

Open Source in accessibility analyses

Modeling of street petrol station accessibility in Germany

by Stefan Neumeier

Johann Heinrich von Thünen-Institut, Federal Research Institute for Rural Areas, Forestry and Fisheries, Germany. stefan.neumeier@ti.bund.de

Abstract

The paper relates two closely intertwined story lines. A socioeconomic one concentrating on street petrol station accessibility in Germany's rural areas and a geospatial one concentrating on the usability of the Open Street Map within rural studies. This is attributed to the fact that the paper builds on findings from applied research within rural studies and is intended to serve as a "practical experiences report" assessing the usability of open source GIS/-data in rural studies. Here we analysed the accessibility of street petrol stations as one core service of general interest important for the overall individual mobility of the population, especially in rural areas, based on an exemplary raster-based GIS accessibility analysis. This analysis builds upon an open source approach using PostgreSQL/PostGIS as well as the Dijkstra shortest path algorithm implemented in the Perl-module „Graph-0.94“. Besides acquiring objective data on street petrol stations accessibility for policy advice, we were also interested in reviewing the usability of OpenStreetMap (OSM) data compared to commercial routing networks (ESM).

Altogether the findings suggest that in Germany, street petrol stations are for the majority of the population (ESM: 99.5%/ OSM: 99.4%) quite accessible. On average the distance to the next street petrol station amounts 5.4 km (ESM)/ 5.5 km (OSM). Regions with disadvantageous accessibility are predominantly sparsely populated.

The comparison of the accessibility values calculated based on ESM and OSM showed that great differences exist in a per cell comparison as well as on the community level, whereas aggregated average accessibility values for greater aggregates like counties proved to be comparable to each other. Against the background of OSM's lower level of completeness in rural areas an interesting but unexpected result is the fact that the accessibility values differ within urban as well as rural areas within the same range. Nevertheless, considering the identi-

fied shortcomings of the OSM, the data set still seems to show a lower performance than commercial data sets.

Keywords: Petrol Station Accessibility, Shortest Path Analysis, OpenStreetMap.

1. Introduction

Amongst other factors, accessibility is one precondition for the economic, social and cultural development of regions (Hemetsberger & Ortner 2008). For example, good accessibility is one precondition for the participation of a region's resident in the economic prosperity of the centres; for participation in potential development opportunities, and the economic capacity of business locations (Hemetsberger & Ortner 2008, Platzer & Gmeinhardt 2003). However, accessibility is not only important for location decisions and regional development but also for the individual life situation of the citizens. The quality of infrastructure accessibility also determines the regional infrastructure supply (BMVBS: Raumordnungsbericht 2011, p. 55). For the current discussion on ensuring the provision of services of general interest (SGI) for Germany's rural areas, held against the background of the demographic change and its potential consequences as well as the normative aim of providing comparable living conditions in all areas, up-to-date information about the accessibility of SGI is important in order to get an objective and realistic view on the current situation that can function as input for policy interventions (Schulz & Bröcker 2007). Accessibility information can also be utilized as input for catchment analyses, potential analyses and econometric analyses, for the building of regional typologies or for regional benchmarking. Objective methods of assessment and comparison are needed for this purpose. Data aggregated in statistical regions (supply indicators) strongly distort the situation as important intraregional disparities are levelled by the aggregation. Activity-based accessibility indicators like the accessibility of regional metropolitan areas mainly permit only indirect conclusions due to the accessibility of SGI (e.g., based on the level of centrality of the different places and their planned provision of infrastructure). Therefore more concrete data on the provision of SGI, is needed to enable scientists as well as politicians to get an idea of the true situation. The modelling of accessibility situations based on a raster-based ac-

cessibility analysis is one solution for obtaining realistic quantitative accessibility data. This information will be presented in the next paragraphs, taking as example the accessibility of street petrol stations in Germany. The analysis itself is based on an open source approach using PostgreSQL/PostGIS as well as the Perl-module „Graph-0.94“ and is based on the Dijkstra shortest path algorithm. In addition to acquiring objective accessibility data for policy advice, the study also aims to compare the usability of the OpenStreetMap (OSM) data as a low/no cost alternative for rural studies, that according to Ludwig et al. (2010) is less complete in comparison to commercial routing networks especially for rural areas. Thus, instead of concentrating on pure GIS technical facts only, the article purposely relates two story lines: a socioeconomic one (street petrol station accessibility in Germany's rural areas) and a geospatial one (usability/constraints of the Open Street Map within rural studies), that are closely intertwined. This is attributed to the fact, that the article builds on findings from applied research within rural studies and is intended to serve as a “practical experiences report”.

The remainder of the article is divided into seven chapters. In sections 2 to 4, key data are introduced on the situation and importance of street petrol station accessibility as well as the core concepts of accessibility. In section 5 the methodology of the accessibility analysis is explained. In section 6 key results are presented comparing accessibility data acquired by using commercial street data and data acquired by using OSM data. In section 7 the lessons learned due to the street networks as well as software used are summarized. The paper concludes with a consideration of the usability of OSM data in research focusing on rural areas.

2. Key data to mobility in rural areas

Mobility is generally defined as movement, whereas a differentiation is made between social mobility (change in the socio-economic status) and spatial, respectively geographical, mobility (migration, traffic, transport) (Leser et al. 1993, p. 409, Johnston et al. 2000, p. 507). The article focuses on geographic mobility. The development of the mobility behaviour is characterised by a rising motorisation of private households; an increasing share of non-commercial traffic; a more and more circadian traf-

fic situation; higher growth rates in shopping and leisure time traffic, as well as increasing time budgets for daily mobility resulting in increasing travel distances (Küpper & Steinrück 2010, p. 14). Reasons are the individualisation, orientation towards leisure-time of society, increasing household incomes, increasing distances between place of residence and the places of other functions of daily life as well as traffic policy and policy areas influencing spatial and settlement patterns (Küpper & Steinrück 2010, p. 14).

Generally only rudimentary public transportation exists in rural areas or sub-optimal station times prevail so that citizens are dependent on their cars if they want to be mobile. Additionally, especially in rural areas, comparatively long distances have to be covered to reach, for example, the next supermarket, place of employment or medical services (Küpper & Steinrück 2010, p. 5). In sparsely populated rural areas nearly two thirds of all travel is covered with passenger cars (Küpper & Steinrück 2010, p. 14).

Therefore, according to Heinze et al. (1982), the availability of a passenger car is a key to individual space and time structuring (Heinze et al. 1982, p. 368). But in order to be mobile, not only the availability of a passenger car, but also the accessibility of a petrol station, is important.

3. Key data to the German petrol station market

Petrol stations can be attributed to the local supply, meaning goods and services for short- and mid-term needs near to the place of residence. They can be seen as services complementing the basic services. In Germany there were 14,744 petrol stations in 2011. These petrol stations can be divided into 377 so called highway petrol stations¹⁹ and 14,367 street petrol stations (ADAC 2012).

