ARTICLE IN PRESS

Remote Sensing of Environment xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Remote Sensing of Environment

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

Assessing the accuracy of multi-temporal built-up land layers across rural-urban trajectories in the United States

Stefan Leyk[a,*], Johannes H. Uhl[a], Deborah Balk[b], Bryan Jones[b]

[a] University of Colorado Boulder, Department of Geography, Boulder, CO 80309, USA
[b] City University of New York, Institute for Demographic Research and Baruch College, New York, NY 10010, USA

ARTICLE INFO

Keywords:
Spatiotemporal uncertainty
Global Human Settlement Layer
Open data
Data integration
Accuracy assessment
Urban land

ABSTRACT

Global data on settlements, built-up land and population distributions are becoming increasingly available and represent important inputs to a better understanding of key demographic processes such as urbanization and interactions between human and natural systems over time. One persistent drawback that prevents user communities from effectively and objectively using these data products more broadly, is the absence of thorough and transparent validation studies. This study develops a validation framework for accuracy assessment of multi-temporal built-up land layers using integrated public parcel and building records as validation data. The framework is based on measures derived from confusion matrices and incorporates a sensitivity analysis for potential spatial offsets between validation and test data as well as tests for the effects of varying criteria of the abstract term built-up land on accuracy measures. Furthermore, the framework allows for accuracy assessments by strata of built-up density, which provides important insights on the relationship between classification accuracy and development intensity to better instruct and educate user communities on quality aspects that might be relevant to different purposes. We use data from the newly-released Global Human Settlement Layer (GHSL), for four epochs since 1975 and at fine spatial resolution (38 m), in the United States for a demonstration of the framework. The results show very encouraging accuracy measures that vary across study areas, generally improve over time but show very distinct patterns across the rural-urban trajectories. Areas of higher development intensity are very accurately classified and highly reliable. Rural areas show low degrees of accuracy, which could be affected by misalignment between the reference data and the data under test in areas where built-up land is scattered and rare. However, a regression analysis, which examines how well GHSL can estimate built-up land using spatially aggregated analytical units, indicates that classification error is mainly of thematic nature. Thus, caution should be taken in using the data product in rural regions. The results can be useful in further improving classification procedures to create measures of the built environment. The validation framework can be extended to data-poor regions of the world using map data and Volunteered Geographic Information.

1. Introduction

How much we know about processes of urbanization and land conversion over the past decades depends heavily on the data that are available for analysis. In developed regions, there is usually an abundance of demographic and land use related data resulting in some confidence in analytical results at national levels for limited periods of time. However, such data are not available for most regions and countries of the world and thus global (and globally consistent) data products on population, land cover and built-up land play an important role for developing a better understanding of global dimensions of demographic key processes.

To date, existing large-scale datasets show significant limitations (e.g., Gong et al., 2013) and important differences in basic parameters such as spatial and temporal resolution, temporal coverage or thematic and regional consistency (for an overview see Grekousis et al., 2015). The most common drawback among all these data products is the persistent lack of knowledge of their spatial and thematic (Strahler et al., 2006) as well as spatio-temporal (Tsutsumida and Comber, 2015) accuracy and thus their fitness for different uses by the research user community. In addition, to date, there is a lack of consistent collection of global information across time impeding reporting efforts related to activities as required by the post-2015 Development Agenda (United Nations, 2012). It is well-known that classifications of built-up land or developed land suffer from lower levels of accuracy in less developed regions and rural settings (Smith et al., 2002; Wickham et al., 2013).

* Corresponding author.
E-mail addresses: stefan.leyk@colorado.edu (S. Leyk), johannes.uhl@colorado.edu (J.H. Uhl), deborah.balk@baruch.cuny.edu (D. Balk), bryan.jones@baruch.cuny.edu (B. Jones).

http://dx.doi.org/10.1016/j.rse.2017.08.035
Received 9 February 2017; Received in revised form 10 August 2017; Accepted 25 August 2017

