

VERGLEICH DER AMTLICHEN HAUSKOORDINATEN UND HAUSUMRINGE DEUTSCHLAND MIT OPENSTREETMAP-DATEN

***Bachelorarbeit zur Erlangung des akademischen Grades
Bachelor of Engineering (B. Eng.)***

vorgelegt von

Saam, Sophia

Geburtsdatum 28. August 1997

Gutachter:

Herr Prof. Dr.-Ing. Andreas Schmitt
Hochschule für angewandte Wissenschaften München

Externer Betreuer:

Florian Siegert
Landesamt für Digitalisierung, Breitband und Vermessung

München, 21.06.2020

I. KURZZUSAMMENFASSUNG

Mit Koordinaten versehene postalische Adressen und geografisch korrekten Hausgrundrisse sind in der heutigen Zeit nicht mehr aus dem Leben wegzu-denken. Sie sind z.B. für Navigation und Infrastrukturplanung wichtig.

In dieser Arbeit werden postalische Adressen und georeferenzierte Hausgrundrisse aus zwei verschiedenen Quellen miteinander hinsichtlich *geografischer / geometrischer Übereinstimmung, Vollständigkeit* und *Informationsgehalt* vergli-chen: die Datensätze Hauskoordinaten Deutschland (HK-DE) und Hausumringe Deutschland (HU-DE) von der Zentralen Stelle Hauskoordinaten und Hausumringe (ZSHH) und die Gebäude- und Punkt-Daten von OpenStreetMap (OSM). Es werden vier Kategorien mit je vier Testgebieten festgelegt: *gebirgige Gebiete, Großstädte und Ballungsräume, ländliche Gebiete* und *mittelgroße Städte*. Es wird geprüft, ob für die Gebäude ein Gegenstück im anderen Datensatz vor-handen ist und wenn, wie gut die Gebäudegrundrisse beider Datenquellen zu-sammenpassen. Zudem wird geprüft, welche Attribute in den Datensätzen vor-handen sind und wie oft diese ausgefüllt sind. Da die Datensätze von OSM und von der ZSHH für unterschiedliche Zwecke erstellt wurden, ist es nicht verwun-derlich, dass es zum Teil große Unterschiede zwischen den Datensätzen gibt. Es zeigt sich, dass die ZSHH-Daten mehr Gebäude und postalische Adressen als die OSM-Daten enthalten und die HU- und HK-Attribute häufiger ausgefüllt sind, als die OSM-Adress-Attribute.

Abstract

Postal addresses with coordinates and geographically accurate building ground views are nowadays indispensable. They are e.g. important for navigation and infrastructure design.

In this bachelor thesis postal addresses and georeferenced building ground views from two different sources are compared, regarding *geographical / geometrical compliance, completeness* and *informational content*. This sources are: the datasets Hauskoordinaten Deutschland (HK-DE) (building coordinates germany) and Hausumringe Deutschland (HU-DE) (building borders germany) from the Zentralen Stelle Hauskoordinaten und Hausumringe (ZSHH) (central department building coordinates and building borders) and the point and building data from openstreetmap (OSM). Four different classes with four test areas each are defined: *mountainous areas, major cities and metropolitan areas, rural areas* and *mid-sized cities*. It will be verified whether all buildings have a counterpart in the other data set and if, how well the ground views of both data sources go together. Furthermore is checked, which attributes in both data sets exist and how often they are filled. Since the data sets from OSM and the ZSHH are created for different purposes, it is hardly surprising that there are partly large differences between the data sets. It shows that the ZSHH data has more buildings and postal addresses than the OSM data and that the HU attributes and HK attributes are more often filled than the OSM address attributes.

II. INHALTSVERZEICHNIS

Inhaltsverzeichnis

I.	Kurzzusammenfassung	2
II.	Inhaltsverzeichnis	4
III.	Tabellenverzeichnis.....	6
IV.	Abbildungsverzeichnis.....	9
V.	Abkürzungsverzeichnis	11
1.	Einleitung	12
2.	Grundlagen und Daten	15
2.1	OpenData und OpenStreetMap	15
2.2	Zentrale Stelle Hauskoordinaten und Hausumringe	17
2.3	Aufbau der Datensätze	18
2.3.1	HK-DE	19
2.3.2	HU-DE	22
2.3.3	OSM	24
3.	Vorgehensweise und Methodik	26
3.1	Ablauf	26
3.2	Vergleich der OSM-Datensätze aus verschiedenen Quellen.....	27
3.3	Vergleichskriterien	35
3.3.1	Geografische / Geometrische Übereinstimmung.....	36
3.3.2	Vollständigkeit	38
3.3.3	Informationsgehalt.....	39
3.4	Auswahl und Begründung der Testgebiete	39
3.5	Festlegung von Toleranzklassen	43
3.6	Umsetzung mit QGIS und Tableau	44
4.	Ergebnisse	50

4.1 Ergebnisse des Vergleichs der geografischen/ geometrischen Übereinstimmung	51
4.2 Ergebnisse des Vollständigkeitsvergleichs	57
4.3 Ergebnisse des Informationsgehaltsvergleichs.....	59
5. Diskussion	63
6. Zusammenfassung / Fazit / Ausblick.....	68
VI. Literaturverzeichnis	69
VII. Anhang.....	71
Anhang 1: Programmschritte mit Screenshots	71
Anhang 2: Ergebnistabellen.....	85

III. TABELLENVERZEICHNIS

TABELLENVERZEICHNIS

Tabelle 1: Attributlisten der OSM-Daten aus verschiedenen Quellen.	29
Tabelle 2: Vorprozessierungsschritte für die verschiedenen OSM-Quellen.	33
Tabelle 3: Übersicht der Vergleichskriterien.....	35
Tabelle 4: Bundesländer mit und ohne OpenData-Initiativen.....	39
Tabelle 5: Zuordnung der Jaccard-Koeffizienten-Intervalle zu den Toleranzklassen.....	43
Tabelle 6: Mittelwert, Median und Maximum der HU-, OSM-Gebäude- und Überdeckungsflächen für die Kategorien.	52
Tabelle 7: Mittelwert, Median und Maximum der Gebäudepunkteanzahl der HU und der OSM-Gebäude für die Kategorien.....	53
Tabelle 8: Jaccard-Koeffizient Mittelwert, Median und für das ganze Testgebiet für die Kategorien.....	53
Tabelle 9: Unterschied Gebäudeanzahl HU und OSM.....	57
Tabelle 10: Relative Häufigkeit [%] der OSM-Attribute name und shop.	60
Tabelle 11: Übereinstimmende postalische Adresse.	62
Tabelle 12: OSM-Punkte die in OSM-Gebäuden fallen.....	86
Tabelle 13: Mittel, Median, Minimum und Maximum Gebäudeflächenvergleich HU und OSM-Gebäude und Überdeckungsfläche.	87
Tabelle 14: Mittel, Median, Minimum und Maximum Gebäudepunktevergleich HU und OSM-Gebäude.	88
Tabelle 15: Jaccard-Koeffizient Mittelwert, Median.....	89
Tabelle 16: Jaccard-Koeffizient absolute Häufigkeit der Toleranzklassen.	90
Tabelle 17: Jaccard-Koeffizient relative Häufigkeit der Toleranzklassen (Prozent).	91
Tabelle 18: Toleranzklassen Jaccard-Koeffizient CDF.....	92
Tabelle 19: Jaccard-Koeffizienten der Testgebiete.	93
Tabelle 20: Toleranzklassen Gebäude absolute Häufigkeit.	94
Tabelle 21: Toleranzklassen Gebäude relative Häufigkeit in Prozent.	95
Tabelle 22: Vergleich Gebäudeanzahl OSM und HU.	96
Tabelle 23: Gebäude ohne Gegenstück im anderen Datensatz.....	97

Tabelle 24: Ausgefüllte HK-Attribute absolute Häufigkeit in gebirgigen Gebieten.	98
Tabelle 25: Ausgefüllte HK-Attribute relative Häufigkeit in Prozent in gebirgigen Gebieten.	99
Tabelle 26: Ausgefüllte HK-Attribute absolute Häufigkeit in Großstädten und Ballungsräumen.	100
Tabelle 27: Ausgefüllte HK-Attribute relative Häufigkeit in Prozent in Großstädten und Ballungsräumen.	101
Tabelle 28: Ausgefüllte HK-Attribute absolute Häufigkeit in ländlichen Gebieten.	102
Tabelle 29: Ausgefüllte HK-Attribute relative Häufigkeit in Prozent in ländlichen Gebieten.	103
Tabelle 30: Ausgefüllte HK-Attribute absolute Häufigkeit in mittelgroßen Städten.	104
Tabelle 31: Ausgefüllte HK-Attribute relative Häufigkeit in Prozent in mittelgroßen Städten.....	105
Tabelle 32: Ausgefüllte OSM-Gebäude-Attribute absolute Häufigkeit in gebirgigen Gebieten.....	106
Tabelle 33: Ausgefüllte OSM-Gebäude-Attribute relative Häufigkeit in Prozent in gebirgigen Gebieten.....	107
Tabelle 34: Ausgefüllte OSM-Gebäude-Attribute absolute Häufigkeit in Großstädten und Ballungsräumen.	108
Tabelle 35: Ausgefüllte OSM-Gebäude-Attribute relative Häufigkeit in Prozent in Großstädten und Ballungsräumen.	109
Tabelle 36: Ausgefüllte OSM-Gebäude-Attribute absolute Häufigkeit in ländlichen Gebieten.	110
Tabelle 37: Ausgefüllte Attribute OSM-Gebäude relative Häufigkeit in Prozent in ländlichen Gebieten.	111
Tabelle 38: Ausgefüllte OSM-Gebäude-Attribute absolute Häufigkeit in mittelgroßen Städten.....	112
Tabelle 39: Ausgefüllte OSM-Gebäude-Attribute relative Häufigkeit in Prozent in mittelgroßen Städten.....	113

Tabelle 40: Ausgefüllte Attribute OSM-Punkte in Gebäuden absolute Häufigkeit in gebirgigen Gebieten.	114
Tabelle 41: Ausgefüllte Attribute OSM-Punkte in Gebäuden relative Häufigkeit in Prozent in gebirgigen Gebieten.	115
Tabelle 42: Ausgefüllte Attribute OSM-Punkte in Gebäuden absolute Häufigkeit in Großstädten und Ballungsräumen.	116
Tabelle 43: Ausgefüllte Attribute OSM-Punkte in Gebäuden relative Häufigkeit in Prozent in Großstädten und Ballungsräumen.	117
Tabelle 44: Ausgefüllte Attribute OSM-Punkte in Gebäuden absolute Häufigkeit in ländlichen Gebieten.	118
Tabelle 45: Ausgefüllte Attribute OSM-Punkte in Gebäuden relative Häufigkeit in Prozent in ländlichen Gebieten.	119
Tabelle 46: Ausgefüllte Attribute OSM-Punkte in Gebäuden absolute Häufigkeit in mittelgroßen Städten.	120
Tabelle 47: Ausgefüllte Attribute OSM-Punkte in Gebäuden relative Häufigkeit in Prozent in mittelgroßen Städten.	121
Tabelle 48: Adressvergleich HU und OSM-Gebäude absolute Häufigkeit.	122
Tabelle 49: Adressvergleich HU und OSM-Gebäude relative Häufigkeit in Prozent.	122

IV. ABBILDUNGSVERZEICHNIS

ABBILDUNGSVERZEICHNIS

Abbildung 1: Screenshot HK-DE.....	19
Abbildung 2: Ablaufdiagramm.....	26
Abbildung 3: Beispiel einer CDF Quelle: (vgl. Cramer & Kamps, 2008, p. 18).	38
Abbildung 4: Karte der Testgebiete.....	42
Abbildung 5: Schnittmenge.....	47
Abbildung 6: Schnittmenge nach OI aufgelöst.....	47
Abbildung 7: Vereinigungsmenge.....	47
Abbildung 8: Vereinigungsmenge nach OI aufgelöst.....	47
Abbildung 9: Beispiel einer Formel für ein berechnetes Feld in Tableau um eine Übereinstimmung der Adresse zu ermitteln.....	49
Abbildung 10: Diagramm Anteil der OSM-Punkte-in-Gebäuden an allen OSM-Punkten in %. Balkenfarben: gebirgige Gebiete: gelb, Großstädte und Ballungsräume: rot, ländliche Gebiete: grün, mittelgroße Städte: blau.....	51
Abbildung 11: Diagramme der Jaccard-Koeffizient-Verteilung der Kategorien.....	54
Abbildung 12: CDF-Stufendiagramme der Jaccard-Koeffizient-Verteilung der Kategorien.....	55
Abbildung 13: Toleranzklassen-Verteilung der Gebäudeübereinstimmung der verschiedenen Kategorien.....	56
Abbildung 14: Gebäude ohne Gegenstück im anderen Datensatz (relative Häufigkeit).....	58
Abbildung 15: OSM-Gebäude (grau), HU-Gebäude (rot) und OSM-Gebäude überlagert mit den HU-Gebäuden.....	64
Abbildung 16: HK-Layer aus Textdatei erzeugen.....	72
Abbildung 17: OSM-Punktelaye mit dem QuickOSM-Plugin erzeugen.....	74
Abbildung 18: OSM-Gebäudelaye mit dem QuickOSM-Plugin erzeugen.....	75
Abbildung 19: Attribut testgebiet erzeugen.....	76
Abbildung 20: OSM-Punkte-in-Gebäuden-Layer erzeugen.....	77
Abbildung 21: Attribut Fläche erzeugen.....	78
Abbildung 22: Layer zu einem deutschlandweiten Layer zusammenfügen.....	80

Abbildung 23: Attribute zusammenfassen.....	81
Abbildung 24: einen Layer für Gebäude ohne Gegenstück im anderen Datensatz erzeugen.	82
Abbildung 25: Tableau Beispiel einer Verknüpfungsbedingung.	83
Abbildung 26: IoU in Tableau berechnen.	83
Abbildung 27: Gruppierung Jaccard-Koeffizient.	84
Abbildung 28: Abfrage übereinstimmender Straßenname.	85

V. ABKÜRZUNGSVERZEICHNIS

AdV	Arbeitsgemeinschaft der Vermessungsverwaltungen der Länder der Bundesrepublik Deutschland
AGNB	Allgemeine Geschäfts- und Nutzungsbedingungen bzw. Allgemeine Bedingungen für die Bereitstellung und Nutzung von Geodaten
ALKIS	Amtliches Liegenschaftskataster Informationssystem
API	Application Programming Interface (Programmierschnittstelle)
CDF	Cumulative distribution function (empirische Verteilungsfunktion)
HU-DE	Hausumringe Deutschland
HK-DE	Hauskoordinaten Deutschland
IoU	Intersection over Union (Jaccard-Koeffizient)
OSM	OpenStreetMap
ZSHH	Zentrale Stelle Hauskoordinaten und Hausumringe

1. EINLEITUNG

Mit Koordinaten versehene postalische Adressen sind in der heutigen Zeit nicht mehr aus dem Leben wegzudenken. Sie ermöglichen zum Beispiel, dass Navigationsgeräte die Route zur richtigen Adresse anzeigen oder Lieferunternehmen eine effiziente Auslieferung planen können (vgl. Kaaranen, 2005, pp. 245–246). Auch die geografisch korrekten Hausgrundrisse sind in den unterschiedlichsten Bereichen von Nöten (vgl. Landesamt für Vermessung und Geoinformation Thüringen, 2014). Wenn z. B. ein neues Glasfaserkabel oder eine neue Wasserleitung geplant werden soll, beeinflussen Lage und Größe von Gebäuden die Verlaufsmöglichkeiten. Aber auch für Versicherungsunternehmen spielt die Lage der Gebäude zusammen mit anderen Informationen hinsichtlich der Risikokalkulation eine Rolle: liegt das zu versichernde Gebäude an der Bachseite eines Grundstücks oder an der Straßenseite etc.

Von den Vermessungsverwaltungen der Länder werden in Deutschland sogenannte Hauskoordinaten Deutschland (HK-DE) und Hausumringe Deutschland (HU-DE) gegen Gebühr zur Verfügung gestellt. Die Hauskoordinaten sind die mit Koordinaten versehenen postalischen Adressen, und die Hausumringe sind die geografisch korrekten Hausgrundrisse. Die beiden Datensätze sind nicht miteinander verknüpft. Durch die in den Daten enthaltenen Koordinaten können sie dennoch, falls gewünscht, einfach in einem GIS kombiniert werden. Die Bereitstellung von Geobasisinformationen basiert auf dem gesetzlichen Auftrag (vgl. Arbeitsgemeinschaft der Vermessungsverwaltungen der Länder der Bundesrepublik Deutschland) durch die INSPIRE-Direktive (INfrastructure for SPatial InfoRmation in Europe|Nfrastructure) der EU und ihrer deutschen Umsetzung der GDI-DE (Geodateninfrastruktur). Allerdings fragen sich einige Kunden, die Hauskoordinaten und Hausumringe benötigen, ob sie diese Daten in Zeiten von OpenData nicht auch aus anderen Quellen kostenlos mit der gleichen oder ähnlichen Qualität beziehen können. OpenStreetMap (OSM) ist so eine freie Datenbezugsquelle.

Es gibt bisher keinen wissenschaftlichen Vergleich der OpenStreetMap-Daten mit den Datensätzen HK-DE und HU-DE hinsichtlich *geografischer / geometrischer Übereinstimmung, Vollständigkeit und Informationsgehalt* für die Bundesrepublik Deutschland. Die HK-DE ist eine von der Zentralen Stelle Hauskoordinaten und Hausumringe (ZSHH) bereitgestellte Textdatei, die die Koordinaten, postalischen Adressen und weitere Attribute sämtlicher im Liegenschaftskataster der verschiedenen Bundesländer enthaltenen Hauptgebäude mit Hausnummer enthält (vgl. Zentrale Stelle Hauskoordinaten und Hausumringe, Landesamt für Digitalisierung, Breitband und Vermessung Bayern, 2019a). Die HU-DE ist eine ebenfalls von der ZSHH bereitgestellte AdV-Shapedatei, die die Gebäudeumringe sämtlicher im Liegenschaftskataster enthaltener Gebäude und Bauwerke enthält. Nicht dargestellt werden Ausgestaltungsgeometrien, Dächer und unterirdische Gebäude. Die ZSHH-Daten werden gegen eine Gebühr unter speziellen Lizenzen vertrieben (vgl. Zentrale Stelle Hauskoordinaten und Hausumringe, Landesamt für Digitalisierung, Breitband und Vermessung Bayern, 2019b). OSM ist ein freies Internet-Kartenwerk, das allen kostenlos zur Verfügung steht. OSM-Daten sind auf XML basierende Daten, die z.B. Straßen, Gebäude und Wasserläufe darstellen. Es ist fraglich, ob die OSM-Daten in allen Bereichen untereinander die gleiche Genauigkeit aufweisen, da sie von freiwilligen Mappern erstellt werden und es vermutlich nicht in allen Regionen und Bereichen Deutschlands gleich viele Mapper gibt. Das Wort Mapper kommt vom englischen Wort „map“, also Karte, und bezeichnet diejenigen, die bei OSM die Kartedaten erstellen oder bearbeiten. Allerdings hat jedes Bundesland eigene Vermessungsvorschriften, weshalb es auch bei den HK und HU zwischen den einzelnen Bundesländern kleinere Unterschiede geben könnte.

Ziel meiner Bachelorarbeit ist es, einen replizierbaren und überprüfbaren Vergleich der ZSHH-Daten und der OSM-Daten zu erstellen. Dies dient dazu, eine Argumentationsbasis für die ZSHH gegenüber ihren Kunden und der Arbeitsgemeinschaft der Vermessungsverwaltungen der Länder der Bundesrepublik Deutschland (AdV), die die von der ZSHH gewünschte Neuerungen absegnen muss, zu schaffen.

Meine Forschungsfrage lautet: „Wie unterscheiden sich von OSM beziehbare Daten von HK-DE und HU-DE hinsichtlich geografischer / geometrischer Übereinstimmung, Vollständigkeit und Informationsgehalt in unterschiedlichen Regionen und unterschiedlichen Bundesländern in Deutschland?“

2. GRUNDLAGEN UND DATEN

In diesem Kapitel werden die Daten- und Literaturgrundlagen behandelt.

2.1 OPENDATA UND OPENSTREETMAP

Nach dem Artikel „Was sind offene Daten?“ von Dietrich aus dem Jahr 2011 können Daten dann als offen gelten, wenn sie von jedem und für jegliche Zwecke genutzt, weiterverarbeitet und weiterverbreitet werden können. Daten, die zwar kostenlos veröffentlicht wurden, aber keine Lizenz besitzen, die eine Weiterverarbeitung gestattet, oder in einem schwer verarbeitbaren oder nicht maschinenlesbaren Format vorliegen, können nicht als offene Daten (OpenData) bezeichnet werden. Offene Lizenzen sind leider nicht immer miteinander kompatibel, daher können nicht alle offenen Daten miteinander kombiniert werden. Vor einer Nutzung von offenen Daten aus verschiedenen Quellen sollten daher die Lizenzen auf Kompatibilität geprüft werden. Unter <http://opendefinition.org/licenses/> findet sich eine Liste mit Lizenzen, die als offen gelten, und Erklärungen zu ihnen. Offene Daten sollten möglichst in einem offenen Standard zur Verfügung gestellt werden, also einem Standard, der für alle leicht zugänglich, einsetzbar und weiterentwickelbar ist, und nicht nur von ausgewählten Programmen gelesen oder verarbeitet werden kann oder Eigentum einer Firma ist. Offene Daten können auf drei Ebenen vorliegen: als Rohdaten, als Programmierschnittstelle (API, „Application Programming Interface“), die den Zugriff auf die Rohdaten ermöglicht, und/oder als Anwendungen, die auf der API basieren (vgl. Dietrich, 2011).

Im Jahr 2004 entstand das Projekt OpenStreetMap (OSM), nachdem 2000 die künstliche Verschlechterung (Selective Availability, SA) des zivilen GPS-Signals endgültig eingestellt wurde, und genaue Kartenwerke nur für relativ viel Geld erhältlich und stellenweise ungenau oder falsch waren. Zudem unterlagen diese Kartenwerke urheberrechtlichem Schutz, sodass eine Weiterbearbeitung unmöglich war (vgl. Koch, Hrsg. [Bill], & Donaubaue[r] [Hrsg.], 2013, pp. 182–188).

Bei OSM arbeiten freiwillige Mitwirkende (Mapper) an einer freien Weltkarte, die anhand selbstaufgezeichneter GNSS-Positionen, amtlicher oder öffentlicher Datenspenden und Luft- und Satellitenbildern erstellt wurde. Im Gegensatz zu anderen Kartenwerken werden die Rohdaten unter einer freien Lizenz (CC-BY-SA bis September 2012, seitdem ODbL) veröffentlicht. Inzwischen hat sich die Community für Importe sehr strenge Regeln gegeben, da es in der Vergangenheit Probleme mit doppelt vorhandenen, gelöschten bestehenden Objekten oder ohne Prüfung ersetzten Objekten gab (vgl. Koch et al., 2013, pp. 182–188).

OSM bietet eine Fülle an Daten zu den verschiedensten Themen, z. B. zu Wegen, Wasserläufen, Eisenbahn, zur Energieversorgung, zu Kunstbauten, zur Nutzung von Gebäuden, zu Einrichtungen, zur Landnutzung, zu Historischem, Grenzen, Ortsangaben oder Adressen (vgl. Seip, Korduan, & Zehner, 2017, pp. 282–284).

OSM verfügt über eigene, aber offen dokumentierte, auf XML basierende, Dateiformate, über die Daten ausgetauscht bzw. heruntergeladen werden können. Die Daten liegen in Key-Value-Paaren, also Schlüssel-Wert-Paaren, vor. Von OSM wurde eine Palette von Softwareprodukten bzw. Werkzeugen zur Datenerzeugung, -nutzung, -editierung, -manipulation und -selektion entwickelt. Viele davon werden auf OSM-Wiki (<http://wiki.openstreetmap.org>) beschrieben (vgl. OSM-Community).

2.2 ZENTRALE STELLE HAUSKOORDINATEN UND HAUSUMRINGE

In Deutschland sind die einzelnen Bundesländer für die amtliche Vermessung zuständig und erlassen in eigenen Landesgesetzen die zu erfüllenden Aufgaben. Um eine weitgehende Harmonisierung und Einheitlichkeit zu erreichen, fortzuentwickeln und zu bewahren, koordiniert die AdV das amtliche deutsche Vermessungswesen. Neben den für das Vermessungswesen zuständigen Fachverwaltungen der Länder wirken in der AdV auch die Bundesministerien des Inneren, der Verteidigung sowie für Verkehr und die digitale Infrastruktur mit. Als Gäste gehören ihr die Deutsche Geodätische Kommission (DGK) als Vertretung der geodätischen Lehre und Forschung sowie die Arbeitsgemeinschaft Nachhaltige Landentwicklung (ArgeLandentwicklung) als Bund-Länder-Vertretung für ländliche Neuordnung an. Das Plenum der AdV definiert die fachliche und strategische Ausrichtung der AdV und fasst Beschlüsse von grundsätzlicher Bedeutung (vgl. Arbeitsgemeinschaft der Vermessungsverwaltungen der Länder der Bundesrepublik Deutschland). Es gibt auch drei zentrale Stellen, die Folgeprodukte auf Grundlage des Liegenschaftskatasters der Vermessungsverwaltungen der Länder bündeln und zu einheitlichen Gebühren- und Lizenzmodellen sowie in bundesweit einheitlichen Datenformaten bereitstellen. Eine dieser zentralen Stellen ist die Zentrale Stelle Hauskoordinaten und Hausumringe (ZSHH). Sie befindet sich am Bayerischen Landesamt für Digitalisierung, Breitband und Vermessung (LDBV) und ist der Ansprechpartner, wenn Hauskoordinaten, Hausumringe oder 3D-Gebäudemodelle in den Ausprägungen LoD1 und LoD2 (Level of Detail) bundesländerübergreifend oder bundesweit bezogen werden sollen (vgl. Zentrale Stelle Hauskoordinaten und Hausumringe, Landesamt für Digitalisierung, Breitband und Vermessung Bayern). Die Regelungen mit den Kunden werden auf Grundlage der Musterlizenzvereinbarung, der AGNB (Allgemeine Geschäfts- und Nutzungsbedingungen bzw. Allgemeine Bedingungen für die Bereitstellung und Nutzung von Geodaten) und der AdV-Gebührenrichtlinien getroffen. Bei der Lizenzierung wird zwischen interner und externer Nutzung der Daten unterschieden. Es sind unterschiedliche Laufzeiten und ein Bezug von Aktualisierungsdaten in einem vereinbarten Turnus möglich (vgl. Arbeitskreis PRM, Arbeitsgemeinschaft der Vermessungsverwal-

tungen der Bundesrepublik Deutschland [AdV], 2019). Alle Rechte an den Geodaten liegen beim Lizenzgeber, also der ZSHH. Eine Nutzung der Daten durch Umarbeitung, Vervielfältigung, Veröffentlichung im Internet etc. muss von der ZSHH genehmigt und mit dem deutlich sichtbaren Quellvermerk © GeoBasis-DE / LDBV 2019 versehen werden (vgl. Zentrale Stelle Hauskoordinaten und Hausumringe, Landesamt für Digitalisierung, Breitband und Vermessung Bayern, 2019c). Die Gebühr wird über die Objektanzahl erhoben und es gibt eine Mindestgebühr von 50€ pro Antrag. Für die Bereitstellung von aktualisierten Geobasisdaten werden pro Jahr 18% der für die erstmalige Bereitstellung dieser Daten geltenden Gebühren erhoben. Für die externe Nutzung der Daten wird zusätzlich eine Verwertungsgebühr erhoben. Hauskoordinaten kosten 0,15 € / Objekt und maximal 85.000 € und Hausumringe kosten 0,12 € / Objekt und maximal 81.000 € nach Stand 2019 (Arbeitsgemeinschaft der Vermessungsverwaltungen der Länder der Bundesrepublik Deutschland, 2019).

2.3 AUFBAU DER DATENSÄTZE

Die ZSHH liefert zwei verschiedene Datensätze: Hauskoordinaten Deutschland (HK-DE) und Hausumringe Deutschland (HU-DE). Die OSM-Daten können aus verschiedenen Quellen bezogen werden. Ein Vergleich der OSM-Datensätze aus verschiedenen Quellen erfolgt im Kapitel 3.2. Hier in diesem Kapitel finden sich lediglich allgemeine Informationen zu OSM-Datensätzen.

2.3.1 HK-DE

Die Informationen zu HK-DE sind der Datenformatbeschreibung Hauskoordinaten Deutschland der ZSHH entnommen. Das Abgabeformat der HK-DE ist eine Textdatei. Es werden drei zusammengehörende Dateien abgegeben: die Hauskoordinatendatei, die Entschlüsselungsdatei und die Umschlüsselungsdatei. Zeilen mit # als erstem Zeichen der Zeile sind Kommentarzeilen. Die einzelnen Datenelemente innerhalb der Datensätze in allen drei Dateien werden durch Semikolon getrennt (zurzeit sind es bei der HK-DE 17 Trennzeichen pro Datensatz, da es 18 Datenelemente gibt).

```
N;DESHPDHK0000KM8a;A:01;0;03;000;0100;00001;1;;32611060,728;5969737,251;Aegidienkirchhof;23552;Lübeck;;Innenstadt
N;DESHPDHK0000KM8A;A:01;0;03;000;0100;00002;3;;32610855,645;5969782,649;Aegidienstraße;23552;Lübeck;;Innenstadt
N;DESHPDHK0000KM8b;A:01;0;03;000;0100;00001;2;;32611059,388;5969732,525;Aegidienkirchhof;23552;Lübeck;;Innenstadt
N;DESHPDHK0000KM8B;A:01;0;03;000;0100;00002;30;;32611030,910;5969739,720;Aegidienstraße;23552;Lübeck;;Innenstadt
N;DESHPDHK0000KM8c;A:01;0;03;000;0100;00001;3;;32611057,924;5969727,211;Aegidienkirchhof;23552;Lübeck;;Innenstadt
N;DESHPDHK0000KM8C;A:01;0;03;000;0100;00002;31;;32610965,800;5969757,064;Aegidienstraße;23552;Lübeck;;Innenstadt
N;DESHPDHK0000KM8d;A:01;0;03;000;0100;00002;1;;32610855,645;5969782,649;Aegidienstraße;23552;Lübeck;;Innenstadt
N;DESHPDHK0000KM8D;A:01;0;03;000;0100;00002;32;;32611035,543;5969739,480;Aegidienstraße;23552;Lübeck;;Innenstadt
N;DESHPDHK0000KM8e;A:01;0;03;000;0100;00002;10;;32610932,006;5969749,724;Aegidienstraße;23552;Lübeck;;Innenstadt
N;DESHPDHK0000KM8E;A:01;0;03;000;0100;00002;33;;32610972,813;5969755,772;Aegidienstraße;23552;Lübeck;;Innenstadt
N;DESHPDHK0000KM8f;A:01;0;03;000;0100;00002;11;;32610877,517;5969777,653;Aegidienstraße;23552;Lübeck;;Innenstadt
N;DESHPDHK0000KM8F;A:01;0;03;000;0100;00002;34;;32611040,247;5969739,391;Aegidienstraße;23552;Lübeck;;Innenstadt
N;DESHPDHK0000KM8g;A:01;0;03;000;0100;00002;12;;32610938,212;5969748,805;Aegidienstraße;23552;Lübeck;;Innenstadt
N;DESHPDHK0000KM8G;A:01;0;03;000;0100;00002;35;;32610981,032;5969754,509;Aegidienstraße;23552;Lübeck;;Innenstadt
N;DESHPDHK0000KM8h;A:01;0;03;000;0100;00002;13;;32610877,517;5969777,653;Aegidienstraße;23552;Lübeck;;Innenstadt
N;DESHPDHK0000KM8H;A:01;0;03;000;0100;00002;36;;32611045,227;5969739,218;Aegidienstraße;23552;Lübeck;;Innenstadt
N;DESHPDHK0000KM8i;A:01;0;03;000;0100;00002;14;;32610944,613;5969748,085;Aegidienstraße;23552;Lübeck;;Innenstadt
N;DESHPDHK0000KM8I;A:01;0;03;000;0100;00002;37;;32610987,921;5969753,591;Aegidienstraße;23552;Lübeck;;Innenstadt
N;DESHPDHK0000KM8j;A:01;0;03;000;0100;00002;15;;32610894,084;5969782,635;Aegidienstraße;23552;Lübeck;;Innenstadt
N;DESHPDHK0000KM8J;A:01;0;03;000;0100;00002;38;;32611049,640;5969738,825;Aegidienstraße;23552;Lübeck;;Innenstadt
N;DESHPDHK0000KM8k;A:01;0;03;000;0100;00002;16;;32610953,409;5969746,707;Aegidienstraße;23552;Lübeck;;Innenstadt
N;DESHPDHK0000KM8K;A:01;0;03;000;0100;00002;39;;32610994,854;5969752,885;Aegidienstraße;23552;Lübeck;;Innenstadt
N;DESHPDHK0000KM8l;A:01;0;03;000;0100;00002;17;;32610916,569;5969780,545;Aegidienstraße;23552;Lübeck;;Innenstadt
N;DESHPDHK0000KM8L;A:01;0;03;000;0100;00002;4;;32610869,683;5969762,990;Aegidienstraße;23552;Lübeck;;Innenstadt
N;DESHPDHK0000KM8m;A:01;0;03;000;0100;00002;18;;32610960,673;5969745,420;Aegidienstraße;23552;Lübeck;;Innenstadt
N;DESHPDHK0000KM8M;A:01;0;03;000;0100;00002;40;;32611053,856;5969738,626;Aegidienstraße;23552;Lübeck;;Innenstadt
N;DESHPDHK0000KM8n;A:01;0;03;000;0100;00002;19;;32610927,552;5969769,495;Aegidienstraße;23552;Lübeck;;Innenstadt
N;DESHPDHK0000KM8N;A:01;0;03;000;0100;00002;41;;32611002,296;5969752,134;Aegidienstraße;23552;Lübeck;;Innenstadt
```

Abbildung 1: Screenshot HK-DE.

