International Journal of Applied Earth Observation and Geoinformation 35 (2015) 329–337

Contents lists available at ScienceDirect

International Journal of Applied Earth Observation and Geoinformation

journal homepage: www.elsevier.com/locate/jag

An assessment of a collaborative mapping approach for exploring land use patterns for several European metropolises

Jamal Jokar Arsanjani[a,*], Eric Vaz[b]

[a] Heidelberg University, Institute of Geography, GIScience Research Group, Heidelberg, Germany
[b] Ryerson University, Department of Geography, Toronto, ON, Canada

ARTICLE INFO

Article history:
Received 26 February 2014
Accepted 17 September 2014
Available online 25 October 2014

Keywords:
Land use mapping
Comparative assessment
GMESUA
OpenStreetMap
Collaborative mapping
Citizen science

ABSTRACT

Until recently, land surveys and digital interpretation of remotely sensed imagery have been used to generate land use inventories. These techniques however, are often cumbersome and costly, allocating large amounts of technical and temporal costs. The technological advances of web 2.0 have brought a wide array of technological achievements, stimulating the participatory role in collaborative and crowd sourced mapping products. This has been fostered by GPS-enabled devices, and accessible tools that enable visual interpretation of high resolution satellite images/air photos provided in collaborative mapping projects. Such technologies offer an integrative approach to geography by means of promoting public participation and allowing accurate assessment and classification of land use as well as geographical features. OpenStreetMap (OSM) has supported the evolution of such techniques, contributing to the existence of a large inventory of spatial land use information. This paper explores the introduction of this novel participatory phenomenon for land use classification in Europe's metropolitan regions. We adopt a positivistic approach to assess comparatively the accuracy of these contributions of OSM for land use classifications in seven large European metropolitan regions. Thematic accuracy and degree of completeness of OSM data was compared to available Global Monitoring for Environment and Security Urban Atlas (GMESUA) datasets for the chosen metropolises. We further extend our findings of land use within a novel framework for geography, justifying that volunteered geographic information (VGI) sources are of great benefit for land use mapping depending on location and degree of VGI dynamism and offer a great alternative to traditional mapping techniques for metropolitan regions throughout Europe. Evaluation of several land use types at the local level suggests that a number of OSM classes (such as anthropogenic land use, agricultural and some natural environment classes) are viable alternatives for land use classification. These classes are highly accurate and can be integrated into planning decisions for stakeholders and policymakers.

Introduction

Land use and land cover mapping have brought accurate understanding of spatial allocation of land use/cover patterns as well as cognition of their attributes (Paneque-Gálvez et al., 2013; Sexton et al., 2013). Despite huge efforts of several projects for land use/cover mapping (e.g., GLC-2000: Fritz et al., 2003; MODIS: Friedl et al., 2002; GlobCover: Bontemps et al., 2011; CORINE 2000: Buettner et al., 2002) however, a substantial disagreement between these is still present (Mayaux et al., 2006; Herold et al., 2008). This divergence resides in the accuracy of land use inventories and the accuracy of land use dynamics, and their often uncertain integration as planning instruments (Mayaux et al., 2006; Pontius and Petrova, 2010; Jokar Arsanjani et al., 2013b; Vaz et al., 2013).

Remotely sensed imagery and remote sensing techniques have substantially facilitated the process of land cover/use mapping, but an essential component is additionally required to produce high quality land use maps, namely the "in-field information" (de Sherbinin, 2002; Flanagin and Metzger, 2008; Gervais et al., 2009; Guptill and Morrison, 1995; Waestefelt and Arnberg, 2013), often an absent property, residing as a culprit in the lack of quality in the land use classification process. While the advances in signal processing algorithms and remote sensing images have allowed significant advances in land use classification (Kandrika and Roy, 2008), which has been diversified in a myriad of current studies (see Pacifici et al., 2009; Qi et al., 2012), the in-field information can be

* Corresponding author at: Institute of Geography, Berliner Strasse 48, 69120, Heidelberg, Germany. Tel.: +49 6221545572.
E-mail addresses: jokar.arsanjani@geog.uni-heidelberg.de (J. Jokar Arsanjani), evaz@ryerson.ca (E. Vaz).

http://dx.doi.org/10.1016/j.jag.2014.09.009

hardly delivered by remotely sensed data. This leads land use managers and practitioners to fund field surveys enabling the collection of additional information to generate land use/cover inventories (Saadat et al., 2011). From a budgetary perspective, the costs of in-field collection are often doubled (Cihlar and Jansen, 2001; De Leeuw et al., 2011). In addition, large areas of the world, particularly in developing countries, are inaccessible to field surveys, thus preventing accurate assessments of land use inventories (Vaz et al., 2014). Resulting from the lack of land use availability, new integrative methods are progressively becoming requested, abridging this asymmetry of land use data sources (Powell et al., 2004; Strahler et al., 2006; Giri, 2012).

