

Institut für Geographie der Universität Potsdam

Fachgruppe Geoinformatik

# **Untersuchungen zur räumlichen Analyse und Visualisierung von Mietpreisdaten für Immobilienportale**

Dissertation zur Erlangung des akademischen Grades  
"doctor rerum naturalium" (Dr. rer. nat.)  
in der Wissenschaftsdisziplin "Geoinformatik".

eingereicht an der  
Mathematisch-Naturwissenschaftlichen Fakultät  
der Universität Potsdam

von

Mag. rer. nat. Harald Schernthanner

Potsdam, den 12. Juni 2015

Online veröffentlicht auf dem  
Publikationsserver der Universität Potsdam:  
URN urn:nbn:de:kobv:517-opus4-89492  
<http://nbn-resolving.de/urn:nbn:de:kobv:517-opus4-89492>



# DANKSAGUNG

Mein ganz besonderer Dank gilt meinem Doktorvater Herrn Prof. Dr. Hartmut Asche, der mich in all den Jahren immer unterstützte. Danke Aura, ohne dich hätte ich diesen oft harten Weg nicht gehen können. Außerdem bedanke ich mich bei allen Kollegen und Freunden für eure Inspiration und euren Optimismus.

# INHALTSVERZEICHNIS

<b>Danksagung</b> .....	<b>ii</b>
<b>Inhaltsverzeichnis</b> .....	<b>iii</b>
<b>Abbildungsverzeichnis</b> .....	<b>vii</b>
<b>Formelverzeichnis</b> .....	<b>x</b>
<b>Tabellenverzeichnis</b> .....	<b>xi</b>
<b>Abkürzungsverzeichnis</b> .....	<b>xii</b>
<b>1 Einleitung</b> .....	<b>1</b>
1.1 Motivation .....	1
1.2 Aufgabenstellung .....	3
1.3 Hypothese .....	3
1.4 Arbeitspakete .....	4
1.4.1 Erstes Arbeitspaket.....	5
1.4.2 Zweites Arbeitspaket .....	6
1.4.3 Drittes Arbeitspaket .....	7
1.5 Gliederung und Aufbau der Arbeit.....	9
<b>2 Stand der Forschung, Entwicklung und Technik</b> .....	<b>11</b>
2.1 Immobilienportale.....	11
2.2 Literatur.....	12
2.2.1 Literatur aus dem Bereich der räumlichen Analyse .....	14
2.2.2 Literatur aus dem Bereich des maschinellen Lernens.....	16
2.2.3 Literatur aus dem Bereich der Geovisualisierung .....	16
2.3 Technologische Recherche .....	16
2.4 Status Quo der Methoden.....	23

2.4.1 Hedonische Modellverfahren .....	24
2.4.2 Räumliche Bezugseinheiten .....	25
2.4.3 Nicht dynamische Legenden .....	28
2.4.4 Punktsignaturen und das „Icon Cluttering“ Problem.....	30
2.4.5 Integration von Distanz .....	31
2.4.6 Integration von Räumlichen Indikatoren .....	32
2.4.7 Fazit .....	33
<b>3 Datenbasis.....</b>	<b>34</b>
3.1 Quantitative Datenbasis.....	34
3.2 Qualitative Datenbasis.....	36
3.2.1 Experteninterviews.....	36
3.2.2 Die geoinformatische und geovisuelle Prozesskette von Immobilienportalen .....	43
3.2.3 Räumliche Optimierungspotentiale .....	43
<b>4 Methoden.....</b>	<b>46</b>
4.1 Methodengruppe der räumlich-statistischen Methoden.....	46
4.1.1 Regionalisierte Variablen.....	47
4.1.2 Räumlichen Autokorrelation .....	47
4.1.3 Hauptkomponentenanalyse (Principal component analysis, PCA) .....	50
4.1.4 Geographisch gewichtete Regression .....	50
4.1.5 Räumliche Schätzverfahren, Interpolation .....	51
4.2 Methodengruppe der Verfahren des Maschinellen Lernens .....	63
4.2.1 Random Forest (RF) .....	64
4.2.2 Künstliche Neuronale Netze.....	65
4.3 Methodengruppe der geovisuellen Methoden.....	66
4.3.1 Grid-Mapping.....	66
4.3.2 Map APIs .....	67

4.3.3 Kartenbasis.....	68
4.3.4 Cloudbasierte Geovisualisierung .....	68
4.4    Fazit .....	69
<b>5    Anwendung der Methoden .....</b>	<b>71</b>
5.1    Anwendung räumlich-statistischer Methoden .....	72
5.1.1 Teilimplementierung der hedonischen Regression.....	72
5.1.2 Teilimplementierung der Auswahl von Sekundärvariablen .....	75
5.1.3 Teilimplementierung der Messung der räumlichen Autokorrelation .....	82
5.1.4 Teilimplementierung der geographisch-gewichteten Regression in Kombination mit Lagemodellierungen.....	85
5.1.5 Teilimplementierung der Modellevaluierung für die Anwendung für Immobilienportale.....	90
5.2    Anwendung der Methoden des maschinellen Lernens .....	94
5.2.1 Teilimplementierung von RF zur Mietpreisschätzung.....	95
5.2.2 Teilimplementierung von ANN zur Mietpreisschätzung.....	98
5.3    Teilimplementierung Geovisualisierungs-Methoden .....	99
5.3.1 Vorüberlegungen zu Map APIs.....	99
5.3.2 Vorüberlegung zu potentiellen Nutzergruppen von Mietpreiskarten ...	100
5.3.3 Zonenlose Preiskarte.....	100
5.3.4 Prototypenentwicklung .....	105
5.3.5 Analyse der Gemeinsamkeiten der Mietpreiskarten .....	105
<b>6    Fazit und Umsetzungsoptionen.....</b>	<b>111</b>
6.1    Fazit .....	111
6.1.1 Raumanalytische Teilkomponente des Fazit.....	111
6.1.2 Geovisuelle Teilkomponente des Fazit .....	112
6.2    Umsetzungsoptionen .....	114
6.2.1 Szenario Mietpreisschätzung I.....	114

6.2.2 Szenario Mietpreisschätzung II.....	114
6.2.3 Szenario Mietpreiskarte .....	115
<b>7 Zusammenfassung und Ausblick .....</b>	<b>116</b>
7.1 Zusammenfassung .....	116
7.2 Beitrag zum Forschungsstand und Ausblick.....	117
7.2.1 Ausblick .....	118
<b>Literaturverzeichnis.....</b>	<b>119</b>
<b>Anhang.....</b>	<b>129</b>
Anhang 1: Leitfaden Expertenbefragungen .....	129
Anhang 2: Auswertematrix technologische Recherche .....	131
Anhang 3: Variablen des Angebotsdatensätze .....	132
Anhang 4: Angebotsdichten und Durchschnittsnettokalmmieten pro Jahr in Potsdam.....	133
Anhang 5: Autokorrelation der Angebotsdaten - Potsdam Erstes Quartal 2013 .....	134
Anhang 6: Random Forest Score für 6 künstliche Angebotswerte aus dem ersten Quartal 2013.....	135
Anhang 7: Fehlerreport RF Modell mit 150 Bäumen .....	136
Anhang 8: Ergebnisse der PCA: Q1/2013.....	136
Anhang 9: Modellierte Einflussfaktoren auf die Nettokaltmiete (Potsdam 22, 2012) .....	136

# ABBILDUNGSVERZEICHNIS

ABB. 1-1: SCHEMATISCHE DARSTELLUNG DES ERSTEN ARBEITSPAKETES. (ENTWURF UND DARSTELLUNG: SCHERNTHANNER 2015).....	5
ABB. 1-2: SCHEMATISCHE DARSTELLUNG DES ERSTEN ARBEITSPAKETES. (ENTWURF UND DARSTELLUNG: SCHERNTHANNER 2015).....	6
ABB. 1-3: SCHEMATISCHE DARSTELLUNG DES ZWEITEN ARBEITSPAKETES. (ENTWURF UND DARSTELLUNG: SCHERNTHANNER 2015). .....	7
ABB. 1-4: SCHEMATISCHE DARSTELLUNG DES DRITTEN ARBEITSPAKETES. (ENTWURF UND DARSTELLUNG, SCHERNTHANNER 2015).....	9
ABB. 1-5: SCHEMATISCHE KAPITELÜBERSICHT DER ERSTEN BEIDEN GLIEDERUNGSEBENEN. (ENTWURF UND DARSTELLUNG: SCHERNTHANNER 2015). .....	10
ABB. 2-1: WORTWOLKE VON LÖSUNGSANSÄTZEN, BASIEREND AUF DER LITERATURRECHERCHE (ENTWURF UND DARSTELLUNG: SCHERNTHANNER 2015).....	14
ABB. 2-2: HÄUFIGKEIT DER ANWENDUNG VON METHODEN DER GEOINFORMATIK UND - VISUALISIERUNG IN 32 UNTERSUCHTEN IMMOBILIENPORTALEN NACH NATIONEN ZUSAMMENGEFASST. (ENTWURF UND DARSTELLUNG: SCHERNTHANNER 2015).....	19
ABB. 2-3: KARTEN-MASHUP VON HOUSINGMAPS.COM AUS DEM JAHR 2005 (HOUSINGMAPS, 2014).....	20
ABB. 2-4: DIAGRAMM DER IN IMMOBILIENPORTALEN VERWENDETEN BASEMAPS UNTERSCHIEDLICHER MAP API ANBIETER. (ENTWURF UND DARSTELLUNG: SCHERNTHANNER 2015).....	21
ABB. 2-5: IMMOBILIENPREISKARTE ZÜRICHS FÜR WOHNHEGENTUM IN SCHWEIZER FRANKEN DES PORTALS HOMEGATE.CH (2014).....	23
ABB. 2-6: LINKS: VERGLEICH VON DREI ZIP CODE BEREICHEN (ROT) MIT EINER ZILLOW- NEIGHBOORHOOD IN NEW YORK (BLAU GESTRICHELTE LINIEN). MABSTAB: 1: 40.000. (ENTWURF UND DARSTELLUNG: SCHERNTHANNER 2015).....	27
ABB. 2-7: GEOMETRIE EINES POSTLEITZAHLENBEZIRKS (ROT) IM VERGLEICH ZU EINEM WOHNQUARTIER DER FIRMA INFAS / NEXIGA. 1: 75.000. (ENTWURF UND DARSTELLUNG: SCHERNTHANNER 2015).....	28
ABB. 2-8: HISTOGRAMM DER NETTOKALTMIETEN DER STADT MÜNCHEN. (ENTWURF UND DARSTELLUNG: SCHERNTHANNER: 2015). .....	30
ABB. 2-9: AUSSCHNITTE AUS MIETPREISKARTEN DES IMMOBILIENPORTALS IS24 LINKS: MÜNCHEN. RECHTS: OSTERRODE IM HARZ. (IMMOBILIENSCOUT 24, 2015).....	30
ABB. 2-10: BURRIGAT UND CHITTAROS (2008) ILLUSTRATION DES ICON CLUTTERING PROBLEMS AM BEISPIEL VON POIS IN YAHOO MAPS.....	31
ABB. 2-11: ANGEBOTSFILTERUNG ÜBER PENDELZEITEN (TRULIA 2014).....	32
ABB. 2-12: KARTE DES KRIMINALITÄTSRISIKOS MIT GEMELDETEN GECLUSTERTEN TATEN (TRULIA, 2014). .....	32
ABB. 3-1: JÄHRLICHE DURCHSCHNITTLICHE ANGEBOTSDICHTE NACH STADTTEILEN IN POTSDAM (2007-2013). (ENTWURF UND DARSTELLUNG: SCHERNTHANNER 2015). .....	35
ABB. 3-2: AUFBAU UND FRAGENBLÖCKE DES EXPERTENINTERVIEWS. (ENTWURF UND DARSTELLUNG: SCHERNTHANNER 2015).....	38
ABB. 3-3: DARSTELLUNG DES STATUS QUO DER PROZESSKETTE VON IMMOBILIENPORTALEN. (ENTWURF UND DARSTELLUNG: SCHERNTHANNER 2015) .....	43
ABB. 4-1: ZUSTÄNDE DER RÄUMLICHEN AUTOKORRELATION NACH LO & YEUNG (2002).....	48
ABB. 4-2: BEISPIEL FÜR POSITIVE UND NEGATIVE RÄUMLICHE AUTOKORRELATION NACH ABELER ET AL. (1992). .....	48

ABB. 4-3: SCHEMA DER AUSWAHL, VALIDIERUNG UND TEILIMPLEMENTIERUNG VON SCHÄTZVERFAHREN. (ENTWURF UND DARSTELLUNG, SCHERNTHANNER: 2015).....	52
ABB. 4-4: EXAKTE, NICHT EXAKTE INTERPOLATION (NIEMEIER, 2008). .....	53
ABB. 4-5: BEISPIEL FÜR EIN SEMIVARIOGRAMM EINES EXPONENTIELLEN MODELLS NACH LI UND HEAP (2008). .....	56
ABB. 4-6: SPHÄRISCHE-, EXPONENTIELLE-, LINEARE- UND GAUSSISCHE MODELFUNKTION NACH LI UND HEAP (2008).....	57
ABB. 4-7: ABLAUF DES KRIGING INTERPOLATIONSVERFAHRENS. (ENTWURF UND DARSTELLUNG: SCHERNTHANNER 2015).....	58
ABB. 4-8: ENTSCHEIDUNGSBAUM VON BIVAND ET AL. (2008) ZUR AUSWAHL EINES GEEIGNETEN SCHÄTZVERFAHRENS. ....	63
ABB. 5-1: ANWENDUNGSFALLDIAGRAMM DES ANWENDUNGSFALLES DER RÄUMLICHEN ANALYSE (ENTWURF UND DARSTELLUNG: SCHERNTHANNER 2015).....	72
ABB. 5-2: ANWENDUNGSFALLDIAGRAMM DES ANWENDUNGSFALLES DER GEOVISUALISIERUNG (ENTWURF UND DARSTELLUNG: SCHERNTHANNER 2015).....	72
ABB. 5-3: SCHEMATISCHER ABLAUF DER HEDONISCHEN REGRESSION. (ENTWURF UND DARSTELLUNG: SCHERNTHANNER 2015).....	74
ABB. 5-4: ERGEBNISDARSTELLUNG DER HEDREG AUF STADTTEILEBENE. (ENTWURF UND DARSTELLUNG: SCHERNTHANNER 2015).....	75
ABB. 5-5: METHODIK DER AUSWAHL DER SEKUNDÄRVARIABLEN. (ENTWURF UND DARSTELLUNG: SCHERNTHANNER 2015).....	76
ABB. 5-6: DIREKTE EINFLUSSFAKTOREN AUF DIE ABSOLUTE NETTOKALTMIETE LAUT EXPERTENGREMIUM POTSDAM 22. DIE LINIEN ZEIGEN DIE EINZELNEN EINFLUSSFAKTOREN AUF DIE NETTOKALTMIETE UND IN WEITERER FOLGE AUF DIE WOHNKOSTEN. ÜBER DAS PLUS-SYMBOL KÖNNEN FAKTOREN NIEDRIGERER HIERACHIEBENENEN VISUALISIERT WERDEN. (POTSDAM 22, 2013) .....	77
ABB. 5-7: SCREE-PLOT DER VARIANZ IN DEN VERSCHIEDENEN HAUPTKOMPONENTEN. (ENTWURF UND DARSTELLUNG: SCHERNTHANNER 2015). ....	79
ABB. 5-8: PLOT DER ERSTEN BEIDEN HAUPTKOMPONENTEN. DIE X-ACHSE STELLT DIE FAKTORENLOADUNG DER ERSTEN HAUPTKOMPONENTE, DIE Y-ACHSE DER ZWEITEN HAUPTKOMPONENTE DAR (ENTWURF UND DARSTELLUNG: SCHERNTHANNER 2015).....	80
ABB. 5-9: MORANS I NACH DISTANZKLASSEN. (ENTWURF UND DARSTELLUNG: SCHERNTHANNER 2015).....	83
ABB. 5-10: NACH GETIS ORD G ERMITTELTE HOT- UND COLDSPOTS IN POTSDAM. DARGESTELLT WIRD DER GERASTERTE Z WERT VON GETIS ORD G. (ENTWURF UND DARSTELLUNG: SCHERNTHANNER 2015).....	84
ABB. 5-11: NETZWERKDISTANZ VON MIETANGEBOTEN ZU BUSHALTESTELLEN UND RESIDUEN DER PREISMODELLIERUNG MITTELS GWR. (ENTWURF UND DARSTELLUNG: SCHERNTHANNER 2015). .....	87
ABB. 5-12: GÜTE (RMSE) UND RECHENDAUER DER SCHÄTZVERFAHREN BEI EINER 5-FOLD CROSS VALIDATION (SCHEELE UND SCHERNTHANNER, 2014). ....	90
ABB. 5-13: SCHEMATISCHER ABLAUF DER EVALUIERUNG DER SCHÄTZVERFAHREN UND DER ÜBERTRAGUNG DER ERGEBNISSE AUF DIE RÄUMLICHE INTERPOLATION. (ENTWURF UND DARSTELLUNG: SCHERNTHANNER 2015). ....	91
ABB. 5-14: VERGLEICH INTERPOLIERTER ANGEBOTSMIETPREISE (50M GRID) MIT EINER HEDONISCHEN REGRESSION IN BEZUG AUF DIE FLÄCHE EINER 90 HEKTAR GROßEN INFAS-NACHBARSCHAFT. HINTER DEM GEWÄHLTEN KRIGING SCHÄTZER FINDET SICH DIE ANGABE DER GEWÄHLTEN MODELLFUNKTION SOWIE DIE NUGGET-/ SILL-/ RANGE-WERTE. (ENTWURF UND DARSTELLUNG: SCHERNTHANNER 2015).....	94
ABB. 5-15: SCHEMATISCHER ABLAUF RF. (ENTWURF UND DARSTELLUNG: SCHERNTHANNER 2015). 95	

ABB. 5-16: KURVE DER FEHLERRATE VON RF-MODELLEN BEI ZUNEHMENDER BAUMANZAHL. (ENTWURF UND DARSTELLUNG: SCHERNTHANNER 2015).	96
ABB. 5-17: DIFFERENZ REALER ZU RF-GESCHÄTZTEN ANGEBOTSPREISEN AN VISUALISIERTEN ANGEBOTSTANDORTEN FÜR EINEN RAUMAUSSCHNITT IN BABELSBERG / POTSDAM. DIE PUNKTE ZEIGEN DIE JEWEILIGEN WOHNUNGSSTANDORTE. KARTENBASIS: STAMEN OPENSTREETMAP STIL. (ENTWURF UND DARSTELLUNG: SCHERNTHANNER 2015)	97
ABB. 5-18: DIFFERENZ REALER ZU ANN-MODELLIERTER MIETPREISE. (ENTWURF UND DARSTELLUNG: SCHERNTHANNER 2015)	99
ABB. 5-19: INTERPOLIERTER 50M-RASTER (ENTWURF UND DARSTELLUNG: SCHERNTHANNER 2015).	104
ABB. 5-20: MIETPREISKARTE FÜR POTSDAM DES PORTALS IS24 FÜR DAS DRITTE QUARTAL 2013 (IMMOBILIENSCOUT 24, 2014)	104
ABB. 5-21: VISUALISIERUNG VON MIETPREISEN MIT CARTODB. (ENTWURF UND DARSTELLUNG: SCHERNTHANNER 2015)	106
ABB. 5-22: MIETPREISKARTEN-PROTOTYP. ALS ZUSÄTZLICHE DATENSCHICHTEN WURDEN SCHULEINRICHTUNGEN (WEIßER MARKER) UND ZONEN MIT HOHER LÄRMBELASTUNG (GRÜN) EINGEBLENDET. (DARSTELLUNG: CHRISTOPHER BRAUNE, 2014)	107
ABB. 5-23: NETTOKALTMIETE IN €/M <sup>2</sup> ALS THEMATISCHER Z WERT DARGESTELLT MIT DEM OSM BUILDINGS PLUGIN VON LEAFLET. (ENTWURF UND DARSTELLUNG: SCHERNTHANNER 2015)..	108
ABB. 5-24: ZOOMSTUFE 17 EINER PROTOTYPISCHEN MIETPREISKARTE (ENTWURF UND DARSTELLUNG: SCHERNTHANNER 2015)	109
ABB. 5-25: ZOOMSTUFE 18: SPRUNG DER DARSTELLUNG VON EINER TRENDOBERFLÄCHE ZUR GEBÄUDEBLOCK-DARSTELLUNG BEI EINEM WECHSEL DER ZOOMSTUFE 17 AUF 18 (ENTWURF UND DARSTELLUNG: SCHERNTHANNER 2015).	109
ABB. 5-26: PROTOTYPISCHE ANGEBOTSDARSTELLUNG IN HEXABINS. GRÖßERE WABEN BEDEUTEN DABEI GRÖßERE ANGEBOTSCLUSTER. (ENTWURF UND DARSTELLUNG: SCHERNTHANNER 2015).	110
ABB. 6-1: SZENARIO 1: BESTEHEND AUS DEM VERFAHREN OK MIT OPTIONALER KOPPLUNG MIT GWR UND RF. (ORANGE OPTIONALE PROZESSKOMPONENTEN, GRÜN: OBLIGATORISCHE PROZESSKOMPONENTEN). (ENTWURF UND DARSTELLUNG: SCHERNTHANNER 2015)	114
ABB. 6-2: SZENARIO 2: BESTEHEND AUS DEM VERFAHREN KED MIT OPTIONALER KOPPELUNG WEITERER VERFAHREN. ORANGE OPTIONALE PROZESSKOMPONENTEN, GRÜN: OBLIGATORISCHE PROZESSKOMPONENTEN). (ENTWURF UND DARSTELLUNG: SCHERNTHANNER 2015)	115
ABB. 6-3: SZENARIO 3: ERSTELLUNG EINER MIETPREISKARTE BESTEHEND AUS DEN OBLIGATORISCHEN PROZESSKOMPONENTEN DER VERFAHREN DES GRIDMAPPING UND DER ERSTELLUNG VON MIETPREISGRIDS UND DER OPTIONALEN PROZESSKOMPONENTE DER VERSCHNEIDUNG VON MIETPREISGRIDS MIT ZUSÄTZLICHEN DATENSCHICHTEN. (ENTWURF UND DARSTELLUNG: SCHERNTHANNER 2015)	115

# FORMELVERZEICHNIS

FORMEL 2-1: SEMILOG VARIANTE DER HEDONISCHEN REGRESSION.....	25
FORMEL 4-1: MORANS I TEST (MORAN, 1947) .....	49
FORMEL 4-2: GEARYS C TEST (GEARY, 1954) .....	49
FORMEL 4-3: KORRELATIONSKOEFFIZIENT NACH GETIS ORD $G$ .....	50
FORMEL 4-4: GEOGRAPHISCH GEWICHTETE REGRESSION.....	51
FORMEL 4-5: GAUSS'SCHE DISTANZ-ABNAHME-FUNKTION .....	51
FORMEL 4-6: SEMIVARIOGRAMM.....	55
FORMEL 4-7: BERECHNUNG DER SEMIVARIANZ NACH <i>LI UND HEAP (2008)</i> .....	56
FORMEL 4-8: DER KRIGING-SCHÄTZER .....	59
FORMEL 4-9: KED-SCHÄTZER .....	59
FORMEL 4-10: BERECHNUNG DER KREUZ-SEMIVARIANZ NACH <i>BORROUGH UND MCDONNELL</i> (1996). .....	60
FORMEL 4-11: OCK .....	60

# TABELLENVERZEICHNIS

TABELLE 2-1: ÜBERSICHT ZU STUDIEN ÜBER PREISSCHÄTZVERFAHREN IN DER IMMOBILIENMARKTFORSCHUNG. ....	15
TABELLE 2-2: KATEGORIEN TECHNOLOGISCHE RECHERCHE. DER PUNKT „IMMOBILIENPREISKARTE“ FÄLLT IN BEIDE KATEGORIEN. BEIM PUNKT BASEMAP/GRUNDKARTE ERFOLGTE EINE AUSWERTUNG NACH DEM GRUNDKARTEN ANBIETER. ....	18
TABELLE 3-1: BEISPIELE FÜR VARIABLEN DES DATENSATZES. ....	35
TABELLE 3-2: IDENTIFIZIERTE THEMENFELDER MIT UNTERGEORDNETEN PROBLEMFELDERN. (ORANGE: GEOINFORMATION, GRÜN: GEOVISUALISIERUNG). ....	39
TABELLE 4-1: VOR- UND NACHTEILE DETERMINISTISCHER- UND GEOSTATISTISCHER INTERPOLATIONSVERFAHREN. ....	62
TABELLE 5-1: STANDARDWOHNUNG IM STADTTEIL BABELSBERG NORD IM ERSTEN QUARTAL 2013. ..	73
TABELLE 5-2: ABGLEICH DER WOHNKOSTENFAKTOREN VON POTSDAM 22 MIT DEM EXEMPLARISCHEN ANGEBOTSDATENSATZ. ....	78
TABELLE 5-3: SCHNITTMENGE ZWISCHEN DEN EINZELNEN VERFAHREN. GRÜN HINTERLEGTE VARIABLEN WURDEN IN ALLEN DREI STUFEN DES VERFAHRENS ALS SEKUNDÄRVARIABLE IDENTIFIZIERT, GELB HINTERLEGTE WURDEN IN RF UND DEM EXPERTENGREMIUM IDENTIFIZIERT. ....	82
TABELLE 5-4: MORAN´S I NACH DISTANZKLASSEN .....	83
TABELLE 5-5: R <sup>2</sup> BEI 5 MODELLÄUFEN. ....	87
TABELLE 5-6: ERGEBNISSE DER MODELLEVALUIERUNG. ....	93
TABELLE 5-7: AUFSTIEGENDER GEMITTELTER QUADRATISCHER FEHLER DER RF-MODELLE BEI ENTSPRECHENDER BAUMANZAHL. ....	97
TABELLE 5-8: MASCHENWEITEN UND MÖGLICHE MABSTÄBE UND ZOOMSTUFEN. ....	102

# ABKÜRZUNGSVERZEICHNIS

Ajax	Asynchrones JavaScript
BSD	Berkeley Software Distribution
CART	Classification and Regression Tree
IS24	Immobilienscout 24
GWR	Geographisch gewichtete Regression
GfK	Gesellschaft für Konsumforschung
GIS -	Geoinformationssystem
FOS GIS	Freies quelloffenes Geoinformationssystem
HedReg	Hedonische Regression
HTML	Hypertext Markup Language
JSON	JavaScript Object Notation
InMSE %	logarithmische MSE in %
Map APIs	Map Application Programming Interfaces)
MSE	Mean squared error
OGC	Open geospatial consortium
Ock	Ordinary Cokriging
OK	Ordinary Kriging
OLS	Ordinary Least Square
ÖPNV	Öffentlicher Personennahverkehr
PCA	Principal component analysis (Hauptkomponentenanalyse)
RF	Random Forest
REST	RESTful Webservices
SAAS	Software as a Service
SVM	Support Vector Machines
TMS	Tile Map Service
WFS	Web Feature Service
WMS	Web Mapping Service
WPS	Web Processing Service

XML

Extensive Markup Language



# 1 EINLEITUNG

## 1.1 Motivation

Mit dem Start der ersten Immobilienportale<sup>1</sup> vor ca. 15 Jahren entstanden erstmals Datensätze, die bundesweite Immobilienanalysen ermöglichen. Davor waren großräumige Analysen nicht oder nur sehr eingeschränkt möglich. Immobilienmarktanalysen beruhten auf der lokalen Marktkenntnis und den eingeschränkten Datensätzen von Maklern (Enderle, 2009). Der Immobilienmarkt war intransparent. Analysen, geschweige denn räumliche Analysen oder Kartendarstellungen, wurden nur in geringem Maß durchgeführt.

Immobilienportale sammeln seit mehr als einer Dekade geokodierte Informationen zu Immobilien. Unterschiedliche Akteure der Immobilienwirtschaft wie Makler, Genossenschaften, private Anbieter, kommunale Immobilienunternehmen etc. veröffentlichen auf Immobilienportalen ihre lokalen Angebote, um diese auf den meist überregionalen Online-Marktplätzen von Immobilienportalen anzubieten. Die Zeitung als primäres Suchmedium für die Immobiliensuche wurde weitgehend von Immobilienportalen abgelöst. Erstmals liegen nun deutschlandweit Angebotsdaten von Immobilien in statistisch relevantem Umfang vor. Gemeinsam ist den seit Jahren eingespeisten Daten eine räumliche Komponente in Form einer sekundären Metrik (Adresse und Hausnummer). Die vorhandene Geokodierung nutzen Immobilienportale bisher nur bei der Verortung ihrer Angebote, obwohl die Datenfülle enorm ist. Die Zahlenangaben der großen Portale über ihr monatliches Angebotsaufkommen ähneln sich in Deutschland. Das deutsche Immobilienportal *Immonet* (Immonet, 2014) gibt z. B. auf seiner Startseite an, dass allein hierüber 1,49 Millionen Wohnimmobilien im Miet- und Kaufbereich in Deutschland angeboten werden. Weitere große Portale wie *Immowelt* geben 1,2 Millionen Angebote an (Immowelt, 2014), 1,5 Millionen Immobilien bietet *ImmobilienScout 24 (IS24)* an (Immobilienportale, 2012). Wird davon ausgegangen, dass die durchschnittliche Verweildauer eines Mietobjektes in einem Immobilienportal ein Quartal beträgt, dann sind bei

---

<sup>1</sup> Immobilienportale sind Online-Marktplätze zur Vermittlung von Immobilien.

den populärsten Portalen jährlich jeweils 5,56 Millionen unterschiedliche Angebote im Bestand. Wird diese Anzahl den laut des Zensus 2011 (*Statistisches Bundesamt, 2011*) deutschlandweit insgesamt verfügbaren rund 40 Millionen Wohnungen<sup>2</sup> gegenübergestellt, dann ergibt sich eine enorme Datenmenge an geokodierten Angebotsinformationen. Vor diesem Hintergrund stellt sich die Frage, ob und wie diese Informationen von Immobilienportalen für räumliche Analysen und Geovisualisierungen genutzt werden.

Der bekannten Maklerweisheit, dass genau drei Dinge bei einer Immobilie wichtig sind: 1. die Lage, 2. die Lage und 3. die Lage (*Stroisch, 2010*), wird von Portalbetreibern nur zu einem kleinen Teil Beachtung geschenkt. Ganz allgemein ergibt sich der Preis einer Immobilie beim Zusammentreffen von Angebot und Nachfrage aus speziellen marktrelevanten Eigenschaften wie Lage, Kaltmiete, Zimmeranzahl, Wohnfläche etc. (*Mussel, 2001*). Immobilienportale nehmen zwar den Raumbezug der vorhandenen Immobilien über die Geokodierung von Adressdaten auf, diese Lageinformationen werden aber überwiegend nur unzulänglich genutzt. Bei der Analyse der Angebotsdaten spielen Methoden und Techniken aus der Geoinformatik, der räumlichen Statistik und der Geovisualisierung bisher kaum eine Rolle. Teils werden die Angebotsdaten mittels Methoden der deskriptiven Statistik analysiert. Die räumliche Autokorrelation der Angebotsdaten wird von den Portalen vernachlässigt, obwohl gerade die räumliche Nähe ein ähnliches Marktgefüge bedingt, das schlussendlich nur durch variierende Ausstattungsmerkmale der Immobilien differenziert wird (*Mussel, 2001*).

Neben diesem größtenteils nicht genutzten Potential der räumlichen Auswertung von Angebotsdaten ist die Visualisierung von Immobilienangeboten und -preisen kartographisch als suboptimal einzustufen. Im Jahr 2005 stellte Google, kurz nach der Veröffentlichung von Google Maps, die erste Programmierschnittstelle (Map API) zu einem Online-Kartendienst bereit (*Peterson, 2014*). Mit dem Aufkommen der ersten Programmierschnittstellen (Map APIs) und entsprechenden Online-Kartendiensten nutzen auch Immobilienportalbetreiber vermehrt Kartendarstellungen in Form von Angebotsdarstellungen und Preiskarten.

---

<sup>2</sup> Laut dem letzten Zensus 2011 gab es in Deutschland mit dem Stichtag 9. Mai 2011 insgesamt 40 577 963 Wohnungen in 18 932 411 Gebäuden (*Statistisches Bundesamt, 2011*).

## 1.2 Aufgabenstellung

Diese Arbeit stellt sich der wissenschaftlichen Aufgabe, das bisher weitestgehend unerschlossene raumanalytische und räumlich-visuelle Potential von Immobilienportalen besser auszuschöpfen. Dafür werden auf konzeptioneller und (teil-)implementierter Ebene geoinformatische Lösungsansätze erarbeitet. Die Perspektive der Arbeit ist dabei eine geoinformatische. Es existiert in der Geoinformatik bereits ein etablierter Methodenapparat, mit einer stetig steigenden Menge von Methoden und Lösungsansätzen. Vor diesem Hintergrund besteht eine besondere Herausforderung dieser Arbeit darin, eine geeignete Methodenauswahl und -anwendung zum Ausgleich der raumanalytischen und geovisuellen Defizite von Immobilienportalen zu erreichen.

## 1.3 Hypothese

Gegenstand dieser Arbeit ist die Optimierung räumlicher Analysemethoden und Methoden der Visualisierung bei Immobilienportalen. Der übergreifende Forschungsansatz der Arbeit besteht einerseits darin, durch einen Transfer raumanalytischer und geovisualer Methoden aus dem Forschungsfeld der Geoinformatik in die Anwendungsdomäne von Immobilienportalen einen Zugewinn für Nutzer von Immobilienportalen zu schaffen. Andererseits sollen ein Rückfluss an Erkenntnissen in die Forschungsdomäne und eine Verbesserung der Transparenz von Immobilienmärkten erreicht werden.

Die vorliegende Arbeit geht von zwei Thesen aus.

1. Die erste These lautet: Verfahren der räumlichen Statistik und des Maschinellen Lernens zur Mietpreisschätzung sind den bisher eingesetzten Verfahren der hedonischen Regression überlegen und eignen sich zur räumlichen Optimierung von Immobilienportalen.
2. Die zweite These lautet: Die von Immobilienportalen publizierten webbasierten Mietpreiskarten geben nicht die tatsächlichen räumlichen Verhältnisse auf Immobilienmärkten wieder. Alternative webbasierte Darstellungsformen, wie z.B. Gridmaps, sind dem Status quo der Mietpreiskarten von Immobilienportalen überlegen und visualisieren die tatsächlichen räumlichen Verhältnisse von Immobilienpreisen effektiver.

Diese Thesen gilt es im Rahmen der vorliegenden Arbeit zu überprüfen. Die Thesen beruhen auf der Grundannahme, dass Immobilienportale signifikante raumanalytische und räumlich-visuelle Defizite aufweisen. Hieraus konstatiert der Autor einen erheblichen Forschungsbedarf. Der größte Forschungsbedarf wird in der Immobilienpreisanalyse und -darstellung gesehen. Das Finden von

Lösungsansätzen für diese Problematiken steht im Vordergrund der vorliegenden Arbeit.

Ausgehend von der formulierten Hypothese ergeben sich zwei zu untersuchende Forschungsfragen:

1. Wie ist der Status quo der Anwendung raumanalytischer und geovisualer Methoden und Techniken bei Immobilienportalen?
2. Welche Methoden aus dem geoinformatischen Methodenspektrum können so adaptiert werden, dass sie den realweltlichen Anforderungen von Immobilienportalen und deren Nutzergruppen gerecht werden? Wie können diese Methoden für eine Nutzung für Immobilienportale adaptiert werden?

Übergeordnetes Ziel (Globalziel) der Arbeit ist die Schaffung einer konzeptionellen Grundlage und ihrer exemplarischen Teilimplementierung zum Transfer von Forschungsergebnissen in das operationelle Geschäft von Immobilienportalen. Es gilt, eine Grundlage zur räumlichen Optimierung von Immobilienportalen zu schaffen. Hierzu ist festzuhalten, dass diese Arbeit nicht nur Potentiale zur Optimierung von Immobilienportalen findet und durch die Teilimplementierung realisiert, sondern dass das Forschungsproblem im Rahmen der Arbeit erstmals grundlegend bearbeitet wird.

### *Operationalisierung zur Zielerreichung*

Zur Zielerreichung wird das übergeordnete Ziel in ausführbare Schritte zerlegt. Hierfür gilt es, insbesondere die Zielgruppe der Arbeit zu definieren. Die Ergebnisse der Arbeit sind an Portalbetreiber adressiert und sollten diesen neue Ansätze und Möglichkeiten zur Optimierung der räumlichen Visualisierung von Wohnungsmietpreisen bieten. Eine weitere Zielgruppe sind Personen, die sich mit der Analyse von Mietpreisdaten beschäftigen, diese bisher nicht räumlich analysieren und nur unzureichend visualisieren. Das sind vor allem in der Immobilienwirtschaft tätige Betriebswirte und Akteure in Kommunen, die beispielsweise für die Erstellung von Mietspiegeln zuständig sind.

## 1.4 Arbeitspakete

Zur Überprüfung der zuvor postulierten Thesen folgt die Arbeit einem mehrstufigen Forschungsdesign. Das Forschungsdesign (siehe Abbildung 1-1) besteht im Kern aus vom Autor durchgeführten qualitativen und quantitativen Analysen.



Abb. 1-1: Schematische Darstellung des ersten Arbeitspaketes. (Entwurf und Darstellung: Schernthanner 2015).

Aus den formulierten Hypothesen, den Forschungsfragen und dem Globalziel der Arbeit ergeben sich drei Arbeitspakete.

### 1.4.1 Erstes Arbeitspaket

Das erste Arbeitspaket umfasst die Erhebung des Status quo raumanalytischer und geovisualer Methoden und Techniken bei Immobilienportalen. Themen- und Problemfelder des Status quo werden umfassend recherchiert. Der erhobene Status quo dient dazu, räumliche Optimierungspotentiale zu identifizieren. Hierfür gilt es, sich auch der Frage zu stellen, was Optimierung im Kontext der in der Arbeit behandelten Thematiken bedeutet.

Ein weiteres Arbeitspaket ist das Finden von Optimierungspotentialen, die den größten Forschungsbedarfen unterliegen. Vorausgreifend sei gesagt, dass die Optimierungspotentiale mit den größten Forschungsbedarfen in den Bereichen räumliche Analytik und geovisuelle Darstellung von Angebotsmieten liegen. Mittels zweier Anwendungsfälle, eines Anwendungsfalls der räumlichen Analyse und eines Anwendungsfalls der Geovisualisierung, werden Lösungsansätze zur räumlichen Optimierung von Immobilienportalen auf konzeptioneller Ebene diskutiert und durch Teilimplementierung realisiert.

#### *Bearbeitungsschritte des ersten Arbeitspaketes*

Zur Bearbeitung des Arbeitspaketes erfolgt eine erstmals durchgeführte Bestandserhebung der in Immobilienportalen zum Einsatz kommenden geoinformatischen und geovisuellen Methoden und Techniken. Hierfür wird ein Kriterienkatalog entworfen, nach dessen geoinformatischen und geovisuellen Gesichtspunkten insgesamt 32 Immobilienportale analysiert werden. Es wurden vom Autor außerdem fünf Experteninterviews mit Experten eines Immobilienportals und Immobilienmarktexperten durchgeführt und ausgewertet. Abbildung 1-2 zeigt eine schematische Darstellung des ersten Arbeitspaketes, dessen Teilpakete sowie zugehöriger Teilschritte zur Zielerreichung.

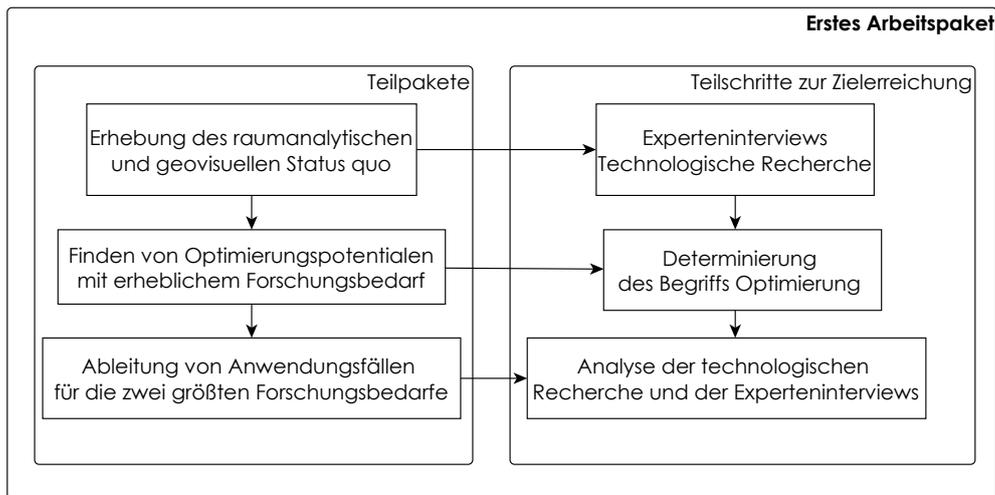


Abb. 1-2: Schematische Darstellung des ersten Arbeitspaketes. (Entwurf und Darstellung: Schernthanner 2015).

#### 1.4.2 Zweites Arbeitspaket

Das zweite Arbeitspaket sieht die Erarbeitung von Lösungsansätzen zur Optimierung der ermittelten raumanalytischen Defizite vor. Um dieses Ziel zu erreichen, werden Alternativen zu den von Immobilienportalen bisher genutzten „nicht räumlichen“ Analyseverfahren zur Preismodellierung untersucht.

Das zweite Arbeitspaket gliedert sich wiederum in verschiedene Teilpakete:

1. Die Erhebung von räumlichen Alternativen zum Preisschätzverfahren der hedonischen Regression als räumlich-statistische Grundlage für die Erstellung von Mietpreiskarten. Im Fokus stehen dabei Alternativen in Form von Verfahren des Maschinellen Lernens, der geographisch gewichteten Regression (GWR) und vor allem räumliche Schätzverfahren aus der Methodenfamilie des Kriging.
2. Das Treffen konzeptioneller Vorüberlegungen zur Autokorrelation von Mietangeboten und die Teilimplementierung der Messung der räumlichen Autokorrelation von Angebotsdaten.
3. Das Treffen konzeptioneller Vorüberlegungen zur Auswahl eines geeigneten Sekundärvariablensets für alternative Preisschätzverfahren und die Teilimplementierung von Verfahren zum Finden von Sekundärvariablen.
4. Die konzeptionellen Vorüberlegungen und die Teilimplementierung von Lösungsansätzen der GWR in Verbindung mit einer Lagemodellierung.
5. Die konzeptionellen Vorüberlegungen und die Teilimplementierung von Verfahren des Maschinellen Lernens.
6. Die konzeptionellen Vorüberlegungen zur Auswahl räumlicher Schätzverfahren und die vergleichende Teilimplementierung von räumlichen Schätzverfahren im Rahmen einer programmiertechnisch umgesetzten Validierung.

7. Die Durchführung eines vergleichenden Tests von Sensitivitätsparametern zur Mietpreisschätzung, mit dem Ziel, validestmögliche räumliche Schätzergebnisse für Immobilienportale zu erreichen.

### Bearbeitungsschritte des zweiten Arbeitspaketes

Im Rahmen des zweiten Arbeitspaketes wurde eine Literaturrecherche zur Erhebung des Status quo und zur Ermittlung räumlicher Alternativen zu nicht-räumlichen Preisschätzverfahren durchgeführt. Es erfolgte des Weiteren eine Teilnahme des Autors an einem Expertengremium zu wohnungspolitischen Vorschlägen der Stadt Potsdam. Diese Teilnahme diente der Gewinnung qualitativer Erkenntnisse, die in die konzeptionellen Vorüberlegungen und darauf aufbauende Teilimplementierungen einer räumlichen Mietpreisanalyse einfließen. Außerdem wurde vom Autor ein Validierungsrahmen für räumliche Schätzverfahren konzipiert und programmiertechnisch umgesetzt. Abbildung 1-3 zeigt eine schematische Darstellung des zweiten Arbeitspaketes, dessen Teilpakete sowie zugehöriger Teilschritte zur Zielerreichung.

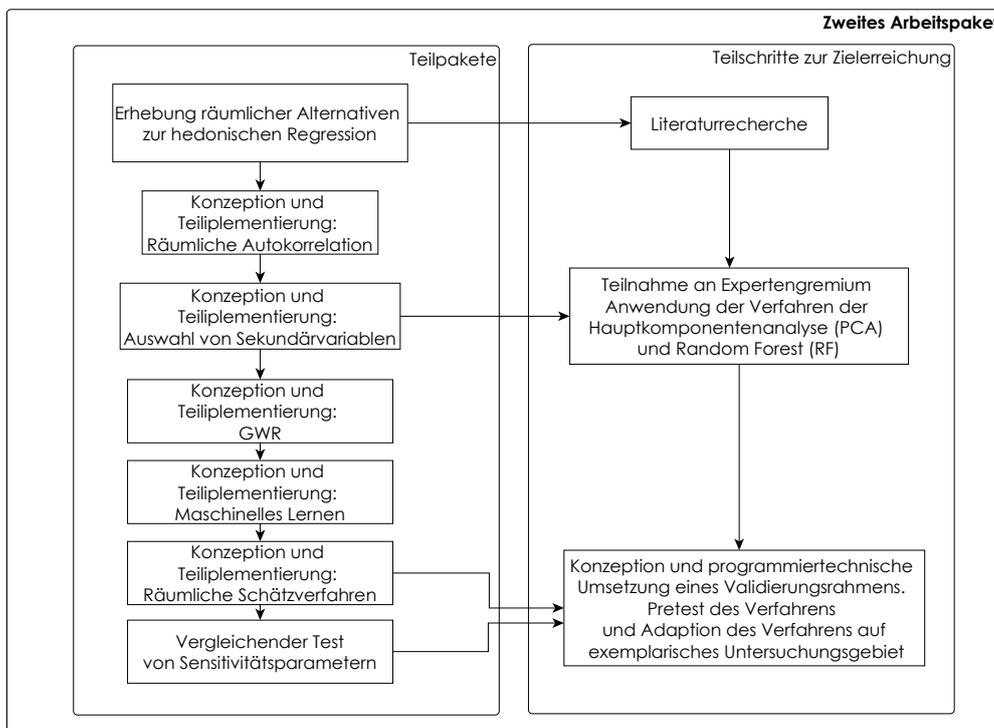


Abb. 1-3: Schematische Darstellung des zweiten Arbeitspaketes. (Entwurf und Darstellung: Scherthanner 2015).

### 1.4.3 Drittes Arbeitspaket

Im dritten Arbeitspaket werden Lösungsansätze der räumlichen Visualisierung von Mietpreisen diskutiert und teilimplementiert. Das dritte Arbeitspaket geht vom vorherrschenden Status quo aus, demnach Mietpreise in Immobilienpreiskarten auf fachfremd verwendete räumliche Einheiten bezogen werden. Teil des Arbeitspaketes ist das Erstellen eines Sets von alternativen Preisdarstel-

lungsmethoden. Die Ergebnisse des zweiten Arbeitspaketes werden in das dritte Arbeitspaket transferiert.

Ziel des Arbeitspaketes ist das Erstellen von Mietreiskarten, die von nicht sachgerecht verwendeten Raumeinheiten losgelöst sind. Zweck der Mietpreiskarten ist das Darstellen der tatsächlichen räumlichen Verhältnisse von Mietpreisen. Nach einer Diskussion der konzeptionellen Anforderungen an eine Mietpreiskarte werden prototypische Mietpreiskarten erstellt und hieraus Empfehlungen für Immobilienportale abgeleitet. Hierzu werden ausschließlich webbasierte, mittels Map APIs umsetzbare Ansätze diskutiert und (teil-)implementiert. Es wird nach Methoden recherchiert, mögliche Methoden werden diskutiert und eine Auswahl an Methoden wird webbasiert umgesetzt.

Das dritte Arbeitspaket umfasst mehrere Teilpakete:

1. Das Finden von Alternativen zum Status quo des Bezugs von Modellierungsergebnissen auf nicht sachgerecht verwendete Raumeinheiten. Zur Zielerreichung bedient sich dieses Arbeitspaket u.a. der Methoden rasterbasierter Ansätze des „Grid Mapping“.
2. Die Ermittlung idealer maßstabsabhängiger Maschenweiten im Kontext der Interpolation zur Darstellung von Mietpreisen.
3. Die Teilimplementierung von prototypischen Mietpreiskarten z.B. als Heatmaps, mit Preisbezug auf Gebäudegeometrien oder skalenabhängige Rasterdarstellungen.