Die HK-DE enthält nur Hauptgebäude mit Hausnummern und besitzt 18 Attribute:

- Kennung des Datensatzes (NBA)
- Eindeutige Nummer des Datensatzes (OI)
- Qualität der georeferenzierten Gebäudeadresse (QUA)
- Schlüssel Land (LAN)
- Schlüssel Regierungsbezirk (RBZ)
- Schlüssel Kreis / kreisfreie Stadt (KRS)
- Schlüssel Gemeinde (GMD)
- Schlüssel Orts- bzw. Gemeindeteil (OTT)
- Schlüssel der Straße (SSS)

- Hausnummer (HNR)
- Adressierungszusatz (ADZ)
- Koordinatenwert ETRS89 / UTM mit Zonenkennung (Easting-Wert)
- Koordinatenwert ETRS89 / UTM (Northing-Wert) (beide Koordinatenwerte sind auf drei Nachkommastellen genau angegeben)
- (unverschlüsselter) Straßename (STN)
- Postleitzahl (PLZ)
- Postalischer Ortsname (ONM)
- Zusatz zum postalischen Ortsnamen (ZON)
- Postalischer Ortsteil (POT)

Es gibt drei Qualitäten der georeferenzierten Adresse:

- A = amtliche Hausnummern, deren Koordinaten sicher innerhalb der erfassten Gebäudegeometrie liegen.
- B = amtliche Hausnummern, deren Koordinaten sicher innerhalb der Flurstücksfläche liegen; ein Gebäude ist hier nicht sicher in der Örtlichkeit vorhanden.
- C = katasterinterne Hausnummern, deren Koordinaten sicher innerhalb der erfassten Gebäudegeometrie liegen.

Die Entschlüsselungsdatei dient zur Entschlüsselung des Gemeindegemeindekennzeichens (LAN/RBZ/KRS/GMD) sowie eventuell des Orts- bzw. Gemeindeteils (OTT). Die Einträge dieser Datei werden jeweils durch eine Kennung beschrieben: der Buchstabe L steht für Ländereinträge, R für Regierungsbezirkseinträge, K für Kreiseinträge, G für Gemeindeeinträge, O für Gemeindeteileinträge und durch die numerischen und langschriftlichen Bedeutungen der Schlüssel erläutert (z.B. L;05; Nordrhein-Westfalen).

Die Umschlüsselungsdatei wird einer Update-Lieferung beigelegt, wenn sich die OI eines Datensatzes zum Vorjahresbestand geändert hat. Die Umschlüsselungsdatei besteht aus aoi und noi mit aoi = bisheriger OI und noi = neuer OI.

Die HK-DE wird derzeit einmal jährlich aktualisiert. Die Daten der Länder werden zum 01.04 geliefert. Die Auslieferung an die Kunden erfolgt ca. ab dem 01.07.

Die Nomenklatur der drei Dateien ist: adressen-<nn>.txt, umschluessel-<nn>.txt und schluessel-de.txt mit nn: Länderkürzel nach GeoInfoDok, Vers. 6.0.1, Hauptdokument, Kapitel 3.3.10 (vgl. Zentrale Stelle Hauskoordinaten und Hausumringe, Landesamt für Digitalisierung, Breitband und Vermessung Bayern, 2019a).

In dieser Arbeit wurde der Datensatz von 2019 verwendet.

2.3.2 HU-DE

Die Informationen zu HU-DE sind der Datenformatbeschreibung Hausumringe Deutschland der ZSHH entnommen. Die HU-DE werden im AdV-Shape Format abgegeben. Es gibt Bildungsregeln für die Objekt-ID. Die Eigentümerangaben werden in einer zusätzlichen CSV-Datei dargestellt, die für diese Anwendung nicht von Belang ist und dem Datenschutz unterliegt. Die Objekte in HU-DE sind georeferenzierte Umring-Polygone, die die Gebäudegrundrisse des Liegenschaftskatasters darstellen. Nicht dargestellt werden Ausgestaltungsgeometrien, Dächer und unterirdische Gebäude. Es werden für die Ableitung aus dem Liegenschaftskataster der Länder verschiedene Objektgruppen, die der Gebäudeart entsprechen, herangezogen:

- AX_Gebaeude,
- AX_Turm,
- AX_BauwerkOderAnlageFuerIndustrieUndGewerbe,
- AX_VorratsbehaelterSpeicherbauwerk,
- AX_BauwerkOderAnlageFuerSportFreizeitUndErholung,
- AX_SonstigesBauwerkOderSonstigeEinrichtung und
- AX_HistorischesBauwerkOderHistorischeEinrichtung.

Eine genaue Auflistung der Gebäude- und Bauwerksdefinitionen findet sich unter <http://repository.gdi-de.org/schemas/adv/citygml/Codelisten/BuildingFunctionTypeAdV.xml>, z.B. Wohngebäude, Wohnhaus oder gemischtgenutztes Gebäude mit Wohnen. HU-Objekte besitzen lediglich zwei Attribute: den Amtlichen Gemeindeschlüssel (AGS) und den Objektidentifikator (OI). Die Koordinaten werden im Raumbezugssystem ETRS 89 / UTM Zone 32 oder Zone 33 Nord und damit in der Einheit Meter angegeben. Das Shape-Format und auch das AdV-Shape Format bestehen aus vier getrennten Dateien mit gleichem Dateinamen und unterschiedlichen Dateierweiterungen („shp“, „shx“, „dbf“, „prj“). Abgegeben werden die Hausumringe entweder nach Verwaltungseinheit oder nach geometrischer Abgrenzung (Rechteck, Polygon) selektiert. Der Datenbestand wird einmal pro Jahr aktualisiert (Länder liefern die Daten zum 01.04; Abgabe der Daten ca. ab dem 01.07) (vgl. Zentrale Stelle Hauskoordinaten und

Hausumringe, Landesamt für Digitalisierung, Breitband und Vermessung Bayern, 2019b).

Die HU-DE werden extra für den Vertrieb an Kunden erstellt und liegen daher im gängigen Geodaten austauschformat Shape vor. Vermessungsverwaltungsintern werden die Gebäudegrundrisse über ALKIS (Amtliches Liegenschaftsinformationssystem) abgerufen und genutzt.

In dieser Arbeit wurde der Datensatz von 2019 verwendet.

2.3.3 OSM

Nach OSM-Wiki können von OSM selbst die Rohdaten nur in den OSM-eigenen Datenformaten heruntergeladen werden. Die am meisten verwendeten OSM-Datenformate sind „osm.bz2“ und „osm.pbf“. „osm.bz2“ ist das bzip2 komprimierte ursprüngliche OSM-XML-Format. Es ist UTF-8 kodiert und die Positionen sind in geographischer Länge und Breite in WGS84 angegeben. „osm.pbf“ steht für das „Protocolbuffer Binary Format“, ist ein von OSM entwickeltes Binärformat und ersetzt in den meisten Fällen das ursprüngliche OSM-XML-Format, da es deutlich weniger Speicherplatz benötigt und schneller verarbeitet werden kann. Eine Datei besteht aus einem Dateikopf, gefolgt von einer Sequenz von Dateiblöcken. Dies ermöglicht einen besseren Zugriff auf die einzelnen Objekte und ein Überspringen von nicht benötigten oder nicht lesbaren Informationen (vgl. OSM-Community).

Das OSM-Datenmodell besteht aus Nodes, Ways und Relations, die Tags besitzen (vgl. Ramm & Topf, 2010, pp. 55–59):

- Tags sind Attribute, die die Objekte näher beschreiben. Sie sind Key-Value-Paare (Key: Schlüssel und Value: Wert).
- Nodes (Knoten, Punkte) enthalten die Positionen in geografischer Länge und Breite, den Zeitpunkt der letzten Änderung und optionalen Tags.
- Ways (Wege) sind geordnete Listen von mehreren Knoten mit optionalen Tags und dem letzten Änderungszeitpunkt. Ways werden meist für linienförmige und zum Teil auch flächenhafte Objekte verwendet. Voraussetzung für eine Verwendung als Fläche ist, dass der Way geschlossen ist. Meist werden Flächen durch das Attribut area gekennzeichnet.
- Relations (Beziehungen) sind ein Datentyp, mit dem Beziehungen zwischen verschiedenen Objekten modelliert werden. Sie sind sortierte Listen von beliebig vielen Knoten, Wegen und / oder anderen Relations. Diese besitzen eine Rollenangabe innerhalb der Relation. Relations können ebenfalls optional beliebig viele Tags besitzen. Sie stellen meist Be-

grenzungen aus mehreren Objekten, wie z.B. Landesgrenzen, Multipolygone oder Routen dar.

Nur Nodes besitzen Positionen, daher enthalten Ways und Relations Listen mit Knoten, wobei diese Knoten in der Regel keine eigenen Tags besitzen. Ways und Relations besitzen zusätzlich topologische Beziehungen, die es erlauben, Verbindungen oder Überschneidungen zwischen Objekten zu erfassen. Jedes Objekt besitzt eine ID, wobei nur die Kombination von Objekttyp und ID eindeutig ist (zusammengesetzter Schlüssel), und kann eine beliebige Anzahl an Tags enthalten (vgl. "Overpass API User's Manual").

Nach dem Overpass API User's Manual kann jedes OSM-Objekt für jeden Key maximal einen Value besitzen. Manchmal kommt es allerdings vor, dass bei einem Objekt mehrere Values für einen Key vergeben wurden. Hier ist die verwendete Notlösung für mehrere Values für einen Key, die Werte innerhalb des Value-Bereichs durch Semikolons zu trennen, wobei dies nicht gerne gesehen wird. Keys und Values liegen im Textformat vor und können maximal 255 Zeichen besitzen. Ansonsten gibt es keine weiteren Formatbeschränkungen. Das Datenmodell gibt keine speziellen Tags oder Keys vor. Die für den Anwendungsfall zweckmäßigsten Tags können spontan gewählt oder neu erstellt werden. Dies ist einer der Gründe für den großen Erfolg von OSM. Es gibt zwei inoffizielle Tagkategorien:

- klassifizierende Tags (z.B. der Key „highway“ für das öffentliche Straßennetz mit den Straßentypen als Values) und
- beschreibende Tags (z.B. die Keys „description“ und „name“).

Für klassifizierende Tags existieren nur wenige Keys und für diese Keys nur wenige Values. Abweichende Values werden als fehlerhaft wahrgenommen. Bei beschreibenden Tags wird bei den Values alles akzeptiert, einschließlich Groß- und Kleinbuchstaben, Zahlen und Sonderzeichen. Die am häufigsten genutzten und auch in dieser Arbeit verwendeten Quellen für Key- und Value-Semantik und -beschreibungen sind die Webseiten OSM-wiki (<https://wiki.openstreetmap.org/wiki>) (unter MapFeatures) und Taginfo (<https://taginfo.openstreetmap.org/>). (vgl. "Overpass API User's Manual").

3. VORGEHENSWEISE UND METHODIK

3.1 ABLAUF

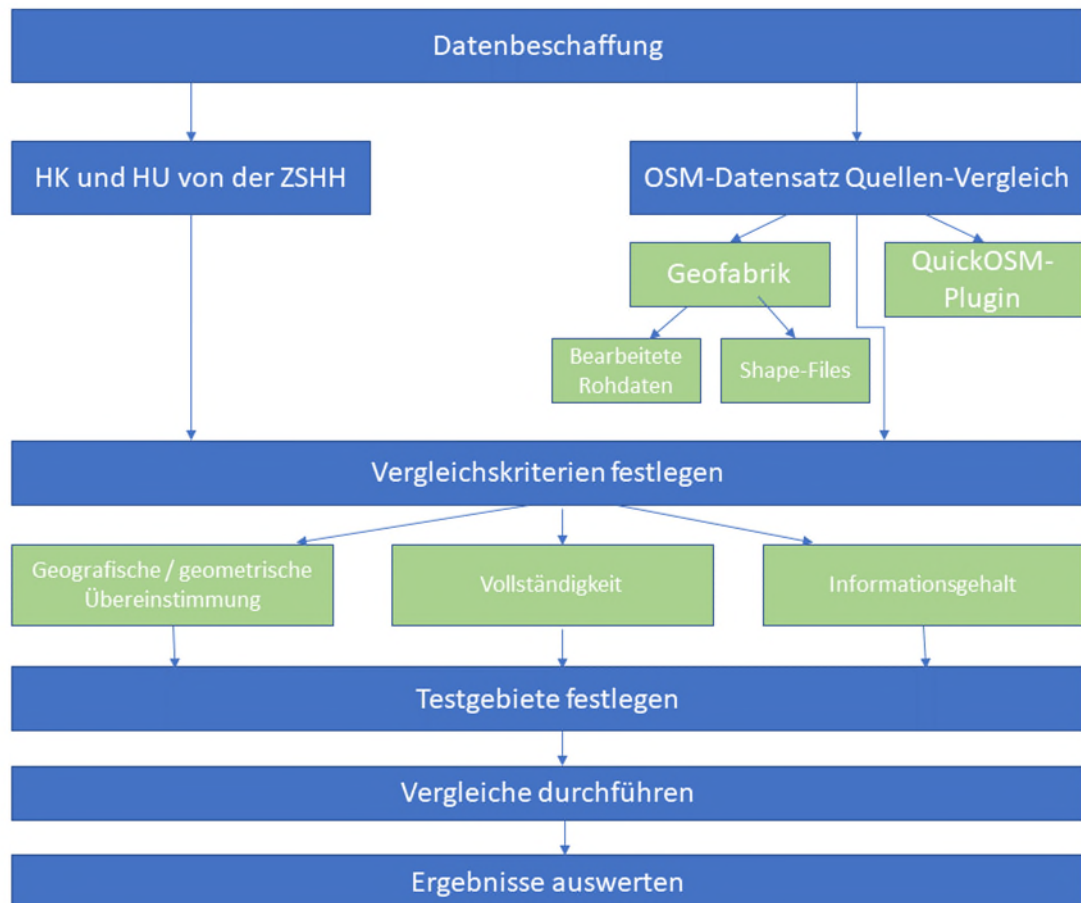


Abbildung 2: Ablaufdiagramm.

Die Abbildung 2 zeigt den groben Ablauf dieser Bachelorarbeit.

3.2 VERGLEICH DER OSM-DATENSÄTZE AUS VERSCHIEDENEN QUELLEN

In diesem Kapitel werden drei verschiedene Datensätze aus zwei Quellen miteinander verglichen: von Geofabrik (<http://download.geofabrik.de/>) bearbeitete Rohdaten und Shapefiles und gespiegelte Daten aus der OSM-Hauptdatenbank mit dem QuickOSM-Plugin von QGIS. Sowohl die von Geofabrik als auch die durch das QuickOSM-Plugin bezogenen Daten sind kostenlos und unterliegen der OSM-Lizenz. Die bekannteste Bezugsstelle ist Geofabrik. Die von Geofabrik bereitgestellten Daten enthalten allerdings nicht alle Tags. Die Geofabrik-Daten werden täglich aktualisiert und können nach Ländern und Regionen der Kontinente selektiert heruntergeladen werden. Um die Daten der gesamten Welt auf einmal herunterzuladen, muss die OSM-Website genutzt werden. Das QuickOSM-Plugin nutzt Overpass-API-Abfragen um OSM-Daten herunterzuladen. Die Overpass-API hält eine minütlich-aktualisierte Kopie der OSM-Hauptdatenbank für Abfragen bereit. Sie ist OpenSource und benötigt keine spezielle Hardware. Sie ist so designt, dass sie Abfragen von anderer Software über das Internet beantwortet (vgl. "Overpass API User's Manual"). Die Overpass API ist eine read-only API, die es ermöglicht, selektierte Daten aus der OSM-Hauptdatenbank herunterzuladen. Selektiert werden kann nach Orten, Keys, Tags, Objekttypen oder Kombinationen von diesen. Die Overpass API besitzt eine eigene Abfragesprache namens OverpassQL. Es kann nach Objekten, die zuletzt von einem bestimmten Nutzer bearbeitet wurden oder die nur neuer als ein bestimmtes Datum sind, selektiert werden (vgl. OSM-Community).

Die für den Anwendungsfall relevanten Layer der bearbeiteten Rohdaten von Geofabrik sind latest-points und latest-multipolygons. Die postalische Adresse findet sich in verschiedenen Key-Value-Paaren in `other_tags`. Sie ist allerdings kein Pflichtattribut. Nicht immer werden die adressbezogenen Tags auch in das Feld `other_tags` übernommen und eine Trennung und Sortierung der in `other_tags` gesammelten Tags ist nur mit viel Aufwand und guten Programmierkenntnissen möglich. Eine Beschreibung der Attribute (Tags) findet sich auf OSM-Wiki unter MapFeatures und auf der Webseite Taginfo (<https://taginfo.openstreetmap.org/>).

Die für den Anwendungsfall relevanten Geofabrik-Shapes sind:

- gis_osm_pois_free_1 (Points of Interest: z.B. öffentliche Gebäude, Restaurants oder Krankenhäuser),
- gis_osm_pofw_free_1 (Points of Worship: z.B. Moscheen und Kirchen) und
- gis_osm_buildings_a_free_1 (Gebäude).

Shapefiles die ein „a“ wie bei gis_osm_buildings_a_free_1 im Namen enthalten sind Polygonlayer. Eine postalische Adresse ist nicht ursprünglich in den Datensätzen enthalten, sie kann nur über umgekehrte Geokodierung erzeugt werden.

Da die durch das QuickOSM-Plugin bezogenen Daten für Gebäudelaye in Städten zum Teil über 150 verschiedenen Attribute und für Punktlayer sogar über 300 Attribute besitzen, wird im Folgenden für das QuickOSM-Plugin nur eine Attributauswahl betrachtet.

Tabelle 1: Attributlisten der OSM-Daten aus verschiedenen Quellen.

Quelle	Bearbeitete Rohdaten (Geofabrik)	Shape-Files (Geofabrik)	QuickOSM-Plugin für QGIS (Auswahl)
Punkt-Attribute	<ul style="list-style-type: none"> • osm_id • name • barrier • highway • ref • is_in • place • man_made • other_tags (enthält weitere Attribute z.B.): <ul style="list-style-type: none"> ○ addr: housenumber ○ addr:street ○ addr:city ○ addr:country ○ addr:postcode ○ addr: housename 	<ul style="list-style-type: none"> • osm_id • code • fclass • name 	<ul style="list-style-type: none"> • full_id • osm_id • addr: housenumber • addr:street • addr:city • addr:country • addr:postcode • postal_code • addr:housename • name • descripton • official_name • tourism • level • wheelchair • amenity • shop • ele • place • alt_name • access • leisure • entrance • toilets:wheelchair • school
Gebäude-Attribute	<ul style="list-style-type: none"> • osm_id • osm_way_id • name • type • aeroway • amenity • admin_level • barrier • boundary • building • craft • geological 	<ul style="list-style-type: none"> • osm_id • code • fclass • name 	<ul style="list-style-type: none"> • full_id • osm_id • addr: housenumber • addr:street • addr:city • addr:country • addr:postcode • addr:housename • building • name • official_name

Quelle	Bearbeitete Rohdaten (Geofabrik)	Shape-Files (Geofabrik)	QuickOSM-Plugin für QGIS (Auswahl)
	<ul style="list-style-type: none"> • historic • land_area • land_use • leisure • man_made • military • natural • office • place • shop • sport • tourism • other_tags (enthält weitere Attribute z.B.): <ul style="list-style-type: none"> ○ addr:housenumber ○ addr:street ○ addr:city ○ addr:country ○ addr:postcode 		<ul style="list-style-type: none"> • tourism • level • wheelchair • amenity • shop • ele • place • alt_name • access • leisure • toilets:wheelchair • school

Bei den Geofabrik Shape-Files enthält das Attribut *osm_id* die aus der OSM-Hauptdatenbank stammende ID. Das Attribut *code* definiert die von Geofabrik vergebene Featureclass als vierstellige Zahl. Eine Liste der verschiedenen Codes findet sich in der Formatbeschreibung von Geofabrik („OpenStreetMap Data in Layerd GIS Format“ von Frederik Ramm), die auf der Geofabrik-Webseite zu finden ist. Das Attribut *fclass* enthält die unverschlüsselten Informationen des Attributs *code* und das Attribut *name* entspricht dem OSM-Tag *name* (wenn im OSM-Tag offensichtlich fehlerhafte Daten enthalten sind, bleibt dieses Attribut leer) (vgl. Ramm, 2019).

Die Beschreibungen bzw. Erklärungen der folgenden Attribute sind der Website TagInfo oder OSM-Wiki entnommen oder lassen sich durch den Namen herleiten. Das Attribut *osm_id* ist nur für den Objekttyp, also Nodes, Ways und Relations, eindeutig, daher wurde für die Overpass-API das Attribut *full_id*, das aus dem ersten Buchstaben des Objekttyps und der *osm_id* besteht. Das Attribut

name ist der Name eines Objekts. Das Attribut *official_name* enthält den offiziellen Namen eines Objekts. Das Attribut *tourism* wird für Plätze und Einrichtungen, die hauptsächlich von Touristen genutzt werden, z.B. Hotels, Touristeninformationen etc. verwendet. Das Attribut *level* gibt die Anzahl der Stockwerke eines Gebäudes wieder. Das Attribut *wheelchair* zeigt an, ob ein Ort mit einem Rollstuhl benutzt werden kann. Das Attribut *amenity* wird zur Markierung nützlicher und wichtiger Einrichtungen für Besucher und Anwohner z.B. Cafés, Restaurants, Schulen etc. verwendet. Das Attribut *shop* enthält die Art oder den Namen eines Geschäfts. Das Attribut *elev* entspricht der Höhe eines Punktes in Metern über dem Meeresspiegel. Das Attribut *place* wird bei bewohnten und unbewohnten Orten, welche einen Namen tragen, verwendet. Das Attribut *building* wird benutzt, um ein Gebäude zu kennzeichnen. Das Attribut *alt_name* wird verwendet, falls es einen alternativen Namen für ein Objekt gibt. Das Attribut *access* enthält die rechtlichen Zugangsregelungen eines Objekts z.B. Privat. Das Attribut *leisure* kennzeichnet Orte, an denen Menschen ihre Freizeit verbringen. Das Attribut *toilets:wheelchair* zeigt die Existenz einer Behindertentoilette an. Das Attribut *school* wird für Schulen verwendet. Das Attribut *addr:housenumber* enthält die Hausnummer eines Gebäudes. Das Attribut *addr:street* enthält den Straßennamen. Das Attribut *addr:city* enthält den Ortsnamen. Das Attribut *addr:country* enthält den Landesnamen. Das Attribut *addr:postcode* enthält die Postleitzahl. Das Attribut *addr:housename* wird verwendet falls ein Gebäude einen Eigennamen besitzt, z.B. Allianz-Arena. Das Attribut *description* enthält eine kurze verständliche Beschreibung. Das Attribut *entrance* wird verwendet, wenn ein Punkt die Position eines Ein- oder Ausgangs markiert. Das Attribut *type* gibt an ob es sich um einen Node, Way oder eine Relation handelt. Das Attribut *aeroway* markiert Einrichtungen und Anlagen der zivilen Luftfahrt. Das Attribut *admin_level* bezieht sich auf die Ebene einer Grenze, z.B. *admin_level* = 2 für Staatsgrenzen. Das Attribut *barrier* kennzeichnet alle Arten von Barrieren, die die Bewegungsfreiheit von Fahrzeugen, Menschen oder Tieren einschränken. Das Attribut *boundary* wird für Grenzen verwendet. Das Attribut *craft* markiert alle Formen von Handwerksbetrieben. Das Attribut *geological* beschreibt geologische Objekte wie z.B. Moränen. Das Attribut *historic* kennzeichnet historische Plätze und Gebäude. Das Attribut *land_area* wird benutzt, um Grenzen von

Landmassen darzustellen. Das Attribut *land_use* gibt die Flächennutzungsart an. Das Attribut *man_made* wird verwendet, wenn es sich um künstlich erschaffene Strukturen handelt. Das Attribut *military* wird für Gebäude und Land von Militäreinrichtungen verwendet. Das Attribut *natural* wird zur Beschreibung von Landschaften benutzt. Das Attribut *office* gibt an, um was für einen Bürotyp es sich handelt z.B. ein Unternehmen oder eine Regierungseinrichtung. Das Attribut *sport* beschreibt was für Sportarten auf bzw. in einem Objekt betrieben werden können. Das Attribut *highway* wird für die verschiedenen Straßentypen verwendet z.B. für Autobahnen oder für Straßen in Wohngebieten. Das Attribut *ref* enthält Referenznummern bzw. -codes und wird hauptsächlich bei Straßen, Ausfahrten etc. verwendet. Das Attribut *is_in* zeigt an, ob sich ein Objekt in einem anderen Objekt befindet.

Wie Tabelle 1 zu entnehmen ist, enthalten die durch das QuickOSM-Plugin bezogenen Daten die meisten adressbezogenen Attribute.

Tabelle 2: Vorprozessierungsschritte für die verschiedenen OSM-Quellen.

Quelle	Vorprozessierungsschritte
Bearbeitete Rohdaten (Geofabrik)	<ol style="list-style-type: none"> 1. Die Daten von der Geofabrik-Webseite herunterladen. 2. Die Daten in QGIS (eine 3.x Version von QGIS nötig) laden. 3. Die Funktion „Geometrien reparieren“ anwenden, da es sonst bei der weiteren Bearbeitung der Layer Probleme geben kann. Dadurch ändert sich die Objektanzahl nicht. 4. Eine Umwandlung der Daten in Shape-Files oder ein anderes gängiges Geodatenformat. 5. Falls nötig die Layer reprojizieren, um das gleiche Koordinatenbezugssystem wie die anderen Daten zuzuweisen. 6. Um einen Gebäudelayer zu erzeugen, werden die Daten im Polygonlayer nach dem buildings-Attribut ausgewählt und die Auswahl als neuer Layer gespeichert.
Shape-Files (Geofabrik)	<ol style="list-style-type: none"> 1. Die Daten von der Geofabrik-Webseite herunterladen. 2. Die Daten in QGIS laden. 3. Falls nötig, die Layer reprojizieren, um das gleiche Koordinatenbezugssystem wie die anderen Daten zuzuweisen.
QuickOSM-Plugin für QGIS (Auswahl)	<ol style="list-style-type: none"> 1. Das QuickOSM-Plugin in QGIS unter dem Reiter Erweiterungen installieren. 2. Das QuickOSM-Plugin öffnen. 3. Die benötigten Daten über die Layerausdehnung der ZSHH-Daten und beim Gebäudelayer nach key = building suchen. Anschließend die auszugebenden Kategorien wählen (z.B. Node, Way, Points etc.), die Ausgabedatei angeben und Query laufen lassen. Die Ausgabedatei ist immer im GeoJson-Format. Falls keine Ausgabedatei angegeben wird, wird ein temporärer Layer geladen, der beim Speichern Probleme machen kann. 4. In der Attributtabelle überflüssige Attributspalten lö-

	<p>schen. Die Attributwerte der gelöschten Spalten werden zunächst noch in den vorhandenen Spalten dargestellt, erst wenn der Layer gelöscht und neu eingeladen wird, wird in QGIS alles richtig dargestellt.</p> <ol style="list-style-type: none"> 5. Falls nötig, den Layer reprojizieren, um das gleiche Koordinatenbezugssystem wie die anderen Daten zuzuweisen. 6. Die Funktion „Geometrien reparieren“ anwenden, da sonst die Analysefunktionen nicht funktionieren.
--	--

Wie der Tabelle 2 zu entnehmen ist, unterscheidet sich der Vorbereitungsaufwand für die verschiedenen Quellen nicht sehr stark. Daher spielen für die Entscheidung für eine Quelle hauptsächlich die Attribute eine Rolle.

Insgesamt für einen adressbezogenen Vergleich mit den ZSHH-Daten am besten geeignet sind die über das Quick-OSM-Plugin bezogenen Daten. Die von hieran verwendeten OSM-Daten wurden am 11.03.2020 mit dem Quick-OSM-Plugin heruntergeladen.

3.3 VERGLEICHSKRITERIEN

Wie in Tabelle 3 zu erkennen ist, werden die Vergleichskriterien grob in drei Kategorien unterteilt: *geografische / geometrische Übereinstimmung*, *Vollständigkeit* und *Informationsgehalt*. Zudem enthält die Tabelle 3 die wichtigsten Fragen, die in den Kategorien behandelt werden.

Tabelle 3: Übersicht der Vergleichskriterien.

Kategorien Vergleichskriterien		
Geografische / Geometrische Übereinstimmung	Vollständigkeit	Informationsgehalt
<ul style="list-style-type: none">• Wie viele OSM-Punkte fallen in Gebäude?• Wie gut passen die Hausumringe beider Datenquellen zusammen?	<ul style="list-style-type: none">• Wie viele Gebäude beider Datenquellen gibt es in den Testgebieten?• Wie viele Gebäude gibt es in beiden Datensätzen, die kein Gegenstück im anderen Datensatz besitzen?	<ul style="list-style-type: none">• Welche Attribute gibt es in den verschiedenen Datensätzen und wie oft sind diese ausgefüllt?• Stimmt die postalische Adresse der Datensätze überein?

3.3.1 GEOGRAFISCHE / GEOMETRISCHE ÜBEREINSTIMMUNG

Die Kategorie *geografische / geometrische Übereinstimmung* behandelt, wie gut die Daten der beiden Quellen hinsichtlich ihrer Lage und Form zusammenpassen. Bei den HK ist durch das Attribut Qualität der georeferenzierten Adresse erkennbar ob die Koordinaten in einem Gebäude liegen. Da viele der OSM-Punkte, die in Gebäuden liegen, kein Attribut *building* besitzen, ist eine Abfrage nötig, um festzustellen welche OSM-Punkte in Gebäude fallen. Die zweite große Fragestellung dieser Kategorie ist, wie gut die Hausumringe der beiden Datenquellen zusammenpassen. Hierfür muss zunächst für die Gebäude das Gegenstück im anderen Datensatz gefunden werden. Anschließend werden die Flächen in Quadratmetern und die Anzahl der Gebäudepunkte dieser Gebäude verglichen. Des Weiteren werden Überdeckung, also die Schnittmenge, und der Jaccard-Koeffizient berechnet.

Der Jaccard-Koeffizient wurde 1907 von Paul Jaccard als Maß für die Verteilung verschiedener Pflanzen in den Alpen entwickelt, allerdings noch ohne den Namen Jaccard-Koeffizient (vgl. Jaccard, 1912). Der Jaccard-Koeffizient ist aufgrund seiner Berechnungsart auch unter dem Namen Intersection over Union (IoU) bekannt. Die Formel für den Jaccard-Koeffizient lautet $IoU = \frac{|A \cap B|}{|A \cup B|}$. Das heißt also, die Schnittmenge wird durch die Vereinigungsmenge geteilt. Die IoU-Werte liegen zwischen 0 und 1. Je näher der Wert an 1 ist, desto ähnlicher sind sich die verglichenen Mengen bzw. Objekte. Der Jaccard-Koeffizient ist die beliebteste Bewertungsmetrik bei der Objekterkennung (vgl. Rezatofighi et al.). Um mögliche Unterschiede zwischen den einzelnen Testgebiete besser erkennen zu können, wird auch ein Jaccard-Koeffizient für jedes Testgebiet berechnet.

Die Ergebnisse werden mithilfe der Absoluten Häufigkeit, der Relativen Häufigkeit und der Cumulative distribution function (CDF) bzw. der empirischen Verteilungsfunktion dargestellt. Die Erklärungen der Absoluten Häufigkeit, der Relativen Häufigkeit und der CDF basieren auf dem Buch „Grundlagen der Wahrscheinlichkeitsrechnung und Statistik“ von Cramer und Kamps von 2008.

Die Anzahl aller Beobachtungswert einer Stichprobe, also aller für die Auswertung eines Merkmals X erstellten Beobachtungen, heißt Stichprobenumfang n . Die m verschiedenen Ausprägungen des Merkmals X werden als u_1, \dots, u_m bezeichnet. Die Absolute Häufigkeit entspricht der Häufigkeit im normalen Sprachgebrauch und gibt die Anzahl aller Beobachtungswerte n_j an, die mit einer bestimmten Merkmalsausprägung u_j übereinstimmen. Die Summe aller Absoluten Häufigkeiten eines Merkmals ist immer der Stichprobenumfang n . Die Relative Häufigkeit f_j ist der Quotient der Absoluten Häufigkeit n_j und des Stichprobenumfangs n . Die Formel für die Relative Häufigkeit lautet $f_j = \frac{n_j}{n}$. Um die Relative Häufigkeit in Prozent zu erhalten, wird die Relative Häufigkeit mit 100 multipliziert:

$$\text{relative Häufigkeit in \%} = \frac{\text{Absolute Häufigkeit}}{\text{Anzahl aller Beobachtungen}} \cdot 100\% .$$

Die Summe aller Relativen Häufigkeiten eines Merkmals ist immer 1 (eins). Tabellarische Zusammenstellungen von Absoluten oder Relativen Häufigkeiten werden als Häufigkeitstabellen bezeichnet. Die CDF F_n ist für diskrete Werte eine Treppenstufenfunktion, d.h. sie besitzt mehrere Sprungstellen. Die Sprungstellen liegen an den Stellen u_1, \dots, u_m . Die Höhe des Sprungs bzw. der Treppenstufe an der Stelle u_j ist gleich der relativen Häufigkeit f_j dieser Merkmalsausprägung. F_n liegt immer zwischen 0 und 1. Die vereinfachte Formel der CDF lautet $F_n = \sum_{j=1}^k f_{(j)}$ das bedeutet $F_1 = f_{(1)}, F_2 = f_{(1)} + f_{(2)}, F_3 = f_{(1)} + f_{(2)} + f_{(3)}$, $F_n = f_{(1)} + f_{(2)} + \dots + f_n$ Die Summe aller relativen Häufigkeiten, also F_n an der Stelle u_m , ist immer 1 (vgl. Cramer & Kamps, 2008, pp. 16–19).

