

Daten aus OSM extrahieren und in QGIS weiterverarbeiten

Einführung

Die OSM-Daten enthalten viele umweltrelevante Informationen, die auf den veröffentlichten Webkarten nicht offensichtlich erkennbar sind. Mit Hilfe von QGIS lassen sich diese Informationen auswerten und in neue aussagekräftigen Karten umsetzen.

Der Vortrag erläutert die Fragestellung am Beispiel der Verteilung von Windkraftanlagen in Deutschland. Angefangen vom Import der deutschlandweiten OSM-Daten in eine PostGis-Datenbank, über die Abfrage der WKA-Standorte bis hin zur Präsentation der Verteilung in einer farbigen Flächendichtekarte, die aus einer interpolierten Rasteroberfläche erzeugt worden ist, welche die Anzahl der Windanlagen im Umkreis von 20km darstellt. Dabei kamen Abfrage- und Geoverarbeitungswerkzeuge aus QGIS sowie das GRASS-Modul zur Spline-Interpolation (v.surf.rst) zur Anwendung.

OSM als Datenquelle

Es gibt aktuell keinen einfach zu beschaffenden bundesweiten Datensatz mit den Standorten sämtlicher Windanlagen. In diesem Kapitel wird am Beispiel der Windanlagen exemplarisch dargestellt, wie die Datenbank des Crowdsourcing Projektes *OpenStreetmap* (OSM) zur Gewinnung naturschutzfachlich relevanter Daten genutzt werden kann.¹ Ziele ist es, eine Karte zur Verteilung der Windkraftanlagen in Deutschland zu erstellen, um Regionen zu erkennen, in denen bisher nur eine geringe Dichte von Anlagen vorhanden ist. Diese resultierenden Daten lassen sich dann für bundesweite Auswertungen z.B. im Bezug zum Vogelschutz nutzen.

Der *Bundesverband Windenergie* gibt die bundesweiten Standortdaten nicht als Geodaten heraus, führt aber eine Statistik, in der für Dezember 2013 insgesamt 23.654 Windanlagen verzeichnet waren.² Das ATKIS-Basis-DLM umfasst nach KOLDRACK ET. AL. im Jahr 2012 insgesamt 21.137 Windanlagen. Eine weitere Datenquelle ist das *EEG-Anlagenregister* mit insgesamt 23.013 Anlagen, dessen Daten jedoch nur über eine Adresscodierung und keine exakten geographischen Verortungen verfügen. Die OSM-Datenbank hatte 2013 insgesamt 21.300 Windanlagen erfasst (90 %).

Es mag eine überraschende Idee sein, in der Datenbank von OpenStreetmap nach WKA-Standorten zu suchen. Doch fällt schon beim Betrachten der OSM-Webkarte auf, dass Windanlagen in großer Zahl verzeichnet sind. OSM begann mit dem Ziel, eine frei verfügbare Straßenkarte der Welt auf Grundlage selbst erfasster GPS-Geodaten zu erschaffen, erfasst jedoch inzwischen jede Art von räumlichen Daten und gewinnt in vielen Gebieten den Charakter einer topographischen Karte. Die Karte des Projektes OpenTopoMap zeigt, dass sich eine der TK 25 ähnliche Darstellung mit OSM-Daten umsetzen lässt.³ Flächennutzung und topographische Einzelobjekte werden in einzelnen Regionen so präzise kartiert, dass die Qualität in Genauigkeit und Fehlerfreiheit in manchen Gegenden die neue ATKIS-DLM-TK 25 hinter sich gelassen hat. Dies lässt sich exemplarisch für den Bereich westlich Kassels sehr gut erkennen. Auch wenn die Flächennutzung im landwirtschaftlichen Bereich noch sehr uneinheitlich erfasst wird, ist das topographische Potential von Karten aus OSM deutlich zu erkennen.