### *Bearbeitungsschritte des dritten Arbeitspaketes*

Im Rahmen des dritten Arbeitspaketes wurden räumliche Darstellungsmethoden für Mietpreiskarten durch eine Literaturrecherche des Autors erhoben. Zur Bearbeitung der Forschungsfragen des dritten Teilpaketes wurden Verfahren des Gridmappings in konzeptioneller und programmiertechnisch teilimplementierter Weise angewendet. Außerdem wurden prototypische Mietpreiskarten auf Basis verschiedener Map APIs entwickelt. Abbildung 1-4 zeigt eine schematische Darstellung des zweiten Arbeitspaketes, dessen Teilpakete sowie zugehöriger Teilschritte zur Zielerreichung.

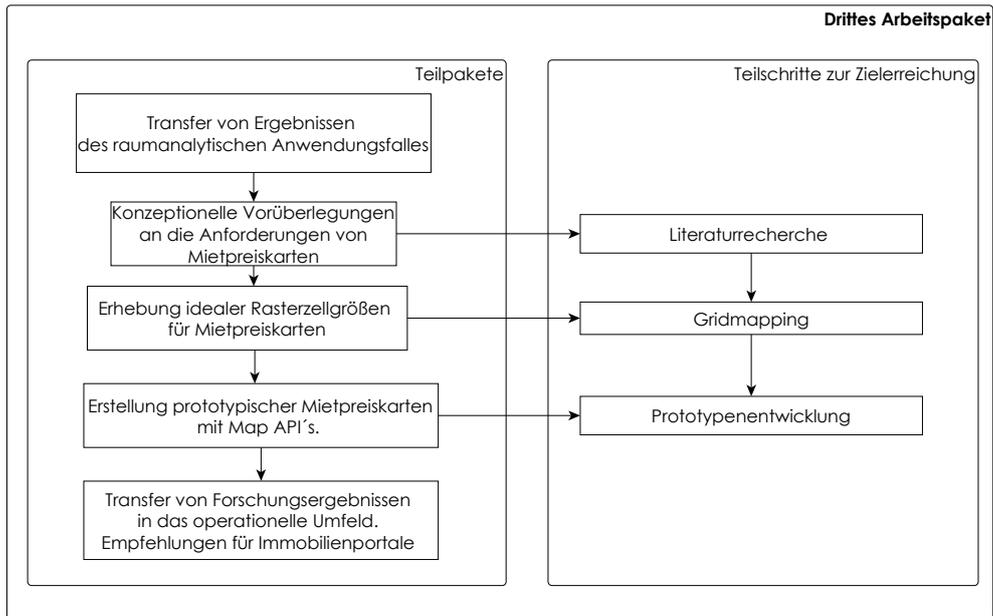


Abb. 1-4: Schematische Darstellung des dritten Arbeitspaketes. (Entwurf und Darstellung, Scherthanner 2015).

## 1.5 Gliederung und Aufbau der Arbeit

Die vorliegende Dissertation besteht aus acht Kapiteln und ist wie folgt aufgebaut. Im einleitenden Kapitel wird die Hypothese formuliert. Das zweite Kapitel gibt den Stand der Forschung, Entwicklung und Technik (den Status quo) wieder. Das dritte Kapitel hat die qualitative und quantitative Datenbasis der Arbeit zum Gegenstand. In Kapitel 4 erfolgt eine theoretische Erläuterung der räumlich-statistischen und geovisuellen Methoden, die in Kapitel 5 angewendet werden. In Kapitel 7 wird ein Fazit gezogen und es werden Umsetzungsoptionen erläutert. Abschließend erfolgen eine Zusammenfassung, eine Einordnung zum Beitrag des Forschungsstandes und ein Ausblick.

Nachfolgende schematische Übersicht (Abbildung 1-5) zeigt den Aufbau und die Gliederung der Arbeit.

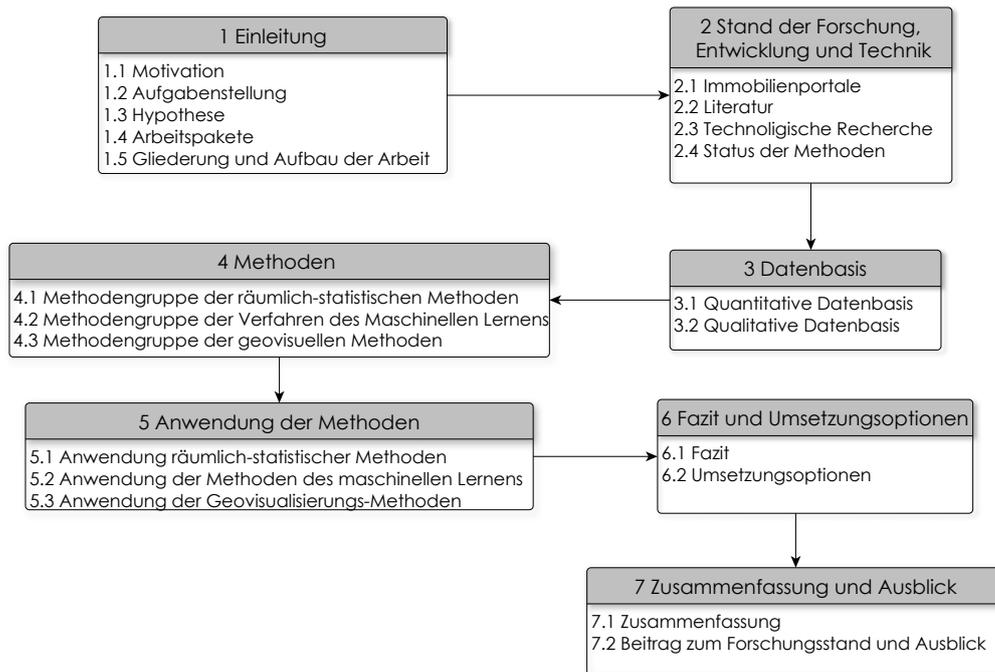


Abb. 1-5: Schematische Kapitelübersicht der ersten beiden Gliederungsebenen.  
(Entwurf und Darstellung: Scherthanner 2015).

## 2 STAND DER FORSCHUNG, ENTWICKLUNG UND TECHNIK

Die Erhebung des Standes der Forschung, Entwicklung und Technik raumanalytischer und geovisualer Methoden und Techniken bei Immobilienportalen ist Inhalt des zweiten Kapitels. Diese Bestandsaufnahme zeigt die Breite und die Vernetzungen eines möglichen neuen Forschungsfeldes, das bisher allein durch die geovisualen und geoinformatischen Methoden der Portalbetreiber abgesteckt wird und zu dessen forschungsbasierter Erweiterung diese Arbeit beitragen möchte. Zur Bestandsaufnahme des State of the Art wird vom Autor eine technologische Recherche zur Analyse der raumanalytischen und geovisualen Methoden und Techniken von Immobilienportalen sowie eine diesbezügliche Literaturrecherche durchgeführt. Ergebnis dieser Recherchen ist die Darstellung des Status quo der vorherrschenden raumanalytischen und geovisualen Methoden und Techniken bei Immobilienportalen.

### 2.1 Immobilienportale

Seit Immobilienportale die Printmedien als primäre Anlaufstelle zur Immobiliensuche verdrängt haben, ist für die Mehrzahl der Wohnungsuchenden eine Suche ohne die Nutzung eines Immobilienportals undenkbar geworden. Das Maklergeschäft wird zunehmend abhängig von Immobilienportalen, deshalb pflegen Makler ihre Angebotsdaten in die Portale ein. Immobilienportale geben einen Überblick über die einzelnen angebotenen Objekte, stellen Immobilienmerkmale in Form von Portfolios dar, visualisieren die Objektlage in Kartenform und analysieren die Angebotspreise. Nutzern werden die Analyseergebnisse in Kartenform bereitgestellt.

Die aktuelle Relevanz von Immobilienportalen lässt sich anhand zahlreicher Statistiken belegen. Im Jahr 2011 publizierte das Hamburger Marktforschungsinstitut *Fittkau und Maaß* Ergebnisse einer im November 2010 durchgeführten Studie von 15.300 Haushalten, die innerhalb eines Zeitraums von zwölf Monaten umgezogen sind. 70,9 % dieser Haushalte gaben an, ihre Immobilie online gesucht zu haben (*Rebhan, 2011*). Wird diese Erhebung der in Kapitel 1.1 vom Autor geschätzten Zahl von jährlich 5,6 Millionen Immobiliengeboten gegen-

übergestellt und werden zusätzlich die Besucherzahlenauswertungen von *Alexa.com*<sup>3</sup> betrachtet, unterstreicht dies die Relevanz von Immobilienportalen für die Immobiliensuche und ganz allgemein für Immobilienmärkte sehr deutlich.

Immobilienportale speichern alle eingestellten Angebotsdaten und haben sich zu großen Datensammlern entwickelt. Allein das 1999 gegründete deutsche Immobilienportal *ImmobilienScout 24* speichert derzeit 1,5 Mio. Immobilienangebote in seiner Datenbank (*Immobilienportale*, 2012). Die Tatsache, dass alle aufgenommenen Daten geokodiert sind, machen Immobilienportale zu einer großen Quelle für raumanalytische Ansätze der Immobilienmarktanalyse. Die enorme verfügbare Datenfülle erlaubt es den Betreibern von Immobilienportalen, umfangreiche, auf Methoden der Statistik basierende Marktanalysen durchzuführen. Maklern und Vertreibern von Immobilien werden diverse kostenpflichtige Reporting-Dienste zur Analyse der Nachfrage ihres Angebots offeriert. Neben Reporting-Diensten wurde zudem das Thema Transparenz als zusätzliches Geschäftsfeld erkannt. So werden die vorhandenen Datenbestände genutzt, um zusätzliche – nicht nur monetäre – Mehrwerte über die reine Immobiliensuche hinaus zu generieren. Beispielsweise führten die *Gesellschaft für Konsumforschung (GfK)* bzw. das *Forschungsinstitut Empirica* diverse Marktanalysen basierend auf Daten von Portalbetreibern durch (*Enderle*, 2009).

## 2.2 Literatur

Trotz der unbestreitbaren Relevanz von Immobilienportalen gibt es bisher kaum Studien hierüber. Es existieren in der Literatur allenfalls Untersuchungen, die Teilaspekte des Forschungsgegenstandes behandeln. Ein Artikel von *Borchert* (2006) beispielsweise beschäftigt sich mit der Anwendung von Geoinformationssystemen (GIS) für die Immobilienwirtschaft. Im Jahr 2009 analysierten *Schernthanner* und *Tyrillová* die Rolle freier quelloffener Geoinformationssysteme (FOS-GIS) als Instrument der Wohnungsmarktanalyse. Aus den Feldern der Wirtschaftswissenschaft und der Immobilienmarktforschung gibt es unter-

---

<sup>3</sup> *Alexa.com* ist eine von *Amazon.com* betriebene Webseite, die Nutzerzahlen von Webseiten analysiert. Das deutsche Immobilienportal *ImmobilienScout 24* liegt bei *Alexa* seit Jahren in den Top 50 der meistbesuchten Internetseiten Deutschlands (Platz 21 am 20.11.2014). Das US-amerikanische Portal *Zillow* liegt auf Platz 37 (Stand 07.05.2015) der meistbesuchten Seiten in den USA (*Alexa*, 2014).

schiedliche Beiträge zur Thematik „Immobilienportale“. *Enderle* (2009) analysiert die durch Immobilienportale veränderte Markttransparenz aus dem Blickwinkel der Wirtschaftswissenschaften. *Rohmert* und *Böhm* (2013) gaben eine Aufsatzsammlung zu den Einflüssen des Internets auf die Immobilienwirtschaft heraus.

Zu den in Portalen angewendeten raumanalytischen und geovisuellen Methoden findet sich eine höhere Anzahl an Beiträgen. Nachfolgend werden diejenigen Studien näher vorgestellt, die für den Forschungsgegenstand von Relevanz sind und als Beweis der Thesen sowie dem Finden von Lösungsvorschlägen dienen. Die Studien lassen sich dabei den Bereichen der räumlichen Analyse bzw. räumlichen Statistik, des Maschinellen Lernens und der Geovisualisierung zuordnen.

Es existiert in der Literatur eine Vielzahl an Studien zur Immobilienpreisschätzung, jedoch wurde bisher keine der darin dargestellten Methoden auf deren Anwendbarkeit in Immobilienportalen untersucht. Abbildung 2-1 verdeutlicht, dargestellt als Wortwolke, die Vielzahl an möglichen anwendbaren Methoden.



Kriging Interpolatoren. Páez (2007) vergleicht Methoden des Moving Window Kriging, der GWR und der Moving Window Regression zum Messen von Effekten der räumlichen Abhängigkeit. Bourrousa et al. (2010) stellen einen umfassenden Modellvergleich von OLS-, geostatistischen und Trendmodellen an und nutzen dabei u.a. ein hedonisches Modell in Kombination mit einem Ordinary Kriging Modell. Eine Anwendung eines Regression-Krigings stellt Tsutsumi (2011) zur Berechnung von Grundstückspreisen in Tokio (Japan) vor. Kuntz und Helbich (2014) stellen einen empirischen Vergleich an, mit dem Schwerpunkt auf der Validierung von Kriging und OCK Interpolatoren zum "Kartographieren" von Einfamilien-Wohneigentum in Wien, Österreich. Die nachfolgende Tabelle 2-1 gibt eine Übersicht über Preisschätzverfahren in der Immobilienmarktforschung.

<b>Autoren</b>	<b>Anwendungsfall Preisschätzung</b>	<b>Verwendete Methoden</b>
<b>Chica Olmo</b> (2007)	Eigentumswohnungen in Granada/Spanien	OK/OCK
<b>Bourassa et al.</b> (2010)	Einfamilienhäuser in Kentucky/USA	HED/OK
<b>Montarro und Larraz</b> (2011)	Gewerbeimmobilien Toledo /Spanien	HED/OCK/OK
<b>Wong et al.</b>	Eigentumswohnungen Apartmenthochhäuser in Hong Kong/China	Hedonische Regression
<b>Páez et al.</b> (2008)	Einzelhäuser in Toronto/Kanada	Moving Window Regression/GWR
<b>Tsutsumi et al.</b> (2011)	Grundpreise in Tokio/Japan	Regression Kriging
<b>Kuntz und Helbich</b> (2014)	Wohnungseigentum Wien/Österreich	Detrend Kriging/Universal Kriging (KED)/Universal Co – Kriging

Tabelle 2-1: Übersicht zu Studien über Preisschätzverfahren in der Immobilienmarktforschung.

Es finden sich in der Literatur zur räumlichen Schätzung von Immobilienpreisen Erweiterungen herkömmlicher Regressionsanalysen um geographische Komponenten. Brundson et al. (1998) diskutieren in ihrem Artikel "Geographically weighted regression modelling spatial non-stationarity" erstmals das Prinzip der geographisch gewichteten Regressionsanalyse (GWR) und demonstrieren damit das erste Mal die Möglichkeit, Regressionskoeffizienten auf Karten darzustellen. Seither sind zahlreiche Anwendungen von GWR zum Schätzen von Immobilienpreisen aufgetaucht. Mc Clusky et al. (2011) versuchen Immobilien-submärkte über geographisch gewichtete Regression zu identifizieren. Eine weitere Anwendung von GWR zur exemplarischen Analyse des Belfaster Mietmarktes stammt von McCor et al. (2014). Weitere Lösungsansätze in Kombination mit GWR finden sich in Studien zu Lagemodellierungen. Beispiele für Lagemodellierungen geben u.a. Prinz und Reithofer (2005), wobei Wohnstandor-

te in der Stadt Salzburg betrachtet und hierfür ein indikatorenbasiertes multikriterielles Modell entwickelt wurden.

### 2.2.2 Literatur aus dem Bereich des Maschinellen Lernens

Neben Lösungsmöglichkeiten der räumlichen Statistik lohnt sich ein Blick auf Verfahren des Maschinellen Lernens, wobei vorauszuschicken ist, dass es sich bei diesen Verfahren um nicht räumliche Verfahren handelt. Es gibt jedoch erste Ansätze in der Forschung, die Methoden des Maschinellen Lernens auch für räumliche Analysen zu nutzen. Große Beliebtheit erlangten die Methoden beispielsweise in der Fernerkundung.

*Antipov und Pokryshevskaya (2010)* stellen einen Ansatz der Preisschätzung von Wohnungen mittels der Methode Random Forest (RF) dar. *Gu et al. (2010)* verbinden einen genetischen Algorithmus mit Random Forest zur Vorhersage von Immobilienpreisen. In ihrem Artikel "Machine Learning and the Spatial Structure of House Prices and Housing Returns" wenden *Caplin et al. (2008)* Methoden des Maschinellen Lernens zur Ermittlung von Immobilienpreisen an und untersuchen mit ihrem Modell die räumliche Struktur von Transaktionsdaten. *Bark und Bae (2014)* vergleichen mehrere Algorithmen aus dem Maschinellen Lernen zur Schätzung von Immobilienpreisen. Ein weiteres Anwendungspotential haben (hierarchische) Clusterverfahren. *Hepsen und Vatansever (2012)* stellen eine Anwendung verschiedener hierarchischer Clusterverfahren für die Clustierung der Mietmärkte von 71 türkischen Großstädten vor.

### 2.2.3 Literatur aus dem Bereich der Geovisualisierung

Studien aus dem Feld der Geovisualisierung behandeln meist Teilaspekte, die für die Überprüfung der zweiten formulierten These in Frage kommen. Für diverse in der Forschung im Diskurs stehende Lösungsansätze sei auf *Huang und Gartner (2012)* verwiesen, die einen Ansatz der kollaborativen Filterung besprechen. *Burigat und Chittaro (2008)* diskutieren Lösungsansätze der Aggregation von Icons. Die Forschung über die Thematiken cloudbasiertes Mapping, Webmapping und Map APIs findet sich bei *Peterson (2012, 2014)*. Eine relevante Studie zur Thematik „Gridmapping“ veröffentlichte *Trainor (2010)*.

## 2.3 Technologische Recherche

Zur Erhebung des Status quo wurde vom Autor, komplementär zur Literaturrecherche, eine technologische Recherche durchgeführt. Nachfolgend werden die Ergebnisse dieser technologischen Recherche über die in Immobilienportalen zum Einsatz kommenden Methoden der räumlichen Analyse und der Geovisualisierung präsentiert. Zur standardisierten Untersuchung verfügbarer Immobilienportale wurde vom Verfasser eine Auswertematrix konzipiert. Insgesamt wurden anhand dieser Matrix 32 Immobilienportale nach jeweils 14 Kriterien untersucht. Einzelne Methoden, die für die räumliche Sicht von Immobili-

enportalen als besonders relevant einzustufen sind, werden im Kapitel 2.3 ausführlicher erläutert.

Bei der Auswahl der Portale wurden verschiedene Gesichtspunkte berücksichtigt. Es wurden die Immobilienportale mit den nach *Alexa.com*<sup>5</sup> höchsten Besucherzahlen ausgewählt. Danach erfolgte eine Selektion der nach *Alexa.com* drei bis fünf größten nationalen Portale in den USA, Großbritannien, Deutschland, Österreich und der Schweiz. Hierbei zeigen die Besucherzahlen, dass im angelsächsischen und im deutschen Sprachraum die meisten Immobilienportale zu finden sind. Die Liste der Portale wurde anschließend noch um einige wenige unterschiedliche innovative Ansätze verfolgende Immobilienportale erweitert. Es fand eine Bereinigung aller ausgewählten Portale um Portale ohne Angebotskarten statt. Übrig blieben 32 Immobilienportale, die anhand von 14 Kriterien untersucht wurden. Für alle spezifizierten Portale wurde jeweils festgestellt, ob sie ein Kriterium aufweisen, teilweise aufweisen oder nicht aufweisen. Im Anhang 2 findet sich die Auswertematrix, mittels derer die gewählten Portale analysiert wurden. Die überprüften Kriterien wurden in Anlehnung an die Ergebnisse der Expertenbefragung gewählt und umfassen zwei Kategorien, eine geoinformatische, raumanalytische Kategorie und eine Geovisualisierungskategorie (siehe Tabelle 2-2).

<b>Geoinformatik</b>	<b>Geovisualisierung</b>
Immobilienpreiskarte	Immobilienpreiskarte
Fachfremd verwendete administrative Grenzen	Dynamische Legende
Statistische Auswertungen	Grundkarte/Basemap
Preismodellierung mittels räumlicher Statistik	API zu Angebotsdaten

---

<sup>5</sup> *Alexa.com* zählt über eine Toolbar im Browser die Seitenaufrufe eines Nutzers und erstellt mit diesen Daten eine quartalsweise Hochrechnung. Auf Grundlage dieser Hochrechnung wird im Dreimonatsintervall eine Liste mit den 100 000 meist aufgerufen Webseiten erstellt (*Alexa, 2014*).

Immobilienindex	Zusatzinformation in Marker/Layer
Hedonische Regression	Pin/Marker Problematik
Indikatorenbasierte Distanzabhängige Suche	Mobile Angebotskarten

Tabelle 2-2: Kategorien technologische Recherche. Der Punkt „Immobilienpreiskarte“ fällt in beide Kategorien. Beim Punkt Basemap/Grundkarte erfolgte eine Auswertung nach dem Grundkarten-Anbieter.

### *Ergebnisse der Recherche*

Ein Ergebnis der Recherche ist eine Unterscheidung von Immobilienportalen in drei Kategorien.

- **Kategorie 1:** Immobilienportale ohne Angebots- und Immobilienpreiskarten. Portale dieser Kategorie wurden nicht weiter berücksichtigt.
- **Kategorie 2:** Immobilienportale mit Angebotskarten und ohne Immobilienpreiskarten.
- **Kategorie 3:** Immobilienportale mit Angebotskarten und auf Basis der hedonischen Regression erstellten Immobilienpreiskarten.

Tageszeitungen spielen bei der webbasierten Immobiliensuche praktisch keine Rolle mehr (Enderle, 2009). Diese Beobachtung von Enderle (2009) kann durch die durchgeführte Analyse bestätigt werden. Zwei Ausnahmen davon sind im deutschen Sprachraum zu nennen. Das ist einerseits die österreichische Tageszeitung *Der Standard*, die in Österreich das nach Nutzerzahlen und Bekanntheit zweitwichtigste Immobilienportal betreibt, sowie in Deutschland die *Bildzeitung*, die über den Axel Springer Verlag das Portal *Immonet.de* führt (Wer zu Wem Firmenverzeichnis, 2014). Abbildung 2-2 zeigt die Häufigkeiten, mit denen bestimmte Kriterien in Portalen festgestellt wurden. Die Portale sind hierbei, ihrer Herkunft entsprechend, nach Nationen zusammengefasst.

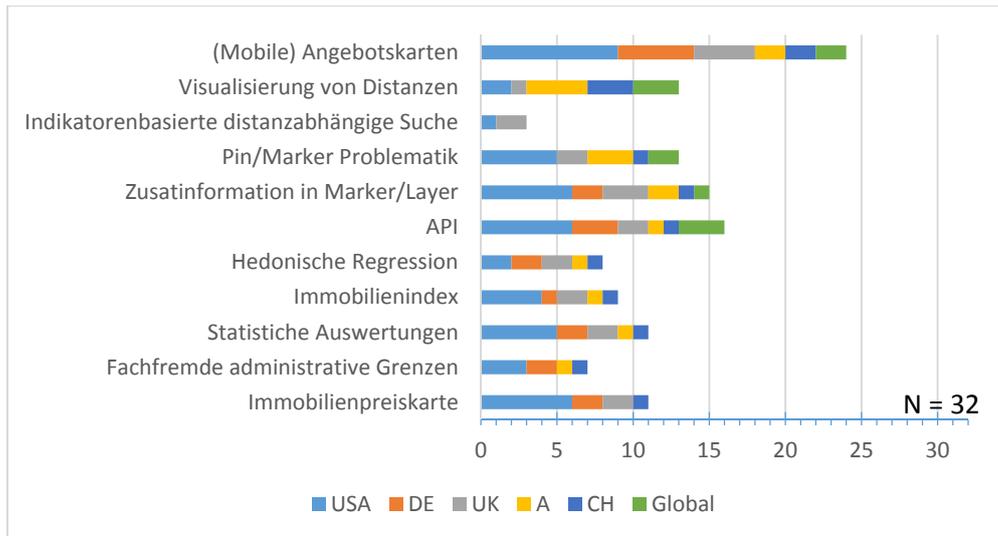


Abb. 2-2: Häufigkeit der Anwendung von Methoden der Geoinformatik und -visualisierung in 32 untersuchten Immobilienportalen nach Nationen zusammengefasst. (Entwurf und Darstellung: Scherthanner 2015).

Eine der auffälligsten Beobachtungen ist, dass 24 der 32 Portale mobile Angebotskarten veröffentlichen. Dies geschieht sowohl in Form mobiler Applikationen für die gängigen mobilen Betriebssysteme (Android/iOS/Windows Phone) als auch plattformübergreifend im Browser. Mobile Angebotskarten und responsives Webdesign<sup>6</sup> werden mittlerweile bei der Mehrheit der beobachteten Portale umgesetzt.

Alle Immobilienportale setzen auf die Verwendung von Grundkarten, die über Map APIs bereitgestellt werden. Als Grundlage (Basemap) wird hierbei der in das Medium des World Wide Web (WWW) portierte Ansatz aus der thematischen Kartographie verstanden, eine topographische Grundlage als geometrische Bezugs- und Orientierungsmöglichkeit bereitzustellen (Arnberger, 1997). Basemaps werden in der Regel als „gekachelte“ Webservices (TMS, Tiled Map-services) mittels Raster oder Vector Tiles über Map-APIs bereitgestellt. Die Webanwendung *Housingmaps.com* (Housingmaps, 2014) ist beispielsweise eine

---

<sup>6</sup> Responsives Webdesign meint das erstmals von Marcotte (2010) erläuterte Paradigma, das Webseiten auf unterschiedliche Bildschirmgrößen hin reagieren und eine auf das jeweilige Endgerät optimierte Darstellung erfolgt.

der frühesten bekannten Karten-Mashups (Peterson, 2014)<sup>7</sup>. Housingmaps stellt Immobilienangebote einer Webkarte von Google maps dar (siehe Abbildung 2-3). Die verwendeten Daten stammten dabei aus dem in den USA verbreiteten Online-Kleinanzeigendienst Craigslist (Craigslist, 2005).

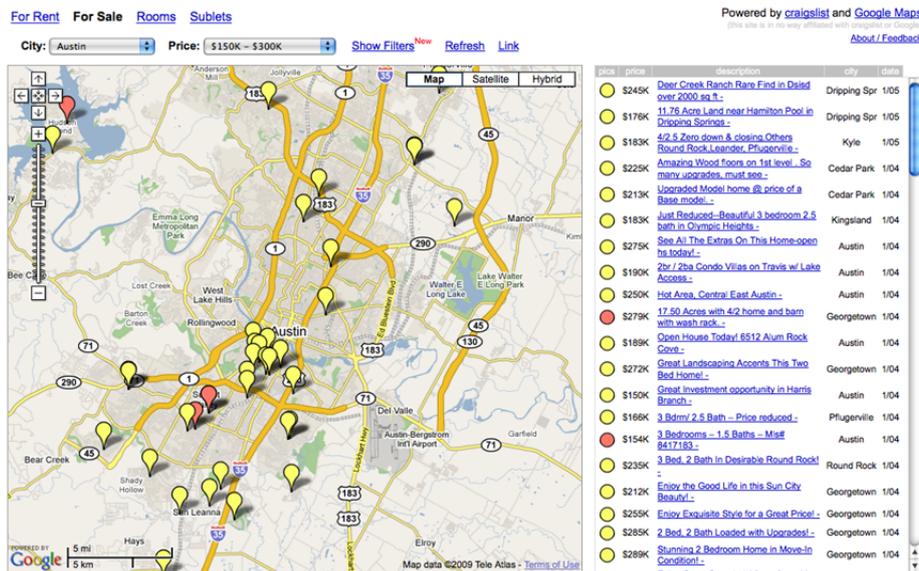


Abb. 2-3: Karten-Mashup von Housingmaps.com aus dem Jahr 2005 (Housingmaps, 2014).

Die Dominanz von Google maps als Grundkarte fiel bei der Untersuchung besonders auf (siehe Abbildung 2-4). So nutzen 21 von 32 Portalen die Google API als Schnittstelle zu Google maps. Alle 31 betrachteten Portale nutzen Grundkarten proprietärer Anbieter.

<sup>7</sup> Der aus der Popmusik stammende Begriff „Mashup“ wird in dieser Arbeit wie von Batty et al. (2010) charakterisiert verwendet. In diesem Verständnis bezieht sich der Begriff Mashup auf Webseiten, die Daten aus verschiedenen Quellen zu neuen Online-Diensten – im Falle von Karten-Mashups zu integrierten Kartendiensten – verflechten.

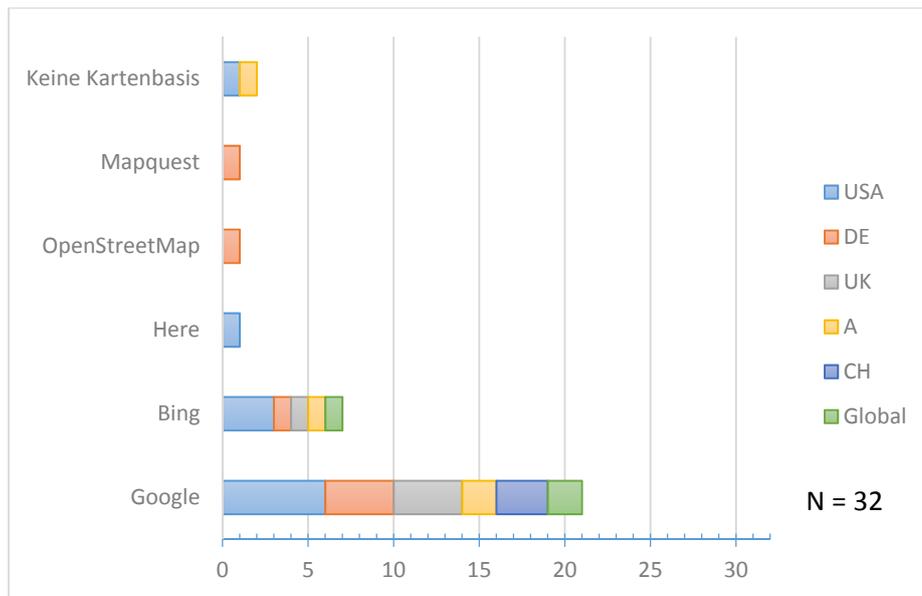


Abb. 2-4: Diagramm der in Immobilienportalen verwendeten Basemaps unterschiedlicher Map-API-Anbieter. (Entwurf und Darstellung: Scherthanner 2015).

Angebote werden bei allen Portalen in ihrer Verortung als Positionssignatur bzw. als „Pin map“ dargestellt. Hierbei wird häufig das Symbol eines Hauses gewählt. Signaturen werden meist nicht maßstabsabhängig skaliert. Eine weitere, oft gewählte Darstellung erfolgt mittels Angabe des Preises als absoluter Zahlenwert (Zahlenwertsignatur) innerhalb geometrischer Formen, wie z.B. innerhalb eines Rechtecks. Auch hierbei treten häufig Probleme mit überlappenden Signaturen auf. Zwar wird eine jeweils von den Map-APIs bereitgestellte maßstabsabhängige Generalisierung der Grundkarte angeboten, bei den Markern selbst fehlt jedoch jegliche maßstabsabhängige Generalisierung. Eine Änderung der Zoomstufe<sup>8</sup> bewirkt bei 13 Immobilienportalen keinerlei und bei neun untersuchten Portalen teilweise starke Überlappungen der Positionsmarker. Ein Fehlen von maßstabsabhängigen Generalisierungen im Zusammenhang mit Map-APIs bzw. Webkarten ist ein bestehendes Forschungsproblem, bekannt als „Icon Cluttering“, ausführlich diskutiert u. a. von Huang und

<sup>8</sup> Auf der Wiki des Openstreetmap Projektes (Openstreetmap, 2014) findet sich eine Übersicht über den Maßstab der jeweiligen Zoomstufen. Beispielsweise bedeutet in der Openstreetmap eine Zoomstufe von 13 einen Kartenmaßstab von 1 : 19 093.

Gartner (2012). In Kapitel 2.3.4 kommt es zu einer weiteren Diskussion der Problematik. Immerhin sechs Portale bieten Lösungsansätze für diese Problematik, meist in Form einer Aggregation durch Angabe einer Ziffer innerhalb des Positionsmarkers oder über eine Clusterung von Angeboten bei kleinerem Maßstab. Es wird teils beim Herauszoomen in Zoomstufen kleinerer Maßstabsebenen die Anzahl der aggregierten Angebote in eine geometrische Signatur geschrieben. Über eine Farbkodierung der Positionsmarker und das Einbinden von Popups integrieren viele Portale Zusatzinformation in die Marker. Einige Portale stellen zusätzliche Datenschichten bereit. Eine distanzbasierte Angebotssuche ist in wenigen Portalen möglich. Teils erfolgt eine Visualisierung von euklidischen Distanzen ähnlich einem Puffer, basierend auf denen von Suchenden festgelegten Umkreisen. Ob hierbei netzwerkbasierend oder über euklidische Distanzen gesucht wird, kann nur vermutet werden. Eine Suche über zusätzliche Indikatoren, z.B. über eine räumliche Selektion von Angeboten nach topologischen Regeln, ist in keinem Portal möglich.

Fast alle großen Portale bieten eine eigene Programmierschnittstelle zu ihren Angeboten. Hierdurch wird es möglich, diese Angebote dann geokodiert oder revers geokodiert für eigene Anwendungen zu beziehen. In den meisten Fällen handelt es sich dabei um APIs die dem Paradigma der RESTful Webservices (REST) unterliegen<sup>9</sup>. Alle größeren Portale stellen ihren Nutzern Immobilienpreiskarten bereit. Die errechneten Preise werden dabei in der Mehrzahl der Fälle auf willkürlich gewählte räumliche Bezugseinheiten umgerechnet und als Choroplethenkarten dargestellt. Die Preise sind dabei reine Durchschnittspreise oder ein auf z. B. einen Stadtteil umgelegter Median, ein Durchschnittspreis oder das auf eine Bezugsgeometrie umgerechnete Ergebnis einer hedonischen Regression. In Abbildung 2-5 wird eine solche Karte des Portals *Homegate.ch* gezeigt. *Homegate* (2014) weist ganz klar aus, dass Preise mittels einer hedonischen Regression gerechnet werden. Die räumlichen Bezugseinheiten werden nicht angegeben.

---

<sup>9</sup> Einem von *Fielding* (2000) vorgeschlagenen Paradigma für Webservices, wonach gewünschte Informationen im Falle von Angebotsdaten der Immobilienportale über Formate wie XML als JavaScript Object Notation (JSON) oder in der Hypertext Markup Language (HTML) ausgegeben werden.

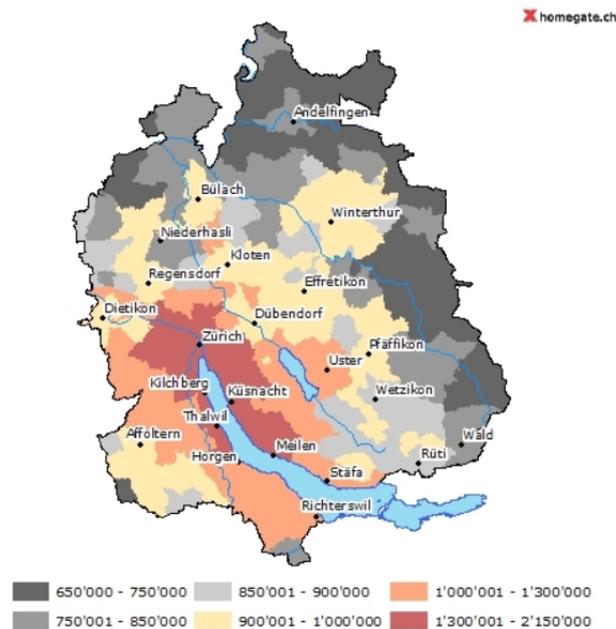


Abb. 2-5: Immobilienpreiskarte Zürichs für Wohneigentum in Schweizer Franken des Portals *homegate.ch* (2014).

Neben einfachen deskriptiven Maßen zur Preisberechnung ist die Methode der hedonischen Regression die am weitesten verbreitete Methode der Preisberechnung. Acht Portale geben ganz klar an, dass diese Form der Immobilienpreisschätzung eingesetzt wird. Es ist jedoch anzunehmen, dass noch mehr Portale auf diese Methode setzen. Räumliche Zusammenhänge finden in der Methode der hedonischen Regressionsgleichungen keine Beachtung. Sie zählen demgemäß zu den nicht räumlichen Verfahren. Die Ergebnisse der hedonischen Regression werden dabei ebenfalls auf fachfremd verwendete Raumeinheiten umgelegt. Die Ergebnisse hedonischer Regressionsgleichungen sind vielfach Basis von Immobilienpreisindizes, die von Portalen veröffentlicht werden. Insgesamt neun Portale bieten Immobilienpreisindizes an, wovon wiederum acht Fällen auf einer hedonischen Regression – teils gekoppelt mit einem, teils ohne ein automatisiertes Immobilienbewertungsverfahren – beruhen.

Keines der Portale nutzt in den veröffentlichten Preiskarten eine dynamische Legende, weder bei den veröffentlichten Immobilienpreiskarten noch bei den Angebotskarten.

## 2.4 Status quo der Methoden

Nachfolgend wird der *Status quo* der Methoden der Immobilienpreisanalyse und -darstellung von Immobilienportalen aus Forschungssicht der Geoinformatik adressiert. Im Einzelnen sind das: die Methode der hedonischen Regression (siehe Kapitel 2.4.1) und die damit im Zusammenhang stehenden fachfremd gewählten räumlichen Bezugseinheiten (siehe Kapitel 2.4.2), die nicht dynamischen Legenden (siehe Kapitel 2.4.3), die Verortung von Angeboten über Punktsignaturen und das damit entstehende Problem der Signaturenüberlage-

rung (Icon Cluttering) (siehe Kapitel 2.4.4) sowie die Integration von Distanz und räumlichen Indikatoren in Immobilienportale (siehe Kapitel 2.4.5).

### 2.4.1 Hedonische Modellverfahren

Hedonische Regressionsgleichungen, als methodische Vergleichswertverfahren, sind ein Standardverfahren der Miet- und Kaufpreisschätzung bei Immobilienportalen. Es erfolgt ganz allgemein eine Preisschätzung von Produkten mit laufender Veränderung des Preis-Leistungs-Verhältnisses. Der Preis wird über eine Qualitätsanpassung ständig neu berechnet (Gordon, 1990). Der Preis einer Immobilie wird über ihre Qualität, ausgedrückt in den intrinsischen Werten (Ausstattungsmerkmale wie z. B. Fläche, Zimmeranzahl, Etage etc.), und über die Verfügbarkeit geschätzt. Die räumliche Autokorrelation<sup>10</sup> von Immobilien wird hierbei vernachlässigt (eine ausführliche Diskussion des Konzeptes der räumlichen Autokorrelation folgt in Kapitel 4). In Bezug auf die Anwendung von Immobilienportalen ist demnach eine Vernachlässigung von Toblers (1970) postuliertem Ersten Gesetz der Geographie: „Everything is related to everything else, but near things are more related than distant things“ festzustellen. Das Feld der hedonischen Regression geht auf Court (1939) und Griliches (1961) zurück. Angewendet wird die hedonische Regression zur Preisschätzung eines Produktes bei laufender zeitlicher Veränderung des Preis-Leistungs-Verhältnisses. Gordon (1990) spricht deshalb von einer Preisberechnung mit Qualitätsanpassung.

Die Grundlage der hedonischen Regression ist die hedonische Funktion, in der verschiedene Varianten eines Produkts  $k$  einem Preis zugeordnet werden (Triplet, 2004). Laut Diewert (2003) werden einzelne Varianten eines Produktes mittels verschiedener Merkmalsausprägungen der dem Produkt zugesprochenen  $m$  Eigenschaften unterschieden, das ergibt einen Tupel in der Form  $k = (x_{k1}, \dots, x_{km})$ . Als Preis wird laut Diewert (2003) immer der Preis gesehen, den ein möglicher Käufer bereit ist zu zahlen. Der Preis wird in verschiedene Komponenten aufgeteilt, d. h., das Verfahren beruht auf der Annahme, dass

---

<sup>10</sup> Bei etlichen räumlichen Prozessen sind die Ähnlichkeiten oder die Unterschiede der von einer gewissen räumlichen Distanz getrennten Umstände größer, als bei zufälligen Prozessen zu erwarten wäre (Legendre, 1993). In Kapitel 4.1.2 erfolgt eine umfassende Erklärung der räumlichen Autokorrelation.

ein Gesamtpreis eines Produktes auf Grundlage unterschiedlicher Einzelkomponenten eines Produktes berechnet werden kann. Eine Preisschätzung erfolgt nur in Abhängigkeit von einer betrachteten Zeitperiode (t). Der Faktor Zeit muss also in die Gleichung integriert werden. Das kann auf zwei Arten geschehen. Man findet für t entweder eine Dummy-Variable<sup>11</sup> oder berechnet die hedonische Regression für einzelne Zeitintervalle. Die betrachteten Immobilienportale schätzten ihre Preise dabei ausschließlich für das Zeitintervall "Quartal", also in vierteljährlichen Zeitschritten.

Exemplarisch wird nachfolgend die Semilog-Variante von Gordon (1990) beschrieben: Für ein Objekt k wird der Preis  $P_k$  ermittelt.

$$\ln P_k = \beta_0 + \sum_{j=1}^m \beta_j X_{kj} + \delta_k$$

Formel 2-1: Semilog-Variante der hedonischen Regression.

Der zu ermittelnde Preis ist die Primärvariable, die Sekundärvariable  $m$  wird zur Preisermittlung herangezogen (Li und Heap, 2008). Es werden Anteile am Gesamtpreis berechnet, die auf einzelne Eigenschaften bzw. Ausprägungen eines Produktes zurückgehen. Es lassen sich mit dem Regressionskoeffizienten der hedonischen Regression Marktpreise berechnen, die noch nicht bekannt sind und unterschiedliche Produktvarianten (z. B. Produktvariante Reihenhaus oder Dachgeschosswohnung) schätzen. Die Marktausprägungen  $m$  sind Ausprägungen des Einzelobjektes. Räumliche Autokorrelation wird bei der hedonischen Regression außer Acht gelassen. Immobilienportale legen die Ergebnisse der hedonischen Regression auf fachfremd verwendete geometrische Bezugseinheiten um.

## 2.4.2 Räumliche Bezugseinheiten

Ergebnisse der hedonischen Regression werden auf fachfremd verwendete räumliche Bezugseinheiten umgelegt. Das können Stadtteile, Postleitzahlenbereiche, statistische Blöcke etc. sein. Im Rahmen der Dissertation vom Autor un-

---

<sup>11</sup> Binäre „Scheinvariable“, die ausdrückt, ob eine Eigenschaft zutrifft oder nicht.

tersuchte US-amerikanische Portale wiesen mehrheitlich von den Portalen selbst definierte Geometrien, genannt Nachbarschaften („Neighborhoods“), Postleitzahlenbezirke („Zip Codes“), Stadtteilgrenzen oder statistische Einheiten des US-Zensus („Core based statistical areas“) auf. *Core based statistical areas* der US-amerikanischen nationalen Statistik bestehen beispielsweise aus Bezirken und Landkreisen die mindestens einen urbanen Kern aufweisen, mindestens 10.000 Einwohner haben und einen hohen Grad an sozioökonomischer Verbundenheit mit dem urbanen Kern zeigen (*United States Census, 2010*).

Das Immobilienportal *Zillow.com* veröffentlicht alle „Nachbarschaften“ unter der freien „Gnu Public License“. Es handelt sich um insgesamt 7.000 Nachbarschaften, die für größere Städte gebildet wurden (*Zillow, 2014*). Abbildung 2-6 zeigt beispielhaft drei Zip-Code-Bereiche in New York Manhattan im Vergleich von der von *Zillow.com* gebildeten „Neighborhood“ namens Gramercy. Teils orientieren sich diese „Neighborhoods“ an den Zip Codes, teils sind sie größer, teils kleiner gewählt. Es ist keine Methodik für die Auswahl der Neighborhoods erkennbar. *Zillow* berechnet jeweils für jede Nachbarschaft den Median der Miet- und Kaufpreise.



Abb. 2-6: Links: Vergleich von drei Zip-Code-Bereichen (rot) mit einer Zillow-Neighborhood in New York (blau gestrichelte Linien). Maßstab: 1 : 40.000. (Entwurf und Darstellung: Scherthanner 2015).

Beim deutschen Portal *Immobilienscout 24 (IS24)* wurden als Wohnquartiere bezeichnete Bezugseinheiten der Firma *Infas Geodaten*<sup>12</sup> gewählt. Laut *Infas Geodaten* umfasst ein Wohnquartier im Durchschnitt 400 Haushalte und weist eine möglichst homogene Struktur auf (*Nexiga, 2014*). Die erste Erfassung der Wohnquartiere erfolgte 1994 auf Grundlage der Stimmbezirke der Bundestagswahl 1987 (alte Bundesländer) und auf Basis der Stimmbezirke der Bundestagswahl 1994 (Neue Bundesländer). 2006 waren 75.000 Wohnquartiere verfügbar (*Infas Geodaten, 2006*). Wohnquartiere beachten keine räumliche Autokorrelation. Die Größe der Wohnquartiere variiert sehr stark und ist im ruralen Raum viel größer als im dicht besiedelten urbanen Raum. Abbildung 2-7 zeigt

<sup>12</sup> *Infas Geodaten* wurde 2014 in *Nexiga GmbH* umbenannt.

einen Vergleich der Größe eines Wohnquartiers in Potsdam-Babelsberg mit einem Postleitzahlenbezirk.

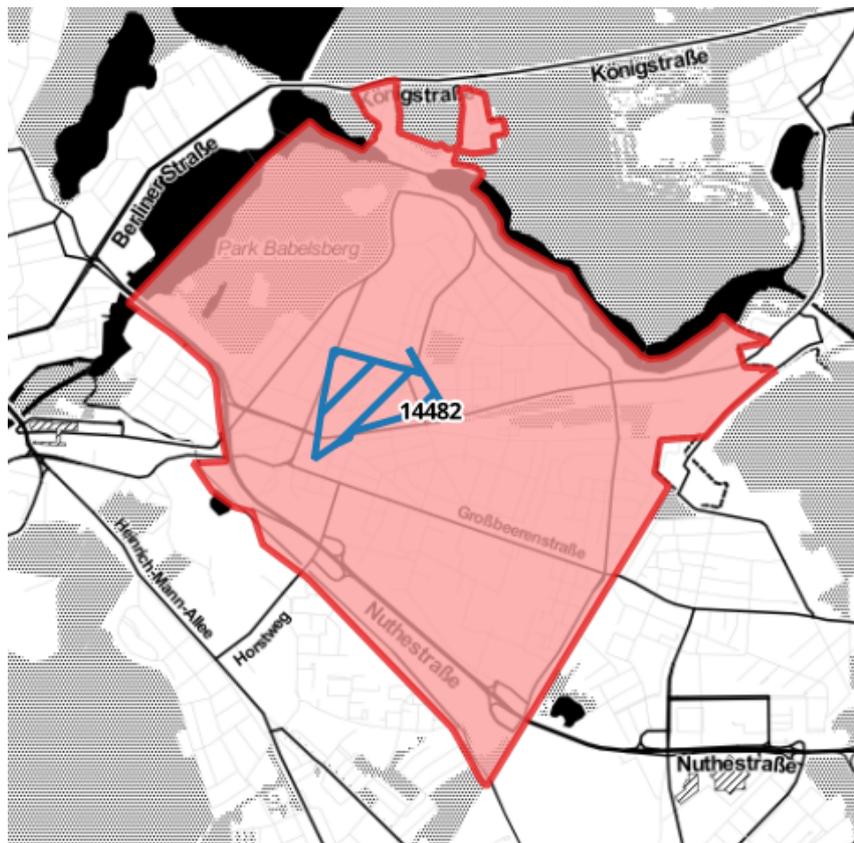


Abb. 2-7: Geometrie eines Postleitzahlenbezirks (rot) im Vergleich zu einem Wohnquartier der Firma *Infas/Nexiga*. 1 : 75.000. (Entwurf und Darstellung: Scherthanner 2015).