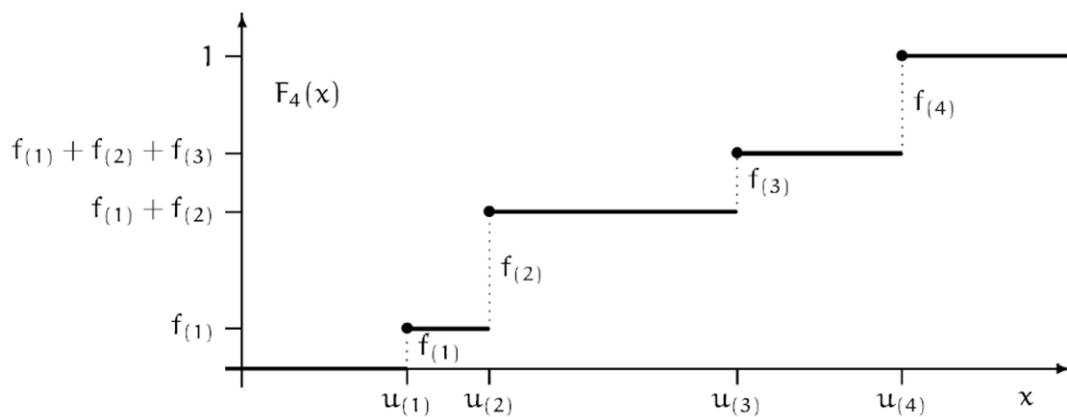


Abbildung 3: Beispiel einer CDF
Quelle: (vgl. Cramer & Kamps, 2008, p. 18).

Abbildung 3 zeigt ein Beispiel einer CDF für vier relative Häufigkeiten.

3.3.2 VOLLSTÄNDIGKEIT

In der Kategorie *Vollständigkeit* wird behandelt, wie viele Gebäude es in beiden Datenquellen für die Testgebiete gibt und wie viele Gebäude kein Gegenstück im anderen Datensatz besitzen. Diese Kategorie ist wichtig, da die Kunden der ZSHH darauf vertrauen, dass sie alle im Liegenschaftskataster enthaltenen Gebäude erhalten und nicht zusätzlich die Daten der für sie relevanten Gebiete verdichten müssen. Da die Kunden der ZSHH in der Regel Abnehmer ganzer Bundesländer oder bundesländerübergreifender Gebiete sind, wäre dies ein sehr großer Aufwand und würde zu erheblichen Mehrkosten führen.

3.3.3 INFORMATIONSGEHALT

In der Kategorie *Informationsgehalt* geht es darum, ob die postalische Adresse der Gebäude der Datensätze übereinstimmt und welche Attribute der Datensätze wie oft ausgefüllt wurden. Da es bei den OSM-Daten so viele Attribute gibt, werden nur die adressbezogenen Attribute sowie das Attribut *name* als ein häufig ausgefülltes Attribut und das Attribut *shop* als ein selten ausgefülltes Attribut behandelt. Bei den amtlichen Daten ist die Anzahl der Attribute vorgegeben. Die HU besitzen 2 Attribute und die HK besitzen 18 Attribute. Sie enthalten allerdings ausschließlich adressbezogene Attribute und IDs und keine zusätzlichen Informationen.

3.4 AUSWAHL UND BEGRÜNDUNG DER TESTGEBIETE

Es werden Bundesländer mit OpenData-Initiative und Bundesländer ohne OpenData-Initiative verglichen.

Tabelle 4: Bundesländer mit und ohne OpenData-Initiativen.

Bundesländer <u>mit</u> OpenData-Initiative	Bundesländer <u>ohne</u> OpenData-Initiative
Berlin	Baden-Württemberg
Hamburg	Bayern
Nordrhein-Westfalen	Bremen
Sachsen	Hessen
Thüringen	Mecklenburg-Vorpommern
seit Januar 2020: Brandenburg	Niedersachsen
	Rheinland-Pfalz
	Saarland
	Sachsen-Anhalt
	Schleswig-Holstein

Zudem gibt es Testgebiete in *gebirgigen Gebieten; Großstädte und Ballungsräume; mittelgroße Städte* und *ländliche Gebieten*.

Eine eigene Kategorie für *gebirgige Gebiete* wurde gewählt, da die meisten OSM-Daten mithilfe von GNSS aufgenommen werden und in Gebirgen die Wahrscheinlichkeit von Abschattungen höher ist. Eine Abschattung ist eine Situation, in der die GNSS-Signale nicht auf die Empfängerantenne treffen können, da sich zwischen dem Satelliten und dem Empfänger ein Hindernis befindet. Aufgrund des Aufbaus der Satellitennetze stehen die GNSS-Satelliten in Deutschland meist recht niedrig, weshalb hier das Risiko von Abschattungen höher ist als in den Teilen der Welt, die näher am Äquator liegen.

In Thüringen werden die Hausumringe aus Luftbild-Befliegungen und nicht durch terrestrische Vermessungen erstellt, daher ist die Wahrscheinlichkeit hoch, dass die Genauigkeit der Hausumringe in Thüringen im Vergleich zu den HU der anderen Bundesländer etwas geringer ist.

Es werden 16 Testgebiete festgelegt: vier Testgebiete für jede Kategorie und möglichst ein Testgebiet pro Bundesland, wobei sich das aufgrund einer möglichst gleichmäßigen Verteilung der Testgebiete einer Kategorie über Deutschland nicht ganz umsetzen lässt.

Als Größe für die Testgebiete werden Bounding-Boxen um Postleitzahlgebiete festgelegt, da Bounding-Boxen um Postleitzahlgebiete gut selektierbar sind und sich in einem Postleitzahlgebiet nicht zu viele Gebäude befinden, als dass die Datenmengen nicht mehr handelbar wären. Für die Auswahl der Testgebiete spielt auch die Form der Postleitzahlgebiete eine Rolle. Da die Bounding-Boxen um die Postleitzahlgebiete verwendet werden, sollten die Postleitzahlgebiete möglichst rechteckig sein, um in den Testgebieten nicht zu viele Gebäude aus anderen Postleitzahlgebieten zu enthalten.

Die Testgebiete in *gebirgigen Gebieten* sind:

- 87541 Bad Hindelang (Bayern, Alpen)
- 38877 Benneckenstein (Harz) (Sachsen-Anhalt, Harz)
- 08309 Eibenstock (Sachsen, Erzgebirge)
- 54608 Bleialf (Rheinland-Pfalz, Eifel)

Die Testgebiete in *Großstädten und Ballungsräumen* sind:

- 45889 Gelsenkirchen (Bismarck) (Nordrhein-Westfalen)
- 22305 Hamburg (Brambek) (Hamburg)
- 10243 Berlin (Friedrichshain) (Berlin)
- 60311 Frankfurt am Main (Altstadt) (Hessen)

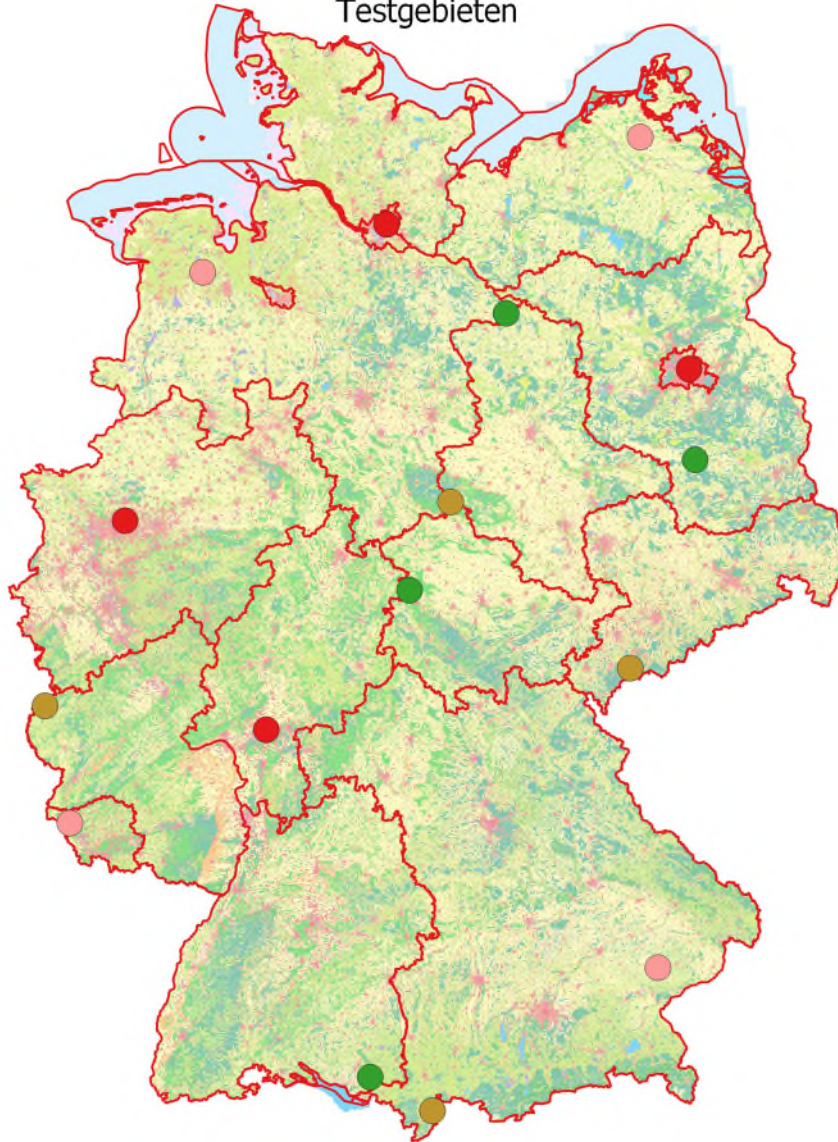
Die Testgebiete in *mittelgroßen Städten* sind:

- 18507 Grimmen (Mecklenburg-Vorpommern, ca. 10.000 Einwohner)
- 66663 Merzig (Saarland, ca. 30.000 Einwohner)
- 26655 Westerstede (Niedersachsen, ca. 20.000 Einwohner)
- 84307 Eggenfelden (Bayern, ca. 15.000 Einwohner)

Die Testgebiete in *ländlichen Gebieten* sind:

- 88279 Amtzell (Baden-Württemberg, nordöstlich des Bodensees)
- 15936 Dahme / Mark (Brandenburg, Spreewald)
- 99831 Creuzburg (Thüringen, nördlich von Eisenach)
- 29491 Prezelle (Niedersachsen, Wendland)

Deutschlandkarte mit den Testgebieten



Legende

0 50 100 150 200 km



Testgebiete

- gebirgisches Gebiet
- Großstadt oder Ballungsraum
- ländliches Gebiet
- mittelgroße Stadt
- Verwaltungsgrenzen250

Corine Landcover 5ha 2018

Die Quelle für die Verwaltungsgrenzen250-Shape ist <https://gdz.bkg.bund.de/index.php/default/digitale-geodaten/verwaltungsgebiete/verwaltungsgebiete-1-250-000-ebenen-stand-01-01-vg250-ebenen-01-01.html>. Die Quelle für den Corine-Landcover-5ha-2018-WMS ist https://sgx.geodatenzentrum.de/wms_clc5_2018.

Abbildung 4: Karte der Testgebiete.

3.5 FESTLEGUNG VON TOLERANZKLASSEN

Die Toleranzklassen sind: bei einem Jaccard-Koeffizienten zwischen 0 und 0,50 handelt es sich nicht um das gleiche Gebäude, bei einem Jaccard-Koeffizienten zwischen 0,50 und 0,70 handelt es sich möglicherweise um das gleiche Gebäude und bei einem Jaccard-Koeffizienten zwischen 0,70 und 1 handelt es sich tatsächlich um das gleiche Gebäude.

Um die Verteilung der Jaccard-Koeffizienten besser interpretieren zu können, werden zudem verschiedene Intervalle bzw. Klassen betrachtet (siehe Tabelle 5).

Tabelle 5: Zuordnung der Jaccard-Koeffizienten-Intervalle zu den Toleranzklassen.

Toleranzklasse	Intervalle Jaccard-Koeffizienten
nicht das gleiche Gebäude	0,00
	>0,00 – <0,50
möglicherweise das gleiche Gebäude	0,50 - <0,70
das gleiche Gebäude	0,70 - <0,80
	0,80 - <0,90
	0,90 - <1,00
	1,00

3.6 UMSETZUNG MIT QGIS UND TABLEAU

Als Vergleichssoftware habe ich mich für QGIS 3.4.14 entschieden, da QGIS derzeit eine der besten kostenlosen und freien GIS-Softwares ist. Die Version 3.4.14 ist die aktuelle Langzeit-Version von QGIS. QGIS ist einfach anzuwenden, kann über selbstprogrammierte oder heruntergeladene Plugins für die persönlichen Zwecke verbessert werden, verfügt über eine ausführliche Dokumentation und viele Diskussionsforen. Bei vielen wissenschaftlich Projekte der letzten Jahre im GIS-Bereich wurde daher QGIS verwendet. In QGIS können Projekte erstellt werden, in die mehrere Datensätze (auch aus verschiedenen Quellen) eingelesen und in Layer umgewandelt werden. Die im Projekt enthaltenen Layer können miteinander verglichen, verschnitten, kombiniert etc. werden. Außerdem sind Analysen der Datenattribute und Metadaten möglich. In QGIS können ab Version 3.x auch OSM-Rohdaten ohne extra add-ins eingeladen werden.

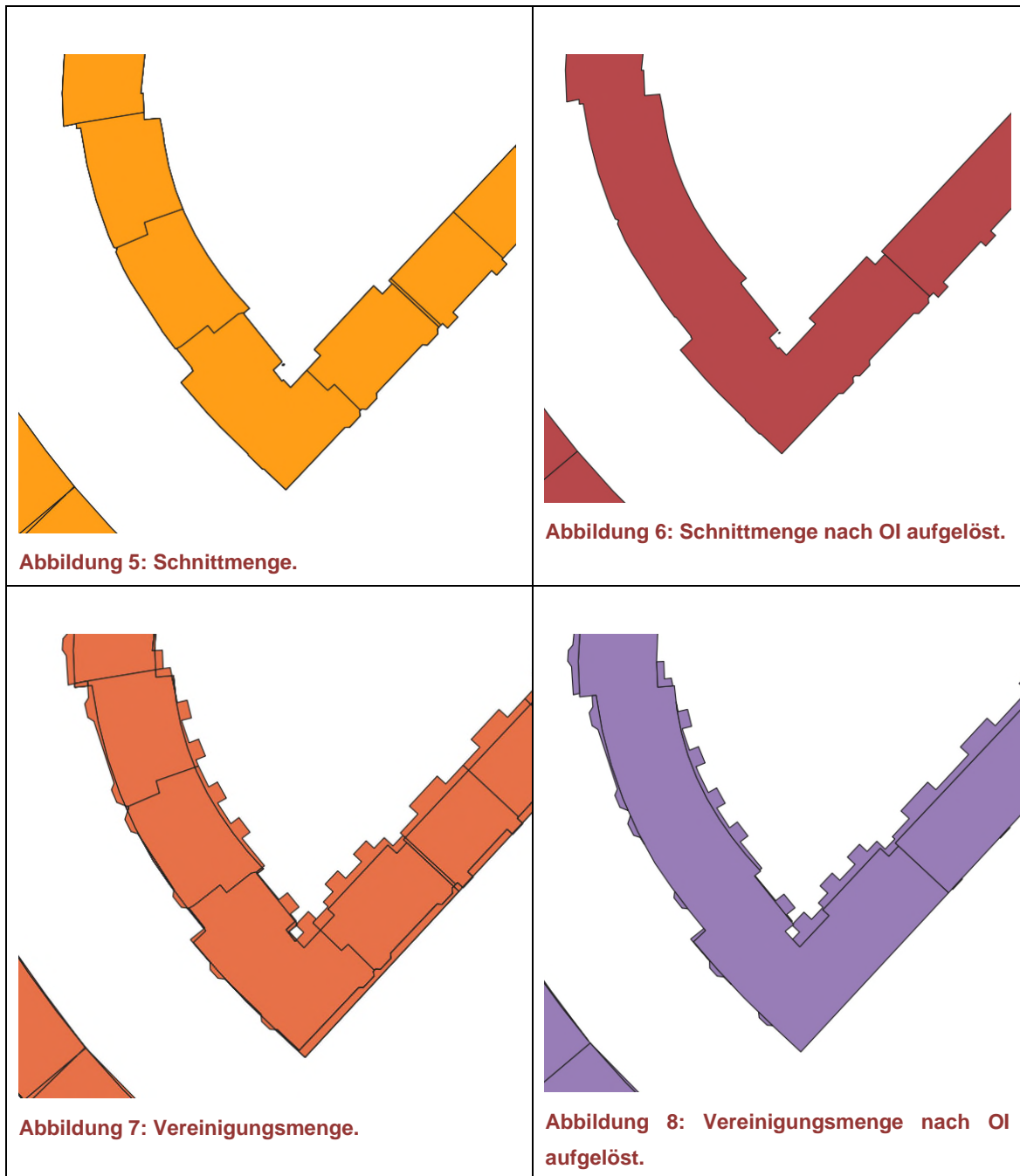
Als zweite Vergleichssoftware habe ich mich für Tableau entschieden, da Tableau einfach anzuwenden ist, in kurzer Zeit sehr große Datenmengen verarbeitet werden können und ein Teil der Analysen in QGIS nur sehr viel extra Aufwand umsetzbar gewesen wären in Tableau leicht umsetzbar sind. Tableau ist ein Programm, das es ermöglicht, aus Daten aus einer oder mehrerer Quellen Grafiken und Tabellen zu erstellen. Hierzu können auch datensatzübergreifende Berechnungen, Analysen und Hierarchien erzeugt werden und vielfältige Filter eingesetzt werden. Zudem werden auch große Datenmengen deutlich schneller als in QGIS verarbeitet.

In Tableau können Shape-Files eingelesen werden. Allerdings können maximal 32 Tabellen bzw. Datensätze gleichzeitig eingelesen werden, daher müssen die verschiedenen Shapes der Testgebiete am Ende zu deutschlandweiten Dateien zusammengefügt werden.

Im Angang findet sich der genaue Ablauf mit Programmschritten und Screenshots.

Zunächst muss in QGIS das Projektkoordinatensystem festlegen. In diesem Fall EPSG 4647 [UTM / ETRS89 Zone 32] für eine deutschlandweite Vergleichbarkeit und da die HK im Gegensatz zu den HU nur in Zone 32 und nicht auch in Zone 33 vorliegen. Danach werden die bundesländerweiten HU-Layer und HK-Layer importiert und je eine Bounding-Box um die Postleitzahlengebiete erzeugt. Die Bounding-Box ist das kleinste umschließende Rechteck einer Objektauswahl. Der nächste Schritt ist die HU- und HK-Layer mit den Bounding-Boxen zu den Testgebieten zuzuschneiden. Nun werden mit dem QuickOSM-Plugin für die einzelnen Testgebiete ein OSM-Gebäude- und ein OSM-Punkte-Layer erzeugt. Diese Layer müssen vor der weiteren Verwendung repariert werden, da sonst einige QGIS-Funktionen Fehlermeldungen erzeugen. Das Reparieren eines Layers beeinflusst dessen Objektanzahl nicht. Da die mit dem QuickOSM-Plugin erzeugten Layer auch einige Objekte außerhalb der Bounding-Boxen enthalten, müssen die OSM-Layer mit den Bounding-Boxen zugeschnitten werden, um eine Vergleichbarkeit mit den ZSHH-Layern herzustellen. Die mit dem OSM-Plugin erstellten Layer enthalten sämtliche OSM-Attribute. Das können je nach Region mehrere Hundert sein, daher ist es notwendig die Attributlisten auf die für den Vergleich mit den ZSHH-Daten nötigen Attribute einzuschränken und die restlichen Attribute zu löschen, um die Rechenzeiten und den Speicherplatz zu reduzieren. Die OSM-Daten liegen als Längen und Breiten vor und müssen vor der weiteren Verwendung nach UTM reprojeziert werden. Um die Layer der einzelnen Testgebiete nach dem zusammenfügen auseinanderhalten zu können, wird ein neues Attribut „*testgebiet*“ für alle Layer eingeführt und mit Hilfe des Feldrechners von QGIS erzeugt. Es enthält die Postleitzahl und den Ortsnamen. Im OSM-Punktlayer sind alle OSM-Punkte im Testgebiet enthalten und nicht nur die in den Gebäuden. Um einen OSM-Punkte-in-Gebäuden-Layer zu erzeugen, ist der nächste Schritt daher den OSM-Punktlayer mit dem OSM-Gebäudelayer zu verschneiden. Für den Vergleich der postalischen Adressen der Gebäude ist es nötig eine Verknüpfung zwischen denselben Gebäuden der beiden Quellen herstellen zu können. Dies ist nur bei den Gebäudelayern gut möglich, daher werden über die QGIS-Funktion „Attribute nach Position zusammenfügen“ die HK-Adressen den HU-Layern hinzugefügt. Für den Vergleich der geometrischen Übereinstimmung der Gebäude werden mit dem Feldrechner die Attribute „*Flä-*

che“ für die Grundrissfläche in Quadratmetern und „*GebPunkte*“ für die Anzahl der Gebäudepunkte für die HU- und OSM-Gebäudelayers erzeugt. Über die QGIS-Funktionen „Verschneidung“ und „Vereinigungen“ werden Schnittmengen und Vereinigungsmengen zwischen den HU- und den OSM-Gebäudelayers erzeugt. Um die Rechenzeiten und den Speicherplatz zu reduzieren, werden in den Attributlisten alle Attribute gelöscht, die sich nicht auf die Objektidentifikation oder das Testgebiet beziehen. Die Beziehung zu den anderen Attributen kann später in Tableau mit den Objektidentifikatoren wiederhergestellt werden.



Wie in Abbildung 5 und Abbildung 6 zu erkennen ist, entstehen beim Erzeugen von Schnittmengen und Vereinigungsmengen in QGIS zum Teil kleine Restpolygone. Um diesem Problem entgegenzuwirken, werden die Objekte der Schnittmengen- und Vereinigungsmengenlayer nach dem Attribut „OI“, also der Objekt-ID der HU-Gebäude, aufgelöst. Auflösen ist in QGIS die Funktion zum Verschmelzen von Objekten desselben Layers. Wie in Abbildung 7 und Abbildung 8 zu erkennen ist, verschwinden durch das Auflösen leider nicht alle

Restpolygone. Ein weiteres auflösen nach dem Attribut „*full_id*“, also der Objekt-ID der OSM-Gebäude, führt allerdings dazu, dass auch weit auseinanderliegende Objekte miteinander verschmolzen werden. Daher werden die Restpolygone für den weiteren Vergleich in Kauf genommen und bei der Interpretation der Ergebnisse mitbehandelt.

Um die Jaccard-Koeffizienten berechnen zu können, werden die Flächen der Schnittmengen und Vereinigungsmengen der Gebäude in Quadratmetern benötigt. Diese werden mit dem Feldrechner als Attribut „*Fläche*“ für die aufgelösten Schnittmengen- und Vereinigungsmengenlayer erzeugt. Nach dem die Vorbereitungen für die einzelnen Testgebiete hiermit abgeschlossen sind, werden die Layer der Testgebiete nach Kategorien, also alle Schnittmengenlayer, alle HU-Layer, etc., zu deutschlandweiten Layern zusammengefügt. Dabei kommt es bei den OSM-Layer vor, dass es mehrere Attribute mit derselben Bedeutung aber unterschiedlichen Namen gibt. Diese Attribute müssen mithilfe des Feldrechners zu einem gemeinsamen Attribut zusammengefasst werden. Für eine Verknüpfung der Daten in Tableau ist es notwendig das HU-Attribut „*OI*“ für den OSM-Gebäudelayer aller Testgebiete zu übernehmen, dies geschieht mithilfe der QGIS-Funktion „Attribut nach Feldwert“. Ebenfalls mit der Funktion „Attribut nach Feldwert“ werden sowohl für den HU- als auch für den OSM-Gebäudelayer ein Layer für Gebäude ohne Gegenstück im anderen Datensatz erzeugt.

Zunächst muss in Tableau eine Verbindung zu den Daten hergestellt werden. Dabei wird jede Shape-Datei einzeln eingelesen und anschließend im Datenbereich mit den anderen Dateien über übereinstimmende Felder verknüpft. Wenn alle Datensätze eingelesen sind, wird eine Hierarchie der Attribute „Gebietart“, also *gebirgiges Gebiet*, *Großstadt* und *Ballungsraum* etc., und „Testgebiet“ erstellt. Dies ermöglicht es, dass bei den Analysen die Ergebnisse der Testgebiete einer Gebietart zusammengefasst werden können. Über die Tableau-Funktion „berechnetes Feld erstellen ...“ wird der Jaccard-Koeffizient aus den aufgelösten Schnittmenge- und Vereinigungsmengelayern aller Testgebiete berechnet. Die Ergebnisse werden mithilfe der Tableau-Funktion „Gruppieren ...“ zu den im Kapitel 3.5 festgelegten Klassen zusammengefasst. Um die Häufigkeit eines ausgefüllten Attributes feststellen zu können, wird ebenfalls mit der Grup-

pierungs-Funktion gearbeitet. Hierzu wird aus 0 oder Null eine Gruppe „nicht ausgefüllt“ und aus allen anderen Werten eine Gruppe „ausgefüllt“ erstellt. Um die Übereinstimmung eines Adress-Attributes der HU- und der OSM-Gebäude zu prüfen, wird zunächst für dieses Attribut ein berechnetes Feld erstellt.

```
IF STR([HK-field10]) = [Addr House] THEN "übereinstimmende Hausnummer"  
ELSE "keine Übereinstimmung" END
```

Die Berechnung ist gültig.

Übernehmen

OK

Abbildung 9: Beispiel einer Formel für ein berechnetes Feld in Tableau um eine Übereinstimmung der Adresse zu ermitteln.

In Abbildung 9 ist die Formel für die Überprüfung der Hausnummer zu erkennen. Mithilfe einer IF-ELSE-Abfrage werden die Attributwerte „übereinstimmende Hausnummer“ und „keine Übereinstimmung“ vergeben. Danach werden die Attributwerte gruppiert, um eine Abfrage nach dem einen oder dem anderen Attributwert zu ermöglichen. Für eine Darstellung der Ergebnisse werden die Felder und Gruppen in die Bereiche Zeilen oder Spalten, oder in die Markierungs- oder Filterfenster gezogen, um Tabellen oder Grafiken bzw. Diagramme zu erzeugen.

4. ERGEBNISSE

Im Ergebnisteil werden die essenziellen Erkenntnisse dieser Bachelorarbeit in kondensierter Form präsentiert. Die meisten der hier zusammengefassten Ergebnisse werden nach den Kategorien zusammengefasst. Die ausführlichen Ergebnisse finden sich in Tabellen im Anhang.

4.1 ERGEBNISSE DES VERGLEICHS DER GEOGRAFISCHEN/ GEOMETRISCHEN ÜBEREINSTIMMUNG

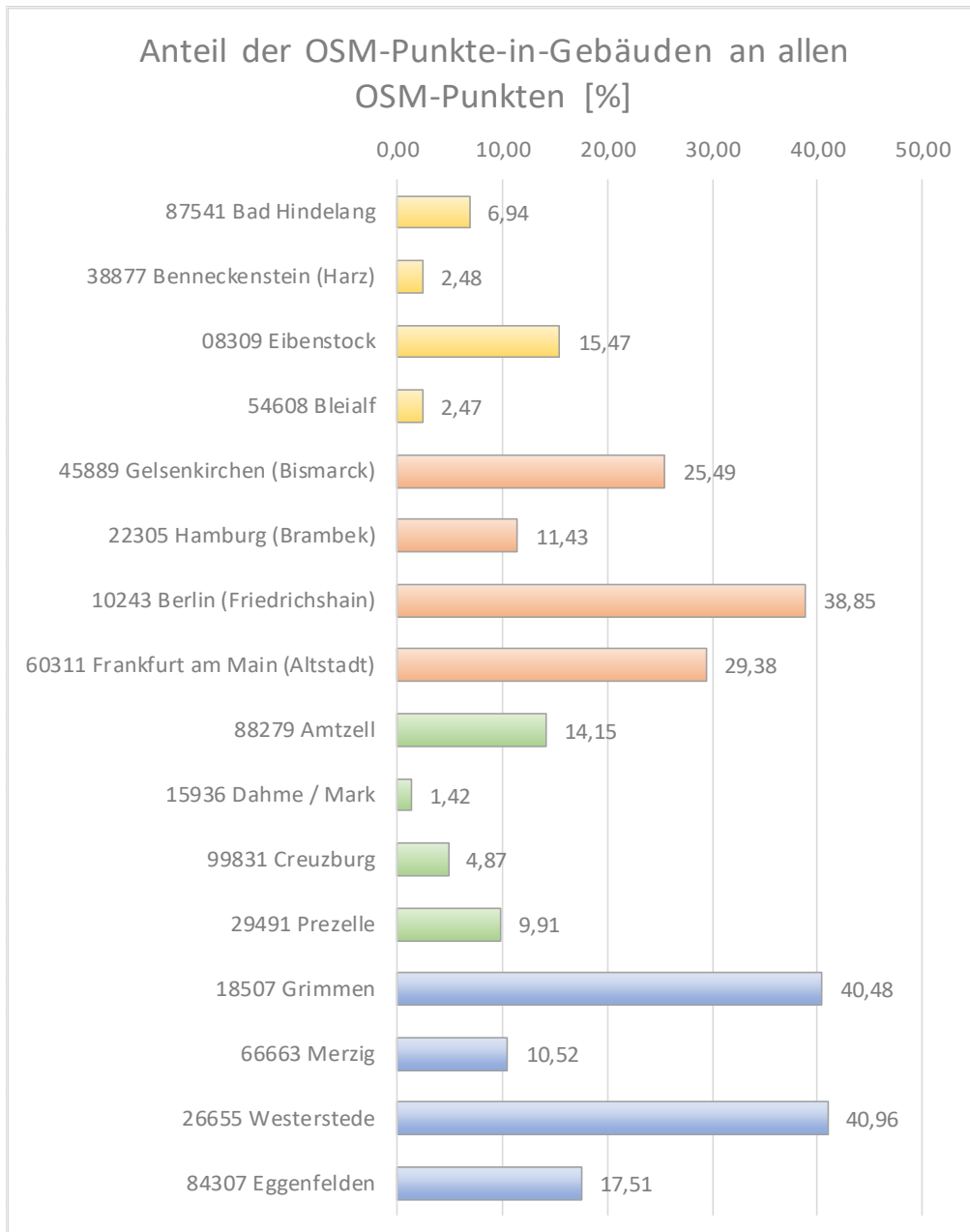


Abbildung 10: Diagramm Anteil der OSM-Punkte-in-Gebäuden an allen OSM-Punkten in %.
Balkenfarben: gebirgige Gebiete: gelb, Großstädte und Ballungsräume: rot, ländliche Gebiete: grün, mittelgroße Städte: blau.

Die Abbildung 10 zeigt die Anteile der OSM-Punkte, die in Gebäuden liegen an allen OSM-Punkten für die einzelnen Testgebiete in Prozent. Innerhalb jeder Kategorie gibt es zwischen den einzelnen Testgebieten große Schwankungen.

Dennoch ergeben sich gewisse Trends. In *gebirgigen Gebieten* und *ländlichen Gebieten* sind erkennbar weniger OSM-Punkte Punkte, die Gebäude kennzeichnen. Insgesamt liegen in allen Testgebieten lediglich zwischen 1,4 % und 41 % aller Punkte in Gebäuden.

Tabelle 6: Mittelwert, Median und Maximum der HU-, OSM-Gebäude- und Überdeckungsflächen für die Kategorien.

Gebietart	Fläche HU [m ²]			Fläche OSM-Gebäude [m ²]			Fläche Überdeckung (Schnittmenge) [m ²]		
	Mittel	Median	Maximum	Mittel	Median	Maximum	Mittel	Median	Maximum
gebirgige Gebiete	143	96	5859	256	158	7865	114	77	5831
Großstädte und Ballungsräume	468	173	25394	421	165	35806	216	79	25394
ländliche Gebiete	162	90	9846	311	167	25216	130	73	9846
mittelgroße Städte	181	90	93260	272	154	46084	123	72	78924

Die Tabelle 6 enthält Mittelwerte, Mediane und Maxima der HU- und OSM-Gebäudeflächen und der Überdeckungsflächen zwischen HU- und OSM-Gebäuden für die verschiedenen Kategorien in m². Die Maxima der verschiedenen Flächenarten in den einzelnen Kategorien lassen darauf schließen, dass die größten Gebäude der Kategorien sowohl im HU- als auch im OSM-Datensatz übereinstimmen und ziemlich gut zusammenpassen. Lediglich bei den *mittelgroßen Städten* gibt es eine größere Differenz zwischen der kleineren maximalen Gebäudefläche und der maximalen Überdeckungsfläche. Dies könnte bedeuten, dass das HU-Gebäude in OSM als zwei oder mehr Gebäude dargestellt wurde. Außer in der Kategorie *Großstädte und Ballungsräume* sind die Mediane und Mittelwerte der HU-Gebäudeflächen immer kleiner als die OSM-Gebäudeflächenmediane.

