in Kapitel 4 beschrieben. In Kapitel 5 wird die Klassifikation der semantischen Gebäudeklassen dargestellt und die Aggregation der Informationen auf der räumlichen Ebene des Baublockes. Das sechste Kapitel stellt die Synthese der Verknüpfung der räumlichen Kontextmerkmale und den Survey-Daten anhand von Beispielen dar und beleuchtet Potenziale und Limitierungen.

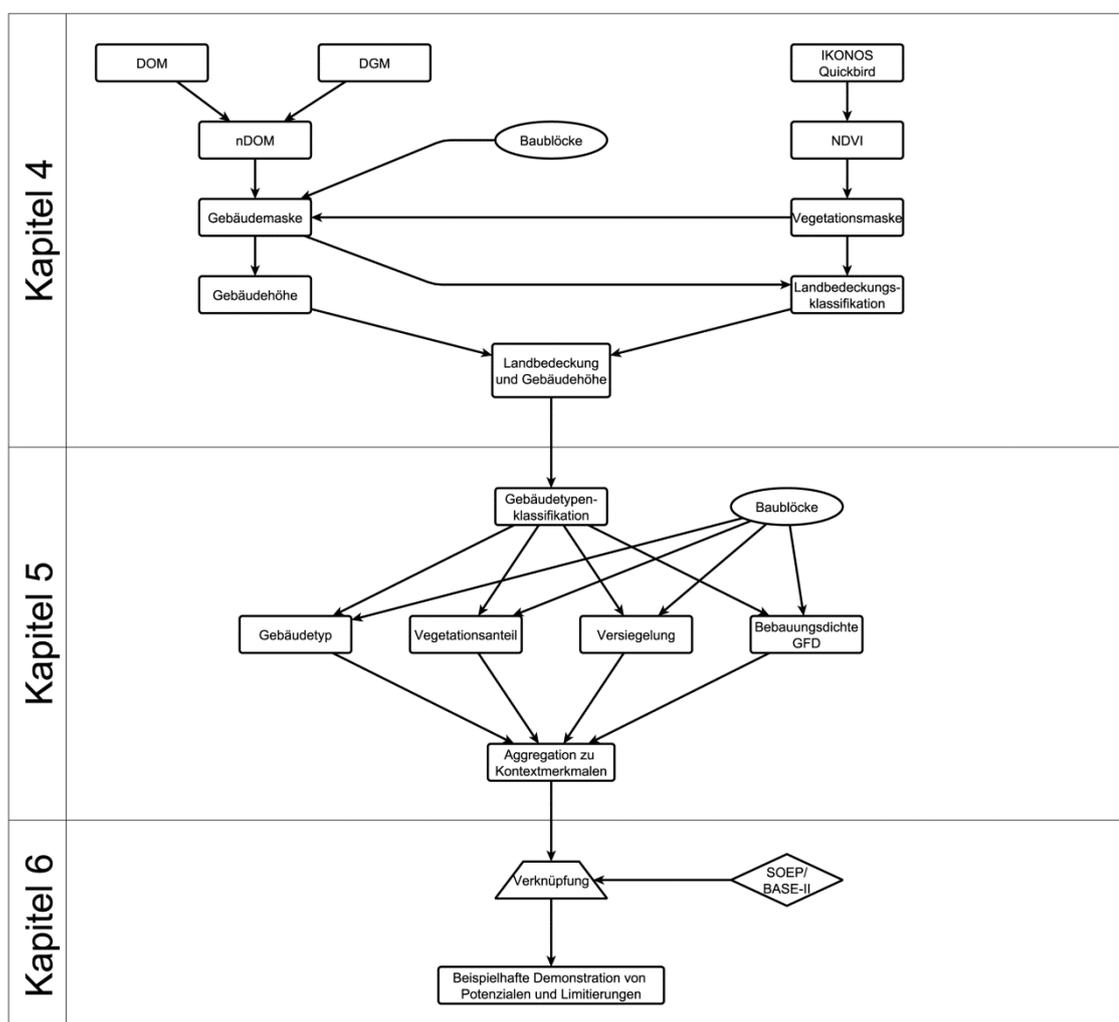


Abbildung 11: Graphische Übersicht über die Arbeitsschritte.

### **2.6. Zusammenfassung**

In diesem Kapitel wurden die Grundlagen der Fernerkundung und die Grundlagen der sozialwissenschaftlichen Paneldaten SOEP und BASE-II vorgestellt, da diese beiden Themengebiete die Grundlagen der vorliegenden Arbeit darstellen. Der zweite Abschnitt reflektiert den aktuellen Stand der Forschung in der urbanen Fernerkundung und ihren für die Arbeit relevanten Teilgebieten der objektbasierten Bildklassifikation, der Ableitung von Bebauungsstrukturen aus digitalen Oberflächenmodellen, der Klassifikation von Bebauungsklassen und der Verwendung von Fernerkundungsdaten für die sozialwissenschaftliche Forschung.

Der dritte Abschnitt geht detailliert auf die Struktur und die Eigenschaften der Längsschnittdaten ein und beschreibt aktuelle raumbezogene Forschungsarbeiten mit diesen Datensätzen. Im letzten Abschnitt werden alle in diesem Kapitel dargestellten Informationen noch einmal aufgegriffen und in den Bezug der vorliegenden Arbeit gestellt und daraus die Ziele und die Innovationen der Arbeit abgeleitet.

Vor dem Hintergrund der Schlüsse beschreibt die vorliegende Arbeit die objektbasierte Auswertung von hochaufgelösten, optischen Satellitenbilddaten und digitalen Oberflächenmodellen für die Erstellung einer Landbedeckungsklassifikation und einer Klassifikation der Gebäudestrukturen zur Ableitung flächendeckender, räumlicher Kontextmerkmale. Sie werden über ein striktes Datenschutzkonzept erstmalig mit Längsschnittdaten der sozialwissenschaftlichen SOEP und BASE-II-Studien verknüpft. Die Teilgebiete der Arbeit spiegeln sich im Aufbau der Arbeit wider.

# Kapitel 3

## Konzeptioneller Rahmen und Geodaten

*“The metropolis is, it seems, a great sifting and sorting mechanism, which [...] infallibly selects out of the population as a whole the individuals best suited to live in a particular region or a particular milieu.”*

Robert Ezra Park (1952)

In diesem Kapitel werden die Rahmenbedingungen für die interdisziplinäre Verknüpfung der Fernerkundungsdaten und der SOEP/BASE-II-Daten erläutert. Dazu wird das räumliche Konzept beschrieben, zu welchem neben dem Datenschutzkonzept auch die technische Verknüpfung der Informationen und die räumliche Bezugsebene zählt. Im zweiten Abschnitt des Kapitels werden die Untersuchungsgebiete und im dritten Abschnitt die Geodatengrundlagen vorgestellt. Ein weiterer Abschnitt widmet sich dem Validierungskonzept für die Klassifikation der Fernerkundungsdaten.

### 3.1. Konzeptioneller Rahmen

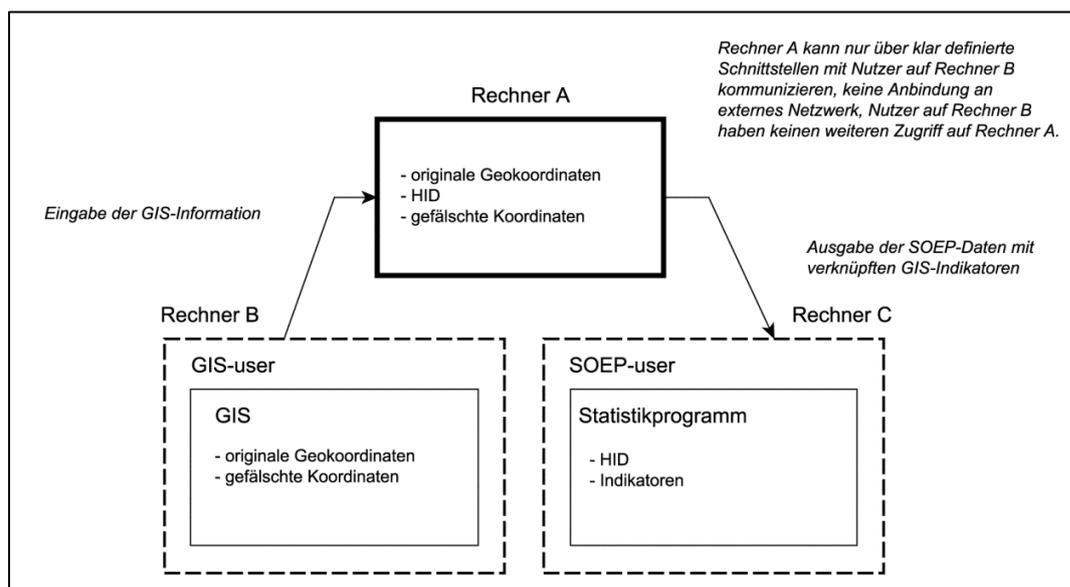
In diesem Abschnitt wird die technische Fusionierung von Fernerkundungsdaten und der SOEP/BASE-II-Studien beschrieben. Dabei wird zunächst auf das Datenschutzkonzept eingegangen und die technische Verknüpfung von Geodaten und SOEP/BASE-II-Daten und danach die räumliche Bezugsebene und die Verknüpfung der beiden Daten beschrieben.

#### 3.1.1. Technische Verknüpfung von Geodaten und SOEP/BASE-II-Daten

Wie bereits in der Einleitung erwähnt, wurden die Wohnadressen der teilnehmenden Haushalte für die Verknüpfung mit Fernerkundungsdaten im Rahmen von *SOEP<sub>geo</sub>* geographisch verortet. Diese Geokoordinate wurde für sämtliche SOEP-Wohnadressen für die Jahre 2000-2010 erzeugt und in einer eigenen Datei, getrennt von den anderen SOEP-Informationen aus der Haushaltsbefragung, abgelegt. Dadurch wird eine Deanonymisierung

der Haushaltsbewohner faktisch unmöglich gemacht, aber trotzdem eine Möglichkeit geschaffen, räumliche Informationen in die SOEP-Daten zu integrieren. In diesem Abschnitt werden die Daten vereinfacht als SOEP-Daten bezeichnet, obwohl das Datenschutzkonzept gleichermaßen für die BASE-II-Daten angewendet wird.

Die einzige Verbindung zwischen der geographischen Koordinate des Haushaltes und den SOEP-Daten besteht über eine eindeutige Haushaltsidentifikationsnummer (HID), über welche in einer gesicherten Netzwerkumgebung die räumlichen Nachbarschaftsinformationen aus den Fernerkundungsdaten automatisiert verknüpft werden können. Des Weiteren wird auf dem Server jeder Zugriff protokolliert. Das technische Konzept von *SOEPgeo* sieht eine eigene Rechnerinfrastruktur vor, in welcher in drei voneinander getrennten Systemen die Verarbeitung der geokodierten SOEP-Daten gemeinsam mit räumlichen Daten erfolgt (Goebel, 2009). Eine Übersicht über das Konzept von *SOEPgeo* ist in Abbildung 12 dargestellt. Sie zeigt, dass räumliche Informationen aus z.B. Fernerkundungsdaten auf einem Rechner B aufbereitet werden können. Über ein gesichertes Netzwerk werden die Informationen an den Rechner A übermittelt. Auf diesem Rechner sind sowohl die originalen Koordinaten der Haushalte als auch die Haushaltsidentifikationsnummern der SOEP-Teilnehmer abgelegt und es wird die Verknüpfung der räumlichen Daten mit den SOEP-Daten durchgeführt. Von diesem Rechner A werden die SOEP-Daten mit den verknüpften räumlichen Informationen an den Rechner C übermittelt. Der Nutzer auf Rechner C hat demnach ausschließlich Zugriff auf die SOEP-Daten mit den verknüpften kleinräumigen Nachbarschaftsmerkmalen. Ein Zugriff auf den Rechner A ist durch die technische Infrastruktur nicht möglich, weshalb zu *keinem* Zeitpunkt die geographischen Koordinaten und die SOEP-Informationen gleichzeitig für einen Nutzer verfügbar sind.



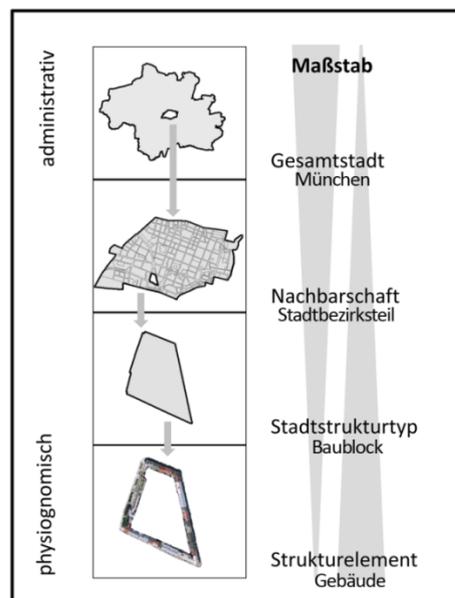
**Abbildung 12:** Schema der Verknüpfung von Geodaten und SOEP/BASE-II-Daten.

(Verändert nach: Goebel, 2009)

### 3.1.2. Die räumliche Bezugsebene für die Datenverknüpfung

Das *Wohnquartier* stellt in der Sozialforschung den Handlungsraum und die Sozialisierungsinstanz zur räumlichen Interpretation von Umfragedaten dar (Hoffmeyer-Zlotnik, 2000). Allerdings hat das Wohnquartier aus sozialwissenschaftlicher Sicht keine feste räumliche Bezugsgröße. Diese kann je nach Fragestellung und Datengrundlage variieren.

In der Stadtökologie ist eine ähnliche räumliche Problematik bekannt. Dort hat sich seit Jahrzehnten das Stadtbiotypen- oder Stadtstrukturtypenkonzept etabliert (Schulte, Sukopp und Werner, 1993). Dieses Konzept beschreibt auf der kleinräumigen Ebene des Baublocks die vorherrschende Nutzung, Landbedeckung und bauliche Struktur. Entsprechend dieses Konzeptes ist es daher möglich die einzelnen stadträumlichen Elemente, wie Einzelgebäude oder Bäume zu räumlichen Kontextmerkmalen zu aggregieren. Der Baublock stellt demnach ein sinnvolles Bindeglied zwischen den einzelnen stadträumlichen Elementen und übergeordneten Raumeinheiten dar, da die städtische Struktur auf einem räumlichen Mehrebenensystem basiert, welches sich aus den einzelnen Teilen der Mikro-, der Meso- und der Makroebene der Gesamtstadt zusammensetzt (Lichtenberger, 1998). Die Einbettung des Strukturtyps in die Hierarchie von räumlichen Strukturen im urbanen Raum ist in Abbildung 13 dargestellt. Sie zeigt auf der untersten Ebene das einzelne Strukturelement, welches beispielsweise durch ein einzelnes Gebäude repräsentiert wird. Die übergeordnete Ebene ist der Stadtstrukturtyp auf der räumlichen Ebene des Baublocks, welcher die aggregierten Informationen der darunter liegenden Ebene vorhält. Darüber hinaus können weitere Raumeinheiten aus der Zusammenfassung von mehreren Baublocken gebildet werden und das Bindeglied zu administrativen Raumeinheiten darstellen.

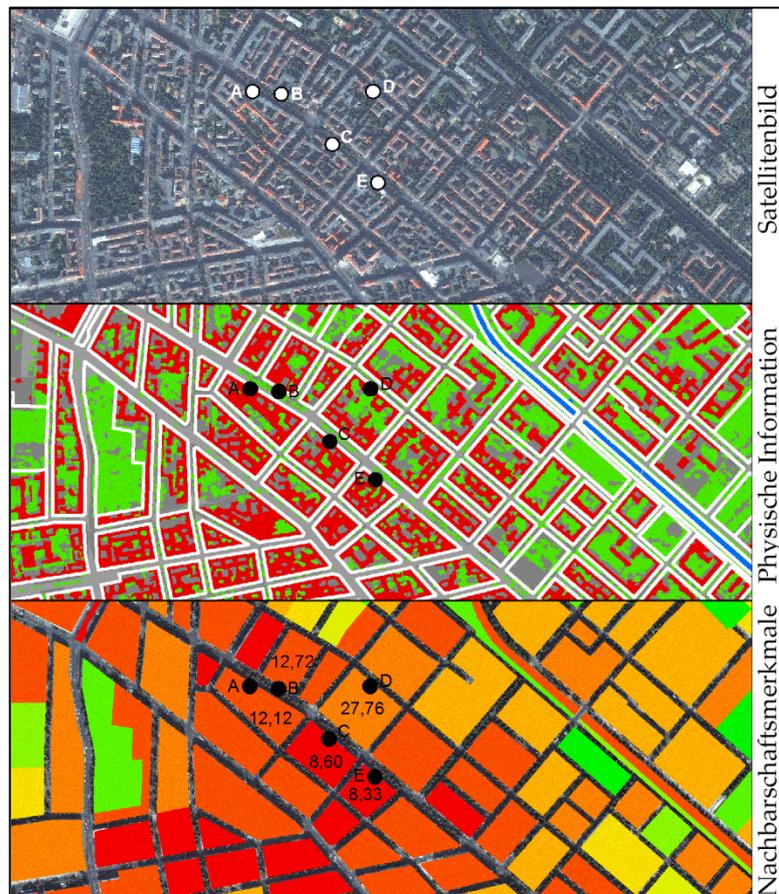


**Abbildung 13:** Maßstabsabhängige Strukturen im urbanen Raum.

(Verändert nach: Wickop et al., 1998)

Auf dieser räumlichen Ebene des Baublocks können entsprechend des oben beschriebenen Konzeptes die physischen Wohnumfeldmerkmale eines Haushaltes zusammengefasst werden. Der Baublock stellt demnach eine natürliche räumliche Untergliederung der Stadt nach einem funktionalen und morphologischen Raummuster dar und ist unabhängig von künstlichen, administrativen Raumeinheiten. Er ist zudem ausreichend kleinräumig, um das direkte Umfeld der Bewohner beschreiben zu können und ist gleichzeitig groß genug, um die Betrachtung über das Einzelhaus hinaus zu ermöglichen. Des Weiteren können auf dieser räumlichen Ebene etwaige Unsicherheiten in der geographischen Verortung der Haushalte kompensiert werden.

Der Baublock stellt demnach in dieser Arbeit die räumliche Bezugsebene für die Datenverknüpfung dar. Auf dieser Ebene werden die Informationen aus den Erdbeobachtungsdaten aggregiert und mit den Survey-Haushalten verknüpft. Ein praktisches Beispiel, wie die räumliche Bildinformation aus Erdbeobachtungsdaten in flächendeckende raumstrukturierende Informationen umgewandelt wird und im Rahmen des Konzeptes in Nachbarschaftsmerkmale zusammengefasst wird, ist in Abbildung 14 dargestellt. Sie zeigt im oberen Teil der Abbildung ein Satellitenbild mit überlagerten, simulierten, geokodierten Haushaltsadressen von Survey-Teilnehmern (Haushaltsidentifikationen A-E). In der Mitte der Abbildung ist die thematische Information aus der Klassifikation des Satellitenbildes mit den überlagerten Baublöcken zu sehen. Der untere Teil zeigt die aggregierte Information der Klassifikation auf der räumlichen Ebene der Baublöcke und beispielhaft die prozentualen Werte des Vegetationsanteils. Für die Auswertung auf Rechner C (vgl. Abbildung 12 auf S. 44) ist dann die verknüpfte Information (z.B. Vegetationsanteil) des Baublocks für jeden Haushalt verfügbar (vgl. Tabelle 22, S. 134).



**Abbildung 14:** Das räumliche Konzept zur Integration von Nachbarschaftsmerkmalen in die Surveydaten. Die Punkte zur geographischen Verortung der Survey-Haushalte dienen lediglich zur Darstellung des Konzeptes und basieren nicht auf tatsächlichen Haushaltsadressen oder Survey-Teilnehmer. Die Lage der Punkte ist zufällig gewählt und eine Deanonimisierung von realen Personen kann auf Basis dieser Punkte nicht durchgeführt werden. Eine räumliche Übereinstimmung mit tatsächlichen Haushaltsadressen ist rein zufällig. (© Baublockgrenzen: Senatsverwaltung für Stadtentwicklung der Stadt Berlin)

## 3.2. Untersuchungsgebiete

Wie bereits dargestellt, wird die Verknüpfung der Fernerkundungsdaten und der SOEP/BASE-II-Daten exemplarisch an zwei Untersuchungsgebieten durchgeführt. Die SOEP-Studie ist zwar eine deutschlandweite Studie, aber die technische Verknüpfung von kleinräumigen Kontextmerkmalen kann am Beispiel von Testgebieten demonstriert werden. Die beiden Testgebiete Berlin und München wurden aus folgenden Gründen ausgewählt: Berlin stellt als bevölkerungsreichste Stadt Deutschlands auch proportional die höchste Anzahl an SOEP-Teilnehmern. Darüber hinaus liegen für Berlin noch zusätzliche Surveydaten aus der BASE-II-Studie vor, welche mit demselben Fragebogen wie die SOEP-Daten erhoben wurden. Dadurch kann die auswertbare Anzahl an Fällen deutlich erhöht werden. Das zweite Testgebiet stellt die Stadt München dar, welche als drittgrößte Stadt Deutschlands ebenfalls eine proportional hohe

Anzahl an Survey-Teilnehmern enthält. Sie steht auch repräsentativ als Beispiel einer westdeutschen Stadt gegenüber Berlin mit seiner besonderen Historie. Die beiden Untersuchungsgebiete werden im Folgenden kurz charakterisiert.

*Berlin* ist sowohl die Bundeshauptstadt als auch einer der drei Stadtstaaten Deutschlands (neben Hamburg und Bremen). Sie beherbergt rund 3,4 Millionen Einwohner (Statistisches Bundesamt, 2009) auf einer Fläche von rund 891 km<sup>2</sup> (Statistisches Bundesamt, 2010) und stellt somit sowohl flächenmäßig als auch nach Einwohnern die größte Stadt Deutschlands dar. Die Stadt gliedert sich in 12 administrative Stadtbezirke, welche im Zuge der Gebietsreform nach der deutschen Wiedervereinigung von ursprünglich 23 Stadtbezirken zusammengefasst wurden. Berlin liegt an den Flüssen Spree und Havel in einem glazial überprägten Urstromtal. Ausgehend vom historischen Zentrum Berlins an der Spreeinsel wuchs die Stadt unter dem Einfluss der Hohenzollern seit dem Mittelalter, aber erst ab dem Beginn des 18. Jahrhunderts zu einer Großstadt heran. Die Mietskasernen prägen das Stadtbild und entstanden im Zuge des einsetzenden Bevölkerungswachstums zur Zeit der Industrialisierung. Besonders charakteristisch für Berlin ist die Ausprägung eines zweiten Stadtzentrums im Bereich des heutigen Kurfürstendamm/Zoologischer Garten im Zeitraum der Teilung Berlins von 1961-1990. Durch die Teilung Berlins und die großräumigen Zerstörungen der Bausubstanz während des zweiten Weltkrieges, sind neben den aktuellen Baumaßnahmen auf den durch die Wiedervereinigung freigewordenen Flächen auch durch die Phase des Sozialismus beeinflusste Baustrukturen prägend für das Stadtbild.

*München* ist die Landeshauptstadt des Freistaates Bayern, dem einwohnerreichsten Bundesland Deutschlands. Die Stadt ist heute mit seinen rund 1,3 Millionen Einwohnern (Statistisches Bundesamt, 2009) die drittgrößte Stadt Deutschlands (nach Berlin und Hamburg) und ist in 25 administrative Stadtbezirke gegliedert. Sie ist eine kreisfreie Stadt mit einer flächenhaften Ausdehnung von rund 310 km<sup>2</sup> (Statistisches Bundesamt, 2010) und hat weitreichenden wirtschaftlichen und kulturellen Einfluss. München liegt im nördlichen Alpenvorland auf tertiären Schotterebenen und wird von der Isar durchquert. In seiner über 850-jährigen Geschichte spielte die Stadt immer wieder eine wichtige Rolle – nicht zuletzt als Residenzstadt der Wittelsbacher. Die Stadt wuchs über die Jahrhunderte zur Großstadt von heute und wird auch in naher Zukunft eine der wenigen deutschen Stadtregionen sein (neben Hamburg, Köln, Rhein-Main und Stuttgart), für welche ein Bevölkerungswachstum prognostiziert wird (Reiß-Schmidt, 2010). Während des zweiten Weltkrieges wurden große Flächen der innerstädtischen historischen Baustruktur zerstört. Sie wurden aber in der Phase des Wiederaufbaus weitestgehend wiederhergestellt. München beherbergt neben seinen ausgedehnten urbanen Strukturen aber auch viele großflächige städtische Grünflächen, worunter der *Englische Garten* den größten unter ihnen ausmacht und nördlich des Stadtzentrums entlang der Isar als innerstädtisches Erholungsgebiet dient. Die Lage der Untersuchungsgebiete in Deutschland ist in Abbildung 15 dargestellt.



Abbildung 15: Lage der Untersuchungsgebiete in Deutschland.

### 3.3. Geodatengrundlagen

In der Einleitung und in Kapitel 2 wurde bereits auf Anforderungen der für die vorliegende Arbeit verwendeten Datengrundlagen verwiesen. Rekapitulierend ist zunächst eine hohe geometrische Auflösung notwendig, um die kleinräumigen Details der einzelnen stadträumlichen Elemente zu erfassen. Für die Ableitung der Landbedeckung sind spektrale Daten erforderlich und für die Ableitung der Gebäude sind Daten notwendig, welche die physische Bebauungsstruktur beinhalten. Diese Informationen sind in digitalen Oberflächenmodellen gegeben. Auf die Eigenschaften der verwendeten Fernerkundungsdatengrundlagen und weitere Geodaten wird im aktuellen Abschnitt eingegangen.

#### 3.3.1. Fernerkundungsdaten

Für die beiden Untersuchungsgebiete werden sowohl geometrisch höchst aufgelöste optische Satellitenbilddaten mit einer Auflösung von einem Meter als auch digitale Oberflächenmodelle mit derselben geometrischen Auflösung verwendet. Zu den geometrisch höchst auflösenden optischen Erdbeobachtungssensoren werden jene gezählt, in welchen der Abstand zwischen zwei Bildelementen in der Aufnahme einem Abstand in der Natur von weniger oder gleich einem Meter entspricht.

### 3.3.1.1. Satellitenbilddaten

Ikonos war der erste kommerzielle Satellit, welcher im Jahr 1999 erfolgreich in die Erdumlaufbahn gebracht wurde und seither multispektrale Aufnahmen in einer Bodenauflösung von 4 m und panchromatische Aufnahmen in einer Bodenauflösung von 1 m erzeugt. Für die gleichzeitige Nutzung der vollen geometrischen und spektralen Auflösung von 4 spektralen Kanälen müssen die Daten fusioniert werden. Der Sensor kann auf seiner sonnensynchronen Erdumlaufbahn aus einer Flughöhe von 681 km Szenen mit einer Breite von etwa 11 km im Nadirbereich und einer Länge von bis zu 1000 km aufzeichnen (GeoEye, 2009). Die Daten sind in einer radiometrischen Auflösung von 11 bit abgelegt.

Neben weiteren geometrisch höchstaflösenden optischen Erdbeobachtungssatelliten zeichnet der Sensor Quickbird seit dem Jahr 2001 detaillierte Aufnahmen mit ähnlichen Spezifikationen wie Ikonos auf. Der Sensor besteht ebenfalls aus 4 multispektralen Aufnahmekanälen und einem panchromatischen Kanal bei einer radiometrischen Auflösung von ebenfalls 11 bit (Digital Globe, 2011a). Der Sensor unterscheidet sich von Ikonos zunächst durch die höhere geometrische Auflösung im panchromatischen Kanal von 0,65 m bzw. 2,62 m in den multispektralen Kanälen. Ferner beträgt die Bodenabdeckung mit einer Aufnahme durch eine Anhebung der Flugbahn von 450 auf 482 km mittlerweile 18 km (Digital Globe, 2011b).

Einige technische Eigenschaften der beiden Satelliten sind in Tabelle 4 dargestellt. Es ist ersichtlich, dass beide Sensoren die vier Spektralkanäle *blau*, *grün*, *rot* und *nahes Infrarot* abdecken. An den Wellenlängenbereichen der einzelnen Spektralkanäle ist allerdings zu erkennen, dass sie zwischen den Sensoren voneinander abweichen. So weist der Sensor von Quickbird für jeden Spektralkanal ein breiteres Spektrum als Ikonos auf. Die weiteren wesentlichen Unterschiede zwischen den Sensoren stellen vor allem die höhere geometrische Auflösung von Quickbird wie auch die größere räumliche Abdeckung und die Unterschiede in der Aufnahmegeometrie dar. Die Unterschiede können im Rahmen der Anwendung der vorliegenden Arbeit allerdings vernachlässigt werden, da beide Daten auf die gleiche geometrische Auflösung von 1 m skaliert werden und die spektrale Information ausschließlich in Form von erstellten Indizes ausgewertet wird (vgl. Abschnitt 4.3).

**Tabelle 4:** Eigenschaften von Ikonos und Quickbird.

	Ikonos	Quickbird
<b>Spektrale Auflösung (nm)</b>		
Blau	445 – 516	430 – 545
Grün	506 – 595	466 – 620
Rot	632 – 698	590 – 710
Nahes Infrarot	757 – 853	715 – 918
Panchromatisch	526 – 929	405 – 1053
<b>Bodenauflösung (m)</b>		
Multispektral	4	2,62
Panchromatisch	1	0,65
<b>Radiometrische Auflösung (bit)</b>		
	11	11
<b>Flughöhe (km)</b>		
	681	450/482
<b>Bodenabdeckung (km)</b>		
	11x11	16,5x16,5 / 18x18

Für das Untersuchungsgebiet *München* wurde eine Ikonos-Szene am 17.09.2003 aufgezeichnet. Die Aufnahme deckt das Zentrum Münchens weitgehend ab (vgl. schwarzer Rahmen in Abbildung 16 oben). In der Aufnahme sind lokale Beeinträchtigungen durch Dunstschleier gegeben (vgl. Abbildung 21, S. 66). Die Daten waren in ihrem Auslieferungszustand bereits radiometrisch korrigiert und fusioniert, weshalb für alle vier spektralen Kanäle die volle geometrische Auflösung von 1 m verfügbar war.

Für das flächenmäßig größere Untersuchungsgebiet *Berlin* wurde eine Quickbird-Szene am 17.09.2007 aufgezeichnet. Die Szene deckt ebenfalls das Zentrum der Stadt ab und wurde bei sehr günstigen atmosphärischen Bedingungen aufgenommen. Die Daten wurden als fusioniertes Produkt mit einer geometrischen Auflösung von 0,6 m in allen vier Spektralkanälen ausgeliefert. Eine visuelle Darstellung der beiden Satellitenbildaufnahmen findet sich in Anhang A.

### 3.3.1.2. Digitale Oberflächenmodelle

Für die Herstellung von digitalen Oberflächenmodellen stehen einige Verfahren zur Verfügung, es haben sich in der Fernerkundung allerdings zwei Verfahren für die operationelle Anwendung etabliert. Ein Verfahren stellt das aktive LiDAR-Verfahren dar, welches die Geländeoberfläche mit Lichtpunkten abtastet und über die Laufzeitmessung der Lichtpunkte die Distanz und entsprechend die Höhe der Oberfläche bestimmen kann. Im zweiten gängigen Verfahren, der Photogrammetrie, werden durch die unterschiedlichen Betrachtungswinkel in Stereoaufnahmen die digitalen Oberflächenmodelle hergestellt. Die Datengrundlage für die

Stereoaufnahmen kann dabei entweder aus mehreren Satellitenaufnahmen oder aus einer Reihe an Luftbildaufnahmen stammen. Für die beiden Untersuchungsgebiete Berlin und München wurden digitale Oberflächenmodelle aus flugzeuggetragenen Sensoren photogrammetrisch erstellt. Die Oberflächenmodelle für Berlin und München wurden auf eine geometrische Auflösung von 1 m skaliert, damit sie dieselbe Auflösung wie die Satellitenbilddaten aufweisen. Die beiden Luftbildkameras, welche für die Erstellung der digitalen Oberflächenmodelle für Berlin und München verwendet wurden, werden an dieser Stelle kurz charakterisiert.

Die *high resolution stereo camera – airborne extended* (HRSC-AX) wurde ursprünglich am Deutschen Zentrum für Luft- und Raumfahrt (DLR) für die russische Mars96 Mission entwickelt und in weiterer Folge für flugzeuggetragene Anwendungen adaptiert (Neukum et al., 1995). Sie besteht aus einem multiplen Zeilenscanner mit einer Kammerkonstante von 151 mm und einer Sichtweite (*field of view*, FOV) von 28,9° (Oczipka, 2007). Die Daten werden in vier Spektralkanälen zwischen 450 und 814 nm aufgezeichnet. Im Kamerasystem ist ein Globales Positionierungssystem und Inertiales Navigationssystem (GPS/INS) angebracht und die Bilder werden mit neun *charge-coupled devices* (CCD) aufgenommen. Die Datenprozessierung geschieht weitgehend automatisiert und beinhaltet die Vorverarbeitung, die radiometrische Anpassung und die Generierung von digitalen Oberflächenmodellen und Orthophotos (Scholten et al., 2003).

Die zweite Kamera, die UltraCamX, wurde im Jahre 2006 als Nachfolger der UltraCamD von Vexcel Imaging© vorgestellt. Sie wurde als digitale Großformatkamera mit einer Kammerkonstante von 100 mm entwickelt und zeichnet bis zu 4700 Einzelbilder auf 13 CCD Einheiten mit jeweils einer Bildauflösung von 16 Megapixeln und einer radiometrischen Auflösung von 12 bit auf (Gruber, 2007). In der Regel werden für die Erstellung von digitalen Oberflächenmodellen in städtischen Untersuchungsgebieten geometrische Auflösungen zwischen 7 und 20 cm verwendet (Brüßhaber, Trosset und Bucher, 2010). Für die Herstellung des DOMs aus den Originaldaten wird ein *semi global matching* Verfahren angewendet (Hirschmüller, 2006).

Die Befliegung für das Untersuchungsgebiet München fand am 15. September 2004 statt. Das digitale Oberflächenmodell wurde mit einer geometrischen Auflösung von 1 m im horizontalen Bereich und einer vertikalen Höhenauflösung im Dezimeterbereich zur Verfügung gestellt. Die Abdeckung des Oberflächenmodells beträgt 378 km<sup>2</sup> und reicht in allen vier Richtungen über die Abdeckung der Satellitenbildszene hinaus (vgl. Abbildung 16 schwarzer Rahmen, oben), weshalb das Untersuchungsgebiet auf die räumliche Schnittmenge der beiden Datensätze reduziert wurde.

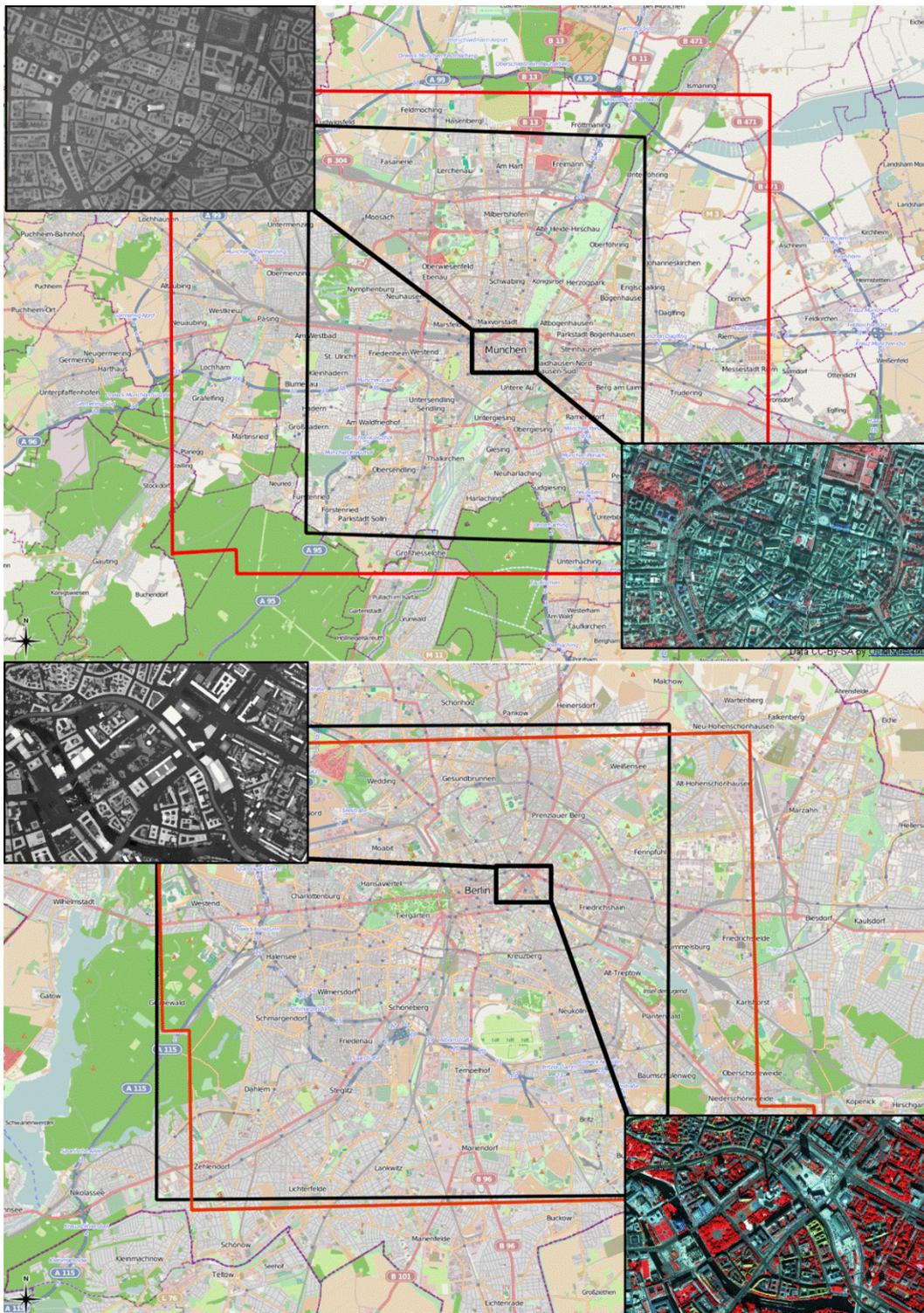
Das digitale Oberflächenmodell für das Untersuchungsgebiet Berlin wurde zunächst in einer geometrischen Auflösung von 15 cm erstellt und für die Anwendungen im Kontext der vorliegenden Arbeit auf die gleiche geometrische Auflösung wie für München (1 m) angepasst.

Die Abdeckung des DOM beträgt rund 312 km<sup>2</sup> und entspricht in etwa jener der QuickBird Aufnahme (vgl. Abbildung 16 schwarzer Rahmen, unten). Das Untersuchungsgebiet wird hier ebenfalls aus der räumlichen Schnittmenge der beiden Datensätze gebildet.

Die Eigenschaften der Satellitenbilddaufnahmen und der verwendeten digitalen Oberflächenmodelle sind in Tabelle 5 zusammengefasst. Abbildungen über die Datensätze finden sich im Anhang A dieser Arbeit.

**Tabelle 5:** Eigenschaften der Fernerkundungsdatengrundlagen für die Untersuchungsgebiete Berlin und München.

	<b>München</b>	<b>Berlin</b>
<b>Satellitenbild</b>	Ikonos	QuickBird
<b>Aufnahmedatum</b>	17.09.2003	17.09.2007
<b>Abdeckung</b>	181 km <sup>2</sup>	267 km <sup>2</sup>
<b>DOM</b>	HRSC-AX	UltraCam-X
<b>Aufnahmedatum</b>	15.09.2004	24.06.2008 02.07.2008
<b>Bodenauflösung</b>	1 m	1 m
<b>Abdeckung</b>	380 km <sup>2</sup>	312 km <sup>2</sup>



**Abbildung 16:** Räumliche Abdeckung der Fernerkundungsdatengrundlagen und ihre beispielhafte Darstellung für die beiden Untersuchungsgebiete München (oben) und Berlin (unten). Der schwarze Rahmen markiert die Abdeckung des Satellitenbildes und der rote Rahmen die Abdeckung des DOMs. (© Openstreetmap, 2011)

### 3.3.2. Weitere Geodaten

Neben den Fernerkundungsdatengrundlagen werden im Rahmen der Arbeit auch weitere Geodaten verwendet, welche in diesem Abschnitt kurz beschrieben werden.

#### 3.3.2.1. Digitale Geländemodelle

Digitale Geländemodelle (DGM) beschreiben im Gegensatz zu digitalen Oberflächenmodellen nur die Geländeformen der Erdoberfläche ohne die darauf befindlichen Objekte wie Vegetation oder Bebauung. Die DGMs wurden vom Bundesamt für Kartographie und Geodäsie (BKG) zur Verfügung gestellt und bestehen aus einer in einem regelmäßigen Gitter angeordneten Punktwolke, wobei die Gitterweite 25 m beträgt. Die geländetypenabhängigen Lage- und Höhengenaugigkeiten der Punkte werden vom BKG mit  $\pm 1-3$  m angegeben (Bundesamt für Kartographie und Geodäsie, 2012a). Die Geländemodelle werden mit unterschiedlichen Methoden wie Laserscanning, Photogrammetrie oder durch Digitalisierung von Höhenlinien aus topographischen Karten im Maßstab von 1:25000 erstellt. Sie werden in der Arbeit für die Normalisierung der digitalen Oberflächenmodelle verwendet (vgl. Abschnitt 4.2.1), weshalb die Lage- und Höhengenaugigkeit dieser Daten einen direkten Einfluss auf die abgeleiteten Gebäudehöhen hat.

#### 3.3.2.2. Baublöcke/Stadtstrukturtypenklassifikation

Wie oben in Abschnitt 3.1.2 beschrieben, wird die räumliche Ebene für die Verknüpfung der Fernerkundungsdaten mit den SOEP/BASE-II-Daten durch Baublöcke repräsentiert. Sie wird für die Aggregation der Landbedeckung und der Bebauungsstruktur und für die Verknüpfung der erstellten räumlichen Kontextmerkmale mit den Surveydaten verwendet.

Die Grundlage für diese Daten bilden *Stadtstrukturtypenkartierungen* von München und Berlin, welche für die vorliegende Arbeit von der Landeshauptstadt München und der Senatsverwaltung für Stadtentwicklung Berlin zur Verfügung gestellt wurden. Diese Kartierungen wurden auf Basis von visuellen Luftbildauswertungen und zusätzlichen Informationen über die Nutzung und den Gebäudebestand hergestellt. Sie stellen aggregierte Informationen über unterschiedliche Bebauungs- und Nutzungstypen auf der Ebene des Baublocks dar.

Die Klassifikationsschlüssel in Stadtstrukturtypenkartierungen in Deutschland können sich sehr stark unterscheiden, da sie an die jeweiligen lokalen Charakteristika angepasst sind. Damit eine harmonisierte Datenbasis im Rahmen der Arbeit verwendet werden kann, werden die Gebäudeklassen für die beiden Untersuchungsgebiete neu klassifiziert. Die bestehenden

Kartierungen werden dabei für das Training und die Validierung der semantischen Gebäudeklassifikation verwendet (vgl. Kapitel 5).

Abbildung 17 zeigt einen Ausschnitt der Stadtstrukturtypenklassifikation für München. Zu sehen sind die Raumabgrenzungen der Baublöcke und die Landbedeckungsklassen *Grünflächen*, *Freiflächen*, *Versiegelt*, *Sondernutzung* und *Gewässer*. Die ursprünglich 45 Klassen wurden zu Darstellungszwecken zusammengefasst.



**Abbildung 17:** Ausschnitt aus der Stadtstrukturtypenklassifikation für München.

(© Landeshauptstadt München)

#### 3.3.2.3. Digitale Orthofotos

Für die geometrische Vorverarbeitung der Satellitenbildszenen (vgl. Abschnitt 4.1) wurden digitale Orthofotos in einer geometrischen Auflösung von 0,2 m vom BKG zur Verfügung gestellt (Bundesamt für Kartographie und Geodäsie, 2012b). Sie dienen der Messung von geometrischen Referenzpunkten anhand welcher die geometrische Korrektur durchgeführt wird.

#### 3.3.2.4. Digitale Gebäudemodelle aus dem Kataster

Für die Validierung der Gebäudeabgrenzung (vgl. Abschnitt 4.2.3) und die empirische Ermittlung der Funktion zur Abschätzung der Geschosshöhen (vgl. Abschnitt 4.2.6) wurden für beide Untersuchungsgebiete amtliche, digitale Gebäudemodelle verwendet, welche von der Landeshauptstadt München und der Senatsverwaltung für Stadtentwicklung Berlin zur Verfügung gestellt wurden. Die Daten beinhalten katastergenaue Gebäudeumgriffe inklusive der Geschosshöhe jedes Gebäudes.

Zusammenfassend sind die verwendeten Geodaten in Tabelle 6 aufgeführt.

**Tabelle 6:** Übersicht über die verwendeten Geodaten.

<b>Datensatz</b>	<b>Eigenschaften</b>	<b>Verwendung</b>
Ikonos Satellitenbild	4 Kanäle, 1 m geom. Auflösung, Aufnahme: 17.09.2003	Landbedeckungsklassifikation
Quickbird Satellitenbild	4 Kanäle, 1 m geom. Auflösung Aufnahme: 17.09.2007	Landbedeckungsklassifikation
DOM München	HRSC, 1 m geom. Auflösung, Aufnahme: 2004	Gebäudeabgrenzung
DOM Berlin	UltraCamX, 1 m geom. Auflösung, Aufnahme: 2004	Gebäudeabgrenzung
Digitale Geländemodelle München und Berlin	25 m geom. Auflösung	Erstellung nDOM
Digitale Orthofotos München und Berlin	0,2 m geom. Auflösung	Geometrische Korrektur der Satellitenbilder
Stadtstrukturtypenkartierung München	19 Bebauungsstrukturtypen	Blockgrenzen, Training und Validierung
Stadtstrukturtypenkartierung Berlin	17 Bebauungsstrukturtypen	Blockgrenzen, Training und Validierung
Digitales Gebäudemodell München	Gebäudegrundflächen und Geschosszahlen, Stand: 2008	Validierung der Gebäudeabgrenzung und zur
Digitales Gebäudemodell Berlin	Gebäudegrundflächen und Geschosszahlen, Stand: 2009	Ermittlung des empirischen Zusammenhanges zwischen Gebäudehöhe und Geschosszahl

### 3.4. Validierung der Klassifikationsergebnisse

In computergestützten Bildklassifikationsverfahren wird ein Bild nach bestimmten, vorab definierten Kriterien oder einem Algorithmus ausgewertet. Viele automatische Verfahren können dadurch sehr hohe Genauigkeiten erzielen, allerdings stimmt in der Regel kein Klassifikationsergebnis mit der Realität vollständig überein. Die Validierung des Klassifikationsergebnisses wird dabei auf Basis von Referenzdaten durchgeführt. Dadurch kann eine Einschätzung der Übereinstimmung des Klassifikationsergebnisses mit den Referenzdaten getroffen werden.

In Kapitel 4 und 5 werden computergestützte Klassifikationen der Landbedeckung und der semantischen Gebäudeklassen durchgeführt. Die Validierung dieser Klassifikationsergebnisse basiert auf den im aktuellen Abschnitt beschriebenen Validierungsmethoden.

Ein etabliertes Verfahren zur Abschätzung der Genauigkeit von Klassifikationsergebnissen ist die Konfusionsmatrix. Sie besteht aus einer Tabelle mit gleich vielen Zeilen und Spalten und repräsentiert die Anzahl der verglichenen Untersuchungselemente (z.B. Pixel oder Objekte),

welche zu einer bestimmten Kategorie (z.B. Landbedeckungsklasse) gehören. Die Übereinstimmungen zwischen der Klassifikation und der Referenz sind in der Diagonale der Konfusionsmatrix aufgetragen und die Werte außerhalb der Diagonale zeigen die Fehlklassifikationen in den jeweiligen Klassen. Durch Aufsummieren der Übereinstimmungen und der Fehlklassifikationen kann sowohl für jede untersuchte Klasse als auch für die Gesamtklassifikation eine Genauigkeitsabschätzung durchgeführt werden (Congalton und Green, 2008).

Tabelle 7 zeigt exemplarisch eine Konfusionsmatrix für fünf verschiedene Landbedeckungsklassen. In den Spalten ist die Referenz angeführt und in den Zeilen die Klassifikationsergebnisse. In den Zellen der grau unterlegten Hauptdiagonalen werden die Übereinstimmungen zwischen der Referenz und der Klassifikation eingetragen. Die Summe der Übereinstimmungen für jede Klasse wird in der letzten Zeile und der letzten Spalte der Tabelle dargestellt.

**Tabelle 7:** Konfusionsmatrix (verändert nach Foody, 2002).

	Referenzdaten				
Klassifikation	Wasser	Vegetation	versiegelt	Gebäude	Summe
Wasser					
Vegetation					
versiegelt					
Gebäude					
Summe					

Um die Übereinstimmung des Klassifikationsergebnisses mit der Referenz zu berechnen, wird die Anzahl der korrekt klassifizierten Pixel durch die Anzahl der untersuchten Pixel dividiert. Die Gesamtgenauigkeit berechnet sich demnach wie folgt:

$$(1) \quad \text{Gesamtgenauigkeit} = \frac{\text{korrekt klassifizierte Pixel}}{\text{Anzahl Pixel in Validierungsgebieten}}$$

Sie gibt allerdings nur den Anteil der korrekt klassifizierten Pixel an der Summe der Pixel in den Validierungsgebieten an. Eine Differenzierung nach Unterlassungsfehlern (*omission error*) und Überlassungsfehlern (*comission error*) gibt für das Ergebnis den Anteil an Pixeln an, die eigentlich zu einer anderen Klasse gehören bzw. den Anteil nicht erkannter Pixel an. Der Unterlassungsfehler wird dabei in der Spalte abgelesen. Alle Werte, die in der Spalte nicht in der entsprechenden Hauptdiagonalen stehen, zeigen wie viele der untersuchten Pixel nicht als richtige Klasse erkannt wurden, sondern in der Klassifikation *unterlassen* wurden. Der Überlassungsfehler für jede Klasse wird in den Zeilen abgelesen. Er gibt an, wie viele der untersuchten Pixel einer anderen Klasse in der Ergebnisklasse vorkommen.

Aus der Konfusionsmatrix können auch noch weitere Maße für die Klassifikation berechnet werden. Die *Herstellergenauigkeit* (HG) stellt das Verhältnis der korrekt klassifizierten Pixel zur Gesamtzahl der Pixel der entsprechenden Referenzklasse dar. Über die Herstellergenauigkeit können Aussagen getroffen werden, wie stark das hergestellte Klassifikationsergebnis mit den Referenzdaten übereinstimmt.

Die *Nutzergenauigkeit* (NG) stellt das Verhältnis der korrekt klassifizierten Pixel zur Gesamtzahl der Ergebnispixel der entsprechenden Klasse dar. Es kann damit die Wahrscheinlichkeit ausgedrückt werden, mit der ein Ergebnispixel den Referenzdaten entspricht (Congalton und Green, 2008):

$$(2) \quad HG = \frac{\text{korrekt klassifizierte Pixel}}{\text{Gesamtzahl der Referenzpixel der Klasse}} = \frac{\text{Element der Hauptdiagonale}}{\text{Spaltensumme}}$$

$$(3) \quad NG = \frac{\text{korrekt klassifizierte Pixel}}{\text{Gesamtzahl der Ergebnispixel der Klasse}} = \frac{\text{Element der Hauptdiagonale}}{\text{Zeilensumme}}$$

Auf Basis der Konfusionsmatrix können auch weitere Gütemaße berechnet werden. Die Kappa-Analyse und der daraus resultierende Kappakoeffizient (Cohen, 1960) stellen ebenfalls etablierte Verfahren und Größen in der Validierung von Klassifikationsergebnissen dar. Der Kappakoeffizient ist eines der am häufigsten verwendeten Maße zur Beschreibung der Güte von Klassifikationsergebnissen und gibt an, inwiefern sich die Werte der Konfusionsmatrix von einem statistischen Zufallsergebnis unterscheiden. Er berücksichtigt neben den Werten der Hauptdiagonalen auch die Über- und Unterlassungsfehler und berechnet sich wie folgt:

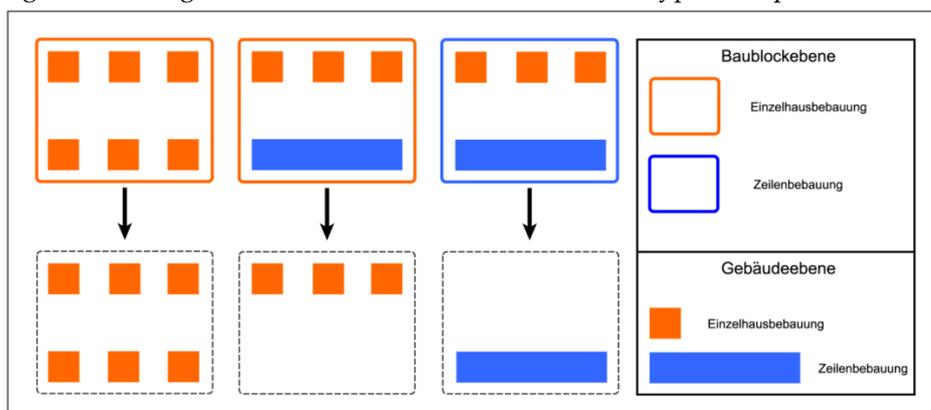
$$(4) \quad \hat{K} = \frac{N \sum_{i=1}^r x_{ii} - \sum_{i=1}^r (x_{i+} \cdot x_{+i})}{N^2 - \sum_{i=1}^r (x_{i+} \cdot x_{+i})}$$

wobei  $r$  die Anzahl der Reihen in der Matrix,  $x_{ii}$  die Anzahl der Beobachtungen in Reihe  $i$  und Spalte  $i$  darstellt,  $x_{i+}$  und  $x_{+i}$  sind die Gesamtwerte von Reihe  $i$  und Spalte  $j$  und  $N$  ist die Summe der Beobachtungen.

In der praktischen Anwendung der Konfusionsmatrix werden in der Regel nicht sämtliche Elemente der Klassifikation mit den Referenzdaten verglichen, sondern eine statistisch repräsentative Stichprobe gezogen. Die Güte der Klassifikation wird auf Basis dieser Stichprobe für die gesamte Klassifikation abgeschätzt. Die untersuchten Referenzpunkte sollten thematisch stratifiziert und räumlich randomisiert über das untersuchte Klassifikationsergebnis verteilt sein. Dabei wird die Anzahl der benötigten Referenzpunkte für eine Klassifikation mit mehr als zwei Klassen über eine multinomiale Verteilung der Klassenpixel berechnet. Eine detaillierte Beschreibung der Berechnung der Anzahl der Stichproben ist in Abschnitt 4.3.3 beschrieben.

Für die Validierung der Landbedeckungsklassifikation werden Referenzdaten benötigt, anhand welcher die Güte des Klassifikationsergebnisses bestimmt werden kann. Diese Referenzdaten stammen üblicherweise aus terrestrischen Erhebungen, aus bestehenden Karten oder Luftbildinterpretationen oder aus anderen Quellen. Die Landbedeckungsklassen, welche in den Satellitenbilddaten klassifiziert werden, sind in den Bilddaten eindeutig identifizierbar und werden auf Basis von etablierten Merkmalen wie dem NDVI (vgl. Abschnitt 4.2.2) klassifiziert. Trotzdem werden von computergestützten Bildklassifikationsverfahren Fehler in der Zuordnung gemacht, die durch die Überlegenheit der visuellen Bildinterpretation von einem erfahrenen Interpreten identifiziert werden können. Als einzige Alternative zur terrestrischen Erhebung von Referenzdaten in den beiden Untersuchungsgebieten wurden daher im Rahmen der Arbeit auf Basis einer Stichprobe für jede Landbedeckungsklasse visuelle Referenzpunkte in den Satellitenbilddaten erhoben und mit den Klassifikationsergebnissen verglichen.

Für die Validierung der Gebäudeklassifikation werden amtliche Referenzdaten über die städtischen Strukturtypen verwendet (vgl. Abschnitt 3.3.2.2), da für beide Städte keine Gebäudetypenklassifikation mit den im Rahmen der Arbeit verwendeten Gebäudeklassen verfügbar ist. Vor diesem Hintergrund wurden die Referenzdaten für die Validierung der semantischen Gebäudeklassifikation aus den Stadtstrukturtypenklassifizierungen erstellt. Dabei wurden die Einzelgebäude mit den Stadtstrukturtypenklassifikationen räumlich überlagert und die Information aus der übergeordneten Stadtstrukturtypenklassifikation auf die darunter liegenden Einzelgebäude übertragen. Dabei wurde durch visuelle Überarbeitung der übergeordnete Stadtstrukturtyp aber nur auf jene Einzelgebäude übertragen, welche auch tatsächlich dem Stadtstrukturtyp entsprechen. Eine graphische Darstellung dieses Arbeitsschrittes ist in untenstehender Abbildung 18 dargestellt. Die Abbildung zeigt in der oberen Hälfte die klassifizierten Baublöcke auf Basis der Stadtstrukturtypenkartierung und die Einzelgebäude mit ihrer Klassenzugehörigkeit. Durch visuelle Überarbeitung bleiben nur mehr die Einzelgebäude übrig, die auch der Klasse der Stadtstrukturtypen entsprechen.



**Abbildung 18:** Erstellung des Trainings- und Validierungsdatensatzes auf Gebäudeebene anhand der Stadtstrukturtypenklassifizierung.

### 3.5. Zusammenfassung

In diesem Kapitel wurden der konzeptionelle Rahmen und die für die Arbeit verwendeten Daten vorgestellt. Der Rahmen beschreibt die technische Verknüpfung zwischen den Fernerkundungsdaten und den Surveydaten. Auf diese Weise können räumliche Daten mit den datenschutzrechtlich höchst sensiblen Surveydaten verknüpft werden, ohne dass die persönlichen Rechte von Umfrageteilnehmern verletzt werden. Im konzeptionellen Rahmen wird darüber hinaus auch die räumliche Ebene, auf welcher die flächendeckenden Fernerkundungsinformationen aggregiert werden, beleuchtet. Diese Ebene stellt das Bindeglied zwischen den einzelnen städtischen Strukturelementen und übergeordneten administrativen Raumeinheiten dar.

Im zweiten Abschnitt wurden die Auswahl der beiden Untersuchungsgebiete München und Berlin begründet und die beiden Städte kurz charakterisiert.

Die Geodatengrundlage für beide Städte wurde im dritten Abschnitt beschrieben. Sie besteht für beide Städte aus geometrisch höchst aufgelösten Satellitenbilddaten der Sensoren Ikonos und Quickbird. Darüber hinaus wurden auch die digitalen Oberflächenmodelle für die beiden Städte beschrieben. Im letzten Teil des Abschnitts wird auf die weiteren Geodaten eingegangen, welche für bestimmte Verfahren im Rahmen der Arbeit verwendet wurden.

Im letzten Teil des Kapitels wird die Vorgehensweise für die Validierung der in den folgenden Kapiteln beschriebenen Klassifikationsergebnisse beschrieben. Die Ergebnisse aus den Validierungen erlauben eine Einschätzung der Güte der einzelnen Klassifikationsergebnisse.



## Kapitel 4

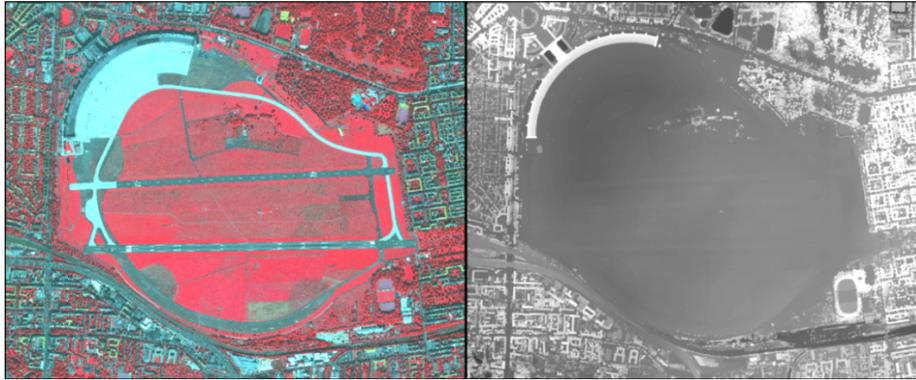
# Erfassung der Gebäudestruktur und der Landbedeckung in den Testgebieten

*“[...] the natural and built environment, the design of space and place, shapes social relation [...].”*

Ann R. Tickamyer (2000)

Fernerkundungsbilder stellen zunächst nur Daten und keine Informationen dar. Deshalb müssen die vom Sensor aufgezeichneten Bilder verarbeitet werden, um ihre Informationen nutzen zu können, beziehungsweise um sie im Rahmen der Arbeit mit den Surveydaten zu verknüpfen. Die Bildklassifikation stellt neben weiteren Verfahren eine Art der thematischen Auswertung dar. Sie dient dazu, um beispielsweise die komplexen urbanen Strukturen in einem Satellitenbild zu erfassen und in die entsprechenden thematischen Klassen einzuordnen. Die relevanten Zielklassen der Landbedeckung sind im Rahmen der Arbeit: *Gebäude*, *Vegetation*, und *versiegelte Flächen*.

