

Quality Assessment of the Contributed Land Use Information from OpenStreetMap Versus Authoritative Datasets

Jamal Jokar Arsanjani, Peter Mooney, Alexander Zipf and Anne Schauss

Abstract Land use (LU) maps are an important source of information in academia and for policy-makers describing the usage of land parcels. A large amount of effort and monetary resources are spent on mapping LU features over time and at local, regional, and global scales. Remote sensing images and signal processing techniques, as well as land surveying are the prime sources to map LU features. However, both data gathering approaches are financially expensive and time consuming. But recently, Web 2.0 technologies and the wide dissemination of GPS-enabled devices boosted public participation in collaborative mapping projects (CMPs). In this regard, the OpenStreetMap (OSM) project has been one of the most successful representatives, providing LU features. The main objective of this paper is to comparatively assess the accuracy of the contributed OSM-LU features in four German metropolitan areas versus the pan-European GMESUA dataset as a reference. Kappa index analysis along with per-class user's and producers' accuracies are used for accuracy assessment. The empirical findings suggest OSM as an alternative complementary source for extracting LU information whereas exceeding 50 % of the selected cities are mapped by mappers. Moreover, the results identify which land types preserve high/moderate/low accuracy across cities for urban LU mapping. The findings strength the potential of collaboratively collected LU

J. Jokar Arsanjani (✉) · A. Zipf · A. Schauss
GIScience Research Group, Institute of Geography, Heidelberg University,
69120 Heidelberg, Germany
e-mail: jokar.arsanjani@geog.uni-heidelberg.de

A. Zipf
e-mail: zipf@uni-heidelberg.de

A. Schauss
e-mail: anneschauss@gmail.com

P. Mooney
Department of Computer Science, Maynooth University,
Maynooth, Co. Kildare, Ireland
e-mail: peter.mooney@nuim.ie

features for providing temporal LU maps as well as updating/enriching existing inventories. Furthermore, such a collaborative approach can be used for collecting a global coverage of LU information specifically in countries in which temporal and monetary efforts could be minimized.

Keywords Land use features · Comparative assessment · Global monitoring for environment and security urban atlas (GMESUA) · OpenStreetMap · Confusion matrix

1 Introduction

The process of mapping land is known as LU mapping (Thenkabail et al. 2005) and land cover (LC) mapping (Kasetkasem et al. 2005), and is reflected in LU/LC maps. LU and LC maps are of great importance for many purposes concerning urban and regional planning, LU policy making, etc. In fact, these two concepts are distinct in essence, as LU maps illustrate human activities, such as artificial surface construction, farming, and forestry that represent the usage of land (Ellis 2007; Wästfelt and Arnberg 2013), whilst LC maps display the physical and biological cover over the land surface regardless of the purpose for which they are used (De Sherbinin 2002; Ellis 2007; Vaz et al. 2012, 2013). In other words, LC maps identify which land types cover the land and LU maps classify the land based on the usage of the land (Paneque-Gálvez et al. 2013; Sexton et al. 2013). For instance, if a particular land parcel is covered by grass, its LC type is labeled as grassland, whilst this parcel might be a part of a meadow LU class.

Employing signal processing algorithms on remote sensing data coupled with in-field measurements and ancillary data have been the main source of collecting LU and LC features (Kandrika and Roy 2008; Pacifici et al. 2009; Saadat et al. 2011; Qi et al. 2012). Remote sensing images and techniques often require in-field surveying for the results' validation process, i.e., in situ measurements as ground-truth data play a great role in delivering final products. Within the field-surveying data collection, experts and native residents' knowledge of the environment are needed to minimize uncertainty of measurements (Cihlar and Jansen 2001; De Leeuw et al. 2011). Contrary to LC mapping, LU mapping requires in-field information collection on the status and current usage of each land parcel, which are scarcely achievable from remote sensing images. Therefore, investigators must collect ancillary data as well to label LU patterns appropriately. Thus, LU mapping is even more complicated than LC mapping, and the information collected by experts from local residents, land managers, and evidence sources plays a vital role for accurate LU mapping (Fritz et al. 2012).

So far, a noticeable amount of efforts have been spent on generating LU maps at global, regional, and local scales. Examples of global scale and coarse resolution datasets comprise Global Land Cover (GLC2000) (Fritz et al. 2003), Moderate-resolution Imaging Spectroradiometer [MODIS; (McIver and Friedl 2002)], and

GlobCover (Arino et al. 2012), among others. At a regional (e.g., European) scale, the CORINE 2000 (Büttner et al. 2002) and Global Monitoring for Environment and Security Urban Atlas (GMESUA; Seifert 2009) deliver LC and LU maps at continental and municipal levels, respectively. High-resolution images including SPOT, Rapideye, and ALOS images have been utilized to attain fine-scale maps of urban areas delivering GMESUA (Kong et al. 2012), while a European coverage of LC maps at a coarse spatial resolution of 100 and 250 m has been provided in the CORINE dataset. But, their accuracies have been a major concern as outlined by Mayaux et al. (2006), Strahler et al. (2006), Herold et al. (2008), Fritz et al. (2012). In terms of accuracy, the lack of sufficient accuracy is even more critical in the case of global LU mapping, because the collection of globally-covered in-field information of LU features is such a huge task (Foody 2002; Foody et al. 2013). Thus, the necessity of finding an alternative and complementary approach for mapping LU features becomes evident. This could presumably be responded by the Web 2.0 innovations.

Lately, the development of Web 2.0 technologies has resulted in the emergence of a large number of CMP projects, which collect information about geographical objects from citizens. The majority of these CMPs offer very high-resolution satellite and aerial images (even less than one meter spatial resolution) through image libraries (e.g., Google Maps, Bing Maps) in their interfaces, which enable people to visualize the whole globe by fine-resolution remote sensing images so that they can map any features and additionally attach respective attributes to them (Rouse et al. 2007). This simple and straightforward way of visual interpretation of remote sensing images can be considered as alternative approach for LU mapping and even achieving fine-resolution LU maps at a global scale. The CMPs as listed in Sester et al. (2014) also provide people with some basic mapping tools in order to mark and digitize the visible objects. Some examples of CMPs are: OSM (Ramm et al. 2011), Geo-wiki (Fritz et al. 2012; Comber et al. 2013; See et al. 2013), Eye into Earth (Birringer 2008), and Wikiloc (Castelein et al. 2010). The individuals are also capable of enriching the attributes of the objects with some personal knowledge about these objects. The capability of importing the recordings of GPS-enabled devices i.e., smart phones and GPS devices, is granted to enable anyone to contribute even if s/he has minimal mapping expertise.

In brief, people interpret and integrate remote sensing images along with their personal information and their GPS-enabled device records. This sort of information has been called Volunteered Geographic Information (VGI): (Goodchild 2007). Among the CMPs, OSM is a unique platform in collecting LU features, because OSM has been so far a pioneer CMP due to attracting a huge amount of public attention and contributions (Ramm et al. 2011) by having almost 1.9 million users until December, 2014 and continue to grow as outlined by Jokar Arsanjani et al. (2015a, 2015c). More interestingly, OSM is highly democratic in receiving contributions by enabling anyone to edit/modify the existing features and sharing the whole data history freely and openly with the public in a structured way (Flanagin and Metzger 2008; Koukoletsos et al. 2012). It should be noted that OSM collects spatial information in GIS vector formats such as points, polylines, and polygons depending on the type of objects and presents them through a number of organized tags as listed in Ramm (2014).