According to a study about the petrol station market, Germany, - with less than two petrol stations per 10,000 citizens and three petrol stations per 10,000 cars - belongs to the countries with the lowest petrol station densities in Europe (Morgenstern & Zimmermann 2012, p. 15). Besides selling petrol, a large number of petrol stations in Germany complement their core business with additional retail trade (Korn 2006, Morgenstern & Zimmermann 2012).

In order to keep services competitive in rural areas they are often concentrated in central locations at the expense of peripheral regions (OECD 2007, p.

¹⁹This are petrol stations that are only accessibly by highway, including the outlets of the mineral oil companies.

18-19). This also applies to the provision of petrol near the place of residence. Since 1970 the number of street petrol stations in Germany has decreased from 46,684 street petrol stations in 1970, to 14,367 street petrol stations in 2011 (ADAC 2012). The thinning of the petrol station network led to an increasing spatial concentration at highly frequented profitable locations (Gyllensvärd 1999, Korn, 2006). Unfortunately regionalised data on street petrol stations is not available, but considering the trend to a thinning of SGI, it can be assumed that especially unprofitable locations with comparatively low numbers of customers, especially in rural areas, have closed.

4. Accessibility

The concept “accessibility” is used in quite a few areas, including infrastructure- and city planning or marketing. Therefore the term “accessibility” has quite a few meanings. But generally accessibility can be understood as the number of opportunities for economic and social life that can be reached with a reasonable effort. Accessibility describes the quality of a point in space constituted by its traffic connections to other attractive points in space. So accessibility is the main product of transport systems (Bleich & Koellreuter 2003, p. 7, Schürmann et al. 1997, Schwarte 2005). Different methods exist to get accessibility values that can roughly be grouped into three categories: (1) approaches common mainly in a regional economy based on spatial interaction models (e.g., gravitation models, logit-models) (Bleich 2005, Schulz & Bröcker 2007); (2) approaches common in transport sciences based on a prognosis of the traffic situation (e.g., gravitation models, opportunity models, random utility theory) (Bleich 2005, Schulz & Bröcker 2007) and (3) approaches focusing on the geographic accessibility (Euclidean distance, distance within route networks) (Hemetsberger & Ortner 2008) as the approach followed in the study. One peculiarity of geographic accessibility is that in regions with a dense road network like metropolitan agglomerations and cities, the Euclidean distance provides accurate enough results whereas in regions with sparse road networks like rural areas, the determination of the Euclidean distance is insufficient. This is because the circumvention of natural and anthropogenic barriers is more likely to extend the distances to be covered (Dahlgren 2008, p. 16). In such cases it is therefore better to determine the geographic accessibility based on distances within the traffic network. This kind of analysis can be sup-

ported by the use of network analysis methods that are part of many Geographic Information Systems (GIS). Thereby two main approaches can be used:

1. In a more traditional, less cost intensive approach for a selected, quite limited number of starting points (e.g., centroids of polygons representing administrative units) the distance to targets is computed based on a shortest path network analysis – often with predetermined time- cost budgets. The results are catchment areas, respectively time/cost matrices. The main drawbacks of this approach are that the resulting isochores do not allow a further internal differentiation and that results are not spatially inclusive and comprehensive.
2. In a quite cost-intensive raster-based approach the region under consideration is overlain by a raster with a pre-defined feed size. The centroids of the raster cells represent the starting points of the analysis. That means for every raster-centroid, the shortest distance to the next target is computed and the resulting distance value is attributed to the raster cell. The advantage of this approach is that the distance can be computed from almost unlimited starting points (raster cells) to a high number of targets. Results are also time/cost matrices but ones which, in contrast to the former approach, are spatially inclusive and comprehensive. Furthermore the resulting isochores can be further differentiated internally. Based on the single results, calculation of areas can easily be achieved, e.g., by computing the arithmetic mean. The feed size of the raster cell determines the accuracy of the model as the smaller the feed size the more accurate is the model. Simultaneously, the smaller the feed size the more computation costs are necessary. Thus the feed size of the raster cells has to be selected carefully considering the available resources and desired/necessary accuracy of the results.

5. Method: GIS accessibility model

As the raster-based approach is more flexible, it was decided to take this approach as a basis for the calculation of the street petrol station accessibility. All data have been transformed to the geographical reference system „DHDN/3-degree Gauss-Krüger Zone 3“ (SRID/EPSC: 31467). The data preparation

and analysis was performed with PostgreSQL 9.1.1, PostGIS 1.5.3 and Perl 5.14.2.

Road network

Besides obtaining accessibility information for policy advice, the study also aimed at assessing the usability of OSM data released under the Open Data Commons Open Database Licence (OdbL) as low/no cost alternative compared to commercial street network data taking as example the Esri StreetMap (ESP) dataset. The OSM is a „collaborative project in which volunteers are creating a free and editable map of the world, to reduce the dependency on commercial data providers“ (Ludwig et al. 2012, p. 148). That means the data of the OSM is based on data collected by volunteers interested in participating at the project, whereas practically everybody is able to participate. One major drawback inherent to the OSM project is that data collection does not follow a specific acquisition plan, meaning it depends on the volunteers participating for which region data is collected. All in all, Ludwig et al. (2010) showed that in Germany the OSM is less complete compared to commercial data sets, whereas the completeness of the OSM is higher in densely populated areas than in sparsely populated ones (including rural areas) (Ludwig et al. 2010, p. 149). Nevertheless because the OSM can be used without having to pay licence fees, the question arises if despite the drawbacks concerning the overall completeness, especially in rural areas, the dataset can be used for accessibility analyses without distorting the results too much. Therefore, in order to evaluate the usability of OSM for accessibility analyses, the analysis of the street petrol station accessibility has been carried out based on ESP data that are based on data from NAVTEQ and Tele Atlas as well as on OSM data. Thereby the download as well as preparation of the OSM street data was performed with the help of the tool „OSM2PO“ developed by Carsten Möller (<http://osm2po.de/>).

Street Petrol Stations

The locations of street petrol stations have been derived from the database „D Tankstellen“ that was provided by the POICON GmbH, a company specialized in points of interest for car navigation systems. The street petrol station dataset is provided for free at the POICON website but is copyrighted by the POICON GmbH. It was deliberately decided

not to use the petrol station data set of the OSM (amenity fuel) as this data set proved not to meet the requirements desired for following reasons: (1) For Germany the OSM contains only information about 12,971 petrol stations (ca. 87 % of the overall German petrol stations). As no information is given on missing petrol stations it is more than likely that systematic missing locations might occur especially in rural areas which would distort the result of the accessibility analysis considerably. (2) The OSM dataset does not differentiate between the street petrol stations in which we were interested and highway petrol stations, which we wanted to exclude from our study. (3) The OSM data set does not allow explicit differentiation between branded and independent street petrol stations. (4) Experiences showed that amenity tags within the OSM seem not to be quality controlled or validated, rising questions about validity. The POICON data set „D Tankstellen“ used contains geographical locations for 14,217²⁰ street petrol stations in Germany. Altogether the data set contains 149 (ca. 1 %) street petrol stations less than those listed in the current German Petrol Station Statistics 2012. So the data set covers approximately 98.96 % of Germany's street petrol stations.