Tabelle 7: Mittelwert, Median und Maximum der Gebäudepunktzahl der HU und der OSM-Gebäude für die Kategorien.

Gebietart	Anzahl Gebäudepunkte HU			Anzahl Gebäudepunkte OSM-Gebäude		
	Mittel	Median	Maximum	Mittel	Median	Maximum
	gebirgige Gebiete	9	7	255	7	5
Großstädte und Ballungsräume	19	9	1709	10	7	182
ländliche Gebiete	8	7	227	7	5	28
mittelgroße Städte	9	7	271	7	5	11

Beim Vergleich der Gebäudepunktzahl in der Tabelle 7 ergibt sich ein anderes Ergebnis. Auch hier wurden Mittelwerte, Mediane und Maxima für die HU- und OSM-Gebäude-Datensätze gebildet, allerdings gibt es in allen Kategorien und bei allen Vergleichskriterien mehr HU-Gebäudepunkte als OSM-Gebäudepunkte. Daraus lässt sich schließen, dass die HU-Gebäude in der Regel detaillierter aufgenommen wurden.

Tabelle 8: Jaccard-Koeffizient Mittelwert, Median und für das ganze Testgebiet für die Kategorien.

Gebietart	Jaccard-Koeffizient	Jaccard-Koeffizient	Jaccard-Koeffizient
	Mittelwert	Median	ganzes Testgebiet
gebirgige Gebiete	0,8413	0,9155	0,8242
Großstädte und Ballungsräume	0,8713	0,9483	0,9246
ländliche Gebiete	0,7800	0,8910	0,5486
mittelgroße Städte	0,7768	0,8701	0,6519

Die Tabelle 8 zeigt die Jaccard-Koeffizient-Mittelwerte und -Mediane und die Mittel der Jaccard-Koeffizienten, die für ein gesamtes Testgebiet berechnet wurden, für die verschiedenen Kategorien. Die Jaccard-Koeffizienten-Mittelwerte und -Mediane liegen alle über 0,70. Der Definition aus dem Kapitel 3.5 nach zeigt ein Jaccard-Koeffizient über 0,70, dass es sich bei den verglichenen Gebäuden eindeutig um das gleiche Gebäude handelt. Bei den Mittelwerten der Jaccard-Koeffizienten für ganze Testgebiet liegen die Mittel der Kategorien *ländliche Gebiete* und *mittelgroße Städte* zwischen 0,50 und 0,70. Nach Kapitel 3.5 definiert dieses Intervall lediglich eine mögliche Übereinstimmung der Gebäude und keine sichere.

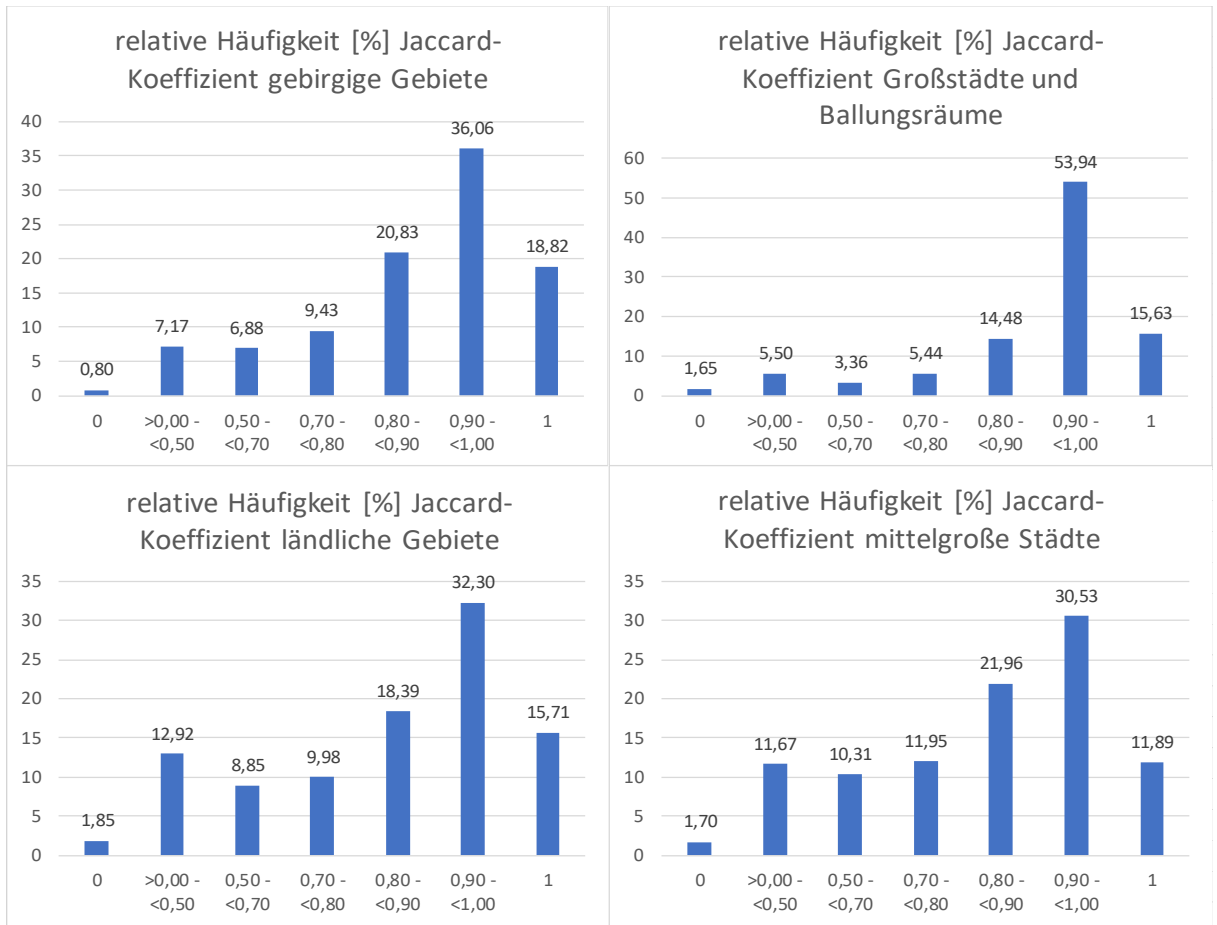


Abbildung 11: Diagramme der Jaccard-Koeffizient-Verteilung der Kategorien.

Die Abbildung 11 zeigt vier Diagramme. Jedes Diagramm steht für eine Kategorie und zeigt die Verteilung der Jaccard-Koeffizienten auf die in Kapitel 3.5 definierten Intervalle in Prozent. Dabei fällt auf, dass die relative Häufigkeit für Jaccard-Koeffizienten, die im Intervall 0,90 - <1,00 liegen, in der Kategorie Großstädte und Ballungsräume bei gerundet 54% liegt, wohingegen die relative Häufigkeit von Jaccard-Koeffizienten in diesem Intervall in den anderen Testgebieten gerundet zwischen 30% und 35% liegt. In allen Kategorien sind die meisten Jaccard-Koeffizienten größer als 0,70 und zwischen 10% und 20% aller Jaccard-Koeffizienten sind 1,00.

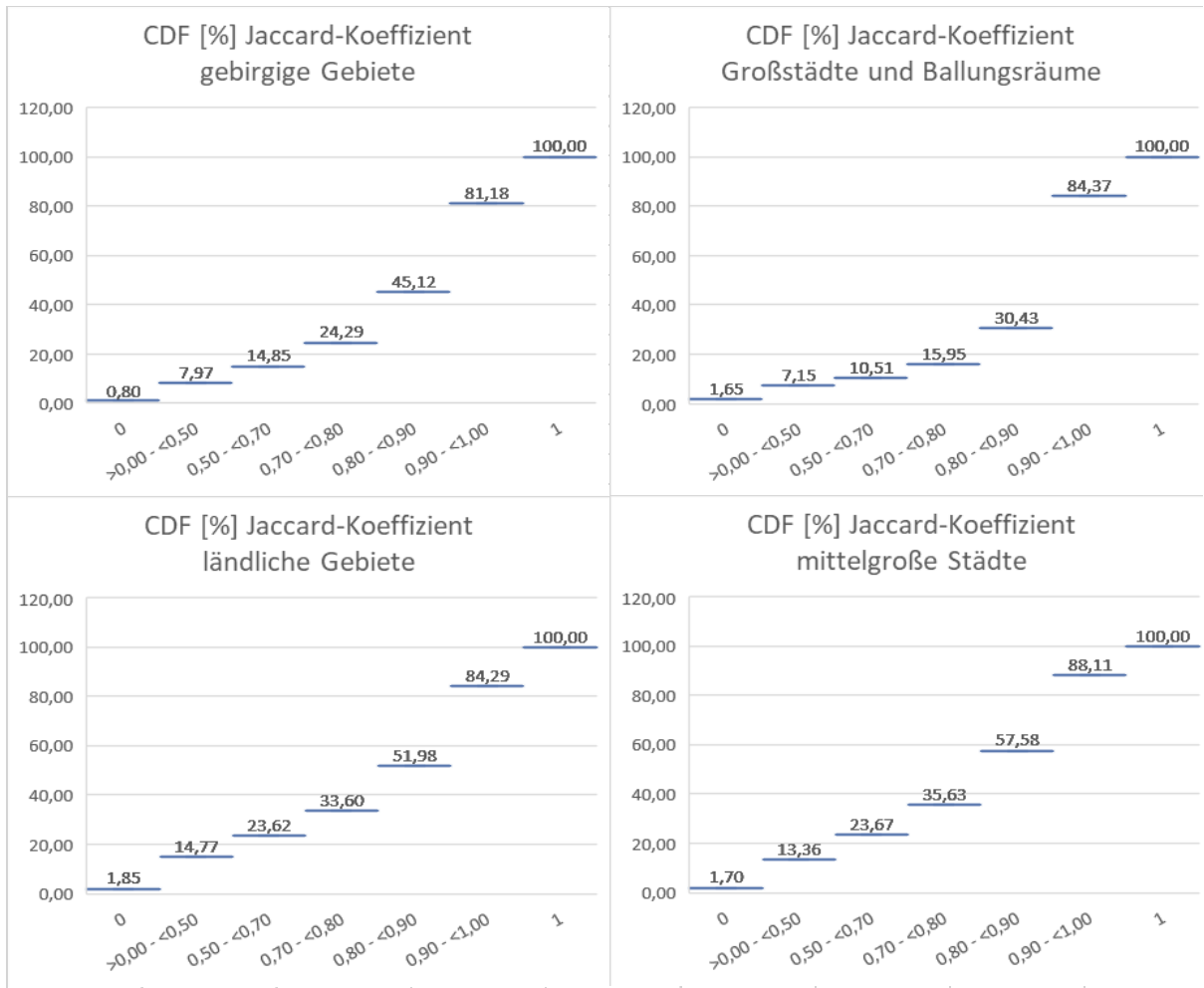


Abbildung 12: CDF-Stufendiagramme der Jaccard-Koeffizient-Verteilung der Kategorien.

Die Abbildung 12 zeigt für jede Kategorie ein CDF-Stufendiagramm. Die Sprünge zwischen den einzelnen Stufen geben die relative Häufigkeit der Jaccard-Koeffizienten für die Intervalle wieder. Die CDF hat im Vergleich zu den relativen Häufigkeiten den Vorteil, dass man an ihr einfacher ablesen kann, wie viel Prozent aller Werte größer oder kleiner als ein bestimmter Schwellwert (Jaccard-Koeffizient-Intervall) sind. Ob die Werte kleiner sind, lässt sich direkt ablesen, indem vom Strich, der das Intervall repräsentiert, waagrecht zur CDF-Achse in Prozent gegangen wird. Ob die Werte größer als ein Schwellwert sind lässt sich über die Formel $1 - \text{CDF}$ berechnen.

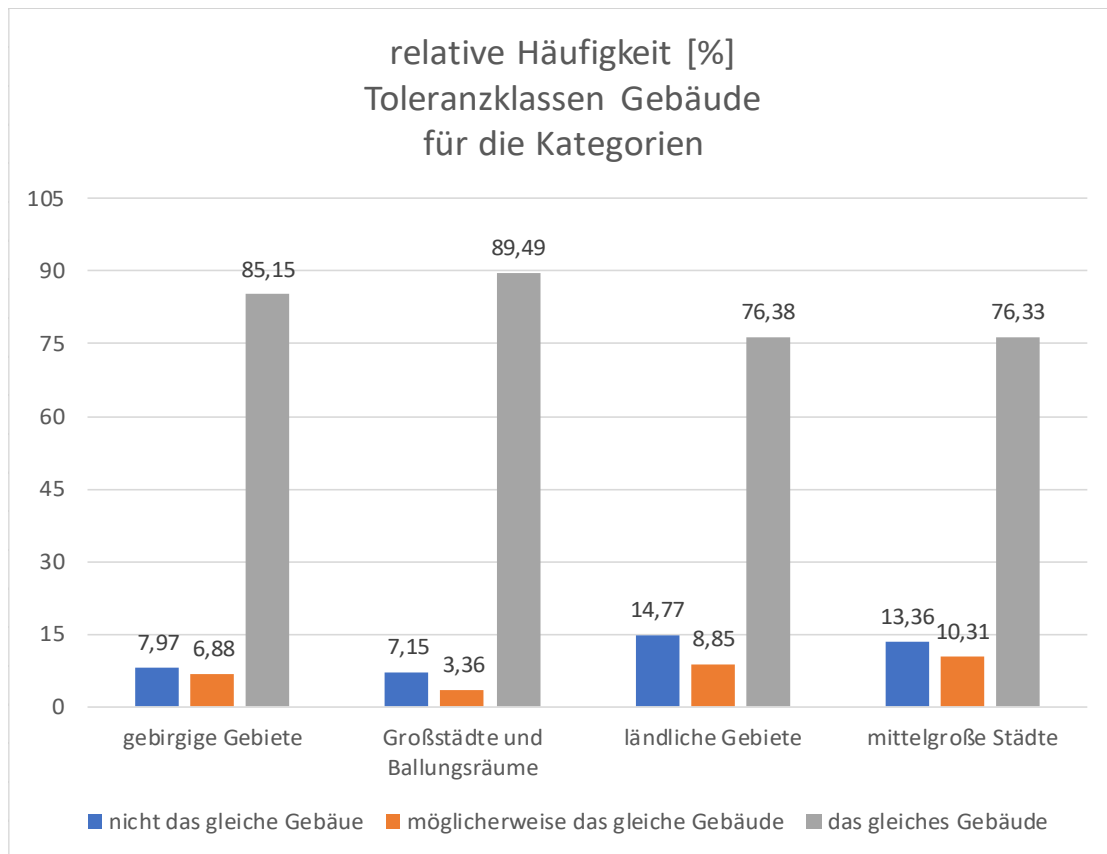


Abbildung 13: Toleranzklassen-Verteilung der Gebäudeübereinstimmung der verschiedenen Kategorien

Die Abbildung 13 zeigt für die verschiedenen Kategorien, wie oft die Jaccard-Koeffizienten in die Klassen *nicht das gleiche Gebäude*, *möglicherweise das gleiche Gebäude* und *das gleiche Gebäude* fallen. Bei den Kategorien *ländliche Gebiete* und *mittelgroße Städte* liegen etwa 10% aller Jaccard-Koeffizienten in der Klasse *möglicherweise das gleiche Gebäude* und etwa 76% aller Jaccard-Koeffizienten in der Klasse *das gleiche Gebäude*. Die Anzahl der Jaccard-Koeffizienten in der Klasse *nicht das gleiche Gebäude* ist in diesen beiden Kategorien etwa doppelt so groß wie in den Kategorien *gebirgige Gebiete* und *Großstädte und Ballungsräume*. In diesen Kategorien fallen ca. 3% bis 7% aller Jaccard-Koeffizienten in die Klasse *möglicherweise das gleiche Gebäude* und ca. 85% bis 90% aller Jaccard-Koeffizienten in die Klasse *das gleiche Gebäude*. Insgesamt lässt sich sagen, dass in den Kategorien *gebirgige Gebiete* und *Großstädte und Ballungsräume* die HU- und OSM-Gebäude-Daten etwas besser zusammenpassen, als in den Kategorien *ländliche Gebiete* und *mittelgroße Städte*.

4.2 ERGEBNISSE DES VOLLSTÄNDIGKEITSVERGLEICHS

Tabelle 9: Unterschied Gebäudeanzahl HU und OSM.

Testgebiete / Gebietart	OSM- Gebäude (Anzahl)	HU (Anzahl)	Unterschied HU und OSM- Gebäude
gebirgige Gebiete			
87541 Bad Hindelang	3594	5561	1967
38877 Benneckenstein (Harz)	1769	1746	23
08309 Eibenstock	4413	5820	1407
54608 Bleialf	1631	3746	2115
Großstädte und Ballungsräume			
45889 Gelsenkirchen (Bismarck)	5736	7036	1300
22305 Hamburg (Brambek)	1750	1570	180
10243 Berlin (Friedrichshain)	1741	2146	405
60311 Frankfurt am Main (Altstadt)	1379	2566	1187
ländliche Gebiete			
88279 Amtzell	3185	4569	1384
15936 Dahme / Mark	3125	14802	11677
99831 Creuzburg	894	3021	2127
29491 Prezelle	190	1128	938
mittelgroße Städte			
18507 Grimmen	5160	8152	2992
66663 Merzig	13264	19367	6103
26655 Westerstede	17715	53355	35640
84307 Eggenfelden	5421	11489	6068

Die Tabelle 9 zeigt für die einzelnen Testgebiete um wie viel Gebäude sich die Anzahl der Gebäude für HU und OSM-Gebäude unterscheidet. Bis auf in *38877 Benneckenstein (Harz)* und *22305 Hamburg (Brambek)* ist die Gebäudeanzahl bei den HU immer größer als bei OSM. Die Anzahl der Gebäude ohne Gegenstück im anderen Datensatz wird in der nächsten Abbildung, also Abbildung 14, behandelt. Der Gebäudeanzahlunterschied ist nicht dasselbe wie die Anzahl der Gebäude ohne Gegenstück im anderen Datensatz.

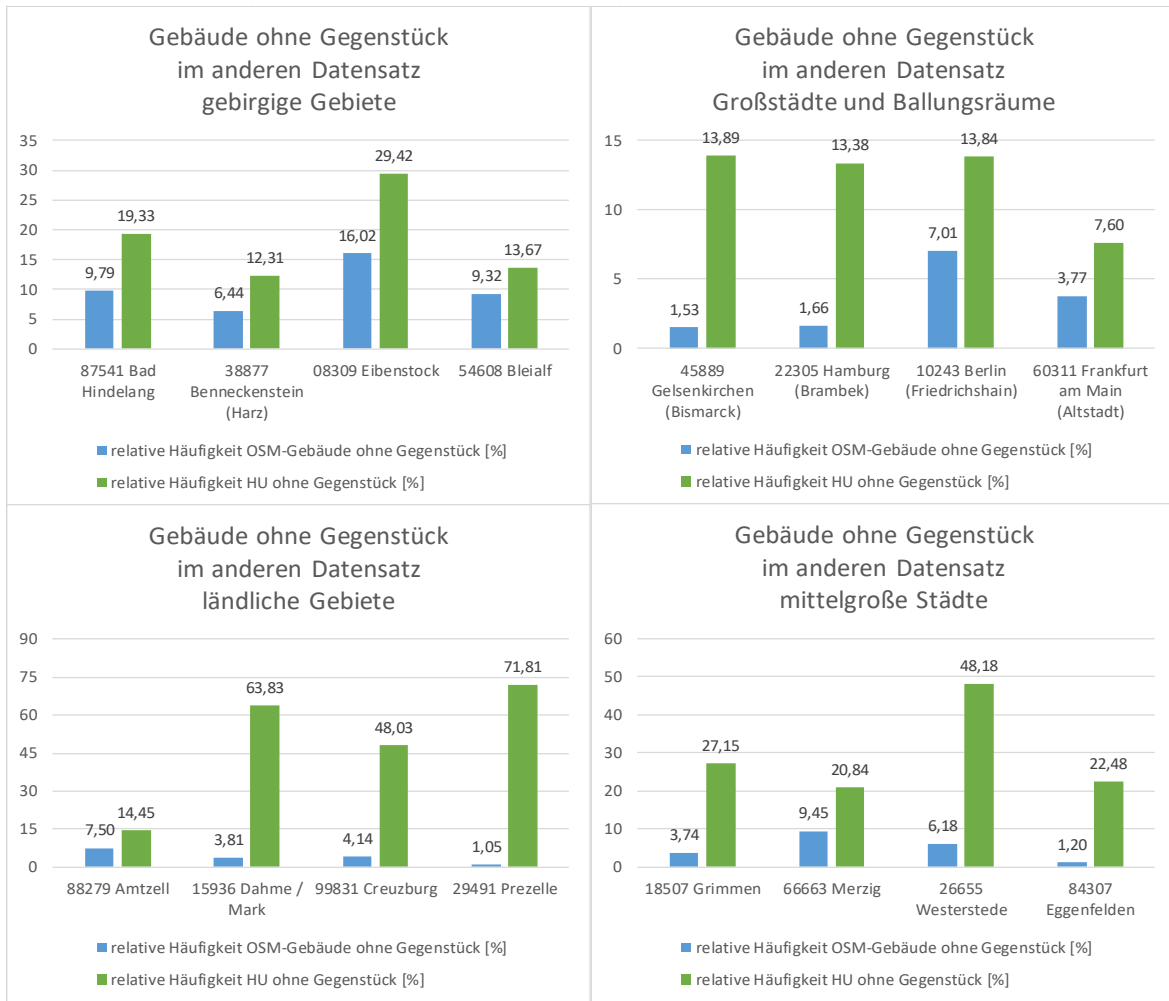


Abbildung 14: Gebäude ohne Gegenstück im anderen Datensatz (relative Häufigkeit)

Die Abbildung 14 zeigt ein Diagramm für jede Kategorie mit den relativen Häufigkeiten der OSM-Gebäude ohne Gegenstück im anderen Datensatz und der HU-Gebäude ohne Gegenstück im anderen Datensatz in Prozent für die Testgebiete der Kategorien. Auffällig ist, dass immer deutlich mehr HU-Gebäude kein Gegenstück im anderen Datensatz besitzen. Dies dürfte größtenteils darauf zurückzuführen sein, dass, wie in Tabelle 9 erkennbar, in den meisten Testgebieten mehr HU-Gebäude als OSM-Gebäude vorhanden sind.

4.3 ERGEBNISSE DES INFORMATIONSGEHALTSVERGLEICHS

Bei den HU-Daten sind immer beide Attribute ausgefüllt, da es sich hierbei um Pflichtattribute handelt, deren Vorhandensein von der ZSHH geprüft wird.

Bei den HK-Daten sind die Attribute *Kennung des Datensatzes*, *eindeutige Nummer des Datensatzes (OI)*, *Qualität der georeferenzierten Gebäudeadresse*, *Schlüssel Land*, *Hausnummer*, *Easting-Wert der UTM-Koordinate*, *Northing-Wert der UTM-Koordinate*, *(unverschlüsselter) Straßenname*, *Postleitzahl* und *postalischer Ortsname* immer ausgefüllt.

Die Attribute *Schlüssel Regierungsbezirk*, *Schlüssel Kreis / kreisfreie Stadt*, *Schlüssel Gemeinde* und *Schlüssel der Straße* sind in manchen Testgebieten oder Bundesländern nicht ausgefüllt. Dies liegt an den leicht unterschiedlichen Regelungen für die Landesvermessung in den Bundesländern und daran, dass bestimmte Attribute für einige Bundesländer oder Ortschaften gar nicht nötig sind. Z.B. bestehen die Bundesländer Hamburg und Berlin nur aus einer einzigen Stadt und benötigen daher das Attribut *Schlüssel Kreis / kreisfreie Stadt* nicht.

Das Attribut *Schlüssel des Orts- bzw. Gemeindeteils* ist in den Testgebieten *38877 Benneckenstein (Harz)*, *54608 Bleialf*, *45889 Gelsenkirchen (Bismarck)*, *88279 Amtzell*, *99831 Creuzburg*, *29491 Prezelle*, *18507 Grimmen* und *26655 Westerstede* nicht und im Testgebiet *08309 Eibenstock* nur zum Teil ausgefüllt. Ein Schlüssel für einen Orts- bzw. Gemeindeteil kann nur dann vergeben werden, wenn auch ein Ortsteil existiert.

Die Attribute *Adressierungszusatz*, *Zusatz zur postalischen Adresse* und *postalischer Ortsteil* sind nur dann ausgefüllt, wenn die Deutsche Post AG einen Zusatz für eine Adresse führt.

Bei den OSM-Daten ist nur die *ID*, also der Schlüssel, ein Pflichtattribut und damit immer ausgefüllt.

Tabelle 10: Relative Häufigkeit [%] der OSM-Attribute name und shop.

Attribut	gebirgige Gebiete		Großstädte und Ballungsräume					
Testgebiet	87541 Hindelang	38877 Benneckenstein (Harz)	08309 Eibenstock	54608 Bleialf	45889 Gelsenkirchen (Bismarck)	22305 Hamburg (Brambek)	10243 Berlin (Friedrichshain)	60311 Frankfurt am Main (Altstadt)
OSM-Gebäude								
rel. Häufigkeit	8,29	4,13	2,11	1,41	1,73	2,57	8,67	13,78
[%] name								
rel. Häufigkeit	0,47	1,19	0,16	0,12	0,12	0,63	0,92	0,51
[%] shop								
OSM-Punkte-in-Gebäuden								
rel. Häufigkeit	65,66	35,00	55,98	91,67	28,72	38,46	30,55	85,55
[%] name								
rel. Häufigkeit	15,66	25,00	25,64	16,67	7,69	16,55	12,34	40,51
[%] shop								
	ländliche Gebiete		mittelgroße Städte					
Testgebiet	88279 Arntzell	15936 Dahme / Mark	99831 Creuzburg	29491 Prezelle	18507 Grimmen	66663 Merzig	26655 Westerstede	84307 Eggenfelden
OSM-Gebäude								
rel. Häufigkeit	1,00	2,27	1,34	0,53	1,88	1,36	2,69	0,77
[%] name								
rel. Häufigkeit	0,03	0,45	0,34	0,00	0,27	0,27	0,41	0,13
[%] shop								
OSM-Punkte-in-Gebäuden								
rel. Häufigkeit	16,67	73,08	80,00	18,18	2,91	85,34	7,69	75,63
[%] name								
rel. Häufigkeit	4,86	3,85	15,00	0,00	0,93	43,10	3,14	38,58
[%] shop								

Die Tabelle 10 zeigt die relative Häufigkeit zweier nicht-adressbezogener OSM-Attribute in Prozent für alle Testgebiete. Als ein OSM-Punkt bzw. OSM-Gebäude-Attribut, das häufig ausgefüllt wurde, wird das Attribut *name*

betrachtet und als Attribut, das selten ausgefüllt wurde, aber in allen Testgebieten vorhanden ist, wurde das Attribut *shop* gewählt. Die Attribute wurden bei den OSM-Punkten immer häufiger ausgefüllt als bei den OSM-Gebäuden. Wie oft die Attribute ausgefüllt wurden, schwankt auch zwischen Testgebieten innerhalb einer Kategorie stark. Das Attribut *name* wurde bei den OSM-Punkten gerundet bis zu 86% und bei den OSM-Gebäuden gerundet bis zu 9% ausgefüllt. Das Attribut *shop* wurde bei den OSM-Punkten gerundet bis zu 43% und bei den OSM-Gebäuden gerundet bis zu 1,2% ausgefüllt.

Bei den adressbezogenen OSM-Gebäude-Attributen ist das Attribut *addr_district* nur im Testgebiet *18507 Grimmen*, das Attribut *addr_hamlet* nur im Testgebiet *26655 Westerstede*, das Attribut *addr_state* nur in den Testgebieten *18507 Grimmen*, *66663 Merzig* und *26655 Westerstede* und das Attribut *addr_subdivision* nur im Testgebiet *15936 Dahme / Mark* ausgefüllt.

Die adressbezogenen OSM-Gebäude-Attribute *addr_suburb* und *addr_housename* sind in den meisten Testgebieten ausgefüllt und die anderen adressbezogenen OSM-Gebäude-Attribute sind zwischen gerundet 0,1% und 69% ausgefüllt.

Bei den adressbezogenen OSM-Punkte-in-Gebäuden-Attributen ist das Attribut *addr_hamlet* nur in den Testgebieten *88279 Amtzell* und *26655 Westerstede*, das Attribut *addr_inclusion* nur im Testgebiet *10243 Berlin (Friedrichshain)*, das Attribut *addr_state* nur im Testgebiet *26655 Westerstede* und das Attribut *addr_place* nur in den Testgebieten *88279 Amtzell* und *18507 Grimmen* ausgefüllt.

Die adressbezogenen OSM-Punkte-in-Gebäuden-Attribute *addr_suburb* und *addr_housename* sind in den meisten Testgebieten ausgefüllt und die anderen adressbezogenen OSM-Punkte-in-Gebäuden-Attribute sind zwischen gerundet 7,7% und 98 % ausgefüllt.

Tabelle 11: Übereinstimmende postalische Adresse.

Testgebiete / Gebietart	Adressvergleich HU und OSM-Gebäude (relative Häufigkeit [%])				
	übereinstimmende Hausnummer	übereinstimmende Straße	übereinstimmender Ortsname	übereinstimmende Postleitzahl	übereinstimmender Ortsteil
gebirgige Gebiete	32,26	36,57	26,84	14,35	10,43
87541 Bad Hindelang	27,54	30,67	27,25	27,32	16,73
38877 Benneckenstein (Harz)	23,23	23,30	0,00	34,18	0,00
08309 Eibenstock	51,98	60,64	47,65	0,00	13,11
54608 Bleialf	3,34	3,54	0,27	3,54	0,00
Großstädte und Ballungsräume	18,49	30,82	27,39	32,24	1,66
45889 Gelsenkirchen (Bismarck)	24,96	35,69	35,80	36,06	0,00
22305 Hamburg (Brambek)	15,01	37,08	39,17	39,28	1,16
10243 Berlin (Friedrichshain)	4,16	8,26	9,35	9,48	7,81
60311 Frankfurt am Main (Altstadt)	19,28	34,44	0,00	39,24	0,00
ländliche Gebiete	15,09	10,65	12,73	10,33	0,74
88279 Amtzell	32,24	20,90	29,49	18,57	0,64
15936 Dahme / Mark	3,40	3,19	0,42	4,54	0,52
99831 Creuzburg	2,80	4,63	3,17	4,76	1,59
29491 Prezelle	4,76	5,95	5,95	6,55	1,79
mittelgroße Städte	14,47	15,70	17,15	16,44	3,43
18507 Grimmen	21,18	9,34	25,35	19,38	1,40
66663 Merzig	16,61	20,88	19,98	20,55	0,08
26655 Westerstede	13,68	16,76	16,46	16,48	7,52
84307 Eggenfelden	6,02	7,51	5,49	4,82	0,02

Die Tabelle 11 zeigt die Übereinstimmung der Adressattribute zwischen dem HU- und dem OSM-Gebäude-Datensatz für übereinstimmende Gebäude als relative Häufigkeit in Prozent. Für den Vergleich der Adressen wurden die OSM-Gebäude mit der OI der HU-Gebäude, mit denen sie sich schneiden, versehen, um eine Verknüpfung in Tableau zu ermöglichen. Die relative Häufigkeit in Prozent bezieht sich auf die Anzahl der Verknüpfungen in den Testgebieten. Auch bei diesem Vergleich gibt es zwischen den Testgebieten einer Kategorie teils starke Schwankungen. Die Werte in den Kategoriezeilen sind Mittelwerte der Werte für die Testgebiete dieser Kategorien. Bei den *Hausnummern* stimmen maximal 51,98%, bei den *Straßennamen* maximal 60,64%, bei den *Ortsnamen* maximal 47,65%, bei den *Postleitzahlen* maximal 39,28% und bei den *Ortsteilen* maximal 16,73% überein.

5. DISKUSSION

Ziel dieser Arbeit ist es, die ZSHH-Daten und die OSM-Daten im Hinblick auf adressbezogene Daten hinsichtlich der *geografischen / geometrischen Übereinstimmung*, der *Datenvollständigkeit* und des *Informationsgehaltes* zu vergleichen, auch wenn die Datensätze für unterschiedliche Zwecke erstellt wurden.

Die Ergebnisse des Vergleichs hinsichtlich der *geografischen / geometrischen Übereinstimmung* lassen sich wie folgt interpretieren:

Da bei den OSM-Punkten, die in Gebäude fallen, selten das Attribut *building* ausgefüllt wurde mussten zunächst alle OSM-Punkte der Testgebiete geladen werden, um die OSM-Punkte in Gebäuden mithilfe einer Verschneidung mit den Gebäudepolygonen herauszufiltern. Aus den gefilterten OSM-Punkten wurde ein neuer Layer erstellt und für die weiteren Analysen verwendet. Dies ist insofern für den Vergleich der OSM-Daten mit den ZSHH-Daten relevant, da dieser zusätzliche Schritt Arbeitszeit und Speicherplatz benötigt.