### 2.4.3 Nicht dynamische Legenden

Bei den veröffentlichten Immobilienpreiskarten dominieren nicht dynamische Legenden. Diese Legenden passen sich nicht an die im Kartenausschnitt vorhandenen Immobilienpreise an, dieselben Klassen werden uniform für ganz Deutschland verwendet. Das führt dazu, dass in Immobilienpreiskarten unabhängig vom tatsächlichen Preisverhältnis des Untersuchungsgebietes diesel-

ben Preisklassen bei Immobilienpreisen angezeigt werden. Das Histogramm der Mietpreise Münchens (Stand Dezember 2012) in Abbildung 8 verdeutlicht diese Problematik<sup>13</sup>. Wohnungen zur Miete in München werden auf dem Immobilienportal IS24 zum Großteil der teuersten Preisklasse von größer als 12 €/m<sup>2</sup> zugewiesen, und das, obwohl laut Angaben auf IS24 der Durchschnittspreis beispielsweise im Stadtteil München-Bogenhausen 13,6 €/m<sup>2</sup> (Stand Dezember 2014) betrug, während rund um den Viktualienmarkt 17,10 €/m<sup>2</sup> Durchschnittsmiete verlangt wurden. Es erfolgt eine fast uniforme Zuweisung zur Preisklasse von größer als 12 €/m<sup>2</sup>. Die Klassenzuweisung eines Durchschnittspreises auf Stadtteilebene zur höchsten Preisklasse ist somit verzerrend. Dieselbe Legende wird im Kontrast dazu auch auf Gebiete mit einer sich gänzlich unterscheidenden Marktzusammensetzung angewendet. Beispielsweise wird ganz Osterode am Harz der niedrigsten Preisklasse von geringer als 4 €/m<sup>2</sup> zugewiesen, obwohl dort der Durchschnittspreis im ersten Quartal 2014 bei 4,3 €/m<sup>2</sup> lag. Die Klassen passen sich nicht an die im Kartenausschnitt tatsächlichen Mietpreise an. Dieselben Preisklassen werden ohne Rücksicht auf regionale Begebenheiten deutschlandweit angewendet. Die Kartendarstellung orientiert sich daher nicht an den tatsächlichen Marktverhältnissen, es erfolgt keine Anpassung der Legenden an die tatsächliche vor Ort herrschende Werteverteilung der Preise. Zur Bestätigung dieser Beobachtung wurden 1.152 Datenpunkte vom 04.12.2014 mit Angebotsmieten der Stadt München über die Rest-API des Immobilienportals IS24 bezogen (*ImmobilienScout 24*, 2014). Die danach gebildete deskriptive Statistik der Stadt München weist Extremwerte von 8,84 €/m<sup>2</sup> bis 36 €/m<sup>2</sup> bei einem Mittelwert von 17,87 €/m<sup>2</sup> und einer Standardabweichung von 4,51 €/m<sup>2</sup> auf. Abbildung 2-8 zeigt ein Histogramm der Preisverteilung. Hiermit wird belegt, dass die Kartendarstellung (siehe Abbildung 2-9) und Klassenbildung der veröffentlichten Mietpreiskarte keine plausible Klasseneinteilung zeigen.

---

<sup>13</sup> Die Daten zur Berechnung des Histogramms wurden über die Rest API des Immobilienportals IS24 bezogen: <http://api.immobilienscout24.de/>.

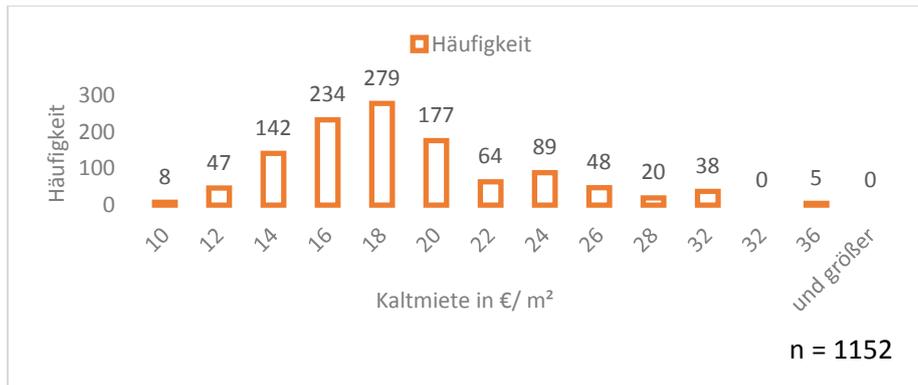


Abb. 2-8: Histogramm der Nettokaltmieten der Stadt München. (Entwurf und Darstellung: Schernthanner 2015).



Abb. 2-9: Ausschnitte aus Mietpreiskarten des Immobilienportals IS24 Links: München. Rechts: Osterode im Harz. (ImmobilienScout 24, 2015)

## 2.4.4 Punktsignaturen und das „Icon-Cluttering“-Problem

Bei einer ungefilterten Darstellung von Immobilienangeboten als Punktsignaturen kommt es zu dem bekannten Problem der Signaturenüberlagerung (Icon Cluttering). Die Positionsmarker überlappen und überlagern sich (Huang und Gartner, 2012). Die Lesbarkeit der Angebotskarten wird stark eingeschränkt. Es ist Nutzern nicht mehr in sinnvoller Weise möglich, Informationen über Immobilienangebote zu lesen. Besonders stark zum Tragen kommt die Problematik bei mobilen Angebotskarten. Burigat und Chittaro (2006) führen das Beispiel einer mobilen Kartendarstellung von Restaurants an. In einem großen Maßstab ist die Darstellung der Immobilienangebote problemlos, die POIs und zur Orientierung eingeblendete Straßenverläufe in der Grundkarte sind gut ersichtlich. Bei einem Wechsel in kleinere Maßstäbe tritt das Icon-Cluttering-Problem durch die

Überlappung von Symbolen deutlich in Erscheinung – Straßenverläufe zur Orientierung werden durch Signaturen überlagert.



Abb. 2-10: Burringat und Chittaros (2008) Illustration des Icon-Cluttering-Problems am Beispiel von POIs in Yahoo maps.

### 2.4.5 Integration von Distanz

Die Relevanz der Integration distanzbasierter Suche steigt bei Immobilienportalen. Jedoch wird Distanz in den meisten Fällen über eine rein euklidische Distanz und nicht über eine Netzwerk-Distanz ausgedrückt. Das heißt, dass beispielsweise die Immobiliensuche auf Angebote innerhalb eines in euklidischer Distanz definierten Puffers eingeschränkt wird. Einige Portale wie *Trulia* oder *Immobilienscout* verknüpfen die Suche mit der Berechnung von Reisezeiten, die über die Nutzung von externen APIs berechnet werden. Im Falle von *IS24* ist es möglich, Angebote über Reisezeiten zu definierten Zielen im Öffentlichen Personen Nahverkehr (ÖPNV) zu filtern. Abbildung 2-11 zeigt als Beispiel das Angebot des Immobilienportals *Trulia*, über definierte Punkte und als Filter definierte Reisezeiten in einem Straßennetzwerk und im ÖPNV zu berechnen und darüber Angebote zu filtern.

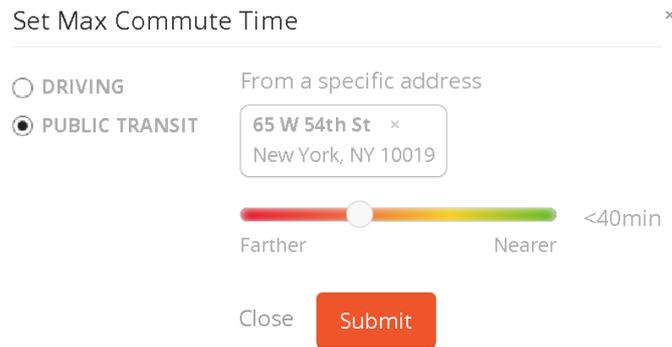


Abb. 2-11: Angebotsfilterung über Pendelzeiten (Trulia, 2014).

Visualisiert werden die Reisezeiten bei Trulia nicht, es erfolgt nur eine Anzeige der gefilterten Angebote. Eine Möglichkeit, Angebote über eine multimodale Verknüpfung von Reisemitteln zu filtern, existiert in Portalen bis dato nicht.

#### 2.4.6 Integration von räumlichen Indikatoren

Immobilienportale setzen vielfach auf die Integration von räumlichen Indikatoren bei ihren Angebotskarten. Jedoch finden sich sehr große räumliche Unterschiede. Bei Immobilienportalen im deutschsprachigen Raum werden als für die Lage wichtig erachtete Punkte als POIs visualisiert. Es kommt zur Darstellung von POIs wie z. B. Gesundheitseinrichtungen, Schulen, ÖPNV-Haltestellen als zusätzlichen Kartenschichten. Bei angelsächsischen Portalen wird mit Lageindikatoren sehr offen umgegangen, es gibt thematische Schichten mit der Bewertung von Schulqualität, Kriminalität (siehe Abbildung 2-12), der Gefahr vor Naturgefahren und vieles mehr.



Abb. 2-12: Karte des Kriminalitätsrisikos mit geclusterten Tatebestände (Trulia, 2014).

In keinem der Portale wird jedoch transparent offengelegt, wie diese Bewertungen zustande kamen, eine Art räumlicher Lageindex wird in keinem der Portale gebildet.

## 2.4.7 Fazit

Die durchgeführte technologische Recherche zeigt, dass die räumliche Sicht von Immobilienportalen eingeschränkt ist. Es werden keine räumlichen Analysemethoden zur Preisschätzung angewandt. Nicht räumliche Schätzergebnisse werden auf fachfremd verwendete Bezugseinheiten bezogen. Es existieren nur wenige Forschungsarbeiten über Immobilienportale, jedoch gibt es eine Reihe an Forschungsarbeiten, die aus wissenschaftlicher Sicht Teilaspekte der identifizierten Fragestellungen behandeln.

Keine der recherchierten Forschungsarbeiten beschäftigt sich mit der Schätzung von Mietpreisen, sondern vielmehr mit der räumlichen Schätzung von Kaufpreisen. Keine der Arbeiten bringt die untersuchten Methoden in einen operationellen Zusammenhang. Die Parameterkonfiguration für Schätzverfahren und die Sekundärvariablenauswahl wurden in der bisherigen räumlichen Immobilienmarktforschung noch unzureichend analysiert. Der Verfasser der vorliegenden Dissertation sieht in Verfahren der räumlichen Statistik (vor allem Schätzverfahren) und Verfahren aus dem Gridmapping mit einer Visualisierung mittels Map APIs das größte Anwendungspotential zur Belegung der formulierten Thesen.

## 3 DATENBASIS

Zur Beantwortung der im ersten Kapitel erhobenen Fragestellungen konnte auf eine quantitative Datenbasis zurückgegriffen werden, die im Kapitel 3.1 vorgestellt wird. Die quantitative Datenbasis reicht jedoch nicht vollends zur umfassenden Beantwortung aller Fragestellungen aus, deshalb war es notwendig, komplementär qualitative Daten zu erheben. Dazu wurden Experten befragt. Die vom Autor durchgeführten Experteninterviews sind Gegenstand des Kapitels 3.2.

### 3.1 Quantitative Datenbasis

Zur Erarbeitung von Lösungsvorschlägen zur räumlichen Optimierung von Immobilienportalen konnten exemplarische Mietangebotsdatensätze gewonnen werden. Es konnte ein multitemporaler und multidimensionaler Angebotsdatensatz als Auszug aus der Datenbank des Immobilienportals *IS24* bezogen werden<sup>14</sup>. Die bezogenen Daten gleichen vom Aufbau her Datensätzen anderer Portale und stehen somit exemplarisch für Angebotsdatensätze von Immobilienportalen. Der Datensatz umfasst quartalsweise die Mietangebote der Stadt Potsdam (Deutschland) von Januar 2007 bis September 2013. Insgesamt beinhaltet der Datensatz über alle Zeitschnitte 74.098 Angebote. Die Daten sind geokodiert und liegen in der Projektion WiGeoEU (Lambert'sche winkeltreue Kegelpjektion) vor. Die Angebotsdaten enthalten 64 verschiedene immobilienpezifische Variablen, von denen 37 numerisch (Flächenangaben, Zimmeranzahl, Kosten), 16 kategorial (Ausstattungsmerkmale) und zehn Strings (vor allem Namensbezeichnungen und interne Kodierungen) sind. Die nachstehende Tabelle (Tabelle 3-1) zeigt Beispiele für Variablen des verwendeten Datenbankauszugs. Anhang 3 zeigt alle Variablen des Datensatzes.

---

<sup>14</sup> Der exemplarische Angebotsdatensatz wurde dem Autor im Rahmen des Programms Programm „TransparenzOffensive“ des Immobilienportals *IS24* zur Verfügung gestellt: <http://www.transparenzoffensive.net/>.

Variable	GEOX	GEOY	Wohnfläche	Zimmeranzahl	ETAGE	COURTAGE	MIETE-KALT
<b>Beschreibung</b>	X-Koordinate	Y-Koordinate	Wohnfläche in m <sup>2</sup>	Zimmeranzahl	Etage des Angebots	Maklercourtage	Kaltmiete

Tabelle 3-1: Beispiele für Variablen des Datensatzes.

Damit ähnelt der Datensatz vom Aufbau her den Angebotsdatensätzen anderer Portale. Hieraus lässt sich ableiten, dass die angewendeten Analysen generisch auf andere Angebotsdatensätze übertragbar sind. Der Datenbankauszug ist insofern ideal für den raumanalytischen Anwendungsfall, als der Datensatz unterschiedliche Angebotsdichten in den jeweiligen Quartalen beinhaltet. Das erhöht zwar die Analysekomplexität, entspricht jedoch realweltlichen Bedingungen von städtischen Mietmärkten mit sehr unterschiedlichen Datendichten (siehe Abbildung 3-1). Der 0,44 km<sup>2</sup> große Potsdamer Stadtteil „Kirchsteigfeld“ hat z. B. eine über den Untersuchungszeitraum gesehene jährliche Angebotsdichte von 962,76 Angeboten/km<sup>2</sup>, während der Stadtteil Uetz-Paaren eine jährliche Angebotsdichte von nur 0,057 Angeboten/km<sup>2</sup> aufweist. Damit existiert in diesem Stadtteil praktisch kein Mietmarkt.

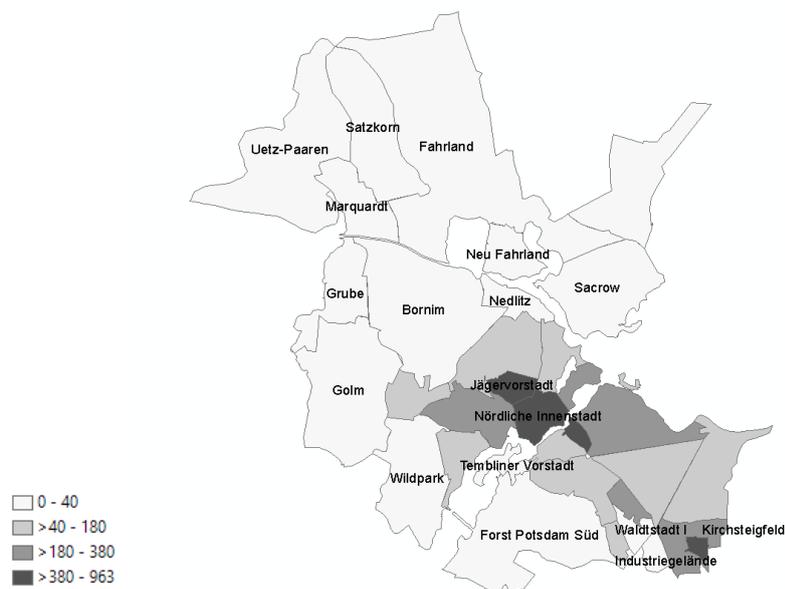


Abb. 3-1: Jährliche durchschnittliche Angebotsdichte nach Stadtteilen in Potsdam (2007 – 2013). (Entwurf und Darstellung: Schernthanner 2015).

Es wurde ausschließlich mit „historischen“ Angebotsdaten gearbeitet. Das ist gerechtfertigt, da die Optimierung des Instrumentariums und nicht die Analyse der aktuellen Marktlage im Vordergrund der Arbeit steht. Hierfür wurden aus dem Gesamtdatensatz die Angebotsdaten des ersten Quartals 2013 gewählt. In diesem Zeitschnitt liegen 1.319 Angebotsdaten vor. Eine Übertragbarkeit der

Implementierungen auf rezente Datensätze und auf andere städtische Untersuchungsräume ist durch die gewählten generischen Ansätze gegeben.

Ein weiterer Grund für die Arbeit mit historischen Daten ist die große Medienpräsenz und Relevanz der Thematik „Mieten“ insbesondere in Potsdam. Die Arbeit untersucht das Instrumentarium zur raumanalytischen Betrachtung von Mietpreisen und distanziert sich damit bewusst von aktuellen Pressediskursen. Eine Suchanfrage bei *Google News* mit dem eingeschränkten Suchzeitraum vom 20.11.2014 bis 20.12.2014 ergab in der Kombination der Suchbegriffe „Miete und Potsdam“ 40 Artikel mit der Thematik Mietpreise, in denen auch der Potsdamer Mietmarkt angesprochen wird.<sup>15</sup>

Es stellt sich die Frage, welche Relevanz die Analyse und Schätzung von Angebotsmieten hat. Angebotsmieten geben die aktuelle Marktlage wieder und sind von der Aussagekraft Neuvertragsmieten gleichzusetzen. Sie spiegeln in angemessener Weise die aktuelle Marktlage wider (BBSR, 2014). Datensätze wie der vorliegende sagen jedoch nichts darüber aus, ob die Wohnungen nach der Veröffentlichung im Immobilienportal tatsächlich vermietet wurden.

## 3.2 Qualitative Datenbasis

### 3.2.1 Experteninterviews

Komplementär zu den gewonnenen qualitativen Daten wurde zur Beantwortung der Forschungsfragen der qualitative Ansatz des Experteninterviews gewählt. Zur Beantwortung der Frage: „Was ist ein Experte?“, folgt die Arbeit der Definition von *Meuser und Nagel (1991)*.

Demgemäß ist ein Experte u. a.:

- „wer in irgendeiner Weise Verantwortung trägt für den Entwurf, die Implementierung oder die Kontrolle einer Problemlösung“ bzw.
- „wer über einen privilegierten Zugang zu Informationen über Personengruppen oder Entscheidungsprozesse verfügt.“

---

<sup>15</sup> Einen Pressespiegel zu der Thematik Mieten und Mieten in Potsdam findet man unter <http://potsdam22.de/presse>.

## *Auswahl der Experten*

Maßgeblich für die Auswahl der Experten war der besonders gute Kenntnisstand der Personen in den Thematiken: Immobilienwirtschaft, Immobilienportale, Methoden der räumlichen Analyse und Geovisualisierung von Mietpreisdaten. Insgesamt wurden fünf Experten befragt.

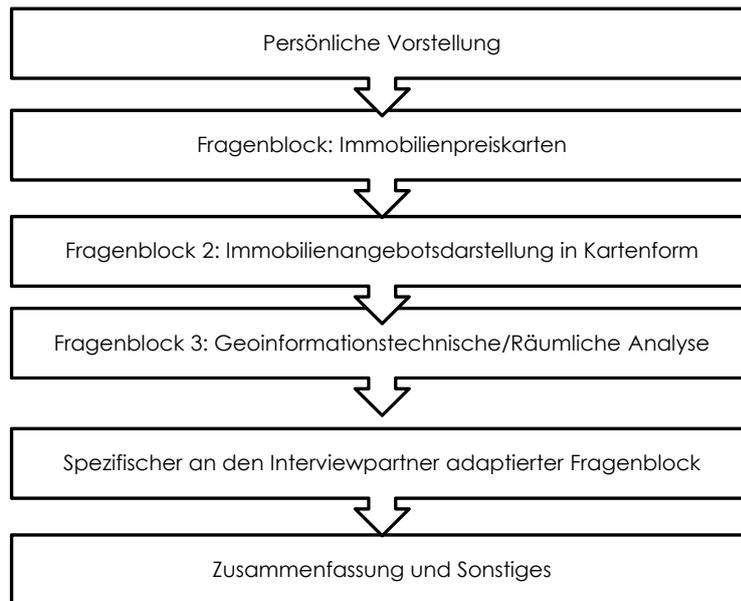
Die interviewten Experten waren: 1. *Peter Ache*, Autor des Immobilienmarktberichts Deutschlands und Leiter der oberen Gutachterausschüsse Deutschlands, 2. *Jörn-Michael Westphal*, Geschäftsführer der Potsdamer städtischen Immobiliengesellschaft *Pro Potsdam*. Es wurden außerdem drei Mitarbeiter des Immobilienportals *IS24* befragt, die mehrjährige Erfahrung in ihren Arbeitsbereichen haben. Die Experten sind 3. *Jan Hebecker*, Leiter des Bereichs „Research und Analysis“ und Leiter der „Transparenzoffensive Immobilienwirtschaft<sup>16</sup>“, 4. *Lars Holger Wilke*, Leiter des Bereichs „Analytics Insights Data“ und 5. *Dr. Jan Arne Schwarz*, Leiter des Produktmanagements im Bereich Immobilienbewertung. Eine grundsätzliche Übertragbarkeit der Meinungen der befragten Experten wird angenommen, da ihr Expertenwissen exemplarisch für Immobilienportale, Immobilienpreisanalysen bzw. die Immobilienwirtschaft allgemein gesehen werden kann.

## *Aufbau der Experteninterviews*

---

<sup>16</sup> Die „Transparenzoffensive Immobilienwirtschaft“ ist ein von *ImmobilienScout 24* initiiertes offenes Forum für Akteure der Immobilienwirtschaft aus Forschung, Politik und Wirtschaft und möchte die Entscheidungsfindung von Marktteilnehmern durch bessere Informationen unterstützen.

Die Interviews wurden im Juli und August 2014 geführt. Insgesamt wurden fünf teilstandardisierte, qualitative Interviews mit einem Interviewleitfaden geführt (siehe Anhang 1). Zur Auswertung lag nach den Interviews ca. vier Stunden Audiomaterial vor. Der Aufbau der Interviews erfolgte unter Berücksichtigung der unterschiedlichen Rollen der Experten in unterschiedlichen Feldern von Immobilienportalen und der Immobilienwirtschaft. Die Interviewdauer betrug durchschnittlich 45 Minuten. Die Interviews enthielten Fragen eines allgemeinen Fragenblocks, der allen Experten in angepasster Weise gestellt wurde, und einen spezifischen, auf den jeweiligen Experten abgestimmten Fragenblock.



In Abbildung 3-2 erfolgt eine schematische Darstellung der Fragenblöcke.

Abb. 3-2: Aufbau und Fragenblöcke des Experteninterviews. (Entwurf und Darstellung: Schernthanner 2015).

### Auswertung und Ergebnisse

Zur Auswertung der Interviews wurden die von Meuser und Nagel (1991) vorgeschlagenen Auswerteschritte gewählt: a) Transkription, b) Paraphrasierung, c) Reduktion durch Überschriften, d) Bereichsspezifische Analyse des Materials, e) Thematischer Vergleich und f) Theoretische Generalisierung und Kategorienbildung. Nach der Transkription der Audiodaten lag ein 50-seitiges auswertbares Interviewtranskript zur Auswertung vor. Es wurden durch die Auswertung der Experteninterviews verschiedene Themenfelder, mit weiteren untergeordneten Problemfeldern, identifiziert. Die folgende Tabelle (Tabelle 3-2) zeigt die in den Experteninterviews identifizierten Themenfelder und deren jeweils untergeordnete Problemfelder. Problemfelder, bei denen davon auszugehen ist, dass Lösungsansätze mit den Methoden und Techniken der Geoinformation und Geovisualisierung gefunden werden können, wurden farblich hervorgehoben. Insgesamt wurden in den Experteninterviews 14 Problemfelder in vier Themenfeldern ausgemacht. Zwölf der Problemfelder haben einen klaren Forschungsbezug zur Geoinformatik und Geovisualisierung. Ein Problemfeld, das der „fach-

fremd verwendeten administrativen Grenzen“, fällt zu Teilen in den Bereich Geoinformation, zu Teilen in die Geovisualisierung.

Immobilienpreiskarten		Immobilienpreismodellierung	Daten	Angebotskarten	
Fachfremd verwendete Grenzen	Fachfremd verwendete Grenzen	Hedonische Regressionsmodelle	Datenschutz	Pin-Marker Problematik	
Legende		Datenpräprozessierung	Datenhomogenisierung	Zusätzliche Informationsebenen	
		Datendifferenzierung			
		Dichteschwankungen	Open Data	Indikatorenbasierte, distanzabhängige Suche	Visualisierung der Distanzen
		Lagemodellierung Ableitung Bestandspreis			
Geoinformatik			Geovisualisierung		

Tabelle 3-2: Identifizierte Themenfelder mit untergeordneten Problemfeldern. (orange: Geoinformation, grün: Geovisualisierung).

Nach der tabellarischen Übersicht der identifizierten Themen- und Problemfelder folgt eine analytische Zusammenfassung der von den Experten angesprochenen Themen- und Problemfelder.

### 1. Themenfeld: Immobilienpreiskarten

Das erste in den Interviews ausgemachte Themenfeld ist das Themenfeld Immobilienpreiskarte und umfasst mehrere Problemfelder:

#### Fachfremde, nicht sachgerecht angewandte räumliche Bezugseinheiten

Das erste dieser Thematik untergeordnete Problemfeld ist gleichzeitig das mit Abstand von allen Experten am häufigsten genannte. Hierbei handelt es sich um die Abbildung eines Lagepreises auf fachfremd, nicht sachgerecht angewandte räumliche Bezugseinheiten. Teilmärkte werden dadurch nicht klar strukturiert. Es gibt keine Einigung der Akteure der Immobilienwirtschaft, wie beispielsweise von Gutachterausschüssen und Portalbetreibern, auf einen einheitlich verwendeten Bereich, z. B. einen Postleitzahlenbereich. Derzeit werden laut der Interviews bei IS24 Analyseergebnisse auf der künstlichen, administrativen Ebene eines „Wohnquartiers“ der Firma Nexiga (vormals Infas Geodaten) abgebildet.

## Legende

Ein weiteres von den Experten genanntes Problemfeld im Themenfeld Immobilienpreiskarte ist das Fehlen dynamischer, adaptiver Legenden, die denselben Preisklassen zugeordnet werden.

## **2. Themenfeld: Immobilienpreismodellierung**

Das zweite in den Interviews identifizierte Themenfeld ist die räumlich-statistische Modellierung von Immobilienpreisen. In diesem Themenfeld wurden in den Interviews mit Abstand am meisten Problemfelder benannt. Es ergeben sich vielfach Überschneidungen mit dem Themenfeld „Immobilienpreiskarten“.

### Hedonische Regressionsmodelle

Das erste direkt diesem Themenfeld untergeordnete Problemfeld der hedonischen Regressionsmodelle steht in direktem Zusammenhang mit dem Themenfeld Immobilienpreiskarten sowie dem Problem der fachfremd angewandten räumlichen Bezugseinheiten. Es wurde in den Interviews die „Nicht-Räumlichkeit“ hedonischer Modelle kritisiert. Des Weiteren wurde kritisiert, dass Immobilienportale ihre Modelle nicht oder nur eingeschränkt veröffentlichen. Das wird von den Experten als ein Mangel an Transparenz gesehen.

### Datenpräprozessierung

Von den befragten Experten wird eingeschätzt, dass neuere Methoden aus dem Maschinellen Lernen und der Statistik ein großes Potential in der Datenvorverarbeitung von Immobilienpreisdaten aufweisen. Es wurden dabei die Methoden Random Forest (RF) und Support Vector Machines (SVM) aus dem Bereich Maschinelles Lernen sowie aus der Statistik die Methode der Hauptkomponentenanalyse genannt. Jedoch wurden bei den Methoden des Maschinellen Lernens Probleme in der Nachvollziehbarkeit der Modelle gesehen. Es wurde aber auch vor Phänomenen wie „Overfitting<sup>17</sup>“, verursacht durch das Rauschen der von einer Webseite stammenden Angebotsdaten, gewarnt. Bei der Hauptkomponentenanalyse, die in deskriptiver Hinsicht von Interesse ist, wird als mögliches Problemfeld eine „Zerstörung“ von Zusammenhängen in

---

<sup>17</sup> Überanpassung eines statistischen Modells an einen vorgegebenen Datensatz.

den Daten genannt. Die Hauptkomponentenanalyse (PCA) ist laut den Expertenaussagen teilweise bereits in die Modelle des Immobilienportals IS24 integriert. Als Forschungsansatz wird u. a. eine Verbesserung der Nachvollziehbarkeit der genannten Methoden und die Modellierung möglicher nichtlinearer Zusammenhänge genannt.

#### Dichteschwankungen

Das Problemfeld, das im Zusammenhang mit der Modellierung von Immobilienangebotsdaten am häufigsten angesprochen wurde, sind auftretende Dichteschwankungen von Angebotsdaten. Man hat es mit Angebotsdaten ganz unterschiedlicher Datendichten – von sehr hohen Datendichten im dicht verbauten urbanen Raum bis zu sehr geringen Datendichten im ruralen Raum – zu tun.

#### Lagemodellierung

Lageeigenschaften, wie z. B. die Nähe zu einer Grünfläche oder zu einer ÖPNV-Haltestelle, werden laut den Experten von vielen Akteuren unterschiedlich und sehr subjektiv behandelt. Eine bessere Lagemodellierung, z. B. in Form einer Lageapproximation von Wohnstandorten, könnte ein Teil der Preismodellierung werden. In diesem Problemfeld spielen auch Fragestellungen unterschiedlicher Lagemodellierungen bei variierenden Datendichten eine Rolle.

#### Datendifferenzierung

In den Interviews wurde mehrmals die Notwendigkeit einer differenzierten Betrachtungsweise von ermittelten Preisen bei Preiskarten nach Gesichtspunkten wie Bestand, Sanierungsgrad, Art der Bebauung und Strukturtyp angeführt.

### **3. Themenfeld: Daten**

Im Themenfeld Daten wurden von den Experten folgende Problemfelder angesprochen:

#### Datenschutz

Mehrfach wurde von den Experten die gebäudescharfe Abbildung von Preismodellierungsergebnissen auf Gebäudegeometrien erwähnt. Jedoch herrschten bei den Experten hierbei große datenschutzrechtliche Bedenken.

#### Datenhomogenisierung

In den Experteninterviews wurde deutlich, dass eine Zusammenführung von Abschlussdaten der Gutachterausschüsse mit Angebotsdaten der Immobilienportale im Fokus des Interesses von Akteuren der Gutachterausschüsse und von Immobilienportalen steht. Die Experten sehen, dass die derzeit bei Portalen

genutzte Gebäudeinventarisierung nicht die Qualität aufweist, die für diverse analytische Zwecke gebraucht wird. Die auf kommunaler Ebene angesiedelte Pflege von Gebäudedaten ist laut Expertenmeinung inhomogen und sehr unterschiedlich. Vielfach fehlen Attribute, die notwendig sind, um immobilienwirtschaftlich relevante Fragen oder auch Zukunftsfragen, wie die Energieeffizienz, zu beantworten. Das Problemfeld der Datenhomogenisierung geht Hand in Hand mit dem nächsten identifizierten Problemfeld, der Frage nach freien, offenen Daten.

#### Open Data

Es wird von den Experten des Immobilienportals *IS24* ein freier Zugang zu Katasterinformationen gefordert. Diese Forderung wird laut *Peter Ache* von den Gutachterausschüssen vielfach aufgegriffen, z. B. in Hamburg, wo es ein dementsprechendes Gesetz gibt, oder in Niedersachsen, wo ein solches in Vorbereitung ist. Wegen der föderalen Struktur der Gutachterausschüsse herrscht hier jedoch große Inhomogenität, was die Realisierung von Open-Data-Bestrebungen allgemein betrifft sowie die Frage, welche und wie viele Daten bereitgestellt werden.

### **4. Themenfeld Angebotskarten**

Ein weiteres Themenfeld ist das Themenfeld der Angebotskarten, in welchem sich ebenfalls mehrere Problemfelder ergeben.

#### Pin-Marker-Problematik

Bei einer ungefilterten Darstellung von Immobilienangeboten kommt es nicht selten zu einer extremen Häufung und teilweisen Überlappung von Pins bzw. Markern. Es gibt von den Experten Überlegungen, statt des herkömmlichen Filters einen maßstabsabhängigen "Kartenfilter" zu generieren, der z. B. Angebote je nach Maßstab über einen Generalisierungsansatz filtert. Jedoch werden bei dieser Idee Bedenken geäußert, dass die Nutzer der Immobiliensuche schon sehr an das Suchfeld von *Google* als Filter gewöhnt sind.

#### Zusätzliche Informationsebenen

Eine Unterbringung weiterer Informationsebenen, wie z. B. einer Preisinformation in die Angebotskarten, wird von allen Experten gewünscht. Beispielsweise wurde die Möglichkeit einer Farbkodierung für Pins angesprochen. Hierbei werden aber Schwierigkeiten einer gemeinsamen Darstellung von Objekteigenschaften, Ergebnissen der Preismodellierung und Lageinformationen in einer Karte gesehen.

#### Indikator-basierte, distanzabhängige Suche und Visualisierung der Distanzen

Diese Problematik zielt auf die Frage ab, wie sich klassische räumliche Selektionsfragen in die Immobiliensuche integrieren und visualisieren lassen. So wün-

schen sich die Experten unter anderem, dass Immobilienangebote in Abhängigkeit von wichtigen "Points of interest" (POIs) eines Suchenden angezeigt werden.

Zusammenfassend lässt sich zu den Experteninterviews sagen, dass das Problemfeld der fachfremd angewandten räumlichen Bezugseinheiten und in diesem Zusammenhang die Nicht-Beachtung räumlicher Zusammenhänge bei der Immobilienpreismodellierung mit Abstand als das dringlichste Problem identifiziert wurde. Dieses Problemfeld ist ein Problemfeld, das sowohl Thematiken der räumlichen Analyse als auch der Geovisualisierung tangiert. Als ein weiteres sehr großes Problemfeld wurde der Umgang mit unterschiedlichen Datendichten angeführt.

### 3.2.2 Die geoinformatische und geovisuelle Prozesskette von Immobilienportalen

Ein Ergebnis der im zweiten Kapitel erfolgten Bestandsaufnahme und der im dritten Kapitel vorgenommenen Auswertung der qualitativen Datenbasis ist die Darstellung der Prozesskette der geoinformatischen und geovisuellen Prozesskomponenten von Immobilienportalen. Diese wird im folgenden Graph (Abbildung 3-3) schematisch dargestellt.

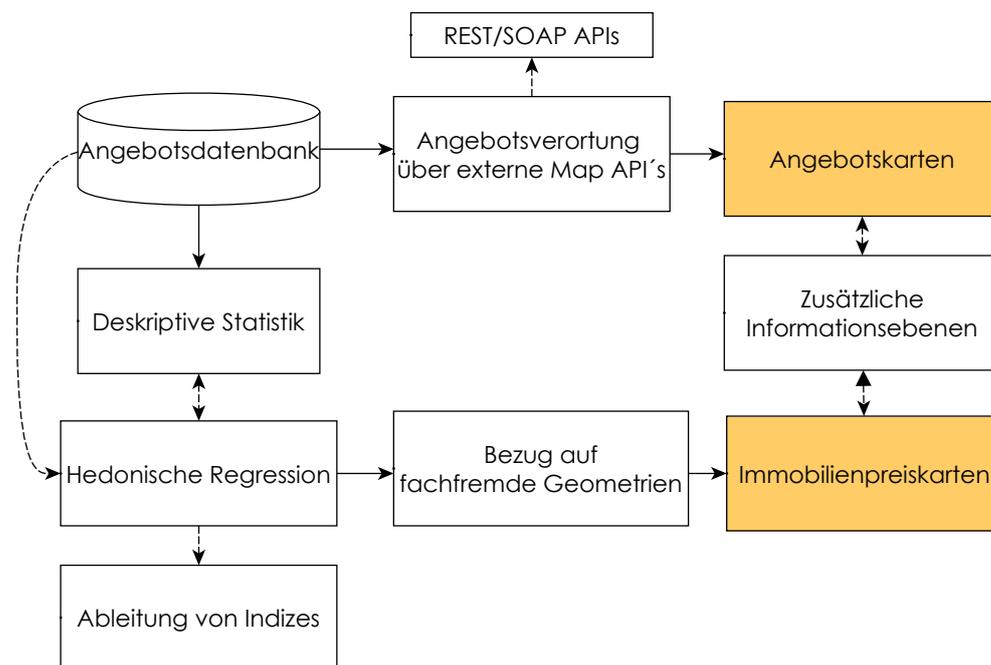


Abb. 3-3: Darstellung des Status quo der Prozesskette von Immobilienportalen. (Entwurf und Darstellung: Schernthanner 2015)

### 3.2.3 Räumliche Optimierungspotentiale

Ausgehend von der Bestandsaufnahme und dem Status quo werden räumliche Optimierungspotentiale von Immobilienportalen definiert. Was meint „Op-

timierungspotential“ im Zusammenhang mit dieser Arbeit? Es handelt sich um das „Ausloten“ der Potentiale der Prozessoptimierung aus geoinformatischer und geovisuelle Sicht. Optimierung ist ein sehr diffuser Begriff. Ganz allgemein bedeutet Optimierung eine Verbesserung bis hin zur Erreichung eines möglichen Optimums (*Wiktionary*, 2014). Hier stellt sich auch gleichzeitig das erste Problem im Zusammengang mit der vorliegenden Forschung: Was ist das "räumliche Optimum" eines Immobilienportals? Der Begriff Optimierung ist deshalb differenziert im Forschungskontext dieser Arbeit zu betrachten und wird nachfolgend charakterisiert. Es geht hierbei insbesondere darum, zu spezifizieren, in welcher Art und Weise der Optimierungsbegriff in dieser Arbeit angewendet wird.

Der Begriff Optimierung unterscheidet sich in seiner Terminologie sehr stark in diversen Wissenschaftsfeldern. In den Sozialwissenschaften determiniert Esser (2002) beispielsweise den Begriff Optimierung als eine Art Orientierung bei Handlungsalternativen des Individuums. In der Mathematik ist die Optimierung ein breites Feld, das sich laut *Oettli und Blum* (1975) mit der Extremwertermittlung einer Funktion befasst. Über die diversen Spielarten der Optimierung in der Mathematik gibt *Papageorgiou* (2012) in seinem Lehrbuch "Optimierungsverfahren" einen Überblick. In der Wirtschaftsinformatik findet die stochastische Optimierung (*Koberstein*, 2013) Anwendung, sie stellt eine Erweiterung der mathematischen Optimierung um Modelle und Methoden der optimalen Entscheidungsfindung dar.

Ein Teilgebiet der Optimierung, dessen sich die Geoinformatik des Öfteren bedient, ist die Heuristik. Optimierung im Sinne der Heuristik versucht Lösungen für reale Entscheidungsprobleme über Optimierungsmodelle zu finden. Ein klassisches Beispiel hierfür ist das "Traveling-salesman" Problem, bei dem ein Handlungsreisender in minimaler Netzwerkdistanz seine Kunden aufsuchen möchte (*Domschke & Scholl*, 2006). In der Heuristik steht die Suche nach guten, nicht unbedingt optimalen Lösungen für bestimmte Planungsprobleme im Vordergrund (*Pear*, 1984).

Im Rahmen dieser Arbeit wird der Begriff der Optimierung als eine „heuristische“ Optimierung angesehen. Eine „optimale“ Lösung für die raumanalytischen Unzulänglichkeiten kann nicht angeboten werden, sondern es können nur Lösungswege eingebettet in Anwendungsfälle aufgezeigt werden. Es handelt sich also in dieser Arbeit um eine Optimierung im Sinne einer geoinformatischen, heuristischen Prozessoptimierung der räumlichen Prozesse von Immobilienportalen.

Wie für den Begriff der Optimierung gibt es für den Begriff "Prozess" verschiedenste Definitionen. Ganz allgemein definiert sich Prozess als der Verlauf einer Entwicklung, als ein Ablauf oder auch als "Workflow" (*Duden*, 2014). Diese Arbeit bedient sich einer aus dem Geschäftsumfeld kommenden Terminologie, die einen Prozess als eine Abfolge von Tätigkeiten mit einem Start und einem Endpunkt sieht (*Geoinformatik Lexikon*, 2006). Für die einzelnen Teilkomponenten der bestehenden geoinformatischen und geovisuellen Prozesskette von Immobilienportalen werden bestehende Analysepotentiale definiert.



## 4 METHODEN

Zur Überprüfung der im Kapitel 1.3 formulierten Thesen gilt es, aus der Fülle existierender raumanalytischer und geovisualer Methoden geeignete Methoden auszuwählen und diese für die räumliche Optimierung von Immobilienportalen zu adaptieren. Das Finden geeigneter Methoden, deren Adaption und Teilimplementierung in den Anwendungskontext von Immobilienportalen, stellt eine wissenschaftliche Herausforderung dar. Auf Grundlage der Literaturrecherche (Kapitel 2.2), der technologischen Recherche (Kapitel 2.3) und der durchgeführten Experteninterviews (Kapitel 3.2.1) wurden unterschiedliche Methoden zur Teilimplementierung ausgewählt und für die Optimierung der räumlichen Analyse und Darstellung von Immobilienportalen in Betracht gezogen. Die Methoden werden in drei Methodengruppen eingeteilt: die Gruppe der räumlich-statistischen, die Gruppe der Verfahren des Maschinellen Lernens und die Methodengruppe der geovisualen Methoden. Im fünften Kapitel der Arbeit erfolgt eine Teilimplementierung der nachfolgend beschriebenen Methoden.

### 4.1 Methodengruppe der räumlich-statistischen Methoden

Die Literaturrecherche im zweiten Kapitel hat aufgezeigt, dass Methoden der Naturwissenschaften, in denen Kriging-Interpolatoren vorhanden sind, welche die räumliche Autokorrelation berücksichtigen, sehr verbreitet sind. Es stellt sich die Frage, ob das von *Krige* (1952) für den Bergbau entwickelte und von *Matherton* (1963) weiterentwickelte Verfahren auch zur Preisschätzung von Immobilien anwendbar ist. In Bereichen wie Hydrologie (*Laha et al.*, 2014), Bodenkunde (*Kerry et al.*, 2012) oder Epidemiologie (*Sampson et al.*, 2013) existieren zahlreiche Anwendungen des Kriging-Schätzverfahrens. Im Bereich der Immobilienmarktforschung gibt es hingegen bisher nur wenige Anwendungen von Kriging zur räumlichen Mietpreisschätzung. Für die Analyse von Immobilienpreisen kommt eine Vielzahl an nicht räumlichen Methoden zur Anwendung. Diese beruhen in der Regel auf der Annahme der statistischen Unabhängigkeit von Datenpunkten, was bei Immobilienpreisdaten jedoch nicht gegeben ist. Eine gemeinsame unumstößliche Charakteristik aller Immobilien ist die Gebundenheit an einen Ort. Die Lage eines Objektes hat also direkten Einfluss auf die Nutzung und den Wert einer Immobilie.

Werden Immobilienpreise mit herkömmlichen nicht räumlichen Methoden der Statistik behandelt, wird von der Annahme ausgegangen, dass jeder Datenwert zufällig, unabhängig von anderen vorliegenden Datenwerten zustande kommt. Bei beobachteten Immobilienpreisen würde das bedeuten, dass sich jeder Preis sowohl unabhängig von der Position als auch unabhängig von den benachbarten Immobilien und deren Nähe ergibt. Werden Immobilienpreise als statistisch unabhängige Zufallsvariablen betrachtet, erfolgt daraus eine Missachtung von *Toblers* erstem Gesetz der Geographie (*Tobler*, 1970), wonach alles mit allem, aber Näheres enger zusammenhängt.

Hieraus ist zu schlussfolgern, dass nicht räumliche Verfahren zur Mietpreisschätzung ungeeignet sind, da sie wichtige Einflussfaktoren räumlicher Charakteristika und Nachbarschaft nicht berücksichtigen.

#### 4.1.1 Regionalisierte Variablen

Wir haben es bei Immobilienangebotsdaten mit regionalisierten Variablen zu tun. Nach *Matherton (1963)* zeigen solche Variablen einen ganz konkreten Ortsbezug, z. B. über ein Koordinatensystem. Dieser Ortsbezug wird bei jeder Messung der Variablen über eine Koordinate mit aufgenommen. Eine regionalisierte Variable  $Z_{(i)}$  nach *Matherton (1963)* kann über drei Komponenten be-

schrieben werden:  $Z_{(i)} = f_{(i)} + s_{(i)} + \mathcal{E}$ , wobei  $f_{(i)}$  eine Trendkomponente ist,

$s_{(i)}$  eine zufällige lokale Ortsabhängigkeit, die es zu finden gilt und  $\mathcal{E}$  eine Fehlervarianz. Als wichtigste Vorüberlegung des raumanalytischen Anwendungsfalles müssen die vorhandenen Daten räumliche Autokorrelation aufweisen und die angewendeten Methoden diese räumliche Autokorrelation berücksichtigen.

Eine zu überprüfende Vorüberlegung ist, ob räumliche Verfahren existieren, die als Alternative zur hedonischen Regression Anwendung finden können. Hierzu existieren im Kontext von Immobilienportalen bisher keine Studien.

#### 4.1.2 Räumlichen Autokorrelation

Die Ermittlung der räumlichen Autokorrelation ist Vorbedingung in einer Vielzahl räumlich-statistischer Methoden zur räumlichen Optimierung der Mietpreisschätzung von Immobilienportalen. Das Konzept der räumlichen Autokorrelation besagt laut *Legendre (1993)*, dass bei räumlichen Prozessen die Ähnlichkeiten oder die Unterschiede der von einer gewissen räumlichen Distanz getrennten Umstände größer sind, als bei zufälligen Prozessen zu erwarten wäre. Autokorrelation kann zwei verschiedene Zustände, positive oder negative Autokorrelation, annehmen. Abbildung 4-1 von *Lo und Yeung (2002)* verdeutlicht an zwei extremen Beispielen positive und negative Autokorrelation im Vergleich zum Fehlen von Autokorrelation.

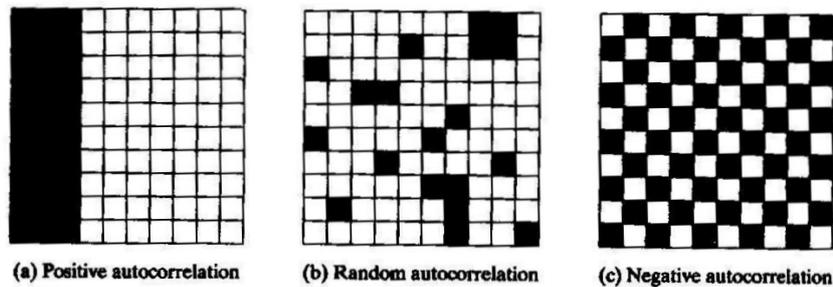


Abb. 4-1: Zustände der räumlichen Autokorrelation nach Lo & Yeung (2002).

Abeler *et al.* (1992) veranschaulichen das Konzept der Autokorrelation anhand eines Vergleichs der Bevölkerungsverteilung in Kalifornien und Iowa (Abbildung 4-2). Kalifornien weist hierbei eine stark positive und Iowa eine stark negative Autokorrelation auf.

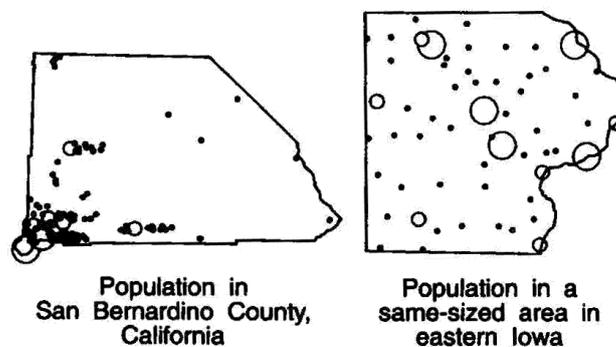


Abb. 4-2: Beispiel für positive und negative räumliche Autokorrelation nach Abeler *et al.* (1992).