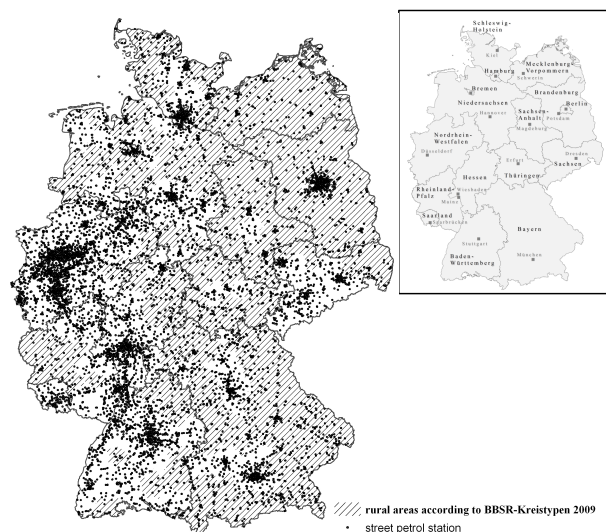


Figure 1: Distribution of street petrol stations in Germany. (Data sources: POICON GmbH; Administrative boundaries: Bundesamt für Kartographie und Geodäsie 2010).

Figure 1 shows the distribution of street petrol stations in Germany. It can clearly be seen that the density of the street petrol station network decreases

²⁰ Altogether the dataset „D Tankstellen“ contains 14,222 street petrol station locations. However, five locations were excluded from the analysis as they proved to be outside Germany.

from the agglomerations to the more rural areas. Furthermore it can also be seen that the density of the network is comparably lower in the new federal states „Mecklenburg-Vorpommern“, „Brandenburg“, and „Sachsen-Anhalt“.

In 2009, the Federal Institute for Research on Building, Urban Affairs and Spatial Development (BBSR) developed the so-called BBSR-Kreistypen 2009 (BBSR 2009a) that are meant to assist in an intraregional comparison of different types of regions within Germany. The system differentiates between cities, defined as urban districts with more than 100,000 citizens and other districts which, including the hinterland, have less than 100,000 citizens. Outside the cities the types are defined based on the population density as well as the settlement structure.

A merging of the street petrol stations with the BBSR-Kreistypen 2009 provides us with a supply indicator due to street petrol stations. This gives us a first impression about potential differences in the street petrol station availability within the different types of regions in Germany, as well as about areas which are potentially disadvantaged in terms of street petrol station accessibility (see Table 1). The resulting data suggests that in Germany, street petrol stations are least accessible in areas identified as rural.

Table 1: Distribution of street petrol stations in Germany according to federal state and BBSR-Kreistyp 2009 (Data sources: POICON GmbH, BBSR).

Federal State	Number of street petrol		street petrol stations per BBSR-K-reistyp 2009																			
	total	%	1		2		3		4		5		6		7		8		9			
			total	%	total	%	total	%	total	%	total	%	total	%	total	%	total	%	total	%		
Baden-Württemberg	1650	11.6	191	12	539	33			22	1.3	97	5.9	625	38	176	11						
Bayern	2319	16.3	304	13	91	3.9	220	9.5	42	1.8	141	6.1	357	15	291	13	682	29	191	8.2		
Berlin	342	2.4	342	100																		
Brandenburg	426	3.0	22	5.2				264	62						67	16				58	14	
Bremen	118	0.8	96	81																		
Hamburg	275	1.9	275	100																		
Hessen	1136	8.0	181	16	289	25	199	18			31	2.7	232	20	123	11	81	7.1				
Mecklenburg-Vorpommern	294	2.1									22	7.5			35	12	55	19	182	62		
Niedersachsen	1472	10.4			214	15	156	11	118	8	159	11	410	28	214	15	138	9.4	63	4.3		
Nordrhein-Westfalen	3415	24.0	1220	36	1233	36	291	8.5			63	1.8	507	15	101	3						
Rheinland-Pfalz	703	4.9	29	4.1	38	5.4	81	12			71	10	365	52	119	17						
Saarland	165	1.2	61	37	84	51	20	12														
Sachsen	607	4.3	130	21			112	18	27	4.4	49	8.1	213	35			76	13				
Sachsen-Anhalt	359	2.5									74	21	25	7	168	47	46	13	46	13		
Schleswig-Holstein	596	4.2			52	8.7	105	18	39	6.5	86	14	90	15	69	12	83	14	72	12		
Thüringen	340	2.4									57	17	76	22	89	26	92	27	26	7.6		
Germany	14217	100	2851	20	2540	18	1184	8.3	512	3.6	887	6.2	2900	20	1452	10	1253	8.8	638	4.5		

- 1 Core cities in agglomerations
- 2 Densely populated districts in agglomerations
- 3 Highly populated districts in agglomerations
- 4 Rural districts in agglomerations
- 5 Core cities in urbanised areas
- 6 Densely populated areas in urbanised areas
- 7 Rural districts in urbanised areas
- 8 Densely populated rural areas
- 9 Sparsely populated rural areas

²¹ This represents a subset of the AIT-GRID for Germany as cells split by clipping methods have been counted as one cell and cells that could not be attributed to a BBSR-Kreistypen 2009, federal state or community have not been included in the analysis. (This affected cells at the border of the data set where, because of minor geographical inconsistencies in the different data sets available the data sets differ from one another due to their geographical accuracy/coverage).

Raster of Reference

The Austrian Institute of Technology (AIT) - Population Density GRID (Steinöcher, Köstl, Weichselbaum, 2001) for Germany has been chosen as base raster for the analysis. The AIT-GRID is based on Eurostat’s European geostatistical 1km grid „Grid_ETRS89_LAEA_1k“, which was built based on Corine Land Cover 2006 data and which covers areas classified as settlement. The AIT-GRID enhances the European geostatistical 1km grid by adding attribute information due to absolute population according to Eurostat 2006 data for every grid cell. Although the AIT-GRID is based on 2006 population as well as settlement data and does therefore not represent the current state of settlement and population development, it was decided to use this raster as base raster of the accessibility analysis, as it was the only approved data set available with gridded population data for Germany. In order to be inherently consistent it was decided to use the AIT-GRID as is and not to update or merge it with other settlement or population data sets. Altogether the raster contains 209,203 cells²¹ with a feed size of 1,000 m. The raster was chosen for two reasons. First, as the main aim of the analysis is to analyse the proximity of street petrol stations to the place of residence or employment, the raster allows computation costs to be reduced by concentrating on the areas of interest -the settlement areas- as starting point only. Second, and more importantly, the gridded population data offers the possibility to evaluate accessibility values in relation to the population affected, allowing a better assessment of the overall accessibility situation.