A literature review reveals that in contrast to extensive analysis of road networks in OSM (Ludwig et al. 2011; Mooney and Corcoran 2012), first attempts at analyzing LU features from the whole OSM datasets have been carried out by Hagenauer and Helbich (2012), Jokar Arsanjani et al. (2013) in which they tried to extract LU features from the shared objects. However, except (Jokar Arsanjani and Vaz 2015b) virtually no studies on comparative assessment of the OSM-LU features with other authoritative datasets have been ever published. Therefore, the idea of using the contributed LU features to OSM arises in order to see how suitably we can collect temporal LU features at a local scale from OSM or even exploit the contributed features for updating the current LU datasets. To do so, a comparative quality assessment analysis must be carried out to gain some insights about it. To conclude, statements about the suitability of voluntarily collected LU data still remain highly speculative and even less is known whether these data might identify mismatches or even complement authoritative LU datasets.

In addition to that, spatial heterogeneity in the data quality increases the complexity of comparative studies as proven by Haklay (2010), Helbich et al. (2012), Koukoletsos et al. (2012). Hence, this research intends to evaluate the quality of OSM-LU features compared to a recent pan-European LU dataset, namely GME-SUA, as a reference, in order to find out how accurate LU features are attributed across four different German metropolitan areas. Besides preparing a LU dataset from OSM contributions, this study aims to cross-compare the degree of completeness and the attribute accuracy of the OSM-based LU features with the GMESUA data by means of a statistical assessment. To be more precise, this research seeks to find out: (a) how complete LU features are contributed to OSM, (b) how well OSM-LU features are attributed, (c) whether or not OSM-LU features are already usable for LU mapping, and (d) how effective the use of OSM data for questions in LU science would be.

The remainder of the paper is structured as follows: an overview of the utilized datasets and the chosen study sites are given in Sect. 2. Section 3 introduces the applied method while key results are presented in Sect. 4. Finally, Sect. 5 draws discussions and conclusions and Sect. 6 provides some recommendations.

2 Materials and Data Processing

2.1 OSM Dataset

The first datasets utilized in the present study is the OSM snapshot for November 5, 2013. The features tagged with “Natural” describe a wide variety of physical features, which are categorized into different categories such as water bodies, forest, etc. as described in Ramm (2014). “Land use” is the human use of land, which represents the purpose a land parcel is being used for (Ramm 2014). To extract relevant LU features, objects labeled with the tags “Land use” and “Natural” are exported from the OSM planet file into a uniform dataset.

2.2 *GMESUA Dataset as a Reference Dataset*

The second dataset, serving as reference data, is the pan-European GMESUA dataset, which comprises LU data for selected metropolitan areas exceeding 100,000 inhabitants. It is adapted to European needs, and contains information that can be derived chiefly from Earth Observation (EO) data supported by other reference data such as commercial-off-the-shelf (COTS) navigation data and topographic maps. Its minimum mapping unit (MMU) is between 0.25 and 1 ha, and a minimum width of linear elements of 100 m with ± 5 m positional accuracy is applied (European Union 2011). Additionally, some complementary data are integrated to improve the accuracy of classification processes namely (a) COTS navigation data such as POIs, LU, LC, water bodies; (b) Google Earth for interpretation; (c) local city maps for certain classes; (d) local zoning data such as cadastral data; (e) field checks (on-site visits); and (f) high-resolution satellite images (finer than 1 m ground resolution) (European Union 2011). At the time of writing this paper, this dataset covers 305 urban regions within Europe. The thematic accuracy for all classes is above 80 %. For more details see the Urban Atlas mapping guide (European Union 2011). Table 1 represents the defined classes in GMESUA and in this article these classes will be recalled by their codes as well e.g., Isolated structures [113].

2.3 *Study Areas*

Four large metropolitan areas of Germany, from different regions, are selected: Berlin, Frankfurt am Main, Munich, and Hamburg. There are multiple reasons for choosing these areas in Germany. First, the OSM community in Germany is very active and dynamic, and therefore it is rational to begin from potentially well-mapped areas. Secondly, no bulk import of authoritative datasets for Germany into the OSM database has yet been reported. Thirdly, according to osmatrix.uni-hd.de (Roick et al. 2011) and (Jokar Arsanjani et al. 2014) these cities have received high rates of contributions. Fourthly, the reference dataset i.e., GMESUA for these cities are available. The selected areas cover approximately 35,000 km² and contain 15 major LU classes, so a wide variety of LU features from heterogeneous areas are identified. Both the selected areas and the input data are shown in Fig. 1.

3 **Methods**

In geodata quality analysis, the quality of geodata should be internally and externally considered (van Oort 2006; Gervais et al. 2009). Internal quality reflects the data production specifications, which recognizes errors in the data. The major standard organizations (e.g., ISO 19157, ICA, FGDC, and CEN) have introduced

Table 1 Classification scheme applied in the preparation of GMESUA datasets

Classification level	Level 1	Level 2	Level 3
Land	Artificial surfaces [100]	Urban fabrics [110]	Continuous urban fabrics [111]
			Discontinuous urban fabrics [112]
			Isolated structures [113]
		Industrial, commercial, public, military, private and transport units [120]	Industrial, commercial, public, military and public units [121]
			Road and rail network and associated lands [122]
			Port areas [123]
			Airports [124]
		Mine, dump and construction sites [130]	Mineral extraction and dump sites [131]
			Construction sites [132]
			Land without current use [133]
Artificial nonagricultural vegetated areas [140]	Green urban areas [141]		
	Sports and leisure facilities [142]		
	Agricultural + seminatural areas + wetlands [200]	–	–
	Forests [300]	–	–
Water	Water [500]	–	–

their diverse quality criteria and the followings five are common amongst them: (1) thematic accuracy, (2) positional accuracy, (3) temporal accuracy, (4) logical consistency, and (5) completeness (Guptill and Morrison 1995). These data properties are introduced to the users through metadata files attached to datasets by producers (Devillers et al. 2007). On the other hand, external quality reflects the suitability of a dataset for a particular purpose and addresses the concept of “Fitness of Use” (FoU): (Guptill and Morrison 1995; Devillers et al. 2007). In this study, the internal aspects of data quality are considered.

The workflow of evaluating the OSM-LU dataset is summarized in Fig. 2 and described as follows. First, OSM features tagged with “Land use” and “Natural” are retrieved and merged together to result in a unique dataset. Second, overlaps and topological errors between the features are then resolved, which is described in details