Die Ableitung dieser Zielklassen beruht auf den in Abschnitt 3.3 beschriebenen Datengrundlagen aus digitalen Oberflächenmodellen und geometrisch höchst aufgelösten, spektralen Satellitenbilddaten. Die beiden Datensätze ergänzen sich gegenseitig und beinhalten komplementäre Informationen für die Ableitung der städtischen Struktur. Die spektrale Information im Satellitenbild wird für die Identifikation der unterschiedlichen Landbedeckungsarten verwendet und die Höheninformation im digitalen Oberflächenmodell dient zur Abgrenzung der Gebäude und der Ableitung ihrer Höhe. Ein Beispiel für die gegenseitige Ergänzung der Informationen aus den fusionierten Daten ist in Abbildung 19 dargestellt. Zu sehen ist auf der linken Seite ein Ausschnitt vom ehemaligen Flughafen Tempelhof in Berlin in einer Falschfarbeninfrarotdarstellung. Es sind deutlich die unterschiedlichen Landbedeckungsarten wie Vegetation und versiegelte Flächen aufgrund ihres unterschiedlichen Rückstreuerverhaltens zu unterscheiden. Auf der rechten Seite zeigt die unterschiedliche Helligkeit der Objekte ihre Höhe, wodurch sich Gebäude von ihrer Umgebung abgrenzen lassen.



**Abbildung 19:** Darstellung des ehemaligen Berliner Flughafens Tempelhof in der Quickbird-Aufnahme in einer Falschfarbeninfrarotdarstellung (links) und im digitalen Oberflächenmodell (rechts).

Quickbird-Aufnahme: © European Space Imaging, Oberflächenmodell: © DLR Berlin-Adlershof

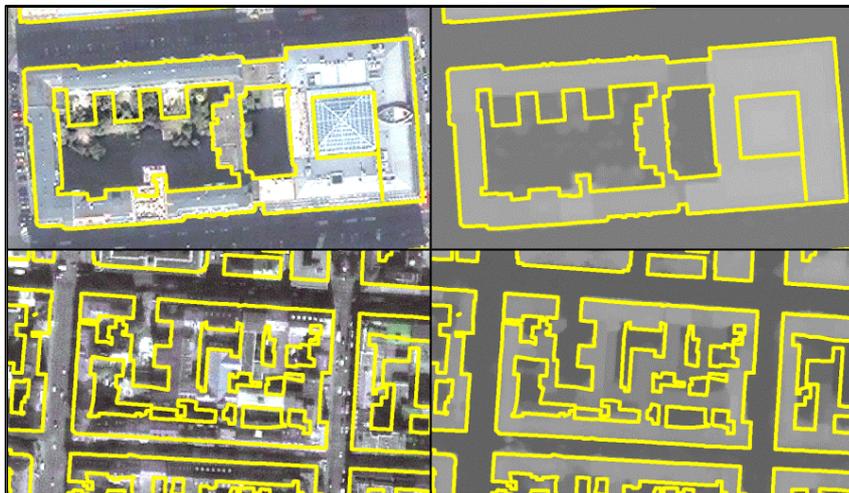
Die Erfassung der Gebäude und der Landbedeckung auf Basis der Datensätze anhand eines Bildklassifikationsverfahrens wird im aktuellen Kapitel beschrieben. Beim Klassifikationsverfahren handelt es sich um einen objektbasierten Klassifikationsansatz, in welchem die Auswertung nicht auf dem Einzelpixel durchgeführt wird, sondern auf zusammenhängenden Pixelgruppen – den Objekten. Diese Objekte erlauben die Verwendung von Formmerkmalen für die semantische Klassifikation von Gebäudeklassen, die im nachfolgenden Kapitel 5 beschrieben wird.

In den folgenden Abschnitten werden die relevanten Vorverarbeitungsschritte, die Abgrenzung der Gebäude und der Klassifikationsablauf für die Ableitung der Landbedeckungsklassen beschrieben. Das Klassifikationsverfahren ist in einen hierarchischen Ablauf integriert, in welchem zunächst die Gebäudestrukturen als Gebäudemasken aus dem digitalen Oberflächenmodell abgeleitet werden und danach in einem weiteren Schritt die Landbedeckungsklassen im optischen Datensatz klassifiziert werden.

### 4.1. Datenvorverarbeitung

Durch unterschiedliche Aufnahmebedingungen wie Sensoreigenschaften, Jahreszeit, Tageszeit, Bewölkung, atmosphärische Bedingungen, die Aufnahmegeometrie oder durch die Dynamik der Landoberfläche stellt jedes Satellitenbild ein Unikat dar. Für einige dieser Einflüsse auf das Satellitenbild gibt es in der praktischen Anwendung von Fernerkundungsdaten Verfahren, um eine einheitliche Datenbasis zu schaffen. Die Lagegenauigkeit der abgebildeten Objekte ist in den Rohdaten der optischen Satellitenbilddaten für eine gemeinsame Verwendung mit den digitalen Oberflächenmodellen nicht ausreichend. Deshalb stellt die geometrische Korrektur der Bilddaten einen wichtigen Vorverarbeitungsschritt bei der Nutzung unterschiedlicher räumlicher Informationsquellen dar. Die Ursachen für geometrische Fehler in digitalen Fernerkundungsdaten sind a) Instrumentenfehler, b) Panoramaverzerrung und

Geländehöhenunterschiede, c) Erdrotation und d) die Instabilität der Plattform (Mather, 1999). Die Korrektur von geometrischen Fehlern in den Bilddaten beinhaltet die Herstellung einer Beziehung zwischen dem Bildkoordinatensystem und einem Kartenkoordinatensystem. Dafür werden im zu korrigierenden Bild die gleichen Objekte wie in einem geometrisch korrigierten Referenzdatensatz identifiziert und eine geometrische Beziehung zwischen den beiden Koordinatensystemen hergestellt. In der vorliegenden Arbeit wurden die Satellitenbilddaten anhand von Passpunkten in Referenzorthofotos korrigiert. Ein Maß zur Beurteilung der Güte der geometrischen Korrektur ist der mittlere quadratische Vorhersagefehler (engl.: *root mean square error*, RMSE). Er wird berechnet durch die Quadratwurzel des durchschnittlichen Prognosefehlers, also der durchschnittlichen Abweichung der Prognose von der tatsächlichen Beobachtung. Er gewichtet hohe Prognosefehler stärker als geringe (Alisch, Winter und Arentzen, 2005). In der Fernerkundungspraxis ist ein üblicher Richtwert für die Güte der Korrektur ein RMSE kleiner als die geometrische Auflösung. Für die beiden Satellitenbildszenen wurde anhand der digitalen Orthofotos ein RMSE kleiner als ein Pixel erzielt. Die Lagegenauigkeit der Objekte liegt also unter 1 m. Die digitalen Oberflächenmodelle weisen bereits eine Lagegenauigkeit kleiner als 1 m auf, weshalb für diese Daten keine geometrische Korrektur durchgeführt wurde. Die trotz der geometrischen Korrektur der spektralen Satellitenbilddaten verbleibenden geometrischen Lageungenauigkeiten sind für die Anwendung in der vorliegenden Arbeit nicht wesentlich, da für die objektbasierte Verarbeitung keine pixelgenaue Überlagerung der Bilddaten zwingend erforderlich ist. Eine Darstellung über die Lage von Gebäuden nach der geometrischen Vorverarbeitung ist in Abbildung 20 dargestellt.

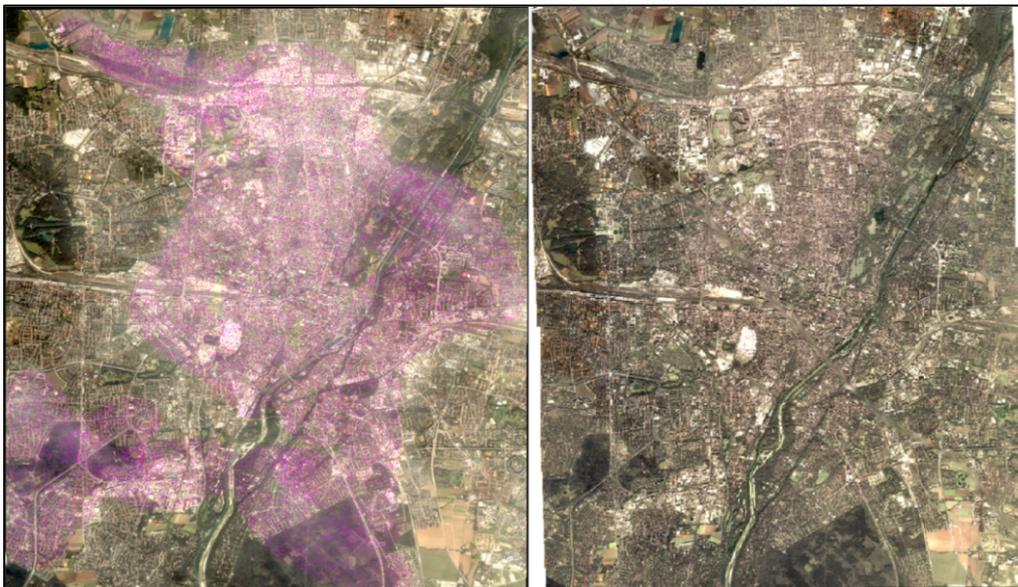


**Abbildung 20:** Lage von Objekten in den Satellitenbilddaten und den digitalen Oberflächenmodellen (oben: Berlin, unten: München; Überlagerung: Digitales Gebäudemodell)

© Landeshauptstadt München, Senatsverwaltung für Stadtentwicklung Berlin

Viele optische Satellitenbildszenen beinhalten Wolken oder Dunstbereiche. Während bei Wolken die optische Dicke so hoch ist, dass die darunter liegenden Bereiche auf der Erdoberfläche vom Sensor nicht erfasst werden können, können für Gebiete unterhalb von Dunstbereichen noch Informationen erfasst werden. In der Ikonos-Aufnahme für München sind über dem Stadtzentrum großflächige Dunstbereiche zu erkennen (vgl. Abbildung 21, links). Sie wurden mit der Funktion *haze removal* im Softwarepaket ATCOR (Atmospheric and Topographic Correction) entfernt (ATCOR, 2010; Richter, 2010). Die Entfernung der Dunstbereiche funktioniert in ATCOR vollautomatisch. Dabei wird die untersuchte Szene in wolken- und dunstfreie Bereiche, dunstige Bereiche und Wolken unterteilt. Der Algorithmus verfährt in fünf Schritten. Im ersten Schritt wird eine Dunstmaske erstellt. Danach wird zwischen dem blauen und dem roten Band eine Regression für die dunstfreien Bereiche angewendet (Richter, 1996). Darauf aufbauend wird für die Dunstbereiche eine *haze optimized transform* (HOT) durchgeführt und die Histogramme für diese Gebiete berechnet. Im letzten Schritt werden die Grauwerte der Dunstpixel um einen aus den Histogrammen der Dunstbereiche und der dunstfreien Bereiche berechneten Wert reduziert. Der Algorithmus funktioniert nur für Pixel, welche über der Landoberfläche liegen, weshalb das infrarote Band (NIR) benötigt wird, um Wasserpixel auszuschließen.

Abbildung 21 zeigt das Ergebnis der Datenvorverarbeitung für die Ikonos-Aufnahme für das Untersuchungsgebiet München. Auf der linken Seite ist die Ausgangsszene vor der geometrischen Korrektur und des *haze removal* abgebildet und auf der rechten Seite das Ergebnis.



**Abbildung 21:** Ikonos-Aufnahme für München vor (links) und nach (rechts) der Datenvorverarbeitung. Die Dunstbereiche sind im linken Bild markiert. © European Space Imaging

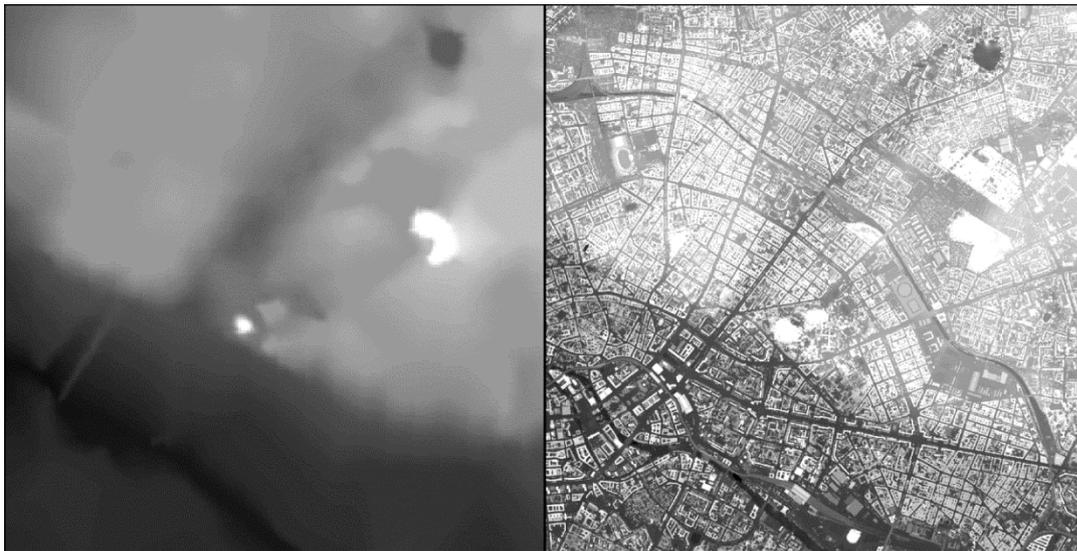
## 4.2. Ableitung der Gebäudestruktur

Die Ableitung der dreidimensionalen Bebauungsstruktur aus dem digitalen Oberflächenmodell ist in einzelne Verarbeitungsschritte unterteilt, welche im Folgenden beschrieben werden.

### 4.2.1. Normalisierung der digitalen Oberflächenmodelle

In digitalen Oberflächenmodellen (DOM) ist die Höhe des Geländes inklusive der erhabenen Objekte – wie Gebäude oder hohe Vegetation – auf der Erdoberfläche dargestellt. Die Grauwerte der Pixel repräsentieren also sowohl die Höhe des Geländes als auch die darauf liegenden Objekte. Im Rahmen der Arbeit sind vor allem die Gebäudeobjekte von Interesse, weshalb die Gesamthöhe der Gebäudeobjekte zuerst um die Geländehöhe reduziert werden muss. Das Resultat dieser Reduktion stellt ein normalisiertes digitales Oberflächenmodell (nDOM) dar, in welchem nur mehr die Höhen der erhabenen Objekte ohne die Geländehöhe erhalten sind. Für die Normalisierung wurden die Höhen aus den digitalen Oberflächenmodellen mit den Höhen aus den digitalen Geländemodellen (DGM; vgl. Abschnitt 3.3.2.1) subtrahiert.

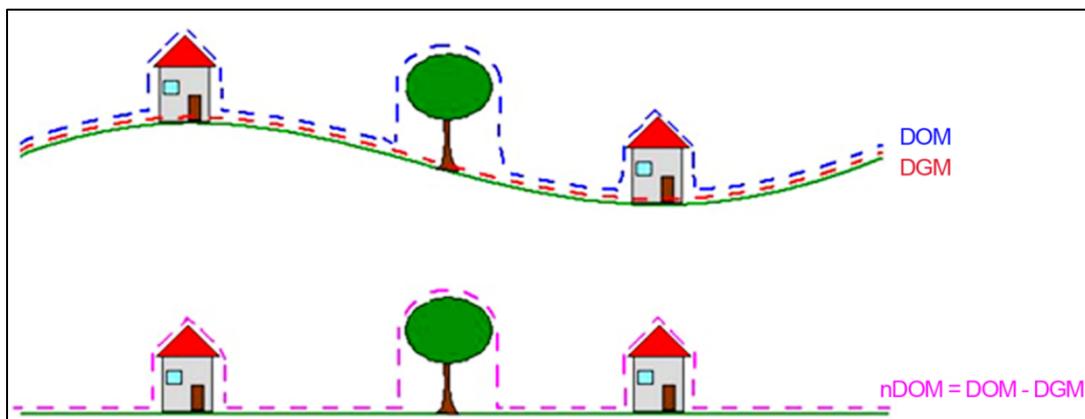
Eine exemplarische Darstellung der beiden Datensätze ist in Abbildung 22 visualisiert. Sie zeigt das DGM und das DOM vor der Normalisierung für einen Ausschnitt des Untersuchungsgebietes Berlin. Auf der linken Seite der Abbildung ist das DGM mit den Geländehöhen dargestellt. Es sind deutlich die tieferen (dunkleren) von den höheren (helleren) Bereichen des Geländes zu unterscheiden. Zu den Geländeerhöhungen zählt auch der Trümmerberg im Volkspark Friedrichshain, welcher im Bild als helles Objekt zu erkennen ist. Tiefer liegende Bereiche finden sich im oberen rechten Bildbereich, wo der Weissensee durch seine dunkle Fläche zu erkennen ist. Das DOM auf der rechten Seite der Abbildung zeigt zusätzlich zu den Geländehöhen auch die Gebäude- und Vegetationshöhen. Sie bleiben nach dem Abzug der Geländehöhen von den Objekthöhen übrig.



**Abbildung 22:** Vergleich zwischen Höheninformation im DGM (links) und im DOM (rechts).

DGM: © Bundesamt für Kartographie und Geodäsie, DOM: © DLR Berlin-Adlershof

Das Prinzip der Erstellung eines normalisierten digitalen Oberflächenmodells ist in Abbildung 23 dargestellt. Es zeigt die Höhen der Geländeoberfläche im DGM und die Höhen der Objekte auf der Geländeoberfläche inklusive der Geländehöhen. Nach der Reduktion der Objekthöhen um die Geländehöhen bleibt eine plane Ebene mit ausschließlich den erhabenen Objekten auf der Geländeoberfläche.



**Abbildung 23:** Normalisierung von digitalen Oberflächenmodellen.

(Verändert nach: Brüßhaber et al. 2010)

Das Ergebnis dieses Normalisierungsschrittes zwischen dem DGM und dem DOM ist in Abbildung 24 dargestellt. Sie zeigt das nDOM für den gleichen Ausschnitt wie in Abbildung 22. Es ist zu erkennen, dass die Einflüsse des Geländes reduziert wurden und die helleren Bildbereiche nur mehr Objekte auf der Geländeoberfläche darstellen. Durch die unterschiedlichen geometrischen Auflösungen von 1 m im DOM und 25 m im DGM und die

unterschiedlichen Höhengenaugkeiten, können durch dieses Vorgehen auch Fehler im resultierenden nDOM entstehen, weshalb unten auf alternative Verfahren eingegangen wird.

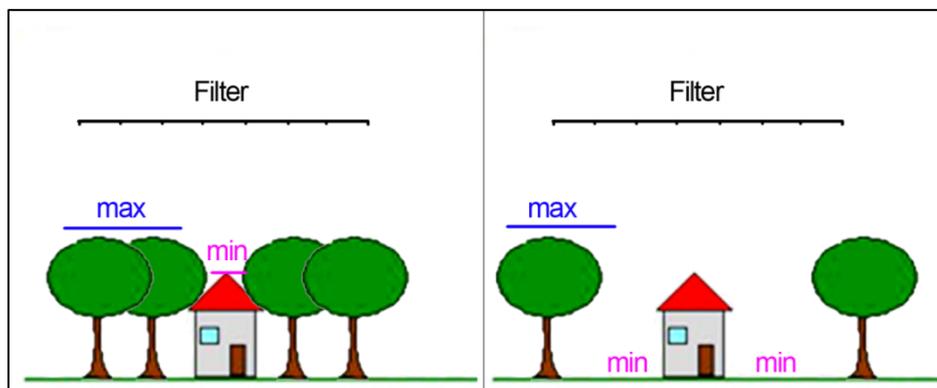


**Abbildung 24:** Normalisiertes digitales Oberflächenmodell (nDOM).

Alternativ zur Verwendung von zusätzlichen digitalen Geländemodellen für die Erstellung von nDOMs werden in einigen Studien Verfahren beschrieben, welche das Geländemodell auf Basis von verschiedenen Filterungsmethoden aus dem Oberflächenmodell berechnen (z.B. Brenner, 2000; Mayer, 2000; Schiewe, 2001; Mayer, 2004; Bochow, 2009). Da in der vorliegenden Arbeit die Normalisierung auf Basis eines digitalen Geländemodells durchgeführt wird, wird keine Filterung des Oberflächenmodells durchgeführt. Es wird an dieser Stelle aber kurz auf Vorteile und Limitierungen der Filterung von Oberflächenmodellen eingegangen. Die Filterung verfolgt das Ziel, die Grauwerte der erhobenen Objekte auf der Geländeoberfläche durch niedrigere Werte von Nachbarpixeln zu ersetzen. Eine einfache Form der Filterung ist das morphologische *Opening*, welches aus einer Aneinanderreihung von Erosionen und Dilatationen besteht. Dabei wird in einem gleitenden Fenster (*moving window*) mit einer ungeraden Anzahl an Zeilen und Spalten dem Zentrumspixel zunächst der kleinste Grauwert des Fensters vergeben (Erosion). Im darauffolgenden Schritt wird der höchste Grauwert des Fensters vergeben (Dilatation). Dadurch wird die Oberfläche auf die im räumlichen Kontext benachbarten kleinsten Grauwerte reduziert. Die Objekte, welche eine kleinere räumliche Ausdehnung als die definierte Filtergröße aufweisen, werden durch diesen Vorgang eliminiert.

Die Filterung des digitalen Oberflächenmodells für die Erstellung des DGMs hat den Vorteil, dass neben dem DOM kein zusätzliches DGM benötigt wird. Allerdings ergeben sich durch die Filteroperationen auch Schwierigkeiten, auf die im Folgenden kurz eingegangen wird. Die Filterungsoperationen können nur dort die erhobenen Objekte auf die Höhe der Geländeoberfläche filtern, wo das Minimum innerhalb des Filterfensters auch die Geländeoberfläche darstellt. Im vorliegenden Fall wurden die beiden Oberflächenmodelle

durch photogrammetrische Verfahren aus stereoskopischen Luftbildbefliegungen gewonnen, welche noch während der Vegetationsperiode aufgezeichnet wurden (vgl. Tabelle 5, S. 53). Dadurch sind hohe Vegetationsflächen auch in den Oberflächenmodellen erkennbar und können vor allem in den städtischen Randgebieten – welche durch aufgelockerte Einzel- und Doppelhausbebauung mit starker Durchgrünung geprägt sind – die niedrigeren Bebauungsstrukturen überlagern. In diesem Fall würde das DOM höchstens bis auf die Höhe der Hausdächer reduziert werden und ein Segmentierungsverfahren würde keine Gebäudeobjekte generieren können. Aber vor allem die gelockerte Stadtrandbebauung ist für die Einbindung in einen sozialwissenschaftlichen Kontext relevant, da sie eine andere sozioökonomische Situation der Haushaltsbewohner implizieren kann, als Mehrfamilienhäuser. Eine schematische Darstellung ist in Abbildung 25 dargestellt. Sie zeigt die Filtergröße und die maximalen und minimalen Höhen im DOM.



**Abbildung 25:** Auswirkungen hoher Vegetation in nächster Nachbarschaft zu Häusern auf die Berechnung des kleinsten (min) und höchsten (max) Grauwertes innerhalb eines Filterfensters.

Eine weitere Schwierigkeit bei der Filterung von digitalen Oberflächenmodellen ergibt sich aus der Wahl der Filtergröße, welche für das gesamte Bild gleichmäßig angewendet wird. Die Entscheidung des Anwenders für eine bestimmte Filtergröße ist demnach immer ein Kompromiss zwischen einem zu großen und einem zu kleinen Filter. Die Auswirkungen des Letzteren sind besonders deutlich bei der Anwendung auf sehr großen Gebäuden oder über großflächigen, dichten Vegetationsgebieten ersichtlich, da sie in diesem Fall erhalten bleiben. Des Weiteren besteht bei der Anwendung von morphologischen Filteroperationen immer die Gefahr, dass sich im gefilterten Bild die Struktur des Filters abzeichnet.

#### 4.2.2. Eliminierung der Vegetation

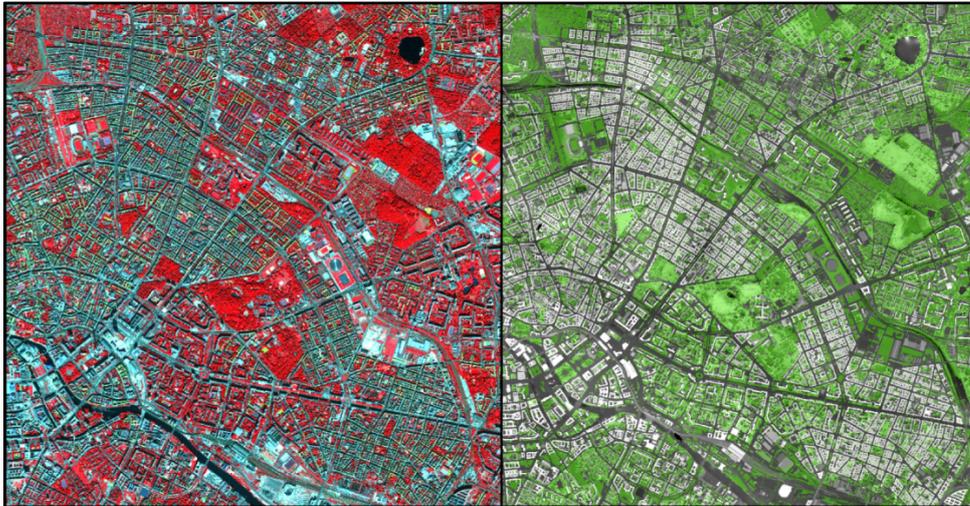
Wie bereits oben dargestellt, finden sich im normalisierten Oberflächenmodell neben den Gebäuden auch Bereiche mit hoher Vegetation. Das Verfahren zur Abgrenzung der Einzelgebäude ist auf die relativen Höhenunterschiede im Bilddatensatz ausgelegt, weshalb

eine Unterscheidung von Vegetation und Gebäuden alleine auf Basis dieser Information nicht möglich ist. Vor diesem Hintergrund werden vor der eigentlichen Gebäudeabgrenzung die Vegetationsbereiche anhand der spektralen Information im optischen Bilddatensatz identifiziert und für die weitere Verarbeitung ausmaskiert. Durch die geometrische, präzise Vorverarbeitung der Satellitenbilddaten (vgl. Abschnitt 4.1) können sie mit den Daten aus den digitalen Oberflächenmodellen überlagert werden und komplementär für die Verarbeitung verwendet werden. Für die Ausweisung der Vegetation in optischen Satellitenbilddaten wird ihr charakteristisches Reflexionsverhalten verwendet. Aktive Vegetation weist in den blauen und roten Wellenlängenbereichen des elektromagnetischen Spektrums relativ geringe Reflektanzen auf, mit einem lokalen Maximum im grünen Bereich. Dies ist auf die Absorption des blauen und roten Lichts vom Chlorophyll in den Pflanzen zurückzuführen, welches als Energie für die Photosynthese benötigt wird (Mather, 1999). Die Reflexion steigt im Wellenlängenbereich von 0,75 µm stark an. Dieser Effekt wird *red edge* genannt (Filella und Penuelas, 1994). Für die Klassifikation von aktiver Vegetation in optischen Fernerkundungsdaten hat sich die Bildung einer Ratio zwischen den Grauwerten des roten Wellenlängenbereichs und jenen des nahen Infrarots (nir) etabliert. Dieser *normalized difference vegetation index* (NDVI) berechnet sich wie folgt:

$$(5) \quad NDVI = \frac{nir - rot}{nir + rot}$$

Für die Klassifikation der Vegetationsflächen wurde das NDVI-Bild mittels einer *multiresolution* Segmentierung (Batz und Schäpe, 2000) in zusammenhängende Objektgruppen unterteilt und anhand eines empirisch ermittelten Schwellwertes in „Vegetationsbereiche“ und „Nicht-Vegetationsbereiche“ unterteilt. Durch die Verwendung des objektbasierten Ansatzes in diesem Zusammenhang werden nur zusammenhängende Pixelgruppen in die Vegetationsmaske aufgenommen und ein *Salz-und-Pfeffer-Effekt*, wie er bei pixelbasierten Klassifikationen vorkommen kann, vermieden. Überdies ist der Ablauf zur Gebäudeabgrenzung und zur Landbedeckungsklassifikation in eine objektbasierte Verarbeitungskette eingebaut, weshalb in den folgenden Arbeitsschritten wieder auf die abgeleiteten Vegetationsbereiche zurückgegriffen werden kann.

Die Klassifikation kann als *Vegetationsmaske* dargestellt werden, welche mit dem nDOM überlagert werden kann. Es werden für die Gebäudeabgrenzung nur mehr jene Bereiche betrachtet, welche nicht von der Vegetationsmaske überlagert werden. Abbildung 26 zeigt eine Falschfarbeninfrarotdarstellung von einem Ausschnitt des Untersuchungsgebietes Berlin mit den charakteristischen roten Vegetationsflächen und die Überlagerung der abgeleiteten Vegetationsmaske über dem nDOM (transparent).



**Abbildung 26:** Ableitung der Vegetationsmaske auf Basis der Reflexion im nahen Infrarot (links) und Überlagerung mit dem nDOM (rechts).

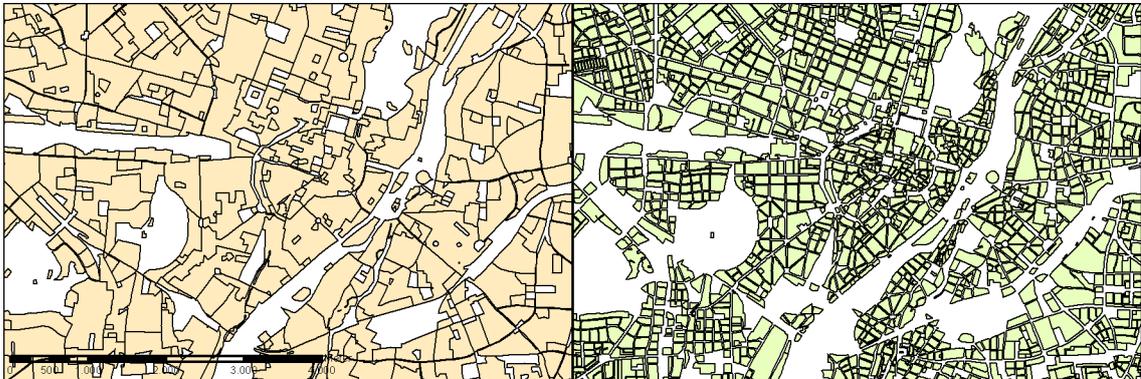
Quickbird-Aufnahme: © European Space Imaging.

### 4.2.3. Abgrenzung der Einzelgebäude

Auf der verbleibenden Information im nDOM mit den ausmaskierten Vegetationsflächen werden auf Basis eines lokal angepassten Segmentierungsverfahrens die Gebäude abgegrenzt. Dieser *local contrast-split* Segmentierer (Trimble, 2010b) untersucht Schritt für Schritt die Grauwertbereiche für einen kleinen Ausschnitt des Bildes und berechnet darauf den lokalen besten Schwellwert für eine Unterteilung in helle (erhöhte) und dunkle (niedrige) Bildobjekte. Die abgegrenzten Bildobjekte werden im weiteren Verlauf in „Gebäude“ und „Nicht Gebäude“ klassifiziert. Der Ablauf für die Segmentierung der Einzelgebäude wird im Folgenden im Detail beschrieben.

Für die Berechnung des lokalen besten Schwellwertes wird das Bild zunächst in kleinere Bereiche auf Basis von Baublöcken unterteilt. Sie haben den Vorteil, dass sie den städtischen Bereich in morphologische Raumeinheiten unterteilen und ihre Grenzen keine Gebäudestrukturen zerschneiden. Diese Baublockgrenzen stellen auch die räumliche Bezugsebene dar, welche im Rahmen des räumlichen Konzeptes für die Aggregation der räumlichen Nachbarschaftsmerkmale verwendet wird (vgl. Abschnitt 3.1). Im Rahmen der vorliegenden Arbeit beinhalten sie auch thematische Informationen über die Baustruktur innerhalb des Baublocks, welche in Kapitel 5 für das Training und die Validierung der Klassifikation der Gebäudetypen verwendet wird. Alternativ zu den Baublöcken können auch Vektordaten aus dem Objektbereich *Siedlung* des Basis-digitalen Landschaftsmodells (Basis-DLM) aus ATKIS verwendet werden. In diesem Datensatz sind zwar grundsätzliche Nutzungsinformationen enthalten, die Geometrie der Blockgrenzen repräsentiert allerdings für

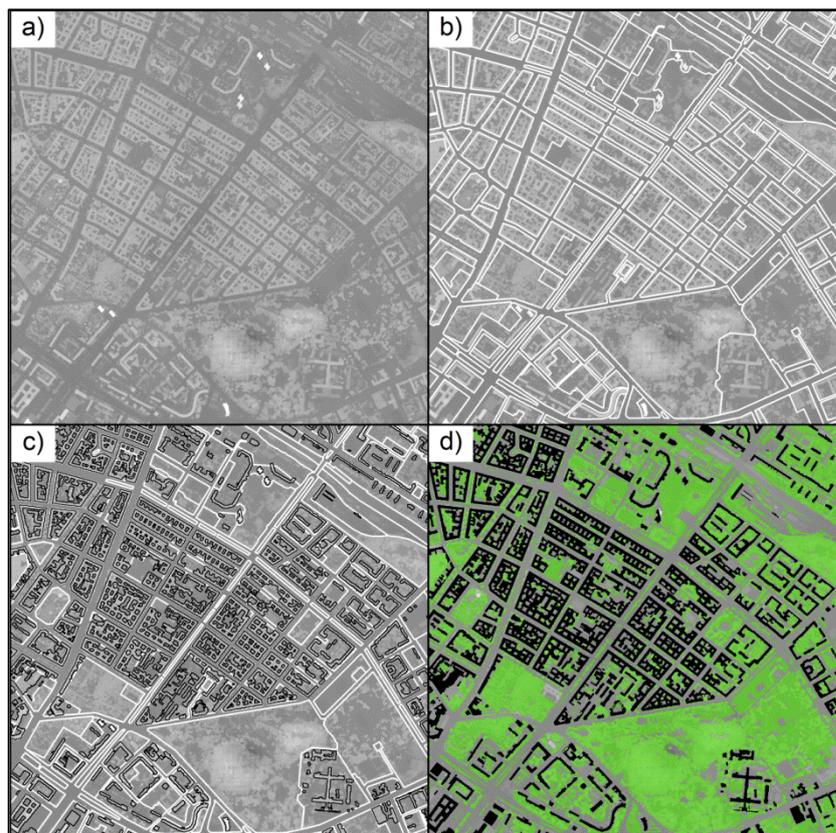
die untersuchten Städte nicht das gleiche Detail wie die verwendeten Baublöcke (vgl. Abbildung 27).



**Abbildung 27:** Vergleich der Blockgrenzen zwischen ATKIS Basis-DLM (links) und Baublöcken (rechts) für einen Ausschnitt im Untersuchungsgebiet München.

(ATKIS: © Bundesamt für Kartographie und Geodäsie, Baublöcke: © Landeshauptstadt München)

Die einzelnen Teilschritte der Abgrenzung der Einzelgebäude sind in Abbildung 28 dargestellt. Sie zeigt: a) das nDOM, b) die Baublöcke, c) die Gebäudepolygone und d) die Gebäudegrundflächen mit der Vegetationsmaske.



**Abbildung 28:** Ablauf der Abgrenzung der Einzelgebäude.

© Baublockgrenzen: Senatsverwaltung für Stadtentwicklung der Stadt Berlin.

Die Segmentierung der Gebäudeobjekte wird wie oben erwähnt auf Basis einer *local contrast-split* Segmentierung durchgeführt. Sie untersucht schrittweise die Grauwertverteilung innerhalb der Baublockgrenzen für die Pixel im nDOM, welche nicht von der Vegetationsmaske überlagert sind. Dabei wird ein lokal angepasster Schwellwert berechnet, welcher den Kontrast zwischen hellen und dunklen Grauwerten maximiert. Die Berechnung dieses höchsten Kontrastes ist ein iteratives Verfahren, welches wiederholt den Kontrast zwischen zwei Testgrauwerten berechnet. Diese Testschwellwerte befinden sich zwischen dem lokal kleinsten und dem lokal größten Grauwert. Zwischen den Grauwertbereichen, für welche der beste Kontrast gefunden wurde, wird ein Schwellwert gesetzt, welcher die erhabenen (helleren) Objekte von den niedrigen (dunklen) Objekten trennt. Dieser Schritt wird für jeden Baublock durchgeführt. Für die Berechnung des Kontrastes wurde im verwendeten Verfahren die *Kantenratio* (engl.: *edge ratio*) verwendet. Sie stellt eine Weiterentwicklung der oben beschriebenen, einfacheren Berechnung des Kontrastes auf Basis der *Objektdifferenzen* dar. Bei der Kantenratio wird der Kontrast zwischen den Testgrauwerten der potenziellen Kanten, welche in helle und dunkle Objekte trennen, berechnet. Die Formel für die Berechnung der Kantenratio lautet:

$$(6) \quad k = \frac{a - b}{a + b}$$

wobei  $k$  der Kontrast ist,  $a$  der Grauwertmittelwert der hellen Kantenobjekte und  $b$  der Grauwertmittelwert der dunklen Objekte. Das Verfahren zur Berechnung des höchsten Kontrastes und des daraus resultierenden besten Schwellwertes ist in Abbildung 29 dargestellt. Sie zeigt auf der linken Seite einen Ausschnitt aus einem digitalen Oberflächenmodell, wobei die helleren Grauwerte höhere Bereiche und die dunkleren Grauwerte niedrigere Bereiche darstellen. Die Pixel sind mit den entsprechenden Grauwerten überlagert. Der mittlere Teil der Abbildung zeigt die helleren und dunkleren Kantenobjekte zwischen der Kante auf Basis des getesteten Schwellwertes aus Gleichung (7). Für sie wird der mittlere Grauwert (vgl. Gleichung (8)) und der Kontrast berechnet (vgl. Gleichung (9)). Nachdem sämtliche vorkommenden Grauwerte im Baublock miteinander verglichen wurden, werden die helleren Objekte von den dunkleren Objekten auf Basis des höchsten berechneten Kontrastes getrennt.

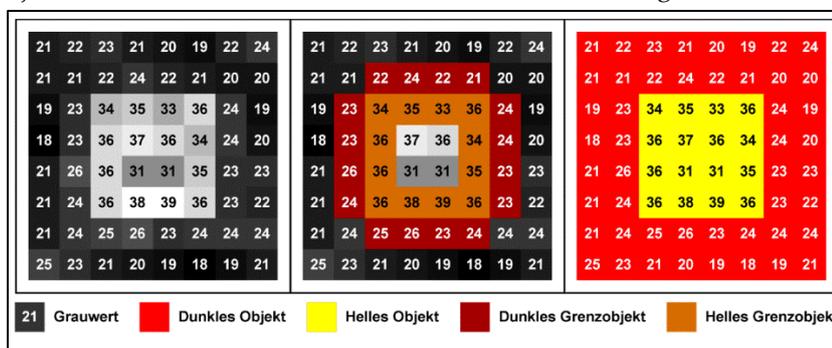


Abbildung 29: Berechnung des trennenden Schwellwertes auf Basis des besten Kontrastes.

$$(7) \quad \text{Testschwellwert} = 26$$

$$(8) \quad \begin{aligned} \text{Grauwertmittelwert des hellen Kantenobjektes} &= a = 35,67 \\ \text{Grauwertmittelwert des dunklen Kantenobjektes} &= b = 23,56 \end{aligned}$$

$$(9) \quad k = \frac{35,67 - 23,56}{35,67 + 23,56} = \frac{12,11}{59,23} = 0,2044$$

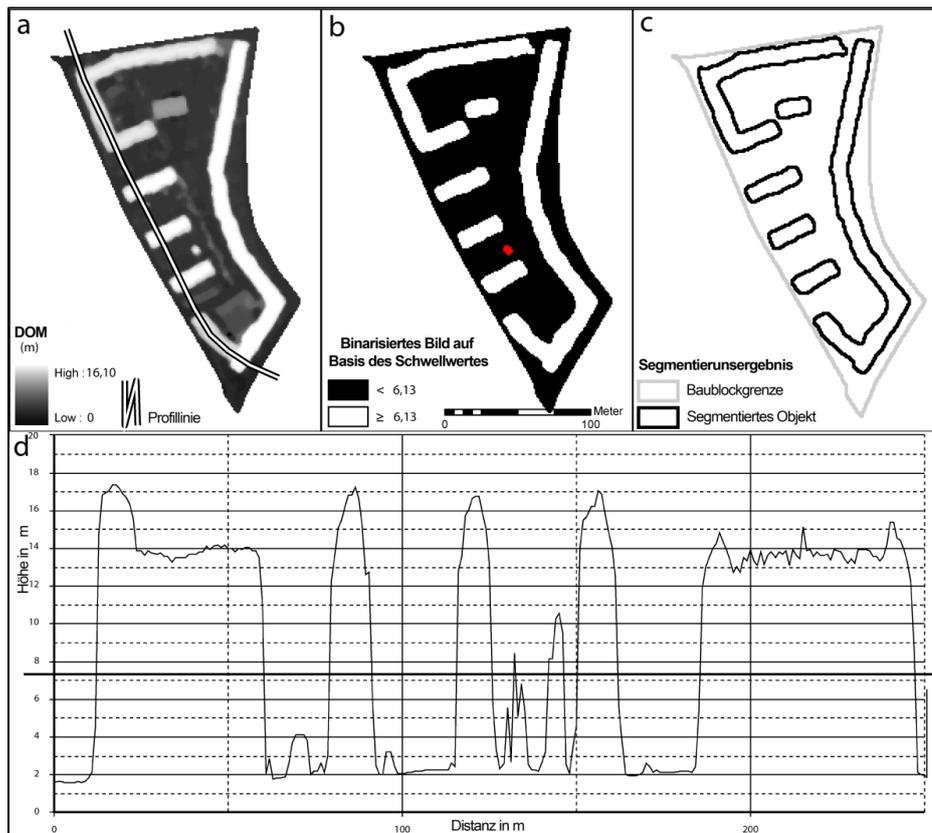
Eine Veranschaulichung des Verfahrens stellen Tabelle 8 und Abbildung 30 dar. Die Tabelle zeigt eine Zusammenfassung der vorkommenden Höhenwerte für den dargestellten Baublock in der Abbildung. Sie zeigt neben dem kleinsten und dem größten Grauwert (Minimum und Maximum) auch verschiedene Quantile ( $Q_x$ ), um die Verteilung der Grauwerte zu beschreiben. Darüber hinaus zeigt die Tabelle ebenfalls den gefundenen Schwellwert auf Basis des höchsten Kontrastes.

**Tabelle 8:** Berechnete Variablenwerte und bester Schwellwert für einen Beispielbaublock (vgl. Baublock in Abbildung 30).

Grauwertbereich	Höhenwert (m)
Minimum	0
Q01	0,31
Q10	0,73
Q20	0,88
Q30	0,96
Q40	1,07
Q50	1,37
Q60	2,46
Q70	7,72
Q80	12,58
Q90	14,47
Q99	15,93
Maximum	16,10
höchster Kontrast	0,045867
<b>Schwellwert</b>	<b>6,13</b>

Die Abfolge für die Segmentierung auf Basis dieses berechneten lokalen Schwellwertes ist in den Teilschritten in Abbildung 30 dargestellt. Im ersten Teil des Bildes (a) sind die Höhenwerte des DOMs für den Ausschnitt eines Baublockes inklusive einer Profillinie dargestellt. Die Profillinie zeigt die Grauwertausprägungen der Höhenwerte für fünf Gebäude. Die Höhenwerte der Profillinie sind im unteren Bereich der Abbildung dargestellt (d). Die Anwendung des in Tabelle 8 dargestellten Schwellwertes auf die Grauwerte des Baublockes ist im Bildbereich (b) als binarisiertes Bild dargestellt: in Schwarz sind die Grauwerte kleiner als

der berechnete lokale Schwellwert und in Weiß sind die Grauwerte größer gleich dem berechneten lokalen Schwellwert abgebildet. Die Abgrenzung der Einzelgebäude ist im Bildbereich (c) dargestellt. Sie zeigt außerdem die Anwendung eines empirisch ermittelten Flächenkriteriums, welches für die Anwendung im Rahmen der Arbeit auf eine Mindestgröße von 150 m<sup>2</sup> festgelegt wurde. Dadurch werden Objekte erst ab dieser Mindestgröße in der Gebäudemasken berücksichtigt. Fällt das Flächenkriterium zu klein aus, werden zu viele Objekte abgegrenzt, welche in der Natur keine Gebäude darstellen. Die Ergebnisse für die Gebäudeabgrenzung werden im folgenden Abschnitt 4.2.4 dargestellt und beschrieben.



**Abbildung 30:** Anwendung des in Tabelle 8 berechneten Schwellwertes für die Abgrenzung von Gebäuden.

#### 4.2.4. Ergebnisse der Gebäudeabgrenzung

Das oben beschriebene Verfahren zur Gebäudeabgrenzung wurde auf die normalisierten Oberflächenmodelle der Untersuchungsgebiete München und Berlin angewendet. Das Ergebnis stellt eine zweidimensionale Gebäudemasken dar, welche die Gebäudestruktur der Städte widerspiegelt. Die Ergebnisse sind in Abbildung 31 dargestellt, welche die Gebäudegrundflächen zeigt.

Im oberen Teil der Abbildung ist ein Ausschnitt für das charakteristische Zentrum von München dargestellt. An der Baustruktur sind unterschiedliche Phasen der Stadtentwicklung zu erkennen. So zeigt beispielsweise das Zentrum der Abbildung das verdichtete historische Altstadtzentrum. Es grenzte sich früher von seinen umliegenden Vorstädten durch eine Stadtmauer ab, welche allerdings gegen Ende des 18. Jahrhunderts weitgehend abgerissen wurde. Der Verlauf der ehemaligen zweiten Stadtmauer ist auch heute noch an der Baustruktur erkennbar. Angrenzend an den historischen Stadtkern finden sich die Stadtteile, welche im Zuge der Stadterweiterungen im 19. Jahrhundert entstanden sind: die Ludwigsvorstadt, die Isarvorstadt und die Maxvorstadt. Im linken unteren Bildbereich ist die Freifläche der Theresienwiese erkennbar.

Der untere Teil der Abbildung zeigt die Gebäudemasse für das östliche Stadtzentrum Berlins um die Spreeinsel und den Alexanderplatz. Charakteristisch für die aktuelle Baustruktur von Berlin ist die geringere Dichte der Bebauung gegenüber vergleichbaren mitteleuropäischen Großstädten und die großflächigen, zentrumsnahen Blockbebauungen, welche im 19. Jahrhundert entstanden. Besonders charakteristisch stellt sich die durch den Sozialismus in der Phase der ehemaligen Deutschen Demokratischen Republik (DDR) im Zeitraum von 1949 – 1990 geprägte Baustruktur dar. Beispiele dafür sind im rechten Bildrand zu erkennen.



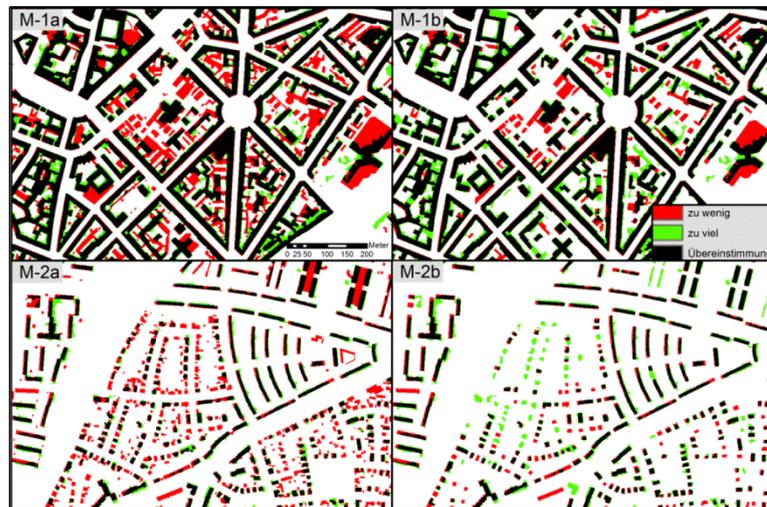
Abbildung 31: Ergebnisse der Gebäudeabgrenzung für München (oben) und Berlin (unten).

Das verwendete Verfahren zur Gebäudeabgrenzung verläuft automatisch und ohne Anpassungen durch den Anwender. Die Güte der abgegrenzten Gebäude ist daher direkt von den Eingangsdaten abhängig. Faktoren, welche Auswirkungen auf die Genauigkeit der abgegrenzten Gebäude haben, sind die geometrische Auflösung der digitalen Oberflächenmodelle und der digitalen Geländemodelle, die Höhen- und Lagegenauigkeit und der Befliegungszeitpunkt (während oder außerhalb der Vegetationsperiode). Durch die Segmentierung können daher auch nur jene Objekte abgegrenzt werden, welche auch tatsächlich in den Datengrundlagen enthalten sind. Die Bilder für die photogrammetrischen Oberflächenmodelle für München und Berlin wurden beide während der Vegetationsperiode befliegen (vgl. Tabelle 5, S. 53), weshalb die Vegetation in den Oberflächenmodellen enthalten ist und für die Abgrenzung der Gebäude neben der Höhe auch der NDVI verwendet wurde (vgl. vorheriger Abschnitt).

Ein Vergleich über die Güte des Verfahrens zur Gebäudeabgrenzung kann anhand von amtlichen Gebäudedaten aus dem Kataster durchgeführt werden, welche als Referenz für die Validierung der Gebäudeabgrenzung verwendet werden (Aktualitätsstand: 2008 München und 2009 Berlin). Sie weisen neben der Gebäudegrundfläche auch die Geschossigkeit der Gebäude auf (vgl. Abschnitt 3.3.2.4). Im Vergleich werden die abgeleiteten Gebäudegrundflächen mit dem digitalen Gebäudemodell aus dem Kataster gegenüber gestellt und ein flächenhafter Vergleich der erfassten und der nicht erfassten Gebäude durchgeführt. Dafür werden die abgeleiteten Gebäudegrundflächen mit den Referenzdaten überlagert und die Flächen quantifiziert. Eine visuelle Darstellung des Vergleichs ist in Abbildung 32 für München und in Abbildung 33 für Berlin dargestellt. Dabei sind die übereinstimmenden Flächen in schwarz dargestellt, die fehlenden Flächen in rot und die zu viel abgegrenzten Gebäude in grün.

Die Abbildungen zeigen jeweils den Vergleich zwischen dem Referenzdatensatz und den abgegrenzten Gebäuden für zwei unterschiedliche Gebäudestrukturen. Im oberen Teil der Abbildungen sind große, blockartige Bebauungsstrukturen und im unteren Teil sind kleine, lockere Bebauungsstrukturen wie Einzel- und Doppelhausbebauungen und Zeilenbebauungen zu erkennen. Im Vergleich der beiden Gebäudemasken wird deutlich, dass die großen Baustrukturen weitgehend richtig abgegrenzt wurden. Vor allem kleinere Gebäude wie Hinter- und Nebengebäude innerhalb von Blockrandbebauungen und kleinere Einzel- und Doppelhäuser wurden jedoch bei der Segmentierung nicht erfasst, da sie nicht das berechnete Schwellwert- oder Flächenkriterium erfüllen. Diese kleineren Hinter- und Nebengebäude stellen je nach Aufnahme in das amtliche Referenzgebäudemodell aus dem Kataster in ihrer Fläche einen bedeutenden Teil der Gesamtfläche dar (~45,9 %). Sie sind allerdings nicht eindeutig im amtlichen Gebäudemodell aus dem Kataster gekennzeichnet, weshalb kein Vergleich eines um diese Gebäude bereinigten Referenzdatensatzes mit den Ergebnissen der Gebäudeabgrenzung durchgeführt werden kann. Eine Betrachtung der Eigenschaften dieser Gebäude im amtlichen Gebäudemodell zeigt jedoch, dass viele dieser Gebäude eine Geschossigkeit von 0 bzw. 1 aufweisen. Vor diesem Hintergrund wird der Vergleich zwischen

der Gebäudeabgrenzung aus dem digitalen Oberflächenmodell und den amtlichen Gebäudedaten aus dem Kataster ein zweites Mal durchgeführt. Es werden beim zweiten Vergleich allerdings nur Gebäude in den Referenzdatensatz aufgenommen, welche eine Geschossigkeit größer als 1 aufweisen. Dadurch fällt ein Großteil der niedrigen Nebengebäude aus dem Referenzdatensatz heraus. Allerdings fallen dabei auch teilweise Lager- und Fabrikhallen heraus, welche ebenfalls mit einer Geschossigkeit kleiner als 2 attribuiert sind. In Abbildung 32 für München und in Abbildung 33 für Berlin ist dieser Vergleich im rechten Teil der Abbildung dargestellt. Es ist erkennbar, dass die unterlassenen Gebäudestrukturen gegenüber dem gesamten Datensatz im linken Teil der Abbildung deutlich reduziert sind.



**Abbildung 32:** Vergleich der abgegrenzten Gebäudegrundflächen mit dem amtlichen Gebäudemodell für München. (M-1a und M-2a: alle Gebäude; M-1b und M-2b: Gebäude mit mehr als einem Geschoss).



**Abbildung 33:** Vergleich der abgegrenzten Gebäudegrundflächen mit dem amtlichen Gebäudemodell für Berlin. (M-1a und M-2a: alle Gebäude; M-1b und M-2b: Gebäude mit mehr als einem Geschoss).

Diese Beobachtung zeigt sich auch im quantitativen Vergleich in Tabelle 9 für München und in Tabelle 10 für Berlin. Sie zeigt die Gesamtübereinstimmungen zwischen dem amtlichen Gebäudemodell aus dem Kataster und den abgegrenzten Gebäuden sowie die Differenzierung nach Nutzer- und Herstellergenauigkeit (vgl. hierzu Validierungsmethoden in Abschnitt 3.4) für den Vergleich mit dem gesamten Datensatz und den Vergleich nur mit Gebäuden, welche mehr als ein Geschoss aufweisen.

Es ist zu erkennen, dass die Herstellergenauigkeit steigt, wenn nur Gebäude mit einer Höhe größer als ein Geschoss miteinander verglichen werden, dabei sinkt jedoch die Nutzergenauigkeit. Das bedeutet, dass im Vergleich mit dem Referenzdatensatz mit Gebäuden mit mehr als einem Geschoss natürlich weniger Gebäude bei der Gebäudeabgrenzung unterlassen werden (Anstieg der Herstellergenauigkeit), an anderen Stellen jedoch Gebäude abgegrenzt wurden, für die es im Referenzdatensatz kein Pendant gibt (Sinken der Nutzergenauigkeit).

**Tabelle 9:** Flächenbasierte Genauigkeitsabschätzung der Gebäudemaske für München.

	Nutzergenauigkeit		Herstellergenauigkeit		Gesamt	Kappa
	Gebäude	Nicht-Gebäude	Gebäude	Nicht-Gebäude		
<b>Alle Gebäude</b>	82,72 %	92,32 %	57,88 %	97,67 %	91,24 %	0,632
<b>Gebäude &gt; 1</b>	70,21 %	96,74 %	73,33 %	96,22 %	93,74 %	0,682

**Tabelle 10:** Flächenbasierte Genauigkeitsabschätzung der Gebäudemaske für Berlin.

	Nutzergenauigkeit		Herstellergenauigkeit		Gesamt	Kappa
	Gebäude	Nicht-Gebäude	Gebäude	Nicht-Gebäude		
<b>Alle Gebäude</b>	83,23 %	94,52 %	70,53 %	97,28 %	92,98 %	0,723
<b>Gebäude &gt; 1</b>	80,25 %	96,22 %	76,99 %	96,87 %	94,05 %	0,751

Zusammenfassend betrachtet, zeigt sich, dass die Abgrenzung der Einzelgebäude auf Basis von digitalen Oberflächenmodellen mit einer geometrischen Auflösung von einem Meter die größeren Bebauungsstrukturen in beiden Städten weitgehend erfasst. Viele der kleineren Bebauungsstrukturen wie Einzelhäuser und Nebengebäude werden bei der Segmentierung nicht berücksichtigt, die sie das Schwellwert- oder Flächenkriterium nicht erfüllen. Der quantitative Vergleich mit Gebäudedaten des amtlichen Gebäudemodells aus dem Kataster kann in diesem Zusammenhang allerdings auch nur für eine generelle Einschätzung der Gebäudeabgrenzung dienen und keine exakten Aussagen liefern, da die Datenbasis der Referenzdaten nicht konsistent ist und der Aktualitätsstand teilweise erheblich von den Oberflächenmodellen abweicht. In München unterscheidet sich die Aktualität um vier Jahre (vgl. hierzu auch Heldens, 2010, S. 44ff.). Dennoch werden rund dreiviertel der Gebäudegrundflächen von allen Gebäuden mit einer Geschossigkeit größer als 1 erfasst, was im

Zusammenhang mit einer der Zielsetzungen der vorliegenden Arbeit, nämlich der Ableitung und Beschreibung der Bebauungsstruktur, ausreichend ist.

#### 4.2.5. Ableitung der Gebäudehöhen

Für die Berechnung der Gebäudehöhen werden den abgegrenzten Gebäudegrundflächen Höhenwerte aus den nDOMs zugewiesen. Die Höhe der Einzelgebäude ist für die Berechnung von 3D-Formmerkmalen, welche für die semantische Gebäudeklassifikation verwendet werden (vgl. folgendes Kapitel 5) und die Berechnung der Bebauungsdichte notwendig. Beide werden als Variablen mit den Surveydaten verknüpft (vgl. Kapitel 6).

Die Gebäudehöhe wird aus den Höhenwerten des nDOMs den Gebäudegrundflächen zugewiesen. Dabei wird aus den einzelnen, pixelbasierten Höhenwerten für jede Gebäudegrundfläche das arithmetische Mittel gebildet und somit jedem Gebäude ein spezifischer Höhenwert zugewiesen. Diese Datensätze, welche die Gebäudegrundflächen und einen zugehörigen Höhenwert aufweisen, werden auch als 3D-Gebäudemodell im LoD1 (*Level of Detail*) oder als „Klötzchenmodell“ bezeichnet. Eine Visualisierung der aus den Oberflächenmodellen für München und Berlin erstellten „Klötzchenmodelle“ ist in Abbildung 31 dargestellt. Sie zeigt oben den Bereich um das Münchener Zentrum in perspektivischer Darstellung aus südlicher Richtung. Im „Klötzchenmodell“ ist die vereinfachte Darstellung von komplexen Baustrukturen zu erkennen. Der untere Teil der Abbildung zeigt einen Ausschnitt der Berliner Spreeinsel aus südöstlicher Perspektive.

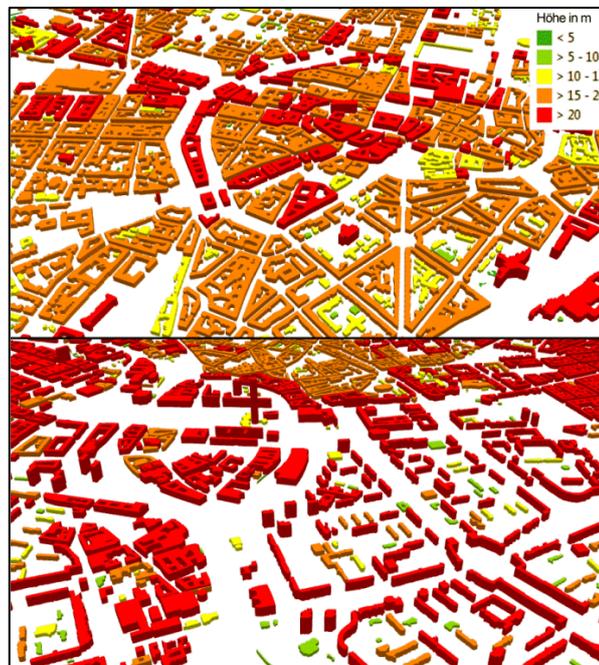


Abbildung 34: Gebäudegrundflächen und Gebäudehöhen für die Stadtzentren von München und Berlin.