Method of raster-based accessibility analysis

Based on the coordinates of the raster centroids of the AIT-GRID, the road network and the coordinates of the street petrol stations for every raster centroid the shortest street distance to the next street petrol station has been calculated. Thus, the accessibility analysis is based implicitly on the assumption that from every start-point the next street petrol station is chosen.

Because of the huge amount of data (street network covering the whole of Germany) it was decided to use PostgreSQL/PostGIS for data preparation prior to the actual accessibility analysis. Neverthe-

less, although PostgreSQL/PostGIS can be extended with network analysis functions with the „pgdijkstra library“ it was decided to perform the shortest path calculation based on the „Dijkstra shortest path“ algorithm (Dijkstra 1959) with the help of the Perl-Module „Graph-0.94“. Two reasons were decisive: First, the hardware available for the analysis proved to run out of resources when trying to perform the whole analysis within PostgreSQL/PostGIS. Second, we wanted to be able to write every completed shortest path calculation to file/database, so that in the case of an error, power shortage, etc. the computation process could easily be resumed with the remaining starting points. Altogether the workflow for the accessibility analysis contained following steps:

1. Building of a network topology

- (a) Determination of the road/line next to every start/target coordinate;
- (b) Determination of the point where the lead from every start/target point crosses the road/line next to the point;
- (c) Adding of a new line to the road network from the start/target point to the point determined by c;
- (d) Splitting of the original line/road in the network at the point determined by c;
- (e) Building of a topological network graph for the resulting street data set by defining the start and end point of every line segment and allocating unique node id's to the defined start and end points;
- (f) Allocation of network node id's to the corresponding start and target points.

2. Shortest path analysis

- (a) Due to performance issues it was decided to perform a catchment area analysis prior to the shortest-path analysis. On the one hand, the three petrol stations nearest to the starting-point under consideration were first defined based on Euclidean distance, and on the other hand the route network to be considered by the shortest path algorithm was limited to only the subset of streets which are within the distance to the point furthest away from the starting-point plus 1,000 meters (e.g., Figure 2).

This allowed the potential 2,974,239,051²² routing requests to be reduced to 627,609²³ requests and the computation time necessary to perform the whole accessibility analysis to be reduced quite considerably²⁴.

- (b) Determination of the shortest distance (in meters) between start point and every potential target point based on the „Dijkstra shortest path“ algorithm;
- (c) Allocation of the shortest paths to the raster cell of the AIT-Grid.

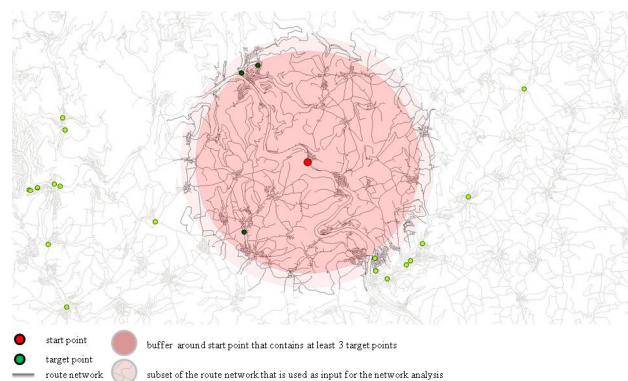


Figure 2: Method of catchment area analysis prior to the routing analysis.

6. Accessibility of Street Petrol Stations – comparing results based on ESM with results based on OSM

After having introduced the data sets as well as the methodology upon which the accessibility analysis is based, I now want to present some of my findings on street petrol station accessibility in Germany. Thereby two aims are pursued: (1) The assessment of the street petrol station accessibility situation in Germany and (2), a comparison of the performance of ESM and OSM in order to approach the question of the usability of OSM within research focusing on rural studies. Due to the accessibility results it has to be kept in mind that accessibility is not static but changes during the course of time. In this sense, the results presented represent a snap-shot for the reference years 2011/2012. Furthermore the results are strongly dependent on the accessibility model used; they do not represent reality but, as common

²²209.203*14.217 (start points * target points).

²³209.203*3 (start points * next three target points).

²⁴Three weeks instead of more than 95 years (estimated based on computation time of single queries).

with models, an ideal typical situation (Johnston et al, 2000, 508 ff.).

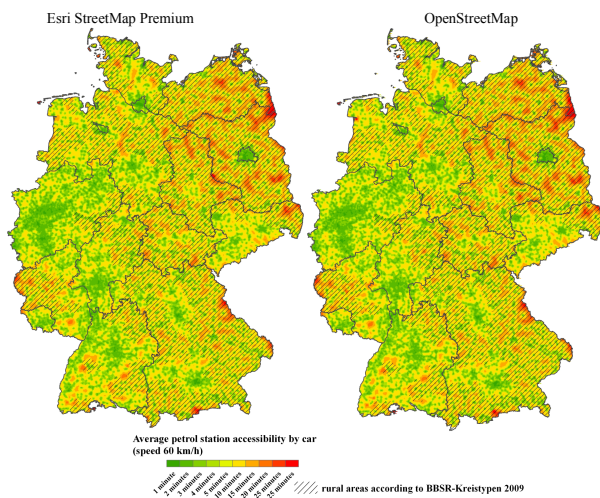


Figure 3: Street petrol station accessibility by car (speed 60 km/h) from ESM data (left) and OSM data (right) (Sources: Administrative boundaries: Bundesamt für Kartographie und Geodäsie 2010; BBSR-Kreistypen 2009: BBSR; Raster: © EFGS, 2009).

Only petrol stations within Germany were included in the accessibility analysis. Petrol stations in neighboring foreign countries were not considered. Therefore in areas close to the country border the model results are likely to diverge more from reality than in regions at greater distances to the country borders. In Figure 3 the accessibility of street petrol stations by car (speed 60 km/h) according to the ESM is contrasted to that acquired with the help of the OSM. Visually, by and large both results are equal to each other. But what do the actual data sets say that is revealed by the map? Does the street network used as a basis for the accessibility analysis make a difference concerning the results obtained? In order to answer this question we have to compare the results with each other. The results of the two accessibility analyses represent related samples as we computed both based on the same starting-points. Furthermore the resulting values proved not to be normally distributed. Therefore it was decided to execute the Wilcoxon-test, a non-parametric statistical hypothesis test, in order to determine if the results differ sta-

tistically from one another. The two-sided Wilcoxon-test²⁵ performed with a confidence interval of 0.95 delivered a V of 9395891083 and a p-value 2.2e-16. So as $p < 0.025$ it can be reasoned that the results gained by the ESP dataset differ from that gained by the OSM in a statistical sense. But what does this mean in practical terms? In order to approach this question we will have a look at the accessibility results computed based on ESP as well as OSM, first.

