

Studienarbeit

Automatisierte Generierung von Postleitzahlgebieten aus OpenStreetMap-Daten unter Verwendung von Open Source GIS Software

eingereicht von Christian Hauck

geboren am 18.02.1984 in Zwickau

Tag der Einreichung 25.07.2011

Betreuer Prof. Dipl.-Phys. Dr.-Ing. habil. Dirk Burghardt
Institut für Kartographie

Dipl.-Ing. (FH) Stefan Hahmann
Institut für Kartographie

Kurzfassung

Das Projekt OpenStreetMap als freie Wiki-Weltkarte gewinnt als Quelle von Geodaten für unterschiedlichste Bedürfnisse innerhalb der Geowissenschaften, des Geomarketings und auch im Alltag immer mehr an Bedeutung. Die kostenlosen, von Freiwilligen einer Community gesammelten geographischen Daten, sogenannte nutzergenerierte Daten, dienen heute vielen Anwendern als Datengrundlage und stehen in der Konkurrenz zu proprietären Geodaten von kommerziellen Anbietern.

Neben Straßendaten sind zahlreiche zusätzliche Daten innerhalb OpenStreetMap verfügbar.

Die aktuelle Technologie des Webmapping 2.0 und die dafür zahlreich verfügbaren Open Source GIS Systeme erlauben dem Anwender eine Vielzahl von Möglichkeiten zu Bearbeitung von Geodaten.

Die freie Verfügbarkeit von Daten und Software machen die Nutzung und Verarbeitung von Geodaten somit auch für kleinere Unternehmen und Privatnutzer bezahlbar.

Die vorliegende Studienarbeit stellt ein Verfahren vor, welches es ermöglicht aus OpenStreetMap-Daten Postleitzahlgebiete zu erstellen. Postleitzahlgebiete sind für viele Bereiche der Wirtschaft sehr wichtige Planungsstrukturen. Als Datengrundlage werden OSM-Adressdaten genutzt, aus denen, unter Nutzung von Open Source GIS Software, die Postleitzahlgebiete erzeugt werden. Die Generierung ist dabei automatisierbar und ohne die Nutzung grafischer Benutzeroberflächen möglich. Sie liefert als Ergebnis die Postleitzahlgebiete Deutschlands. Diese werden anschließend, unter Berücksichtigung der ISO-Normen für Geoinformation, mit einem kommerziellen Datensatz verglichen und auf ihre Nutzbarkeit für Geomarketing und andere nützliche Anwendungen hin überprüft.

Abstract

The OpenStreetMap project as the Free Wiki World Map as a source of gains for a wide variety of geospatial data needs within the geosciences, the geomarketing and in everyday life is becoming increasingly important. The free, a community of volunteers gathered geo-graphical data, so-called user-generated data, now serve many users as basic data and are in competition with proprietary spatial data from commercial providers. In addition to road data within OpenStreetMap numerous additional data is available.

The current technology of the Web Mapping 2.0 and the many available Open Source GIS systems provide the user with a variety of options for managing spatial data. The free availability of data and software make the use and processing of geospatial data thus affordable for small businesses and home users.

The current work presents a method that allows to create postcode areas from OpenStreetMap data. Postcode areas are very important planning structures for many areas of the economy. The OSM address data are used as data base, out of which the zip code areas are produced by taking advantage of Open Source GIS software. The creation is automated and without the use of graphical user interfaces. It provides as result the postal code areas of Germany. Taking into account the ISO standards for geoinformation, the postal code areas are later compared with a commercial data set and their usability for geomarketing and other useful application is tested.

Inhaltsverzeichnis

| | |
|---|------|
| Kurzfassung | I |
| Abstract | II |
| Abbildungsverzeichnis..... | V |
| Tabellenverzeichnis..... | VI |
| Formeln | VII |
| Abkürzungsverzeichnis..... | VIII |
| 1 Einleitung..... | 1 |
| 1.1 Motivation..... | 1 |
| 1.2 Aufbau der Arbeit..... | 2 |
| 1.3 Ziel der Arbeit..... | 3 |
| 2 Theorie..... | 4 |
| 2.1 Postleitzahlen | 4 |
| 2.2 Nutzung von Postleitzahlen in der Privatwirtschaft..... | 6 |
| 2.3 Postleitzahlen in OpenStreetMap | 8 |
| 2.4 Qualität von Geodaten | 13 |
| 2.5 Vergleich von Daten unterschiedlicher Herkunft..... | 18 |
| 2.5.1 Qualität von OSM-Daten | 19 |
| 2.5.2 Vergleich von OSM-Daten mit Daten kommerzieller Anbieter..... | 20 |
| 2.5.3 Vergleichsmethoden für Polygondatensätze | 22 |
| 2.5.3.1 Vergleichsmethoden | 23 |
| 2.5.3.2 Vergleichskriterien und Qualitätsmaße..... | 24 |
| 2.6 Gebietsgenerierung aus Punktdaten..... | 29 |
| 2.6.1 Allgemeines und Literatur | 29 |
| 2.6.2 Voronoi-Verfahren | 32 |
| 2.7 Open Source GIS Software | 35 |
| 3 Praxis | 41 |
| 3.1 Technische Voraussetzungen | 41 |
| 3.2 Datengrundlage..... | 41 |
| 3.3 Allgemeiner Arbeitsablauf..... | 43 |
| 3.3.1 Datenvorverarbeitung | 43 |
| 3.3.2 Erzeugung der Gebiete | 46 |
| 3.3.3 Datennachbearbeitung..... | 47 |
| 3.3.4 Export der Daten als Shapefile | 47 |
| 3.4 Praktische Umsetzung | 48 |

| | | |
|-------|--|--------|
| 3.4.1 | Datenvorverarbeitung | 48 |
| 3.4.2 | Erzeugung der Gebiete | 50 |
| 3.4.3 | Datennachbearbeitung..... | 51 |
| 3.4.4 | Export der Daten als Shapefile | 51 |
| 3.5 | Ergebnisse..... | 51 |
| 3.6 | Vergleich der Daten..... | 52 |
| 4 | Fazit | 63 |
| 5 | Ausblick..... | 65 |
| | Quellenverzeichnis | IX |
| | Literaturquellen..... | IX |
| | Internetquellen..... | XV |
| | Anhang | XIX |
| A | Anhang Quellcodes..... | XIX |
| A.1 | Quellcode Import Deutschlandgrenze | XIX |
| A.2 | Quellcode Vorverarbeitung | XIX |
| A.3 | Quellcode Erzeugung Polygone..... | XXI |
| A.4 | Quellcode Nachbearbeitung..... | XXII |
| A.5 | Quellcode Export | XXIII |
| B | Anhang Screenshots PDF-Karten..... | XXV |
| B.1 | Postleitzonen (PLZ1) Deutschland OSM | XXV |
| B.2 | Postleitregionen (PLZ2) Deutschland OSM | XXVI |
| B.3 | Postleitzahlen 3-stellig Deutschland OSM..... | XXVII |
| B.4 | Postleitzahlgebiete (PLZ5) Deutschland OSM | XXVIII |
| B.5 | Vergleich der Postleitzahlen..... | XXIX |
| B.6 | Vergleich PLZ5 Hamburg | XXX |
| B.7 | Punktdichte PLZ5-Centroide OpenStreetMap..... | XXXI |
| B.8 | Punktdichte PLZ5-Centroide TeleAtlas..... | XXXII |
| B.9 | Euklidische Distanz PLZ3-Centroide OpenStreetMap | XXXIII |
| B.10 | Euklidische Distanz PLZ3-Centroide TeleAtlas..... | XXXIV |
| B.11 | Euklidische Distanz PLZ5-Centroide OpenStreetMap | XXXV |
| B.12 | Euklidische Distanz PLZ5-Centroide TeleAtlas..... | XXXVI |
| C | Anhang Tabelle | XXXVII |
| C.1 | Tabelle Vergleich PLZ5-Gebiete Hamburg..... | XXXVII |
| D | Anhang..... | XLI |
| D.1 | CD | XLI |

Abbildungsverzeichnis

| | | |
|----------|--|----|
| Abb. 2.1 | PLZ-Suche Deutsche Post AG (DEUTSCHE POST, 2011b)..... | 5 |
| Abb. 2.2 | PLZ-Gebiete Sachsen, Daten © TeleAtlas 2010..... | 6 |
| Abb. 2.3 | Karte zu PLZ-Gebieten in Großbritannien (RANDOMJUNK, 2011)..... | 11 |
| Abb. 2.4 | Beispiel Puffervergleichsmethode..... | 27 |
| Abb. 2.5 | Punktmenge und Voronoi-Diagramm (KANG, 2008)..... | 33 |
| Abb. 2.6 | Ausschnitt Voronoi-Diagramm aus Abb. 2.5 (KANG, 2008)..... | 33 |
| Abb. 2.7 | Ablauf Fortune's Algorithmus (ODGAARD & NIELSEN, 2000) | 35 |
| Abb. 3.1 | allgemeiner Arbeitsablauf | 43 |
| Abb. 3.2 | schematischer Arbeitsablauf mit Funktionsnamen aus GRASS GIS | 48 |
| Abb. 3.3 | Ausschnitt Euklidische Distanz PLZ5 OSM | 57 |
| Abb. 3.4 | Ausschnitt Punktdichte PLZ5 OSM | 57 |
| Abb. 3.5 | Ausschnitt Hamburg | 59 |
| Abb. 3.6 | Ausschnitt PLZ1 | 60 |
| Abb. 3.7 | Ausschnitt PLZ2 | 60 |
| Abb. 3.8 | Ausschnitt PLZ3 | 60 |
| Abb. 3.9 | Ausschnitt PLZ5 | 60 |

Tabellenverzeichnis

| | | |
|-----------|--|----|
| Tab. 2.1 | Subelemente zur Positionsgenauigkeit (PAS 1071, 2007)..... | 15 |
| Tab. 2.2 | Subelemente zur thematischen Genauigkeit (PAS 1071, 2007)..... | 15 |
| Tab. 2.3 | Subelemente zur Vollständigkeit (PAS 1071, 2007) | 15 |
| Tab. 2.4 | Subelemente zur logischen Konsistenz (PAS 1071, 2007)..... | 16 |
| Tab. 2.5 | Subelemente zur zeitlichen Genauigkeit (PAS 1071, 2007) | 16 |
| Tab. 2.6 | Zusammenfassung Vergleichskriterien | 29 |
| Tab. 3.1 | Anzahl PLZ..... | 52 |
| Tab. 3.2 | Anzahl nicht übereinstimmender und übereinstimmender PLZ..... | 53 |
| Tab. 3.3 | Anzahl nicht übereinstimmender PLZ nach Anbieter..... | 53 |
| Tab. 3.4 | Vergleich Fläche PLZ-Gebiete gesamt | 54 |
| Tab. 3.5 | Vergleich Einzelfläche PLZ1 und Abstand Centroide PLZ1 | 54 |
| Tab. 3.6 | Vergleich Einzelfläche PLZ2 und Abstand Centroide PLZ2 | 54 |
| Tab. 3.7 | Vergleich Einzelfläche PLZ3 und Abstand Centroide PLZ3 | 54 |
| Tab. 3.8 | Vergleich Einzelfläche PLZ5 und Abstand Centroide PLZ5 | 54 |
| Tab. 3.9 | Anzahl OSM-Centroide innerhalb eines Puffers um die TeleAtlas-Centroide..... | 56 |
| Tab. 3.10 | Genauigkeit und Vollständigkeit der Daten in Hamburg..... | 58 |
| Tab. 3.11 | fehlende bzw. fehlerhafte PLZ5..... | 62 |

Formeln

| | | |
|----------|---|----|
| Formel 1 | Abplattung der Erde | 26 |
| Formel 2 | Äquatorradius der Erde in Kilometern | 26 |
| Formel 3 | Parameter F, G und I..... | 26 |
| Formel 4 | Umrechnung in Bogenmaß..... | 26 |
| Formel 5 | Berechnung Abstand D..... | 26 |
| Formel 6 | Korrektur Abstand D mit H_1 und H_2 | 26 |
| Formel 7 | Abstand s in Kilometern | 26 |

Abkürzungsverzeichnis

| | |
|-----------|--|
| API | Application Programming Interface |
| ATKIS | Amtliches Topographisch-Kartographisches Informationssystem |
| DDGI | Deutscher Dachverband für Geoinformation |
| DIN | Deutsche Industrienorm bzw. Deutsches Institut für Normung e. V. |
| DLM | Digitales Landschaftsmodell |
| EPSG | European Petroleum Survey Group |
| FOSS | Free and Open Source Software |
| FOSS4G | Free and Open Source Software for Geospatial |
| GDAL | Geospatial Data Abstraction Library |
| GeoRSS | Geographic Really Simple Syndication |
| GIS | Geoinformationssystem |
| GPS | Global Positioning System |
| GPX | GPS Exchange Format |
| GRASS GIS | Geographic Resources Analysis Support System GIS |
| ICA | International Cartographic Association |
| ID | Identifikator |
| ISO | International Organization for Standardization |
| KML | Keyhole Markup Language |
| ODBL | Open Database License |
| OGR | OpenGIS Simple Features Reference Implementation |
| ÖPNV | Öffentlicher Personennahverkehr |
| OSGEO | Open Source Geospatial Foundation |
| OSM | OpenStreetMap |
| PAS | Publicly Available Specification |
| PDF | Portable Document Format |
| PGSQL | Postgre Structured Query Language |
| PLZ | Postleitzahl |
| POI | Point of Interest |
| RMSE | Root Mean Square Error |
| SQL | Structured Query Language |
| TA | TeleAtlas |
| UMN | University of Minnesota |
| WGS | World Geodetic System |
| XML | Extensible Markup Language |

1 Einleitung

Nutzergenerierte Daten gewinnen immer mehr an Wichtigkeit. So ist auch das 2004 gegründete Projekt *OpenStreetMap*¹ (*OSM*), welches innerhalb Europas mittlerweile einen großen Teil an Geodaten gesammelt hat (ZIELSTRA & ZIPF, 2010), ein Projekt, welches sich auf das Sammeln nutzergenerierter Geodaten weltweit spezialisiert hat.

Unter diesen Daten befinden sich nicht nur topographische Daten wie Straßen, Flüsse, Eisenbahnlinien oder auch Grünflächen, es werden auch immer mehr thematische Daten wie zum Beispiel Supermärkte, Straßenbahnhaltestellen oder Postleitzahlen im Projekt *OpenStreetMap* gesammelt.

Diese Daten lassen sich alle kostenfrei nutzen, da sie alle unter der *Creative Commons Attribution-Share-Alike-Lizenz* (kurz: *CC BY-SA, Version 2.0*)² stehen. So lassen sich aus den Daten weitere neue Daten wie zum Beispiel Postleitzahlengebiete erzeugen.

Mit Hilfe sogenannter Open Source GIS Software, eine Zusammenstellung findet man unter *OSGEO*³, *GISWIKI*⁴ oder auch unter *Open Source GIS*⁵, welche mittlerweile zahlreich über das Internet verfügbar ist, lassen sich diese *OSM*-Daten weiterverarbeiten. Es ist somit möglich kostenfreie Daten mit kostenfreier Software zu verwalten und zu verarbeiten um daraus diverse kartographische Produkte zu erstellen. Im Speziellen sind mit diesen Daten auch Analysen bis hin zum Geomarketing möglich.

1.1 Motivation

Durch die Möglichkeit der freien Nutzung der Geodaten von *OSM* und der Open Source GIS Software lassen sich verschiedene Folgedaten erzeugen. Ein Beispiel hierfür sind die Postleitzahlgebiete, welche in *OSM* noch nicht flächendeckend vorhanden sind.

Postleitzahlgebiete dienen für viele Wirtschaftsbereiche, insbesondere dem Bereich des Geomarketings, als wichtige Grundlage für eine Planungsstruktur. Es lassen sich mit dieser Abgrenzung statistische Werte wie Arbeitslosigkeit, Einwohnerzahlen, Anzahl der Haushalte oder ähnliche Werte anschaulich darstellen.

Aufgrund der stetig steigenden Anzahl von *OSM*-Mitgliedern und der somit steigenden Anzahl von Nutzerbeiträgen (ZIELSTRA & ZIPF, 2009) lassen sich immer mehr Daten zur Generierung von Postleitzahlgebieten im Datenbestand von *OSM* finden. Hier wird insbesondere

¹ <http://www.openstreetmap.org/>

² <http://creativecommons.org/licenses/by-sa/2.0/>

³ <http://www.osgeo.org/>

⁴ http://www.giswiki.org/wiki/Open_Source_GIS

⁵ <http://opensourcegis.org/>

ein Augenmerk auf die Adressdaten gelegt, welche die Postleitzahl eines Punktes wiedergeben.

Es ist angestrebt die Postleitzahlgebiete mit Hilfe der Adressdaten von *OSM* zu erzeugen, sprich eine Gebietsgenerierung aus Punktdaten durchzuführen.

Durch die Nutzung der *OSM*-Daten ist hier eine mögliche Alternative zu Datenprodukten kommerzieller Anbieter wie die *Deutsche Post*⁶, *TeleAtlas*⁷ oder *digital data services*⁸ vorhanden. Dies ist im Besonderen aus wirtschaftlichen Gründen ein interessantes Thema.

Zur Erzeugung dieser Daten wird verschiedene Open Source Software verwendet. Des Weiteren soll diese Erzeugung automatisierbar sein, damit man die vielen Änderungen am Geodatenbestand von *OSM* jederzeit mit einbeziehen kann.

Außerdem wird ein Vergleich der aktuellen Daten aus *OSM* und der aktuellen Daten von *TeleAtlas*, unter Betrachtung diverser Vergleichswerte, erfolgen.

1.2 Aufbau der Arbeit

Die Arbeit untergliedert sich in die Kapitel Theorie, Praxis, Fazit und Ausblick. Im Theorie-Teil wird auf die notwendigen Grundlagen zur Erstellung der Postleitzahlgebiete eingegangen. Das Thema Postleitzahl wird allgemein näher erläutert und über die Darstellung dieser Daten in *OSM* diskutiert. Des Weiteren wird ein kurzer Überblick über die Anbieter solcher Daten gegeben. Neben den Postleitzahlen sind Vergleichsmethoden für Polygondatensätze unterschiedlicher Herkunft ein wichtiges Thema. Dabei wird auf die Qualitätsgrundsätze nach *ISO 19113* (ISO, 2005a) und die Verfahren zur Ermittlung der Datenqualität nach *ISO 19114* (ISO, 2005b) eingegangen. Hierbei gilt es auch aktuelle Literatur bzw. Forschungsarbeiten und deren Methoden zum Thema zu betrachten.

Die automatisierte Gebietsgenerierung aus Punktdaten soll ebenfalls thematisiert werden. Hierfür gibt es diverse Ansätze, die kurz erläutert werden. Der für die Erzeugung genutzte Algorithmus wird dabei detaillierter erklärt.

Als letzter Punkt des Theorie-Teils wird ein Überblick über aktuelle und verfügbare Open Source GIS Software gegeben. Dabei wird auf die leider noch sehr wenig verfügbaren Fachbücher eingegangen.

Der Praxisteil der Arbeit erläutert die Vorgehensweise bei der Erstellung von Postleitzahlgebieten aus *OpenStreetMap*-Daten unter Verwendung von Open Source GIS Software. Es werden anfangs die Voraussetzungen an die Software, aber auch an das System auf welchem die

⁶ <http://www.deutschepost.de>

⁷ <http://www.teleatlas.com/>

⁸ <http://ddsgeo.de/>

Software installiert wird, erläutert. Im Folgenden wird kurz auf den Datensatz und die Besonderheiten dieses Datensatzes, sowie das Testgebiet eingegangen. Anschließend folgt die Beschreibung des allgemeinen Programmablaufes von der Datenvorverarbeitung bis zur Ausgabe der Daten. Nach diesen einleitenden Worten wird die praktische Umsetzung mittels *PostgreSQL*⁹, *PostGIS*¹⁰ und *GRASS GIS*¹¹ beschrieben. Dabei wird auf die einzelnen Schritte des Programmablaufes in Bezug auf diese Software näher eingegangen. Nachfolgend werden die Möglichkeiten der Ausgabe der erzeugten Daten benannt und es wird kurz auf die Ergebnisse eingegangen werden. Im letzten Kapitel des Praxisteils werden die erzeugten Postleitzahlgebiete mit denen des kommerziellen Anbieters *TeleAtlas* verglichen. Hier werden die im Kapitel 2.5.3.2 genannten Vergleichskriterien auf die beiden Datensätze angewendet.

Im nachfolgenden Kapitel wird die Methode der Generierung und deren Ergebnisse unter verschiedenen Punkten diskutiert und ein Fazit gezogen. Am Ende wird ein kurzer Ausblick über zukünftige Arbeiten bzw. Verbesserungen erfolgen und die gesamte Arbeit kurz zusammengefasst.

1.3 Ziel der Arbeit

Das Ziel der Arbeit ist die Erstellung eines Datensatzes der Postleitzahlgebiete Deutschlands aus *OSM*-Daten unter Verwendung von Open Source GIS Software. Hierbei ist ein weitgehend automatisierbares Verfahren zu entwickeln, welches möglichst ohne die Bedienung graphischer Benutzeroberflächen auskommt, sodass im Anschluss an die Arbeit ein vollständig automatischer Einsatz des Verfahrens mit Hilfe eines Skriptes möglich wird. Dies ist von Interesse, da man somit eine schnelle, an die Aktualisierung der *OSM*-Daten angepasste Erstellung der Postleitzahlgebiete ermöglichen kann. Die Ergebnisse dieses Verfahrens sollen vorgestellt und bewertet werden.

Weiteres Ziel der Arbeit ist der Vergleich des erzeugten Datensatzes mit dem Datensatz des kommerziellen Anbieters *TeleAtlas*. Dabei soll auf Flächen- und Entfernungsabweichungen zwischen den Datensätzen eingegangen werden.

Des Weiteren sollen die erzeugten Postleitzahlgebiete mittels Karten im PDF-Format dargestellt und somit deren möglicher Einsatz für weitere Nutzungen diskutiert werden.

⁹ <http://www.postgresql.org/>

¹⁰ <http://postgis.refractor.net/>

¹¹ <http://grass.fbk.eu/>

2 Theorie

2.1 Postleitzahlen

Dieses Kapitel beschreibt die Postleitzahlen und ihre Eigenschaften, sowie die Nutzung dieser. Es wird dabei auf das Postleitzahlensystem von Deutschland eingegangen, da dessen PLZ-Gebiete als Ziel der Arbeit erzeugt werden soll.

Postleitzahlen gibt es in Deutschland bereits seit ca. 150 Jahren. Im Jahr 1961 wurden diese bundesweit vereinheitlicht. Damit besaß Deutschland das erste „vollständige“ Postleitzahlensystem weltweit.

Ein paar Eckdaten der Geschichte der Postleitzahl zeigen die Entwicklung. So wurden 1853 die Ringnummernstempel der Postverwaltung der Thurn und Taxis eingeführt. Diese Stempel ermöglichten die Erkennung eines Ortes an der jeweiligen Zahl. Im Jahr 1941 war die Einführung von Päckchenleitgebieten und 2 Jahre später, im Jahr 1943, gab es eine Anweisung für den Briefverteildienst, in der die Postleitzahlen mit 32 Leitgebieten als verbindlich im Postverkehr eingeführt wurden. Aufgrund wachsenden Postaufkommens wurde im Jahr 1961 das System geändert und es wurden die vierstelligen Postleitzahlen gültig. Nach der Wiedervereinigung Deutschlands kam es 1993 zur Einführung der fünfstelligen Postleitzahlen.

*Dr. Klaus Zumwinkel*¹², kündigte damals an: «Das neue System wird für die Kunden einfacher, übersichtlicher und verwechslungssicher, der Postbetrieb wird zuverlässiger, rationaler und schneller.»

Aufgrund der mehr als 63.000 Änderungen seit dem Jahr 1993 wurde ein neues Postleitzahlenbuch im Jahr 2005 herausgegeben. Ebenso gibt es die Möglichkeit eine CD-ROM mit den Postleitzahlen zu erwerben oder auch die Daten im Internet abzufragen.

Seit 2010 kann man sich die Postleitzahlgebiete auch kartographisch darstellen lassen. Dafür hat die *Deutsche Post* eigens eine dynamische Suche im Internet, Abbildung 2.1, eingerichtet. Diese neue Funktion findet man unter Postleitzahlensuche¹³ auf der Seite der *Deutschen Post* (DEUTSCHE POST, 2011a).

¹² dt. Unternehmensberater, Geschäftsführer der Deutschen Bundespost von 1990-1994, Vorstandsvorsitzender der Deutschen Post von 1995-2008

¹³ <http://www.postdirekt.de/plzserver/>

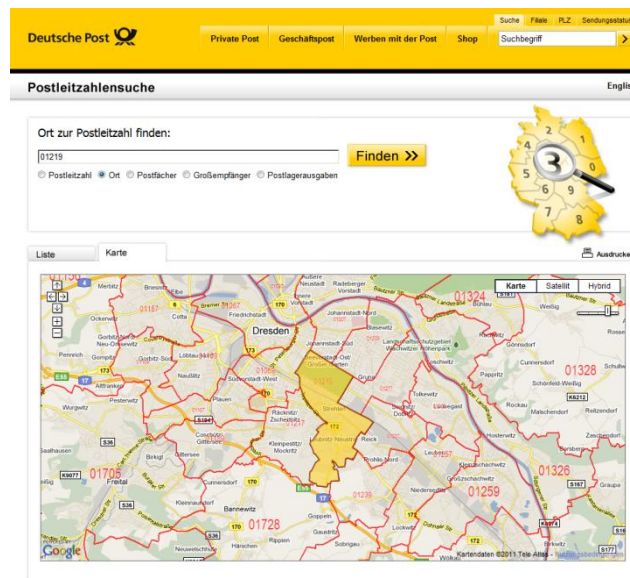


Abb. 2.1 PLZ-Suche Deutsche Post AG (DEUTSCHE POST, 2011b)

Die deutschen Postleitzahlen lassen sich in mehrere Kategorien teilen. So gibt es die einstelligen Postleitzonen (PLZ1), die zweistelligen Postleitregionen (PLZ2) und die fünfstelligen Postleitzahlen (PLZ5). In einigen Bereichen werden auch dreistellige Postleitzahlen (PLZ3) verwendet.

In Deutschland gibt es insgesamt 10 PLZ1-Gebiete. Diese bezeichnen die Postleitzone und damit den Verkehrsflughafen, ab dem die Post ausgeliefert wird. Die Postleitregionen, sprich die PLZ2-Gebiete, geben die Region an in der der Ort liegt, an den die Post geliefert werden soll (GD POSTDIENST, 1993). In Deutschland gibt es laut Daten von *TeleAtlas* aktuell 95 PLZ2-Gebiete. Innerhalb einer Leitregion werden zwischen 20 und 200 Nummern für die Leitbereiche (PLZ5-Gebiete) vergeben. Dabei erhält jede „Gemeinde“ einen Nummernbereich. Begonnen wird mit dem Hauptort des Leitbereichs, danach wird teilweise alphabetisch oder nach Einwohnerzahl verteilt. Die niedrigsten Nummern sind für Postfächer, die Nummern danach für Großempfänger und die höchsten Nummern schließlich für Zustellbezirke vergeben (WIKIPEDIA, 2011a).

Die Postleitzahlengebiete, auch Zustell-Postleitzahlen-Gebiete genannt (DEUTSCHE POST, 2010), sind die PLZ5-Gebiete, welche die wichtigste Funktion, die der Zuweisung der Zustellbezirke, besitzen. In Deutschland gibt es aktuell 8242 PLZ5-Gebiete (TELEATLAS, 2010). Hier entsteht ein Unterschied zur Gesamtzahl der Postleitzahlen der *Deutschen Post*, die bei ca. 29.000 (DEUTSCHE POST, 2010) liegt. Das hängt damit zusammen, dass die Postleitzahlen in verschiedene Gruppen eingeteilt sind. So gibt es zusätzlich PLZ für Postfächer und PLZ für Großkunden (DEUTSCHE POST, 2003), welche innerhalb der jeweiligen PLZ5-Gebiete liegen und somit keine eigenen PLZ5-Gebiete darstellen.

Abbildung 2.2 zeigt die verschiedenen Gebiete im Bereich Sachsen auf Grundlage der *TeleAtlas*-Daten von 2010 (TELEATLAS, 2010). Es sind die PLZ1-, PLZ2- und PLZ3-Gebiete beschriftet. Die PLZ5-Gebiete sind lediglich mittels Grenzlinien dargestellt.



Abb. 2.2 PLZ-Gebiete Sachsen, Daten © TeleAtlas 2010

Eigentlich gibt es diese Postleitzahlengebiete nicht als Fläche, sondern nur als Adressen bzw. Punkte denen eine PLZ zugeordnet ist. Diese lassen sich aber in Flächen zusammenfassen, wobei darauf zu achten ist, dass diese Grenzen dann relativ willkürlich verlaufen. So gibt es teilweise auf der linken Straßenseite andere PLZ als auf der rechten (ARNULF.US, 2011).

Zusammenfassend lässt sich zu den fünfstelligen Postleitzahlen sagen: *«Die ersten beiden Ziffern ermöglichen es, einen Ort geografisch einzuordnen. Sie bezeichnen die Zielregion und damit das Briefzentrum einer Region. Die Ziffern drei bis fünf zeigen an, wo der Kunde wohnt, ob er seine Post per Zustellung oder Postfach erhält und ob er Großkunde ist.»* (DEUTSCHE POST, 2011a)

2.2 Nutzung von Postleitzahlen in der Privatwirtschaft

Neben der Nutzung der Postleitzahlen für die Zustellung von Sendungen werden die Postleitzahlgebiete auch häufig in der Privatwirtschaft genutzt. Hier ergibt sich eine für viele Anwendungsfelder günstige Gebietseinteilung nach diesen Gebieten. So dienen die PLZ-Gebiete vor allem für Geomarketing und Vertriebsplanungen. Ebenso lassen sich anhand dieser Gebiete Geschäftsdaten und viele weitere Daten visualisieren, die sich auf Adressen beziehen. Dies erleichtert die Orientierung im Raum (DDS, 2011). Die PLZ-Gebiete werden als eine Art sta-

tistische Bezirke zur Planung oder Veranschaulichung diverser Daten wie Anzahl der Haushalte, Anzahl der Minderjährigen oder ähnlichen Themen verwendet.

Hier sei erwähnt, dass die Postleitzahlen zwar intern als Zustellsystematik der *Deutschen Post* gedacht sind, sich aber im Laufe der Jahre ein immer größer werdendes Interesse der Privatwirtschaft an diesen Daten eingestellt hat. Dabei war es nicht die *Deutsche Post* die als erstes Postleitzahlgebiete (Flächen) erstellt haben. In diesem Bereich waren die kommerziellen Anbieter wie *TeleAtlas* oder auch *digital data services* die ersten Anbieter von digitalen Daten solcher Flächen.

Der Vorteil der Postleitzahlenflächen liegt darin, dass diese Gebiete eindeutig sind. Des Weiteren bilden diese Flächen ein heterogenes, feinmaschiges Netz. An diesem Netz lassen sich sogar die wirtschaftlichen Ballungsräume eines Landes erkennen. So ist in dicht besiedelten Gebieten die Einteilung feiner als in unbewohnten Bereichen. Außerdem entsprechen die Grenzen der PLZ-Gebiete meist der Topographie, da sie in der Regel entlang von Straßen, Flüssen oder Ortsteilen verlaufen. Ein weiterer Punkt der Eignung von Postleitzahlen für die Nutzung in der Privatwirtschaft ist die hierarchische Einteilung dieser. So zeigt die erste Stelle eine gröbere Gebieteinteilung als zum Beispiel die vollständige Postleitzahl (WIKIPEDIA, 2011b).

Neben der mittlerweile durch die *Deutsche Post* angebotenen geocodierten Flächen der Postleitzahlgebiete in Deutschland gibt es auch weitere Anbieter die diese Daten verkaufen. Die *Deutsche Post* bietet die Daten über das Produkt DATAFACTORY GEOCODE¹⁴ an. Dabei erhält der Kunde die geocodierten Flächen der Postleitzahlgebiete sowie die Leitregionen und Leitzonen. Dieser Datensatz enthält rund 8.200 PLZ-Gebiete und wird jährlich aktualisiert. Er wird als Vektordatensatz im *MapInfo*- oder Shape-Dateiformat ausgeliefert (DEUTSCHE POST, 2010). *DDS* oder auch *TeleAtlas* liefern ebenfalls PLZ-Gebiete in den heutzutage üblichen Formaten aus, welche jährlich aktualisiert werden. Dabei besitzt jedes Unternehmen seinen eigenen Datensatz. Die Deutsche Post erklärte auf Anfrage (E-Mail von Frank Krämer, Deutsche Post, Team Datenservice, Januar 2011), dass die PLZ5-Gebiete durch die Post selbst erzeugt wurden und nicht auf einen bereits vorhandenen Datensatz der anderen Anbieter zurückgegriffen wurde. Dies wäre möglich gewesen, da es erst seit kurzem dieses Angebot bei der *Deutschen Post* gibt.

Durch die eigene Erzeugung dieser Datensätze durch die jeweiligen Anbieter lassen sich Unterschiede in der Anzahl und in der Abgrenzung der PLZ-Gebiete feststellen. Diese Unterschiede sollen in dieser Arbeit nicht betrachtet werden.

¹⁴ http://www.deutschepost.de/dpag?tab=1&skin=lo&check=no&lang=de_DE&xmlFile=link1020331_1020325

2.3 Postleitzahlen in OpenStreetMap

Bevor auf die Darstellung der Postleitzahlen und deren Besonderheiten in *OSM* eingegangen wird, will ich kurz die Besonderheiten der Postleitzahlen in der Datenverarbeitung erläutern. Postleitzahlen sollten stets als Zeichenketten angesehen werden, da sie nicht nur aus Ziffern sondern in anderen Ländern auch aus Buchstaben und/oder Sonderzeichen bestehen. Ein weiterer Grund für die Speicherung der Postleitzahlen als Zeichenketten ist die führende Null, diese würde bei einer Speicherung als Zahl verloren gehen. Außerdem hat der numerische Wert einer Postleitzahl keine Funktion, da arithmetische Operationen auf Postleitzahlen keinen Sinn ergeben (WIKIPEDIA, 2011b). Somit werden Postleitzahlen, wie auch in *OpenStreetMap*, immer als Text gespeichert.

In *OpenStreetMap* gibt es drei verschiedene Objekttypen. Es gibt Knoten (Nodes), Wege (Ways) und Relationen (Relations). Aus diesen drei Objekttypen bestehen alle Objekte des Datenbestandes von *OpenStreetMap*, egal ob Straßen, Bahnhöfe oder Waldflächen. Diese Daten sind in einem eigenen *OSM*-Dateiformat, mit der Endung *.osm*, verfügbar. Dabei handelt es sich um ein *XML*-Format. Jedes Objekt besitzt dabei eine eigene *XML*-Repräsentation (LUDWIG, 2010).

Damit die Daten eindeutig sind, erhält jedes Objekt bei der Erstellung eine ID, einen Zeitstempel und den Namen des Benutzers, der dieses Objekt erstellt hat. Relationen (Relations) dienen zur Gruppierung von einzelnen Elementen, um zum Beispiel aus mehreren Wegen eine Buslinie zu erstellen. Um die verschiedenen Objekte in ihren Attributen zu unterscheiden und um vor allen die Darstellung innerhalb der Karte festzulegen, werden sogenannte Tags vergeben. Diese Tags geben an, um was es sich bei den Objekten handelt, sprich ob es eine Autobahn, eine Straßenbahnhaltestelle oder Ähnliches ist, wie der Name dieses Objektes ist und diverse andere Dinge, wie zum Beispiel die zulässige Höchstgeschwindigkeit auf einer Straße. Auf Basis dieser zusätzlich verfügbaren Informationen eines geographischen Objektes wurden neben der *OpenStreetMap*-Standardkarte noch weitere nützliche Karten, wie *OpenSeaMap*¹⁵, *OpenCycleMap*¹⁶ oder auch die *ÖPNV-Karte*¹⁷, erstellt.

Zu diesen Tags gehört auch die Angabe der Postleitzahl. Diese lässt sich über das Tag *postal_code*¹⁸ oder über das Tag *addr:postcode*¹⁹ angeben. Dabei ist vorgesehen, dass man das

¹⁵ <http://openseamap.org>

¹⁶ <http://opencyclemap.org>

¹⁷ <http://www.opnvkarte.de>

¹⁸ http://wiki.openstreetmap.org/wiki/Key:postal_code

Tag *postal_code* nutzt, um einem Gebiet, einem Gebäude oder anderen Objekten eine Postleitzahl zuzuordnen. Das Tag *addr:postcode* sollte hierbei nur alternativ genutzt werden, ist aber mittlerweile auch stark verbreitet (OSM-WIKI²⁰, 2011). Einem Weg ist beispielweise das Tag *postal_code=01219* oder auch *addr:postcode=01219* zugeordnet. Das heißt, der Weg besitzt die Postleitzahl 01219. Daraus lässt sich schließen, dass sich der Weg in Dresden-Strehlen befindet.

In *OpenStreetMap* gibt es weltweit aktuell 334.645 Objekte mit dem Tag *postal_code*. Dabei werden 48.595 Werte dargestellt. Diese Summe teilt sich in 84.528 Knoten (38.602 Werte), 239.060 Wege (6.758 Werte) und 11.057 Relationen (9.924 Werte) (Stand 23.06.2011, TAGINFO OPENSTREETMAP²¹, 2011). Hier sei erwähnt, dass die Anzahl der Werte nicht die Anzahl unterschiedlicher Postleitzahlen angibt, sondern die Anzahl der jeweiligen Objekte, denen eine Postleitzahl zugeordnet ist. So ist es häufig der Fall, dass innerhalb eines PLZ-Bereiches mehrere Punkte und Wege liegen, da es beispielweise mehrere Straßen innerhalb eines PLZ-Bereiches gibt, welche jeweils gezählt wurden. Ebenfalls sei hier gesagt, dass es auch viele Straßen gibt, die durch mehrere unterschiedliche Postleitzahlgebiete verlaufen.