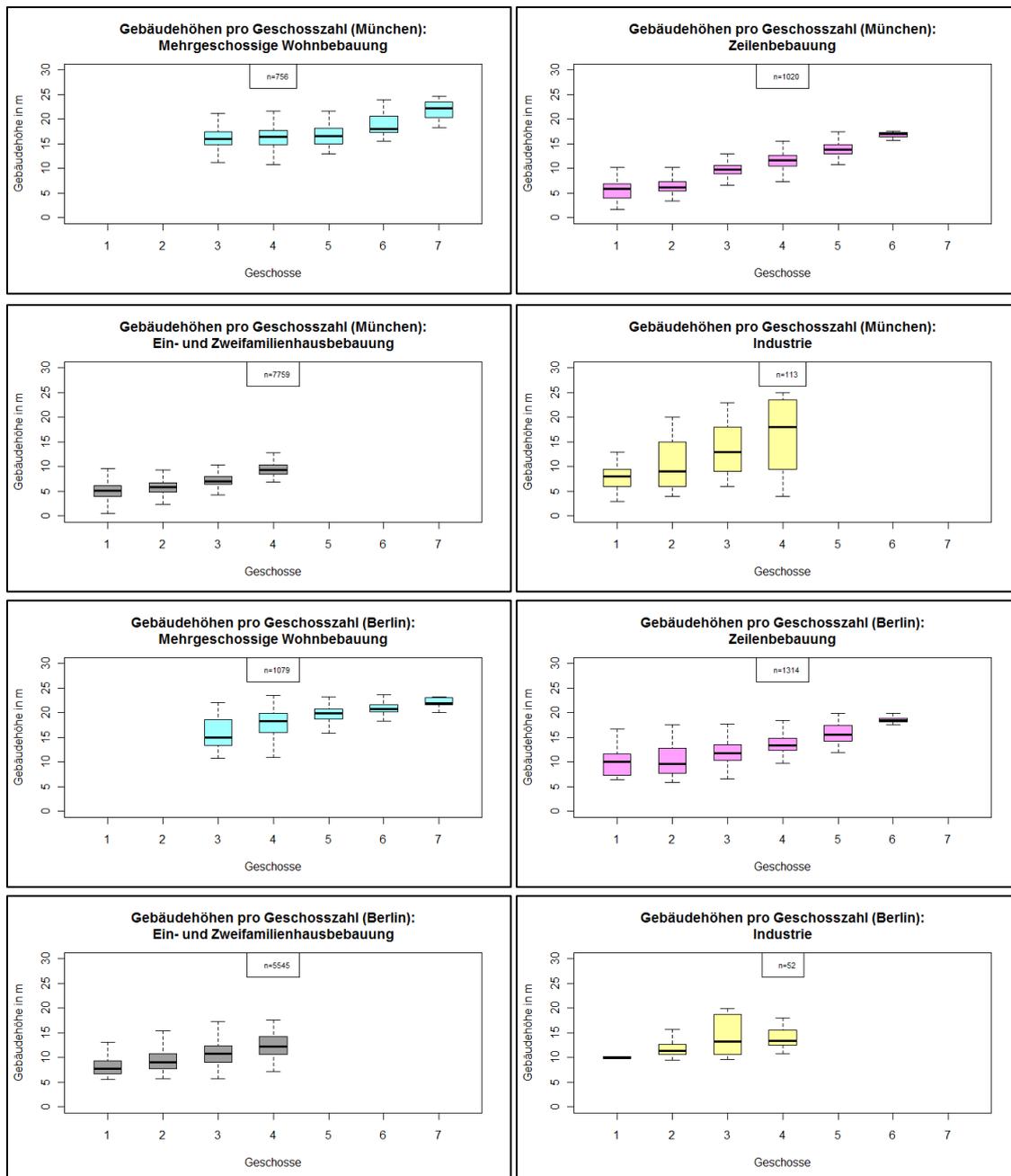
### 4.2.6. Abschätzung der Geschosszahl

Für die Berechnung von verschiedenen Maßen der baulichen Nutzung wird neben der metrischen Gebäudehöhe auch die Anzahl der Geschosse pro Gebäude benötigt. Die Bebauungsdichte und die Geschossflächendichte sind in Deutschland in der Baunutzungsverordnung (BauNVO) festgelegt. In ihr sind neben der Grundflächenzahl (GRZ) und der Geschossflächenzahl (GFZ) auch die Baumassenzahl (BMZ) definiert. Während die Grundflächenzahl (hier auch: Bebauungsdichte) den flächenmäßigen Anteil der bebauten Grundstücksfläche angibt, stellt die Baumassenzahl die Kubikmeter Baumasse je Grundstücksfläche dar. Sie wird allerdings primär für Bebauungspläne von Gewerbe- und Industriegebieten verwendet. Demgegenüber stellt die Geschossflächenzahl ein gebräuchliches Maß für die Dichte der Bebauung dar. Für ihre Berechnung wird das Verhältnis der Summe der Flächen sämtlicher Vollgeschosse pro Fläche des Grundstückes berechnet. Da im vorliegenden Fall die exakten Grundstücksgrenzen nicht bekannt sind, wird die Geschossflächendichte (GFD) als Maß für die Dichte der Bebauung verwendet. Sie berechnet sich ähnlich der GFZ, allerdings wird anstatt der Grundstücksfläche die Fläche des Baublocks als Bezugsfläche verwendet. Damit die Geschossflächendichte berechnet werden kann, wird also die Anzahl der Geschosse pro Gebäude benötigt. Vor diesem Hintergrund wird im aktuellen Abschnitt ein Ansatz untersucht, um die Geschosszahl auf Basis der abgeleiteten „Klötzchenmodelle“ abzuschätzen.

Die Anzahl der Geschosse für die einzelnen Gebäude im „Klötzchenmodell“ kann nicht direkt aus den Oberflächenmodellen gemessen werden. Es besteht allerdings ein ermittelbarer Zusammenhang zwischen der metrischen Gebäudehöhe und der Geschosszahl, denn je höher ein Gebäude ist, desto mehr Geschosse hat es in der Regel. Diese Hypothese kann für die Erstellung eines empirischen Modells genutzt werden, welches den statistischen Zusammenhang zwischen Gebäudehöhe und Geschosszahl quantifiziert. Für die Herstellung dieses statistischen Zusammenhangs werden für eine ausreichende Anzahl an Gebäuden sowohl die Gebäudehöhe als auch die Geschosszahl untersucht. Die Information über die Gebäudehöhe ist in den abgeleiteten Gebäuden aus dem digitalen Oberflächenmodell enthalten („Klötzchenmodell“) und die Information über die Geschosszahl ist im amtlichen Gebäudemodell aus dem Kataster enthalten (vgl. Abschnitt 3.3.2.4). Zusätzlich ist in den Referenzdaten auch die Information der vier Gebäudetypen *mehrgeschossige Wohnbebauung*, *Zeilenbebauung*, *Ein- und Zweifamilienhausbebauung* und *industrielle Bebauung* enthalten (vgl. Abbildung 18, S.60). Somit kann für jede Stadt und für vier verschiedene Gebäudetypen der Zusammenhang zwischen Gebäudehöhe und Geschosszahl untersucht werden. Dies ist für die Generierung eines Basiswissens notwendig und auf Basis dieses Wissens wird die empirische Funktion ermittelt. Die Vorgehensweise verfährt wie folgt: zunächst werden die Abhängigkeiten zwischen Gebäudehöhe und Geschossigkeit für beide Städte anhand von deskriptiven Statistiken untersucht und in einem weiteren Schritt, die Funktion aufgestellt.

Dafür werden zuerst die Verteilungen der Gebäudehöhen für die einzelnen Gebäudetypen betrachtet. Dafür werden *Boxplots* erstellt, welche für die Gebäudetypen die Verteilungen der Gebäudehöhen für jede Geschoszahl darstellen. Boxplots sind graphische Darstellungen, anhand welcher sehr einfach die Verteilung und die Lage der Werte einer Beobachtungsreihe interpretiert werden können. Sie setzen sich zusammen aus der *Box* und den *Antennen*. Dabei repräsentiert die Box den Interquartilsbereich, also die 50 % aller vorkommenden Werte zwischen dem ersten und dem dritten Quartil. Die Linie innerhalb der Box stellt den Median der Werte dar. Die Antennen zeigen den 1,5-fachen Interquartilsabstand an. Weitere Informationen über Boxplots können McGill, Tukey und Larsen (1978) entnommen werden.

In Abbildung 35 sind die Boxplots für die vier Gebäudeklassen *mehrgeschossige Wohnbebauung*, *Zeilenbebauung* und *Ein- und Zweifamilienhausbebauung* und *Industrie* für die beiden Städte Berlin und München dargestellt. Es ist zu erkennen, dass die Gebäudehöhe mit der Geschossigkeit zunimmt. Besonders deutlich ist der Zusammenhang für die Zeilenbebauung zu sehen. Für die Gebäude der Klasse *mehrgeschossige Wohnbebauung* ist zu erkennen, dass die Gebäude mit 3, 4 und 5 Geschossen in München sehr ähnliche Gebäudehöhen aufweisen, in Berlin steigt die Gebäudehöhe mit jedem Geschoss zunächst stärker an, um dann abzuflachen. Auch ist die Streuung für diese Klasse in Berlin vor allem für 3-4 geschossige Gebäude höher als in München. Diese Gebäudeklasse ist ausschließlich durch Block- und Blockrandbebauungen besetzt. Diese Gebäudetypen sind vorwiegend zur Gründerzeit im Stil des Historismus entstanden, also in der zweiten Hälfte des 19. und auch zu Beginn des 20. Jahrhunderts. Charakteristisch für diesen Baustil sind vier bis sechsgeschossige Gebäude mit großen Raumhöhen und hohen Dächern, welche gelegentlich zu Wohnräumen ausgebaut werden. Die Streuung der Gebäudehöhen für die einzelnen Geschoszahlen ist für die ersten drei Gebäudetypen gering, für den Gebäudetyp *Industrie* ist die Streuung allerdings größer. Dies liegt daran, dass Hallen und Fabriksgebäude sehr unterschiedliche Höhen und Geschoszahlen aufweisen.

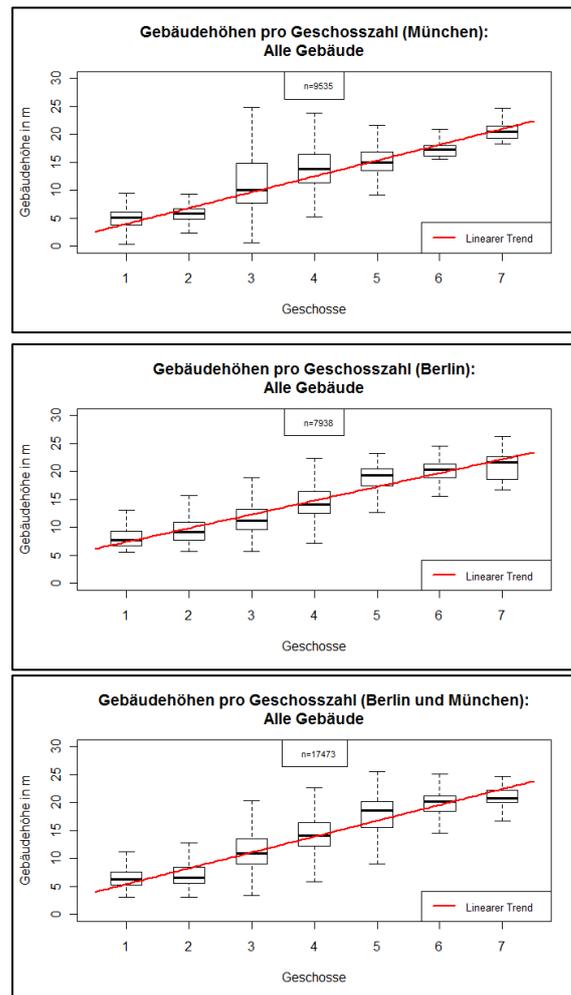


**Abbildung 35:** Gebäudetypenabhängige Darstellung der Gebäudehöhen pro Geschoszahl für München und Berlin.

Anhand dieses deskriptiven Vergleichs für die einzelnen Gebäudetypen kann die Hypothese, dass die Geschoszahl mit der Gebäudehöhe zunimmt, bestätigt werden. Es kann auch beobachtet werden, dass der Zusammenhang für alle vier Gebäudetypen in beiden Städten tendenziell ähnlich ist, aber nicht identisch ist. Es bestehen also kleine Unterschiede in

der Bebauungsstruktur zwischen München und Berlin. Entsprechend lässt sich daraus ableiten, dass es keine einzelne Funktion geben kann, welche für alle Städte eines Kulturkreises und alle auftretenden Gebäudetypen den Zusammenhang zwischen Geschossigkeit und Gebäudehöhe ideal beschreibt. In Anbetracht dieser Erkenntnis kann konkludiert werden, dass für die exakte Abschätzung der Geschossigkeit von Gebäuden in Städten der empirische Zusammenhang für jede Stadt neu berechnet werden muss, oder eine Funktion ermittelt wird, welche den grundsätzlichen Trend im empirisch ermittelten Zusammenhang beschreibt. Eine solche Funktion kann auf allen Städten angewendet werden, wobei berücksichtigt werden muss, dass sie keine idealtypische, lokal angepasste Funktion beschreibt, sondern eine Annäherung an eine allgemeine Funktion darstellt und es dabei zu lokalen Über- bzw. Unterschätzungen der tatsächlichen Geschossigkeit kommt.

Für die Berechnung dieser Funktion werden zunächst wie oben die Verteilungen der Gebäudehöhen für die einzelnen Geschosszahlen zwischen 1 und 7 betrachtet. In Abbildung 36 sind daher wiederum Boxplots für alle Gebäude der beiden Städte München und Berlin dargestellt. Sie zeigt neben den herkömmlichen Elementen von Boxplots auch den linearen Trend, der den linearen Zusammenhang zwischen der Geschosszahl und der Gebäudehöhe verdeutlicht. Der lineare Trend ist eine auf Basis einer linearen Regression berechnete Funktion, welche den statistischen Zusammenhang zwischen den beiden Größen *Gebäudehöhe* und *Geschosszahl* ausdrückt. Es ist auch denkbar, einen nichtlinearen Zusammenhang im Sinne eines Polynoms höherer Ordnung zu schätzen, der erklärte Anteil der Varianz steigt bei einem Trend 4. Ordnung aber lediglich um 5 % (vgl. Anhang 5). Vor diesem Hintergrund und im Sinne einer einfachen Übertragbarkeit und einer logischen Hypothese, nämlich dass ein Gebäude höher ist, je mehr Geschosse es aufweist, stellt der lineare Zusammenhang ein probates Mittel für die Geschosszahlenabschätzung aus der Gebäudehöhe für die Städte München und Berlin dar.



**Abbildung 36:** Darstellung der Gebäudehöhen pro Geschosszahl für alle Gebäude.

Aus der allgemeinen Formel für die lineare Regression (vgl. zu linearer Regression Abschnitt 6.3.1) kann folgender linearer Zusammenhang aufgestellt werden, in dem  $x$  und  $k$  die zu ermittelnden Unbekannten sind:

$$(10) \quad \text{Gebäudehöhe} = \text{Geschosszahl} \cdot x + k$$

Die Gebäudehöhe und die Geschosszahl sind für die Ermittlung dieses linearen Zusammenhangs bekannt und der Trend für München und Berlin kann durch Schätzung der kleinsten Quadrate ermittelt werden (Backhaus et al., 2003):

$$(11) \quad h = g \cdot 2,841 + 1,997$$

und weiter für die Berechnung der Geschosszahl umgeformt werden:

$$(12) \quad g = \frac{h - 1,997}{2,841}$$

Auf Basis dieser Funktion werden die Geschossezahlen für alle abgeleiteten Gebäude durch Einsetzen der metrischen Gebäudehöhe  $h$  geschätzt.

#### 4.2.7. Zusammenfassung für die Gebäudestruktur

Das in Abschnitt 4.2 dargestellte Verfahren zur Ableitung der Gebäudestruktur stellt den ersten Schritt für die Erstellung der Kontextmerkmale dar, welche mit den Surveydaten verknüpft werden. In diesem Abschnitt wurde zunächst das Verfahren beschrieben, wie aus digitalen Oberflächenmodellen die Gebäudegrundflächen mit objektbasierten Bildverarbeitungsmethoden extrahiert werden. Die Ableitung der Gebäudegrundflächen wurde mit amtlichen Referenzdaten verglichen und zeigt Genauigkeiten zwischen fast 60 % und über 80 %. Bei der Gebäudeabgrenzung werden vor allem sehr große Gebäude erfasst, flächenmäßig kleine und niedrige Gebäude wie kleinere Einfamilienhäuser, Nebengebäude oder Garagen weisen in den DOMs keinen ausreichenden Kontrast der Grauwerte auf, um von der Nachbarschaft als eigenständiges Objekt separiert zu werden.

Den abgeleiteten Gebäudegrundflächen wurden auf Basis der Höhen in den normalisierten digitalen Oberflächenmodellen Gebäudehöhenwerte zugewiesen. Diese Höhenwerte sind einerseits für die semantische Klassifikation von Gebäudetypen auf Basis von Formmerkmalen relevant (vgl. Abschnitt 5.1) und andererseits für die Berechnung der Geschossezahl der Einzelgebäude. Diese Geschossezahl wird benötigt, um die Geschossflächendichte zu berechnen (vgl. Abschnitt 5.2). Die Geschossezahl ist nicht in den Daten der digitalen Oberflächenmodelle enthalten, es kann allerdings auf Basis eines empirisch ermittelten statistischen Zusammenhanges eine Beziehung zwischen den beiden Größen hergestellt werden. Dafür wurde für über 10000 Gebäude, für die sowohl die Gebäudehöhe als auch die Geschossezahl bekannt sind, ein linearer Trend berechnet und die Geschossezahl für sämtliche Gebäude geschätzt.

### 4.3. Ableitung der urbanen Landbedeckung

Die Landbedeckungsklassen, welche aus den Satellitenbilddaten abgeleitet werden, beschreiben neben den klassifizierten *Gebäuden* (vgl. Abschnitt 4.2) die physische Ausstattung des städtischen Raumes. Im aktuellen Abschnitt wird daher die Klassifikation der *Vegetation* und der *restlichen versiegelten Flächen* ohne Gebäude beschrieben. Dazu gehören versiegelte Freiflächen wie Parkplätze und Straßen. Da aber im urbanen Raum neben diesen drei Landbedeckungsarten auch noch weitere Klassen vorkommen, werden zusätzlich die Klassen *Oberflächengewässer* und *offener Boden* klassifiziert. Sie werden allerdings nicht direkt für die Verknüpfung mit den SOEP/BASE-II-Daten verwendet, sondern dienen dem Zweck einer differenzierten und flächendeckenden Landbedeckungsklassifikation. Die klassifizierten Flächen über die urbane Landbedeckung repräsentieren die Basisinformationen über die urbane Struktur. Diese Informationen werden im folgenden Kapitel 5 auf der räumlichen Ebene des Baublocks aggregiert und stellen die räumlichen Kontextmerkmale dar, welche mit den Surveydaten verknüpft werden (vgl. Abschnitt 3.1.2).

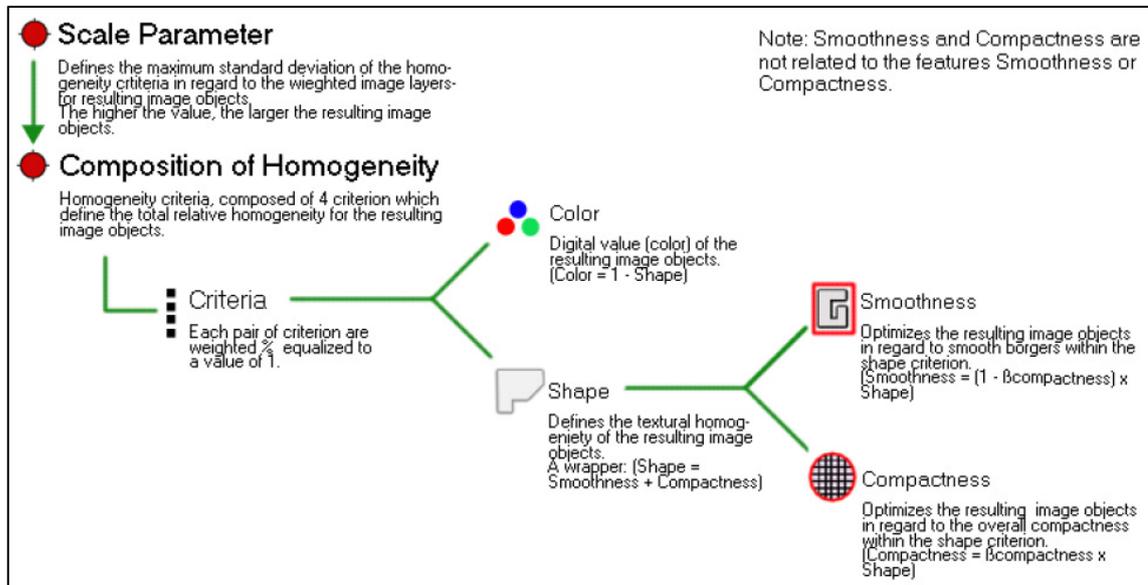
Die Landbedeckung wird im Rahmen eines hierarchischen, objektbasierten Klassifikationsablaufes abgeleitet. Das Verfahren ist deshalb hierarchisch, weil die einzelnen Landbedeckungsarten nacheinander klassifiziert werden. Somit stehen am Beginn der Klassifikation die im vorangegangenen Abschnitt bereits abgeleiteten Klassen *Gebäude* und die *Vegetation*, welche für die Gebäudeableitung erstellt wurde.

#### 4.3.1. Objektbasierte Bildklassifikation

Der aktuelle Abschnitt fokussiert auf die Zuweisung der Flächen, welche weder zur Klasse *Gebäude* noch zur Klasse *Vegetation* gehören. Diese Flächen werden zunächst mit einem Bildsegmentierungsverfahren in Objekte segmentiert und danach klassifiziert. Der objektbasierte Auswertevorgang wurde in der Arbeit in das Softwarepaket eCognition© implementiert (Trimble, 2010a) und wird im Folgenden im Detail beschrieben.

Das verwendete Segmentierungsverfahren *Multiresolutionsegmentierung* stellt eines der gängigsten Segmentierungsverfahren dar. Es teilt ein Bild in zusammenhängende Objekte auf Basis eines spektralen Homogenitätsschwellwertes auf, welcher durch den Nutzer festgelegt wird. Dabei werden benachbarte Pixel so lange miteinander fusioniert, bis der Homogenitätsschwellwert überschritten wird. Der Homogenitätsschwellwert setzt sich aus den Parametern *color* und *shape* zusammen, welche neben dem *scale parameter* definiert werden müssen. Der *scale parameter* legt die maximale Standardabweichung des Homogenitätskriteriums der resultierenden Bildobjekte fest und bestimmt folglich auch die Größe der resultierenden Objekte. Die Festlegung des Homogenitätsschwellwertes und des *scale*

*parameters* wird entweder durch den Nutzer in einem iterativen, empirischen Verfahren gefunden oder durch automatische Verfahren (z.B. Zhang et al., 2010; LVC, 2008). Der Zusammenhang zwischen den relevanten Kriterien und Parametern für die Multiresolutionsegmentierung in eCognition© ist in Abbildung 37 dargestellt.



**Abbildung 37:** Zusammenhang der Parameter für die Multiresolutionsegmentierung in eCognition.

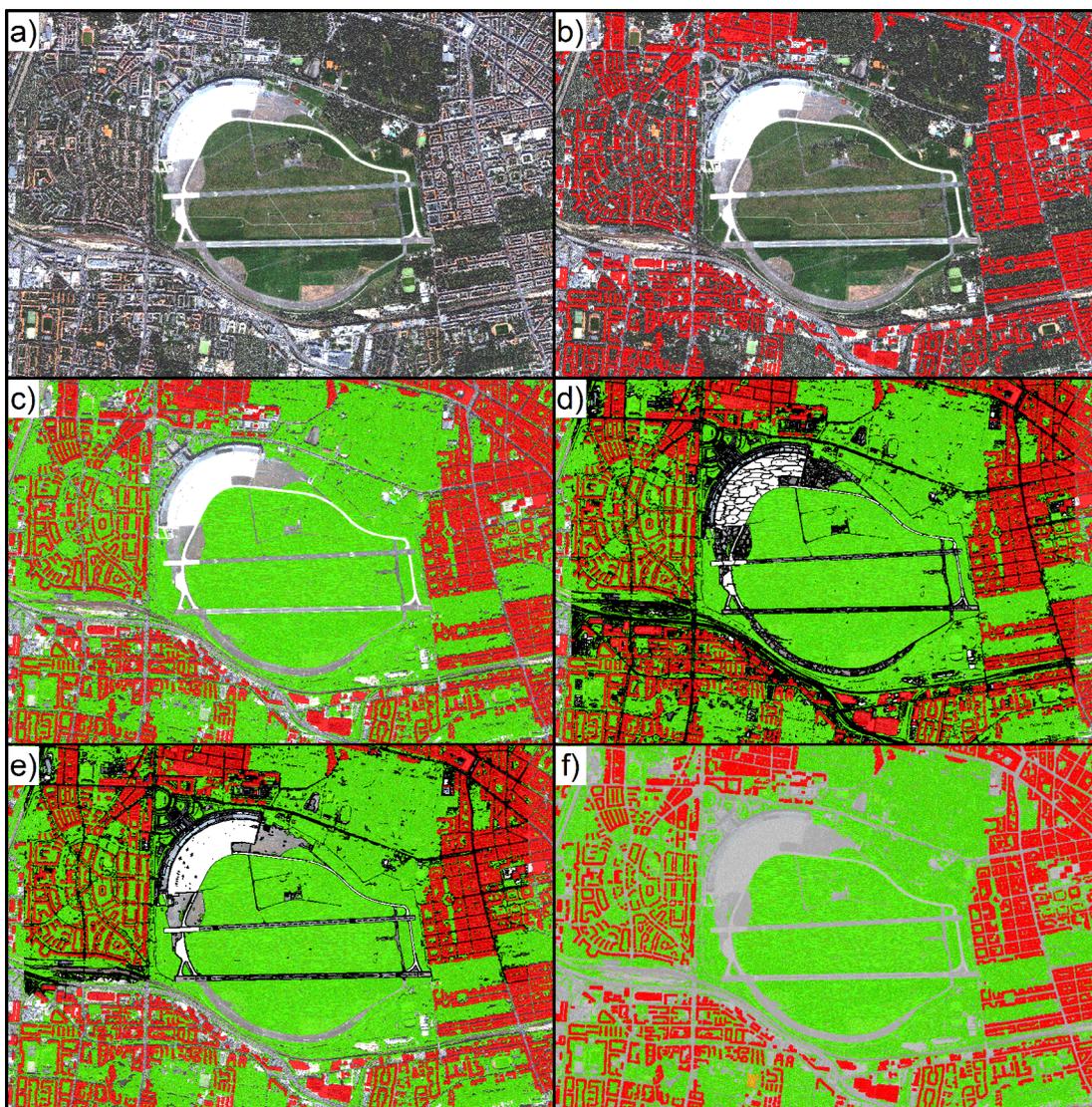
(Quelle: Trimble, 2010a)

Ausgehend von den bereits klassifizierten Flächen, werden die verbleibenden Flächen auf Basis des Vegetationsindex NDVI segmentiert. Im NDVI sind sowohl Gewässerflächen, versiegelte Flächen und Vegetationsflächen sehr gut zu unterscheiden (vgl. Gleichung (1), S. 71). Die Flächen werden in diesem Segmentierungsschritt übersegmentiert, das bedeutet, dass die gebildeten Objekte kleiner sind, als die Objekte in der Natur. Dies erlaubt eine kontrollierte Weiterverarbeitung der Objekte. So werden im folgenden Schritt spektral sehr ähnliche und nebeneinanderliegende Objekte zu größeren Objekten zusammengefasst. Somit existieren große, spektral homogene Objekte und kleine, spektral heterogene Objekte. Sie werden im letzten Schritt einer der Klassen *versiegelte Flächen*, *Oberflächengewässer* und *offener Boden* zugewiesen.

Für die Klassifikation der erstellten Objekte wurde in eCognition© ein Fuzzy Logik-Ansatz verwendet, bei welchem die Zuweisung der Objekte zu den thematischen Klassen auf Basis eines individuellen Zugehörigkeitswertes (*membership value*) durchgeführt wird. Die Fuzzy Logik Methode wurde entwickelt, um mathematisch präzise mit Ungenauigkeiten umzugehen (Zadeh, 1965). Ungenauigkeiten sind in diesem Zusammenhang als unscharf definierte Zugehörigkeiten eines Elements zu einer Gruppe zu verstehen. Somit können ungenaue Zugehörigkeiten wie „ein bisschen“ oder „in der Nähe von“ zwischen *wahr* oder *falsch* adressiert

werden. Eine gute Einleitung und Übersicht über die Fuzzy Logik Theorie und ihre Anwendung für die Klassifikation von Fernerkundungsdaten findet sich bei Bochow (2009).

Die Klassifikation verfährt in einem iterativen Verfahren. Dabei werden zunächst *Saatobjekte* identifiziert, welche aufgrund ihrer Objekteigenschaften eindeutig einer Klasse zugewiesen werden können. Im nächsten Schritt werden angrenzende Objekte klassifiziert, die ähnliche Objekteigenschaften wie die Saatobjekte aufweisen. Dieser Schritt wird solange wiederholt, bis keine Objekte mehr den Kriterien für die Klassifikation entsprechen. Nicht klassifizierte Objekte wurden für eine vollständige Klassifikation manuell einer der Landbedeckungsklassen zugewiesen. Die einzelnen Schritte des beschriebenen Ablaufs sind in Abbildung 38 dargestellt.



**Abbildung 38:** Die einzelnen Ablaufschritte zur Klassifikation der Landbedeckung: a) Satellitenbild b) Überlagerung mit der Gebäudemaske c) Überlagerung mit der Vegetationsmaske d) Multiresolutionsegmentierung e) Fusionierung von benachbarten Objekten f) Klassifikation.

### 4.3.2. Klassifikationsergebnisse

Der oben beschriebene Ablauf für die Klassifikation der Landbedeckung wurde auf die Daten der beiden Untersuchungsgebiete Berlin und München angewendet. Die Ergebnisse der Landbedeckungsklassifikationen für die beiden Städte sind in Abbildung 39 und Abbildung 40 dargestellt. Sie zeigen das physische Erscheinungsbild der Städte in den fünf thematischen Klassen *Gebäude*, *Vegetation*, *versiegelte Flächen (ohne Gebäude)*, *offener Boden* und *Oberflächengewässer*. Die Anordnung dieser physischen Elemente spiegelt die städtische Struktur der untersuchten Städte im Wechselspiel zwischen bebauten Flächen, Freiflächen und Vegetationsbereichen wider. Charakteristisch für Berlin (Abbildung 39) ist der Tiergarten als städtisches Naherholungsgebiet in Zentrumsnähe, neben weiteren, ausgedehnten Grünflächen zu erkennen. Daneben ist der Wandel von dichter Bebauung mit hohem Versiegelungsanteil in den innerstädtischen Bereichen hin zu lockeren Bauungsformen mit starker Durchgrünung in den städtischen Randbereichen zu erkennen.

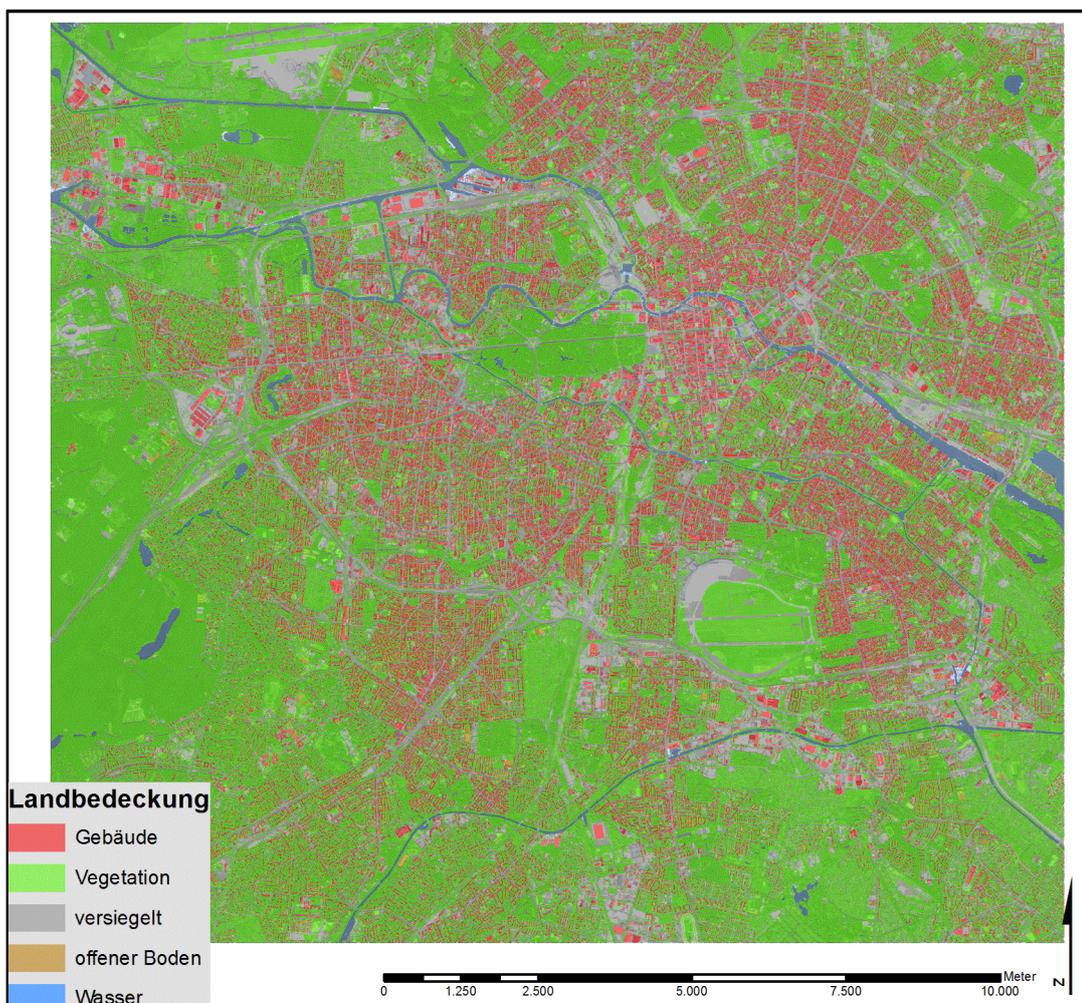


Abbildung 39: Ergebnis der Landbedeckungsklassifikation für Berlin.

Das Klassifikationsergebnis für München (Abbildung 40) zeigt die dichte Bebauung in Zentrumsnähe, welche durch Parks und Grünflächen aufgelockert wird. Vor allem der Englische Garten als große Grünanlage nördlich des Stadtzentrums und die Grünflächen entlang der Isar im Süden des Stadtzentrums treten hervor. Im Gegensatz dazu ist vor allem der westliche Bereich der Stadt durch versiegelte Flächen des Eisenbahnverkehrs gekennzeichnet. An der baulichen Struktur können die zentrumsnahen Wohnbaugebiete von den im Norden der Stadt angesiedelten, großen Industrie- und Gewerbeflächen identifiziert werden.

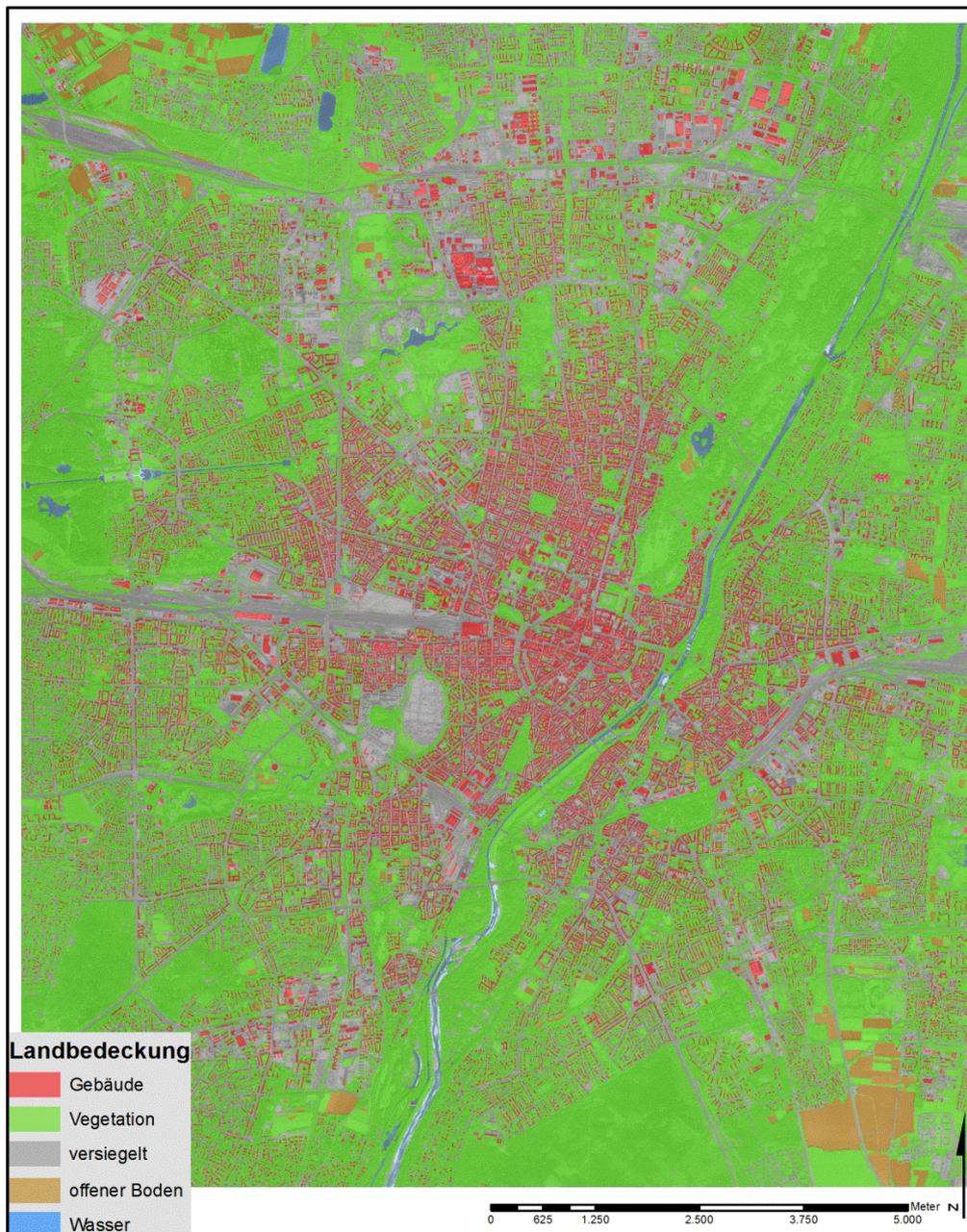


Abbildung 40: Ergebnis der Landbedeckungsklassifikation für München.

Diese Klassifikationen stellen nun die Grundlage für die aggregierten Kontextmerkmale dar, welche im folgenden Kapitel 5 beschrieben werden. Die Landbedeckungsklassen *Vegetation* und *versiegelte Flächen (ohne Gebäude)* werden im Rahmen der in Abschnitt 3.1.2 beschriebenen räumlichen Bezugsebene auf der Ebene des Baublocks aggregiert und mit den Surveydaten verknüpft.

### 4.3.3. Genauigkeitsabschätzung

Für die beiden Landbedeckungsklassifikationen wurde eine Genauigkeitsabschätzung auf Basis von Referenzpunkten durchgeführt. Die Punkte wurden durch ein stratifiziertes, randomisiertes Samplingverfahren in den Satellitenbildern, welche die Referenz darstellen, verteilt und für alle Punkte die zugehörige Landbedeckungsklasse erfasst. Die Anzahl der Referenzpunkte richtet sich nach dem Ansatz von Tortora (1978), der bei Congalton und Green (2008) erläutert wird: die Anzahl der benötigten Referenzpunkte für eine Klassifikation mit mehr als zwei Klassen wird dabei über eine multinomiale Verteilung der Klassenpixel berechnet:

$$(13) \quad n = B \Pi_i \cdot (1 - \Pi_i) / b_i^2$$

wobei  $n$  die Anzahl der gesuchten Referenzpunkte ist und  $B$  das  $(1-\alpha/k)$ -Quantil einer  $X^2$ -Verteilung mit einem Freiheitsgrad ist. Dabei ist  $k$  die Anzahl der Klassen und der  $\alpha$ -Fehler wurde auf 5 % festgelegt. Demnach erhält man für  $X^2_{(1,0,99)}=6,63$ .  $\Pi_i$  ist der Anteil der Pixel in der  $i$ -ten Klasse und  $b_i$  ist der gewünschte Fehler für die Klasse. Wenn die Anzahl der Pixel der einzelnen Klassen nicht gleichverteilt ist, wird der Flächenanteil der Klasse mit dem höchsten Flächenanteil gewählt. In der Landbedeckungsklassifikation stellt die Klasse *Vegetation* mit 55 % den höchsten Flächenanteil dar. Trägt man die bekannten Werte in die Formel ein, erhält man:

$$(14) \quad n = 6,63 \cdot (0,55) \cdot (1 - 0,55) / 0,05^2 = 656$$

Die Summe der Referenzpunkte aus oben ermittelter Formel beträgt demnach 656. Wenn die Punkte auf die einzelnen Landbedeckungsklassen verteilt werden, ergibt sich eine mittlere Anzahl von etwa 131 Referenzpunkten pro Landbedeckungsklasse. Im Rahmen der Genauigkeitsabschätzung wurde die Anzahl der Punkte auf 200 aufgerundet. Dadurch verringert sich der  $\alpha$ -Fehler auf etwa 4 %.

Für jede der Klassen wurden die Punkte zwischen der Referenz und der Klassifikation miteinander verglichen. Die Validierungen sind als Konfusionsmatrizen (vgl. Tabelle 7, S. 58) in Tabelle 11 und Tabelle 12 für Berlin und München dargestellt und weisen die Übereinstimmungen der untersuchten Punkte aus. In den Tabellen sind auch für jede der fünf thematischen Klassen die Nutzer- und die Herstellergenauigkeiten angegeben. Die beiden Klassen *Wasser* und *Vegetation* weisen sowohl hohe Nutzer- als auch Herstellergenauigkeiten

auf, da sie sich aufgrund ihrer spektralen Merkmale deutlich von den anderen Klassen unterscheiden. Die Klasse *Gebäude* weist hohe Nutzer-, aber geringere Herstellergenauigkeiten auf. Dies ist dadurch begründet, dass in der Gebäudemasken nicht alle Gebäude erfasst wurden (vgl. Tabelle 9, S. 80) und die nicht erfassten Gebäude als *versiegelt* klassifiziert wurden. Die geringeren Herstellergenauigkeiten in der Klasse *offener Boden* resultieren ebenfalls aus der spektralen Ähnlichkeit zwischen *offenem Boden* und den *versiegelten Flächen (ohne Gebäude)*. Zusätzlich zu diesen beiden Maßen ist auch die Gesamtgenauigkeit angegeben. Sie berechnet sich aus dem Verhältnis der Summe aller Übereinstimmungen zur Gesamtzahl der Beobachtungen und beträgt für beide Klassifikationen über 90%. Zusätzlich ist der Kappakoeffizient als Maß zur Güte des Klassifikationsergebnisses angegeben (Cohen, 1960). Er gibt an, inwiefern sich die Werte aus der Konfusionsmatrix von einem statistischen Zufallsergebnis unterscheiden. Der Kappakoeffizient berechnet sich aus Gleichung (4) auf Seite 59. Er zeigt für beide Klassifikationen Werte von 0,88 bzw. 0,89. Kappa-Werte über 0,75 bzw. über 0,8 werden in der Literatur als sehr gute Übereinstimmungen bezeichnet (Greve und Wentura, 1997 bzw. Landis und Koch, 1977).

**Tabelle 11:** Genauigkeitsabschätzung für die Klassifikation von Berlin.

Referenzdaten							
Klassifikation	Wasser	Vegetation	versiegelt (ohne Gebäude)	offener Boden	Gebäude	Summe	Nutzergen.%
Wasser	188	2	8	-	2	200	94,00
Vegetation	-	188	6	4	2	200	94,00
versiegelt (ohne Gebäude)	1	2	172	1	24	200	86,00
offener Boden	-	6	16	175	3	200	87,50
Gebäude	-	-	3	-	197	200	98,50
Summe	189	198	205	180	228	1000	
Herstellergen.%	99,47	94,95	83,90	97,22	84,55		<b>Gesamt: 92,00</b> <b>Kappa: 0,89</b>

**Tabelle 12:** Genauigkeitsabschätzung für die Klassifikation von München.

Referenzdaten							
Klassifikation	Wasser	Vegetation	versiegelt (ohne Gebäude)	offener Boden	Gebäude	Summe	Nutzergen.%
Wasser	184	5	11	-	-	200	92,00
Vegetation	-	187	6	1	6	200	93,50
versiegelt (ohne Gebäude)	1	2	170	8	19	200	85,00
offener Boden	-	13	19	167	1	200	83,50
Gebäude	-	4	1	1	194	200	97,00
Summe	185	211	207	177	220	1000	
Herstellergen.%	99,46	88,63	82,13	94,35	88,18		<b>Gesamt: 90,20</b> <b>Kappa: 0,88</b>

#### **4.4. Zusammenfassung zur Erfassung der Gebäudestruktur und der Landbedeckung**

In diesem Kapitel wurden die im Rahmen der Arbeit entwickelten Verfahren beschrieben, anhand welcher die Gebäudestruktur und die urbane Landbedeckung aus digitalen Oberflächenmodellen und optischen Satellitenbilddaten abgeleitet werden. Die Produkte aus den angewendeten Verfahren ermöglichen eine detaillierte Beschreibung der kleinräumigen städtischen Struktur und stellen die Grundlage für die Verknüpfung mit den Surveydaten dar. Sie beinhalten eine Beschreibung der physischen Bebauungsstruktur als Klötzchenmodell und eine Klassifikation der für die Verknüpfung relevanten thematischen Klassen *Vegetation* und *versiegelte Flächen*.

Für die abgeleiteten Einzelgebäude wurde auf Basis der Gebäudehöhe anhand eines empirisch ermittelten Regressionsmodells die Geschosszahl abgeschätzt, welche für die Berechnung der Bebauungsdichte notwendig ist. Im folgenden Kapitel wird die Berechnung der Kontextmerkmale auf der Ebene des Baublocks beschrieben, welche im letzten Kapitel mit den Surveydaten verknüpft werden.



# Kapitel 5

## Gebäudeklassifikation und Aggregation von Kontextmerkmalen

*“Cities are made up of social and physical features.”*

Schell und Ulijaszek (1999)

In diesem Kapitel wird aufbauend auf die Klassifikationen der städtischen Struktur und der Landbedeckung aus dem vorangehenden Kapitel die Klassifikation der Gebäude und die Aggregation der Informationen zu den Baublöcken beschrieben (vgl. Abschnitt 3.1.2). Für die Verknüpfung werden die kleinräumigen Informationen aus dem 3D-Gebäudemodell und der Landbedeckungsklassifikation für die beiden Untersuchungsgebiete München und Berlin auf der Ebene des Baublocks zusammengefasst (vgl. Abbildung 11, S. 41). Wie in der Zielsetzung der Arbeit in Abschnitt 1.3 dargestellt, sind die kleinräumigen Nachbarschaftsmerkmale:

- die dominierende Gebäudeklasse,
- die Bebauungsdichte,
- die Geschossflächendichte,
- der Vegetationsanteil,
- der Versiegelungsanteil,
- und die Lage,

welche in den folgenden Abschnitten beschrieben werden.

### 5.1. Klassifikation der dominierenden Gebäudeklasse

Wie bereits dargestellt gehört zur Beschreibung der Nachbarschaft in städtischen Räumen neben quantitativen Merkmalen über die Gebäudestruktur auch die Beschreibung von semantischen Gebäudeklassen, da die Gebäudeklasse in der Verknüpfung mit sozialwissenschaftlichen Daten einen zusätzlichen Erklärungsgehalt in der Interpretation sozialwissenschaftlicher Fragestellungen darstellt.

Die Gebäudeklassen werden zunächst für jedes Einzelgebäude auf Basis von Formmerkmalen bestimmt und in der Folge auf der Ebene des Baublocks zusammengefasst. Die Formmerkmale bestehen aus den elementaren physischen Gebäudeeigenschaften wie *Form* und *Größe* und werden auf Basis des 3D-Gebäudemodells berechnet. Im Rahmen der Arbeit werden sechs unterschiedliche Gebäudeklassen abgeleitet. Sie stellen Grundtypen der Bebauungsstrukturen in deutschen Städten dar, da sie für bestimmte Stadtentwicklungsphasen charakteristische Formen aufweisen. Die Bebauungstypen sind:

- Einzel- und Doppelhausbebauung
- Zeilen-/Reihenhausbebauung
- Mehrgeschossige Wohnbebauung (Block-/Blockrandbebauung)
- Hochhausbebauung
- Industriell bzw. gewerblich genutzte Bebauung
- Sonstige.

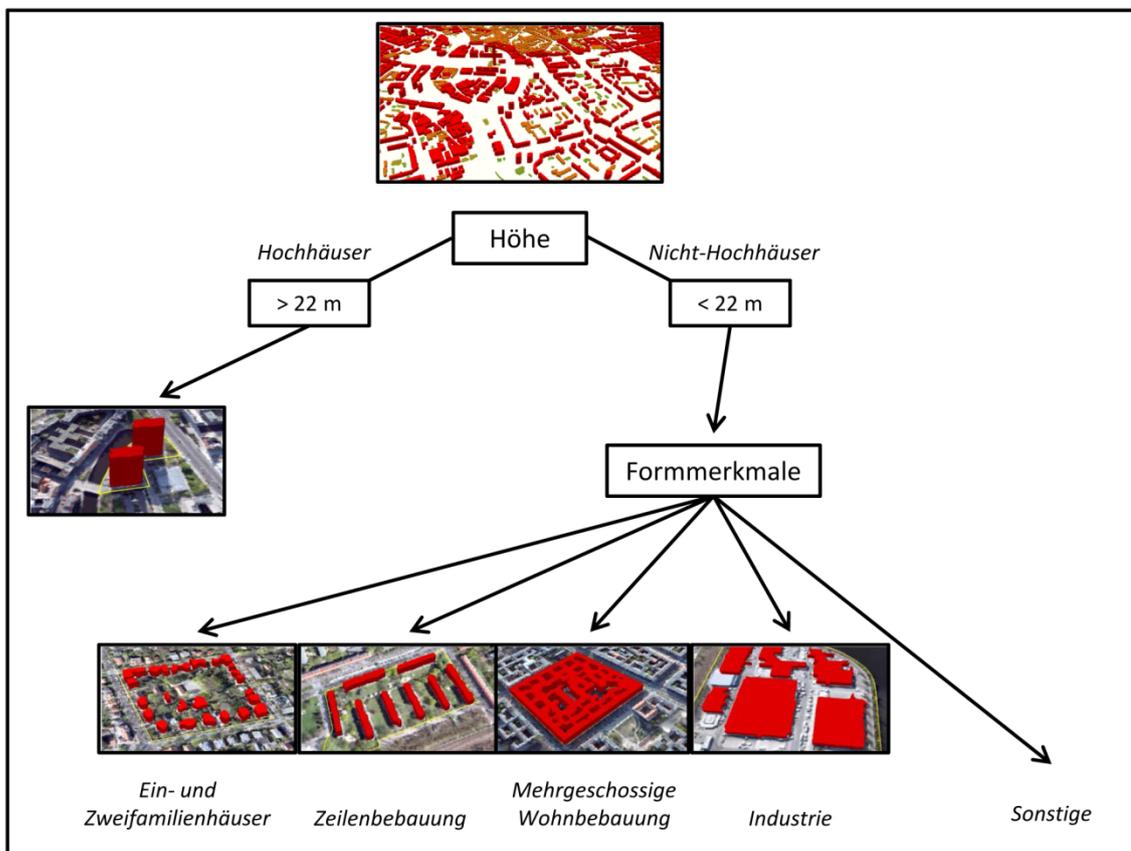
Eine Visualisierung dieser unterschiedlichen Gebäudeklassen anhand der Daten des 3D-Gebäudemodells und ein Vergleich mit Schrägluftbildern ist in der oben dargestellten Reihenfolge in Abbildung 41 dargestellt.



**Abbildung 41:** Vergleich der fünf unterschiedlichen Gebäudetypen zwischen Schrägluftbildaufnahme (links) und 3D-Gebäudemodell. (Schrägluftbilder: © Microsoft).

Von oben nach unten: Einzel- und Doppelhausbebauung, Zeilen-/Reihenhausbebauung, Mehrgeschossige Wohnbebauung (Block-/Blockrandbebauung), Hochhausbebauung und industriell bzw. gewerblich genutzte Bebauung.

Der Arbeitsablauf für die Klassifikation der Gebäudetypen ist hierarchisch in drei Schritte gegliedert und in Abbildung 42 dargestellt. Im ersten Schritt werden die zu klassifizierenden Gebäude nach ihrer Höhe in *Hochhäuser* bzw. *Nicht-Hochhäuser* unterteilt. Hochhäuser sind in der *Hochhausverordnung* definiert als Gebäude, in welchen der Fußboden für das höchste Geschoss eine größere Höhe aufweist als 22 m (MBO, 2002). Im zweiten Schritt der Klassifikation werden die verbleibenden vier Gebäudetypen unter 22 m nach weiteren Formmerkmalen in einem überwachten Klassifikationsverfahren klassifiziert. Die verbleibenden Gebäude, welche keinem der Gebäudetypen zugeordnet werden können, werden als Klasse *Sonstige* ausgewiesen.



**Abbildung 42:** Hierarchischer Ablauf für die Klassifikation der Gebäude.

Die Klassifikation der Einzelgebäude in die vier Gebäudeklassen basiert im zweiten Schritt auf Formmerkmalen. Diese beschreiben die Gebäudephysiognomie in quantifizierbaren, geometrischen Maßzahlen. Wie in Abschnitt 2.2.4 dargelegt, stützen sich aktuelle Arbeiten über die Klassifikation von Einzelgebäuden auf bestimmte gebäudespezifische Merkmale. Die Diversität dieser Merkmale in den Arbeiten ist allerdings sehr hoch und es besteht noch kein standardisiertes Merkmalsset für die objektive Analyse von Bebauungsstrukturen (Herold, 2004). Vor diesem Hintergrund werden verschiedene Formmerkmale aus der Literatur und aus Implementierungen in gängigen Softwarepaketen auf ihre Eignung für die Klassifikation von

Gebäudetypen untersucht. Weitere Informationen über die verwendeten Maßzahlen finden sich bei McGarigal und Marks (1995), McGarigal et al. (2011), Lang und Blaschke (2007), Angel, Parent und Civco (2010) und Trimble (2010b). Viele der vorgestellten Maßzahlen sind miteinander korreliert, da sie aus den gleichen Basismerkmalen *Fläche*, *Umfang* und *Höhe* berechnet werden. Durch die Analyse nach ihrem Erklärungsgehalt wird die redundante Information verringert und es werden nur mehr jene Merkmale für die Klassifikation ausgewählt, welche den größten Erklärungsgehalt für die Zielklassen aufweisen. In den folgenden Abschnitten werden die Eigenschaften der Formmerkmale, die Auswahl nach ihrem Erklärungsgehalt und der überwachte Klassifikationsablauf beschrieben.

### 5.1.1. Formmerkmale für die Beschreibung der Gebäudephysiognomie

Die für die Beschreibung der Gebäudephysiognomie verwendeten geometrischen Maßzahlen werden im Folgenden beschrieben. Es handelt sich dabei sowohl um zweidimensionale als auch dreidimensionale Merkmale. Neben einfachen Formmerkmalen wie der *Fläche* oder der *Höhe*, welche einfache geometrische Eigenschaften der Objekte beschreiben, verwenden andere Merkmale Approximationen von verschiedenen geometrischen Figuren wie Kreisen, Quadraten und Rechtecken oder auch von Körpern wie Würfeln, um die Ähnlichkeit des Objektes mit ihnen zu beschreiben. Den Großteil der vorgestellten Merkmale stellen 2D-Merkmale dar, für einige von ihnen existiert auch das zugehörige 3D-Pendant. Auf diese Weise kann der Einfluss der dritten Dimension für die Klassifikation der Gebäude untersucht werden. Die Formeln für die Merkmale sind aus den oben zitierten Quellen entnommen.

#### 5.1.1.1. Einfache Formmerkmale

##### Fläche

Der Flächeninhalt  $A$  gibt die Grundfläche des Gebäudeobjekts in Quadratmetern wieder. Er berechnet sich aus der Summe aller Pixel eines Gebäudes, multipliziert mit der Grundfläche eines Pixels.

$$(15) \quad A = n \cdot A_p$$

wobei  $n$  die Anzahl der Pixel ist und  $A_p$  die Fläche eines Pixels in Quadratmetern.

##### Umfang

Der Umfang  $U$  eines Gebäudes beschreibt die Länge seiner Begrenzungslinie in Metern. Falls das Gebäudeobjekt ringförmige Geometrien aufweist, ergibt sich der Umfang aus der äußeren und inneren Begrenzungslinie.

$$(16) \quad U = U_{\text{au\ss en}} + U_{\text{innen}}$$

wobei  $U_{\text{au\ss en}}$  die Lange der au\ss eren und  $U_{\text{innen}}$  die Lange der inneren Begrenzungslinie ist.

### Hoh e

Die Hoh e  $h$  eines Gebau des in Metern wird gebildet aus dem Mittelwert der Hoh en jedes Pixels  $h_j$  des Gebau deobjektes.

$$(17) \quad h = \frac{1}{n} \cdot \sum_{j=1}^n h_j$$

### Volumen

Das Volumen  $V$  in Kubikmetern eines Objektes berechnet sich aus der Multiplikation seiner Grundflach e  $A$  mit seiner Hoh e  $h$ .

$$(18) \quad V = A \cdot h$$

### Hoh envarianz

Sie ist ein Ma\ss fur die Gleichma\ss igkeit der Hoh en innerhalb eines Objektes. Sie wird aus dem Quadrat der Standardabweichung  $\sigma_z$  aller Hoh enwerte  $z_j$  eines Objektes berechnet. Je hoh er die Hoh envarianz eines Objektes ist, desto unregelma\ss iger ist die Dachform.

$$(19) \quad \text{Var}Z = \sigma_z^2$$

#### 5.1.1.2. Kreisapproximation

##### Normalisierter Umfangindex (nUI)

Dieser Index ist an den einfachen Umfangindex angelehnt, jedoch wird hier der Wertebereich mit einem Kreis der gleichen Flach e zwischen  $0 < \text{nUI} < 1$  normalisiert, da der Kreis die kompakteste Form darstellt und somit der Einfluss der Flach e auf das Ma\ss eliminiert wird.

$$(20) \quad \text{nUI} = \frac{U_{\text{AKreis}}}{U}$$

wobei  $U_{\text{AKreis}}$  den Umfang eines Kreises mit dem gleichen Flach eninhalt des Objektes darstellt.

##### Normalisierter Abstandsindex (*proximity index*)

Der normalisierte Abstandsindex beruht auf der Berechnung der euklidischen Distanzen zwischen den einzelnen Pixeln eines Objektes zum Objektmittelpunkt. Er ist angelehnt an den Proximity-Index zur Berechnung von raumlichen Verteilungen von Habitaten (Gustafson und

Parker, 1994). Er wird ebenfalls auf Basis der Proximität eines Kreises mit dem gleichen Flächeninhalt wie das betrachtete Objekt normalisiert.

$$(21) \quad nAI = \frac{P_{AKreis}}{P_{Objekt}}$$

wobei die Proximität des Kreises:

$$(22) \quad P_{AKreis} = \frac{2}{3} \cdot r_{AKreis}$$

mit  $r_{AKreis}$  den Radius des Kreises und für  $P_{Objekt}$  gilt:

$$(23) \quad P_{Objekt} = \sum_{j=1}^n d_j \cdot \frac{1}{n}$$

wobei  $d$  die euklidische Distanz der Pixel zum Objektmittelpunkt darstellt.

#### Normalisiertes Trägheitsmoment (nTM)

Dieser Index ähnelt dem Proximitäts-Index, jedoch wird beim Trägheitsmoment ein stärkeres Gewicht auf die Extremitäten des Polygons gelegt. Dieses Maß ist ebenfalls für die Berechnung der Kompaktheit eines Objektes geeignet, gegenüber anderen Kompaktheitsmaßen aber auch in der Lage stark zerklüftete Objekte zu quantifizieren.

$$(24) \quad nTM = \frac{J_{AKreis}}{J_{Objekt}}$$

wobei das Trägheitsmoment  $J$  für den Kreis:

$$(25) \quad J_{AKreis} = \frac{1}{2} \cdot r_{AKreis}^2$$

und für das Objekt:

$$(26) \quad J_{Objekt} = \frac{1}{n} \cdot \sum_{j=1}^n d_j^2$$

und  $d^2$  die quadrierte euklidische Distanz der Pixel zum Objektmittelpunkt darstellt.