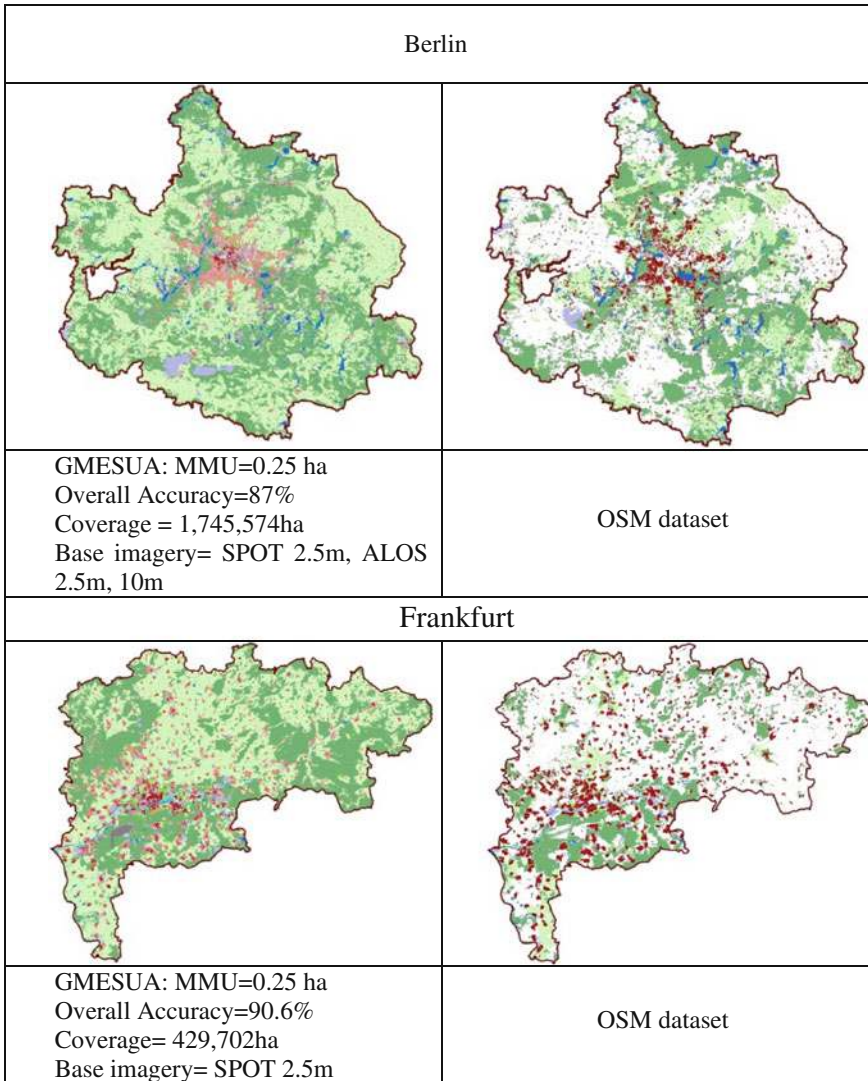


Fig. 1 The physical extent of the selected cities—Berlin, Frankfurt, Hamburg, and Munich, represented by the GMESUA datasets accompanied with their metadata (*left panels*) and contributed OSM features (*right panels*)

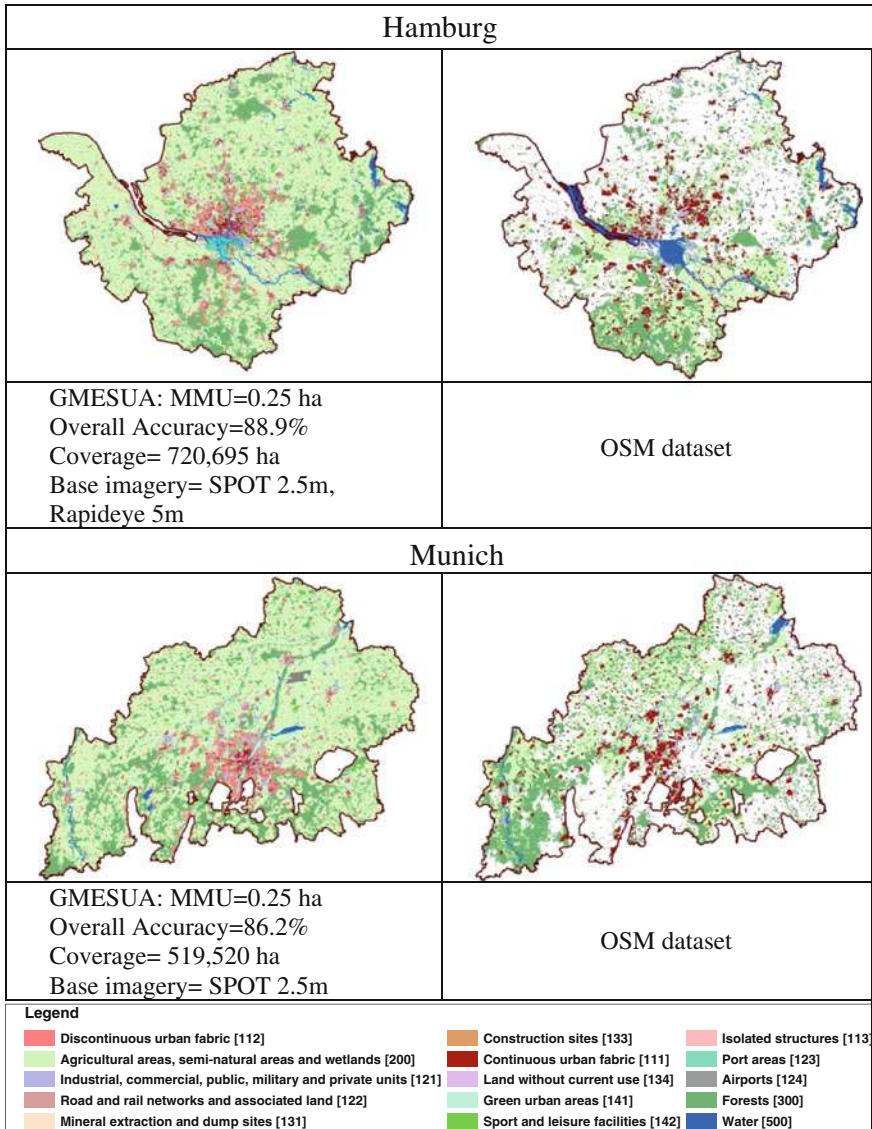


Fig. 1 (continued)

in Sect. 3.1. Third, the OSM features are re-labeled and matched according to the GMESUA nomenclature as will be explained in Sect. 3.2. Fourth, the degree of completeness for each city is determined to measure how much of the area is mapped. Finally, an error matrix between the OSM and GMESUA datasets is computed to measure the overall thematic accuracy of the OSM features along with a detailed per-class analysis accompanied with a map of agreement/disagreement values.

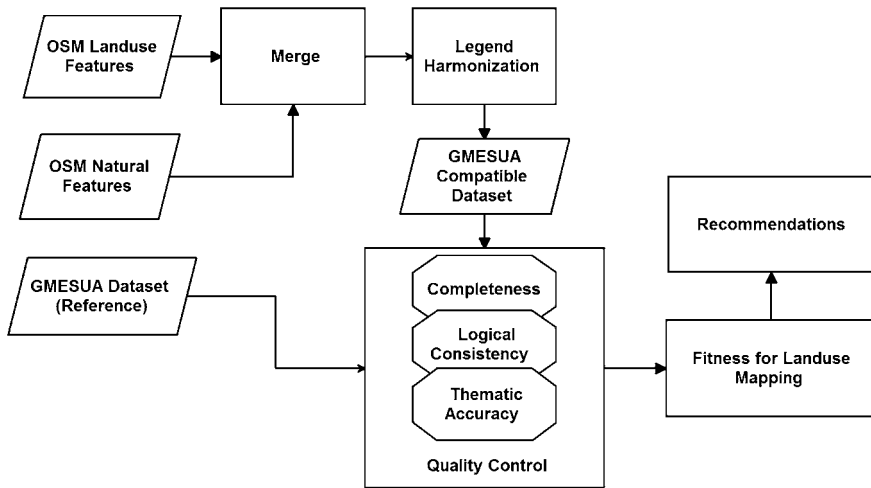


Fig. 2 The flowchart of evaluating OSM land use features

3.1 Logical Consistency and Topology

Logical consistency addresses how well logical relationships between the elements of the dataset are defined. If the objects are not topologically defined to each other, the dataset will fail in having proper internal relation between the objects (van Oort 2006). This issue is even more problematic when the features are polygons. There is no indicator to measure it quantitatively and in the metadata attached to the data is indicated with a Boolean value whether the data set has been cleaned from topological errors or not (Devillers et al. 2007). This concern is a challenge for the collected OSM polygon features, because the OSM contributions are mapped at different zoom levels, which result in dissimilar data scales and some features might overlap each other and consequently some areas possess more than one label (Sester et al. 2014).