Neben den zwei genannten Tags, welche für die Erstellung der Postleitzahlgebiete genutzt wurden, gibt es seit kurzer Zeit in *OSM* zusätzlich auch die PLZ-Gebiete. Hierbei handelt es sich um Flächen, welche zusätzlich zum Tag *postal_code* noch das Tag *boundary=postal_code* besitzen. Dieser Ausdruck bedeutet, dass dem Objekt das Attribut *Grenze/Grenzlinie=Postleitzahl* zugeordnet wurde. Es handelt sich also um eine Abgrenzung des PLZ-Gebietes. Um diese Postleitzahlgebiete in *OpenStreetMap* einzuspielen, wurde durch *Arnulf Christ*²² ein Ende 1999 erworbener Datensatz von nicht genanntem Anbieter käuflich erworben und als *Public Domain*²³, mit dem Hinweis, dass die Daten weiterhin als *Public Domain* verfügbar sind, zum Download²⁴ als Shapefile bereitgestellt (ARNULF.US, 2011). Auf Basis dieser Daten wurde dann ein Arbeitsplan durch *OSM* erstellt, der die Einarbeitung dieser Daten vorantreiben soll. Hierfür wurde in der *OSM-Wiki* eine Anleitung²⁵ geschrieben, die den Nutzern erklärt wie sie die Daten importieren sollen. Insbesondere wurde erklärt, wie Sonderfälle, zum Beispiel unterschiedliche PLZ auf den unterschiedlichen Straßenseiten, zu

¹⁹ <http://wiki.openstreetmap.org/wiki/Key:addr:postcode>

²⁰ <http://wiki.openstreetmap.org>

²¹ <http://taginfo.openstreetmap.org>

²² <http://arnulf.us>

²³ <http://creativecommons.org/licenses/publicdomain/>

²⁴ <http://www.metaspacial.net/download/plz.tar.gz>

²⁵ http://wiki.openstreetmap.org/wiki/Import/Catalogue/Postleitzahlen_Deutschland_2010

behandeln sind. Dabei wurde ausdrücklich erwähnt, dass der Datensatz veraltet und ungenau ist und deshalb nicht einfach übernommen werden soll, sondern manuell eingearbeitet und neu angepasst werden soll. Der vorhin genannte Arbeitsplan dient zur Einteilung des Imports in verschiedene PLZ-Bereiche, geordnet nach den Postleitzonen. Dieser Plan enthält die unterschiedlichen PLZ-Bereiche mit zugeordnetem Bundesland, Landkreis, Bearbeiter, Fortschritt und Bemerkungen. Zum heutigen Stand sind, bis auf wenige Bereiche, vor allem in Schleswig-Holstein und Niedersachsen, diese Daten komplett importiert und können genutzt werden.

Der Import dieser Daten sei hier nur der Vollständigkeit halber erwähnt. Für die Generierung der Postleitzahlgebiete wurden diese Daten nicht verwendet. Die Generierung erfolgte ausschließlich aus Daten, die nur die vorher genannten Tags besitzen, damit man die PLZ-Gebiete auch für alle anderen Länder mit Hilfe dieser Methode erstellen kann.

Für die genutzten Daten gibt es in *OSM* keine genauen Quellenangaben. In der *OSM*-Wiki wird zum Thema Quellen der Postleitzahl Daten erwähnt, dass es im Normalfall so sein sollte, dass Postleitzahl Daten durch eine staatliche Einrichtung verwaltet werden und frei zugänglich sein sollten. Leider ist das aber in den meisten Ländern nicht der Fall, weshalb diese Daten, wie auch die anderen Geodaten in *OSM*, nach der bekannten Arbeitsweise gesammelt werden und nicht von anderen Quellen kopiert werden. Daten von anderen Herstellern können nur benutzt werden wenn dies durch deren Lizenz ausdrücklich gestattet ist.

Weiterhin wird explizit erwähnt, dass man die PLZ-Suche der *Deutschen Post* (DEUTSCHE POST, 2011b) nicht nutzen darf um die PLZ für eine bestimmte Straße zu finden. Diese Suche darf nur zur Kontrolle verwendet werden.

Als gutes Beispiel einer Quelle für Postleitzahl Daten bzw. deren Sammlung sei Großbritannien genannt. Hier wurde das Projekt *Free the Postcode*²⁶ gestartet, wobei es sich um eine Datenbank für Postleitzahlen handelt. Mit Hilfe des Projektes sollen Postleitzahlen gesammelt werden. Weiterhin lassen sich Postleitzahlen in geographische Koordinaten und zurück umwandeln. Dies war zum damaligen Zeitpunkt bei kommerziellen Anbietern sehr teuer. Aufgrund des Druckes durch dieses Projekt auf die proprietären Anbieter der PLZ-Daten, hat die amtliche Landesvermessung Großbritanniens, *Ordnance Survey*²⁷, die Daten für die Postleitzahlen innerhalb ihres offenen Datensatzes freigegeben (OSM-WIKI, 2011a). So gibt es in Großbritannien das Produkt *Code-Point Open*. Dieses Produkt enthält allerdings nur Punkte und keine Polygone (ORDNANCE SURVEY, 2011), weshalb es für die direkte Nutzung

²⁶ <http://www.freethepostcode.org/>

²⁷ <http://www.ordnancesurvey.co.uk/>

nicht sehr sinnvoll erscheint, da die PLZ-Gebiete erst erzeugt werden müssen. Diese kann man allerdings mit der in dieser Arbeit erläuterten Methode erzeugen.

Durch die Freigabe dieser Daten verliert das Projekt *Free the Postcode* zwar an Wert, aber immerhin hat es dazu geführt, dass diese Daten überhaupt veröffentlicht wurden. Ebenso werden Daten für Großbritannien über die *Open Postcode Database*²⁸ gesammelt und zum Download bereitgestellt (FREE THE POSTCODE, 2011). In diesem Zusammenhang sei die Seite <http://random.dev.openstreetmap.org/postcodes/> (RANDOMJUNK, 2011) genannt, welche die Postleitzahlgebiete Großbritanniens auf einer interaktiven Karte, Abbildung 2.3, darstellt. Die Daten werden dabei wöchentlich aktualisiert und es kann zwischen den verschiedenen Quellen, wie *OSM*, *Free the Postcode* oder auch dem *Code-Point Open*-Datensatz gewählt werden, aus deren Daten dann jeweils die PLZ-Gebiete dargestellt werden. In Deutschland gibt es kein vergleichbares Projekt.

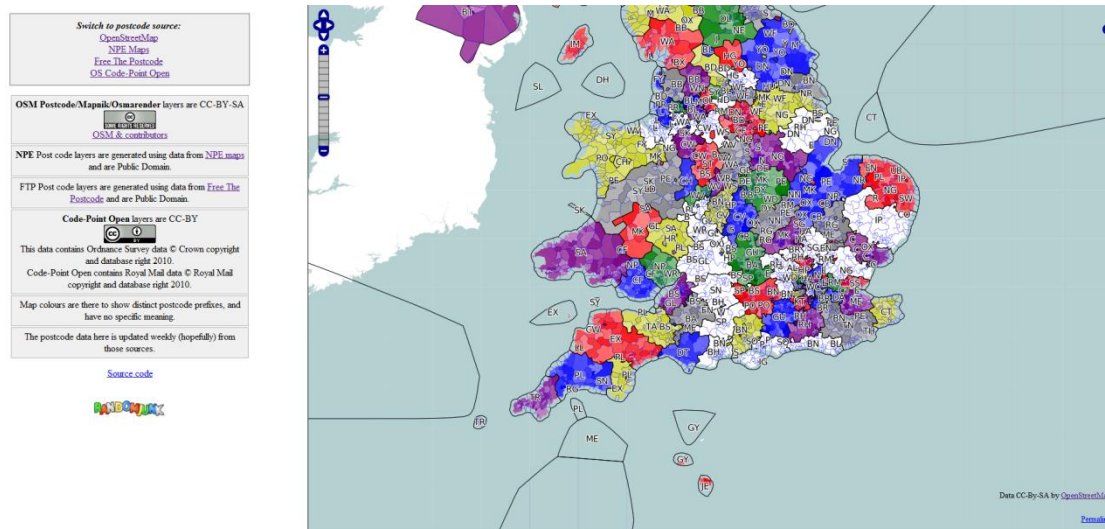


Abb. 2.3 Karte zu PLZ-Gebieten in Großbritannien (RANDOMJUNK, 2011)

Die Postleitzahlendaten von *OSM* stehen wie alle anderen freien Daten von *OSM* unter der *Creative Commons Attribution-Share-Alike 2.0-Lizenz (CC BY-SA)*. Die deutsche Bezeichnung für diese Lizenz ist: *Namensnennung-Weitergabe unter gleichen Bedingungen*. Die Lizenz erlaubt eine Vervielfältigung, Verbreitung und öffentliche Freigabe der Daten. Es ist ebenso gestattet Abwandlungen und Bearbeitungen der Daten anzufertigen. Dies geschieht unter der Bedingung, dass man den Namen des Autors/Rechteinhabers nennt und das selbst erstellte Werk unter der gleichen oder einer vergleichbaren Lizenz steht (CREATIVE COMMONS, 2011). Im Falle von *OpenStreetMap* ist mindestens © *OpenStreetMap und Mitwirkende, CC BY-SA* als Quelle anzugeben. Wenn man ausschließlich Geodaten nutzt, so ist mindestens

²⁸ <http://postcodedb.sourceforge.net/>

Geodaten © *OpenStreetMap und Mitwirkende*, *CC BY-SA* anzugeben. Des Weiteren sollte ein Hyperlink auf *OpenStreetMap* und die Lizenz *CC BY-SA*²⁹ gesetzt werden.

Aktuell ist eine Diskussion über die Änderung der Lizenz im Gange. Dabei sollen die Daten zukünftig unter der *Open Database License (ODBL)* veröffentlicht werden. Diese Lizenz baut nicht auf dem Urheberrecht, sondern auf dem europäischen Datenbankrecht auf (RAMM & TOPF, 2010).

Da es sich bei *OpenStreetMap* um ein Internetprojekt handelt, welches mit Hilfe einer großen Community versucht individuell erhobene geographische Daten oder erstellte Karten über das Internet verfügbar zu machen und über das Internet auszutauschen (GOODCHILD, 2007), spricht man in diesem Sinne auch von nutzergenerierten Daten bzw. nutzergeneriertem Inhalt (user generated content). Vorreiter der user generated content-Projekte ist *Wikipedia*³⁰. In Bezug auf user generated maps, wobei es sich in einem gewissen Sinne bei *OSM* handelt, spricht HOFFMANN (2011) von „kollaborativen“, durch mehrere Nutzer erstellte, und „personalisierten“, nur von einer Person erstellte, Karten (HOFFMANN, 2011). Neben dem nutzergenerierten Inhalt handelt es sich bei *OSM*-Daten um Geodaten aus einer gewollten Bereitstellung individuell erstellter Informationen, *Volunteered Geographic Information* (GOODCHILD, 2007). Diese Begriffe werden innerhalb des Web 2.0 häufig angewendet und beschreiben eine neue Form der Informationsgewinnung über die Internet-Community, welche aufgrund der Individualität und Bandbreite große Chancen bietet. Allerdings stellt sich die Frage nach der Qualität dieser Daten (MAUÉ & SCHADE, 2008).

Unter Web 2.0 versteht man eine substantielle Veränderung der Art und Weise wie das Internet von Nutzern und Entwicklern verwendet werden kann. Hauptpunkte sind die Verwendung kollektiver Intelligenz, die Architektur der Partizipation, einfache Programmiermodelle und offene Standards sowie die Geräteunabhängigkeit (GARTNER & SCHMIDT, 2010).

Bezüglich der Verarbeitung von raumbezogenen Informationen innerhalb des Internets gab es einen Wandel vom Webmapping 1.0 zum Webmapping 2.0. Dabei wandelte sich der Kartennutzer zum Kartenproduzenten. Es ist dabei auch oft von Prosumenten, eine Mischung aus Konsument und Produzent, die Rede (HOFFMANN, 2010). *OSM* ist hierfür das beste Beispiel. Webmapping 2.0 ist dabei gekennzeichnet durch offene Programmierschnittstellen (APIs), standardisierte Datenformate (z.B. *KML*, *GeoRSS*, *GPX*) und eine einfache Bedienung und Handhabung. Als weiteres wichtiges Merkmal von Webmapping 2.0 sei die Möglichkeit

²⁹ <http://creativecommons.org/licenses/by-sa/2.0/>

³⁰ <http://wikipedia.org>

der einfachen Einbindung selbst erzeugter Karten in Webseiten genannt (HOFFMANN, 2010).

Wie bereits angemerkt, stellt sich die Frage der Qualität oder auch der Nutzbarkeit, solcher im Web 2.0 bereitgestellten Daten und Informationen. So gibt es einige Vor- und Nachteile von nutzergenerierten Inhalten allgemein und auch speziell in Bezug auf *OSM*. Vorteile sind die geringen Kosten, die Werbung für ein Projekt innerhalb der Internetcommunity und die Aktualität der Daten. Die Aktualität ist dabei der entscheidende Punkt, da die Nutzer Änderungen sehr schnell einspielen. Als Profiteur der sehr hohen Aktualität der Daten sei das Crisis Mapping genannt. So konnte man beobachten, wie nach dem Erdbeben in Haiti³¹ im Januar 2010 in kürzester Zeit eine Karte auf Basis von *OSM* mit Informationen zur Lage im Katastrophengebiet erstellt wurde (HOFFMANN, 2011).

Leider hat diese Art der Informationssammlung auch Nachteile. So wird zum Beispiel von vielen kritisiert, dass nur das gesammelt wird was auch den einzelnen Nutzer interessiert. Weitere Nachteile sind die Möglichkeit des Vandalismus, so besteht die Gefahr, dass die Daten gelöscht oder manipuliert werden und es somit zu einer Verfälschung der Ergebnisse kommt (POLCHAU, 2009). Das Hauptproblem liegt allerdings in der nicht vorhandenen systematischen Qualitätskontrolle. Eventuelle Fehler werden zwar durch die große Community entdeckt, dies geschieht aber leider meist nur zufällig.

Die Qualität von Geodaten bzw. räumlichen Daten im Allgemeinen, die Qualitätskriterien für Geodaten und die Qualität von Daten innerhalb *OSM* sollen in den nächsten Kapiteln diskutiert und aktuelle Forschungsergebnisse betrachtet werden. Ebenso sollen verschiedene Vergleichsmethoden für Daten unterschiedlicher Herkunft erläutert werden. Dabei soll speziell auf den Vergleich von Polygondatensätzen Bezug genommen werden, da im praktischen Teil die aus *OSM*-Daten erstellten PLZ-Polygone mit den Daten von *TeleAtlas* verglichen werden sollen.

2.4 Qualität von Geodaten

Seit der Entwicklung geographischer Informationssysteme und der Nutzung von räumlichen Daten spielt die Qualität dieser Daten eine immer größer werdende Rolle. Aufgrund der unterschiedlichen Erzeugung der Daten durch verschiedene Personen, Institutionen und Firmen ist es notwendig einen gemeinsamen Qualitätsstandard zu nutzen, damit diese Daten sinnvoll verwendet werden können.

³¹ <http://haiti.openstreetmap.nl>

Es gibt bereits mehrere Normen, die versuchen die Qualitätsbeschreibung für Geodaten zu standardisieren. So wurden 2002 von der *Internationalen Normungsorganisation*³² (ISO) einige Elemente für die Qualitätsbeschreibung von Geodaten standardisiert. Diese Elemente sind in der *ISO 19113 Geoinformation – Qualitätsgrundsätze* (ISO, 2005a) definiert. Um die Qualität der Daten zu ermitteln wurde zusätzlich die *ISO 19114 Geoinformation – Verfahren zur Ermittlung der Datenqualität* (ISO, 2005b) erstellt. Seit 2006 gibt es neben diesen Normen noch Qualitätsmaße zur Bewertung der Qualitätselemente, welche in der *ISO 19138 Geoinformation – Maßeinheiten der Datenqualität* (ISO, 2006) beschrieben sind (LUDWIG, 2010). Aufbauend auf diesen Normen wurde vom *Deutschen Dachverband für Geoinformation*³³ (DDGI) ein Qualitätsmodell für die Beschreibung von Geodaten entwickelt. Dieses Modell wurde in der *PAS 1071*, einem deutschsprachigen Dokument, niedergeschrieben. Hierbei handelt es sich um eine Publicly Available Specification (PAS). Mittels der *PAS 1071* soll eine Struktur hergestellt werden, die zur einfachen und einheitlichen Beschreibung der Qualität von Geodatenprodukten dient. Insbesondere wird die Beziehung zwischen Hersteller, Anbieter und Nutzer berücksichtigt (LUDWIG, 2010). Die *PAS 1071* orientiert sich an Inhalten der Normenfamilie *ISO 191XX Geoinformation*. In der Norm wird die Qualität eines Geodaten-satzes durch quantitative Merkmale (Qualitätselemente) und durch beschreibende Merkmale (Metadaten) beschrieben. Beschreibende Merkmale sind dabei die Angabe der Produktklasse und die Produktbeschreibung, welche als Metadaten bezeichnet werden. Die Produktbeschreibung enthält Angaben über den Hersteller und die Herkunft der Daten sowie Produktmerkmale. Unter quantitativen Merkmalen sind hauptsächlich die Qualitätsgrundsätze aus der *ISO 19113* (ISO, 2005a) aufgeführt. Dies sind die Positionsgenauigkeit, die thematische Genauigkeit, die Vollständigkeit, die logische Konsistenz und die zeitliche Genauigkeit der Geodaten. In folgenden Tabellen sollen diese Qualitätsmerkmale näher erläutert werden. Die Tabellen wurden auf Basis der *ISO*-Normen erzeugt und in der *PAS 1071* zusammengefasst wiedergegeben.

Tabelle 2.1 zeigt die verschiedenen Subelemente der Positionsgenauigkeit. Die Positionsgenauigkeit ist das Maß der Übereinstimmung zwischen dem Ergebnis der Positionsbestimmung und dem wahren Wert der Position eines Objekts (PAS 1071, 2007).

³² <http://www.iso.org/>

³³ <http://www.ddgi.de/>

| | |
|--|--|
| Absolute (äußere) Positionsgenauigkeit | Maß der Übereinstimmung des festgestellten Koordinatenwertes mit dem wahren oder als wahr angenommenen Koordinatenwert |
| Relative (innere) Positionsgenauigkeit | Maß der Übereinstimmung der festgestellten relativen Positionen von <i>Objekten</i> in einem <i>Datensatz</i> zueinander mit ihren entsprechenden wahren oder als wahr angenommenen relativen Positionen |
| Rasterdatengenauigkeit | Maß der Übereinstimmung von <i>Rasterdaten</i> positionswerten mit den wahren oder als wahr angenommenen Werten |

Tab. 2.1 Subelemente zur Positionsgenauigkeit (PAS 1071, 2007)

In Tabelle 2.2 werden die Subelemente zur thematischen Genauigkeit näher erläutert. Unter thematischer Genauigkeit ist die Richtigkeit der Klassifizierung, die Richtigkeit nichtquantitativer und die Genauigkeit quantitativer Attribute zu verstehen (PAS 1071, 2007).

| | |
|--|--|
| Richtigkeit der Klassifikation | Vergleich der den Objekten oder ihren Attributen zugewiesenen Klassen mit einer allgemein anerkannten Aussage oder mit einem Referenzdatenbestand Beispiel: Zuordnung einer Straße zu Verkehrswegen und nicht zu Vegetation |
| Richtigkeit nichtquantitativer Attribute | Richtigkeit nichtquantitativer Attribute Beispiel: Nutzungsart, wie Wohngebäude oder gewerbliches Gebäude |
| <i>Genauigkeit</i> quantitativer Attribute | <i>Genauigkeit</i> quantitativer Attribute Beispiel: Höhe eines Gebäudes |

Tab. 2.2 Subelemente zur thematischen Genauigkeit (PAS 1071, 2007)

Tabelle 2.3 erläutert die Subelemente der Vollständigkeit. Datenüberschuss wird in diesem Zusammenhang auch als *Error of commision*, Datenmangel als *Error of ommision* bezeichnet.

| | |
|-----------------|---|
| Datenüberschuss | Der <i>Datensatz</i> enthält mehr Daten als (in der zugehörigen Spezifikation) angegeben |
| Datenmangel | Der <i>Datensatz</i> enthält weniger Daten als (in der zugehörigen Spezifikation) angegeben |

Tab. 2.3 Subelemente zur Vollständigkeit (PAS 1071, 2007)

In Tabelle 2.4 werden die Subelemente der logischen Konsistenz näher erläutert. Unter logischer Konsistenz ist der Grad der Einhaltung der logischen Regeln der Datenstruktur, der Attributierung und der Beziehungen, die in der zugehörigen Spezifikation vorgegeben sind, zu verstehen (PAS 1071, 2007).

| | |
|--------------------------------|--|
| Konzeptuelle <i>Konsistenz</i> | Grad der Einhaltung der Regeln des <i>konzeptuellen Schemas</i> Beispiel: Bei der Aktualisierung einzelner <i>Objekte</i> eines <i>Datensatzes</i> müssen die neu eingefügten <i>Objekte</i> der Spezifikation der ursprünglichen Daten entsprechen. Ein aktualisiertes Gebäude eines 3D-Stadtmodells darf nicht mit erheblich abweichender <i>Genauigkeit</i> erfasst werden. |
| Wertekonsistenz | Grad der Einhaltung der in den Wertebereichen angegebenen Werte Beispiel: Bei einer Bevölkerungskartierung dürfen keine negativen Werte auftreten, die nicht explizit mit einer besonderen Bedeutung spezifiziert sind. |
| Formatkonsistenz | Grad, mit dem der <i>Datensatz</i> in Übereinstimmung mit der physikalischen Datenstruktur gespeichert ist Beispiel: Nicht belegte Werte müssen konsistent gekennzeichnet sein, entweder als „NULL“-Werte (fehlender Wert), als „Blank“ (Leerzeichen) oder als „0“. |
| Topologische <i>Konsistenz</i> | Richtigkeit der explizit codierten topologischen Charakteristiken des <i>Datensatzes</i> Beispiel: Die Nachbarschaftsbeziehungen einzelner Datenelemente müssen auch dann erhalten bleiben, wenn diese Elemente einem Veränderungsprozess, insbesondere einem Generalisierungsprozess, unterworfen worden waren. So darf der nördlich einer Straße stehende Fernsehturm nach einer Generalisierung nicht südlich der Straße stehen. |
| Geometrische <i>Konsistenz</i> | Grad der Übereinstimmung des <i>Geodatensatzes</i> mit den geometrischen Bedingungen der zugehörigen Spezifikation Beispiele: Folgende Aussagen müssen zutreffen: „Es gibt keine doppelten Digitalisierungspunkte“, „Jedes <i>Objekt</i> der Realität existiert im <i>Datensatz</i> nur einmal“. |
| Thematische <i>Konsistenz</i> | Grad der Übereinstimmung des <i>Geodatensatzes</i> mit den semantischen Bedingungen in der zugehörigen Spezifikation Beispiele: Linienpolygone, die eine Fläche bilden, müssen geschlossen sein. Kanten eines Vektornetzes müssen mit einem Knoten beginnen. |

Tab. 2.4 Subelemente zur logischen Konsistenz (PAS 1071, 2007)

Tabelle 2.5 zeigt die Subelemente der zeitlichen Genauigkeit. Die Angabe der Zeit, sprich des Aufnahmedatums oder auch die zeitliche Genauigkeit sind wichtige Angaben über die Geodaten.

| | |
|------------------------------------|--|
| <i>Genauigkeit</i> der Zeitmessung | <i>Genauigkeit</i> der zeitlichen Angabe zu einem <i>Datensatz</i> oder einem <i>Objekt</i> Beispiel: taggenau, monatgenau, Jahrhundert |
| Zeitliche <i>Konsistenz</i> | Richtigkeit der zeitlichen Ereignisse und Abfolgen Beispiel: Reihenfolge der Pegelstände |
| Zeitliche Gültigkeit | Richtigkeit des <i>Datensatzes</i> in Bezug auf den geforderten Zeitpunkt Beispiel: Zeitangabe vom Typ <Jahr-Monat-Tag> |

Tab. 2.5 Subelemente zur zeitlichen Genauigkeit (PAS 1071, 2007)

Diese Qualitätsmerkmale, mit ihren jeweiligen Subelementen, spielen für den Vergleich von Geodaten eine große Rolle. Einige der Merkmale, wie die Positionsgenauigkeit und die Vollständigkeit, werden, soweit dies möglich ist, für den späteren Vergleich der erzeugten Postleitzahlgebiete angewandt.

Ebenfalls in diesen Normen werden die Möglichkeiten zur Qualitätsevaluation beschrieben. Diese Evaluation soll die jeweiligen Qualitätskriterien näher untersuchen und eine objektive Beschreibung der Datenqualität liefern. Der erste Schritt einer Qualitätsevaluation ist, nachdem man sich für einen Datensatz und einen geographischen Bereich entschieden hat, die Auswahl der Qualitätskriterien die man überprüfen will. Hat man diese Qualitätskriterien gewählt, muss man sich Qualitätsmaße und eine Qualitätsmethode auswählen. Qualitätsmaße dienen hierbei zu Messung der Qualität. Hier ist je nach Qualitätskriterium ein passendes Maß zu wählen. So empfiehlt sich zum Beispiel für die Positionsgenauigkeit die Differenz zwischen dem Wert des zu untersuchenden Datenbestandes und dem passenden wahren Wert oder Referenzwert (HAINING, 2005). Diese Differenz wird dann in einer Längeneinheit angegeben. Nach der Berechnung des Qualitätsmaßes wird schließlich darüber entschieden, ob die Qualität ausreichend ist und der „Test“ bestanden wurde. Dieser ganze Ablauf wird auch als *fitness-for-use* bezeichnet. So geht es darum, ob die gewählten Daten „fit“ genug sind um für die jeweilige Anwendung genutzt werden zu können.

Weitere Qualitätsmaße und deren Berechnung sowie verschiedene Methoden zur Kontrolle bzw. Beschreibung der Qualität werden im Kapitel 2.5 Vergleich von Daten unterschiedlicher Herkunft erläutert.

Neben diesen Normen gibt es mittlerweile auch zahlreiche Fachliteratur zum Thema Qualität georäumlicher Daten. Ein oft zitiertes Werk zu diesem Thema stammt von GUPTILL in Zusammenarbeit mit der *International Cartographic Association*³⁴ (ICA) (1995). In dem Werk, *Elements of spatial data quality* (GUPTILL & ICA, 1995), werden die verschiedenen Qualitätselemente beschrieben. Das Buch präsentiert die Ergebnisse jahrelanger Arbeit der *Commission for Spatial Data Quality* der ICA, welche als Grundlage für die ISO-Normen genutzt wurden.

Des Weiteren beschäftigen sich DEVILLERS und JEANSOULIN in *Spatial Data Quality: Concepts* (DEVILLERS & JEANSOULIN, 2006) mit vielen Bereichen, in denen die Qualität von Geodaten relevant ist. Es wird das Konzept der Qualität von Geodaten, theoretische

³⁴ <http://icaci.org/>

Grundlagen, aber auch Anwendungsbeispiele beschrieben. Bei dieser Literatur handelt es sich um eine Sammlung verschiedener Kapitel, die durch Spezialisten auf ihrem Gebiet geschrieben und durch die zwei Autoren zusammengetragen wurden. Dabei wird für beide Sichtweisen, die des Datenproduzenten und die des Datennutzers, erklärt wie man die Qualität von Vektor- und Rasterdaten evaluiert.

Ein weiteres wichtiges Werk zum Thema Qualität von räumlichen Daten *Spatial data quality* (SHI ET AL., 2002) stammt von SHI, FISHER und GOODCHILD. In diesem Buch werden die Standardtechniken zur Analyse von räumlichen Daten betrachtet. Es ist dabei in drei Teile gegliedert: Teil 1 beinhaltet die Theorie bezüglich der Qualität, Teil 2 die Methoden zur Handhabung der Qualität räumlicher Daten und Teil 3 verschiedene Anwendungsbeispiele. Das Werk gibt den aktuellen Stand der Forschung zur Qualität räumlicher Daten wieder (SHI ET AL. 2002).

Eine weitere Arbeit stammt von JOOS (1999), der in seiner Dissertation die Qualität von objektstrukturierten Geodaten näher betrachtet. Dabei wird die Modellierung der Daten und deren Metadaten erläutert. Des Weiteren stellt er ein Qualitätsmodell auf und gibt Hinweise zum Qualitätsmanagement bei der Datenerfassung. Joos erläutert ebenso Konsistenzprüfungen und Stichprobenprüfungen zur Kontrolle der Qualität von Geodaten (JOOS, 1999).

Auch VAN OORT schreibt in seiner Dissertation *Spatial data quality: from description to application* (VAN OORT, 2005) über die Qualitätsbeschreibung von räumlichen Daten in Bezug auf verschiedene Anwendungsfälle und betont, dass man eine Qualitätsbeschreibung nutzen sollte bevor man räumliche Daten einsetzt. Er beschreibt dabei ebenfalls die Qualitätselemente und nennt Methoden zur Kontrolle dieser. Des Weiteren stellt er die Frage, ob Nutzer die Qualität von räumlichen Daten ignorieren (VAN OORT, 2005).

2.5 Vergleich von Daten unterschiedlicher Herkunft

Das Ziel dieser Arbeit ist die Erstellung der PLZ-Gebiete aus *OSM*-Daten und der anschließende Vergleich mit dem PLZ-Datensatz von *TeleAtlas*. Dabei handelt es sich um Daten unterschiedlicher Herkunft. Die Qualitätsmerkmale sind in erster Linie auf den Vergleich von aufgenommenen Daten zur Wirklichkeit gedacht, aber auch für den Vergleich zweier Datensätze können sie angewandt werden. Durch diese Merkmale lassen sich sehr gut die Unterschiede oder auch Gemeinsamkeiten der Datensätze feststellen und mittels verschiedenen Qualitätsmaßen bewerten. Im Folgenden soll die Qualitätssicherung innerhalb *OSM* und die bisherigen Vergleiche von *OSM*-Daten mit Daten kommerzieller Anbieter analysiert werden. Dabei sollen vor allem die Vergleichsmethoden und die Ergebnisse dieser Arbeiten diskutiert werden.

2.5.1 Qualität von OSM-Daten

Als erstes seien kurz die Maßnahmen zur Qualitätskontrolle innerhalb *OSM* genannt. Wie bereits erwähnt, gibt es keine systematische Kontrolle der Datenqualität. Die Daten werden eher durch zufällig erkannte Fehler der Community korrigiert.

In *OSM* gibt es zur Vorbeugung von falschen Einträgen und zur Wahrung eines gewissen Grades an Richtigkeit die sogenannten *Map Features*³⁵. Das ist eine Liste mit Eigenschaften, die den verschiedenen Elementen in der Karte zugewiesen werden können. Man spricht dabei auch von Basiseigenschaften. Sie sollen dazu dienen, dass die Eigenschaften der Daten einheitlich benannt werden. Neben diesen *Map Features* gibt es die *Proposed Features*³⁶. Sie dienen zur Erweiterung der Basiseigenschaften. So kann jeder Nutzer Vorschläge zu noch nicht vorhandenen *Map Features* machen. Diese Vorschläge werden dann durch die Community diskutiert und bei Zustimmung übernommen. Man kann also auch von einer Art Kontrolle sprechen.

Neben den genannten „Qualitätskontrollen“ gibt es noch einzelne automatische Konsistenzprüfungen innerhalb *OSM*. Sie dienen dazu, die logische Konsistenz der Geodaten zu untersuchen. Diese meistens als Programme implementierten Kontrollen prüfen den Datenbestand auf Fehler in der Struktur. Innerhalb *OSM* wird dies als Qualitätssicherung³⁷ (Quality Assurance) bezeichnet. Dazu gibt es eine WIKI-Seite, die Links und Werkzeuge zum Thema Qualitätssicherung sammelt die zur Verbesserung, Sicherung und Kontrolle der Qualität der Kartendaten beitragen sollen (OSM-WIKI, 2011b). Diese Werkzeuge werden in verschiedene Kategorien unterteilt. Es gibt Hinweise zur Meldung in der Karte gefundener Fehler, die man über *OpenStreetBugs*³⁸ in die Karte eintragen kann, damit die Community diese Fehler beheben kann. Neben *OpenStreetBugs* gibt es *MapDust*³⁹. Diese Seite dient zur Eintragung von Fehlern, auch Bugs genannt, in der Karte mit einer kurzen Beschreibung und der Angabe, ob dieser Fehler noch offen oder schon behoben ist. Des Weiteren wurde eine Wichtigkeitsangabe gemacht damit man den Fehlern unterschiedliche Prioritäten geben kann. Weitere Programme zum Testen der Qualität sind *Waycheck*⁴⁰ zur Kontrolle von Wegen und *Relation Check* zur Kontrolle von Relationen.

³⁵ http://wiki.openstreetmap.org/wiki/DE:Map_Features

³⁶ http://wiki.openstreetmap.org/wiki/DE:Proposed_features

³⁷ http://wiki.openstreetmap.org/wiki/DE:Quality_Assurance

³⁸ <http://wiki.openstreetmap.org/wiki/DE:OpenStreetBugs>

³⁹ <http://www.mapdust.com/>

⁴⁰ <http://wiki.openstreetmap.org/wiki/DE:WayCheck>

Zusätzlich gibt es noch Karten zur Fehlersuche, die dem Finden von Fehlern durch graphische Darstellung der fehlerhaften Daten dienen. Die wichtigsten sind *OSM Inspector*⁴¹, *Keep Right*⁴², *Maplint*⁴³ (nicht mehr aktualisiert), *Osmose*⁴⁴ und *Duplicate Nodes*⁴⁵. All diese Karten zeigen einem die Fehler direkt auf der *OSM*-Karte. Die Fehler lassen sich somit leichter finden und beheben.

Mit Hilfe dieser Tools kann ein gewisser Grad an Qualität gesichert werden. Ob diese Qualität ausreicht um den Daten der kommerziellen Anbieter Konkurrenz zu machen soll im nächsten Kapitel diskutiert werden.

2.5.2 Vergleich von OSM-Daten mit Daten kommerzieller Anbieter

Kurze Zeit nach der Entstehung des Projekts *OpenStreetMap* gab es auch erste Kritiken zur Qualität und Vollständigkeit der Daten innerhalb *OSM*. So gibt es zahlreiche Arbeiten zum Vergleich von *OSM*-Daten mit Daten kommerzieller Anbieter, wie *TeleAtlas*, *Navteq*⁴⁶ oder auch öffentlicher Anbieter, wie *Ordnance Survey* oder den Daten aus *ATKIS*⁴⁷.

Eine der ersten Arbeiten hierzu wurde von HAKLAY (2008) verfasst. Er verglich die *OSM*-Daten in England mit den amtlichen Datensätzen *OS Meridian 2* (ORDNANCE SURVEY, 2009). Ein Jahr später verglich ATHER (2009) den *OSM*-Datensatz mit dem *Ordnance Survey MasterMap*-Datensatz (NEIS ET AL., 2010). Neben diesen englischsprachigen Veröffentlichungen gab es auch zahlreiche deutschsprachige Arbeiten. So befasste sich bereits ZIELSTRA 2009 mit dem Thema *Datenqualität und Anwendbarkeit von Volunteered Geographic Information* (ZIELSTRA, 2009) und verglich in seiner Diplomarbeit proprietäre mit frei verfügbaren Geodaten. Weitere wichtige Arbeiten zum Thema gab es von STRUNCK (2010), LUDWIG (2010) und ganz aktuell STROH (2010). Als Hauptinitiator dieser Werke sei Zipf⁴⁸ genannt, welcher Autor, Co-Autor und Betreuer diverser Arbeiten zum Thema Qualität von *OSM*-Daten bzw. Qualität von frei verfügbaren Geodaten allgemein und den Vergleich dieser Daten ist.

Das Hauptthema der genannten Arbeiten war der Vergleich des Straßennetzes von *OpenStreetMap* mit den Daten eines kommerziellen Anbieters. Erst die späteren Arbeiten verglichen

⁴¹ <http://tools.geofabrik.de/osmi/>

⁴² http://wiki.openstreetmap.org/wiki/DE:Keep_Right

⁴³ <http://wiki.openstreetmap.org/wiki/DE:Maplint>

⁴⁴ <http://osmose.openstreetmap.fr/>

⁴⁵ http://matt.dev.openstreetmap.org/dupe_nodes/

⁴⁶ <http://www.navteq.com/>

⁴⁷ <http://www.atkis.de/>

⁴⁸ Zipf, Alexander, Professor der Geoinformatik am Geographischen Institut der Ruprecht-Karls-Universität Heidelberg

auch die POIs. So beschäftigte sich STRUNCK (2010) erstmals mit diesem Thema. STROH (2010) beschäftigt sich in ihrer Arbeit verstärkt mit der Rolle des Mappers (NEIS ET AL., 2010).

Neben HAKLAY (2008) und ATHER (2009) gibt es auch eine deutschsprachige Arbeit von SCHOOF (2010) in der ein amtlicher *ATKIS-Basis-DLM*-Datensatz mit den *OSM*-Daten verglichen wurde. Schoof verglich dabei auch verschiedene Flächenobjekte (SCHOOF, 2010).

Zusammenfassend lässt sich sagen, dass alle Arbeiten annähernd gleiche Qualitätsmerkmale und Methoden für den Vergleich verwendet haben. Es wurde die logische Konsistenz, die Vollständigkeit, die Positionsgenauigkeit, die zeitliche Genauigkeit und die thematische Genauigkeit geprüft. So wurde zum Beispiel der Unterschied in der Streckenlänge betrachtet und ein Puffer um die kommerziellen Daten gelegt (Puffer-Vergleichsmethode von GOODCHILD & HUNTER, 1997), um zu sehen wie viel Prozent der *OSM*-Daten in diesem Puffer liegen, sich mit den kommerziellen Daten überlappen. Teilweise wurden diese Ergebnisse dann mittels eines Grids dargestellt, um zu veranschaulichen welcher Teil besonders gut und welcher weniger gut in *OSM* erfasst ist. Besonders bekannt sind die Vergleichskarten, welche die Differenz des Wegenetzes zwischen *OSM* und dem kommerziellen Anbieter pro km² in Prozent vergleichen (ZIELSTRA & ZIPF, 2010). Des Weiteren wurde erläutert, wie die Daten gesammelt und verarbeitet wurden. Neben der Qualitätskontrolle wurde meist eine Erklärung gegeben wie man diese Daten am besten nutzen kann und ob sie für diverse Anwendung sinnvoll sind.