1 <http://www.openstreetmap.de/>

2 <http://www.wind-energie.de/infocenter/statistiken/deutschland/windenergieanlagen-deutschland>

3 <http://opentopomap.org>

Datenbeschaffung und Datenimport

Viele Objekte, die in der OSM-Datenbank erfasst werden, sind nicht zwangsläufig in der über das Web veröffentlichten Karte sichtbar. Jedoch lassen sich die öffentlich verfügbaren Rohdaten mit Hilfe üblicher GIS-Software auswerten und visualisieren. Die zu verarbeitenden Datenmengen sind sehr groß und strukturell sperrig. OSM-Daten von Deutschland werden tagesaktuell über den Server der Geofabrik als Shapefile-Extrakt und als Rohdaten angeboten.⁴ Die dort aufbereiteten Shapefiles enthalten nur einen Teil der Daten und sind für viele Fragestellungen deshalb nicht geeignet. Für die Analyse der WKA werden die Originaldaten im PBF-Format benötigt. Die ca. 2 GB große Datei *germany-latest.osm.pbf* enthält die Gesamtdatenbank für OSM in Deutschland.⁵ Dieser Datenbestand ist nur handhabbar, wenn er in eine performante Geodatenbank importiert wird, so dass die PBF-Datei in einem ersten Schritt mit Hilfe des Programms *ogr2ogr* in eine PostGis-Datenbank geladen worden ist.⁶ Das Kommando zum Import in die Datenbank OSM sieht folgendermaßen aus:

```
ogr2ogr --config PG_USE_COPY YES -f PostgreSQL PG:"dbname='OSM' host='localhost' port='5432'
user='postgres' password='passwort'" -s_srs EPSG:4326 -t_srs EPSG:25832 -skipfailures germany.pbf
```

Die resultierende Datenbank war Anfang 2014 ca. 12 GB groß und enthält einen Polygonlayer, einen Linienlayer mit dem Verkehrs- und Gewässernetz sowie einen Punktlayer mit insgesamt 5.270.054 Punkten.

Windanlagen auswählen

In jedem GIS-Programm lassen sich die Windanlagen mit einer SQL-Where-Abfrage über den Abfrageeditor auswählen und anschließend als eigenen Datensatz speichern. Auf der Internetseite des OSM-Projektes findet man die Attributierungsregeln für Windanlagen.⁷ Die Attributierung, das sogenannte *Tagging*, lautet für Windanlagen: „*generator:source=wind*“ und findet sich nach dem PostGis-Import in der Spalte *other_tags*. Mit der Abfrage

```
"other_tags" like '%wind%' AND "other_tags" like '%generator%'
```

lassen sich die Windanlagen auswählen. Zum Zeitpunkt der Auswertung waren 21.301 Windanlagen in der Datenbank erfasst, dies entspricht 90 % der vom Bundesverband für Windkraft registrierten Anlagen. Auf einer solchen Grundlage lässt sich eine räumliche Analyse zur Verteilung der Windanlagen in Deutschland umsetzen.

Geodatenverarbeitung

Abbildung 1 stellt das Verteilungsmuster der Windanlagen in Deutschland da. Die grundlegende Verteilung ist gut zu erkennen. Ziel der Analyse ist jedoch eine kategorisierte Zonierung der Verteilung, die zu anderen räumlichen Datensätzen mit Polygon-Geometrien in Bezug gesetzt werden kann.

Wenn die gesamte Fläche Deutschlands im Bezug zur Dichte der Windanlagen kategorisiert werden soll, muss zunächst eine Methodik ausgewählt werden, mit der die Verteilung zu analysieren ist. So ließe sich ein quadratisches Raster mit 10 km Zellausdehnung über Deutschland legen, um anschließend die Anzahl der Windanlagen in den Zellen ermitteln zu lassen.