Dubin (1998) untersuchte eingehend die räumliche Autokorrelation von Immobilienpreisen. Er veranschaulichte, dass die Lage, die er als "Nachbarschaft" bezeichnet, einen über die Autokorrelation messbaren Effekt auf Immobilienpreise hat. Das Phänomen der räumlichen Autokorrelation im Kontext der Immobilienpreisanalyse zeigt sich insbesondere im Verhalten von Mietern und Käufern. Deren Kaufverhalten wird maßgeblich von der Ähnlichkeit von Eigenschaften nah beieinanderliegender Immobilien und somit von deren Preis beeinflusst. Räumlich nah beieinanderliegende Immobilien teilen dabei ähnliche Lageparameter wie Erreichbarkeit oder Zugang zur Infrastruktur.

Wird die Stärke der räumlichen Autokorrelation gemessen? Hypothesentests der klassischen Statistik können durch eine fehlende Unabhängigkeit von Datenpunkten nicht angewendet werden (Legendre, 1993). In der räumlichen Statistik und der räumlichen Ökonometrie wurden jedoch Methoden zur Messung der räumlichen Autokorrelation entwickelt. So besteht eine Möglichkeit der Veranschaulichung der räumlichen Autokorrelation in der Erstellung von Variogrammen bzw. synonym Semivariogrammen (Legendre, 1993). Über sie kann die Beziehung von regionalisierten Variablen (Datenpunkten) zu Nachbarpunkten dargestellt werden.

Neben Variogrammen wurden lokal, d. h. für ein Subset der Datenpunkte, bezogene und globale, d. h. auf die Grundgesamtheit der Datenpunkte, anwendbare räumliche Hypothesentests zur Ermittlung der Stärke der Autokorrelation entwickelt. Zu den globalen Modellen zählen dabei die sehr ähnlichen Tests Moran's I und Geary's C. Beiden Tests gemeinsam ist die Nullhypothese. Diese lautet, dass die Daten zufällig im Raum verteilt sind bzw. dass keine räumliche Autokorrelation in den Daten vorhanden ist. Moran (1948) erläuterte erstmals die Möglichkeiten der Berechnung der räumlichen Autokorrelation. Formel 4-1 zeigt den von Moran aufgestellten Test;  $n$  ist hierbei die Anzahl der Beobachtungspunkte und das Gewicht für die Beziehung zwischen  $i$ -ten und  $j$ -ten Beobachtungspunkt ist  $W_{ij}$  (vgl. Li et al., 2007, Cliff und Ord, 1970).

$$I = \frac{n}{2A} \frac{\sum_{i=1}^n \sum_{j=1}^n z_i W_{ij} z_j}{\sum_{i=1}^n z_i^2}$$

Formel 4-1: Morans I Test (Moran, 1947)

$z_i$  steht für  $z_i = x_i - \bar{x} = x_i - \sum_{j=1}^n \frac{x_j}{n}$ . Dabei gilt  $x_i$  als Wert der  $i$ -ten Beobachtung

für die um den Mittelwert  $\bar{x}$  zentrierten Beobachtungen. Positive Werte  $> 0$  stehen für eine positive, negative Werte  $< 0$  für negative räumliche Autokorrelation.

Im Vergleich dazu zeigt Formel 4-2 die Berechnung des Korrelationskoeffizienten mittels Gearys C-Test (Geary, 1954). Für diesen Test liegt laut (Lo & Yeung, 2002) eine positive räumliche Autokorrelation vor, wenn  $c < 1$  ist. Ist  $c > 1$  ist die räumliche Autokorrelation negativ.

$$c = \frac{n-1}{4A} \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} (z_i - z_j)^2}{\sum_{i=1}^n z_i^2}$$

Formel 4-2: Gearys C-Test (Geary, 1954)

De Jong et al. (1984) geben eine ausführliche Übersicht über die beiden gängigen globalen Methoden räumlicher Hypothesentests.

Getis und Ord (1992) entwickelten ein weiterführendes lokales Verfahren zur Ermittlung der räumlichen Autokorrelation. Dieses Verfahren erlaubt es, nicht nur einen Korrelationskoeffizienten zu bestimmen, sondern ermöglicht darüber hinaus auch Clusteranalysen durchzuführen. Formel 4-3 zeigt die Berechnung der Autokorrelation nach Getis und Ord, wobei  $n$  die Anzahl der Beobachtungspunkte,  $x_i$  der Wert der  $i$ -ten Beobachtung und  $W_{ij}$  das Gewicht zwischen der  $i$ -ten und  $j$ -ten Beobachtung ist.

$$G = \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} x_i x_j}{\sum_{i=1}^n \sum_{j=1}^n x_i x_j} \quad \text{für alle } j \neq i$$

Formel 4-3: Korrelationskoeffizient nach *Getis Ord*

### 4.1.3 Hauptkomponentenanalyse (Principal component analysis, PCA)

Für mehrere im fünften Kapitel teilimplementierte Methoden ist es notwendig, geeignete Ko-Variablensets auszuwählen. Eine Methode, die als Vorverarbeitungsschritt zur Auswahl geeigneter Ko-Variablen angewendet wird, ist die Hauptkomponentenanalyse (Principal component analysis, PCA). Die Grundannahme der PCA besteht darin, dass die Richtung der Varianz in den Daten der größte Informationsträger ist. Beim Standardverfahren der Hauptkomponentenanalyse wird von linearen Beziehungen zwischen numerischen Variablen ausgegangen. Es erfolgt eine n-dimensionale Drehung des Datenraums in Richtung der größten Varianz. Die Richtung der größten Varianz wird zu den Koordinatenachsen des Datenraums. Unter den Hauptkomponenten werden die normal zueinander stehenden, nach absteigender Varianz geordneten neuen Achsen des Datenraums verstanden. Die maximale Varianz der Daten ist in der ersten Hauptkomponente ersichtlich. Die orthogonal zueinander stehenden Hauptkomponenten zeigen die geringer werdende Varianz in den einzelnen Hauptkomponenten.

Als Ergebnis werden sogenannte Ladungswerte erhalten. Ein hoher Ladungswert bedeutet, dass die Hauptkomponentenachse stark in die Richtung dieser Originalvariable zeigt. Die Originalvariable mit dem höchsten Ladungswert ist am wichtigsten für die Richtung der Hauptkomponente (*Kessler, 2007*). Je höher dabei der Absolutwert einer Komponentenladung ist, desto stärker ist auch der Zusammenhang des Indikators mit der jeweiligen Hauptkomponente. Werte nahe Null spielen bei der Erklärung der Hauptkomponente hingegen keine Rolle (*Schendera, 2010*).

### 4.1.4 Geographisch gewichtete Regression

Grundlage für die von *Brunsdon et al. (1996)* bzw. *Forsterhingham et al. (1996, 2002)* entwickelte Methode des GWR ist wie in der hedonischen Regressionsanalyse ein multivariates Regressionsmodell mit einer Erweiterung um einen zusätzlichen geographischen Gewichtungsfaktor. Neuere Anwendungen der GWR zur Preisschätzung finden sich in der Literatur beispielsweise bei *Geniaux et al. (2012)*, die eine Einführung in semi-parametrische Modelle für kombinierte „geoadditiv“ und GWR-Modelle geben. Laut *Lu et al. (2011)* und *Bourassa et al. (2010)* stellt die Anwendung von GWR gegenüber der hedonischen Regression eine Verbesserung der Immobilienpreisschätzung dar. Die räumliche Autokorrelation der Datenpunkte findet bei der GWR Beachtung. Einziger Unter-

schied zu einer linearen oder multiplen Regression ist ein in die Gleichung eingesetzter lokaler Gewichtungsfaktor, womit für im Raum verortete Punkte Aussagen getroffen werden können. Die Methode kann wie folgt (Formel 5) ausgedrückt werden:

$$y_i = \alpha_0(\mu_i, \nu_i) + \sum_k \alpha_k(\mu_i, \nu_i) X_{ij} + \varepsilon_i$$

Formel 4-4: Geographisch gewichtete Regression

wobei  $\mu_i, \nu_i$  die Koordinaten am Beobachtungspunkt darstellen,  $\alpha_0(\mu_i, \nu_i)$  und  $\alpha_k(\mu_i, \nu_i) X_{ij}$  geben den lokal geschätzten Parameter für die unabhängige Variable  $X_{ij}$  am Beobachtungspunkt  $i$  wieder und  $\varepsilon$  ist der Fehlerterm bzw. das Residuum am  $i$ -ten Beobachtungspunkt. Die geographisch gewichtete Regression (GWR) ist eine Erweiterung globaler Regressionsmethoden, wie z. B. der Methode der kleinsten Quadrate (OLS, Ordinary Least Squares) (Brundson et al., 1996; Fortheringham et al., 1996). Die geographische Lage eines jeden Beobachtungspunktes ist in der Regressionsgleichung integriert. Somit ist GWR in der Lage, für lokale, im geographischen Raum verortete Punkte Aussagen zu treffen. Räumlich nähere Beobachtungspunkte haben einen größeren Einfluss auf lokale Parameter. Die Gewichtung dieses Einflusses kann mittels einer Gauß'schen Distanz-Abnahme-Funktion wie folgt (Formel 4-5) ausgedrückt werden:

$$w_{ij} = \exp\left(-\frac{d_{ij}^2}{h^2}\right)$$

Formel 4-5: Gauß'sche Distanz-Abnahme-Funktion

#### 4.1.5 Räumliche Schätzverfahren, Interpolation

Generell wird unter Interpolation die Berechnung unbekannter Datenwerte verstanden, die zwischen bekannten Datenwerten liegen, bzw. das Finden von Näherungswerten für unbekannte Orte (Kappa, 2001). Ganz allgemein ist eine Interpolation also die Ermittlung eines unbekanntes Funktionswertes zu einem Argument, das zwischen zwei Argumenten mit bekannten Funktionswerten (Stützwerten) liegt“ (Geoinformatik-Service, 2002).

Es existiert eine Vielzahl an Interpolationsmethoden. Es gilt aus dieser Vielzahl an Interpolationsmethoden, Methoden auf ihre Anwendbarkeit im Zusammenhang mit dem raumanalytischen Anwendungsfall zu überprüfen. Entscheidend ist also die Wahl der geeignetsten Interpolationsmethode für die Mietpreisschätzung. Die Methodenauswahl hängt von verschiedenen Phänomenen ab (Vassilopoulos, 2008):

1. Räumlicher Verbreitung des Phänomens
2. Methode der Datenerfassung
3. Datendichte und -homogenität

Je nach Anwendung ist es notwendig, verschiedene Interpolationsverfahren auf einem Datensatz zu testen sowie abschließend eine Bewertung und Validitätsprüfung der Methoden durchzuführen. Erst nach erfolgter Methodenauswahl kann eine Feinabstimmung der Parametrisierung vorgenommen werden (Vassilopoulos, 2008).

Diesem Ansatz folgend werden Interpolationsverfahren im raumanalytischen Anwendungsfall nach dem in Abbildung 4-3 dargestellten Schema konzeptionell diskutiert. In Kapitel 5 erfolgen die Teilimplementierung und Validierung der Methoden, um schlussendlich eine Aussage über ein für Immobilienportale geeignetes Interpolationsverfahren zu treffen.

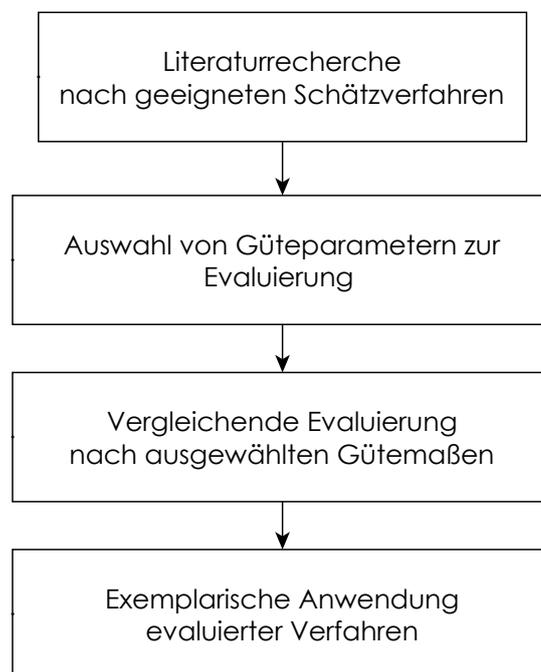


Abb. 4-3: Schema der Auswahl, Validierung und Teilimplementierung von Schätzverfahren. (Entwurf und Darstellung, Schernthanner: 2015).

Zur Unterstützung der Methodenauswahl erfolgt eingangs eine erklärende Einteilung der unterschiedlichen Interpolationsarten. Sehr gute Einführungen zu räumlichen Schätzverfahren finden sich in Bivand et al. (2013), Li und Heap (2008) sowie Chilés und Delfiner (1999).

### *Globale und lokale Interpolationsmethoden*

Der erste grundlegende Unterschied zwischen den diversen Interpolationsmethoden besteht in ihrer Charakteristik als lokale bzw. globale Methoden. Globale Methoden nutzen alle vorhandenen Datenpunkte für die Interpolation, lokale Methoden nur eine Teilmenge der vorhandenen Beobachtungspunkte für eine möglichst exakte Wertschätzung. Zur Wertschätzung werden lokale

Methoden eingesetzt. Globale Methoden werden zur Bewertung globaler Strukturen verwendet (Albrecht, 2007).

### Exakte und nicht exakte Interpolationsmethoden

Eine weitere Differenzierung liegt in der Unterscheidung zwischen exakten und nicht exakten Methoden. Bei exakten Methoden verläuft die interpolierte Fläche genau durch den Beobachtungspunkt, bei der nicht exakten Methode wird auch für die bekannten Beobachtungspunkte ein Wert geschätzt und somit eine glatte Oberfläche gebildet (Albrecht, 2007). Abbildung 4-4 illustriert den Unterschied zwischen exakten und nicht exakten Interpolationsmethoden.

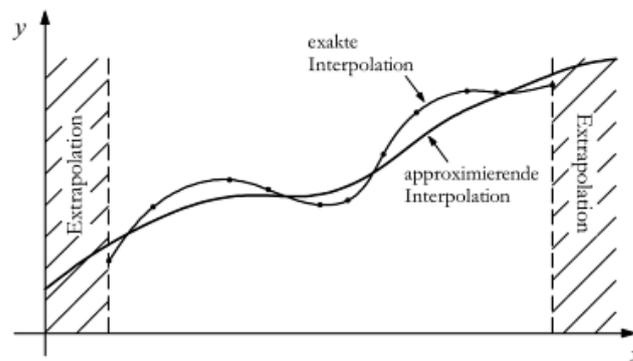


Abb. 4-4: Exakte, nicht exakte Interpolation (Niemeier, 2008).

### Deterministische Interpolationsverfahren

Deterministische Interpolatoren basieren auf vorab festgelegten Funktionen. Richtungsspezifische Informationen (Anisotropie) finden keinen Eingang in das Schätzverfahren. Die bekanntesten deterministischen Interpolatoren sind Spline und Natural Neighbor. Das wohl in der Geoinformatik am häufigsten verwendete Verfahren ist die Methode IDW. IDW verfügt zwar über den Vorteil, dass die Berechnung sehr schnell möglich ist. Als Nachteile dieser Methode entstehen jedoch rundliche Artefakte rund um Inverse distance weighted (IDW), bekannte Datenpunkte und es wird keine Anisotropie beachtet. De Lange (2006) gibt eine umfassende Erklärung der Methode IDW.

### Geostatistische Interpolationsmethoden

Die Beschreibung der Methoden ist im Rahmen dieser Arbeit als eine Diskussion der Grundlagen im Kontext des raumanalytischen Anwendungsfalles zu verstehen. Die Komplexität der Kriging-Schätzer, mit allen Teilaspekten, erschließt sich vollends durch die Konsultation der im Folgenden genannten Quellen.

Geostatistische Methoden analysieren nicht nur die Dimensionen im Merkmalsraum der Daten, sondern auch im geographischen Raum (Rossiter, 2012; Wackernagel, 2003) sowie Wertausprägungen benachbarter Datenpunkte (Li und Heap, 2008). Richtungsspezifische Informationen (Anisotropie) in den Daten

werden explizit betrachtet. Laut *Chiles und Delifiner (2012)* weisen Methoden aus der Methodenfamilie des Kriging klare Vorteile gegenüber deterministischen Methoden auf, da die Möglichkeit besteht, die Kovarianz von Raumausschnitten mit vielen Beobachtungspunkten auf Raumausschnitte mit wenigen Beobachtungspunkten zu übertragen. Diese Eigenschaft macht Kriging-Ansätze für Datensätze mit unterschiedlichen Datendichten und unregelmäßiger Punkteverteilung, wie sie bei Immobilienangebotsdatensätzen üblich sind, anwendbar.

Kriging-Verfahren gehen auf den südafrikanischen Bergbauingenieur *Krige (1951)* zurück. Zur Schätzung von Mineralressourcen entwickelte *Krige* die Grundlagen der heutigen Geostatistik. Der französische Mathematiker und Geologe *Matherton (1963)* stellte die Theorie der regionalisierten Variablen (siehe Kapitel 4.1.1) auf und entwickelte mittels dieser Theorie *Kruges* Verfahren weiter.

Verfahren aus der Kriging-Methodenfamilie sind heute hauptsächlich im Forschungsfeld und in Anwendungsbereichen der Geowissenschaften/Geologie zu finden. Beispielsweise werden Kriging-Schätzer nach wie vor eingesetzt, um Rohstoffvorkommen zu schätzen. Eine weitere sehr häufig zu findende Anwendung ist die Erstellung von Temperaturkarten oder Niederschlagskarten in der Meteorologie. Anwendungen im Bereich der Immobilienwirtschaft finden sich nur in der Forschung.

Alle Verfahren aus der Kriging-Familie haben gemeinsam, dass es sich um minimalvarianter, linearer und erwartungstreuer Schätzer handelt. Die Schätzer sind Regressionsalgorithmen nach dem verallgemeinerten Kleinste-Quadrate-Ansatz (*Sarma, 2009; Şen, 2009*). Die Kleinste-Quadrate-Regression minimiert zur Beschreibung der Verteilung von Datenpaaren die quadratische Abweichung der Modellkurve zu den Datenpunkten (*Hedderich und Sachs, 2012*). Kriging ist ein exaktes Interpolationsverfahren, das den besten unverzerrten Schätzwert liefert, d. h., es gibt keine systematischen Verzerrungen. Der Erwartungswert des Schätzers beträgt bei einem unkorrelierten Fehler und konstanter Varianz 0.

Um ein Schätzverfahren aus der Kriging-Methodenfamilie anwenden zu können, müssen mehrere Bedingungen erfüllt sein (*Harvey, 2008*):

1. Räumliche Autokorrelation muss gegeben sein.
2. Eine Normalverteilung der Daten ist zwar nicht grundlegend notwendig, jedoch zeigen sich bei normalverteilten Daten bessere Schätzergebnisse. Über Techniken wie eine logarithmische Transformation oder Box-Cox-Transformation von Daten können nicht normalverteilte Daten annähernd in eine Normalverteilung gebracht werden.

3. Allein durch den Abstand der Datenwerte und nicht durch die absolute Position im Raum ergibt sich der räumliche Zusammenhang der Daten. Dementsprechend besitzen Wertepaare mit demselben Abstand dieselbe Varianz (Stationarität).

### Variographie

Wichtigster Bestandteil aller Schätzverfahren der Kriging-Familie ist die Variographie. Unter Variographie wird die Modellierung des räumlichen Zusammenhangs zwischen Attributwerten benachbarter Beobachtungspunkte in Abhängigkeit von ihrer Entfernung verstanden. Mittels der Variographie wird also die Semivarianz zwischen zwei Datenpunkten analysiert und daraus ein Variogramm erstellt, welches den funktionellen Zusammenhang zwischen den Datenpunkten zeigt. Das Semivariogramm gibt Aufschluss über den Werteunterschied der Beobachtungspunkte in Distanzabhängigkeit und lässt sich vereinfacht ausdrücken durch (Formel 4-6):

$$\frac{(Z_{(x+h)} - Z_x)^2}{2}$$

Formel 4-6: Semivariogramm

wobei  $Z$  den beobachteten Wert und  $h$  den Distanzunterschied darstellt.

Eine bestmögliche Parameterwahl ist entscheidend für die Schätzung von Angebotspreisen. Im Zusammenhang mit der Schätzung von Immobilienpreisen ist dies eine noch ungeklärte Forschungsfrage.

Mittels des Semivariogramms lassen sich folgende Parameter ablesen:

1. Range: Distanzwert, der angibt, bis zu welcher Entfernung ein räumlicher Zusammenhang zwischen den einzelnen Werten besteht. Eine Abflachung der Kurve zeigt dabei eine Abflachung des räumlichen Zusammenhangs an.
2. Sill: zeigt die Gesamtvarianz der Daten an.
3. Partial Sill: gibt die Differenz zwischen Sill und Range an.
4. Nugget: Maß für die Streuung der Daten, die beispielsweise durch Messfehler entstehen kann.

Folgende Abbildung 4-5 zeigt exemplarisch ein Semivariogramm nach *Li und Heap* (2008).

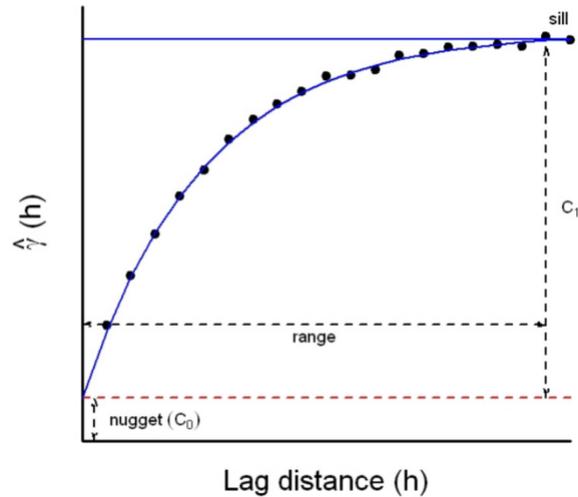


Abb. 4-5: Beispiel für ein Semivariogramm eines exponentiellen Modells nach Li und Heap (2008).

Eine Möglichkeit, die Semivarianz der Daten zu analysieren, ergibt sich laut Li und Heap (2008) durch (Formel 4-7):

$$\gamma(x_i, x_0) = \gamma(h) = \frac{1}{2} \text{var} [Z(x_i) - Z(x_0)]$$

Formel 4-7: Berechnung der Semivarianz nach Li und Heap (2008)

Mittels der Formel wird die Semivarianz  $\gamma(h)$  für eine Wertauswahl bzw. Intervalle verwendet.  $Z(x_i)$  steht hierbei für den Wert am Punkt  $x_i$ ,  $n$  für die Anzahl der Punktpaare  $(x_i, x_0)$  mit  $x_i - x_0 = h$ . Diese Möglichkeit ist isotropisch. Das entstehende empirische Variogramm kann durch einfache Modellfunktionen wie z. B. Gauß'sche, Nugget-, sphärische, exponentielle oder Power-Modelle repräsentiert werden (Li und Heap, 2008; zitiert nach: Burrough and McDonnell, 1998; Pebesma, 2004; Webster und Oliver, 2001). Abbildung 4-6 verdeutlicht den Verlauf der vier oft verwendeten Modellfunktionen: sphärische, exponentielle, lineare und Gauß'sche Modelle.

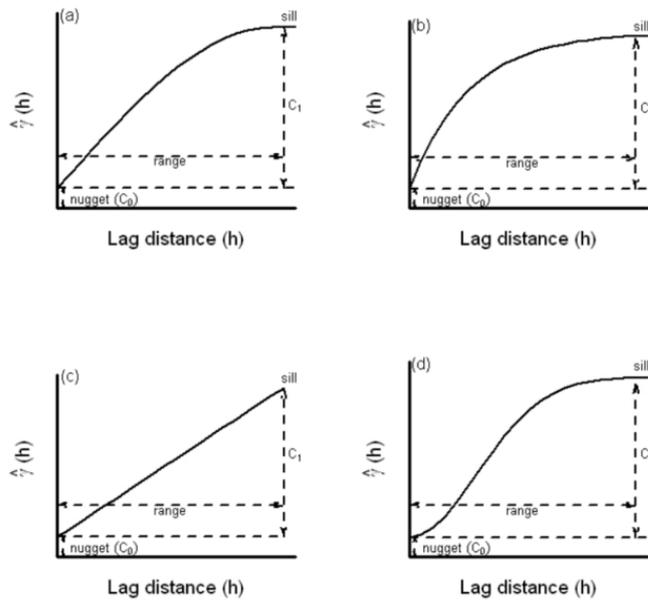


Abb. 4-6: sphärische, exponentielle, lineare und Gauß'sche Modelfunktion nach Li und Heap (2008).

Mittels der Variographie muss, unabhängig von der gewählten Kriging-Methode, geklärt werden, bis zu welcher Distanz ein räumlicher Zusammenhang zwischen den Angebotsdaten besteht. Hierbei ist zu beachten, dass nur bei einem sphärischen Modell der tatsächliche Range bzw. die tatsächliche Reichweite angegeben wird. Bei Gauß'schen und exponentiellen Modellen hat man es theoretisch mit einer rechnerisch unendlichen Reichweite zu tun. Eine rechnerische Reichweite, genannt effektive Reichweite, kann jedoch angegeben werden; ab dieser bleibt das funktionale Verhalten annähernd gleich.

### Unterschiedliche Kriging Schätzverfahren

Neben der Variographie, der Auswahl der Parameter und der Wahl der Sekundärvariablen ist die Auswahl des geeigneten Kriging-Schätzverfahrens ein wichtiger Bestandteil der Mietpreisschätzung im Rahmen des raumanalytischen Anwendungsfalles. Die folgende Abbildung 4-7 zeigt den schematischen Ablauf der Kriging-Interpolation mit allen wichtigen Teilschritten.

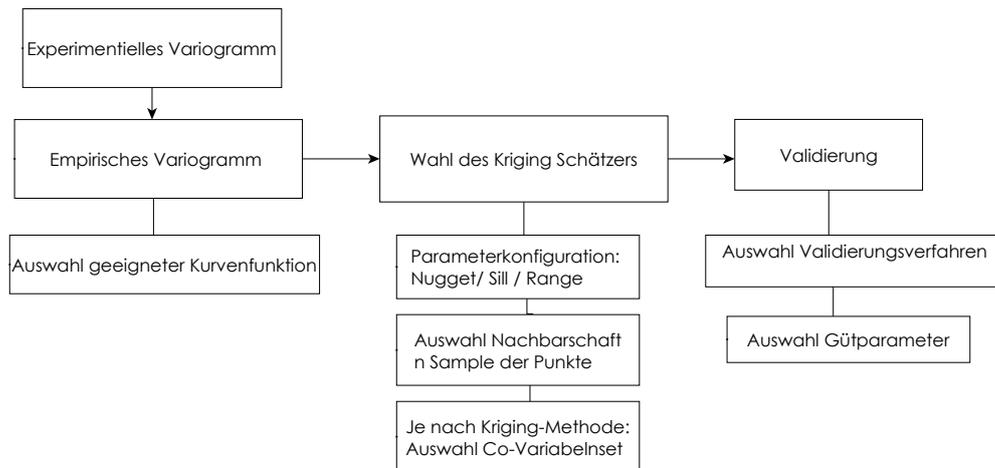


Abb. 4-7: Ablauf des Kriging Interpolationsverfahrens. (Entwurf und Darstellung: Scherthanner 2015).

Geostatistische Schätzverfahren wurden in der Literatur der Forschungsfelder der Geoinformatik (Burrough und McDonnell, 1998), der Geostatistik (Bivand, 2008, Chiles und Delifiner, 1999; Wackernagel, 1998) und der Geowissenschaften (Li und Heap, 2008, Goovaerts, 1997) ausführlich diskutiert. Die Theorie der Kriging-Schätzer und ausgewählter Methoden der Methodenfamilie des Krigings werden nachfolgend zusammenfassend diskutiert. Diese Diskussion stützt sich auf die oben genannten Quellen. Zwar finden sich in der Immobilienmarktforschung Lösungsansätze zur Schätzung von Immobilienpreisen, im operativen Umfeld der Immobilienwirtschaft gibt es bis dato hingegen keine bekannte Anwendung des Kriging-Verfahrens.

### Kriging-Schätzer

Grundlage eines jedes Kriging-Schätzverfahrens ist der Kriging-Schätzer, mit dem der Wert an der Position  $Z(x_0)$  geschätzt wird (siehe Formel 4-8). Jedes weitere Kriging-Schätzverfahren stellt eine Modifikation dieses grundlegenden Kriging-Schätzers dar, wobei  $n$  für die Anzahl der in die Schätzung mit einfließenden Datenpunkte, abhängig von der verwendeten Suchnachbarschaft, steht.  $\lambda_i$  steht für das Kriging-Gewicht und  $\mu(x_0)$  ist der lokale Mittelwert der Datenpunkte innerhalb der gewählten Suchnachbarschaft. Die Gewichtung ist hierbei von der relativen Lage der Datenpunkte abhängig und die Gewichte sind so zu berechnen, dass die Schätzvarianz minimiert wird. Der Kriging-Schätzer besteht weiter aus einer Trendkomponente  $\mu$  und Residuenkomponente  $R = (Z - \mu)$ . Der Umgang mit der Trendkomponente ist entscheidend für die Auswahl der Kriging-Methode (Li und Heap, 2008; Bivand et al. 2008).

$$Z(x_0) = \sum_{i=1}^n \lambda_i [Z(x_i) - \mu(x_0)] + \mu$$

Formel 4-8: Der Kriging-Schätzer

Es existieren unterschiedliche Methoden und Abwandlungen von Kriging-Schätzverfahren. Basierend auf Lösungsansätzen in der Literatur (siehe Kapitel 2.2) werden nachfolgend die drei Methoden Ordinary Kriging (OK), Kriging with external drift (KED) und Ordinary Cokriging (OCK) näher vorgestellt. Es handelt sich dabei um jene drei Methoden, die im Rahmen der Teilimplementierung des raumanalytischen Anwendungsfalles ausgewählt wurden (vgl. Auswahl der Schätzverfahren: Kapitel 5.1.5).

### Ordinary Kriging

OK stellt die Basis für eine Reihe anderer Kriging-Methoden dar. Die globale Trendkomponente  $\mu$  (siehe Abbildung 38) wird durch das lokale Mittel  $\mu(x_0)$ , den Mittelwert der Datenpunkte innerhalb der gewählten Suchnachbarschaft, ersetzt. Das ist der grundlegende Unterschied zur Methode Simple Kriging. OK funktioniert unter der Prämisse, dass sich die Kriging-Gewichte, ausgedrückt durch  $\sum_{i=1}^n \lambda_i = 1$  zu 1, aufsummieren lassen.

Ordinary Kriging nutzt zur Schätzung nur Informationen aus den Primärvariablen, dabei wird der lokal konstante Mittelwert geschätzt und auf den sich ergebenden Residuen ein Simple Kriging angewendet (Li und Heap, 2008, zitiert nach Goovaerts, 1997).

### Kriging with external drift

Kriging with external drift (synonym mit Universal Kriging) stellt eine Erweiterung des „Kriging-mit-Trend-Verfahrens“ dar, das bereits von Matherton (1969) vorgeschlagen wurde. Anstelle der Nutzung eines lokalen Mittelwerts wird eine lineare Regression definierter Sekundärvariablen innerhalb der festgelegten Suchnachbarschaft durchgeführt. Die Abweichung des Regressionsergebnisses, also die Residuen der linearen Regression, werden mittels Simple Kriging interpoliert.

$$\sum_{i=1}^n \lambda_i q_k(x_i) = q_k(x_0) \quad \text{für } k = 1, \dots, m$$

Formel 4-9: KED-Schätzer

$q_k$  ist der Wert der k-ten Sekundärvariable am zu schätzenden Punkt  $x_0$  bzw. am Stützpunkt  $x_i$ .  $m$  ist die Anzahl der Sekundärvariablen.

Für KED gilt, dass die Sekundärvariablen an allen Beobachtungs- und an allen zu schätzenden Punkten vorhanden sein müssen.

## Ordinary Cokriging

Ordinary Cokriging (OCK) basiert auf der Verwendung von Kreuzvariogrammen, die eine Erweiterung des Variogramms darstellen. Während Variogramme nur Auskunft über räumliche Informationen einer Variable geben, kann über Kreuzvariogramme zusätzlich Auskunft über die räumliche Korrelation von zwei Variablen erhalten werden. Die Kreuz-Semivarianz wird mittels der Formel 4-10 berechnet.

$$\hat{\gamma}_{12}(h) = \frac{1}{2n} \sum_{i=1}^n [z_1(x_i) - z_1(x_i+h)][z_2(x_i) - z_2(x_i+h)]$$

Formel 4-10: Berechnung der Kreuz-Semivarianz nach *Borrough und McDonnell (1996)*.

$n$  ist hierbei die Anzahl der Punktpaare der Variablen  $z_1$  und  $z_2$  mit dem Abstand  $h$ .

Der Kriging-Schätzer wird bei OCK zu Formel 4-11 erweitert.

$$\hat{z}_1(x_0) = \sum_{i=1}^{n_1} \lambda_i [Z_1(x_{i_1}) - \mu_1(x_{i_1})] + \sum_{j=2}^{n_v} \sum_{i_j=1}^{n_j} \lambda_{i_j} [Z_j(x_{i_j}) - \mu_j(x_{i_j})] + \mu_1$$

Formel 4-11: OCK

$\mu_1$  = Stationärer Mittelwert der Zielvariablen

$Z_1(x_{i_1})$  = Wert der Zielvariable an Punkt  $x_{i_1}$

$\mu_1(x_{i_1})$  = Mittelwert der Zielvariablen Einflussradius

$n_1$  = Anzahl der Beobachtungspunkte im Einflussradius  $x_0$

$\lambda_{i_1}$  = Gewicht zur Minimierung der Varianz der Zielvariable

$n_v$  = die Anzahl der Sekundärvariablen.

$n_j$  = Anzahl der Beobachtungspunkte der  $j$ -ten Sekundärvariable innerhalb des Einflussradius  $\lambda_{i_j}$

des zum  $i_j$ -ten Beobachtungspunkt der  $j$ -ten Sekundärvariable gehörende Gewicht

$Z_j(x_{i_j})$  = Wert der  $j$ -ten Sekundärvariable an dem  $i_j$ -ten Beobachtungspunkt

Der wesentliche Vorteil von OCK liegt darin, dass Primär- und Sekundärvariablen in unterschiedlicher räumlicher Dichte vorliegen können und dass nicht alle Sekundärvariablen aller Beobachtungs- und Modellierungspunkte vorliegen müssen.



## Auswahl der Schätzverfahren

Zur endgültigen Auswahl der Schätzverfahren dient eine tabellarische Übersicht (Tabelle 4-1). Hieraus ersichtlich sind die einzelnen Vor- und Nachteile der vorgestellten geostatistischen Interpolationsmethoden.

	Verfahren	Vorteile	Nachteile
Deterministische Interpolationsverfahren	IDW	<ul style="list-style-type: none"> <li>Schnelle Trendbestimmung. Distanzgewichtung ist möglich.</li> </ul>	<ul style="list-style-type: none"> <li>Bull's eye-Effekte.</li> <li>Keine Beachtung von Anisotropie.</li> <li>Keine Nutzung von Sekundärvariablen.</li> </ul>
	Splines	<ul style="list-style-type: none"> <li>Bildung glatter Trendoberflächen.</li> </ul>	<ul style="list-style-type: none"> <li>Keine Beachtung von Anisotropie.</li> <li>Keine Nutzung von Sekundärvariablen.</li> </ul>
	Natural Neighbor	<ul style="list-style-type: none"> <li>Schnelle Trendbestimmung.</li> </ul>	<ul style="list-style-type: none"> <li>Keine Beachtung von Anisotropie.</li> <li>Keine Nutzung von Sekundärvariablen.</li> </ul>
Geostatistische Interpolationsverfahren	OK	<ul style="list-style-type: none"> <li>Beachtung richtungsabhängiger Information.</li> <li>Untersuchung räumlicher Autokorrelation.</li> </ul>	<ul style="list-style-type: none"> <li>Keine Beachtung von Sekundärvariablen.</li> </ul>
	KED	<ul style="list-style-type: none"> <li>Beachtung von Sekundärvariablen.</li> <li>Untersuchung räumlicher Autokorrelation.</li> </ul>	<ul style="list-style-type: none"> <li>Sekundärvariablen müssen an allen Beobachtungspunkten vorhanden sein.</li> </ul>
	OCK	<ul style="list-style-type: none"> <li>Beachtung von Sekundärvariablen.</li> <li>Untersuchung räumlicher Autokorrelation.</li> <li>Ein Fehlen von Sekundärvariablen ist möglich.</li> </ul>	<ul style="list-style-type: none"> <li>Lange Rechenzeiten.</li> </ul>

Tabelle 4-1: Vor- und Nachteile deterministischer und geostatistischer Interpolationsverfahren.

Hilfreich für die Auswahl eines geeigneten Interpolators, je nach Vorhandensein eines Variogramms bzw. Trends, ist der von *Bivand* (2008) aufgestellte Entscheidungsbaum zur Wahl eines Schätzers in der R-Bibliothek *gstat* (Abbildung 4-8).

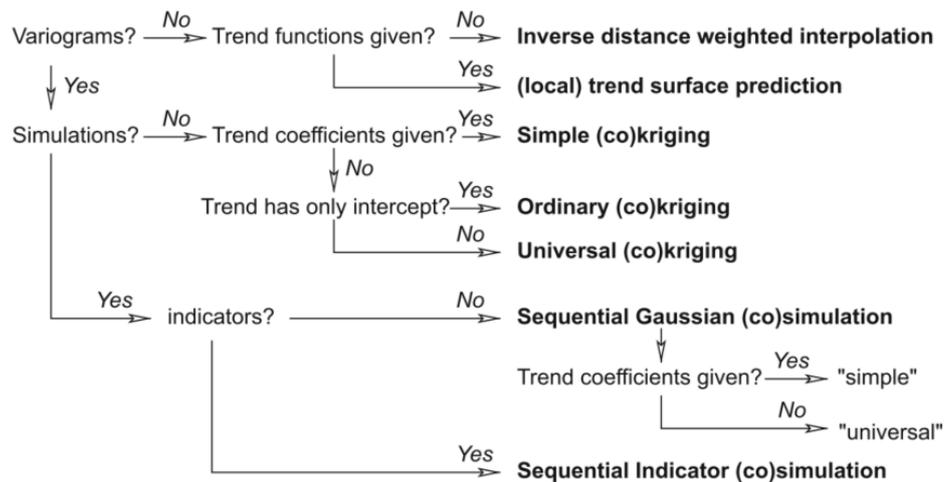


Abb. 4-8: Entscheidungsbaum von Bivand et al. (2008) zur Auswahl eines geeigneten Schätzverfahrens.

Die Verfahren OK/KED und OCK werden ausgewählt und dem Standardverfahren der hedonischen Regression gegenübergestellt. Eine Einschränkung in Bezug auf die Verwendung der Verfahren KED und OCK liegt darin, dass die Verfahren keine kategorialen Variablen zur Schätzung nutzen können. Laut Hengl (2009) sind Interpolationsverfahren für die Anwendung auf numerische Variablen ausgelegt. Zwar gibt es laut Hengl (2009) auch Verfahren, die kategoriale Variablen bei der Schätzung berücksichtigen können, diese Verfahren unterliegen jedoch unterschiedlichen großen computertechnischen Problemen. Kategoriale Variablen sind nur in einer geringen Anzahl vorhanden (16 von 64) und spielen bei der Auswahl der Sekundärvariablen lediglich eine geringe Rolle, wie in Kapitel 5 dargestellt wird.

## 4.2 Methodengruppe der Verfahren des Maschinellen Lernens

Die zweite Methodengruppe, die für eine Teilimplementierung zur räumlichen Optimierung der Mietpreisschätzung von Portalen in Frage kommt, ist die Methodengruppe des Maschinellen Lernens. Beim Maschinellen Lernen wird versucht anhand der Eigenschaften der vorhandenen Daten und mittels Trainingsdaten Vorhersagen zu treffen. Verfahren aus dem Maschinellen Lernen nutzen dabei das überwachte Festlegen von Trainingsdaten, wohingegen beim Datamining das unüberwachte Auffinden von Eigenschaften im Datensatz im Vordergrund steht (Fürnkranz et al., 2012). Exemplarisch werden im fünften Kapitel zwei Verfahren auf den Angebotsdatensatz angewendet. Dabei handelt es sich zum einen um den von Breiman (2001) entwickelten Random-Forest-Ansatz und zum anderen um den Ansatz der Neuronalen Netze.

## 4.2.1 Random Forest (RF)

Der von *Breiman* (2001) entwickelte Algorithmus "Random Forest" (RF) ist eine Weiterentwicklung herkömmlicher einfacher Entscheidungsbäume. RF trifft Vorhersagen aus einer Vielzahl zufällig (randomisiert) zu einem „Wald“ gewachsener Entscheidungsbäume. Die "Entscheidung" über das endgültige Resultat eines Random-Forest-Modells wird über eine Anzahl an willkürlich gewachsenen Entscheidungsbäumen bestimmt (*Statsoft*, 2014). Es gibt nur wenige Studien, die RF als Modell zur Schätzung von Mietpreisen nutzen. Eine der wenigen dieser Anwendungen ist eine vergleichende Untersuchung mehrerer Algorithmen des Maschinellen Lernens von *Antipov und Pokryshevskaya* (2010). Das Ergebnis ihrer vergleichenden Modellierung von Wohnungspreisen der Stadt St. Petersburg zeigt, dass sich RF-Modelle im Vergleich mit anderen Modellen, wie z. B. Künstlichen Neuronalen Netzen, als sehr robust gegenüber Ausreißern und fehlenden Datenwerten erwiesen. Beim vorliegenden exemplarischen Angebotsdatensatz hat man es vielfach mit Ausreißern und Lücken in den Daten zu tun, womit RF als Methode der Mietpreismodellierung Potentiale aufweist. Random Forest ist laut *Breiman* (2001) ein CART (Classification and Regression Tree) und kann folglich in zwei Fällen angewendet werden. Das ist im ersten Fall die Klassifikation unabhängiger Variablen anhand der in einer abhängigen kategorialen Variablen vorhandenen Klassen bzw. Kategorien. Der Algorithmus weist den einzelnen Bäumen Klassen zu, das "Ensemble" an Bäumen bestimmt dann per Mehrheitsentscheidung, welche Klasse die größte Stimmanzahl aufweist (*Lüthy*, 2009). Der zweite Fall der Anwendung, wie sie im Kontext der Mietpreisschätzung denkbar ist, ist die Regression einer unabhängigen Variable über die zufällig gebildeten Entscheidungsbäume. Im Ergebnis ergibt sich ein gemittelter Schätzwert einer kontinuierlichen Zielvariablen.

Mehrere Faktoren beeinflussen das Random-Forest-Modell:

- die Art des RF-Modells (Klassifikation oder Regression),
- die Tiefe des RF oder dessen Baumanzahl,
- das Pruning, als die Beschneidung der Bäume auf eine vernünftige Größe<sup>18</sup>,
- die Variablenanzahl sowie

---

<sup>18</sup> Pruning wird von *Breiman* (2001) nicht empfohlen, wird in der Praxis jedoch teilweise angewendet.

- die Arbeit mit oder ohne Stichprobe.

Eine dritte Möglichkeit der Anwendung liegt in der Datenvorverarbeitung zum Finden möglicher Sekundärvariablen, eine Teilimplementierung dieser Art der Anwendung wird in Kapitel 5.1.2 diskutiert.

Des Weiteren bemerken *Antipov und Pokryshevskaya (2010)*, dass die Konfiguration des RF-Algorithmus hinsichtlich z. B. der geeigneten Variablenauswahl äußerst wichtig ist und dass für die Wahl der optimalen Parameter weiterhin Forschungsbedarf besteht.

## 4.2.2 Künstliche Neuronale Netze

Als zweites Verfahren des Maschinellen Lernens wurde exemplarisch das Verfahren der *Künstlichen Neuronalen Netze (ANN)* angewendet. Künstliche Neuronale Netze (Artificial neural network) sind laut *Zell (1994)* informationsverarbeitende Systeme, bestehend aus einer großen Anzahl von Einheiten, Zellen oder Neuronen. Die Neuronen senden sich über die "Aktivierung" von Zellen Informationen zu. Wir haben es laut *Bock (1995)* mit einer abstrakten Simulation natürlicher neuronaler Netze zu tun.

ANN sind selbst lernende Systeme, die ähnlich den Lebewesen von Beispielen lernen können. Beim Trainieren neuronaler Netze kommt es zu einer Anpassung der Gewichte der Verbindungen und der Parameter der Neuronen an die Trainingsdaten. Eine der wichtigsten Eigenschaften von ANN ist die Parallelität, wobei eine Vielzahl an Neuronen parallel arbeitet und damit die Kapazität des ANN erhöht. Per se sind Neuronen im ANN einfache Einheiten, die erst durch ihr Zusammenschalten einen hohen Komplexitätsgrad gewinnen. ANN bestehen aus einer Eingabeschicht, ein bis zwei verdeckten Schichten (Hidden Layer) und einer Ausgabeschicht. Diese Schichten wiederum bestehen aus Neuronen. Die einzelnen Neuronen werden durch gerichtete Verbindungen zusammengehalten. Jedes Neuron besitzt eine Eingabefunktion, welche die Summe der Eingabesignale sowie deren Multiplikation mit den Gewichten der Verbindungen aus den vorangegangenen Schichten darstellt. Des Weiteren verfügt jedes Neuron über eine an einen Schwellwert gekoppelte Aktivierungsfunktion (*Bock, 1995*). Anhand der vorhandenen Attribute im Datensatz wird die Anzahl der Neuronen ermittelt. Ein wichtiges Element eines neuronalen Netzes ist die Propagierungsfunktion. Diese gibt an, wie sich die Netzeingabe eines Neurons aus den Ausgaben anderer Neuronen und den Verbindungsgewichten der Vorgängerneuronen berechnet (*Harbich & Peter, 2007*). Backpropagation ist ein weiterer Bestandteil von ANN und eine Form eines Gradientenabstiegsverfahrens. Mit dieser wird das globale Minimum der Fehlerfunktion ermittelt und eine die Fehlersumme minimierende Konfiguration der Gewichte gesucht. Bis ein vorab festgelegter Fehlerwert unterschritten oder erreicht wird, durchläuft das ANN die Propagation und die Backpropagation (*Bock 1995*).

Neben dem Vorteil der Lernfähigkeit weist das Verfahren einige Nachteile auf. *Dixon und Candade (2008)* geben dabei lange, mit der Anzahl an Trainingsda-

ten ansteigende Trainingszeiten und die Neigung zu Overfitting und Underfitting<sup>19</sup> bei nicht optimaler Parameterwahl als Nachteile von ANN, an.

### 4.3 Methodengruppe der geovisuellen Methoden

Die dritte Methodengruppe dient zur Überprüfung der zweiten These, die gemäß Mietpreiskarten nicht die tatsächlichen räumlichen Verhältnisse auf Immobilienmärkten wiedergibt. Es besteht eine Querbeziehung zwischen den ersten beiden Methodengruppen. Die Methoden aus der ersten und zweiten Methodengruppe bilden das räumlich-statistische Fundament, auf dem anschließend webbasierte Darstellungsformen von Immobilienpreisen entwickelt werden. Zur Umsetzung alternativer Darstellungsformen werden Webkarten mittels Map APIs umgesetzt. Bei der Erarbeitung von alternativen Darstellungsmethoden müssen deshalb Darstellungsformen gefunden werden, die Mietpreise – losgelöst von den fachfremd verwendeten Zonen – in zonenlosen Preiskarten darstellen. Es sind diesbezüglich Überlegungen anzustellen, wie – basierend auf Ergebnissen einer räumlichen Schätzung von Mietpreisen – eine Loslösung von Zonen erreicht werden kann. Eine Methode, die im Rahmen dieser Arbeit teilimplementiert wird, ist das Gridding. Beim Gridding werden Schätzergebnisse auf ein vorab zu definierendes Raster (Grid) bezogen. Dieser Bezug räumlich-statistischer Schätzergebnisse auf die Geometrie eines Grids und die Darstellung in Form einer Gridmap erlauben die Wiedergabe der tatsächlichen Verteilung von Mietpreisen im Raum, unabhängig von administrativen Grenzen. Die teilimplementierten Mietpreiskarten basieren also wesentlich auf den räumlichen Schätzergebnissen und auf Methoden des Gridding. Nachfolgend werden die theoretischen Grundlagen der teilimplementierten Methoden erläutert.