All diese Arbeiten zum Vergleich von Straßendaten kommen auf ähnliche Ergebnisse. So stellte man fest, dass die *OSM*-Datenqualität in weniger dicht besiedelten Gebieten niedriger ist als in dicht besiedelten Gebieten (NEIS ET AL., 2010). ZIELSTRA (2009) stellte fest, dass es starke Differenzen bei der relativen Positionsgenauigkeit wie auch bei der relativen Vollständigkeit zwischen Daten von Groß- und Mittelstädten gibt. Weiterhin wurde bei allen Untersuchungen deutlich, dass die Datenmenge in *OSM* sehr schnell wächst und die Länge des Gesamtstraßennetzes von *OSM* mittlerweile gleich der eines kommerziellen Anbieters ist. In Ballungsgebieten wird diese Länge sogar schon durch *OSM* übertroffen. Allerdings sieht man auch eine klare Abnahme der *OSM*-Datenmenge von den Ballungsgebieten und Stadtzentren hin zum Umland, welche sich mit der abnehmenden Bevölkerungsdichte und der somit geringeren Anzahl an Mappern erklären lässt. (ZIELSTRA, 2009).

Der Vergleich der POIs hatte als Ergebnis, dass die Gesamtzahl an POIs sehr stark differiert. So ergab sich bei Betrachtung aller POIs in der Quadratkilometerrasterung, dass die Anzahl der POIs von *TeleAtlas* nur auf 39.960 km² höher ist, während die von *OSM* auf 55.239 km² höher ist. Betrachtet man allerdings nur die zwischen beiden Datensätzen vergleichbaren POIs, so besitzt *TeleAtlas* mehr POIs auf 40.606 km² und *OSM* mehr auf 19.311 km² (NEIS ET AL., 2010). Auf der restlichen Fläche war die Anzahl der POIs jeweils gleich. Hier lässt sich feststellen, dass es noch große Unterschiede gibt. Trotzdem nimmt die Menge der POIs, wie auch die der Straßendaten, in *OSM* rapide zu und wird sicher in kurzer Zeit den Stand von *TeleAtlas* erreichen, wenn nicht sogar überschreiten.

Da es sich bei diesen Untersuchungen meist um Vergleiche zwischen Linien- bzw. Punktoobjekten handelt, kann man die verwendeten Methoden nur teilweise auf das Problem des Vergleiches von Polygondatensätzen unterschiedlicher Herkunft anwenden. Aus diesem Grund werden diese Methoden hier nicht weiter erläutert, sondern im folgenden Kapitel einige Methoden zum Vergleich von Polygondatensätzen sowie die später angewandten Qualitätselemente und Qualitätsmaße genannt.

2.5.3 Vergleichsmethoden für Polygondatensätze

Leider gibt es zum heutigen Zeitpunkt kaum Arbeiten zum Vergleich von *OSM*-Polygondatensätzen mit Polygondatensätzen kommerzieller Anbieter. Ein solcher Vergleich könnte zum Beispiel für die Hausobjekte oder auch die Waldflächen sehr interessante Ergebnisse liefern. In den genannten Arbeiten wurde fast ausschließlich das Straßennetz betrachtet. Dies ergibt insofern einen Sinn, da die Straßen die ersten Objekte in *OpenStreetMap* waren, was den Namen des Projektes erklärt. Neben den Straßen wurden aktuell auch die POIs verglichen.

Einzig SCHOOF (2010) verglich das *ATKIS-Basis-DLM* mit den *OpenStreetMap*-Daten. Hierbei wurden Linien-, Punkt- und Flächendaten verglichen. Leider lag auch hier das Hauptaugenmerk auf den Straßenobjekten, sodass die Punkt- und Flächenobjekte nur am Rand betrachtet wurden. Sie verglich dabei das Gebiet einer Großstadt, einer Mittelstadt und eines ländlichen Ortes. Die Betrachtung der Flächenobjekte bezog sich dabei nur auf deren Vollständigkeit. Schoof berechnet für den Vergleich die Flächen in ha und versucht damit zu beschreiben, wie vollständig die gewählten Gebiete abgedeckt werden. Dazu erzeugte sie auch Karten welche die Flächenüberlappung zwischen *ATKIS-Basis-DLM* und *OSM* darstellen. Sie betrachtete dabei die unterschiedlichen Nutzungsklassen von Flächen und die gesamte Fläche.

Im Ergebnis wurde gezeigt, dass *OSM* nur einen Bruchteil der Flächen des *ATKIS-Basis-DLM* beinhaltet (SCHOOOF, 2010).

Ein direkter Vergleich der PLZ-Gebiete mit denen von *TeleAtlas* ist leider auch nicht zu 100% möglich, da die PLZ-Gebiete erst aus Punktdaten erzeugt werden müssen und somit nicht direkt als Polygonobjekte in *OSM* vorhanden sind. Hier könnte man in Zukunft allerdings auf die momentan in der Erstellung bzw. Übernahme befindlichen PLZ-Gebiete zurückgreifen.

Durch die Erstellung der PLZ-Gebiete aus den Punktdaten ergibt sich automatisch ein Unterschied zu den Daten von *TeleAtlas*. Diesen Unterschied gilt es zu untersuchen und zu bewerten. Dafür ist es notwendig allgemeine Methoden zum Vergleich von Polygonen verschiedener Datensätze zu finden und ihre Nutzbarkeit zu sondieren.

2.5.3.1 Vergleichsmethoden

In der Literatur findet man zum Vergleich von Polygondatensätzen unterschiedlicher Herkunft aktuell nur wenige nützliche Ausarbeitungen.

CHRISMAN und YANDELL (1988) versuchten in ihrem Artikel zu beschreiben, wie sich der Lagefehler der Vertices eines Polygons auf dessen Fläche auswirkt. Sie betrachteten dabei nur planare Polygone, welche aus Punkten (Vertices) bestehen, die mit geraden Linien verbunden sind. Die Hauptaussage der beiden Autoren ist, dass der Fehler eines Polygons in den Vertices liegt. Hierfür wurden die wahren Koordinaten mit den gemessenen verglichen und über die Berechnung der Varianz, der Kovarianz und des Erwartungswertes Fehler festgestellt. Für die Berechnung von Fehlern wurde hier auch der *RMSE* betrachtet. CHRISMAN und YANDELL wollten ein statistisches Modell für das oben genannte Problem erstellen (CHRISMAN & YANDELL, 1988).

Einen ähnlichen Ansatz wählten auch PRISLEY, GREGOIRE und SMITH (1989) für den Vergleich aufgenommener Koordinaten von Polygonen und deren realen Positionen. Dabei wurde hauptsächlich der Mittelwert und die Varianz der Polygonflächenschätzung, sowie die Kovarianz der Flächen von benachbarten Polygonen betrachtet. In der Ausarbeitung hat man dabei die Polygonvertices mit dem Erwartungswert der Vertices verglichen und dabei jeweils die Fläche betrachtet. Die Fragestellung war, wie sich der Flächeninhalt ändern würde, wenn sich die Koordinaten der Vertices ändern. Hierbei wurde auch auf den Centroid der jeweiligen Fläche Bezug genommen (PRISLEY ET AL., 1989).

Die Varianz- und Kovarianzgleichung für die Flächenschätzung von Polygonen ohne die Nutzung des Centroides beschreibt VAN OORT (2005) in seiner Dissertation. Er nimmt dabei die

Arbeiten von CHRISMAN und YANDELL (1988), sowie die Arbeit von PRISLEY ET AL. (1989) als Grundlage und betont die Unabhängigkeit seiner Ausführungen von den Koordinaten des Polygoncentroides. VAN OORT schreibt, dass die Ungenauigkeit der Polygonfläche durch die Ungenauigkeit der gemessenen Koordinaten und deren Unvollständigkeit entsteht (VAN OORT, 2005). All diese Arbeiten bzw. Artikel untersuchen somit den gleichen Sachverhalt und versuchen mehr eine Erklärung dafür zu finden, warum die Polygondatensätze unterschiedlich sind.

Ein weiterer Versuch Polygondatensätze zu vergleichen bzw. zu beschreiben wird von MOONEY ET AL. (2010) erläutert. Es wird dabei versucht die Vertices und den Umfang eines Polygons innerhalb *OSM* als Vergleichsmerkmal heranziehen. Dabei werden zwei Möglichkeiten genannt. Erstens wird die mittlere Distanz in Metern zwischen benachbarten Vertices eines Polygons berechnet um zu bestimmen wie flächendeckend die Datenpunkte sind. Zweitens wird der Umfang der Polygone durch die Anzahl der Vertices geteilt um die äquidistante Entfernung zwischen allen benachbarten Vertices zu berechnen (MOONEY ET AL., 2010). Es werden also *OSM*-Polygone analysiert, was für die weitere Arbeit ein interessanter Punkt ist. Das Ergebnis ist, dass manche Polygone in *OSM* zu wenig, manche zu viele Vertices besitzen und somit beide ungenau sind (MOONEY ET AL., 2010).

Zusätzlich sei hier nochmals SCHOOFF (2010) genannt, die in ihrer Arbeit Polygondatensätze unterschiedlicher Herkunft miteinander verglichen hat, aber, wie bereits erwähnt, nur die Vollständigkeit unter Nutzung der Fläche betrachtete und keine weiteren Vergleichskriterien für die Untersuchung auswählte (SCHOOFF, 2010).

2.5.3.2 Vergleichskriterien und Qualitätsmaße

Die verschiedenen Vergleichskriterien zum Vergleich der beiden Polygondatensätze zu den PLZ-Gebieten Deutschlands sollen jeweils für die PLZ1-, PLZ2-, PLZ3- und PLZ5-Gebiete erfolgen, um eine Wertung abzugeben, in welchem Maßstab die Daten angewendet werden können. Die verschiedenen Kriterien sind dabei in Tabelle 2.6 nochmal tabellarisch zusammengefasst. Die Zuordnung der *TeleAtlas*-Daten zu den *OSM*-Daten erfolgte über das Attribut Postleitzahl.

Als erstes Qualitätselement sei die Vollständigkeit genannt. Hier soll kontrolliert werden, ob der *OSM*-Datensatz einen Datenüberschuss (*Error of commision*) oder einen Datenmangel (*Error of omission*) aufweist. Dieser Vergleich basiert auf der Anzahl der jeweiligen PLZ-

Gebiete. Dabei wird immer die Gesamtanzahl an Elementen betrachtet. Hier gilt es zu beachten, dass evtl. vorhandene Enklaven oder Exklaven eines PLZ-Gebietes nicht einzeln gezählt werden. Als Qualitätsmaß kann hier auch der Prozentanteil der zugeordneten *OSM*-Objekte gemessen an der Gesamtanzahl der *TeleAtlas*-Objekte gewählt werden. Dabei wird berechnet, wie viel Prozent der in *TeleAtlas* vorhandenen PLZ-Gebiete auch in dem aus *OSM* erzeugten Datensatz vorhanden sind.

Als zweites Vergleichsmerkmal soll die Fläche der PLZ-Gebiete herangezogen werden. Es sollen die Gesamtfläche der PLZ-Gebiete und die jeweiligen Einzelflächen betrachtet werden. Als Maß soll dabei der Unterschied in km^2 angegeben werden. Des Weiteren kann man hier den Mittelwert und die mittlere Abweichung vom Mittelwert der Differenz der Flächen mittels *MS EXCEL*, *MITTELWERT* und *MITTELABW*, berechnen. Die Flächen wurden mit einer Standardfunktion innerhalb *ArcGIS*⁴⁹, *Calculate Area*, berechnet.

Die ersten beiden Vergleichsmerkmale sollen sich, bis auf die Einzelflächen, auf die gesamten PLZ-Gebiete beziehen, während die nachfolgenden Merkmale nur auf die PLZ-Gebiete angewendet werden, die auch innerhalb des *TeleAtlas*-Datensatzes vorhanden sind. Es soll also die Anzahl der PLZ-Gebiete ausgeglichen werden. Das bedeutet, dass der Datenmangel bzw. -überschuss vor den weiteren Vergleichen bereinigt werden muss. Es wird eine 1:1-Beziehung zwischen *TeleAtlas* und *OSM* aufgebaut.

Ein drittes Vergleichsmerkmal sind die Centroide der PLZ-Gebiete. Es handelt sich dabei um den jeweiligen Flächenmittelpunkt. Hier soll die Positionsgenauigkeit betrachtet werden. Dabei wird der Abstand der Centroide der *OSM*-Daten zu denen der *TeleAtlas*-Daten berechnet und dessen Mittelwert und mittlere Abweichung vom Mittelwert angegeben. Die Koordinaten der *TeleAtlas*-Daten sollen dabei als wahre Werte angesehen werden. Es sollen dabei die Koordinaten der *OSM*-Daten mit den Daten einer externen Quelle verglichen werden. Dementsprechend soll die absolute Positionsgenauigkeit berechnet werden. Die Centroide der Flächen wurden mittels *ArcGIS*, *Feature to Point*, erstellt. Die Berechnung der Abstände in Metern erfolgt dabei mittels einer Orthodromen (WIKIPEDIA, 2011c), um die genauen Werte zu erhalten. Die folgenden Formeln (Formel 1 bis Formel 7) zeigen die genaue Berechnung der Abstände, der das WGS84-Ellipsoid zu Grunde liegt. Dabei werden folgende Werte benötigt: b_1 als geografische Breite von Centroid 1, l_1 als geografische Länge von Centroid 1,

⁴⁹ <http://www.esri-germany.de/products/arcgis/index.html>

b_2 als geografische Breite von Centroid 2, l_2 als geografische Länge von Centroid 2 im Gradmaß (PC-HEAVEN, 2011). Diese Werte wurden mittels *ArcGIS* unter Berechnung der X- bzw. Y-Koordinate des jeweiligen Centroides ermittelt.

$$f = \frac{1}{298,257223563}$$

Formel 1 Abplattung der Erde

$$a = \frac{6378137}{1000}$$

Formel 2 Äquatorradius der Erde in Kilometern

$$F = \frac{b_1 + b_2}{2}; G = \frac{b_1 - b_2}{2}; l = \frac{l_1 - l_2}{2}$$

Formel 3 Parameter F, G und l

$$F = \frac{\pi}{180} * F; G = \frac{\pi}{180} * G; l = \frac{\pi}{180} * l$$

Formel 4 Umrechnung in Bogenmaß

$$S = (\sin G)^2 * (\cos l)^2 + (\cos F)^2 * (\sin l)^2$$

$$C = (\cos G)^2 * (\cos l)^2 + (\sin F)^2 * (\sin l)^2$$

$$\omega = \arctan \sqrt{\frac{S}{C}}$$

$$D = 2 * \omega * a$$

Formel 5 Berechnung Abstand D

$$R = \frac{\sqrt{S * C}}{\omega}$$

$$H_1 = \frac{3 * R - 1}{2 * C}$$

$$H_2 = \frac{3 * R + 1}{2 * S}$$

Formel 6 Korrektur Abstand D mit H_1 und H_2

$$s = D * (1 + f * H_1 * (\sin F)^2 * (\cos G)^2 - f * H_2 * (\cos F)^2 * (\sin G)^2)$$

Formel 7 Abstand s in Kilometern

In Bezug auf die Centroide soll zusätzlich die Puffer-Vergleichsmethode (GOODCHILD & HUNTER, 1997) angewendet werden. Hierbei soll ein Puffer mit unterschiedlicher Größe um die *TeleAtlas*-Centroide erstellt werden. Dabei sollen die Puffer kreisförmig sein und einen

Radius von 1, 2, 5 und 10 km besitzen (Abbildung 2.4). Anschließend soll betrachtet werden, wie viele der *OSM*-Centroide sich innerhalb dieser Entfernung zu den *TeleAtlas*-Centroiden befinden. Dieser Vergleich ermöglicht dabei eine weitere Aussage zum Abstand der Centroide. Für diesen Vergleich werden die Shapefiles benötigt. Diese sind nicht bereinigt, das heißt es wird mit dem erzeugten Datensatz gearbeitet, ohne dass fehlerhafte Centroide berücksichtigt wurden.

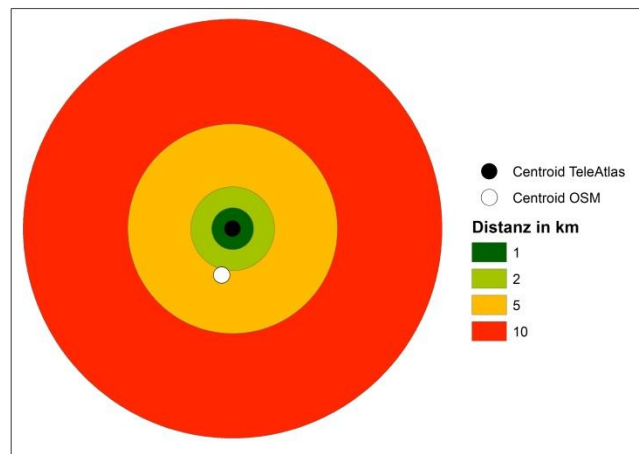


Abb. 2.4 Beispiel Puffervergleichsmethode

Als weitere Vergleichsmethode kann noch die Euklidische Distanz der Centroide zur nächsten Rasterzelle (Zellgröße 1 km) dargestellt werden. Hierfür kann eine Funktion innerhalb *ArcGIS*, *Euclidean Distance*, genutzt werden, welche diese Distanz berechnet und als Raster ausgibt. Somit kann für die zwei Datensätze jeweils ein Raster erstellt werden, welches einen visuellen Vergleich und somit auch eine Aussage über die Punktdichte bzw. die Lage der Punkte zueinander ermöglicht. Diese Funktion soll auf die PLZ3- und PLZ5-Centroiden angewendet werden. Im Ergebnis können die Abstände im *TeleAtlas*-Datensatz mit denen des *OSM*-Datensatzes visuell verglichen werden. Auch hier wird der nicht bereinigte Datensatz verwendet.

Ein zweiter visueller Vergleich soll für die PLZ5-Gebiete erstellt werden. Hier soll mittels *ArcGIS*, *Point Density*, die Punktdichte in einem Raster dargestellt werden. Dabei wird ein Kreis mit einem Radius von 10 km um die jeweiligen Centroiden gelegt und geprüft, ob sich diese Kreise schneiden. Je mehr Kreise übereinander liegen bzw. je mehr sich überlappen, umso mehr Punkte sind in diesem Raum zu finden. Die Dichte ist demzufolge an solchen Stellen höher.

Ein weiteres Vergleichsmerkmal wäre die unterschiedliche Anzahl der Vertices des jeweiligen PLZ-Gebietes, die zur Beschreibung des Grenzverlaufes dienen. Hierfür könnten einfach die Vertices der Polygone mittels *ArcGIS*, `Feature Vertices to Points`, erstellt und jeweils addiert werden. Dieser Vergleich wird hier aufgrund der Datenmenge und der damit zusammenhängende Rechenzeit nicht durchgeführt.

Im Zusammenhang mit der Anzahl der Vertices würde sich auch ein Vergleich der Umfänge, wie von MOONEY ET AL. (2010) beschrieben, anbieten. Dieser Vergleich lohnt sich bei diesen Datensätzen allerdings nicht, da sich die Grenzverläufe der PLZ-Gebiete von *TeleAtlas* von denen aus *OSM* stark unterscheiden, so dass hier von vornherein feststeht, dass dieser Vergleich eindeutig zugunsten *TeleAtlas* ausfällt bzw. keinen richtigen Sinn ergibt, da feststeht, dass der *TeleAtlas*-Datensatz detailliertere Grenzverläufe aufweist und diese somit auch aus mehr Vertices bestehen.

Neben diesen Berechnungen soll versucht werden, die Genauigkeit der Daten in Bezug auf die Bevölkerungsdichte zu betrachten. Dabei soll ein Vergleich zwischen Stadt und Land erfolgen. Im gleichen Zusammenhang soll auch die Genauigkeit der Daten in einer Region betrachtet werden, die als sogenannte *OSM*-Hochburg bekannt ist. Dabei handelt es sich um eine Region, die fast komplett erfasst ist. Als Beispiel soll die Region Hamburg gewählt werden. Diese beiden Vergleiche, bei denen die vorher berechneten und betrachteten Vergleichskriterien zwischen den Gebieten bzw. für das eine Gebiet betrachtet werden, kann man in Bezug zum Raum sehen.

Nach der Berechnung und Betrachtung der unterschiedlichen Qualitätselemente und -maße bzw. Vergleichskriterien soll eingeschätzt werden für welchen Maßstab die Nutzung der aus den *OSM*-Daten generierten PLZ-Gebiete aktuell sinnvoll erscheint. Es erfolgt ein Vergleich über die verschiedenen Kategorien der PLZ-Gebiete und deren Nutzbarkeit für spätere Anwendungen.

Als visuelle Vergleichsmöglichkeit sollen die PLZ-Datensätze übereinander gelegt werden und mittels PDF-Karten visualisiert werden. Hier sollen die Unterschiede oder Gemeinsamkeiten der beiden Datensätze anschaulich dargestellt werden. Mittels dieser PDF-Karten lassen sich auch Bewertungen zum Maßstab und im Bezug zum Raum treffen.

Ein stichprobenartiger Vergleich der fehlerhaften PLZ-Daten in *OSM* soll mittels der PLZ-Suche der Deutschen Post durchgeführt werden. Es soll dabei festgestellt werden, ob die PLZ-Daten wirklich falsch sind oder ob diese evtl. in *TeleAtlas* fehlen.

Tabelle 2.6 dient als Zusammenfassung der gewählten Vergleichskriterien.

| Vergleichskriterium | Unterkriterium | Bestimmung - Qualitätsmaß |
|----------------------|--|---|
| Vollständigkeit | Datenüberschuss/Datenmangel | Anzahl der PLZ-Gebiete |
| Vollständigkeit | Datenüberschuss/Datenmangel | Prozentanteil gegenüber TA |
| Fläche | gesamt/einzeln | Differenz Fläche in km ² |
| Fläche | einzeln | Mittelwert Differenz in km ² |
| Fläche | einzeln | mittlere Abweichung vom Mittelwert in km ² |
| Positionsgenauigkeit | Abstand der Centroide | Formel 1-8 |
| Positionsgenauigkeit | Abstand der Centroide | Mittelwert in m |
| Positionsgenauigkeit | Abstand der Centroide | mittlere Abweichung vom Mittelwert in m |
| Positionsgenauigkeit | Lage der Centroide | Puffer-Vergleichsmethode |
| Positionsgenauigkeit | Euklidische Distanz der PLZ3-/PLZ5-Centroide | ArcGIS-Funktion → Raster |
| Punktdichte | PLZ5 | ArcGIS-Funktion → Raster |
| Raum | Stadt/Land | Betrachtung vorhergehender Kriterien |
| Raum | OSM-Hochburg Hamburg | Betrachtung vorhergehender Kriterien |
| Maßstab | PLZ-Kategorie | Betrachtung vorhergehender Kriterien |
| Visuell | Gebietsüberlagerung | PDF-Karten |
| Stichprobe | fehlerhafte PLZ | PLZ-Suche Deutsche Post |

Tab. 2.6 Zusammenfassung Vergleichskriterien

2.6 Gebietsgenerierung aus Punktdaten

2.6.1 Allgemeines und Literatur

Die automatisierte Gebietsgenerierung aus Punktdaten ist eine Methode um aus einem vorhandenen Punktdatensatz einen Polygondatensatz zu bilden. Es gibt dabei Methoden, welche aus Punkten, die als Vektordaten vorliegen, Polygone erzeugen, die dann ebenfalls als Vektordaten vorliegen. Andere Methoden erzeugen dagegen ein Raster aus den Punktvektordaten. Man spricht bei der Gebietsgenerierung auch von räumlicher Interpretation bzw. Regionalisierung.

Verfahren zur Gebietsgenerierung dienen dabei häufig der Interpolation von Messwerten. So liegen als Grundlage im Raum verteilte Punkte mit bestimmten Messwerten vor. Aus diesen Punkten soll eine Interpolation für den zwischen den Punkten befindlichen Raum berechnet werden. Der Wert zwischen den Messpunkten soll also geschätzt werden um Rückschlüsse

auf das gesamte Messgebiet ziehen zu können. Es wird also aus Punkten eine Fläche bzw. mehrere Flächen erzeugt. Dabei wird davon ausgegangen, dass die Messwerte zweier Messpunkte umso ähnlicher sind, je näher diese Messpunkte zusammen liegen.

Die verschiedenen Verfahren werden dabei in verschiedene Klassen eingeteilt. Es gibt kontinuierliche oder abrupte, exakte oder approximierte, lokale oder globale und deterministische oder stochastische Verfahren (HEBELER, 2004). Approximiert bedeutet, dass die Interpolation auch für die originären Messwerte berechnete Werte liefert. Exakt heißt, dass der Originalwert erhalten bleibt. Bei lokalen Methoden wird nur eine Teilmenge aller Punkte zur Interpolation benutzt, bei globalen hingegen alle vorhandenen Punkte. Deterministisch heißt, dass die Interpolation mit einer festgelegten Funktion durchgeführt wird, wohingegen stochastisch bedeutet, dass die Interpolation mit einer von vielen möglichen Varianten einer Verteilungsfunktion durchgeführt wird (HEBELER, 2004). Unter kontinuierlich ist zu verstehen, dass die Schätzwerte stetig verlaufen, wogegen die Messwerte bei abrupt schlagartig wechseln können.

Die am häufigsten verwendeten Verfahren zur Gebietsgenerierung aus Punktdaten bzw. Interpolation sind dabei Inverse Distance Weighting, Kriging, Spline und das Bilden von Voronoi- oder auch Thiessen-Polygonen.

Das spätere verwendete Verfahren der Voronoi-Polygone besitzt dabei die Eigenschaften exakt, lokal, deterministisch, abrupt und ist besonders für nominale Daten geeignet (HEBELER, 2004), weshalb es für die Erzeugung der PLZ-Gebiete die beste Wahl ist.

Die Verfahren werden in zahlreichen Wissenschaften angewandt. So gibt es im Bereich Biologie die Nutzung von gewichteten Voronoi-Diagrammen zur Analyse des Wachstums von Pflanzen und deren Ausbreitung. Dabei wird berechnet, welchen Platz diese Pflanzen benötigen und wie das Größen-Entfernungs- und Größen-Flächen-Verhältnis ist (GAVRILOVA, 1997).

Neben dem Bereich der Biologie werden auch im Bereich der mobilen Robotik Voronoi-Diagramme genutzt. So wird mit Hilfe von generalisierten Voronoi-Diagrammen ein Pfad für einen mobilen Roboter berechnet. Dabei wird als Ausgangsbasis ein Plan eines Campus benutzt und die Gebäude bzw. diverse Flächen als Hindernisse angesehen, die der Roboter auf einem optimalen Pfad umgehen soll um an das gewünschte Ziel zu kommen (BLAER, 2009).

Eine Auflistung verschiedener Anwendungen von Voronoi-Diagrammen, geordnet nach Themenbereichen, findet man dabei in der *VoroWiki*⁵⁰.

Außerdem gibt es zahlreiche Anwendungen von Verfahren zur automatisierten Gebietsgenerierung bzw. Interpolation in den Bereichen Astronomie, Geologie, Medizin, Routenplanung und Geographie.

Kriging wird dabei speziell in der Kartographie, den Agrarwissenschaften, der Hausverwaltung, der Epidemiologie, der Hydrologie und im Bergbau verwendet (DUBIN, 2008). Voronoi-Diagramme verwendet man dagegen eher für Online Maps, Routenplanung, Supply-Chain-Management, Krisenmanagement, in der Graphentheorie und zur Lösung des Nearest Neighbor Problems (KANG, 2008). Inverse Distance Weighting wird häufig in der Geostatistik zur Interpolation räumlicher Abhängigkeiten von georeferenzierten Daten genutzt.

Die Untersuchung einer fundamentalen geometrischen Datenstruktur führte AURENHAMMER (1991) an Voronoi-Diagrammen durch. Dabei betrachtet er diese aus verschiedenen Gesichtspunkten und nennt algorithmische Anwendungen (AURENHAMMER, 1991). Später beschreiben AURENHAMMER und KLEIN (1996) in ihrer Ausarbeitung die Grundlagen, verschiedene Algorithmen, die Generalisierung, strukturelle Eigenschaften und geometrische Anwendungen von Voronoi-Diagrammen. Auch OKABE ET AL. schreiben ausführlich über die Konzepte und Anwendungen von Voronoi-Diagrammen (OKABE ET AL., 2000).

Eine weitere Einführung zu Voronoi-Diagrammen liefert KANG (2008) in seinem Artikel in der *Encyclopedia of GIS* (SHEKAR & XIONG, 2008). Hier definiert KANG Voronoi-Diagramme, beschreibt den historischen Hintergrund und erläutert die wissenschaftlichen Grundlagen. Des Weiteren nennt er auch einige Hauptanwendungsgebiete von Voronoi-Diagrammen (KANG, 2008).

Neben eher grafischen Anwendungen von Voronoi-Diagrammen werden diese auch für die Abfrageverarbeitung in räumlichen Datenbanken verwendet. Diese Möglichkeit beschreiben SHAHABI und SHARIFZADEH (2008). Sie geben dabei eine Definition und erläutern das Problem und die Nutzung des Voronoi-Verfahrens in diesem Zusammenhang. Dabei wird vor allem gezeigt, wie effizient diese Verfahren ist, um das Nearest Neighbor Problem und das Problem der Optimal Sequenced Route zu lösen. Es handelt sich dabei meist um eine Art Pfadberechnung (SHAHABI & SHARIFZADEH, 2008).

⁵⁰ http://www.voronoi.com/wiki/index.php?title=Voronoi_Applications

Die Methode des Kriging wird von DUBIN (2008) erläutert. Er schreibt dabei über die historische Entstehung und die wissenschaftlichen Grundlagen. DUBIN (2008) erläutert hierbei die verschiedenen Methoden des Kriging; Simple Kriging, Ordinary Kriging und Universal Kriging und nennt diverse Anwendungsgebiete (DUBIN, 2008).

Erläuterungen zum Thema Inverse Distance Weighting, Kriging und zu verschiedenen anderen Interpolations- und Approximationsmethoden liefert WÄLDER (2008). Sie schreibt über mathematische Methoden für Ingenieure und Geowissenschaftler und zeigt in ihrem Buch verschiedene Lösungen zu realen Problemen innerhalb dieser Wissenschaften. Des Weiteren nennt sie praktische Beispiele für mathematische Modelle, so zum Beispiel zur Generalisierung von Höhenprofilen (WÄLDER, 2008).

Auch LU und WONG (2008) erläutern in ihrem Artikel zu Inverse Distance Weighting dieses Verfahren. Dabei versuchen sie diese Verfahren zu verbessern und mit Kriging zu vergleichen (LU & WONG, 2008). Einen weiteren Vergleich von den Verfahren Kriging und Inverse Distance Weighting an einem praktischen Beispiel erbringen BALMER ET AL. (2009).

Erläuterungen zum Thema räumliche Interpolation allgemein erhält man von MITAS und MITASOVA (1999), die verschiedene Methoden vorstellen (MITAS & MITASOVA, 1999). Eine weitere Beschreibung diverser räumlicher Interpolationsmethoden geben LI und HEAP (2008). Sie erläutern dabei kurz eine Vielzahl von nicht-geostatistischen und geostatistischen, sowie kombinierten Verfahren. Diese werden später auf ihre Eigenschaften untersucht und schließlich miteinander verglichen. Des Weiteren werden Faktoren zur Performance der Verfahren genannt, die Methoden klassifiziert, deren Anwendungsgebiete erläutert und ein Überblick über die verschiedenen Softwarepakete gegeben, die solche Verfahren ermöglichen (LI & HEAP, 2008).

2.6.2 Voronoi-Verfahren

In diesem Kapitel soll das bei der Erzeugung der PLZ-Gebiete verwendete Verfahren kurz näher beschrieben werden. Es handelt sich dabei um die Gebietsgenerierung aus Punktdaten mit Hilfe von Voronoi-Diagrammen. Diese Diagramme werden oft auch als Thiessen-Polygone oder auch als Dirichlet-Zerlegung bezeichnet. Die Bezeichnung geht auf *Georgi Feodosjewitsch Woronoi*⁵¹ zurück, der 1905 in der Ukraine lebte. Allerdings wurde die Be-

⁵¹ russischer Mathematiker, ukrainischer Herkunft, lebte vom 28. April 1868 bis 20. November 1908

zeichnung Voronoi bereits 1644 von *René Descartes*⁵² und 1850 von *Johann Peter Gustav Lejeune Dirichlet*⁵³ verwendet (KLEIN, 2005).

Mit Voronoi-Diagramm wird eine Zerlegung des Raumes in Regionen bezeichnet. Das Voronoi-Diagramm zerlegt dabei einen gegebenen Raum bei einer ebenfalls gegebenen Punktmenge in Regionen. Jedem dieser Punkte – genannt Voronoi-Punkt – ist genau eine Region zugeordnet. Die Regionen haben die wesentliche Eigenschaft, dass alle Punkte einer Region ihrem Voronoi-Punkt näher, im Sinne der Euklidischen Distanz, sind als allen anderen Voronoi-Punkten (RASKOB, 2004).

In Abbildung 2.5 sieht man eine Punktmenge und das zugehörige Voronoi-Diagramm. Abbildung 2.6 zeigt einen Ausschnitt aus Abbildung 2.5 zur näheren Erläuterung.

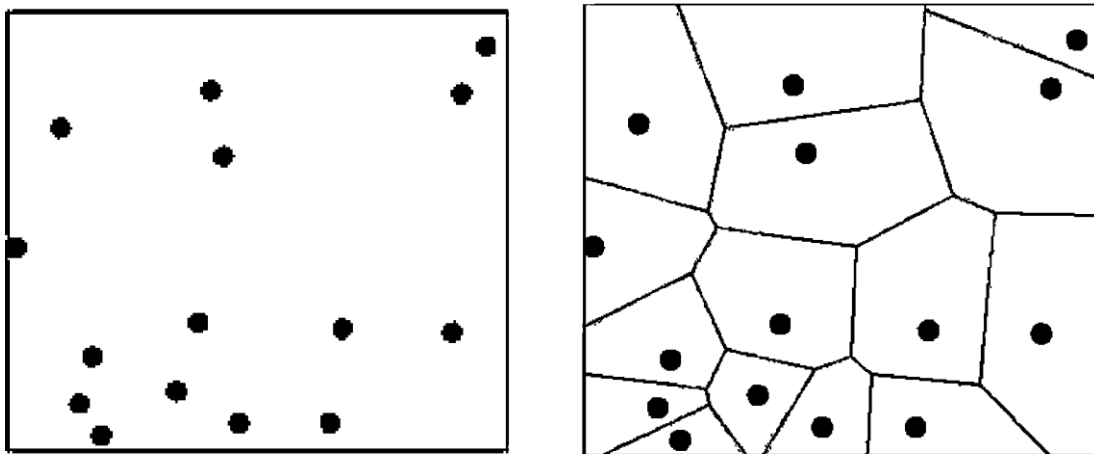


Abb. 2.5 Punktmenge und Voronoi-Diagramm (KANG, 2008)

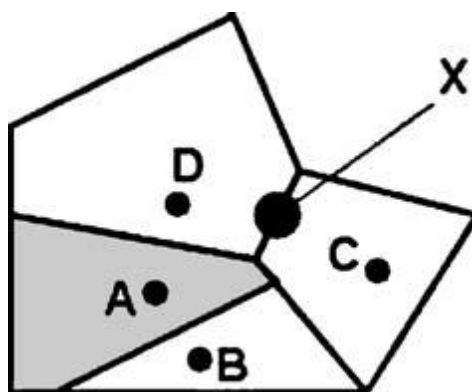


Abb. 2.6 Ausschnitt Voronoi-Diagramm aus Abb. 2.5 (KANG, 2008)

Jeder Punkt innerhalb des grauen Polygons, welches den Punkt A enthält ist näher an A als jeder anderer Punkt B, C oder D. Die Punkte auf der Außenlinie eines Polygons sind dabei

⁵² französischer Mathematiker, Philosoph und Naturwissenschaftler, lebte vom 31. März 1596 bis 11. Februar 1650

⁵³ deutscher Mathematiker, lebte vom 13. Februar 1805 bis 5. Mai 1859

Mittelpunkte zwischen zwei Punkten. Zum Beispiel besitzt Punkt X die gleiche Distanz zu C wie zu D (KANG, 2008).

Umgangssprachlich formuliert wird mittels Voronoi-Verfahren aus einer Menge an Punkten eine Menge an Flächen erstellt. Dabei ist die Anzahl der Flächen gleich der Anzahl der Punkte. Jeder Punkt liegt in einer Fläche (Polygon). Die Polygonaußenlinie ist dabei die Grenze zwischen zwei Punkten und liegt genau im Mittelpunkt zweier benachbarter Punkte.

Zur Berechnung dieser Voronoi-Diagramme gab es zahlreiche verschiedene Verfahren und Algorithmen. Darunter Durchschnitt von Halbebenen, Inkrementelle Konstruktion, Divide & Conquer und Sweep (AURENHAMMER & KLEIN, 1996). Im Folgenden soll dabei ein Sweep-Algorithmus näher erläutert werden, da dieser innerhalb *GRASS GIS* implementiert ist.

Der in *GRASS GIS* zur Erzeugung der PLZ-Gebiete genutzte Algorithmus heißt Fortune's Algorithmus und wurde 1986 von *Steven Fortune*⁵⁴ entwickelt. Bei diesem Algorithmus handelt es sich um einen sogenannten „plane-sweep“-Algorithmus. Die Vorteile sind, dass dieser Algorithmus einfacher als der inkrementelle ist und im schlechtesten Fall eine Komplexität von $O(n \log n)$ aufweist (FORTUNE, 1986).