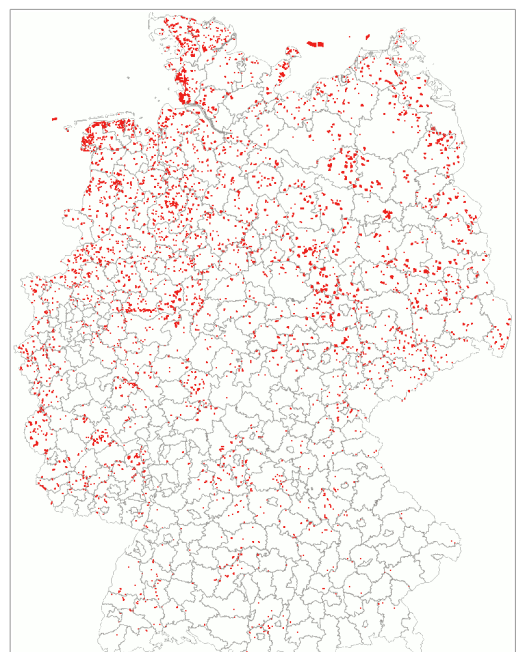


Abbildung 1: Verteilungsmuster WKA

4 <http://download.geofabrik.de/europe/germany.html>

5 <http://download.geofabrik.de/europe/germany-latest.osm.pbf>

6 Dies Programm ist teils der GDAL/OGR-Software die beispielsweise mit QGIS ausgeliefert wird

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

Direkt zum Thema: <http://wiki.openstreetmap.org/wiki/DE:Tag:generator:source%3Dwind>

Interessanter ist, eine kontinuierliche Oberfläche mit 1 qkm Pixel-Auflösung zu generieren, die über eine farbige Signatur symbolisiert, wie viele WKA sich im Umkreis von 20 km um jeden Pixel befinden. Dies lässt sich mit Hilfe von Interpolationsverfahren erreichen, wenn die WKA-Punkte mit Werten zur Anzahl der benachbarten Windanlagen versehen werden. Jede WKA bekommt als Wert die Anzahl der im Umkreis von 20 km stehenden Windanlagen. Dies wird umgesetzt, in dem

- um die Windanlagen 20 km-Pufferpolygone mit der ID der Ursprungspunkte gelegt werden,
- anschließend die Anzahl der im Pufferbereich liegenden WKA-Punkte ermittelt und als Attribut auf die Pufferfläche übertragen werden (Werkzeug: *SpatialJoin*), um anschließend
- die Werte über *AttributJoin* anhand der ID auf die WKA-Punkte zurückzuübertragen.

Anschließend ist für jeden WKA-Punkt die Anzahl der in 20 km befindlichen Windanlagen als Attribut verzeichnet. Es gibt allerdings Bereiche, in denen sich im Umkreis von 20 km keine einzige Windanlage befindet. Da jedoch nur die WKA zum Sammeln der Punkt-Dichtedaten verwendet worden sind, gibt es keine Punkte mit dem Wert 0, so dass bei einer anschließenden Interpolation auch keine Bereiche ermittelt werden können, in denen sich keine Windanlagen im Umkreis von 20 km befinden. Aus diesem Grunde muss ein Hilfs-Punktraster erstellt werden, um in Bereichen ohne Windanlagen Datenpunkte mit dem Wert 0 für die Interpolation zu erzeugen.

Zu diesem Zweck wird eine regelmäßige Punktmenge mit 20 km-Abstand generiert, auf die im vorher beschriebenen Verfahren die Anzahl der benachbarten Windanlagen übertragen wird. Die resultierende Punktmenge wird mit den WKA Punkten vereinigt, so dass eine Punktmenge entsteht, die auch Punkte mit dem Attributwert 0 enthält. Abschließend wird aus dem Punktshapefile mit den Dichtewerten eine Rasteroberfläche mit 1 km² Auflösung im Spline-Verfahren (GRASS v.surf.rst) interpoliert. Das Ergebnis ist in Abbildung 2 zu sehen.

Die resultierende Oberfläche stellt ein nachvollziehbares Bild der Verteilung von Windanlagen in Deutschland da. Aus der Rasteroberfläche lassen sich in einem GIS-Rasterrechner gewünschte Wertebereiche abfragen und in einzelne Zonen-Shapefiles für konkrete bundesweite Fragestellungen exportieren. So ist es möglich, extrahierte Zonen besonders hoher Anlagendichte mit dem Vorkommen spezifischer Arten, mit Vogelzugrouten oder mit bestimmten Schutzgebieten zu verschneiden.