However, depending on the type of data, the degree of goodness varies; for instance, this problem is not encountered for point datasets e.g., POIs. In the case of polyline datasets, such topological inconsistency can be observed at the road junctions as well as at the beginning and ending of the road segments. This issue is schematically illustrated in Fig. 3. Therefore, using this layer for any external application generally demands applying topological cleaning of features in order to clean them from overlaps and dangle errors. These issues are resolved automatically by applying topology for removing errors of our dataset including unclosed gaps, gaps between polygons, and overlapping polygons.

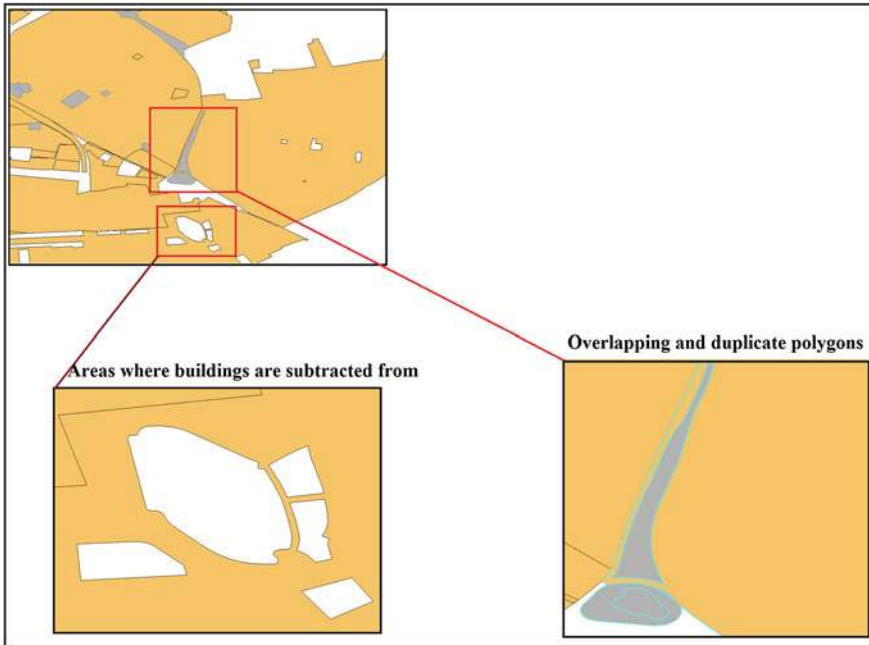


Fig. 3 A sample representation of duplicate polygons (*right-down*) and subtracted areas from land use dataset into building dataset (*left-down*)

3.2 Harmonization of the Datasets Nomenclatures

The OSM has its own classification nomenclature, which is designed for fine-scale LU classification, therefore, the OSM-LU features must be translated to globally known nomenclatures and in this case, the GMESUA nomenclature. This was carried out through adding an additional label compatible with the first and third level of classification of the GMESUA legend for artificial surfaces and non-artificial surfaces, respectively (see Estima and Painho 2013; Jokar Arsanjani et al. 2013). This helps to make a common LU nomenclature and also creates a dictionary for translating the contributions of individuals to the GMESUA adjusted LU types. For instance, water-related features are assigned as water [500]. Table 2 represents an exemplary dictionary of translating OSM features to GMESUA nomenclature.

Several difficulties were encountered such as (i) semantic understanding of terms in domestic language i.e., German terms with typos which had to be translated into English, (ii) the use of unidentified types of features by contributors and the incomplete attributes of features, which were between 10 and 15 % of area and removed from the analysis. This implies that the handling of the heterogeneous and unstructured contributions cannot be automatically handled.

Table 2 Translation of OSM tags (origin) to GMESUA nomenclature (target)

Origin: OSM tags	Target: GMESUA classes [CODE]
Wohngebiet, residential, apartments, residential, dorfgemeinschaft_breitenfurt	Continuous urban fabric [111]
House, hut, villa	Discontinuous urban fabric [112]
Isolated	Isolated structures [113]
Warehouse, university, social, school, sauna, retail, religious, public, power_station, power, place_of_worship, palace, office, museum, mosque, manufacture, kindergarten, Industrial, hotel, hospital, historic, greenhouse, glasshouse, factory, embassy, commercial, clubhouse, club, cinema, café, allotments, cemetery, fortress, greenhouse_horticulture, industrial, military, nursery, wayside_shrine, ruins, monument, monastery, memorial, industrial, grave, city_gate, castle, archaeological_site, tank, water_tower, warehouse, temple, storage_tank, library, church, chapel, cathedral, castle	Industrial, commercial, public, military and private units [121]
Bridge, railway, traffic_island, bicycle_parking, bus_station, fuel, motorcycle parking, parking entrance, parking space, taxi	Road and rail networks and associated land [122]
Ship	Port areas [123]
Airfield, airport	Airports [124]
Wastewater_plant, coal_heap, landfill	Mineral extraction and dump sites [131]
Construction	Construction sites [133]
Collapsed, greenfield, brownfield	Land without current use [134]
Park, grass, nature_reserve, recreation_ground, recreation_ground, zoo	Green urban areas [141]
Swimming_pool, leisure, alpine_hut, artwork, camp_site, caravan_site, information, picnic_site, theme_park, trail_riding_station, viewpoint, swimming_pool	Sport and leisure facilities [142]
Village_green, vineyard, scrub, orchard, meadow, green, grassland, grass, farmyard, farmland, farm, agriculture, agricultural, hühnerfarm	Agricultural areas, semi-natural areas and wetlands [200]
Forest, wood	Forests [300]
Water_basin, water_protected area, pond	Water [500]

3.3 Completeness

According to the published literature (Koukoletsos et al. 2012; Hecht et al. 2013), completeness is the major criterion for using OSM datasets as it is an indicator of how much of the whole has been mapped. Although measuring the degree of completeness for polyline and point features in OSM datasets requires a reference dataset, the completeness for LU features within a certain area can be measured

even without having any reference dataset. Because every piece of the land should have an attribute, so the total coverage area is the maximum area to be mapped. The degree of completeness was measured by calculating a completeness index, which calculates the overall area (ha) mapped by contributors out of the whole area (ha) for each individual city.

3.4 Thematic Accuracy

The most important criterion to judge the quality of the contributed LU features is to find out how correctly the land parcels are attributed. This criterion i.e., thematic accuracy, is generally called “accuracy assessment” in the LU/LC classification literature (Congalton 1991; Foody 2002; Foody et al. 2013). The accuracy assessment reflects the difference between the target dataset and the reference dataset. The accuracy assessment process usually summarizes all data in a confusion matrix and reports several indicators such as “overall/per class accuracies”, “Kappa index of agreement”, “user’s accuracy” and “producer’s accuracy”. In this study, a confusion matrix analysis is applied to achieve these measures. These measures have been the most straightforward and practical statistical tools for checking the degree of match between two thematic datasets as outlined in (Foody 2002; Herold et al. 2008). A measure for the overall accuracy is calculated by dividing the number of identical pixels by the total number of pixels. However, it does not identify how well individual classes between the two datasets match. Hence, the user’s accuracy and producer’s accuracy should be calculated to measure the accuracy of each class. The user’s accuracy indicates the probability that a pixel from the OSM-LU map actually matches the GMESUA dataset, while the producer’s accuracy refers to the probability that a specific LU type from the reference dataset is classified as such. These two measurements, typically for any given LU type, are not equal. For instance, if for a specific land type of ‘A’, with accuracies achieved of 89 and 78 % for user’s accuracy and producer’s accuracy respectively, it implies that as a user of the data, roughly 89 % of all the pixels classified as A are the same in the reference dataset and, as a producer, only 78 % of all A pixels are classified as such.