Als Sweep wird ein Verfahren der Informatik verstanden, dass zum Design von Algorithmen verwendet wird. Solch ein Algorithmus besitzt eine Sweep-Line. Mit Hilfe dieser bewegt man sich solange durch den Raum bis alle Objekte des Problems berührt und verarbeitet wurden. Dazu wird eine Datenstruktur verwendet die die berührten Objekte speichert. Meistens werden dadurch Probleme der Algorithmischen Geometrie gelöst. Allgemein wird dabei ein n-dimensionales statisches Problem in ein (n-1)-dimensionales dynamisches Problem umgewandelt (KLEIN, 2005).

Der Fortune's Algorithmus besitzt dabei eine sogenannte Sweep-Line und eine Beach-Line, die sich beide über die Ebene bewegen, wenn der Algorithmus abläuft. Die Sweep-Line ist eine gerade Linie die vertikal ist und von links nach rechts über die Ebene läuft. Zu jeder Zeit, während der Algorithmus läuft, werden die Eingabepunkte links der Sweep-Line in das Voronoi-Diagramm integriert, während die Punkte rechts der Sweep-Line noch nicht betrachtet werden. Die Beach-Line ist keine Linie sondern eine komplex Kurve links der Sweep-Line. Sie besteht aus Parabeln und trennt den Teil der Ebene in der das Voronoi-Diagramm bekannt ist, ungeachtet der Punkte rechts der Sweep-Line, vom Rest der Ebene. Für jeden Punkt links der Sweep-Line kann eine Parabel aus gleichabständigen Punkten zu diesem Punkt und zur

⁵⁴ technischer Manager der Algorithms Research Group of Bell Laboratories, Muray Hill, USA

Sweep-Line bestimmt werden. Die Beach-Line ist dabei die Grenze der Vereinigung dieser Parabeln. Wenn sich die Sweep-Line weiter bewegt erzeugen die Vertices der Beach-Line, an der Stelle wo sich die Parabeln schneiden, die Kanten des Voronoi-Diagramms (FORTUNE, 1986). Abbildung 2.7 zeigt dabei den Ablauf des Algorithmus an einem Beispiel. Die rote Linie ist die Sweep-Line, die schwarzen Linien bilden die Beach-Line und die blauen Linien sind die erzeugten Kanten des Voronoi-Diagramms.

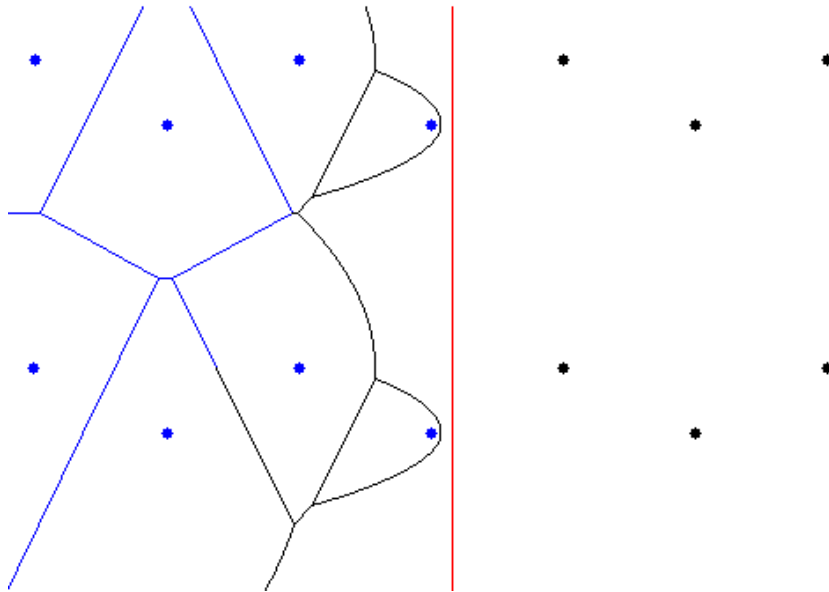


Abb. 2.7 Ablauf Fortune's Algorithmus (ODGAARD & NIELSEN, 2000)

Die Funktion zur Berechnung von Voronoi-Polygonen in *GRASS GIS* `v.voronoi` verwendet diesen Algorithmus. Diese Funktion wurde genutzt um die Gebiete aus den Punktdaten zu erzeugen.

2.7 Open Source GIS Software

Open Source Software sind Programme, die durch eine Lizenz geschützt werden, deren Bedingungen vorschreiben, dass man diese Programme ohne Einschränkungen nutzen darf (OSGEO, 2011). Wie der Name Open Source schon andeutet ist der Quelltext bei solcher Software offen und für jeden zugänglich und modifizierbar. So kann zum Beispiel auch jeder einfach neue Plug-Ins oder Module für solch eine Software schreiben und zur Verfügung stellen. In fast allen Bereichen findet man heutzutage Open Source Software. Seit langem gibt es auch Open Source Software im Bereich der Geoinformationsverarbeitung. Hier findet man häufig Programme zur Erfassung, Bearbeitung, Organisation, Analyse und Präsentation von räumlichen Daten, sogenannte Open Source GIS Software. Im Internet gibt es hierzu Webseiten, welche eine Liste der aktuell verfügbaren Open Source oder auch Free GIS Software zeigen.

*Open Source GIS*⁵⁵ versucht einen kompletten Index dieser Software zur Bearbeitung von räumlichen Daten bereitzustellen. *FreeGIS*⁵⁶ führt in seiner Datenbank ebenfalls eine Liste unterschiedlicher Software zur Bearbeitung solcher Daten.

Neben diesen Internetseiten gibt es mittlerweile einige Fachliteratur zum Thema Open Source GIS. Im Folgenden soll ein Überblick über die aktuell vorhandene Fachliteratur zu diesem Thema gegeben werden.

Als erstes Werk sei *PostGIS in Action* (OBE & HSU, 2011) genannt, welches sich mit *PostGIS*⁵⁷ beschäftigt. *PostGIS* ist eine Erweiterung zu *PostgreSQL*⁵⁸, die es ermöglicht geographische Objekte in einer objektrelationalen Datenbank zu speichern und zu bearbeiten, so dass man *PostgreSQL* als räumliche Datenbank nutzen kann. *PostGIS* wurde unter anderem zur Bearbeitung der *OSM*-Daten genutzt, aus denen die PLZ-Gebiete innerhalb dieser Arbeit erzeugt wurden.

Das Buch bringt eine Einführung in die Grundlagen der räumlichen Daten und der Datenmodellierung und hinterlegt dies mit vielen Beispielen. Ebenso führt es in *PostgreSQL* und räumliche Datenbanken ein, erklärt Geometrietypen, geometrische Funktionen und geometrische Abfragen, erklärt die Nutzung von *PostGIS* an Beispielen der realen Welt und zeigt wie man *PostGIS* in Internet- und Desktopanwendungen integrieren kann. Dabei soll das Buch kein Ersatz zur offiziellen *PostGIS*-Dokumentation⁵⁹ sein. *PostGIS in Action* ist in drei Teilen aufgebaut. Teil 1 diskutiert die Grundlagen, Teil 2 erläutert die Arbeit mit *PostGIS* und erklärt wie man räumliche Probleme löst, Teil 3 beschreibt wie man *PostGIS* mit anderen Tools nutzen kann. Neben dem Textteil gibt es zahlreiche Codebeispiele und Geodaten zum Download, damit man die Beispiele am eigenen Rechner selbst durchführen kann (OBE & HSU, 2011).

Ein zweites großes Werk, welches sich auf ein einzelnes Open Source GIS Programm bezieht, ist *Open Source GIS: A GRASS GIS Approach* (NETELER & MITASOVA, 2007). Dieses Buch bezieht sich auf die Open Source Software *GRASS GIS*, die zur Erstellung der PLZ-Gebiete innerhalb dieser Arbeit genutzt wurde. Es beschreibt dabei *GRASS GIS*, ein GIS zur Bearbeitung von Raster- und Vektordaten, kombiniert mit Bildverarbeitung und Datenvisualisierung, welches als das weltweit führende Free Software GIS bezeichnet wird und bereits seit

⁵⁵ <http://opensourcegis.org/>

⁵⁶ <http://www.freegis.org/>

⁵⁷ <http://postgis.refractive.net/>

⁵⁸ <http://www.postgresql.org/>

⁵⁹ <http://postgis.refractive.net/documentation/manual-1.5/>

den 1980ern entwickelt wird. *GRASS GIS* ist ein offizielles Projekt von *OSGEO* und besteht aus mehr als 350 Modulen für das Management, die Verarbeitung, die Analyse und die Visualisierung von georeferenzierten Daten. Wie auch *PostGIS in Action* (OBE & HSU, 2011) beschreibt dieses Buch nicht alle Funktionen des Programms, da es auch bei *GRASS GIS* eine ausführliche Dokumentation⁶⁰ gibt. Am Anfang erläutert das Buch die Begriffe Open Software und GIS, beschreibt die Beispieldaten, welche für *GRASS GIS* existieren, und wie man dieses Buch nutzen soll. Dabei wird beschrieben, dass sich dieses Buch auf die Basisprinzipien und die Basisfunktionen von *GRASS GIS* konzentriert. Weiter beschreibt das Buch die Konzepte und Grundlagen von GIS, sprich Datenmodelle, Kartenprojektionen und so weiter. Es bringt einem die ersten Schritte in *GRASS GIS* näher und erklärt die Datenmodelle und den Datenaustausch in *GRASS GIS*. Das Buch beschreibt wie man Raster- und Vektordaten bearbeitet und diese Daten ausgibt und visualisiert. Weitere Kapitel erläutern die Bildverarbeitung, bringen Hinweise zur Programmierung in *GRASS GIS* und erklären wie man *GRASS GIS* mit anderen Open Source Tools nutzt. So kann man *GRASS GIS* zum Beispiel in *Quantum GIS*⁶¹ integrieren (NETELER & MITASOVA, 2007).

Zur Kombination von *PostgreSQL* bzw. *PostGIS* mit *GRASS GIS* gibt es ein Werk mit dem Thema *Einrichten von PostgreSQL / PostGIS als Datenbankserver für GRASS: Einrichten einer SQL-Datenbank als Datenbankserver für ein Geoinformationssystem unter Verwendung freier Software* (DEE, 2008). Darin wird die Einrichtung einer solchen Datenbank als Server für ein GIS beschrieben. Dee beschreibt dabei, wie man *GRASS GIS* mit einer *PostgreSQL*-Datenbank, mit *PostGIS*, verbindet. Dem Werk liegt die Diplomarbeit von Dee zum Thema zugrunde.

Neben diesen Werken gibt es noch weitere Literatur zum Thema Open Source GIS und deren Nutzung zur Bearbeitung von räumlichen Daten bzw. zur Erstellung von Karten.

Open Source Approaches in Spatial Data Handling (HALL & LEAHY, 2008) diskutiert dabei über *FOSS4G*⁶², der Konferenz für Freie und Open Source Software für Geodaten, im Allgemeinen und nennt eine paar bekannte *FOSS4G*-Projekte. Der zweite Teil des Buches beschreibt verschiedene universitäre *FOSS4G*-Projekte. Das Buch will Lesern die Arbeit und die Projekte von *FOSS4G* näherbringen und ein paar Mainstreamprogramme der webbasierten und Desktop-*FOSS4G*-Tools vorstellen. Dabei konnten nicht alle bekannten Programme dis-

⁶⁰ http://grass.fbk.eu/grass64/manuals/html64_user/

⁶¹ <http://www.qgis.org/>

⁶² <http://foss4g.org>

kutiert werden. So wird in diesem Buch nichts über Programme wie *Quantum GIS*, *MapBender*⁶³, *OpenJump*⁶⁴ und *uDig*⁶⁵ geschrieben. Allerdings findet sich in diesem Werk ein Kapitel über *UMN MapServer*⁶⁶, *GRASS GIS*, *GeoVista Studio*⁶⁷, *MapGuide*⁶⁸ und andere Tools, sowie ein interessanter Vergleich von Open Source Datenbanken und ihren räumlichen Erweiterungen (HALL, 2008).

Bei den meisten Beiträgen handelt es sich nur um Zusammenfassungen. So ist das Kapitel zu *GRASS GIS* eine stark gekürzte Einführung in das Programm, eher sogar eine Art Produktvorstellung ohne tief in Details zu gehen. Das Buch bringt zwar einen Überblick zu Open Source GIS Projekten und Software, deckt allerdings nur einen kleinen Teil der vorhandenen Programme und Projekte ab.

Bücher zur Anwendung von Open Source GIS Software zur Erstellung von Karten oder Webanwendungen lassen sich ebenfalls auf dem Markt finden. SHERMAN (2008) schreibt dabei in seinem Werk *Desktop GIS: Mapping the Planet with Open Source Tools* (SHERMAN, 2008) über die verschiedene Open Source GIS Software und gibt Hinweise wie man das richtige Programm für seine Aufgaben findet. Er gibt dabei eine kurze Einführung in viele erhältliche Softwarepakete und zeigt diverse Beispiele. Das Buch zeigt einen Überblick über die bekanntesten Open Source GIS Desktopprogramme wie *GRASS GIS*, *Quantum GIS*, *uDIG* und auch räumliche Datenbanken. Des Weiteren wird die Erstellung von Skripten in *Python*⁶⁹, *UNIX Shell* und anderen Sprachen erklärt, damit man seine Daten visualisieren, digitalisieren und analysieren kann (SHERMAN, 2008). Es ist ein Überblick über das gesamte Gebiet des Desktop Mapping im Zusammenhang mit Open Source GIS Software. *Sherman* blickt dabei auf jahrelange Erfahrung zurück. Er selbst ist Gründer des *Quantum GIS* Projektes.

Ein weiteres Buch welches die Nutzung von Open Source GIS Software zu Erstellung von Karten in Bezug auf das Internet zeigt ist *Web Mapping Illustrated: Using Open Source GIS Toolkits* (MITCHELL, 2005). Die deutsche Version *Web-Mapping mit Open Source-GIS-Tools* wurde von EMDE und CHRISTL (2008) überarbeitet. Dieses Buch zeigt dabei anschaulich, wie man mit frei erhältlichen Programmen interaktive Karten erstellen kann. Es

⁶³ <http://www.mapbender.org>

⁶⁴ <http://www.openjump.org/>

⁶⁵ <http://udig.refractions.net/>

⁶⁶ <http://mapserver.org/>

⁶⁷ <http://www.geovistastudio.psu.edu/jsp/index.jsp>

⁶⁸ <http://mapguide.osgeo.org/>

⁶⁹ <http://www.python.org/>

wird dabei auf die Programme und Werkzeuge *UMN MapServer*, *Quantum GIS*, *GDAL*⁷⁰, *OGR*⁷¹, *MapBender* und auch *PostGIS* eingegangen und diese anschaulich erklärt. Des Weiteren wird erläutert, wo man Geodaten finden kann, wie man diese Daten aufbereitet und sie schließlich auf einer Webseite als Karte präsentiert. So wird Schritt für Schritt erklärt, wie man solch eine interaktive Internetkarte erstellt. Am Ende findet man eine Zusammenfassung, die angibt, welche Aufgabe man mit welchem Programm oder Werkzeug erledigen kann. So sind die Aufgaben in Visualisierung, Analyse, Manipulation, Konvertierung und Datenbereitstellung untergliedert (*MITCHELL ET AL. 2008*). Alles in allem ist dieses Buch ähnlich dem Buch von *SHERMAN (2008)*. Es wird hier allerdings mehr Wert auf die Praxis und das Erstellen einer Internetkarte gelegt, wohingegen bei *SHERMAN (2008)* die Desktopkarten und die theoretischen Aspekte näher erläutert werden.

Zum Thema Open Source GIS Software gibt es auch einige wissenschaftliche Artikel. So schreiben *ANDERSON* und *MORENO-SANCHEZ (2003)* in ihrem Artikel *Building Web-Based Spatial Information Solutions around Open Specifications and Open Source Software* wie man mit Hilfe dieser Spezifikationen und dieser Software eine Webanwendung erstellen kann. Es geht darum, dass sich Geoinformationssysteme immer mehr ins Internet verlagern und sich dadurch eine höhere Interoperabilität ergibt. Bei proprietären Systemen wurden die Nutzer durch die hohen Kosten abgeschreckt und somit wollen die beiden Autoren das Potential eines Prototyps eines Web-GIS testen, dass nur mit Open Source Software erstellt wurde. Sie kamen dabei auf das Ergebnis, dass man aktuelle webbasierte Geoprozessierungslösungen mit der aktuell erhältlichen Open Source Software und Open Specifications erstellen kann (*ANDERSON & MORENO-SANCHEZ, 2003*).

Ein weiterer Beitrag stammt von *Donnelly (2010)*, der in seinem Artikel *Evaluating open source GIS for libraries* (*DONNELLY, 2010*) beschreibt, wie man diese Programme in Bibliotheken einsetzen kann. Er gibt dabei einen Überblick über Freie und Open Source GIS Software und stellt die Frage, ob Bibliotheken oder akademische Einrichtungen FOSS GIS Software nutzen sollten. Es wurde dabei getestet, wie man in sechs verschiedenen FOSS GIS Softwarepaketen eine thematische Karte erstellt. Dieser Test sollte zeigen, wie im Vergleich zu proprietärer Software eine solche Karte in Open Source Software erstellt werden kann. Das Ergebnis zeigt, dass jedes der einzelnen Softwarepakete seine Eigenarten besitzt und etwas schlechter im Vergleich zu *ArcGIS* abschnitt. In der Zusammenfassung kann man lesen, dass

⁷⁰ <http://www.gdal.org/>

⁷¹ <http://www.gdal.org/ogr/>

FOSS GIS Software noch nicht ganz an proprietäre Software heranreicht, sich aber in einer positiven Entwicklung befindet und somit aufschließen kann (DONNELLY, 2010).

Auch *Refractions Research Inc.*⁷² hat einen Beitrag zum Status von Open Source GIS herausgegeben. In dem Dokument *The State of Open Source GIS* (RAMSEY, 2004) werden neben der Erklärung der Begriffe Open Source und Open Source GIS verschiedene Programme aufgeteilt nach ihren Programmiersprachen vorgestellt. Es wird dabei erwähnt, dass man mit Open Source GIS Software mittlerweile jede Art von räumlichen Problemstellungen lösen kann und diese Software eine komplette Alternative zu proprietärer Software ist. Das Dokument stellt die aktuellen Programme vor und erläutert diese kurz mit anschaulichen Screenshots oder Abbildungen. Dabei wird zwischen C-Projekten und Java-Projekten unterschieden (RAMSEY, 2004).

In der *Encyclopedia of GIS* von SHEKAR und XIONG (2008) wird ebenfalls über Open Source GIS Software geschrieben. Es werden dabei Programme wie *GRASS GIS*, *PostGIS*, *Quantum GIS* und *UMN MapServer*. Dabei handelt es sich um die bekannteste Open Source GIS Software, die bereits durch viele andere Autoren vorgestellt wurde. Des Weiteren findet man auch Beiträge über Open Source GIS Libraries wie *GDAL* oder *OGR*. In diesen Beiträgen werden hauptsächlich die Funktionen und die Anwendungsmöglichkeiten der Software betrachtet. Die *Encyclopedia of GIS* (SHEKAR & XIONG, 2008) ist eine Sammlung verschiedener Beiträge zum Thema GIS. Neben den genannten Beiträgen finden sich weitere Beiträge über die unterschiedlichsten Themen. Dabei kann man etwas über Basic Concepts, Road Network Databases bis hin zu Photogrammetry, Voronoi oder auch Indoor Positioning erfahren.

Aufgebaut ist dieses Buch wie ein Lexikon, in dem die Einträge alphabetisch geordnet sind und man sich das jeweilige Schlagwort sucht. Zu jedem Schlagwort gibt es dann die Angabe vorhandener Synonyme, eine Definition, einen historischen Hintergrund, wissenschaftliche Grundlagen, Hauptanwendungen, einen Ausblick, Referenzen und eine Liste weiterer Artikel zum Thema (SHEKAR & XIONG, 2008). Das Werk ist eher als ein Nachschlagewerk anzusehen, welches sehr umfangreich das Thema GIS und dessen verschiedene Wissenschaftsbereiche abdeckt.

⁷² <http://refractions.net/>

3 Praxis

Im vorhergehenden Kapitel wurden wichtige Definitionen getroffen, Begriffe erläutert und Methoden erklärt. Dieses Kapitel beschreibt, wie man die Aufgabenstellung *Automatisierte Generierung von Postleitzahlgebieten aus OpenStreetMap-Daten unter Verwendung von Open Source GIS Software* praktisch umsetzt und wie der Vergleich der erzeugten Daten durchgeführt wird. Dabei sollen die technischen Voraussetzungen betrachtet werden. Ebenso soll der Ablauf schrittweise erläutert werden, wobei kurz auf die Funktionen der genutzten Programme eingegangen werden soll und eine Art Workflow erstellt wird. Nachdem die Gebiete erstellt wurden, werden diese unter Benutzung der genannten Vergleichskriterien untersucht, um diese Gebiete qualitativ zu bewerten.

3.1 Technische Voraussetzungen

Die Erstellung der Postleitzahlgebiete aus *OSM*-Daten setzt keine hohen Voraussetzungen an die Hard- bzw. Software. Es genügt hierfür ein aktueller Computer mit ausreichend vorhandenem Speicherplatz für die Daten. So benötigen die Europadaten⁷³ von *OSM* allein ca. 7,5 GB (stetig steigend) für den Download vom Server. Für die *PostgreSQL*-Datenbank mit den eingespielten Europadaten wird dann zusätzlich noch ein Speicherplatz von ca. 70 GB benötigt. Sollte man sich für die Daten der ganzen Welt (Planetfile⁷⁴) entscheiden, so sind hierfür ca. 17 GB für den Download und ca. 200 GB für die Datenbank einzuplanen.

Neben ausreichend Speicherplatz wird lediglich spezielle Software, *PostgreSQL* mit *PostGIS* und *GRASS GIS* benötigt. Diese Programme laufen auf den gängigen Betriebssystemen wie *MS Windows*, *Mac OS X* und *Linux*, sodass hierfür keine speziellen Voraussetzungen notwendig sind. *GRASS GIS* und *PostgreSQL* mit *PostGIS* sind Open Source Software und können frei genutzt werden.

Lediglich für den Vergleich der Datensätze wurde proprietäre Software genutzt. Die statistischen bzw. räumlichen Vergleiche wurden mit *MS Excel* und *ArcGIS* durchgeführt. Hier sei erwähnt, dass man diese Vergleiche auch mittels Open Source Software durchführen hätte können.

3.2 Datengrundlage

Bei der Bearbeitung wird vorausgesetzt, dass eine *PostgreSQL*-Datenbank mit *PostGIS*-Erweiterung vorliegt, in die bereits die *OSM*-Daten mittels *osm2pgsql*⁷⁵ importiert wurden. In dieser Arbeit wurde dabei der Europadatensatz (Stand 27.03.2011) von *OSM* verwendet. Für

⁷³ <http://download.geofabrik.de/osm/europe.osm.bz2>

⁷⁴ <http://ftp5.gwdg.de/pub/misc/openstreetmap/planet.openstreetmap.org/planet-latest.osm.bz2>

⁷⁵ <http://wiki.openstreetmap.org/wiki/Osm2pgsql>

die Bearbeitung hätte ebenso der Deutschlanddatensatz⁷⁶ oder der Weltdatensatz (Planetfile) genutzt werden können. All diese Daten können unter den genannten Links heruntergeladen werden. Es wird dabei immer der aktuelle Stand heruntergeladen.

Als Testgebiet wurde Deutschland gewählt. Hier kann ebenfalls ein beliebiges anderes Land gewählt werden. Wählt man ein anderes Land ist aber darauf zu achten, dass man die Postleitzahlen in der Vorverarbeitung (siehe nächstes Kapitel) richtig bearbeitet, da es in verschiedenen Ländern verschiedene Postleitzahlformate gibt. Darauf soll aber nicht weiter eingegangen werden, da hier ausschließlich Deutschland betrachtet werden soll.

Da das Gebiet auf Deutschland begrenzt wurde, wurde hierfür noch die Deutschlandgrenze benötigt. Diese wurde als Shapefile von *Natural Earth Data*⁷⁷ heruntergeladen. Dabei wurden die *Admin 0 – Countries*⁷⁸ (Version 1.3.0, Stand 29.01.2011), die Grenzen der Länder der Welt, als Shapefile heruntergeladen und die Grenze von Deutschland als einzelnes Shapefile extrahiert. Bei diesen Daten handelt es sich um frei verfügbare Daten.

Man hätte die Deutschlandgrenze ebenso von *OSM* beziehen können. Da sind die Grenzen der Länder der Welt, *world_boundaries*⁷⁹, als Shapefile verfügbar (RAMM & TOPF, 2010). Allerdings sind diese Daten nicht auf dem aktuellen Stand, sondern Stand März 2008, weshalb diese Daten hier nicht verwendet wurden.

Neben den *OSM*-Daten sind für den späteren Vergleich auch die Daten eines proprietären Anbieters notwendig. Dabei wurden die PLZ-Daten von *TeleAtlas* verwendet. Es handelt sich hierbei um die PLZ5-Gebiete Deutschlands (Stand 05.07.2010), welche als Shapefile vorhanden sind. Da es sich hier lediglich um die PLZ5-Gebiete handelt, aber auch die PLZ1-, PLZ2- und PLZ3-Gebiete verglichen werden sollten, wurden diese mit Hilfe von ArcGIS und der Funktion `Dissolve` erzeugt. Sodass diese vier verschiedenen PLZ-Gebiete als Shapefile zum Vergleich vorliegen.

Bei allen Daten ist darauf zu achten, dass sie in der gleichen Projektion vorhanden sind, damit die Verwendung bzw. der Vergleich auch ohne Probleme durchgeführt werden kann. Für die Bearbeitung wurde jeweils *WGS 1984* als geographisches und *WGS 1984 World Mercator* als projiziertes Koordinatensystem verwendet.

⁷⁶ <http://download.geofabrik.de/osm/europe/germany.osm.bz2>

⁷⁷ <http://www.naturalearthdata.com/>

⁷⁸ <http://www.naturalearthdata.com/downloads/10m-cultural-vectors/10m-admin-0-countries/>

⁷⁹ http://tile.openstreetmap.org/world_boundaries-spherical.tgz

3.3 Allgemeiner Arbeitsablauf

Die Erstellung der PLZ-Gebiete mit der genannten Software erfolgte in verschiedenen Einzelschritten. Diese Einzelschritte werden im Folgenden kurz erläutert. Die Durchführung wird später anhand von *GRASS GIS* näher beschrieben. Es ist möglich sämtliche Arbeitsschritte mittels *GRASS GIS* zu tätigen. Dabei wird davon ausgegangen, dass die Ausgangsdaten alle in der Datenbank vorliegen, das betrifft die *OSM*-Daten und die Deutschlandgrenze.

Der Programmablauf wird in Datenvorverarbeitung, Erzeugung der Gebiete und Datennachbearbeitung gegliedert. Abbildung 3.1 stellt den Programmablauf graphisch dar. Hierbei ist sicher strittig welche Schritte man in welchen Bereich einordnet. Die folgende Aufteilung trennt mehr oder minder die Bearbeitung mittels SQL-Befehlen von der direkten Bearbeitung in *GRASS GIS*. Anschließend wird noch kurz ein Hinweis zum Export der Daten als Shapefile gegeben, da dies für den Vergleich notwendig ist.

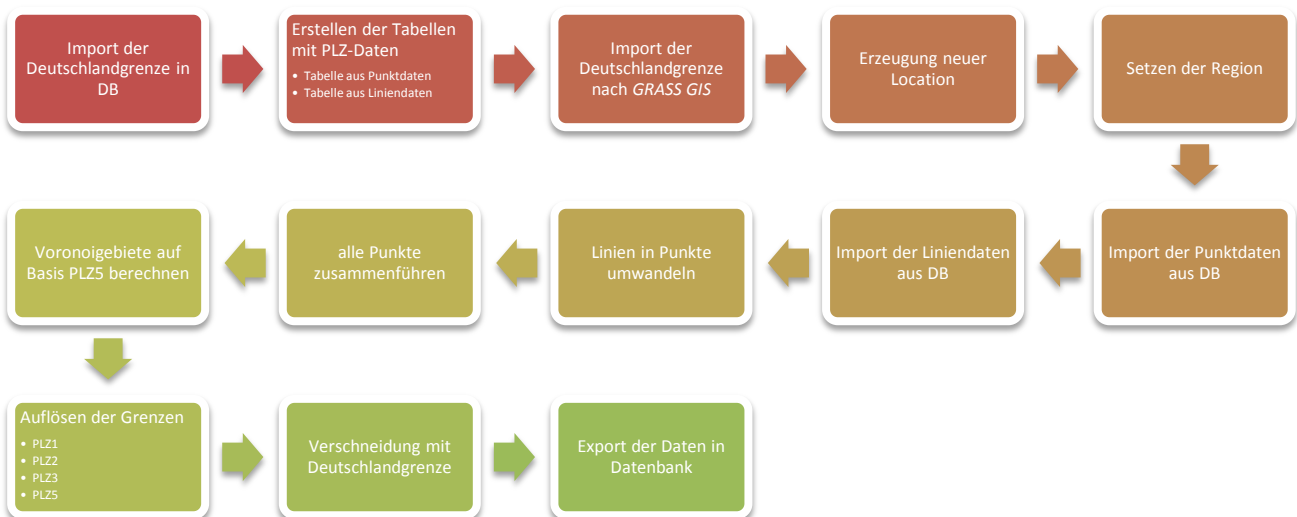


Abb. 3.1 allgemeiner Arbeitsablauf

3.3.1 Datenvorverarbeitung

Nachdem alle Daten in der richtigen Projektion vorhanden sind, kann mit der Datenvorverarbeitung begonnen werden. Als erstes muss dabei die Deutschlandgrenze in die *PostgreSQL*-Datenbank importiert werden, da diese nicht im Europadatensatz von *OSM* vorhanden ist. Dieser Import wird mit Hilfe von *shp2pgsql*⁸⁰, welches in der *PostGIS*-Installation enthalten ist, durchgeführt. Die Deutschlandgrenze wird dabei als neue Tabelle in der Datenbank angelegt. Nach diesem Import befinden sich nun alle notwendigen Daten in der Datenbank.

⁸⁰ http://www.bostongis.com/pgsql2shp_shp2pgsql_quickguide.bqg

Als nächster Schritt werden mittels SQL neue Tabellen in der Datenbank erzeugt, welche zur weiteren Bearbeitung benötigt werden und lediglich Daten mit Postleitzahlenangabe innerhalb Deutschlands beinhalten. Dabei ist darauf zu achten, dass beide Tags, *postal_code* und *addr:postcode*, betrachtet werden.

Durch den Import der Daten mittels *osm2pgsql* liegen in der Datenbank folgende vier Tabellen vor (RAMM & TOPF, 2010):

- *europe_point* → alle punktförmigen Objekte
- *europe_line* → alle linienförmigen Objekte
- *europe_polygon* → alle flächigen Objekte
- *europe_roads* → einige linienförmige Objekte für niedrige Zoom-Level

Die Bezeichnung der Tabellen erfolgt beim Import automatisch. Lediglich das Präfix, in diesem Fall *europe*, kann durch den Nutzer gewählt werden. So würde man hier bei Nutzung des Planetfiles als Präfix *planet* wählen bzw. würde *osm2pgsql* den Namen automatisch anhand des *OSM*-Dateinamens wählen lassen.

Für die weitere Bearbeitung wurde lediglich die Tabelle *europe_point* und *europe_roads* verwendet, da sonst die Datenmenge zu groß wird und nicht mehr durch *GRASS GIS* verarbeitet werden kann.

Die neu zu erstellenden Tabellen sollen dabei die Punkte und Linien aus den genannten Tabellen enthalten, die entweder einen Eintrag bei *postal_code* oder einen Eintrag bei *addr:postcode* besitzen. Um dies zu testen muss vorher die Spalte *addr:postcode* in *addr_postcode* umbenannt werden, da es sonst zu Problemen bei der weiteren Bearbeitung kommt. Der Test ist notwendig, dass es in *OSM* die zwei verschiedenen Möglichkeiten gibt, einem Objekt eine Postleitzahl zuzuweisen.

Des Weiteren sollen diese Einträge bestimmte Eigenschaften besitzen. Diese Eigenschaften werden bei der Erstellung abgefragt. Dabei wird eine *WHERE*-Bedingung mit einer zweiten *WHERE*-Bedingung verknüpft. Die erste Bedingung prüft sozusagen die Einträge in der Spalte *addr_postcode*, während die zweite Bedingung die Einträge in der Spalte *postal_code* prüft. Dabei wird beachtet, dass der Eintrag nicht leer ist, der Eintrag genau fünfstellig ist, der Eintrag nur aus Ziffern zwischen 0 und 9 besteht, der Eintrag keine Buchstaben oder Sonderzeichen enthält und ob das jeweilige Objekt innerhalb Deutschlands liegt. Die Prüfung auf Sonderzeichen oder Buchstaben, sowie die Einhaltung der Regel nur Ziffern zwischen 0 und 9,

wird mittels regulären Ausdrücken durchgeführt. Die Prüfung der Länge wird mittels der Funktion `length`, die Prüfung auf Inhalt mittels `IS NOT NULL` durchgeführt. Zur Prüfung der Lage innerhalb der Grenze Deutschland wird abgefragt, ob sich das jeweilige Objekt mit der Deutschlandgrenze schneidet (`Intersects`). Hierbei wird auf eine Funktion von *PostGIS* zugegriffen, während die anderen Prüfungen auch ohne *PostGIS* möglich wären. Zum Prüfen der Lage wird hierbei die Geometrie des jeweiligen Objektes betrachtet. Es wird also getestet, ob die Geometrie des Objektes sich mit der Geometrie der Deutschlandgrenze schneidet bzw. innerhalb dieser liegt.

Bei der Erstellung der neuen Tabellen wird lediglich die Postleitzahl und die Geometrie aus den Ausgangstabellen übernommen, alle anderen Werte werden nicht benötigt. Die Postleitzahl wird dabei als `PLZ1`, `PLZ2`, `PLZ3` und `PLZ5` übernommen, sodass die neuen Tabellen diese Spalten besitzen. Dabei wird die geprüfte `PLZ5` auf ein, zwei bzw. drei Stellen gekürzt, um die anderen Spalten zu erzeugen und alle `PLZ`-Kategorien abzudecken.

Neben diesen Prüfungen muss den neuen Tabellen noch eine `ID` hinzugefügt werden, damit jeder Eintrag eindeutig zugeordnet werden kann. Es wird also ein Primärschlüssel erstellt. Dies ist notwendig, damit die Daten später verarbeitet werden können. Zusätzlich zum Primärschlüssel ist auch noch der Eintrag in die Geometrietabelle der Datenbank erforderlich. Dieser Eintrag dient dazu, die Art der Geometrie der jeweiligen Tabelle, die Dimension der Daten und das Koordinatensystem, sowie weitere Attribute festzulegen. Das ist erforderlich, damit ein auf die Daten zugreifendes Programm, zum Beispiel *GRASS GIS*, weiß, in welchem Koordinatensystem die Daten gespeichert sind und um welchen Geometriety es sich handelt. Hier gibt es Punkt, Linie, Fläche oder auch allgemein Geometrie als möglichen Typ. Das Koordinatensystem wird mittels *EPSG-Code*⁸¹ gespeichert. Ein *EPSG-Code* ist eine eindeutige vier- bis fünfstellige Schlüsselnummer für Koordinatenreferenzsysteme.

Unter Betrachtung der genannten Bedingungen wurden nun die Tabellen auf Grundlage der Tabellen *europa_point* und *europa_roads* erzeugt. Man kann die Erstellung der Tabellen dabei als SQL-Abfrage abspeichern und später direkt über *GRASS GIS* aufrufen.

Diese Vorverarbeitung ist notwendig, um die Daten sinnvoll weiter nutzen zu können und nicht relevante bzw. fehlerhafte Daten auszuschließen.

⁸¹ <http://www.epsg.org/>

3.3.2 Erzeugung der Gebiete

Die Erzeugung der Gebiete aus den vorverarbeiteten Daten ist nun der nächste Schritt. Als erstes wurde dabei *GRASS GIS* mit der *PostgreSQL*-Datenbank verbunden, um auf diese Daten zugreifen zu können. Diese Verbindung ist auch notwendig, wenn man die SQL-Abfrage direkt aus *GRASS GIS* aufrufen möchte.

Als nächstes wurde die Deutschlandgrenze in *GRASS GIS* geladen. Dabei wurde gleich eine neue *Location* erzeugt, die als Grundlage für die weiteren Arbeiten verwendet wird. Beim Import der Deutschlandgrenze wurde dabei zusätzlich ein neues *Mapset* in *GRASS GIS* angelegt. Die *Location* dient zur Angabe der Projektion bzw. des Koordinatensystem, welches hierbei gleich aus der Datenbank übernommen wurde. Das *Mapset* dagegen gibt das Verzeichnis mit den GIS-Dateien an. Diese werden standardmäßig, soweit nicht anders konfiguriert, im Nutzerverzeichnis gespeichert. Nachdem Import wechselt man zur neuen *Location*, damit die weiteren Daten ebenfalls diese *Location* erhalten und im gleichen *Mapset* gespeichert werden.

Für die spätere Berechnung der Gebiete ist es notwendig, eine *Region*, sprich einen Bereich bzw. eine Ausdehnung der Daten anzugeben. Hierbei wurde die *Region* der Deutschlandgrenze gewählt. Würde man keine *Region* setzen, so würden die späteren Gebiete evtl. falsch erzeugt, da zum Beispiel eine zu kleine oder zu große *Region* standardmäßig eingestellt ist. Die *Region* dient zur Abgrenzung des Bearbeitungsbereiches innerhalb *GRASS GIS*.

Wurden diese Schritte durchgeführt folgt nun der Import der Daten der beiden neu erstellten Tabellen aus der Datenbank. Hierbei handelt es sich um die Punktdaten aus der *europe_point*-Tabelle und die Liniendaten aus der *europe_roads*-Tabelle, welche die vorher genannten Bedingungen erfüllen.