Beim Vergleich der OSM-Punktzahl mit der Anzahl der OSM-Punkte in Gebäuden zeigte sich, dass in den *gebirgigen und ländlichen Gebieten* weniger OSM-Punkte in Gebäude fallen, als in *mittelgroßen Städten und Großstädten*. Das liegt daran, dass diese Gebiete in der Regel nicht so dicht besiedelt sind wie *mittelgroße Städte* oder *Großstädte und Ballungsräume* und daher dort grundsätzlich mehr andere Objekte als Gebäude existieren. Dass allerdings auch in *mittelgroßen Städten und Großstädten und Ballungsräumen* nicht einmal 50% aller OSM-Punkte in Gebäude fallen liegt daran, dass ein Großteil aller OSM-Punkte andere Objekte kennzeichnet, z.B. Abfalleimer, Parkbänke, Briefkästen und Ampeln. Auch werden bei OSM Punkte dazu verwendet, um Infrastrukturobjekte wie Fußgängerübergänge, Kreuzungen, Bahnsignale oder Weichen von Bahngleisen zu kennzeichnen.

Der Vergleich der Gebäudeflächen in m² zwischen OSM und HU ergab, dass die grundflächenmäßig größten Gebäude in OSM und HU übereinstimmen. Der Sichtvergleich der beiden Datensätze hat zudem ergeben, dass die gleichen Gebäude in beiden Datensätzen oft etwas unterschiedlich dargestellt werden, wie in Abbildung 15 erkennbar ist.

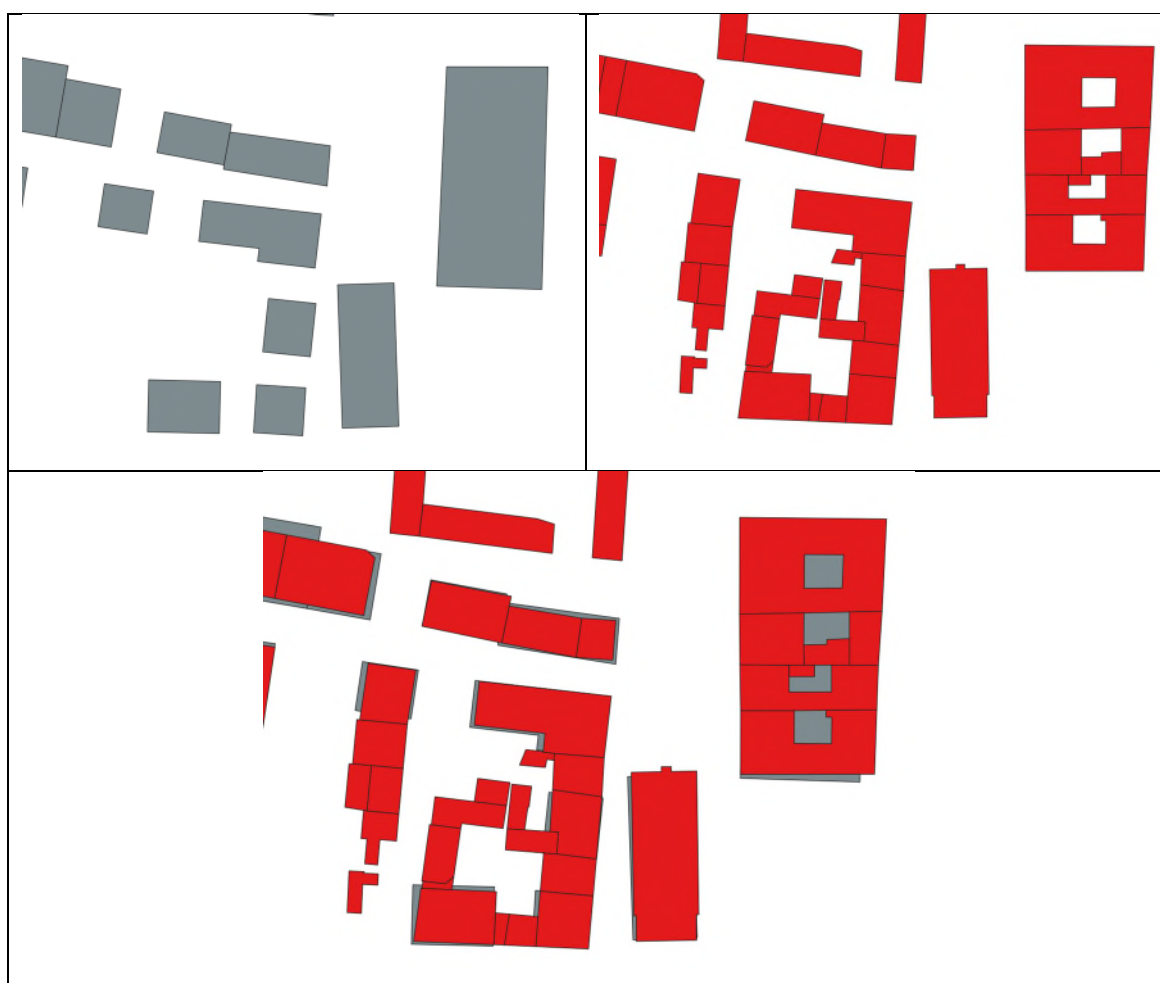


Abbildung 15: OSM-Gebäude (grau), HU-Gebäude (rot) und OSM-Gebäude überlagert mit den HU-Gebäuden.

Vor allem bei Doppelhäusern, Reihenhäusern oder Häuserblocks unterscheidet sich die Abgrenzung der Gebäude zwischen den Datensätzen. Zum Teil werden sie als ein großes Gebäude und zum Teil als mehrere aneinander angrenzende kleine Gebäude dargestellt. Die Mittel und Mediane der HU-Gebäudeflächen sind kleiner als die der OSM-Gebäudeflächen. Das liegt daran, dass die Vermessungsverwaltungen der Länder Liegenschaften aufnehmen, zu denen auch die darauf befindlichen Gebäude gehören. Sobald also Doppelhaushälften etc. zu mehreren Flurstücken gehören, werden sie in den amtlichen Daten, anders als

bei OSM, als mehrere Gebäude dargestellt. Daraus, dass die HU-Gebäude sowohl im Mittel als auch nach den Medianen immer mehr Gebäudepunkte als die OSM-Gebäude besitzen, lässt sich schließen, dass die HU-Gebäude in der Regel auch detaillierter aufgenommen wurden. Dies ist z.B. für Infrastrukturunternehmen relevant, die für eine exakte Planung von Trassen die genaue Form und Lage der Gebäude benötigen.

Die Betrachtung der Jaccard-Koeffizienten ergab, dass die Mittelwerte und Mediane der Jaccard-Koeffizienten für die einzelnen Testgebiete und Kategorien in der Klasse *gleiches Gebäude* liegen, d.h. die Übereinstimmung der Gebäude zwischen den beiden Datenquellen ist hoch. Allerdings gibt es bei den für die gesamten Testgebiete berechneten Jaccard-Koeffizienten Unstimmigkeiten. Sie sind bei den *ländlichen Gebieten* und *mittelgroßen Städten* lediglich in der Klasse einer mögliche und keine sichere Gebäudeübereinstimmung, obwohl die Mittelwerte und Mediane in den Testgebieten besser sind. Da zwar die Mittelwerte und Mediane der Jaccard-Koeffizienten der Testgebiete, aber nicht der Jaccard-Koeffizient des gesamten Testgebietes das Ergebnis einer exakten Übereinstimmung liefern, müssen die Summen der Schnittmengen und der Vereinigungsmengen für die ganzen Testgebiete größere Unterschiede besitzen als die Unterschiede der Schnittmengen und Vereinigungsmengen für einzelne Gebäude.

Der Grund dafür, dass in allen Kategorien zwischen 10% und 20% der Jaccard-Koeffizienten gleich 1 sind und somit eine perfekte Übereinstimmung der Gebäude beider Datenquellen suggeriert wird, ist systematisch bedingt. Bei der Erzeugung der Schnittmengen und der Vereinigungsmengen in QGIS wurden zunächst für jede Schnittfläche zwischen den Gebäuden eine eigene Geometrie erzeugt, also bei Schnittmengen eine oder mehr und bei Vereinigungsmengen meist vier oder mehr. Erst in einem weiteren Schritt wurden die zusammengehörenden Teilflächen wieder verschmolzen. Dieser Schritt war allerdings QGIS-bedingt nicht ganz fehlerfrei und manche Teilflächen wurden nicht verschmolzen (siehe Abbildung 5 – 8 im Kapitel 3.6). Es ist durchaus möglich, dass es bei einzelnen Gebäuden eine perfekte Übereinstimmung zwischen beiden Datensätzen gibt, allerdings sicher nicht 10% bis 20%.

Dennoch fallen in allen Kategorien zwischen 75% und 90% aller Jaccard-Koeffizienten in das Intervall 0,70 bis 1 und gelten daher als Übereinstimmung zwischen den dazugehörigen Gebäuden. Lediglich bei 7% bis 15% aller Jaccard-Koeffizienten handelt es sich sicher nicht um dasselbe Gebäude.

Insgesamt lässt sich also sagen, dass die Gebäudegrundrisse beider Datenquellen trotz etwas unterschiedlicher Datenmodelle, wenn ein Gegenstück im anderen Datensatz vorhanden ist, ziemlich gut zusammenpassen.

Zu den Ergebnissen des *Vollständigkeitsvergleichs* lässt sich sagen, dass bis auf in zwei Testgebieten immer deutlich mehr HU-Gebäude als OSM-Gebäude existieren. Dies ist auch der Grund dafür, dass die Zahl der HU-Gebäude ohne Gegenstück im anderen Datensatz so viel höher ist, als die Zahl der OSM-Gebäude ohne Gegenstück im anderen Datensatz. Dennoch gibt es in allen Testgebieten auch OSM-Gebäude ohne Gegenstück. Das bedeutet, dass bei OSM zum Teil andere Gebäude als bei HU aufgenommen wurden.

Insgesamt wurden von den Vermessungsämtern mehr Gebäude aufgenommen als von den freiwilligen Mappern von OSM. Dies liegt an der Koppelung von Gebäude und Grundstück (vgl. § 94 BGB), wodurch eine amtliche Gebäudeeinmesspflicht entsteht. Der amtlichen Gebäudedefinition, die sich zwischen den Bundesländern zum Teil ein wenig unterscheidet, nach sind bei den HU also sämtliche Gebäude Deutschlands enthalten. Bei OSM besteht diese Gebäudeaufnahmepflicht nicht.

Die Ergebnisse des *Informationsgehaltsvergleichs* zeigen, dass die Attribute der beiden Datenquellen sehr verschieden ausgefüllt sind. Der HU-Datensatz besteht aus den Geometrien und lediglich zwei Attributen, die immer ausgefüllt sind. Wegen unterschiedlicher Regelungen in den verschiedenen Bundesländern sind manche HK-Attribute in bestimmten Bundesländern gar nicht ausgefüllt. Die Zusätze zu Ortsbezeichnungen etc. sind in allen Testgebieten nur dann ausgefüllt, wenn die Deutsch Post AG für die Adresse einen Zusatz führt und die anderen Attribute sind immer ausgefüllt. HU und HK sind zusammen somit sehr gute Quellen für Gebäudegrundrisse und vollständige postalische Adressen.

Bei den OSM-Daten ist hingegen nur die ID immer ausgefüllt. Die OSM-Gebäude besitzen insgesamt weniger Attribute als die OSM-Punkte. Diese Attribute sind auch seltener ausgefüllt. Dennoch sind die Attribute der OSM-Punkte auch deutlich seltener ausgefüllt als die Attribute der amtlichen Datensätze.

Die Übereinstimmung der Adressen der gleichen Gebäude ist in den meisten Testgebieten ziemlich gering. Das dürfte größtenteils daran liegen, dass die Attribute der OSM-Gebäude deutlich seltener ausgefüllt sind als die der HU-Gebäude.

Zusammenfassend für den Informationsgehalt lässt sich sagen, dass die ZSHH-Datensätze den OSM-Datensätzen in dieser Hinsicht weit überlegen sind, da die Attribute viel öfter ausgefüllt sind und sie mehr Adressattribute besitzen.

Die grundlegende Überlegung für diese Bachelorarbeit war die Frage, ob die freien OSM-Daten die ZSHH-Daten ersetzen können. Dies ist nicht der Fall: es gibt weniger OSM-Gebäude als HU-Gebäude, die OSM-Gebäude sind in der Regel nicht so detailliert erstellt wie die HU-Gebäude, bei OSM gibt es weniger adressbezogene Attribute und diese sind nicht annähernd so oft ausgefüllt wie die HK-Attribute.

6. ZUSAMMENFASSUNG / FAZIT / AUSBLICK

In dieser Arbeit wurden Daten aus zwei verschiedenen Quellen miteinander hinsichtlich *geografischer / geometrischer Übereinstimmung, Vollständigkeit und Informationsgehalt* verglichen. Die Datenquellen waren die Zentralen Stelle Hauskoordinaten und Hausumringe mit den Datensätzen Hauskoordinaten Deutschland und Hausumringe Deutschland und OpenStreetMap mit Gebäude- und Punkt-Daten. Es wurden vier Kategorien mit je vier Testgebieten festgelegt: *gebirgige Gebiete, Großstädte und Ballungsräume, ländliche Gebiete und mittelgroße Städte*. Wenn ein Gegenstück im anderen Datensatz vorhanden war, passten die Gebäudegrundrisse beider Datenquellen gut zusammen, allerdings waren die ZSHH-Daten den OSM-Daten sowohl hinsichtlich der Vollständigkeit als auch des Informationsgehaltes überlegen. Auch bei den Abgabeformaten sind die ZSHH-Daten den OSM-Daten überlegen, da es sich um offizielle und weitverbreitete Standards handelt. Die OSM-Daten hingegen sind in einem von OSM entwickelten Format erstellt und müssen erst in ein gängiges Format konvertiert werden.

Die Datensätze wurden für unterschiedliche Zwecke erstellt. Die ZSHH-Daten dienen der Bereitstellung von deutschlandweiten lagerichtigen Gebäudegrundrissen und korrekten postalischen Adressen und OSM wurde als opensource Online-Weltkarte, die durch freiwillige Mitwirkende erstellt und gepflegt wird, konzipiert. Eine vollständige und korrekte Adressdatensammlung und Gebäudegrundrisse aller Gebäude in Deutschland sind nicht das Hauptziel des OSM-Projektes, weshalb OSM derzeit weder HU noch HK ersetzen kann.

Da sich Gemeinschaftsprojekte wie OSM jedoch stetig weiterentwickeln und an Daten gewinnen, könnte es interessant sein, eventuell in 10 Jahren noch einmal einen ähnlichen Vergleich durchzuführen.

VI. LITERATURVERZEICHNIS

Arbeitsgemeinschaft der Vermessungsverwaltungen der Länder der Bundesrepublik Deutschland. AdV. Retrieved from <http://www.adv-online.de/Wir-ueber-uns/>

Arbeitsgemeinschaft der Vermessungsverwaltungen der Länder der Bundesrepublik Deutschland. AdV-Produkte. Retrieved from <http://www.adv-online.de/AdV-Produkte/>

Arbeitsgemeinschaft der Vermessungsverwaltungen der Länder der Bundesrepublik Deutschland (2019, December 19). *Richtlinie über Gebühren für die Bereitstellung und Nutzung von Geobasisdaten der Vermessungsverwaltungen der Länder der Bundesrepublik Deutschland: AdV-Gebührenrichtlinie - AdV-GR.*

Arbeitskreis PRM, Arbeitsgemeinschaft der Vermessungsverwaltungen der Bundesrepublik Deutschland (2019, April 2). *Musterlizenzvereinbarungen "Geodaten"*.

Cramer, E., & Kamps, U. (2008). *Grundlagen der Wahrscheinlichkeitsrechnung und Statistik: Ein Skript für Studierende der Informatik, der Ingenieur- und Wirtschaftswissenschaften* (zweite überarbeitete Auflage). Berlin, Heidelberg: Springer-Verlag.

Dietrich, D. (2011, October 26). Was sind offene Daten? Retrieved from <https://www.bpb.de/gesellschaft/medien/opendata/64055/was-sind-offene-daten?p=all>

Jaccard, P. (1912). THE DISTRIBUTION OF THE FLORA IN THE ALPINE ZONE. *The New Phytologist*, Vol. XI(2), 37–50. Retrieved from <https://nph.onlinelibrary.wiley.com/doi/epdf/10.1111/j.1469-8137.1912.tb05611.x>

Kaarainen, H. (2005). *Umts networks: Architecture, mobility, and services* (2nd ed.). Chichester, West Sussex, England, Hoboken, NJ: J. Wiley & Sons. Ret-

rieved from

<http://site.ebrary.com/lib/alltitles/docDetail.action?docID=10308118>

Koch, A. (H.), Hrsg., & Donaubauer (Eds.) (2013). *Geoinformationssysteme 2013: Beiträge zum 18. Münchner Fortbildungsseminar 2013 ; [der Runder Tisch GIS e.V. An der Technischen Universität München ... Lädt vom 8. Bis 11. April 2013 ... Zum 18. Münchner Fortbildungsseminar Geoinformationssysteme ein]*. Berlin: Wichmann.

Landesamt für Vermessung und Geoinformation Thüringen (2014). *Geodaten Anwendungen*. Erfurt. Retrieved from

<https://www.thueringen.de/de/publikationen/pic/pubdownload1548.pdf>

OSM-Community. OSM-Wiki. Retrieved from <https://wiki.openstreetmap.org/wiki>

Overpass API User's Manual. Retrieved from

<https://dev.overpass-api.de/overpass-doc/en/>

Ramm, F. (2019, November 6). *OpenStreetMap Data in Layered GIS Format: Free shapefiles - 2019-11-06*. Retrieved from Geofabrik website:

<https://www.geofabrik.de/de/data/shapefiles.html>

Ramm, F., & Topf, J. (2010). *OpenStreetMap: Die freie Weltkarte nutzen und mitgestalten* (3., überarb. und erw. Aufl.). Berlin: Lehmanns Media.

Rezatofighi, H., Tsoi, N., Gwak, J., Sadeghian, A., Reid, I., & Savarese, S. *Generalized Intersection over Union: A Metric and A Loss for Bounding Box Regression*. Retrieved from Computer Science Department, Stanford University, United states; School of Computer Science, The University of Adelaide, Australia; Aibee Inc, USA website: <https://giou.stanford.edu/GIoU.pdf>

Seip, C., Korduan, P., & Zehner, M. L. (2017). *Web-GIS: Grundlagen, Anwendungen und Implementierungsbeispiele*. Berlin, Offenbach: Wichmann.

Zentrale Stelle Hauskoordinaten und Hausumringe, Landesamt für Digitalisierung, Breitband und Vermessung Bayern. Zentrale Stelle Hauskoordinaten und Hausumringe. Retrieved from

<https://www.ldbv.bayern.de/ueberuns/zshh.html>

Zentrale Stelle Hauskoordinaten und Hausumringe, Landesamt für Digitalisierung, Breitband und Vermessung Bayern (2019a, July 1). *Datenformatbeschreibung Hauskoordinaten Deutschland (HK-DE)*. Retrieved from Zentrale Stelle Hauskoordinaten und Hausumringe, Landesamt für Digitalisierung, Breitband und Vermessung Bayern website:
<http://www.adv-online.de/AdV-Produkte/Vertriebsstellen/ZSHH/broker.jsp?uCon=37a705cc-190b-2941-44c9-10d7072e13d6&uBasVariant=11111111-1111-1111-1111-111111111111>

Zentrale Stelle Hauskoordinaten und Hausumringe, Landesamt für Digitalisierung, Breitband und Vermessung Bayern (2019b, July 1). *Datenformatbeschreibung Hausumringe Deutschland (HU-DE)*. Retrieved from Zentrale Stelle Hauskoordinaten und Hausumringe, Landesamt für Digitalisierung, Breitband und Vermessung Bayern website:
<http://www.adv-online.de/AdV-Produkte/Vertriebsstellen/ZSHH/broker.jsp?uCon=37a705cc-190b-2941-44c9-10d7072e13d6&uBasVariant=11111111-1111-1111-1111-111111111111>

Zentrale Stelle Hauskoordinaten und Hausumringe, Landesamt für Digitalisierung, Breitband und Vermessung Bayern (2019c, September 2). *Allgemeine Bedingungen für die Bereitstellung und Nutzung von Geodaten: Allgemeine Geschäfts- und Nutzungsbedingungen*.

VII. ANHANG

ANHANG 1: PROGRAMMSCHRITTE MIT SCREENSHOTS

QGIS:

- neues QGIS-Projekt erstellen
- Projektkoordinatensystem festlegen:
 - „Projekt“
 - „Eigenschaften“

- „KBS“
- Koordinatensystem auswählen
- HU-Layer einladen (über Layer-Browser-Fenster)
- HK-Layer einladen
 - "Layer"
 - "Layer hinzufügen"
 - "Textdatei als Layer importieren ..."

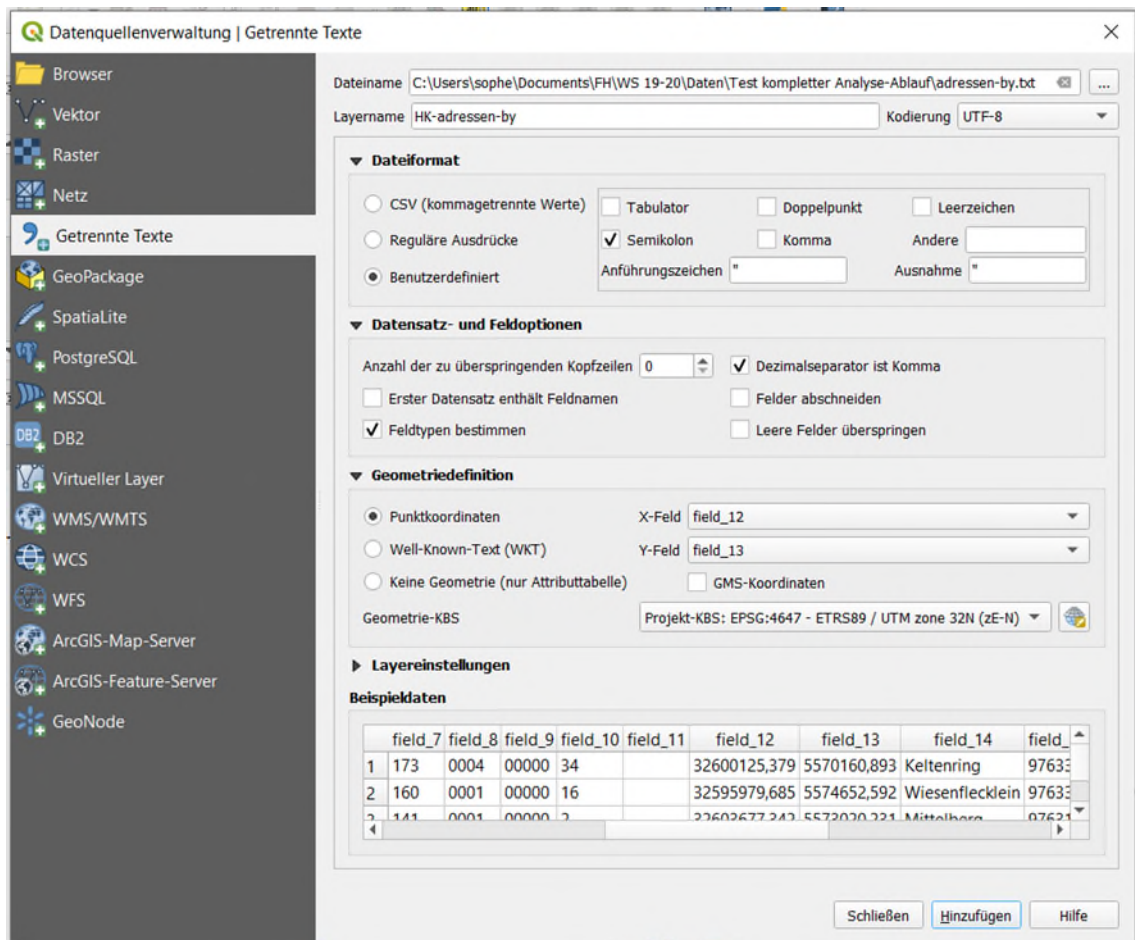


Abbildung 16: HK-Layer aus Textdatei erzeugen.

- Bounding-Box um Postleitzahlengebiet erzeugen
 - Attributtabelle HK-Layer öffnen
 - "Objekte mit Formular wählen / filtern"
 - Filter: Postleitzahl des Gebietes
 - "Objekte wählen"
 - "Vektorverarbeitungswerkzeuge"-Fenster
 - "Vektorengeometrie"

- "Minimale begrenzen Geometrie"
- Eingabelayer angeben, "Nur gewählte Objekte", Geometrietyp: "Einhüllende (Bounding Box)", Begrenzende Geometrie: Ausgabebayer
- HU- und HK-Layer mit Bounding-Box zuschneiden
 - "Vektorverarbeitungswerkzeuge"-Fenster
 - "Vektorlayerüberlagerung"
 - "Zuschneiden"
- OSM-Gebäude- und OSM-Punkt-Layer mit QuickOSM erzeugen
 - QuickOSM öffnen
 - "Layer Extent" und BoundingBox-Layer auswählen
 - Bei Punktlayer: Nodes, Ways, Relations und Points auswählen und Speicherort angeben
 - Bei Gebäudelayer: für Key „building“ angeben, Nodes, Ways, Relations und Multipolygons auswählen und Speicherort angeben
 - „Run query“

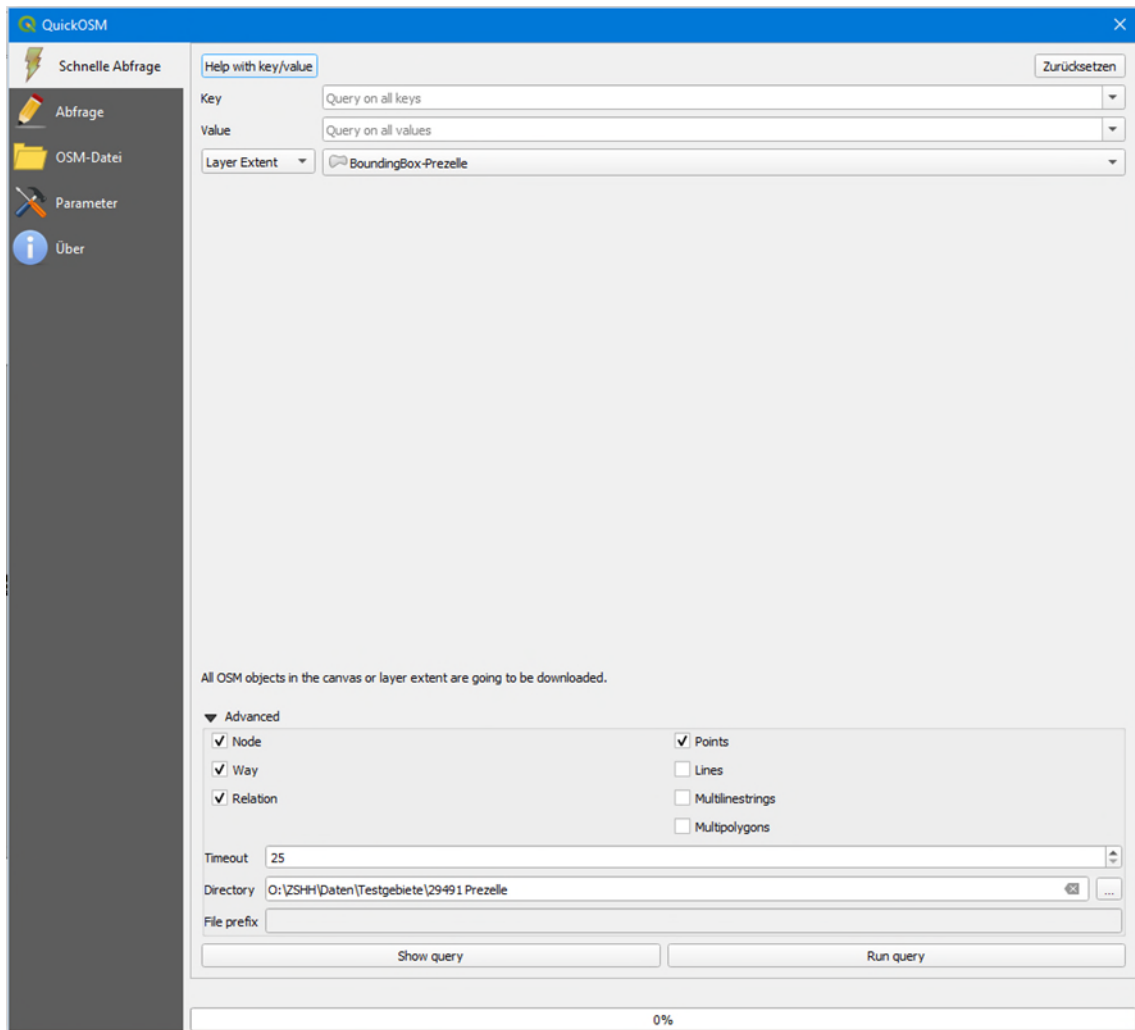


Abbildung 17: OSM-Punktelayer mit dem QuickOSM-Plugin erzeugen.

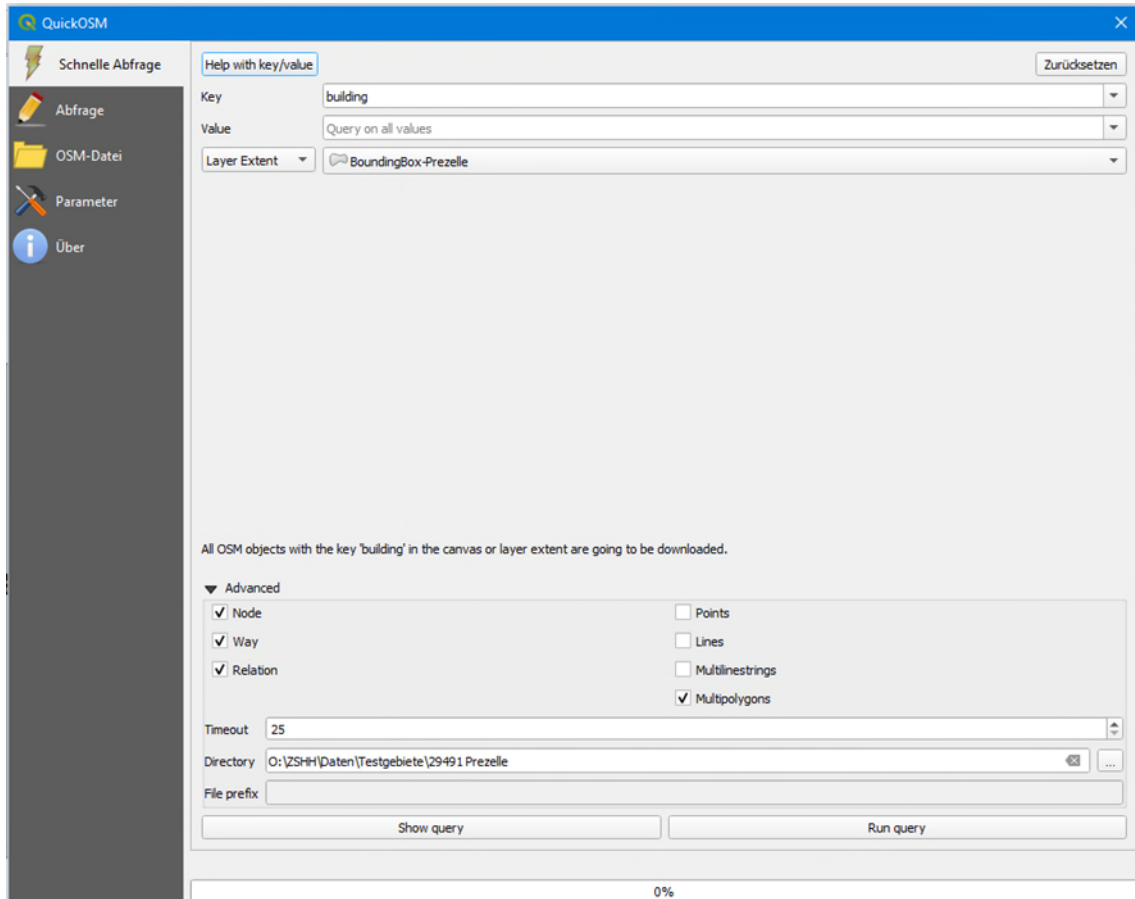


Abbildung 18: OSM-Gebäudelayer mit dem QuickOSM-Plugin erzeugen.