In Germany the nearest street petrol station is accessible in 5.4 km on average according to ESP and 5.5 km according to OSM (see Table 2).

Considered according to the BBSR-Kreistypen 2009 it can be observed that the average distance to be covered to reach the next street petrol station increases from „Core cities in agglomerations“ (ESP: 1.9 km; OSM: 2.0 km) to „Sparsely populated rural areas“ (ESP: 7.8 km; OSM: 7.9 km). This observation suggests that a correlation exists between BBSR-Kreistyp 2009 and distance to be covered to reach the next petrol station. The existence of such a correlation is statistically proven by the measure of association η ²⁶, which amounts to 0.4 for the ESM as well as for the OSM data and indicates the existence of a medium statistical correlation. All in all it can be concluded that with increasing peripherality, the distance to be covered to reach the next petrol station in Germany increases, too.

Similar to the BBSR-Kreistypen 2009 the average distances to be covered to reach the next street petrol station differ within the single federal states between Berlin (ESP/OSM: 2 km) and Mecklenburg-Vorpommern (ESP/OSM: 8 km) (see Table 2).

Here, too, a possible statistical correlation between distance to be covered and federal state can be determined by considering the measure of association η , which amounts 0.3 for the ESM as well as OSM data and indicates the existence of a medium statistical correlation. All in all, similar to the BBSR-Kreistypen 2009, it can be concluded that the federal state in which a person lives influences the distance to be covered to reach the next petrol station in Germany. Comparing the average accessibility values for the statistical units „Germany“, „BBSR-Kreistypen 2009“ and federal states, the difference between the values based on ESP and those based on OSM are considered less than 150 m in all regions

²⁵The Wilcoxon-test was performed with the statistics program R (p: probability value; R's V corresponds the Wilcoxon W that is the sum of positive and negative ranks).

²⁶ η (Eta) is a measurement of association that enables to compute the statistical correlation between an independent nominal variable and a metric dependent variable. The value range is between 0 and 1 whereas $\eta=0$ means that no correlation exists. As a rule of thumb η can be interpreted as follows: $0 < \eta < 0.2$: extremely weak correlation; $0.2 < \eta < 0.4$: weak correlation; $0.4 < \eta < 0.6$ medium correlation; $0.6 < \eta < 0.8$: strong correlation; $0.8 < \eta < 1$: very strong correlation; $\eta=1$: perfect correlation. The measurement of association η does not allow to determine the direction of the correlation.

(see Table 2). So, considering the accessibility values aggregated for the BBSR-Kreistypen 2009, as well as the federal states, no noteworthy differences can be identified between the results obtained based on ESM or OSM.

Table 2: Comparison of average street petrol station accessibility for selected statistical units based on ESP data set compared to Open StreetMap data set (Source: Own calculation).

statistical units	Esri StreetMap Premium						Open StreetMap						Difference (OSM-ESM)			
	stdev.	var.	meters				stdev.	var.	meters				meters			
			m	avg.	min.	max.			m	avg.	min.	max.	m	avg.	min.	max.
Germany	3772.64	61.42	3637.91	5394.12	18.06	591.69	3850.18	62.05	4693.65	5466.32	19.54	51569.00	4.76	72.20	-42516.21	41396.30
BBSR-Kreistypen 2009																
1 Core cities in agglomerations	1378.13	37.12	1563.76	1899.02	30.78	1263.91	1484.37	38.53	1570.77	1953.26	35.01	14805.26	0.69	54.238	-4945.61	7570.84
2 Densely populated districts in agglomerations	2075.27	45.56	2725.93	3159.54	19.60	18007.39	2155.58	46.43	2774.53	3231.46	19.54	23637.26	4.11	71.922	-15276.83	11485.88
3 Highly populated districts in agglomerations	2970.48	54.50	4114.43	4639.36	29.46	34251.03	3006.94	54.84	4180.20	4696.68	28.17	23543.51	5.15	57.319	-26516.17	16244.04
4 Rural districts in agglomerations	4311.71	65.66	6162.09	6828.79	11.16	59169.00	4357.36	66.01	6191.18	6896.80	110.33	51569.20	3.94	68.009	-42516.21	26675.94
5 Core cities in urbanised areas	1797.60	42.40	2093.92	2522.40	22.51	13943.58	1901.13	43.60	2126.83	2579.00	22.18	23205.99	1.70	56.604	-5650.77	20281.78
6 Densely populated areas in urbanised areas	2907.19	53.92	4148.79	4650.75	18.06	31460.88	3025.52	55.00	4196.99	4719.52	32.28	42198.00	4.70	68.768	-17134.09	31348.41
7 Rural districts in urbanised areas	3716.53	60.96	5722.11	6217.40	27.42	40980.98	3859.70	62.13	5792.24	6340.81	25.54	46953.76	8.38	123.41	-24421.86	41396.30
8 Densely populated rural areas	3519.52	59.33	5369.68	5868.62	27.33	54141.78	3570.80	59.76	5407.01	5913.90	28.10	47972.24	4.61	45.276	-33265.31	34186.50
9 Sparsely populated rural areas	4889.49	69.92	7053.66	7811.93	39.07	48879.79	4939.45	70.28	7092.72	7863.12	33.48	44368.30	3.53	51.197	-26359.42	24496.69
Federal states																
Schleswig-Holstein	3472.45	58.93	4778.67	5365.07	36.88	33634.92	3390.72	58.23	4021.67	5350.77	36.88	32132.01	1.58	-14.29	-10293.86	26675.94
Hamburg	1553.69	39.42	1436.14	1955.31	59.76	9993.74	1676.56	40.95	1304.51	2037.02	55.16	9834.29	1.80	81.71	-4886.56	5479.54
Niedersachsen	3151.31	56.14	4695.43	5134.94	37.17	34251.03	3213.79	56.69	4718.59	5166.57	39.37	42198.00	1.68	31.63	-26516.17	31348.41
Bremen	1603.93	40.05	1599.32	2039.44	116.73	10090.76	1614.48	40.18	1618.67	2070.27	75.52	11454.69	0.69	30.83	-4119.07	4683.78
Nordrhein-Westfalen	2131.25	46.17	2930.77	3341.34	26.99	19478.35	2194.90	46.85	2990.91	3410.53	34.32	23205.99	3.92	69.19	-11694.94	20281.78
Hessen	2883.67	53.70	3624.75	4227.78	19.60	17754.22	3015.69	54.92	3690.49	4309.20	19.54	21991.40	4.35	81.42	-15276.83	16118.24
Rheinland-Pfalz	3807.99	61.71	4899.88	5677.62	22.51	28155.06	3996.36	63.22	4947.42	5781.62	22.18	30710.24	9.68	104.00	-14676.53	14600.52
Baden-Württemberg	3442.84	58.68	4173.37	4911.94	21.29	40980.98	3539.88	59.50	4270.08	5001.60	25.54	38568.59	9.16	89.67	-24421.86	23992.70
Bayern	3497.67	59.14	4992.52	5545.26	18.06	54141.78	3548.47	59.57	5016.87	5591.63	28.10	32617.54	5.85	46.37	-33265.31	20889.38
Saarland	3251.03	57.02	3410.23	4264.37	122.16	20165.05	3404.42	58.35	3495.18	4395.56	114.51	23543.51	6.79	129.19	-5233.76	7422.76
Berlin	1288.63	35.90	1387.31	1707.66	30.78	12631.98	1299.30	36.05	1322.94	1655.72	35.01	14805.26	-4.41	-51.94	-4945.61	7541.02
Brandenburg	5065.29	71.17	7173.29	7954.05	111.16	59169.00	5186.06	72.01	7273.02	8097.70	110.33	51569.20	6.10	143.65	-42516.21	41396.30
Mecklenburg-Vorpommern	5008.15	70.77	7564.53	8118.41	55.70	48879.79	4987.06	70.62	7610.80	8181.88	46.18	44368.30	3.96	63.47	-26359.42	24496.69
Sachsen	3568.31	59.74	4854.92	5490.21	61.58	43971.34	3713.20	60.94	4973.87	5634.56	61.75	32290.57	8.00	144.35	-12857.00	11330.54
Sachsen-Anhalt	4187.20	64.71	6305.46	6888.08	50.91	28473.93	4326.43	65.78	6361.00	6981.82	50.91	47972.24	5.84	93.75	-15176.64	34186.50
Thüringen	3694.95	60.79	5710.73	6197.92	74.62	23420.83	3850.81	62.05	5835.30	6332.63	69.73	24879.75	9.49	135.71	-11385.68	3863.07