#### 5.1.1.3. Quadratapproximation

##### Flächen-Umfang-Verhältnis (*shape index*)

Das Flächen-Umfang-Verhältnis beschreibt ähnlich dem *Umfangindex* die Glätte der äußeren Form eines Objektes. Die Berechnung basiert auf dem Verhältnis zwischen dem tatsächlichen Umfang  $U$  eines Objektes und einem approximierten Quadrat mit der gleichen Fläche wie das Objekt. Das Verhältnis ist allerdings von der absoluten Flächengröße abhängig, weshalb auch

standardisierte Derivate verwendet werden. Der absolute *shape index* (Forman und Godron, 1986) ist in Verbindung mit weiteren Maßzahlen aussagekräftiger. In der Berechnung mit Rasterdaten gilt: je näher die Form einem Quadrat, desto kleiner ist sein *shape index*.

$$(27) \quad FI2D = \frac{U}{4 \cdot \sqrt{A}}$$

### Fraktale Dimension (*fractal dimension*)

Die fraktale Dimension *FD* (Mandelbrot, 1982) beschreibt die Komplexität und Fragmentierung eines Objektes durch ein Verhältnis zwischen Umfang und Fläche. Sie ermöglicht die Beschreibung von komplexen Formen über einen sehr großen Wertebereich hinaus, weshalb es durch die Darstellung dieses Verhältnisses möglich ist, eine der größten Limitierungen des direkten Umfang:Fläche-Verhältnisses zu überbrücken, da sie nicht durch die Größe des Objektes verändert wird. Sie ist ähnlich dem *shape index*, doch durch die Normalisierung der Werte auf Basis des natürlichen Logarithmus ist der Wertebereich zwischen 1 und 2 festgelegt:  $1 \leq FD \leq 2$ .

$$(28) \quad FD = \frac{\ln\left(\frac{U}{4}\right)^2}{\ln A}$$

### Dichte (2D)

Die flächenmäßige Dichte *D2D* eines Objektes beschreibt die räumliche Verteilung seiner Bildelemente. Dabei beschreiben hohe Dichtewerte eine größere Ähnlichkeit mit einem Quadrat, kleinere Werte entsprechen eher länglichen Objekten.

$$(29) \quad D2D = \frac{\sqrt{A}}{1 + \sqrt{\sigma_x^2 + \sigma_y^2}}$$

Dabei ist  $\sigma_x^2$  die Varianz der Verteilung der Bildelemente in Richtung *X* und  $\sigma_y^2$  die Varianz der Verteilung der Bildelemente in Richtung *Y*:

$$(30) \quad \sigma_x^2 = \sqrt{\frac{1}{n-1} \cdot \sum_{j=1}^n (x_j - \bar{x})^2}$$

mit dem Mittelwert  $\bar{x}$  aus:

$$(31) \quad \bar{x} = \frac{1}{n} \cdot \sum_{j=1}^n x_j$$

#### 5.1.1.4. Rechteckapproximation

##### Umfangindex

Der Umfangindex *UI* beschreibt, wie zerklüftet der Umfang *U* eines Objektes ist. Je höher der Wert, desto unregelmäßiger ist die Form des Objektes. Dieser Index ist ähnlich der

*Kompaktheit (2D)*, jedoch vergleicht er den Umfang anstelle der Fläche. Er drückt das Verhältnis zwischen dem Rechteckumfang und dem tatsächlichen Umfang  $U$  aus.

$$(32) \quad UI = \frac{2 \cdot (l + b)}{U}$$

wobei die Länge  $l$  die 1. Hauptträgheitsachse  $\lambda_{max}$  des Objektes und die Breite  $b$  die 2. Hauptträgheitsachse  $\lambda_{min}$  des Objektes ist.

### **Kompaktheit (2D)**

Sie ist ähnlich dem *Umfangindex*, sie wird jedoch basierend auf der Fläche des Objektes anstelle des Umfangs berechnet. Das Verhältnis zwischen der Rechteckfläche eines Objektes zu seiner Fläche  $A$  beschreibt wie kompakt ein Objekt ist.

$$(33) \quad K2D = \frac{l \cdot b}{A}$$

### **Rechteckigkeit**

Die Rechteckigkeit beschreibt wie gut das betrachtete Objekt in ein Rechteck mit der gleichen Größe passt, wobei ein Wert 0 keine Übereinstimmung beschreibt und 1 eine ideale Übereinstimmung. Für die Berechnung dieses Indexes wird ein Rechteck mit der gleichen Länge und Breite des betrachteten Objektes erstellt, wobei die Fläche des Objektes innerhalb des Rechteckes mit der Fläche des Objektes außerhalb des Rechteckes verglichen wird.

$$(34) \quad RE = \frac{\sum_p A_p \in \text{Rechteck}}{\sum_p A_p \notin \text{Rechteck}}$$

#### **5.1.1.5. Würfelapproximation**

##### **Flächen-Umfang-Verhältnis (3D shape index)**

Ähnlich der Berechnung des flächenmäßigen *Formindex* wird bei der Berechnung des Formindexes 3D *FI3D* die Glätte der Oberfläche des Objektes beschrieben. Allerdings berechnet er sich aus dem Verhältnis zwischen dem Umfang eines approximierten Würfels mit dem gleichen Volumen  $V$  wie das Objekt und dem tatsächlichen Umfang des Objektes.

$$(35) \quad FI3D = \frac{U}{4 \cdot \sqrt[3]{V}}$$

##### **Dichte (3D)**

Der dreidimensionale Dichtewert wird ähnlich seinem zweidimensionalen Pendant berechnet. Seine Berechnung unterscheidet sich dadurch, dass für seine Berechnung die idealisierte Kantenlänge eines Würfels mit dem gleichen Volumen wie das Objekt durch die Längen der drei Hauptträgheitsachsen geteilt wird.

$$(36) \quad D3D = \frac{\sqrt[3]{V}}{\sqrt{\sigma_x^2 + \sigma_y^2 + \sigma_z^2}}$$

#### 5.1.1.6. Weitere Merkmale

##### Asymmetrie (2D)

Die flächenmäßige Asymmetrie  $A2D$  eines Objektes beschreibt die relative Länge des Objektes verglichen mit einem regelmäßigen Objekt. Dabei wird die Länge  $l$  mit der Breite  $b$  in einen direkten Bezug gebracht, wobei die Länge und die Breite ähnlich dem Umfangindex den beiden Hauptträgheitsachsen entsprechen. Je höher der Wert ist, desto asymmetrischer ist das Objekt.

$$(37) \quad A2D = \frac{l}{b}$$

##### Asymmetrie (3D)

Die dreidimensionale Asymmetrie beschreibt ähnlich wie ihr zweidimensionales Pendant die relative Länge eines Objektes, jedoch in drei Dimensionen. Sie berechnet sich aus dem Verhältnis zwischen der kleinsten und der größten Hauptträgheitsachse des Objektes.

$$(38) \quad A3D = 1 - \sqrt{\frac{\lambda_{min}}{\lambda_{max}}}$$

wobei  $\lambda_{min}$  die kleinste Hauptträgheitsachse des Objektes ist:

$$(39) \quad \lambda_{min} = \min(l; b; h)$$

und  $\lambda_{max}$  die größte Hauptträgheitsachse des Objektes ist:

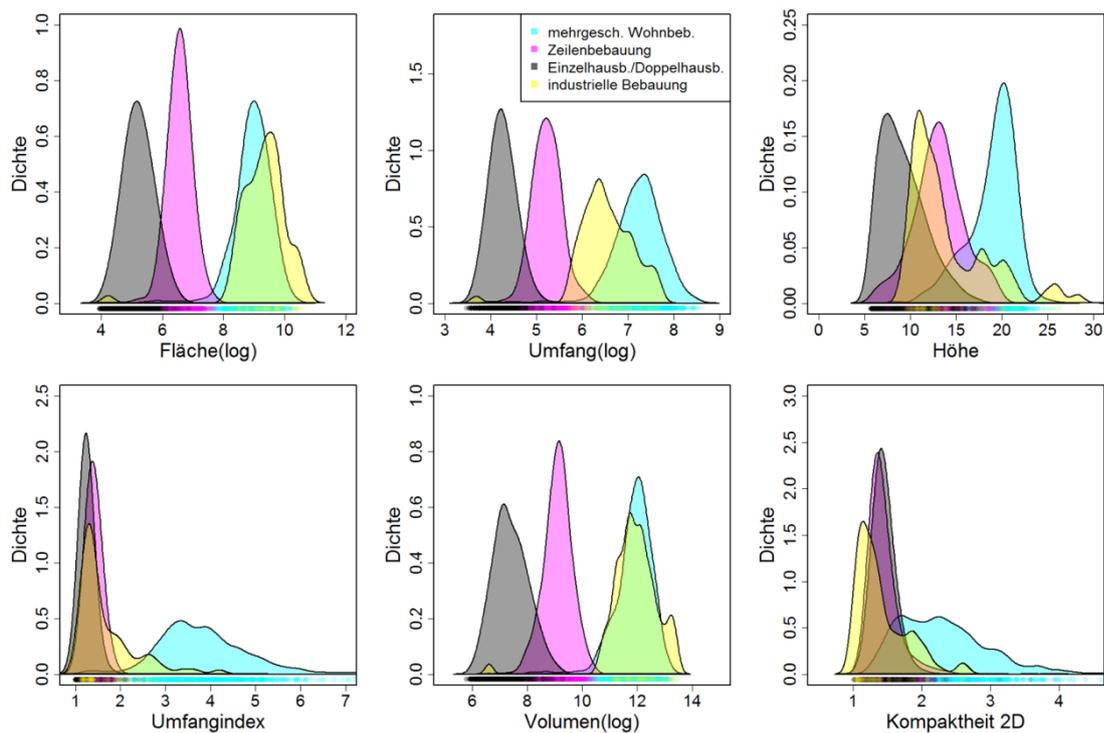
$$(40) \quad \lambda_{max} = \max(l; b; h)$$

#### 5.1.2. Häufigkeitsverteilung der Formmerkmale

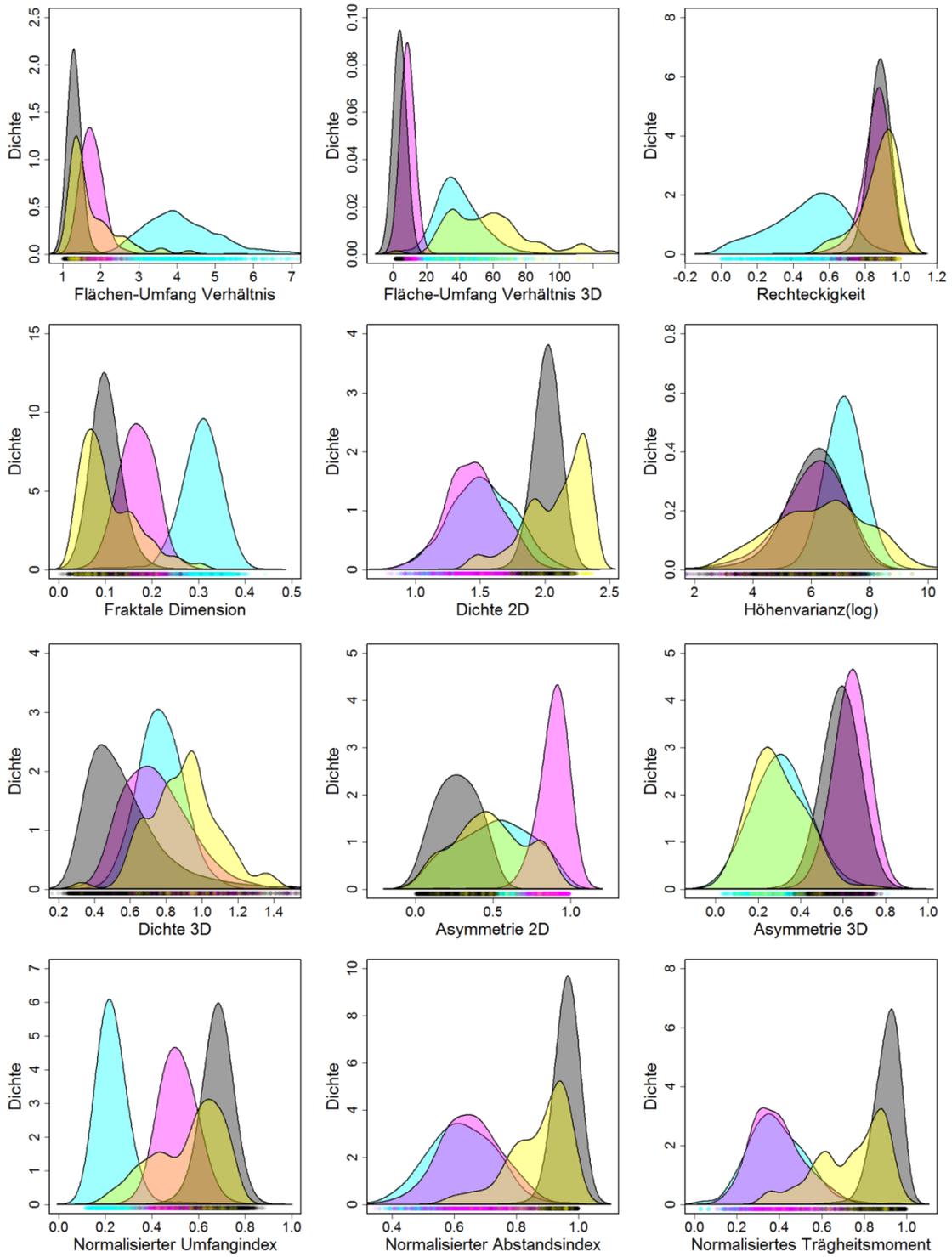
Die im vorigen Abschnitt beschriebenen Formmerkmale dienen der Beschreibung der Gebäudephysiognomie. Sie werden dazu verwendet, um unterschiedliche Gebäudeklassen auf Basis dieser Merkmale zu klassifizieren. Die Berechnung der Merkmale wird für jedes einzelne Gebäudeobjekt durchgeführt. Für die Analyse der Merkmalsausprägungen bzw. für die überwachte Klassifikation der Gebäude ist Vorwissen über die definierten Zielklassen notwendig. Damit die Merkmalsausprägungen für die Zielklassen untersucht werden können, muss ein Datensatz sowohl die Zielklassen als auch die Merkmalsausprägungen enthalten. Für die Untersuchung der Einflüsse auf die Klassifikation der einzelnen Merkmale und für das

Training der Klassifikation wurde ein Referenzdatensatz aufbereitet, welcher auf Gebäudeebene sowohl die Gebäudeklassen als auch die Formmerkmale enthält (vgl. Abbildung 18, S.60).

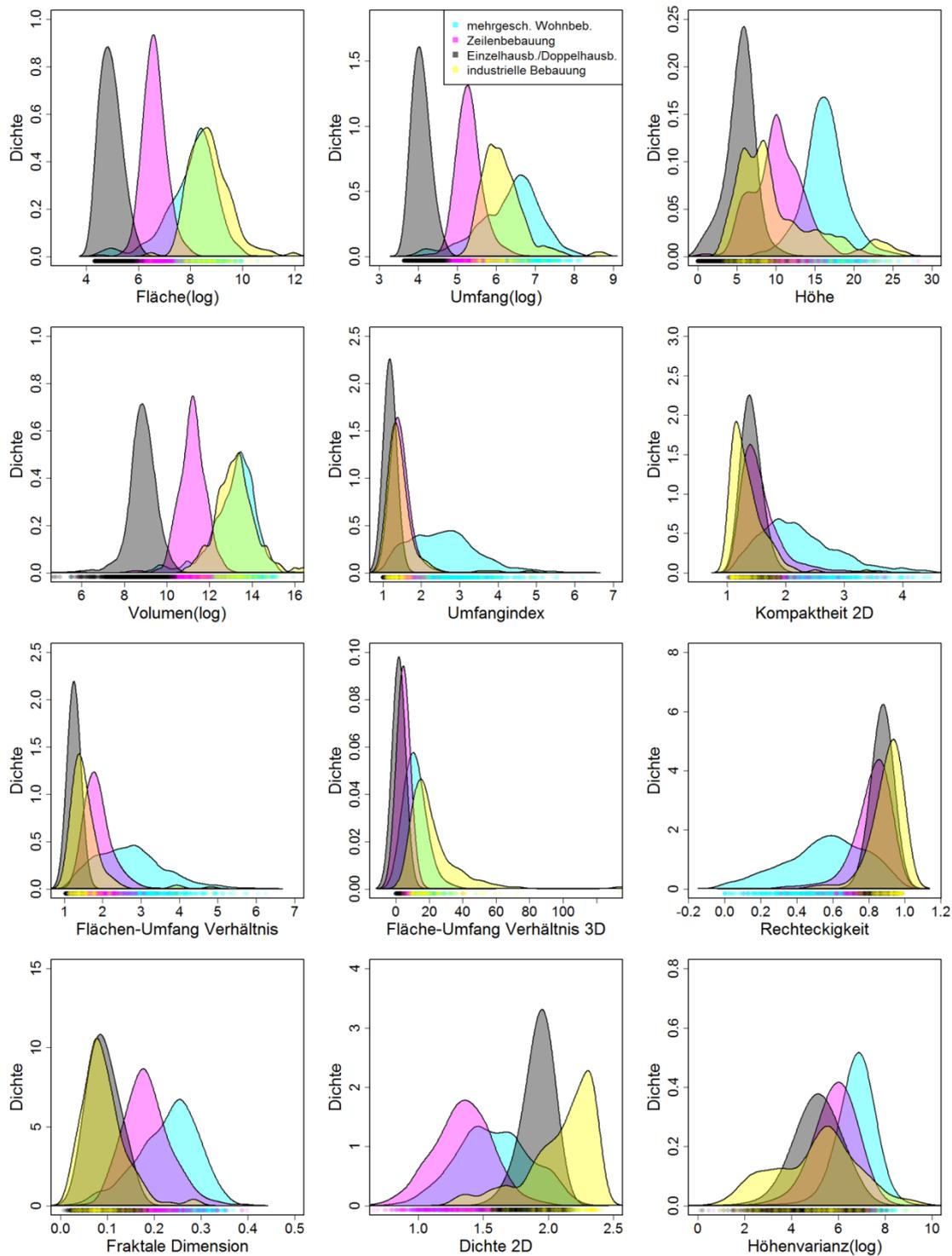
Der Einfluss auf die Klassifikation und die Relevanz der beschriebenen Formmerkmale wird im folgenden Abschnitt auf Basis der linearen Diskriminanzfunktion berechnet. Es kann aber bereits davor eine visuelle Untersuchung der Häufigkeitsverteilungen für die einzelnen Klassen durchgeführt werden. In der folgenden Abbildung 43 (Berlin) und Abbildung 44 (München) sind die Merkmalsausprägungen der Formmerkmale für die vier *Nicht-Hochhaus*-Gebäudeklassen in kontinuierlichen Histogrammen dargestellt. Diese Art von Histogrammen wird auch Kerndichteschätzerdiagramm genannt (Parzen, 1962; Sheather, 2004). Der Kerndichteschätzer stellt die Häufigkeitsverteilung der vorkommenden Werte als stetige Schätzung anstelle von diskreten Wertebereichen wie in einem Histogramm dar. Die Dichte (ähnlich der Häufigkeit in einem Histogramm) ist auf der Ordinate dargestellt, während der Wertebereich auf der Abszisse abgebildet ist.



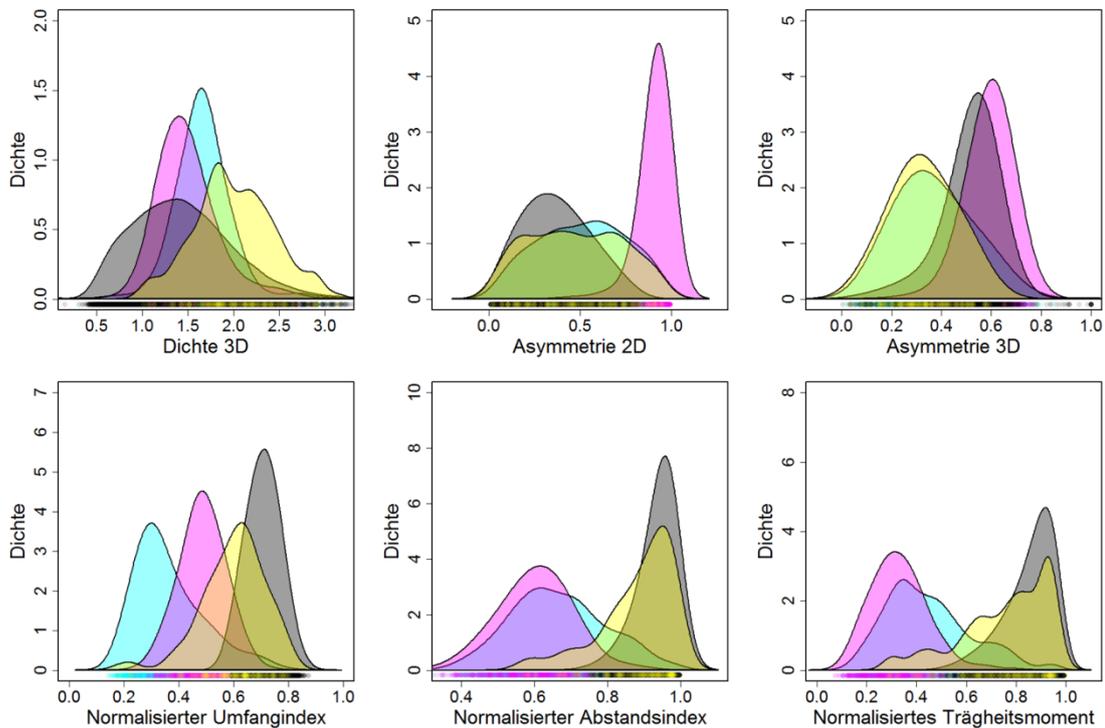
**Abbildung 43:** Häufigkeitsverteilung der Formmerkmale für die 4 Gebäudeklassen für das Untersuchungsgebiet Berlin (Fortsetzung auf der nächsten Seite).



**Abbildung 41:** Häufigkeitsverteilung der Formmerkmale der 4 Gebäudeklassen für das Untersuchungsgebiet Berlin (Fortsetzung von voriger Seite).



**Abbildung 44:** Häufigkeitsverteilung der Formmerkmale für die 4 Gebäudeklassen für das Untersuchungsgebiet München (Fortsetzung auf der nächsten Seite).



**Abbildung 42:** Häufigkeitsverteilung der Formmerkmale für die 4 Gebäudeklassen für das Untersuchungsgebiet München (Fortsetzung von voriger Seite).

Die Abbildungen zeigen, dass einige der geometrischen Maßzahlen eine klare Trennbarkeit zwischen den vier Gebäudeklassen ermöglichen, wie zum Beispiel der *Umfang* oder die *Höhe*. Andere Merkmale hingegen eignen sich nur für die Trennung zwischen zwei oder drei Gebäudeklassen wie die *Fläche*, das *Volumen* oder die *Asymmetrie 3D*. Weitere Merkmale erscheinen zunächst nicht besonders geeignet für die Trennung zwischen den unterschiedlichen Gebäudeklassen, da sich die Merkmalsausprägungen der Klassen sehr stark überschneiden. Es kann aber erst durch die Analyse der Diskriminanzkraft der einzelnen Merkmale im folgenden Abschnitt eine Aussage über den Einfluss auf die Klassifikation getroffen werden.

Die Häufigkeitsverteilungen der Merkmalsausprägungen zeigen im Vergleich der beiden Städte Ähnlichkeiten in der Form der Histogramme. Allerdings können Unterschiede festgestellt werden, welche auf keine eindeutigen Grenzen zwischen den Gebäudeklassen zurückzuführen sind. Inwieweit sich die Formmerkmale der beiden Städte unterscheiden und für eine Klassifikation verwendet werden können, wird im folgenden Abschnitt auf Basis der Diskriminanzanalyse behandelt. Für die Klassifikation der Gebäudeklassen wird deshalb ein Klassifikationsverfahren angewendet, welche die Ausprägungen von mehreren Merkmalen berücksichtigt, da bereits in der visuellen Interpretation der Häufigkeitsverteilungen zu beobachten ist, dass ein einziges Merkmal nicht in der Lage ist, mehrere Gebäudeklassen eindeutig und für andere Untersuchungsgebiete zu unterscheiden.

### 5.1.3. Semantische Klassifikation von Gebäuden auf Basis von Formmerkmalen und linearer Diskriminanzfunktion

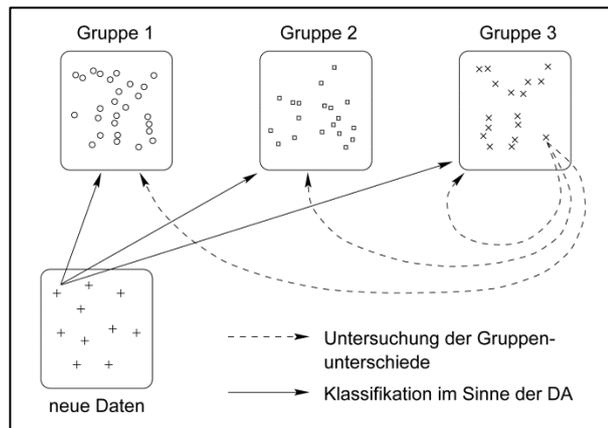
Viele klassische Klassifikationsverfahren, welche auch in der Fernerkundung verwendet werden, haben ihren Ursprung in statistischen, multivariaten Analysemethoden (Krzanowski, 1988) und in den letzten Jahren in den Methoden der Mustererkennung (Ripley, 1996). Einige der bekanntesten Verfahren sind nearest neighbors, classification and regression trees - CART (Breiman et al., 1984), neuronale Netze (Bishop, 1995) und vor allem Support Vector Machines (Vapnik, 2010). Letztere sind sehr rechenintensive Verfahren und werden vor allem durch die Steigerung von Rechnerkapazitäten und -leistungen in den letzten Jahren verstärkt angewendet. In der Fernerkundung werden neben diesen Verfahren allerdings noch immer einige klassische Methoden verwendet (Venables und Ripley, 2002). Zu diesen klassischen Methoden gehört auch die Diskriminanzanalyse, welche ein etabliertes multivariates, statistisches Verfahren darstellt. Sie wird in der Statistik verwendet, um den Einfluss von einzelnen Merkmalen auf die Klassifikation zu überprüfen. In der Fernerkundung wird die Diskriminanzfunktion zusätzlich auch zur Klassifikation von Objekten in multidimensionalen Merkmalsräumen verwendet (z.B. Lobo, 1997; Herold, Liu und Clarke, 2003; Landgrebe und Biehl, 2011).

Das Verfahren wurde zunächst für pflanzentaxonomische Untersuchungen von Fisher (1936) beschrieben: „*When two or more populations have been measured in several characters  $x_1, \dots, x_s$ , special interest attaches to certain linear functions of the measurements by which the populations are best discriminated.*“

Allgemein formuliert ermöglicht die Diskriminanzanalyse die Untersuchung von folgenden Fragestellungen:

- *Unterscheiden sich die Gruppen signifikant voneinander hinsichtlich der Merkmale?*
- *Welche Merkmale bzw. Merkmalskombinationen sind zur Unterscheidung zwischen den Gruppen geeignet bzw. ungeeignet?*
- *In welche Gruppe kann ein weiteres Objekt, dessen Gruppenzugehörigkeit nicht bekannt ist, aufgrund seiner Merkmalsausprägungen eingeordnet werden?*

Sie wird im Kontext der Arbeit verwendet, da sie ein etabliertes, statistisches Standardverfahren darstellt und alle Aufgabenstellungen der Arbeit damit behandelt werden können. In der Arbeit ist es dadurch möglich den Einfluss der Formmerkmale für die Klassifikation der vier Gebäudeklassen zu untersuchen und die Gebäude zu klassifizieren. Die Diskriminanzfunktion eignet sich daher sowohl für die Untersuchung der Gruppenunterschiede als auch zur Klassifikation neuer Objekte (vgl. Abbildung 45).



**Abbildung 45:** Untersuchung der Gruppenunterschiede und Klassifikation neuer Objekte.

(Quelle: Kraft, 2000)

Theorie und Anwendung der Diskriminanzanalyse wird in den statistischen Lehrbüchern ausführlich diskutiert, auf welche sich auch die folgenden Ausführungen beziehen (Brosius, 1998; Bortz, 1999; Kraft, 2000; Handl, 2002; Venables und Ripley, 2002; Backhaus et al., 2003; Hastie, Tibshirani und Friedman, 2011).

Allgemein formuliert dient die *Diskriminanzanalyse* der Untersuchung von  $n$  Objekten, die zu  $g$  bekannten Gruppen gehören anhand von  $p$  Merkmalen. Es gilt:  $n_1 + n_2 + \dots + n_g = n$  sowohl als auch  $1 \leq i \leq p$ ,  $1 \leq k \leq g$  und  $1 \leq j \leq n_i$ , wobei  $n_k$  die Anzahl der Objekte in jeder Gruppe,  $x_i$  die  $i$ -te Merkmalsvariable und  $x_{ijk}$  die Ausprägung des  $i$ -ten Merkmals des  $j$ -ten Objektes der  $k$ -ten Gruppe ist. Darüber hinaus müssen mindestens zwei Gruppen bestehen und mindestens zwei Objekte pro Gruppe enthalten sein. Es muss auch mindestens ein Merkmal existieren und es müssen mindestens drei Objekte mehr existieren als die Anzahl der Merkmale, da sie für die Lösung der Gleichung benötigt werden. Die Merkmale müssen auch mit einer Intervallskala erfasst sein und die Merkmale dürfen sich nicht als Linearkombination anderer Merkmale schreiben lassen. Zuletzt müssen die Objekte jeder Gruppe aus einer normalverteilten Grundgesamtheit stammen.

Bei der Verwendung einer großen Anzahl an Merkmalen, die alle auf ähnlichen Eigenschaften aufbauen, ist von einer hohen Korrelation der einzelnen Merkmale zueinander auszugehen, aber nicht grundsätzlich davon, dass sich die Merkmale durch Linearkombination anderer Merkmale schreiben lassen können. Im Anwendungsfall der Arbeit wurde eine Korrelationsmatrix erstellt, welche die Korrelationen zwischen den einzelnen Merkmalen beschreibt und in Tabelle 13 dargestellt ist. Sie zeigt sehr unterschiedliche Zusammenhänge zwischen den Merkmalen, einige sind sehr stark negativ korreliert und einige sind sehr hoch positiv korreliert.

**Tabelle 13:** Korrelationsmatrix der Formmerkmale.

	Fläche	Umfang	Höhe	Volumen	VarZ	nUI	nAI	nTM	FI2D	FD	D2D	UI	K2D	RE	FI3D	D3D	A2D	A3D
<b>Fläche</b>	1	0,946	0,662	0,985	0,345	-0,764	-0,520	-0,509	0,875	0,738	-0,386	0,874	0,547	-0,638	0,983	0,236	-0,718	0,208
<b>Umfang</b>	0,946	1	0,695	0,956	0,327	-0,835	-0,593	-0,575	0,968	0,832	-0,472	0,965	0,646	-0,735	0,944	0,225	-0,732	0,228
<b>Höhe</b>	0,662	0,695	1	0,694	0,398	-0,781	-0,674	-0,672	0,738	0,729	-0,534	0,709	0,482	-0,592	0,703	0,324	-0,395	0,411
<b>Volumen</b>	0,985	0,956	0,694	1	0,383	-0,768	-0,512	-0,501	0,886	0,747	-0,381	0,886	0,541	-0,631	0,958	0,235	-0,702	0,199
<b>VarZ</b>	0,345	0,327	0,398	0,383	1	-0,324	-0,198	-0,194	0,320	0,291	-0,093	0,327	0,254	-0,268	0,349	-0,270	-0,298	0,067
<b>nUI</b>	-0,764	-0,835	-0,781	-0,768	-0,324	1	0,842	0,835	-0,926	-0,985	0,725	-0,889	-0,707	0,806	-0,829	-0,272	0,670	-0,488
<b>nAI</b>	-0,520	-0,593	-0,674	-0,512	-0,198	0,842	1	0,992	-0,711	-0,807	0,951	-0,621	-0,538	0,665	-0,609	-0,300	0,368	-0,767
<b>nTM</b>	-0,509	-0,575	-0,672	-0,501	-0,194	0,835	0,992	1	-0,690	-0,797	0,946	-0,598	-0,500	0,619	-0,596	-0,319	0,342	-0,817
<b>FI2D</b>	0,875	0,968	0,738	0,886	0,320	-0,926	-0,711	-0,690	1	0,931	-0,599	0,987	0,729	-0,824	0,904	0,238	-0,738	0,310
<b>FD</b>	0,738	0,832	0,729	0,747	0,291	-0,985	-0,807	-0,797	0,931	1	-0,721	0,899	0,733	-0,816	0,792	0,242	-0,643	0,442
<b>D2D</b>	-0,386	-0,472	-0,534	-0,381	-0,093	0,725	0,951	0,946	-0,599	-0,721	1	-0,494	-0,466	0,576	-0,459	-0,265	0,182	-0,786
<b>UI</b>	0,874	0,965	0,709	0,886	0,327	-0,889	-0,621	-0,598	0,987	0,899	-0,494	1	0,757	-0,839	0,897	0,220	-0,797	0,189
<b>K2D</b>	0,547	0,646	0,482	0,541	0,254	-0,707	-0,538	-0,500	0,729	0,733	-0,466	0,757	1	-0,863	0,596	0,068	-0,699	0,081
<b>RE</b>	-0,638	-0,735	-0,592	-0,631	-0,268	0,806	0,665	0,619	-0,824	-0,816	0,576	-0,839	-0,863	1	-0,702	-0,143	0,744	-0,151
<b>FI3D</b>	0,983	0,944	0,703	0,958	0,349	-0,829	-0,609	-0,596	0,904	0,792	-0,459	0,897	0,596	-0,702	1	0,262	-0,772	0,272
<b>D3D</b>	0,236	0,225	0,324	0,235	-0,270	-0,272	-0,300	-0,319	0,238	0,242	-0,265	0,220	0,068	-0,143	0,262	1	-0,127	0,264
<b>A2D</b>	-0,718	-0,732	-0,395	-0,702	-0,298	0,670	0,368	0,342	-0,738	-0,643	0,182	-0,797	-0,699	0,744	-0,772	-0,127	1	0,084
<b>A3D</b>	0,208	0,228	0,411	0,199	0,067	-0,488	-0,767	-0,817	0,310	0,442	-0,786	0,189	0,081	-0,151	0,272	0,264	0,084	1

Es kann anhand der Korrelationsmatrix auch überprüft werden, ob sich ein Merkmal durch Linearkombination anderer Merkmale schreiben lassen kann. Wenn es in einer Korrelationsmatrix lineare Abhängigkeiten gibt, so zeigt sich dies unter anderem darin, dass der Rang um die Anzahl der linearen Abhängigkeiten in der Matrix vermindert ist. Umgekehrt heißt es, wenn die Matrix vollen Rang hat, bedeutet dies, dass zwischen den Vektoren keine lineare Beziehung besteht (Hackl, 2008). Die Matrix aus Tabelle 13 hat den Rang 18, also vollen Rang. Daraus kann geschlossen werden, dass sich die Merkmale nicht als Linearkombination anderer Merkmale schreiben lassen.

In der praktischen Anwendung erfolgt der Ablauf der Diskriminanzanalyse in der Regel in sechs Teilschritten: *a) Definition der Gruppen, b) Formulierung der Diskriminanzfunktion, c) Schätzung der Diskriminanzfunktion, d) Prüfung der Diskriminanzfunktion, e) Prüfung der Merkmalsvariablen* und zuletzt *f) die Klassifikation neuer Elemente*.

Die Anzahl der Gruppen und ihre möglichen Ausprägungen sind bereits am Beginn dieses Kapitels festgelegt worden. Für die vier definierten Gebäudeklassen *g*: *Einzel- und Doppelhausbebauung, Zeilenbebauung, mehrgeschossige Wohnbebauung* und *industrielle Bebauung* werden auf Basis der *p* geometrischen Maßzahlen drei lineare Diskriminanzfunktion formuliert, da die Anzahl der gesuchten Diskriminanzfunktionen *g-1* beträgt. Die Formulierung der einzelnen Diskriminanzfunktionen geschieht auf Basis der Form:

$$(41) \quad y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_4x_4$$

wobei *y* die Diskriminanzvariable, *x<sub>i</sub>* die Merkmalsvariablen, *b<sub>1</sub>*, *b<sub>2</sub>*, *b<sub>3</sub>* und *b<sub>4</sub>* die Diskriminanzkoeffizienten und *b<sub>0</sub>* ein konstantes Glied ist. Folglich repräsentiert die Diskriminanzfunktion eine Transformation aus dem *p*-dimensionalen Merkmalsraum in den 1-dimensionalen Raum der Diskriminanzvariablen. Auf Basis dieser Gleichung kann eine Schätzung der unbekanntenen Koeffizienten *b<sub>i</sub>* durchgeführt werden. Sie verfolgt das Ziel, eine optimale Trennung zwischen den betrachteten Gruppen zu finden. Dies geschieht dadurch, dass das Diskriminanzkriterium, welches die Unterschiedlichkeit der Gruppen misst, maximiert wird. Das Diskriminanzkriterium wird folgendermaßen formuliert:

$$(42) \quad \Gamma = \frac{\text{Streuung zwischen den Gruppen}}{\text{Streuung innerhalb der Gruppen}} = \frac{SQ_{\text{zwischen}}}{SQ_{\text{innerhalb}}}$$

wobei

$$(43) \quad SQ_{\text{zwischen}} = \sum_{k=1}^g n_k (\bar{y}_k - \bar{y})^2$$

und

$$(44) \quad SQ_{\text{innerhalb}} = \sum_{k=1}^g \sum_{j=1}^{n_k} (\bar{y}_{jk} - \bar{y}_k)^2$$

Der Maximalwert  $\gamma = \max(\Gamma)$  heißt *Eigenwert* der Diskriminanzfunktion. Nachdem die erste Diskriminanzfunktion mit dem Ziel ermittelt wird, dass ihr Eigenwert und damit auch ihr Erklärungsanteil maximal werden, werden die verbleibenden Diskriminanzfunktionen so ermittelt, dass sie den maximalen Anteil der verbleibenden Reststreuung erklären. Der *Eigenwertanteil* oder auch *Varianzanteil* lautet:

$$(45) \quad EA_k = \frac{\gamma_k}{\gamma_1 + \gamma_2 + \dots + \gamma_k}$$

Auf Basis der geschätzten Diskriminanzfunktionen ist eine Klassifizierung der für die Berechnung der Diskriminanzfunktionen verwendeten Elemente möglich. Da die Elemente, welche für die Erstellung der Diskriminanzfunktionen relevant sind, bereits eine Gruppenzugehörigkeit aufweisen, ist durch die erneute Klassifizierung der Eingangsdaten ein Vergleich zwischen der Ausgangsklasse und der auf Basis der Diskriminanzfunktionen geschätzten Klassenzugehörigkeit möglich. Dieser Vergleich bietet eine Möglichkeit, die Güte einer Klassifikation von neuen, noch nicht klassifizierten Elementen durch die aufgestellten Diskriminanzfunktionen abzuschätzen. Für die Berechnung der Diskriminanzfunktionen wurden zunächst alle Merkmale verwendet. Eine Auswahl der Merkmale auf Basis ihrer Diskriminanzkraft wird weiter unten beschrieben. Der Vergleich ist in Tabelle 14 für Berlin und in Tabelle 15 für München dargestellt.

**Tabelle 14:** Überprüfung der Diskriminanzfunktionen auf Basis der Referenzdaten für Berlin.

Referenzdaten (Berlin)					
Klassifikation	Einzel-/Doppelh.	Zeilenbebauung	Mehrg. Wohnbeb.	Industrie	Nutzergen.
Einzel-/Doppelh.	1169	1	1	5	99,49 %
Zeilenbebauung	22	1356	7	0	97,91 %
Mehrg. Wohnbeb.	8	26	6409	1	99,44 %
Industrie	0	0	0	69	100 %
<i>Herstellergen.</i>	97,50 %	98,05 %	99,88 %	92,00 %	
<b>Gesamtgenauigkeit: 99,22 %    Anzahl der Beobachtungen: 9074</b>					

**Tabelle 15:** Überprüfung der Diskriminanzfunktionen auf Basis der Referenzdaten für München.

Referenzdaten (München)					
Klassifikation	Einzel-/Doppelh.	Zeilenbebauung	Mehrg. Wohnbeb.	Industrie	Nutzergen.
Einzel-/Doppelh.	881	21	2	15	95,97 %
Zeilenbebauung	59	899	4	7	92,68 %
Mehrg. Wohnbeb.	61	35	7450	4	98,68 %
Industrie	1	0	1	109	98,20 %
<i>Herstellergen.</i>	87,92 %	94,14 %	99,91 %	80,74 %	
<b>Gesamtgenauigkeit: 97,80 %    Anzahl der Beobachtungen: 9549</b>					

Ein bedeutender Vorteil bei der Verwendung der Diskriminanzanalyse im Rahmen der Arbeit ist, dass der gesamte Arbeitsablauf mit demselben Verfahren durchgeführt werden kann. Es kann sowohl der Einfluss der einzelnen Merkmalsvariablen auf die Trennbarkeit der Klassen überprüft werden als auch die Klassifikation von neuen Elementen durchgeführt werden. Die Überprüfung des Einflusses ist relevant, um unbedeutende oder redundante Merkmale zu entfernen. Der diskriminatorische Einfluss der einzelnen Merkmale wird durch den jeweiligen *Diskriminanzkoeffizienten* ausgedrückt. In der praktischen Anwendung von unterschiedlichen Merkmalen können unterschiedliche Skalierungen auftreten, weshalb zur Bestimmung der Diskriminanzkraft des einzelnen Merkmals ein *standardisierter Diskriminanzkoeffizient* berechnet wird. Für die Standardisierung werden die Werte der Merkmale so transformiert, dass sie einen Mittelwert von 0 und eine Standardabweichung von 1 haben (Standardnormalverteilung).

$$(46) \quad b_i^* = b_i \cdot s_i$$

wobei  $b_i^*$  der standardisierte Diskriminanzkoeffizient,  $b_i$  der Diskriminanzkoeffizient des Merkmals  $i$  und  $s_i$  die Standardabweichung des Merkmals  $i$  ist.

In der durchgeführten Analyse werden mehrere Diskriminanzfunktionen geschätzt, weshalb für jedes Merkmal mehrere Diskriminanzkoeffizienten existieren. Für die Bestimmung der gesamten diskriminatorischen Bedeutung des einzelnen Merkmals über alle Diskriminanzfunktionen wird der *mittlere standardisierte Diskriminanzkoeffizient* berechnet. Er wird über den in Gleichung (36) berechneten Eigenwertanteil gewichteten absoluten Wert der Koeffizienten berechnet:

$$(47) \quad \bar{b}_i = \sum_{k=1}^K |b_{ik}^*| \cdot EA_k$$

wobei  $\bar{b}_i$  der mittlere standardisierte Diskriminanzkoeffizient des Merkmals  $i$ ,  $b_{ik}^*$  der standardisierte Diskriminanzkoeffizient des Merkmals  $i$  für die  $k$ -te Diskriminanzfunktion und  $EA_k$  der Eigenwertanteil der  $k$ -ten Diskriminanzfunktion ist.  $K$  ist die Anzahl der Diskriminanzfunktionen.

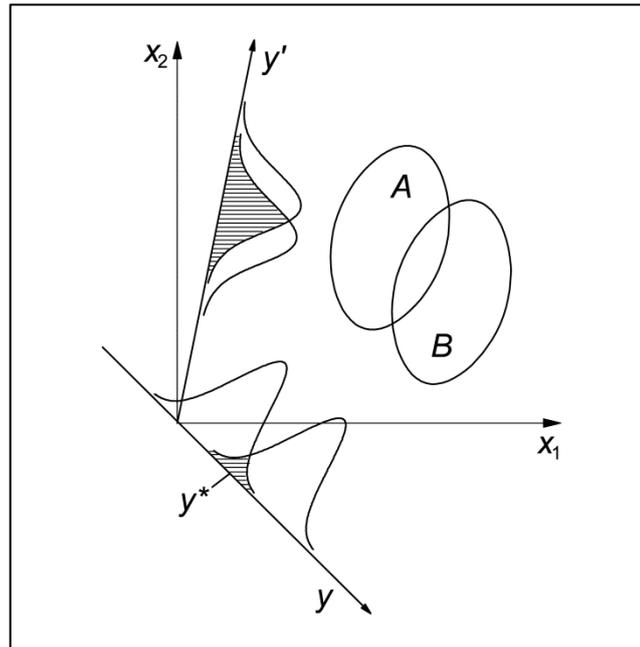
Die berechneten mittleren standardisierten Diskriminanzkoeffizienten für die Untersuchungsgebiete Berlin und München sind in Tabelle 16 dargestellt. Sie zeigen den Einfluss jedes Formmerkmals für die Klassifikation der Gebäude zu einer der vier Gebäudeklassen. Der Tabelle ist ebenfalls zu entnehmen, dass sich sowohl die Ausprägungen der mittleren standardisierten Diskriminanzkoeffizienten als auch die diskriminatorische Bedeutung jedes einzelnen Merkmals für beide Untersuchungsgebiete unterscheiden. Dies untermauert die Hypothese, dass keine allgemeingültigen Grenzen dieser Merkmalsausprägungen für unterschiedliche Untersuchungsgebiete angenommen werden können, sondern ein wiederholtes *Trainieren* des Klassifikationsverfahrens anhand von Trainingsgebieten notwendig ist. Des Weiteren kann aufgrund der Höhe der

Diskriminanzkoeffizienten entnommen werden, dass vor allem die 2D-Merkmale die höchsten diskriminatorischen Eigenschaften haben. Von den 3D-Merkmalen haben vor allem die einfachen Merkmale wie *Höhe* und *Volumen* einen relevanten Einfluss auf die Trennung der Gebäudeklassen. Für die Klassifikation der Gebäudeklassen werden auf Basis dieser Untersuchung die 10 Merkmale mit der höchsten Trennkraft ausgewählt.

**Tabelle 16:** Mittlere standardisierte Diskriminanzkoeffizienten für Berlin und München.

Berlin		München	
norm. TM	3.7308	norm. Umfangindex	2.8343
3D Shape index	3.4472	Fraktale Dimension	2.4582
Fläche	3.2762	Umfangindex	1.0471
Umfangindex	3.0475	Umfang	0.9386
Proximitätsindex	2.6329	norm. TM	0.8727
Fraktale Dimension	2.5286	Höhe	0.7516
norm. Umfangindex	2.1786	2D Dichte	0.6340
Umfang	1.4533	Rechteckigkeit	0.5485
2D Dichte	1.3798	2D Asymmetrie	0.4647
Volumen	1.3632	Fläche	0.2790
Shape index	1.0364	Proximitätsindex	0.2383
2D Asymmetrie	0.4837	Volumen	0.2169
Kompaktheit	0.3770	Shape index	0.1590
Höhe	0.2563	3D Asymmetrie	0.1523
Rechteckigkeit	0.1237	Kompaktheit	0.1472
Höhenvarianz	0.0850	3D Shape index	0.0472
3D Asymmetrie	0.0601	Höhenvarianz	0.0439
3D Dichte	0.0601	3D Dichte	0.0089

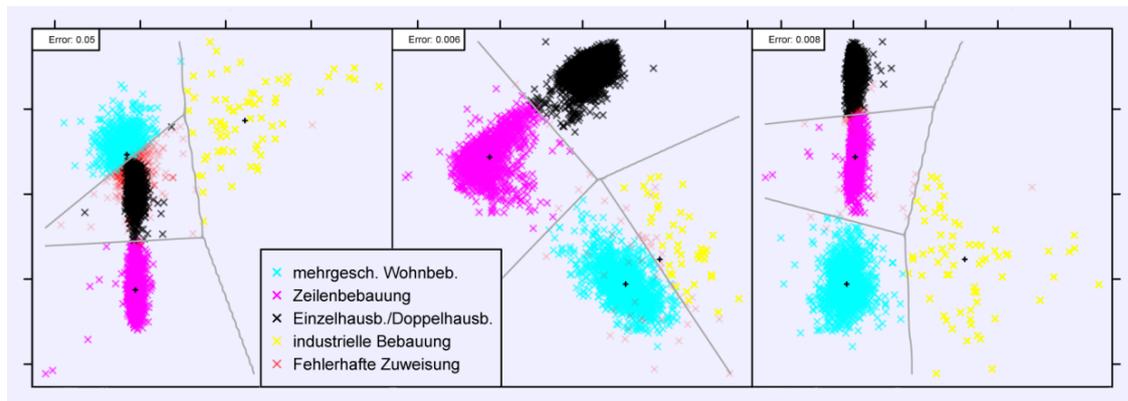
In einem Fall mit zwei Merkmalen bildet die Diskriminanzfunktion eine Ebene bzw. im Fall mit mehr als zwei Merkmalen eine Hyperebene im Raum. Die Diskriminanzvariable lässt sich als *Diskriminanzachse* in Form einer Geraden darstellen. In einer allgemeinen Form ist die Trennung von zwei unterschiedlichen Merkmalsgruppen am Beispiel eines Zwei-Gruppen-Falls in Abbildung 46 dargestellt. Sie zeigt zwei Ellipsen A und B, welche die Punktwolken der einzelnen Fälle repräsentieren. Zusätzlich sind die Häufigkeitsverteilungen der beiden Merkmalsgruppen dargestellt, die durch die Abbildung der Punktwolken auf die Diskriminanzachse entstehen würden. Es ist zu erkennen, dass eine Trennung der beiden Gruppen auf Basis der Diskriminanzachse  $y'$  zu größeren Überschneidungen im Merkmalsbereich führt und die geringste Überschneidung durch eine Trennung der beiden Gruppen durch den *kritischen Wert*  $y^*$  auf der Diskriminanzachse  $y$  geschieht. Es können unendlich viele Diskriminanzachsen durch Drehung um den Ursprung gebildet werden, wobei es gilt, diejenige zu finden, bei der sich die Verteilungen der Merkmale am geringsten überschneiden.



**Abbildung 46:** Geometrische Darstellung der Diskriminanzachse an einem Zwei-Gruppen-Fall.  
(Quelle: Kraft, 2000)

Die geometrische Darstellung der Diskriminanzfunktionen ermöglicht eine einfache Veranschaulichung, wie die Gruppen auf Basis der unterschiedlichen Merkmalsgruppen voneinander getrennt werden können. Wie die Trennung der einzelnen Gebäudeklassen auf Basis der drei geschätzten Diskriminanzfunktionen (LD1, LD2 und LD3) durchgeführt wird, ist in Abbildung 47 für das Untersuchungsgebiet Berlin und in Abbildung 48 für München dargestellt. Die Abbildungen zeigen jeweils die Gegenüberstellung von zwei Diskriminanzfunktionen, die trennenden Geraden und die Klassifizierung der einzelnen Merkmale auf Basis der geschätzten Diskriminanzfunktionen.

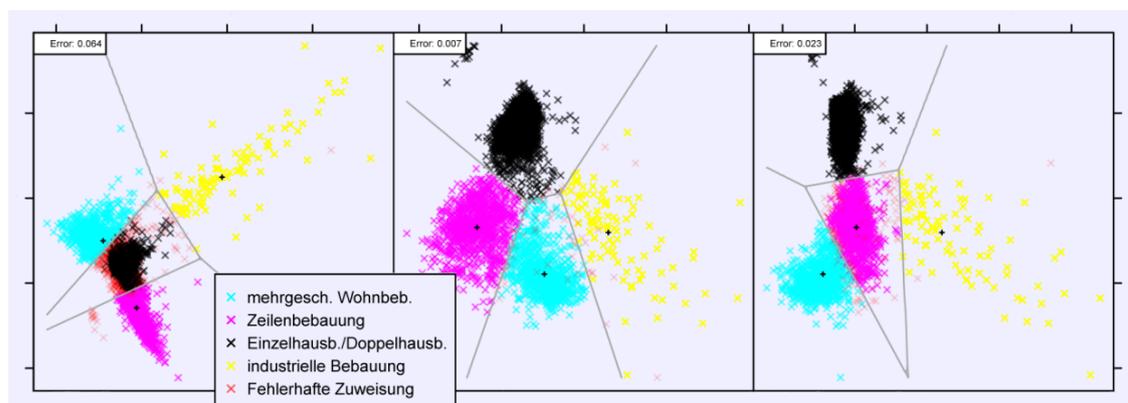
Da die Trefferquote der Klassifikation im Allgemeinen überhöht ist, wenn die Überprüfung auf Basis der selben Stichprobe wie die Eingangsstichprobe geschieht, wird aus dem gesamten erstellten Referenzdatensatz für beide Untersuchungsgebiete eine kleine Teilstichprobe (Trainingsdatensatz) erstellt. Auf dieser Basis werden die einzelnen Diskriminanzfunktionen aufgestellt und die Überprüfung mit den verbleibenden Referenzdaten durchgeführt. Durch diese überwachte Klassifikation können auf Basis eines zahlenmäßig kleineren Trainingsdatensatzes die verbleibenden Gebäudeobjekte klassifiziert werden. Die Ergebnisse dieser überwachten Klassifikation für beide Untersuchungsgebiete sind in Tabelle 17 für Berlin und in Tabelle 18 für München dargestellt.



**Abbildung 47:** Geometrische Abgrenzung zwischen den Gebäudeklassen auf Basis der 3 linearen Diskriminanzfunktionen (LD1, LD2 und LD3) für Berlin.

**Tabelle 17:** Genauigkeitsabschätzung für die Klassifikation auf Basis von Trainingssamples für Berlin.

Referenzdaten					
Klassifikation	Einzel-/Doppelh.	Zeilenbebauung	Mehrg. Wohnbeb.	Industrie	Nutzergen.
Einzel-/Doppelh.	1133	1	1	4	99,47 %
Zeilenbebauung	56	1368	11	2	95,20 %
Mehrg. Wohnbeb.	9	14	6405	1	99,63 %
Industrie	1	0	0	68	98,55 %
<i>Herstellergen.</i>	94,50 %	98,92 %	99,81 %	90,67 %	
<b>Gesamtgenauigkeit: 98,90 %    Anzahl der Beobachtungen: 9074, Anzahl der Samples: 332</b>					



**Abbildung 48:** Geometrische Abgrenzung zwischen den Gebäudeklassen auf Basis der 3 Diskriminanzfunktionen (LD1, LD2 und LD3) für München.

**Tabelle 18:** Genauigkeitsabschätzung für die Klassifikation auf Basis von Trainingssamples für München.

Klassifikation	Referenzdaten				Nutzerger.
	Einzel-/Doppelh.	Zeilenbebauung	Mehrg. Wohnbeb.	Industrie	
Einzel-/Doppelh.	840	16	0	12	96,77 %
Zeilenbebauung	104	919	61	12	83,85 %
Mehrg. Wohnbeb.	40	16	7394	0	99,25 %
Industrie	18	4	2	111	82,22 %
Herstellerger.	83,83 %	96,23 %	99,16 %	82,22 %	
<b>Gesamtgenauigkeit: 97,02 % Anzahl der Beobachtungen: 9549, Anzahl der Samples: 285</b>					

Inwiefern es möglich ist, allgemeingültige Trainingssamples für die Klassifikation der vier Gebäudeklassen zu erheben, welche direkt für die Klassifikation in unterschiedlichen Untersuchungsgebieten verwendet werden können, wird anhand der Übertragung der erhobenen Trainingssamples für Berlin und München auf das jeweilige andere Untersuchungsgebiet dargestellt. In der folgenden Tabelle 19 ist die Genauigkeitsabschätzung für die in München ausgewählten Trainingssamples und ihre Anwendung auf die Referenzdaten in Berlin dargestellt und in Tabelle 20 der umgekehrte Fall. Es ist zu erkennen, dass nur für die beiden Gebäudeklassen *Zeilenbebauung* und *mehrgeschossige Wohnbebauung* in der Übertragung eine ausreichende Trefferquote erzielt werden kann. Die anderen beiden Klassen *Einzel-/Doppelhausbebauung* und *industrielle Bebauung* weisen sehr hohe Überschneidungen untereinander auf. Vor dem Hintergrund dieser Beobachtung ist es also erforderlich, für jedes Untersuchungsgebiet eigene, an die lokalen baulichen und stadtstrukturellen Gegebenheiten angepassten Trainingssamples für die Klassifikation zu erheben.

**Tabelle 19:** Genauigkeitsabschätzung für die Übertragung der Trainingssamples von München auf Berlin.

<b>Referenzdaten (Berlin)</b>					
<b>Klassifikation</b>	Einzel-/Doppelh.	Zeilenbebauung	Mehrg. Wohnbeb.	Industrie	<i>Nutzergen.</i>
Einzel-/Doppelh.	138	152	132	0	32,70 %
Zeilenbebauung	15	1178	0	0	98,74 %
Mehrg. Wohnbeb.	4	53	6285	1	99,09 %
Industrie	1042	0	0	74	6,63 %
<i>Herstellergen.</i>	11,51 %	85,18 %	97,94 %	98,67 %	
<b>Gesamtgenauigkeit: 84,58 %    Anzahl der Beobachtungen: 9074, Anzahl der Samples: 285</b>					

**Tabelle 20:** Genauigkeitsabschätzung für die Übertragung der Trainingssamples von Berlin auf München.

<b>Referenzdaten (München)</b>					
<b>Klassifikation</b>	Einzel-/Doppelh.	Zeilenbebauung	Mehrg. Wohnbeb.	Industrie	<i>Nutzergen.</i>
Einzel-/Doppelh.	377	97	29	20	72,08 %
Zeilenbebauung	35	848	383	4	66,77 %
Mehrg. Wohnbeb.	36	10	7043	1	99,34 %
Industrie	554	0	2	110	16,52 %
<i>Herstellergen.</i>	37,62 %	88,80 %	94,45 %	81,48 %	
<b>Gesamtgenauigkeit: 87,74 %    Anzahl der Beobachtungen: 9549, Anzahl der Samples: 332</b>					

Nachdem die Gebäude einer der vier Gebäudeklassen zugewiesen wurden, werden alle Gebäude, welche nicht zuzuordnen sind, bzw. jene bei denen die Zuordnungswahrscheinlichkeit  $\leq 0,7$  beträgt, der Klasse *Sonstige* zugewiesen (vgl. Abbildung 42, S. 99).

Für die Verknüpfung der Gebäudeklassen mit den Surveydaten werden sie auf der räumlichen Ebene des Baublocks zu Bebauungstypen zusammengefasst (vgl. Abschnitt 3.1). Die Aggregation der thematischen Information erfolgt dabei nach der physischen Dominanz der einzelnen Gebäudeklasse. Dabei wird der Baublock jener Klasse zugewiesen für welche der Anteil der Summe der Gebäudevolumina am gesamten Gebäudevolumen aller Gebäude am größten ist. Die Ergebnisse der Klassifikation der Bebauungstypen sind in Abbildung 49 für Berlin und in Abbildung 50 für München abgebildet. In beiden Untersuchungsgebieten ist die Abfolge unterschiedlicher Bebauungstypen vom Zentrum bis zum Stadtrand zu erkennen. Beide Stadtzentren und ihre unmittelbaren Umgebungen sind durch überwiegend blockartige, mehrgeschossige Wohnbebauungen charakterisiert und in der Nähe des Stadtrandes in beiden Untersuchungsgebieten durch Einzel- und Doppelhausbebauungen abgelöst.

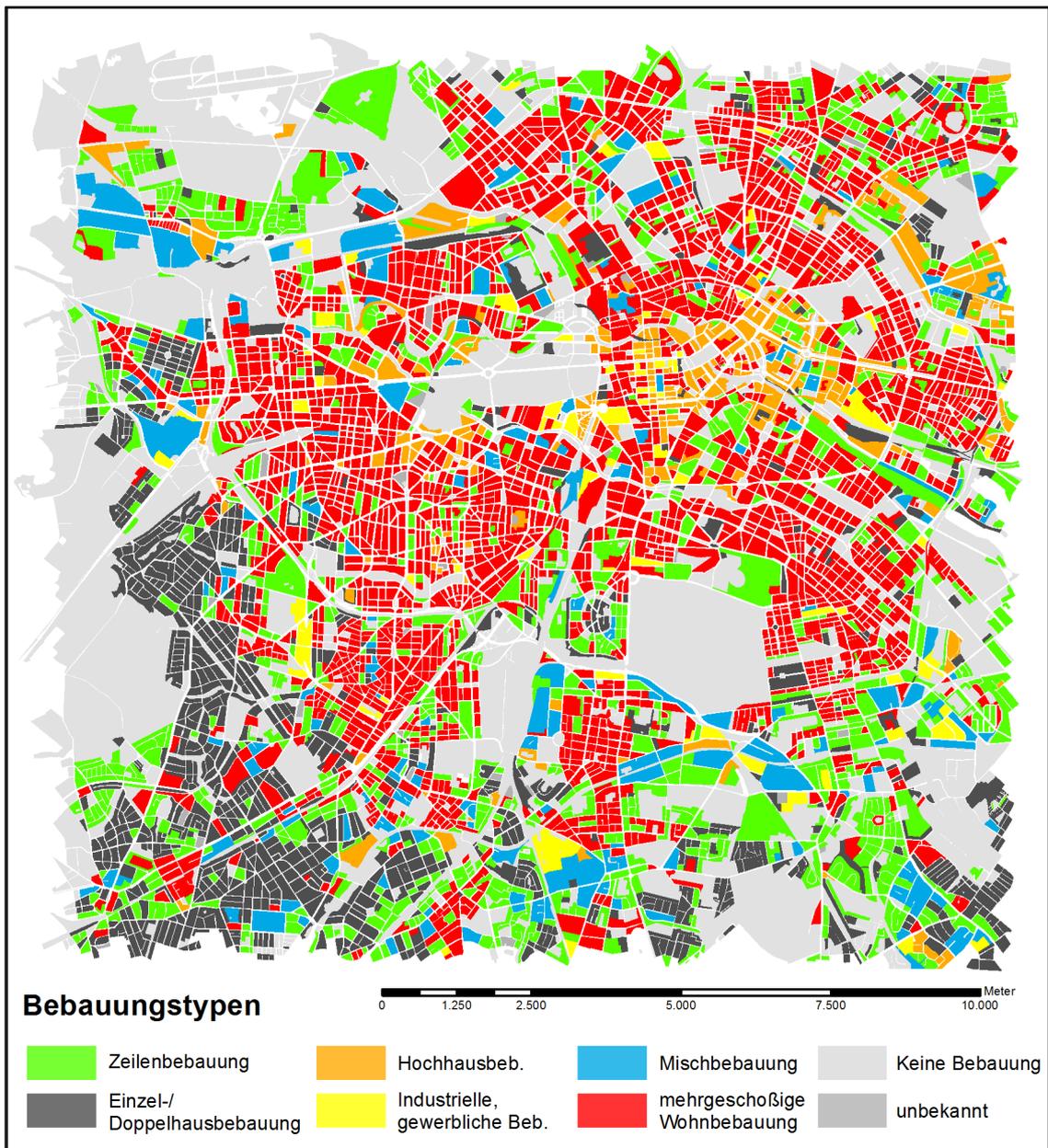


Abbildung 49: Ergebnis der Klassifikation der Bebauungstypen für Berlin.