The combination of community-driven spatial information and attributes with remote sensing data through a participatory process, which could bring the citizens into play and create a cumulative shared knowledge, allows for an alternative approach of better understanding of land use/cover classification (Fritz et al., 2012). Such an approach could reduce the classification errors often found in land use/cover inventories (Strahler et al., 2006; Bontemps et al., 2011), and respond to the monetary constraints and promote cost efficiency. The idea of involving citizens to land mapping through web 2.0 technologies has been recently implemented through initiatives such as the Geo-wiki project (Fritz et al., 2012; Foody et al., 2014), as well as the Landspotting project (Sturn et al., 2013). These integrative approaches are particularly advantageous in parts of the world that lack land use/cover maps and where local knowledge can enhance land use classification (Foody, 2002). The accumulation of local knowledge has advanced drastically in recent years given the advance of web 2.0 technologies. These technologies have allowed for a number of different user generated content, as well as opportunities for collaborative mapping solutions, as explored by Haklay (2013). Thanks to the wide release of image libraries (e.g., Bing maps) to citizen science-based projects, volunteers have access to a set of user-friendly interfaces with some map editing capabilities to delineate geometrical representation of any feature/area of interest according to the provided high resolution images/aerial photos. Additionally, mapped features can be enriched with individual attributes. This has led to significant advances in importing preliminary prepared digital maps as well as highly accurate GPS-enabled technology (e.g., smartphones) in promoting a better and more complete mapping experience. The result fosters better digital content with much more diversified spatial and attribute information resulting from multiple volunteers. To this pertains the core philosophy of collaborative mapping, creating what Goodchild refers to as "citizens as sensors" (Goodchild, 2007), enabling much larger sets of information and digital content thanks to crowdsourcing (Heipke, 2010). As such, location has become a diversified source adequate for planning of regions, and projects can benefit greatly from this approach (Fritz et al., 2012). Concerning its practicality, recent findings on streets networks of OpenStreetMap (OSM) suggest that OSM datasets are occasionally close to complete in relation to proprietary sources (e.g., Ludwig et al., 2011; Neis and Zipf, 2012; Corcoran and Mooney, 2013; Koukoletsos et al., 2012). Thus, the new paradigm of using collective spatial data can greatly enhance several issues of missing information in spatial datasets regarding land use/cover. Additionally, the constant augmentation of spatial content in collaborative mapping brings unprecedented potential to multi-temporal land use assessments. This is in contrast to proprietary inventories that are temporally static and heavily dependent on arranged surveys. In contrast to the advantages of VGI, it should be noted that the degree of dynamism in VGI across time and space is a crucial point, because the more volunteers get involved in contributing to VGI platforms, the better data quality is achieved. Similarly, the more often volunteers contribute, the more up-to-date VGI proceeds (Jokar Arsanjani et al., 2014).

This study explores the completeness and thematic accuracy of the contributed land use features to the OSM. It is assumed that the given land use features to the OSM for metropolitan areas can be partially used as a source for effective land use classification. A quality assessment on how well land use features have been delivered to OSM in different regions throughout Europe is conducted. The main objective is thus to prepare land use information from OSM contributions and extent these to the GMESUA land use datasets to create a semantic compatibility through harmonizing the legend of both data sources. We then organize a thematic accuracy assessment over the OSM features versus the GMESUA. Next, we addressed the following questions: (i) how complete are the land use features in OSM in relation to GMESUA dataset?; (ii) how effective is the use of OSM data for challenges in land use science specifically for unclassified, misclassified, and invalidated parcels?; (iii) what methods are effective for exacting land use information from OSM and assessing their accuracy for use in land use science?; and (iv) what are the current challenges in collective land use sources and how can these be addressed for more accurate global mapping purposes. We further explore the used datasets and frame the study regions in *Materials* section. *Methodology* section depicts our methodology. *Results and discussion* section adopts an analytical perspective on the achieved results, and *Conclusions* section advances with concluding remarks and draws future recommendations.

Materials

OpenStreetMap dataset

In this section we describe how the input data from OSM is prepared. It is necessary to mention how OSM data for this study is acquired and how the data is pre-processed for the analysis. OSM features tagged with "*land use*", "*natural*", and "*buildings*" were extracted from the available OSM database acquired from the planet.osm file downloaded on September 5, 2013. In contrast to anthropogenic features, a wide array of physical features were designated as "Natural". The feature "Land use" represents the use of anthropogenic land, adding on the information of the land use for each single land segment. "Buildings", on the other hand, represents the geometrical footprint of anthropogenic constructs over space. These features were compiled accordingly and represent a homogenous land use inventory with different land use types.

Comparative reference dataset: GMESUA

The European Environment Agency (EEA) provides pan-European land use datasets for urban areas containing more than 100,000 inhabitants. These datasets are adapted to European needs, and contain information that is extracted from diverse sources, such as Earth Observation (EO) data, commercial-off-the-shelf (COTS) navigation data, as well as topographic maps. Its minimum mapping unit (MMU) fluctuates between 0.25–1 ha, and a minimum width of linear elements of 100 m with $\pm$5 m positional accuracy is used (European Union, 2011). Quality of the datasets is assessed through integration of ancillary data such as (a) COTS navigation data, e.g., POIs, land use, land cover, 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) very-high-resolution aerial imagery (finer than 1 m ground resolution) (European Environment Agency, 2010). At present, 305 urban regions in Europe are covered. The thematic accuracy for all classes is over 80% (Seifert, 2009; European Union, 2011). Table 1 shows the land use classes found in the Urban Atlas. For the purpose of this paper and consistency of data, we adopt a