Investigations on Locational Accuracy of Volunteered Geographic Information Using OpenStreetMap Data

Marco Helbich, Christof Amelunxen, Pascal Neis and Alexander Zipf

GIScience, Department of Geography, University of Heidelberg, Berliner Strasse 48, 69120 Heidelberg, Germany. Correspondence: helbich@uni-heidelberg.de

In recent years, profound changes have taken place in Geographic Information Science. The formerly centralized data collection processes as well as its maintenance and distribution, mainly embossed by land surveying offices and commercial vendors, have changed dramatically. Nowadays, a massive increase of data provided by citizens, working in a collaborative fashion, is noticeable. Goodchild (2007) calls this "Volunteered Geographic Information (VGI)", which is furthermore one emerging characteristic of the Web 2.0.

Among a broad list of initiatives dealing with user generated geographic information, OpenStreetMap (OSM) is one of the most promising crowd sourced products. Its primary goal is to generate a free map of the world through volunteered participation. Everyone can use this data set for his or her purposes (e.g. OpenRouteService, Neis & Zipf 2008). However, using these data means accepting their limitations concerning spatial accuracy, which is affected by a technological bias (e.g. GPS-receiver), different data acquisition techniques (e.g. digitising), subjective knowledge about the data gathering process etc. The presence and utilization of such alternative data sources has far-reaching consequences and affects almost all sciences working with empirical data. For instance, missing and imprecise data effect model calibrations and in the worst case leads to false conclusions.

Nevertheless, research concerning different kinds of accuracies (e.g. positional, topological) of VGI has yet not gained much interest. A first descriptive attempt was conducted by Haklay (in press) who analyzed the positional accuracy of OSM compared to commercial data (OS Meridian 2) for the United Kingdom. Further research has addressed the spatial accuracy using OSM for geocoding purposes (Amelunxen 2010) or the completeness of OSM compared to Tele Atlas (TA, Zielstra & Zipf 2010). Generally, there is consensus about the usefulness of OSM, but a statistical analysis of the positional accuracy is still lacking so far.

This on-going research attends this issue and compares OSM, TA and official survey data from a city in Germany, as a case study. Therefore, highly accurate survey data serves as reference data, to which the other data sets are relatively evaluated.

Data sets from all three sources are preprocessed in a PostgreSQL/PostGIS relational database. Spatial SQL queries are used to locate and extract the positions of identical road junctions in either data set (i.e. two roads crossing each other at a distinct point coordinate). The deviation of the junction point coordinates compared with the corresponding points in the defined reference data set is then used as a first measure of positional accuracy. Based on this, a scatter diagram of positional errors is investigated to inspect errors. Subsequently, to gain knowledge about the distortion between the point patterns their global spatial geometry is analyzed in a bidimensional regression framework (Tobler 1994, Friedman & Kohler 2003). Finally, areas with high and low accuracy are detected with the G_i^* statistic (Getis & Ord 1992).

Preliminary results showed that both data sets, OSM and TA, have a highly locational accuracy. TA scattered clearly more westward around the true location of each road junction. However, the bidimensional regression estimates refered to highest correlation between OSM/TA and their true position, but TA data had less distortion than OSM. The G_i^* statistic resulted in some clusters with high an low positional accuracy, interpretable as spatial heterogeneity. Overall, OSM is a suitable alternative but to get confidence future research is needed, especially more case studies must be analyzed and other methodological approaches must be tested. Comparison between urban and rural areas seems fruitful, because rural areas are mapped with significantly less completeness (Zielstra & Zipf 2010).

References

Amelunxen, C. (2010), An Approach to Geocoding Based on Volunteered Spatial Data. In Zipf, A. et al.: Geoinformatik 2010. "Die Welt im Netz", 7–12.

Friedman, A. & Kohler, B. (2003), Bidimensional Regression: A Method for Assessing the Configural Similarity of Cognitive Maps and Other Two-Dimensional Data. *Psychological Methods, 8, 468–491.*

Goodchild, M. F. (2007), Citizens as Sensors: the World of Volunteered Geography. *GeoJournal, 69, 211-221.*

Getis, A. & Ord, J.K. (1992), The Analysis of Spatial Association by Use of Distance Statistics. *Geographical Analysis, 24, 189–206.*

Haklay, M. (in press), How Good is OpenStreetMap Information? A Comparative Study of OpenStreetMap and Ordnance Survey Datasets for London and the Rest of England. *Environment & Planning B.*

Neis, P. & Zipf, A. (2008), OpenRouteService.org is Three Times "Open": Combining OpenSource, OpenLS and OpenStreetMaps. GIS Research UK, Manchester.

Tobler, W. (1994), Bidimensional Regression. *Geographical Analysis, 26, 187–212.*

Zielstra, D. & Zipf, A. (2010), A Comparative Study of Proprietary Geodata and Volunteered Geographic Information for Germany. 13th AGILE International Conference on Geographic Information Science. Guimarães, Portugal.