Fazit

Wo es keine offiziellen Daten gibt, lohnt es sich in öffentlich zugänglichen Datenbanken nachzusehen. *OpenStreetmap* kann eine ertragreiche Quelle sein. Es bedarf in jedem Einzelfall einer sorgfältigen Prüfung, ob die Daten für den jeweiligen Zweck nutzbar sind. Im Bereich Verkehrswege, erneuerbarer Energien und Stromleitungstrassen ist OSM eine gute Quelle. Anzumerken ist, dass in OSM Daten zu finden sind, für die es ansonsten gar keine Quellen gibt. So werden viele inoffizielle MountainBike-DownHill-Abfahrten in OSM eingetragen.

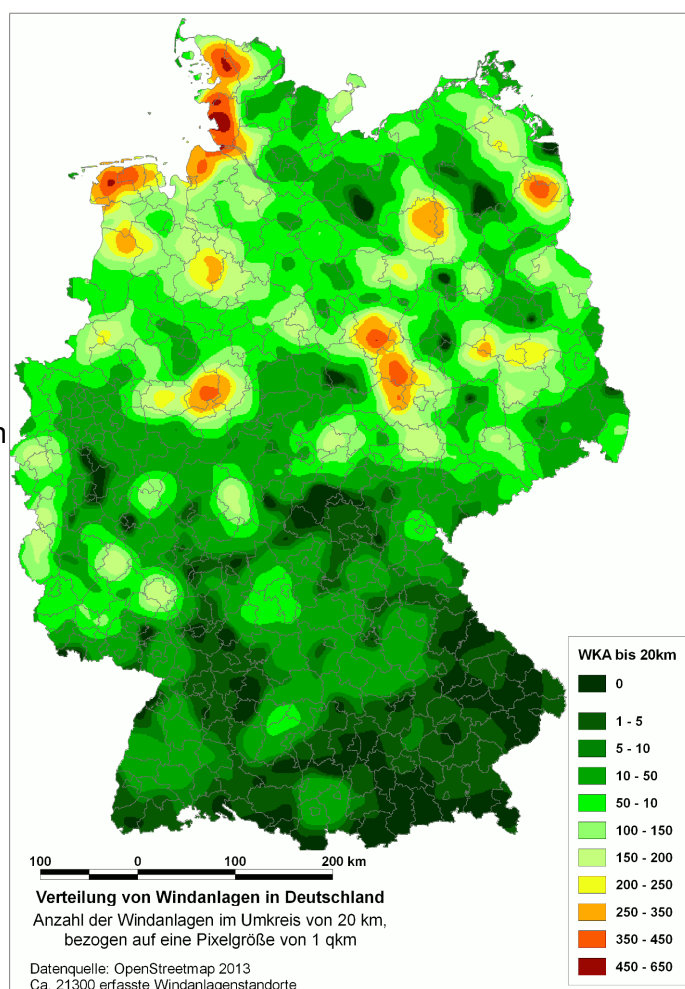


Abbildung 2: Windanlagendichte in Deutschland

OpenStreetmap als Datenquelle für besondere Verkehrsarten

Das Straßennetz war von Anfang an Schwerpunkt der Datenerfassung im OSM-Projekt. Die Vollständigkeit des in OSM erfassten Straßennetzes erreicht bis zur Ebene der Kreisstraßen flächendeckend das Niveau des ATKIS-DLM. Auf der Ebene von Gemeindestraßen und Wegen ist die Vollständigkeit regional sehr unterschiedlich. In vielen Gegenden sind kleinste Pfade und Fußwege erfasst, in anderen fehlen selbst einzelne asphaltierte Feldwege.

Die Attributierung ist auf Kategorien des amtlichen ATKIS-DLM-Straßennetzes übertragbar, wie die Modellierung beider Netze in Abbildung 3, S. 4 bei einer Gegenüberstellung für den Bereich der Stadt Kassel zeigt.