- OSM-Layer reparieren (sonst Fehlermeldung bei weiterer Bearbeitung)
 - "Verarbeitungswerkzeuge"-Fenster
 - "Vektorgeometrie"
 - "Geometrie reparieren"
- OSM-Layer mit Bounding Box zuschneiden
 - „Vektor“
 - „Geoverarbeitungswerkzeuge“
 - „Zuschneiden“
 - Eingabelayer, überlagerter Layer (Bounding-Box-Layer) und Ausgabelayer angeben
- überzählige Attribute der Attributlisten der OSM-Layer löschen
 - Rechtsklick auf Layer
 - „Attributtabelle öffnen“
 - „Spalten löschen“
 - Zu löschende Attribute markieren

- OSM-Layer nach UTM reprojezieren
 - „Verarbeitungswerkzeuge“-Fenster
 - „Vektoren allgemein“
 - „Layer reprojezieren“
 - Eingabelayer, neues Koordinatensystem und Ausgabelayer angeben
- Attribut testgebiet (Name des Testgebietes) für alle Layer erzeugen
 - Rechtsklick auf Layer
 - „Attributtabelle öffnen“
 - Ausgabefeldname und Ausgabefeldtyp angeben, Ausdruck: Postleitzahl und Ortsname

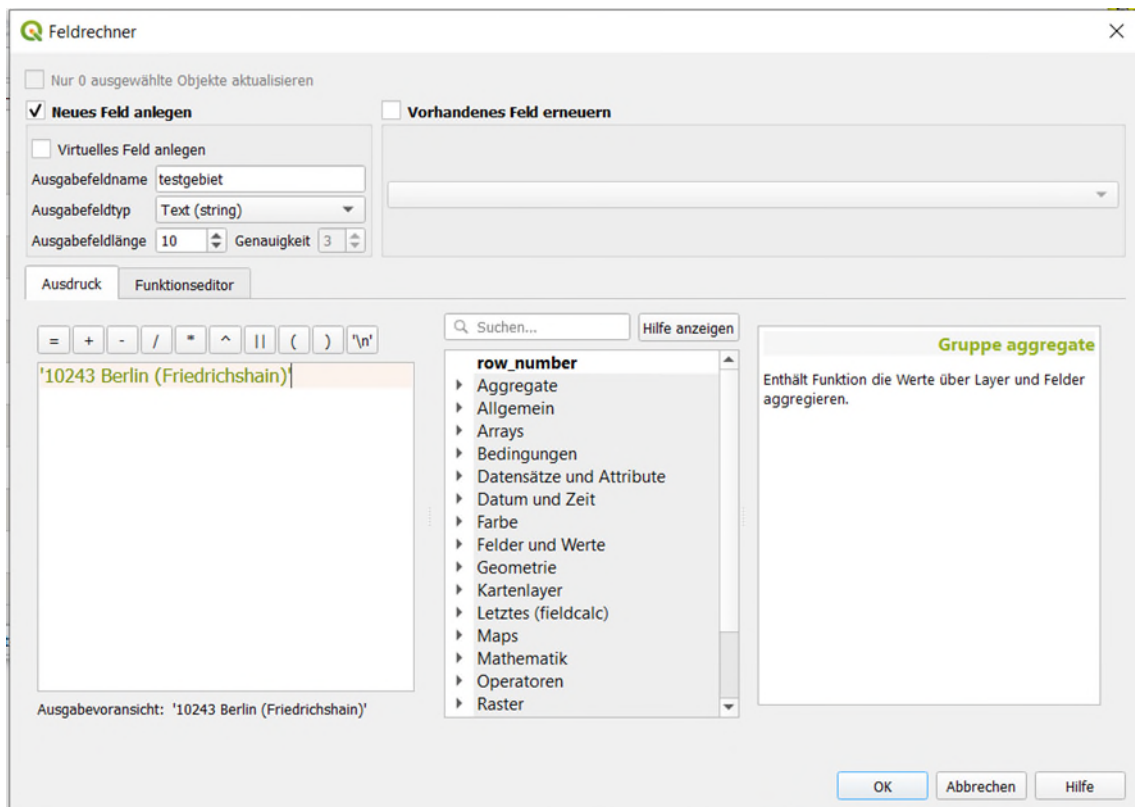


Abbildung 19: Attribut testgebiet erzeugen.

- OSM-Gebäudepunkte-Layer erzeugen: OSM-Punktlayer mit OSM-Gebäudelayer verschneiden
 - „Verarbeitungswerkzeuge“-Fenster
 - „Vektoren allgemein“
 - „Attribute nach Position zusammenfügen“

- Eingabelayer (OSM-Punktlayer), verknüpfter Layer (OSM-Gebäudelayer), geometrisches Prädikat: „innerhalb“, „Alle Datensätze verwerfen, die nicht verknüpft werden konnten“, evtl. hinzuzufügende Felder einschränken und zusammengefassten Layer angeben

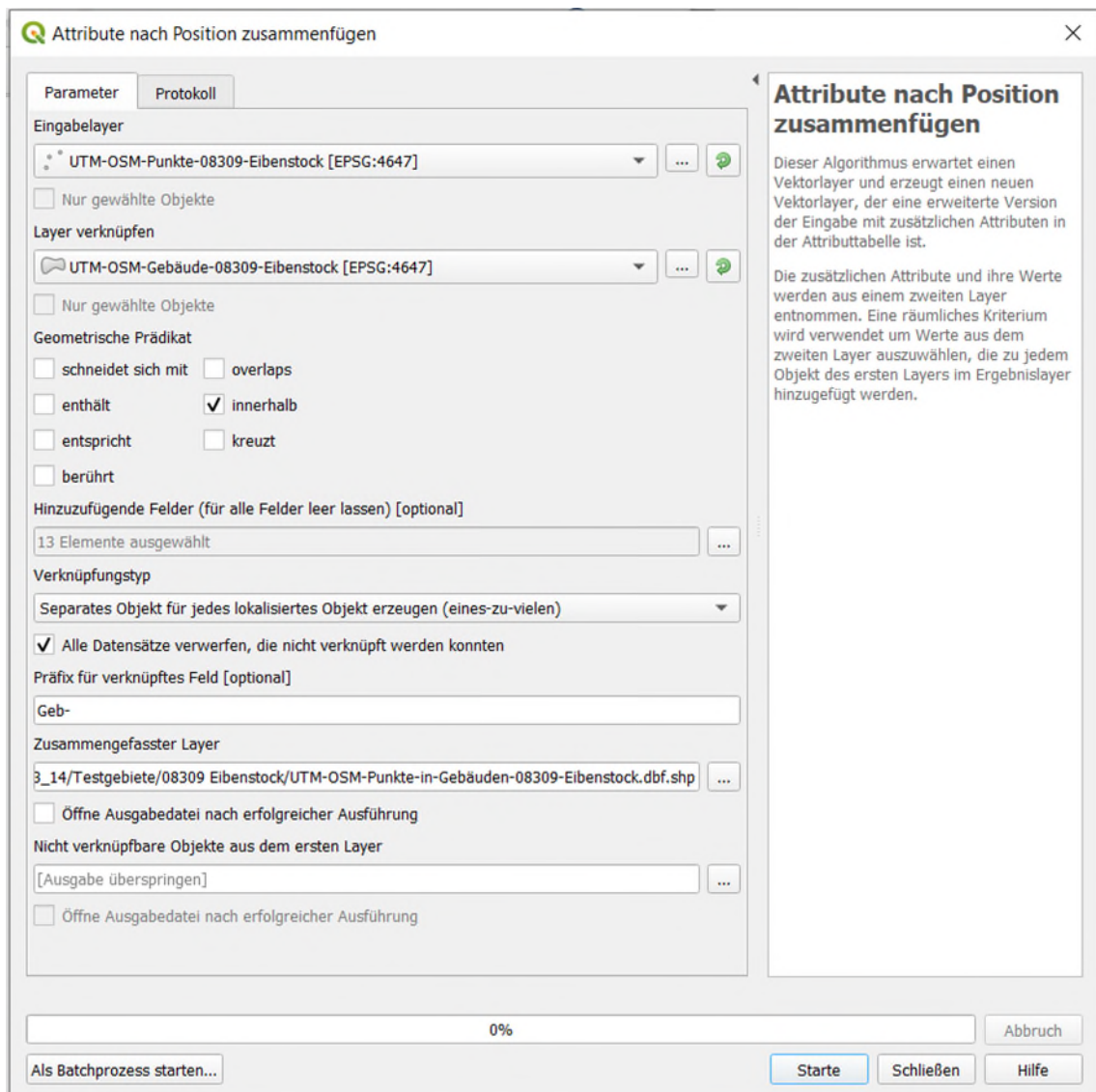


Abbildung 20: OSM-Punkte-in-Gebäuden-Layer erzeugen.

- HK-Adressen zum HU-Layer hinzufügen
 - „Verarbeitungswerkzeuge“-Fenster
 - „Vektoren allgemein“
 - „Attribute nach Position zusammenfügen“

- Eingabelayer (HU), verknüpfter Layer (HK), Geometrisches Prädikat: „enthält“ und zusammengefassten Layer angeben
- Attribut Fläche für OSM-Gebäudelayer und HU-Layer erzeugen
 - Rechtsklick auf Layer
 - „Attributtabelle öffnen“
 - „Feldrechner öffnen“
 - Ausgabefeldname und Ausgabefeldtyp angeben, Ausdruck: \$area (in der Liste unter Geometrie)

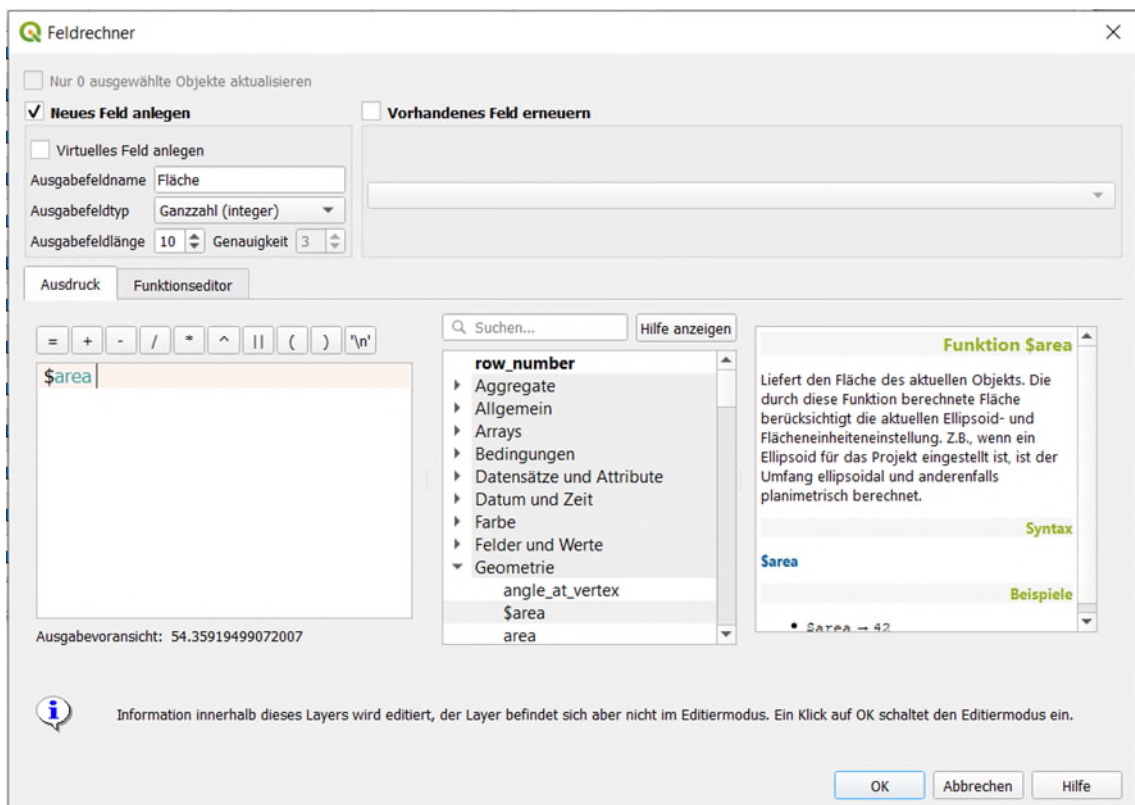


Abbildung 21: Attribut Fläche erzeugen.

- Attribut Anzahl Gebäudestützpunkte für OSM-Gebäudelayer und HU-Layer erzeugen
 - Rechtsklick auf Layer
 - „Attributtabelle öffnen“
 - Ausgabefeldname und Ausgabefeldtyp angeben, Ausdruck: num_points(\$geometry) (beides in der Liste unter Geometrie)
- Schnittmenge HU- und OSM-Gebäudelayer erzeugen
 - „Vektor“

- „Geoverarbeitungswerkzeuge“
 - „Verschneidung“
 - Eingabelayer, überlagerter Layer, zu übernehmende Attribute einschränken und Ausgabelayer angeben
- Vereinigungsmenge HU- und OSM-Gebäudelayer erzeugen
 - „Vektor“
 - „Geoverarbeitungswerkzeuge“
 - „Vereinigungen“
 - Eingabelayer, überlagerter Layer und Ausgabelayer angeben
- Überflüssige Attribute im Vereinigungsmenge-Layer in der Attributtabelle löschen
- Schnittmenge- und Vereinigungsmenge-Layer nach OI auflösen:
 - „Verarbeitungswerkzeuge“-Fenster
 - „Vektorgeometrie“
 - „Auflösen“
 - Eingabe- und Ausgabelayer und aufzulösendes Feld angeben
- Attribut Fläche für die aufgelösten und nicht-aufgelösten Schnittmenge- und Vereinigungsmenge-Layer erzeugen
 - Rechtsklick auf Layer
 - „Attributtabelle öffnen“
 - „Feldrechner öffnen“
 - Ausgabefeldname und Ausgabefeldtyp angeben, Ausdruck: \$area (in der Liste unter Geometrie)
- Die Layer der Testgebiete nach Kategorien zu deutschlandweiten Layern zusammenfügen
 - „Vektorverarbeitungswerkzeuge“-Fenster
 - „Vektoren allgemein“
 - „Vektorlayer zusammenführen“
 - Eingabelayer und Koordinatensystem auswählen und Ausgabelayer angeben

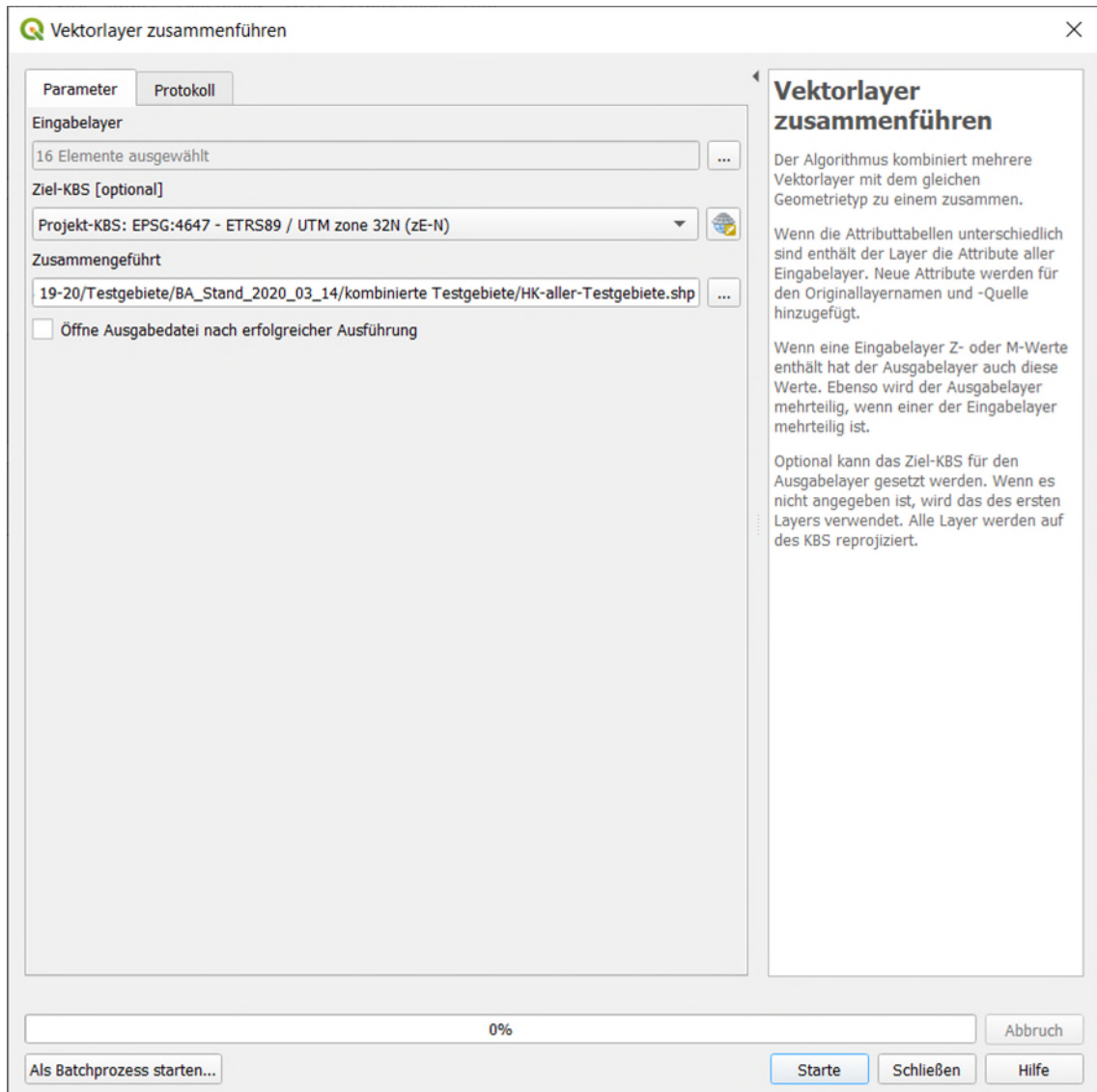


Abbildung 22: Layer zu einem deutschlandweiten Layer zusammenfügen.

- Attribute mit unterschiedlichen Namen aber gleicher Bedeutung zusammenfügen:
 - Rechtsklick auf Layer
 - „Attributtabelle öffnen“
 - „Feldrechner öffnen“
 - Ausgabefeldname und Ausgabefeldtyp angeben, Ausdruck: case when „attribut 1“ is not null then „attribut 1“ when „attribut 2“ is not null then „attribut 2“ end

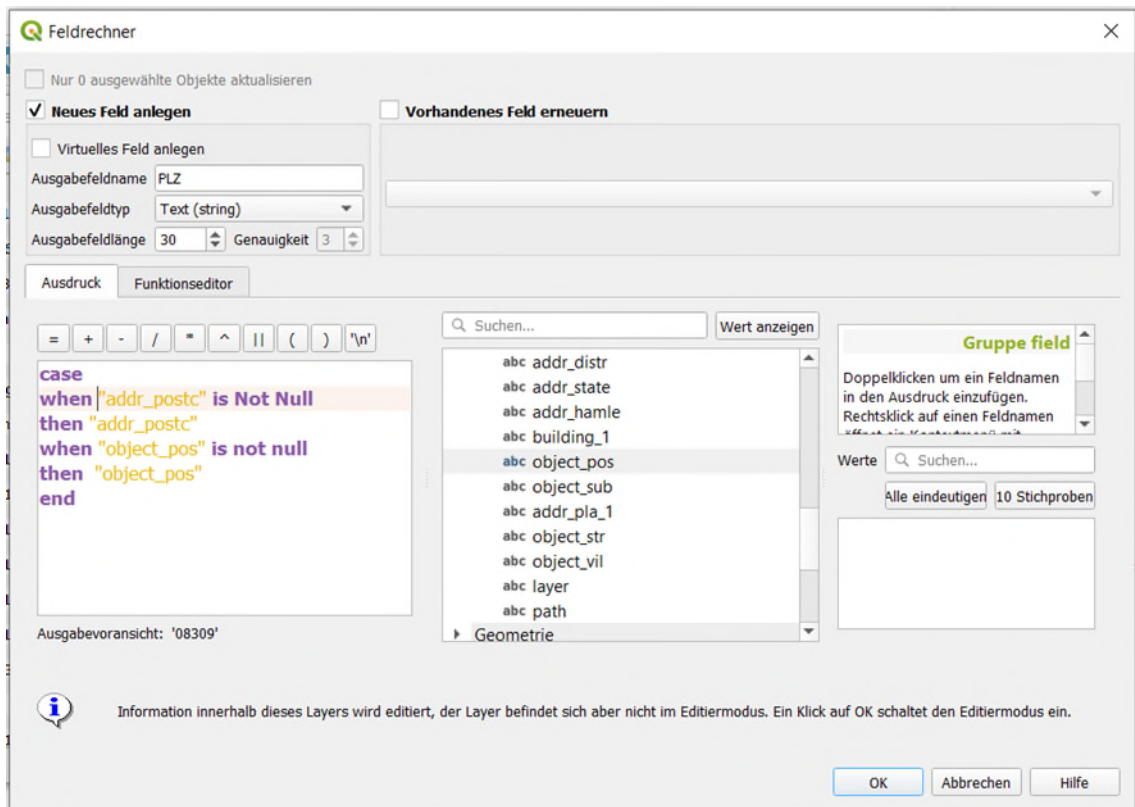


Abbildung 23: Attribute zusammenfassen.

- HU-OI für UTM-OSM-Gebäudelayer-aller-Testgebiete übernehmen:
 - „Vektorverarbeitungswerkzeuge“-Fenster
 - „Vektoren allgemein“
 - „Attribut nach Feldwert“
 - Eingabelayer: „UTM-OSM-Gebäudelayer-aller-Testgebiete“, Tabellenspalte: „full_id“, Eingabelayer 2: „Schnittmenge-Gebäudelayer“, Tabellenspalte „full_id“, zu kopierendes Feld Layer 2: „oi“, Verknüpfungstyp und zusammengefasster Layer angeben
- Layer Gebäude-ohne-Gegenstück erzeugen (für HU und für OSM):
 - „Vektorverarbeitungswerkzeuge“-Fenster
 - „Vektoren allgemein“
 - „Attribut nach Feldwert“
 - Eingabelayer: „UTM-OSM-Gebäudelayer-aller-Testgebiete“ / „HU-aller-Testgebiete“, Tabellenspalte: „full_id“ / „oi“, Eingabelayer 2: „Schnittmenge-Gebäudelayer“, Tabellenspalte „full_id“ / „oi“,

zusammengefasster Layer und Layer mit nichtverknüpfbaren Objekten aus dem ersten Layer angeben

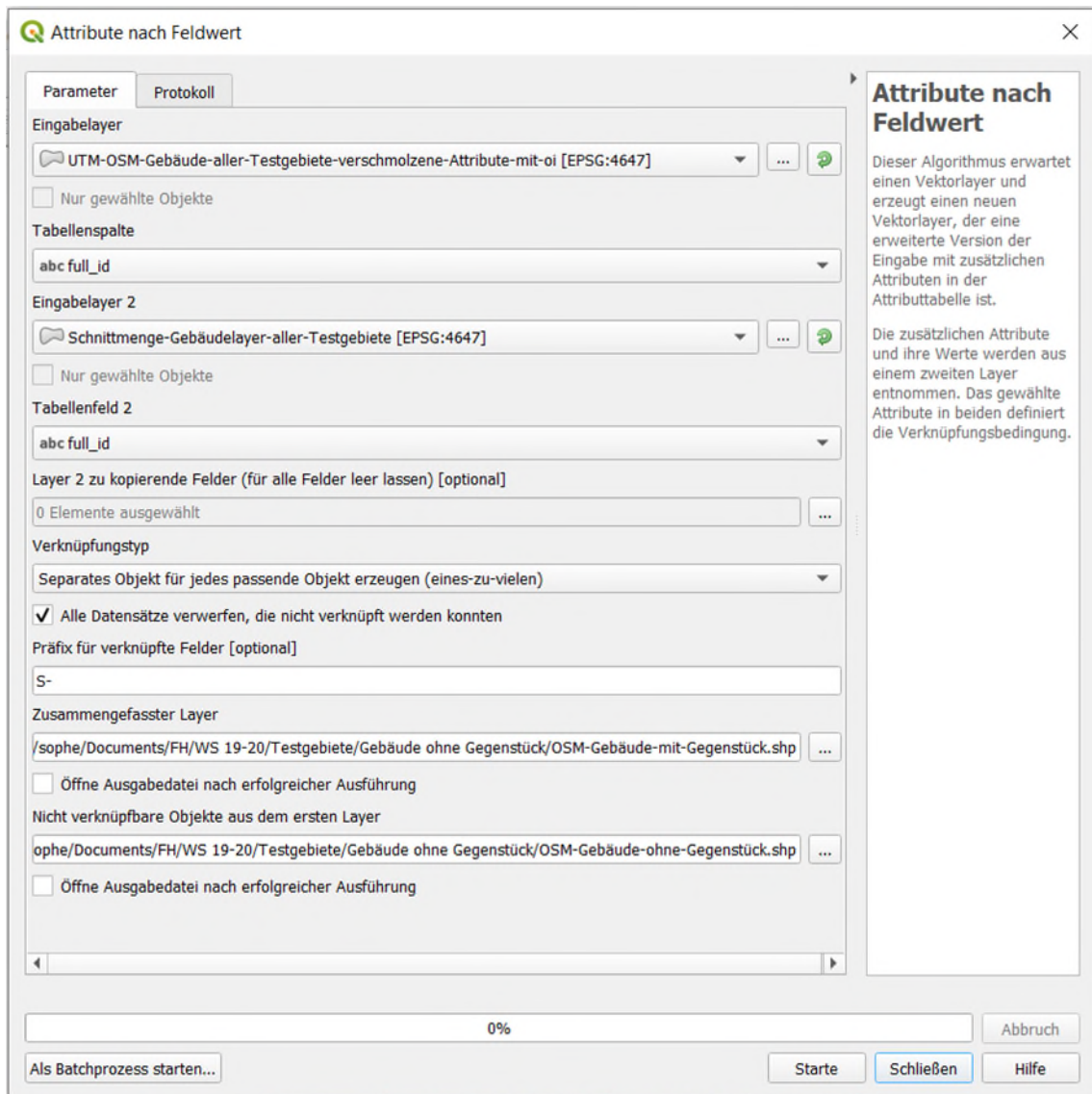


Abbildung 24: einen Layer für Gebäude ohne Gegenstück im anderen Datensatz erzeugen.

Tableau:

- Tableau Desktop öffnen und neues Projekt erstellen und speichern
- Verbindungen zu den verschiedenen Daten herstellen
- Verknüpfungs-Bedingungen für Verknüpfungen der Daten angeben

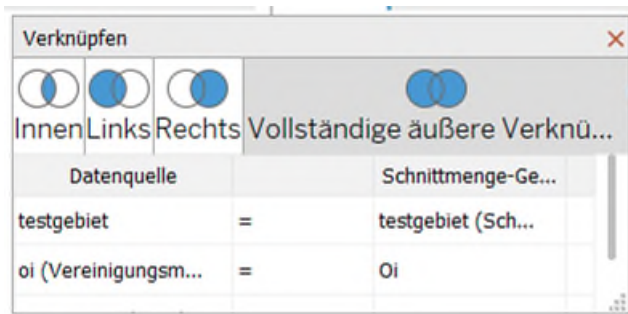


Abbildung 25: Tableau Beispiel einer Verknüpfungsbedingung.

- Hierarchie von Gebietart und Testgebiet erstellen:
 - Im Bereich Daten das Feld Testgebiet auf das Feld Gebietart
 - Namen der Hierarchie angeben
- Intersection over Union (Jaccard-Koeffizient) berechnen:
 - „Analyse“
 - „berechnetes Feld erstellen ...“
 - Name und Formel angeben



Abbildung 26: IoU in Tableau berechnen.

- Intersection over Union in Bereiche gruppieren:
 - Im Bereich Daten Rechtsklick aufs Feld Intersection over Union
 - „Erstellen“
 - „Gruppieren ...“
 - Mit der Option Suchen die gewünschten Elemente suchen und daraus Gruppen erstellen

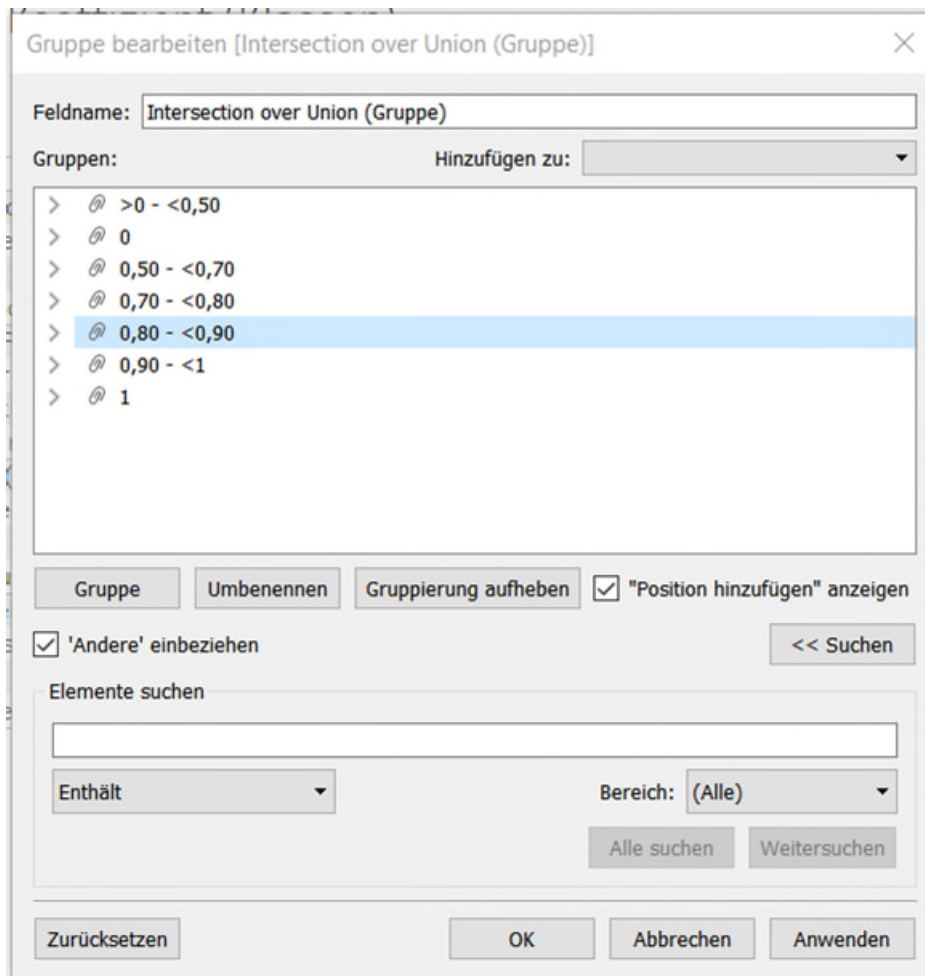


Abbildung 27: Gruppierung Jaccard-Koeffizient.

- Häufigkeit ausgefüllter Attribute:
 - Im Bereich Daten Rechtsklick auf das gewünschte Attribut
 - „Erstellen“
 - „Gruppieren ...“
 - 0 oder Null anklicken und eine Gruppe „nicht ausgefüllt“ erstellen
 - „Andere einbeziehen“ anklicken und die Gruppe in „ausgefüllt“ umbenennen
- Übereinstimmende Adressen abfragen:
 - „Analyse“
 - „berechnetes Feld erstellen ...“
 - Für jedes vergleichbare Adress-Attribut Name und Formel angeben
 - Im Bereich Daten Rechtsklick auf die berechneten Felder
 - „Erstellen“

- „Gruppieren ...“
- Gruppe für „keine Übereinstimmung“ und Gruppe für „übereinstimmendes Attribut“ erstellen

```
IF [HK-field14] = [Stasse] THEN "übereinstimmende Straße"  
ELSE "keine Übereinstimmung" END
```

Die Berechnung ist gültig.

Abbildung 28: Abfrage übereinstimmender Straßennamen.

- Felder und Gruppen auf Zeilen, Spalten, ins Markierungs- und / oder Filterfenster ziehen, um gewünschte Tabellen oder Grafiken zu erzeugen

ANHANG 2: ERGEBNISTABELLEN

Tabelle 12: OSM-Punkte die in OSM-Gebäuden fallen.

Testgebiete / Gebietart	alle OSM-Punkte	OSM-Punkte in-Gebäuden	Anteil der OSM-Punkte in-Gebäuden an allen OSM-Punkten [%]
gebirgige Gebiete			
87541 Bad Hindelang	2391	166	6,94
38877 Benneckenstein (Harz)	806	20	2,48
08309 Eibenstock	1513	234	15,47
54608 Bleialf	485	12	2,47
Großstädte und Ballungsräume			
45889 Gelsenkirchen (Bismarck)	765	195	25,49
22305 Hamburg (Brambek)	6026	689	11,43
10243 Berlin (Friedrichshain)	6193	2406	38,85
60311 Frankfurt am Main (Altstadt)	2403	706	29,38
ländliche Gebiete			
88279 Amtzell	1018	144	14,15
15936 Dahme / Mark	1831	26	1,42
99831 Creuzburg	411	20	4,87
29491 Prezelle	111	11	9,91
mittelgroße Städte			
18507 Grimmen	2122	859	40,48
66663 Merzig	2205	232	10,52
26655 Westerstede	9711	3978	40,96
84307 Eggenfelden	1125	197	17,51

Tabelle 13: Mittel, Median, Minimum und Maximum Gebäudeflächenvergleich HU und OSM-Gebäude und Überdeckungsfläche.