#### 4.3.1 Grid-Mapping

In der technologischen Recherche am Anfang der Arbeit (Kapitel 2.3) wurde festgestellt, dass Immobilienportale Immobilienpreise fast ausschließlich in Bezug zu fachfremd verwendeten administrativen Einheiten – also zonengebunden – darstellen. In den geführten Experteninterviews wurde das Problemfeld der fachfremd angewandten Bezugsgeometrien, neben dem Problemfeld der

---

<sup>19</sup> Über- und Unteranpassung eines Modells.

fehlenden räumlich statistischen Preismodellierung, als das Dringlichste zu bearbeitende genannt. Die Darstellung von Mietpreisen innerhalb fachfremd verwendeter Grenzen (nachfolgend Zonen genannt) stellt in keiner Weise die wahre Verteilung von Mietpreisen im Raum dar.

Das Grid-Mapping wird als Lösungsansatz seit mehreren Jahren von unterschiedlichen nationalen und internationalen Behörden vor allem aus dem Statistikbereich verfolgt. Vorreiter für Ansätze des Grid-Mapping ist insbesondere das U.S. Zensus Büro, welches schon in den frühen 1990er Jahren die irregulär geformten statistischen Blöcke (US Census Blocks) in ein nationales Grid umwandelte. Die „Statistik Austria“ (Statistics Austria, 2013) entwickelte ein österreichweites Grid mit unterschiedlichen Maschenweiten (100 m, 250 m, 500 m, 1.000 m, 2.000 m, 5.000 m und 10.000 m) zur Darstellung von Bevölkerungs-, Gebäude- und Arbeitsmarktstatistiken. Trainor (2010) fasste mehrere Argumente des Für und Wider zu Ansätzen des Bezugs von statistischen Ergebnissen auf Gridzellen zusammen. Zwei von Trainors dafürsprechenden Argumenten (Trainor, 2010) sind, dass dadurch eine größere Vergleichbarkeit gegenüber Ansätzen mit einem Bezug auf administrative Einheiten gegeben ist und dass es dadurch möglich ist, Cluster und Verteilungen im Raum zu erkennen.

### 4.3.2 Map APIs

Unter Webkarte wird in dieser Arbeit eine webbasierte graphische Repräsentierung eines Datenmodells verstanden. Es handelt sich um Präsentationskarten, in denen ein digitaler Datenbestand durch Attribuierung mit graphischen Merkmalen sichtbar gemacht wird. Map APIs sind JavaScript-Bibliotheken, die Funktionalitäten zur Interaktion mit der Kartengrafik bereitstellen und die Integration von Karteninhalten von Kartenservern (meist Tiled Map Services) ermöglichen. „Tiled Map Serving“ beruht dabei auf zwei Lösungsansätzen der webbasierten Kartenbereitstellung (Peterson, 2014) – einerseits dem „Image Tiling“ als einem Lösungsansatz, nach dem das Bild in viele kleine Segmente unterteilt und jedes dieser Segmente, je nach aufgerufenem Raumausschnitt und gewähltem Maßstab (Zoomstufe), über das Internet gesendet wird. Der zweite Lösungsansatz beruht auf dem asynchronen JavaScript (Ajax) und der Extensive Markup Language (XML). Mit Einführung dieser beiden Technologien im Jahr 2005 durch Google (Peterson, 2014) wurde eine bis dato neue Art der Server-Client-Interaktion ermöglicht. In Anwendung auf Webkarten bedeutet dies, dass eine Möglichkeit geschaffen wird, eigene Karteninhalte – im Fall von Mietpreiskarten die geschätzten Mietpreise – mit bereitgestellten Inhalten unterschiedlicher Anbieter zu kombinieren.

Der Status quo der Visualisierung von Mietpreisen ist die Darstellung der aus deskriptiver oder hedonischer Regression stammenden Schätzergebnisse mittels Map API und Grundkarte der Firma Google. Map APIs sind eine junge Entwicklung (Peterson, 2014) und leiden noch häufig an Kinderkrankheiten. Zu diesen Defiziten gehören beispielsweise das in Kapitel 2.2 geschilderte „Icon Cluttering“, die häufig unreflektierte Verwendung der „Webmerkator“-Projektion, aber auch die bei der Entwicklung der Prototypen aufgetretene Problematik, sinnvoll Klassengrenzen darzustellen. Verbesserungen der Darstel-

lung von Mietpreisschätzungen sollten daher unbedingt unter der Prämisse der Anwendung von Map APIs gefunden werden.

### 4.3.3 Kartenbasis

Über die Wahl der geeigneten Map API hinaus stellt sich die Frage nach einer geeigneten Kartenbasis zur Darstellung von Mietpreisen. Das Problem einer geeigneten Kartenbasis gilt im Übrigen auch für die Angebotsdarstellung. Laut *Hake et al.* (1994) dient eine Basiskarte oder Grundkarte als geometrisches Gerüst zur Verortung der thematischen Angaben bzw. zum sachlichen Verständnis des Themas. Laut *Imhof* (1972) sollte die Basiskarte das eingeführte Thema stützen, ergänzen und so wenig als möglich stören oder belasten.

Es stellt sich somit die Frage: Wird die mehrheitlich vorzufindende Verwendung von *Google maps* als Grundkarte diesen von *Imhof* (1972) formulierten qualitativen Ansprüchen gerecht? Bei der Suche nach einer Alternative zu *Google maps* muss ein gewisser "Gewöhnungseffekt" von Kartennutzern an *Google maps* beachtet werden. Es gilt also Alternativen zu finden, die einerseits *Imhofs* (1972) formuliertem Anspruch genügen und andererseits an *Google maps* gewöhnte Nutzer nicht überfordern. Eine weitere konzeptionelle Vorüberlegung betrifft den unbedingt darzustellenden Karteninhalt, genauer die Frage, ob und wenn ja welche zusätzlichen Informationen in die Grundkarte integriert werden oder als zusätzliche Kartenschichten der Mietpreiskarte vorhanden sein müssen.

### 4.3.4 Cloudbasierte Geovisualisierung

Es gibt aktuell zahlreiche Technologien zur Umsetzung cloudbasierter Geovisualisierung; drei davon – nach Einschätzung des Autors die drei wichtigsten: CartoDB, Geoserver und Mapbox – werden nachfolgend kurz charakterisiert. CartoDB ist ein cloudbasierter Ansatz, um die Ansprüche an eine Immobilienpreiskarte umzusetzen. CartoDB wird als Software as a Service (SaaS) angeboten und ermöglicht dadurch Webmapping sowie die Nutzung von GIS-Werkzeugen im Webbrowser. CartoDB folgt dem Gedanken des Open Source Stack und schaltet unterschiedliche freie, quelloffene Komponenten (z. B. PostGIS, PostSQL, Mapnik) zu einer cloudbasierten Plattform zusammen (*Github*, 2015). CartoDB kann dabei entweder auf einem eigenen Server installiert oder in Form unterschiedlich gestaffelter Hosting Pakete als vorinstallierte CartoDB-Instanz abonniert werden (*CartoDB*, 2015). Der Geoserver ist ein Java basierter OGC-(Open Geospatial Consortium-)konformer, freier und quelloffener „Kartenserver“, welcher Web Map Service (WMS), Web Coverage Service (WCS), Web Feature Service (WFS, auch transaktional) und Web Processing Service (WPS) bereitstellen kann (*Geoserver*, 2015). Mapbox (2015) ist ein proprietärer Anbieter eines TMS-Dienstes. Mapbox erfüllt die Anforderungen an Quelloffenheit nicht, da der Stack zwar wie CartoDB auf freien Komponenten aufsetzt, jedoch nur proprietär, über Hostingpakete, genutzt werden kann.

## 4.4 Fazit

Insgesamt lassen sich mögliche Lösungsansätze für eine räumliche Optimierung der Mietpreisschätzung in drei Gruppen einteilen:

1. Räumlich-statistische Methoden
2. Methoden des Maschinellen Lernens
3. Geovisuelle Methoden

Diese drei Methodengruppen werden im fünften Kapitel der Arbeit auf ihre Eignung zur räumlichen Optimierung von Mietpreisschätzungen von Immobilienportalen untersucht. Der ersten Methodengruppe wird vom Autor ein großes Potential zur räumlichen Optimierung der Mietpreisschätzung von Immobilienportalen zugeschrieben.

Es existieren innerhalb der Verfahren der räumlichen Interpolation zahlreiche Methoden. Eine Auswahl eines optimalen Schätzverfahrens im Kontext von Immobilienportalen ist daher entscheidend für die räumliche Optimierung der Mietpreisschätzung. Aus diesem Grund ist bei der Teilimplementierung von Methoden der ersten Verfahrensgruppe, der Validierung, große Aufmerksamkeit zu schenken. Das gilt im Übrigen auch für die Methodengruppe des Maschinellen Lernens. Die bestmögliche Parameterkonfiguration und Sekundärvariablenauswahl bei räumlichen Schätzverfahren ist – wie dargestellt – entscheidend für die räumliche Interpolation von Angebotspreisen. Sie stellt im Zusammenhang mit der Schätzung von Mietpreisen eine noch ungeklärte Forschungsfrage dar.

GWR ist ein geeigneter Ansatz, Mietpreisschätzungen mit Lagemodellierungen zu kombinieren, jedoch sind dem Autor keine Arbeiten bekannt, in denen Lageindikatoren aus einer Lagemodellierung abgeleitet und in ein GWR-Modell integriert wurden. Im Rahmen der geführten Experteninterviews stieß die Möglichkeit auf Interesse, indikatorbasierte Standortbewertungen in Schätzmodelle zu integrieren. Diesem Interesse stand jedoch die Kritik entgegen, dass es kaum möglich ist, Wohnpräferenzen in ein räumlich-statistisches Modell der Preisschätzung zu integrieren.

Hinsichtlich der zweiten Methodengruppe des Maschinellen Lernens stellt sich die Frage: Sind diese Verfahren auf einzelne Quartalsdatensätze mit dünner Datenlage anwendbar oder sind hierfür Verfahren aus der räumlichen Statistik geeigneter? Es muss außerdem geklärt werden, ob diese Verfahren möglicherweise optimaler für die Datenpräprozessierung eines zeitlich aggregierten Massendatensatzes, ähnlich der Isoclastierung in der Fernerkundung, geeignet sind, um die räumliche Struktur eines Angebotsdatensatzes zu charakterisieren.

Für eine Teilimplementierung der Methodengruppe der Visualisierung von Mietpreisen gilt es, geeignete Maschenweiten für Gridmaps zu finden. Um die optimale graphische Darstellungsform für Mietpreise zu ermitteln, werden zur Teilimplementierung von Geovisualisierungsmethoden Prototypen entworfen (siehe Kapitel 5.3). Mit Hilfe der Prototypen sollten mehrere Fragen geklärt wer-

den, u. a.: Was ist die geeignete Kartenbasis für eine Mietpreiskarte? Welche Darstellungsform ist geeignet, um die tatsächliche räumliche Verteilung von Mietpreisen zu visualisieren? Welche Kartenschichten und Funktionalitäten benötigt eine webbasierte Mietpreiskarte? Welche Mietpreisklassen und welche Farben sind für die Darstellung geeignet? Darüber hinaus ergibt sich die technologiebezogene Problematik der Wahl einer geeigneten technologischen Plattform zur Umsetzung cloudbasierter Geovisualisierung.

## 5 ANWENDUNG DER METHODEN

Das fünfte Kapitel ist der Darstellung der vom Autor entwickelten Teilimplementierungen zur räumlichen Optimierung von Immobilienportalen hinsichtlich der Mietpreisschätzung und der Darstellung von Mietpreisen gewidmet. Die Teilimplementierungen werden dabei in zwei Anwendungsfälle (Use cases) gegliedert. Das ist zum einen ein Anwendungsfall der räumlichen Analyse und zum anderen ein Anwendungsfall der Geovisualisierung. Über diese beiden Anwendungsfälle erfolgt in konzeptioneller und teilimplementierter Form eine adaptive Konfektionierung bestehender Methoden mit dem Ziel der räumlichen Optimierung von Immobilienportalen. Im ersten Anwendungsfall findet eine Änderung und Erweiterung des Status quo von Verfahren der Preismodellierung um räumliche Verfahren statt. Der zweite Anwendungsfall ist der Anwendungsfall der Geovisualisierung, hier werden Lösungsansätze der Visualisierung von Immobilienpreisen (konkret Mietpreise) diskutiert. Nachfolgend werden die beiden Anwendungsfälle näher erläutert.

### *Anwendungsfälle*

Folgende Abbildungen (5-1 und 5-2) stellen die Anwendungsfälle der räumlichen Analyse und der Geovisualisierung in einem Anwendungsfalldiagramm stark abstrahiert dar. Die Rechtecke zeigen die wesentlichen analytischen Anforderungen an den Anwendungsfall, die abgerundeten Felder die wesentlichen Prozesse, die notwendig sind, um die einzelnen Anforderungen zu erfüllen. Der Anwendungsfall der räumlichen Analyse besteht aus zwei Komponenten: 1. der Datenpräprozessierung und 2. der räumlichen Preisschätzung. Beide Komponenten umfassen unterschiedliche Prozesse, die für das Ziel einer räumlich optimierten Mietpreisschätzung wesentlich sind. Der Anwendungsfall der Geovisualisierung baut hierbei auf den Ergebnissen der räumlichen Analyse auf und geht der Frage nach, wie geschätzte Mietpreise in Form von Immobilienpreiskarten visualisiert werden können. Er besteht aus den beiden Teilkomponenten: Wahl der Bezugsgeometrien und Visualisierung mittels Map APIs.

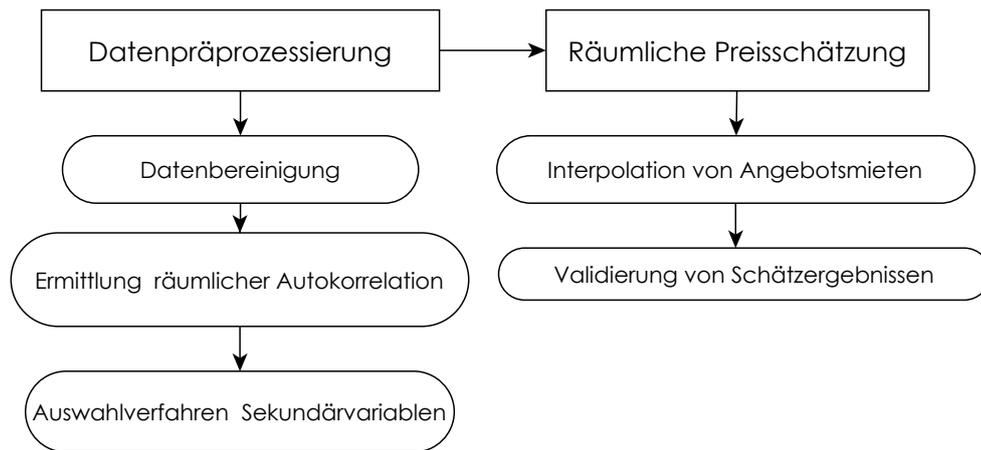


Abb. 5-1: Anwendungsfalldiagramm des Anwendungsfalles der räumlichen Analyse (Entwurf und Darstellung: Schernthanner 2015).

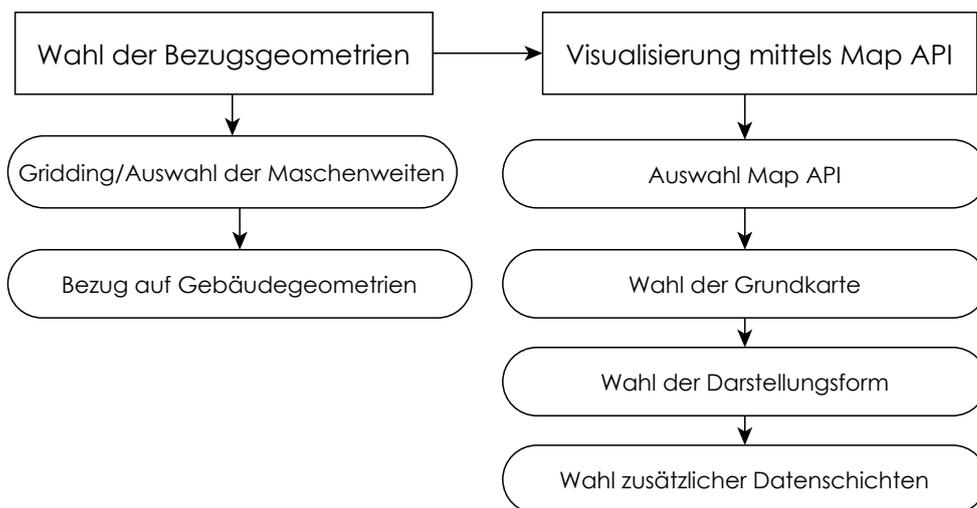


Abb. 5-2: Anwendungsfalldiagramm des Anwendungsfalles der Geovisualisierung (Entwurf und Darstellung: Schernthanner 2015).

Die entstandenen Teilimplementierungen des Anwendungsfalles der räumlichen Analyse werden in Kapitel 5.1 und 5.2, die des Anwendungsfalles der Geovisualisierung in Kapitel 5.3 diskutiert.

## 5.1 Anwendung räumlich-statistischer Methoden

### 5.1.1 Teilimplementierung der hedonischen Regression

Zur Demonstration des Status quo der Mietpreisschätzung von Immobilienportalen wird die hedonische Regression in der Semilog-Variante beispielhaft auf den vorhandenen aus dem ersten Quartal 2013 stammenden Angebotsdaten-

satz angewendet. Hierbei wird der Preis für die Datenpunkte in einem Stadtteil mittels hedonischer Regression (HedReg) ermittelt. Dazu werden alle Datenpunkte innerhalb der administrativen Grenzen eines Stadtteils selektiert. Aus diesen Daten wird dann die durchschnittliche oder "Standardwohnung" ermittelt. Diese sieht für den Untersuchungszeitraum im Potsdamer Stadtteil Babelsberg-Nord wie folgt aus (siehe Tabelle 5-1).

<b>Ø Wohnfläche m<sup>2</sup></b>	<b>Ø Objektalter</b>	<b>Keller</b>	<b>Aufzug</b>	<b>Garten</b>	<b>Balkon</b>	<b>Einbauküche</b>	<b>Ø Zimmer</b>	<b>Ø Etage</b>
78,9	42	Ja	Nein	Nein	Ja	Ja	2,67	1,56

Tabelle 5-1: Standardwohnung im Stadtteil Babelsberg Nord im ersten Quartal 2013.

Aus dem Gesamtdatensatz für den zu untersuchenden Zeitschnitt werden dann diejenigen Angebotsdaten selektiert, die innerhalb des jeweiligen Zeitschnittes und des zu untersuchenden Stadtteils liegen. Über ein festzulegendes Ko-Variablenset wird anschließend der Mietpreise pro Quadratmeter für eine Standardwohnung in dem jeweiligen Stadtteil modelliert. Für die beispielhafte hedonische Regression der Angebotsdaten des ersten Quartals 2013 wurde folgendes Sekundärvariablenset verwendet:

*Wohnfläche/Objektalter/Zimmeranzahl/Etage*

Mittels der OLS-Methode erfolgt eine Schätzung der Koeffizienten mit dem Ziel, den Einfluss von Störgrößen innerhalb der Quadratsumme zu minimieren. In einigen Stadtteilen fehlen Beobachtungspunkte gänzlich oder sind in so geringem Maße – statistisch nicht signifikant – vorhanden, dass keine Regression möglich ist. In anderen Stadtteilen kann durch fehlende Werte im Datensatz keine Regression berechnet werden. Folgende Abbildung (5-3) zeigt einen schematischen Ablauf der HedReg.

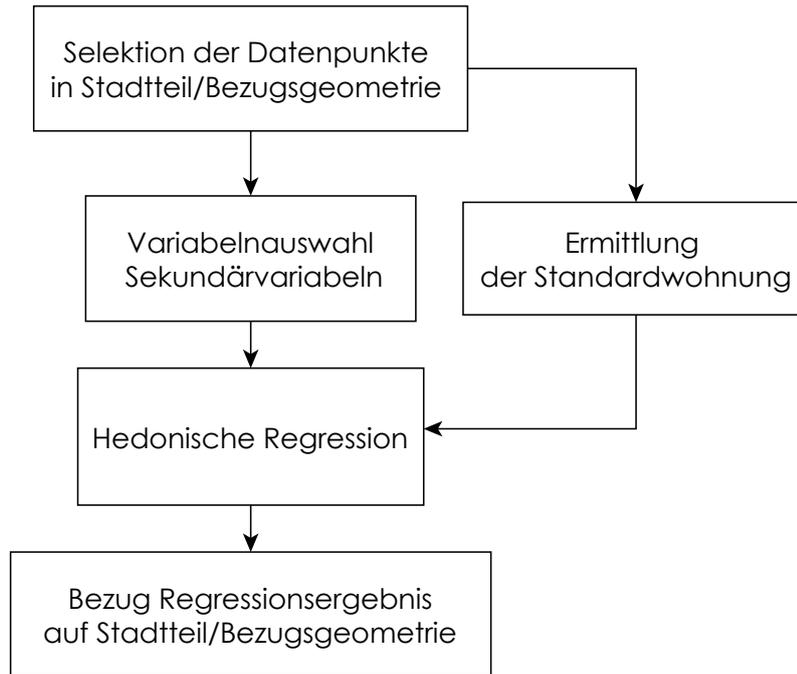


Abb. 5-3: Schematischer Ablauf der hedonischen Regression. (Entwurf und Darstellung: Scherthanner 2015).

Es sei nochmals betont, dass die Methode der HedReg in keiner Weise räumlich ist und dass die Ergebnisse der vorliegenden Regression in die Raumeinheit der administrativen Grenze der Stadt Potsdam "gepresst" wurden. Für jeden der Stadtteile bzw. der darin liegenden Angebotsdaten ist ein separates Modell zu rechnen. Ein großer Vorteil der Berechnung mittels hedonischer Regression ist die Rechengeschwindigkeit der Modellläufe, die im Durchschnitt bei drei Minuten lag. Abbildung 5-4 zeigt das Ergebnis der gerechneten hedonischen Regression für verschiedene Potsdamer Stadtteile.

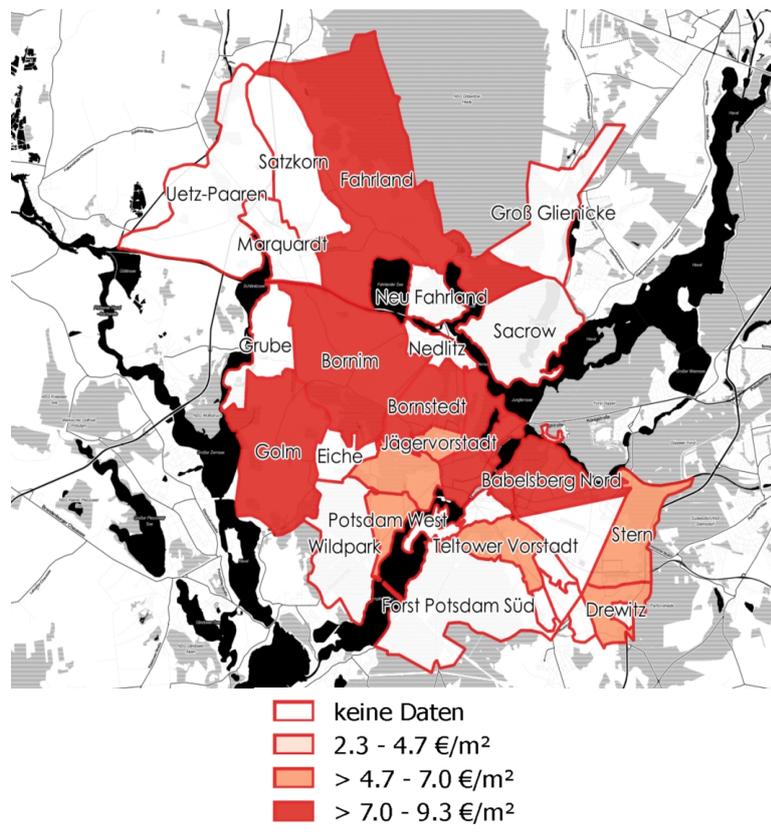


Abb. 5-4: Ergebnisdarstellung der HedReg auf Stadtteilebene. (Entwurf und Darstellung: Schernthanner 2015).

### 5.1.2 Teilimplementierung der Auswahl von Sekundärvariablen

Mehrere Verfahren aus den identifizierten Methodengruppen der räumlich-statistischen Methoden und aus der Gruppe der Verfahren des Maschinellen Lernens benötigen geeignete Sekundärvariablen. Deshalb ist deren Auswahl ein geeigneter wichtiger Bestandteil jeder räumlichen Optimierung einer Mietpreisschätzung.

Die Auswahl geeigneter Sekundärvariablen erfolgte in einem dreistufigen Ansatz (siehe Abbildung 5-5). In der ersten Stufe wurden von einem Expertengremium ermittelte Einflussfaktoren auf die Nettokaltmiete mit dem vorhandenen exemplarischen Angebotsdatensatz abgeglichen. Daran anschließend folgten eine Hauptkomponentenanalyse (PCA) sowie eine Analyse der Variablenge-wichtigkeit mittels der Methode Random Forest (RF).

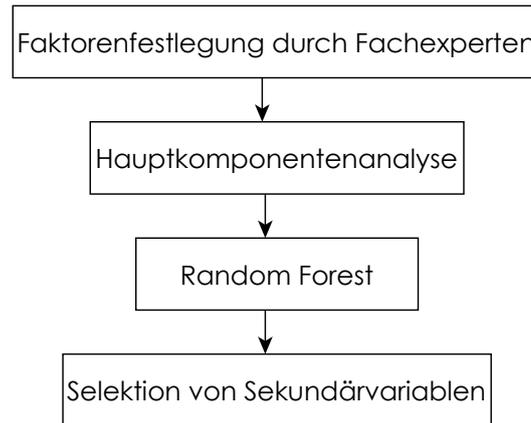


Abb. 5-5: Methodik der Auswahl der Sekundärvariablen. (Entwurf und Darstellung: Scherthanner 2015).

### *Faktorenauswahl durch Fachexperten*

Den Anstoß zur Auswahl geeigneter Sekundärvariablen gab eine über das Jahr 2012 verteilte Gastteilnahme an einem Runden Tisch der Stadt Potsdam. Die Stadt Potsdam entschied mit einem Stadtratsbeschluss vom 28.09.2011 die Bildung eines Expertengremiums zur Erarbeitung eines Konzeptes: „Erschwingliche Mieten für Alle“ (Landeshauptstadt Potsdam, 2011). Auf Grundlage dieses Beschlusses wurde ein Expertengremium, bestehend aus fünf Experten der Bereiche Mieterschutz (Deutscher Mieterbund), Demographischer Wandel und Soziale Strukturen (Stadtverwaltung Potsdam), Wohnungswirtschaft (AK Stadtspuren), Energie (Energieforum Potsdam e.V.) und Klimaschutz (Potsdam Institut für Klimafolgenforschung), gebildet (Potsdam 22, 2013). Die Empfehlungen der Expertengruppe wurden der Stadt Potsdam am 21.02.2013 vorgelegt und fließen derzeit in das wohnungspolitische Konzept der Stadt ein (Potsdam 22, 2013).

Der Autor der vorliegenden Dissertation nahm von 2012 bis 2013 als Gast an den Sitzungen der Expertenrunde teil. Hierdurch konnte ein tiefer Einblick in das Funktionieren eines städtischen Mietmarkts und in das Zustandekommen eines Mietpreises gewonnen werden. In einer der Sitzungen wurde die Expertenrunde gebeten, alle aus ihrer Sicht vorhandenen Einflussfaktoren auf Wohnkosten

insgesamt zu nennen. Diese sollten des Weiteren Themen zugeordnet und in ihren Querbeziehungen aufgezeigt werden. An diesem Prozess konnte der Autor mitwirken. Der Mitwirkungsprozess erfolgte im ersten Schritt analog mittels Flipchart und Markern. Die analogen Ergebnisse wurden anschließend digitalisiert und den Teilnehmern mittels der Software „Consideo<sup>20</sup>“ als Basis wirtschaftlicher Entscheidungsprozesse modelliert und neu gewichtet zur Verfügung gestellt. Insgesamt wurden von den Experten 142 Einflussfaktoren auf Wohnkosten genannt, dem stehen 64 im exemplarischen Angebotsdatensatz vorhandene Sekundärvariablen gegenüber.

Das im Ergebnis des Expertengremiums entstandene Modell hat insgesamt vier Hierarchieebenen. Jeder dieser Ebene wurden Einflussfaktoren zugeordnet, die miteinander verlinkt sind. Abbildung 5-6 zeigt die in der ersten – der höchsten – Hierarchieebene vorhandenen Einflussfaktoren auf die Nettokaltmiete. Im Anhang 9 findet sich eine Abbildung aller in *Potsdam 22* festgestellten Einflussfaktoren auf die Nettokaltmiete.

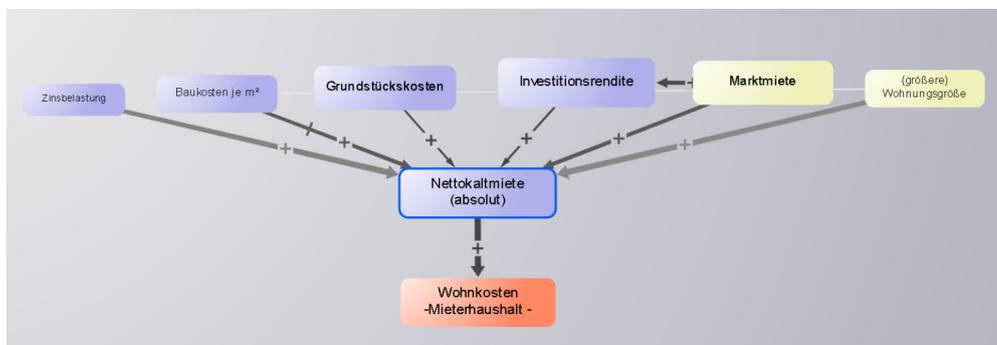


Abb. 5-6: Direkte Einflussfaktoren auf die absolute Nettokaltmiete laut Expertengremium *Potsdam 22*. Die Linien zeigen die einzelnen Einflussfaktoren auf die Nettokaltmiete und in weiterer Folge auf die Wohnkosten. Über das Plus-Symbol können Faktoren niedrigerer Hierarchieebenen visualisiert werden. (*Potsdam 22*, 2013)

Im Modell wurden neben unterschiedlichen Faktoren auf die Wohnkosten insgesamt Faktoren genannt, die unmittelbar Einfluss auf die Nettokaltmiete haben (z. B. Lagefaktoren, Baukosten etc.)

<sup>20</sup> Software zur visuellen Prozessmodellierung: <http://www.consideo.de/>

Da es im raumanalytischen Anwendungsfall um die Schätzung von Nettoangebotsmieten geht, wurden nur diejenigen Faktoren herausgegriffen, die laut Experten einen unmittelbaren Einfluss auf die Nettokaltmieten ausüben (Subfaktoren). Nur einige wenige der insgesamt benannten Einflussfaktoren lassen sich aus den in den exemplarischen Angebotsdaten vorliegenden Variablen herauslesen. Die vom Expertengremium aufgestellten Faktoren geben einen umfassenden Einblick dahingehend, wie Wohnkosten zustande kommen. Als Grundlage für eine Mietpreisschätzung sind die aufgestellten Faktoren jedoch nur in geringem Maße verwendbar. Es gibt lediglich eine geringe Übereinstimmung der Faktoren mit denen in den realweltlichen Mietangebotsdatensätzen vorhandenen Variablen. Es wurde ein Abgleich der im Beispieldatensatz vorhandenen Sekundärvariablen vorgenommen (siehe Tabelle 5-2). Sehr gut abgedeckt werden von den vorhandenen Angebotsdaten Daten über vorhandene Wohnungsgrößen bzw. Ausstattungsmerkmale von Wohnungen.

Der Abgleich der in *Potsdam 22* festgestellten Faktoren mit den vorhandenen Angebotsdaten stellte einen sinnvollen ersten Schritt im Auswahlverfahren von Sekundärvariablen dar. Darüber hinaus bot dieser Prozess einen erheblichen Zugewinn an Expertise über das Zustandekommen von Wohnkosten allgemein und ermöglichte eine erste Selektion von Ko-Variablen.

<b>Faktoren <i>Potsdam 22</i></b>	<b>Variablen IS24</b>	<b>Datentyp IS24</b>
<b>Wohnungsgröße</b>	Wohnungsfläche	Numerisch
	Nutzfläche	Numerisch
	Zimmeranzahl	Numerisch
	Schlafzimmer	Numerisch
	Grundfläche	Numerisch
<b>Ausstattung Wohnung</b>	Einbauküche	Kategorial
	Badezimmer	Numerisch
	Möbliert	Kategorial
	GästeWC	Kategorial
	Balkon	Kategorial
	Garten	Kategorial
	Parkplatz	Kategorial
	Parkplatzpreis	Numerisch
	Etage	Numerisch
	<b>Barrierefreiheit</b>	Aufzug
Rollstuhlgerecht		Kategorial
<b>Gebäudestandard</b>	Objektzustand	Numerisch
	Immobilientyp	Numerisch
	Baujahr	Numerisch

Tabelle 5-2: Abgleich der Wohnkostenfaktoren von *Potsdam 22* mit dem exemplarischen Angebotsdatensatz.

## Hauptkomponentenanalyse (PCA)

Im zweiten Schritt der Variablenauswahl wurde eine PCA durchgeführt. Die PCA wurde jeweils einmal auf den gesamten Datenbankauszug sowie zu Vergleichszwecken auf den Datensatz aus dem Quartal 1, 2013 angewendet. Zur Anwendung der PCA wurde mit der R Bibliothek *princomp* ein entsprechendes R-Skript geschrieben.

Der Datenbankauszug mit insgesamt 74.099 Beobachtungen und 64 Variablen wurde auf alle numerischen Variablen exklusiv potenziert bzw. zur internen Systematik auf vorliegende Variablen reduziert. Die Variablen "Baujahr", "Etage", "Kaltmiete", "Objektalter", "Verweildauer der Anzeige", "Wohnfläche" und "Zimmeranzahl" wurden danach in eine Kovarianzmatrix überführt. Die ersten drei Hauptkomponenten haben an der Varianz einen kumulativen Anteil von mehr als 80 %. Dies lässt die Annahme zu, dass diese drei Hauptkomponenten ausreichen, die Daten zusammenzufassen. Die Analyse der aus den ersten drei Hauptkomponenten resultierenden Ladungswerte zeigt, dass in der ersten Hauptkomponente die Variablen "Etage" und "Baujahr", in der zweiten Hauptkomponente die "Kaltmiete" und die "Zimmeranzahl" die größten Faktorenladungen aufweisen. Selbiges gilt für die dritte Hauptkomponente.

Eine weitere PCA des Datensatzes aus dem ersten Quartal 2013 zeigt ähnliche Ergebnisse. Die ersten drei Hauptkomponenten haben zu 80 % Anteil an der Varianz und die Variablen mit den größten Faktorenladungen sind "Baujahr" und "Etage" in der ersten Hauptkomponente, "Kaltmiete", "Wohnfläche" und "Zimmeranzahl" in der zweiten Hauptkomponente, "Verweildauer" und "Etage" in der dritten Hauptkomponente. Ein in R erstellter Scree-Plot (Plot der Varianz versus Hauptkomponenten, Abbildung 5-7) zeigt, dass bereits die ersten beiden Hauptkomponenten den Großteil der Variabilität in den Daten erklären.

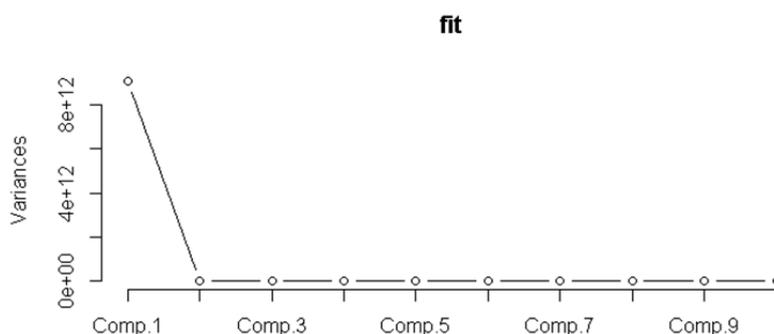


Abb. 5-7: Scree-Plot der Varianz in den verschiedenen Hauptkomponenten. (Entwurf und Darstellung: Scherthanner 2015).

Ein weiterer Plot der PCA (Abbildung 5-8) zeigt die Projektion der Datenpunkte des ersten Quartals 2013 auf die ersten beiden Hauptkomponenten.

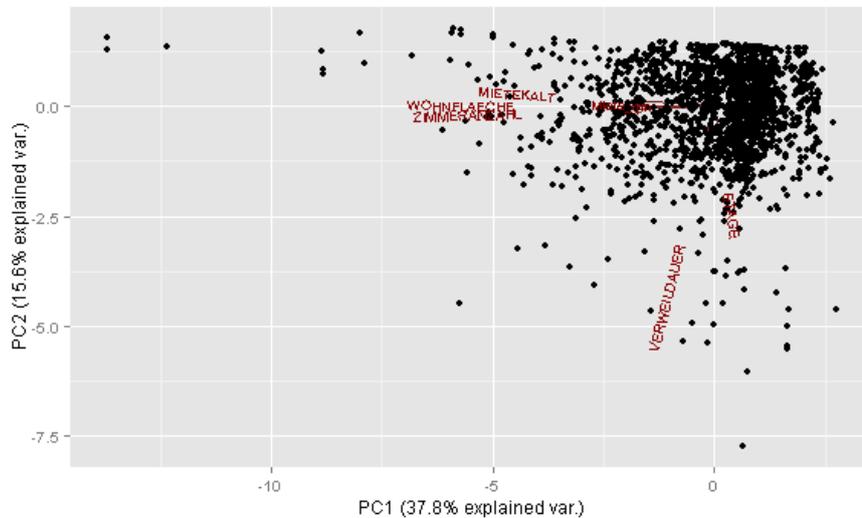


Abb. 5-8: Plot der ersten beiden Hauptkomponenten. Die x-Achse stellt die Faktorenladung der ersten Hauptkomponente, die y-Achse der zweiten Hauptkomponente dar (Entwurf und Darstellung: Scherthanner 2015).

Resultierend aus der Hauptkomponentenanalyse werden folgende Sekundärvariablen als Grundlage für Ko-Variablen sets vorgeschlagen:

*Baujahr/Etage/Objektalter/Verweildauer/Wohnfläche.*

Als dritte Stufe in der Auswahl von Sekundärvariablen wurde die Methode RF aus dem Maschinellen Lernen angewendet. Diese wird auch als eine Alternative zur hedonischen Regression in Betracht gezogen.

### *RF zur Auswahl von Sekundärvariablen*

Die Anwendung von RF ist eine weitere Entscheidungshilfe für die Auswahl von Kovariablen. Random Forest, entwickelt von Breiman (2001), wird oft zum Auffinden von Variablen mit dem höchsten Gewichtungsgrad in multivariaten Datensätzen verwendet (Williams, 2009). Ein Vorteil der Anwendung von Random Forest ist, dass eine Anwendung auch auf kategoriale Variablen erfolgen

kann. Die Teilimplementierung von RF erfolgte in *Rattle*<sup>21</sup>, einer in R verfügbaren Benutzeroberfläche zur Anwendung von Verfahren des Maschinellen Lernens. Hierbei handelt es sich um eine von *Williams (2009)* initiierte Bibliothek, die eine Benutzeroberfläche für R-Pakete aus dem Bereich des Maschinellen Lernens und dem Datamining bereitstellt. *Rattle* greift dabei wiederum auf unterschiedliche R-Bibliotheken zum Datamining und Maschinellen Lernen (z. B. das R-Paket *RandomForest*) zurück. Vor der eigentlichen Berechnung erfolgte eine manuelle Reduktion der 64 vorhandenen Variablen um künstliche Regionscodes, vorhandene logarithmische Variablen und numerische Variablen, die potenziert vorlagen. Nach dieser Bereinigung blieben insgesamt zwölf Variablen übrig, deren Wichtigkeit in Anhang 7 ersichtlich ist. Die fünf wichtigsten Variablen in der Anwendung von RF auf 150 Entscheidungsbäume waren (Nennung in der Reihenfolge ihrer Wichtigkeit): *Baujahr/Einbauküche/Wohnfläche/Zimmeranzahl/Etage*. Nur die Variable *Einbauküche* ist kategorial und unter den gewichtigen Variablen sind alle weiteren numerischen Typs.

### *Ergebnisse der Auswahl der Sekundärvariablen*

Tabelle 5-3 zeigt die fünf mittels PCA und RF extrahierten wichtigsten Sekundärvariablen und gibt über die farbliche Markierung wieder, ob diese im Faktorenmodell von *Potsdam 22* vorhanden waren. Die Schnittmenge der Methoden ist Ausgangsbasis für die Auswahl an Sekundärvariablen.

---

<sup>21</sup> *Rattle* basiert auf dem RGtk2 Toolkit, einem Werkzeug zur Erstellung graphischer Benutzeroberflächen: <http://www.ggobi.org/rgtk2/>.

RF	PCA	Schnittmenge mit Potsdam-22-Faktoren
Baujahr	Baujahr	Baujahr
Einbauküche	Etage	Zimmeranzahl
Wohnfläche	Objektalter	Einbauküche
Zimmeranzahl	Verweildauer der Anzeige	Etage
Etage	Wohnfläche	Wohnfläche

Tabelle 5-3: Schnittmenge zwischen den einzelnen Verfahren. Grün hinterlegte Variablen wurden in allen drei Stufen des Verfahrens als Sekundärvariable identifiziert, gelb hinterlegte wurden in RF und dem Expertengremium identifiziert.

Einzig die Einbauküche ist nicht in der Schnittmenge enthalten und kommt, bedingt durch ihren Variablentyp, ohnehin in nur wenigen Verfahren als Sekundärvariable in Betracht. Der Abgleich der Verfahren ergibt eine fundierte Grundlage für die schlussendliche Variablenauswahl. Hierbei ist zu beachten, dass die Variable „Baujahr“ zwar numerisch, jedoch eine Datumsangabe ist. Um Problemen bei den Teilimplementierungen vorzubeugen, wurde anstelle des Baujahrs mit der Sekundärvariable „Objektalter“ gearbeitet<sup>22</sup>.

### 5.1.3 Teilimplementierung der Messung der räumlichen Autokorrelation

Für die vorhandenen Angebotsdatensätze muss ermittelt werden, ob Autokorrelation vorhanden ist und wie stark diese ist. Hierfür kommen die in Kapitel 4.2.3 beschriebenen gängigen Tests zur Anwendung. Im exemplarischen Untersuchungsraum Potsdam wurde die Autokorrelation mit den drei Tests: Morans I, Gearys C und Getis Ord G auf Basis der Angebotsdaten des ersten Quartals 2013 berechnet. Hierbei konnte eine Normalverteilung der Datenpunkte festgestellt werden. Mit den R Paketen ape und spatstat konnte unter Miteinbe-

<sup>22</sup> Die Variable Baujahr ist ein numerisches Datum. Die Variable Objektalter ist ebenfalls numerisch und gibt das Alter eines Objektes in der Anzahl der Jahre wieder.

ziehung der jeweils vier um einen Beobachtungspunkt liegenden, Nachbarpunkte eine Gewichtsmatrix von inversen Distanzen gebildet werden. Dabei wurde näher beieinanderliegenden Punkten ein höheres Gewicht zugewiesen als weiter auseinanderliegenden. Danach wurden der Moran's I und der Geary's C Hypothesentest jeweils auf Nachbarpunkte rund um den Beobachtungspunkt angewendet. Die Autokorrelation war im Ergebnis beider globaler Tests positiv. Sie lag beim globalen Moran's I bei 0,81 und bei Geary's C bei 0,2.

Zu einem weiteren Test der Nullhypothese wurde der Moran's I zusätzlich innerhalb von binären Distanzmatrizen der Beobachtungspunkte in 300 m/500 m/1.000 m/5.000 m und 15.000 m Distanz vom jeweiligen Beobachtungspunkt gebildet. Durch die Anwendung der räumlichen Autokorrelation auf eine Teilmenge der Beobachtungspunkte (lag) innerhalb der definierten Distanzklassen ist es möglich, das Verhalten der Autokorrelation in unterschiedlichen Distanzen zu untersuchen und in einem Korrelogramm (siehe Abbildung 5-9) zu betrachten. Die jeweiligen Moran's I Werte der Distanzklassen zeigt die Tabelle 5-4.

Die Nullhypothese konnte beim vorliegenden Datensatz in vier von fünf gebildeten Distanzklassen abgelehnt werden. Erst in der letzten Distanzklasse von 15.000 m war nachweislich keine räumliche Autokorrelation mehr gegeben.

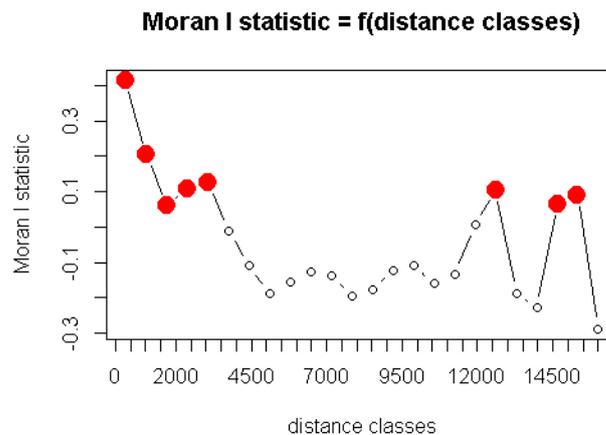


Abb. 5-9: Morans I nach Distanzklassen. (Entwurf und Darstellung: Scherthanner 2015).

| Morans I   |
|----------|----------|----------|----------|------------|
| 300 m    | 500 m    | 1000 m   | 5000 m   | 15 000 m   |
| 0.4598   | 0.5531   | 0.3263   | 0.08578  | - 0.001532 |

Tabelle 5-4: Moran's I nach Distanzklassen

Neben den globalen Tests wurde Getis Ord G als lokaler Test angewendet. Getis Ord G wird als globaler Z-Wert ausgewiesen und tendierte in den Untersuchungen gegen Null. Auf Grund dessen konnte keine Aussage über eine vorliegende positive bzw. negative räumliche Autokorrelation getroffen werden. Differenzierte Aussagen ermöglicht lediglich eine lokale Betrachtung, die klar Räume mit negativer und stark positiver räumlicher Autokorrelation ausweist. Der Test ermöglicht weitergehende Analysen wie z. B. die Berechnung von Hot- und Coldspots<sup>23</sup> und deren räumlicher Darstellung. Eine solche Berechnung wurde unter Einbeziehung der euklidischen Distanzen beispielhaft durchgeführt. Die errechneten Hot- und Coldspots wurden mittels der Nächste-Nachbarschaft-Methode in einen 500-m-Raster (siehe Abbildung 5-10) konvertiert.

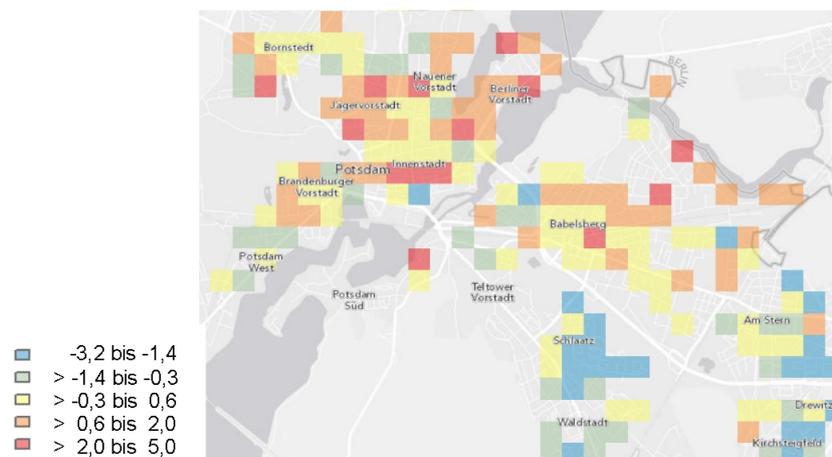


Abb. 5-10: Nach Getis Ord G ermittelte Hot- und Coldspots in Potsdam. Dargestellt wird der gerasterte Z Wert von Getis Ord G. (Entwurf und Darstellung: Schernthanner 2015).