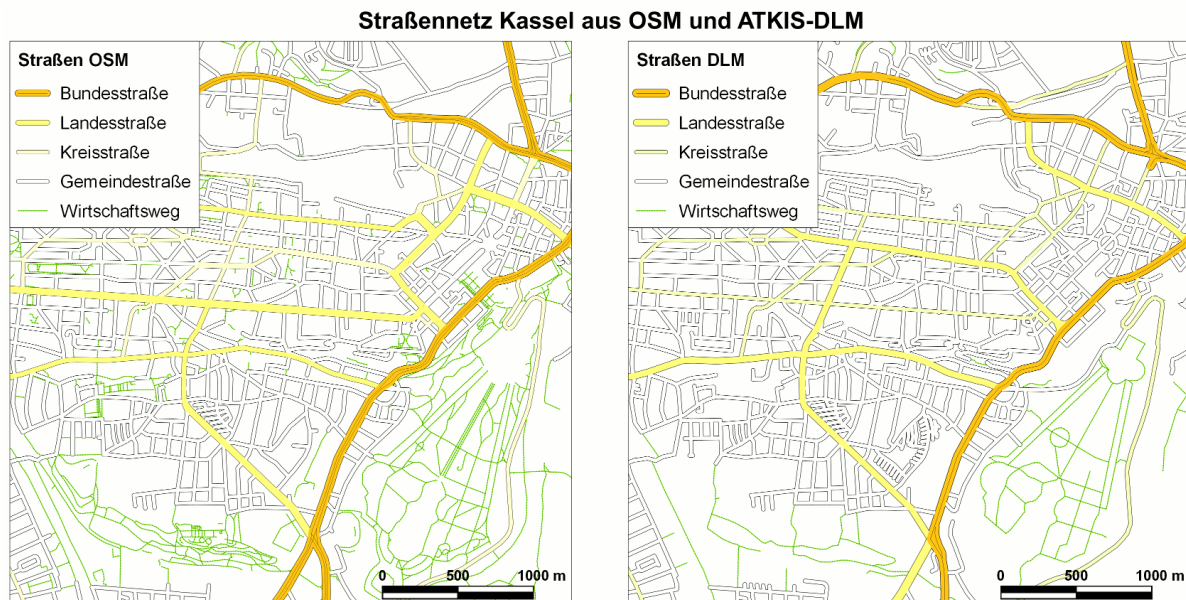


Abbildung 3: Vergleich des Straßennetzes Kassel auf Basis von ATKIS und OSM

OSM-Daten sind als Datenquelle brauchbar, bieten jedoch keine wesentlichen Vorteile bei Analysen überregionaler Straßennetze, wenn der Zugriff auf amtlichen Daten vorhanden ist. Die Aktualität der OSM-Daten ist jedoch in vielen Gebieten höher als die der amtlichen Daten.

Bei der Suche nach ganz speziellen Daten ist OSM eine erstaunlich ergiebige Quelle. Im folgenden Beispiel werden Mountain-Bike-Strecken auf unbefestigten Wegen deutschland-weit ausgewählt und visualisiert. Stellt man das Ergebnis naturschutzfachlich bedeutenden Waldlandschaften gegenüber, ist es möglich Regionen einzugrenzen, in denen Konflikte zwischen Naturschutz und Mountainbikern potentiell häufig auftreten können.

Mountainbike-Routen sind in der OSM-Datenbank im Linien-Layer über eine Suche nach dem *tag mtb* zu finden. Gleichzeitig wird die Suche auf unbefestigte Wege (*highway = path*) eingegrenzt.

```
"other_tags" LIKE '%mtb%' AND „highway“ = 'path'
```

visualisiert die Verteilung der in der OSM-Datenbank auf vollständig unbefestigten Wegen erfassten Mountainbike-Strecken in zwei verschiedenen Darstellungsarten. Es ist zu erkennen, dass Harz und Pfälzerwald eine auffällige Dichte von Mountainbikestrecken aufweisen.