Nachdem beide Datensätze in *GRASS GIS* importiert wurden, wurden anschließend die Linien in Punkte konvertiert und beide Datensätze, die Punkte aus der Datenbank und die in Punkte konvertieren Linien, zusammengefügt und somit ein Datensatz für die weitere Bearbeitung erhalten.

Aus diesem Punktdatensatz wurden nun die Voronoi-Polygone auf Basis der PLZ5 erzeugt. Zur Erinnerung, der Datensatz enthält die PLZ1-, PLZ2-, PLZ3- und PLZ5-Daten. Bei der Erzeugung der Gebiete wird dabei für jedes Objekt mit der jeweiligen PLZ5 eine Fläche erzeugt. Es entstand also ein neuer Datensatz mit der gleichen Anzahl an Flächen wie der Eingabedatensatz Punkte hatte.

Da es Punkte gab, welche die gleiche Postleitzahl besitzen, gibt es nun auch Flächen, welche die gleiche Postleitzahl besitzen. Um diese Flächen zu verbinden wurden die Grenzen zwischen Flächen mit gleichen Attributen aufgelöst. Dies geschah auf Basis der PLZ1-, PLZ2-, PLZ3- und PLZ5-Daten, damit man als Ergebnis die jeweiligen Postleitzahlgebiete erhält.

Die vier erhaltenen Datensätze wurden anschließend jeweils mit der Deutschlandgrenze verschnitten, damit diese Daten grenzscharf vorliegen und nicht über die Grenze Deutschlands hinausgehen, was bei der Erzeugung der Gebiete durch die Voronoi-Funktion passieren kann. Aus diesem Grund wurde anfangs auch die Region gesetzt, damit diese Gebiete nicht unendlich groß werden.

Nach diesen Schritten befindet sich nun der Datensatz der PLZ1-, PLZ2-, PLZ3- und PLZ5-Gebiete Deutschlands im Datenraum von *GRASS GIS*. Für die Nutzung in *PostgreSQL* sind dabei noch einige wenige Schritte notwendig, die im nächsten Abschnitt genannt werden.

3.3.3 Datennachbearbeitung

Da sich die Daten jetzt lediglich in *GRASS GIS* befinden und somit noch nicht in der *PostgreSQL*-Datenbank, wurden diese Daten nun in die Datenbank exportiert. Es wurden also die vier Datensätze in die Datenbank geschrieben.

Durch die Verschneidung mit der Deutschlandgrenze enthalten die nun in der Datenbank vorhandenen neuen Tabellen zu den Postleitzahlgebieten zusätzliche Spalten, welche für eine spätere Verwendung nicht notwendig sind. Diese Spalten wurden mittels einer SQL-Abfrage gelöscht und die übriggebliebene Spalte mit der jeweiligen PLZ-Kategorie wurde sinnvoll umbenannt.

Nach diesen Schritten befinden sich schließlich die Daten, mit sinnvollen Bezeichnungen und den notwendigen Spalten, in der *PostgreSQL*-Datenbank. Man könnte jetzt die Daten innerhalb *GRASS GIS*, sowie die anfangs erzeugten Tabellen mit den Punkten und Linien der *OSM*-Daten löschen um Speicherplatz zu sparen. Das Löschen der Tabellen wurde dabei in der SQL-Abfrage zum Löschen der Spalten mit durchgeführt. Das Löschen der Daten aus *GRASS GIS* erfolgte anschließend innerhalb *GRASS GIS*.

3.3.4 Export der Daten als Shapefile

Für den Vergleich der erzeugten Daten aus dem *OSM*-Datensatz mit dem Datensatz von *TeleAtlas* wurden die Daten nun anschließend aus der *PostgreSQL*-Datenbank exportiert. Dieser

Export wurde mittels *pgsql2shp*⁸², welches in der *PostGIS*-Installation enthalten ist, durchgeführt. Man hätte hier aber auch den direkten Export aus *GRASS GIS* wählen können. Beim Export wurde auf die Datenbank zugegriffen und das jeweilige Shapefile in der gewünschten Datei abgespeichert. Es entstanden also vier neue Shapefiles, jeweils eines pro PLZ-Kategorie. Diese Shapefiles wurden später zum Vergleich der Daten herangezogen, siehe hierzu Kapitel 3.6 Vergleich der Daten.

3.4 Praktische Umsetzung

Im vorhergehenden Kapitel wurde der Ablauf nur allgemein beschrieben. Dieser soll nun speziell in Bezug auf *GRASS GIS* nochmals erläutert werden. Hierbei sollen kurz die genutzten Funktionen genannt werden. Der Ablauf ist dabei wie oben gegliedert. Es wurde bei der Bearbeitung darauf geachtet, dass sämtliche genutzte Funktionen ohne graphische Benutzeroberfläche auskommen und aus diesem Grund über die Kommandozeile aufgerufen wurden. Der genaue Code ist dem Anhang zu entnehmen. Weitere Hinweise zu den einzelnen Funktionen findet man in der jeweiligen Dokumentation.

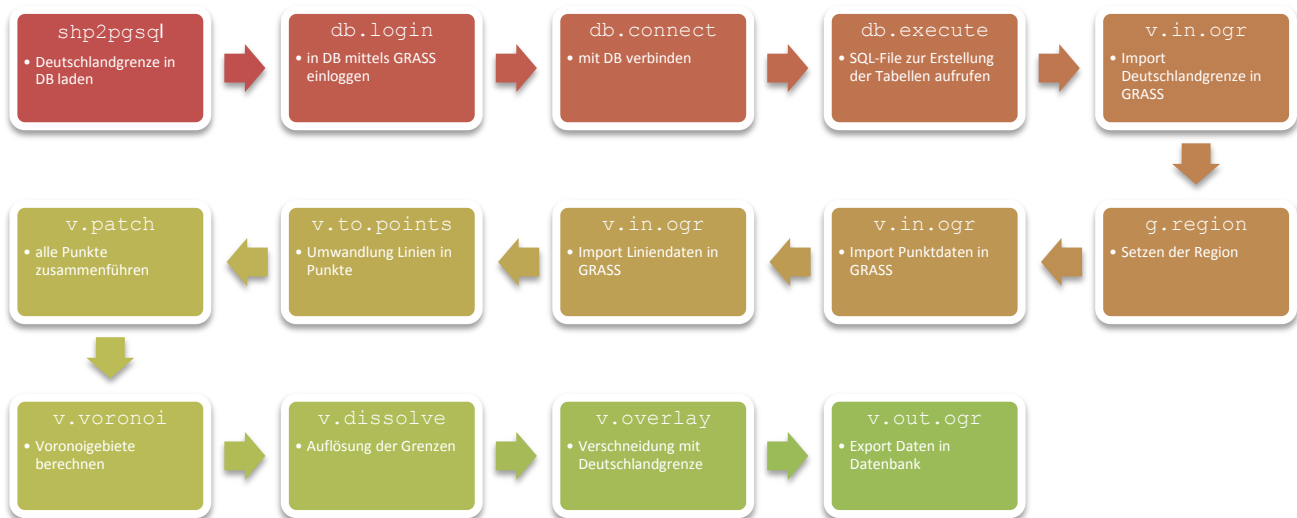


Abb. 3.2 schematischer Arbeitsablauf mit Funktionsnamen aus GRASS GIS

3.4.1 Datenvorverarbeitung

Der Import der Deutschlandgrenze erfolgte mit *shp2pgsql*, einem Tool zum Import von Shapefiles in eine *PGSQL*-Datenbank. Der Aufruf des Importes erfolgte direkt in der Kommandozeile. Beim Aufruf sind dabei die Projektion, das zu importierende Shapefile mit Speicher-

⁸² http://www.bostongis.com/pgsql2shp_shp2pgsql_quickguide.bqg

ort und die Daten zur Datenbank, wie Tabellename, Host, Datenbank und Nutzernamen anzugeben.

Nach dem anschließenden Start von *GRASS GIS* (Textvariante) wurde die Verbindung zur Datenbank hergestellt. Dies wurde über den Befehl `db.connect` und die Angabe des Datenbanktreibers, des Hosts und des Datenbanknamens erledigt. Anschließend meldete man sich mit `db.login` unter Angabe des Nutzernamen und des Passwortes an. Die Verbindung von *GRASS GIS* zur *PostgreSQL*-Datenbank war somit hergestellt.

Mit Hilfe des Befehles `db.execute`, welcher ausführende Operationen auf die Datenbank erlaubt, wurde anschließend die SQL-Abfrage zur Erstellung der Tabellen aufgerufen. Hierbei muss man lediglich die SQL-Datei aufrufen. Die Abfrage wird anschließend im Hintergrund durchgeführt und die erfolgreiche Durchführung bestätigt. Die Datei zur SQL-Abfrage wurde vorher erstellt und beinhaltet die Funktionen zum Erstellen der neuen Tabellen.

Bevor neue Tabellen erzeugt wurden, wurden alte evtl. vorhandene Tabellen mit gleichem Namen gelöscht (`DROP TABLE IF EXISTS ...`) und auch aus der Geometrietabelle entfernt (`DELETE FROM ... WHERE ...`). Diese Funktionen wurden genutzt, da der Datensatz aktuell gehalten werden soll und somit bei der Aktualisierung die veralteten Daten automatisch entfernt werden können.

Die Spalten der mittels *osm2pgsql* eingespielten Daten wurden ebenfalls umbenannt. Hierzu wurde die Funktion `ALTER TABLE ... RENAME COLUMN ... TO ...` verwendet.

Die in der SQL-Abfrage hauptsächlich verwendete Funktion ist `CREATE TABLE ... AS SELECT ... FROM ... WHERE ...`. Dabei handelt es sich um eine SQL-Funktion zum Erstellen einer neuen Tabelle aus bereits vorhandenen Daten unter Angabe des neuen Tabellennamens, der zu nutzenden Spalten (`SELECT`), der Tabellen (`FROM`) und der Bedingungen (`WHERE`), die die Daten erfüllen sollen. Hier wurde im Speziellen noch die Funktion `substring` genutzt, welche die erste, die ersten zwei und die ersten drei Ziffern der PLZ5-Daten zurückgeben soll. Mit Hilfe dieser Funktion wurden die PLZ1-, PLZ2- und PLZ3-Daten erzeugt, so dass jeder PLZ5 diese Daten zugeordnet werden konnten. Innerhalb der Bedingung (`WHERE`) wurde geprüft, dass die PLZ nicht leer ist (`postal_code IS NOT NULL`). Des Weiteren wurden nur Daten verwendet, deren Länge gleich fünf ist (`length(postal_code)=5`), die nur aus Ziffern bestehen (`postal_code ~ '[0-9]'`), die keine Buchstaben oder Sonderzeichen enthalten

(`postal_code !~ '[a-zA-Z.?!]'`) und die innerhalb der Deutschlandgrenze liegen, spricht sich mit dieser schneiden (`Intersects (germany.geometry, europe_point.way)`). Für die letzte Abfrage wurde die Geometrie der Deutschlandgrenze und die der Punktdaten verwendet, weshalb innerhalb des `FROM`-Befehls beide Tabellen genannt werden müssen. Diese Abfragen wurden für `postal_code` und `addr_postcode` durchgeführt und miteinander vereint (`UNION`).

Zusätzlich wurden für die neuen Tabellen neue Primärschlüssel erzeugt. Dafür wurde eine neue Spalte hinzugefügt (`ALTER TABLE ... ADD COLUMN ...`) und danach der Schlüssel erzeugt und eingetragen (`ALTER TABLE ... ADD CONSTRAINT ... PRIMARY KEY ...`). Danach wurden die Daten zur Tabelle noch in die Geometrietabelle eingetragen. Hier wurden einfach die Werte mittels `INSERT INTO ... VALUES ...` übergeben.

Diese SQL-Abfrage wurde auf die Tabelle `europe_point` und `europe_roads` angewendet.

Anschließend an die Erstellung der neuen Tabellen wurde noch `VACUUM` und `ANALYZE` durchgeführt. `VACUUM` dient dabei zur Wiedergewinnung von belegtem Speicherplatz und schützt vor Verlust alter Daten. `ANALYZE` dient zur Aktualisierung der Statistiken über die Tabellen.

3.4.2 Erzeugung der Gebiete

Nachdem nun die SQL-Abfrage komplett durchgeführt wurde, wurde die Deutschlandgrenze mittels `v.in.ogr` in *GRASS GIS* importiert. Hierbei wurde gleich eine neue Location erstellt, die dann mittels `g.gisenv` als aktuelle Location gesetzt wurde. Mittels `g.region` wurde außerdem eine neue Region gesetzt.

Anschließend wurden mittels `v.in.ogr` die beiden Tabellen mit den Postleitzahl-daten importiert. Die Liniendaten wurden danach mittels `v.to.points` in Punkte konvertiert. Dabei wurden lediglich die Verbindungspunkte gewählt, da es sonst am Ende zu viele Punkte geworden wären und *GRASS GIS* diese nicht hätte verarbeiten können. Diese Punkte wurden dann mittels `v.patch` mit den Punktdaten aus der Datenbank zusammengefügt. Dabei ist darauf zu achten, dass auch die Attributtabelle mit zusammengefügt wird.

Nachdem nun alle Punkte in einer Datei vorhanden waren, wurden aus diesen Punktdaten die Voronoi-Polygone berechnet. Hierfür wurde die Funktion `v.voronoi` verwendet. Danach wurden die Grenzen dieser Gebiete mittels `v.dissolve` aufgelöst und daraus die PLZ1-, PLZ2-, PLZ3- und PLZ5-Gebiete erzeugt.

Die vier Datensätze wurden schließlich mit Hilfe von `v.overlay` mit der Deutschlandgrenze verschnitten.

3.4.3 Datennachbearbeitung

Nachdem die Erzeugung der Gebiete abgeschlossen war, wurden die vier Datensätze nun wieder in die *PostgreSQL*-Datenbank exportiert. Dies ermöglicht die Funktion `v.out.ogr`.

Anschließend wurden unnötige Spalten gelöscht und der Spaltenname angepasst. Hierfür wurde wieder eine SQL-Abfrage aufgerufen. Dabei wurden mittels `ALTER TABLE ... DROP COLUMN ...` die unnötigen Spalten entfernt und mit `ALTER TABLE ... RENAME COLUMN ... TO ...` die PLZ-Spalte umbenannt. Ebenso wurden die anfangs erstellten Tabellen mittels `DROP TABLE IF EXISTS ...` gelöscht.

Als letzter Schritt wurden nun noch die Daten aus *GRASS GIS* entfernt, damit es bei einer Aktualisierung und der dadurch verbundenen neuen Erzeugung der PLZ-Gebiete nicht zu Problemen kommt. Um die Daten aus *GRASS GIS* zu löschen wurde `g.remove` genutzt. Nach diesem letzten Schritt wurde *GRASS GIS* geschlossen.

3.4.4 Export der Daten als Shapefile

Der für den Vergleich notwendige Export der Daten als Shapefile wurde mittels `pgsql2shp`, einem Tool zum Export von *PGSQL*-Daten als Shapefile, durchgeführt. Der Aufruf des Exportes erfolgte direkt in der Kommandozeile. Beim Aufruf sind dabei der Name des Shapefiles samt Speicherort, die Daten zur Datenbank, wie Tabellename, Host, Datenbank und Nutzernamen, sowie das Passwort und die zu exportierende Tabelle anzugeben.

Diesen Export hätte man auch mittels `v.out.ogr` direkt über *GRASS GIS* durchführen können.

3.5 Ergebnisse

Als Ergebnis des beschriebenen Arbeitsablaufes erhält man verschiedene Daten. So liegen einerseits die Daten in der *PostgreSQL*-Datenbank als neue Tabellen vor, andererseits liegen die Daten als Shapefile vor. Die Daten innerhalb der Datenbank kann man dabei zum Beispiel für ein WMS nutzen oder sie mittels eines Query Layers über ein GIS abfragen. Die Shapefiles kann man für die Weiterverarbeitung innerhalb eines GIS nutzen oder auch für die Weitergabe an eventuelle Interessenten.

In dieser Arbeit wurden dabei die Daten in der Datenbank nicht weiter betrachtet. Die Shapefiles wurden dagegen für weitere Analysen herangezogen. So wurden diese für den Vergleich der erzeugten PLZ-Gebiete mit den PLZ-Gebieten von *TeleAtlas* genutzt. Des Weiteren wurden die Shapefiles genutzt um PDF-Karten zu erzeugen, welche die PLZ-Gebiete graphisch darstellen sollen und somit auch eine Einschätzung der Qualität ermöglichen sollen. Im Zusammenhang zu den PDF-Karten und dem Vergleich, wurden auch Vergleichskarten erstellt.

Beide Datensätze, der in der Datenbank, wie auch die Shapefiles, liegen für die verschiedenen PLZ-Kategorien vor. Es gibt also bei beiden die PLZ1-, PLZ2-, PLZ3- und PLZ5-Gebiete, welche beliebig genutzt werden können.

3.6 Vergleich der Daten

Der Vergleich der aus den *OSM*-Daten erzeugten mit den von *TeleAtlas* erstellten PLZ-Gebieten erfolgt auf Grundlage der bereits genannten Qualitätselemente bzw. Vergleichskriterien.

Als erstes soll dabei die Vollständigkeit betrachtet werden. Tabelle 3.1 zeigt die Anzahl der Postleitzahlen geordnet nach der PLZ-Kategorie und nach Anbieter, die Differenz der Anzahl und den Prozentsatz, der in *OSM* enthaltenen PLZ im Vergleich zu *TeleAtlas*.

| PLZ-Kategorie | Anzahl TeleAtlas | Anzahl OSM | Differenz TA-OSM | Prozent OSM/TA |
|---------------|------------------|------------|------------------|----------------|
| PLZ1 | 10 | 10 | 0 | 100 |
| PLZ2 | 95 | 98 | -3 | 103,2 |
| PLZ3 | 671 | 709 | -38 | 105,7 |
| PLZ5 | 8.242 | 7.425 | 817 | 90,1 |

Tab. 3.1 Anzahl PLZ

Es ist festzustellen, dass der *OSM*-Datensatz 90,1 % der Gesamtzahl der PLZ5-Gebiete von *TeleAtlas* beinhaltet. Das ist ein Unterschied von 817 Stück. Andererseits zeigt der Vergleich, dass im erzeugten *OSM*-Datensatz mehr PLZ2- und PLZ3-Gebiete vorhanden sind. Lediglich die Anzahl der PLZ1-Gebiete ist bei beiden Datensätzen gleich. Die unterschiedliche Anzahl der anderen Gebiete ist auf Datenüberschuss bzw. Datenmangel zurückzuführen, wobei der Datenüberschuss im *OSM*-Datensatz auf fehlerhaften PLZ-Angaben beruht, wohingegen der Datenmangel mit fehlenden PLZ-Angaben zu begründen ist. Diesbezüglich sind in Tabelle 3.2 und Tabelle 3.3 die PLZ-Gebiete nochmal näher untersucht und die Anzahl der nicht übereinstimmenden mit den übereinstimmenden PLZ-Angaben verglichen.

| PLZ-Kategorie | nicht übereinstimmende PLZ | übereinstimmende PLZ |
|---------------|----------------------------|----------------------|
| PLZ1 | 0 | 10 |
| PLZ2 | 3 | 95 |
| PLZ3 | 38 | 671 |
| PLZ5 | 1392 | 7137 |

Tab. 3.2 Anzahl nicht übereinstimmender und übereinstimmender PLZ

| PLZ-Kategorie | OSM | TeleAtlas |
|---------------|-----|-----------|
| PLZ1 | 0 | 0 |
| PLZ2 | 3 | 0 |
| PLZ3 | 38 | 0 |
| PLZ5 | 288 | 1105 |

Tab. 3.3 Anzahl nicht übereinstimmender PLZ nach Anbieter

Tabelle 3.2 zeigt, dass die Anzahl der nicht übereinstimmenden PLZ-Gebiete in der PLZ5-Kategorie am größten ist. Hier gibt es immerhin 1392 nicht übereinstimmende PLZ-Angaben. Auf diesen Angaben beruhen schließlich auch die Fehler in den anderen Kategorien, da diese aus den PLZ5-Angaben erzeugt wurden. Allerdings ist die Anzahl der nicht übereinstimmenden Angaben in der PLZ2- und PLZ3-Kategorie nicht mehr so hoch. Hier zeigt sich bereits, dass die OSM-Daten für kleinere Maßstäbe (PLZ1-PLZ3) besser geeignet sind als für größere Maßstäbe (PLZ5).

Tabelle 3.3 zeigt die nicht übereinstimmenden PLZ nach Anbieter. Dies ist so zu verstehen, dass der jeweilige Anbieter diese Anzahl an PLZ liefert, der andere diese PLZ aber nicht liefert. Der *OSM*-Datensatz besitzt zum Beispiel 288 PLZ5 die nicht im *TeleAtlas*-Datensatz vorhanden sind, während der *TeleAtlas*-Datensatz 1105 PLZ5 besitzt, die nicht in *OSM* vorhanden sind. Diese fehlerhaften PLZ-Angaben werden in einem späteren Abschnitt stichprobenartig kontrolliert um festzustellen bei welchem Anbieter ein Fehler vorliegt.

Das nächste Vergleichskriterium ist die Fläche. Hierbei wurde in Tabelle 3.4 die Gesamtfläche betrachtet und deren Differenz errechnet. Die Tabelle 3.5 – 3.8 liefern die Ergebnisse des Vergleiches der einzelnen Flächen und der einzelnen Abstände der Centoride der jeweiligen PLZ-Kategorie. Bei den Einzelflächen bzw. Abständen wurden jeweils die Mittelwerte und die mittlere Abweichung vom Mittelwert berechnet.

| | TeleAtlas | OSM | Differenz TA-OSM | Deutschland |
|---|------------|------------|------------------|-----------------------|
| Fläche gesamt in km ² | 354.359,87 | 356.584,97 | -2.225,10 | 357.104 ⁸³ |
| Gesamtflächendifferenz (TA-OSM) PLZ1 in km ² | | | | -2.225,00 |
| Gesamtflächendifferenz (TA-OSM) PLZ2 in km ² | | | | -2.223,85 |
| Gesamtflächendifferenz (TA-OSM) PLZ3 in km ² | | | | -2.110,78 |
| Gesamtflächendifferenz (TA-OSM) PLZ5 in km ² | | | | -2.225,10 |

Tab. 3.4 Vergleich Fläche PLZ-Gebiete gesamt

| PLZ1 | Fläche TeleAtlas in km ² | Fläche OSM in km ² | Differenz TA-OSM in km ² | Abstand Centroide in m |
|--|-------------------------------------|-------------------------------|-------------------------------------|------------------------|
| 0 | 37.296,70 | 37.301,60 | -4,90 | 850,48 |
| 1 | 47.936,90 | 47.825,10 | 111,80 | 1.361,15 |
| 2 | 44.412,40 | 44.467,10 | -54,70 | 1.451,35 |
| 3 | 45.495,30 | 46.639,20 | -1143,90 | 600,88 |
| 4 | 20.197,10 | 20.307,70 | -110,60 | 1.235,75 |
| 5 | 28.859,30 | 28.769,30 | 90,00 | 829,72 |
| 6 | 17.188,10 | 17.439,80 | -251,70 | 1.605,20 |
| 7 | 27.918,30 | 27.880,80 | 37,50 | 1.094,54 |
| 8 | 36.761,30 | 36.700,70 | 60,60 | 1.714,77 |
| 9 | 48.294,50 | 49.253,60 | -959,10 | 384,05 |
| Mittelwert Differenz Fläche in km² | | | | -222,50 |
| mittlere Abweichung vom Mittelwert Differenz Fläche in km² | | | | 337,44 |
| Mittelwert Abstand Centroide in m | | | | 1.112,79 |
| mittlere Abweichung vom Mittelwert Abstand Centroide in m | | | | 360,86 |

Tab. 3.5 Vergleich Einzelfläche PLZ1 und Abstand Centroide PLZ1

| | |
|--|-----------------|
| Mittelwert Differenz Fläche in km² | -23,41 |
| mittlere Abweichung vom Mittelwert Differenz Fläche in km² | 89,66 |
| Mittelwert Abstand Centroide in m | 1.451,67 |
| mittlere Abweichung vom Mittelwert Abstand Centroide in m | 1.078,55 |

Tab. 3.6 Vergleich Einzelfläche PLZ2 und Abstand Centroide PLZ2

| | |
|--|-----------------|
| Mittelwert Differenz Fläche in km² | -3,15 |
| mittlere Abweichung vom Mittelwert Differenz Fläche in km² | 37,60 |
| Mittelwert Abstand Centroide in m | 2.049,42 |
| mittlere Abweichung vom Mittelwert Abstand Centroide in m | 1.990,88 |

Tab. 3.7 Vergleich Einzelfläche PLZ3 und Abstand Centroide PLZ3

| | |
|--|-----------------|
| Mittelwert Differenz Fläche in km² | -3,09 |
| mittlere Abweichung vom Mittelwert Differenz Fläche in km² | 13,58 |
| Mittelwert Abstand Centroide in m | 1.973,11 |
| mittlere Abweichung vom Mittelwert Abstand Centroide in m | 1.904,14 |

Tab. 3.8 Vergleich Einzelfläche PLZ5 und Abstand Centroide PLZ5

⁸³ <http://www.deutschland.de/>

Die Flächendifferenz des gesamten Gebietes beträgt dabei 2.225,10 km². Dabei deckt *TeleAtlas* eine geringere Fläche ab als *OSM*. Dieser Flächenunterschied ist drauf zurückzuführen, dass im *TeleAtlas*-Datensatz teilweise Seen (z.B. Starnberger See) und gemeindefreie Gebiete (z.B. Forstenrieder Park südlich von München) ausgespart sind. Eine weitere Differenz entsteht, da im *OSM* -Datensatz die Wasserflächen an den Grenzen Deutschlands (z.B. Elbemündung Hamburg bis Nordsee) ausgespart sind. Dies ist wiederum auf die genutzte Deutschlandgrenze zurückzuführen, da bei dieser diese Stellen bereits nicht enthalten sind. Dadurch ist auch die Differenz zur Fläche von Deutschland zu erklären. Zum Bereich des Starnberger Sees sei gesagt, dass die *Deutsche Post* hier keine Aussparungen getätigt hat, sondern der See zu einem Postleitzahlgebiet gehört und somit diese Fläche abgedeckt ist. Es ergibt sich also auch eine Differenz zwischen *TeleAtlas* und *Deutsche Post*. Sieht man die Daten der *Deutschen Post* als amtlich an, so ergibt sich für die erzeugten Gebiete aus den *OSM*-Daten sogar eine bessere Abdeckung, da hier die Seen mit abgedeckt sind. Der Vergleich der Einzelflächen zeigt, dass die Abweichungen kleiner werden, je höher die PLZ-Kategorie wird. So sind die Abweichungen im PLZ5-Bereich bedeutend kleiner als im PLZ1-Bereich. Das ist allerdings damit zu begründen, dass die PLZ5-Flächen allgemein kleiner sind als die PLZ1-Flächen und somit nicht solch große Abweichungen entstehen können. Zusammenfassend kann man sagen, dass die Flächeninhalte, sowohl gesamt wie auch einzeln, doch merkbar voneinander abweichen und hier eine bessere Methode zur Erzeugung der PLZ-Gebiete bzw. eine Verbesserung der genutzten Methode erforderlich ist.

Zu den Abständen der Centroide sei gesagt, dass der mittlere Abstand ca. 2 km beträgt. Betrachtet man hierzu noch die mittlere Abweichung kann man sagen, dass der Abstand bei den meisten Centroide nicht mehr als 4 km beträgt. Dieses Ergebnis zeigt, dass die *OSM*-Daten relativ nah an denen von *TeleAtlas* liegen. Speziell die PLZ1-Centroide liegen sehr dicht beieinander.

Tabelle 3.9 liefert bezüglich der Abstände die Ergebnisse aus der Anwendung der Puffer-Vergleichsmethode.

| PLZ-Kategorie | Anzahl OSM-Centroide | 1 km Puffer | 2 km Puffer | 5 km Puffer | 10 km Puffer |
|---------------|----------------------|-------------|-------------|-------------|--------------|
| PLZ1 | 10 | 2 | 5 | 10 | 10 |
| PLZ2 | 98 | 28 | 71 | 85 | 93 |
| PLZ3 | 709 | 215 | 446 | 628 | 667 |
| PLZ5 | 7425 | 2460 | 4865 | 7137 | 7412 |

Tab. 3.9 Anzahl OSM-Centroide innerhalb eines Puffers um die TeleAtlas-Centroide

Wie bereits erwähnt zeigt sich hier der geringe Abstand der Centroide voneinander. So befinden sich innerhalb des 2 km-Puffers rund 65 % der *OSM-PLZ5*-Centroide, beim 5 km-Puffer sind es sogar rund 96 %. Ähnliche Werte ergab auch der Vergleich der anderen PLZ-Kategorien. Die Mittelpunkte der erzeugten PLZ-Gebiete stimmen somit relativ gut überein.

Die detaillierten Tabellen zur Berechnung der Flächen und Abstände sind der CD im Anhang zu entnehmen. Aufgrund der Vielzahl der PLZ-Gebiete wurden diese nicht mit abgedruckt.

Neben der Berechnung der Abstände und der Anwendung der Puffer-Vergleichsmethode, welche beide zur Bestimmung der Positionsgenauigkeit dienen, wurde noch die Euklidische Distanz berechnet. Ein Ausschnitt dieses Ergebnisses zeigt Abbildung 3.3. Hier wurde die Euklidische Distanz der *OSM-PLZ5*-Centroide in km dargestellt. Dabei stellt man fest, dass diese Distanz im Süden Deutschland sehr gering ist, wogegen in Brandenburg größere Distanzen auftreten. Das ist darauf zurückzuführen, dass die Punktdichte (Abb. 3.4) im Süden beachtlich höher ist als in Brandenburg. Die Unterschiede entstehen hierbei nicht durch fehlerhafte bzw. fehlende PLZ-Angaben, sondern durch die niedrigere Anzahl der PLZ-5-Gebiete in Brandenburg im Vergleich zum Süden von Deutschland. Man kann auch die Distanz bzw. die Punktdichte von *OSM* mit derer von *TeleAtlas* anhand dieser Darstellungen vergleichen. Dabei sei hier auf die im Anhang befindlichen Darstellungen verwiesen.

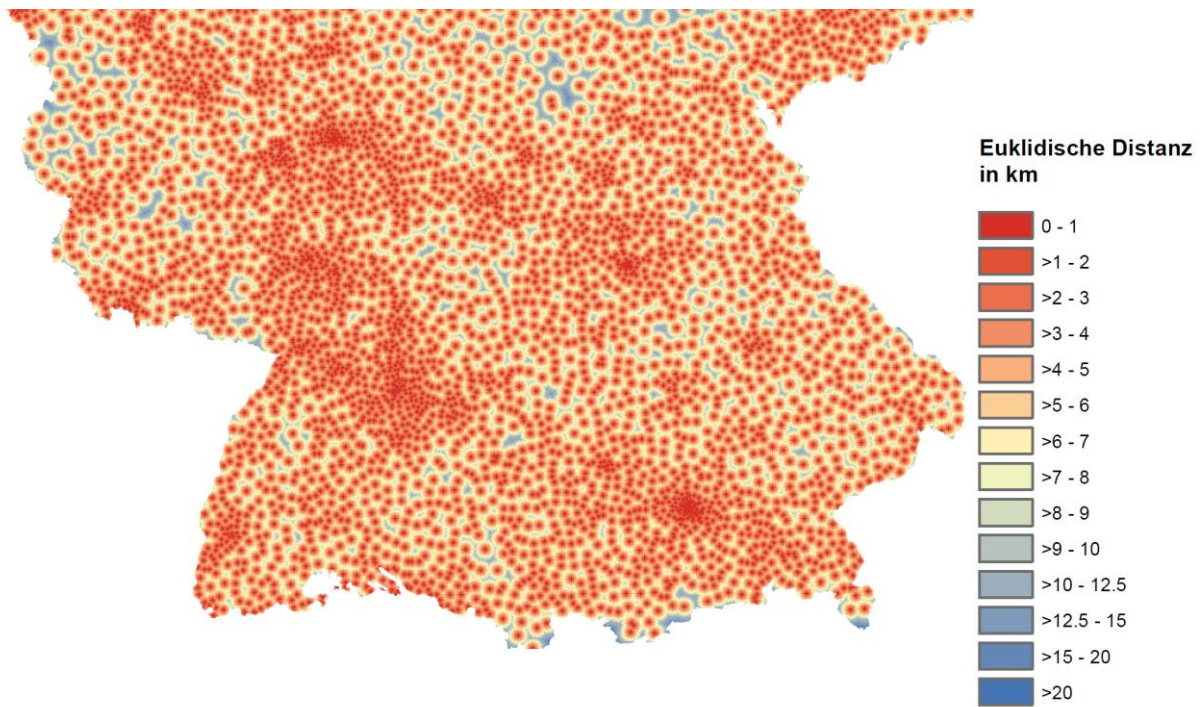


Abb. 3.3 Ausschnitt Euklidische Distanz PLZ5 OSM

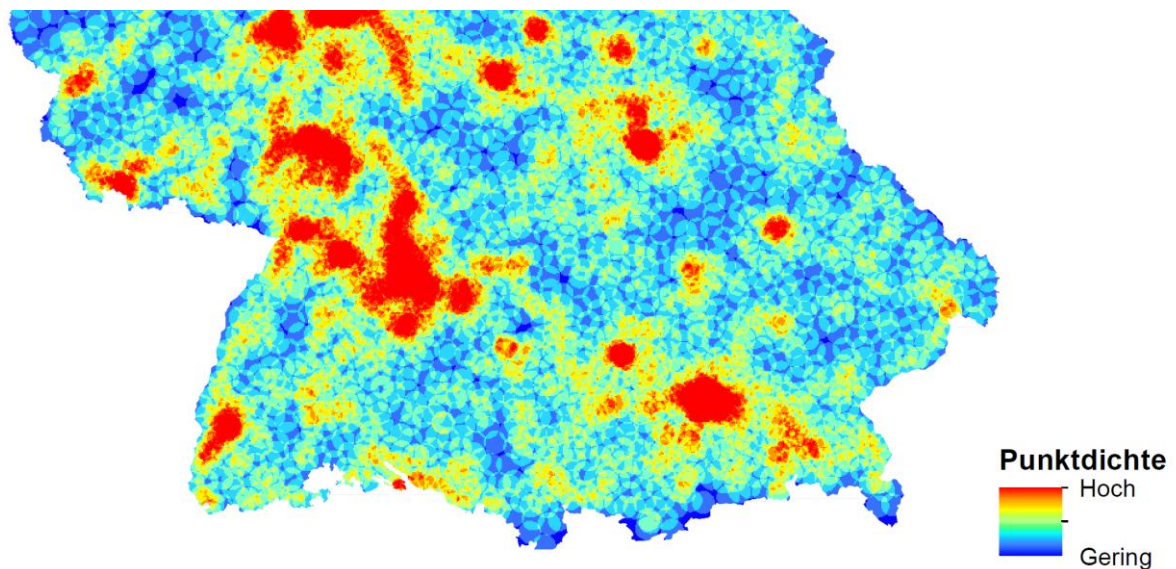


Abb. 3.4 Ausschnitt Punktdichte PLZ5 OSM

Ein weiteres Vergleichskriterium ist der Unterschied zwischen Stadt und Land, sowie die Genauigkeit der Daten im Gebiet einer *OSM*-Hochburg wie Hamburg. Tabelle 3.10 zeigt den Vergleich der PLZ5-Daten für Hamburg. Hier werden die fehlerhaften Daten von *OSM* und die statistischen Werte des gesamten Gebietes gezeigt.

| PLZ5 OSM | Fläche OSM in km ² | in beiden Datensätzen vorhanden? | Bemerkung |
|--|-------------------------------|---|------------------------------------|
| 20246 | 0,40 | nein | Nicht vorhanden/keine deutsche PLZ |
| 20350 | 0,01 | nein | Nicht vorhanden/keine deutsche PLZ |
| 20531 | 0,06 | nein | Nicht vorhanden/keine deutsche PLZ |
| 20882 | 6,71 | nein | Nicht vorhanden/keine deutsche PLZ |
| 22057 | 0,09 | nein | Nicht vorhanden/keine deutsche PLZ |
| 22088 | 0,02 | nein | Nicht vorhanden/keine deutsche PLZ |
| 22097 | 0,01 | nein | Nicht vorhanden/keine deutsche PLZ |
| 22099 | 0,04 | nein | Nicht vorhanden/keine deutsche PLZ |
| 22291 | 0,43 | nein | Nicht vorhanden/keine deutsche PLZ |
| 22407 | 6,11 | nein | Nicht vorhanden/keine deutsche PLZ |
| Gesamtfläche OSM: 850,80 km ² | | Gesamtfläche TA: 796,22 km ² | |
| Anzahl OSM: | 110 | Anzahl TA: | 100 |
| Mittelwert Differenz Fläche in km ² | | | -0,41 |
| mittlere Abweichung vom Mittelwert Differenz Fläche in km ² | | | 2,48 |
| Mittelwert Abstand Centroide in m | | | 4.523,07 |
| mittlere Abweichung vom Mittelwert Abstand Centroide in m | | | 7.827,63 |

Tab. 3.10 Genauigkeit und Vollständigkeit der Daten in Hamburg

Der *TeleAtlas*-Datensatz beinhaltet alle von der *Deutschen Post* für dieses Gebiet vergebenen Postleitzahlen. Auch der *OSM*-Datensatz beinhaltet alle diese Postleitzahlgebiete, allerdings befinden sich hier 10 zusätzliche PLZ-Angaben, die laut *Deutscher Post* (DEUTSCHE POST, 2011b) nicht korrekt bzw. in Deutschland gar nicht vorhanden sind. Hierbei handelt es sich also um fehlerhafte Daten bzw. fehlerhafte Angaben innerhalb des *OSM*-Datensatzes. Bezüglich der Fläche lässt sich im Gebiet Hamburg eine sehr geringe Differenz im Vergleich zu ganz Deutschland feststellen. Dies gilt nur bedingt für die Abstände der Centroide. Der Grund hierfür ist die falsche Lage der *OSM*-Centroide. Zum Beispiel befindet sich der Centroid zur PLZ 22457 bei *OSM* in Hessen, da hier eine falsche PLZ-Angabe gemacht wurde und somit eine Exklave entstanden ist, die bei der Erzeugung der Centroide das Ergebnis enorm verfälscht hat. Betrachtet man lediglich die korrekt erzeugten Centroide, so ist der Abstand dieser in Hamburg im Mittel rund 500 m, was ein sehr gutes Ergebnis ist. Auch die Vollständigkeit der *OSM*-Daten ist für diesen Bereich ist sehr gut. Man kann also feststellen, dass die *OSM*-Daten in einer Hochburg wie Hamburg relativ genau sind und somit auch ein guter *PLZ*-Datensatz erzeugt werden kann. Allerdings muss man hier sagen, dass der Grenzverlauf der *OSM*-Gebiete nicht denen von *TeleAtlas* entspricht. Abbildung 3.5 zeigt den Ausschnitt von Hamburg mit den überlagerten *PLZ*-Gebieten. Hier sieht man, dass der Grenzverlauf der aus *OSM*-Daten erzeugten Gebiete zwar grob mit denen von *TeleAtlas* übereinstimmt, aber es

doch diverse Differenzen gibt. So lassen sich die PLZ-Gebiete zwar zueinander zuordnen, würden aber bei der Betrachtung von Straßenverläufen etc. unterschiedliche PLZ-Angaben machen. Dieser nicht genügend detaillierte Grenzverlauf ist im gesamten Gebiet Deutschlands zu finden und erstreckt sich auch über alle PLZ-Kategorien. Dies ist darauf zurückzuführen, dass die *TeleAtlas*-Daten an Straßenverläufe angepasst wurden und somit auch mehr Vertices besitzen, während die *OSM*-Daten lediglich mit der Bildung von Voronoi-Polygonen erzeugt wurden.