<sup>23</sup> Hotspot = räumliche Clusterung von hohen Werten, Coldspots = räumliche Clusterung von niedrigen Werten.

#### 5.1.4 Teilimplementierung der geographisch gewichteten Regression in Kombination mit Lagemodellierungen

Nachfolgend wird das in der Methodengruppe der räumlich-statistischen Verfahren identifizierte (siehe Kapitel 4.1.4) Verfahren der geographisch gewichteten Regression in Kombination mit und ohne Integration eines Lageindikandums<sup>24</sup> für den raumanalytischen Anwendungsfall diskutiert. In den Experteninterviews kam heraus, dass ein Ansatz der Preisschätzung über eine Lagemodellierung als verfolgenswert erachtet wird. Jedoch wird die subjektive Einschätzung von Lagequalität betont, die es schwermacht, Immobilienpreise über einen Lageindikator zu modellieren. Mögliche Lageparameter wie z. B. Grünflächen, vorhandene Infrastruktur, Bildungseinrichtungen und deren Qualität bzw. Nähe zu ÖPNV werden als sehr subjektiv beurteilt. Wissenschaftlich scheint es fragwürdig, diese Parameter zu einem Lageindikator für eine Mietpreisschätzung zusammenzufassen. Diese Arbeit wird keinen weiteren Ansatz der Modellierung der Lage(-qualität) von Wohnstandorten verfolgen; es musste also ein Lageparameter gefunden werden, der relativ unstrittig ist.

##### *Teilimplementierung der GWR mit Bildung eines Lageindikandums*

Die fußläufige Distanz zu ÖPNV-Haltestellen kann als ein relativ unstrittiger Lageparameter bezeichnet werden. Im aktuellen Mietspiegel der Stadt Potsdam (*Landeshauptstadt Potsdam, 2014*) werden erstmals Lageeigenschaften und nicht nur Ausstattungseigenschaften berücksichtigt. In den Erläuterungen des Mietspiegels wird als Nicht-Erreichbarkeit von ÖPNV definiert, wenn Busse oder Bahn „innerhalb von ca. 1 km von 6.00 bis 20.00 Uhr“ bzw. „werktags seltener als 2 mal stündlich“ erreicht werden können (*Landeshauptstadt Potsdam, 2014*). Diese Zeit- und Distanzangaben werden im Mietspiegel nicht wissenschaftlich begründet.

Als Maß der Erreichbarkeit von Objekten wird vom Autor die Netzwerkdistanz des exemplarischen Angebotsdatensatzes auf Basis der 472 Bushaltestellen der Stadt Potsdam berechnet. Die ermittelten Netzwerkdistanzen fließen in das

---

<sup>24</sup> Laut *Prinz* (2007) sind Indikatoren Kerngrößen die zur Abbildung eines bestimmten, nicht direkt messbaren und oftmals komplexen Sachverhalts (Indikandum) ausgewählt werden.

Mietpreisschätzmodell ein. Die verwendeten Bushaltestellen wurden mittels der "Overpass API<sup>25</sup>" aus der OpenStreetMap gefiltert. Hierbei wurden Schwellwerte an den letzten Mietspiegel der Stadt Potsdam angeglichen. Zur Berechnung der Distanzen wurde ein routingfähiges Netzwerk basierend auf einem Ausschnitt der OpenStreetMap gebildet. Hierfür wurde das Werkzeug OSM2NetworkAnalyst genutzt (Codeplex, 2013). Dieses Werkzeug bereitet OpenStreetMap-Straßengraphen für die Verwendung im Arc GIS Network Analyst auf. Diese Aufbereitung ist notwendig, um anschließend die Netzwerkdistanzen in Arc GIS zu berechnen. In Anlehnung an die Mietspiegelfelder der Stadt Potsdam wurde die Netzwerkdistanz mittels einer „Closest facility“ Analyse der Wohnungsangebote in 100 m, 200 m und 500 m zu der jeweils nächstgelegenen Bushaltestelle berechnet. Es existieren zahlreiche Lagemodellierungen von Wohnstandorten über indikatorbasierte Ansätze. In diesem Zusammenhang stellt sich die Frage, ob man einen Lagewert oder in weiterer Folge einen aus mehreren Lagewerten ermittelten Lageindikator in ein räumliches Modell zur Mietpreisschätzung überführen kann.

Hierfür bietet sich die Methode der geographisch-gewichteten Regression (GWR) als eine Möglichkeit der Einbindung eines Lagewertes in ein Modell an. Beispielhaft wurde eine solche GWR-basierte Mietpreisschätzung durchgeführt. Als Zielvariable diente die Miete pro Quadratmeter, übereinstimmend zur hedonischen Regression wurden nachfolgend genannte Kovariablen verwendet: *Wohnfläche/Objektalter/Zimmeranzahl/Etage* zuzüglich der errechneten Netzwerkdistanz. Diese Variationen wurden in fünf Modellläufen getestet. Wird der globale  $R^2$  Wert betrachtet, dann zeigt sich, dass das GWR-Modell die abhängige Variable, je nach gewählter Kombination von abhängigen Variablen, in unterschiedlicher Güte erklärt. Tabelle 5-5 zeigt die  $R^2$  Werte für die unterschiedlichen Kovariablensets. Das geringste Bestimmungsmaß  $R^2$  von 0,33 wurde in der Kombination *Objektalter/Wohnfläche/Zimmeranzahl/Etage/Netzwerkdistanz* erreicht. Mit diesem Kovariablenset ist das Vorhersagemodell also untauglich. Wird die Netzwerkdistanz  $R^2$  als Lageparameter aus der GWR herausgenommen, erreichten drei Modellläufe Werte von größer als 0,5. Der bestmögliche  $R^2$  Wert wurde jedoch in einer Kombination mit der Netzwerkdistanz erreicht.

---

<sup>25</sup> Eine API, die den selektiven Datendownload aus der OpenStreetmap ermöglicht: <http://www.overpass-api.de/>.

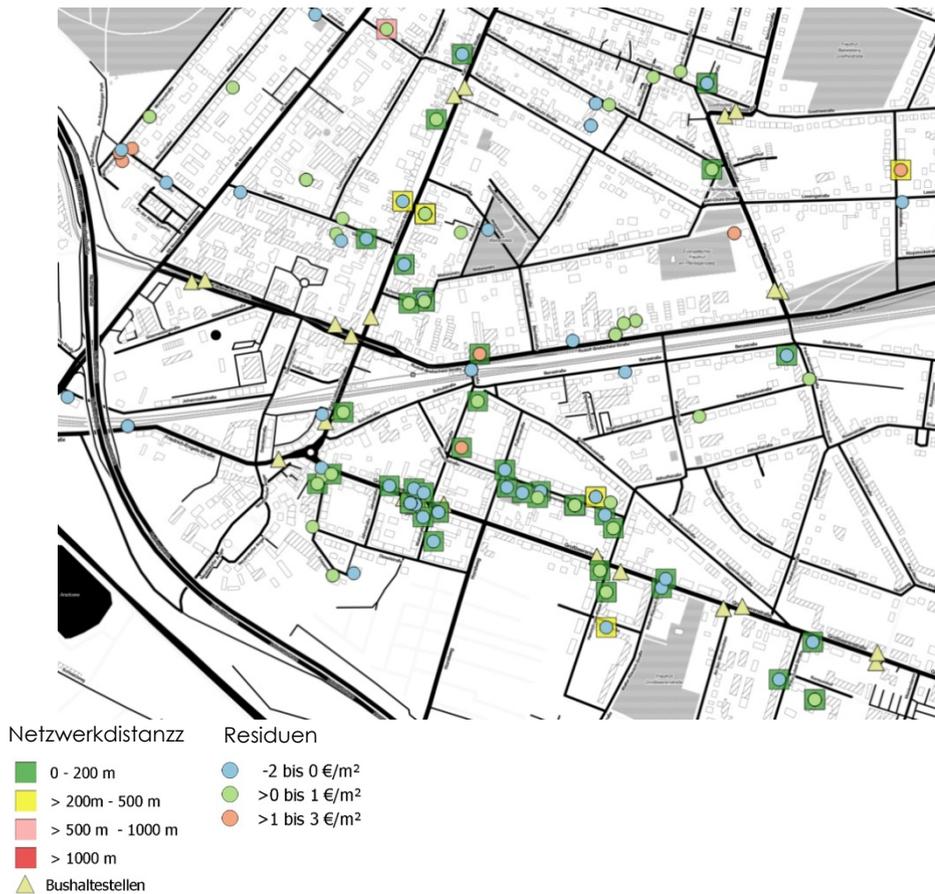


Abb. 5-11: Netzwerkdistanz von Mietangeboten zu Bushaltestellen und Residuen der Preismodellierung mittels GWR. (Entwurf und Darstellung: Scherthanner 2015).

Tabelle 5-5 zeigt die  $R^2$  Werte der 5 Modellläufe:

$R^2$	Ko-Variablen-set
0,33	Objektalter, Wohnfläche, Zimmeranzahl, Etage, Netzwerkdistanz
0,58	Objektalter, Wohnfläche, Zimmeranzahl
0,60	Wohnfläche, Etage, Netzwerkdistanz
0,61	Wohnfläche, Objektalter, Zimmeranzahl, Etage
0,63	Wohnfläche, Zimmeranzahl, Etage, Netzwerkdistanz

Tabelle 5-5:  $R^2$  bei fünf Modellläufen.

$R^2$  lässt sich räumlich betrachten. Eine lokale Betrachtung des  $R^2$  ergab große Variationen. Hierbei ist nicht nur eine omnidirektionale Betrachtung des Fehlerwerts, sondern eine Betrachtung in geographische Richtung möglich. Als weiteres Gütemaß diente die Standardabweichung des beobachteten vom

vorhergesagten Wert. Eine Betrachtung der Differenz von beobachtetem Wert zu geschätztem Wert zeigt jedoch einen geringen durchschnittlichen Fehlerwert von nur 0,2 €/m<sup>2</sup>. Eine differenziertere Auswahl des Ko-Variablensets ist daher als weitere Betrachtung von GWR unabdingbar. Die Residuen des GWR-Modells wurden im Raum dargestellt. Abbildung 5-11 zeigt die Residuen (Punkte) und die an den Angebotsdaten berechneten Netzwerkdistanzen (Quadrate). Aus der Abbildung wird ersichtlich, dass an einigen Datenpunkten keine Netzwerkdistanzen gemessen wurden. Das verwendete GWR-Modell weist im Rahmen des raumanalytischen Anwendungsfalles ein Anwendungspotential zur Optimierung der Mietpreisschätzung auf. Eine Anwendung im Kontext von Immobilienportalen bedarf jedoch noch mehr Feinabstimmung hinsichtlich der Parameterwahl.

In einem zweiten exemplarischen Versuch wurden die Angebotsdaten zusätzlich nach Gebäudestruktur (z. B. Einzelhaus, Reihenhaus etc.) eingeteilt. Außerdem wurden Daten aus dem Bereich Infrastruktur (z. B. vorhandene Apotheken) in die GWR eingebunden. Infolge dieser Erweiterungen lag der errechnete R<sup>2</sup> bei 0,93, was sich jedoch nicht in der Standardabweichung widerspiegelt. Diese lag in diesem Modelllauf bei > 2 €/m<sup>2</sup>.

### *Validierungsverfahren und Gütemaße*

Um die Eignung der ausgewählten Schätzverfahren im Rahmen des raumanalytischen Anwendungsfalles zu prüfen, ist es notwendig, die Güte der Mietpreisschätzungen bei variierenden Parameterkombinationen zu messen. Hierzu wurden alle getesteten Schätzverfahren einer Kreuzvalidierung unterzogen. Für die Durchführung der Kreuzvalidierung und Bestimmung der Gütemaße erfolgte eine programmiertechnische Umsetzung mittels R-Skripten.

Die entstandenen Evaluierungsskripte dienen dazu, die Validierung und die Gütemessung der Schätzverfahren automatisiert durchzuführen. Zur Validierung der getesteten Schätzverfahren wurde die Methode der Kreuzvalidierung (k-fold Crossvalidation) angewendet. Bei der Kreuzvalidierung werden aus der Menge aller Stützpunkte n Untermengen entfernt. Danach werden die übriggebliebenen Stützpunkte mittels des gewählten Schätzverfahrens (Bivand, 2013) interpoliert und die entfernten Untermengen als Prüfgrößen genutzt. k gibt dabei an, in wie viele Untermengen der Datensatz unterteilt wird. Im Rahmen des raumanalytischen Anwendungsfalles wurde einheitlich eine fünffache Kreuzvalidierung (5-fold Crossvalidation) angewendet (d. h. k = 5).

Basierend auf der Literaturrecherche (vgl. Tab. 6 sowie Lee und Heap (2008) bzw. Hengl (2009)) werden statistische Gütemaße zur Bewertung der Modelle ausgewählt. Hengl (2009) nennt im Zuge der Modellauswahl die entscheidenden Kriterien: Gesamtgenauigkeit, Verzerrungen des Modells, Sensitivität und Zuverlässigkeit des Modells und den Rechenaufwand.

Schlussendlich wurden im Rahmen des raumanalytischen Anwendungsfalles folgende Verfahren zur Modellevaluierung angewendet:

1. Mean Error (ME): Mittlerer Fehler als arithmetischer Mittelwert der Schätzfehler =  $1/n \sum e_t$
2. RMSE (Root Mean Square Error): Quadratwurzel des Mean Squared Errors (MSE), wobei der MSE das arithmetische Mittel der Summe der Quadrate der Schätzfehler angibt.  
 $MSE = 1/n \sum e_t^2$   
 $RMSE = \sqrt{MSE}$
3. Mittlere Relative Abweichung in %: Als weiterer vom ME abgeleiteter Fehlerwert.
4. Anteil der Datenpunkte Abweichung < 10 %: zeigt in % wie groß der prozentuale Anteil der Daten, die eine Fehlerabweichung <10 % aufweisen, an den Gesamtdaten ist.
5. Rechendauer der Modelle in Sekunden.

Der ME dient zur Anzeige des Grades der Fehlerverzerrung (Lee und Heap, 1998). Die Verwendung des ME als Indikator der Genauigkeit ist durchaus kritisch zu sehen. Entgegenwirkende negative und positive Residuen können den ME geringer ausfallen lassen, als der eigentliche Fehler groß ist (Nalder und Wein, 1998 zitiert nach Lee und Heap, 2008).

Der RMSE gilt neben dem Mittleren Absoluten Fehler also als das beste taugliche Gesamtmaß der Modellgüte räumlicher Schätzverfahren (Willmott, 1982 zitiert nach Lee und Heap, 2008). Der RMSE ist jedoch sehr sensibel gegenüber Ausreißern, denen der RMSE eine sehr hohe Gewichtung zuweist (Lee und Heap, 2008).

Die abgeleiteten statistischen Gütemaße erlauben einen Vergleich der Methoden bei unterschiedlichen Parameterkombinationen. Hauptziel der Implementierung ist der Test, der die Sensitivität der einzelnen Methoden bei unterschiedlichen Nugget-/Sill-/Range-Parameterkombinationen bestimmt.

### *Prätest der ausgewählten Verfahren*

Vor der Teilimplementierung der gewählten Schätzverfahren auf das ausgewählte exemplarische Untersuchungsgebiet erfolgte ein Prätest zur Feststellung der Eignung der ausgewählten Schätzverfahren. In diesem von Scheele und Schernthanner (2014) durchgeführten, dem eigentlichen Schätzverfahren vorangestellten Test (Prätest) wurden 637 Anzeigedaten für Mietwohnungen innerhalb eines 4,5 km<sup>2</sup> großen rechteckigen Testgebietes im Berliner Ortsteil Kreuzberg aus dem dritten Quartal 2007 ausgewählt. Der Untersuchungsraum und der Angebotsdatensatz wurden zu Testzwecken bewusst kleingehalten. Die räumliche Autokorrelation wurde bei dem Angebotsdatensatz mittels der Verfahren Geary's c und Moran's I festgestellt, sie lag bei 0,35 (Geary's C) sowie 0,67 (Moran's I). Insgesamt wurden in diesem Vorabtest 5.130 Modellläufe für die Verfahren OK und KED und 45 Modellläufe

für das Verfahren OCK durchgeführt. Der erste Test zeigte, dass alle geostatistischen Schätzverfahren bessere Schätzergebnisse als die hedonische Regression lieferten. Einzig in der Rechengeschwindigkeit ist die hedonische Regression den anderen Verfahren überlegen. Die Rechendauer lag bei etwa einer Minute. Im Gegensatz dazu hat OCK sehr lange Rechenlaufzeiten, die bei dem kleingehaltene Testdatensatz bei > 100 Minuten lagen. Die Güteschätzung der OCK war jedoch trotz der hohen Rechenlaufzeiten kaum valider als bei KED und OK (siehe Abbildung 5-12).

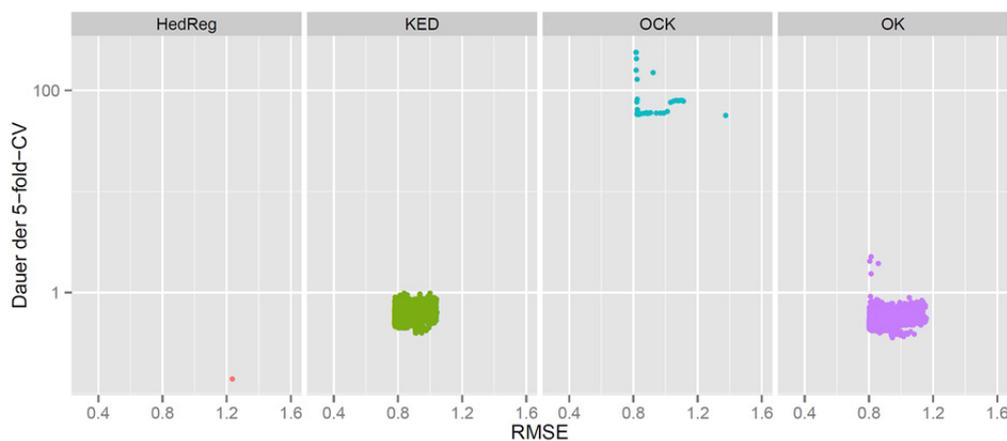


Abb. 5-12: Güte (RMSE) und Rechendauer der Schätzverfahren bei einer 5-fold Cross Validation (Scheele und Schernthanner, 2014).

### 5.1.5 Teilimplementierung der Modellevaluierung für die Anwendung für Immobilienportale

Ziel des Modellvergleichs ist es, die Sensitivität der Nugget-/Sill-/Range-Parameter von (OK, KED und OCK) in Kombination mit geeigneten Kovariablensets in ihrer Anwendung auf ein größeres sich an den realweltlichen Verhältnissen eines Immobilienportals orientierenden exemplarischen Untersuchungsgebietes zu überprüfen. Dies geschieht im Vergleich zu dem Standardverfahren der HedReg. In der Teilimplementierung der Verfahren ist es notwendig, die drei Parameter Nugget, Sill und Range an die exemplarischen Angebotsdaten anzupassen. Des Weiteren muss für unterschiedliche Kriging-Modelle geklärt werden, wie sensitiv die Parameter auf unterschiedliche Konfigurationen reagieren. Es erfolgte eine programmiertechnische Adaption der in den Prättests durchgeführten Validierungen auf den Untersuchungsraum Potsdam. Alle programmiertechnischen Umsetzungen in R greifen auf Bivands R-Bibliothek „gstat“ zurück. Das konstruierte Testumfeld erlaubt es, unter Nutzung einer in den Skripten vorhandenen Schleife, die Nugget-/Sill-/Range-Parameterkonfigurationen automatisiert in beliebigen Konfigurationen zu validieren. Somit ist eine Prüfung der Modellvalidität über theoretisch unendlich viele Modellläufe gegeben. Die Validierungsergebnisse werden anschließend in einzelne CSV-Dateien geschrieben, welche mittels weiterer R-Skripte zu einer CSV-Datei kombiniert und abschließend im Excel-Tabellenformat exportiert werden.

Um die Angebotsmietpreise im ausgewählten Untersuchungsraum bestmöglich zu interpolieren, wurden nur die validesten Verfahren mit den validesten Parameterkombinationen angewendet. Der schematische Ablauf der Evaluierung und der Übertragung der geeignetsten Parameterkombinationen zur Angebotsmietpreisschätzung wird in Abbildung 5-13 gezeigt.

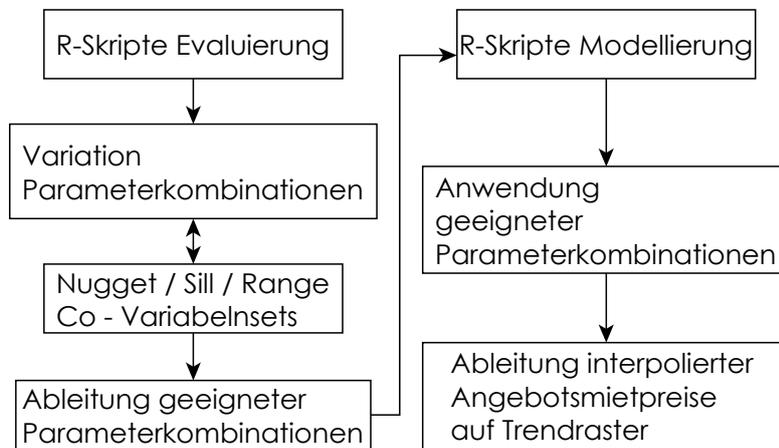


Abb. 5-13: Schematischer Ablauf der Evaluierung der Schätzverfahren und der Übertragung der Ergebnisse auf die räumliche Interpolation. (Entwurf und Darstellung: Schernthanner 2015).

In jedem Evaluierungsdurchlauf wird jeweils der Nugget, Sill und Range der vorab gewählten exponentiellen, sphärischen und Gauß'schen Modelle auf deren Güteparameter getestet.

### Resultate des Modellvergleichs

Nach 120 Modellläufen konnten bei den Methoden OK und KED keine signifikanten Änderungen der Güteparameter mehr beobachtet werden. OCK wurde in neun Kombinationen bei fix vorgegebenen Nugget- und Sill-Werten evaluiert. Der Grund für die geringere Anzahl der Evaluierungen von OCK liegt in der Rechenzeit. Von weiteren Testläufen wurde abgesehen, da sich bei der Anwendung auf das exemplarische Untersuchungsgebiet in den Güteparametern keine signifikanten Steigerungen der Validität mehr erreichen ließen. Schlussendlich konnten durch die Evaluierungen Verfahrensempfehlungen für einen Untersuchungsraum mit sehr unterschiedlichen Datendichten gegeben werden.

Der methodenübergreifende Modellvergleich zeigt, dass sämtliche evaluierten Parameterkombinationen aller verglichenen Kriging-Verfahren bessere Schätzgüten als die nicht räumliche HedReg liefern. Die Ergebnisse fielen jedoch insofern überraschend aus, als dass das Verfahren OK ohne die Verwendung von Sekundärvariablen in allen Parameterkombinationen validere Ergebnisse als KED und OCK lieferte. Eine Variation der Kovariablensets ergab dabei kaum validere Ergebnisse.

Nachfolgend werden die drei modellübergreifend besten Parameterkombinationen (alle OK) im Vergleich mit den validesten Modellläufen der Verfahren OK, KED und HedReg angeführt.

Das valideste Modell ist ein exponentielles OK-Modell mit der Parameterkonfiguration: Nugget 0,001/Partial Sill 0,1/Range 700 m. Am sensitivsten reagierten die Modelle auf Änderungen der Reichweite der angewendeten Kriging-Schätzverfahren.

Beobachtungen im Prätest hatten gezeigt, dass bei sphärischen OK-Modellen die beste Schätzgüte bei höheren Range-Werten (ab 1 km) erreicht wurde. Entgegen dieser Beobachtung wurden bei der Adaption auf den Untersuchungsraum Potsdam validere Ergebnisse bei mittleren Range-Werten und exponentiellen Modellanpassungen erreicht. KED fiel mit etwas geringeren Gütewerten als OK bei sehr geringer Rechenlaufzeit auf. HedReg war in der Mehrheit der Parameterkombinationen weniger valide als die vergleichenden Methoden. OCK hatte durchwegs schlechtere Gütewerte als die OK und KED bei einer vielfach höheren Rechenlaufzeit. Eine Verbesserung der Gütewerte von OCK ließ sich auch durch die Verwendung geänderter Sekundärvariablensets nicht erreichen. Die Verwendung aller als geeignet identifizierter Sekundärvariablen führte zu Rechenzeiten von bis zu 10 h, jedoch ohne eine Verbesserung der Modellgüte zu erreichen. Auch eine Reduktion auf die laut PCA und RF gewichtigsten zwei Variablen brachte keine Verkürzung der Laufzeit bei gleichbleibend geringer Modellgüte (ME 0,11 und RMSE 1,235). *Chica-Olmo (2007)* und *Montero und Laraz (2011)* erzielten in einem Modellvergleich zwischen OCK und OK mit OCK bessere Ergebnisse. Die verwendete Datenstruktur war jedoch bei *Chica-Olmo (2007)* und *Montero und Laraz (2011)* durch die Verwendung weiterer Datenpunkte, an denen Sekundärvariablen vorhanden waren, wesentlich komplexer als die Datenstruktur des exemplarischen Angebotsdatensatzes. Trotz großer Sorgfalt und der Anwendung einer PCA + RF zur Auswahl der Sekundärvariablen führte OCK zu keinen besseren Ergebnissen in der Modellierung. Von einer Anwendung von OCK zur Interpolation von Angebotsdaten, die direkt von Immobilienportalen gewonnen wurden, ist aufgrund der Natur der Datensätze (vielfach fehlende Dateneinträge und somit vielfach fehlende Sekundärvariablen) abzuraten. Folgende Tabelle 5-6 gibt eine tabellarische Übersicht der Ergebnisse der Modellevaluierung.

Me- thode	Nugget	Par- tial Sill	Ran- ge	Modell	ME	RM SE	Abwei- chung < 10 % [%]	Rechen- dauer (sek)
OK	0,001	0,1	700	Expo- nentiell	0,04 0	0,78 9	84,448	45,480
OK	0,001	0,1	600	Expo- nentiell	0,04 0	0,78 9	84,646	50,440
OK	0,001	0,09 5	600	Expo- nentiell	0,04 0	0,78 9	84,613	60,030
KED	0,001	0,06	200	Expo- nentiell	0,05 1	1,06 9	70,887	4,250
OCK	Alle Sekun- därvariablen		1100	Sphä- risch	0,10 9	1,20 7	58,226	34350,440
Hed Reg					0,14 3	1,56 3	42,381	0,250
OCK	Auswahl Se- kundärvariab- len		700	Expo- nentiell	0,14 0			39600 (11 h)

Tabelle 5-6: Ergebnisse der Modellevaluierung.

### *Interpolation der Angebotsmieten mit den validesten Parameterkombinationen*

Unter Anwendung idealer Güteparameter wurden die Angebotsmietpreise im exemplarischen Untersuchungsraum interpoliert. Abbildung 5-14 stellt den Status quo einer hedonischen Regression, ermittelt für eine beispielhafte „Infas Geodaten“ Nachbarschaft im Stadtteil Babelsberg, den interpolierten Ergebnistrastern gegenüber. Im Vergleich zu den interpolierten Ergebnistrastern kann in der mittels HedReg ermittelten Preiskarte des Portals IS24 nur ein Preis für eine Nachbarschaft innerhalb der identischen 90 Hektar großen Raumeinheit gesehen werden. In dem interpolierten 50-m-Ergebnistrastern sind wesentlich feinere Differenzierungen innerhalb derselben Nachbarschaft zu erkennen. Anzumerken ist, dass die Rechendauer einer Interpolation mittels OK bei nur drei Minuten und bei KED bei sieben Minuten lag. Demgegenüber stand die Rechendauer für die Berechnung eines Rasters mittels OCK mit einer Rechendauer von elf Stunden (zehn Stunden für die Evaluierung) und weniger validen Gütemaßen. Diese Beobachtungen entsprechen den Beobachtungen von Monterro und Larraz (2011).

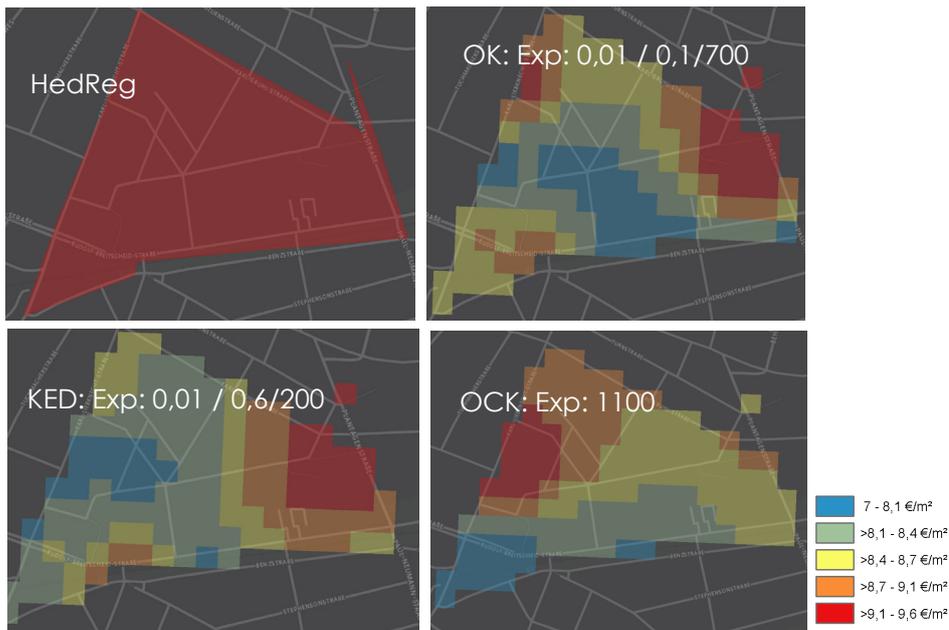


Abb. 5-14: Vergleich interpolierter Angebotsmietpreise (50-m-Grid) mit einer hedonischen Regression in Bezug auf die Fläche einer 90 Hektar großen Infas-Nachbarschaft. Hinter dem gewählten Kriging-Schätzer finden sich die Angabe der gewählten Modellfunktion sowie die Nugget-/Sill-/Range-Werte. (Entwurf und Darstellung: Schernthanner 2015).

## 5.2 Anwendung der Methoden des Maschinellen Lernens

Die Verfahren des Maschinellen Lernens wurden erneut mittels der freien, quelloffenen für statistisches Rechnen geeigneten Programmierumgebung "R" umgesetzt. Zur Anwendung kam hierfür die bereits für die Teilimplementierung von RF genutzte R-Bibliothek *Rattle*. Die im ersten Schritt mit *Rattle* erstellten Skripte wurden anschließend in R weiter angepasst und verfeinert. *Rattle* ist nicht die einzige quelloffene Datamining-Software, bietet jedoch durch die direkte Integration in R viele Vorteile. Als wichtigster Vorteil ist hierbei die Möglichkeit der Nutzung in Kombination mit anderen R-Paketen zu nennen. Alter-

nativen zu *Rattle*<sup>26</sup> sind beispielweise das auf der Programmiersprache Java basierende *Weka* (*Weka 3*, 2014) oder der von der Universität Yale entwickelte *Rapid Miner* (*Rapid Miner*, 2014).

### 5.2.1 Teilimplementierung von RF zur Mietpreisschätzung

RF wurde exemplarisch als Regressionsmodell zur Schätzung der Kaltmiete pro Quadratmeter angewendet. Im Rahmen der vom Autor durchgeführten Teilimplementierung von RF wurde den Empfehlungen *Breimans* (2001) gefolgt und es wurden Samples zu Testzwecken des RF zurückgelegt. Eine schematische Darstellung der Teilimplementierung von RF folgt in Abbildung 5-15.

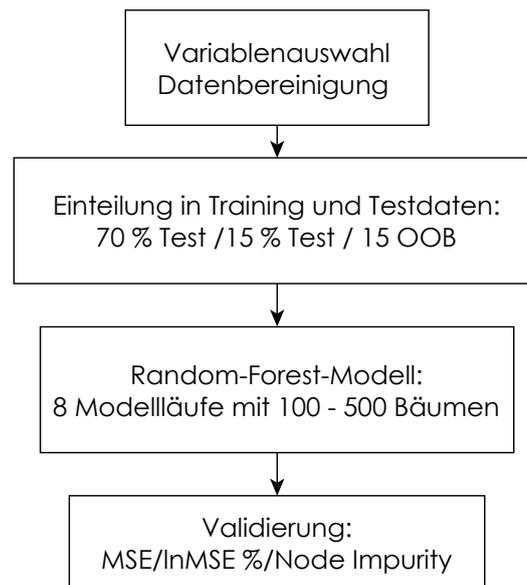


Abb. 5-15: Schematischer Ablauf RF. (Entwurf und Darstellung: *Schernthanner* 2015).

Der Datensatz wurde – im Verhältnis 70 % Testdaten, 15 % Trainingsdaten und 15 % zurückgelegte Daten – geteilt. Die zurückgelegten Daten werden bei RF-

---

<sup>26</sup> Die Webseite: <http://www.predictiveanalyticstoday.com/top-15-free-data-mining-software/> bietet eine Übersicht über freie Software für Maschinelles Lernen und Datamining.

Modellen in der Regel zur Ermittlung des Out-of-bag-Fehlers genutzt, einem Leistungsmaß zur Beurteilung des RF-Modells. Randomisiert wird das Modell bei der Auswahl der Trainingssamples und bei der Entscheidung, an welcher Stelle im Merkmalsraum am besten geteilt bzw. partitioniert wird. Zur Bildung der randomisierten Entscheidungsbäume wurden 14 Variablen des Datensatzes des ersten Quartals 2013 ausgewählt. Künstliche Regionscodes, vorhandene logarithmische Variablen und potenziert vorliegende numerische Variablen wurden bereinigt. Die abhängige Variable war die Kaltmiete pro Quadratmeter. Der Datensatz wurde in zehn Modellläufen in 50er-Schritten von 100 – 500 Entscheidungsbäumen modelliert. Die Berechnung mit 500 Bäumen und die Nutzung von minimal 100 Bäumen richteten sich hierbei nach den Empfehlungen von *Breiman* (2001). Als Maß der Gesamtgüte der RF-Modelle dient der "Mean squared error" (MSE), als das gemittelte Residuum des Vorhersagewertes zum Quadrat bzw. der mittlere quadratische Fehler. Als weiteres Gütemaß wird, basierend auf den out-of-bag-Samples, für jeden Modelllauf der logarithmische MSE in % (lnMSE%) als die Abnahme der Vorhersagegenauigkeit je Variable berechnet. Die Berechnung des lnMSE% erfolgt unter der Annahme, dass eine Variable aus dem Modell entfernt wird. Der als lnMSE% ermittelte Fehlerwert kann als Faktor für die Beurteilung der Wichtigkeit einer Variablen in einem RF-Modell dienen. Als weiteres Gütemaß wird die „Node Impurity“ – die Knotenunreinheit als Maß der Heterogenität innerhalb eines Knotens – gemessen. Die durchgeführten Modellläufe zeigten, dass die Fehlerraten der Prognosen sehr schnell abnehmen, was schon bei ca. 50 randomisierten Bäumen der Fall ist (siehe Abbildung 5-16). Der MSE nimmt danach nur noch sehr gering ab. Der beste MSE von 1,1385 wurde bei 150 Bäumen erreicht (siehe Tabelle 5-7).

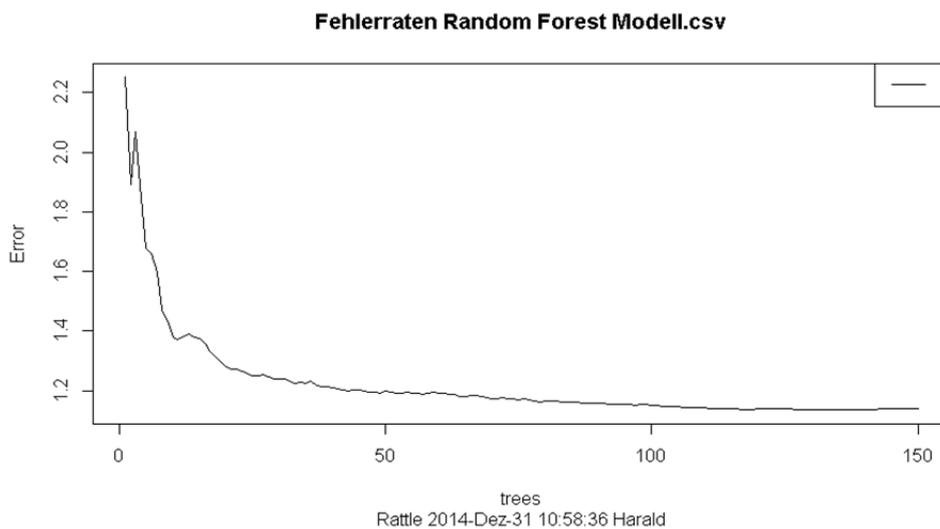


Abb. 5-16: Kurve der Fehlerrate von RF-Modellen bei zunehmender Baumanzahl. (Entwurf und Darstellung: *Schernthanner* 2015).

n tree	150	350	400	300	450	250	500	200	100
MSE	1,1385	1,1391	1,1392	1,1397	1,1404	1,1407	1,1412	1,1436	1,1508

Tabelle 5-7: Aufsteigender gemittelter quadratischer Fehler der RF-Modelle bei entsprechender Baumanzahl.

Wird die Variablengewichtigkeit mittels des  $\ln\text{MSE}\%$  betrachtet, ergibt sich, dass das Baujahr, gefolgt von der kategorialen Variable „Einbauküche“ die höchste Gewichtigkeit aufweist. In Anhang 7 findet sich der Fehlerreport des besten RF-Modellaufs. Insgesamt ist eine stärkere Unter- als Überschätzung der durch RF modellierten Angebotspreise zu beobachten. Abbildung 5-17 zeigt die Differenz von tatsächlichen zu RF-geschätzten Angebotspreisen in €/m<sup>2</sup> für einen Raumausschnitt des Potsdamer Stadtteils Babelsberg.

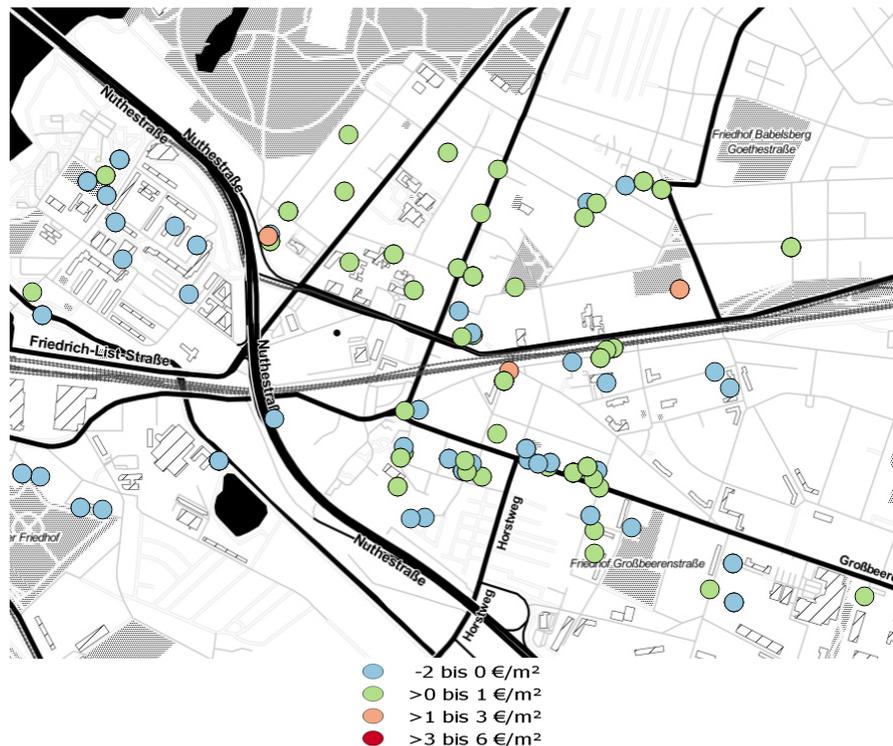


Abb. 5-17: Differenz realer zu RF-geschätzten Angebotspreisen an visualisierten Angebotsstandorten für einen Raumausschnitt in Babelsberg/Potsdam. Die Punkte zeigen die jeweiligen Wohnungsstandorte. Kartenbasis: *Stamen OpenStreetmap* Stil. (Entwurf und Darstellung: *Schernthanner* 2015)

RF kann als Regression genutzt werden, um den Mietpreis für Datenpunkte zu schätzen, deren Sekundärvariablen bekannt, deren Mietpreis jedoch unbekannt ist. Diese Nutzungsmöglichkeit birgt im Rahmen des ersten Anwendungsfalles das Potential, Stützpunkte für eine räumliche Interpolation in Untersuchungsgebieten mit einer geringen Angebotsdatendichte zu generieren. Diese mittels RF generierten Stützpunkte können später in einer räumlichen Interpolation weiterverwendet werden. Testweise wurde für sechs künstlich geschaffene (nicht reale) Angebote und eine willkürliche Vergabe von Sekundärvariablen ein Mietpreis als „Scorewert“ ermittelt. Dieses Vorgehen ermöglicht es, das Potential einer Stützpunktgenerierung aufzuzeigen (siehe Anhang 6).

RF verhält sich im Allgemeinen äußerst robust gegenüber Ausreißern und fehlenden Datenwerten. Somit eignet sich RF sehr gut zur Modellierung von Da-

ten/Datensätzen, in denen sehr oft Datenlücken oder Ausreißer vorhanden sind. Dies trifft insbesondere auf Datensätze zu, die von einem Immobilienportal kommen.

## 5.2.2 Teilimplementierung von ANN zur Mietpreisschätzung

Zur exemplarischen Anwendung wurde wiederum das R-Paket Rattle verwendet. Rattle nutzt zur Bildung des ANN die R-Bibliotheken multinom und nnet. Bei der Berechnung des ANN werden nur numerische Daten berücksichtigt. Dies kann als Nachteil gegenüber dem RF-Verfahren angesehen werden. Im vorliegenden Fall werden zur Bildung des ANN die Variablen: Verweildauer der Anzeige/Wohnfläche/Baujahr/Objektalter/Zimmeranzahl und Etage verwendet, um als Antwort vom Netz die Zielvariable „Miete in €/Quadratmeter“ zu bekommen. Wie bei der Anwendung des RF-Modells wurde der Datensatz im Verhältnis 70 % : 15 % : 15 % von Test-, Trainings- und Validierungsdaten unterteilt.

Als Validierungsmaß wird das Pseudo  $R^2$ <sup>27</sup> der vorhergesagten versus der beobachteten Daten gemessen. Der im Ergebnis gemessene Pseudo  $R^2$  Wert von 0,192 ist sehr gering. Auch der Vergleich der modellierten mit den tatsächlichen Daten legt den Schluss nahe, ANN nicht als Lösungsansatz im Rahmen des Anwendungsfalles der räumlichen Analyse zu verwenden. Abbildung 5-18 zeigt die Differenz zwischen realen und ANN-modellierten Mietpreisen.

---

<sup>27</sup> Bestimmtheitsmaß, welches die Varianz einer abhängigen Variable beschreibt.

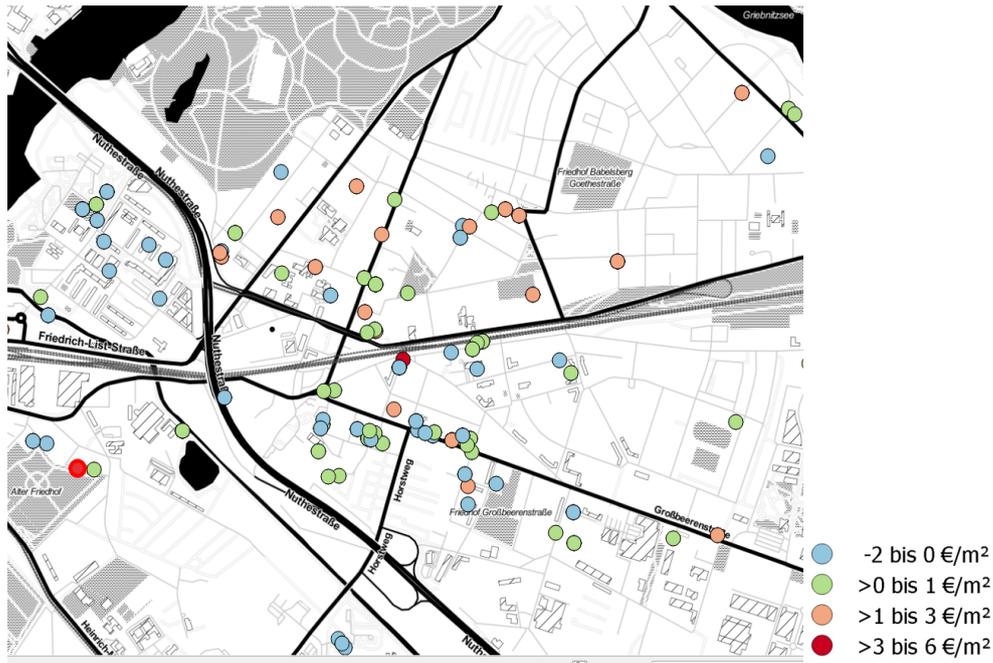


Abb. 5-18: Differenz realer zu ANN-modellierter Mietpreise. (Entwurf und Darstellung: Scherthanner 2015).

### 5.3 Teilimplementierung Geovisualisierungsmethoden

Um Schätzpreise adäquat darzustellen, müssen konzeptionelle Vorüberlegungen bezüglich der Anforderungen einer Mietpreiskarte getroffen werden. Diese Vorüberlegungen dienen als Ausgangsbasis des geovisuellen Anwendungsfalles, in dem funktionierende Prototypen von Mietpreiskarten erstellt werden. Gegenstand eines vom Verfasser der Dissertation geleiteten diesbezüglichen Projektes war die Entwicklung prototypischer „Mietpreis-Mashups“. Im Rahmen dieses Projektes entstanden insgesamt 19 funktionierende Mietpreis-Webkarten. Hierbei wurden die geschätzten Kaltmietenpreise jeweils in €/m<sup>2</sup> für eine 77-m<sup>2</sup>-Standardwohnung in Potsdam dargestellt. Die entstandenen Prototypen gehen von der Annahme aus, dass eine Wohnimmobilie neu vermietet wird, und zeigen den geschätzten Preis einer Wohnung in „Standardgröße“ bei einer möglichen Neuvermietung an.

Auch beim geovisuellen Anwendungsfall gibt es eine Vielzahl an Methoden und Techniken, die mit unterschiedlichsten Komponenten umsetzbar sind.

#### 5.3.1 Vorüberlegungen zu Map APIs

Eine Map API zur Umsetzung sollte möglichst externe TMS (Tile Map Services) Anbieter unterstützen und nicht auf einen proprietären Anbieter limitiert sein. Die API sollte möglichst quelloffen sein, eine Vielzahl adaptierbarer Grundfunktionen bieten und erweiterbar sein. Die bei Immobilienportalen meist im Einsatz befindliche „Google Maps“ API erfüllt diese Anforderungen nicht. Sie erlaubt ausschließlich die Verwendung von Grundkarten der Firma Google. Google Maps ist proprietär und eine Integration externer Dienste (z. B. Open geospatial

consortium<sup>28</sup> (OGC) konformen Services) ist laut Google nur in experimenteller Form möglich (Google, 2015). Die Google Maps API erfüllt die vom Autor formulierten Bedingungen nicht.