Table 3: Difference of street petrol station accessibility between ESM and OSM.

spatial category	nr of cells	distance in m				
		≤ 150	> 150 to 300	> 300 to 500	> 500 to 1000	> 1000
		% cells per cells in category				
all cells	209,203	59.6	10.8	8.1	9.7	11.7
rural areas	112,357	58.0	10.6	8.1	9.9	13.4
urban areas	96,846	61.4	11.2	8.2	9.5	9.7
1 Core cities in agglomerations	7,863	65.9	11.7	8.4	8.1	5.9
2 Densely populated districts in agglomerations	19,678	60.6	11.9	8.8	10.1	8.6
3 Highly populated districts in agglomerations	18,618	60.2	11.0	8.4	9.5	10.9
4 Rural districts in agglomerations	13,956	59.1	10.7	7.7	9.2	13.3
5 Core cities in urbanised areas	3,484	64.5	10.6	8.1	8.6	8.2
6 Densely populated areas in urbanised areas	47,203	61.3	10.9	7.9	9.6	10.4
7 Rural districts in urbanised areas	40,904	55.6	11.0	8.9	10.8	13.8
8 Densely populated rural areas	31,963	60.5	10.5	7.6	9.6	11.7
9 Sparsely populated rural areas	25,534	58.3	9.9	7.5	9.4	15.0

Nevertheless comparing the average distances at the community level it can be observed that the dis-

tances acquired by ESM differ quite considerably from those acquired by OSM. For 47 % of Germany's 11,589 communities, the difference is greater than 150 meters, and for 27 % of the communities greater than 300 meters.

Thereby 72% of the communities with differences of 150 meters belong to the BBSR-Kreistypen 2009 „rural districts in urbanised areas“ (1516), „densely populated areas in urbanised areas“ (1345), „densely populated rural areas“ (763) and „sparsely populated rural areas“ (727). Last but not least, a comparison based on the single raster cells of the AIT-GRID reveals that the computed accessibility values differ, sometimes quite considerably, from one another. In 69% of the rural cells (compare Table 3) the difference between ESM and OSM (absolute value) is less than 300 m (a value, we think is still acceptable when considering driving distances, but not when considering walking distances). Surprisingly, urban areas do not differ much from rural areas as a whole.

Also, considering differences greater than 1,000 m, no noteworthy rural (13 % of the „rural“ raster cells) or urban (10 % of the „urban“ raster cells) differences can be found.

Nevertheless, as can be seen in Figure 4 as well, the cells/regions with the greatest differences can be found in the federal states Mecklenburg-Vorpommern, Brandenburg, Thüringen and Sachsen-Anhalt, northern Bayern, western Rheinland-Pfalz, eastern Niedersachsen and the Alps.

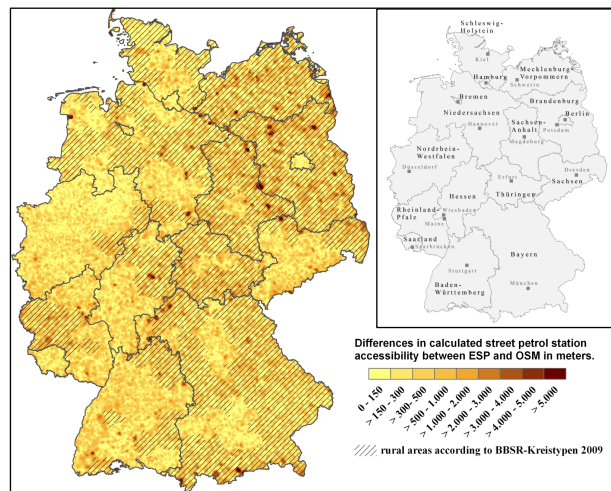


Figure 4: Differences in street petrol station accessibility between ESM and OSM. (Administrative boundaries: Bundesamt für Kartographie und Geodäsie 2010; BBSR-Kreistypen 2009: BBSR; Raster: © EFGS, 2009).

Population and Street Petrol Station accessibility

The identification of the areas with the greatest distances to be covered present a first indication for areas with potential deficits with regard to accessibility, respectively supply shortage, but they do not allow the population affected to be inferred. But this is important for an assessment of the accessibility situation, as a below average accessibility may be rated differently in regions with a high share of population than in regions with a low share of population. Such a consideration is enabled by merging the accessibility results with the population data of the AIT-GRID. Against the background that petrol stations are operated under economic aspects it is not surprising that a modest negative correlation²⁷ can be found between the accessibility of petrol stations and population (ESM/OSM r of -0.32). By trend, the longer the distances to be covered to reach the next petrol station are, the lower the population value is within a raster cell. This finding suggests that especially regions with low population are affected by disadvantageous street petrol station accessibility.