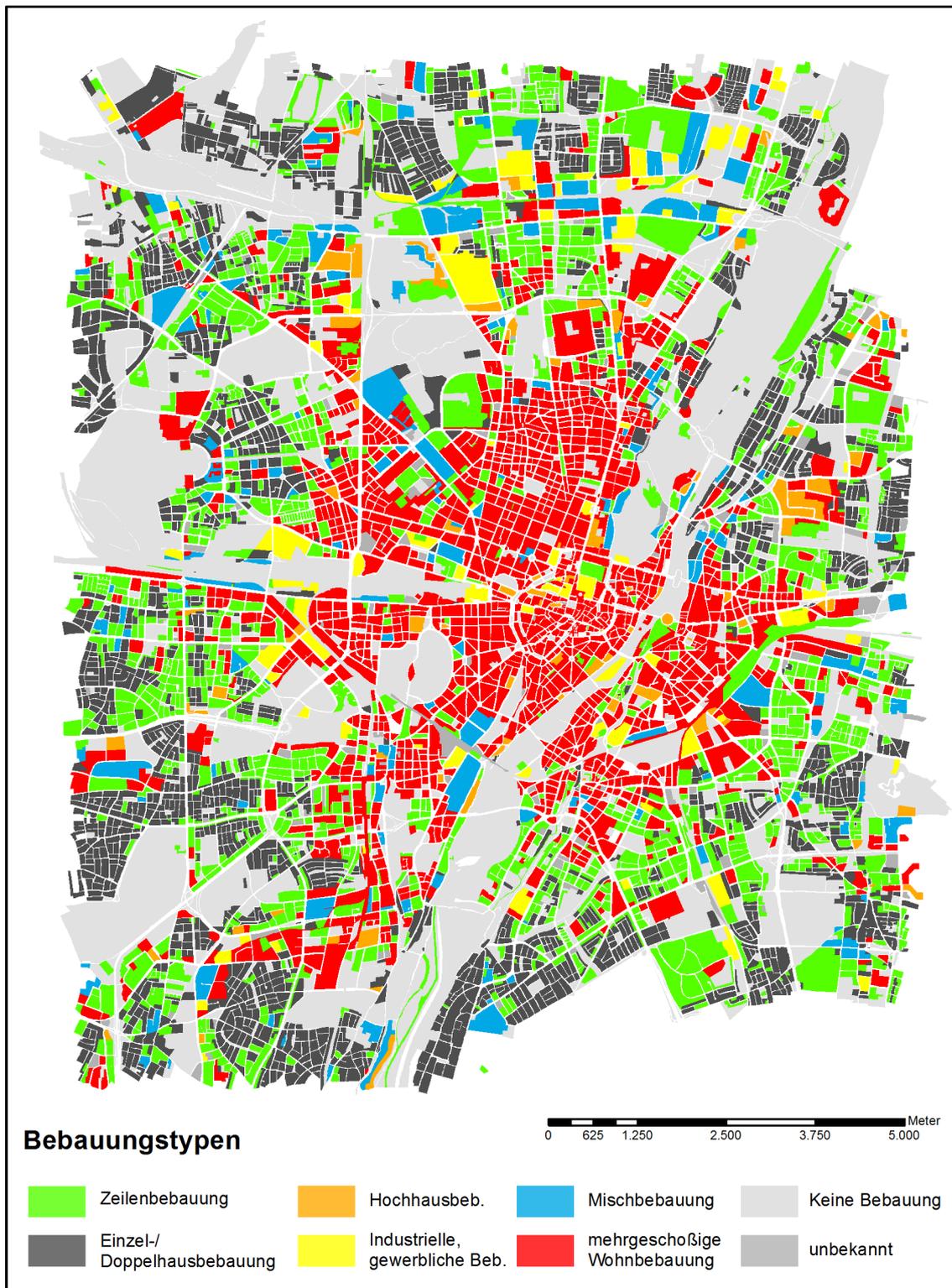
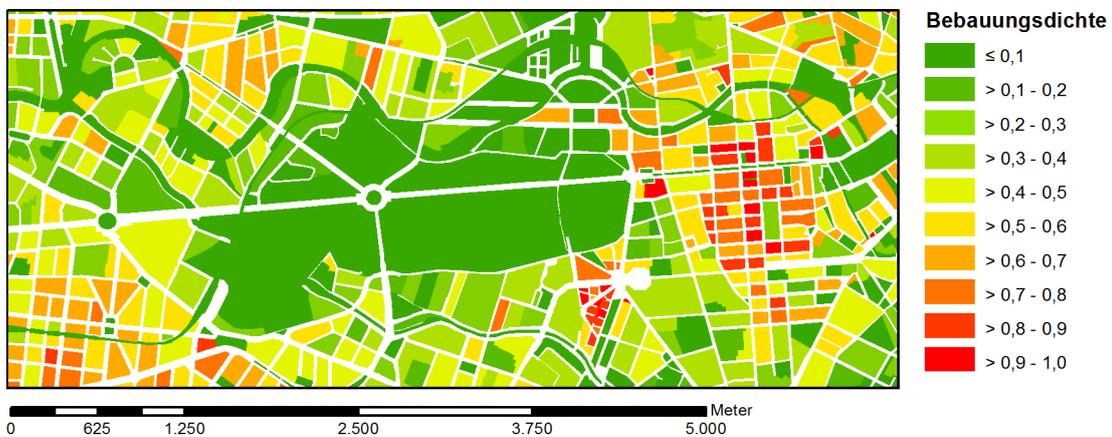


Abbildung 50: Ergebnis der Klassifikation der Bebauungstypen für München.

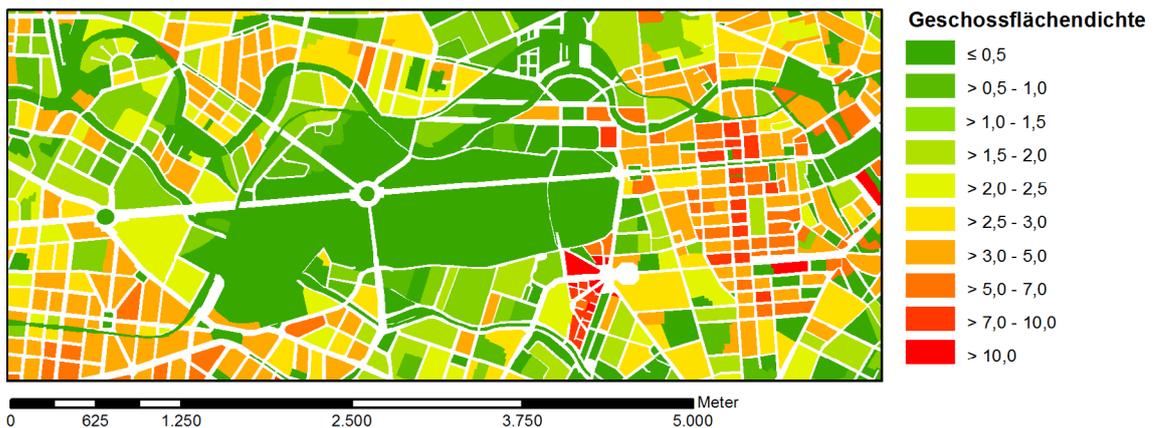
## 5.2. Bebauungs- und Geschossflächendichte

Die beiden Merkmale, welche die Dichte der Bebauung sowohl flächenmäßig als auch dreidimensional darstellen, sind die Bebauungs- und die Geschossflächendichte. Die Bebauungsdichte wird als Anteil der bebauten Fläche, also jener, die von Gebäuden belegt ist, zur Blockfläche berechnet. Die Geschossflächendichte wird aus dem Verhältnis der Summe der Fläche aller Vollgeschosse zur Blockfläche berechnet. Die Ergebnisse für diesen Arbeitsschritt sind exemplarisch für einen Ausschnitt des Berliner Tiergartens in Abbildung 51 und Abbildung 52 dargestellt.



**Abbildung 51:** Bebauungsdichte am Beispiel Berlins.

Baublockgrenzen: © Senatsverwaltung für Stadtentwicklung der Stadt Berlin.



**Abbildung 52:** Geschossflächendichte am Beispiel Berlins.

Baublockgrenzen: © Senatsverwaltung für Stadtentwicklung der Stadt Berlin.

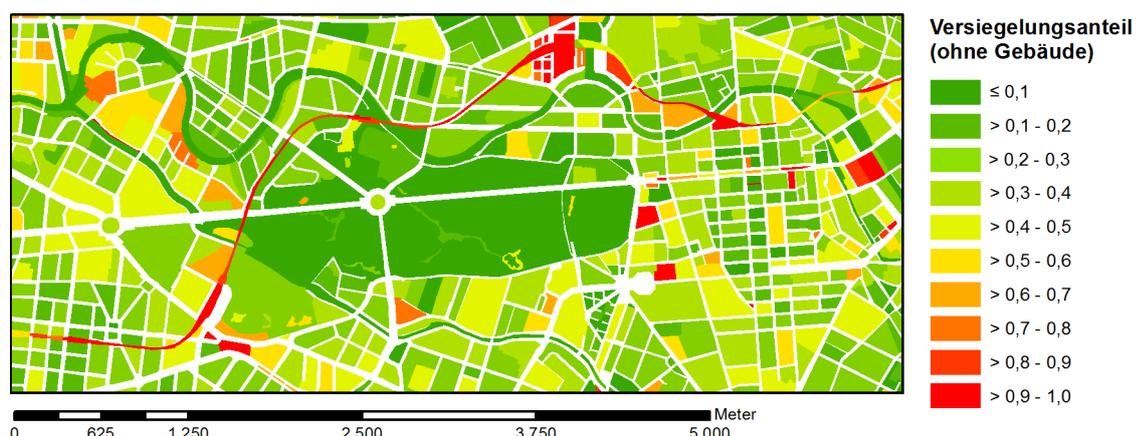
### 5.3. Vegetations- und Versiegelungsanteil

Neben den Dichten der bebauten Flächen werden auch die abgeleiteten Informationen über die Landbedeckung auf der Ebene des Baublocks aggregiert. Für die Berechnung des Vegetations- und Versiegelungsanteils werden die Flächenanteile der beiden Klassen für jeden Baublock berechnet (vgl. dazu auch Abbildung 14, S. 47). Der Vegetationsanteil stellt demnach den prozentualen Anteil der Vegetationsflächen pro Baublock dar. Der Versiegelungsanteil (ohne Gebäude) ist der Anteil der versiegelten Flächen ohne die Gebäudeflächen. Der gesamte Anteil an versiegelten Flächen pro Baublock ergibt sich aus der Summe der Bebauungsdichte (vgl. Abschnitt 5.2) und des Versiegelungsanteils (ohne Gebäude). Die Ergebnisse dieses Aggregationsschrittes sind exemplarisch für einen Ausschnitt um den Berliner Tiergarten in untenstehender Abbildung 53 und Abbildung 54 dargestellt.



**Abbildung 53:** Vegetationsanteil am Beispiel Berlins.

Baublockgrenzen: © Senatsverwaltung für Stadtentwicklung der Stadt Berlin.

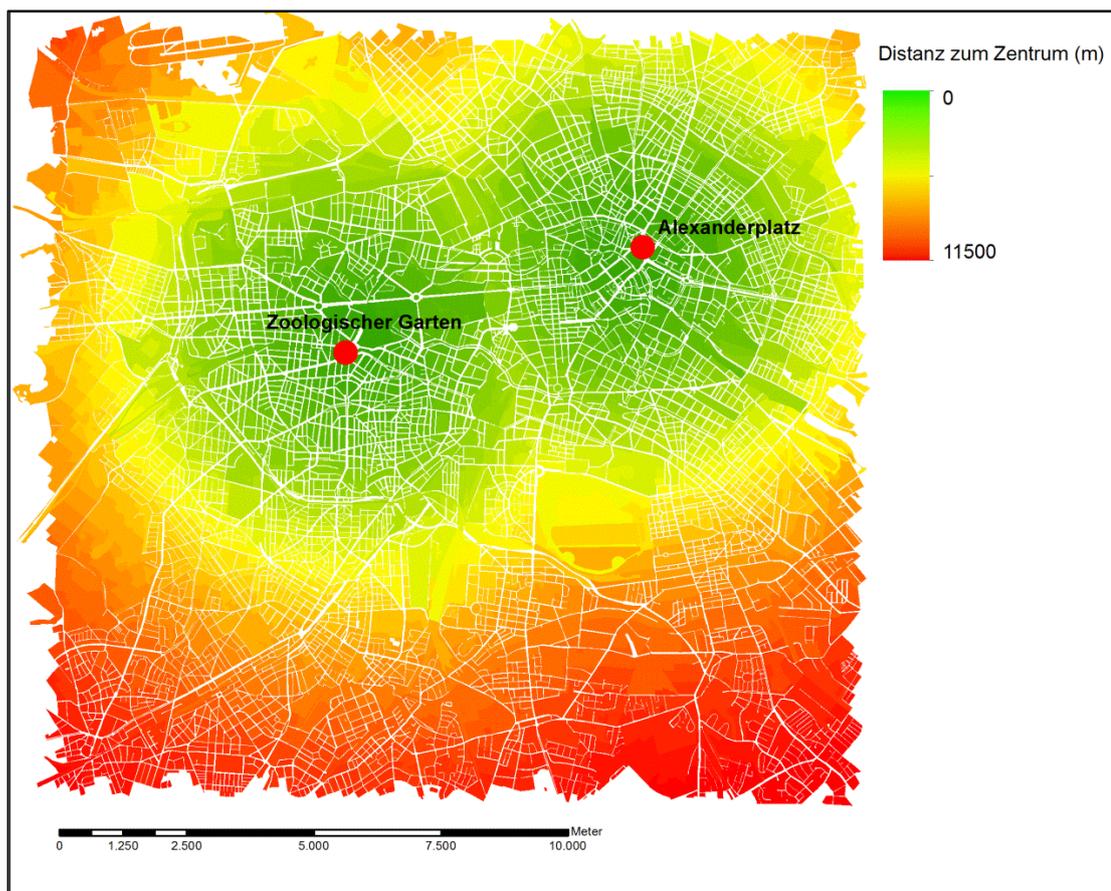


**Abbildung 54:** Versiegelungsanteil am Beispiel Berlins.

Baublockgrenzen: © Senatsverwaltung für Stadtentwicklung der Stadt Berlin.

## 5.4. Lage

Ein weiteres Merkmal, welches zwar nicht aus Fernerkundungsdaten abgeleitet wird, aber durch die geographische Verortung der Survey-Haushaltsadressen verfügbar gemacht werden kann, ist die *Lage* der Haushalte innerhalb des Stadtgebietes. Wie bereits in Abschnitt 2.3.3 dargestellt, weisen die Surveydaten nur einen indirekten Raumbezug auf, indem die Haushalte nach ihrer räumlichen Distanz zum nächsten Großstadtgeschäftszentrum kategorisiert werden. Durch die Verortung der Haushaltsadressen ist es möglich, verschiedenste räumliche Bezüge herzustellen. In der Arbeit wird das Merkmal *Lage* als direkte, absolute metrische Distanz zwischen dem Baublock und dem Geschäftszentrum ermittelt. Als Geschäftszentrum wurde für München der *Marienplatz* definiert. Für Berlin, das aufgrund seiner besonderen historischen Entwicklung eine Sonderstellung einnimmt, wurde ein Zentrum *West* (Bahnhof Zoologischer Garten) und ein Zentrum *Ost* (Alexanderplatz) definiert. Die Distanz des jeweiligen Baublockes zum Geschäftszentrum ist dann die jeweils kürzeste Distanz zu einem der beiden Zentren (vgl. Abbildung 55).



**Abbildung 55:** Distanz der Baublöcke zum Stadtzentrum am Beispiel Berlins.

Baublockgrenzen: © Senatsverwaltung für Stadtentwicklung der Stadt Berlin.

## 5.5. Zusammenfassung zur Gebäudeklassifikation und Aggregation zu Kontextmerkmalen

In diesem Kapitel wurde die Klassifikation von Kontextmerkmalen auf der räumlichen Ebene des Baublocks beschrieben. Die Merkmale wurden auf Basis der im vorangegangenen Kapitel beschriebenen abgeleiteten Informationen über das 3D-Gebäudemodell und die Landbedeckungsklassifikation auf der Ebene des Baublocks zusammengefasst. Die räumlichen Kontextmerkmale stellen aggregierte Informationen über die einzelnen Nachbarschaften im Stadtgebiet dar.

Diese Kontextmerkmale wurden flächendeckend für die beiden Untersuchungsgebiete erhoben und können über die geographische Verortung mit den einzelnen Befragungshaushalten aus den SOEP/BASE-II-Studien verknüpft werden. Es wurden verschiedene Merkmale zur *Dichte*, wie die *Bebauungsdichte*, und die *Geschossflächendichte* sowie *Vegetations-* und *Versiegelungsanteile* berechnet. Darüber hinaus wurde auch die *Lage* des Haushaltes innerhalb des Stadtgebietes als metrische Distanz zum Hauptgeschäftszentrum über die geographische Verortung der Wohnadressen ermittelt.

Einen zentralen Punkt in diesem Kapitel stellt die Klassifikation von semantischen Gebäudeklassen dar. Die Klassifikation wurde über die Form und Größe der einzelnen Gebäudeobjekte anhand eines überwachten Klassifikationsverfahrens durchgeführt. Es wurden 18 Formmerkmale auf ihre diskriminatorischen Eigenschaften für die Klassifikation untersucht. Die Formmerkmale reichen von einfachen, zweidimensionalen geometrischen Merkmalen, wie *Fläche* oder *Umfang* des Gebäudes, bis zu komplexen, dreidimensionalen Merkmalen. Auf Basis von repräsentativen Gebäuden für jeden Gebäudetyp wurden zunächst über lineare Diskriminanzfunktionen die relevanten Formmerkmale analysiert und ausgewählt. Im Rahmen der Untersuchung hat sich gezeigt, dass vorwiegend die zweidimensionalen Merkmale einen höheren Einfluss auf die Trennung zwischen den Gebäudeklassen haben. Die Formmerkmale wurden in weiterer Folge für die Schätzung der Diskriminanzfunktionen verwendet. Anhand dieser Diskriminanzfunktionen wurden alle Gebäude in beiden Untersuchungsgebieten klassifiziert. Vergleiche mit Referenzdaten zeigen hohe Genauigkeiten des Klassifikationsansatzes. Es konnte auch festgestellt werden, dass für jedes der Untersuchungsgebiete ein eigener Trainingsdatensatz erhoben werden muss, da eine Übertragbarkeit der Trainingsamples zwischen den Städten nur eingeschränkt möglich ist. Auf Basis der Gebäudeklassifikation wurde für jeden Baublock der dominierende Gebäudetyp ermittelt.

Die relevanten Kontextmerkmale *Lage*, *Dichte* und *Bebauung* wurden ebenfalls für jeden Baublock des Stadtgebietes berechnet. Diese Informationen können über das in Abschnitt 3.1 vorgestellte Konzept mit den Surveydaten verknüpft werden und im folgenden Kapitel

gemeinschaftlich ausgewertet werden. Zusammenfassend werden in Tabelle 21 die abgeleiteten Kontextmerkmale aufgelistet.

**Tabelle 21:** Räumliche Merkmale für die Einbindung in SOEP/BASE-II.

Variable	Beschreibung	Ausprägung
<b>Lage</b>	Distanz zum Geschäftszentrum	absolute Distanz in m
<b>Dichte</b>	Geschossflächendichte	Verhältnis der Summe der Flächen aller Vollgeschosse zur Bezugsfläche (Baublock)
	Bebauungsdichte	Prozentualer Anteil der bebauten Fläche (nur Gebäude)
	Versiegelungsanteil	Prozentualer Anteil der versiegelten Fläche (ohne Gebäude)
	Vegetationsanteil	Prozentualer Anteil der Vegetationsflächen
<b>Bebauung</b>	Bebauungstyp	Einzel-/Doppelhausbebauung Zeilenbebauung Mehrgeschossige Wohnbebauung Hochhausbebauung Industrielle/gewerbliche Bebauung Gemischte Bebauung



## Kapitel 6

# Sozialwissenschaftliche Analysepotenziale der Kontextmerkmale und der SOEP/BASE-II-Daten

*“Jeder Akteur ist charakterisiert durch den Ort, an dem er mehr oder minder dauerhaft situiert ist, sein Domizil [...], und durch die Position seiner Lokalisationen [...].”*

Pierre Bourdieu (1991)

Im aktuellen Kapitel werden die in den beiden vorangehenden Kapiteln abgeleiteten Informationen aus den Fernerkundungsdaten mit den SOEP/BASE-II-Daten verknüpft. Die abgeleiteten räumlichen Kontextmerkmale *Versiegelungsanteil*, *Vegetationsanteil*, *Bebauungsdichte*, *Geschossflächendichte* und *der Gebäudetyp* sowie auch die *Lage* wurden als räumliche Kontextmerkmale auf der räumlichen Ebene des Baublocks aggregiert und können über die verorteten Wohnadressen der Surveyteilnehmer mit den SOEP/BASE-II-Daten verknüpft werden. Die Adressen bleiben dabei dem Nutzer verborgen, da sie ausschließlich im Rahmen eines eigenen Datenschutzkonzeptes ausgewertet werden können (vgl. Abschnitt 3.1.1).

Die abgeleiteten Kontextmerkmale stehen im SOEP/BASE-II-Datensatz nach der Verknüpfung als zusätzliche Merkmale zur Verfügung. Sie können dann gemeinsam mit den anderen SOEP/BASE-II-Merkmalen ausgewertet werden. Dadurch ist es möglich, den Raum als Kontext, Ursache oder als Ergebnis von sozialen Prozessen zu untersuchen (Tickamyer, 2000).

Die Anwendungsmöglichkeiten von räumlichen Informationen in Verbindung mit sozialwissenschaftlichen Daten und Fragestellungen können sehr vielfältig sein und richten sich immer nach bestimmten Hypothesen, den verfügbaren Datengrundlagen und Auswerteverfahren. Vor diesem Hintergrund werden in diesem Kapitel Beispiele präsentiert, welche Untersuchungen auf Basis des erstellten Datensatzes möglich sind. Dabei soll anhand von exemplarischen Untersuchungen die Relevanz von fernerkundungsbasierten Methoden für sozialwissenschaftliche Studien aufgezeigt werden. Die Beispiele wurden ausgewählt, da sie zunächst die technische Verknüpfung der zusätzlichen räumlichen Daten demonstrieren und darüber hinaus die Potenziale für die inhaltliche Verknüpfung von fernerkundungsbasierter

räumlicher Informationen aufzeigen können. Sie sind aber auch nicht uneingeschränkt anwendbar bzw. bringt die Verknüpfung auch Limitierungen mit sich.

Dieses Kapitel ist in drei Abschnitte gegliedert. Im ersten Abschnitt wird der erstellte Datensatz aus der Verknüpfung der Kontextmerkmale mit den Surveydaten beschrieben. Dabei wird detailliert auf die Stichprobe und die Merkmale aus dem verknüpften Datensatz eingegangen. Der zweite Abschnitt fokussiert auf die Besonderheiten des räumlichen Bezuges des erstellten Datensatzes und vergleicht die räumlichen Merkmale aus den Fernerkundungsdaten und den Surveydaten. Im letzten Abschnitt werden anhand von statistischen Methoden der empirischen Sozialforschung Zusammenhänge und Abhängigkeiten zwischen bestimmten Eigenschaften oder Verhaltensweisen von Menschen mit der physischen Ausstattung der Stadt untersucht.

## 6.1. Verknüpfung der Kontextmerkmale und der Surveydaten

### 6.1.1. Stichprobenumfang

Von allen SOEP und BASE-II-Teilnehmern, welche in Berlin und München wohnen, wurden im Rahmen der Arbeit nur jene Haushalte mit den räumlichen Merkmalen verknüpft, welche innerhalb der in Abschnitt 3.3.1 definierten Ausschnitte der beiden Untersuchungsgebiete wohnen. Die Untersuchungsgebiete sind durch die räumlichen Abdeckungen der Erdbeobachtungsdaten begrenzt und decken daher nicht das gesamte administrative Gebiet der beiden Städte ab (vgl. Abbildung 16, S. 54).

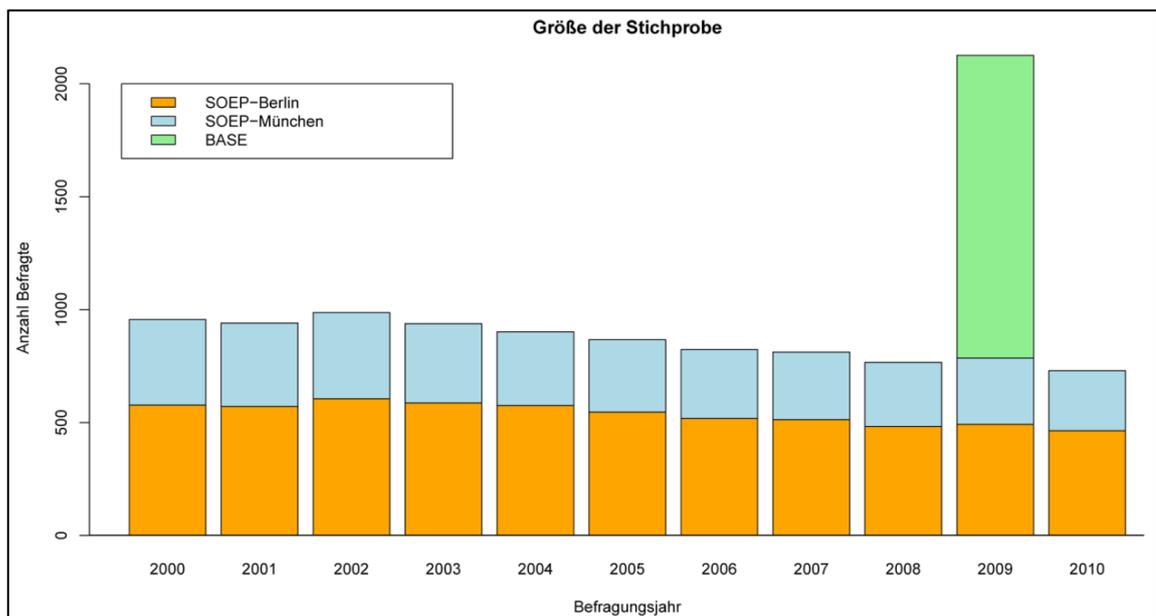
Es wurden die räumlichen Merkmale für insgesamt 895 unterschiedliche Haushalte aus der SOEP-Stichprobe für den Befragungszeitraum der Jahre 2000 bis 2010 verknüpft. Davon entfallen 329 Haushalte auf das Untersuchungsgebiet München und 566 auf Berlin.

Im Rahmen der Arbeit wurden zusätzlich zu den Daten des SOEP auch Daten aus den Haushaltsinterviews der *Berliner Altersstudie II (BASE-II)* aufgenommen. Diese Studie verwendet dieselben Personen- und Haushaltsfragebögen wie das SOEP und auch die Adressen wurden mit demselben Verfahren verortet. Deswegen ist es möglich, die BASE-II-Surveydaten zur Erhöhung der Stichprobe einzubinden. Es wurden aus BASE-II weitere 896 Haushalte aufgenommen, wodurch der gesamte erstellte Datensatz 1791 unterschiedliche Haushalte enthält, für welche die räumlichen Kontextmerkmale verknüpft wurden.

In diesen 1791 Haushalten wurden innerhalb des Beobachtungszeitraumes der Jahre 2000 bis 2010 insgesamt 3006 Einzelpersonen erfasst. Davon entfallen 2385 Einzelpersonen auf Berlin, wobei 1044 aus der SOEP-Studie und 1341 aus der BASE-Studie sind. Für München entfallen 621 Personen auf das Untersuchungsgebiet. Insgesamt weist der erstellte Datensatz 10849

Beobachtungen für 132 Variablen auf. Für München liegen 3576 Beobachtungen vor und für Berlin 7273 Beobachtungen, wobei 5932 aus dem SOEP und 1341 aus BASE-II stammen.

Wie bereits in Abschnitt 2.3.1 dargestellt, nimmt die Anzahl der Surveyteilnehmer kontinuierlich ab. Die Änderung der Anzahl der Befragungsteilnehmer für beide Untersuchungsgebiete ist in Abbildung 56 dargestellt. Sie zeigt eine tendenzielle Abnahme der Teilnehmer für die beiden Untersuchungsgebiete Berlin und München über den Beobachtungszeitraum zwischen den Jahren 2000 bis 2010, da das SOEP als Längsschnittstudie kontinuierlich Teilnehmer verliert. Die dadurch eventuell entstehenden Strukturverzerrungen werden durch Gewichtungen korrigiert. Die Vergrößerung der Stichprobe im Jahr 2009 ist auf die Daten der BASE-Studie zurückzuführen, welche in diesem Jahr erhoben wurden.



**Abbildung 56:** Anzahl der teilnehmenden Haushalte über den Untersuchungszeitraum für die einzelnen Stichproben.

### 6.1.2. Die Merkmale im verknüpften Datensatz

Die räumlichen Kontextmerkmale aus dem vorhergehenden Kapitel 5 wurden auf der räumlichen Ebene des Baublockes über die verorteten Haushaltsadressen der Surveyteilnehmer mit den SOEP/BASE-II-Daten verknüpft und stellen zusätzliche Variablen dar. Diese zusätzlichen räumlichen Merkmale können nach der Verknüpfung gemeinsam mit den Variablen aus den Surveydatensätzen ausgewertet werden<sup>2</sup>.

<sup>2</sup> Die Verknüpfung der Surveydaten mit den räumlichen Merkmalen wurde am DIW Berlin von Dr. Jan Goebel durchgeführt.

Für den in der Auswertung von sozialwissenschaftlichen Surveydaten nicht kundigen Leser wird an dieser Stelle der durch die Fusionierung erstellte Datensatz kurz beschrieben. Dazu ist in Tabelle 22 ein Auszug aus diesem Datensatz mit einigen beispielhaften Surveyvariablen auf der linken Seite (vgl. Abschnitt 2.3.3, S. 35) und einigen räumlichen Kontextmerkmalen auf der rechten Seite (vgl. Tabelle 21, S. 127) dargestellt. Eine zentrale Rolle bei der Bearbeitung des erstellten Datensatzes kommt der Personennummer (*P-ID*) zu. Sie ermöglicht eine eindeutige Identifikation der Person im Surveydatensatz. Jede Person im SOEP/BASE-II ist auch einem eindeutigen Haushalt und somit auch einer eindeutigen Haushaltsidentifikationsnummer (*H-ID*) zugeordnet. Somit können sämtliche Personen, welche im gleichen Haushalt wohnen, untersucht werden.

Die *P-ID* in der Tabelle ist verfälscht dargestellt, um eine Deanonymisierung der Person zu verhindern. Zusätzlich zu diesen Identifikatoren ist in der Tabelle auch das Jahr der Erhebung der jeweiligen Frage dargestellt, da nicht alle Merkmale in jedem Jahr erhoben werden. Sie werden entweder in unterschiedlichen Abständen in Form von Schwerpunktfragebögen (vgl. Wagner et al., 2008), bei Änderung der Haushaltssituation (z.B. Umzug) oder nur beim ersten Interview mit dem Haushalt erhoben. Merkmale, welche konstant bleiben (z.B. Distanz zum nächsten Großstadtzentrum), sind in der Aktualität der letzten Erhebung dargestellt.

Das *Jahr* der Erhebung der jeweiligen Information ist ebenfalls im Datensatz abgelegt. Die besondere Eigenschaft einer Längsschnittstudie ist, dass die teilnehmenden Personen über mehrere Jahre hinweg befragt werden. Einige Fragen zu den Variablen werden Jahr für Jahr gestellt, einige Variablen werden jedoch nur im Abstand von mehreren Jahren aktualisiert. Deshalb ist das Jahr der Information für die Auswertung von erheblicher Bedeutung, da sich nicht nur objektive Lebensumstände der Teilnehmer, sondern auch subjektive Einstellungen ändern können.

Aus den Surveyvariablen sind in der Tabelle noch die beiden Merkmale *Grünflächen* und *Zentrum* dargestellt. Die Variable *Grünflächen* stellt eine subjektive Einschätzung des Interviewten dar, inwiefern er sich durch den fehlenden Zugang zu Grünflächen in der Nachbarschaft beeinträchtigt fühlt (vgl. Abbildung 10, S. 37). Die Variable *Zentrum* stellt ebenfalls eine subjektive Einschätzung über die Lage der Wohnung des Interviewten in Bezug auf das räumlich nächstliegende Großstadtzentrum dar (vgl. Abbildung 6, S. 35). Auf diese beiden Variablen wird in den Beispielen in den folgenden Abschnitten wiederholt zurückgegriffen, weshalb sie an dieser Stelle nicht weiter beschrieben werden.

Auf der rechten Seite der Tabelle sind die räumlichen Kontextmerkmale aus den Fernerkundungsdaten und die Lage dargestellt. Der *Vegetationsanteil*, der *Versiegelungsanteil*, die *Bebauungsdichte* und die *Geschossflächendichte* beziehen sich auf die Anteile der jeweiligen Landbedeckungsart bzw. der Gebäude für den Baublock, in dem die Surveyteilnehmer wohnen. Die *Distanz* stellt die direkte metrische Distanz zwischen dem Baublock und dem Großstadtzentrum dar (vgl. Abschnitt 5.4).

Die Tabelle kann beispielhaft wie folgt interpretiert werden. Den ersten Eintrag stellt die Person mit der P-ID 1792 dar. Die Person ist weiblich und die raumbezogenen Merkmale des SOEP (vgl. Abschnitt 2.3.3) wurden das letzte Mal im Jahr 2004 erfragt. Die Person fühlt sich gar nicht durch fehlende zugängliche Grünflächen in der Nachbarschaft beeinträchtigt und die Wohnung liegt laut der Befragten im Zentrum der Stadt. Aus den verknüpften räumlichen Merkmalen aus den Fernerkundungsdaten und der Lage geht hervor, dass in dem Baublock der Wohnung beinahe ein Drittel der Fläche von Vegetation überlagert ist. Weitere 26,9 % der Fläche sind versiegelt und auf 40,9 % der Fläche stehen Gebäude. Die Distanz zum Stadtzentrum beträgt etwa 3600 m.

Die darauffolgenden drei Personeneinträge weisen darauf hin, dass alle Personen im gleichen Haushalt wohnen, da sie alle die gleiche H-ID besitzen. Es ist zu erkennen, dass bestimmte Merkmale und Merkmalsausprägungen für alle Personen im Haushalt gleich sind, während andere Merkmale sich zwischen den Personen unterscheiden. Dies sind *Haushaltsmerkmale*, welche nur durch den Haushaltsvorstand erhoben werden. Die personenbezogenen Merkmale werden von allen Personen im Haushalt einzeln erhoben. Die räumlichen Kontextmerkmale aus den Fernerkundungsdaten werden ebenfalls wie die Haushaltsmerkmale für alle Personen im Haushalt gleich vergeben. Dieser Umstand ist vor allem bei der quantitativen Auswertung der Daten relevant, da die Untersuchungsebene bei der Interpretation von erheblicher Bedeutung ist und die Daten in der statistischen Analyse gegebenenfalls angepasst werden müssen.

**Tabelle 22:** Auszug aus den verknüpften Einzeldatensätzen: SOEP/BASE-II (links) und die fernerkundungsbasierten räumlichen Merkmale und die Distanz (rechts) der Baublöcke, in denen die Surveyteilnehmer wohnen.

P-ID	Jahr	m/w	H-ID	Grünfl.	Zentrum	Veg.%	Vers.%	Beb.%	Dist.(m)	GFD
1792	2004	w	2248	gar nicht	Zentrum	32,2	26,9	40,9	36xx	2,39
3192	2004	m	2321	gering	Zentrum	12,7	17,2	70,0	41xx	4,23
3194	2004	w	2321	gering	Zentrum	12,7	17,2	70,0	41xx	4,23
3196	2004	m	2321	gering	Zentrum	12,7	17,2	70,0	41xx	4,23
4592	2004	m	2384	gar nicht	Zentrum	12,8	13,2	74,0	24xx	5,28
4592	2009	m	2384	gar nicht	Zentrum	69,7	19,2	11,2	49xx	0,15
4594	2004	w	68043	stark	Zentrum	28,0	25,6	46,4	19xx	2,72
4594	2009	w	68043	ger. ertr.	Zentrum	28,0	25,6	46,4	19xx	2,72
4598	2004	m	2384	gar nicht	Zentrum	12,8	13,2	74,0	24xx	5,28
5992	2004	m	2456	gering	Zentrum	18,4	15,8	65,7	8xx	3,94
5994	2004	w	2456	gering	Zentrum	18,4	15,8	65,7	8xx	3,94
5996	2004	m	2456	gering	Zentrum	18,4	15,8	65,7	8xx	3,94
5996	2009	m	2456	gar nicht	Zentrum	18,4	15,8	65,7	8xx	3,94
5992	2009	m	127547	gar nicht	Zentrum	66,8	22,0	11,2	86xx	0,19
5994	2009	w	127547	gar nicht	Zentrum	66,8	22,0	11,2	86xx	0,19
9392	2004	w	2631	gar nicht	< 10 km	70,8	11,2	18,0	77xx	0,35
9392	2009	w	2631	k. A.	Zentrum	70,8	11,2	18,0	77xx	0,35
10594	2004	w	2685	gar nicht	Zentrum	9,8	34,5	55,7	9xx	4,49
11792	2009	w	2744	gar nicht	< 10 km	69,0	6,9	24,1	51xx	0,65
11992	2004	m	116715	gar nicht	Zentrum	32,2	32,1	35,8	36xx	1,45
11994	2004	w	2758	gering	< 10 km	27,4	23,5	49,1	50xx	1,97
11994	2009	w	2758	gar nicht	< 10 km	44,3	21,3	34,4	60xx	1,62
11996	2004	w	123009	ger. ertr.	Zentrum	47,8	25,3	26,9	83xx	1,41
11996	2009	w	123009	gering	Zentrum	34,6	33,1	32,3	17xx	2,91
13192	2004	w	2817	gar nicht	< 10 km	74,1	16,0	9,6	47xx	0,28
17198	2004	w	123288	stark	Zentrum	33,1	22,8	44,0	22xx	2,67
...	...	...	...	...	...	...	...	...	...	...

**P-ID:** Personennummer (verfälscht), **Jahr:** Erhebungsjahr, **m/w:** Geschlecht, **H-ID:** Haushaltsnummer (verfälscht), **Grünfl.:** „Wie sehr fühlen Sie sich hier in dieser Wohngegend durch folgende Umwelteinflüsse beeinträchtigt?: Durch fehlende zugängliche Grünflächen“. „ger. ertr.“ = gerade erträglich, **Zentrum:** „Wie weit ist es von hier bis ins Zentrum der nächsten Großstadt?“, **Veg. %:** prozentualer Anteil der Vegetation im Baublock des befragten Haushaltes, **Vers. %:** prozentualer Anteil der versiegelten Fläche im Baublock des befragten Haushaltes, **Beb. %:** prozentualer Anteil der bebauten Fläche (Gebäude) im Baublock des befragten Haushaltes, **Dist.(m):** direkte Distanz zwischen dem Baublock des befragten Haushaltes und dem Geschäftszentrum der Großstadt in Metern. Die letzten beiden Stellen wurden aus Datenschutzgründen unkenntlich gemacht, **GFD:** Geschossflächendichte des Baublocks des befragten Haushaltes.

## 6.2. Der räumliche Bezug der Surveydaten und der räumlichen Kontextmerkmale

Der Großteil der personenbezogenen und der haushaltsbezogenen Merkmale aus den Surveys SOEP und BASE-II beziehen sich auf Eigenschaften, welche direkt die einzelne Person oder den Haushalt charakterisieren. Einige der Merkmale beziehen sich zudem auf die Lage des Haushaltes innerhalb des Stadtgebietes und die räumliche Ausstattung im Umfeld der Wohnung. Einige Beispiele hierfür wurden in Abschnitt 2.3.3 dargestellt. Sie beschreiben beispielsweise die räumliche Lage des Haushaltes als die subjektive Einschätzung des Befragten auf einer 6-stufigen Skala oder die subjektive Beeinflussung durch fehlende Grünflächen in der Nachbarschaft.

Da der Raumbezug der Kontextinformationen und auch der Surveydaten im Rahmen der Arbeit eine wesentliche Rolle spielen, wird im aktuellen Abschnitt auf die unterschiedlichen Raumbezüge eingegangen. Dazu werden die Variablen aus den Surveydaten mit den räumlichen Kontextmerkmalen aus den Fernerkundungsdaten gegenübergestellt.

### 6.2.1. Lage

Im Rahmen der haushaltsbezogenen Merkmale, welche in den Fragebögen erhoben werden, wird die Frage nach der Lage der Wohnung gestellt. Diese Frage wird entweder beim Erstinterview, bei Wohnortwechsel oder im Abstand von fünf Jahren beantwortet. Auf die Frage: „Wie weit ist es von hier bis ins Zentrum der nächsten Großstadt?“ antwortet der Teilnehmer mit einer der folgenden Möglichkeiten: „keine Angabe“, „Wohnung liegt im Stadtzentrum“, „unter 10 km“, „10 bis unter 25 km“, „25 bis unter 40 km“, „40 bis unter 60 km“ oder „60 km und mehr“ (vgl. Abbildung 6, S. 35). Diese Angaben zur Lage innerhalb des Stadtgebietes sind sehr grob, da eine räumliche Differenzierung der Surveyteilnehmer auf Basis dieser Angaben nur in Städten mit einer großen räumlichen Ausdehnung sinnvoll ist.

Darüber hinaus ist die räumliche Abgrenzung des Stadtzentrums nicht genau definiert. Der Begriff des *Stadtzentrums* kann sehr vielfältig interpretiert werden und hat in der Stadtgeographie unterschiedliche Bedeutungen. Des Weiteren ist die räumliche Abgrenzung des Stadtzentrums häufig von lokalen Strukturen abhängig und wird durch Individuen unterschiedlich aufgefasst. Vor diesem Hintergrund sind die Angaben der Lage in den Surveydaten subjektiv. Im Gegensatz zu diesen subjektiven Angaben können die gemessenen Distanzen über die Verortung der Haushaltsadressen objektive Aussagen über die Lage der Wohnungen im Stadtgebiet liefern (vgl. Abbildung 55, S. 125).

Im verknüpften Datensatz ist es möglich, die subjektiven Aussagen über die Lage innerhalb des Stadtgebietes den gemessenen Distanzen über die verorteten Haushaltsadressen

gegenüberzustellen. Es kann dabei untersucht werden, inwiefern sich die subjektiven Antworten mit den gemessenen Distanzen überschneiden und wo das subjektive Stadtzentrum liegt. Dafür werden zunächst die Einzeldatensätze von Personen in Tabelle 22 untersucht und danach aggregierte Informationen auf der Haushaltsebene betrachtet.

### 6.2.1.1. Personenebene

Bei Betrachtung der dargestellten Variablen der Person mit der (verfälschten) P-ID 4592, ist alleine auf Basis der SOEP-Variablen auf der linken Seite der Tabelle keine Änderung zwischen den beiden Erhebungsjahren 2004 und 2009 zu erkennen. Der Teilnehmer gibt als Lage seiner Wohnung in beiden Jahren an, dass die Wohnung im Zentrum liegt. Bei Betrachtung der Lage aus der Verortung der Wohnadresse ist erkennbar, dass sich die Distanz zum Großstadtzentrum zwischen den beiden Zeitpunkten von 2492 m auf 4931 m verdoppelt hat. Der Haushalt ist demnach innerhalb des Stadtgebietes umgezogen. Anhand der Koordinaten der verorteten Haushaltsadressen wäre es zusätzlich möglich, die exakten Positionen des Umzuges zu bestimmen und somit nicht nur die relative Distanz sondern auch die Lage des Haushaltes im Stadtgebiet zu untersuchen. Alleine auf Basis der SOEP-Variablen ist diese Information nicht verfügbar. Neben den dargestellten SOEP-Merkmalen in der Tabelle wird in den Interviews zwar erfragt, ob es gegenüber der vorherigen Befragung einen Umzug gegeben hat und wie groß die Distanz zum nächsten Großstadtzentrum ist bzw. wie die Nachbarschaft eingeschätzt wird, aber die Lage der beiden Wohnorte kann nicht genau bestimmt werden. Wenn sich das subjektive Merkmal *Zentrum* nicht ändert, kann lediglich die Aussage getroffen werden, ob es einen Umzug gegeben hat oder nicht. Wenn sich die reale Distanz aber nicht stark verändert hat, ist es nicht nachvollziehbar von wo nach wo der Haushalt umgezogen ist. Erst durch die exakte Verortung können diese räumlichen Beziehungen hergestellt werden.

Das folgende Beispiel beschreibt die Subjektivität der subjektiven Merkmale, die im Rahmen der Interviews erfragt werden. Da das subjektive Empfinden über einen Sachverhalt stets den persönlichen Eindruck des Befragten repräsentiert, ist es nur auf Basis der Surveydaten nicht nachvollziehbar, ob es sich dabei um eine tatsächliche Änderung des Sachverhaltes oder des subjektiven Empfindens handelt. Bei Betrachtung der Merkmale der SOEP-Teilnehmerin mit der P-ID 9392 in Tabelle 22, ist erkennbar, dass die Person im Jahr 2004 angegeben hat, weniger als 10 km vom Großstadtzentrum entfernt zu wohnen. Bei der erneuten Beantwortung der Frage im Jahr 2009 gibt die Teilnehmerin an, dass sich die Wohnung im Zentrum befindet. Wenn man dazu die metrische Distanz aus verorteten Haushaltsadressen betrachtet, erkennt man, dass die Distanz unverändert geblieben ist. Auch die Variable, welche angibt, ob ein Umzug stattgefunden hat, zeigt, dass es in der Zwischenzeit keinen Umzug gegeben hat (diese Variable ist in der Tabelle nicht aufgeführt). Es ist also deutlich, dass

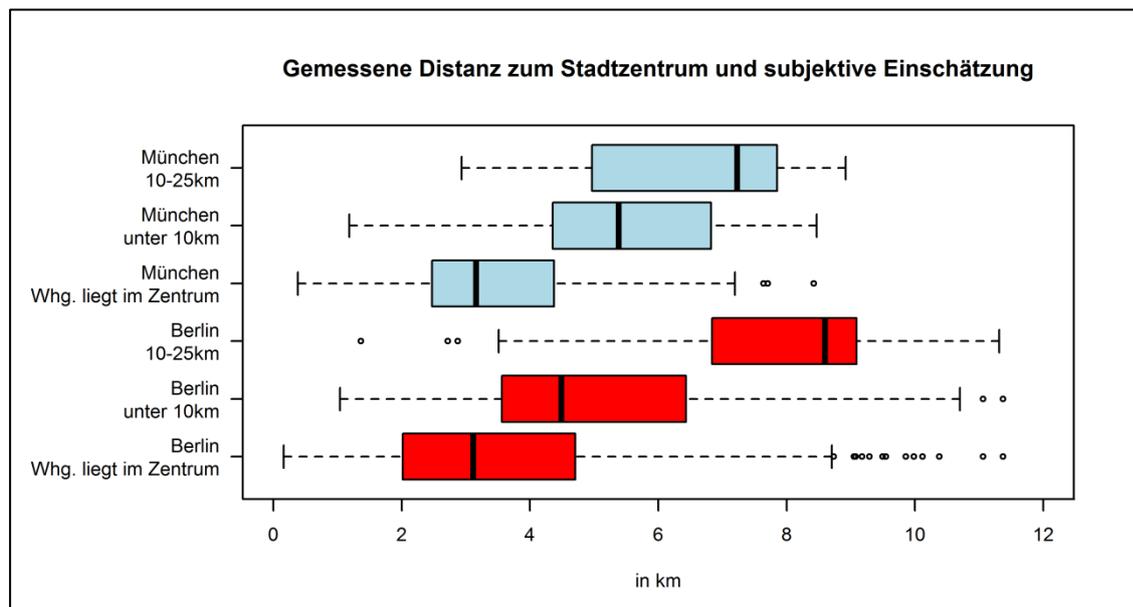
Aussagen über subjektive Merkmale der Befragten immer eine persönliche Einschätzung der Befragten darstellen, die sich auch über die Zeit ändern kann.

### 6.2.1.2. Haushaltsebene

Die oben dargestellten Beispiele über die räumlichen Lagebezüge zwischen den verorteten Haushaltsadressen und den Angaben der Umfrageteilnehmer basieren auf der Personenebene und auf Basis der Betrachtung von einzelnen Datensätzen. Es ist überdies hinaus möglich, die räumlichen Lagebezüge der Wohnungen auf der Haushaltsebene und für alle Umfrageteilnehmer zu untersuchen. Dies ist in Abbildung 57 in einem Boxplotdiagramm dargestellt, in dem die gemessenen Distanzen mit den gegebenen Antworten gegenübergestellt sind. Die Abbildung ist aufgeteilt nach den beiden Städten Berlin und München und zeigt die Verteilung der Distanzen für die ersten Antwortkategorien. Die Box repräsentiert den Interquartilsbereich, also die Hälfte aller gegebenen Antworten zwischen dem ersten und dem dritten Quartil. Der Median ist durch die vertikale Linie innerhalb der Box dargestellt und die Antennen repräsentieren den 1,5-fachen Interquartilsabstand. Die Punkte außerhalb zeigen die statistischen „Ausreißer“, also jene Werte, welche über oder unter dem 1,5-fachen Interquartilsabstand liegen.

Aus der Verteilung der Werte für beide Städte ist zu entnehmen, dass, je größer die subjektive Distanz zum Stadtzentrum ist, desto größer ist auch die gemessene Distanz. Im Detail betrachtet ist zu erkennen, dass in München die Hälfte aller Personen, welche angeben im Zentrum zu wohnen, unter 3,16 km vom gemessenen Stadtzentrum entfernt wohnen. Die Spannweite des Interquartilsabstandes bewegt sich dabei von 2,47 km bis 4,37 km. Die gemessenen Distanzen für jene Personen, welche geantwortet haben, dass ihre Wohnung nicht im Zentrum liegt, aber weniger als 10 km davon entfernt, liegen im Median bei 5,38 km. Der Interquartilsbereich bewegt sich dabei von 4,45 km bis 6,82 km. Es ist also eine deutliche Differenzierung der gemessenen Distanzen zwischen den beiden ersten Antwortkategorien zu erkennen. Weniger deutlich ist die Differenzierung zwischen der zweiten und dritten Antwortkategorie, da sich die Interquartilsbereiche stark überlappen. Jene Personen, welche angeben zwischen 10 und 25 km vom Zentrum entfernt zu wohnen, befinden sich tatsächlich zwischen 4,97 und 7,77 km vom Stadtzentrum entfernt. Im Median sind es 7,23 km.

In Berlin erweist sich die räumliche Abgrenzung des Zentrums schwieriger. Zwar liegen die Wohnungen der Hälfte aller Personen, welche angeben im Zentrum zu wohnen, ähnlich wie in München bei 3,11 km, allerdings ist die Spannweite des Interquartilsabstandes größer: zwischen 2,02 und 4,70 km. Auch ist die Abgrenzung zur nächsten Antwortkategorie nicht so deutlich ausgeprägt: Der Median der Wohnungen „unter 10 km“ liegt bei 4,49 km, das erste Quartil bei 3,57 km und das dritte Quartil bei 6,41 km. Das Stadtzentrum ist in Berlin demnach nicht so deutlich abgrenzbar wie für München.



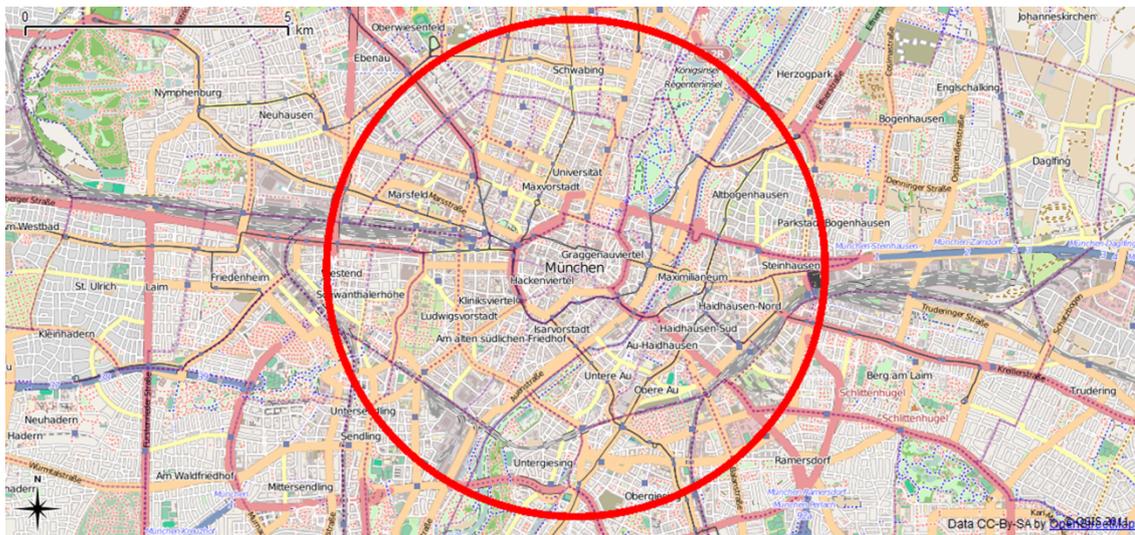
**Abbildung 57:** Vergleich zwischen gemessener Distanz der Wohnung zum Stadtzentrum und subjektiver Einschätzung durch die Surveyteilnehmer (SOEP und BASE-II).

### 6.2.1.3. Interpretation des emotionalen Raumbezuges für München

Über die oben angeführten, quantitativen Möglichkeiten die räumlichen Informationen aus den Surveys und die verorteten Haushaltsadressen zu analysieren, stellt der aktuelle Unterabschnitt einen sozialgeographischen Interpretationsansatz des Beispiels über das Großstadtzentrum dar. Die quantitativen, empirischen Analysepotenziale des verknüpften Datensatzes werden im folgenden Abschnitt 6.3 beschrieben.

Auf Basis der oben angeführten Betrachtungen ist die räumliche Abgrenzung des Stadtzentrums in München nicht so subjektiv ausgeprägt wie in Berlin. Dies ist auf Basis der historischen Entwicklung der Stadt und der darauf basierenden physischen städtischen Struktur auch nachvollziehbar. Die Bundesstraße 2R, oder auch *Mittlerer Ring*, gehört in München neben den historisch gewachsenen Strukturen auch zu den städtischen Ordnungssystemen, die die Stadt München gliedern (Landeshauptstadt München, 2000). Beide Seiten des Rings haben sich unabhängig voneinander entwickelt. Der Ring stellt demnach eine Grenze im städtischen Gliederungssystem dar, welche neben der physischen Ausprägung auch eine emotionale, subjektive Grenze in der Stadt darstellt. Diese Grenze ist auch im subjektiven Empfinden der städtischen Bewohner messbar. Die mittlere Distanz des mittleren Rings vom Stadtzentrum deckt sich weitgehend mit der gemessenen Distanz der Bewohner, welche nach ihrem subjektiven Empfinden im Zentrum Münchens wohnen. In Abbildung 58 ist der Median der Distanz für alle Zentrumsbewohner für München auf einer Karte aufgetragen (Median: 3,16 km). Die Distanz deckt sich weitgehend mit dem mittleren Ring. Auf diese Weise ist ein

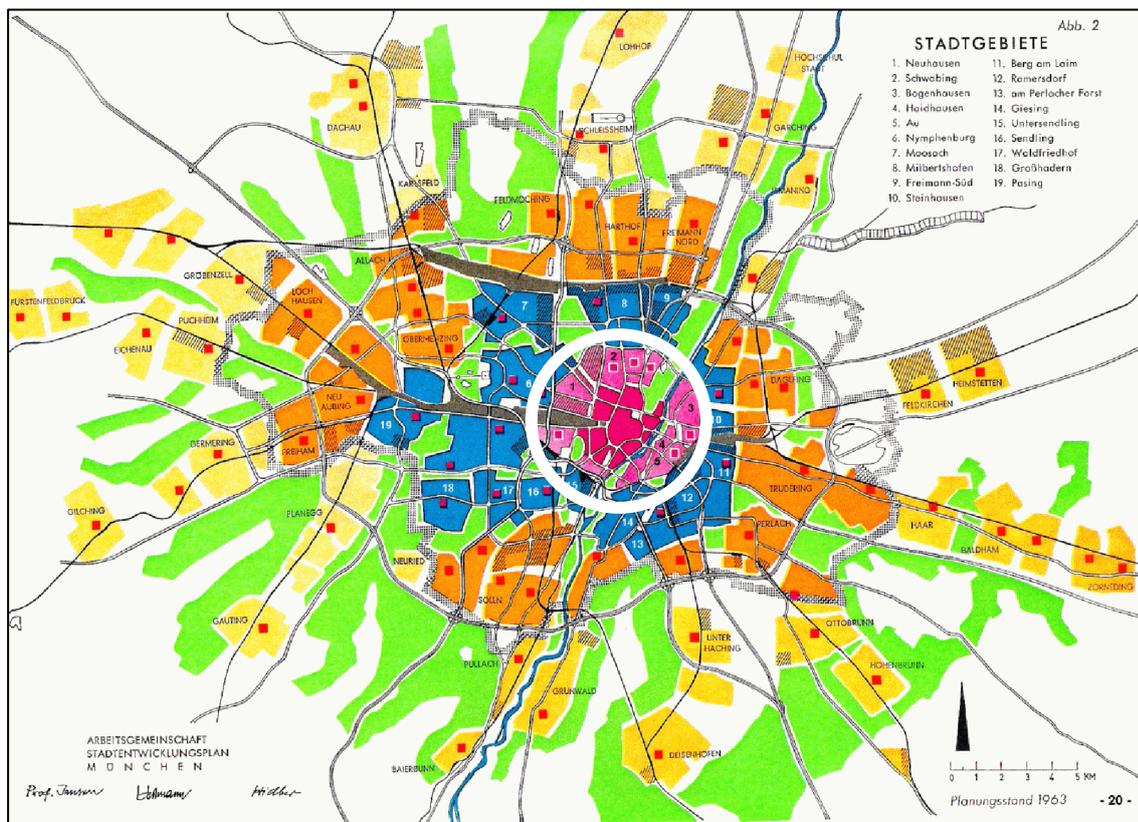
Zusammenhang zwischen der physischen städtischen Struktur und dem subjektiven Empfinden der Bewohner herstellbar. Eine vergleichbare Untersuchung für die Stadt Köln stellt ein ähnliches Bild dar (Wurm, Goebel und Wagner, 2010). Dort lässt sich auch die stadträumlich trennende Struktur des inneren Grüngürtels als ehemalige historische Grenze in der subjektiven Abgrenzung des Stadtzentrums nachvollziehen.



**Abbildung 58:** Mediantanz (3,16 km) der Haushalte im Großstadtzentrum (nach subjektiver Selbsteinschätzung) für das Untersuchungsgebiet München.

(© Openstreetmap, 2011)

Die räumliche Abgrenzung des Stadtzentrums für München deckt sich auch mit dem Stadtentwicklungskonzept für München aus dem Jahr 1963, das in Abbildung 59 dargestellt ist. Das *Citygebiet* und die *Innenstadtgebiete* sind die radialen Erweiterungen um das historische Stadtzentrum. Sie sind dargestellt in den Rottönen im zentralen Bereich der Abbildung (vgl. Reiß-Schmidt, 2010, S. 19). Zum Vergleich mit der vorhergehenden Abbildung ist auch die Mediantanz der Haushalte als weißer Kreis dargestellt.



**Abbildung 59:** Stadtentwicklungsplan von München aus dem Jahr 1963.

Der weiße Kreis markiert die Mediandistanz aus Abbildung 58.

(Quelle: Landeshauptstadt München, 1963)

Durch die Gegenüberstellung der unterschiedlichen räumlichen Lagebezüge zwischen den Surveydaten und den gemessenen Distanzen ist ein Mehrwert für das Analysepotenzial der räumlichen Auswertung darstellbar. Durch die Verortung der Wohnadressen der Haushalte sind eindeutige Beziehungen im Raum herstellbar. Die räumliche Lage der Haushalte ist auf Basis der Antworten in den Surveydaten subjektiv und birgt demnach entsprechende Schwierigkeiten bei der Auswertung dieser Informationen, da der Begriff des Zentrums in verschiedenen Städten unterschiedlich klar ausgeprägt sein kann. In der Kombination der subjektiven Angaben mit den gemessenen Distanzen können sich allerdings weitere Analysepotenziale ergeben, da somit die räumliche Lage des emotionalen Stadtzentrums auf Basis der Angaben von vielen Haushalten untersucht werden kann.

### 6.2.2. Grünanteil

Eine weitere Angabe in den Surveydaten zu subjektiven räumlichen Bezügen wird in einer Frage aus dem Haushaltsfragebogen erfragt. Sie behandelt die subjektive Umweltbeeinträchtigung in der Nachbarschaft. Die Frage lautet: „*Wie sehr fühlen Sie sich hier in dieser Wohngegend durch folgende Umwelteinflüsse beeinträchtigt: durch fehlende zugängliche Grünflächen?*“. Auf einer fünfstufigen Nominalskala können die Interviewteilnehmer zwischen folgenden subjektiven Beeinträchtigungen wählen: „*gar nicht*“, „*gering*“, „*gerade erträglich*“, „*stark*“ und „*sehr stark*“ (vgl. Abbildung 10, S. 37). Die Frage wird ebenfalls alle fünf Jahre gestellt, für die Haushalte in den Untersuchungsgebieten zum letzten Mal 2009 und davor im Jahr 2004.