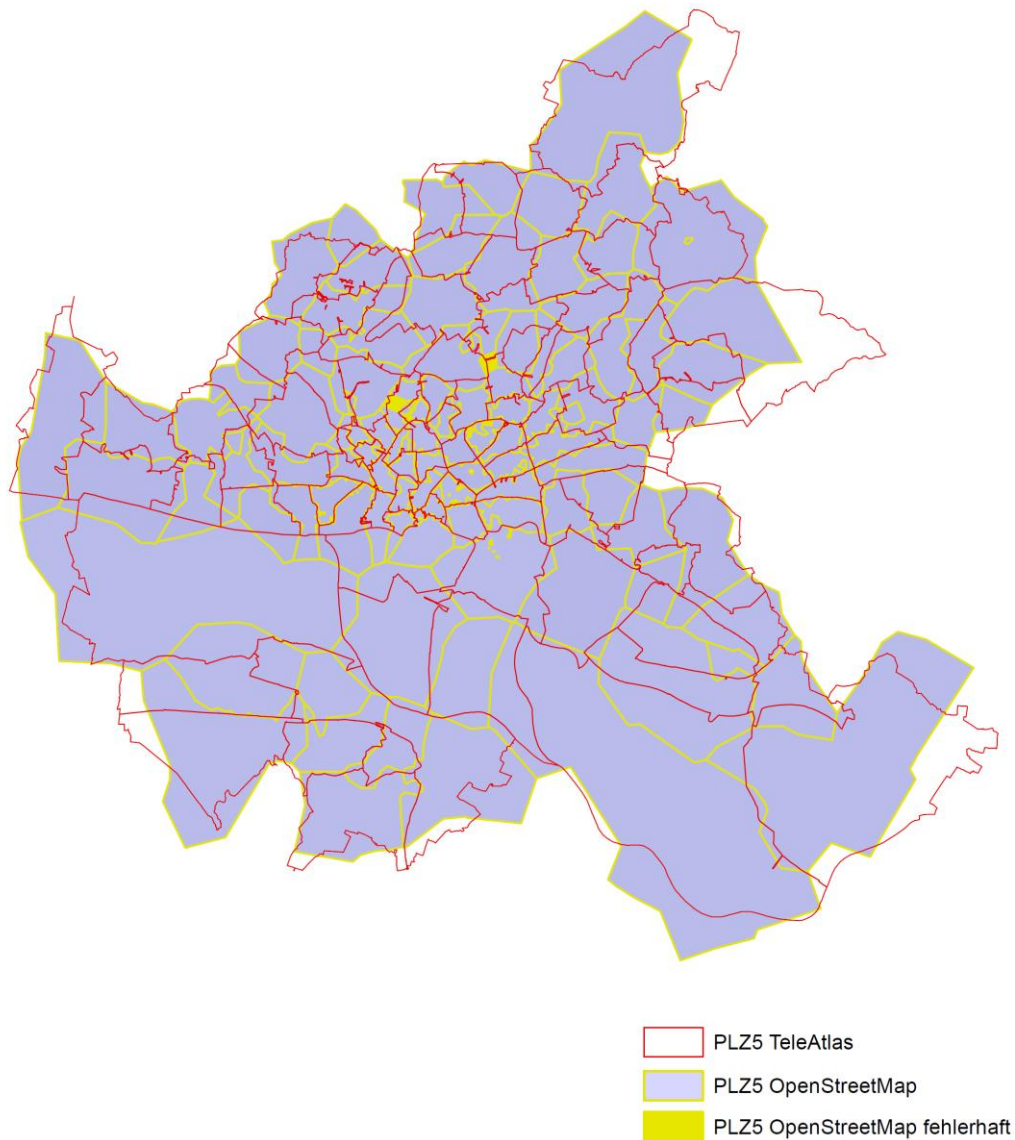


Abb. 3.5 Ausschnitt Hamburg

Die unterschiedlichen Gebietsverläufe seien in den folgenden Abbildungen (Abb. 3.6– 3.9) nochmals dargestellt. Durch diese Abbildungen kann man einen visuellen Vergleich durchführen. Der Vergleich zeigt ebenfalls die ungenauen Grenzverläufe der *OSM*-Daten. Allerdings sieht man hier sehr gut, dass sich die erzeugten PLZ-Gebiete einfach zu denen von *Te-*

leAtlas zuordnen lassen. Diese Tatsache spricht für die erzeugten *OSM*-Daten, da die Abgrenzungen denen von *TeleAtlas* sehr ähnlich sind.

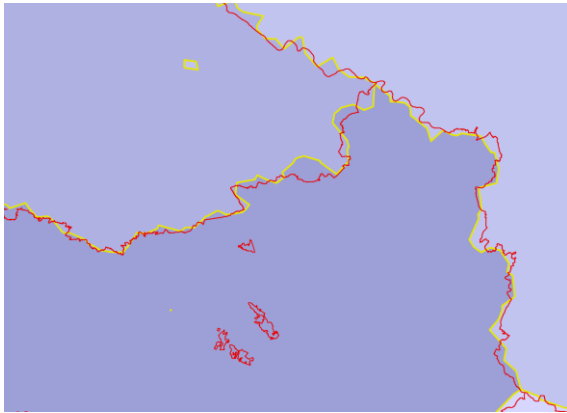


Abb. 3.6 Ausschnitt PLZ1

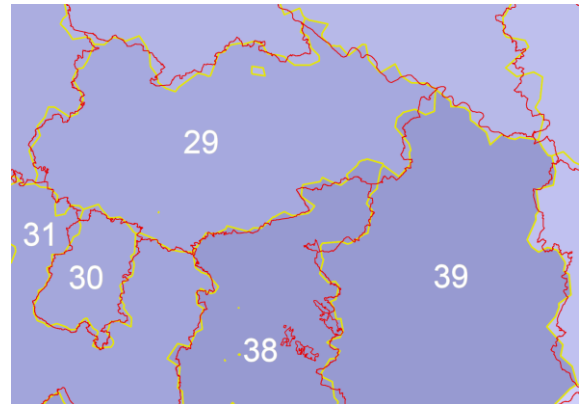


Abb. 3.7 Ausschnitt PLZ2

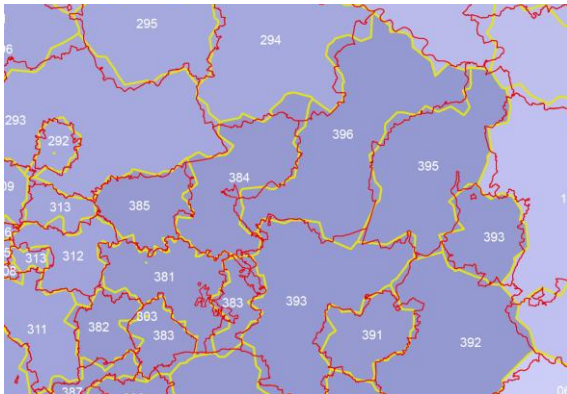


Abb. 3.8 Ausschnitt PLZ3

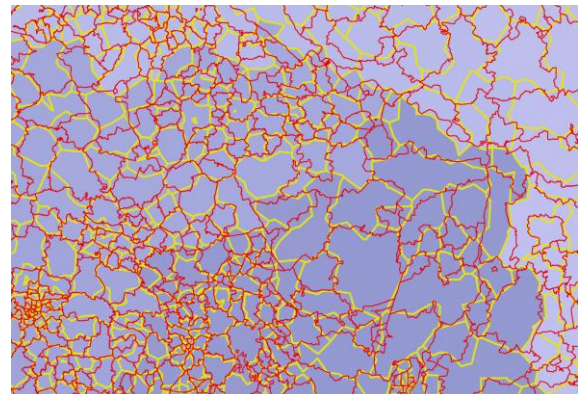


Abb. 3.9 Ausschnitt PLZ5

Zum Vergleich zwischen einem Gebiet mit höherer und einem niedrigerer Bevölkerungsdichte sei gesagt, dass die Genauigkeit in einer Stadt höher ist als im ländlichen Bereich. Diese Aussage kann man am Beispiel der PLZ5-Gebiete der Stadt Dresden und deren Umland zeigen. Innerhalb des Stadtgebietes zeigt sich eine genauere Abgrenzung der *OSM*-PLZ-Gebiete im Vergleich mit denen von *TeleAtlas* als in der ländlichen Gegend. Dies ist aber auf die unterschiedlich vorhandenen Adresspunkte zurückzuführen und nicht auf mangelnde Daten innerhalb *OSM*. Hier sei erwähnt, dass innerhalb einer Stadt von Grund auf mehr Adresspunkte vorhanden sind, da es hier einfach mehr Adressen gibt. In einer ländlichen Gegend gibt es aufgrund der geringeren Anzahl von Häusern und Straßen auch weniger Adressen. Ein direkter Vergleich der Unterschiede zwischen Stadt und Land ist somit für den Bereich der Postleitzahlen eigentlich gar nicht möglich.

Zusammenfassend zu den bisher durchgeführten Vergleichen kann man sagen, dass die erzeugten PLZ-Gebiete nicht direkt maßstabsabhängig sind. Die Grenzverläufe der PLZ1- bis PLZ3-Gebiete sind aus den PLZ5-Gebieten entstanden und somit nicht genauer oder ungenauer als die der PLZ5-Gebiete. Zu den Flächen der Gebiete sei gesagt, dass die Abweichung in Bezug auf die Gesamtfläche eines PLZ-Gebietes bei den PLZ5-Gebieten größer ist. Die Abweichungen nehmen also mit größer werdendem Maßstab zu. Die Lage der Centroide ist bei allen Kategorien verhältnismäßig gleich einzuschätzen und zeigt eine sehr gute Näherung an die Lage der Centroide der *TeleAtlas*-Daten.

Würde man nur diese Aspekte betrachten, so liefert der erzeugte Datensatz eine gute Alternative zum proprietären Datensatz. Allerdings ist die Anzahl der fehlerhaften und fehlenden PLZ-Angaben innerhalb *OSM* noch gravierend hoch, sodass die Verwendung der PLZ5-Gebiete für ganz Deutschland nicht sinnvoll erscheint. Einen Teil der fehlerhaften PLZ5 zeigt Tabelle 3.11.

| PLZ5 OSM | in beiden Datensätzen vorhanden? | PLZ5 TA | fehlend oder fehlerhaft? |
|----------|----------------------------------|---------|--------------------------|
| 80008 | nein | | fehlerhaft in OSM |
| 80327 | nein | | fehlerhaft in OSM |
| 80355 | nein | | fehlerhaft in OSM |
| 80525 | nein | | fehlerhaft in OSM |
| 80635 | nein | | fehlerhaft in OSM |
| 80643 | nein | | fehlerhaft in OSM |
| 80793 | nein | | fehlerhaft in OSM |
| 80835 | nein | | fehlerhaft in OSM |
| 81289 | nein | | fehlerhaft in OSM |
| 81337 | nein | | fehlerhaft in OSM |
| 81593 | nein | | fehlerhaft in OSM |
| 81732 | nein | | fehlerhaft in OSM |
| | nein | 82208 | fehlerhaft in TA |
| | nein | 82276 | fehlend in OSM |
| | nein | 82279 | fehlend in OSM |
| | nein | 82287 | fehlend in OSM |
| | nein | 82290 | fehlend in OSM |
| | nein | 82294 | fehlend in OSM |
| | nein | 82297 | fehlend in OSM |
| 82353 | nein | | fehlerhaft in OSM |
| | nein | 82387 | fehlend in OSM |
| | nein | 82389 | fehlend in OSM |
| | nein | 82392 | fehlend in OSM |
| | nein | 82399 | fehlend in OSM |
| | nein | 82404 | fehlend in OSM |
| | nein | 82409 | fehlend in OSM |
| | nein | 82435 | fehlend in OSM |
| | nein | 82436 | fehlend in OSM |
| | nein | 82438 | fehlend in OSM |

Tab. 3.11 fehlende bzw. fehlerhafte PLZ5

Die meisten fehlenden PLZ-Angaben innerhalb *OSM* im Vergleich zu *TeleAtlas* sind dabei im Bereich der Postleitzonen 8 (270 Stück), 9 (199 Stück) und 7 (198 Stück) zu finden. Diese Postleitzonen erstrecken sich vor allem über Bayern, Baden-Württemberg und Thüringen.

Aus diesen Ergebnissen lässt sich ableiten, dass die Vollständigkeit der Daten im Süden Deutschlands nicht so hoch ist wie im Norden.

Aber auch im Datensatz von *TeleAtlas* lassen sich beim Vergleich über die *Deutsche Post* (DEUTSCHE POST, 2011b) fehlerhafte PLZ-Angaben finden. Allerdings scheint es sich hierbei um alte Daten zu handeln, da die *TeleAtlas*-Daten aus dem Jahr 2010 stammen, die *Deutsche Post* aber die aktuellen Daten bei ihrer PLZ-Suche nutzt.

4 Fazit

Für die automatisierte Generierung von Postleitzahlgebieten aus *OpenStreetMap*-Daten unter Verwendung von Open Source GIS Software wurde eine sinnvolle kostengünstige Lösung gefunden. Es wurden mit überschaubarem Aufwand die unterschiedlichen PLZ-Gebiete erstellt und für eine weitere Benutzung in gängigen Datenformaten abgespeichert. Die dabei genutzten Adressdaten von *OpenStreetMap* stellten sich als gute, aber nicht hundertprozentig korrekte Grundlage für diese Gebietsgenerierung heraus. So sind diese Daten unvollständig, nicht gleichmäßig über das gesamte Gebiet verteilt und teilweise fehlerhaft.

Mit der Nutzung von *GRASS GIS* und *PostgreSQL* mit *PostGIS* als Open Source GIS Software konnte eine Methode gefunden werden, die die PLZ-Gebiete mit relativ geringem Zeitaufwand ermöglicht und keine Softwarekosten erzeugt. Hierbei sei erwähnt, dass die Operationen auf der Datenbank, sprich die Erzeugung neuer Tabellen unter Berücksichtigung der fehlerhaften Attributeingaben der *OSM*-Daten und die Abfrage der Lage der Objekte innerhalb der Deutschlandgrenze, die meiste Zeit in Anspruch nahm. Die Bearbeitung der Daten innerhalb *GRASS GIS* benötigte nur einen Bruchteil dieser Zeit. Dies ist eventuell darauf zurückzuführen, dass die Datenmenge innerhalb der Datenbank beachtlich größer ist als die Datenmenge, welche in *GRASS GIS* benutzt wird.

Die Umsetzung der Aufgabe ohne Bedienung graphischer Benutzeroberflächen war dabei eine gute Wahl, da dadurch eine spätere Automatisierung einfacher möglich ist. Durch die Nutzung von *GRASS GIS* konnte dabei auch ohne Probleme eine Verbindung zur *PostgreSQL*-Datenbank hergestellt werden, was eine alleinige Bearbeitung innerhalb der Textversion von *GRASS GIS* ermöglichte und somit auch weniger Zeit in Anspruch nahm.

Mit Hilfe der Bildung von Voronoi-Polygonen aus den Punktdaten und dem anschließenden Auflösen der Grenzen gleicher Attributwerte konnten somit die PLZ-Gebiete unterschiedlicher Kategorien erzeugt werden. Diese Daten wurden schließlich mit einem Datensatz eines kommerziellen Anbieters verglichen. Der Vergleich zeigt dabei einen Unterschied zwischen den Datensätzen. So gibt es allein bereits bei der Gesamtfläche Differenzen, da es in Deutschland auch Gebiete gibt, welche laut *TeleAtlas* keine PLZ besitzen, diese aber mit der genutzten Methode nicht beachtet werden können. Neben der Fläche findet man auch Differenzen in der Anzahl der PLZ-Gebiete, dem Grenzverlauf und der Position des Flächenmittelpunktes (Centroid). Dabei werden diese Unterschiede deutlicher, je mehr man den Maßstab erhöht. Es ist sichtbar, dass die erzeugten PLZ1-Gebiete am ähnlichsten denen von *TeleAtlas* sind, wogegen die weitere Aufgliederung in PLZ2, PLZ3 oder auch PLZ5 deutlichere Unterschiede

zum Vorschein bringt. Am deutlichsten sieht man diesen Unterschied im Grenzverlauf, dieser ist bei *TeleAtlas* viel detaillierter dargestellt. Ein Vergleich der Daten im Bereich einer *OSM*-Hochburg zeigte, dass diese Daten, aufgrund der größeren Anzahl der Nutzer und der dadurch größeren Anzahl an Daten und deren höhere Genauigkeit, weniger Differenzen aufweisen als andere Daten. Der Vergleich der Daten war somit eine wichtige Aufgabe zur Prüfung der Nutzbarkeit und auch zur Bewertung der Qualität der erzeugten Daten.

Eine Nutzung der Daten im Bereich des Geomarketings zur Darstellung von statistischen Werten auf Grundlage der erzeugten PLZ-Gebiete scheint meiner Meinung nach noch nicht sinnvoll. Hierfür müssten die PLZ-Gebiete genauer erzeugt werden und noch andere Verfahren, wie zum Beispiel die Anpassung an topographische Linienverläufe, angewendet werden. Die als Ergebnis vorliegenden PLZ-Gebiete wurden trotz der Unterschiede zum proprietären Datensatz auf verschiedenen PDF-Karten dargestellt.

Zusammenfassend sei erwähnt, dass diese Methode der Generierung von PLZ-Gebieten auf Grundlage von *OSM*-Daten, trotz der Unterschiede, ein gutes Ergebnis liefert. Man kann diese Methode dabei als Grundlage für weitere Verfeinerungen nutzen. Aufgrund der nicht allzu langen Laufzeit der Generierung, eignet sich diese Methode gut für regelmäßige Aktualisierungen. Das gesetzte Ziel, einen Weg zu finden, diese Gebiete automatisch zu generieren wurde somit erreicht. Die Gebiete wurden erzeugt und stehen für eine weitere Nutzung zur Verfügung. Auch der Vergleich zwischen den Datensätzen wurde vollzogen und somit eine Qualitätseinschätzung der erzeugten Gebiete geliefert.

5 Ausblick

Da mehrere Unterschiede zwischen dem aus *OSM*-Daten erzeugten Datensatz der PLZ-Gebiete und dem proprietären Datensatz von *TeleAtlas* vorhanden sind, soll an dieser Stelle ein Ausblick auf mögliche Verbesserungen bzw. zusätzliche Arbeiten bei der Gebietsgenerierung gegeben werden. Ebenso soll auf eine stärkere Automatisierung der Erstellung dieser Gebiete eingegangen werden.

Ein komplett automatischer Ablauf der Gebietsgenerierung sollte in Zukunft entwickelt werden. Hierfür ist es lediglich notwendig, den genannten Arbeitsablauf, sprich die genannten Funktionsaufrufe, in einem Skript zusammenzufassen. *GRASS GIS* bietet dafür die Möglichkeit der Skriptprogrammierung. Diese ermöglicht es einzelne Rechenschritte automatisiert ablaufen zu lassen. Die Skripte sind dabei als *UNIX-Shell*-Skripte zu schreiben. Es gibt allerdings auch die Möglichkeit der *C*-Programmierung sowie eine *Java*-Schnittstelle (NETELER, 2005). Ein komplett automatisierter Ablauf ist aufgrund der ständigen Erweiterung des Datensatzes von *OSM* wünschenswert. So könnte man nach erfolgter Aktualisierung der *OSM*-Daten innerhalb der *PostgreSQL*-Datenbank die PLZ-Gebiete durch Aufruf des Skriptes neu erzeugen und würde somit immer den aktuellen Stand zur Verfügung haben.

Ein weiterer Punkt, welcher in einer weiteren Arbeit durchgeführt werden könnte, ist die Verbesserung der Gebietsgenerierung. Hierbei könnten die erzeugten Voronoi-Polygone als Grundlage dienen. Diese Polygone müssten dann allerdings im Grenzverlauf geändert werden. So wäre eine Anpassung an Straßenobjekte sinnvoll, da die Postleitzahlen Punkten zugeordnet sind und sich diese Punkte meist rechts und links einer Straße befinden. Bei solch einer Anpassung würde sich der Grenzverlauf dann realer darstellen und somit auch eher den proprietären Daten entsprechen.

Um eventuelle Verbesserungen der Qualität zu gewährleisten wäre es, neben einer Anpassung an den Straßenverlauf, sinnvoll weitere Adressdaten zu nutzen. Dabei wäre zu untersuchen, ob die Nutzung von *OpenAdresses*⁸⁴ als zusätzliche Datenquelle für Punktdaten mit Postleitzahlangebe eine Verbesserung ermöglicht.

Als ein weiterer Vergleich von automatisch erzeugten PLZ-Gebieten zu „originär“ vorhandenen PLZ-Gebieten bietet sich in Zukunft auch die Nutzung der in *OSM* eingespielten PLZ-Gebiete an. Diese wurden bereits angesprochen und sind auf Basis von Daten eines kommerziellen Anbieters erzeugt. Ein Vergleich dieser, neu in *OSM* verfügbaren PLZ-Gebiete mit

⁸⁴ <http://openaddresses.org/>

den aus Punktdaten erzeugten PLZ-Gebieten wäre ein interessantes Thema, zumal diese „originären“ Daten nicht einfach nur übernommen wurden, sondern nachgearbeitet wurden. Neben diesem Vergleich ist sicher auch der Vergleich mit dem Datensatz der PLZ-Gebiete der Deutschen Post eine sinnvolle zukünftige Aufgabe.

Ein etwas kühner, aber sehr interessanter Punkt wäre der Versuch die mittlerweile im Geomarketing genutzten PLZ8-Gebiete aufgrund von *OSM*-Daten und weiteren statistischen Daten zu erzeugen. Hierfür sind aber vorerst genauerer PLZ5-Gebiete notwendig. PLZ8-Gebiete dienen heutzutage als kleinräumige Gebietseinteilung für das Vertriebsmanagement. Dabei sind in einem PLZ8-Gebiet jeweils ca. 500 Haushalte zusammengefasst wurden. Dies ergibt dann eine Unterteilung der PLZ5 Gebiete in noch detailliertere Gebiete mit einem bekannten Wert an Haushalten. PLZ8-Gebiete besitzen dabei achtstellige Nummern, das heißt sie werden im Gegensatz zu den PLZ5-Gebieten mit drei zusätzlichen Ziffern codiert (DDS, 2011).

Quellenverzeichnis

Literaturquellen

- ANDERSON, G. und R. MORENO-SANCHEZ (2003): *Building Web-Based Spatial Information Solutions around Open Specification and Open Source Software*. In: *Transactions in GIS* 7 (4), S. 447–466
- ATHER, A. (2009): *A Quality Analysis of OpenStreetMap Data*. London: University College London, DEPARTMENT of Civil, Environmental & Geomatic Engineering, Dissertation, 2009
- AURENHAMMER, F. (1991): *Voronoi Diagrams - A Survey of a Fundamental Geometric Data Structure*. In: *ACM Computing Surveys*, Vol. 23, Issue 3, Seite 345 – 405, 1991
- BALMER, D. J., B. K. HÖCK, M. O. KIMBERLEY, M. S. WATT, D. J. LOWE und T. W. PAYN (2009): *Comparison of spatial prediction techniques for developing Pinus radiata productivity surfaces across New Zealand*. In: *Forest Ecology and Management*, Vol. 258, Seite 2046-2055, Elsevier, 2009
- CHRISMAN, N. R. und B. S. YANDELL (1988): *Effects of Point Error on Area Calculations: A Statistical Model*, In: *Surveying and Mapping*, Vol. 48, No. 4, Seite 241-246 1988
- DEE, M. (2008): *Einrichten von PostgreSQL/PostGIS als Datenbankserver für GRASS: Einrichten einer SQL-Datenbank als Datenbankserver für ein Geoinformationssystem unter Verwendung freier Software*. Saarbrücken: Vdm Verlag Dr. Müller, 2008
- DEUTSCHE POST (2010): *Daten integrieren, besser planen – mit Datafactory*, Broschüre Datafactory. Bonn: Deutsche Post Direkt GmbH, 2010. URL http://www.deutschepost.de/mlm.nf/dpag/images/d/datafactory/20100205_datafactory_internet.pdf

- DEVILLERS, R. (2010): *Spatial data quality: from process to decisions*. CRC Press, 2010
- DEVILLERS, R. und R. JEANSOULIN (2006): *Spatial Data Quality: Concepts*. In: Rodolphe Devillers (Hg.): *Fundamentals of spatial data quality*. ISTE (Geographical information systems series), S. 31–43, 2006
- DONELLY, F. P. (2010): *Evaluating open source GIS for libraries*. In: *Library Hi Tech* Vol. 28, No. 1, S. 131–151, 2010
- DUBIN, R. (2008): *Kriging*. In: *Encyclopedia of GIS*, 2008
- EHLERS, M., J. SCHIEWE, S. KLONUS und P. ROSSO (2008): *Prüfung von Luftbilddaten zweier unterschiedlicher Aufnahmesensoren hinsichtlich eines optimalen Aufnahmesystems zur Erfüllung von Aufgaben von Vermessungs- und Umweltverwaltung für das Landesamt für Natur und Umwelt*. Osnabrück: Institut für Geoinformatik und Fernerkundung, *gi-reports@igf*, Band 9, 2008
- FORTUNE, S. (1986): *A sweepline algorithm for Voronoi diagrams*. *Proceedings of the second annual symposium on Computational geometry*. New York, Seite 313–322, 1986
- GARTNER, G. und M. SCHMIDT (2010): *Moderne Kartographie - Technologische Entwicklungen und Implikationen*. In: *KN - Kartographische Nachrichten*, 60 (6), Seite 299 – 305, Deutsche Gesellschaft für Kartographie, Kirschbaum Verlag Bonn, 2010
- GD POSTDIENST (1993): *Hintergründe zu den neuen Postleitzahlen - Das geheime Leben der Zahlen* (10), S. 735
- GOODCHILD, M. F. (2007): *Citizens as sensors: the world of volunteered geography*. In: *GeoJournal* 69 (4), S. 211, 221
- GOODCHILD, M. F. und G. J. HUNTER (1997): *A simple positional accuracy measure for linear features*. In: *International Journal of Geographical Information Science* 11 (3), S. 299–306

- GUPTILL, S. C. und INTERNATIONAL CARTOGRAPHIC ASSOCIATION (ICA) (1995): *Elements of spatial data quality*. Elsevier Science, 1995
- HAINING, R. P. (2005): *Spatial data analysis: Theory and practice*. Cambridge: Cambridge Univ. Press, 2005
- HAKLAY, M. (2008): *How good is OpenStreetMap information? A comparative study of OpenStreetMap and Ordnance Survey datasets for London and the rest of England*. Under review in *Environment & Planning B*. In: *Planning and Design*.
- HALL, B. und M. G. LEAHY (2008): *Open Source Approaches in Spatial Data Handling*. *Advances in Geographic Information Science*. Springer, 2008
- HOFFMANN, K. (2011): *Nutzergenerierte Karten und kartographische Kommunikation im Web 2.0*. In: *KN - Kartographische Nachrichten*, 61 (2), Seite 72-78, Deutsche Gesellschaft für Kartographie, Kirschbaum Verlag Bonn, 2011
- ISO (2005a): *ISO 19113 Geoinformation - Qualitätsgrundsätze (ISO 19113:2002); Englische Fassung EN ISO 19113:2005*. Brüssel: Europäisches Komitee für Normung, 2005
- ISO (2005b): *ISO 19114 Geoinformation - Verfahren zur Ermittlung der Datenqualität (ISO 19114:2003); Englische Fassung EN ISO 19114:2005*. Brüssel: Europäisches Komitee für Normung, 2005
- ISO (2006): *ISO 19138 Geoinformation - Qualitätsmaße für Geodaten*. Brüssel: Europäisches Komitee für Normung, 2006
- JOOS, G. (1999): *Zur Qualität von objektstrukturierten Geodaten*. München: Universität der Bundeswehr, Geodäsie und Geoinformation, Dissertation, 1999
- KANG, J.M. (2008): *Voronoi Diagram*. In: *Encyclopedia of GIS*, 2008
- KLEIN, R. (2005): *Algorithmische Geometrie*. Springer, 2005
- LI, J. und A. D. HEAP (2008): *A Review of Spatial Interpolation Methods for Environmental Scientists*. Canberra: Geoscience Australia, Record 2008/23, 2008

- LU, G. Y. und D. W. WONG (2008): *An adaptive inverse-distance weighting spatial interpolation technique*. In: *Computers & Geosciences*, Vol. 34, Issue 9, Seite 1044-1055, Elsevier, 2008
- LUDWIG, I. (2010): *Abbildung von Straßendaten für Qualitätsuntersuchungen. Ein Vergleich von OpenStreetMap mit Navteq*. Bonn: Rheinische Friedrich-Wilhelms Universität Bonn, Geographisches Institut, Diplomarbeit, 2010
- MAUÉ, P. und S. SCHADE (2008): *Quality Of Geographic Information Patchworks*. 11th AGILE International Conference on Geographic Information Science, Girona, Spain, 2008
- MITAS, L. und H. MITASOVA (1999): *Spatial interpolation*. In: P. Longley, M. F. Goodchild, D.J. Maguire, D. W. Rhind (Eds.), *Geographical Information Systems: Principles, Techniques, Management and Applications*, GeoInformation International, Wiley, 2005
- MITCHELL, T. (2005): *Web Mapping Illustrated: Using Open Source GIS Toolkits*. O'Reilly Media, 2005
- MITCHELL, T., A. EMDE und A. CHRISTL (2008): *Web Mapping mit Open Source-GIS-Tools*. Deutsche Übersetzung von J. W. LANG. O'Reilly, 2008
- NETELER, M. und H. MITASOVA (2007): *Open Source GIS: A GRASS GIS Approach*. Third Edition. New York: Springer, 2007
- NEUMANN, A. (2008): *Open Source GIS Software*. In: *Encyclopedia of GIS*, 2008
- OBE, R. O. und L. S. HSU (2011): *PostGIS in Action*. Unter Mitarbeit von Paul Ramsey. Manning Publications, 2011
- OKABE, A., B. BOOTS, K. SUGIHARA und S. NOK CHIU (2000): *Spatial tessellations: concepts and applications of Voronoi diagrams*, Wiley Series in Probability and Statistics. Second Edition. Wiley, 2000
- PAS 1071 (2007): *Qualitätsmodell für die Beschreibung von Geodaten*. Deutscher Dachverband für Geoinformation (DDGI), 2007

- POLCHAU, J. (2009): *OpenStreetMap im Vergleich zu anderen Geodatenanbietern*.
Wedel: FH Wedel, Seminararbeit, 2009
- PRISLEY, S. P., T. G. GREGOIRE und J. L. SMITH (1989): *The Mean and Variance of Area Estimates Computed in an Arc-Node Geographic Information System*. In:
Photogrammetric Engineering and Remote Sensing, Vol. 55, No. 11, Seite 1601-1612, American Society for Photogrammetry and Remote Sensing, 1989
- RAMM, F. und J. TOPF (2010): *OpenStreetMap. Die freie Weltkarte nutzen und mitgestalten*. 3. überarb. und erw. Auflage. Berlin: Lehmanns Media, 2010
- RAMSEY, P. (2004): *The State of Open Source GIS*. Victoria: Refrations Research Inc., 2004
- RITTMEIER, W. (1993): *GD Postdienst legt die Briefregionen fest*. In: *DBZ* (3), S. 158ff.
- SCHOOF, M. (2010): *ATKIS-Basis-DLM und OpenStreetMap – Ein Datenvergleich anhand ausgewählter Gebiete in Niedersachsen*. Osnabrück: Universität Osnabrück, Institut für Geoinformatik und Fernerkundung, Bachelorarbeit, 2010
- SHAHABI, C. und M. SHARIFZADEH (2008): *Voronoi Diagrams for Query Processing*. In: *Encyclopedia of GIS*, 2008
- SHEKHAR, S. und H. XIONG (2008): *Encyclopedia of GIS*. New York: Springer Science + Business Media LLC, 2008
- SHERMAN, G. E. (2008): *Desktop GIS: Mapping the Planet with Open Source Tools*. 1st Edition. Pragmatic Bookshelf, 2008
- SHI, W., P. F. FISHER und M. F. GOODCHILD (2002): *Spatial Data Quality*. London und New York: Taylor and Francis, 2002
- STROH, H. (2010): *OpenStreetMap – Raumzeitliche Verhaltensmuster und potenzielle Entwicklungspfade einer freien Nutzergemeinschaft*. Heidelberg: Universität Heidelberg, Diplomarbeit, 2010

- STRUNCK, A. (2010): *Raumzeitliche Qualitätsuntersuchung von OpenStreetMap*. Bonn: Rheinische Friedrich-Wilhelms Universität Bonn, Geographisches Institut, Diplomarbeit, 2010
- VAN OORT, P. (2005): *Spatial data quality: from description to application*. Wageningen: Universiteit Wageningen, Dissertation, 2005
- WÄLDER, O (2008): *Mathematical Methods for Engineers and Geoscientist*. Berlin: Springer, 2008
- ZIELSTRA, D. (2009): *Datenqualität und Anwendbarkeit von Volunteered Geographic Information. Vergleich von proprietären und frei verfügbaren Geodaten*. Diplomarbeit, 2009
- ZIELSTRA, D. und A. ZIPF (2009): *Datenqualität von OpenStreetMap - Erste Ergebnisse empirischer Untersuchungen*. Poster, AGIT 2009, Salzburg, 2009
- ZIELSTRA, D. und A. ZIPF (2010): *A Comparative Study of Proprietary Geodata and Volunteered Geographic Information for Germany*. AGILE 2010, The 13th AGILE International Conference on Geographic Information Science, Guimarães, 2010

Internetquellen

ARNULF.US (2011): PLZ. Zugriff am 13.07.2011, URL <http://arnulf.us/PLZ>

AURENHAMMER, F. und R. KLEIN (1996): *Voronoi diagrams*. Hagen: FernUniversität, *Praktische Informatik VI*, 1996, *Informatik-Bericht Nummer: 198*. Zugriff am 10.07.2011, URL <http://www.wpi6.fernuni-hagen.de/Publikationen/tr198.pdf>

BLAER, P. (2009): *Robot Path Planning Using Generalized Voronoi Diagrams*. Zugriff am 10.07.2011, URL http://www.cs.columbia.edu/~pblaer/projects/path_planner/

CREATIVE COMMONS (2011): *Namensnennung-Weitergabe unter gleichen Bedingungen 2.0 (CC BY-SA 2.0)*. Zugriff am 13.07.2011, URL <http://creativecommons.org/licenses/by-sa/2.0/de/>

DDS DIGITAL DATA SERVICES GMBH (2011): *PLZ8 DEUTSCHLAND – Kleinräumige Gebietseinteilung für das Vertriebsmanagement*. Zugriff am 14.06.2011 URL <http://ddsgeo.de/produkte/plz8.html>

DEUTSCHE POST (2003): *10 Jahre fünfstellige Postleitzahl*, *Pressemitteilung 2003*. Zugriff am 13.07.2011, URL http://www.dp-dhl.com/de/presse/pressemitteilungen/2003/10_jahre_fuenfstellige_postleitzahl.html

DEUTSCHE POST (2011): *Hauptseite*. Zugriff am 30.06.2011, URL <http://www.deutschepost.de/>

DEUTSCHE POST (2011a): *Die Geschichte der PLZ*. Zugriff am 13.07.2011, URL http://www.deutschepost.de/dpag?skin=hi&check=yes&lang=de_DE&tab=1&xmIFile=link1016092_1007923

DEUTSCHE POST (2011b): *Postleitzahlensuche*. Zugriff am 13.07.2011, URL <http://www.postdirekt.de/plzserver/>

FREE THE POSTCODE (2011): *Hauptseite*. Zugriff am 13.07.2011, URL <http://www.freethepostcode.org/>

