

# Spatial-temporal Evolution and Completeness Analysis of OpenStreetMap Building Data in China from 2012 to 2017

Qi Zhou\*, YuanJian Tian

*School of Information Engineering, China University of Geosciences, Wuhan, Hubei, P.R. China*

\* Corresponding author, E-mail: sevenlaand@hotmail.com

**Keywords:** OpenStreetMap, China; Completeness, Evolution, Building data

## Abstract:

OpenStreetMap (OSM), as a typical volunteered geographic information project, is an online map with free content and everyone can edit and use it (Goodchild 2007). A range of applications has been proposed using OSM data, including routing and navigation, crisis mapping, 3D modelling, land use/cover mapping. This is because the OSM data is not only free of use, but also has a global coverage and high currentness. In despite of the above advantages, however, most of the OSM data were contributed by ‘non-professional’ or ‘amateur geographers’ (Goodchild 2008; Haklay 2010). Therefore, a lot of concerns have been paid attention to the quality issue of the OSM data. Assessing the quality of OSM data has become a hot topic in the field of geographic information science.

Extensive studies have been carried out on assessing various quality measures (e.g. positional accuracy, completeness and attribute accuracy) of OSM datasets in different countries or districts such as Germany, England, France, Italy, Canada and the United States. In the meanwhile, the road feature of an OSM dataset has been paid much attention to. To our knowledge, however, not any study has been focused on assessing the data quality of OSM building data in China, although it may be an essential data source for urban planning and management, 3D modelling and indoor navigation. Therefore, the aim of this study is to investigate the OSM building data in China. More precisely, an analysis of the spatial-temporal evolution and completeness of OSM building data in China from 2012 to 2017 was carried out. The tenet of our study was to employ two quality indicators, i.e. building count (Gröchenig et al. 2013, Barron et al. 2014, Fan 2016) and building density (Zhou 2018), for the analyses. First, the numbers of OSM building data from 2012 to 2017 were calculated in terms of both provincial- and prefecture-level divisions in China; The OSM building count were then compared among different divisions and also different years (2012-2017) for analyzing the evolution of OSM building data in China in both temporal dimension and spatial scale. Moreover, the correlations between OSM building counts and four potential factors (i.e. gross domestic product (GDP), population, urban land area, OSM road length), which may influence the development of OSM building data in China, were respectively investigated. Second, a  $1 \times 1$  km<sup>2</sup> regular grid was overlapped onto the OSM building datasets in urban areas for calculating the OSM building density of each grid cell; Moreover, high-density grid cells (whose OSM building data were almost complete) were extracted and analyzed through a simple clustering method, in order to investigate the spatial pattern of OSM building data in urban areas. Results showed that,

- 1) The OSM building data in China increased almost 20 times from the years 2012 to 2017, especially for those located in the eastern coastal regions of China (e.g. the provincial-level divisions: Jinagsu, Zhejiang, Guangzhou and Shandong and the prefecture-level divisions: Beijing, Nantong, Shanghai, Tianjin, Suzhou, Yangzhou and Dalian). In most cases, both the GDP and OSM road length factors had a moderate correlation with the OSM building count.
- 2) Most of the grid cells in urban areas still had no building or their building densities were equal to 0%, which indicated that the OSM building dataset in China was far from being complete. From analyzing the high density grid cells, two typical spatial distribution modes, i.e. dispersion and aggregation, were found in different prefecture-level divisions. As an example, the high-density grid cells for some prefecture-level divisions (e.g. Luoyang, Yueyang and Dalian) were mostly aggregated in the city cores; while those for some (e.g. Beijing Tianjin and Shanghai) were located in the hot spots such as business districts, attractions and transportation hubs.

The above results may benefit for users (especially those researchers and educators) to choose appropriate study area(s) from the OSM building dataset in China. In the meanwhile, the volunteers around the world may be motivated to contribute more OSM building data in this region. Further research work may include: developing quality indicators for quantitative completeness estimation of OSM building data, especially in rural areas; and investigating other quality measures (e.g. positional accuracy and semantic accuracy) or geographical features (e.g. railways, land uses, and points of interest) in China’s OSM dataset.

## **Acknowledgments**

This project was supported by the National Natural Science Foundation of China (Nos. 41771428 and 71874165), Fundamental Research Funds for the Central Universities, China University of Geosciences (Wuhan), and China Institute of Geo-Environment Monitoring (No. 0001212016CC60013). The authors want to thank the OpenStreetMap © contributors for making available as open their databases.