### 5.3.2 Vorüberlegung zu potentiellen Nutzergruppen von Mietpreiskarten

Wer nutzt Mietpreiskarten und welche Nutzergruppen gibt es? Nutzer sind einerseits private Wohnungssuchende, die sich über Preise informieren möchten. Andererseits lassen sich hier auch Akteure der Immobilienwirtschaft wie Immobilienanbieter nennen, die für wirtschaftliche Überlegungen, wie beispielsweise potentielle Investitionen in einem Mietmarkt, Preisbeobachtungen betreiben. Weitere Nutzer sind Makler, die sich über die Marktlage und mögliche Neuvermietungspreise informieren möchten. Durch den Ansatz der zonenlosen Preiskarte können neben den genannten Nutzergruppen zusätzliche Nutzergruppen erschlossen werden. Denkbar sind beispielsweise Nutzer der öffentlichen Verwaltung (Kommunen, Städte etc.), die Mietpreiskarten zum Monitoring ihres kommunalen Wohnungsmarkts nutzen.

### 5.3.3 Zonenlose Preiskarte

Ein wichtiger Aspekt bei der Darstellung einer zonenlosen Preiskarte als Grid-Map ist die Wahl der geeigneten Maschenweite des Rasters. Hier ist anzumerken, dass es per se keine ideale Maschenweite gibt und dass die Maschenweite immer vor dem Hintergrund der geplanten Anwendung, in diesem Fall der Erstellung einer Mietpreiskarte, zu sehen ist. Im vorliegenden Fall muss sowohl die Zellgröße des zu interpolierenden Rasters als auch der Darstellungsmaßstab des interpolierten Rasters in einer Webkarte gefunden werden.

GIS-Softwarepakete wie das proprietäre ArcGIS nutzen zur Berechnung der Zellgröße einen pragmatischen, wissenschaftlich aber nicht nachvollziehbaren Ansatz. Demnach wird die Ausdehnung eines Untersuchungsraumes per „default“ durch den Wert 250 geteilt und somit eine entsprechende Ma-

---

<sup>28</sup> Die OGC ist laut Selbstauskunft eine Konsortium, bestehend aus 511 Mitgliedern (Stand 07.02.2015), mit dem Ziel, allgemein zugängliche Standards der Interoperabilität von Geodaten zu schaffen (OGC, 2015).

schenweite festgelegt. Hengl (2010) schlägt für unterschiedliche Anwendungen verschiedene Herangehensweisen zur Ermittlung einer optimalen Maschenweite von Gridzellen vor. Für den geovisuellen Anwendungsfall wird eine Herangehensweise Hengls (2010) adaptiert, indem unter Einbeziehung der räumlichen Datendichte, des Nuggets und der ermittelten räumlichen Autokorrelation nutzbare Maschenweiten bestimmt werden. Der Autor folgt dabei Hengls (2010) Überlegungen, dass es keine „ideale“ Gridzellengröße gibt. Im Rahmen der Arbeit werden drei mögliche Grid-Auflösungen – eine feinstmögliche, mittlere sowie eine grobe Auflösung – vorgeschlagen. Für die Ermittlung der drei Auflösungsstufen wurde ein R-Skript von Hengl (*Spatial-analyst.net*, 2010) angepasst und verwendet, um die Grundfläche durchschnittlicher Wohngebäude im exemplarischen Untersuchungsraum zu berechnen. Das vom Autor adaptierte Skript unterstützt bei der Ermittlung möglicher Zellgrößen. Unter Zugriff auf die R-Bibliothek „spatstat“ erfolgten eine Berechnung der Punktdichte, die Ermittlung der durchschnittlichen Distanz zwischen den vorhandenen Stützpunkten sowie der kumulativen Distanz zwischen den geokodierten Mietpreisangeboten.

Zur Ermittlung der durchschnittlichen Wohngebäudegrundfläche wurde der Datenbestand der OpenStreetmap genutzt. Mittels der Overpass API<sup>29</sup> wurden im Untersuchungsraum der OpenStreetmap diejenigen Flächen herausgefiltert, deren Nutzungsart als „Wohnen“ (Objektart „Residential“ in der OpenStreetMap) ausgewiesen war. Die gefilterten Flächen mit der Nutzungsart Wohnen wurden mit in der OpenStreetmap der im Untersuchungsraum vorhandenen Gebäudeblöcke verschnitten. Von den hieraus resultierenden Gebäudeblöcken wurde die durchschnittliche Grundfläche ermittelt, die dann als Basis der Maschenweite der Gridzellen dient.

Die durchschnittliche Distanz zwischen den Angeboten liegt bei 110 m, bei einer gleichzeitigen Punktdichte von 8,22 Angeboten pro km<sup>2</sup>. Die durchschnittliche Grundfläche von Wohngebäuden im Untersuchungsraum beträgt 247 m<sup>2</sup>, was, auf die Zellgröße umgelegt, eine Maschenweite von 15,72 m ergeben würde. Bei einem so feinen Grid sind sehr lange Rechenzeiten zu erwarten. Die Berechnung der kumulativen Distanz zeigte, dass 90 % der Angebote

---

<sup>29</sup> Die Overpass ist eine API zur gezielten Filterung nach OpenStreetMap Daten (*Openstreetmap*, 2015).

einen Abstand von höchstens 20 m aufweisen, alle weiteren Angebotspunkte sind weiter voneinander entfernt.

Als weitere Näherung an die ideale Maschenweite wurde die Grundfläche Potsdams durch die Anzahl der Gebäudeblöcke mit Wohnnutzung (14.079 Blöcke) geteilt. Dies ergab eine Maschenweite von 115,55 m. Mit Hilfe des Range-Wertes des validesten OK-Modells wurde eine weitere Näherung vorgenommen, deren Resultat bei 700 m lag. 700 m ist die größte bestimmte Maschenweite, die jedoch die Anwendung eines räumlichen Interpolationsverfahrens als Alternative zur hedonischen Regression unmöglich macht. Der Abstraktionsgrad räumlicher Trends ist bei einer so großen Maschenweite zu hoch. Es ist anzunehmen, dass über diesem Wert eine signifikante Abnahme der Autokorrelation zu finden ist. Somit stellt der ermittelte Range-Wert den größten möglichen Maschenwert dar. Dies wird jedoch als zu grobmaschig für eine sinnvolle Darstellung erachtet.

Optimale Maschenweiten für die vorliegende Anwendung liegen also bei 15,75 m, 75 m und 110 m. Bei der Berechnung eines 15,75 m großen Grids sind Performanzprobleme hinsichtlich der Rechenzeit zu erwarten und die Darstellung eines solch feinmaschigen Grids scheint nicht mehr sinnvoll. Schlussendlich wurden in Anlehnung an die angestellten Berechnungen folgende Maschenweitenbereiche für unterschiedliche Anwendungen ausgewählt und Maßstäben zugeordnet (siehe Tabelle 5-8).

Zellgrößen (m)	Maßstab	Zoomstufe in Webmap (Maßstab entspricht Zoomstufen nach OpenStreetMap).
> 75 – 110	1 : 35 000 und 1 : 70 000	14 und 13
> 50 – 75	1 : 8 000 und 1 : 15 00	16 und 15
20 – 50	Nur zur Berechnung/ feinmaschige Grids	

Tabelle 5-8: Maschenweiten und mögliche Maßstäbe und Zoomstufen.

### *Prototypische Darstellung einer Mietpreiskarte*

Die folgende prototypische Darstellung (Abbildung 5-20) zeigt exemplarisch eine Mietpreistrendoberfläche, die mittels der Leaflet.js API und der Funktion „image overlay“ als Datenschicht eingebunden wurde. Die Darstellung ist an eine Heatmap angelegt und soll das Auffinden teurer bzw. günstiger Wohngegenden intuitiv ermöglichen. Abbildung 19 stellt einen mittels OK interpolierten Mietpreistraster als „Image Overlay“ in einer mit Leaflet.js entwickelten Webmap dar (Zoomstufe 14/1 : 35 000). Die Angebotsdaten aus dem ersten Quartal 2008 des Immobilienportals IS24 wurden hierfür mittels eines exponentiellen Ordinary-Kriging-Modells mit der Parameterkonfiguration: Partial Sill: 0.01, Nug-

get: 0,001 und Range: 700 m interpoliert. Der Mietpreisraster wurde mit der Overpass-API nach Flächen mit Wohnnutzung (landuse = residential) gefiltert. Das hat den Vorteil, dass der geschätzte Mietpreis ausschließlich für Flächen mit tatsächlicher Wohnnutzung visualisiert wird. Jedoch ist der OpenStreetMap-Datenbestand der Flächen mit Wohnnutzung für Potsdam lückenhaft. Vor allem der Stadtteil „Am Stern“ – ein sehr dicht besiedelter und relativ junger Stadtteil – ist nicht vorhanden. Eine Verbesserung der Visualisierung könnte grundsätzlich dadurch erreicht werden, dass nur diejenigen Flächen dargestellt werden, die tatsächlich besiedelt sind. Dies könnte durch den Einsatz von Landnutzungs- und Landbedeckungsdatensätzen wie den „Urban Atlas“ Daten 30 der „European Space Agency“ (ESA) erreicht werden. Dieser Ansatz wurde im Rahmen der vorliegenden Arbeit jedoch nicht weiter verfolgt.

Auffälliger in der Darstellung sind Probleme bei den Klassengrenzen. Prägnante räumliche Muster wie die niedrigen Mietpreise im Stadtteil „Am Schlaatz“ oder die hohen Mietpreise der „Berliner Vorstadt“ lassen sich in der Darstellung auf einen Blick erkennen. Die prototypische Mietpreiskarte wird dem Status quo einer Mietpreiskarte des Immobilienportals *IS24* gegenübergestellt (Abbildung 5-20). Die Zeitschnitte der Darstellung korrespondieren nicht.

---

<sup>30</sup> Urban Atlas ist eine Landnutzungs- und Landbedeckungsklassifikation der ESA für alle urbanen Räume Europas die mehr als 100.000 Einwohner haben (*European Environmental Agency*, 2015).

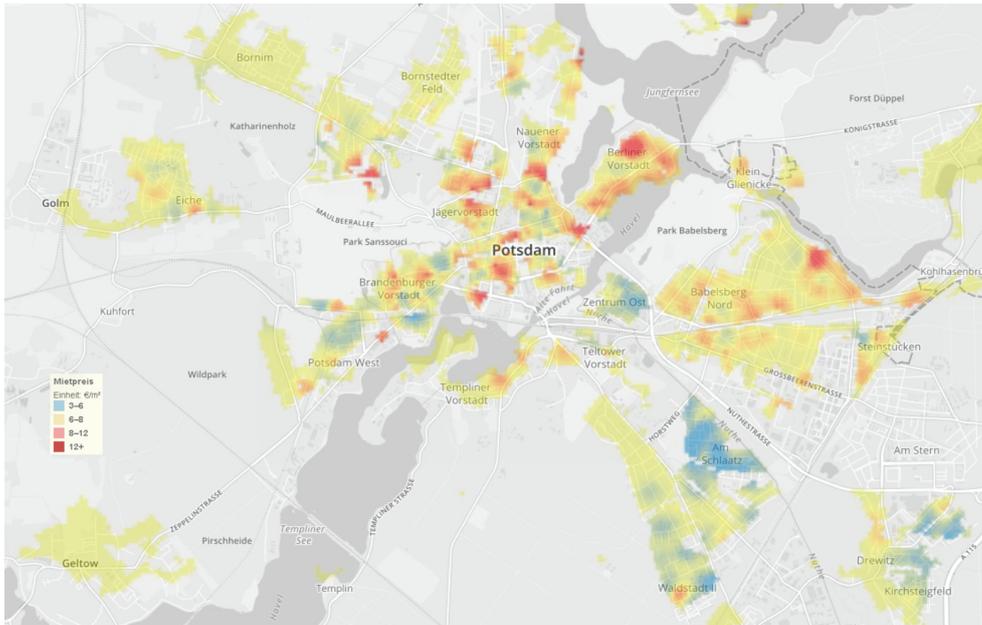


Abb. 5-19: Interpolierter 50-m-Raster (Entwurf und Darstellung: Schernthanner 2015).

### Mietpreise für Berliner Vorstadt

Hier finden Sie eine deutschlandweite Übersicht der aktuellen Miet- und Kaufpreise.  
Die Kartendarstellung unterscheidet sowohl Miet- und Kaufpreise als auch Wohnungen Einfamilienhäuser.

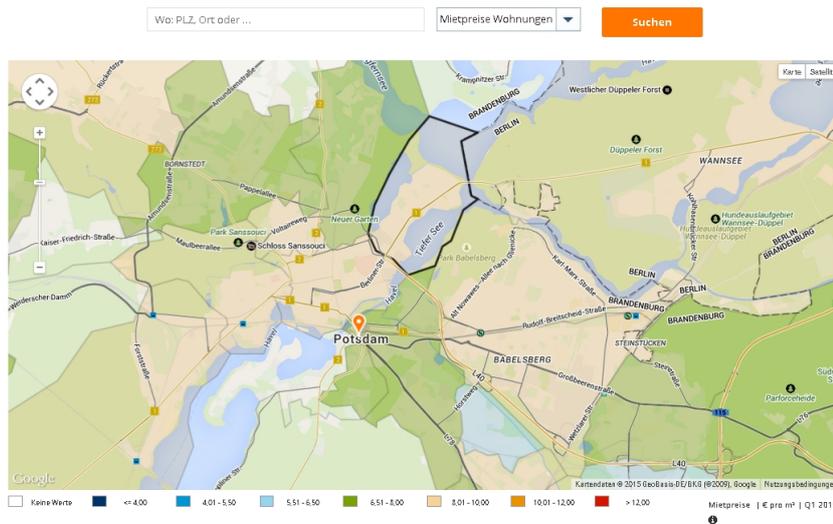


Abb. 5-20: Mietpreiskarte für Potsdam des Portals IS24 für das dritte Quartal 2013 (ImmobilienScout 24, 2014)

Für die Darstellung geschätzter Mietpreise in Maßstäben größer als 1 : 8.000 wird ein Bezug der geschätzten Mietpreise auf Gebäudeblöcke vorgeschlagen. Dies entspricht dem Einstiegsmaßstab eines typischen privaten Suchenden, der Immobilien oft sehr gezielt, in einer hohen Zoomstufe, in bestimmten Nachbarschaften sucht und sich sehr lokal nach Preisen informiert.

### 5.3.4 Prototypenentwicklung

Um auf Basis der bisherigen Erkenntnisse und unter Nutzung optimaler Methoden und Techniken die „bestmögliche“ Mietpreiskarte zu entwickeln, wurde vom Autor ein diesbezügliches Projekt initiiert und geleitet. Teilnehmer dieses Projektes waren Studierende des Masterstudiengangs Geoinformation und Visualisierung der Universität Potsdam. Im Rahmen des Projektes wurden „gebäudescharfe“ Mietpreiskarten erstellt, wozu Leaflet.js, CartoDB, Mapbox bzw. der Geoserver eingesetzt wurden. Neben der in den konzeptionellen Vorüberlegungen identifizierten, Map API Leaflet.js, bieten die Komponenten CartoDB, Geoserver und Mapbox weitere vielversprechende Funktionen zur prototypischen Teilimplementierung einer Mietpreiskarte. Als Datengrundlage im Rahmen des Projektes wurden geschätzte Mietpreise im Untersuchungsraum zum einen auf den Zentroid der Gebäudeblöcke und zum anderen als 50-m-Trendraster abgebildet.

Die entstandenen Mietpreiskarten wurden vom Autor auf folgende Gemeinsamkeiten hin untersucht:

1. zur Prototyp-Erstellung genutzte Komponenten,
2. Funktionen der Webkarte sowie
3. zusätzlich genutzte Plugins,
4. gewählte Grundkarte,
5. Darstellungsform,
6. Kartenschichten,
7. Klassifizierung der Daten,
8. Farbskalierung.

### 5.3.5 Analyse der Gemeinsamkeiten der Mietpreiskarten

#### *Zur Prototyp-Erstellung genutzte Komponenten*

Die eine Hälfte der während der Projektlaufzeit generierten Prototypen wurde mittels CartoDB, die andere Hälfte unter Nutzung von Leaflet.js umgesetzt. Die in Leaflet.js entstandenen Prototypen integrierten vielfach mit dem Geoserver bereitgestellte Web feature services (WFS). Geoserver zeigte hierbei eine besonders hohe Performanz für die Darstellung der Mietpreise auf Gebäudeblocklevel. Über das Format GeoJson erfolgte eine direkte Einbindung des Geoservers in den JavaScript Code. Eine besondere Herausforderung bestand darin, über 13.000 Vektorgeometrien mit Gebäudeblöcken mit einer Gesamt-speichergröße von 16 MB in den Browser zu laden. Trotz Kompression bzw. Überführung der Daten in Formate wie GeoJson war ein ruckelfreies schnelles Nachladen der Webkarten nicht möglich. Dieses Problem konnte nur durch die Bereitstellung der Daten als Geoserver-basierten WFS-Dienst und die Einbindung der Daten mittels Asynchronem JavaScript(Ajax) entschärft werden.

## Darstellungsform, Farbskalierung und Klassifizierung

In allen Prototypen werden die geschätzten Mietpreise auf die vorhandenen Gebäudeblöcke umgelegt dargestellt. Zur Unterstützung einer optimalen Farbwahl der Mietpreisklassen wurde im Seminar vielfach auf das Werkzeug des „Color Brewer“ von Brewer und Harrower (2014) der Pennsylvania State University zurückgegriffen. Die Klassifizierung der auf die Gebäudezentroide geschätzten Mietpreise wurde vielfach mittels der „Faustformel“ von Sturges.  $k = 1 + 3,32 \lg(n)$  (Schulze, 2007) ermittelt. Das Ergebnis würde jedoch 14 Klassen bedeuten. Die Mehrzahl der Prototypen wurde mit vier bis sechs Mietpreisklassen erstellt. Hinsichtlich der Klassenbreite wurden die Daten der Prototypen meist nach der Methode der natürlichen Unterbrechung von Jenks (Jenks & Caspall, 1971) unterteilt. Die Klassifizierung unter Nutzung der Leaflet API muss manuell oder vorab in einem Geoinformationssystem ermittelt werden. CartoDB stellt eigenständige Werkzeuge zur Klassifizierung bereit, diese sind jedoch nicht zufriedenstellend. Die Wertspannen zu den jeweiligen Klassen werden hierbei nur bei der größten und niedrigsten Klasse angezeigt, was eine effektive Nutzung der so entstandenen Webkarte nicht möglich macht (siehe Abbildung 5-21). Der Zeitraum der in Abbildung 5-21 visualisierten Mietpreise weicht von den übrigen Prototypen ab. Visualisiert werden mittels OK interpolierte Angebotsdaten aus dem ersten Quartal 2013.



Abb. 5-21: Visualisierung von Mietpreisen mit CartoDB. (Entwurf und Darstellung: Schernthanner 2015).

Leaflet.js weist ebenfalls Probleme bei der Darstellung der Klassenbreiten auf. Da keine Größer-/Kleiner-Zeichen (<>) verwendet werden können, ergeben sich zwangsläufig Überlappungen der Klassen. Die Farbassoziation für die Farbgebung der Preisklassen wurde meist durch einen Farbverlauf von Gelb für niedrigere Mietpreise nach Rot für höhere Mietpreise visualisiert. Mehrere Proto-

typen nutzten eine bipolare Farbskala von Grün oder Blau für niedrige nach Rot für hohe Mietpreise. Diese Darstellung lässt trotz des Nicht-Vorhandenseins einer zweigeteilten Datenreihe eine gute Differenzierung von niedrigen und hohen Nettokaltmieten zu. Abbildung 5-22 zeigt exemplarisch einen der entstandenen Prototypen.



Abb. 5-22: Mietpreiskarten-Prototyp. Als zusätzliche Datenschichten wurden Schuleinrichtungen (weißer Marker) und Zonen mit hoher Lärmbelastung (grün) eingeblendet. (Darstellung: Christopher Braune, 2014)

In einigen Fällen wurde neben den nach Mietpreisklassen farbkodierten Gebäudeblöcken ein interpolierter Trendraster mit 50-m-Maschenweite in die Webkarte aufgenommen. Dies wurde mit der Absicht verfolgt, das Finden von Mietpreistrends in mittleren Maßstäben (1 : 35.000 – 1 : 70.000) zu unterstützen.

### *Funktionen und zusätzliche Kartenschichten der Webkarte*

Alle Webkarten verfügen über Zoom-in- und Zoom-out Funktionen. Als zusätzliche Funktion wurde mehrheitlich eine „Mouse-Hoover“-Funktion verwendet, die bei der Bewegung der Maus über einen Gebäudeblock den geschätzten Mietpreis anzeigt. Als weitere Funktion nutzen viele Prototypen die Geolocation API (W3C, 2014), die auch von Leaflet.js angesprochen werden kann. Diese API liest die aktuelle Position des zur Kartenbetrachtung genutzten Endgerätes aus und passt den Kartenmittelpunkt an die jeweilige Position des Kartennutzers an.

Als zusätzliches Plugin nutzte die Mehrzahl der Prototypen einen Geokodierer zur Adresssuche. Die Projektteilnehmer erachteten diese Funktion als essentiell und bestätigten somit die in den Interviews gemachte Aussage, dass in den meisten Fällen adressgenau oder nach bestimmten Straßenabschnitten gesucht wird. Bei dem in die Prototypen integrierten Geokodierer halten sich die Nutzung von „Nominatim“ (OpenStreetMap, 2014) und des auf Nominatim aufbauenden Geocoder „Photon“ die Waage (Github, 2015). Beides sind freie Services, die über Adressangaben von OpenStreetMap geokodieren.

Bei zusätzlich genutzten thematischen Ebenen wurden mehrheitlich die Standorte von Bildungseinrichtungen und Haltestellen des ÖPNV (Bus- und S-Bahnhaltestellen) eingebunden. Teils wurden zusätzliche thematische Ebenen wie Arbeitslosigkeit als Choroplethenkarte, Ergebnisse einer Lärmkartierung oder vorhandene Parkflächen in die Webmap eingebracht. Einige Prototypen stellen den Mietpreis 2,5-dimensional als Z-Wert dar. Dies ermöglicht in Kombination mit einer Farbkodierung des Preises eine schnelle Erfassung des Preisgefüges. Für diese Darstellungsform wurde das Leaflet.js-Plugin „OSMBuildings“ (Osmbuildings, 2014) genutzt, das es ermöglicht, performant 2,5-D-Gebäude direkt in einem Webbrowser darzustellen (siehe Abbildung 5-23).

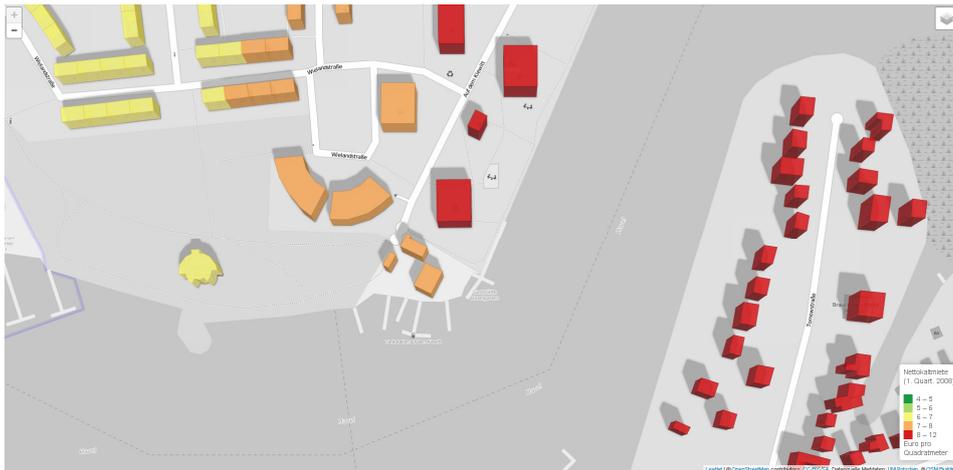


Abb. 5-23: Nettokaltmiete in €/m<sup>2</sup> als thematischer Z-Wert dargestellt mit dem OSM Buildings Plugin von Leaflet. (Entwurf und Darstellung: Scherthanner 2015).

Die prototypischen Abbildungen 5-24 und 5-25 (siehe Folgeseite) zeigen den Sprung von der Zoomstufe 17 zur Zoomstufe 18. Es erfolgt in der prototypischen Webmap ein Wechsel der Darstellung von der Visualisierung einer Mietpreistrendoberfläche zu einer gebäudeblockbezogenen Mietpreisdarstellung.



Abb. 5-24: Zoomstufe 17 einer prototypischen Mietpreiskarte (Entwurf und Darstellung: Schernthanner 2015).

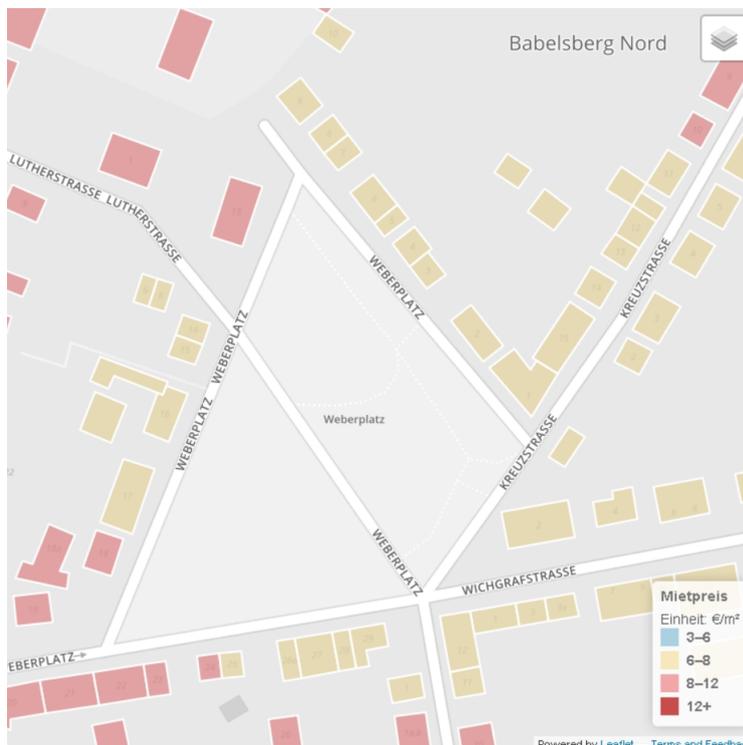


Abb. 5-25: Zoomstufe 18: Sprung der Darstellung von einer Trendoberfläche zur Gebäudeblockdarstellung bei einem Wechsel der Zoomstufe 17 auf 18 (Entwurf und Darstellung: Schernthanner 2015).

Nicht im Fokus dieser Forschungsarbeit liegt es, Lösungsmöglichkeiten für das „Icon-Cluttering“-Problem zu suchen. Potentiell mögliche Lösungen für diese Forschungsfrage bieten Map APIs mittlerweile an, jedoch gilt es, diese noch eingehender zu untersuchen. Als eine beispielhafte Visualisierung wurde eine Hexabin-Darstellung erstellt (Abbildung 5-26).

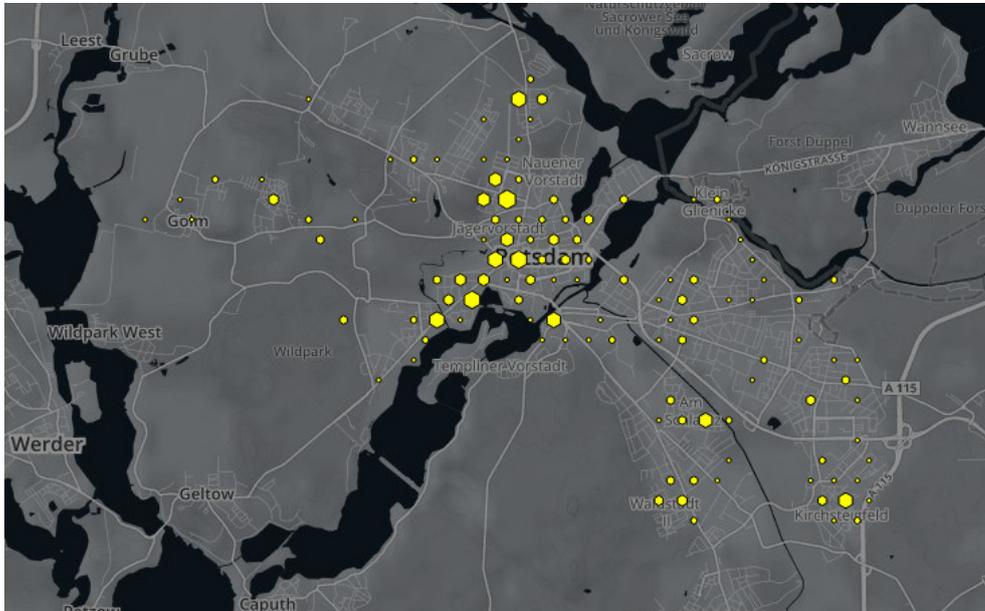


Abb. 5-26: Prototypische Angebotsdarstellung in Hexabins. Größere Waben bedeuten dabei größere Angebotscluster. (Entwurf und Darstellung: Scherthanner 2015).

# 6 FAZIT UND UMSETZUNGSOPTIONEN

## 6.1 Fazit

Die durchgeführte technologische Recherche hat gezeigt, dass die räumliche Sicht von Immobilienportalen eingeschränkt ist. Diese eingangs in der Arbeit konstatierte eingeschränkte räumliche Sicht existierender Immobilienportale kann durch die Teilimplementierung der unterschiedlich dargestellten Methoden teils aufgehoben werden. Das Fazit der Arbeit lässt sich in zwei Teilkomponenten – eine raumanalytische und eine geovisuelle Teilkomponente – gliedern, die sich an den zu Beginn der Arbeit formulierten Thesen orientieren. Zwischen beiden Teilkomponenten ergeben sich Wechselbeziehungen, die bei einem Transfer der Forschungsergebnisse in das operationelle Umfeld von Immobilienportalen berücksichtigt werden müssen. Nachfolgend wird ein Fazit für diese beiden Teilkomponenten, ausgehend von denen am Beginn der Arbeit (Kapitel 1.3) formulierten Thesen, gezogen.

### 6.1.1 Raumanalytische Teilkomponente des Fazits

Die erste These der Arbeit lautete, dass Verfahren der räumlichen Statistik und des Maschinellen Lernens zur Mietpreisschätzung hinsichtlich der räumlichen Optimierung von Immobilienportalen besser geeignet sind als die bisher eingesetzten Verfahren der hedonischen Regression. Insgesamt kann die erste These der Arbeit verifiziert werden, jedoch ist im Rahmen des Fazits eine differenzierte Betrachtung der Resultate der durchgeführten Teilimplementierungen notwendig.

Bei der Teilimplementierung räumlicher Schätzverfahren aus der Methodenfamilie des Kriging kann folgendes Fazit gezogen werden: Trotz intensiver Bemühungen, wie der Auswahl der Sekundärvariablen durch das Einbringen von Expertenwissen, durch die Anwendung unterschiedlicher Kovariablensets und das Testen der Sekundärvariablen mithilfe der Verfahren PCA und RF konnten die beiden Methoden KED und OCK keine valideren Schätzergebnisse als OK liefern. Für die Anwendung im Kontext eines Immobilienportals ist die Anwendung von OK die Alternative der Wahl zur hedonischen Regression. Es ist in allen Testläufen hinsichtlich der Güte und der Rechengeschwindigkeit den Verfahren KED und OCK überlegen. OK ist eine schnelle und effektive Alternative zur HedReg, um räumliche Schätzungen von Angebotsmieten für das operationelle Tagesgeschäft eines Immobilienportals durchzuführen. Gegebenenfalls sind eine weitere Untersuchung und Adaption der Nugget-, Sill- und Range-Parameter bei noch größeren als dem exemplarischen Untersuchungsraum notwendig. KED bietet sich für weitere Untersuchungen an. In der vorliegenden beispielhaften Teilimplementierung von KED wird die räumliche Variabilität von Ko-Variablensets nur sehr eingeschränkt berücksichtigt, da das KED-Modell auf der Annahme beruht, dass überall im Raum eine Standardwohnung vorhanden ist. Diese Standardwohnung hat im exemplarischen Untersuchungsraum Potsdam eine Größe von 77 m<sup>2</sup>. Alle geschätzten Angebotsmieten basieren

auf den Sekundärvariablen einer Standardwohnung im Untersuchungsraum. Hier ist ein Forschungsansatz, KED mit einem Verfahren aus dem Maschinellen Lernen wie dem vorgestellten RF zu koppeln, wie es beispielsweise *Li et al.* (2011) in einer Anwendung zur Interpolation des Schlammgehaltes testen.

Für die Mietpreisschätzung mittels des Verfahrens GWR und der Integration von Lageindikatoren lässt sich folgendes Fazit ziehen: Aufgrund des subjektiven Charakters von Lage- bzw. Lageindikatormodellierungen rät der Autor dazu, nur empirisch abgesicherte Lageindikatoren in ein GWR-Modell zu übernehmen. GWR hat durchaus ein Anwendungspotential im Kontext von Immobilienportalen, jedoch bedarf es hierzu weiterer Forschung, insbesondere die Auswahl von Ko-Variablen betreffend.

Als Fazit zur Anwendung von Methoden des Maschinellen Lernens ist zu sagen, dass sich vor allem die Methode RF als für die Modellierung von Angebotsdaten von Immobilienportalen als sehr gut geeignet gezeigt hat. Die Methode kann sehr gut mit lückenhaften Datensätzen und Ausreißern umgehen und ermöglicht außerdem eine Anwendung bei kategorialen Variablen. RF besitzt zwar ebenfalls den Mangel der Nicht-Beachtung räumlicher Autokorrelation, jedoch ist die Vorhersagegenauigkeit um einiges genauer als die der hedonischen Regression. Im Rahmen der Immobilienpreismodellierung wird die Anwendung von RF in zwei Fällen vorgeschlagen:

1. Für die Berechnung der Bedeutung von Variablen als Entscheidungshilfe zur Auswahl von Ko-Variablen für räumliche Modelle.
2. Für die Nutzung als Vorhersagemodell zur Modellierung einzelner Stützpunkte bei einer Interpolation von Immobilienpreisen.

Von der Anwendung von ANN ist abzuraten, die Teilimplementierungen der Verfahren waren nur in geringem Umfang valide. Insgesamt gesehen birgt vor allem eine Kombination von RF mit Methoden des Kriging zur Ermittlung zusätzlicher Stützpunkte ein großes Potential.

### 6.1.2 Geovisuelle Teilkomponente des Fazits

Die zweite These der Arbeit lautete, dass die von Immobilienportalen publizierten webbasierten Mietpreiskarten nicht die tatsächlichen räumlichen Verhältnisse auf Immobilienmärkten wiedergeben. Alternative webbasierte Darstellungsformen, wie z. B. Gridmaps, sind dem Status quo der Mietpreiskarten von Immobilienportalen überlegen und visualisieren die tatsächlichen räumlichen Verhältnisse von Immobilienpreisen zweckmäßiger.

Die Wechselwirkung der raumanalytischen mit der geovisuellen Teilkomponente wird bei den teilimplementierten Methoden des Gridmapping besonders deutlich. Die hier angewandten Methoden sind sowohl für die Preisschätzung als auch für die Darstellung der Mietpreisschätzungen relevant. Methoden des Gridmapping dienen in den durchgeführten Teilimplementierungen dazu, die optimale Maschenweite für das Bilden eines räumlichen Kontinuums der als

Diskreta vorliegenden räumlichen Schätzergebnisse zu finden. Diese optimalen Maschenweiten sind dann wiederum bei der Geovisualisierung der Schätzergebnisse in Form von Mietpreiskarten relevant. Sie können einerseits als räumliche Kontinua in Trendoberflächen abgebildet werden. Sie stellen andererseits die Grundlage dar, die gebildeten Raster, unter Einsatz von Gebäudegeometrien, zusätzlicher Daten, wie Flächen der Wohnnutzung oder zusätzlicher Landnutzungsbedeckungsklassen, wieder zu diskretisieren. Es kann also durch die Anwendung von Verfahren des Gridmapping ein Dreischritt von räumlichen Diskreta (geschätzten Punkten) über räumliche Kontinua (Mietpreistrendoberflächen) zu wiederum räumlichen Diskreta (Abbildung der Schätzergebnisse auf z. B. Gebäudegeometrien) gegangen werden. Die Anwendung des Gridmapping hilft entscheidend, die räumliche Sicht von Immobilienportalen zu optimieren. Anstatt Analyseergebnisse auf fachfremd verwendete Geometrien zu beziehen, erfolgen eine Analyse und Darstellung der realen räumlichen Verteilung der Mietpreise im Raum.

Zur Darstellung der interpolierten Gridmaps mittels Map APIs ist zu sagen, dass diese Art der Visualisierung in vielen Fällen noch am Anfang ihrer Entwicklung steht. Ansätze der cloudbasierten Kartendarstellung (z. B. mittels Leaflet.js oder CartoDB etc.) weisen ein großes Potential zur Darstellung von Mietpreisen auf. Hierbei muss jedoch in vielerlei Hinsicht das Potential der Visualisierung ausgebaut werden, um z. B. Probleme bei der Darstellung von Klassengrenzen, wie aufgezeigt, zu lösen.

Bezüglich der optimalen Darstellung von Mietpreisschätzungen ergibt sich als Fazit, dass für einen mittleren Maßstabsbereich zwischen 1 : 35.000 – 1 : 70.000 die Darstellung in Form einer interpolierten Trendoberfläche einen sehr schnellen Überblick über die vorherrschende Mietpreislandschaft gibt. Für größere Maßstäbe und Zoomstufen sollte eine Abbildung von Preisschätzergebnissen auf die Zentroide von Gebäudeblöcken erfolgen. Zu den Funktionen, die aus Forschungssicht für eine Mietpreiskarte essentiell sind, zählen: eine Zoom-in- und Zoom-out-Funktion zum Wechsel der Maßstabsebenen, eine Mouse-hoover-Funktion zur Anzeige von Mietpreisen einzelner Gebäudeblöcke und ein Suchfeld zur Adresssuche. Als Farbverlauf zur Visualisierung von Preisklassen wird Gelb – für niedrigere Mietpreise – nach Rot – für höhere Mietpreise – als farbsemantisch korrekt angesehen. Als Klassenanzahl sollten vier bis sieben Mietpreisklassen verwendet werden. Die Darstellung von Klassengrenzen und Klassenbreiten stellt eines der größten Probleme bei der Verwendung von Leaflet.js und CartoDB dar. Hinsichtlich zusätzlicher Kartenschichten sind als Fazit Bildungseinrichtungen und ÖPNV-Standorte zu integrieren. Bei der Wahl der technologischen Plattform bietet Leaflet.js den Vorteil, dass es durch zahlreiche Plugins erweiterbar und hierdurch eine einfache Optimierung responsiver Darstellungen möglich ist. Insbesondere Letztere wurden in den Experteninterviews mehrfach gefordert.

Insgesamt kann festgestellt werden, dass die entwickelten Mietpreiskartenprototypen räumliche Zusammenhänge in einer bisher nicht verfügbaren räumlichen Schärfe und Genauigkeit zeigen. Die zweite These der Arbeit konnte so-

mit auf Basis der angestellten konzeptionellen Vorüberlegungen und der Teilimplementierungen ebenfalls verifiziert werden.

## 6.2 Umsetzungsoptionen

Wesentliches Ergebnis der durchgeführten Teilimplementierungen der Mietpreisschätzung sind zwei Szenarien der Umsetzung der Methoden im Rahmen von Immobilienportalen sowie ein Szenario der Umsetzung einer Mietpreiskarte. Diese drei Szenarien werden nachfolgend schematisch veranschaulicht. Die schematisierten Darstellungen dienen dazu, Immobilienportalbetreibern die relevanten Prozesskomponenten in abstrahierter Form darzustellen, die für eine räumliche Optimierung der Mietpreisschätzung und -visualisierung aus Forschungssicht notwendig sind. Unterschieden wird bei der Darstellung jeweils zwischen obligatorischen und optionalen Prozesskomponenten.

### 6.2.1 Szenario Mietpreisschätzung I

Das erste Szenario weist einen geringeren Komplexitätsgrad auf und zielt auf eine schnelle und genaue räumliche Bestimmung von Mietpreistrends ab. Das erste Szenario empfiehlt den Einsatz der OK mit der optionalen Kopplung von Verfahren zur Ermittlung fehlender Stützpunkte wie RF oder mittels GWR für Teilräume mit geringerer Dichte an Angebotsdaten. Dieses Szenario ist optimal für den Transfer in das operationelle Umfeld eines Immobilienportals geeignet. Abbildung 6-1 stellt das Szenario schematisch dar.

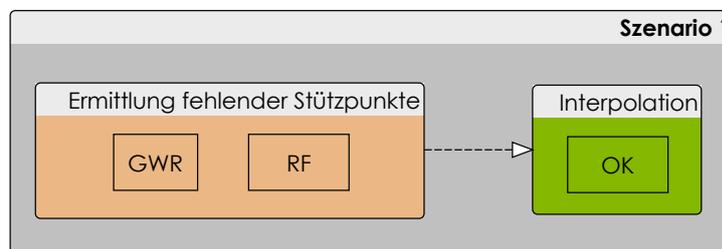


Abb. 6-1: Szenario 1: Bestehend aus dem Verfahren OK mit optionaler Kopplung mit GWR und RF (orange optionale Prozesskomponenten, grün: obligatorische Prozesskomponenten). (Entwurf und Darstellung: Scherthanner 2015).

### 6.2.2 Szenario Mietpreisschätzung II

Das zweite Szenario weist einen höheren Komplexitätsgrad auf und erfordert für eine Übertragung in die operationellen Abläufe eines Immobilienportals eine weitere Feinabstimmung, insbesondere die Auswahl an Sekundärvariablen betreffend. Dieses Szenario sieht die Anwendung des Verfahrens KED mit Koppelung von Verfahren zur Ermittlung bestmöglicher Ko-Variablensets (PCA und RF) und mit dem Verfahren RF zur „Konstruktion“ fehlender Stützpunkte vor. Um eine Einbettung dieses Szenarios in operationelle Abläufe eines Immobilienportals zu erreichen, ist weiterer Forschungsaufwand erforderlich. Abbildung 6-2 stellt das Szenario schematisch dar.

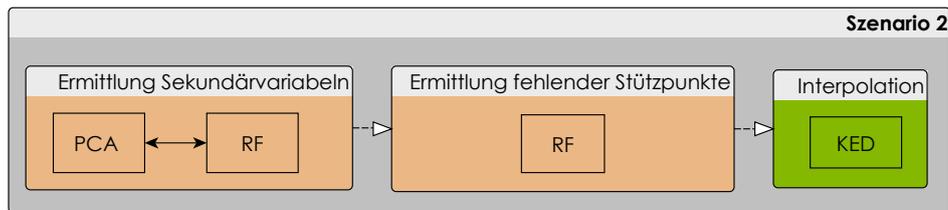


Abb. 6-2: Szenario 2: Bestehend aus dem Verfahren KED mit optionaler Koppelung weiterer Verfahren (orange optionale Prozesskomponenten, grün: obligatorische Prozesskomponenten). (Entwurf und Darstellung: Scherthanner 2015).

Ein Ergebnis der vorliegenden Arbeit ist, dass das im Rahmen dieser Arbeit vorgeschlagene und umgesetzte OK-Verfahren beim Immobilienportal IS24 in der Kombination mit einer herkömmlichen HedReg stabil läuft. Das programmtechnisch erarbeitete Verfahren befindet sich bei IS24 in einer Testphase bzw. Betaphase (Stand Dezember 2014). Die Grundlage für eine erste operationelle Anwendung einer räumlichen Preisschätzung im Rahmen eines Immobilienportals wurde somit gelegt.

### 6.2.3 Szenario Mietpreiskarte

Das dritte Szenario zeigt Prozesskomponenten, die aus Forschungssicht zur Erstellung einer webbasierten Mietpreiskarte notwendig sind. Als Bindeglied zwischen den beiden Szenarien der Mietpreisschätzung wird die Anwendung von Methoden des Gridmapping als essentiell erachtet. Abbildung 6-3 stellt das Szenario schematisch dar.

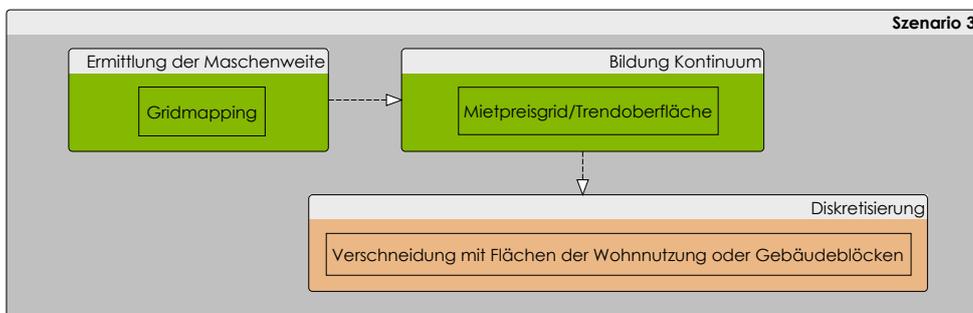


Abb. 6-3: Szenario 3: Erstellung einer Mietpreiskarte bestehend aus den obligatorischen Prozesskomponenten der Verfahren des Gridmapping und der Erstellung von Mietpreisgrids und der optionalen Prozesskomponente der Verschneidung von Mietpreisgrids mit zusätzlichen Datenschichten. (Entwurf und Darstellung: Scherthanner 2015).

# 7 ZUSAMMENFASSUNG UND AUSBLICK

## 7.1 Zusammenfassung

Die vorliegende Arbeit verfolgte das Ziel, aus geoinformatischer Sicht eine konzeptionelle Grundlage zur räumlichen Optimierung von Immobilienportalen zu schaffen. Die Arbeit geht dabei von zwei Hypothesen aus:

1. *Verfahren der räumlichen Statistik und des Maschinellen Lernens zur Mietpreisschätzung sind besser zur räumlichen Optimierung von Immobilienportalen geeignet als die bisher eingesetzten Verfahren der hedonischen Regression und*
2. *die von Immobilienportalen publizierten webbasierten Mietpreiskarten geben nicht die tatsächlichen räumlichen Verhältnisse auf Immobilienmärkten wieder. Alternative webbasierte Darstellungsformen, wie z. B. Gridmaps, sind dem Status quo der Immobilienpreiskarten von Immobilienportalen überlegen und visualisieren die tatsächlichen räumlichen Verhältnisse von Immobilienpreisen zweckmäßiger.*

Beide Thesen können bewiesen werden.

Es erfolgt zunächst eine umfangreiche Erhebung des Forschungsbedarfs mittels Literaturstudien und technologischer Recherche. Zur Beantwortung der Forschungsfragen wird als quantitative Datenbasis ein 74.098 Mietangebote umfassender Datensatz (von Januar 2007 bis September 2013) eines Immobilienportals akquiriert. Dieser reicht jedoch nicht in vollem Umfang zur Beantwortung der Fragestellungen aus. Deshalb führt der Autor Experteninterviews zur Erhebung einer qualitativen Datenbasis. Deren Analyse ergibt in Kombination mit der Literaturstudie und der technologischen Recherche ein umfassendes, bisher so nicht verfügbares Bild. Es stellt den Status quo der räumlichen Sicht sowie der raumanalytischen und geovisuellen Defizite von Immobilienportalen dar.

Zur Optimierung der raumanalytischen und geovisuellen Defizite werden forschungsbasierte Lösungsansätze herausgearbeitet und teilimplementiert. Methoden des Maschinellen Lernens und räumliche Schätzverfahren werden als Alternativen zu den von Immobilienportalen bisher genutzten „nicht räumlichen“ Analyseverfahren zur Preismodellierung untersucht. Auf Grundlage eines hierfür konzipierten Validierungsrahmens werden diese Methoden für die Nutzung im Kontext von Immobilienportalen adaptiert. Die prototypische Teilimplementierung zeigt die programmiertechnische Umsetzung des Konzeptes auf. Eine umfassende Analyse geeigneter Sekundärvariablen zur Mietpreisschätzung liefert als methodisches Resultat, dass Interpolatoren, die Sekundärvariablen benötigen (*Kriging with external drift, Ordinary Cokriging*), kaum zu valideren Mietpreisschätzergebnissen gelangen als die Methode des *Ordinary Kriging*, die keine Sekundärvariablen benötigt. Die Methoden Random Forest aus dem Maschinellen Lernen und die Geographisch Gewichtete Regression hingegen bergen großes Potential zur Nutzung der räumlichen Mietpreisschät-

zung im Kontext von Immobilienportalen. Die Forschungsergebnisse der räumlichen Preismodellierung werden in die räumliche Visualisierung von Mietpreisen transferiert.