- GAVRILOVA, M. (1997): *Weighted Voronoi Diagrams in Biology*. Zugriff am 10.07.2011, URL <http://pages.cpsc.ucalgary.ca/~marina/vpplants/>
- HAKLAY, M. (2009): *OpenStreetMap and Ordnance Survey Meridian 2 - Progress maps*. 2009. Zugriff am 30.01.2011, URL <http://povesham.wordpress.com/2009/11/14/openstreetmap-and-ordnance-survey-meridian-2-progress-maps/>
- HEBELER, F. (2004): *Räumliche Interpolation & Hydrologische Funktionen in ArcGIS9, GIS 2 Übung 4 & 5, WS 2004/2005*. Zürich: University of Zurich, Department of Geography, Vorlesungsfolien, 2004, Zugriff am 10.07.2011, URL www.geo.unizh.ch/~fhebeler/interpolation_gis2.pdf
- HOFFMANN, K. (2010): *Webmapping 2.0 mit freier Software und freien Daten. Freie und Open Source Software für Geoinformationssysteme, FOSSGIS 2010 – 22, Osnabrück, Präsentation, 2010*, URL http://www.fossgis.de/konferenz/2010/attachments/75_hoffmann_fossgis2010.pdf
- MOONEY, P., C. PADRAIG und A. WINSTANTLEY (2010): *A study of data representation of natural features in OpenStreetMap*. Zürich: Sixth International conference on Geographic Information Science, GIScience, Paper, 2010 URL http://www.giscience2010.org/pdfs/paper_150.pdf
- NEIS, P., D. ZIELSTRA, A. ZIPF und A. STRUNCK (2010): *Empirische Untersuchungen zur Datenqualität von OpenStreetMap - Erfahrungen aus zwei Jahren Betrieb mehrerer OSM-Online-Dienste*. 2010. Zugriff am 30.06.2011, URL <http://www.geog.uni-heidelberg.de/giscience.html>
- NETELER, M. (2005): *Hinweise zur Programmierung in GRASS*. 2005. Zugriff am 08.06.2011, URL http://www.gdf-hannover.de/lit_html/grasshandbuch_v12/node146.html
- ODGAARD, A. und B. K. NIELSEN (2000): *A visual implementation of Fortune's Voronoi algorithm*. Zugriff am 10.07.2011, URL <http://www.diku.dk/hjemmesider/studerende/duff/Fortune/>

- ORDNANCE SURVEY (2009): OS Meridian 2 Datensatz, 2009. Zugriff am 13.07.2011,
URL <http://www.ordnancesurvey.co.uk/oswebsite/products/meridian2/>*
- ORDNANCE SURVEY (2011): Code-Point Open. Zugriff am 13.07.2011,
URL <http://www.ordnancesurvey.co.uk/oswebsite/products/code-point-open/>*
- OSGEO (2011): Open Source. Zugriff am 10.07.2011,
URL http://www.osgeo.org/content/faq/foundation_faq.html#Open_Source*
- OSM-WIKI (2011): Key: postal_code. Zugriff am 13.07.2011,
URL http://wiki.openstreetmap.org/wiki/Key:postal_code*
- OSM-WIKI (2011a): Free The Postcode. Zugriff am 13.07.2011,
URL http://wiki.openstreetmap.org/wiki/Free_The_Postcode*
- OSM-WIKI (2011b): Qualitätssicherung. Zugriff am 13.07.2011,
URL http://wiki.openstreetmap.org/wiki/DE:Quality_Assurance*
- PC-HEAVEN (2011): Weitere PHP-Funktionen, Abstandberechnung am WGS84 Ellipsoid. Zugriff am 12.07.2011, URL <http://www.pc-heaven.de/index.php?page=Attachment&attachmentID=2853>*
- RANDOMJUNK (2011): Postcode Map, Postcodes. Zugriff am 13.07.2011,
URL <http://random.dev.openstreetmap.org/postcodes/>*
- RASKOB, C. (2004): Schwerpunkt-Voronoi-Diagramme. Zugriff am 13.07.2011, URL <http://www.raskob.de/fun/d/?PHPSESSID=ff4fefde5baf839e402f97fc7b296405#d>*
- TAGINFO OPENSTREETMAP (2011): Keys: postal_code. Zugriff am 23.06.2011,
URL http://taginfo.openstreetmap.org/keys/postal_code#values*
- TELEATLAS (2010): Datensatz TeleAtlas MultiNet, deu_pd_dissolved, Shapefile, PLZ5-Gebiete Deutschland, Stand 05.07.2010*
- WIKIPEDIA (2011a): Postleitzahl Deutschland. Zugriff am 30.01.2011,
URL [http://de.wikipedia.org/wiki/Postleitzahl_\(Deutschland\)](http://de.wikipedia.org/wiki/Postleitzahl_(Deutschland))*

WIKIPEDIA (2011b): Postleitzahl. Zugriff am 13.07.2011,

URL <http://de.wikipedia.org/wiki/Postleitzahl>

WIKIPEDIA (2011c): Orthodrome. Zugriff am 12.07.2011,

URL <http://de.wikipedia.org/wiki/Orthodrome>

Anhang

A Anhang Quellcodes

A.1 Quellcode Import Deutschlandgrenze

/* Einladen der Deutschlandgrenze in Datenbank mittels shp2pgsql*/

```
D:\PostgreSQL\9.0\bin\shp2pgsql -s 4326 C:\Daten\Ländergrenzen\germany.shp
public.germany | psql -h localhost -d postgis -U postgres
```

A.2 Quellcode Vorverarbeitung

/*Starten von GRASS GIS in Textvariante*/

```
C:\OSGeo4W\bin\grass.bat -text
```

/*mit Datenbank verbinden*/

```
db.connect driver=pg database="host=localhost, dbname=postgis"
db.login user=postgres pass=postgres
```

/*SQL-File zum Erstellen der Tabellen mit den PLZ-Daten laden und ausführen*/

```
db.execute input=C:\SQL_Abfragen\Tabellen_erstellen.sql
```

/*alte Zeilen aus geometry_columns löschen*/

```
DELETE FROM geometry_columns WHERE f_table_name='post_points';
DELETE FROM geometry_columns WHERE f_table_name='post_roads';
DELETE FROM geometry_columns WHERE f_table_name='plz1';
DELETE FROM geometry_columns WHERE f_table_name='plz2';
DELETE FROM geometry_columns WHERE f_table_name='plz3';
DELETE FROM geometry_columns WHERE f_table_name='plz5';
```

/*alte PLZ-Tabellen löschen, falls vorhanden*/

```
DROP TABLE IF EXISTS plz1;
DROP TABLE IF EXISTS plz2;
DROP TABLE IF EXISTS plz3;
DROP TABLE IF EXISTS plz5;
```

/*Spalten umbenennen*/

```
ALTER TABLE europe_point RENAME COLUMN "addr:postcode" TO addr_postcode;
ALTER TABLE europe_roads RENAME COLUMN "addr:postcode" TO addr_postcode;
```

/*Tabelle aus europe_point mit Primärschlüssel erzeugen und in Geometrietabelle eintragen*/

```
DROP TABLE IF EXISTS post_points;
```

```
CREATE TABLE post_points AS
SELECT
substring(addr_postcode from 1 for 1) AS PLZ1,
substring(addr_postcode from 1 for 2) AS PLZ2,
substring(addr_postcode from 1 for 3) AS PLZ3,
addr_postcode AS PLZ5,
Way as geometry
FROM europe_point, germany
```

```

Where addr_postcode IS NOT NULL AND length(addr_postcode)=5 AND
addr_postcode ~ '[0-9]' AND addr_postcode !~ '[a-zA-Z.]' AND Intersects
(germany.geometry, europe_point.way)
UNION
SELECT
substring(postal_code from 1 for 1) AS PLZ1,
substring(postal_code from 1 for 2) AS PLZ2,
substring(postal_code from 1 for 3) AS PLZ3,
postal_code AS PLZ5,
Way as geometry
FROM europe_point, germany
Where postal_code IS NOT NULL AND length(postal_code)=5 AND postal_code ~
'[0-9]' AND postal_code !~ '[a-zA-Z.]' AND Intersects (germany.geometry,
europe_point.way);

```

```

ALTER TABLE post_points ADD COLUMN id serial;
ALTER TABLE post_points ADD CONSTRAINT id_pkey_post_points PRIMARY KEY
(id);

```

```

INSERT INTO geometry_columns (f_table_catalog, f_table_schema, f_table_name,
f_geometry_column, coord_dimension, srid, type)
VALUES ('', 'public', 'post_points', 'geometry', 2, 900913, 'POINT');

```

/*Tabelle aus europe_roads mit Primärschlüssel erzeugen und in Geometrietabelle eintragen*/

```

DROP TABLE IF EXISTS post_roads;

```

```

CREATE TABLE post_roads AS
SELECT
substring(addr_postcode from 1 for 1) AS PLZ1,
substring(addr_postcode from 1 for 2) AS PLZ2,
substring(addr_postcode from 1 for 3) AS PLZ3,
addr_postcode AS PLZ5,
Way as geometry
FROM europe_roads, germany
Where addr_postcode IS NOT NULL AND length(addr_postcode)=5 AND
addr_postcode ~ '[0-9]' AND addr_postcode !~ '[a-zA-Z.]' AND Intersects (
germany.geometry, europe_roads.way)
UNION
SELECT
substring(postal_code from 1 for 1) AS PLZ1,
substring(postal_code from 1 for 2) AS PLZ2,
substring(postal_code from 1 for 3) AS PLZ3,
postal_code AS PLZ5,
Way as geometry
FROM europe_roads, germany
Where postal_code IS NOT NULL AND length(postal_code)=5 AND postal_code ~
'[0-9]' AND postal_code !~ '[a-zA-Z.]' AND Intersects ( germany.geometry,
europe_roads.way);

```

```

ALTER TABLE post_roads ADD COLUMN id serial;

```



```
ALTER TABLE post_roads ADD CONSTRAINT id_pkey_post_roads PRIMARY KEY (id);
```

```
INSERT INTO geometry_columns (f_table_catalog, f_table_schema, f_table_name, f_geometry_column, coord_dimension, srid, type) VALUES ('', 'public', 'post_roads', 'geometry', 2, 900913, 'LINESTRING');
```

/*Vacuum und Analyze durchführen*/

```
VACUUM;  
ANALYZE;
```

A.3 Quellcode Erzeugung Polygone

/*Deutschlandgrenze aus Datenbank in GRASS GIS laden und neue Location erstellen*/

```
v.in.ogr dsn="PG:host=localhost dbname=postgis user=postgres password=postgres"  
output=germany_gg layer=germany location=germany
```

/*zur Location germany wechseln*/

```
g.gisenv set="LOCATION=germany"  
g.gisenv set="LOCATION_NAME=germany"
```

/*Region auf Grenze Deutschland setzen*/

```
g.region vect=germany_gg@PERMANENT
```

/*post_points und post_roads aus Datenbank in GRASS GIS laden*/

```
v.in.ogr dsn="PG:host=localhost dbname=postgis user=postgres password=postgres"  
output=post_points_gg layer=post_points  
v.in.ogr dsn="PG:host=localhost dbname=postgis user=postgres password=postgres"  
output=post_roads_gg layer=post_roads
```

/*Umwandeln Linien post_roads in Punkte (nur Verbindungspunkte)*/

```
v.to.points -n input=post_roads_gg@PERMANENT output=post_roads_points
```

/*alle Punkte in eine Datei schreiben*/

```
v.patch -e input=post_roads_points@PERMANENT,post_points_gg@PERMANENT  
output=post_all
```

/*Voronoigebiete erstellen*/

```
v.voronoi input=post_points_gg@PERMANENT output=post_voronoi
```

/*Auflösen der Grenzen der Voronoigebiete pro PLZ-Kategorie*/

```
v.dissolve input=post_voronoi@PERMANENT output=dissolve_plz1 column=plz1  
v.dissolve input=post_voronoi@PERMANENT output=dissolve_plz2 column=plz2  
v.dissolve input=post_voronoi@PERMANENT output=dissolve_plz3 column=plz3  
v.dissolve input=post_voronoi@PERMANENT output=dissolve_plz5 column=plz5
```

/*Verschneiden mit Deutschlandgrenze pro PLZ-Kategorie*/

```
v.overlay ainput=dissolve_plz1@PERMANENT binput=admin_germany_gg@PERMANENT  
output=intersect_plz1 operator=and  
v.overlay ainput=dissolve_plz2@PERMANENT binput=admin_germany_gg@PERMANENT  
output=intersect_plz2 operator=and
```

```
v.overlay ainput=dissolve_plz3@PERMANENT binput=admin_germany_gg@PERMANENT
output=intersect_plz3 operator=and
v.overlay ainput=dissolve_plz5@PERMANENT binput=admin_germany_gg@PERMANENT
output=intersect_plz5 operator=and
```

A.4 Quellcode Nachbearbeitung

/*Export der PLZ-Gebiete in die Datenbank*/

```
v.out.ogr input=intersect_plz1@PERMANENT type=area dsn="PG:host=localhost
dbname=postgis user=postgres password=postgres" olayer=plz1 format=PostgreSQL
v.out.ogr input=intersect_plz2@PERMANENT type=area dsn="PG:host=localhost
dbname=postgis user=postgres password=postgres" olayer=plz2 format=PostgreSQL
v.out.ogr input=intersect_plz3@PERMANENT type=area dsn="PG:host=localhost
dbname=postgis user=postgres password=postgres" olayer=plz3 format=PostgreSQL
v.out.ogr input=intersect_plz5@PERMANENT type=area dsn="PG:host=localhost
dbname=postgis user=postgres password=postgres" olayer=plz5 format=PostgreSQL
```

/*SQL-File zum Löschen der unnötigen Spalten laden und ausführen*/

```
db.execute
input=C:\Users\Christian\Studium\Studienarbeit\SQL_Abfragen\Spalten_loeschen.sql
```

/*unnötige Spalten löschen und Umbenennung*/

```
ALTER TABLE plz1 DROP COLUMN cat;
ALTER TABLE plz1 DROP COLUMN a_cat;
ALTER TABLE plz1 DROP COLUMN b_cat;
ALTER TABLE plz1 DROP COLUMN b_name;
ALTER TABLE plz1 RENAME COLUMN a_plz1 TO plz1;
ALTER TABLE plz2 DROP COLUMN cat;
ALTER TABLE plz2 DROP COLUMN a_cat;
ALTER TABLE plz2 DROP COLUMN b_cat;
ALTER TABLE plz2 DROP COLUMN b_name;
ALTER TABLE plz2 RENAME COLUMN a_plz2 TO plz2;
ALTER TABLE plz3 DROP COLUMN cat;
ALTER TABLE plz3 DROP COLUMN a_cat;
ALTER TABLE plz3 DROP COLUMN b_cat;
ALTER TABLE plz3 DROP COLUMN b_name;
ALTER TABLE plz3 RENAME COLUMN a_plz3 TO plz3;
ALTER TABLE plz5 DROP COLUMN cat;
ALTER TABLE plz5 DROP COLUMN a_cat;
ALTER TABLE plz5 DROP COLUMN b_cat;
ALTER TABLE plz5 DROP COLUMN b_name;
ALTER TABLE plz5 RENAME COLUMN a_plz5 TO plz5;
```

/*erstellte Tabellen löschen*/

```
DROP TABLE IF EXISTS post_points;
DROP TABLE IF EXISTS post_roads;
```

/*alte Vektorkarten aus GRASS GIS löschen*/

```
g.remove
vect=germany_gg@PERMANENT,post_points_gg@PERMANENT,post_roads_gg@PERMANENT,
post_roads_points@PERMANENT,
```

post_all@PERMANENT,post_voronoi@PERMANENT,dissolve_plz1@PERMANENT,dissolve_plz2@PERMANENT,dissolve_plz3@PERMANENT,dissolve_plz5@PERMANENT,intersect_plz1@PERMANENT,intersect_plz2@PERMANENT,intersect_plz3@PERMANENT,intersect_plz4@PERMANENT

A.5 Quellcode Export

/*Shapefiles exportieren mittels pgsq2shp*/

```
D:\PostgreSQL\9.0\bin\pgsq2shp -f
```

```
C:\Users\Christian\Studium\Studienarbeit\Daten\PLZ1.shp -h localhost -u postgres -P  
postgres postgis public.plz1
```

```
D:\PostgreSQL\9.0\bin\pgsq2shp -f
```

```
C:\Users\Christian\Studium\Studienarbeit\Daten\PLZ2.shp -h localhost -u postgres -P  
postgres postgis public.plz1
```

```
D:\PostgreSQL\9.0\bin\pgsq2shp -f
```

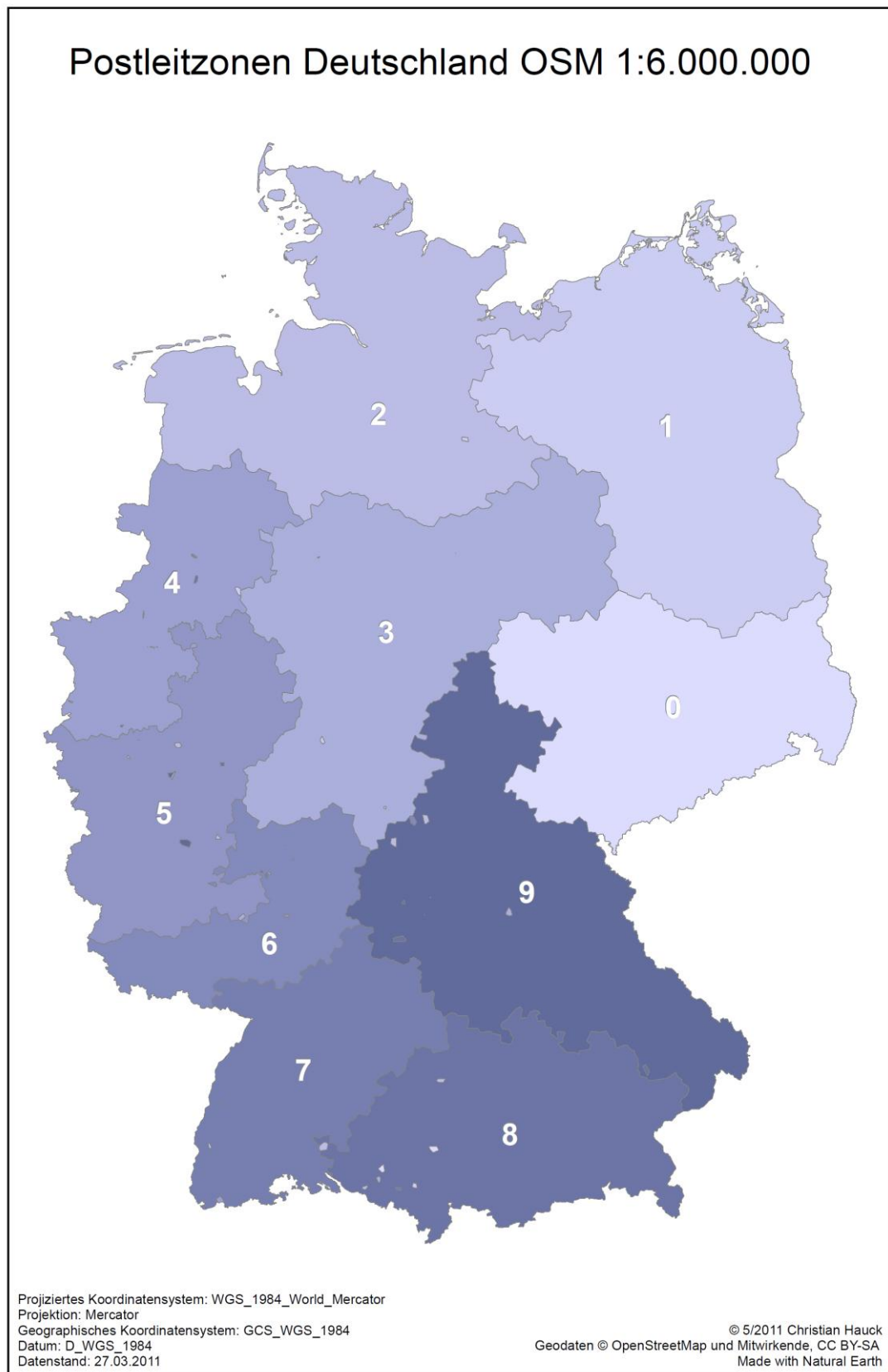
```
C:\Users\Christian\Studium\Studienarbeit\Daten\PLZ3.shp -h localhost -u postgres -P  
postgres postgis public.plz1
```

```
D:\PostgreSQL\9.0\bin\pgsq2shp -f
```

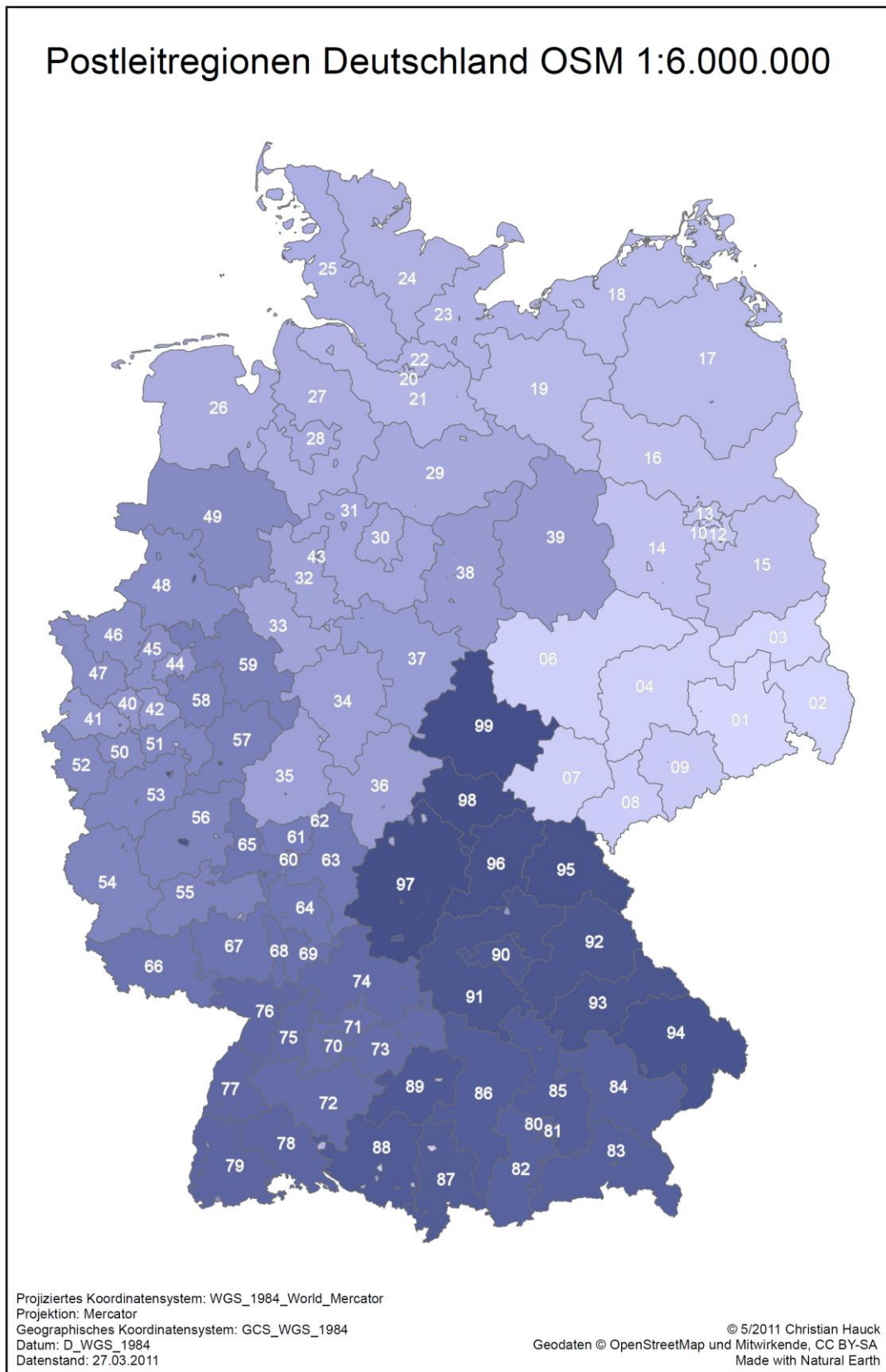
```
C:\Users\Christian\Studium\Studienarbeit\Daten\PLZ5.shp -h localhost -u postgres -P  
postgres postgis public.plz1
```


B Anhang Screenshots PDF-Karten

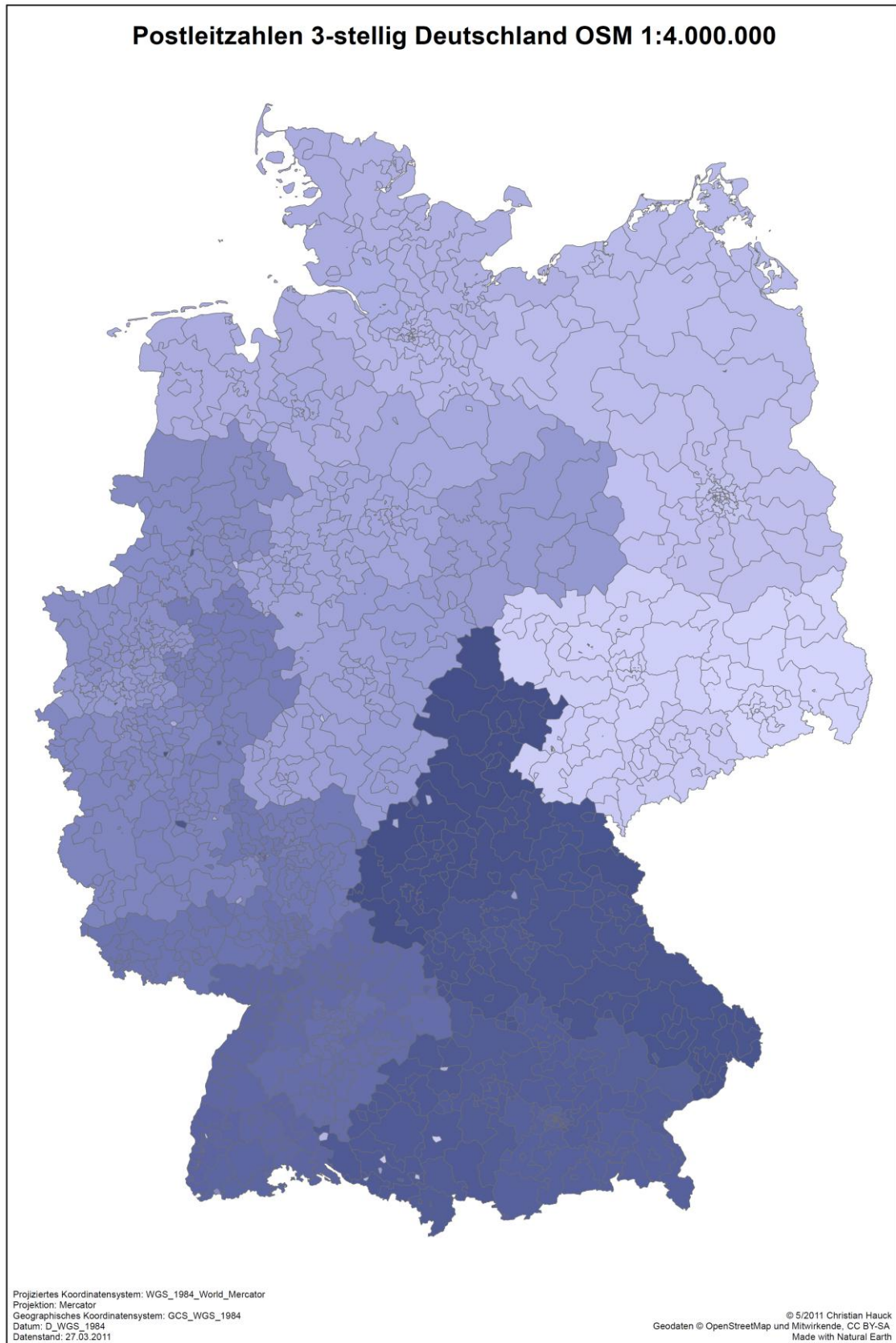
B.1 Postleitzonen (PLZ1) Deutschland OSM



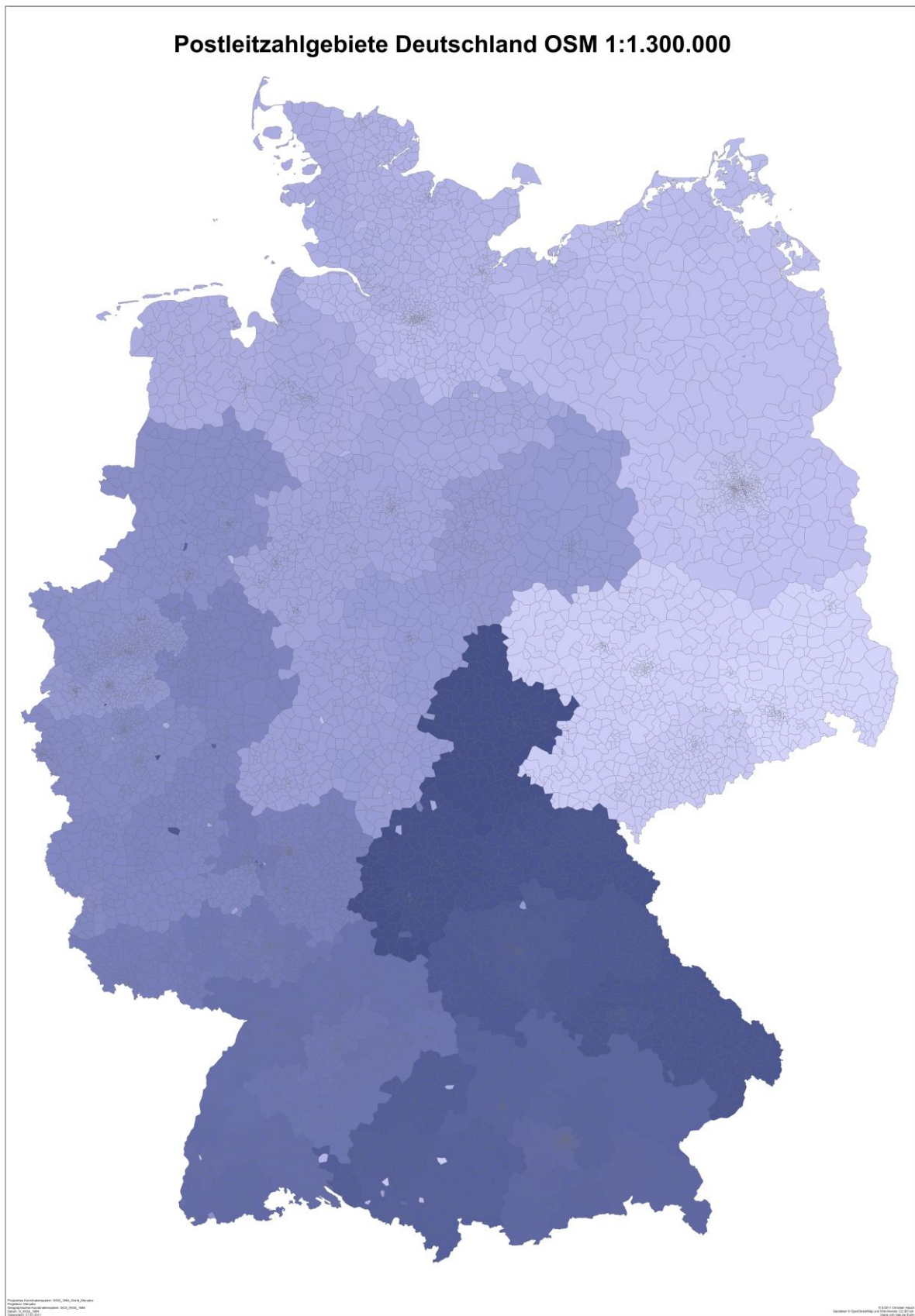
B.2 Postleitregionen (PLZ2) Deutschland OSM



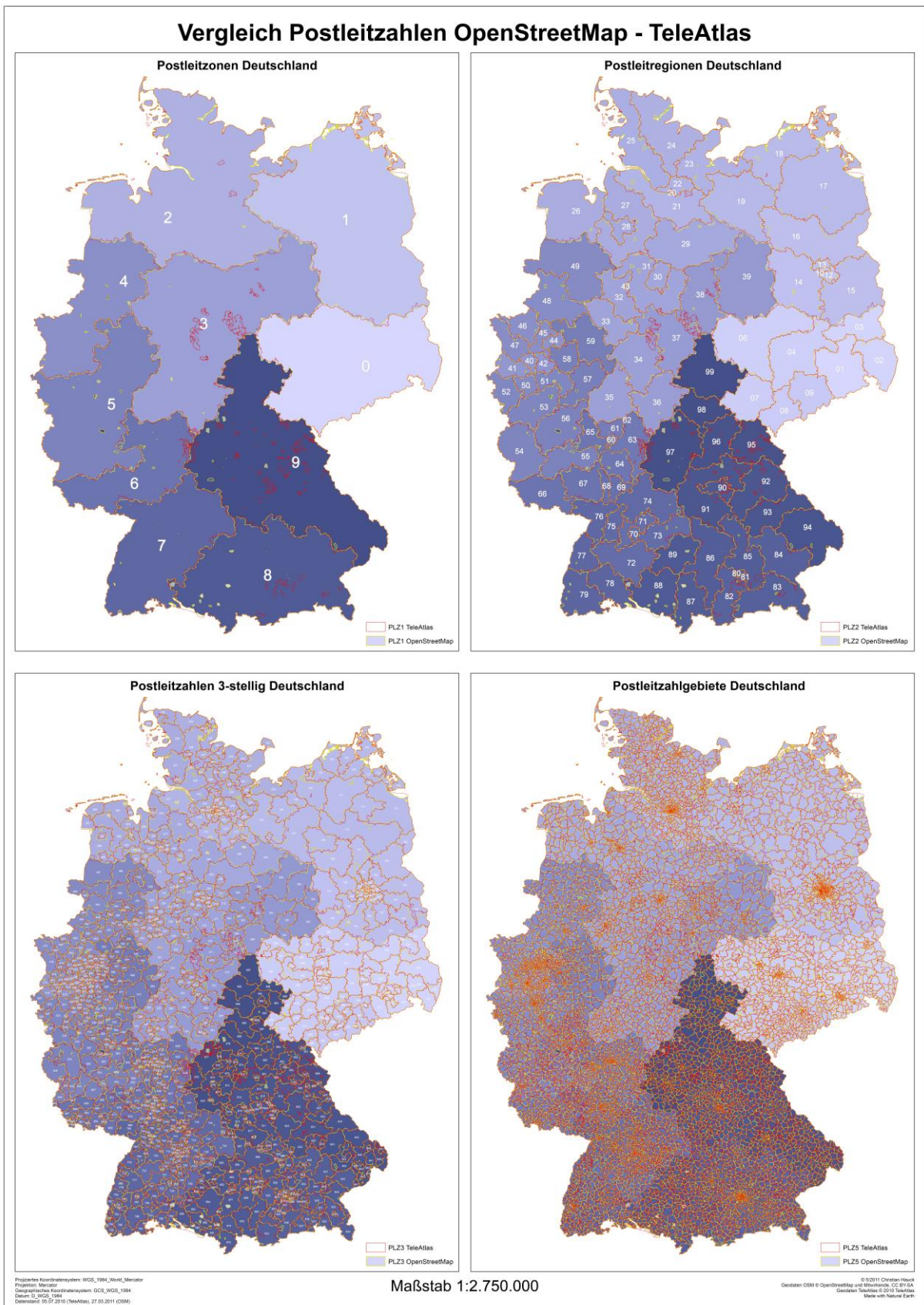
B.3 Postleitzahlen 3-stellig Deutschland OSM



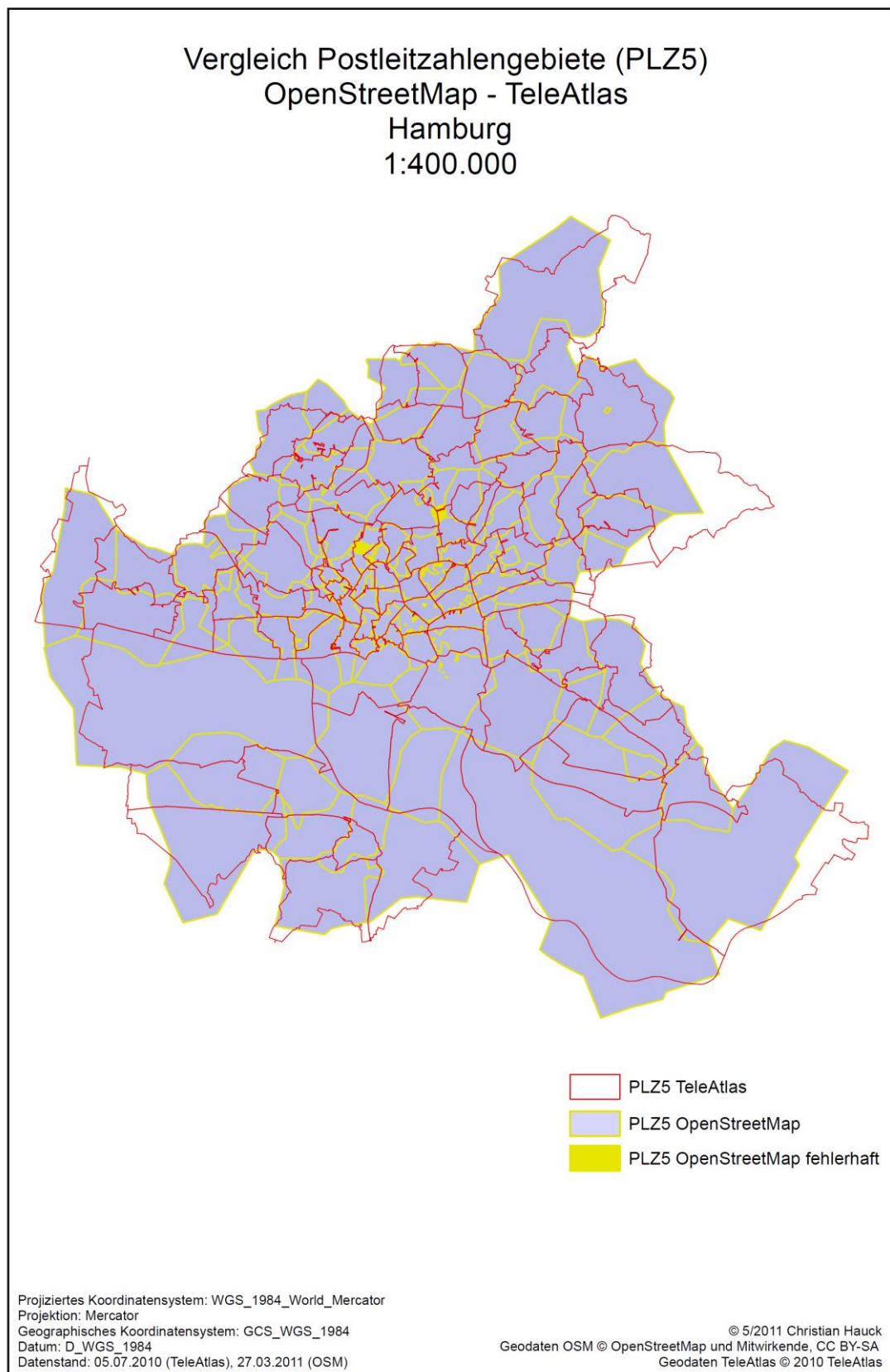
B.4 Postleitzahlgebiete (PLZ5) Deutschland OSM