4 Results

4.1 Sensitivity to Pixel Size

The Kappa index proposed by Cohen (1960) intends to evaluate the degree of agreement between two or more datasets/observations and consequently provides an overall guide to quality of the map (see Landis and Koch 1977; Foody 2002).

Table 3 Kappa index analysis of the contributed land use features in comparison with GMESUA dataset at different pixel sizes for each city

City	Kappa index (pixel size in meter)				
	5	10	15	20	50
Frankfurt	0.36	0.362	0.361	0.361	0.36
Hamburg	0.402	0.402	0.408	0.403	0.402
Berlin	0.518	0.518	0.525	0.521	0.518
Munich	0.453	0.452	0.455	0.452	0.451

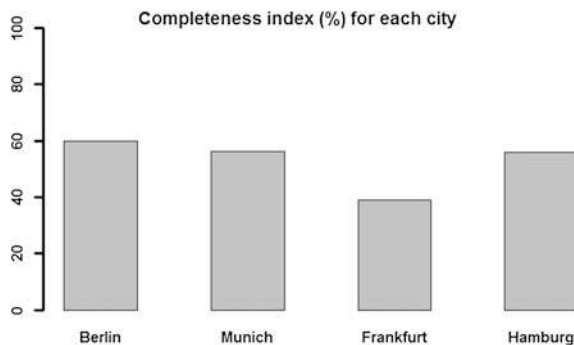
The datasets were converted from vector to raster format at different pixel sizes smaller than 50 m i.e., 5, 10, 15, 20, 30, and 50 m in order to find the most optimal pixel size. This analysis is done and represented in Table 3.

Table 3 demonstrates the computed Kappa indices of agreement between the two datasets at different pixel sizes. The pixel sizes between 10 and 15 m result in a slightly higher degree of match. Consequently, the confusion matrices for the selected cities were designed for the most optimal pixel size applicable for every city at 15 m.

4.2 Degree of Data Completeness

This measure for each city is shown in Fig. 4, which indicates that Berlin has reached the highest degree of completeness, and Frankfurt has the lowest degree of completeness. This value is much higher than the reported value for the Portugal continent at 3 % by Estima and Painho (2013) and confirms how greatly completeness index varies and how heterogeneous the quantity of contributions is. It should be noted that the selected areas consist of urban and rural areas, and the measured completeness index measures the degree of completeness for both regions in which the contributions in urban areas are relatively more than rural areas as illustrated in Fig. 4.

Fig. 4 Calculation of completeness index for each city



4.3 Overall and Per-class Analysis of Thematic Accuracy

As mentioned earlier, the GMESUA dataset is considered as the reference dataset and it will be cross compared with the OSM-LU dataset. In addition to calculating the Kappa index, the overall accuracy of the OSM-LU features, as well as user's and producer's accuracies, are calculated in order to discuss how LU types in each city are contributed as shown in Table 4. Due to heterogeneous accuracies across cities, interpretation of the confusion matrices is discussed for each city separately as follows.

4.3.1 Frankfurt

According to Table 4, among the selected cities the highest Kappa index of 56 % and overall accuracy of 76.5 %. This means that the OSM-LU map and GMESUA map match at a “moderately” rank according to Landis and Koch (1977). Per-class analysis of user's and producer's accuracies reveals that although roughly 98 % of Continuous urban fabric [111] and Water [500] classes have been correctly identified as such, only 20 % of the areas labelled as Continuous urban fabric [111] are actually Continuous urban fabric [111], while 88 % of the areas labelled as Water [500] are actually water [500]. Furthermore, while 71, 62, and 60 % of Forests [300], Industrial, commercial, public, military and private units [121], Agricultural areas + semi-natural areas + wetlands [200] classes have been correctly recognized as such, these classes have been mapped correctly at 94, 77, and 73 % rates. To sum up, the achieved user's accuracies confirm that the Forests [300] and Water [500] classes have a “very high” degree of accuracy with the reference data, while classes like Industrial, commercial, public, military and private units [121], Road and rail networks and associated land [122], and Agricultural areas + semi-natural areas + wetlands [200] are classified as “high” degree. Therefore, it could be concluded that these five classes could be of used for LU mapping purposes at a relatively good level of reliability. Despite the high value of producer's accuracy of Continuous urban fabrics [111] class, its low user's accuracy value (20.3 %) confirms that this class retains as those which are not reliable. On contrary, the contributions to the remaining classes confirm disagreements between the two data sources.

4.3.2 Munich

A Kappa index of 46 % ranked as “moderately” and overall accuracy of 67.1 % for contributed features in Munich are achieved according to Table 4. This means that 67.1 % of contributions are correctly classified. Analysis of the achieved per-class user's accuracies reveals that roughly 100 and 96 % of the Isolated structures [113] and Forests [300] classes have been correctly labelled by contributors, which are ranked as “very high” by Landis and Koch (1977). Moreover, classes such as Industrial, commercial, public, military and private units [121], Agricultural

Table 4 The values of confusion matrices between OSM and GMESUA land use features per city

Land class	City											
	Berlin			Munich			Frankfurt			Hamburg		
	Producer's accuracy (%)	Accuracy (%)	Producer's accuracy (%)	Accuracy (%)	Producer's accuracy (%)	Accuracy (%)	Producer's accuracy (%)	Accuracy (%)	Producer's accuracy (%)	Accuracy (%)	Producer's accuracy (%)	Accuracy (%)
Continuous urban fabric [111]	93.7	5.9	93.2	7.9	98	20	66	11				
Discontinuous urban fabric [112]	0	0	0	0	0	0	0	0				
Isolated structures [113]	0.1	52.4	0	100	0	0	0	0				
Industrial, commercial, public, military and private units [121]	40.8	50.8	56.6	69	62	78	51	66				
Road and rail networks and associated land [122]	4.2	66.3	6.5	66	11	72	7.4	79				
Port areas [123]	0.0	77	0.0	0.0	0.0	0.0	0.0	0.0				
Mineral extraction and dump sites [131]	19.1	0	68.3	0	57	0	61	66				
Construction sites [133]	4.4	5.9	0.0	0.0	0.0	0.0	12	22				
Land without current use [134]	1.7	0.1	2.2	5.3	0.7	1.2	2.3	4.2				
Green urban areas [141]	6.3	41.3	0.8	2	3	4.2	52	7.1				
Sport and leisure facilities [142]	54.3	68.9	8.5	49	11	40	36	83				
Agricultural areas, semi-natural areas and wetlands [200]	76.3	94.6	45	68	61	74	70	95				
Forests [300]	94.2	81.1	69.1	96	71	94	90	71				
Water [500]	84.5	86.1	97.7	60	98	89	87	71				
Overall accuracy (%)	75.9		67.1		76.5		63.9					
Index (%)	52.3		46		56.2		41					

areas + semi-natural areas + wetlands [200], Road and rail networks and associated land [122], and Water [500] are positioned at the second rank i.e., “high” in this city with values of roughly 69, 67, 66, and 60 %, respectively. The remaining classes possess a low level of reliability to be used for LU mapping, so that they are not recommended for usage.