Während die Fragestellung eindeutig auf einen räumlichen Bezug und dadurch entstehende subjektive Beeinträchtigungen in der Nachbarschaft abzielt, ist der räumliche Kontext der Fragestellung unklar. Die Frage bezieht sich auf eine subjektive Beeinträchtigung und ein subjektives räumliches Umfeld der Wohngegend. Verstärkt wird die Subjektivität durch die Formulierung der Fragestellung, da explizit auch auf die *Zugänglichkeit* der Grünflächen abgehoben wird. Ob sich demnach eine Person durch fehlende zugängliche Grünflächen in der Wohngegend beeinträchtigt fühlt, hängt demnach davon ab, wie die Person die räumliche Ausdehnung der Wohngegend definiert, ob die Person ein Bedürfnis für die Nutzung von Grünflächen hat und ob die Person Kenntnis über die Zugänglichkeit der Grünflächen hat.

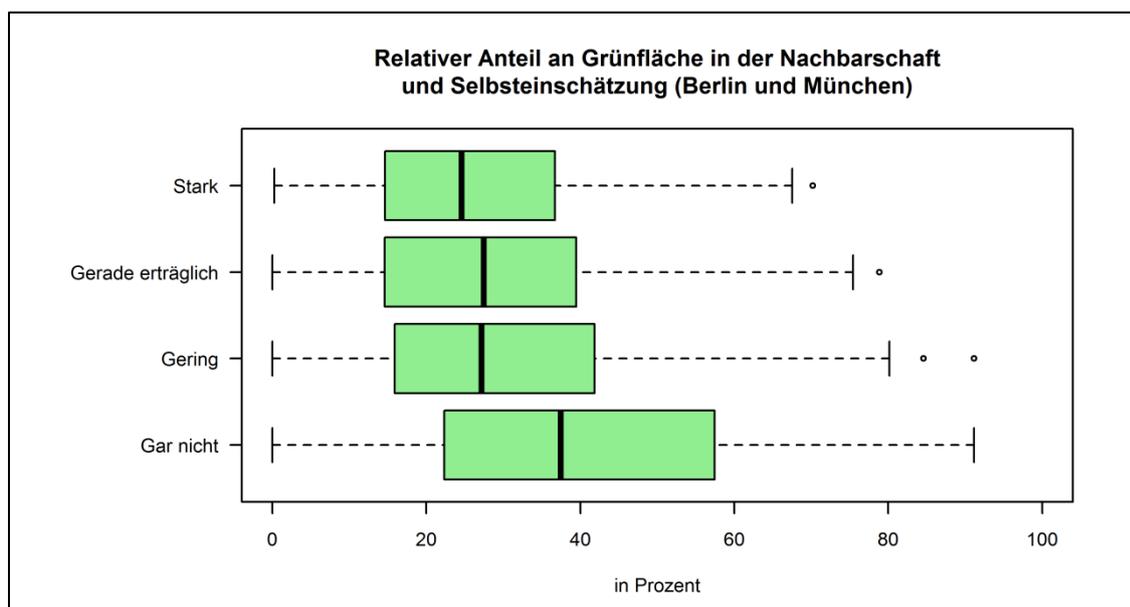
In den abgeleiteten räumlichen Kontextmerkmalen aus den Fernerkundungsdaten sind Informationen über den Vegetationsanteil für jeden Baublock festgehalten. Sie stellen den prozentualen Anteil von Vegetation für jeden Baublock dar (vgl. Abschnitt 5.3). Diese Messungen aus den Fernerkundungsdaten sind objektiv erhobene Informationen über die Verteilung von Vegetation und es können auf dieser Basis quantifizierbare räumliche Beziehungen der Vegetationsverteilung angestellt werden. Diese Berechnungen sind unabhängig von subjektiven Beurteilungen von Personen und können überdies flächenhaft durchgeführt werden. Dies stellt den entscheidenden Mehrwert der flächendeckenden Erhebung von räumlichen Merkmalen dar, da auch für jene Bereiche Informationen verfügbar sind, für welche keine Surveydaten vorliegen. Es kann also der räumliche Bezug der Daten in Abhängigkeit einer sozialwissenschaftlichen Fragestellung individuell definiert werden. Die Vegetationsanteile können demnach nicht nur für die unmittelbare Wohngegend der Bewohner berechnet werden, sondern auch für größere Nachbarschaften oder Stadtteile. Auch können direkte Distanzen zwischen Surveyteilnehmern und bestimmten Grünflächen im Stadtgebiet berechnet werden.

Die Implikationen aus der Subjektivität der Fragestellung nach der Beeinträchtigung durch fehlende zugängliche Grünflächen lassen sich nicht aus objektiven Messungen darstellen. Es

kann allerdings in diesem Zusammenhang untersucht werden, inwieweit sich die subjektiven Angaben mit den objektiven Messungen überschneiden.

Für den Vergleich sind in Abbildung 60 Boxplots dargestellt. Sie zeigen die Gegenüberstellung zwischen dem gemessenen Vegetationsanteil und der subjektiven Beeinträchtigung durch fehlende Grünflächen gemeinsam für Berlin und München. Auf der Abszisse ist der prozentuale Anteil der Vegetation pro Baublock aufgetragen und auf der Ordinate die vier unterschiedlichen, nominal skalierten Antwortmöglichkeiten aus SOEP/BASE-II. Für jede der vier Antwortkategorien ist die Verteilung der Vegetationsflächen der Baublöcke der Surveyteilnehmer dargestellt.

Für jene Haushalte, welche keine Beeinträchtigungen mangels zugänglicher Grünflächen in der Wohngegend empfinden, ist auch der Vegetationsanteil pro Baublock mit rund 37,5 % im Median am höchsten. Für die Hälfte aller gegebenen Antworten liegt der Anteil der Vegetation zwischen 22,3 und 57,4 %. Bei den weiteren Antwortkategorien liegt der Median 27,2 % für *geringe* Beeinträchtigungen, bei 27,5 % für *gerade erträgliche* Beeinträchtigungen und bei 24,6 % für *starke* Beeinträchtigungen. In der Betrachtung der Verteilungen der Messungen und der subjektiven Angaben ist erkennbar, dass mit steigender Beeinträchtigung durch fehlende Grünflächen auch der Vegetationsanteil pro Baublock sinkt. Eine nach den beiden Untersuchungsgebieten differenzierte Darstellung ist in Anhang C abgebildet. Sie zeigt eine ähnliche Tendenz für die beiden Untersuchungsgebiete wobei in München der Vegetationsanteil tendenziell höher liegt als in Berlin, allerdings ist auch die Streuung des Vegetationsanteils in München höher als in Berlin.



**Abbildung 60:** Vergleich zwischen Grünanteil in der Nachbarschaft basierend auf Klassifikationsergebnissen aus Erdbeobachtungsdaten und subjektiver Einschätzung (SOEP/BASE-II).

Die Frage zur subjektiven Beeinträchtigung aufgrund fehlender zugänglicher Grünflächen in der Wohngegend ist in mehreren Punkten subjektiv und vor allem räumlich nicht exakt definiert. Die Angaben der Befragten zur Beeinträchtigung reflektieren daher immer das persönliche Empfinden der Personen zur räumlichen Abgrenzung der Wohngegend, zum Bedürfnis der Nutzung zugänglicher Grünflächen und zur Größe sowie zur Verteilung von Grünflächen. Die Messung der Vegetationsflächen aus den Fernerkundungsdaten stellt ein objektives Verfahren dar, um Anzahl, Verteilung und Größe von Vegetation im Raum eindeutig zu quantifizieren. Über diese objektiven Angaben können allerdings keine Aussagen über die Nutzung der Grünflächen getroffen werden. Durch die Kombination von objektiver, flächendeckender Information und subjektiven Aussagen können die subjektiven Aussagen mit einem objektiven Bezug interpretiert werden.

### 6.2.3. Art der Bebauung

Das dritte Merkmal in den Surveydaten mit einem expliziten räumlichen Bezug, für welches vergleichbare Informationen in den räumlichen Kontextmerkmalen bestehen, bezieht sich auf das Wohnumfeld. Der räumliche Bezug der Wohnumfeldmerkmale in den Surveydaten ist sehr unterschiedlich. Sie reichen von punktuell erhobener Information bis zu einer undefinierten Nachbarschaft. Diese *Wohnumfeldmerkmale* werden durch den Interviewer erhoben (vgl. Abbildung 7-Abbildung 9, S. 36f). Da diese Informationen ohne den für die anderen Fragen notwendigen Kontakt zwischen dem Interviewer und dem Befragungsteilnehmer erhoben werden, werden diese Daten als sogenannte *Brutto-Informationen* erhoben. Sie stehen somit für alle durch die Stichprobe gezogenen Haushalte zur Verfügung. Da die Informationen nicht bei den Haushalten erfragt werden können, die verweigern, machen vielmehr die Interviewer diese Angaben. Der Gebäudetyp wird in acht unterschiedlichen Kategorien erhoben: *Landwirtschaftliches Wohngebäude, freistehendes Ein-/Zweifamilienhaus, Ein-/Zweifamilienreihenhaus, Wohnhaus mit 3-4 Wohnungen, Wohnhaus mit 5-8 Wohnungen, Wohnhaus mit mehr als 9 Wohnungen, Hochhaus* und *sonstiges Gebäude*.

Der Haustyp wird exakt für jenes Haus erhoben, in welchem der Surveyteilnehmer wohnt. Er wird also als punktuelle Information erhoben, ohne Informationen über die Nachbarschaft oder die umliegenden Gebäude zu erheben. Aus den Surveydaten ergibt sich somit ein unvollständiges, lückenhaftes Bild über die Bebauungsstrukturen im Stadtgebiet. Die einzige Differenzierung über die Bebauung in der Nachbarschaft wird als die überwiegende Nutzung der Wohngegend beschrieben, wobei zwischen Altbauten und Neubauten unterschieden wird. Eine weitere Unterscheidung der Bebauungsstruktur in der Nachbarschaft des Haushaltes wird nicht vorgenommen. Darüber hinaus ist, wie in den beiden vorangehenden Beispielen der räumliche Bezug der *Wohngegend* nicht genau definiert.

Die Klassifikation der Bebauungsstruktur auf Basis der Fernerkundungsdaten wurde flächendeckend für die beiden Untersuchungsgebiete durchgeführt und es wurde ein Verfahren verwendet, welches auf andere Städte angewendet werden kann. Somit ist es möglich, einen flächendeckenden, objektiven Datensatz über die Bebauungsstruktur im gesamten Stadtgebiet zu erstellen und für die Auswertung mit den Surveydaten zu verwenden.

Wie in den vorangehenden Beispielen, in denen eine Gegenüberstellung der räumlichen Bezüge aus den Surveydaten und der räumlichen Kontextmerkmale durchgeführt wird, ist auch in der folgenden Gegenüberstellung weder die taxonomische noch die räumliche Dimension der beiden Informationen kohärent, weshalb eine vollständige Übereinstimmung der beiden Daten a priori ausgeschlossen werden kann. Die Bebauungsstruktur auf Basis der Fernerkundungsdaten wurde als dominierender Bebauungstyp für jeden Baublock klassifiziert, der Haustyp in den Surveydaten bezieht sich auf eine punktuelle Information. Trotz der räumlichen und thematischen Unterschiede werden an dieser Stelle die Informationen aus den Surveydaten und der räumlichen Kontextmerkmale gegenübergestellt, da eine teilweise Übereinstimmung der Informationen zu erwarten ist. Die Gegenüberstellung zwischen den beiden Informationen ist in Tabelle 23 dargestellt, wobei horizontal die Haustypen aus den Surveys und vertikal die dominierenden Bebauungstypen eingetragen sind.

**Tabelle 23:** Vergleich zwischen klassifizierten Bebauungsstrukturen und Wohnumfeld aus den Surveys für Berlin und München.

	k.A.	sonst.	Landw.	1/2	1/2 Reihe	WH 3-4	WH 5-8	WH >9	HH	%
<b>MGW</b>	481	29	3	2	55	84	<b>1250</b>	<b>4008</b>	100	55,42
<b>Zeile</b>	149	12	5	87	<b>300</b>	107	744	889	57	21,66
<b>1/2</b>	61	2	11	<b>379</b>	<b>481</b>	161	263	226	5	14,65
<b>HH</b>	29	1	0	0	0	0	9	88	<b>224</b>	3,24
<b>I/G</b>	2	0	0	0	0	0	16	17	22	0,53
<b>Misch.</b>	6	0	0	0	29	17	67	59	11	1,74
<b>Unbek.</b>	43	8	0	2	0	6	61	167	14	2,76
<b>%</b>	7,11	0,48	0,18	4,33	7,97	3,46	22,21	50,27	3,99	$\Sigma$ 10849

k.A.: keine Angabe, sonst.: sonstige Gebäude, Landw.: Landwirtschaftliches Wohngebäude, 1/2: freistehendes Ein- oder Zweifamilienhaus, 1/2 Reihe: Ein- oder Zweifamilienreihenhaus, WH 3-4: Wohnhaus mit 3-4 Wohnungen, WH 5-8: Wohnhaus mit 5-8 Wohnungen, WH >9: Wohnhaus mit mehr als 9 Wohnungen, HH: Hochhaus, MGW: Mehrgeschossige Wohnbebauung, Zeile: Zeilenbebauung, 1/2: Einzel-/Doppelhausbebauung, HH: Hochhausbebauung, I/G: industrielle / gewerbliche Nutzung, Misch.: Mischbebauung, Unbek.: unbekannt, keiner der anderen Klassen zugehörig.

Der Tabelle ist zu entnehmen, dass für jene Bebauungstypen bzw. Haustypen, welche sinngemäß in beiden Datensätzen vorkommen, auch jeweils die größte Übereinstimmung vorhanden ist. So findet sich der Großteil der als *Einzel-/Doppelhausbebauung* klassifizierten Gebäude auch in der Kategorie *freistehende Ein- oder Zweifamilienhausbebauung*. Ebenso finden sich die meisten Übereinstimmungen zwischen *mehrgeschossiger Wohnbebauung* und *Wohnhaus mit 5-8 bzw. Wohnhaus mit mehr als 9 Wohnungen*, sowie für die Kategorie *Hochhaus*. Für die Kategorie *Ein- und Zweifamilienreihenhaus* sind die meisten Übereinstimmungen sowohl in der

Klasse *Zeilenbebauung* als auch in der Klasse *Einzel-/Doppelhausbebauung* zu finden. In dieser Kategorie ist die physische Ähnlichkeit zwischen den Gebäudestrukturen am größten.

#### 6.2.4. Zusammenfassung zum räumlichen Bezug

Durch die Gegenüberstellung der Surveydaten und der abgeleiteten Kontextmerkmale wird der unterschiedliche räumliche Bezug der Daten erkennbar. Die wenigen Fragestellungen in den Surveydaten, welche einen direkten oder indirekten Raumbezug aufweisen, lassen sehr unterschiedliche Interpretationen des räumlichen Bezuges zu. Dies ist auch vielmehr dem Umstand geschuldet, dass für die Auswertung der Surveydaten vordergründig die individuellen Auswirkungen aus der Nachbarschaft auf die jeweilige Person von Interesse sind. Diese Beeinflussungen sind allerdings besonders subjektiv, da einige Fragen ein breites Interpretationsfeld offen lassen.

So ist wie im Beispiel der Lage die Begriffsdefinition des Stadtzentrums je nach Forschungsdisziplin unterschiedlich, es hat aus soziologischer Sicht eine andere Bedeutung als aus ökonomischer Sicht. Die räumliche Abgrenzung des Stadtzentrums ist darüber hinaus sehr individuell. Über die genauere Verortung der Lage der Haushalte kann eine eindeutige Distanz zum Stadtzentrum oder zu anderen geographischen Punkten sowie zu den anderen Survey-Teilnehmern berechnet werden und als Grundlage für die Beschreibung objektiver, räumlicher Beziehungen dienen. Am Beispiel für München wurde dargestellt, inwiefern eine Verbindung zwischen dem emotionalen, subjektiven geographischen Bezug einer Person zu seiner städtischen Umwelt hergestellt werden kann. Im Vergleich der subjektiven Angaben mit stadtstrukturellen Daten kann eine räumliche Übereinstimmung des Empfindens mit den *City-* und *Innenstadtgebieten* der Stadt festgestellt werden. Gegenüber dem Surveymerkmal *Lage*, welches nur die vom Surveyteilnehmer geschätzte Distanz zum nächsten Großstadtzentrum beinhaltet, ist es auf Basis einer exakten geographischen Verortung der Lage des Haushaltes möglich, auch Umzüge innerhalb des Stadtgebietes genauer als bisher zu lokalisieren. Wie an einem Beispiel in Tabelle 22 in Abschnitt 6.1 dargestellt, unterliegen die Teilnehmerangaben der Subjektivität der Befragten, sie können sich ohne Änderung der Lebensumstände alleine durch die Subjektivität des Interviewten ändern. Auch ohne tatsächliche räumliche Änderung der Wohnung, kann die Angabe des Teilnehmers über die Jahre variieren. Darüber hinaus ist der geographische Bezugspunkt in der Fragestellung nicht exakt definiert. Augenscheinlich wird dieser Umstand am Beispiel Berlins: durch die besondere historische Entwicklung gab es über viele Jahrzehnte zwei Stadtzentren. Erst durch eine exakte Verortung der Wohnung kann die Lage des Haushaltes genau bestimmt werden.

Doch nicht nur die Lage bleibt in den Surveydaten sehr subjektiv, sondern auch weitere Merkmale mit räumlichem Bezug. In der Frage nach der Beeinträchtigung durch fehlende zugängliche Grünflächen in der Nachbarschaft, bleibt die räumliche Abgrenzung der

Nachbarschaft unklar. Durch die Subjektivität der Interpretationsmöglichkeiten kann aus den Antworten nur die subjektive Beeinflussung berücksichtigt werden. Eine Veränderung der tatsächlichen Grünflächen in der Nachbarschaft muss sich nicht zwangsläufig in den gegebenen Antworten widerspiegeln und umgekehrt. Die flächendeckende Erhebung der Vegetationsflächen aus Fernerkundungsdaten bietet in diesem Zusammenhang eine objektive Möglichkeit, die Informationen lückenlos zu erheben und mit den weiteren Informationen der Surveyteilnehmer zu verbinden. In der Gegenüberstellung der relativen Vegetationsanteile aus den Fernerkundungsdaten und der subjektiven Angaben der Befragten, ist ein Zusammenhang zwischen gemessener Grünfläche und subjektiver Einschätzung feststellbar: je höher der Vegetationsanteil pro Baublock ist, desto geringer ist auch die subjektive Beeinträchtigung durch mangelnde Grünflächen.

Ähnlich verhält es sich mit dem Wohnumfeld, welches in den Surveydaten durch den Interviewer erhoben wird. Es wird zwar der exakte Haustyp für das Gebäude, in dem der Surveyteilnehmer wohnt, punktuell erhoben, die Nachbarschaft des Wohnquartiers ist allerdings räumlich nicht eindeutig definiert. Aus den abgeleiteten Bebauungstypen der Fernerkundungsdaten können flächendeckende Informationen über die Bebauungsstruktur erhoben werden und mit den Surveydaten gegenübergestellt werden. Es kann beobachtet werden, dass die Daten grundsätzlich Ähnlichkeiten miteinander aufweisen und die meisten Beobachtungen der Wohnumfeldmerkmale aus den Surveys auch dem jeweiligen Pendant aus der Bebauungsklassifikation entsprechen.

Zusammenfassend kann festgestellt werden, dass die räumlichen Bezüge in den dargestellten Surveydaten nicht eindeutig sind und viel Subjektivität zulassen. Die Subjektivität der Informationen stellt zwar grundsätzlich keinen Nachteil in der Auswertung dar, es bestehen allerdings keine Vergleichsmöglichkeiten über Referenzdaten. Darüber hinaus sind diese räumlichen Merkmale auch lückenhaft, da sie nur in den Gebieten erhoben werden, in denen auch Surveyteilnehmer wohnen. Eine Analyse der Daten auf einer anderen räumlichen Ebene, als in jener, in der sie erhoben wurden, ist somit nicht möglich. Demgegenüber stellen die abgeleiteten räumlichen Kontextmerkmale flächendeckende Informationen für ganze Stadtgebiete dar und sie können auf Basis von objektiven Kriterien erhoben werden. Sie können die subjektiven Aussagen aus den Surveydaten nicht ersetzen, aber wichtige zusätzliche Informationen liefern und vor allem einen eindeutigen räumlichen Bezug aufweisen.

### 6.3. Quantitative empirische Analysepotenziale

Das Hauptaugenmerk der vorliegenden Arbeit liegt in der Untersuchung der Vereinbarkeit von flächendeckenden räumlichen Informationen aus Erdbeobachtungsdaten und Surveydaten. Im aktuellen Kapitel werden aber auch Analysepotenziale aus den verknüpften Daten vorgestellt. In der quantitativen Sozialforschung werden neben deskriptiven Verfahren auch numerische Verfahren genutzt, um sozialwissenschaftliche Sachverhalte zu überprüfen. Wie bereits in der Einleitung des aktuellen Kapitels erwähnt, sind die Analysemöglichkeiten mit dem erstellten Datensatz sehr vielfältig und der Datensatz muss in der Regel für jede empirische Analyse individuell angepasst werden. In den beiden vorhergehenden Abschnitten dieses Kapitels wurden die räumlichen Kontextmerkmale mit den Surveydaten verknüpft und beschrieben und für ausgewählte Surveymerkmale mit räumlichem Bezug eine Gegenüberstellung mit den räumlichen Kontextmerkmalen durchgeführt. Der Vergleich zeigt, dass, obwohl in den Surveymerkmalen häufig nur ein indirekter oder subjektiver räumlicher Bezug besteht, eine messbare Ähnlichkeit mit vergleichbaren Kontextmerkmalen festgestellt werden kann. Inwiefern die räumlichen Kontextmerkmale aber tatsächlich eine objektive Beschreibung der Nachbarschaft der Surveyteilnehmer darstellen und inwiefern sie auch einen messbaren Zusammenhang mit sozialem Verhalten oder bestimmten Eigenschaften darstellen, kann erst durch die quantitative Analyse untersucht werden.

Das Ziel einer quantitativen sozialwissenschaftlichen Untersuchung ist die Überprüfung einer Hypothese und die quantitative Beschreibung von Zusammenhängen mit statistischen Methoden. Dafür werden Kausalbeziehungen zwischen bestimmten Merkmalen untersucht. Für die Untersuchung einer Hypothese ist dabei die theoretische Relevanz der eingehenden Größen genauso von erheblicher Bedeutung wie die Zielvariable als Untersuchungsmerkmal theorie- bzw. hypothesenbezogen ausgewählt und auch aufbereitet werden muss. Die sozialwissenschaftliche Theorie für den im aktuellen Abschnitt dargestellten, exemplarischen Anwendungsfall ist rudimentär formuliert. Es ist allerdings möglich, ohne kausale Wirkungszusammenhänge die Korrelationen zwischen den Variablen in der Regressionsanalyse festzustellen.

Die räumliche Definition und Abgrenzung der *Nachbarschaft* stellt in sozialwissenschaftlichen Studien immer wieder eine große Herausforderung dar. Vor allem Ökonomen und Soziologen bedienen sich häufig eines konzeptionellen Ansatzes und vermeiden eine exakte geographische Beschreibung der Nachbarschaft. In quantitativen Analysen ist die geographische Ausprägung der Nachbarschaft meistens datengetrieben und je nach Datenlage werden verschiedene statistische Raumeinheiten als Nachbarschaft definiert. Zwei Kernprobleme sind bei der räumlichen Auswertung dieser Daten immanent: zunächst ist die Definition des betrachteten Raumausschnittes von Bedeutung. Sie wird gemeinhin auch als *modifiable areal unit problem* (MAUP) bezeichnet und bei Openshaw und Taylor (1981) diskutiert.

Zweitens sind Nachbarschaftseffekte in der Sozialwissenschaft relevant. Eine ausführliche Übersicht darüber findet sich bei Dietz (2002). Beide Probleme sind weithin bekannt, werden allerdings in der folgenden, exemplarischen empirischen Modellschätzung nicht berücksichtigt, da sie nicht im Fokus der vorliegenden Arbeit liegen.

### 6.3.1. Die lineare Regression

Die lineare Regression wird in empirischen sozialwissenschaftlichen Analysen verwendet, um den Erklärungsgehalt des erstellten Modells und die Korrelationen zwischen den Merkmalen zu untersuchen. Sie stellt ein statistisches Standardverfahren sowohl in der Fernerkundung als auch in den Sozialwissenschaften dar. Für ein gemeinsames Verständnis werden aber die Elemente der linearen Regression an dieser Stelle wiederholt beschrieben.

Anhand einer linearen Regression kann der Einfluss von unabhängigen Variablen auf eine abhängige Variable untersucht werden. Die unabhängigen Variablen stellen demnach erklärende Größen dar, die einen bestimmten Einfluss auf die zu erklärende Größe ausüben. Wenn der erklärende Einfluss einer Variablen nicht zufälliger Natur ist, wird er als *statistisch signifikant* bezeichnet. Für die Überprüfung der *Signifikanz* existieren ebenfalls statistische Standardverfahren, welche hier angewendet werden.

Mit dem Modell der einfachen linearen Regression können reale sozialwissenschaftliche Fragestellungen in eine formale Form gebracht werden. Zwischen den Elementen  $X$  und  $Y$  wird ein Zusammenhang durch die allgemeine Gleichung:

$$(48) \quad Y = \alpha + \beta X + \varepsilon$$

zum Ausdruck gebracht, wobei  $X$  die *exogene* Variable (Regressor) und  $Y$  die *endogene* Variable (Regressand) darstellt. Jeder Wert von  $Y$  kann über die beiden Koeffizienten  $\alpha$  und  $\beta$  aus  $X$  berechnet werden, wenn der Zusammenhang zwischen  $X$  und  $Y$  linear ist. Die Abweichung, um die  $Y$  von seiner deterministischen Komponente abweicht, wird als Störgröße  $\varepsilon$  oder als *Residuum* bezeichnet. Es nimmt den stochastischen Teil des Zusammenhangs auf und stellt die fallweise Abweichung von einer linearen Regressionsgeraden dar. Um die „beste“ Regressionsgerade zu bestimmen, muss die Summe der Schätzfehler 0 sein, was bedeutet, dass sich die einzelnen Fehler aufheben. Um große Abweichungen möglichst zu vermeiden werden nicht die einfachen Residuen, sondern deren Quadrate betrachtet. Deswegen wird diese lineare Regression auch OLS-Methode (engl.: ordinary least squares; Methode der kleinsten Quadrate) genannt (Backhaus et al., 2003).

Damit der Einfluss von bestimmten Merkmalen auf eine Eigenschaft  $E$  untersucht werden kann, werden in der Regel mehrere Regressionsmodelle geschätzt, welche sich voneinander durch Hinzunahme oder Weglassen von einigen Variablen unterscheiden. Dieses Vorgehen

wird auch im Beispiel im folgenden Unterabschnitt gewählt, wo mehrere Modelle geschätzt werden: ein Modell, in welchem ausschließlich die Surveymerkmale verwendet werden und weitere Modelle, in denen schrittweise bestimmte räumliche Kontextmerkmale hinzugenommen werden. Auf diese Weise kann der Zusammenhang zwischen den räumlichen Kontextmerkmalen und der Zielvariablen untersucht werden. Diesem Ansatz folgend, nehmen die ersten Modelle folgende allgemeine Formen an:

$$(49) \quad E_{ik} = \alpha + \beta_0 x_{ik} + \varepsilon$$

wobei  $x$  einen Vektor aus den sozioökonomischen und demographischen Charakteristika aus den Surveys für die Person  $i$  der Lage  $k$  darstellt. In den weiteren Modellen:

$$(50) \quad E_{ik} = \alpha + \beta_0 x_{ik} + \beta_1 z_{ik} + \varepsilon$$

ist  $x$  wie im ersten Modell definiert und  $z$  ein Vektor aus den verknüpften, räumlichen Kontextmerkmalen. Durch Vergleich der verschiedenen geschätzten Modelle ist eine empirische Analyse der statistischen Zusammenhänge zwischen den Merkmalen möglich und die Modelle können nach ihrem Erklärungsgehalt interpretiert werden. Die Maße für die Interpretation des Erklärungsgehaltes sind das adjustierte  $R^2$  und der Regressionskoeffizient  $\beta$ .

Das Maß *adjustiertes  $R^2$*  gibt an, wie viel Prozent der beobachteten Varianz der abhängigen Variablen mit Hilfe der unabhängigen Variablen erklärt werden kann. Bei einem Wert von Null erfolgt keinerlei Varianzerklärung (von null Variablen ist nur die Konstante in Gleichung (48) signifikant). Bei einem  $R^2$  von 1 wird die gesamte Varianz erklärt. Alle Störterme haben dann den Wert Null.  $R^2$  kann also einen Wert zwischen 0 und 1 annehmen, je näher sich  $R^2$  an 0 befindet, desto geringer ist der beobachtete Zusammenhang, je näher sich  $R^2$  an 1 befindet, desto höher ist der Zusammenhang. Es gibt in der Literatur keine festgelegten Werte, ab welcher Höhe  $R^2$  als gut zu bezeichnen ist (Witt, 2008). Die Bewertung ist in der Regel von der untersuchten Fragestellung abhängig und ein niedriges  $R^2$  muss nicht zwangsläufig als mangelnde Modellqualität interpretiert werden (Backhaus et al., 2003). Orientierungshilfen zur Beurteilung des Bestimmtheitsmaßes finden sich aber dennoch. So beschreibt Bortz (1999) die Zusammenhänge in drei Kategorien: bei einem  $R^2$  von 0,01 liegt ein schwacher Effekt vor, bei einem  $R^2$  von 0,09 ein mittlerer und bei einem  $R^2$  von 0,25 ein starker Effekt. Chin (1998) beurteilt die Qualität der Bestimmtheitsmaße folgendermaßen: ab 0,19 stellt er einen schwachen Zusammenhang fest, ab 0,33 einen durchschnittlichen und ab 0,67 einen substanziellen Erklärungsgehalt.

Neben der Untersuchung der gesamten erklärten Varianz werden auch die einzelnen Regressionskoeffizienten  $\beta$  ausgewertet, da an ihnen abgelesen werden kann, um welchen Wert sich die abhängige Variable ändert, wenn sich die unabhängige Variable um eine Einheit ändert. Bei einem positiven Vorzeichen des Regressionskoeffizienten erhöht sich die Zielvariable, bei einem negativen Vorzeichen ist der Einfluss negativ. Weiterführende

Informationen über multivariate Regressionsanalysen finden sich in der statistischen Standardliteratur und in Lehrbüchern über empirische Sozialwissenschaften (z.B. Bahrenberg, Giese und Nipper, 1992; Bortz, 1999; Backhaus et al., 2003).

### 6.3.2. Exemplarische Analyse zum Haushaltseinkommen

Damit potenzielle statistische Zusammenhänge zwischen den abgeleiteten räumlichen Kontextmerkmalen und den Surveydaten untersucht werden können, wird im aktuellen Abschnitt eine empirische Analyse durchgeführt. Anhand der Analyse kann überprüft werden, ob es messbare statistische Einflüsse zwischen dem Raum und bestimmten Eigenschaften der Surveyteilnehmer gibt. Anhand der exemplarischen Untersuchung wird beispielhaft das Analysepotenzial durch die im Rahmen der Arbeit abgeleiteten Kontextmerkmale für die sozialwissenschaftliche Forschung untersucht.

#### 6.3.2.1. Hintergrund

Im konkreten Beispiel wird anhand eines Modells analysiert, ob es einen messbaren Zusammenhang zwischen der räumlichen Segregation und dem Haushaltseinkommen gibt. Aus sozialgeographischer Sicht kann damit untersucht werden, ob die räumliche Struktur einer Region mit der sozioökonomischen Struktur dieser Region korreliert. Da im Rahmen der Arbeit die physische Struktur der beiden Städte München und Berlin auf Basis von flächendeckenden, objektiven Merkmalen aus Fernerkundungsdaten beschrieben wird, dienen diese Merkmale zur physischen Beschreibung der lokalen Nachbarschaften in den beiden Städten. Sie beinhalten quantifizierbare Informationen zur physischen städtischen Struktur als Dichtemerkmale zur *Vegetation*, zur *Versiegelung*, zur *Bebauungsdichte*, zur *Geschossflächendichte*, zur *Gebäudeklasse* und zur *Lage* innerhalb des Stadtgebietes. Alle diese Merkmale sagen mehr oder weniger über die Ausstattung einer Wohngegend aus und es kann die Hypothese untersucht werden, ob ein statistischer Zusammenhang zwischen der physischen Ausstattung eines Wohnquartiers und dem Einkommen in diesem Wohnquartier besteht. Insbesondere kann untersucht werden ob eine räumliche Trennung der Wohnquartiere besteht und ob sich diese Segregation über die physische Ausstattung der Stadt und das finanzielle Einkommen numerisch feststellen lassen können.

Die Segregation beschreibt die räumliche Trennung der Wohnbevölkerung in einer Stadt aufgrund von sozialem Status, demographischen Merkmalen, oder ethnischen, religiösen oder sprachlich-kulturellen Kriterien (Alich, Winter und Arentzen, 2005). Sie kann freiwilliger oder unfreiwilliger Natur sein, wobei die gewollte Segregation darauf beruht, dass bewusst ein gleiches soziales Umfeld gesucht wird und gleichzeitig bewusst eine soziale und räumliche

Distanz zu anderen Gruppen gesucht wird. Die unfreiwillige Segregation „[...] resultiert aus dem Wohnungsmarkt, der entsprechend den Bodenpreisen und der Mietzahlungsfähigkeit über die Bauform, Wohnungsdichte und Wohnumfeldausstattung räumlich unterschiedliche Wohngebiete schafft und aus staatlich-planerischen Interventionen, die von der Zonierung durch die Bauleitplanung über die staatliche Infrastrukturstandortpolitik bis zu strukturellen Eingriffen in den Wohnungsmarkt (Mietpreisstopp, Wohngeld u.a.) reichen können“ (Alisch, Winter und Arentzen, 2005).

In diesem Beispiel zur Demonstration des Analysepotenzials der Kontextmerkmale und der Surveydaten wird also versucht, einen statistischen Zusammenhang zwischen einer potenziellen räumlichen Segregation von Haushalten mit ähnlichem Haushaltseinkommen auf Basis der physischen Wohnumfeldausstattung, welche durch die physischen Kontextmerkmale und der Lage dargestellt sind, herzustellen.

### 6.3.2.2. Datenaufbereitung

Für die multivariate Analyse zur räumlichen Segregation in München und Berlin ist das Haushaltseinkommen die Zielvariable, welche auf Abhängigkeiten von physischen Kontextmerkmalen untersucht wird. Die Höhe des Einkommens ist von einer Reihe bekannter, aber auch von vielen unbekanntem Faktoren abhängig. Die wichtigsten der grundsätzlich messbaren Faktoren sind das Alter, der Bildungsstand und auch das Geschlecht der betroffenen Person. Zusätzliche Merkmale können einen weiteren Erklärungsgehalt zum aufgestellten Modell beitragen. In SOEP/BASE-II werden im Haushaltsfragebogen vom Haushaltsvorstand verschiedene Informationen über die Zusammenstellung des Haushaltes, die Nachbarschaft und die ökonomische Situation des Haushaltes beantwortet, darunter auch über das Haushaltsnettoeinkommen (vgl. Abbildung 61).

**52. Wenn man mal alle Einkünfte zusammennimmt:  
Wie hoch ist das monatliche Haushaltseinkommen aller Haushaltsmitglieder heute?**

 Bitte geben Sie den monatlichen Netto-Betrag an, also nach Abzug von Steuern und Sozialabgaben. Regelmäßige Zahlungen wie Renten, Wohngeld, Kindergeld, BAföG, Unterhaltszahlungen usw. rechnen Sie bitte dazu!

 Falls nicht genau bekannt:  
Bitte schätzen Sie den monatlichen Betrag. ....  Euro im Monat

**Abbildung 61:** Frage zum monatlichen Haushaltseinkommen.

Quelle: SOEP Haushaltsfragebogen 2010.

Das Haushaltseinkommen ist ein Indikator für die Möglichkeiten der Lebensgestaltung und der Erwerbsbeteiligung der im Haushalt lebenden Personen (BPB, 2004). Die größenmäßige und strukturelle Zusammensetzung der Haushalte kann sehr unterschiedlich sein, weil die Anzahl der Erwerbstätigen und die Anzahl der Kinder im Haushalt einen Einfluss auf die Möglichkeiten der Verwendung des Einkommens haben. Ein Haushalt mit zwei Erwerbstätigen

ohne Kinder hat relativ ein höheres Einkommen als ein Haushalt mit nur einem Erwerbstätigen und Kleinkindern. Deshalb wird das Haushaltseinkommen nach der Anzahl der Personen bedarfsgewichtet. Die Gewichtung erfolgt dabei nach der modifizierten OECD-Skala (engl.: Organisation for Economic Co-operation and Development; deutsch: Organisation für wirtschaftliche Entwicklung und Zusammenarbeit). In der neuen OECD-Skala wird der Hauptbezieher des Haushaltseinkommens mit dem Faktor 1,0 gewichtet, die anderen Haushaltsmitglieder mit einem Alter von 14 und mehr Jahren werden mit einem Faktor 0,5 und alle weiteren Personen mit einem Faktor von 0,3 gewichtet (DeStatis, 2012).

Der erstellte Datensatz aus den verknüpften räumlichen Kontextmerkmalen und den Surveydaten für Berlin und München enthält insgesamt Informationen von 3006 Personen in 1791 Haushalten. Allerdings verringert sich die Untersuchungspopulation noch weiter, da nicht für sämtliche Personen und Haushalte alle untersuchten Merkmale zur Verfügung stehen, weil sie entweder nicht von der expliziten Fragestellung betroffen sind oder keine Angaben dazu gemacht haben.

Für die empirische Modellschätzung wurde jeder Survey-Haushalt nur einmal aufgenommen, auch wenn es über die Jahre mehrere Beobachtungen vom selben Haushalt gibt. Die Auswahl wird deshalb gemacht, damit die wiederholt teilnehmenden Haushalte nicht höher gewichtet werden, als Haushalte, welche nur einmal teilnehmen. Deshalb bleiben für die Untersuchung 1604 unterschiedliche Personen über die Jahre in 968 Haushalten für die beiden Städte München und Berlin übrig (vgl. Tabelle 24).

**Tabelle 24:** Anzahl der beobachteten Personen und Haushalte für jedes Beobachtungsjahr nach Auswahl auf ein einzelnes Beobachtungsjahr für jeden Haushalt für Berlin und München.

Jahr	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	Summe
Personen	53	61	54	45	43	50	46	59	48	707	438	1604
Haushalte	26	40	27	23	25	23	29	26	23	470	256	968

Die Einkommen werden über einen Zeitraum von 10 Jahren betrachtet. Sie müssen vor der Modellschätzung harmonisiert werden, da sich sowohl die Kaufkraft als auch das Einkommen selbst durch die Inflation über die Jahre ändern. Das Statistische Bundesamt Deutschland veröffentlicht laufend aktuelle Tabellen, mit welchen die Deflationierung der Einkommen durchgeführt werden kann (Statistisches Bundesamt, 2012). Tabelle 25 zeigt die Verbraucherpreisindizes (VPI), anhand welcher die Einkommen angeglichen, das heißt auf das Jahr 2005 normiert, werden.

**Tabelle 25:** Anpassung des Haushaltsäquivalenzeinkommens mittels VPI.

Quelle: Statistisches Bundesamt (2012).

Jahr	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
VPI (%)	92,7	94,5	95,9	96,9	98,5	100,0	101,6	103,9	106,6	107,0	108,2

Die ökonomische Situation eines Haushaltes hat einen direkten Einfluss auf die Wahl des Wohngebietes innerhalb einer Stadt. Das Merkmal *Lage* gilt in der Immobilienwelt als wichtigster Indikator für die Gestaltung der Mietpreise. Die Lage beinhaltet in diesem Fall sowohl die geographische Position der Wohnung als auch die Ausstattung des Wohnumfeldes. Durch das Einkommen eines Haushaltes können sich für ihn Möglichkeiten oder Limitierungen in der Wahl der Wohnlage ergeben. Neben den eingangs erwähnten bekannten Faktoren zur empirischen Erklärung des Haushaltseinkommens, sind auch weitere Faktoren von Bedeutung.

Wie bereits ausführlich in Abschnitt 6.2 ausgeführt, werden die räumlichen Informationen in den Surveys nur punktuell für die teilnehmenden Haushalte erhoben. Die Informationen über die Lage der Wohnung innerhalb des Stadtgebietes sind nur indirekt über die gruppierte Distanz zum nächsten Großstadtzentrum oder über eines der Wohnumfeldmerkmale im Datensatz enthalten. Ebenso wird der Haustyp ausschließlich für den untersuchten Haushalt erhoben und keine direkten Informationen über die umliegenden Gebäudetypen erhoben. Erst über die erhobenen räumlichen Kontextmerkmale aus den Fernerkundungsdaten können flächendeckende Informationen über die Wohnumfeldausstattung in die Analyse einbezogen werden und somit der Zusammenhang zwischen der räumlichen Nachbarschaft und dem Haushaltseinkommen untersucht werden. Die Merkmale, welche für die statistische Analyse verwendet werden, sind in Tabelle 26 beschrieben. Im oberen Teil der Tabelle sind die Variablen dargestellt, welche aus den Surveydaten gewonnen werden und im unteren Teil sind die räumlichen Kontextmerkmale aus den Fernerkundungsdaten und die Lage beschrieben.

Als Datengrundlage für die Untersuchung wurden die Informationen der beiden Städte München und Berlin zusammengefasst. Um einen stadtspezifischen Effekt zu überprüfen, wurde eine 0,1-Dummyvariable *Stadt: München* erstellt. Anhand einer *Dummyvariable* können die Auswirkungen qualitativer Unterschiede überprüft werden (Stocker, 2011: 270ff.). Die Schätzung des Nettoäquivalenzeinkommens wurde auch nur für Personen durchgeführt, welche potenziell am Berufsleben teilhaben können. Deshalb sind nur Personen in die Modellschätzung mit aufgenommen, welche zum Befragungszeitpunkt ein Alter von mindestens 25 und höchstens 65 Jahren haben.

**Tabelle 26:** Beschreibung der Surveyvariablen und Nachbarschaftsmerkmale für die empirische Modellschätzung des Nettoäquivalenzeinkommens.

Variablenname	Beschreibung
<i>Bildungszeit</i>	Ausbildungszeit in Jahren
<i>WUM 1</i>	Wohnumfeld 1: Gebäudetyp (Ref.: 1-2fam. Haus). Weitere Ausprägungen: 1-2 Familien Reihenhaushaus, Wohnhaus mit 3-4 Wohnungen, Wohnhaus mit 5-8 Wohnungen, Wohnhaus mit mehr als 9 Wohnungen, Hochhaus
<i>WUM 3</i>	Wohnumfeld 3: Wohngebietstyp (Ref.: Reines Wohngebiet Altbau). Weitere Ausprägungen: reines Wohngebiet Neubau, Mischgebiet, Geschäftszentrum
<i>Alter</i>	Alter in Jahren
<i>Alter<sup>2</sup></i>	Alter in Jahren * Alter in Jahren
<i>Geschlecht</i>	männlich / weiblich (Referenz: männlich)
<i>Fam.st.</i>	Familienstand (Referenz: verheiratet, zusammenlebend), Weitere Ausprägungen: verheiratet getrennt lebend, ledig, geschieden, verwitwet
<i>Anzahl der Kinder &gt; 14 J.</i>	Anzahl der Kinder über 14 Jahren
<i>Ausländer</i>	Referenz: kein Ausländer
<i>Vegetationsanteil</i>	Anteil der Vegetationsfläche an der Blockfläche
<i>Versiegelungsanteil (4er)</i>	Anteil der Versiegelung an der Blockfläche der vier umliegenden Baublöcke
<i>GFD</i>	Geschossflächendichte
<i>Lage</i>	Referenz: 0-1000m
<i>Beb. (Block)</i>	Bebauung (Block): Dominanter Bebauungstyp des Baublocks, Referenz: 1/2 Familienhaus
<i>Beb. (4er)</i>	Bebauung (4er): Dominanter Bebauungstyp der vier umliegenden Baublöcke, Referenz: 1/2 Familienhaus
<i>Stadt: München</i>	Dummyvariable für die Stadt München. 1=München

Eine Auflistung der verwendeten Variablen und deskriptive Statistiken dazu sind in Anhang C dargestellt, wo neben der Anzahl der Beobachtungen für jedes Merkmal auch die Verteilung der Werte dargestellt ist.

### 6.3.2.3. Modellschätzungen

Für die Untersuchung der Korrelationen zwischen der Wohnumfeldausstattung und dem Haushaltseinkommen werden mehrere Regressionsmodelle geschätzt, um die Korrelationen der einzelnen Merkmale in Abhängigkeit des Haushaltseinkommens zu untersuchen. Dafür werden in mehreren Modellen verschiedene Variablen hinzugefügt oder weggelassen. Damit kann bestimmt werden, wie hoch der Erklärungsbedarf der einzelnen Modelle unter Berücksichtigung der ausgewählten Variablen ist und es können Vergleiche zwischen den erstellten Modellen und ihren Variablen gezogen werden.

In einem ersten Modell werden ausschließlich Variablen aus den Surveys verwendet, um die Korrelationen zum Haushaltseinkommen zu untersuchen und den gesamten Erklärungsgehalt des Modells alleine auf Basis von wenigen, grundlegenden sozioökonomischen Merkmalen zu untersuchen. In einem zweiten Modell werden zusätzlich die Wohnumfeldmerkmale aus den Surveys hinzugefügt, um den Einfluss des Wohnumfeldes, wie es in den Surveydaten hinterlegt ist, zu untersuchen. In weiteren vier Modellen werden zusätzlich zu den Surveymerkmalen auch die räumlichen Kontextmerkmale zur physischen Beschreibung der Wohnumfeldausstattung hinzugenommen und unterschiedliche Kombinationen an Merkmalen untersucht.

Diese Modellschätzungen mit den Variablen aus den Surveys und den räumlichen Kontextmerkmalen sind in der folgenden Tabelle 27 dargestellt. Sie zeigt die Koeffizienten der fünf aufgestellten Modellschätzungen mit den unterschiedlichen Kombinationen der Merkmale in (a-f). Die Koeffizienten beschreiben den quantitativen statistischen Zusammenhang zwischen der unabhängigen Variable und der Zielvariable des Haushaltseinkommens. Das Vorzeichen des Koeffizienten beschreibt, ob es einen positiven oder einen negativen Zusammenhang gibt. In der Ergebniszeile jedes Modells ist das adjustierte  $R^2$  dargestellt, welches den gesamten Erklärungsgehalt des aufgestellten Modells beschreibt (vgl. Abschnitt 6.3.1).

Tabelle 28 und Tabelle 29 stellen weitere Modellschätzungen dar, welche ausschließlich auf Basis der räumlichen Kontextmerkmale durchgeführt wurden, um die Korrelationen zwischen der physischen Ausstattung des Wohnumfeldes mit dem Haushaltseinkommen zu untersuchen.

Die Ergebnisse aus den Modellen werden im folgenden Unterabschnitt 6.3.2.4 diskutiert. Des Weiteren finden sich die detaillierten Tabellen zu den einzelnen Modellen in Anhang D.

**Tabelle 27:** Regressionsschätzungen zum Nettoäquivalenzeinkommen für sechs verschiedene Modelle mit und ohne kleinräumige Nachbarschaftsmerkmale.

	(a)		(b)		(c)		(d)		(e)		(f)	
	Koeff.		Koeff.		Koeff.		Koeff.		Koeff.		Koeff.	
<i>Bildungszeit</i>	73,9	***	62,4	***	58,4	***	57,5	***	57,4	***	69,3	***
<i>WUM 1: 1-2 Fam.Rh.</i>	-		-195,5		-231,1		-275,3		-146,8		-	
<i>WUM 1: Wh. 3-4 Whg</i>	-		735,8	.	423,0		358,9		402,1		-	
<i>WUM 1: Wh. 5-8 Whg</i>	-		-611,3	*	-896,9	**	-1095,8	***	-1073,9	**	-	
<i>WUM 1: Wh. &gt; 9 Whg</i>	-		-539,0	.	-896,7	**	-1103,1	***	-1076,4	**	-	
<i>WUM 1: Hh.</i>	-		-1055,2	*	-1461,5	*	-1462,2	**	-1515,3	*	-	
<i>WUM 3: Neub</i>	-		-62,9		6,7		4,2		7,4		-	
<i>WUM 3: Misch.</i>	-		-7,7		8,1		28,6		20,9		-	
<i>WUM 3: Geschäft</i>	-		-846,6		-732,1		-752,8		-582,4		-	
<i>Alter</i>	109,9	*	96,0	*	104,3	*	102,8	*	109,7	*	124,5	**
<i>Alter<sup>2</sup></i>	-1,3	*	-1,0	*	-1,1	*	-1,1	*	-1,2	**	-1,34	**
<i>Geschlecht weiblich</i>	0,7		-23,9		-38,4		-44,8		-40,5		5,8	
<i>Fam.st. verheiratet, getr.</i>	-509,9	.	-280,6		-186,5		-128,6		-127,9		-360,5	
<i>Fam.st. ledig</i>	-557,6	**	-495,8	**	-486,7	**	-457,3	**	-489,2	**	-589,7	***
<i>Fam.st. geschieden</i>	-439,9	*	-477,8	**	-493,9	**	-507,7	**	-524,0	**	-533,7	**
<i>Fam.st. verwitwet</i>	263,6		439,7		475,4		393,7		383,9		403,1	
<i>Anzahl Kinder &gt; 14 Jahre</i>	-704,6	***	-668,3	***	-687,8	***	-653,2	***	-654,7	***	-693,6	***
<i>Ausländer</i>	-352,3	.	-305,1	.	-274,0		-276,6	.	-241,6		-285,6	
<i>Vegetationsanteil (%)</i>	-		-		3,3		5,4		8,1	.	5,6	
<i>Versiegelung (4er) (%)</i>	-		-		-2,02		0,30		1,9		4,2	
<i>GFD</i>	-		-		75,3		83,2		111,6		78,8	
<i>Lage &gt; 1 &lt; 2,5km</i>	-		-		-402,5		-191,1		-115,9		501,3	
<i>Lage &gt; 2,5 &lt; 5km</i>	-		-		-879,9		-690,9		-619,0		-212,7	
<i>Lage &gt; 5 &lt; 7,5km</i>	-		-		-522,4		-288,0		-250,6		-123,8	
<i>Lage &gt; 7,5km</i>	-		-		-651,0		-483,6		-410,3		-75,9	
<i>Beb. (Block) Zeile</i>	-		-		-		634,8	**	805,1	**	480,2	.
<i>Beb. (Block) Block</i>	-		-		-		686,4	**	716,5	*	116,1	
<i>Beb. (Block) Hochhaus</i>	-		-		-		246,9		26,8		-923,5	.
<i>Beb. (Block) Sonstige</i>	-		-		-		705,8	*	655,2	.	47,1	
<i>Beb. (4er Block) Zeile</i>	-		-		120,9		-		-331,1		-511,9	.
<i>Beb. (4er Block) Block</i>	-		-		421,1		-		-45,7		-327,2	
<i>Beb. (4er Block) Hh.</i>	-		-		390,6		-		225,4		-332,0	
<i>Beb. (4er Block) Sonstige</i>	-		-		841,4	*	-		617,7		1286,0	***
<i>Stadt: München</i>	358,3	**	294,6	*	406,4	**	428,5	**	431,2	**	390,1	**
<i>adj. R<sup>2</sup></i>	0,1421		0,1966		0,2231		0,2302		0,2381		0,2148	

N=1197

Signif. codes: 0 '\*\*\*' 0,001 '\*\*' 0,01 '\*' 0,05 '.' 0,1 ' ' 1

**Tabelle 28:** Regressionsschätzungen zum Nettoäquivalenzeinkommen für drei verschiedene Modelle nur auf Basis der kleinräumigen Nachbarschaftsmerkmale (ohne Dummyvariable für die Stadt).

	(a) Koeff.	(b) Koeff.	(c) Koeff.
Vegetationsanteil (%)	-0,314	1,720	4,805
Versiegelungsanteil (4er) (%)	-2,191	-1,989	2,218
GFD	-75,740	-38,300	9,895
Lage > 1 < 2,5km	1061,540	920,060	1117,890
Lage > 2,5 < 5km	295,120	198,510	393,790
Lage > 5 < 7,5km	480,290	254,280	409,704
Lage > 7,5km	289,970	81,560	234,514
Bebauung (Block) Zeile	-112,840	-	329,167
Bebauung (Block) Block	-271,730	-	142,979
Bebauung (Block) Hochhaus	-863,080	-	-976,821
Bebauung (Block) Sonstige	-367,340	-	-168,135
Bebauung (4er Block) Zeile	-	-496,860 *	-697,016 *
Bebauung (4er Block) Block	-	-559,170 *	-753,386 *
Bebauung (4er Block) Hochhaus	-	-735,210	-390,532
Bebauung (4er Block) Sonstige	-	674,560	748,796
adj. R <sup>2</sup>	0,03017	0,0601	0,0683
N=1197	Signif. codes: 0 '***' 0,001 '**' 0,01 '*' 0,05 '.' 0,1 ' ' 1		

**Tabelle 29:** Regressionsschätzungen zum Nettoäquivalenzeinkommen für drei verschiedene Modelle nur auf Basis der kleinräumigen Nachbarschaftsmerkmale.

	(a) Koeff.	(b) Koeff.	(c) Koeff.
Vegetationsanteil (%)	-0,197	1,599	5,051
Versiegelungsanteil (4er) (%)	-2,298	-2,697	1,845
GFD	-5,670	15,310	73,370
Lage > 1 < 2,5km	824,950	753,530	944,120
Lage > 2,5 < 5km	75,991	46,760	235,810
Lage > 5 < 7,5km	285,907	146,270	295,580
Lage > 7,5km	142,517	24,480	175,820
Bebauung (Block) Zeile	-4,204	-	390,470
Bebauung (Block) Block	-170,162	-	173,070
Bebauung (Block) Hochhaus	-797,300	-	-1016,870
Bebauung (Block) Sonstige	-291,654	-	-161,900
Bebauung (4er Block) Zeile	-	-422,120	-652,330 *
Bebauung (4er Block) Block	-	-414,620	-624,790
Bebauung (4er Block) Hochhaus	-	-549,130	-184,450
Bebauung (4er Block) Sonstige	-	756,410 *	832,240 *
Stadt: München	462,021 **	404,080 **	445,500 **
adj. R <sup>2</sup>	0,04862	0,07353	0,08502
N=1197	Signif. codes: 0 '***' 0,001 '**' 0,01 '*' 0,05 '.' 0,1 ' ' 1		

### 6.3.2.4. Ergebnisse

In diesem Unterabschnitt werden die einzelnen Modelle betrachtet und die Korrelationen zum Haushaltseinkommen analysiert. Dabei werden sowohl die Regressionskoeffizienten als auch der erklärte Anteil des Modells untersucht. Dieses Maß ist korrigiert auf die Anzahl der unabhängigen Variablen, da die erklärte Varianz grundsätzlich steigt, wenn mehr unabhängige Variablen berücksichtigt werden. Rein statistisch betrachtet, bedeutet das, dass die Gleichung diesen Anteil der Varianz erklärt und die Höhe des Nettoäquivalenzeinkommens zu etwa einem Fünftel von diesen Merkmalen abhängt. Kausal betrachtet hat natürlich nicht jedes dieser Merkmale einen realen Einfluss auf das Haushaltseinkommen, allerdings können auf diese Weise Zusammenhänge zwischen einer räumlichen sozialen Segregation auf Basis unterschiedlicher ökonomischer Haushaltssituationen quantifiziert werden. Bei der Betrachtung der statistischen Zusammenhänge sei an dieser Stelle darauf hingewiesen, dass der erklärte Modellanteil in sozialwissenschaftlichen Analysen nicht wie in physikalischen Modellen immer einen sehr hohen Anteil aufweisen muss. Die statistischen und besonders die kausalen Zusammenhänge zwischen bestimmten sozialwissenschaftlichen Merkmalen weisen oft eine deutlich höhere Komplexität und weitläufigere Wirkungszusammenhänge auf, als es physikalische Zusammenhänge wie zum Beispiel die Variablen Lufttemperatur und Seehöhe zulassen. Menschen haben individuelle Geschichten und entsprechend groß ist die Variabilität der erhobenen Daten. Die Interpretation des Bestimmtheitsmaßes und der Signifikanz ist in Abschnitt 6.3.1 beschrieben.

Im ersten Modell (vgl. Tabelle 27 a) wurden ausschließlich Informationen aus den Surveys aufgenommen. Die Merkmale wie *Bildungszeit* und *Alter* zeigen beide einen positiven Einfluss auf das Haushaltseinkommen: pro Lebensjahr erhöht sich das Haushaltseinkommen um 110 € und pro Bildungsjahr um rund 74 €. Die Bildungszeit ist darüber hinaus noch hoch signifikant mit einer Irrtumswahrscheinlichkeit  $\leq 0,01$  %. Eine gleich geringe Irrtumswahrscheinlichkeit trifft auf das Merkmal *Anzahl der Kinder über 14 Jahre* zu. Nicht so stark signifikant, aber immerhin mit einer Irrtumswahrscheinlichkeit von  $\leq 0,1$  % sind die Merkmale zum Familienstand: sowohl für ledige als auch geschiedene Personen im Haushalt führt das Zutreffen dieses Merkmals zu einer Reduzierung des bedarfsgewichteten Haushaltseinkommens von etwa 500 €. Das adjustierte Bestimmtheitsmaß für dieses Modell zeigt einen Erklärungsanteil von 0,14. Wenn die SOEP-Wohnumfeldmerkmale über den Gebäudetyp und den Wohngebietstyp zusätzlich in das Modell mit aufgenommen werden, steigt der erklärte Anteil des Modells auf 0,19 (vgl. Tabelle 27 b).

In weiteren Modellen (vgl. Tabelle 27 c-f) wurden zusätzlich zu den oben aufgeführten erklärenden Variablen noch die räumlichen Kontextmerkmale aus den Fernerkundungsdaten und die Distanz mit aufgenommen. In Betrachtung des adjustierten Bestimmtheitsmaßes kann dadurch statistisch gesehen ein zusätzlicher Anteil des Modells von knapp 4,5% erklärt werden. Der korrigierte Anteil der erklärten Varianz schwankt bei den Modellen zwischen etwa 21 und

knapp 24 %. Es kann also knapp ein Viertel des Modells statistisch erklärt werden. Dabei stellt sich der Vegetationsanteil als gering signifikant heraus, wie auch einige der zusätzlich hinzugenommenen Bebauungsklassen des Baublocks und der großräumigeren Nachbarschaft, bestehend aus dem eigentlichen Baublock des Umfrageteilnehmers und zusätzlich den vier umliegenden Baublöcken.

Die Regressionsergebnisse für die Modellschätzungen ohne die sozioökonomischen Merkmale aus den Surveys sondern nur mit den räumlichen Nachbarschaftsmerkmalen sind in Tabelle 28 und Tabelle 29 dargestellt. Sie sollen ausschließlich den statistischen Zusammenhang zwischen der physischen Ausstattung des Wohnumfeldes und der geographischen Lage des Befragungshaushaltes darstellen. Sie zeigen bei reiner Betrachtung des korrigierten Bestimmtheitsmaßes einen erklärten Anteil von knapp 8,5 % (vgl. Tabelle 29 c). Einen Beitrag zur Erklärung des Modells liefert offenbar auch die Stadt der befragten Haushalte. Wenn die Dummyvariable für die Stadt weggelassen wird, werden noch knapp über 6,8 % des Modells erklärt (vgl. Tabelle 28 c). Interessant ist in diesem Zusammenhang, dass die geographische Lage des Haushaltes innerhalb des Stadtgebietes als Distanz zum Großstadtzentrum in allen geschätzten Modellen keine statistische Signifikanz aufweist (vgl. Tabelle 28 und Tabelle 29).

### 6.3.2.5. Diskussion

Die zusätzlichen, räumlichen Kontextmerkmale aus Fernerkundungsdaten können trotz der kleinen Stichprobe für die beiden Städte einen Beitrag zur Erhöhung des Erklärungsanteiles leisten. Dieser zusätzlich erzielte Erklärungsanteil durch die räumlichen Variablen deckt sich mit einer Untersuchung, welche auf einem ähnlichen Datensatz für die beiden Untersuchungsgebiete München und Köln zur Erklärung der Bruttokaltmiete durchgeführt wurde (Wurm, Goebel und Wagner, 2010) und mit einer Studie über eine australische Längsschnittstudie (Ambrey und Fleming, 2011). Die Autoren haben den Einfluss der naturräumlichen Ausstattung auf die Lebenszufriedenheit in Australien untersucht. In der Studie finden die Autoren ebenfalls einen kleinen, zusätzlichen Erklärungsgehalt des Modells bei einem gesamten Erklärungsgehalt von nur rund 6 % und statistisch signifikante Einflüsse der aufgenommenen geographischen Variablen.