Table 4: Accessibility of street petrol stations according to BBSR-Kreistypen 2009 and population per BBSR-Kreistyp 2009 according to the AIT-GRID (Source: Own calculation).

BBSR-Kreistyp 2009	population per BBSR-Kreistyp 2009 in % according to the AIT- Population Density Grid							
	<=5,400 m		> 5,400 m <=10,000 m		>10,000 m <= 15,000 m		> 15,000 m	
	ESM	OSM	ESM	OSM	ESM	OSM	ESM	OSM
1 Core cities in agglomerations	99.54	99.44	0.43	0.52	0.03	0.04	0.00	0.00
2 Densely populated districts in agglomerations	96.49	96.03	3.33	3.78	0.18	0.19	0.00	0.01
3 Highly populated districts in agglomerations	86.36	85.63	11.75	12.22	1.64	1.91	0.25	0.24
4 Rural districts in agglomerations	73.51	73.30	18.34	18.36	6.73	6.62	1.42	1.72
5 Core cities in urbanised areas	98.53	98.24	1.40	1.70	0.07	0.06	0.00	0.00
6 Densely populated areas in urbanised areas	84.78	84.04	13.12	13.72	1.91	2.02	0.18	0.21
7 Rural districts in urbanised areas	69.36	68.43	23.05	23.06	6.48	7.16	1.12	1.35
8 Densely populated rural areas	76.12	75.64	18.36	18.46	4.97	5.23	0.55	0.67
9 Sparsely populated rural areas	64.04	63.97	21.83	21.63	10.31	10.38	3.82	4.02

All in all the portion of the population that is able to reach a street petrol station by car (60 km/h) within 5.4 minutes (average time/distance according to ESM) is based on the ESP data 87.43 %, based on the OSM data 86.93 % Within 15 minutes²⁸ according to the ESM data set 99.56 % and according to the OSM data set 99.50 % of the German population is able to reach the next petrol station by car (60 km/h). Comparing the street petrol station accessibility within the BBSR-Kreistypen 2009 (Table 4)

²⁷ Pearsons's product moment correlation coefficient..

²⁸ This is a time-span that according to a study of the „Amt für Raumentwicklung und Geoinformation“ of the canton St. Gallen, Switzerland (2008) is commonly accepted to reach SGI.

²⁹ Former German Democratic Republic.

³⁰ Federal States of the Federal Republic of Germany before the reunification with the German Democratic Republic.

it becomes obvious that in rural regions the share of population that has to cover greater distances to reach the next street petrol station is greater than in urban regions.

But all in all, in all BBSR-Kreistypen 2009, except in sparsely populated rural areas (96 % ESM/OSM), over 98 % (ESM/OSM) of the population can reach the next street petrol station by car (60 km/h) within 15 minutes. In summary, comparing the average accessibility of street petrol stations in Germany according to the BBSR-Kreistypen 2009 no noteworthy differences or deficits can be found.

Synthesis – Street Petrol Station accessibility in Germany

The results on street petrol station accessibility in Germany as key service supporting the mobility of citizens in rural areas can be summarized as follows:

- In Germany, a street petrol station can on average be reached within 5.4 km (ESM)/ 5.5 km (OSM);
- The higher the absolute population within a region, the shorter the distances to the next street petrol station;
- With increasing peripherality, the distance to the next street petrol station increases;
- The distance to the next street petrol station is dependent on one's location within the federal states, whereby in tendency the distances are longer in the „new“ federal states²⁹ than in the „old“ federal states³⁰;
- At least 87.4 % (ESM)/ 86.9 % (OSM) of the German population is able to reach the next street petrol station within 5.4 minutes by car (60 km/h), and another 12.1 % (ESM)/12.5 (OSM) within 15 minutes.

Altogether these findings suggest that in Germany, despite of the decreasing number of street petrol stations since 1970 and the fact that Germany belongs to the European countries with the lowest density of petrol stations, street petrol stations are for the majority of the population, quite accessible. Regions with disadvantageous street petrol station accessibility are

predominantly sparsely populated. These findings show that, based on the model results, no need for action or intervention can be identified.

7. Lessons learned

The following résumé summarizes the experiences as well as first subjective impressions due to the usability of the data sets as well as the software used. Following this consideration we will conclude with a consideration of the usability of OSM data compared to commercial routing networks within rural studies.

ESRI StreetMap Premium

The commercial ESP dataset proved to be quite comprehensible. The dataset available (Esri StreetMap Premium for mobile) contained streets as lines as well as street crossings as points allowing to differentiate between street crossings and bridges/tunnels/dead-ends. Unfortunately, in order to obtain a topological street network the street crossings had to be projected into the street-lines dataset first. Subjected to the size of the street network under consideration this process can consume quite a lot of computation time. Nevertheless the process could be performed without greater difficulties. Minor topology inconsistencies like unconnected lines and street crossings lead to a number of 1,484 (ca. 0.7 %) start points for which no routing could be performed. For these points the simple Euclidean distance had been determined in order to compensate the missing values.

OpenStreetMap

Compared to the ESP dataset the OSM dataset, proved to be less comprehensible and required more time to figure out its internal structure and how to obtain the data necessary for building a street network fit for performing network analyses. Especially the presentation and description of the OSM in the form of a „wiki“ makes it, according to my perception, quite difficult to become quickly acquainted with its structure. The availability of a concise and printable document describing how to get the data step-by-step and what to do in order to use it would have been very helpful – especially if one only wants to use the data without getting too involved in all the

technical detail necessary to participate in the OSM project.

One major drawback experienced was the fact, that after downloading and importing the route data in the database, because of topological inconsistencies/errors, the dataset proved not to be suitable for routing purposes. That means lines obviously connected in reality are split because of dangling nodes, one aspect we notice because only for 2,800 of the 209,203 starting points were routes to targets (petrol stations) identifiable using the OSM obtained out of the box. This behaviour could be corrected by connecting all dangling nodes in the network graph within a distance of 1 m, but at the cost of keeping information about dead ends between the nodes affected.

Another major drawback is the seemingly missing or insufficient standardization, respectively quality check, concerning the tags used for identifying/differentiating feature from one another³¹. This makes it quite difficult to determine all the lines to be included in the street network, always leaving some kind of uncertainty if something has been left out or added, and which may in consequence distort the network under consideration. In consequence, one gets the impression that the resulting street dataset represents a less exact model of reality than that based on the Esri StreetMap Premium data set.

Because of the difficulties in easily getting concise information about the data set and its use, but especially because of the fact that practically everybody is able to edit or change the dataset and that the dataset contains different regional states of data acquisition – raising questions about its reliability – using the data in scientific analysis causes some uneasiness due to the reliability of the analysis results.

Similar to the ESM dataset, because of topology inconsistencies that could not easily be compensated for 2,441 of the starting points under consideration (ca. 1.2 %), no routing could be performed based on the network. For these points the simple Euclidean distance was determined in order to compensate the missing values.