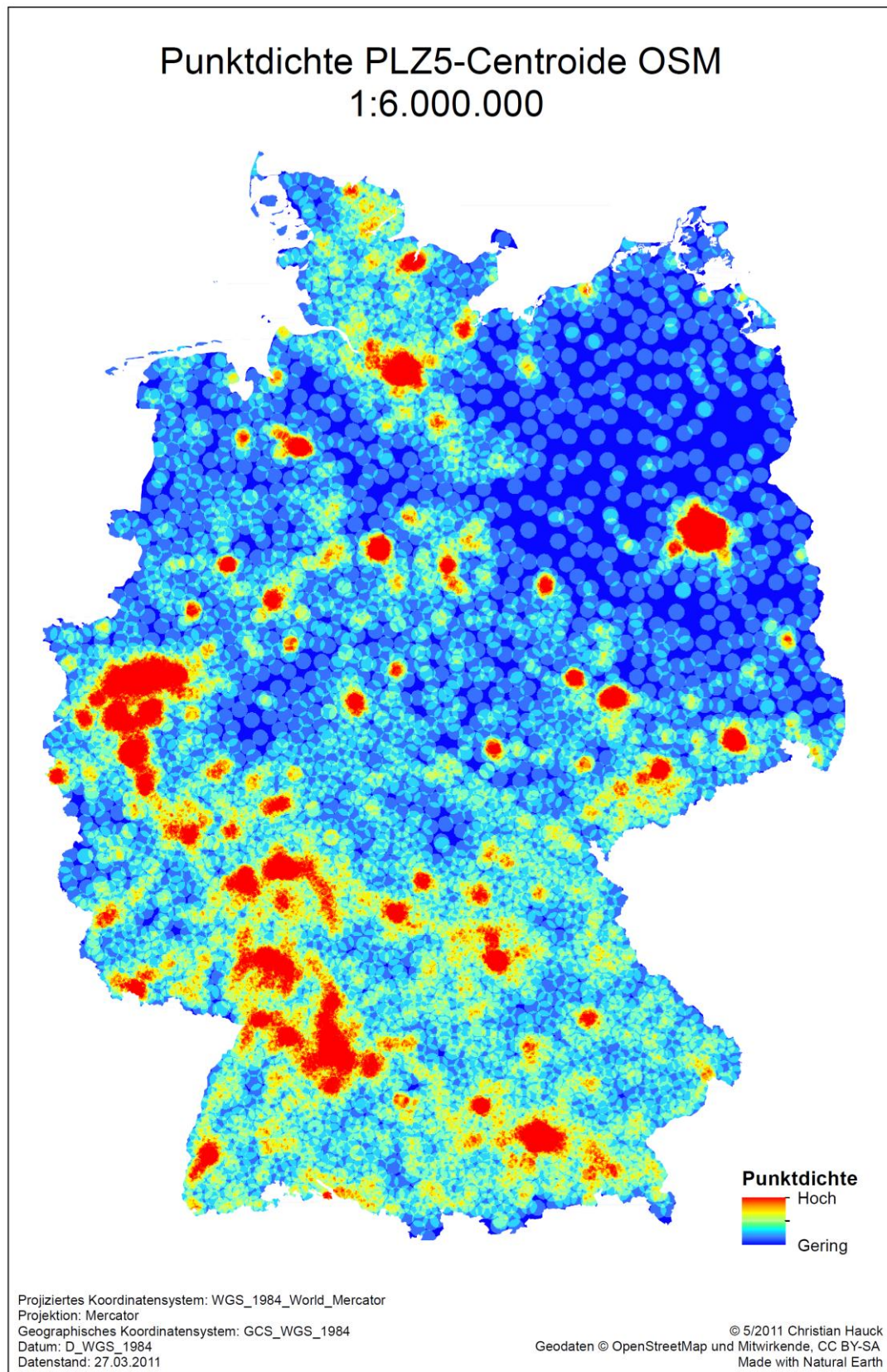
B.5 Vergleich der Postleitzahlen



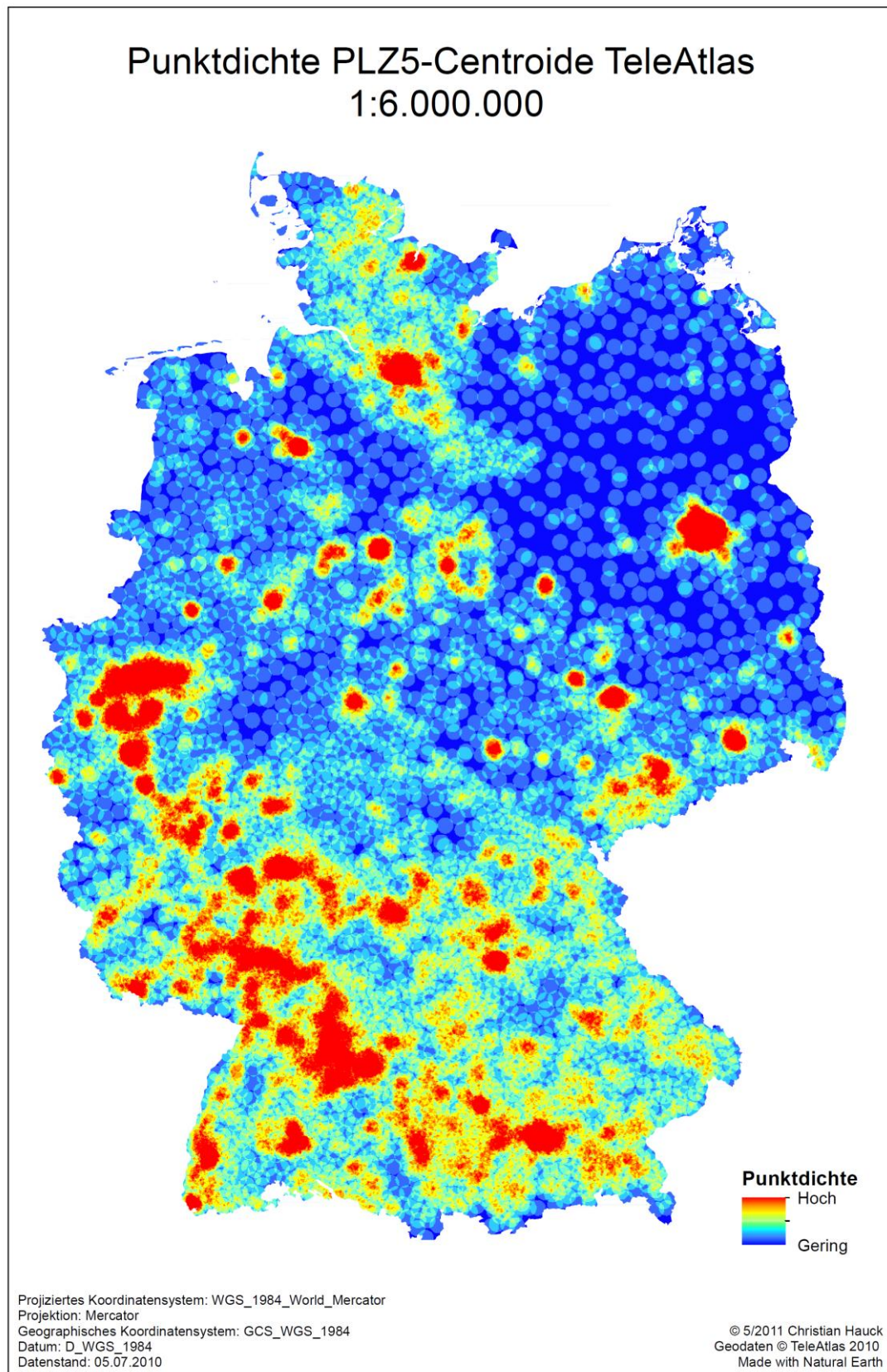
B.6 Vergleich PLZ5 Hamburg



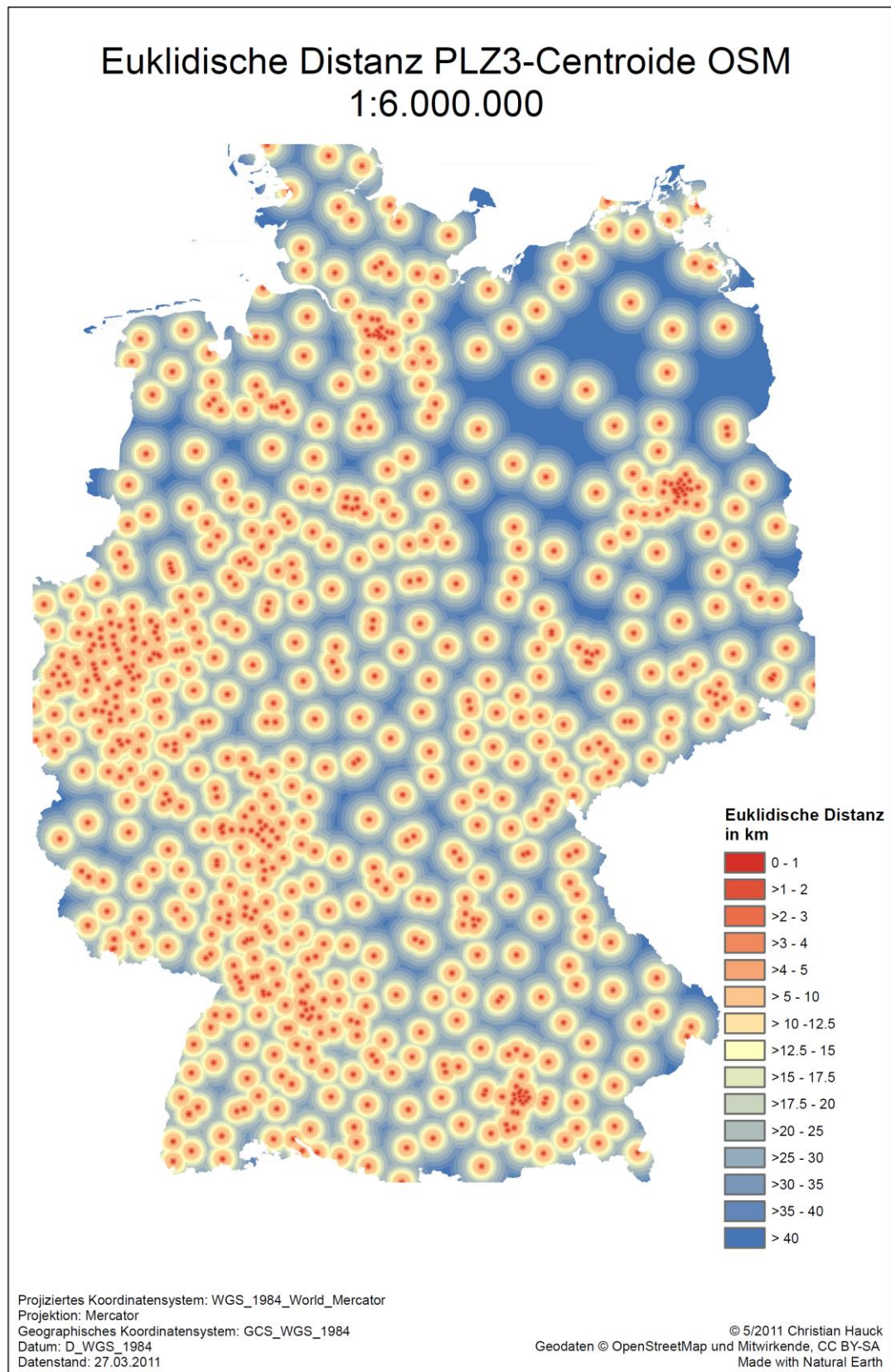
B.7 Punktdichte PLZ5-Centroide OpenStreetMap



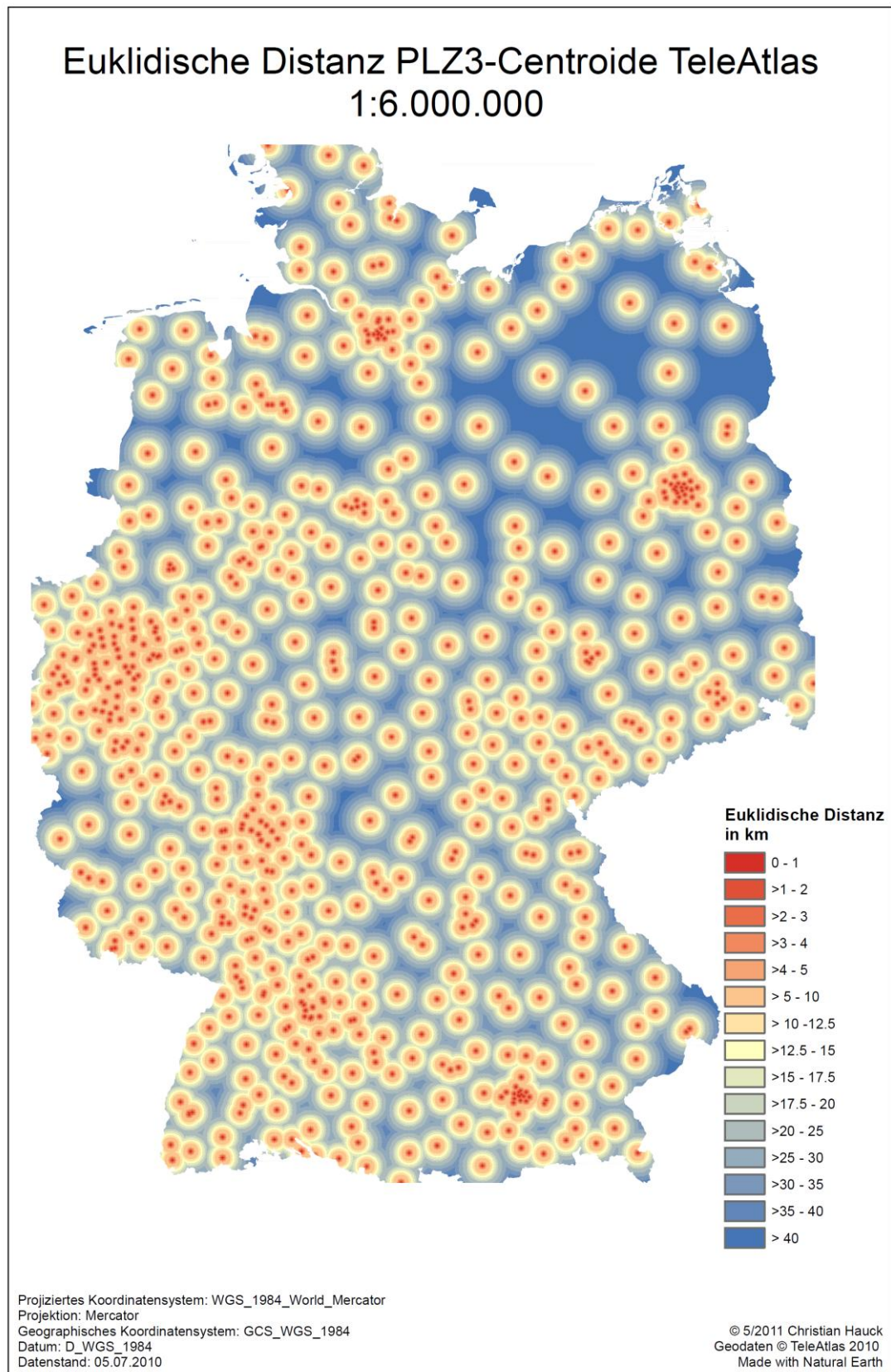
B.8 Punktdichte PLZ5-Centroide TeleAtlas



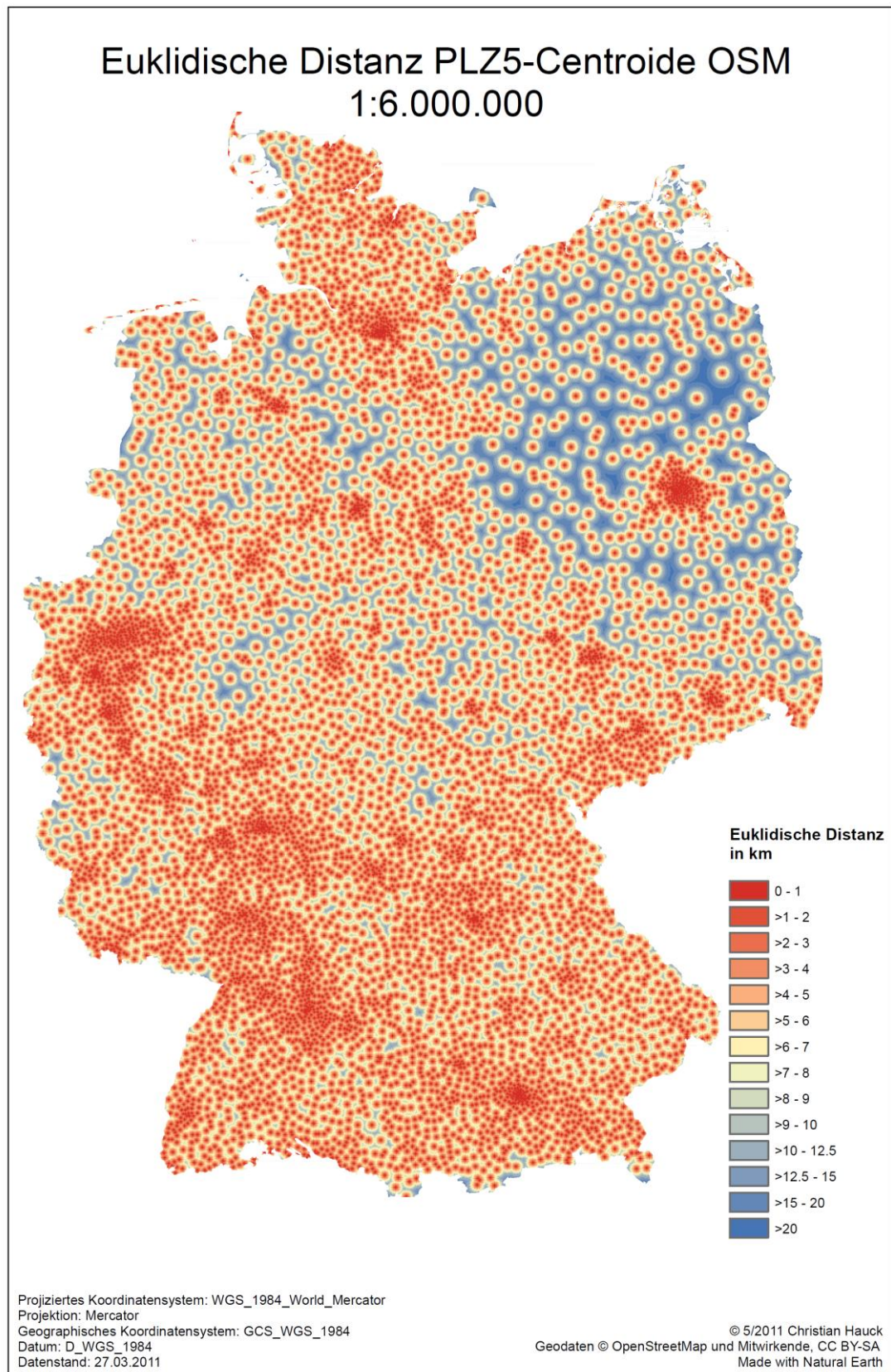
B.9 Euklidische Distanz PLZ3-Centroide OpenStreetMap



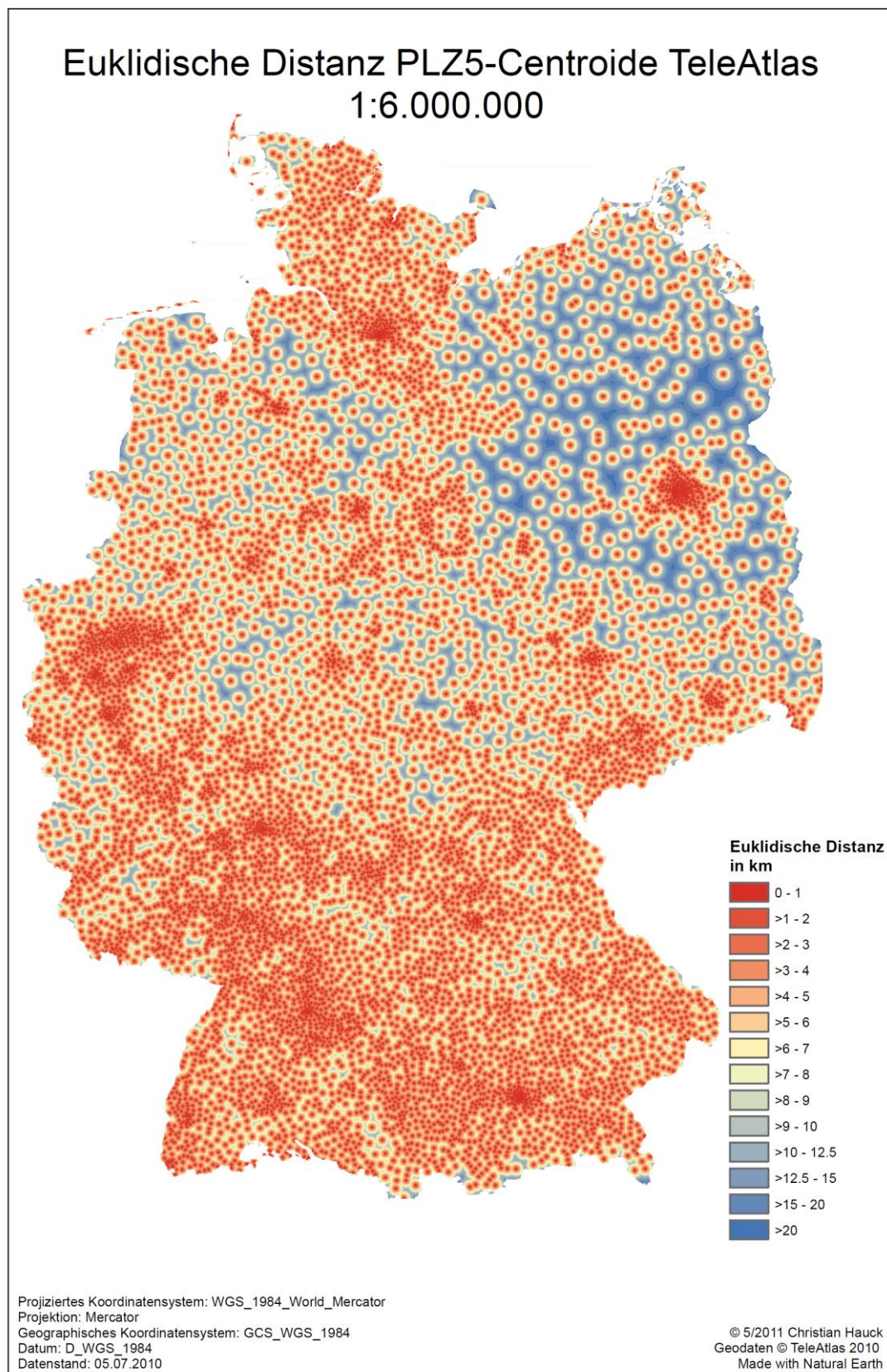
B.10 Euklidische Distanz PLZ3-Centroide TeleAtlas



B.11 Euklidische Distanz PLZ5-Centroide OpenStreetMap



B.12 Euklidische Distanz PLZ5-Centroide TeleAtlas



C Anhang Tabelle

C.1 Tabelle Vergleich PLZ5-Gebiete Hamburg

| PLZ5 OSM | Fläche OSM in km ² | in beiden Datensätzen vorhanden? | PLZ5 TA | Fläche TA in km ² | Differenz Fläche (TA-OSM) in km ² | Abstand Centroide in m | PLZ5 Hamburg (Deutsche Post) | |
|----------|-------------------------------|----------------------------------|------------------------------------|------------------------------|--|------------------------|------------------------------|--|
| 20095 | 1,05 | ja | 20095 | 0,73 | -0,32 | 86,88 | 20095 | |
| 20097 | 2,32 | ja | 20097 | 1,99 | -0,33 | 110,39 | 20097 | |
| 20099 | 1,26 | ja | 20099 | 1,60 | 0,34 | 80,30 | 20099 | |
| 20144 | 0,73 | ja | 20144 | 0,96 | 0,23 | 52,42 | 20144 | |
| 20146 | 1,29 | ja | 20146 | 0,96 | -0,33 | 61,62 | 20146 | |
| 20148 | 1,27 | ja | 20148 | 1,06 | -0,21 | 116,71 | 20148 | |
| 20149 | 1,64 | ja | 20149 | 1,24 | -0,40 | 135,23 | 20149 | |
| 20246 | 0,40 | nein | Nicht vorhanden/keine deutsche PLZ | | | | | |
| 20249 | 0,76 | ja | 20249 | 0,98 | 0,22 | 71,13 | 20249 | |
| 20251 | 1,62 | ja | 20251 | 2,01 | 0,39 | 78,31 | 20251 | |
| 20253 | 1,44 | ja | 20253 | 0,92 | -0,52 | 282,14 | 20253 | |
| 20255 | 0,72 | ja | 20255 | 0,72 | 0,00 | 38,09 | 20255 | |
| 20257 | 0,98 | ja | 20257 | 0,71 | -0,27 | 493,83 | 20257 | |
| 20259 | 0,93 | ja | 20259 | 0,87 | -0,06 | 123,26 | 20259 | |
| 20350 | 0,01 | nein | Nicht vorhanden/keine deutsche PLZ | | | | | |
| 20354 | 1,14 | ja | 20354 | 1,33 | 0,20 | 113,96 | 20354 | |
| 20355 | 1,06 | ja | 20355 | 1,17 | 0,11 | 135,23 | 20355 | |
| 20357 | 1,31 | ja | 20357 | 1,18 | -0,13 | 34,48 | 20357 | |
| 20359 | 2,54 | ja | 20359 | 1,28 | -1,25 | 602,41 | 20359 | |
| 20457 | 4,01 | ja | 20457 | 14,36 | 10,35 | 1.106,55 | 20457 | |
| 20459 | 1,26 | ja | 20459 | 0,93 | -0,33 | 396,68 | 20459 | |
| 20531 | 0,06 | nein | Nicht vorhanden/keine deutsche PLZ | | | | | |
| 20535 | 2,61 | ja | 20535 | 2,14 | -0,47 | 28,75 | 20535 | |
| 20537 | 3,08 | ja | 20537 | 3,40 | 0,33 | 160,63 | 20537 | |
| 20539 | 14,00 | ja | 20539 | 11,54 | -2,46 | 555,07 | 20539 | |
| 20882 | 6,71 | nein | Nicht vorhanden/keine deutsche PLZ | | | | | |
| 21029 | 9,47 | ja | 21029 | 6,30 | -3,17 | 346,61 | 21029 | |
| 21031 | 10,05 | ja | 21031 | 5,47 | -4,58 | 72,48 | 21031 | |
| 21033 | 2,12 | ja | 21033 | 13,39 | 11,28 | 2.166,47 | 21033 | |
| 21035 | 26,13 | ja | 21035 | 10,32 | -15,81 | 357,69 | 21035 | |
| 21037 | 79,52 | ja | 21037 | 73,60 | -5,92 | 997,06 | 21037 | |
| 21039 | 50,74 | ja | 21039 | 55,47 | 4,74 | 557,15 | 21039 | |
| 21073 | 6,59 | ja | 21073 | 3,28 | -3,31 | 617,90 | 21073 | |
| 21075 | 7,74 | ja | 21075 | 8,28 | 0,54 | 731,22 | 21075 | |
| 21077 | 14,67 | ja | 21077 | 14,73 | 0,06 | 202,99 | 21077 | |
| 21079 | 25,55 | ja | 21079 | 31,37 | 5,82 | 1.758,71 | 21079 | |
| 21107 | 24,45 | ja | 21107 | 16,07 | -8,38 | 36.873,83 | 21107 | |
| 21109 | 10,60 | ja | 21109 | 19,79 | 9,19 | 509,21 | 21109 | |

| | | | | | | | | |
|-------|-------|------|------------------------------------|-------|--------|-----------|-------|--|
| 21129 | 60,98 | ja | 21129 | 65,83 | 4,85 | 3.699,49 | 21129 | |
| 21147 | 16,82 | ja | 21147 | 15,14 | -1,68 | 1.707,22 | 21147 | |
| 21149 | 27,93 | ja | 21149 | 18,04 | -9,88 | 272,45 | 21149 | |
| 22041 | 3,83 | ja | 22041 | 3,05 | -0,78 | 273,96 | 22041 | |
| 22043 | 4,92 | ja | 22043 | 5,15 | 0,23 | 55,49 | 22043 | |
| 22045 | 3,58 | ja | 22045 | 5,02 | 1,43 | 289,37 | 22045 | |
| 22047 | 4,07 | ja | 22047 | 3,86 | -0,22 | 336,90 | 22047 | |
| 22049 | 2,88 | ja | 22049 | 2,09 | -0,79 | 120,08 | 22049 | |
| 22057 | 0,09 | nein | Nicht vorhanden/keine deutsche PLZ | | | | | |
| 22081 | 1,55 | ja | 22081 | 1,70 | 0,15 | 57,20 | 22081 | |
| 22083 | 1,14 | ja | 22083 | 1,64 | 0,50 | 72,86 | 22083 | |
| 22085 | 1,74 | ja | 22085 | 1,37 | -0,37 | 59,09 | 22085 | |
| 22087 | 1,78 | ja | 22087 | 1,60 | -0,19 | 267,44 | 22087 | |
| 22088 | 0,02 | nein | Nicht vorhanden/keine deutsche PLZ | | | | | |
| 22089 | 1,60 | ja | 22089 | 1,90 | 0,30 | 172,62 | 22089 | |
| 22097 | 0,01 | nein | Nicht vorhanden/keine deutsche PLZ | | | | | |
| 22099 | 0,04 | nein | Nicht vorhanden/keine deutsche PLZ | | | | | |
| 22111 | 4,34 | ja | 22111 | 4,25 | -0,09 | 535,25 | 22111 | |
| 22113 | 29,18 | ja | 22113 | 27,48 | -1,69 | 331,78 | 22113 | |
| 22115 | 4,83 | ja | 22115 | 3,49 | -1,34 | 1.082,18 | 22115 | |
| 22117 | 6,74 | ja | 22117 | 7,31 | 0,58 | 592,81 | 22117 | |
| 22119 | 3,86 | ja | 22119 | 6,13 | 2,27 | 34,04 | 22119 | |
| 22143 | 7,24 | ja | 22143 | 5,70 | -1,54 | 421,31 | 22143 | |
| 22145 | 12,61 | ja | 22145 | 27,86 | 15,26 | 1.580,70 | 22145 | |
| 22147 | 5,35 | ja | 22147 | 3,74 | -1,61 | 88,87 | 22147 | |
| 22149 | 3,04 | ja | 22149 | 6,79 | 3,75 | 810,58 | 22149 | |
| 22159 | 7,18 | ja | 22159 | 8,35 | 1,17 | 370,20 | 22159 | |
| 22175 | 5,70 | ja | 22175 | 4,30 | -1,40 | 278,67 | 22175 | |
| 22177 | 3,06 | ja | 22177 | 1,67 | -1,40 | 692,75 | 22177 | |
| 22179 | 2,86 | ja | 22179 | 3,95 | 1,09 | 182,24 | 22179 | |
| 22291 | 0,43 | nein | Nicht vorhanden/keine deutsche PLZ | | | | | |
| 22297 | 6,02 | ja | 22297 | 4,06 | -1,96 | 165,48 | 22297 | |
| 22299 | 1,62 | ja | 22299 | 1,72 | 0,11 | 158,89 | 22299 | |
| 22301 | 0,94 | ja | 22301 | 1,31 | 0,37 | 289,87 | 22301 | |
| 22303 | 2,33 | ja | 22303 | 2,85 | 0,52 | 248,06 | 22303 | |
| 22305 | 1,66 | ja | 22305 | 2,01 | 0,35 | 119,72 | 22305 | |
| 22307 | 2,13 | ja | 22307 | 2,01 | -0,12 | 285,62 | 22307 | |
| 22309 | 3,87 | ja | 22309 | 2,50 | -1,36 | 323,74 | 22309 | |
| 22335 | 6,65 | ja | 22335 | 8,58 | 1,93 | 305,19 | 22335 | |
| 22337 | 2,11 | ja | 22337 | 6,68 | 4,56 | 1.168,18 | 22337 | |
| 22339 | 9,32 | ja | 22339 | 5,05 | -4,27 | 330,87 | 22339 | |
| 22359 | 19,87 | ja | 22359 | 12,08 | -7,79 | 161,94 | 22359 | |
| 22391 | 5,09 | ja | 22391 | 6,09 | 1,01 | 204,39 | 22391 | |
| 22393 | 4,87 | ja | 22393 | 6,51 | 1,65 | 537,59 | 22393 | |
| 22395 | 32,17 | ja | 22395 | 8,82 | -23,34 | 47.706,01 | 22395 | |
| 22397 | 27,91 | ja | 22397 | 31,87 | 3,96 | 693,98 | 22397 | |
| 22399 | 5,94 | ja | 22399 | 7,17 | 1,23 | 274,09 | 22399 | |

| | | | | | | | |
|--|-------|--------------------------------------|------------------------------------|-------|--------|------------|----------|
| 22407 | 6,11 | nein | Nicht vorhanden/keine deutsche PLZ | | | | |
| 22415 | 5,64 | ja | 22415 | 4,56 | -1,08 | 333,62 | 22415 |
| 22417 | 4,68 | ja | 22417 | 7,36 | 2,67 | 604,52 | 22417 |
| 22419 | 5,70 | ja | 22419 | 4,65 | -1,04 | 135,02 | 22419 |
| 22453 | 5,59 | ja | 22453 | 9,04 | 3,45 | 263,75 | 22453 |
| 22455 | 4,94 | ja | 22455 | 4,50 | -0,44 | 587,97 | 22455 |
| 22457 | 18,53 | ja | 22457 | 7,00 | -11,53 | 320.370,64 | 22457 |
| 22459 | 5,71 | ja | 22459 | 4,45 | -1,26 | 685,16 | 22459 |
| 22523 | 6,21 | ja | 22523 | 6,54 | 0,33 | 384,54 | 22523 |
| 22525 | 8,96 | ja | 22525 | 8,18 | -0,78 | 450,53 | 22525 |
| 22527 | 2,61 | ja | 22527 | 4,46 | 1,85 | 1.254,27 | 22527 |
| 22529 | 5,95 | ja | 22529 | 5,09 | -0,87 | 420,10 | 22529 |
| 22547 | 6,88 | ja | 22547 | 4,34 | -2,55 | 3.743,80 | 22547 |
| 22549 | 6,09 | ja | 22549 | 6,91 | 0,83 | 255,67 | 22549 |
| 22559 | 17,13 | ja | 22559 | 16,65 | -0,48 | 476,30 | 22559 |
| 22587 | 15,06 | ja | 22587 | 8,95 | -6,11 | 557,51 | 22587 |
| 22589 | 7,62 | ja | 22589 | 8,51 | 0,89 | 568,99 | 22589 |
| 22605 | 4,72 | ja | 22605 | 5,15 | 0,43 | 723,48 | 22605 |
| 22607 | 4,77 | ja | 22607 | 4,28 | -0,49 | 674,21 | 22607 |
| 22609 | 4,74 | ja | 22609 | 5,66 | 0,92 | 613,56 | 22609 |
| 22761 | 6,11 | ja | 22761 | 3,85 | -2,26 | 414,44 | 22761 |
| 22763 | 3,79 | ja | 22763 | 2,52 | -1,27 | 378,41 | 22763 |
| 22765 | 2,03 | ja | 22765 | 2,27 | 0,23 | 50,83 | 22765 |
| 22767 | 5,88 | ja | 22767 | 2,87 | -3,02 | 898,52 | 22767 |
| 22769 | 1,64 | ja | 22769 | 1,52 | -0,12 | 46,41 | 22769 |
| Gesamtfläche: 850,80 km ² | | Gesamtfläche: 796,22 km ² | | | | | |
| Anzahl: 110 | | Anzahl: 100 | | | | | |
| Mittelwert Differenz Fläche in km ² | | | | | | | -0,41 |
| mittlere Abweichung vom Mittelwert Differenz Fläche in km ² | | | | | | | 2,48 |
| Mittelwert Abstand Centroide in m | | | | | | | 4.523,07 |
| mittlere Abweichung vom Mittelwert Abstand Centroide in m | | | | | | | 7.827,63 |

D Anhang

D.1 CD

Die CD enthält die originären und bearbeiteten Daten von *TeleAtlas*, die erzeugten Daten von *OpenStreetMap*, die Deutschlandgrenze von *Natural Earth*, die Quellcodes, die Vergleichstabellen, die berechneten Raster, die PDF-Karten und die Arbeit als PDF-Version.