4.3.3 Berlin

Based on the presented values in Table 4 for Berlin, a Kappa index of 52 % ranked as “moderate” and an overall accuracy of 75.9 % is measured. Per-class analysis of the user’s accuracies shows that roughly 94, 86, and 81 % of the Agricultural areas + semi-natural areas + wetlands [200], Water [500], and Forests [300] classes have been correctly labelled by contributors, while classes such as Port areas [123], Sport and leisure facilities [142], and Road and rail networks and associated land [122] have been mapped at 77, 69, and 66 % rates of accuracy, respectively. Therefore, the remaining classes retain at moderate to low level of agreement with the reference dataset.

4.3.4 Hamburg

According to Table 4, among the selected cities the lowest Kappa index of 41 % and overall accuracy of 63.9 % are achieved, which means the mapped LU features to OSM for Hamburg match at a “moderately” rank with the GMESUA dataset. Analysis of the achieved per-class user’s accuracies reveals that Agricultural areas + semi-natural areas + wetlands [200], Sport and leisure facilities [142] classes have been mapped at approximately 94 and 83 % rate of match i.e., “very high” with the reference dataset, which are followed by Road and rail networks and associated land [122], Water [500], Forests [300], Industrial, commercial, public, military and private units [121], and Mineral extraction and dump sites [131] classes ranked as “high”. The remaining classes lack of sufficient accuracy.

4.4 *Spatial Distribution of Agreements and Disagreements*

Spatial distribution of agreement and disagreement between the OSM and reference datasets is visualized in Fig. 5. Pink pixels identify areas where the two datasets agree on having the same land type whilst blue pixels indicate areas where the two datasets represent dissimilar LU types. In each city map, a black bounding box marks the clustered urban areas of each city. Generally speaking, depending on the city the Isolated structures [113], Industrial, commercial, public, military and private units [121], Road and rail networks and associated land [122], Sport and leisure facilities [142], Agricultural + semi-natural + wetlands [200], Forests [300], and Water [500] contain the highest level of agreement between the two datasets for

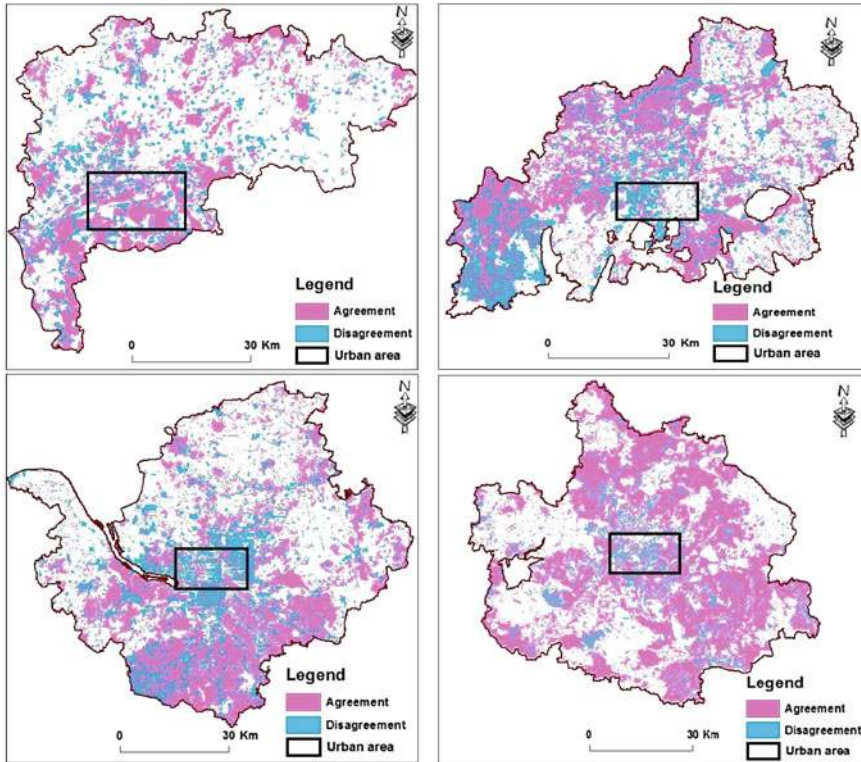


Fig. 5 Spatial distribution of agreement and disagreement between OSM land use features and GMESUA dataset for Frankfurt (*top-left*), Munich (*top-right*), Hamburg (*down-left*), and Berlin (*down-right*). White areas indicate unmapped areas

all cities. In contrast, the remaining classes have a large lack of agreement, assuming that they are correctly reflected in the reference dataset. However, it is an invalid statement to relate these mismatches with the certainty of OSM features as they might represent the correct information. Visual analysis of Fig. 5 reveals that in urban areas of Hamburg and Munich the amount of disagreement increases, while in Berlin and Frankfurt a mixture of both are existent. Furthermore, it is concluded that the contributed OSM-LU features are heterogeneously distributed over inside/outside urban areas, however, the density in urban areas is higher.

5 Discussions and Conclusions

The recent rapid emergence of online CMPs, such as OSM, has attracted large numbers of individuals to share their personal knowledge, as well as records from their GPS-enabled devices, with the public. This bottom-up process of collecting

individuals' contributions has resulted in the generation of a tremendous amount of geolocated information. This new type of user generated geoinformation has been leveraged new applications such as indoor mapping (Goetz and Zipf 2010), routing applications (Bakillah et al. 2014a) and tourism recommendations. Environmental monitoring through extracting LU features from the VGI are still few. While (Hagenauer and Helbich 2012; Jokar Arsanjani et al. 2013) extracted LU information from VGI at acceptable level of accuracy, the capacities to integrate VGI into LU science have not yet been fully discovered. The findings from the Geo-wiki project in terms of users engagement and reliability of crowdsourced information reveal that although the crowdsourced information from users are reliable enough, the number of involved users is not significant. Although the question on how to attract users and how to keep them active in the crowdsourcing process is still unanswered, OSM has so far attracted around 1.7 million users. More importantly, quality analysis of the contributed streets network (Hochmair et al. 2014), buildings footprints (Fan et al. 2014), and POIs (Bakillah et al. 2014b) in OSM proves the promising avenues of using OSM features for multiple applications. Therefore, a huge potential in OSM exists to be explored. Thus, in this study, we comparatively evaluated the completeness and thematic accuracy aspects of the contributed OSM-LU features in four large German metropolitan areas (Berlin, Frankfurt, Hamburg, and Munich) and see how reliable we could start using them.

Some of the lessons learned from this investigation are as follows. The analysis of completeness index reveals that, between 40 and 60 % of the selected areas are mapped. From a logical consistency perspective, in some cases overlapping polygons and topological issues are found that cause additional data processing. Furthermore, the existing differences in LU legends highlight an obvious inconsistency that makes harmonizing LU datasets difficult. From a thematic accuracy viewpoint, the contributed features have, in general, a "moderate" rank of Kappa indices and their overall accuracies are between 63 and 77 %. It must be noted that the overall accuracy of GMESUA datasets barely exceeds 90 % and, therefore, the computed accuracies are noticeable at the current stage of OSM. Per-class analysis of the LU types show that, depending on the city, Isolated structures [113], Industrial, commercial, public, military and private units [121], Road and rail networks and associated land [122], Sport and leisure facilities [142], Agricultural + semi-natural + wetlands [200], Forests [300], and Water [500] reach substantial level of accuracies, which implies that these classes are highly exploitable. It is worth mentioning that integrating ground-truth information with other reference data for accuracy assessment could be highly beneficial for producing hybrid LU products.