PostgreSQL/PostGIS

PostgreSQL with the PostGIS extension proved to be a very valuable tool for data preparation especially when huge amounts of data are to be analysed and edited. In contrast to other non open source GIS tools, PostgreSQL had no problems in handling quite

³¹ A good example is here the tag for the amenity „bank“ (financial institute) that is sometimes mixed up with the amenity „park bench“ which is, at least in the German OSM dataset sometimes also tagged as „bank“.

huge data sets. Nevertheless, dependent on the GIS-functions utilized, computing times for completing queries can be quite long (hours/days). For example it took approximately 12 hours to connect the raster centroids as well as petrol stations with the road network.

The accessibility calculation took approximately two weeks. So in order to effectively handle and analyse large sets of data it proved to be important to use indices as often as possible as this speeds up the data processing within PostgreSQL quite considerably. Provided that database, SQL and GIS basics are known, the specific PostGIS syntax for performing GIS analyses can be quite easily learned (cp. <http://trac.osgeo.org/postgis/> (13.05.2013)). Nevertheless random-access-memory (RAM) proved to be a limiting factor when using PostgreSQL with huge datasets and complex long-running queries. So it was not possible to perform the actual shortest path analysis with PostgreSQL tools alone.

Shortest path analysis

Although PostgreSQL/PostGIS can be extended with shortest-path routing functionalities with the pgRouting extension, it was decided to perform the routing outside the database via the Perl-Module „Graph-0.94. This proved to be necessary as calculating the shortest path for every start-point proved to be so memory-intensive that the analysis failed to complete within PostgreSQL and the hardware available because of memory shortage.

Therefore an external Perl-Script has been developed that is able to read the source, target and route information from the database, perform the catchment as well as network analysis and write back the analysis result to the database. Thereby, the Perl-Module „Graph-0.94“ turned out to be quite easy to use and offer a range of interesting graph analysis functions. One advantage of performing the routing via an external Perl script was the ability to use parallel processing abilities which proved to be quite handy when analysing, for example, different categories of petrol stations. Furthermore status messages could easily be implemented allowing the progress of the analysis to be followed.

Performance of OSM compared to ESM

The direct comparison of the accessibility values calculated based on ESM with those calculated based on OSM showed that the resulting values differ, partly considerably, from one another. Nevertheless, some

kind of a pattern seems to exist as the regions/cells with the greatest distance differences can be found in the federal states Mecklenburg-Vorpommern, Brandenburg, Thüringen and Sachsen-Anhalt, northern Bayern, western Rheinland-Pfalz, eastern Niedersachsen and the Alps. What does this all say about the usability of OSM data for rural studies?

Considering the findings of Ludwig et al (2010), as well as Zielstra and Zipf (2010), due to OSM completeness as well as the fact that tags for identifying features within the OSM dataset (e.g., residential streets, highways, etc.) are not really checked for quality, and the fact that the data set has to be topologically „corrected“ in order to be able to build a graph capable for routing at the cost of topological information (dead ends, bridges, tunnels) it can be assumed that the OSM still represents a less accurate model of reality than the ESM.

Therefore it can be assumed that the distances calculated based on OSM are less accurate than those based on commercial routing networks like for example ESM.

Considering the findings of Ludwig et al. (2010), as well as Zielstra and Zipf (2010) due to OSM completeness it is surprising that the accessibility values differ within urban as well as rural areas within the same range.

This finding was unexpected because we assumed that a clear distinction between urban and rural areas would exist, with greater differences in the accessibility values computed in rural areas. Nevertheless although great differences between the distances calculated could be found on a per cell comparison as well as on community level, aggregated average accessibility values obtained by ESM as well as OSM for greater aggregates like country, federal states or BBSR-Kreistypen 2009 proved to be comparable to each other.

The shortcomings of the OSM – especially the uncertainty about accuracy and completeness – together with the comparatively high learning level necessary to be able to use the OSM makes it, in my opinion, a less useful product within scientific studies focusing especially on rural areas, especially for small-scale analyses at the county level and below or the deduction of a small-scale accessibility indicator. But if one is not interested in the single cell values or small-scale analyses, the OSM-data proved to perform similarly to that of commercial alternatives. In such cases, the results suggest that the OSM may indeed provide an interesting low/no cost alternative.

-
- zur Bewertung und Analyse der Angebotsqualität im Öffentlichen Verkehr', in: „CORP 2003“ Computergestützte Raumplanung Beiträge zum 8. Symposium zur Rolle der Informationstechnologie in der und für die Raumplanung, Wien, 2003.
- Reudenbach C., Schulze U. (2010), GIS für Geographen - eine Einführung. <http://gisbsc.gis-ma.org/GISBScL2/de/text/GISBScL2.pdf>. (23.08.2011)
- Schulz, A.-C.; Bröcker, J. (2007), 'Die Erreichbarkeit der Arbeitsmärkte für Berufspendler aus den Gemeinden Schleswig-Holsteins', in: IAB regional, 1.
- Schürmann, C.; Spiekermann, K.; Wegener, M. (1997), 'Accessibility indicators', Berichte aus dem Institut für Raumplanung 39, Institut für Raumplanung, Dortmund.
- Schürmann, C. (2008), 'Berechnung verschiedener Erreichbarkeitsindikatoren für den Ostseeraum', http://www.brrg.de/_content/documents/publications/dak08_erreichbarkeit.pdf.
- Schwarze, B. (2005), 'Erreichbarkeitsindikatoren in der Nahverkehrsplanung', Institut für Raumplanung, Universität Dortmund – Fakultät Raumplanung, Arbeitspapier 184.
- Spiekermann, K.; Wegener, M. (2008), 'Modelle in der Raumplanung I. 5 Erreichbarkeit I', http://www.spiekermann-wegener.de/mir/pdf/MIR1_5_181108.pdf.
- Steinocher, K.; Köstl, M., Weichselbaum, J. (2011), 'Grid-based population and land take trend indicators – new approaches introduced by the geoland2 core Information Service for Spatial Planning', NTTs conference, Brussels, Feb 2011, CD-ROM.
- Streit, U. (2011), http://ifgivor.uni-muenster.de/vorlesungen/Geoinformatik/kap/kap4/k04_6.htm (18.04.2011).
- Winkel, R.; Greiving, S. Pietschman, H (2007), 'Sicherung der Daseinsvorsorge und Zentrale-Orte-Konzepte – gesellschaftspolitische Ziele und räumliche Organisation in der Diskussion. Stand der Fachdiskussion', BMVBS-Online-Publikation12/2010, http://www.bbsr.bund.de/cln_016/nn_21918/BBSR/DE/Veroeffentlichungen/BMVBS/Online/2010/ON122010.html (24.03.2011).
- Zielstra, A.; Zipf A. (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.