Weiterführende, tiefergehende inhaltliche Interpretationen der dargestellten exemplarischen Modellrechnung werden in dieser Arbeit nicht mehr weiter verfolgt, da sie nicht im Fokus der Arbeit liegen. Es sei an dieser Stelle allerdings noch einmal erwähnt, dass die alleinige Betrachtung des Bestimmtheitsmaßes in diesen Modellschätzungen keine kausale Aussagekraft besitzt. Der Zusammenhang soll in diesem Bezug rein statistisch betrachtet werden und die Korrelationen zwischen Nachbarschaftsmerkmalen und dem Haushaltseinkommen quantifizieren, indem die Änderung der abhängigen Variablen bei Änderung der unabhängigen Variablen betrachtet wird.

Das dargestellte Beispiel einer empirischen Untersuchung zeigt die Verknüpfung von flächendeckender Information aus Fernerkundungsdaten und punktueller, sozioökonomischer Information auf Haushaltsebene. Durch die Eigenschaft der Fernerkundungsdaten, dass sie flächenhaft vorliegen, können für die Verknüpfung mit sozialwissenschaftlichen Daten variierende räumliche Einheiten als Bezugsflächen verwendet werden. Im aktuellen Beispiel wird die räumliche Information um den Baublock der Umfrageteilnehmer als Kontextinformation verwendet. In einem weiteren Modell in Tabelle 27 (e) wird zusätzlich noch die nächste Nachbarschaft, bestehend aus den vier umliegenden Baublöcken zur Charakterisierung der umliegenden Bebauungsstruktur verwendet und der Versiegelungsanteil ebenfalls auf dieser Nachbarschaftseinheit in das Modell aufgenommen. Es können also die ausgewerteten Fernerkundungsdaten mit geographischen Informationssystemen beliebig aufbereitet werden und über das räumliche Konzept mit den Befragungsdaten verknüpft werden. Die Beschreibung der Nachbarschaft kann daher einerseits individueller auf die Fragestellung angepasst werden und es können zusätzliche, quantifizierbare Informationen über die Nachbarschaft erhoben werden.

Anhand der deskriptiven Merkmalsvergleiche und der beispielhaften Modellschätzungen konnte aufgezeigt werden, dass eine Verknüpfung von räumlichen Merkmalen aus Fernerkundungsdaten einen sinnvollen Beitrag für die flächendeckende Erfassung der physischen Ausstattung des städtischen Raumes leisten kann. Wichtig bei der Betrachtung scheint, wie in Abschnitt 6.2 bereits dargestellt, die Abgrenzung des betrachteten Raumausschnittes. Für tiefergehende Analysen ist die Erarbeitung eines kausalen Wirkungszusammenhangs in Form einer konkreten Hypothese erforderlich, welche eine darauf abgestimmte Fernerkundungsdatenauswahl und -auswertung unbedingt erforderlich macht.

# Kapitel 7

## Zusammenfassung und Ausblick

*“Alle Aussagen einer empirischen Wissenschaft müssen prinzipiell an der Erfahrung scheitern können.”*

Sir Karl Raimund Popper (1971)

In der vorliegenden Arbeit wird die Verknüpfung von hoch aufgelösten Fernerkundungsdaten und Surveydaten (SOEP und BASE-II) für sozialwissenschaftliche Analysen beschrieben. Die Aufgabenstellung kann nicht durch einen einzigen Verarbeitungsschritt bearbeitet werden, weshalb die Herangehensweise eine Abfolge an verschiedenen konzeptionellen Überlegungen und Methoden darstellt, welche in den einzelnen Kapiteln der Arbeit behandelt werden. Neben der Entwicklung eines konzeptionellen, räumlichen Ansatzes für die Verknüpfung von flächenhaften Fernerkundungsdaten mit punktuellen Personendaten, stand die Entwicklung von Verfahren zur Ableitung und Charakterisierung der urbanen Struktur aus multisensoralen Fernerkundungsdaten im Vordergrund. Die Verfahren wurden beispielhaft anhand von Daten der Untersuchungsgebiete Berlin und München getestet, um sowohl die räumliche Übertragbarkeit der Klassifikationsansätze als auch die Übertragbarkeit zwischen verschiedenen Datenquellen zu überprüfen. Die inhaltliche Analysemöglichkeit der verknüpften Daten mit empirischen, sozialwissenschaftlichen Methoden wird beispielhaft dargestellt. Die Vereinbarkeit und technische Integration der abgeleiteten räumlichen Kontextmerkmale aus Fernerkundungsdaten in sozialwissenschaftliche Analysen wird anhand von Analysepotenzialen untersucht.

### 7.1. Zusammenfassende Betrachtung

Die Erdbeobachtung stellt eine Quelle für die Erhebung von räumlichen, flächendeckenden und aktuellen Datengrundlagen von Stadtregionen dar. Vor dem Hintergrund des Ziels dieser Arbeit, der *empirischen Untersuchung der technischen Einbindung von flächenhaften, räumlichen Kontextmerkmalen aus Erdbeobachtungsdaten in die SOEP/BASE-II-Studien*, zeigen die Resultate, dass:

- durch die Kombination von unterschiedlichen Erdbeobachtungsdaten eine detaillierte Charakterisierung des städtischen Raumes durchgeführt werden kann und daraus räumliche Kontextmerkmale abgeleitet werden können

- die Potenziale der Fernerkundung dabei vor allem in der Erfassung von objektiven, physischen Merkmalen liegen, aber keine subjektiven, sozialen Merkmale erfasst werden.

Im Folgenden werden die zentralen Fragestellungen aus Kapitel 1 noch einmal aufgegriffen und im Kontext der Erkenntnisse betrachtet. Aus der Betrachtung der Grundlagen in Kapitel 2 und dem konzeptionellen Rahmen für die Verknüpfung der beiden Datenquellen in Kapitel 3 geht hervor, dass:

- Fernerkundung und Sozialwissenschaften ein unterschiedliches Verständnis der der Arbeit zugrundeliegenden Begriffe „Gesellschaft“ und „Raum“ aufweisen
- die Verknüpfung dieser Daten sich nicht auf die Fusionierung der Daten beschränkt, sondern eines interdisziplinären Konzeptes bedarf, welches sowohl den räumlichen, zeitlichen, als auch taxonomischen Betrachtungsmaßstab berücksichtigt.

Die physische Ausprägung des städtischen Raums ist die Aneinanderreihung von physischen, stadträumlichen Elementen, wie Gebäuden, Freiflächen und Vegetation. Die räumliche Abfolge und die physiognomische Ausdehnung dieser städtischen Elemente zeichnet dabei ein hoch strukturiertes, heterogenes Bild der urbanen Landschaft, welche sehr große, flächenmäßige Areale umfassen kann. Für die Erfassung dieses komplexen Gebildes sind daher großflächige, detaillierte Datengrundlagen erforderlich. Satellitenbilddaten der aktuellen Generation erfüllen dabei die Anforderungen einer preiswerten, flächendeckenden und detaillierten Datenquelle für die flächenhafte Erfassung der urbanen Struktur. Die Einschränkungen von zweidimensionalen Daten, um die physische urbane Struktur auch in seiner vertikalen Ausdehnung zu erfassen, bedürfen der Integration von zusätzlichen, fernerkundungsgestützten Datengrundlagen. In digitalen Oberflächenmodellen ist die Morphologie der Geländeoberfläche und damit auch die physische, urbane Struktur abgebildet. Aus der Verknüpfung von spektraler Information und Höheninformation geht hervor, dass:

- die hohe geometrische Auflösung der optischen Satellitenbilddaten eine detaillierte Erfassung der flächenmäßigen kleinräumig urbanen Strukturelemente zulässt
- zusätzliche Höhendaten aus digitalen Oberflächenmodellen wichtige komplementäre Informationen für die Abgrenzung von stadträumlichen Elementen, wie Gebäuden und Informationen über ihre Höhe beinhalten
- die Erfassung von großräumigen Stadtregionen oder darüber hinaus stark von der Datenverfügbarkeit und ihrer räumlichen Ausdehnung abhängt.

Vor dem Hintergrund der computergestützten Auswertung von Fernerkundungsbilddaten geht hervor, dass objektbasierte Klassifikationsmethoden ein geeignetes Werkzeug für die Auswertung von geometrisch hoch aufgelösten, multisensoralen Fernerkundungsdaten darstellen. Die Betrachtung der Bildinhalte auf der übergeordneten Informationsebene von

Pixelverbänden (oder: Regionen, Objekten) erlaubt die Verarbeitung auf einer semantischen Ebene und neben den spektralen Informationen der Einzelpixel können auch Nachbarschafts- oder Formmerkmale für die Klassifikation verwendet werden. Die in der Softwareumgebung eCognition© implementierte proprietäre Programmiersprache CNL (Cognition Network Language) erlaubt dabei die Erstellung eines festgelegten Auswertevorganges, welcher die szenenspezifischen oder geographischen Besonderheiten von Fernerkundungsdaten durch die Interaktion des Auswerters berücksichtigen kann. Ferner zeigt sich, dass:

- objektbasierte Verfahren die Auswertung von multisensoralen Datensätzen erlauben
- durch die Betrachtung der Bildinhalte als Objekte Abweichungen in der geometrischen Bildregistrierung einen geringeren Einfluss auf die Auswertegenauigkeit haben als pixelbasierte Methoden
- neben verschiedenen Rasterdatenquellen auch Vektorinformationen wie Baublockgrenzen in der Auswertung berücksichtigt werden können.

Die automatisierte Auswertung von Bildinformationen geht häufig mit einer Verallgemeinerung oder Abstraktion der Bildinhalte einher. Besondere Eigenschaften bestimmter Datenquellen oder geographischer Räume können dabei nicht immer berücksichtigt werden. Bei der Entwicklung und Anwendung von übertragbaren Bildauswerteverfahren ist dabei zu berücksichtigen, dass die Genauigkeiten hinter manuellen Auswertungen oder lokalen Anpassungen zurückbleiben werden. Allerdings stellt die automatische Auswertung ein effizientes und rasches Verfahren dar, mit welchem über 90 Prozent der Landbedeckung für fünf Klassen für die beiden Untersuchungsgebiete Berlin und München richtig klassifiziert wurden.

Die Ableitung der Gebäudegrundflächen aus digitalen Oberflächenmodellen zeigt, dass die Bildauflösung und die Kontraste der Höhenunterschiede für die Abgrenzung von flächenmäßig größeren und auch höheren Gebäuden besser geeignet sind als für die Abgrenzung von kleineren Gebäudestrukturen wie Einzel- oder Doppelhausbebauung. Die Klassifikation der Landbedeckung stellt in der aktuellen Arbeit die Grundlageninformation für die weitere Ableitung von kleinräumigen Nachbarschaftsmerkmalen dar, welche mit den hier benützten Surveydaten verknüpft werden. Auch wenn die Resultate für geometrisch hoch aufgelöste Klassifikationen in urbanen Räumen durchaus die Genauigkeiten von vergleichbaren Studien aufweisen, ist die Güte dieser Information für die weitere Verarbeitung von erheblicher Bedeutung. Fehlklassifikationen auf der Ebene der Einzelobjekte wirken sich auf die darauf aufbauende Klassifikation der Merkmale auf der Baublockebene aus und können bei der Verknüpfung mit sozioökonomischer Information auf der Personen- oder Haushaltsebene zu Abweichungen in der Modellschätzung oder Fehlinterpretationen der Ergebnisse führen.

Die Klassifikation von unterschiedlichen Bebauungstypen auf Basis von geometrischen Merkmalen zeigt ebenfalls vielversprechende Ergebnisse bezüglich ihrer Genauigkeiten. Die

verwendete Diskriminanzanalyse stellt dabei eine probate Methode dar, um die Trennbarkeit von unterschiedlichen Gebäudetypen auf Basis von Formmerkmalen zu untersuchen. Die relevanten Merkmale können damit ausgewählt werden und zur überwachten Klassifikation der verbleibenden Gebäude verwendet werden. Die Untersuchung der vielfältigen geometrischen Merkmale auf der räumlichen Ebene der Einzelgebäude zeigt dabei, dass:

- viele der in der Literatur aufscheinenden geometrischen Merkmale eine Approximation auf eine der geometrischen Grundformen *Kreis*, *Quadrat*, *Rechteck* oder *Würfel* darstellen
- die Trennbarkeit der Gebäudetypen auf Basis von nur einem der Merkmale zu gering ist, bzw. dass sich die Merkmalsausprägungen zwischen den Gebäudetypen sehr stark überlappen können
- durch die Reduktion des multidimensionalen Merkmalsraums auf die erstellten Diskriminanzfunktionen eine deutliche Verbesserung der Trennbarkeit zwischen den Gebäudetypen erzielt werden kann
- die Klassifikationsgüte innerhalb der beiden Untersuchungsgebiete hohe Genauigkeiten zwischen 80 und über 90 % aufweist
- der auf Basis der Diskriminanzanalyse ausgewählte Merkmalsatz der besten trennenden Merkmale für die beiden Untersuchungsgebiete München und Berlin nicht identisch ist
- eine Übertragbarkeit der Diskriminanzfunktionen von einem Untersuchungsgebiet auf das andere nur bedingt möglich ist.

Die räumlichen Kontextmerkmale werden über das räumliche Aggregat der erstellten Landbedeckungsklassifikationen und der abgeleiteten und klassifizierten Bebauungsstruktur gebildet. Die räumliche Ebene des Baublocks stellt im hierarchischen Gefüge des städtischen Systems eine geeignete räumliche Ebene dar, für welche aussagekräftige Nachbarschaftsmerkmale zusammengefasst werden können und schließt die Lücke zwischen dem einzelnen städtischen Strukturelement und den übergeordneten administrativen Raumeinheiten. Die Berechnung von quantifizierbaren Nachbarschaftsmerkmalen, wie Landbedeckungs- und Dichteanteilen auf dieser räumlichen Ebene, ist einfacher als die Aggregation von semantischer Information wie Gebäudetypen.

Die Verknüpfung der abgeleiteten kleinräumigen Nachbarschaftsmerkmale aus Fernerkundungsdaten mit den SOEP/BASE-II-Daten wird über ein spezielles Datenschutzkonzept durchgeführt. Es erlaubt die Verknüpfung von räumlicher Information mit den sensiblen Informationen aus den Surveys. Dabei wird permanent gewährleistet, dass der Nutzer zu keinem Zeitpunkt gleichzeitig auf die räumlichen Daten und die Survey-Informationen zugreifen kann.

Die Verknüpfung findet vollends automatisch statt und erweitert die bestehenden Surveydaten um die verknüpften räumlichen Merkmale. Eine vergleichende Betrachtung der verknüpften Daten zeigt, dass:

- einige räumliche Merkmale im SOEP/BASE-II abgebildet sind, aber der Raum unklar abgegrenzt ist. Der Vergleich mit ähnlichen räumlichen Merkmalen aus Fernerkundungsdaten zeigt allerdings Übereinstimmungen zwischen den unterschiedlich erhobenen Daten
- der wesentliche Vorteil in der Verknüpfung mit Fernerkundungsdaten in der räumlichen Natur der Daten liegt. Der betrachtete Raumausschnitt kann auf der Grundlage der ausgewerteten Satellitenbilder je nach Fragestellung unterschiedlich definiert werden
- alle weiteren Vor- und Nachteile der Fernerkundung zum Tragen kommen: die Daten können flächendeckend, aktuell und kosteneffizient akquiriert und ausgewertet werden. Eine automatische Auswertung ist allerdings mit einem zu berücksichtigenden Fehlerpotenzial behaftet. Darüber hinaus können keine Aussagen über den Zustand oder die Qualität von Stadtteilen oder Wohnungen gemacht werden
- die Fernerkundungsdaten im Hinblick auf die zu bearbeitende sozialwissenschaftliche Fragestellung hin ausgewertet werden müssen
- das räumliche Verständnis der Datengrundlage bzw. das entwickelte räumliche Konzept nicht unbedingt mit dem Raumverständnis von Umfrageteilnehmern einhergehen muss.

Zusammenfassend kann festgehalten werden, dass Fernerkundungsdaten einen relevanten Baustein für die flächenhafte Erfassung und Charakterisierung von Städten darstellen können. Im Rahmen der vorliegenden Arbeit konnten wichtige methodische und technische Grundlagen untersucht und geschaffen werden, auf welche fortführende Arbeiten im thematischen Schnittpunkt der Fernerkundung und sozialwissenschaftlichen Datenanalyse aufbauen können.

## **7.2. Zukünftiger Forschungsbedarf**

Aus den gewonnenen Erkenntnissen dieser Arbeit leitet sich der zukünftige Forschungsbedarf hinsichtlich der synergetischen Analyse von Fernerkundungsdaten und sozialwissenschaftlichen Daten ab. In diesem Zusammenhang spielen neben alternativen fernerkundlichen Datenquellen auch gänzlich unterschiedliche Geodaten eine Rolle. Des Weiteren sind mögliche Anwendungen, methodische Entwicklungen und die Erkenntnis über den Bedarf an aktueller, räumlicher Information für aktuelle sozialwissenschaftliche Fragestellungen und die Vernetzung der Forschungsdisziplinen von Bedeutung.

Die Datengrundlagen in der vorliegenden Arbeit ergaben sich aus der Überlegung, kleinräumige Informationen über hoch strukturierte städtische Räume zu erheben. Entsprechend wurden geometrisch hoch aufgelöste optische Bilder und Höhenaten verwendet. Die Einschränkungen, die diese Datengrundlagen mit sich bringen, liegen in der Natur der Daten selbst: Je höher die geometrische Auflösung ist, desto geringer ist die flächenmäßige Abdeckung der Daten.

Eine flächendeckende Auswertung dieser hoch aufgelösten Daten für die gesamte Bundesrepublik Deutschland ist mit den bekannten Schwierigkeiten verbunden. Die Datenmengen in dieser Größenordnung implizieren, dass die Aufnahmebedingungen von optischen Daten stark von äußeren Einflüssen (z.B. atmosphärische Bedingungen, Sonnenstand, Bewölkung, Jahreszeit) und den Sensoren (z.B. Aufnahmewinkel) abhängig sind, wodurch eine gleichbleibende Datenqualität für die knapp 360000km<sup>2</sup> Fläche der BRD eine große Herausforderung bedeutet. Fernerkundungsdaten, welche tatsächlich für das gesamte Bundesgebiet verfügbar sind, stehen bereits heute mit Image2000 (European Commission, 2012), einem Mosaik aus Landsataufnahmen mit einer Bodenauflösung von 25 m multispektral und 12,5 m panchromatisch und flächendeckenden RapidEye-Daten (RapidEye, 2010), einem Mosaik aus RapidEye-Aufnahmen mit einer Bodenauflösung von 5 m, zur Verfügung. Eine erste Annäherung, inwiefern diese flächendeckenden Daten mit bundesweit repräsentativen Surveydaten verknüpft werden können, um raumbezogene Aussagen mit sozialwissenschaftlicher Relevanz herzustellen, zeigen Goebel und Wurm (2010) beispielhaft mit der Verknüpfung von SOEP-Survey Daten mit CORINE Land Cover Daten (Corine Land Cover, 2010).

Nicht untersucht wurde in dieser Dissertation die Anwendbarkeit von satellitengestützten Radardaten. Die aktuellsten Radarsatelliten TerraSAR-X, TanDEM-X, RADARSAT-2 und ALOS-PaISAR bieten neben einer hohen geometrischen Auflösung von bis zu einem Meter auch beinahe wetter- und tageszeitunabhängige Aufnahmen. Die Komplexität der Bildinhalte in Radardaten ist allerdings durch die besondere Aufnahmegeometrie und die Bildinhalte gegenüber spektralen Bilddaten weitaus höher. Erste globale Produkte wird es mit dem *Global Urban Footprint* (Esch et al., 2011) und weiteren, darauf aufbauenden Produkten geben. Die Verknüpfung dieses Produktes mit Surveydaten wird in Zukunft eine interessante Herausforderung darstellen. Daneben werden in den nächsten Jahren viele weitere Erdbeobachtungssatelliten in die Erdumlaufbahn gebracht und kontinuierlich Daten der Erdoberfläche erheben. Das europäische *SENTINEL*-Programm wird im Rahmen von GMES (*Global Monitoring for Environment and Security*) hochauflösende optische, Radiometrie- und RADAR-Satelliten betreiben, um die unterschiedlichsten Anforderungen an die Erdbeobachtung zu erfüllen.

Im Bereich der Nutzung von weiteren Geodaten bzw. Derivaten aus fernerkundlich erhobenen Informationen wird die Verknüpfung von frei zugänglichen Daten für die

Verknüpfung mit Surveydaten eine interessante Rolle spielen. Das OpenStreetMap-Projekt stellt nutzergenerierte Karteninformationen zur Verfügung, die laufend aktualisiert werden (OpenStreetMap, 2011; Haklay und Weber, 2008). Zu beachten ist in diesem Zusammenhang allerdings die Abdeckung und der Grad der Informationserhebung, welcher regional sehr unterschiedlich sein kann. Stadtstrukturinformationen für europäische Stadtregionen mit mehr als 100000 Einwohnern werden seit 2011 im Rahmen des europäischen Projektes *Urban Atlas* bereitgestellt (EESA, 2012). Auf einer räumlich höher aufgelösten Ebene werden auch in naher Zukunft Gebäudeinformationen als 3D-Gebäudemodell in der Detailstufe LoD1 (Blockmodell bzw. Klötzchenmodell mit Flachdach) von der AdV (Arbeitsgemeinschaft der Vermessungsverwaltungen der Länder der Bundesrepublik Deutschland) ab 2013 bundesweit zur Verfügung gestellt (Gerschwitz, 2011). Auf Basis dieser Datengrundlage und der Weiterentwicklungen des in dieser Arbeit angewendeten Verfahrens zur Klassifikation der Gebäudetypen, wird eine einheitliche Beschreibung von Gebäuden für das gesamte Siedlungsgebiet Deutschlands möglich sein.

Auf der inhaltlichen Ebene ist zu erwarten, dass in der nahen Zukunft durch den breiter werdenden Zugang an flächendeckenden Fernerkundungsdaten und Geodaten die steigende Relevanz von räumlichen Daten für die Erforschung von sozialwissenschaftlichen Phänomenen Synergien zwischen den Forschungsdisziplinen schaffen wird. Die steigende Zahl an sozialwissenschaftlichen Publikationen mit räumlichem Bezug wird auch die Möglichkeiten und Limitierungen der Fernerkundung für die Verknüpfung dieser Daten weiter ausloten und ein Bewusstsein für die gegenseitigen Bedürfnisse der Wissenschaftsdisziplinen schaffen. Anhand von flächendeckenden Daten ist auch die MAUP-Problematik untersuchbar. National denkbar sind in Zukunft neben der Weiterentwicklung von empirischen Studien auch die Unterstützung von Gebäude- und Wohnungszählungen im Rahmen von Zensen (Wurm et al., 2011). Im globalen Umfeld ist denkbar, dass gerade in geographischen Räumen mit geringer oder gar keiner Information über die Bevölkerung die Fernerkundung einen wichtigen Beitrag zur Erfassung und Erforschung von sozialrelevanten Themen leisten kann.

Vor dem Hintergrund, dass in Deutschland immer mehr sozialrelevante Daten für räumliche Analysen aufbereitet werden, wird die Informationsgewinnung aus der Erdbeobachtung an Bedeutung gewinnen. Als Beispiel sei an dieser Stelle der aktuell veröffentlichte Datensatz über die integrierten Erwerbsbiografien des Instituts für Arbeitsmarkt- und Berufsforschung genannt (Scholz et al., 2011). Ebenfalls nicht untersucht wurde die Möglichkeit der Einbindung von mikrogeographischen Marketingdaten (z.B. microm ©), welche ebenfalls georeferenziert verfügbar sind.

Durch die steigende Menge an zur Verfügung stehenden Daten werden in Zukunft auch vermehrt zeitliche Analysen möglich sein. Vor allem im Zusammenhang mit dem SOEP, welches als Längsschnittstudie ausgelegt ist und räumlich Daten im gesamten Bundesgebiet beinhaltet, werden flächendeckende Fernerkundungsdaten mit mehreren zeitlichen

Datenaufnahmen interessant werden. Eine Studie, welche deutschlandweite, satellitengestützte Umweltbelastungsdaten aus einem Erhebungsjahr mit Gesundheitsaspekten im SOEP verknüpft, zeigen Voigtländer et al. (2011). Wenn die Messreihe der Daten lange genug sein wird, wird es auch möglich sein, die Expositionen gegenüber Umweltbelastungen über einen längeren Zeitraum zu überprüfen.

In einem allgemeinen Zusammenhang mit der weiteren Entwicklung für die Geodatenbereitstellung empfiehlt der Rat für Sozial- und Wirtschaftsdaten (RatSWD, 2010):

*„Die große Herausforderung besteht darin, den Zugang zu Regional- und Geodaten auf eine Art zu ermöglichen, die eine einfache Kombination mit anderen Daten erlaubt. Sowohl aktuelle als auch ältere Daten müssen für Längsschnittstudien zugänglich gemacht werden. Insbesondere im Hinblick auf den [...] Zensus (Volkszählung) ist es wichtig, dass die zu Haushalten und Gebäuden erhobenen Daten einen direkten räumlichen Bezug erhalten. Für die Zukunft wird eine Intensivierung der Zusammenarbeit zwischen Sozialwissenschaftlerinnen und Wissenschaftlern von Institutionen der – bisher segregierten – Bereich Geoinformation und informationelle Infrastruktur empfohlen.“*

Resümierend kann festgehalten werden, dass diese Dissertation einen Baustein zur weiteren Entwicklung von multi- bzw. interdisziplinären Datenanalysen auf Basis von Erdbeobachtungsdaten und sozialwissenschaftlichen (Survey-)Daten beiträgt.

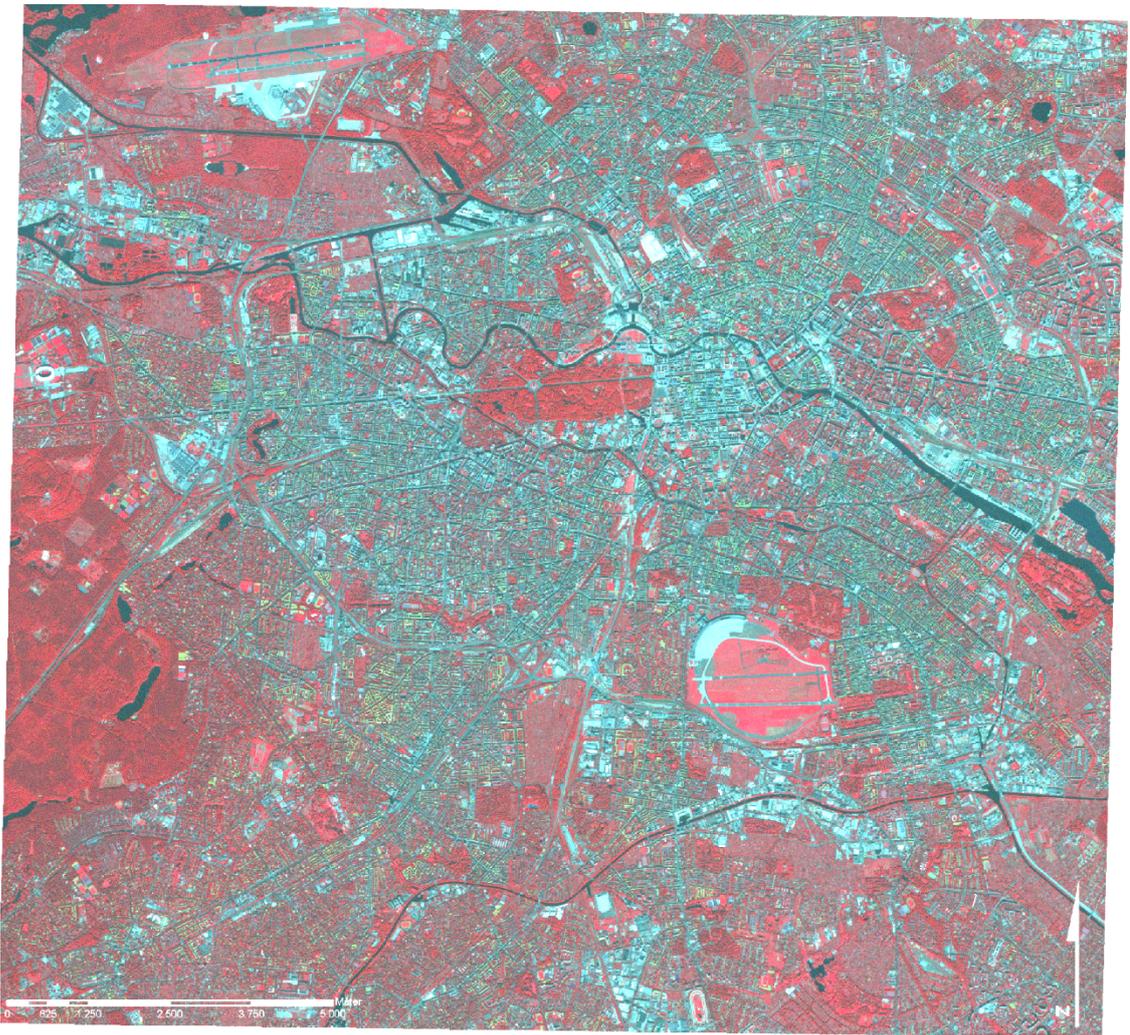
# Anhang A: Fernerkundungsdaten



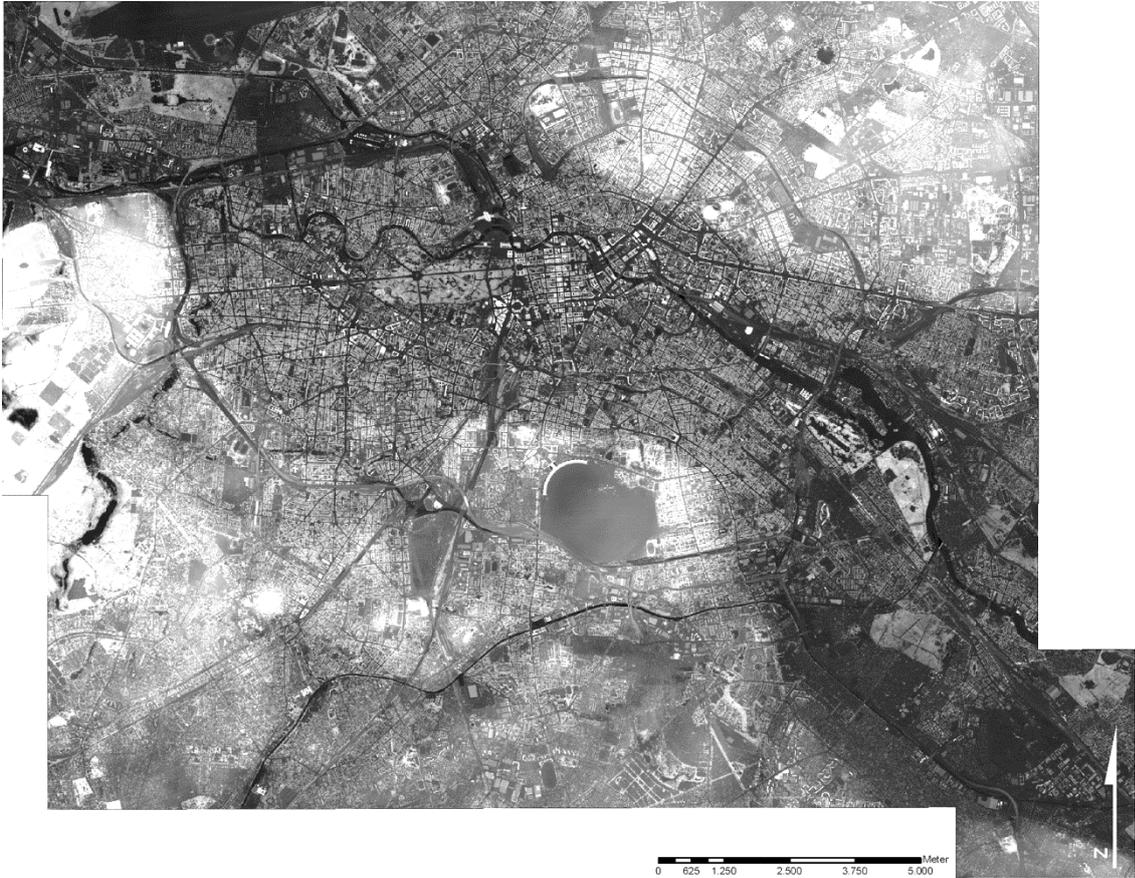
Anhang 1: Ikonos Falschfarbeninfrarotdarstellung für München.



Anhang 2: Digitales Oberflächenmodell für München.

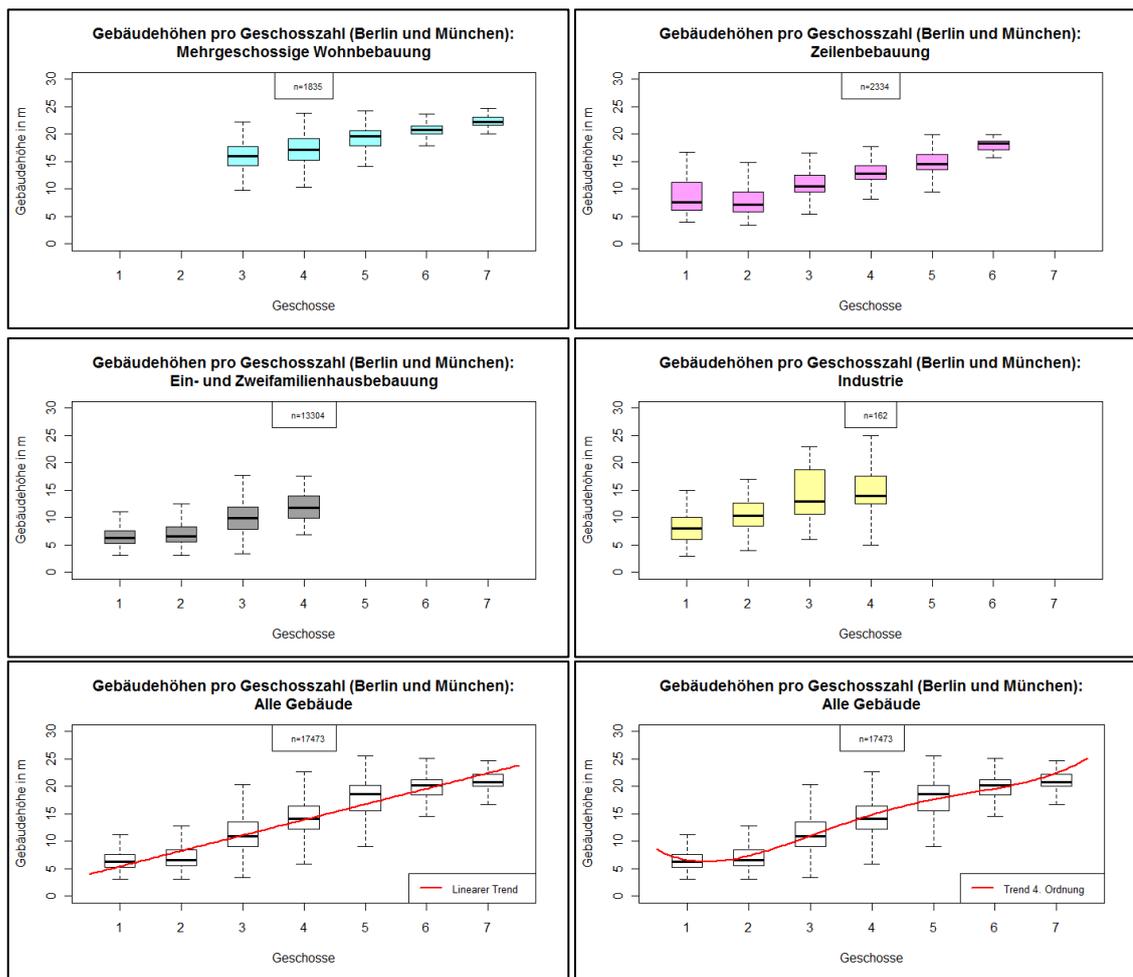


Anhang 3: Quickbird Falschfarbeninfrarotdarstellung für Berlin.



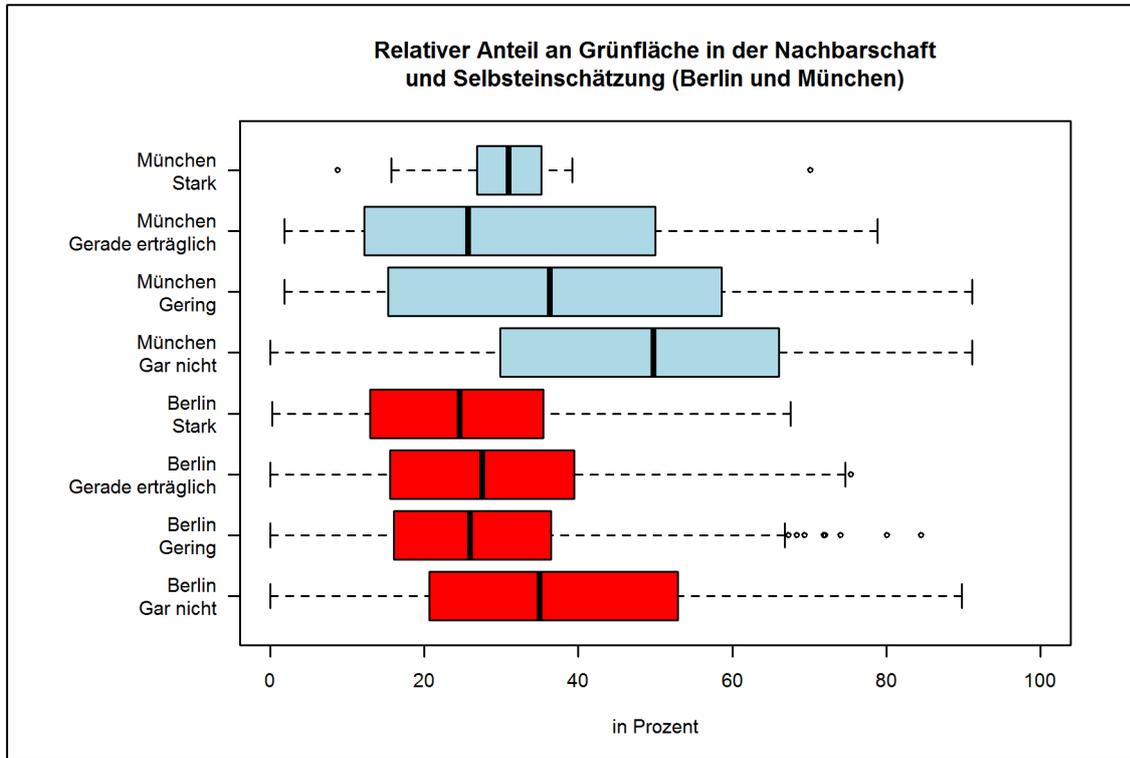
Anhang 4: Digitales Oberflächenmodell für Berlin.

# Anhang B: Boxplots zur Geschosszahlenabschätzung für Berlin und München



Anhang 5: Boxplots zur Geschosszahlenabschätzung für Berlin und München.

# Anhang C: Addendum zu Kapitel 6



**Anhang 6:** Vergleich zwischen Grünanteil in der Nachbarschaft basierend auf Klassifikationsergebnissen aus Erdbeobachtungsdaten und subjektiver Einschätzung getrennt nach München und Berlin.

**Anhang 7:** Deskriptive Statistiken der verwendeten Variablen.

<b>Variable</b>	<b>Beobachtungen</b>	<b>Mittelwert</b>	<b>Mittlerer Fehler</b>	<b>Min.</b>	<b>Max.</b>
Bildungszeit	4627	13,42	3,05	7	18
Wohnumfeld 1					
<i>Freist. 1-2 Fam. Haus</i>	213				
<i>1-2 Fam. Reihenhau</i>	430				
<i>Wohnhaus 3-4</i>	208				
<i>Wohnhaus 5-8</i>	1442				
<i>Wohnhaus &gt;9</i>	3262				
<i>Hochhaus</i>	249				
Wohnumfeld 3					
<i>Altbau</i>	2142				
<i>Neubau</i>	1225				
<i>Mischgebiet</i>	2385				
<i>Geschäftszentrum</i>	37				
Alter	6280	44,69	11,97	26	65
Geschlecht					
<i>Männlich</i>	2966				
<i>Weiblich</i>	3311				
Familienstand					
<i>Verheiratet zus.</i>	2331				
<i>Verheiratet getr.</i>	191				
<i>Ledig</i>	1610				
<i>Geschieden</i>	715				
<i>verwitwet</i>	77				
Anz. Kinder > 14 J.	1293	1,48	0,67	1	4
Ausländer					
<i>Deutscher</i>	4887				
<i>Ausländer</i>	569				
Vegetationsanteil (%)	6280	35,29	21,24	0	92,87
Versiegelungsanteil (4er) (%)	6271	24,29	10,01	0,6	68,85
GFD	6271	1,70	1,16	1,50	6,39
Lage (Distanz) (km)	6280	4,36	2,35	0,16	11,37
Bebauung (Block)					
<i>Ein-Zweifamilienhaus</i>	800				
<i>Zeile</i>	1269				
<i>Block</i>	3719				
<i>Hochhaus</i>	206				
<i>Sonstige</i>	286				
Bebauung (4er)					
<i>Ein-Zweifamilienhaus</i>	601				
<i>Zeile</i>	1055				
<i>Block</i>	4143				
<i>Hochhaus</i>	255				
<i>Sonstige</i>	217				
Stadt					
<i>München</i>	2100				
<i>Berlin</i>	4180				

# Anhang D: Tabellen

**Anhang 8:** Regressionsschätzung zum Nettoäquivalenzeinkommen nur auf Basis der Surveyinformationen ohne Wohnumfeldmerkmale (vgl. Tabelle 27 (a)).

	Koeffizient	Std. Fehler	t Wert	Pr(> t )	
<i>Bildungszeit</i>	73,906	14,869	4,970	0,000	***
<i>Alter</i>	109,993	46,907	2,345	0,019	*
<i>Alter<sup>2</sup></i>	-1,264	0,519	-2,434	0,015	*
<i>Geschlecht weiblich</i>	0,680	120,357	0,006	0,995	
<i>Familienstand verheiratet, getr.</i>	-509,968	284,101	-1,795	0,073	.
<i>Familienstand ledig</i>	-557,551	172,167	-3,238	0,001	**
<i>Familienstand geschieden</i>	-439,929	194,124	-2,266	0,023	*
<i>Familienstand verwitwet</i>	263,510	522,269	0,505	0,614	
<i>Anzahl Kinder &gt; 14 Jahre</i>	-704,658	153,835	-4,581	0,000	***
<i>Ausländer</i>	-352,310	182,877	-1,926	0,054	.
<i>Stadt: München</i>	358,259	125,284	2,860	0,004	**

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

N=1197 adj. R<sup>2</sup>: 0,1421

**Anhang 9:** Regressionsschätzung zum Nettoäquivalenzeinkommen nur auf Basis der Surveyinformationen mit Wohnumfeldmerkmalen (vgl. Tabelle 27 (b)).

	Koeffizient	Std. Fehler	t Wert	Pr(> t )	
<i>Bildungszeit</i>	62,457	13,855	4,508	0,000	***
<i>Wohnumfeld 1: 1-2 Fam.Reihenh.</i>	-195,514	327,241	-0,597	0,551	
<i>Wohnumfeld 1: Wohnhaus 3-4 Whg</i>	735,834	396,141	1,858	0,064	.
<i>Wohnumfeld 1: Wohnhaus 5-8 Whg</i>	-611,326	283,982	-2,153	0,032	*
<i>Wohnumfeld 1: Wohnhaus &gt; 9 Whg</i>	-539,075	276,086	-1,953	0,052	.
<i>Wohnumfeld 1: Hochhaus</i>	-1055,280	426,545	-2,474	0,014	*
<i>Wohnumfeld 3: Wohngeb.Neub</i>	-62,958	157,087	-0,401	0,689	
<i>Wohnumfeld 3: Mischgebiet</i>	-7,737	128,762	-0,060	0,952	
<i>Wohnumfeld 3: Geschäftszentrum</i>	-846,600	682,788	-1,240	0,216	
<i>Alter</i>	96,035	43,214	2,222	0,027	*
<i>Alter<sup>2</sup></i>	-1,088	0,479	-2,269	0,024	*
<i>Geschlecht weiblich</i>	-23,942	110,055	-0,218	0,828	
<i>Familienstand verheiratet, getr.</i>	-280,651	260,810	-1,076	0,283	
<i>Familienstand ledig</i>	-495,800	157,192	-3,154	0,002	**
<i>Familienstand geschieden</i>	-477,847	179,780	-2,658	0,008	**
<i>Familienstand verwitwet</i>	439,776	474,115	0,928	0,354	
<i>Anzahl Kinder &gt; 14 Jahre</i>	-668,302	140,937	-4,742	0,000	***
<i>Ausländer</i>	-305,129	168,838	-1,807	0,071	.
<i>Stadt: München</i>	294,677	125,427	2,349	0,019	*

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

N=1197 adj. R<sup>2</sup>: 0,1966

**Anhang 10:** Regressionsschätzung zum Nettoäquivalenzeinkommen auf Basis der Surveyinformationen und Nachbarschaftsmerkmale aus Fernerkundungsdaten (nur 4er Nachbarschaft) (vgl. Tabelle 27 (c)).

	Koeffizient	Std. Fehler	t Wert	Pr(> t )	
<i>Bildungszeit</i>	58,445	13,908	4,202	0,000	***
<i>Wohnumfeld 1: 1-2 Fam.Reihenh.</i>	-231,139	347,330	-0,665	0,506	
<i>Wohnumfeld 1: Wohnhaus 3-4 Whg</i>	423,076	435,678	0,971	0,332	
<i>Wohnumfeld 1: Wohnhaus 5-8 Whg</i>	-896,910	334,119	-2,684	0,008	**
<i>Wohnumfeld 1: Wohnhaus &gt; 9 Whg</i>	-896,774	344,967	-2,600	0,010	**
<i>Wohnumfeld 1: Hochhaus</i>	-1461,506	580,551	-2,517	0,012	*
<i>Wohnumfeld 3: Wohngeb.Neub</i>	6,719	177,217	0,038	0,970	
<i>Wohnumfeld 3: Mischgebiet</i>	8,138	131,052	0,062	0,951	
<i>Wohnumfeld 3: Geschäftszentrum</i>	-732,141	676,632	-1,082	0,280	
<i>Alter</i>	104,306	43,002	2,426	0,016	*
<i>Alter<sup>2</sup></i>	-1,186	0,477	-2,488	0,013	*
<i>Geschlecht weiblich</i>	-38,496	109,415	-0,352	0,725	
<i>Familienstand verheiratet, getr.</i>	-186,514	261,628	-0,713	0,476	
<i>Familienstand ledig</i>	-486,735	156,228	-3,116	0,002	**
<i>Familienstand geschieden</i>	-493,948	182,955	-2,700	0,007	**
<i>Familienstand verwitwet</i>	475,459	471,710	1,008	0,314	
<i>Anzahl Kinder &gt; 14 Jahre</i>	-687,898	142,242	-4,836	0,000	***
<i>Ausländer</i>	-274,068	167,350	-1,638	0,102	
<i>Vegetationsanteil</i>	334,807	407,743	0,821	0,412	
<i>Versiegelungsanteil (4er)</i>	-201,517	662,486	-0,304	0,761	
<i>GFD</i>	75,350	67,323	1,119	0,264	
<i>Lage &gt; 1 &lt; 2,5km</i>	-402,569	720,289	-0,559	0,577	
<i>Lage &gt; 2,5 &lt; 5km</i>	-879,910	711,668	-1,236	0,217	
<i>Lage &gt; 5 &lt; 7,5km</i>	-522,405	730,919	-0,715	0,475	
<i>Lage &gt; 7,5km</i>	-651,048	732,896	-0,888	0,375	
<i>Bebauung (4er Block) Zeile</i>	120,968	232,304	0,521	0,603	
<i>Bebauung (4er Block) Block</i>	421,114	265,492	1,586	0,113	
<i>Bebauung (4er Block) Hochhaus</i>	390,610	530,732	0,736	0,462	
<i>Bebauung (4er Block) Sonstige</i>	841,467	353,560	2,380	0,018	*
<i>Stadt: München</i>	406,477	135,023	3,010	0,003	**

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

N=1197 adj. R<sup>2</sup>: 0,2231

**Anhang 11:** Regressionsschätzung zum Nettoäquivalenzeinkommen auf Basis der Surveyinformationen und Nachbarschaftsmerkmale aus Fernerkundungsdaten (nur Baublock) (vgl. Tabelle 27 (d)).

	Koeffizient	Std. Fehler	t Wert	Pr(> t )	
<i>Bildungszeit</i>	57,534	13,788	4,173	0,000	***
<i>Wohnumfeld 1: 1-2 Fam.Reihen.</i>	-275,341	329,778	-0,835	0,404	
<i>Wohnumfeld 1: Wohnhaus 3-4 Whg</i>	358,954	417,764	0,859	0,391	
<i>Wohnumfeld 1: Wohnhaus 5-8 Whg</i>	-1095,896	322,381	-3,399	0,001	***
<i>Wohnumfeld 1: Wohnhaus &gt; 9 Whg</i>	-1103,105	330,602	-3,337	0,001	***
<i>Wohnumfeld 1: Hochhaus</i>	-1462,283	536,818	-2,724	0,007	**
<i>Wohnumfeld 3: Wohngeb.Neub</i>	4,252	171,087	0,025	0,980	
<i>Wohnumfeld 3: Mischgebiet</i>	28,610	129,904	0,220	0,826	
<i>Wohnumfeld 3: Geschäftszentrum</i>	-752,886	694,806	-1,084	0,279	
<i>Alter</i>	102,824	43,118	2,385	0,018	*
<i>Alter<sup>2</sup></i>	-1,171	0,478	-2,450	0,015	*
<i>Geschlecht weiblich</i>	-44,872	108,800	-0,412	0,680	
<i>Familienstand verheiratet, getr.</i>	-128,628	260,320	-0,494	0,621	
<i>Familienstand ledig</i>	-457,300	155,218	-2,946	0,003	**
<i>Familienstand geschieden</i>	-507,766	182,556	-2,781	0,006	**
<i>Familienstand verwitwet</i>	393,716	468,470	0,840	0,401	
<i>Anzahl Kinder &gt; 14 Jahre</i>	-653,212	141,518	-4,616	0,000	***
<i>Ausländer</i>	-276,655	167,220	-1,654	0,099	.
<i>Vegetationsanteil</i>	545,522	402,484	1,355	0,176	
<i>Versiegelungsanteil (4er)</i>	30,262	660,247	0,046	0,963	
<i>GFD</i>	83,202	71,305	1,167	0,244	
<i>Lage &gt; 1 &lt; 2,5km</i>	-191,180	738,496	-0,259	0,796	
<i>Lage &gt; 2,5 &lt; 5km</i>	-690,973	727,886	-0,949	0,343	
<i>Lage &gt; 5 &lt; 7,5km</i>	-288,018	752,469	-0,383	0,702	
<i>Lage &gt; 7,5km</i>	-483,611	758,551	-0,638	0,524	
<i>Bebauung (Block) Zeile</i>	634,826	212,754	2,984	0,003	**
<i>Bebauung (Block) Block</i>	686,405	251,270	2,732	0,007	**
<i>Bebauung (Block) Hochhaus</i>	246,948	525,347	0,470	0,639	
<i>Bebauung (Block) Sonstige</i>	705,840	309,012	2,284	0,023	*
<i>Stadt: München</i>	428,599	133,992	3,199	0,001	**

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

N=1197 adj. R<sup>2</sup>: 0,2302

**Anhang 12:** Regressionsschätzung zum Nettoäquivalenzeinkommen auf Basis der Surveyinformationen und Nachbarschaftsmerkmale aus Fernerkundungsdaten (Baublock und 4er Nachbarschaft) (vgl. Tabelle 27 (e)).

	Koeffizient	Std. Fehler	t Wert	Pr(> t )	
<i>Bildungszeit</i>	57,497	13,793	4,169	0,000	***
<i>Wohnumfeld 1: 1-2 Fam.Reihenh.</i>	-146,818	345,154	-0,425	0,671	
<i>Wohnumfeld 1: Wohnhaus 3-4 Whg</i>	402,144	432,185	0,930	0,353	
<i>Wohnumfeld 1: Wohnhaus 5-8 Whg</i>	-1073,969	336,290	-3,194	0,002	**
<i>Wohnumfeld 1: Wohnhaus &gt; 9 Whg</i>	-1076,423	347,274	-3,100	0,002	**
<i>Wohnumfeld 1: Hochhaus</i>	-1515,306	599,383	-2,528	0,012	*
<i>Wohnumfeld 3: Wohngeb.Neub</i>	7,403	175,679	0,042	0,966	
<i>Wohnumfeld 3: Mischgebiet</i>	20,907	130,059	0,161	0,872	
<i>Wohnumfeld 3: Geschäftszentrum</i>	-582,474	694,082	-0,839	0,402	
<i>Alter</i>	109,743	43,055	2,549	0,011	*
<i>Alter<sup>2</sup></i>	-1,250	0,477	-2,620	0,009	**
<i>Geschlecht weiblich</i>	-40,546	108,711	-0,373	0,709	
<i>Familienstand verheiratet, getr.</i>	-127,914	260,074	-0,492	0,623	
<i>Familienstand ledig</i>	-489,268	155,222	-3,152	0,002	**
<i>Familienstand geschieden</i>	-524,069	183,182	-2,861	0,004	**
<i>Familienstand verwitwet</i>	383,982	468,421	0,820	0,413	
<i>Anzahl Kinder &gt; 14 Jahre</i>	-654,792	141,828	-4,617	0,000	***
<i>Ausländer</i>	-241,658	166,848	-1,448	0,148	
<i>Vegetationsanteil</i>	812,078	433,882	1,872	0,062	.
<i>Versiegelungsanteil (4er)</i>	196,218	668,752	0,293	0,769	
<i>GFD</i>	111,680	72,166	1,548	0,122	
<i>Lage &gt; 1 &lt; 2,5km</i>	-115,942	751,913	-0,154	0,878	
<i>Lage &gt; 2,5 &lt; 5km</i>	-619,010	741,281	-0,835	0,404	
<i>Lage &gt; 5 &lt; 7,5km</i>	-250,613	766,930	-0,327	0,744	
<i>Lage &gt; 7,5km</i>	-410,319	767,834	-0,534	0,593	
<i>Bebauung (Block) Zeile</i>	805,132	247,774	3,249	0,001	**
<i>Bebauung (Block) Block</i>	716,577	295,961	2,421	0,016	*
<i>Bebauung (Block) Hochhaus</i>	26,896	552,407	0,049	0,961	
<i>Bebauung (Block) Sonstige</i>	655,200	339,816	1,928	0,055	.
<i>Bebauung (4er Block) Zeile</i>	-331,177	267,740	-1,237	0,217	
<i>Bebauung (4er Block) Block</i>	-45,772	311,958	-0,147	0,883	
<i>Bebauung (4er Block) Hochhaus</i>	225,434	552,996	0,408	0,684	
<i>Bebauung (4er Block) Sonstige</i>	617,740	381,737	1,618	0,106	
<i>Stadt: München</i>	431,239	134,452	3,207	0,001	**

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

N=1197 adj. R<sup>2</sup>: 0,2381

**Anhang 13:** Regressionsschätzung zum Nettoäquivalenzeinkommen auf Basis der Surveyinformationen und Nachbarschaftsmerkmale aus Fernerkundungsdaten (Baublock und 4er Nachbarschaft, ohne Survey-Wohnumfeldmerkmale) (vgl. Tabelle 27 (f)).

	Koeffizient	Std. Fehler	t Wert	Pr(> t )	
<i>Bildungszeit</i>	69,365	14,601	4,751	0,000	***
<i>Alter</i>	124,566	45,690	2,726	0,007	**
<i>Alter<sup>2</sup></i>	-1,385	0,505	-2,744	0,006	**
<i>Geschlecht weiblich</i>	5,825	116,193	0,050	0,960	
<i>Familienstand verheiratet, getr.</i>	-360,464	279,434	-1,290	0,198	
<i>Familienstand ledig</i>	-589,771	168,060	-3,509	0,000	***
<i>Familienstand geschieden</i>	-533,689	193,975	-2,751	0,006	**
<i>Familienstand verwitwet</i>	403,059	505,028	0,798	0,425	
<i>Anzahl Kinder &gt; 14 Jahre</i>	-693,593	152,407	-4,551	0,000	***
<i>Ausländer</i>	-285,607	177,801	-1,606	0,109	
<i>Vegetationsanteil</i>	563,820	448,466	1,257	0,209	
<i>Versiegelungsanteil (4er)</i>	422,608	674,486	0,627	0,531	
<i>GFD</i>	78,849	76,638	1,029	0,304	
<i>Lage &gt; 1 &lt; 2,5km</i>	501,338	646,912	0,775	0,439	
<i>Lage &gt; 2,5 &lt; 5km</i>	-212,719	632,664	-0,336	0,737	
<i>Lage &gt; 5 &lt; 7,5km</i>	-123,821	649,301	-0,191	0,849	
<i>Lage &gt; 7,5km</i>	-75,987	649,295	-0,117	0,907	
<i>Bebauung (Block) Zeile</i>	480,197	250,880	1,914	0,056	.
<i>Bebauung (Block) Block</i>	116,074	294,525	0,394	0,694	
<i>Bebauung (Block) Hochhaus</i>	-923,509	548,623	-1,683	0,093	.
<i>Bebauung (Block) Sonstige</i>	47,122	343,837	0,137	0,891	
<i>Bebauung (4er Block) Zeile</i>	-511,934	272,426	-1,879	0,061	.
<i>Bebauung (4er Block) Block</i>	-327,196	320,610	-1,021	0,308	
<i>Bebauung (4er Block) Hochhaus</i>	-332,005	474,031	-0,700	0,484	
<i>Bebauung (4er Block) Sonstige</i>	1286,010	370,164	3,474	0,001	***
<i>Stadt: München</i>	390,072	138,060	2,825	0,005	**

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
N=1197 adj. R<sup>2</sup>: 0,2148

**Anhang 14:** Regressionsschätzung zum Nettoäquivalenzeinkommen nur auf Basis der Nachbarschaftsmerkmale aus Fernerkundungsdaten (Baublock, ohne Stadt) (vgl. Tabelle 28 (a)).

	Koeffizient	Std. Fehler	t Wert	Pr(> t )
<i>Vegetationsanteil</i>	-31,450	452,340	-0,070	0,945
<i>Versiegelungsanteil (4er)</i>	-219,090	707,930	-0,309	0,757
<i>GFD</i>	-75,740	77,120	-0,982	0,327
<i>Lage &gt; 1 &lt; 2,5km</i>	1061,540	569,230	1,865	0,063
<i>Lage &gt; 2,5 &lt; 5km</i>	295,120	551,660	0,535	0,593
<i>Lage &gt; 5 &lt; 7,5km</i>	480,290	573,510	0,837	0,403
<i>Lage &gt; 7,5km</i>	289,970	583,820	0,497	0,620
<i>Bebauung (Block) Zeile</i>	-112,840	214,390	-0,526	0,599
<i>Bebauung (Block) Block</i>	-271,730	245,470	-1,107	0,269
<i>Bebauung (Block) Hochhaus</i>	-863,080	489,160	-1,764	0,078
<i>Bebauung (Block) Sonstige</i>	-367,340	324,480	-1,132	0,258

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

N=1197 adj. R<sup>2</sup>: 0,03017**Anhang 15:** Regressionsschätzung zum Nettoäquivalenzeinkommen nur auf Basis der Nachbarschaftsmerkmale aus Fernerkundungsdaten (4er Nachbarschaft, ohne Stadt) (vgl. Tabelle 28 (b)).

	Koeffizient	Std. Fehler	t Wert	Pr(> t )
<i>Vegetationsanteil</i>	172,040	442,240	0,389	0,697
<i>Versiegelungsanteil (4er)</i>	-198,870	699,750	-0,284	0,776
<i>GFD</i>	-38,300	72,630	-0,527	0,598
<i>Lage &gt; 1 &lt; 2,5km</i>	920,060	590,940	1,557	0,120
<i>Lage &gt; 2,5 &lt; 5km</i>	198,510	574,750	0,345	0,730
<i>Lage &gt; 5 &lt; 7,5km</i>	254,280	587,810	0,433	0,666
<i>Lage &gt; 7,5km</i>	81,560	590,560	0,138	0,890
<i>Bebauung (4er Block) Zeile</i>	-496,860	232,880	-2,134	0,033 *
<i>Bebauung (4er Block) Block</i>	-559,170	257,660	-2,170	0,030 *
<i>Bebauung (4er Block) Hochhaus</i>	-735,210	413,960	-1,776	0,076
<i>Bebauung (4er Block) Sonstige</i>	674,560	354,060	1,905	0,057

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

N=1197 adj. R<sup>2</sup>: 0,0601

**Anhang 16:** Regressionsschätzung zum Nettoäquivalenzeinkommen nur auf Basis der Nachbarschaftsmerkmale aus Fernerkundungsdaten (Baublock und 4er Nachbarschaft, ohne Stadt) (vgl. Tabelle 28 (c)).