Die OSM-Daten im Originalformat und als informationsreduzierte Shapefiles finden sich auf <http://download.geofabrik.de/europe/germany.html>. Das Straßennetz ist auch in den Shapefiles vollständig und attribuiert enthalten.

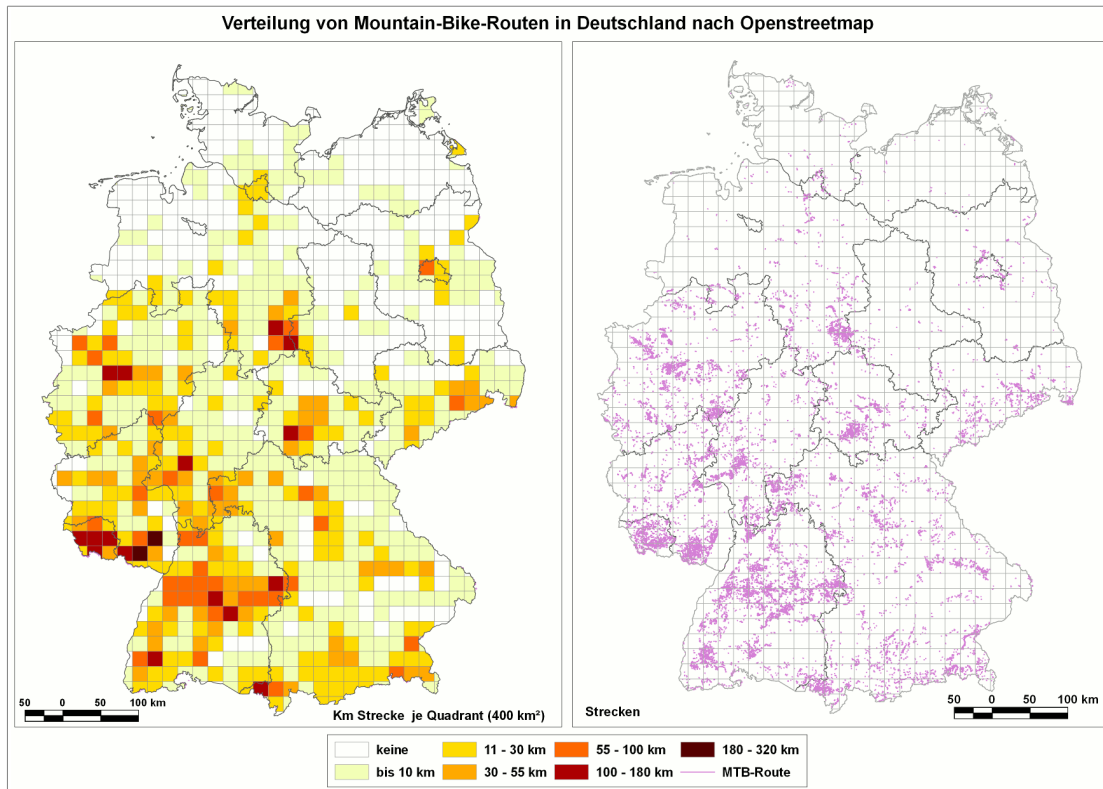


Abbildung 4: Verteilung der von OSM auf unbefestigten Wegen erfassten Mountainbikestrecken

Stromleitungen aus OSM

Auch das Stromleitungsnetz ist in den OSM-Daten nahezu vollständig erfasst. Ein Vergleich des Stromleitungsnetzes aus dem ATKIS-DLM Hessen mit den OSM-Daten zeigt, dass sämtliche im ATKIS-DLM erfassten Stromleitungen dieses Gebietes auch von den OSM-Daten repräsentiert werden. Es handelt sich um Leitungen mit mindestens 110.000 Volt Spannung. Niedrigspannige Freileitungen werden nicht im ATKIS-DLM aber in den OSM-Daten aufgeführt. Zudem sind die Leitungen im OSM-Datensatz zusätzlich nach KV-Spannung attribuiert.