According to European Union (2011), archived images from 2005 until 2010 have been used for LU mapping and this could have caused a major source of disagreement, whereas the OSM-LU contributions have mainly been uploaded after 2009, and therefore, some information from OSM might be even more accurate than the reference data. Moreover, the MMU of the GMESUA datasets is 0.25–1 ha and, therefore, land parcels smaller than this MMU are neglected from mapping, while in OSM even smaller parcels are mapped, i.e., a smaller MMU in OSM is achieved. This means that in some parts while a polygon in GMESUA dataset is

representing a specific LU type, the same area in OSM-LU dataset is covered by multiple small polygons representing different polygons. From the mappers' viewpoint, the citizens' perception of LU types should be further investigated to see how people visually interpret LU types.

As a final conclusion, the OSM-LU features suggest a promising input source of updating LU inventories and also LU mapping. Certainly, the longer OSM is in existence, more contributions will be collected and the higher accuracy LU maps can be retrieved.

6 Recommendations

This study attempts to draw some recommendations that will lead future research possibly in the interesting directions. The following recommendations are proposed to the environmentalists and LU researchers based on the findings in this paper. According to the indicated completeness indices of the four cities, as well as the accuracy statistics, the contributed LU features to OSM in the selected areas account for a potential alternative data source for mapping LU features. However, further studies on other areas must be conducted to see the heterogeneity of completeness and thematic accuracy across space. Furthermore, applying data mining techniques as well as data fusion with other available datasets GMESUA for extracting the LU features of incomplete areas are of high importance. Additionally, the land types with the highest accuracies can be separately incorporated for respective applications.

This investigation of the OSM-LU features will be posted on the OSM blogs, and mailing and discussion lists, to inform the OSM community about current academic progress in the area of LU features. This enables experts to: (a) possibly find ways to draw the attention of mappers to LU features, (b) determine the possible existing problems with the OSM ontology of the LU dataset, or (c) it is the case that users are not able to add further features in the urban areas, because the volume of mapped objects (e.g., POIs, roads, building, etc.) do not let users to add new features including LU polygons, or (d) ultimately the LU attributes are not easily understandable for people and they have insufficient interest in mapping them.

Acknowledgments Jamal Jokar Arsanjani acknowledges the funding of the Alexander von Humboldt foundation.

References

- Arino O, Ramos Perez JJ, Kalogirou V, Bontemps S, Defourny P, Van Bogaert E (2012) Global land cover map for 2009 (GlobCover 2009)
- Bakillah M, Lauer J, Liang SHL, Zipf A, Jokar Arsanjani J, Mobasheri A, Loos L (2014a) Exploiting big VGI to improve routing and navigation services. *Big Data Tech Technol Geoinformatics* 177–192

- Bakillah M, Liang S, Mobasheri A, Jokar Arsanjani J, Zipf A (2014b) Fine-resolution population mapping using OpenStreetMap points-of-interest. *Int J Geogr Inf Sci* 28: 1940–1963
- Birringer J (2008) Eye into Earth. *Space Culture* 11:59
- Büttner G, Feranec J, Gabriel J (2002) Corine land cover update 2000
- Castelein W, Grus L, Crompvoets J, Bregt AA (2010) Characterization of volunteered geographic information. In: 13th AGILE international conference on geographic information science 2010, Guimarães, pp 1–10
- Cihlar J, Jansen LJM (2001) From land cover to land use: a methodology for efficient land use mapping over large areas. *Prof Geogr* 53:275–289
- Cohen J (1960) A coefficient of agreement for nominal scales. *Educ Psychol Meas* 20:37–46
- Comber A, See L, Fritz S, Van der Velde M, Perger C, Foody G (2013) Using control data to determine the reliability of volunteered geographic information about land cover. *Int J Appl Earth Obs Geoinf* 23:37–48
- Congalton RG (1991) A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sens Environ* 37:35–46
- De Leeuw J, Said M, Ortegah L, Nagda S, Georgiadou Y, DeBlois M (2011) An assessment of the accuracy of volunteered road map production in western Kenya. *Remote Sens* 3:247–256
- De Sherbinin A (2002) A CIESIN thematic guide to land land-use and land land-cover change (LUCC), NY, pp 10–20
- Devilleers R, Bédard Y, Jeansoulin R, Moulin B (2007) Towards spatial data quality information analysis tools for experts assessing the fitness for use of spatial data. *Int J Geogr Inf Sci* 21:261–282
- Ellis E (2007) Land-use and land-cover change. *Earth*
- Estima J, Painho M (2013) Exploratory analysis of OpenStreetMap for land use classification. In: Proceedings of the second ACM SIGSPATIAL international workshop on crowdsourced and volunteered geographic information GEOCROWD'13, ACM, New York pp 39–46
- European Union (2011) Mapping guide for a european urban atlas
- Fan H, Zipf A, Fu Q, Neis P (2014) Quality assessment for building footprints data on OpenStreetMap. *Int J Geogr Inf Sci* 28:700–719
- Flanagin AJ, Metzger MJ (2008) The credibility of volunteered geographic information. *GeoJournal* 72:137–148
- Foody GM (2002) Status of land cover classification accuracy assessment. *Remote Sens Environ* 80:185–201
- Foody GM, See L, Fritz S, Van der Velde M, Perger C, Schill C, Boyd DS (2013) Assessing the accuracy of volunteered geographic information arising from multiple contributors to an internet based collaborative project. *Trans GIS* 17(6):847–860
- Fritz S, Bartholomé E, Belward A, Hartley A, Stibig HJ, Eva H, Mayaux P, Bartalev S, Latifovic R, Kolmert S et al (2003) Harmonisation, mosaicing and production of the global land cover 2000 database (beta version). Office for Official Publications of the European Communities Luxembourg, Luxembourg
- Fritz S, McCallum I, Schill C, Perger C, See L, Schepaschenko D, van der Velde M, Kraxner F, Obersteiner M (2012) Geo-Wiki: an online platform for improving global land cover. *Environ Model Softw* 31:110–123
- Gervais M, Bédard Y, Levesque M, Bernier E, Devillers R (2009) Data quality issues and geographic knowledge discovery. *Geogr Data Min Knowl Discov* pp 99–115
- Goetz M, Zipf A (2010) Extending OpenStreetMap to indoor environments : bringing volunteered geographic information to the next level. CRC Press, Delft
- Goodchild MF (2007) Editorial: citizens as voluntary sensors: spatial data infrastructure in the world of web 2.0 2: 24–32
- Guptill SC, Morrison JL (1995) Elements of spatial data quality. Elsevier Science, Oxford
- Hagenauer J, Helbich M (2012) Mining urban land-use patterns from volunteered geographic information by means of genetic algorithms and artificial neural networks. *Int J Geogr Inf Sci* 26:963–982