	Koeffizient	Std. Fehler	t Wert	Pr(> t )
<i>Vegetationsanteil</i>	480,468	474,666	1,012	0,312
<i>Versiegelungsanteil (4er)</i>	221,724	714,172	0,310	0,756
<i>GFD</i>	9,895	78,276	0,126	0,900
<i>Lage &gt; 1 &lt; 2,5km</i>	1117,890	631,723	1,770	0,077 .
<i>Lage &gt; 2,5 &lt; 5km</i>	393,790	614,435	0,641	0,522
<i>Lage &gt; 5 &lt; 7,5km</i>	409,704	631,265	0,649	0,517
<i>Lage &gt; 7,5km</i>	234,514	633,179	0,370	0,711
<i>Bebauung (Block) Zeile</i>	329,167	265,870	1,238	0,216
<i>Bebauung (Block) Block</i>	142,979	311,063	0,460	0,646
<i>Bebauung (Block) Hochhaus</i>	-976,821	566,459	-1,724	0,085 .
<i>Bebauung (Block) Sonstige</i>	-168,135	363,935	-0,462	0,644
<i>Bebauung (4er Block) Zeile</i>	-697,016	287,647	-2,423	0,016 *
<i>Bebauung (4er Block) Block</i>	-753,386	336,819	-2,237	0,026 *
<i>Bebauung (4er Block) Hochhaus</i>	-390,532	491,841	-0,794	0,428
<i>Bebauung (4er Block) Sonstige</i>	748,796	391,095	1,915	0,056 .

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

N=1197 adj. R<sup>2</sup>: 0,0683

**Anhang 17:** Regressionsschätzung zum Nettoäquivalenzeinkommen nur auf Basis der Nachbarschaftsmerkmale aus Fernerkundungsdaten (Baublock, mit Stadt) (vgl. Tabelle 29 (a)).

	Koeffizient	Std. Fehler	t Wert	Pr(> t )
<i>Vegetationsanteil</i>	-19,741	448,035	-0,044	0,965
<i>Versiegelungsanteil (4er)</i>	-229,759	701,171	-0,328	0,743
<i>GFD</i>	-5,670	79,487	-0,071	0,943
<i>Lage &gt; 1 &lt; 2,5km</i>	824,950	568,650	1,451	0,148
<i>Lage &gt; 2,5 &lt; 5km</i>	75,991	550,693	0,138	0,890
<i>Lage &gt; 5 &lt; 7,5km</i>	285,907	571,291	0,500	0,617
<i>Lage &gt; 7,5km</i>	142,517	580,085	0,246	0,806
<i>Bebauung (Block) Zeile</i>	-4,204	215,056	-0,020	0,984
<i>Bebauung (Block) Block</i>	-170,162	245,207	-0,694	0,488
<i>Bebauung (Block) Hochhaus</i>	-797,300	484,927	-1,644	0,101
<i>Bebauung (Block) Sonstige</i>	-291,654	322,258	-0,905	0,366
<i>Stadt: München</i>	462,021	144,972	3,187	0,002 **

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
N=1197 adj. R<sup>2</sup>: 0,04862

**Anhang 18:** Regressionsschätzung zum Nettoäquivalenzeinkommen nur auf Basis der Nachbarschaftsmerkmale aus Fernerkundungsdaten (4er Nachbarschaft, mit Stadt) (vgl. Tabelle 29 (b)).

	Koeffizient	Std. Fehler	t Wert	Pr(> t )
<i>Vegetationsanteil</i>	159,960	439,090	0,364	0,716
<i>Versiegelungsanteil (4er)</i>	-269,710	695,200	-0,388	0,698
<i>GFD</i>	15,310	74,600	0,205	0,837
<i>Lage &gt; 1 &lt; 2,5km</i>	753,530	589,710	1,278	0,202
<i>Lage &gt; 2,5 &lt; 5km</i>	46,760	573,200	0,082	0,935
<i>Lage &gt; 5 &lt; 7,5km</i>	146,270	584,870	0,250	0,803
<i>Lage &gt; 7,5km</i>	24,480	586,680	0,042	0,967
<i>Bebauung (4er Block) Zeile</i>	-422,120	232,740	-1,814	0,070
<i>Bebauung (4er Block) Block</i>	-414,620	260,970	-1,589	0,113
<i>Bebauung (4er Block) Hochhaus</i>	-549,130	416,330	-1,319	0,188
<i>Bebauung (4er Block) Sonstige</i>	756,410	352,730	2,144	0,033 *
<i>Stadt: München</i>	404,080	144,310	2,800	0,005 **

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
N=1197 adj. R<sup>2</sup>: 0,07353

**Anhang 19:** Regressionsschätzung zum Nettoäquivalenzeinkommen nur auf Basis der Nachbarschaftsmerkmale aus Fernerkundungsdaten (Baublock und 4er Nachbarschaft, mit Stadt) (vgl. Tabelle 29 (c)).

	Koeffizient	Std. Fehler	t Wert	Pr(> t )
<i>Vegetationsanteil</i>	505,140	470,450	1,074	0,284
<i>Versiegelungsanteil (4er)</i>	184,470	707,840	0,261	0,795
<i>GFD</i>	73,370	80,240	0,914	0,361
<i>Lage &gt; 1 &lt; 2,5km</i>	944,120	628,550	1,502	0,134
<i>Lage &gt; 2,5 &lt; 5km</i>	235,810	611,040	0,386	0,700
<i>Lage &gt; 5 &lt; 7,5km</i>	295,580	626,660	0,472	0,637
<i>Lage &gt; 7,5km</i>	175,820	627,760	0,280	0,780
<i>Bebauung (Block) Zeile</i>	390,470	264,220	1,478	0,140
<i>Bebauung (Block) Block</i>	173,070	308,410	0,561	0,575
<i>Bebauung (Block) Hochhaus</i>	-1016,870	561,500	-1,811	0,071 .
<i>Bebauung (Block) Sonstige</i>	-161,900	360,660	-0,449	0,654
<i>Bebauung (4er Block) Zeile</i>	-652,330	285,420	-2,285	0,023 *
<i>Bebauung (4er Block) Block</i>	-624,790	336,370	-1,857	0,064 .
<i>Bebauung (4er Block) Hochhaus</i>	-184,450	491,950	-0,375	0,708
<i>Bebauung (4er Block) Sonstige</i>	832,240	388,510	2,142	0,033 *
<i>Stadt: München</i>	445,500	144,130	3,091	0,002 **

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

N=1197 adj. R<sup>2</sup>: 0,08502

# Literaturverzeichnis

- ABS (2005) *Australian Standard Geographical Classification (ASGC)*, 28 Sep, [Online], Abrufbar unter:  
<http://abs.gov.au/AUSSTATS/abs@.nsf/lookup/50F6E014E67EBB7CCA256F1900127934?opendocument> [03 Aug 2011].
- Albertz, J. (2001) *Einführung in die Fernerkundung. Grundlagen der Interpretation von Luft- und Satellitenbildern*, Darmstadt: Wissenschaftliche Buchgesellschaft, 249 S.
- Alisch, K., Winter, E. und Arentzen, U. (2005) *Gabler Wirtschaftslexikon*, Gabler, 3840 S.
- Ambrey, C.L. und Fleming, C.M. (2011) 'Valuing scenic amenity using life satisfaction data', *Ecological Economics*, 72, pp. 106-115.
- Aminipouri, M., Sliuzas, R.V. und Kuffer, M. (2009) 'Object oriented analysis of very high resolution orthophotos for estimating the population of slum areas, case of Dar Es Salaam, Tanzania', Proceedings of the ISPRS conference: High-Resolution Earth Imaging for Geospatial Information, ISPRS XXXVIII 1-4-7/WS, Hannover (02.-05.Juni).
- Angel, S., Parent, J. und Civco, D.L. (2010) 'Ten compactness properties of circles: measuring shape in geography', *Canadian Geographer / Le Géographe canadien*, 54 (4), p. 441-461.
- ATCOR (2010) *ATCOR (7.1) [Computersoftware]*, Oberpfaffenhofen: DLR.
- Aubrecht, C., Steinnocher, K., Hollaus, M. und Wagner, W. (2009) 'Integrating earth observation and GIScience for high resolution spatial and functional modeling of urban land use', *Computers, Environment and Urban Systems*, 33 (1), pp. 15-25.
- Avelar, S., Zah, R. und Tavares-Corre, C. (2009) 'Linking socioeconomic classes and land cover data in Lima, Peru: Assessment through the application of remote sensing and GIS', *International Journal of Applied Earth Observation and Geoinformation*, 11 (1), pp. 27-37.
- Baatz, M. und Schäpe, A. (2000) 'Multiresolution segmentation – an optimization approach for high quality multi-scale image segmentation', In: Strobl, J., Blaschke, T. und Griesebner, G. (eds.) *Angewandte Geographische Informationsverarbeitung XII. Beiträge zum AGIT-Symposium*, Salzburg (05.-07. Juli), pp. 12-23.
- Backhaus, K., Erichson, B., Plinke, W. und Weiber, R. (2003) *Multivariate Analysemethoden. Eine anwendungsorientierte Einführung*, Berlin Heidelberg New York: Springer, 715 S.
- Bahrenberg, G., Giese, E. und Nipper, J. (1992) *Statistische Methoden in der Geographie. Band 2: Multivariate Statistik*, Stuttgart: Teubner, 137 S.
- Banzhaf, E. und Höfer, R. (2008) 'Monitoring urban structure types as spatial indicators with CIR aerial photographs for a more effective urban environmental management', *IEEE Journal of selected topics in applied earth observations and remote sensing (J-STARS)*, 1 (2), pp. 129-138.
- BASE (2007) *Berliner Altersstudie (BASE)*, Oktober, [Online], Abrufbar unter: <http://www.base-berlin.mpg.de> [06 Jul 2011].

- BASE II (2010) *Berliner Altersstudie II (Base II): Körperliche und geistige Gesundheit über die Lebensspanne*, Dezember, [Online], Abrufbar unter: [http://www.base-berlin.mpg.de/de/BASE\\_II.html](http://www.base-berlin.mpg.de/de/BASE_II.html) [06 Jul 2011].
- Baud, I., Kuffer, M., Pfeffer, K., Sliuzas, R.V. und Karuppanan, S. (2010) 'Understanding heterogeneity in metropolitan India: the added value of remote sensing data for analyzing sub-standard residential areas', *International Journal of Applied Earth Observation and Geoinformation*, 12 (5), pp. 359-374.
- Baudot, Y. (2000) 'Geographical analysis of the population of fast-growing cities in the third world', in Donnay, J.P., Barnsley, M.J. und Longley, P.A. (ed.) *Remote Sensing and urban analysis*, London: Taylor and Francis, S.229-246.
- Baumgartner, H.J. (2004) 'Are there any class size effects on early career earnings in West Germany?', *DIW Discussion paper*, 417 (April), 23 S.
- BBSR (2011a) *Laufende Raumbewertung - Raumabgrenzungen. Raumordnungsregionen (Analyseräume)*, [Online], Abrufbar unter: [http://www.bbsr.bund.de/nn\\_103086/BBSR/DE/Raumbewertung/Werkzeuge/Raumabgrenzungen/Raumordnungsregionen/raumordnungsregionen.html](http://www.bbsr.bund.de/nn_103086/BBSR/DE/Raumbewertung/Werkzeuge/Raumabgrenzungen/Raumordnungsregionen/raumordnungsregionen.html) [05 Jul 2011].
- BBSR (2011b) *Laufende Raumbewertung - Raumabgrenzungen. Kreise und Kreisregionen*, [Online], Abrufbar unter: [http://www.bbsr.bund.de/nn\\_103086/BBSR/DE/Raumbewertung/Werkzeuge/Raumabgrenzungen/Kreise/kreise.html](http://www.bbsr.bund.de/nn_103086/BBSR/DE/Raumbewertung/Werkzeuge/Raumabgrenzungen/Kreise/kreise.html) [05 Jul 2011].
- Bellmann, L., Kohaut, S. und Lahner, M. (2002) 'Das IAB-Betriebspanel – Ansatz und Analysepotentiale', in Kleinhenz, G. (ed.) *IAB-Kompendium Arbeitsmarkt und Berufsforschung. Beiträge zur Arbeitsmarkt und Berufsforschung* 250, S. 13-20.
- Bertram, L., Demuth, I., Eckardt, R., Heekeren, H., Li, S.-C., Lindenberger, U., Lövdén, M.L., Nietfeld, W., Pawelec, G., Schmiedek, F., Siedler, T., Wagner, G.G. und Steinhagen-Thiessen, E. (2012) 'Cohort profile: The Berlin Ageing Study II (BASE-II)', *unter Begutachtung*.
- Bhatta, B. (ed.) (2010) *Analysis of Urban Growth and Sprawl from Remote Sensing Data*, Heidelberg Dordrecht London New York: Springer, 170 S.
- BHPS (2011) *British household panel study*, [Online], Abrufbar unter: <http://www.iser.essex.ac.uk/bhps> [28 Jun 2011].
- BIK (2001) *BIK Regionen. Ballungsräume, Stadtregionen, Mittel-/Unterzentrengelände*, Oktober, [Online], Abrufbar unter: <http://www.bik-gmbh.de/texte/BIK-Regionen2000.pdf> [07 Jul 2011].
- Bishop, C.M. (1995) *Neural Networks for Pattern Recognition*, Oxford: Clarendon Press, 504 S.
- Blaschke, T. (2010) 'Object based image analysis for remote sensing', *ISPRS Journal of Photogrammetry and Remote Sensing*, 65 (1), pp. 2-16.
- Blaschke, T., Lang, S. und Hay, G.J. (ed.) (2008) *Object-based image analysis. Spatial Concepts for Knowledge-Driven Remote Sensing Applications*, Berlin, Heidelberg: Springer, 817 S.
- Blaschke, T. und Strobl, J. (2001) 'What's wrong with pixels? Some recent developments interfacing remote sensing and GIS', *GIS-Zeitschrift für Geoinformationssysteme*, 14 (6), pp. 12-17.

- Blumberg, D. und Jacobson, D. (1997) 'New frontiers: Remote Sensing in Social Science Research', *The American Sociologist*, 28 (3), pp. 62-68.
- Bochow, M. (2009) 'Automatisierungspotenzial von Stadtbiotopkartierungen durch Methoden der Fernerkundung' Dissertation Universität Osnabrück, 185 S.
- Bortz, J. (1999) *Statistik für Sozialwissenschaftler*, Berlin Heidelberg New York: Springer, S. 585-605.
- Bourdieu, P. (1991) 'Physischer, sozialer und angeeigneter physischer Raum', in Wentz, M. (ed.) *Stadt-Räume.*, Frankfurt a.M., New York: Campus Verlag, S. 25-34.
- Boustedt, O. (1953) 'Die Stadtregion. Ein Beitrag zur Abgrenzung städtischer Agglomerationen', *Allgemeines Statistisches Archiv*, 37 (1), pp. 13-26.
- Boyle, B., Cullis, A., Feng, Z., Flowerdew, R. und Gayle, V. (2004) *Adding Geographical Variables and Identifiers to Longitudinal Datasets*, A Report to the National Longitudinal Strategy Committee and ESRC, 44 S.
- BPB (2004) *Bundezentrale für politische Bildung: Monatliches Haushaltseinkommen pro Kopf nach Haushaltstyp*, Nov, [Online], Abrufbar unter: [http://www.bpb.de/wissen/I500U6,0,Monatliches\\_Haushaltseinkommen\\_pro\\_Kopf\\_nach\\_Haushaltstyp.html](http://www.bpb.de/wissen/I500U6,0,Monatliches_Haushaltseinkommen_pro_Kopf_nach_Haushaltstyp.html) [18 Mar 2012].
- Breiman, L., Friedman, J.H., Olshen, R.A. und Stone, C.J. (1984) *Classification and Regression Trees*, Belmont, Kalifornien, 368 S.: Wadsworth.
- Brenner, C. (2000) 'Dreidimensionale Gebäuderekonstruktion aus digitalen Oberflächenmodellen und Grundrissen' Dissertation, Universität Stuttgart, 122 S.
- Breuste, J., Wächter, M. und Bauer, B. (ed.) (2001) *Beiträge zur umwelt- und sozialverträglichen Entwicklung von Stadtregionen*, CD-ROM, Leipzig - Halle.
- Brosius, F. (1998) *SPSS 8: Professionelle Statistik unter Windows*, Bonn : International Thomson Publishing, S.591-637.
- Brüßhaber, C., Trosset, A.-M. und Bucher, T. (2010) 'Possibilities and constraints in the use of very high spatial resolution UltraCamX airborne imagery and Digital Surface Models for classification in densely built-up areas – a case study Berlin', In: SPIE Europe Remote Sensing 2010, Toulouse (20.-23. September).
- Bundesamt für Kartographie und Geodäsie (2012a) *Digitales Geländemodell Gitterweite 25m*, 16 Apr, [Online], Abrufbar unter: <http://www.geodatenzentrum.de/docpdf/dgm25.pdf> [03 Jul 2012].
- Bundesamt für Kartographie und Geodäsie (2012b) *Digitale Orthophotos Bodenauflösung 40 cm oder 20 cm*, 20 Feb, [Online], Abrufbar unter: <http://www.geodatenzentrum.de/docpdf/dop.pdf> [03 Jul 2012].
- Burgess, E.W. (1925) 'The growth of the city: an introduction to a research project', in Park, R.E., Burgess, E.W. und McKenzie, R.D. (ed.) *The city*, Chicago: Univ. of Chicago Press.
- Burgess, E.W. (1929) 'Urban areas', in Smith, T.V. und White, L.D. (ed.) *Chicago: an experiment in social science research.*, Chicago: Univ. of Chicago Press.
- Cablk, M.E. und Minor, T.B. (2003) 'Detecting and discriminating impervious cover with highresolution IKONOS data using principal component analysis and morphological operators', *International Journal of Remote Sensing*, 24 (23), p. 4627-4645.

- Chen, K. (2002) 'An approach to linking remotely sensed data and areal census data', *International Journal of Remote Sensing*, 23 (1), pp. 37-48.
- Chen, X. und Nordhaus, W.D. (2011) 'Using Luminosity Data as a Proxy for Economic Statistics', *Proceedings of the National Academy of Sciences*, 108 (21), pp. 8589-8594.
- Chen, Y., Su, W., Li, J. und Sun, Z. (2009) 'Hierarchical object oriented classification using very high resolution imagery and LIDAR data over urban areas', *Advances in Space Research*, 43 (7), pp. 1101-1110.
- Chin, W.W. (1998) 'The partial least squares approach for structural equation modeling', in Marcoulides, G.A. (ed.) *Modern Methods for Business Research*, Psychology Press, S. 295-337.
- CNEF (2011) *CNEF*, [Online], Abrufbar unter: <http://www.human.cornell.edu/pam/research/centers-programs/german-panel/cnef.cfm> [02 Aug 2011].
- Cohen, J. (1960) 'A coefficient of agreement for nominal scales', *Educational and Psychological Measurement*, 20 (1), pp. 37-46.
- Congalton, R.G. und Green, K. (2008) *Assessing the accuracy of remotely sensed data: principles and practices*, Boca Raton, FL: CRC Press, 183 S.
- Corine Land Cover (2010) *Bodenbedeckungskarten für Deutschland*, [Online], Abrufbar unter: [http://www.corine.dfd.dlr.de/intro\\_de.html](http://www.corine.dfd.dlr.de/intro_de.html) [15 Feb 2012].
- De Jong, S.M. und van de Meer, F.D. (ed.) (2004) *Remote Sensing Image Analysis. Including the spatial domain*, Dordrecht: Kluwer, Springer, 359 S.
- De Sherbinin, A., Balk, D., Yager, K., Jaiteh, M., Pozzi, F., Giri, G. und Wannebo, A. (2006) *A CIESIN thematic guide of social science applications of remote sensing*, CIESIN - Center for International Earth Science Information Network, Palisades, NY, USA, 68 S.
- Dech, S. (1997) *Anwendung der Satellitenfernerkundung. Von der geowissenschaftlichen Forschung zum operationellen Einsatz*, Habilitationsschrift, Würzburg, 319 S.
- DeStatis (2012) *OECD-Skala*, [Online], Abrufbar unter: [https://www.destatis.de/DE/ZahlenFakten/GesellschaftStaat/Soziales/Sozialberichterstattung/Begriffserlaeuterungen/OECD\\_Skala.html](https://www.destatis.de/DE/ZahlenFakten/GesellschaftStaat/Soziales/Sozialberichterstattung/Begriffserlaeuterungen/OECD_Skala.html) [18 Mar 2012].
- Dietz, R.D. (2002) 'The estimation of neighborhood effects in the social sciences: An interdisciplinary approach', *Social Science Research*, 31, pp. 539-575.
- Digital Globe (2011a) *Quickbird Spacecraft Data Sheet*, Apr, [Online], Abrufbar unter: <http://www.digitalglobe.com/digitalglobe2/file.php/784/QuickBird-DS-OB.pdf> [13 Jun 2011].
- Digital Globe (2011b) *DigitalGlobeBlog*, 18 April, [Online], Abrufbar unter: <http://www.digitalglobeblog.com/2011/04/18/digitalglobe-completes-quickbird-satellite-orbit-raise/> [08 Aug 2011].
- Dittmann, J. und Goebel, J. (2010) 'Your House, Your Car, Your Education: The Socioeconomic Situation of the Neighborhood and its Impact on Life Satisfaction in Germany', *Social Indicators Research*, 96 (3), pp. 497-513.
- Donnay, J.-P., Barnsley, M.F. und Longley, P.A. (ed.) (2001) *Remote Sensing and Urban Analysis*, London: Taylor & Francis, 272 S.

- Downey, L. (2006) 'Using Geographic Information Systems to reconceptualize spatial relationships and ecological context', *American Journal of Sociology*, 112 (2), pp. 567-612.
- Drever, A.I. (2004) 'Separate spaces, separate outcomes? Neighbourhood impacts on minorities in Germany', *Journal of Urban Studies*, 25, pp. 3-21.
- EESA (2012) *EESA – European environment agency. GMES European Atlas*, [Online], Abrufbar unter: <http://www.eea.europa.eu/data-and-maps/data/urban-atlas> [15 Feb 2012].
- Emmanuel, R. (1997) 'Urban Vegetational Change as an Indicator of Demographic Trends in Cities: the Case of Detroit', *Environment and Planning*, 24, pp. 415-426.
- Esch, T. (2006) 'Automatisierte Analyse von Siedlungsflächen auf der Basis höchstauflösender Radardaten' Dissertation, Bayerische Julius-Maximilians Universität Würzburg, 202 S.
- Esch, T., Taubenböck, H., Felbier, A., Roth, A., Müller, A. und Dech, S. (2011) 'The path to mapping the global urban footprint using TandDEM-X data', Proceedings of ISRSE, Sydney, Australien (10.-15. April).
- European Commission (2012) *European Commission – Joint Research Centre. Institute for Environment and Sustainability - Image2000 and Corine Land Cover Project*, [Online], Abrufbar unter: <http://image2000.jrc.ec.europa.eu/> [15 Feb 2012].
- Faes, G. (2010) *LinDA: Diskriminanzanalyse, R package version 1.1*, [Online], Abrufbar unter: <http://www.faes.de> [13 September 2011].
- Filella, I. und Penuelas, J. (1994) 'The red edge position and shape as indicators of plant chlorophyll content, biomass and hydric status', *International Journal of Remote Sensing*, 15 (7), pp. 1495-1470.
- Filin, S. (2004) 'Surface classification from airborne laser scanning data', *Computers & Geosciences*, 30 (9-10), pp. 1033-1041.
- Fisher, R.A. (1936) 'The use of multiple measurements in taxonomic problems', *Annals of Eugenics*, 7 (1), pp. 179-188.
- Foody, G. (2002) 'Status of land cover classification accuracy assessment', *Remote Sensing of Environment*, 80 (1), pp. 185 - 201.
- Forman, R.T.T. und Godron, M. (1986) *Landscape Ecology*, Cambridge: Cambridge University Press, 640 S.
- FORS (2011) *Schweizer Haushalts-Panel (SHP)*, [Online], Abrufbar unter: <http://www.swisspanel.ch/spip.php?lang=de> [28 Jul 2011].
- Förster, P. und Berth, H. (2011) *Die Sächsische Längsschnittstudie*, [Online], Abrufbar unter: <http://www.wiedervereinigung.de/> [28 Jun 2011].
- Gamba, P., Dell'Acqua, F. und Dasarathy, B. (2005) 'Urban remote sensing using multiple data sets: Past, present and future', *Information Fusion*, 6 (4), pp. 319-326.
- Gamba, P. und Herold, M. (ed.) (2009) *Global Mapping of Human Settlement: Experiences, Datasets and Prospects*, Boca Raton, FL: CRC Press, 374 S.
- Gamba, P., Houshmand, B. und Sacconi, M. (2000) 'Detection and extraction of buildings from interferometric SAR data', *IEEE Transactions on Geoscience and Remote Sensing*, 38 (1), p. 611–618.

- Geiß, C., Taubenböck, H., Wurm, M., Esch, T., Nast, M., Schillings, C. und Blaschke, T. (2011) 'Remote sensing-based characterization of settlement structures for assessing local heating potentials', *Remote Sensing*, 3 (7), pp. 1447-1471.
- GeoEye (2009) *GeoEye Product Guide v1.0.1*, [Online], Abrufbar unter: [http://www.geoeye.com/CorpSite/assets/docs/brochures/GeoEye\\_Product\\_Guide.pdf](http://www.geoeye.com/CorpSite/assets/docs/brochures/GeoEye_Product_Guide.pdf) [06 Mar 2010].
- Gerschwitz, A. (2011) *Die dritte Dimension im ALKIS®*, [Online], Abrufbar unter: [www.dvw.de/sites/default/files/./2/./09-Gerschwitz-3D-ALKIS.pdf](http://www.dvw.de/sites/default/files/./2/./09-Gerschwitz-3D-ALKIS.pdf) [15 Feb 2012].
- Gerstorff, D., Ram, N., Goebel, J., Schupp, J., Lindenberger, U. und Wagner, G.G. (2010) 'Where people live and die makes a difference: Individual and geographic disparities in well-being progression at the end of life', *Psychology and Aging*, 25 (3), pp. 661-676.
- Gibson, C.C., Ostrom, E. und Ahn, T.K. (2000) 'The concept of scale and the human dimensions of global change: a survey', *Ecological Economics*, 32 (2), pp. 217-239.
- Goebel, J. (2007) *Kurzinformationen zum SOEP-Geocode*, Berlin: Das Sozio-oekonomische Panel. Unveröffentlichtes Manuskript, 10 S.
- Goebel, J. (2009) *Datenschutzkonzept zur Nutzung von SOEPgeo am DIW Berlin*, Berlin: Das Sozio-oekonomische Panel. Unveröffentlichtes Manuskript, 5 S.
- Goebel, J., Grabka, M.M., Krause, P., Kroh, M., Pischner, R. und Sieber, I. (2008) 'Mikrodaten, Gewichtung und Datenstruktur der Längsschnittstudie Sozio-oekonomisches Panel (SOEP)', *Vierteljahrshefte zu Wirtschaftsforschung*, 77 (3), pp. 77-109.
- Goebel, J., Spieß, C.K., Witte, N.R.J. und Gerstenberg, S. (2007) *Die Verknüpfung des SOEP mit MICROM-Indikatoren: Der MICROM-SOEP Datensatz.*, Berlin: DIW, Data Documentation 26, 43 S.
- Goebel, J. und Wurm, M. (2010) 'Räumliche Unterschiede im Armutsrisiko in Ost- und Westdeutschland', in Krause, P. und Ostner, I. (ed.) *Leben in Ost- und Westdeutschland*, Campus, S. 673-692.
- Goodchild, M.F. (2007) 'The Morris Hansen Lecture 2006 Statistical Perspectives on Spatial Social Science', *Journal of Official Statistics*, 23 (3), pp. 269-283.
- Goodchild, M.F. und Janelle, D.G. (2004a) 'Thinking spatially in the social sciences', in Goodchild, M.F. und Janelle, D.G. (ed.) *Spatially integrated social science*, New York: Oxford University Press, 480 S.
- Goodchild, M.F. und Janelle, D.G. (2004b) *Spatially integrated social science*, New York: Oxford University Press, 480 S.
- Goodwin, N.R., Coops, N.C. und Culvenor, D.S. (2006) 'Assessment of forest structure with airborne lidar and the effects of platform altitude', *Remote Sensing of Environment*, 103 (2), pp. 140-152.
- Greve, W. und Wentura, D. (1997) *Wissenschaftliche Beobachtung. Eine Einführung*, Weinheim: Beltz/Psychologie Verlags Union, 182 S.
- Griffiths, P., Hostert, P., Gruebner, O. und van der Linden, S. (2010) 'Mapping megacity growth with multi-sensor data', *Remote Sensing of Environment*, 114 (2), pp. 426-439.

- Gruber, M. (2007) 'UltraCamX, the new Digital Aerial Camera System by Microsoft Photogrammetry', *Photogrammetrische Woche '07, Stuttgart* (03.-07.September.), 137-146.
- Gustafson, E.J. und Parker, G.R. (1994) 'Using an index of habitat patch proximity for landscape design', *Landscape and Urban Planning*, 29 ( 1-2 ), pp. 117- 130.
- Haala, N. und Brenner, C. (1999) 'Extraction of buildings and trees in urban environments', *ISPRS Journal of Photogrammetry and Remote Sensing*, 54 (2-3), pp. 130-137.
- Hackl, P. (2008) *Einführung in die Ökonometrie*, Bonn: Addison-Wesley Verlag, 448 S.
- Häder, M. (2006) *Empirische Sozialforschung. Eine Einführung*, Wiesbaden: VS Verlag für Sozialwissenschaften, 497 S.
- Haisken-DeNew, J. (2001) 'A hitchhiker's guide to the world's household panel data sets', *The Australian Economic Review*, 34 (3), pp. 356-366.
- Haklay, M. und Weber, P. (2008) 'OpenStreetMap: User-generated street maps', *Pervasive Computing*, pp. 12-18.
- Hall, O. (2010) 'Remote Sensing in Social Science Research', *The Open Remote Sensing Journal*, pp. 1-16.
- Handl, A. (2002) *Multivariate Analysemethoden*, Berlin Heidelberg New York: Springer, S. 315-361.
- Hank, K. und Kreyenfeld, M. (2004) 'A multilevel analysis of child care and women's fertility decisions in Western Germany', *Journal of Marriage and Family*, 33, pp. 228-244.
- Haralick, R.M. und Shapiro, L.G. (1985) 'Survey: Image segmentation techniques', *Computer Vision, Graphics, and Image Processing*, 29 (1), pp. 100-132.
- Harris, D.R. (1999) 'Property values drop when blacks move in, because: Racial and socioeconomic determinants of neighborhood desirability', *American Sociological Review*, 64 (3), p. 461–479.
- Harris, C.D. und Ullman, E.L. (1945) 'The nature of cities', *Annals of the American Academy for Political Science*, November, pp. 7-17.
- Harris, P.M. und Ventura, S.J. (1995) 'The Integttation of Geographic Data with Remotely sensed imagery to improve Classification in an Urban Area', *Photogrammetric Engineering & Remote Sensing*, 61 (8), pp. 993-99.
- Hastie, T., Tibshirani, R. und Friedman, J. (2011) *The elements of statistical learning*, New York: Springer, S.101-137.
- Heiden, U., Heldens, U., Roessner, S., Segl, K., Esch, T. und Müller, A. (2012) 'Urban structure type characterization using hyperspectral remote sensing and height information', *Landscape and Urban Planning*, 105 (4), pp. 361-375.
- Heldens, W. (2010) 'Use of airborne hyperspectral data and height information to support urban micro climate characterisation' Dissertation Universität Würzburg, 180 S.
- Henderson, J.V., Storeygard, A. und Weil, D.N. ( 2012) 'Measuring Economic Growth from Outer Space', *American Economic Review*, 102 (2), p. 994–1028.

- Henn, A., Römer, C., Gröger und Plümer, L. (2012) 'Automatic classification of building types in 3D city models. Using SVMs for semantic enrichment of low resolution building data', *Geoinformatica*, 16 (2), pp. 281-306.
- Herold, M. (2004) 'Remote Sensing and Spatial Metrics for Mapping and Modeling of Urban Structures and Growth Dynamics' Dissertation, University of California, 325 S.
- Herold, M., Liu, X.H. und Clarke, K.C. (2003) 'Spatial metrics and image texture for mapping urban land use', *Photogrammetric Engineering & Remote Sensing*, 69 (9), pp. 991-1001.
- Herold, M., Scepan, J. und Clarke, K.C. (2002) 'The use of remote sensing and landscape metrics to describe structures and changes in urban land uses', *Environment and Planning A*, 34 (8), pp. 1443-1458.
- Hildebrandt, G. (1996) *Fernerkundung und Luftbildmessung*, Heidelberg: Wichmann, 676 S.
- Hintze, P. und Lakes, T. (2009) 'Geographically referenced data for social science', *RatSWD Working paper series*, Oktober 2009 (125), pp. 1-50.
- Hirschmüller, H. (2006) 'Stereo Vision in Structured Environment by Consistent Semi-Global Matching', Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition (CVPR 2006), New York, NY, USA.
- Hoffmeyer-Zlotnik, J.H.P. (2000) *Regionalisierung sozialwissenschaftlicher Umfragedaten*, Wiesbaden: Westdeutscher Verlag, 249 S.
- Hoyt, H. (1939) *The structure and growth of residential neighborhoods in American cities*, Washington: Federal Housing Association.
- Jensen, J.R. und Cowen, D.C. (1999) 'Remote Sensing of Urban/Suburban Infrastructure and Socio-Economic Attributes', *Photogrammetric Engineering & Remote Sensing*, 65 (5), pp. 611-622.
- Jensen, R., Gatrell, J., Boulton, J. und Harper, B. (2004) 'Using Remote Sensing and Geographic Information Systems to Study Urban Quality of Life and Urban Forest Amenities', *Ecology and Society*, 9 (5), 10 S.
- Kettig, R.L. und Landgrebe, D.A. (1976) 'Classification of multispectral image data by extraction and classification of homogeneous objects', *IEEE Transactions on Geoscience Electronics*, 14 (1), pp. 19-26.
- Kraft, R. (2000) *Biometrische und Ökonometrische Methoden: Diskriminanzanalyse*, Unveröffentlichtes Manuskript: TU München-Weihenstephan, Mathematik und Statistik, 25 S.
- Kraus, K. (2004) *Photogrammetrie*, Berlin, New York: Walter de Gruyter, 516 S.
- Krzanowski, W.J. (1988) *Principles of Multivariate Analysis. A User's Perspective*, Oxford: Oxford University Press, 586 S.
- Lakes, T. (2010) 'Geodata', in RatSWD (ed.) *Building on Progress. Expanding the Research Infrastructure for the Social, Economic, and Behavioral Sciences*, Opladen & Farmington Hills, MI: Budrich UniPress Ltd.
- Landeshauptstadt München (1963) *Stadtentwicklungsplan mit Gesamtverkehrsplan. Gekürzte Fassung*, München: <http://www.ba-muenchen.info/planref/SEP1963/index.html>.

- Landeshauptstadt München (2000) *Mittlerer Ring. Stadt- und freiraumplanerische Studie*, München, 144 S.
- Landgrebe, D.A. und Biehl, L. (2011) *Multispec. A Freeware Multispectral Image Data Analysis System*, 2 September, [Online], Abrufbar unter: <http://www.ece.purdue.edu/~biehl/MultiSpec/> [07 September 2011].
- Landis, J.R. und Koch, G.G. (1977) 'The measurement of observer agreement for categorical data', *Biometrics*, 33, pp. 159-174.
- Lang, S. (2008) 'Object-based image analysis for remote sensing applications: modeling reality – dealing with complexity', in Blaschke, T., Lang, S. und Hay, G.J. (ed.) *Object-based image analysis*, Berlin Heidelberg: Springer, S. 3-28.
- Lang, S. und Blaschke, T. (2007) *Landschaftsanalyse mit GIS*, Stuttgart: UTB, 404 S.
- Lichtenberger, E. (1998) *Stadtgeographie. Band 1 - Begriffe, Konzepte, Modelle, Prozesse.*, Stuttgart und Leipzig: Teubner, 366 S.
- Lillesand, T.M., Kiefer, R.W. und Chipman, J.W. (2004) *Remote Sensing and image interpretation*, Hoboken, NJ: John Wiley & Sons, 763 S.
- Lindenberger, U., Smith, J., Mayer, K.U. und Baltes, P.B. (ed.) (2010) *Die Berliner Altersstudie*, Berlin: Akademie Verlag, 748 S.
- Liu, X.H., Clarke, K. und Herold, M. (2006) 'Population Density and Image Texture: A Comparison Study', *Photogrammetric Engineering & Remote Sensing*, 72 (2), pp. 187-196.
- Liverman, D.M. und Cuesta, R.M.R. (2008) 'Human interactions with the Earth system: people and pixels revisited', *Earth Surface Processes and Landforms*, 33 (9), p. 1458-1471.
- Lobo, A. (1997) 'Image Segmentation and Discriminant Analysis for the Identification of Land Cover Units in Ecology', *IEEE Transactions on Geoscience and Remote Sensing*, 35 (5), pp. 1136-1145.
- Lo, C.P. und Faber, B.J. (1997) 'Integration of Landsat Thematic Mapper and Census Data for Quality of Life Assessment', *Remote Sensing of Environment*, 62 (2), pp. 143-157.
- Luechinger, S. (2009) 'Valuing air quality using the life satisfaction approach', *The Economic Journal*, 119 (536), p. 482-515.
- LVC (2008) *Segmentation Parameter Tuner - SPT*, Verfügbar unter: [www.lvc.ele.puc-rio.br](http://www.lvc.ele.puc-rio.br).
- Maktav, D., Erbek, F.S. und Jürgens, C. (2005) 'Remote sensing of urban areas', *International Journal of Remote Sensing*, 26 (4), pp. 655-659.
- Mandelbrot, B. (1982) *The fractal geometry of nature*, New York: W.H. Freeman, 460 S.
- Mather, P.M. (1999) *Computer Processing of Remotely-Sensed Images*, Chichester, New York, Weinheim, Brisbane, Singapore, Toronto: John Wiley & Sons, 292 S.
- Mayer, S. (2000) 'Retrieving object borders and height from high-resolution digital surface models', *Proceedings of the World Multiconference on Systems, Cybernetics and Informatics*, Orlando, FL, USA, 433-438.
- Mayer, S. (2004) 'Automatisierte Objekterkennung zur Interpretation hochauflösender Bilddaten in der Erdfernerkundung' *Dissertation*, Humboldt Universität, Berlin, 191 S.
- Mayer, K.U. und Baltes, P.B. (ed.) (1996) *Die Berliner Altersstudie. Akademie Verlag*, Berlin, 748 S.: Akademie Verlag.

- MBO (2002) *Musterbauordnung*, [Online], Abrufbar unter: <http://www.umweltdaten.de/rup/Musterbauordnung.pdf> [02 September 2011].
- McGarigal, K., Cushman, S.A., Neel, M.C. und Ene, E. (2011) *FRAGSTATS: Spatial Pattern Analysis Program for Categorical Maps*, [Online], Abrufbar unter: [www.umass.edu/landeco/research/fragstats/fragstats.html](http://www.umass.edu/landeco/research/fragstats/fragstats.html) [14 Jun 2011].
- McGarigal, K. und Marks, B.J. (1995) *FRAGSTATS: spatial pattern analysis program for quantifying landscape structure*, Portland, OR: U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station, 122 S.
- McGill, R., Tukey, J.W. und Larsen, W.A. (1978) 'Variations of Box Plots', *The American Statistician*, 32 (1), pp. 12-16.
- Meinel, G., Hecht, R., Herold, H., Schiller, G., Spangenberg, M., Burgdorf, M. und Kawka, R. (2008) *Automatische Ableitung von stadtstrukturellen Grundlagendaten und Integration in einem Geographischen Informationssystem*, (= Schriftenreihe Forschungen des Bundesministeriums für Verkehr, Bau und Stadtentwicklung und das Bundesamt für Bauwesen und Raumordnung (BBR) Nr. 134): BMVBS und BBR (Hrsg.), 108 S.
- Melbourne Institute (2011) *The household, income and labour dynamics in Australia study*, 23 Jun, [Online], Abrufbar unter: <http://melbourneinstitute.com/hilda/> [28 Jun 2011].
- Mesev, V. (ed.) (2003) *Remotely Sensed Cities*, London New York: Taylor & Francis, 433 S.
- Miliaresis, G. und Kokkas, N. (2007) 'Segmentation and object based classification for the extraction of the building class from LIDAR DEMs', *Computers & Geosciences*, 33 (8), pp. 1076-1087.
- Møller-Jensen, L. (1990) 'Knowledge-Based Classification of an Urban Area Using Texture and Context Information in Landsat-TM Imagery', *Photogrammetric Engineering and Remote Sensing*, 56 (6), pp. 899-904.
- Müller, M. (2001) 'Extraktion großflächiger Strukturen aus hochauflösenden panchromatischen Satellitendaten' Dissertation, Universität Bielefeld, 158 S.
- Murakami, H., Nakagawa, K., Hasegawa, H., Shibata, T. und Iwanami, E. (1999) 'Change detection of buildings using an airborne laser scanner', *ISPRS Journal of Photogrammetry and Remote Sensing*, 54 (2-3), pp. 148-152.
- Netzband, M. (ed.) (2007) *Applied Remote Sensing for Urban Planning, Governance and Sustainability*, Heidelberg Dordrecht London New York: Springer, 278 S.
- Neubert, M. (2005) 'Bewertung, Verarbeitung und segmentbasierte Auswertung sehr hoch auflösender Satellitenbilddaten vor dem Hintergrund landschaftsplanerischer und landschaftsökologischer Anwendungen' Dissertation, Technische Universität Dresden, 172 S.
- Neukum, G., Oberst, J., Schwarz, G., Flohrer, J., Sebastian, I., Jaumann, R., Hoffmann, H., Carsenty, U., Eichertopf, K. und Pischel, R. (1995) 'The multiple line scanner camera experiment for the Russian Mars 96 Mission: Status report and prospects for the future', *Photogrammetrische Woche '95*, Stuttgart (11.-15. September), 45-61.
- Nilsson, M. (1996) 'Estimation of tree heights and stand volume using an airborne lidar system', *Remote Sensing of Environment*, 56 (1), pp. 1-7.

- O'Neill, R.V., Krummel, J.R., Gardner, R.H., Sugihara, G., Jackson, B., Deangelis, D.L., Milne, B.T., Turner, M.G., Zygmunt, B., Christensen, S.W., Dale, V.H. und Graham, R.L. (1988) 'Indices of landscape pattern', *Landscape Ecology*, 1 (1), pp. 153-162.
- Oczipka, M. (2007) 'Objektbasierte Klassifizierung hochauflösender Daten in urbanen Räumen unter besonderer Berücksichtigung von Oberflächenmodellen' Dissertation, Freie Universität Berlin, 166 S.
- Openshaw, S. und Taylor, P.J. (1981) 'The modifiable areal unit problem', in Wrigley, N. und Bennett, R.J. (ed.) *Quantitative Geography: A British View*, London: Routledge and Kegan Paul.
- OpenStreetMap (2011) *OpenStreetMap*, 13 August, [Online], Abrufbar unter: <http://www.openstreetmap.org/> [13 August 2011].
- Owen, D., Green, A. und Elias, P. (2009) 'Review of spatial data needs and resources', *RatSWD Working paper series*, Juli 2009 (88), pp. 1-45.
- Pacifici, F., Chini, M. und Emery, W.J. (2009) 'A neural network approach using multi-scale textural metrics from very high-resolution panchromatic imagery for urban land-use classification', *Remote Sensing of Environment*, 113 (6), p. 1276-1292.
- Park, R.E. (1952) *Human communities: the city and human ecology*, New York: Free Press, 278 S.
- Parzen, E. (1962) 'On Estimation of a Probability Density Function and Mode', *Annals of Mathematical Statistics*, 33 (3), pp. 1065-1076.
- Popper, K.R. (1971) *Logik der Forschung*, Tübingen.
- PSID (2011) *A national study of socioeconomic and health over lifetimes and across generations*, [Online], Abrufbar unter: <http://psidonline.isr.umich.edu/> [28 Jun 2011].
- Puissant, A., Hirsch, J. und Weber, C. (2005) 'The utility of texture analysis to improve per-pixel classification for high to very high spatial resolution imagery', *International Journal of Remote Sensing*, 26 (4), p. 733-745.
- R Development Core Team (2010) *R: A language and environment for statistical computing*, [Online], Abrufbar unter: [URL http://www.R-project.org/](http://www.R-project.org/) [11 Dec 2011].
- RapidEye (2010) *RapidEye German Mosaic now available*, 21 Dec, [Online], Abrufbar unter: <http://www.rapideye.net/news/pr/2010-german%20mosaic.htm> [15 Feb 2012].
- Rashed, T. und Jürgens, C. (ed.) (2010) *Remote Sensing of Urban and Suburban Areas*, Dordrecht Heidelberg London New York: Springer, 352 S.
- RatSWD (2010) *Auf Erfolgen aufbauend. Zur Weiterentwicklung der Forschungsinfrastruktur für die Sozial-, Verhaltens- und Wirtschaftswissenschaften. Empfehlungen des Rates für Sozial- und Wirtschaftsdaten.*, Opladen & Farmington Hills, MI: Budrich UniPress Ltd.
- Rehdanz, K. und Maddison, D. (2009) 'The amenity value of climate to German households', *Oxford Economic Papers*, 61, pp. 150-167.
- Reibel, M. (2007) 'Geographic Information Systems and spatial data processing in demography: a review', *Population Research and Policy Review*, 26 (5-6), pp. 601-618.
- Reiß-Schmidt, S. (2010) 'München: "Wachstum nach Innen" - Nachhaltige Stadtentwicklung zwischen Wandel und Identität', in Taubenböck, H. und Dech, S. (ed.) *Fernerkundung im urbanen Raum*, Darmstadt: Wissenschaftliche Buchgesellschaft, S. 15-30.

- Richards, J.A. und Jia, X. (2006) *Remote Sensing Digital Image Analysis*, Berlin, Heidelberg: Springer, 439 S.
- Richter, R. (1996) 'A spatially adaptive fast atmospheric correction algorithm', *International Journal of Remote Sensing*, 17 (6), pp. 1201-1214.
- Richter, R. (2010) *Atmospheric/topographic correction for satellite imagery (ATCOR-2/3 User Guide, Version 7.1)*, [Online], Abrufbar unter: [http://www.rese.ch/pdf/atcor3\\_manual.pdf](http://www.rese.ch/pdf/atcor3_manual.pdf) [14 November 2010].
- Rindfuss, R.R. und Stern, P.C. (1998) 'Linking remote sensing and social science: the need and the challenges', in Liverman, D., Moran, E.F., Rindfuss, R.R. und Stern, P.C. (ed.) *People and Pixels: Linking remote sensing and social science*, Washington, D.C.: National Academy Press, S. 1-27.
- Ripley, B.D. (1996) *Pattern Recognition and Neural Networks*, Cambridge: Cambridge University Press, 416 S.
- Roth, M., Oke, T.R. und Emery, W.J. (1989) 'Satellite-derived urban heat islands from three coastal cities and the utilization of such data in urban climatology', *International Journal of Remote Sensing*, 10 (11), pp. 1699-1720.
- Schell, L.M. und Ulijaszek, S.J. (ed.) (1999) *Urbanism, health and human biology in industrialized countries*, Cambridge, England: Cambridge University Press, 344 S.
- Schiewe, J. (2001) 'Ein regionen-basiertes Verfahren zur Extraktion der Geländeoberfläche aus digitalen Oberflächen-Modellen', *Photogrammetrie - Fernerkundung - Geoinformation*, 2/2001, pp. 81-90.
- Schiewe, J. (2002) 'Segmentation of high-resolution remotely sensed data – concepts, applications and problems', In: Armenakis, C. and Lee, Y.C. (eds.), *Symposium on Geospatial Theory, Processing and Applications*, Ottawa (9.-12. Juli).
- Scholten, F., Gwinner, K., Tauch, R. und Boulgakova, O. (2003) 'HRSC-AX - High-resolution orthoimages and digital surface models for urban regions', 2. GRSS/ISPRS Joint Workshop on Remote Sensing and Data Fusion over Urban Areas (URBAN 2003), Berlin (22.-23. Mai).
- Scholz, T., Bachteler, T., Bender, S. und Erler, C. (2011) *Die Georeferenzierung der IEB – Erste Ergebnisse*, Nürnberg, Vortrag.
- Schopenhauer, A. (1851) *Parerga und Paralipomena. Kleine philosophische Schriften*, A.W. Hahn, 467 S.
- Schulte, W., Sukopp, W., Voggenreiter, V. und Werner, P. (1986) 'Flächendeckende Biotopkartierung im besiedelten Bereich als Grundlage einer ökologisch bzw. am Naturschutz orientierten Planung', *Natur und Landschaft*, 61 (10), pp. 371-389.
- Schulte, W., Sukopp, W. und Werner, P. (1993) 'Flächendeckende Biotopkartierung im besiedelten Bereich als Grundlage einer am Naturschutz orientierten Planung', *Natur und Landschaft*, 68 (10), pp. 491-526.
- Sheather, S.J. (2004) 'Density Estimation', *Statistical Science*, 19 (4), pp. 588-597.
- Sirmacek, B., Taubenböck, H., Reinartz, P. und Ehlers, M. (2012) 'Performance evaluation for 3-D city model generation of six different DSMs from air- and spaceborne sensors', *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 5(1), pp. 59-70.

- Sliuzas, R.V., Kerle, N. und Kuffer, M. (2008) 'Object-oriented mapping of urban poverty and deprivation', Proceedings of the 4th EARSeL workshop on remote sensing for developing countries in conjunction with GISDECO 8, Istanbul, Türkei (04.-07.Juni).
- Sliuzas, R.V. und Kuffer, M. (2008) 'Analysing the spatial heterogeneity of poverty using remote sensing: typology of poverty areas using selected RS based indicators', EARSeL, Remote sensing: new challenges of high resolution, Bochum (05.-07. März).
- Socher, W. (1999) *Umweltatlas Dresden. Karte 1.4 Stadtstrukturtypen*, [Online], Abrufbar unter: [www.dresden.de](http://www.dresden.de) [10 Jul 2008].
- Soergel, U. (ed.) (2010) *Radar Remote Sensing of Urban Areas*, Heidelberg Dordrecht London New York: Springer, 296 S.
- Spieß, C.K. (2005) 'Das Sozio-oekonomische Panel (SOEP) und die Möglichkeiten regionalbezogener Analysen', in Grözinger, G. und Matiaske, W. (ed.) *Deutschland regional. Sozialwissenschaftliche Daten im Forschungsverbund.*, München, Mering: Rainer Hampp, S.57-64.
- Spieß, C.K. und Wrohlich, K. (2008) 'Does distance determine who attends a university in Germany?', *SOEPpapers on Multidisciplinary Panel Data Research*, Juli 2008 (118), pp. 1-32.
- Statistics Canada (2011) *Survey of Labour and Income Dynamics (SLID) – 2009 Survey Overview*, 15 Jun, [Online], Abrufbar unter: <http://www.statcan.gc.ca/pub/75f0011x/75f0011x2011001-eng.htm> [28 Jul 2011].
- Statistisches Bundesamt (2009) *Fortschreibung des Bevölkerungsstandes*, 31 Dec, [Online], Abrufbar unter: <https://www-genesis.destatis.de/genesis/online> [07 Aug 2011].
- Statistisches Bundesamt (2010) *Feststellung des Gebietsstandes*, 31 Dec, [Online], Abrufbar unter: <https://www-genesis.destatis.de/genesis/online> [07 Aug 2011].
- Statistisches Bundesamt (2012) *Verbraucherpreisindex für Deutschland - Lange Reihen ab 1948*, Jan, [Online], Abrufbar unter: <http://www.destatis.de/jetspeed/portal/cms/Sites/destatis/Internet/DE/Content/Publikationen/Fachveroeffentlichungen/Preise/Verbraucherpreise/VerbraucherpreisindexLangeReihen.psmi> [25 Feb 2012].
- Steinnocher, K., Petrini, F., Tötzer, T. und Weichselbaum, J. (2005) 'Räumliche Disaggregation von sozio-ökonomischen Daten', *Angewandte Geographische Informationsverarbeitung XVII- Beiträge zum AGIT-Symposium*, Salzburg (6.-8.Juli).
- Stocker, H. (2011) *Einführung in die angewandte Ökonometrie*, Unveröffentlichtes Skriptum der Universität Innsbruck.
- Sucoff, C.A. und Upchurch, D.M. (1998) 'Neighborhood context and the risk of childbearing among metropolitan-area black adolescents', *American Sociological Review*, 63 (4), p. 571–585.
- Sutton, P., Roberts, D., Elvidge, C. und Baugh, K. (2001) 'Census from Heaven: an estimate of the global human population using night-time satellite imagery', *International Journal of Remote Sensing*, 22 (16), pp. 3061-3076.
- Sutton, P., Roberts, D., Elvidge, C. und Melj, H. (1997) 'A Comparison of Nighttime Satellite Imagery and Population Density for the Continental United States', *Photogrammetric Engineering & Remote Sensing*, 63 (11), pp. 1303-1313.

- Tapiador, F.J., Avelar, S., Tavares-Corrêa, C. und Zah, R. (2011) 'Deriving fine-scale socioeconomic information of urban areas using very high-resolution satellite imagery', *International Journal of Remote Sensing*, 32 (21), pp. 6437-6456.
- Taubenböck, H. und Dech, S. (ed.) (2010) *Fernerkundung im urbanen Raum. Erdbeobachtung auf dem Weg zur Planungspraxis.*, Darmstadt: Wissenschaftliche Buchgesellschaft, 192 S.
- Taubenböck, H., Roth, A. und Dech, S. (2007) 'Linking structural urban characteristics derived from high resolution satellite data to population distribution', in Coors, Rumor, Fendel und Zlatanova (ed.) *Urban and Regional Data Management*, London, United Kingdom: Taylor & Francis Group.
- Taubenböck, H., Wurm, M., Setiadi, N., Gebert, N., Roth, A., Strunz, G., Birkmann, J. und Dech, S. (2009) 'Integrating Remote Sensing and Social Science - The correlation of urban morphology with socioeconomic parameters', In: Urban Remote Sensing Join Event, Shanghai, China (20.-22. Mai).
- Thomas, N., Hendrix, C. und Congalton, R.G. (2003) 'A comparison of urban mapping methods using high-resolution digital imagery', *Photogrammetric Engineering & Remote Sensing*, 69 (9), pp. 963-972.
- Tickamyer, A.R. (2000) 'Space Matters! Spatial Inequality in Future Sociology', *Contemporary Sociology*, 29 (6), pp. 805-813.
- Tortora, R. (1978) 'A note on sample size estimation for multinomial populations', *The American Statistician*, 32 (3), pp. 100-102.
- Tournachon, G.-F. (1864) *A terre et en l'air. Mémoires du Géant par Nadar*, Paris: E. Dentu, 439 S.
- Trimble (2010a) *eCognition (v. 8.64) [Computersoftware]*, München.
- Trimble (2010b) *eCognition Developer 8.64.0 Reference book*, München, 410 S.
- Trosset, A.-M., Bucher, T.U. und Lehmann, F. (2009) 'Adaption of building extraction rule sets derived from MFC3 and UltraCamD aerial image data sets', In: SPIE Europe Remote Sensing, Berlin (31.August - 03.September).
- Tuia, D., Pacifici, F., Kanevski, M. und Emery, W.J. (2009) 'Classification of Very High Spatial Resolution Imagery Using Mathematical Morphology and Support Vector Machines', *IEEE Transactions on Geoscience and Remote Sensing*, 47 (11), pp. 3866-3879.
- Turner, I.B.L. (1998) 'Frontiers of exploration: Remote sensing and social science research', Proceedings of Pecora 13, Symposium on human interactions with the environment., Bethesda, MD.: American Society for Photogrammetry and Remote Sensing.
- Van de Voorde, T., De Genst, W. und Canters, F. (2007) 'Improving pixel-based VHR land-cover classifications of urban areas with post-classification techniques', *Photogrammetric Engineering & Remote Sensing*, 73 (9), pp. 1017-1027.
- Vapnik, V.N. (2010) *The Nature of Statistical Learning Theory*, New York: Springer, 333 S.
- Venables, W.N. und Ripley, B.D. (2002) *Modern Applied Statistics with S*, New York: Springer.
- Voigtländer, S., Goebel, J., Claßen, T., Wurm, M., Berger, U., Strunk, A. und Elbern, H. (2011) 'Using geographically referenced data on environmental exposures for public health research: a feasibility study based on the German Socio-Economic Panel Study (SOEP)', *SOEPpapers on Multidisciplinary Panel Data Research*, Juni 2011 (386), p. 22.

- Vollmer, C., Michel, U. und Randler, C. (2012) 'Outdoor Lights at Night (LAN) is correlated with eveningness in adolescents', *Chronobiology International*, In press, pp. 1-7.
- Wagner, G.G., Frick, J.R. und Schupp, J. (2007) 'The German Socio-Economic Panel Study (SOEP) - Scope, Evolution and Enhancements', *Journal of Applied Social Science Studies - Schmollers Jahrbuch*, 127 (1), pp. 139-169.
- Wagner, G.G., Goebel, J., Krause, P., Pischner, R. und Sieber, I. (2008) 'Das Sozio-oekonomische Panel (SOEP): Multidisziplinäres Haushaltspanel und Kohortenstudie für Deutschland - Eine Einführung (für neue Datennutzer) mit einem Ausblick (für erfahrene Anwender)', *AStA - Wirtschafts- und Sozialstatistisches Archiv*, 2 (4), pp. 301-328.
- Weihls, C., Ligges, U., Luebke, K. und Raabe, N. (2005) 'klaR Analyzing German Business Cycles', in Baier, D., Decker, R. und Schmidt-Thieme, L. (ed.) *Data Analysis and Decision Support*, Berlin: Springer-Verlag, S. 335-343.
- Weng, Q. (ed.) (2006) *Urban Remote Sensing*, Boca Raton, FL: CRC Press, 432 S.
- Werlen, B. (2000) *Sozialgeographie: eine Einführung*, Bern, Stuttgart, Wien: Haupt, 400 S.
- Wickop, E., Böhm, P., Eitner, K. und Breuste, J. (1998) *Qualitätszielkonzept für Stadtstrukturtypen am Beispiel der Stadt Leipzig*, UFZ Bericht (14), Leipzig, 156 S.
- Witt, H. (2008) *Dynamic Capabilities im Strategischen Electronic Business-Management: Bestimmungsgrößen, Ausprägungen und Erfolgsfaktoren*, Wiesbaden: Gabler Verlag, 357 S.
- Wright, J.K. (1936) 'A method of mapping densities of population with Cape Cod as an example', *Geographical Review*, 26, pp. 103-110.
- Wurm, M., Goebel, J. und Wagner, G.G. (2010) 'Integration raumrelevanter Indikatoren in sozial- und verhaltenswissenschaftliche Analysen', in Taubenböck, H. und Dech, S. (ed.) *Fernerkundung im urbanen Raum.*, Darmstadt: Wissenschaftliche Buchgesellschaft, S.153-162.
- Wurm, M., Taubenböck, H., Goebel, J., Esch, T. und Wagner, G.G. (2011) 'Zensus der Zukunft? Erdbeobachtung zur räumlichen Abschätzung der Bevölkerungsverteilung', *Standort – Zeitschrift für angewandte Geographie*, 35 (4), pp. 169-175.
- Wurm, M., Taubenböck, H., Schardt, M., Esch, T. und Dech, S. (2011) 'Object-based image information fusion using multisensor earth observation data over urban areas', *International Journal of image and data fusion*, 2 (2), pp. 121-147.
- Yang, X. (ed.) (2011) *Urban Remote Sensing: Monitoring, Synthesis and Modeling in the Urban Environment*, Oxford Malden, Mass.: Wiley, 408 S.
- Yoon, J.-s. und Shan, J. (2002) 'Urban DEM generation from raw airborne LiDAR data', ASPRS Annual conference, Washington, DC. (19.-25. April).
- Zadeh, L.A. (1965) 'Fuzzy sets', *Information and Control*, 8 (3), pp. 338-353.
- Zhang, J. (2010) 'Multi-source remote sensing data fusion: status and trends', *International Journal of Image and Data Fusion*, 1 (1), p. 5-24.
- Zhang, Y., Maxwell, T., Tong, H. und Dey, V. (2010) 'Development of supervised software tool for automated determination of optimal segmentation parameters for ecognition', ISPRS TC VII symposium: 100 years ISPRS, Wien (05.-07. Juli).

Zhou, G., Song, C., Simmers, J. und Cheng, P. (2004) 'Urban 3D GIS from LiDAR and digital aerial images', *Computers & Geosciences*, 30 (4), pp. 345-353.

Auszüge dieser Dissertation wurden mit freundlicher Genehmigung des Verlages aus folgender Publikation gestattet:

Wurm, M., Taubenböck, H., Schardt, M., Esch, T. und Dech, S. (2011): Object-based image information fusion using multisensor earth observation data over urban areas. In: *International Journal of image and data fusion* 2 (2), 121-147. Taylor & Francis. Veröffentlicht online am: 23. Mai 2011

Die statistischen Auswertungen und Abbildungen in der vorliegenden Arbeit wurden mit dem freien Softwarepaket *R* (R Development Core Team, 2010) und den zusätzlichen Programmpaketen *klaR* (Weihs et al., 2005) und *LinDA* (Faes, 2010) durchgeführt.

Aus Gründen der leichteren Lesbarkeit wird in dieser Arbeit auf eine geschlechtsspezifische Differenzierung, wie z.B. Teilnehmer/Innen, verzichtet. Entsprechende Begriffe gelten im Sinne der Gleichbehandlung für beide Geschlechter.