Testgebiete / Gebietart	Fläche HU [m ²]			Fläche OSM-Gebäude [m ²]			Fläche Überdeckung (Schnittmenge) [m ²]					
	Mittel	Median	Maximum	Mittel	Median	Maximum	Mittel	Median	Maximum			
gebirgige Gebiete	143	96	0	5859	256	158	1	7865	114	77	0	5831
87541 Bad Hindelang	123	71	2	5616	244	187	4	5774	99	57	0	5325
38877												
Benneckenstein (Harz)	151	111	5	5859	126	95	7	5919	114	82	0	5831
08309 Eibenstock	176	113	0	5565	220	140	1	7865	133	91	0	5091
54608 Bleialf	120	83	1	2491	413	247	12	3681	108	74	0	2436
Großstädte und Ballungsräume	468	173	0	25394	421	165	0	35806	216	79	0	25394
45889 Gelsenkirchen (Bismarck)	126	60	1	25394	159	61	1	25413	107	47	0	25394
22305 Hamburg (Brambek)	874	543	0	13821	423	186	0	17586	399	232	0	13763
10243 Berlin (Friedrichshain)	784	419	0	20236	851	423	1	35806	497	271	0	19629
60311 Frankfurt am Main (Altstadt)	372	170	0	7960	521	232	1	12869	172	72	0	7779
ländliche Gebiete	162	90	0	9846	311	167	0	25216	130	73	0	9846
88279 Amtzell	165	89	0	9846	244	142	0	10366	142	82	0	9846
15936 Dahme / Mark	166	99	0	8305	354	204	7	9457	128	73	0	8052
99831 Creuzburg	123	64	0	9458	324	120	1	25216	105	49	0	9128
29491 Prezelle	186	143	2	1104	365	277	22	2201	138	109	0	1062
mittelgroße Städte	181	90	0	93260	272	154	1	46084	123	72	0	78924
18507 Grimmen	174	73	1	15242	217	99	1	9645	132	63	0	7480
66663 Merzig	215	104	0	93260	206	134	5	46084	125	84	0	78924
26655 Westerstede	151	77	0	30529	289	165	1	32705	112	57	0	30427
84307 Eggenfelden	216	93	1	12743	384	212	5	16212	153	79	0	11876

Tabelle 14: Mittel, Median, Minimum und Maximum Gebäudepunktevergleich HU und OSM-Gebäude.

Testgebiete / Gebietart	Anzahl Gebäudepunkte HU				Anzahl Gebäudepunkte OSM-Gebäude			
	Mittel	Median	Minimum	Maximum	Mittel	Median	Minimum	Maximum
gebirgige Gebiete	9	7	4	255	7	5	4	90
87541 Bad Hindelang	9	7	4	255	7	5	4	43
38877 Benneckenstein (Harz)	10	9	5	71	6	5	4	35
08309 Eibenstock	10	8	4	168	6	5	4	90
54608 Bleialf	8	6	4	97	8	7	5	31
Großstädte und Ballungsräume	19	9	4	1709	10	7	4	182
45889 Gelsenkirchen (Bismarck)	8	6	4	172	8	6	4	159
22305 Hamburg (Brambek)	27	13	4	500	10	7	4	136
10243 Berlin (Friedrichshain)	24	16	4	229	15	12	4	182
60311 Frankfurt am Main (Altstadt)	27	10	4	1709	11	9	4	102
ländliche Gebiete	8	7	4	227	7	5	4	28
88279 Amtzell	9	7	4	227	7	6	4	25
15936 Dahme / Mark	7	6	4	155	7	5	4	28
99831 Creuzburg	7	6	4	73	6	5	4	25
29491 Prezelle	8	6	5	65	6	5	5	27
mittelgroße Städte	9	7	4	271	7	5	4	11
18507 Grimmen	8	6	4	121	7	5	5	111
66663 Merzig	9	7	4	208	6	5	4	67
26655 Westerstede	9	7	4	271	8	6	4	61
84307 Eggenfelden	9	7	4	137	8	7	4	80

Tabelle 15: Jaccard-Koeffizient Mittelwert, Median.

Testgebiete / Gebietart	Jaccard-Koeffizient	
	Mittelwert	Median
gebirgige Gebiete	0,8413	0,9155
87541 Bad Hindelang	0,8351	0,9302
38877 Benneckenstein (Harz)	0,8359	0,8817
08309 Eibenstock	0,8362	0,9063
54608 Bleialf	0,8588	0,9302
Großstädte und Ballungsräume	0,8713	0,9483
45889 Gelsenkirchen (Bismarck)	0,8484	0,9149
22305 Hamburg (Brambek)	0,9513	0,9782
10243 Berlin (Friedrichshain)	0,9020	0,9947
60311 Frankfurt am Main (Altstadt)	0,8597	0,9769
ländliche Gebiete	0,7800	0,8910
88279 Amtzell	0,8774	0,9586
15936 Dahme / Mark	0,7266	0,8241
99831 Creuzburg	0,7307	0,8986
29491 Prezelle	0,7243	0,8096
mittelgroße Städte	0,7768	0,8701
18507 Grimmen	0,8802	0,9600
66663 Merzig	0,7491	0,8381
26655 Westerstede	0,7496	0,8400
84307 Eggenfelden	0,8398	0,9214

Tabelle 16: Jaccard-Koeffizient absolute Häufigkeit der Toleranzklassen.

Testgebiete / Gebietart	Toleranzklassen Jaccard-Koeffizient (absolute Häufigkeit)							
	0	>0,00 - <0,50	0,50 - <0,70	0,70 - <0,80	0,80 - <0,90	0,90 - <1,00	1	
gebirgige Gebiete	13358	107	958	919	1260	2783	4817	2514
87541 Bad Hindelang	4486	27	434	330	389	726	1378	1202
38877								
Benneckenstein (Harz)	1531	12	61	129	209	450	565	105
08309 Eibenstock	4107	28	274	302	417	946	1650	490
54608 Bleialf	3234	40	189	158	245	661	1224	717
Großstädte und Ballungsräume	11618	192	639	390	632	1682	6267	1816
45889 Gelsenkirchen (Bismarck)	6059	59	338	263	556	1386	3002	455
22305 Hamburg (Brambek)	1359	10	13	10	9	60	1123	134
10243 Berlin (Friedrichshain)	1848	84	70	13	12	66	1216	387
60311 Frankfurt am Main (Altstadt)	2352	39	218	104	55	170	926	840
ländliche Gebiete	11144	206	1440	986	1112	2049	3600	1751
88279 Amtzell	3908	51	218	164	190	528	1797	960
15936 Dahme / Mark	5353	99	869	676	779	1214	1222	494
99831 Creuzburg	1565	53	302	95	98	238	516	263
29491 Prezelle	318	3	51	51	45	69	65	34
mittelgroße Städte	57818	981	6745	5961	6911	12696	17651	6873
18507 Grimmen	5939	117	276	232	259	781	3057	1217
66663 Merzig	15327	303	1891	1782	2295	4067	4058	931
26655 Westerstede	27646	464	3888	3546	3698	6002	6665	3383
84307 Eggenfelden	8906	97	690	401	659	1846	3871	1342

Tabelle 17: Jaccard-Koeffizient relative Häufigkeit der Toleranzklassen (Prozent).

Testgebiete / Gebietart	Toleranzklassen Jaccard-Koeffizient (relative Häufigkeit [%])							
	Summe	0 >0,00 - <0,50	0,50 - <0,70	0,70 - <0,80	0,80 - <0,90	0,90 - <1,00	1	
gebirgige Gebiete	100	0,80	7,17	6,88	9,43	20,83	36,06	18,82
87541 Bad Hindelang	100	0,60	9,67	7,36	8,67	16,18	30,72	26,79
38877								
Benneckenstein (Harz)	100	0,78	3,98	8,43	13,65	29,39	36,90	6,86
08309 Eibenstock	100	0,68	6,67	7,35	10,15	23,03	40,18	11,93
54608 Bleialf	100	1,24	5,84	4,89	7,58	20,44	37,85	22,17
Großstädte und Ballungsräume	100	1,65	5,50	3,36	5,44	14,48	53,94	15,63
45889 Gelsenkirchen (Bismarck)	100	0,97	5,58	4,34	9,18	22,88	49,55	7,51
22305 Hamburg (Brambek)	100	0,74	0,96	0,74	0,66	4,42	82,63	9,86
10243 Berlin (Friedrichshain)	100	4,55	3,79	0,70	0,65	3,57	65,80	20,94
60311 Frankfurt am Main (Altstadt)	100	1,66	9,27	4,42	2,34	7,23	39,37	35,71
ländliche Gebiete	100	1,85	12,92	8,85	9,98	18,39	32,30	15,71
88279 Amtzell	100	1,31	5,58	4,20	4,86	13,51	45,98	24,56
15936 Dahme / Mark	100	1,85	16,23	12,63	14,55	22,68	22,83	9,23
99831 Creuzburg	100	3,39	19,30	6,07	6,26	15,21	32,97	16,81
29491 Prezelle	100	0,94	16,04	16,04	14,15	21,70	20,44	10,69
mittelgroße Städte	100	1,70	11,67	10,31	11,95	21,96	30,53	11,89
18507 Grimmen	100	1,97	4,65	3,91	4,36	13,15	51,47	20,49
66663 Merzig	100	1,98	12,34	11,63	14,97	26,53	26,48	6,07
26655 Westerstede	100	1,68	14,06	12,83	13,38	21,71	24,11	12,24
84307 Eggenfelden	100	1,09	7,75	4,50	7,40	20,73	43,47	15,07

Tabelle 18: Toleranzklassen Jaccard-Koeffizient CDF.

Testgebiete / Gebietart	Toleranzklassen Jaccard-Koeffizient (CDF [%])						
	0	>0,00 - <0,50	0,50 - <0,70	0,70 - <0,80	0,80 - <0,90	0,90 - <1,00	1
gebirgige Gebiete	0,80	7,97	14,85	24,29	45,12	81,18	100,00
87541 Bad Hindelang	0,60	10,28	17,63	26,30	42,49	73,21	100,00
38877							
Benneckenstein (Harz)	0,78	4,77	13,19	26,85	56,24	93,14	100,00
08309 Eibenstock	0,68	7,35	14,71	24,86	47,89	88,07	100,00
54608 Bleialf	1,24	7,08	11,97	19,54	39,98	77,83	100,00
Großstädte und Ballungsräume	1,65	7,15	10,51	15,95	30,43	84,37	100,00
45889 Gelsenkirchen (Bismarck)	0,97	6,55	10,89	20,07	42,94	92,49	100,00
22305 Hamburg (Brambek)	0,74	1,69	2,43	3,09	7,51	90,14	100,00
10243 Berlin (Friedrichshain)	4,55	8,33	9,04	9,69	13,26	79,06	100,00
60311 Frankfurt am Main (Altstadt)	1,66	10,93	15,35	17,69	24,91	64,29	100,00
ländliche Gebiete	1,85	14,77	23,62	33,60	51,98	84,29	100,00
88279 Amtzell	1,31	6,88	11,08	15,94	29,45	75,44	100,00
15936 Dahme / Mark	1,85	18,08	30,71	45,26	67,94	90,77	100,00
99831 Creuzburg	3,39	22,68	28,75	35,02	50,22	83,19	100,00
29491 Prezelle	0,94	16,98	33,02	47,17	68,87	89,31	100,00
mittelgroße Städte	1,70	13,36	23,67	35,63	57,58	88,11	100,00
18507 Grimmen	1,97	6,62	10,52	14,88	28,04	79,51	100,00
66663 Merzig	1,98	14,31	25,94	40,91	67,45	93,93	100,00
26655 Westerstede	1,68	15,74	28,57	41,94	63,65	87,76	100,00
84307 Eggenfelden	1,09	8,84	13,34	20,74	41,47	84,93	100,00

Tabelle 19: Jaccard-Koeffizienten der Testgebiete.

Testgebiete / Gebietart	Vereinigungsmenge- Gebäudelayer aufgelöst [m ²]	Schnittmenge- Gebäudelayer aufgelöst [m ²]	Jaccard-Koeffizient ganzes Testgebiet
gebirgige Gebiete	1840399	1516858	0,8242
87541 Bad Hindelang	535745	445164	0,8309
38877 Benneckenstein (Harz)	206442	175016	0,8478
08309 Eibenstock	685929	547025	0,7975
54608 Bleialf	412283	349653	0,8481
Großstädte und Ballungsräume	2722774	2517346	0,9246
45889 Gelsenkirchen (Bismarck)	732212	647322	0,8841
22305 Hamburg (Brambek)	575336	542544	0,9430
10243 Berlin (Friedrichshain)	963435	919395	0,9543
60311 Frankfurt am Main (Altstadt)	451791	408085	0,9033
ländliche Gebiete	2636258	1446238	0,5486
88279 Amtzell	639304	554279	0,8670
15936 Dahme / Mark	1597540	683726	0,4280
99831 Creuzburg	270268	164212	0,6076
29491 Prezelle	129146	44021	0,3409
mittelgroße Städte	10941236	7133057	0,6519
18507 Grimmen	1083579	781224	0,7210
66663 Merzig	2476002	1911635	0,7721
26655 Westerstede	5670204	3081808	0,5435
84307 Eggenfelden	1711452	1358390	0,7937

Tabelle 20: Toleranzklassen Gebäude absolute Häufigkeit.

Testgebiete / Gebietart	Toleranzklassen Gebäude (absolute Häufigkeit)		
	nicht das gleiche Gebäude	möglicherweise das gleiche Gebäude	das gleiches Gebäude
gebirgige Gebiete	1065	919	11374
87541 Bad Hindelang	461	330	3695
38877 Benneckenstein (Harz)	73	129	1329
08309 Eibenstock	302	302	3503
54608 Bleialf	229	158	2847
Großstädte und Ballungsräume	831	390	10397
45889 Gelsenkirchen (Bismarck)	397	263	5399
22305 Hamburg (Brambek)	23	10	1326
10243 Berlin (Friedrichshain)	154	13	1681
60311 Frankfurt am Main (Altstadt)	257	104	1991
ländliche Gebiete	1646	986	8512
88279 Amtzell	269	164	3475
15936 Dahme / Mark	968	676	3709
99831 Creuzburg	355	95	1115
29491 Prezelle	54	51	213
mittelgroße Städte	7726	5961	44131
18507 Grimmen	393	232	5314
66663 Merzig	2194	1782	11351
26655 Westerstede	4352	3546	19748
84307 Eggenfelden	787	401	7718

Tabelle 21: Toleranzklassen Gebäude relative Häufigkeit in Prozent.

Testgebiete / Gebietart	Toleranzklassen Gebäude (relative Häufigkeit [%])		
	nicht das gleiche Gebäude	möglicherweise das gleiche Gebäude	das gleiches Gebäude
gebirgige Gebiete	7,97	6,88	85,15
87541 Bad Hindelang	10,28	7,36	82,37
38877 Benneckenstein (Harz)	4,77	8,43	86,81
08309 Eibenstock	7,35	7,35	85,29
54608 Bleialf	7,08	4,89	88,03
Großstädte und Ballungsräume	7,15	3,36	89,49
45889 Gelsenkirchen (Bismarck)	6,55	4,34	89,11
22305 Hamburg (Brambek)	1,69	0,74	97,57
10243 Berlin (Friedrichshain)	8,33	0,70	90,96
60311 Frankfurt am Main (Altstadt)	10,93	4,42	84,65
ländliche Gebiete	14,77	8,85	76,38
88279 Amtzell	6,88	4,20	88,92
15936 Dahme / Mark	18,08	12,63	69,29
99831 Creuzburg	22,68	6,07	71,25
29491 Prezelle	16,98	16,04	66,98
mittelgroße Städte	13,36	10,31	76,33
18507 Grimmen	6,62	3,91	89,48
66663 Merzig	14,31	11,63	74,06
26655 Westerstede	15,74	12,83	71,43
84307 Eggenfelden	8,84	4,50	86,66

Tabelle 22: Vergleich Gebäudeanzahl OSM und HU.

Testgebiete / Gebietart	OSM-Gebäude	HU	Differenz HU und OSM-Gebäude
gebirgige Gebiete			
87541 Bad Hindelang	3594	5561	1967
38877 Benneckenstein (Harz)	1769	1746	-23
08309 Eibenstock	4413	5820	1407
54608 Bleialf	1631	3746	2115
Großstädte und Ballungsräume			
45889 Gelsenkirchen (Bismarck)	5736	7036	1300
22305 Hamburg (Brambek)	1750	1570	-180
10243 Berlin (Friedrichshain)	1741	2146	405
60311 Frankfurt am Main (Altstadt)	1379	2566	1187
ländliche Gebiete			
88279 Amtzell	3185	4569	1384
15936 Dahme / Mark	3125	14802	11677
99831 Creuzburg	894	3021	2127
29491 Prezelle	190	1128	938
mittelgroße Städte			
18507 Grimmen	5160	8152	2992
66663 Merzig	13264	19367	6103
26655 Westerstede	17715	53355	35640
84307 Eggenfelden	5421	11489	6068

Tabelle 23: Gebäude ohne Gegenstück im anderen Datensatz.

Testgebiete / Gebietart	OSM- Gebäude	HU	OSM- Gebäude ohne Gegenstück	HU ohne Gegenstück	relative Häufigkeit OSM- Gebäude ohne Gegenstück [%]	relative Häufigkeit HU ohne Gegenstück [%]
gebirgige Gebiete						
87541 Bad Hindelang	3594	5561	352	1075	9,79	19,33
38877 Benneckenstein (Harz)	1769	1746	114	215	6,44	12,31
08309 Eibenstock	4413	5820	707	1712	16,02	29,42
54608 Bleialf	1631	3746	152	512	9,32	13,67
Großstädte und Ballungsräume						
45889 Gelsenkirchen (Bismarck)	5736	7036	88	977	1,53	13,89
22305 Hamburg (Brambek)	1750	1570	29	210	1,66	13,38
10243 Berlin (Friedrichshain)	1741	2146	122	297	7,01	13,84
60311 Frankfurt am Main (Altstadt)	1379	2566	52	195	3,77	7,60
ländliche Gebiete						
88279 Amtzell	3185	4569	239	660	7,50	14,45
15936 Dahme / Mark	3125	14802	119	9448	3,81	63,83
99831 Creuzburg	894	3021	37	1451	4,14	48,03
29491 Prezelle	190	1128	2	810	1,05	71,81
mittelgroße Städte						
18507 Grimmen	5160	8152	193	2213	3,74	27,15
66663 Merzig	13264	19367	1254	4036	9,45	20,84
26655 Westerstede	17715	53355	1095	25705	6,18	48,18
84307 Eggenfelden	5421	11489	65	2583	1,20	22,48

Tabelle 24: Ausgefüllte HK-Attribute absolute Häufigkeit in gebirgigen Gebieten.

Attribut	gebirgige Gebiete			
	87541 Bad Hindelang	38877 Benneckenstein (Harz)	08309 Eibenstock	54608 Bleielf
HK				
Objekte gesamt	1617	765	2990	1378
abs. Häufigkeit Datenelement 1 (Kennung des Datensatzes)	1617	765	2990	1378
abs. Häufigkeit Datenelement 2 (OI)	1617	765	2990	1378
abs. Häufigkeit Datenelement 3 (Qualität der georef. Adresse)	1617	765	2990	1378
abs. Häufigkeit Datenelement 4 (Schlüssel Land)	1617	765	2990	1378
abs. Häufigkeit Datenelement 5 (Schlüssel Regierungsbezirk)	1617	0	2990	1378
abs. Häufigkeit Datenelement 6 (Schlüssel Kreis/kreisfreie Stadt)	1617	765	2990	1378
abs. Häufigkeit Datenelement 7 (Schlüssel Gemeinde)	1617	765	2990	1378
abs. Häufigkeit Datenelement 8 (Schlüssel Orts- /Gemeindeteil)	1617	0	2402	0
abs. Häufigkeit Datenelement 9 (Schlüssel Straße)	0	765	2990	1378
abs. Häufigkeit Datenelement 10 (Hausnummer)	1617	765	2990	1378
abs. Häufigkeit Datenelement 11 (Adressierungszusatz)	167	109	335	48
abs. Häufigkeit Datenelement 12 (Koordinaten-Ost-Wert)	1617	765	2990	1378
abs. Häufigkeit Datenelement 13 (Koordinaten-Nord-Wert)	1617	765	2990	1378
abs. Häufigkeit Datenelement 14 (Straßenname)	1617	765	2990	1378
abs. Häufigkeit Datenelement 15 (Postleitzahl)	1617	765	2990	1378
abs. Häufigkeit Datenelement 16 (Postalischer Ortsname)	1617	765	2990	1378
abs. Häufigkeit Datenelement 17 (Zusatz zum postalischen Ortsnamen)	13	765	498	495
abs. Häufigkeit Datenelement 18 (Postalischer Ortsteil)	1617	0	2401	165

Tabelle 25: Ausgefüllte HK-Attribute relative Häufigkeit in Prozent in gebirgigen Gebieten.

Attribut	gebirgige Gebiete			
	87541 Bad Hindelang	38877 Benneckenstein (Harz)	08309 Eibenstock	54608 Bleiälf
HK				
Objekte gesamt	1617	765	2990	1378
rel. Häufigkeit [%] Datenelement 1 (Kennung des Datensatzes)	100,0	100,0	100,0	100,0
rel. Häufigkeit [%] Datenelement 2 (OI)	100,0	100,0	100,0	100,0
rel. Häufigkeit [%] Datenelement 3 (Qualität der georef. Adresse)	100,0	100,0	100,0	100,0
rel. Häufigkeit [%] Datenelement 4 (Schlüssel Land)	100,0	100,0	100,0	100,0
rel. Häufigkeit [%] Datenelement 5 (Schlüssel Regierungsbezirk)	100,0	0,0	100,0	100,0
rel. Häufigkeit [%] Datenelement 6 (Schlüssel Kreis/kreisfreie Stadt)	100,0	100,0	100,0	100,0
rel. Häufigkeit [%] Datenelement 7 (Schlüssel Gemeinde)	100,0	100,0	100,0	100,0
rel. Häufigkeit [%] Datenelement 8 (Schlüssel Orts-/Gemeindeteil)	100,0	0,0	80,3	0,0
rel. Häufigkeit [%] Datenelement 9 (Schlüssel Straße)	0,0	100,0	100,0	100,0
rel. Häufigkeit [%] Datenelement 10 (Hausnummer)	100,0	100,0	100,0	100,0
rel. Häufigkeit [%] Datenelement 11 (Adressierungszusatz)	10,3	14,2	11,2	3,5
rel. Häufigkeit [%] Datenelement 12 (Koordinaten-Ost-Wert)	100,0	100,0	100,0	100,0
rel. Häufigkeit [%] Datenelement 13 (Koordinaten-Nord-Wert)	100,0	100,0	100,0	100,0
rel. Häufigkeit [%] Datenelement 14 (Straßenname)	100,0	100,0	100,0	100,0
rel. Häufigkeit [%] Datenelement 15 (Postleitzahl)	100,0	100,0	100,0	100,0
rel. Häufigkeit [%] Datenelement 16 (Postalischer Ortsname)	100,0	100,0	100,0	100,0
rel. Häufigkeit [%] Datenelement 17 (Zusatz zum postalischen Ortsnamen)	0,8	100,0	16,7	35,9
rel. Häufigkeit [%] Datenelement 18 (Postalischer Ortsteil)	100,0	0,0	80,3	12,0

Tabelle 26: Ausgefüllte HK-Attribute absolute Häufigkeit in Großstädten und Ballungsräumen.

Attribut	Großstädte und Ballungsräume			
	45889 Gelsenkirchen (Bismarck)	22305 Hamburg (Brambek)	10243 Berlin (Friedrichshain)	60311 Frankfurt am Main (Altstadt)
HK				
Objekte gesamt	2637	2085	2001	1189
abs. Häufigkeit Datenelement 1 (Kennung des Datensatzes)	2637	2085	2001	1189
abs. Häufigkeit Datenelement 2 (OI)	2637	2085	2001	1189
abs. Häufigkeit Datenelement 3 (Qualität der georef. Adresse)	2637	2085	2001	1189
abs. Häufigkeit Datenelement 4 (Schlüssel Land)	2637	2085	2001	1189
abs. Häufigkeit Datenelement 5 (Schlüssel Regierungsbezirk)	2637	2085	0	1189
abs. Häufigkeit Datenelement 6 (Schlüssel Kreis/kreisfreie Stadt)	2637	0	0	1189
abs. Häufigkeit Datenelement 7 (Schlüssel Gemeinde)	0	2085	2001	0
abs. Häufigkeit Datenelement 8 (Schlüssel Orts-/Gemeindeteil)	0	2085	2001	1189
abs. Häufigkeit Datenelement 9 (Schlüssel Straße)	2637	2085	2001	1189
abs. Häufigkeit Datenelement 10 (Hausnummer)	2637	2085	2001	1189
abs. Häufigkeit Datenelement 11 (Adressierungszusatz)	186	385	239	27
abs. Häufigkeit Datenelement 12 (Koordinaten-Ost-Wert)	2637	2085	2001	1189
abs. Häufigkeit Datenelement 13 (Koordinaten-Nord-Wert)	2637	2085	2001	1189
abs. Häufigkeit Datenelement 14 (Straßenname)	2637	2085	2001	1189
abs. Häufigkeit Datenelement 15 (Postleitzahl)	2637	2085	2001	1189
abs. Häufigkeit Datenelement 16 (Postalischer Ortsname)	2637	2085	2001	1189
abs. Häufigkeit Datenelement 17 (Zusatz zum postalischen Ortsnamen)	44	0	0	1189
abs. Häufigkeit Datenelement 18 (Postalischer Ortsteil)	2637	2085	2001	1189

Tabelle 27: Ausgefüllte HK-Attribute relative Häufigkeit in Prozent in Großstädten und Ballungsräumen.

Attribut	Großstädte und Ballungsräume			
	45889 Gelsenkirchen (Bismarck)	22305 Hamburg (Brambek)	10243 Berlin (Friedrichshain)	60311 Frankfurt am Main (Altstadt)
HK				
Objekte gesamt	2637	2085	2001	1189
rel. Häufigkeit [%] Datenelement 1 (Kennung des Datensatzes)	100,0	100,0	100,0	100,0
rel. Häufigkeit [%] Datenelement 2 (OI)	100,0	100,0	100,0	100,0
rel. Häufigkeit [%] Datenelement 3 (Qualität der georef. Adresse)	100,0	100,0	100,0	100,0
rel. Häufigkeit [%] Datenelement 4 (Schlüssel Land)	100,0	100,0	100,0	100,0
rel. Häufigkeit [%] Datenelement 5 (Schlüssel Regierungsbezirk)	100,0	100,0	0,0	100,0
rel. Häufigkeit [%] Datenelement 6 (Schlüssel Kreis/kreisfreie Stadt)	100,0	0,0	0,0	100,0
rel. Häufigkeit [%] Datenelement 7 (Schlüssel Gemeinde)	0,0	100,0	100,0	0,0
rel. Häufigkeit [%] Datenelement 8 (Schlüssel Orts-/Gemeindeteil)	0,0	100,0	100,0	100,0
rel. Häufigkeit [%] Datenelement 9 (Schlüssel Straße)	100,0	100,0	100,0	100,0
rel. Häufigkeit [%] Datenelement 10 (Hausnummer)	100,0	100,0	100,0	100,0
rel. Häufigkeit [%] Datenelement 11 (Adressierungszusatz)	7,1	18,5	11,9	2,3
rel. Häufigkeit [%] Datenelement 12 (Koordinaten-Ost-Wert)	100,0	100,0	100,0	100,0
rel. Häufigkeit [%] Datenelement 13 (Koordinaten-Nord-Wert)	100,0	100,0	100,0	100,0
rel. Häufigkeit [%] Datenelement 14 (Straßenname)	100,0	100,0	100,0	100,0
rel. Häufigkeit [%] Datenelement 15 (Postleitzahl)	100,0	100,0	100,0	100,0
rel. Häufigkeit [%] Datenelement 16 (Postalischer Ortsname)	100,0	100,0	100,0	100,0
rel. Häufigkeit [%] Datenelement 17 (Zusatz zum postalischen Ortsnamen)	1,7	0,0	0,0	100,0
rel. Häufigkeit [%] Datenelement 18 (Postalischer Ortsteil)	100,0	100,0	100,0	100,0

Tabelle 28: Ausgefüllte HK-Attribute absolute Häufigkeit in ländlichen Gebieten.

Attribut	ländliche Gebiete			
	88279 Amtzell	15936 Dahme / Mark	99831 Creuzburg	29491 Prezelle
HK				
Objekte gesamt	1648	3010	822	252
abs. Häufigkeit Datenelement 1 (Kennung des Datensatzes)	1648	3010	822	252
abs. Häufigkeit Datenelement 2 (OI)	1648	3010	822	252
abs. Häufigkeit Datenelement 3 (Qualität der georef. Adresse)	1648	3010	822	252
abs. Häufigkeit Datenelement 4 (Schlüssel Land)	1648	3010	822	252
abs. Häufigkeit Datenelement 5 (Schlüssel Regierungsbezirk)	1648	0	0	252
abs. Häufigkeit Datenelement 6 (Schlüssel Kreis/kreisfreie Stadt)	1648	3010	822	252
abs. Häufigkeit Datenelement 7 (Schlüssel Gemeinde)	1648	3010	822	252
abs. Häufigkeit Datenelement 8 (Schlüssel Orts-/Gemeindeteil)	0	3010	0	0
abs. Häufigkeit Datenelement 9 (Schlüssel Straße)	1648	3010	822	252
abs. Häufigkeit Datenelement 10 (Hausnummer)	1648	3010	822	252
abs. Häufigkeit Datenelement 11 (Adressierungszusatz)	129	252	80	10
abs. Häufigkeit Datenelement 12 (Koordinaten-Ost-Wert)	1648	3010	822	252
abs. Häufigkeit Datenelement 13 (Koordinaten-Nord-Wert)	1648	3010	822	252
abs. Häufigkeit Datenelement 14 (Straßenname)	1648	3010	822	252
abs. Häufigkeit Datenelement 15 (Postleitzahl)	1648	3010	822	252
abs. Häufigkeit Datenelement 16 (Postalischer Ortsname)	1648	3010	822	252
abs. Häufigkeit Datenelement 17 (Zusatz zum postalischen Ortsnamen)	304	2438	0	0
abs. Häufigkeit Datenelement 18 (Postalischer Ortsteil)	1592	3010	822	252

Tabelle 29: Ausgefüllte HK-Attribute relative Häufigkeit in Prozent in ländlichen Gebieten.

Attribut	ländliche Gebiete			
	88279 Amtzell	15936 Dahme / Mark	99831 Creuzburg	29491 Prezelle
HK				
Objekte gesamt	1648	3010	822	252
rel. Häufigkeit [%] Datenelement 1 (Kennung des Datensatzes)	100,0	100,0	100,0	100,0
rel. Häufigkeit [%] Datenelement 2 (OI)	100,0	100,0	100,0	100,0
rel. Häufigkeit [%] Datenelement 3 (Qualität der georef. Adresse)	100,0	100,0	100,0	100,0
rel. Häufigkeit [%] Datenelement 4 (Schlüssel Land)	100,0	100,0	100,0	100,0
rel. Häufigkeit [%] Datenelement 5 (Schlüssel Regierungsbezirk)	100,0	0,0	0,0	100,0
rel. Häufigkeit [%] Datenelement 6 (Schlüssel Kreis/kreisfreie Stadt)	100,0	100,0	100,0	100,0
rel. Häufigkeit [%] Datenelement 7 (Schlüssel Gemeinde)	100,0	100,0	100,0	100,0
rel. Häufigkeit [%] Datenelement 8 (Schlüssel Orts-/Gemeindeteil)	0,0	100,0	0,0	0,0
rel. Häufigkeit [%] Datenelement 9 (Schlüssel Straße)	100,0	100,0	100,0	100,0
rel. Häufigkeit [%] Datenelement 10 (Hausnummer)	100,0	100,0	100,0	100,0
rel. Häufigkeit [%] Datenelement 11 (Adressierungszusatz)	7,8	8,4	9,7	4,0
rel. Häufigkeit [%] Datenelement 12 (Koordinaten-Ost-Wert)	100,0	100,0	100,0	100,0
rel. Häufigkeit [%] Datenelement 13 (Koordinaten-Nord-Wert)	100,0	100,0	100,0	100,0
rel. Häufigkeit [%] Datenelement 14 (Straßenname)	100,0	100,0	100,0	100,0
rel. Häufigkeit [%] Datenelement 15 (Postleitzahl)	100,0	100,0	100,0	100,0
rel. Häufigkeit [%] Datenelement 16 (Postalischer Ortsname)	100,0	100,0	100,0	100,0
rel. Häufigkeit [%] Datenelement 17 (Zusatz zum postalischen Ortsnamen)	18,4	81,0	0,0	0,0
rel. Häufigkeit [%] Datenelement 18 (Postalischer Ortsteil)	96,6	100,0	100,0	100,0

Tabelle 30: Ausgefüllte HK-Attribute absolute Häufigkeit in mittelgroßen Städten.