Der Vergleich bezieht sich auf die Objektklasse „DLM_51005_AX_Leitung (1100 = Freileitung)“ des ATKIS-Basis-DLM und auf die mit „power=line“ sowie „voltage“ attribuierten **ways** (Linien) des OSM-Datensatzes, die über eine entsprechende Attributabfrage extrahiert werden können.

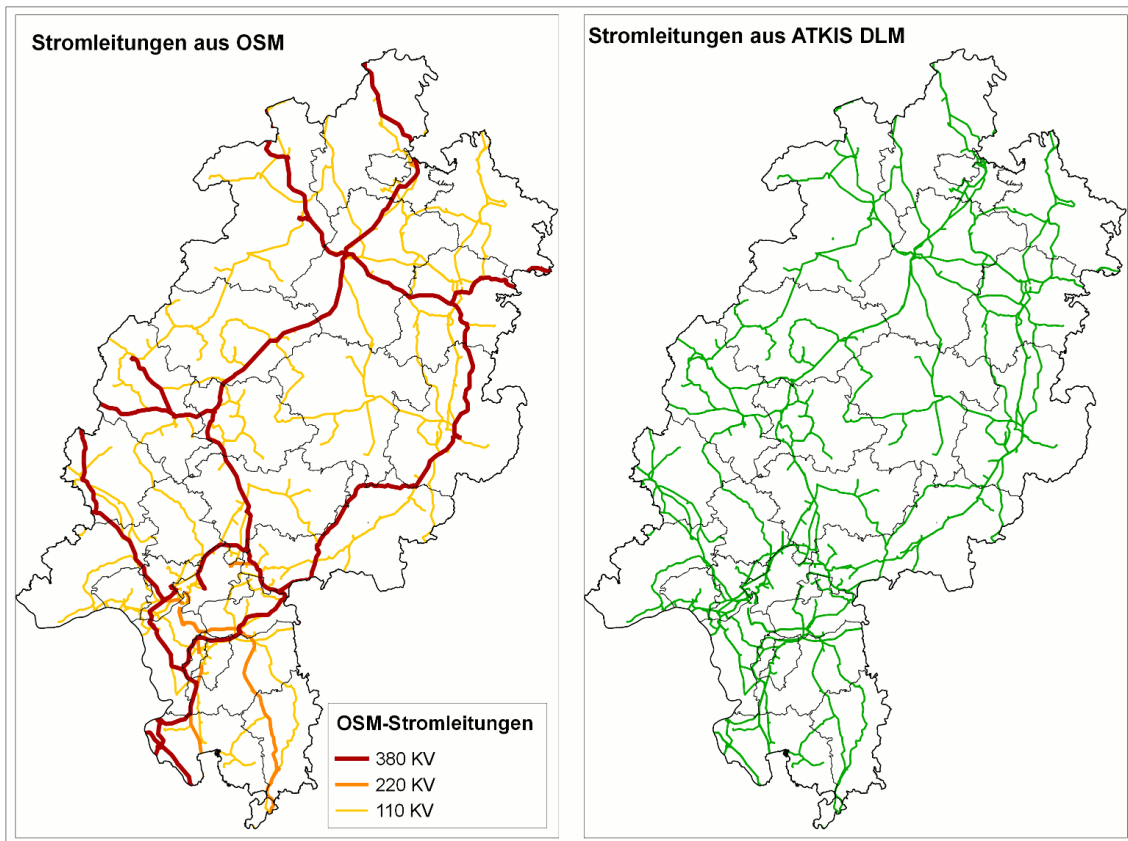


Abbildung 5: Hochspannungsleitungen in Hessen: DLM und OSM im Vergleich

GKG-Kassel - Dr.-Ing. Claas Leiner
Geodatenservice, Kartenwerkstatt &
GIS-Schule Kassel

Wilhelmshöher Allee 304 E
34131 Kassel
Tel. 0561/56013445
claas.leiner@gkg-kassel.de
<http://www.gkg-kassel.de>