- Haklay M (2010) How good is volunteered geographical information? A comparative study of OpenStreetMap and ordnance survey datasets. *Environ Plan B Plan Des* 37:682–703
- Hecht R, Kunze C, Hahmann S (2013) Measuring completeness of building footprints in OpenStreetMap over space and time. *ISPRS Int J Geo-Information* 2:1066–1091
- Helbich M, Amelunxen C, Neis P (2012) Comparative spatial analysis of positional accuracy of OpenStreetMap and proprietary Geodata. In: *International GI_Forum, Salzburg*
- Herold M, Mayaux P, Woodcock CE, Baccini A, Schmullius C (2008) Some challenges in global land cover mapping : an assessment of agreement and accuracy in existing 1 km datasets. *Remote Sens Environ* 112:2538–2556
- Hochmair HH, Zielstra D, Neis P (2014) Assessing the completeness of bicycle trail and lane features in OpenStreetMap for the United States. *Trans GIS*. doi:10.1111/igis.12081
- Jokar Arsanjani J, Helbich M, Bakillah M, Loos L (2015a) The emergence and evolution of OpenStreetMap: a cellular automata approach. *Int J Digital Earth* 8(1):74–88. <http://www.tandfonline.com/doi/abs/10.1080/17538947.2013.847125>
- Jokar Arsanjani J, Vaz E (2015b) An assessment of a collaborative mapping approach for exploring land use patterns for several European metropolises. *Int J Appl Earth Obs Geoinf* 35:329–337
- Jokar Arsanjani J, Helbich M, Bakillah M, Hagenauer J, Zipf A (2013) Toward mapping land-use patterns from volunteered geographic information. *Int J Geogr Inf Sci* 27:2264–2278
- Jokar Arsanjani J, Mooney P, Helbich M, Zipf A (2015c) An exploration of future patterns of the contributions to OpenStreetMap and development of a Contribution Index, *Trans GIS*
- Jokar Arsanjani J, Vaz E, Bakillah M, Mooney P (2014) Towards initiating OpenLandMap founded on citizens' science: the current status of land use features of OpenStreetMap in Europe. In: Huerta Schade G (ed) *Proceedings of the AGILE'2014 international conference on geographic information science*, 3–6 June 2014, AGILE digital editions, Castellón
- Kandrika S, Roy PSS (2008) Land use land cover classification of Orissa using multi-temporal IRS-P6 Awifs data: a decision tree approach. *Int J Appl Earth Obs Geoinf* 10:186–193
- Kasetkasem T, Arora MK, Varshney PK (2005) Super-resolution land cover mapping using a markov random field based approach. *Remote Sens Environ* 96:302–314
- Kong F, Yin H, Nakagoshi N, James P (2012) Simulating urban growth processes incorporating a potential model with spatial metrics. *Ecol Indic* 20:82–91
- Koukoletsos T, Haklay M, Ellul C (2012) Assessing data completeness of VGI through an automated matching procedure for linear data. *Trans GIS* 16(4):477–498
- Landis JR, Koch GG (1977) The measurement of observer agreement for categorical data. *Biometrics* 33:159–174
- Ludwig I, Voss A, Krause-Traudes M (2011) A comparison of the street networks of Navteq and OSM in Germany. In: Geertman S, Reinhardt W, Toppen F (eds) *Advancing geoinformation science for a changing world SE-4 lecture notes in Geoinformation and cartography*. Springer, Berlin, pp 65–84
- Mayaux P, Eva H, Gallego J, Strahler AH, Herold M, Member S, Agrawal S, Naumov S, De Miranda EE, Di Bella CM et al (2006) Validation of the global land cover 2000 map. *IEEE Trans Geosci Remote Sens* 44:1728–1739
- McIver D, Friedl M (2002) Using prior probabilities in decision-tree classification of remotely sensed data. *Remote Sens Environ* 81:253–261
- Mooney P, Corcoran P (2012) The Annotation Process in OpenStreetMap. *Trans GIS* 16:561–579
- Pacifici F, Chini M, Emery WJ (2009) A neural network approach using multi-scale textural metrics from very high-resolution panchromatic imagery for urban land-use classification. *Remote Sens Environ* 113:1276–1292
- Paneque-Gálvez J, Mas J-F, Moré G, Cristóbal J, Orta-Martínez M, Luz AC, Guèze M, Macía MJ, Reyes-García V (2013) Enhanced land use/cover classification of heterogeneous tropical landscapes using support vector machines and textural homogeneity. *Int J Appl Earth Obs Geoinf* 23:372–383
- Qi Z, Yeh AG-O, Li X, Lin Z (2012) A novel algorithm for land use and land cover classification using RADARSAT-2 polarimetric SAR data. *Remote Sens Environ* 118:21–39

- Ramm F (2014) OpenStreetMap data in layered GIS format <http://www.geofabrik.de/data/geofabrik-osm-gis-standard-0.6.pdf>
- Ramm F, Names I, Files SS, Catalogue F, Features P, Features N, Related T, Infrastructure T, Generation P, Features L, et al (2011) OpenStreetMap data in layered GIS format pp 1–21
- Roick O, Hagenauer J, Zipf A (2011) OSMMatrix—grid-based analysis and visualization of OpenStreetMap. In: State of the map EU 2011, Vienna, Austria
- Rouse LJ, Bergeron SJ, Harris TM (2007) Participating in the geospatial web: collaborative mapping, social networks and participatory GIS. In: Scharl A, Tochtermann K (eds) The geospatial web advanced information and knowledge processing. Springer, London, pp 153–158
- Saadat H, Adamowski J, Bonnell R, Sharifi F, Namdar M, Ale-Ebrahim S (2011) Land use and land cover classification over a large area in iran based on single date analysis of satellite imagery. *ISPRS J Photogramm Remote Sens* 66:608–619
- See L, Comber A, Salk C, Fritzz S, van der Velde M, Perger C, Schill C, McCallum I, Kraxner F, Obersteiner M (2013) Comparing the quality of crowdsourced data contributed by expert and non-experts. *PLoS ONE* 8:e69958
- Seifert F (2009) Improving urban monitoring toward a European urban atlas. In: Global mapping of human settlement; remote sensing applications series. CRC Press, USA
- Sester M, Jokar Arsanjani J, Klammer R, Burghardt D, Haunert J-H (2014) Integrating and generalising volunteered geographic information. In: Burghardt D, Duchêne C, Mackaness W (eds) Abstracting geographic information in a data rich world, in series: lecture notes in geoinformation and cartography. Springer, Berlin, pp 119–155
- Sexton JO, Urban DL, Donohue MJ, Song C (2013) Long-term land cover dynamics by multi-temporal classification across the landsat-5 record. *Remote Sens Environ* 128:246–258
- Strahler AH, Boschetti L, Foody GM, Friedl MA, Hansen MC, Herold M, Mayaux P, Morisette JT, Stehman SV, Woodcock CE (2006) Global land cover validation: recommendations for evaluation and accuracy assessment of global land cover maps. Office for Official Publications of the European Communities, Luxembourg
- Thenkabail PS, Schull M, Turrall H (2005) Ganges and Indus river basin land use/land cover (LULC) and irrigated area mapping using continuous streams of MODIS data. *Remote Sens Environ* 95:317–341
- Van Oort P (2006) Spatial data quality: from description to application. Wageningen University
- Vaz E, Nijkamp P, Painho M, Caetano M (2012) A multi-scenario forecast of urban change: a study on urban growth in the Algarve. *Landsc Urban Plan* 104:201–211
- Vaz E, Walczynska A, Nijkamp P (2013) Regional challenges in tourist wetland systems: an integrated approach to the Ria Formosa in the Algarve, Portugal. *Reg Environ Change* 13:33–42
- Wästfelt A, Arnberg W (2013) Local spatial context measurements used to explore the relationship between land cover and land use functions. *Int J Appl Earth Obs Geoinf* 23:234–244