Für die webbasierte Mietpreisdarstellung wird ein Set alternativer Darstellungsmethoden entwickelt, um Mietpreiskarten-Prototypen abzuleiten. Ein methodisches Ergebnis der Entwicklung der Mietpreiskarten-Prototypen ist die Entwicklung eines geeigneten Ansatzes der Loslösung des Preisbezugs von fachfremd verwendeten Bezugsgeometrien. Hierfür wird vom Autor der Begriff der zonenlosen Preiskarte geprägt. Diese werden mit Methoden des Gridmapping erstellt. Es werden optimale Rasterauflösungen zur Darstellung interpolierter Rastergrößen ermittelt. Zonenlose Preiskarten mit Methoden des Gridmapping, gepaart mit einer optionalen gebäudescharfen Darstellung in größeren Maßstäben, sind als Resultate der Forschung die bestmögliche sich an realen Verhältnissen orientierende räumliche Mietpreisdarstellung. Die entstandenen Prototypen sind eine Annäherung der wahren Verteilung des Mietpreises im Raum und um einiges schärfer als die auf der hedonischen Regression basierenden Darstellungen. Somit kann die wahre „Topographie“ der Mietpreislandschaft abgebildet werden. Ein Einsatz der Karten für Nutzergruppen wie Makler, Investoren oder Kommunen zur Analyse städtischer Mietmärkte ist denkbar. Alle entstandenen Prototypen sind unter der Nutzung von Map APIs umgesetzt. Ein Ergebnis dessen ist, dass Map APIs noch an diversen „Kinderkrankheiten“ leiden und derart umgesetzte Mietpreiskarten einen weiten Weg vor sich haben, bis sie das Niveau thematischer Karten von *Imnhof* oder *Arnberger* erreichen.

Die konzeptionellen Überlegungen und Teilimplementierungen münden in drei Prozessketten, die Umsetzungsoptionen für eine räumliche Optimierung von Immobilienportalen darstellen. Dabei werden zwei Szenarien für eine räumlich optimierte Mietpreisschätzung und ein Szenario für eine räumlich optimierte Mietpreisdarstellung herausgearbeitet.

## 7.2 Beitrag zum Forschungsstand und Ausblick

Einen wesentlichen Beitrag der Arbeit zum Stand der Forschung stellt die erstmalige Adaption räumlicher Analytik und Visualisierung an die realen Verhältnisse der Immobilienwirtschaft bzw. Immobilienportale dar. Ein weiterer Beitrag ist die sehr intensive Auseinandersetzung mit raumanalytischen und geovisuellen Problemen von Immobilienportalen. Neben der Erhebung der Probleme mit dem dringlichsten Forschungsbedarf erfolgt eine Untersuchung bisher nur im Wissenschaftsumfeld eingesetzter Methoden auf ihre Nutzbarkeit für Portale. In diesem Zusammenhang werden wichtige Parameter des Kriging-Schätzers intensiv auf ihre Sensitivität zur Interpolation untersucht. Dies erweitert das Forschungsfeld der Immobilienmarktforschung um geoinformatische Mittel. Aus der Perspektive der Geoinformatik erfolgt darüber hinaus eine erstmalige wissenschaftliche Auslotung von Verfahren des Maschinellen Lernens und der Geographisch Gewichteten Regression zur Anwendung im Kontext von Immobilienportalen. Es werden Instrumentarien für Immobilienportale vorgeschlagen, die validere räumliche Aussagen als bisher zu Mietpreisen ermöglichen. Es erfolgt nicht nur ein Vorschlag der Instrumentarien, es werden auch Szenarien

in Form von Prozessketten betrachtet, die räumliche Auswertungsprozesse in Immobilienportale integrieren. Die ausgewählten und teilimplementierten Instrumentarien haben ein Potential, das über die Anwendung in Immobilienportalen hinausgeht. Damit leistet die Arbeit nicht nur einen Beitrag zur räumlichen Optimierung von Immobilienportalen, sondern darüber hinaus auch einen Beitrag zum Wohnraummonitoring insgesamt.

### 7.2.1 Ausblick

Ein Transfer von Methoden der räumlichen und geovisuellen Analyse in das Umfeld von Immobilienportalen hat eine Bedeutung, die weit über den Rahmen der Arbeit hinausgeht. Die von den Portalen publizierten Indizes basieren auf nicht räumlichen Analysemethoden. Ausgehend von diesen werden unscharfe und verzerrte Aussagen über Wohnungsmärkte getroffen. Diese werden meist unreflektiert von Medien übernommen, um scheinbar valide Aussagen über einen der weltweit wichtigsten Wirtschaftsbereiche zu treffen. Ein Transfer von Forschungsergebnissen der räumlichen Immobilienmarktforschung gepaart mit Methoden der Geovisualisierung ermöglicht es Portalbetreibern bzw. Akteuren, die mit Portalen eng zusammenarbeiten (z. B. Gutachterausschüssen), räumlich viel differenziertere Aussagen zu Immobilienmärkten bzw. Mietmärkten zu treffen. Als Ausblick für zukünftige Forschung gilt es, den Transfer der raumanalytischen und geovisuellen Instrumentarien noch vor den Instrumentarien selbst zu verbessern.

# LITERATURVERZEICHNIS

- Abler, R., Marcus, M. G. & Olson, J. M. (Eds.). (1992). *Geography's inner worlds: pervasive themes in contemporary American geography* (Vol. 2). Springer Science & Business.
- Ache, P. (2014) Experteninterview von Harald Schernthanner, Universität Potsdam, Fachgruppe Geoinformatik. 27.08.2015, 38:46.
- Alexa.com (2014) [online] Information. Insight. Advantage.  
<http://www.alexacom/about> [letzter Aufruf: 02.12.2014].
- Alexa.com (2014) [online] How popular is zillow.com?  
<http://www.alexacom/siteinfo/zillow.com> [letzter Aufruf: 20.11.2014].
- Alexa.com (2014) [online] How popular is immobilienscout.de ?  
<http://www.alexacom/siteinfo/immobilienscout.de> [letzter Aufruf: 20.11.2014].
- Albrecht, J. (2007). *Key concepts and techniques in GIS*. Sage.
- Antipov, E. A. & Pokryshevskaya, E. B. (2012). Mass appraisal of residential apartments: An application of Random forest for valuation and a CART-based approach for model diagnostics. *Expert Systems with Applications*, 39(2), 1772 – 1778.
- Arnberger, E. (1997) *Thematische Kartographie. Das geographische Seminar*. Westerman. 4. Auflage.
- Axel Springer Verlag (2014)[Online] Wer gehört zu wem? <http://www.wer-zu-wem.de/firma/ASV.html> [letzter Aufruf: 03.12.2014].
- Baranzini, A., Ramirez, J., Schaerer, C. & Thalmann, P. (2008). *Hedonic methods in housing markets: pricing environmental amenities and segregation*. Springer Science & Business Media.
- Batty, M., Hudson-Smith, A., Milton, R. & Crooks, A. (2010). Map mashups, Web 2.0 and the GIS revolution. *Annals of GIS*, 16(1), 1 – 13.
- Bivand, R. S., Pebesma, E. & Gómez-Rubio, V. (2013). *Applied Spatial Data Analysis with R* (Vol. 10). Springer.
- Bock, S. (1995). Ein Ansatz zur polygonbasierten Klassifikation von Luft- und Satellitenbildern mittels künstlicher neuronaler Netze. Selbstverlag Geographisches Institut der Universität Kiel.
- Borchert, A. (2006). Geographische Informationssysteme für die Immobilienwirtschaft. *Standort*, 30(3), 127 – 131.

- Bourassa, S. C., Cantoni, E. & Hoesli, M. (2010). Predicting house prices with spatial dependence: A comparison of alternative methods. *Journal of Real Estate Research*, 32(2), 139 – 159.
- Burrough, P. A. & McDonnell, R. A. (2011). Principles of geographical information Systems (Vol. 19988). Oxford University Press.
- Breiman, L. (2001). Random forests. *Machine learning*, 45(1), 5 – 32.
- Brunsdon, C., Fotheringham, A. S. & Charlton, M. E. (1996). Geographically weighted regression: a method for exploring spatial nonstationarity. *Geographical analysis*, 28(4), 281 – 298.
- Bundesinstitut für Bau-, Stadt- und Raumforschung (2014). [Online] Überblick über Mietdaten.  
<http://www.wohnungsmarktbeobachtung.de/wissensdatenbank/indikatoren/bilanz/preise/miete-und-mietbelastung/ueberblick> [letzter Aufruf: 18.12.2014].
- Burigat, S. & Chittaro, L. (2008). Decluttering of icons based on aggregation in mobile maps. In *Map-based Mobile Services* (pp. 13 – 32). Springer Berlin Heidelberg.
- CARTODB (2014). [Online] Create amazing maps with your data.  
<http://cartodb.com/> [letzter Aufruf: 10.11.2014].
- Craigslist (2005). [online] <https://forums.craigslist.org/?ID=26638141> [letzter Aufruf: 18.11.2014].
- Chiles, J. P. & Delfiner, P. (2009). *Geostatistics: modeling spatial uncertainty* (Vol. 497). John Wiley & Sons.
- Codeplex (2013) [online] Create a Network Dataset from OSM Data.  
<http://esriosmeditor.codeplex.com/wikipage?title=Create%20a%20Network%20Dataset%20from%20OSM%20Data> [letzter Aufruf: 30.01.2015].
- Colorbrewer 2.0 (2015) [Online] <http://colorbrewer2.org/> [letzter Zugriff: 06.02.2015].
- Dangschat, J. S. (1988). Gentrification: Der Wandel innenstadtnaher Wohnviertel. In *Soziologische Stadtforschung* (pp. 272 – 292). VS Verlag für Sozialwissenschaften.
- Dixon, B. & Candade, N. (2008). Multispectral landuse classification using neural networks and support vector machines: one or the other, or both?. *International Journal of Remote Sensing*, 29(4), 1185 – 1206.
- Dombrowski, K. (2013). Eine Trendwende in der Marketingforschung – Aktuelle Chancen und Risiken von Web Surveys. In *Handel in Theorie und Praxis* (pp. 439 – 465). Springer Fachmedien Wiesbaden.

- Dubin, R. A. (1998). Predicting house prices using multiple listings data. *The Journal of Real Estate Finance and Economics*, 17(1), 35 – 59.
- Dubin, R. A. (1988). Estimation of regression coefficients in the presence of spatially autocorrelated error terms. *The Review of Economics and Statistics*, 466 – 474.
- Duden [Online] Prozess, der. <http://www.duden.de/rechtschreibung/Prozess> [letzter Aufruf: 05.12.2013].
- Enderle, M. (2009). Verbesserte Markttransparenz durch Immobilienportale im Internet. In *Kapitalanlage mit Immobilien* (pp. 361 – 376). Gabler.
- Enzyklopädie der Wirtschaftsinformatik (2013) [Online] Stochastische Optimierung. <http://www.enzyklopaedie-der-wirtschaftsinformatik.de/wi-enzyklopaedie/lexikon/technologien-methoden/Operations-Research/Mathematische-Optimierung/Stochastische-Optimierung> [letzter Aufruf: 04.12.2014].
- Fielding, R. (2000) [Online] Architectural Styles and the Design of Network-based Software Architectures. <http://www.ics.uci.edu/~fielding/pubs/dissertation/top.htm>. [letzter Aufruf: 04.12.2014].
- Fittkau und Maaß (2014) [Online] Reports und Studien. <http://www.fittkaumaass.de/reports-und-studien> [letzter Aufruf: 18.11.2014].
- Fotheringham, A. S., Brunson, C. & Charlton, M. (2003). Geographically weighted regression: the analysis of spatially varying relationships. John Wiley & Sons.
- Fotheringham, A. S., Charlton, M. & Brunson, C. (1996). The geography of parameter space: an investigation of spatial non-stationarity. *International Journal of Geographical Information Systems*, 10(5), 605 – 627.
- Fürnkranz, J., Gamberger, D. & Lavrač, N. (2012). Machine learning and data mining. In *Foundations of Rule Learning* (pp. 1 – 17). Springer Berlin Heidelberg.
- Gandin, L. S. (1965). *Objective analysis of meteorological fields* (Vol. 242). R. Hardin (Ed.). Jerusalem: Israel program for scientific translations.
- García, A. G. & Filgueira, M. N. (2012). The regression model in the House Price Index. *BEIO, Boletín de Estadística e Investigación Operativa*, 28(3), 247 – 260.
- Geniaux, G. & Napoléone, C. (2008). Semi-parametric tools for spatial hedonic models: an introduction to mixed geographically weighted regression

- and geoaddivitive models. In *Hedonic Methods in Housing Markets* (pp. 101 – 127). Springer New York.
- Geoinformatik-Service der Universität Rostock, Lexikon. [Online] Prozess.  
<http://www.geoinformatik.uni-rostock.de/einzel.aspID=1101985565> [letzter Aufruf: 01.12.2014].
- Geoserver (2015) Geoserver <http://geoserver.org/> [letzter Aufruf: 06.02.2015].
- Github (2015) CartoDB <https://github.com/CartoDB/cartodb> [letzter Aufruf: 06.02.2015].
- Github (2014) [online] komoot/phoon <https://github.com/komoot/phoon> [letzter Aufruf: 06.02.2015].
- Google (2015) [online] Support for OGC Standards  
<https://support.google.com/mapsengine/answer/4562234?hl=en> [letzter Aufruf: 07.02.2015].
- Goovaerts, P. (1997). *Geostatistics for natural resources evaluation*. Oxford university press.
- Gordon, R. J. (1990)). *The measurement of durable goods prices*. University of Chicago Press.
- Gu, J., Zhu, M. & Jiang, L. (2011). Housing price forecasting based on genetic algorithm and support vector machine. *Expert Systems with Applications*, 38(4), 3383 – 3386.
- Hake, G., Grünreich, D. & Meng, L. (1994). *Kartographie: Visualisierung raumzeitlicher Informationen*. Walter de Gruyter.
- Harbich, M. & Peter, K. D. (2007) [Online]. Künstliche Neuronale Netze/Artificial Neural Networks, Universität Trier, Institut für Geographie, Abteilung Fernerkundung. [http://fern39.uni-trier.de/feut/studium/hyperspektral\\_ss07](http://fern39.uni-trier.de/feut/studium/hyperspektral_ss07) [letzter Aufruf: 10.12.2014].
- Hengl, T. (2006). Finding the right pixel size. *Computers & Geosciences*, 32(9), 1283 – 1298.
- Hepsen, A. und Vatansever, M. (2012) Using Hierarchical Clustering Algorithms for Turkish Residential Market. *International Journal of Economics and Finance* Vol. 4, No. 1; January 2012.
- Homegate (2014) [Online] Online Schätzung.  
<http://www.homegate.ch/kaufen/wohneigentum-schaetzen/online-schaetzung/methode> [letzter Aufruf: 09.12.2014].
- Housingmaps (2014) [Online] Housing Maps.  
<http://www.housingmaps.com/> [letzter Aufruf: 18.11.2014].

- Huang, H. & Gartner, G. (2012). A Technical Survey on Decluttering of Icons in Online Map-Based Mashups. In *Online Maps with APIs and WebServices* (pp. 157 – 175). Springer Berlin Heidelberg.
- Infas (2015) [Online] Unse-Geodatenkatalog.  
<http://infas360.de/geodaten/unser-geodatenkatalog/>. [letzter Aufruf: 18.11.2014].
- Immobilienportale (2012) [online] Immobilienscout 24.  
<http://www.immobilienportale.com/uebersicht-immobilienportale/20084-immobilienscout24/> [letzter Aufruf: 18.11.2014].
- Immobilienscout 24 (2014) [online] Welcome to the ImmobilienScout24 Developer Center. <http://api.immobilienscout24.de/> [letzter Aufruf: 10.12.2014].
- Immonet (2014) [online] <http://www.immonet.de/> [letzter Aufruf: 18.11.2014].
- Immowelt (2014) [online] <http://www.immowelt.de/> [letzter Aufruf: 18.11.2014].
- Jenks, G. F. & Caspall, F. C. (1971). Error on choroplethic maps: definition, measurement, reduction. *Annals of the Association of American Geographers*, 61 (2), 217 – 244.
- Jermann, J. (2002) [online] Vergleich verschiedener Ansätze der GIS-gestützten Potenzialberechnung. Conference Paper STRC 2002.  
<http://www.strc.ch/conferences/2002/jermann.pdf> [letzter Aufruf: 14.12.2014].
- Jong, P. D., Sprenger, C. & Veen, F. (1984). On extreme values of Moran's I and Geary's c. *Geographical Analysis*, 16(1), 17 – 24.
- Krige, D., G. (1951). A statistical approach to some basic mine valuation problems on the Witwatersrand. 52 (6): S. 119 – 139.
- Kessler, W. (2007). Multivariate Datenanalyse: für die Pharma, Bio-und Prozessanalytik. John Wiley & Sons.
- Kerry, R., Goovaerts, P., Rawlins, B. G. & Marchant, B. P. (2012). Disaggregation of legacy soil data using area to point kriging for mapping soil organic carbon at the regional scale. *Geoderma*, 170, 347 – 358.
- Kuntz, M. & Helbich, M. (2014). Geostatistical mapping of real estate prices: an empirical comparison of kriging and cokriging. *International Journal of Geographical Information Science*, (ahead-of-print), 1 – 18.
- Laaha, G., Skøien, J. O. & Blöschl, G. (2014). Spatial prediction on river networks: comparison of top-kriging with regional regression. *Hydrological Processes*, 28(2), 315 – 324.
- Landeshauptstadt Potsdam (2011) [Online] BESCHLUSS der 35. öffentlichen Sitzung der Stadtverordnetenversammlung der

- Landeshauptstadt Potsdam am 28.09.2011.  
[http://www.potsdam22.de/files/beschluss\\_erarbeitung\\_eines\\_konzeptes\\_erschwingliche\\_mieten\\_fuer\\_alle.pdf](http://www.potsdam22.de/files/beschluss_erarbeitung_eines_konzeptes_erschwingliche_mieten_fuer_alle.pdf) [letzter Aufruf: 10.01.2015].
- Landeshauptstadt Potsdam (2014) [Online] Mietspiegel 2014.  
[http://vv.potsdam.de/vv/produkte/17301010000007870.php.media/13062/Mietspiegel\\_2014\\_Stand\\_07.08.2014.pdf](http://vv.potsdam.de/vv/produkte/17301010000007870.php.media/13062/Mietspiegel_2014_Stand_07.08.2014.pdf) [letzter Aufruf: 09.02.2015].
- Lange, N. (2013). *Geoinformatik: in Theorie und Praxis*. Springer-Verlag.
- Legendre, P. (1993), 'Spatial autocorrelation: trouble or new paradigm?', *Ecology* 74(6), S. 1659 – 1673.
- Leaflet (2015) [Online] Leaflet <http://leafletjs.com/> [letzter Aufruf: 07.02.2015].
- Li, J. (2008). A review of spatial interpolation methods for environmental scientists (Vol. 137). Canberra: Geoscience Australia.
- Li, J., Heap, A. D., Potter, A. & Daniell, J. J. (2011). Application of machine learning methods to spatial interpolation of environmental variables. *Environmental Modelling & Software*, 26(12), 1647 – 1659.
- Lo, C. P. & Yeung, A. K. (2002). *Concepts and Techniques of Geographic Information Systems*, 492 pp. New Jersey.
- Lüthy, S. (2009) [Online] Merkmalswichtigkeit im Random Forest. Masterarbeit, Eidgenössische Technis Hochschule Zürich, Department für Mathematik. <http://bit.ly/1EUBCTy> [letzter Aufruf: 30.01.2015].
- Mapbox (2015) [Online] About Mapbox <https://www.mapbox.com/about/> (letzter Aufruf: 07.02.2015).
- Marcotte, E. (2010) [online] Responsive Webdesign <http://alistapart.com/article/responsive-web-design> [letzter Aufruf: 03.12.2014].
- Matheron, G. (1963), 'Principles of geostatistics', *Economic Geology* 58(8), S. 1246 – 1266
- McCluskey, W. J. & Borst, R. A. (2011). Detecting and validating residential housing submarkets: A geostatistical approach for use in mass appraisal. *International Journal of Housing Markets and Analysis*, 4(3), 290 – 318.
- McCord, M., Davis, P. T., Haran, M., McIlhatton, D. & McCord, J. (2014). Understanding rental prices in the UK: a comparative application of spatial modelling approaches. *International Journal of Housing Markets and Analysis*, 7(1), 98 – 128.
- Meuser, M. & Nagel, U. (1991). ExpertInneninterviews – vielfach erprobt, wenig bedacht. Ein Beitrag zur qualitativen Methodendiskussion. In Detlev Garz & Klaus Kraimer(Hrsg.), *Qualitativ-empirische, Sozialforschung. Konzepte,*

- Methoden, Analysen. (S. 441 – 471). Opladen: Westdeutscher Verlag, S. 443.
- Montero, J. M. & Larraz, B. (2011). Interpolation methods for geographical data: Housing and commercial establishment markets. *Journal of Real Estate Research*, 33(2), 233 – 244.
- Nalder, I. A. & Wein, R. W. (1998). Spatial interpolation of climatic normals: test of a new method in the Canadian boreal forest. *Agricultural and forest meteorology*, 92(4), 211 – 225.
- Nexiga (2014) [Online] Datenkatalog Marktinformationen. <http://bit.ly/1ySxf4i> [letzter Aufruf: 05.12.2014].
- Niemeier, W. (2008). Ausgleichsrechnung. de Gruyter.
- Olmo, J. C. (1995). Spatial estimation of housing prices and locational rents. *Urban Studies*, 32(8), 1331 – 1344.
- Open Geospatial Consortium (2015) [Online] About OGC <http://www.opengeospatial.org/ogc> [letzter Zugriff: 05.12.2014],
- OpenLayers (2015) [Online] OpenLayers <http://openlayers.org/> [letzter Aufruf: 07.02.2015].
- Openstreetmap Wiki (2014) Zoom Levels [Online] [http://wiki.openstreetmap.org/wiki/Zoom\\_levels](http://wiki.openstreetmap.org/wiki/Zoom_levels) [letzter Aufruf: 03.12.2014].
- Openstreetmap [online] Nominatim <http://wiki.openstreetmap.org/wiki/Nominatim> [letzter Aufruf: 06.02.2015].
- Osmbuildings (2015) [online] <http://osmbuildings.org/?lat=52.52052&lon=13.41160&zoom=16> [letzter Aufruf: 06.02.2015].
- Openstreetmap (2015) [online] Overpass API. [http://wiki.openstreetmap.org/wiki/Overpass\\_API](http://wiki.openstreetmap.org/wiki/Overpass_API) [letzter Aufruf: 07.02.2015].
- Papageorgiou, M. (2012). Optimierung: statische, dynamische, stochastische Verfahren. Springer-Verlag.
- Park, B. & Bae, J. K. (2015). Using machine learning algorithms for housing price prediction: The case of Fairfax County, Virginia housing data. *Expert Systems with Applications*, 42(6), 2928 – 2934.
- Pebesma, E. J. (2004). Multivariable geostatistics in S: the gstat package. *Computers & Geosciences*, 30(7), 683 – 691.
- Peterson, M. P. (2014). Mapping in the Cloud (p. 421). Guilford Publications.

- Peterson, M. P. (2012). Online maps with APIs and WebServices. Springer Science & Business Media.
- Potsdam22 (2013). [Online] Kommunale Wohnungspolitik und die soziale Dimension steigender Wohnkosten. [http://www.potsdam22.de/files/bericht\\_potsdam\\_22\\_web.pdf](http://www.potsdam22.de/files/bericht_potsdam_22_web.pdf) [letzter Aufruf: 09.01.2015].
- Prinz, T. (2007). Räumliche Nachhaltigkeitsindikatoren als Planungsgrundlage. Integrative Bewertung von Siedlungsflächen in der Stadt Salzburg. Dissertation, Universität Salzburg.
- Rebhan, C. (2011) [online] W3B-Studie: Online-Wohnimmobiliensuche auf Rekordstand. Immobilienzeitung.de Fachzeitung für Immobilien. <http://www.immobilien-zeitung.de/1000001829/w3b-studie-online-wohnmobiliensuche-auf-rekordstand> [letzter Aufruf: 18.11.2014].
- Rohmert, W. & Böhm, J. (Eds.). (2013). E-Business in der Immobilienwirtschaft: neue Chancen, Märkte und Marktteilnehmer. Springer-Verlag.
- Rosen, S. (1974). Hedonic prices and implicit markets: product differentiation in pure competition. *The journal of political economy*, 34 – 55.
- Rossiter, D. G. (2007). Technical Note: Co-kriging with the gstat package of the R environment for statistical computing.
- Spatial-analyst.net (2010) [online] Grid size calculator [http://spatial-analyst.net/wiki/index.php?title=Grid\\_size\\_calculator](http://spatial-analyst.net/wiki/index.php?title=Grid_size_calculator) [letzter Aufruf: 07.02.2015].
- Statistisches Bundesamt (2011). [online] Gebäude und Wohnungen sowie Wohnverhältnisse der Haushalte Bundesrepublik Deutschland am 9. Mai 2011. <http://bit.ly/14CL6Xe> [letzter Aufruf: 18.11.2014]
- Schendera, C. (2010). Clusteranalyse mit SPSS: mit Faktorenanalyse. Oldenbourg Verlag.
- Scheele, H. & Schernthanner H. (2014). Eine Räumliche Alternative der Modellierung und Darstellung von Immobilienpreisen für Immobilienportale. Angewandte Geoinformatik 2014. In: Strobl, J., Blaschke, T., Griesebner, G. (eds.) Angewandte Geoinformatik 2014. Beiträge zum 26. AGIT-Symposium Salzburg, July 2 – 4, Universität Salzburg, Heidelberg (2014).
- Schernthanner, H. & Tyrillová, L. (2010). FOS-GIS und Immobilienportale als Instrumente der Wohnungsmarktanalyse. Angewandte Geoinformatik 2010. In: Strobl, J., Blaschke, T., Griesebner, G. (eds.) Angewandte Geoinformatik 2010. Beiträge zum 22. AGIT-Symposium Salzburg, July 7 – 9, Universität Salzburg, Heidelberg (2010).
- Schulze, P. M. (2007). *Beschreibende Statistik*. Oldenbourg Verlag.

- Selim, H. (2009). Determinants of house prices in Turkey: Hedonic regression versus artificial neural network. *Expert Systems with Applications*, 36, 2843 – 2852
- Sibson, R. (1981). A brief description of natural neighbor interpolation (Chapter 2)". In V. Barnett. *Interpreting Multivariate Data*. Chichester: John Wiley. pp. 21 – 36.
- Späth, H. (1995). *One Dimensional Spline Interpolation Algorithms*. Wellesley: A K Peters.
- Statistics Austria (2013)[Online]Regional statistical grid units  
[http://www.statistik.at/web\\_en/classifications/regional\\_breakdown/grid/index.html](http://www.statistik.at/web_en/classifications/regional_breakdown/grid/index.html) [letzter Aufruf: 07.20.2015] Stroisch, J. (2010). *Immobilien bewerten leicht gemacht*. Haufe-Lexware.
- Trainor, T. (2014) [online] Common Geographic Boundaries: Small Area Geographies, Administrative, and Grid-based Geographies – One or Many? Global Forum on the Integration of Statistical and Geospatial Information. 4. – 5. August 2014 New York. <http://bit.ly/1C3JkHJ> [letzter Aufruf: 07.02.2015].
- Triplet, J. (2004). Handbook on Hedonic price indexes and quality adjustments in price indexes: Special application to information technology products (No. 9). STI OECD Working Paper. Trulia (2014) [Online] New York Crime. [http://www.trulia.com/real\\_estate/New\\_york-New\\_York/crime/](http://www.trulia.com/real_estate/New_york-New_York/crime/) [letzter Aufruf: 05.12.2014].
- United States Census (2010) [Online] Geographic Terms and Concepts - Core Based Statistical Areas and Related Statistical Areas.  
[www.census.gov/geo/reference/gtc/gtc\\_cbsa.html](http://www.census.gov/geo/reference/gtc/gtc_cbsa.html) [letzter Aufruf: 05.12.2014].
- W3C (2015) [Online] Geolocation API <http://dev.w3.org/geo/api/spec-source.html> [letzter Aufruf: 06.02.2015].
- Wackernagel, H. (1996). Multivariate geostatistics: an introduction with applications. In *International Journal of Rock Mechanics and Mining Sciences and Geomechanics Abstracts* (Vol. 33, No. 8, pp. 363A – 363A). Elsevier.
- Willmott, C. J. (1982). Some comments on the evaluation of model performance. *Bulletin of the American Meteorological Society*, 63(11), 1309 – 1313.
- Webster, R. & Oliver, M. A. (2007). *Geostatistics for environmental scientists*. John Wiley & Sons.

- Weka 3: Data Mining Software in Java (2014)[online]  
<http://www.cs.waikato.ac.nz/ml/weka/> [letzter Aufruf: 10.12.2014].
- Worzala, E., Lenk, M. & Silva, A. (1995). An exploration of neural networks and its application to real estate valuation. *Journal of Real Estate Research*, 10(2), 185 – 201.
- Wiktionary (2013) [Online] Optimierung  
<http://de.wiktionary.org/wiki/Optimierung> [letzter Aufruf: 05.12.2014].
- Williams, G. J. (2009). Rattle: a data mining GUI for R. *The R Journal*, 1(2), 45 – 55.
- Zangemeister, C. (1976). Nutzwertanalyse in der Systemtechnik: eine Methodik zur multidimensionalen Bewertung und Auswahl von Projektalternativen. Auflage: 4., Aufl. (1976).
- Zell, A. (1997). Simulation Neuronaler Netze.
- Zillow. (2014) [Online] Zillow Neighborhood Boundaries.  
<http://www.zillow.com/howto/api/neighborhood-boundaries.htm> [letzter Aufruf: 05.12.2014].
- Zillow (2014) [Online] Gramercy Home Prices & Values.  
<http://www.zillow.com/gramercy-new-york-ny/home-values/> [letzter Aufruf: 05.12.2014].

# ANHANG

## Anhang 1: Leitfaden Expertenbefragungen

Allgemeine Fragen:

- Könnten Sie sich kurz persönlich vorstellen?
- Was ist Ihr derzeitiger Arbeitsbereich/Tätigkeitsschwerpunkt?
- Inwiefern hat die Darstellung von Immobilienpreisen in Form von Immobilienpreiskarten bei Immobilienportalen eine Bedeutung?
- Wie bewerten Sie die Qualität von Immobilienpreiskarten?
- Nennen Sie die aus Ihrer Sicht drei bis vier größten Problemfelder bei der Kartendarstellung von Immobilienpreisen.
- Inwiefern spielen Kartendarstellungen bei der Vermittlung von Immobilien auf Immobilienportalen eine Rolle?
- Nennen Sie die aus Ihrer Sicht drei bis vier größten Problemfelder bei der Kartendarstellung von Immobilienangeboten?
- Inwiefern spielt bei Immobilienportalen die räumliche geoinformationstechnische Analyse von Immobiliendaten eine Rolle.
- Nennen Sie die aus Ihrer Sicht drei bis vier größten Problemfelder bei der räumlichen Analyse von Immobilienpreisdaten bei Immobilienportalen.
- Welche räumlichen Optimierungspotentiale bestehen aus Ihrer Sicht bei Immobilienportalen und welche sind die dringlichsten?

Spezifische Fragen:

Dr. Arne Schwarz:

- Welcher Grad an Interaktivität ist bei der Darstellung von Angebots- und Immobilienpreiskarten sinnvoll?
- Welche Maßstabsebenen sind bei der Preis- und Angebotsdarstellung für Nutzer von Portalen interessant?
- Welche räumlichen Analysemethoden nutzen Sie bei Ihren Angeboten „Mietpreischeck“ und „Immobilienbewertung/Wertgutachten“?
- Wo sehen Sie Optimierungspotentiale bei den genannten Angeboten?
- Welche Klassengrenzen und wie viele Klassen sind bei der Kartendarstellung von Mietpreisen aus Ihrer Sicht sinnvoll?

Herr Lars Holger Wilke:

- Welche Methoden zur Immobilienpreisschätzung verwendet ihr Portal? Was ist der „räumliche Anteil“ an diesen Methoden?
- Welche der folgenden Methoden halten Sie für relevant, um die räumliche Modellierung von Immobilienportalen zu optimieren: PCA Hauptkomponentenanalyse/räumliche Schätzverfahren/Verfahren aus dem Maschinellen Lernen?

- Entspricht die räumliche Preisverteilung auf Ihren Karten den realen Verhältnissen?
- Wo sehen Sie Optimierungspotentiale, um derzeitige räumliche Modelle zu verbessern?
- Welche Sekundärvariablen sind die aus Ihrer Sicht relevantesten zur Analyse und Modellierung von Immobilienpreisen?

#### Jan Hebecker

- Erhöht die Darstellung von Immobilienpreisen und Immobilienangeboten die Transparenz von Immobilienportalen?
- Laut Ihrem CEO Greg Ellis sollten Daten besser aufbereitet und für Kunden als Mehrwert präsentiert werden. Spielen webbasierte Karten und Methoden und Techniken der Geoinformatik/räumlichen Analyse bei diesen Überlegungen eine Rolle?

#### Adaptiertes Experteninterview mit Herrn Peter Ache

Allgemeine Fragen:

- Könnten Sie sich kurz persönlich vorstellen?

Was ist Ihr derzeitiger Arbeitsbereich/Tätigkeitsschwerpunkt?

- Inwiefern hat die Darstellung von Immobilienpreisen in Form von Karten für einen Gutachterausschuss Bedeutung?
- Kann man sich vorstellen, dass die Datengrundlage der Gutachterausschüsse zur Erstellung der Bodenrichtwertkarten durch Daten von Immobilienportalen erweitert wird?
- Sehen Sie Problemfelder bei der Datenerhebung bzw. Datenanalyse von Immobilienportalen?
- Spielen Immobilienpreiskarten eine Rolle in der Verbesserung der Transparenz von Immobilienmärkten?
- Wie bewerten Sie die Qualität von Immobilienpreiskarten auf Immobilienportalen im Allgemeinen?
- Wie bewerten Sie die Qualität der Bodenrichtwertkarten in Deutschland?
- Nennen Sie die aus Ihrer Sicht drei bis vier größten Problemfelder bei der Kartendarstellung von Immobilienpreisen.
- Die Kartendarstellung erfolgt bei Bodenrichtwerten auf Grundlage der amtlichen Daten. Kann man sich hier eine Änderung der Kartengrundlage vorstellen?
- Inwiefern spielt bei der Erstellung der Bodenrichtwertkarten eine räumliche geoinformationstechnische Analyse eine Rolle?
- Nennen Sie die aus Ihrer Sicht drei bis vier größten Problemfelder bei der räumlichen Analyse.
- Ist es aus Ihrer Sicht möglich, über die Schätzung von Angebotsmieten Rückschlüsse auf den Bestand zu ziehen oder Prognosemodelle aufzustellen?

## Anhang 2: Auswertematrix technologische Recherche

Nation	Immobilienportal	Immobilienpreiskarte	Dynamische Legende	Künstliche Administrative Grenzen	Preismodellierung mittels räumlicher Statistik	Statistische Auswertungen	Hedonische Regression	API	Grundkarte / Basemap	Pin / Marker Problematik	Zusätzliche Informationsebenen / Zusatzinformation im Marker	Indikatorenbasierte Distanzabhängige Suche	Visualisierung von Distanzen	Immobilienindex	Mobile Angebotskarten	Alexa Rank Global (27.11.14)	Alexa Rank national
USA	zillow.com	J	N	J	T	J	J	J	bing	J	J	N	N	J	J	210	49
USA	trulia.com	J	N	J	J	J	J	J	google	J	J	T	J	J	J	502	130
UK	zoopla.com	J	N	N	N	J	J	J	google	J	J	J	N	J	J	735	65
UK	realtor.com	J	N	J	N	J	N	J	N	K	J	N	N	N	N	813	226
DE	ImmobilienScout24.de	J	N	J	N	J	J	J	google	J	N	T	T	J	J	838	22
USA	redfin.com	N	N	K	N	N	N	N	google	N	N	N	N	N	J	1320	327
AT	Willhaben.at	N	N	K	N	N	N	N	bing	N	J	N	N	N	J	2075	8
AT	DerStandard.at	N	N	K	N	N	N	J	google	K	K	K	N	N	T	2397	12
DE	Immowelt.de	N	N	K	N	J	J	J	bing	T	N	T	T	N	J	3584	129
DE	Immonet.de	N	N	K	N	N	N	J	google	J	J	N	N	N	J	3678	133
Global	Movoto.com	N	N	N	N	N	N	N	google	J	J	N	N	N	J	4514	967
USA	apartmentguide.com	T	N	J	N	J	N	N	google	T	N	N	N	N	J	7155	1672
USA	hotpads.com	N	N	N	N	N	N	N	google	T	J	N	N	N	J	8403	2257
DE	http://www.wg-gesucht.de	N		K	N	N	N	N	google	N	N	T	T	N	N	9981	486
CH	Homegate.ch	J	N	J	N	J	J	J	google	N	N	N	N	J	J	10582	43
USA	remax.com	T	N	K	N	N	N	N	bing	J	J	N	N	N	N	12025	3140
USA	rent.com	T	N	N	N	N	N	N	google	N	N	J	N	N	J	13346	2960
USA	ziprealty.com	N	N	J	N	K	N	N	google	J	J	N	N	N	J	14021	3306
CH	ImmobilienScout24.ch	N	N	K	N	N	N	N	google	K	N	N	N	N	J	20553	68
UK	foxtons.co.uk	N	N	N	N	N	N	K	Statisch	J	J	N	N	N	J	30274	1816
USA	homes.com	N	N	J	N	J	K	J	Bing	N	N	N	N	J	J	33307	1127
Global	LaMudi.com	N	N	N	N	N	N	J	google	J	N	N	N	N	N	34972	N
AT	Immobilien.net	N	N	J	N	J	J	T	N	N	N	N	N	J	N	55932	305
UK	vebra.com	N	N	N	N	N	N	J	google	J	N	N	N	N	J	57819	2108
UK	mouseprice.com	J	N	K	N	J	J	N	Bing	J	N	N	N	J	J	63953	2249
AT	ImmobilienScout24.at	N	N	N	N	N	N	N	google	N	J	N	N	N	J	64149	230
CH	home.ch	N	N	K	N	N	N	N	google	K	J	N	N	N	N	138310	631
UK	findproperly.co.uk	N	N	K	N	N	N	N	google	J	J	J	J	N	N	402223	23550
AT	Nestoria.com	N	N	N	N	N	N	N	bing	T	N	N	N	N	J	807000	N
DE	buerosuche.de	J	N	J	N	N	N	N	google	N	N	N	N	N	J	14707254	K
DE	wohnungskarte.net	N	N	K	N	N	N	N	eetmap	N	J	N	N	N	J	K	K
USA	homes.yahoo.com	N	N	K	N	N	N	T	here	J	T	N	J	N	J	K	K
J	Ja																
N	Nein																
T	Teilweise																
K	Keine Angaben / Informationen																

## Anhang 3: Variablen des Angebotsdatensätze

<b>Variable</b>	<b>Beschreibung</b>	<b>Variable</b>	<b>Beschreibung</b>
<b>GEOX</b>	X Koordinate	<b>GAESTEWC</b>	Gästezimmer vorhanden J/N
<b>GEOY</b>	Y Koordinate	<b>MOEBLIERT</b>	Möbliert J/N
<b>PLZ</b>	Postleitzahl	<b>NUTZFLAECH</b>	Wohnfläche inklusive Nutzfläche (z. B. Abstellraum)
<b>ORT</b>	Ort	<b>GRUNDFLAEC</b>	Grundfläche – bei Wohnungen nicht vorhanden
<b>STRASSE</b>	Straße	<b>BALKON</b>	Balkon J/N
<b>HAUSNRVON</b>	Hausnummernbereich	<b>GARTEN</b>	Garten oder Mitnutzung J/N
<b>HAUSNRBIS</b>	Hausnummernbereich	<b>PARKPLATZ</b>	Parkplatz vorhanden J/N
<b>WOHNFLAECH</b>	Wohnfläche in m <sup>2</sup>	<b>PARKPLATZP</b>	Parkplatzpreis – Kauf- oder Mietpreis
<b>ZIMMERANZA</b>	Zimmeranzahl	<b>ROLLSTUHLG</b>	Rollstuhlgerecht J/N
<b>BAUJAHR</b>	Baujahr	<b>BETREUT</b>	Betreutes Wohnen J/N
<b>COURTAGE</b>	Maklercourtage	<b>HTID</b>	Haustiere erlaubt J/N
<b>MIETEKALT</b>	Kaltmiete	<b>FOERDERUNG</b>	Förderung J/N
<b>MIETEWARM</b>	Warmmiete	<b>SCHLAFZIMM</b>	Anzahl der Schlafzimmer
<b>MIETEKAUTI</b>	Mietkaution	<b>WOHNGELD</b>	Höhe des Wohngeldes
<b>EDATUM</b>	Einstellungsdatum des Angebotes	<b>AUFZUG</b>	Aufzug vorhanden J/N
<b>KAUFVERMIE</b>	Kaufvermietung		
<b>NEBENKOSTE</b>	Nebenkosten		
<b>FREIAB</b>	Datum der Verfügbarkeit eines Objektes		
<b>ETAGE</b>	Etage, in der das Objekt liegt		
<b>ANZAHEETAG</b>	Anzahl der Etagen des Objektes		

#### Anhang 4: Angebotsdichten und Durchschnittsnetto- kaltmieten pro Jahr in Potsdam

<b>Jahr</b>	<b>Anzeigen</b>	<b>Angebotsdichte /km<sup>2</sup></b>	<b>Miete qm/mean /km<sup>2</sup></b>
<b>2007</b>	11502,00	66,85	7,00
<b>2008</b>	10882,00	63,25	7,21
<b>2009</b>	10741,00	62,43	7,39
<b>2010</b>	10619,00	61,72	7,41
<b>2011</b>	9531,00	55,40	7,89
<b>2012</b>	10922,00	63,48	8,23
<b>2013</b>	9902,00	57,55	8,34

## Anhang 5: Autokorrelation der Angebotsdaten – Potsdam Erstes Quartal 2013

Test	Distanzmatrix 15000 m	Distanzmatrix 5000m	Distanzmatrix 1000m	Distanzmatrix 500m	Distanzmatrix 300m
<b>Morans' I</b>	-0.001532212	0.08578627	0.3263379	0.5531951	0.459847
<b>p – Werte</b>	1,20E-97	0	0	0	0
	Index	Erwartungswert	Varianz	Standardabweichung	p-Wert
<b>Global Moran's I</b>	0.8077773677	-0.0003295979	1,41E+02	6.801.507	0.000000e+00
<b>Geary's C</b>	0.2026486354	10.000.000.000	2,12E+02	5.471.684	0.000000e+00
<b>Getis Ord G</b>	0.0003529526	0.0003295979	3,34E-06	1.278.581	9,84E-32

## Anhang 6: Random Forest Score für sechs künstliche Angebotswerte aus dem ersten Quartal 2013

Typ	Jahr	VERWEILDAU	WOHNFLAECH	BAUJAHR_OR	ZIMMERANZA	Miete RF 500	Adresse
Neubau/Platte	2013	6	80	2010	3	6.22107632078123	
Wohnung	2013	1	110	NA	3,5	7.28064988430286	Zum Exerzierhaus 1b
Reihenhaus	2013	1	115,74	2013	4	6.86935843521461	Brunno Trautstrasse 11
Reihenhaus	2013	1	125	2013	4	6.84855159937863	Brunno Trautstrasse 8a
Studentenwohnhei	2013	1	20,1	NA	1	8.48228927516285	Gegenüber FH / Youni
Wohnung / Neubau	2013	1	103,32	NA	3	7.12738636755719	Voltaireweg 7

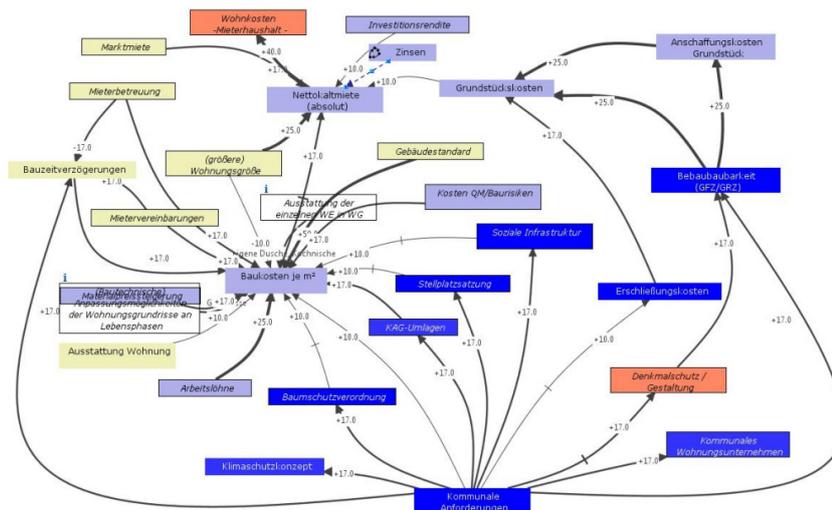
## Anhang 7: Fehlerreport RF Modell mit 150 Bäumen

Summary of the Random Forest Model		
=====		
Number of observations used to build the model: 923		
Missing value imputation is active.		
Call:		
randomForest(formula = Miete_qm ~ .,		
data = crs\$dataset[crs\$sample, c(crs\$input, crs\$target)],		
ntree = 150, mtry = 3, importance = TRUE, replace = FALSE, na.action = na.roughfix)		
Type of random forest: regression		
Number of trees: 150		
No. of variables tried at each split: 3		
Mean of squared residuals: 1.138482		
% Var explained: 62.42		
Variablenwichtigkeit		
=====		
%IncMSE	IncNodePurity	
BAUJAHR_OR	30.54	366.99
EINBAUKUEC	29.07	143.69
WOHNFLAECH	26.06	308.81
ZIMMERANZA	20.32	106.51
ETAGE	18.04	117.71
AUFZUG	17.88	82.06
LETZTE_MOD	17.72	113.04
BALKON	16.04	45.06
GARTEN	15.49	80.44
KELLER	11.95	52.14
VERWEILDAU	8.37	67.67
DENKMALOBJ	0.34	0.55
Benötigte Zeit: 0.49 Sek		
Rattle Zeitstempel: 2014-12-31 11:09:13 Harald		
=====		

## Anhang 8: Ergebnisse der PCA: Q1/2013

	Comp, 1	Comp, 2	Comp, 3	Comp, 4	Comp, 5
BAUJAHR_ORIGINAL	0,27658143	-0,64990300	0,00305574	0,01350651	-0,03248185
ETAGE	0,08275550	0,05954699	0,72114720	0,66452356	-0,16688160
MIETEKALT	-0,53635046	-0,20866533	-0,42328910	-0,02129057	-0,57141997
OBJEKT_ALTER	-0,04475435	0,64981287	-0,00005577	-0,01597760	0,02685929
VERWEILDAUER	-0,56004780	-0,03620905	0,68589060	-0,72360006	0,04852489
WOHNFLAECHE	-0,48655787	-0,22132039	0,01949654	0,03366031	-0,18717745
ZIMMERANZAHL	-0,48655787	-0,24080790	0,08555414	0,18107283	0,77875747
Rcmdr> ,PC\$sd^2 # component variances					
	Comp, 1	Comp, 2	Comp, 3	Comp, 4	Comp, 5
	2,70576163	1,87526662	1,04836721	0,95129503	0,34187775

## Anhang 9: Modellierte Einflussfaktoren auf die Nettokaltmiete (Potsdam 22, 2012)



# ERKLÄRUNG

Hiermit versichere ich, dass diese Arbeit selbstständig und unter ausschließlicher Verwendung der angegebenen Mittel angefertigt wurde. Diese Dissertation ist bisher nicht veröffentlicht und wurde an keiner weiteren Hochschule eingereicht.

Gez. Harald Scherthanner, gez. Potsdam den 12.06.2015