Attribut	mittelgroße Städte			
	18507 Grimmen	66663 Merzig	26655 Westerstede	84307 Eggenfelden
HK				
Objekte gesamt	3039	11095	17388	4542
abs. Häufigkeit Datenelement 1 (Kennung des Datensatzes)	3039	11095	17388	4542
abs. Häufigkeit Datenelement 2 (OI)	3039	11095	17388	4542
abs. Häufigkeit Datenelement 3 (Qualität der georef. Adresse)	3039	11095	17388	4542
abs. Häufigkeit Datenelement 4 (Schlüssel Land)	3039	11095	17388	4542
abs. Häufigkeit Datenelement 5 (Schlüssel Regierungsbezirk)	0	0	17388	4542
abs. Häufigkeit Datenelement 6 (Schlüssel Kreis/kreisfreie Stadt)	3039	11095	17388	4542
abs. Häufigkeit Datenelement 7 (Schlüssel Gemeinde)	3039	11095	17388	4542
abs. Häufigkeit Datenelement 8 (Schlüssel Orts-/Gemeindeteil)	0	11095	0	4542
abs. Häufigkeit Datenelement 9 (Schlüssel Straße)	3039	11095	17388	0
abs. Häufigkeit Datenelement 10 (Hausnummer)	3039	11064	17388	4542
abs. Häufigkeit Datenelement 11 (Adressierungszusatz)	344	774	3148	608
abs. Häufigkeit Datenelement 12 (Koordinaten-Ost-Wert)	3039	11095	17388	4542
abs. Häufigkeit Datenelement 13 (Koordinaten-Nord-Wert)	3039	11095	17388	4542
abs. Häufigkeit Datenelement 14 (Straßenname)	3039	11095	17388	4542
abs. Häufigkeit Datenelement 15 (Postleitzahl)	3039	11095	17388	4542
abs. Häufigkeit Datenelement 16 (Postalischer Ortsname)	3039	11095	17388	4542
abs. Häufigkeit Datenelement 17 (Zusatz zum postalischen Ortsnamen)	125	117	312	69
abs. Häufigkeit Datenelement 18 (Postalischer Ortsteil)	3039	11095	17388	4542

Tabelle 31: Ausgefüllte HK-Attribute relative Häufigkeit in Prozent in mittelgroßen Städten.

Attribut	mittelgroße Städte			
	18507 Grimmen	66663 Merzig	26655 Westerstede	84307 Eggenfelden
HK				
Objekte gesamt	3039	11095	17388	4542
rel. Häufigkeit [%] Datenelement 1 (Kennung des Datensatzes)	100,0	100,0	100,0	100,0
rel. Häufigkeit [%] Datenelement 2 (OI)	100,0	100,0	100,0	100,0
rel. Häufigkeit [%] Datenelement 3 (Qualität der georef. Adresse)	100,0	100,0	100,0	100,0
rel. Häufigkeit [%] Datenelement 4 (Schlüssel Land)	100,0	100,0	100,0	100,0
rel. Häufigkeit [%] Datenelement 5 (Schlüssel Regierungsbezirk)	0,0	0,0	100,0	100,0
rel. Häufigkeit [%] Datenelement 6 (Schlüssel Kreis/kreisfreie Stadt)	100,0	100,0	100,0	100,0
rel. Häufigkeit [%] Datenelement 7 (Schlüssel Gemeinde)	100,0	100,0	100,0	100,0
rel. Häufigkeit [%] Datenelement 8 (Schlüssel Orts-/Gemeindeteil)	0,0	100,0	0,0	100,0
rel. Häufigkeit [%] Datenelement 9 (Schlüssel Straße)	100,0	100,0	100,0	0,0
rel. Häufigkeit [%] Datenelement 10 (Hausnummer)	100,0	99,7	100,0	100,0
rel. Häufigkeit [%] Datenelement 11 (Adressierungszusatz)	11,3	7,0	18,1	13,4
rel. Häufigkeit [%] Datenelement 12 (Koordinaten-Ost-Wert)	100,0	100,0	100,0	100,0
rel. Häufigkeit [%] Datenelement 13 (Koordinaten-Nord-Wert)	100,0	100,0	100,0	100,0
rel. Häufigkeit [%] Datenelement 14 (Straßenname)	100,0	100,0	100,0	100,0
rel. Häufigkeit [%] Datenelement 15 (Postleitzahl)	100,0	100,0	100,0	100,0
rel. Häufigkeit [%] Datenelement 16 (Postalischer Ortsname)	100,0	100,0	100,0	100,0
rel. Häufigkeit [%] Datenelement 17 (Zusatz zum postalischen Ortsnamen)	4,1	1,1	1,8	1,5
rel. Häufigkeit [%] Datenelement 18 (Postalischer Ortsteil)	100,0	100,0	100,0	100,0

Tabelle 32: Ausgefüllte OSM-Gebäude-Attribute absolute Häufigkeit in gebirgigen Gebieten.

Attribut	gebirgige Gebiete			
	87541 Bad Hindelang	38877 Benneckenstein (Harz)	08309 Eibenstock	54608 Bleialf
OSM-Gebäude				
Objekte gesamt	3594	1769	4413	1631
abs. Häufigkeit full_id	3594	1769	4413	1631
abs. Häufigkeit osm_id	3594	1769	4413	1631
abs. Häufigkeit addr_country	1513	700	619	2
abs. Häufigkeit addr_district	0	0	0	0
abs. Häufigkeit addr_hamlet	0	0	0	0
abs. Häufigkeit addr_housename	4	0	3	0
abs. Häufigkeit addr_housenumber	1500	702	2675	110
abs. Häufigkeit addr_state	0	0	0	0
abs. Häufigkeit addr_subdivision	0	0	0	0
abs. Häufigkeit building	3594	1769	4413	1631
abs. Häufigkeit name	298	73	93	23
abs. Häufigkeit addr_city	1328	703	2076	8
abs. Häufigkeit addr_place	18	0	32	1
abs. Häufigkeit addr_postcode	1325	703	2034	91
abs. Häufigkeit addr_street	1495	704	2644	108
abs. Häufigkeit shop	17	21	7	2
abs. Häufigkeit addr_suburb	818	703	567	0

Tabelle 33: Ausgefüllte OSM-Gebäude-Attribute relative Häufigkeit in Prozent in gebirgigen Gebieten.

Attribut	gebirgige Gebiete			
	87541 Bad Hindelang	38877 Benneckenstein (Harz)	08309 Eibenstock	54608 Bleialf
OSM-Gebäude				
Objekte gesamt	3594	1769	4413	1631
rel. Häufigkeit [%]				
full_id	100,0	100,0	100,0	100,0
rel. Häufigkeit [%]				
osm_id	100,0	100,0	100,0	100,0
rel. Häufigkeit [%]				
addr_country	42,1	39,6	14,0	0,1
rel. Häufigkeit [%]				
addr_district	0,0	0,0	0,0	0,0
rel. Häufigkeit [%]				
addr_hamlet	0,0	0,0	0,0	0,0
rel. Häufigkeit [%]				
addr_housename	0,1	0,0	0,1	0,0
rel. Häufigkeit [%]				
addr_housenumber	41,7	39,7	60,6	6,7
rel. Häufigkeit [%]				
addr_state	0,0	0,0	0,0	0,0
rel. Häufigkeit [%]				
addr_subdivision	0,0	0,0	0,0	0,0
rel. Häufigkeit [%]				
building	100,0	100,0	100,0	100,0
rel. Häufigkeit [%]				
name	8,3	4,1	2,1	1,4
rel. Häufigkeit [%]				
addr_city	37,0	39,7	47,0	0,5
rel. Häufigkeit [%]				
addr_place	0,5	0,0	0,7	0,1
rel. Häufigkeit [%]				
addr_postcode	36,9	39,7	46,1	5,6
rel. Häufigkeit [%]				
addr_street	41,6	39,8	59,9	6,6
rel. Häufigkeit [%]				
shop	0,5	1,2	0,2	0,1
rel. Häufigkeit [%]				
addr_suburb	22,8	39,7	12,8	0,0

Tabelle 34: Ausgefüllte OSM-Gebäude-Attribute absolute Häufigkeit in Großstädten und Ballungsräumen.

Attribut	Großstädte und Ballungsräume			
	45889 Gelsenkirchen (Bismarck)	22305 Hamburg (Brambek)	10243 Berlin (Friedrichshain)	60311 Frankfurt am Main (Altstadt)
OSM-Gebäude				
Objekte gesamt	5736	1750	1741	1379
abs. Häufigkeit full_id	5736	1750	1741	1379
abs. Häufigkeit osm_id	5736	1750	1741	1379
abs. Häufigkeit addr_country	2340	826	181	646
abs. Häufigkeit addr_district	0	0	0	0
abs. Häufigkeit addr_hamlet	0	0	0	0
abs. Häufigkeit addr_housename	1	0	20	18
abs. Häufigkeit addr_housenumber	2426	992	217	933
abs. Häufigkeit addr_state	0	0	0	0
abs. Häufigkeit addr_subdivision	0	0	0	0
abs. Häufigkeit building	5736	1750	1738	1379
abs. Häufigkeit name	99	45	151	190
abs. Häufigkeit addr_city	2378	849	209	923
abs. Häufigkeit addr_place	0	0	0	0
abs. Häufigkeit addr_postcode	2409	856	207	906
abs. Häufigkeit addr_street	2427	992	204	939
abs. Häufigkeit shop	7	11	16	7
abs. Häufigkeit addr_suburb	0	20	177	0

Tabelle 35: Ausgefüllte OSM-Gebäude-Attribute relative Häufigkeit in Prozent in Großstädten und Ballungsräumen.

Attribut	Großstädte und Ballungsräume			
	45889 Gelsenkirchen (Bismarck)	22305 Hamburg (Brambek)	10243 Berlin (Friedrichshain)	60311 Frankfurt am Main (Altstadt)
OSM-Gebäude				
Objekte gesamt	5736	1750	1741	1379
rel. Häufigkeit [%]				
full_id	100,0	100,0	100,0	100,0
rel. Häufigkeit [%]				
osm_id	100,0	100,0	100,0	100,0
rel. Häufigkeit [%]				
addr_country	40,8	47,2	10,4	46,8
rel. Häufigkeit [%]				
addr_district	0,0	0,0	0,0	0,0
rel. Häufigkeit [%]				
addr_hamlet	0,0	0,0	0,0	0,0
rel. Häufigkeit [%]				
addr_housename	0,0	0,0	1,1	1,3
rel. Häufigkeit [%]				
addr_housenumber	42,3	56,7	12,5	67,7
rel. Häufigkeit [%]				
addr_state	0,0	0,0	0,0	0,0
rel. Häufigkeit [%]				
addr_subdivision	0,0	0,0	0,0	0,0
rel. Häufigkeit [%]				
building	100,0	100,0	99,8	100,0
rel. Häufigkeit [%]				
name	1,7	2,6	8,7	13,8
rel. Häufigkeit [%]				
addr_city	41,5	48,5	12,0	66,9
rel. Häufigkeit [%]				
addr_place	0,0	0,0	0,0	0,0
rel. Häufigkeit [%]				
addr_postcode	42,0	48,9	11,9	65,7
rel. Häufigkeit [%]				
addr_street	42,3	56,7	11,7	68,1
rel. Häufigkeit [%]				
shop	0,1	0,6	0,9	0,5
rel. Häufigkeit [%]				
addr_suburb	0,0	1,1	10,2	0,0

Tabelle 36: Ausgefüllte OSM-Gebäude-Attribute absolute Häufigkeit in ländlichen Gebieten.

Attribut	ländliche Gebiete			
	88279 Amtzell	15936 Dahme / Mark	99831 Creuzburg	29491 Prezelle
OSM-Gebäude				
Objekte gesamt	3185	3125	894	190
abs. Häufigkeit full_id	3185	3125	894	190
abs. Häufigkeit osm_id	3185	3125	894	190
abs. Häufigkeit addr_country	446	89	16	4
abs. Häufigkeit addr_district	0	0	0	0
abs. Häufigkeit addr_hamlet	0	0	0	0
abs. Häufigkeit addr_housename	0	1	0	0
abs. Häufigkeit addr_housenumber	1388	232	55	14
abs. Häufigkeit addr_state	0	0	0	0
abs. Häufigkeit addr_subdivision	0	1	0	0
abs. Häufigkeit building	3185	3125	894	190
abs. Häufigkeit name	32	71	12	1
abs. Häufigkeit addr_city	1285	156	99	14
abs. Häufigkeit addr_place	555	100	0	0
abs. Häufigkeit addr_postcode	749	258	100	15
abs. Häufigkeit addr_street	833	166	102	14
abs. Häufigkeit shop	1	14	3	0
abs. Häufigkeit addr_suburb	46	27	56	4

Tabelle 37: Ausgefüllte Attribute OSM-Gebäude relative Häufigkeit in Prozent in ländlichen Gebieten.

Attribut	ländliche Gebiete			
	88279 Amtzell	15936 Dahme / Mark	99831 Creuzburg	29491 Prezelle
OSM-Gebäude				
Objekte gesamt	3185	3125	894	190
rel. Häufigkeit [%]				
full_id	100,0	100,0	100,0	100,0
rel. Häufigkeit [%]				
osm_id	100,0	100,0	100,0	100,0
rel. Häufigkeit [%]				
addr_country	14,0	2,8	1,8	2,1
rel. Häufigkeit [%]				
addr_district	0,0	0,0	0,0	0,0
rel. Häufigkeit [%]				
addr_hamlet	0,0	0,0	0,0	0,0
rel. Häufigkeit [%]				
addr_housename	0,0	0,0	0,0	0,0
rel. Häufigkeit [%]				
addr_housenumber	43,6	7,4	6,2	7,4
rel. Häufigkeit [%]				
addr_state	0,0	0,0	0,0	0,0
rel. Häufigkeit [%]				
addr_subdivision	0,0	0,0	0,0	0,0
rel. Häufigkeit [%]				
building	100,0	100,0	100,0	100,0
rel. Häufigkeit [%]				
name	1,0	2,3	1,3	0,5
rel. Häufigkeit [%]				
addr_city	40,3	5,0	11,1	7,4
rel. Häufigkeit [%]				
addr_place	17,4	3,2	0,0	0,0
rel. Häufigkeit [%]				
addr_postcode	23,5	8,3	11,2	7,9
rel. Häufigkeit [%]				
addr_street	26,2	5,3	11,4	7,4
rel. Häufigkeit [%]				
shop	0,0	0,4	0,3	0,0
rel. Häufigkeit [%]				
addr_suburb	1,4	0,9	6,3	2,1

Tabelle 38: Ausgefüllte OSM-Gebäude-Attribute absolute Häufigkeit in mittelgroßen Städten.

Attribut	mittelgroße Städte			
	18507 Grimmen	66663 Merzig	26655 Westerstede	84307 Eggenfelden
OSM-Gebäude				
Objekte gesamt	5160	13264	17715	5421
abs. Häufigkeit full_id	5160	13264	17715	5421
abs. Häufigkeit osm_id	5160	13264	17715	5421
abs. Häufigkeit addr_country	1109	3026	3530	139
abs. Häufigkeit addr_district	8	0	0	0
abs. Häufigkeit addr_hamlet	0	0	407	0
abs. Häufigkeit addr_housename	11	2	12	0
abs. Häufigkeit addr_housenumber	1480	3159	3998	616
abs. Häufigkeit addr_state	8	250	797	0
abs. Häufigkeit addr_subdivision	0	0	0	0
abs. Häufigkeit building	5160	13264	17715	5421
abs. Häufigkeit name	97	180	477	42
abs. Häufigkeit addr_city	1483	3081	3980	421
abs. Häufigkeit addr_place	21	0	1	17
abs. Häufigkeit addr_postcode	1120	3151	3984	374
abs. Häufigkeit addr_street	1463	3159	4143	641
abs. Häufigkeit shop	14	36	73	7
abs. Häufigkeit addr_suburb	81	12	2094	4

Tabelle 39: Ausgefüllte OSM-Gebäude-Attribute relative Häufigkeit in Prozent in mittelgroßen Städten.

Attribut	mittelgroße Städte			
	18507 Grimmen	66663 Merzig	26655 Westerstede	84307 Eggenfelden
OSM-Gebäude				
Objekte gesamt	5160	13264	17715	5421
rel. Häufigkeit [%]				
full_id	100,0	100,0	100,0	100,0
rel. Häufigkeit [%]				
osm_id	100,0	100,0	100,0	100,0
rel. Häufigkeit [%]				
addr_country	21,5	22,8	19,9	2,6
rel. Häufigkeit [%]				
addr_district	0,2	0,0	0,0	0,0
rel. Häufigkeit [%]				
addr_hamlet	0,0	0,0	2,3	0,0
rel. Häufigkeit [%]				
addr_housename	0,2	0,0	0,1	0,0
rel. Häufigkeit [%]				
addr_housenumber	28,7	23,8	22,6	11,4
rel. Häufigkeit [%]				
addr_state	0,2	1,9	4,5	0,0
rel. Häufigkeit [%]				
addr_subdivision	0,0	0,0	0,0	0,0
rel. Häufigkeit [%]				
building	100,0	100,0	100,0	100,0
rel. Häufigkeit [%]				
name	1,9	1,4	2,7	0,8
rel. Häufigkeit [%]				
addr_city	28,7	23,2	22,5	7,8
rel. Häufigkeit [%]				
addr_place	0,4	0,0	0,0	0,3
rel. Häufigkeit [%]				
addr_postcode	21,7	23,8	22,5	6,9
rel. Häufigkeit [%]				
addr_street	28,4	23,8	23,4	11,8
rel. Häufigkeit [%]				
addr_suburb	0,3	0,3	0,4	0,1
rel. Häufigkeit [%]				
addr_suburb	1,6	0,1	11,8	0,1

Tabelle 40: Ausgefüllte Attribute OSM-Punkte in Gebäuden absolute Häufigkeit in gebirgigen Gebieten.

Attribut	gebirgige Gebiete			
	87541 Bad Hindelang	38877 Benneckenstein (Harz)	08309 Eibenstock	54608 Bleialf
OSM-Punkte-in-Gebäuden				
Objekte gesamt	166	20	234	12
abs. Häufigkeit full_id	166	20	234	12
abs. Häufigkeit osm_id	166	20	234	12
abs. Häufigkeit addr_city	45	11	96	1
abs. Häufigkeit addr_hamlet	0	0	0	0
abs. Häufigkeit addr_housename	1	0	2	1
abs. Häufigkeit addr_housnumber	44	13	113	1
abs. Häufigkeit addr_inclusion	0	0	0	0
abs. Häufigkeit addr_state	0	0	0	0
abs. Häufigkeit addr_street	48	13	114	1
abs. Häufigkeit addr_suburb	17	11	49	0
abs. Häufigkeit building	0	0	0	0
abs. Häufigkeit entrance	0	0	0	0
abs. Häufigkeit addr_country	41	11	14	0
abs. Häufigkeit name	109	7	131	11
abs. Häufigkeit addr_place	0	0	0	0
abs. Häufigkeit addr_postcode	47	11	80	2
abs. Häufigkeit shop	26	5	60	2

Tabelle 41: Ausgefüllte Attribute OSM-Punkte in Gebäuden relative Häufigkeit in Prozent in gebirgigen Gebieten.

Attribut	gebirgige Gebiete			
	87541 Bad Hindelang	38877 Benneckenstein (Harz)	08309 Eibenstock	54608 Bleialf
OSM-Punkte-in-Gebäuden				
Objekte gesamt	166	20	234	12
rel. Häufigkeit [%]				
full_id	100,0	100,0	100,0	100,0
rel. Häufigkeit [%]				
osm_id	100,0	100,0	100,0	100,0
rel. Häufigkeit [%]				
addr_city	27,1	55,0	41,0	8,3
rel. Häufigkeit [%]				
addr_hamlet	0,0	0,0	0,0	0,0
rel. Häufigkeit [%]				
addr_housename	0,6	0,0	0,9	8,3
rel. Häufigkeit [%]				
addr_housnumber	26,5	65,0	48,3	8,3
rel. Häufigkeit [%]				
addr_inclusion	0,0	0,0	0,0	0,0
rel. Häufigkeit [%]				
addr_state	0,0	0,0	0,0	0,0
rel. Häufigkeit [%]				
addr_street	28,9	65,0	48,7	8,3
rel. Häufigkeit [%]				
addr_suburb	10,2	55,0	20,9	0,0
rel. Häufigkeit [%]				
building	0,0	0,0	0,0	0,0
rel. Häufigkeit [%]				
entrance	0,0	0,0	0,0	0,0
rel. Häufigkeit [%]				
addr_country	24,7	55,0	6,0	0,0
rel. Häufigkeit [%]				
name	65,7	35,0	56,0	91,7
rel. Häufigkeit [%]				
addr_place	0,0	0,0	0,0	0,0
rel. Häufigkeit [%]				
addr_postcode	28,3	55,0	34,2	16,7
rel. Häufigkeit [%]				
shop	15,7	25,0	25,6	16,7

Tabelle 42: Ausgefüllte Attribute OSM-Punkte in Gebäuden absolute Häufigkeit in Großstädten und Ballungsräumen.

Attribut	Großstädte und Ballungsräume			
Testgebiet	45889 Gelsenkirchen (Bismarck)	22305 Hamburg (Brambek)	10243 Berlin (Friedrichshain)	60311 Frankfurt am Main (Altstadt)
OSM-Punkte-in- Gebäuden				
Objekte gesamt	195	689	2406	706
abs. Häufigkeit full_id	195	689	2406	706
abs. Häufigkeit osm_id	195	689	2406	706
abs. Häufigkeit addr_city	140	224	1691	148
abs. Häufigkeit addr_hamlet	0	0	0	0
abs. Häufigkeit addr_housename	0	0	8	2
abs. Häufigkeit addr_housnumber	151	383	1715	164
abs. Häufigkeit addr_inclusion	0	0	45	0
abs. Häufigkeit addr_state	0	0	0	0
abs. Häufigkeit addr_street	151	385	1732	170
abs. Häufigkeit addr_suburb	0	0	1523	0
abs. Häufigkeit building	0	0	0	1
abs. Häufigkeit entrance	1	0	33	6
abs. Häufigkeit addr_country	119	147	1576	44
abs. Häufigkeit name	56	265	735	604
abs. Häufigkeit addr_place	0	0	0	0
abs. Häufigkeit addr_postcode	147	236	1706	154
abs. Häufigkeit shop	15	114	297	286

Tabelle 43: Ausgefüllte Attribute OSM-Punkte in Gebäuden relative Häufigkeit in Prozent in Großstädten und Ballungsräumen.

Attribut	Großstädte und Ballungsräume			
Testgebiet	45889 Gelsenkirchen (Bismarck)	22305 Hamburg (Brambek)	10243 Berlin (Friedrichshain)	60311 Frankfurt am Main (Altstadt)
OSM-Punkte-in-Gebäuden				
Objekte gesamt	195	689	2406	706
rel. Häufigkeit [%]				
full_id	100,0	100,0	100,0	100,0
rel. Häufigkeit [%]				
osm_id	100,0	100,0	100,0	100,0
rel. Häufigkeit [%]				
addr_city	71,8	32,5	70,3	21,0
rel. Häufigkeit [%]				
addr_hamlet	0,0	0,0	0,0	0,0
rel. Häufigkeit [%]				
addr_housename	0,0	0,0	0,3	0,3
rel. Häufigkeit [%]				
addr_housnumber	77,4	55,6	71,3	23,2
rel. Häufigkeit [%]				
addr_inclusion	0,0	0,0	1,9	0,0
rel. Häufigkeit [%]				
addr_state	0,0	0,0	0,0	0,0
rel. Häufigkeit [%]				
addr_street	77,4	55,9	72,0	24,1
rel. Häufigkeit [%]				
addr_suburb	0,0	0,0	63,3	0,0
rel. Häufigkeit [%]				
building	0,0	0,0	0,0	0,1
rel. Häufigkeit [%]				
entrance	0,5	0,0	1,4	0,8
rel. Häufigkeit [%]				
addr_country	61,0	21,3	65,5	6,2
rel. Häufigkeit [%]				
name	28,7	38,5	30,5	85,6
rel. Häufigkeit [%]				
addr_place	0,0	0,0	0,0	0,0
rel. Häufigkeit [%]				
addr_postcode	75,4	34,3	70,9	21,8
rel. Häufigkeit [%]				
shop	7,7	16,5	12,3	40,5

Tabelle 44: Ausgefüllte Attribute OSM-Punkte in Gebäuden absolute Häufigkeit in ländlichen Gebieten.

Attribut	ländliche Gebiete			
	88279 Amtzell	15936 Dahme / Mark	99831 Creuzburg	29491 Prezelle
OSM-Punkte-in-Gebäuden				
Objekte gesamt	144	26	20	11
abs. Häufigkeit full_id	144	26	20	11
abs. Häufigkeit osm_id	144	26	20	11
abs. Häufigkeit addr_city	107	6	7	10
abs. Häufigkeit addr_hamlet	1	0	0	0
abs. Häufigkeit addr_housename	0	1	0	0
abs. Häufigkeit addr_housnumber	108	6	7	10
abs. Häufigkeit addr_inclusion	0	0	0	0
abs. Häufigkeit addr_state	0	0	0	0
abs. Häufigkeit addr_street	95	7	7	10
abs. Häufigkeit addr_suburb	7	3	0	6
abs. Häufigkeit building	0	0	0	0
abs. Häufigkeit entrance	0	0	0	0
abs. Häufigkeit addr_country	32	3	6	6
abs. Häufigkeit name	24	19	16	2
abs. Häufigkeit addr_place	19	0	0	0
abs. Häufigkeit addr_postcode	48	6	8	10
abs. Häufigkeit shop	7	1	3	0

Tabelle 45: Ausgefüllte Attribute OSM-Punkte in Gebäuden relative Häufigkeit in Prozent in ländlichen Gebieten.

Attribut	ländliche Gebiete			
	88279 Amtzell	15936 Dahme / Mark	99831 Creuzburg	29491 Prezelle
OSM-Punkte-in-Gebäuden				
Objekte gesamt	144	26	20	11
rel. Häufigkeit [%]				
full_id	100,0	100,0	100,0	100,0
rel. Häufigkeit [%]				
osm_id	100,0	100,0	100,0	100,0
rel. Häufigkeit [%]				
addr_city	74,3	23,1	35,0	90,9
rel. Häufigkeit [%]				
addr_hamlet	0,7	0,0	0,0	0,0
rel. Häufigkeit [%]				
addr_housename	0,0	3,8	0,0	0,0
rel. Häufigkeit [%]				
addr_housnumber	75,0	23,1	35,0	90,9
rel. Häufigkeit [%]				
addr_inclusion	0,0	0,0	0,0	0,0
rel. Häufigkeit [%]				
addr_state	0,0	0,0	0,0	0,0
rel. Häufigkeit [%]				
addr_street	66,0	26,9	35,0	90,9
rel. Häufigkeit [%]				
addr_suburb	4,9	11,5	0,0	54,5
rel. Häufigkeit [%]				
building	0,0	0,0	0,0	0,0
rel. Häufigkeit [%]				
entrance	0,0	0,0	0,0	0,0
rel. Häufigkeit [%]				
addr_country	22,2	11,5	30,0	54,5
rel. Häufigkeit [%]				
name	16,7	73,1	80,0	18,2
rel. Häufigkeit [%]				
addr_place	13,2	0,0	0,0	0,0
rel. Häufigkeit [%]				
addr_postcode	33,3	23,1	40,0	90,9
rel. Häufigkeit [%]				
shop	4,9	3,8	15,0	0,0

Tabelle 46: Ausgefüllte Attribute OSM-Punkte in Gebäuden absolute Häufigkeit in mittelgroßen Städten.

Attribut	mittelgroße Städte			
	18507 Grimmen	66663 Merzig	26655 Westerstede	84307 Eggenfelden
OSM-Punkte-in-Gebäuden				
Objekte gesamt	859	232	3978	197
abs. Häufigkeit full_id	859	232	3978	197
abs. Häufigkeit osm_id	859	232	3978	197
abs. Häufigkeit addr_city	843	17	3792	72
abs. Häufigkeit addr_hamlet	0	0	9	0
abs. Häufigkeit addr_housename	2	0	14	0
abs. Häufigkeit addr_housnumber	847	20	3827	76
abs. Häufigkeit addr_inclusion	0	0	0	0
abs. Häufigkeit addr_state	0	0	10	0
abs. Häufigkeit addr_street	823	20	3826	76
abs. Häufigkeit addr_suburb	71	1	1430	1
abs. Häufigkeit building	1	0	0	0
abs. Häufigkeit entrance	0	0	0	0
abs. Häufigkeit addr_country	709	4	3711	27
abs. Häufigkeit name	25	198	306	149
abs. Häufigkeit addr_place	24	0	0	2
abs. Häufigkeit addr_postcode	716	18	3794	73
abs. Häufigkeit shop	8	100	125	76

Tabelle 47: Ausgefüllte Attribute OSM-Punkte in Gebäuden relative Häufigkeit in Prozent in mittelgroßen Städten.

Attribut	mittelgroße Städte			
	18507 Grimmen	66663 Merzig	26655 Westerstede	84307 Eggenfelden
OSM-Punkte-in-Gebäuden				
Objekte gesamt	859	232	3978	197
rel. Häufigkeit [%]				
full_id	100,0	100,0	100,0	100,0
rel. Häufigkeit [%]				
osm_id	100,0	100,0	100,0	100,0
rel. Häufigkeit [%]				
addr_city	98,1	7,3	95,3	36,5
rel. Häufigkeit [%]				
addr_hamlet	0,0	0,0	0,2	0,0
rel. Häufigkeit [%]				
addr_housename	0,2	0,0	0,4	0,0
rel. Häufigkeit [%]				
addr_housnumber	98,6	8,6	96,2	38,6
rel. Häufigkeit [%]				
addr_inclusion	0,0	0,0	0,0	0,0
rel. Häufigkeit [%]				
addr_state	0,0	0,0	0,3	0,0
rel. Häufigkeit [%]				
addr_street	95,8	8,6	96,2	38,6
rel. Häufigkeit [%]				
addr_suburb	8,3	0,4	35,9	0,5
rel. Häufigkeit [%]				
building	0,1	0,0	0,0	0,0
rel. Häufigkeit [%]				
entrance	0,0	0,0	0,0	0,0
rel. Häufigkeit [%]				
addr_country	82,5	1,7	93,3	13,7
rel. Häufigkeit [%]				
name	2,9	85,3	7,7	75,6
rel. Häufigkeit [%]				
addr_place	2,8	0,0	0,0	1,0
rel. Häufigkeit [%]				
addr_postcode	83,4	7,8	95,4	37,1
rel. Häufigkeit [%]				
shop	0,9	43,1	3,1	38,6

Tabelle 48: Adressvergleich HU und OSM-Gebäude absolute Häufigkeit.

Testgebiete / Gebietart	Adressvergleich HU und OSM-Gebäude (absolute Häufigkeit)						Anzahl Adressen gesamt
	übereinstimmende Hausnummer	Straße	Ortsname	Postleitzahl	Ortsteil	übereinstimmender Ortsteil	
gebirgige Gebiete	3015	3418	2509	1341	975	9347	
87541 Bad Hindelang	853	950	844	846	518	3097	
38877 Benneckenstein (Harz)	301	302	0	443	0	1296	
08309 Eibenstock	1812	2114	1661	0	457	3486	
54608 Bleialf	49	52	4	52	0	1468	
Großstädte und Ballungsräume	1585	2642	2348	2764	142	8572	
45889 Gelsenkirchen (Bismarck)	1064	1521	1526	1537	0	4262	
22305 Hamburg (Brambek)	259	640	676	678	20	1726	
10243 Berlin (Friedrichshain)	65	129	146	148	122	1562	
60311 Frankfurt am Main (Altstadt)	197	352	0	401	0	1022	
ländliche Gebiete	985	695	831	674	48	6526	
88279 Amtzell	856	555	783	493	17	2655	
15936 Dahme / Mark	98	92	12	131	15	2883	
99831 Creuzburg	23	38	26	39	13	820	
29491 Prezelle	8	10	10	11	3	168	
mittelgroße Städte	5197	5636	6157	5902	1232	35904	
18507 Grimmen	1011	446	1210	925	67	4773	
66663 Merzig	1780	2237	2141	2202	9	10714	
26655 Westerstede	2101	2573	2528	2531	1155	15354	
84307 Eggenfelden	305	380	278	244	1	5063	

Tabelle 49: Adressvergleich HU und OSM-Gebäude relative Häufigkeit in Prozent.

Testgebiete / Gebietart	Adressvergleich HU und OSM-Gebäude (relative Häufigkeit [%])					
	übereinstimmende Hausnummer	übereinstimmende Straße	übereinstimmende Ortsname	übereinstimmende Postleitzahl	übereinstimmende Ortsteil	übereinstimmender Ortsteil
gebirgige Gebiete	32,26	36,57	26,84	14,35	10,43	10,43
87541 Bad Hindelang	27,54	30,67	27,25	27,32	16,73	16,73
38877						
Benneckenstein (Harz)	23,23	23,30	0,00	34,18	0,00	0,00
08309 Eibenstock	51,98	60,64	47,65	0,00	13,11	13,11
54608 Bleialf	3,34	3,54	0,27	3,54	0,00	0,00
Großstädte und Ballungsräume	18,49	30,82	27,39	32,24	1,66	1,66
45889 Gelsenkirchen (Bismarck)	24,96	35,69	35,80	36,06	0,00	0,00
22305 Hamburg (Brambek)	15,01	37,08	39,17	39,28	1,16	1,16
10243 Berlin (Friedrichshain)	4,16	8,26	9,35	9,48	7,81	7,81
60311 Frankfurt am Main (Altstadt)	19,28	34,44	0,00	39,24	0,00	0,00
ländliche Gebiete	15,09	10,65	12,73	10,33	0,74	0,74
88279 Amtzell	32,24	20,90	29,49	18,57	0,64	0,64
15936 Dahme / Mark	3,40	3,19	0,42	4,54	0,52	0,52
99831 Creuzburg	2,80	4,63	3,17	4,76	1,59	1,59
29491 Prezelle	4,76	5,95	5,95	6,55	1,79	1,79
mittelgroße Städte	14,47	15,70	17,15	16,44	3,43	3,43
18507 Grimmen	21,18	9,34	25,35	19,38	1,40	1,40
66663 Merzig	16,61	20,88	19,98	20,55	0,08	0,08
26655 Westerstede	13,68	16,76	16,46	16,48	7,52	7,52
84307 Eggenfelden	6,02	7,51	5,49	4,82	0,02	0,02