The main reason for this inherent property is the poor performance of remote sensing based classifiers if the data are of rather coarse spatial and spectral resolution in relation to the object of interest, which can result in mixed pixel effects. Such effects are amplified by the use of natural construction material and smaller rooftop surfaces in less developed and rural regions as compared to more developed regions. Population data that is often used as an ancillary variable to improve such classifications is usually available at relatively coarse resolution, and if population distributions are to be created the demographic data is usually allocated to grid cells classified as developed or populated. Because of the lack of validation data for global land cover classifications in general (Zhao et al., 2014) such procedures introduce and propagate uncertainty, and this inherent uncertainty has rarely been quantified, modelled or evaluated, thoroughly. This shortcoming is even more impactful for large geographic areas and in less developed regions, as well as for earlier time periods since uncertainty is expected to be higher under all these conditions. This is considered a serious limitation of such databases because there is very limited understanding of the reliability of estimates and statistics derived from such data and this evokes potential misuse of the data. Some examples of datasets exposed to these issues are GlobCover (Bontemps et al., 2011), GlobLand30 (Chen et al., 2015) the Gridded Population of the World (GPW) and the related Global Rural Urban Mapping Project (GRUMP), which are snap-shots in time on population counts (Balk et al., 2005, 2006; Deichmann et al., 2001; CIESIN, 2005), Landscan (Dobson et al., 2000), which represents the ambient population per grid cell, the WorldPop project (Sorichetta et al., 2015) and the recent Global Human Settlement Layer (GHSL) (Pesaresi et al., 2015) which represents a first attempt to account for global built-up areas from decametric-resolution satellite data using a consistent multi-temporal data classification approach. Klotz et al. (2016) propose a cross-comparison framework to assess the accuracy of some of the above mentioned data products including the GHSL and the Global Urban Footprint dataset (GUF) (Esch et al., 2013) within European settings and for one point in time.

It is thus of great importance to derive further in-depth knowledge of the accuracy of such global datasets to better understand how confident researchers can be in using them in different regions of the world for demographic analysis, land cover change modeling or the characterization of relevant demographic processes such as urbanization. Focusing on built-up land layers, this urgent research need resonates with three main goals of the current study. *First*, exemplified for the United States, a validation dataset will be constructed through data integration procedures incorporating public parcel records and building footprints from several regions that can be used as an extensive sample for a national assessment at different points in time. *Second*, an analytical framework for accuracy assessment will be developed that allows evaluating multi-temporal spatial datasets on built-up or developed land at the national, continental and potentially global scale using such validation datasets. This framework will include a sensitivity analysis to address potential spatial offsets and differences in class definition. Most importantly, this framework will also support accuracy assessment within strata defined by different development intensities (i.e., loosely related to rural, peri-urban and urban land) to better understand relationships between development intensity and classification accuracy. *Third*, we will demonstrate the operationality of the developed analytical framework using the validation data layers constructed and testing the GHSL for the selected U.S. counties as our target dataset. This experiment will shed light on the efficacy of the framework to assess classification accuracy and its variation across the study regions, at different points in time, and across regions of varying development intensities.

Validation of multi-temporal data such as the GHSL is complex and challenging due to difficulties in creating historical versions of the test data at fine spatial resolution, issues of spatial and thematic inconsistency, and temporal mismatches between the validation data and target data. The GHSL is exemplified as the target data in this study because it represents a promising new public data product at fine spatial resolution with extensive temporal coverage of > 40 years (Pesaresi et al., 2015). This data product is already in high demand for multiple target applications in different disciplines including population projections (Linard et al., 2017), disaster management and risk assessment (Freire et al., 2015, 2016), as well as land use change modeling (Small and Sousa, 2016). However, the analytical framework is intended to be applicable as a general protocol for accuracy assessment to other datasets with properties and embedded class abstraction comparable to the GHSL and other regions. Furthermore, accuracy assessment results derived for one region may also provide first insights on expected accuracies in other countries in which validation data may be difficult to obtain or are non-existent. Such an accuracy assessment framework will not only enable the analyst to carry out more reliable analyses but will also present a pathway to future improvements of such data products through the detection of problematic regions or contexts which may need particular attention, different classification procedures or the inclusion of additional ancillary data.

2. Data and preprocessing

2.1. Global Human Settlement Layer (GHSL)

The Global Human Settlement Layer (GHSL) project aims to assess the human presence on the planet through analysis of evidences as collected from earth observation data, census data and volunteered geographic information. The GHSL information is shaped by a scalable abstraction schema overcoming the traditional land cover paradigm (Pesaresi et al., 2009) and is currently structured in three basic geo-information layers of increasing abstraction and decreasing spatial resolution: A) built-up areas, B) population grids, and C) the settlement classification model. In Pesaresi et al. (2013) the GHSL information production workflow targeting the "built-up area" class abstraction was defined and the automatic recognition was tested for a large set of sensors in the spatial resolution range of 0.5–10 m. These sensors may perform very well in detecting built-up areas using textural and morphological image-derived descriptors as input to the automatic classification but they are typically constrained regarding data access and processing and redistribution rights, which makes the scientific use of the derived products difficult or unsustainable. Moreover, they are typically available only for more recent years, and acquired in rather scattered ways for arbitrary points in time, which makes these data difficult to use for uniform and systematic information extraction and analysis of global, regional or even national trends.

In order to mitigate some of these issues, the GHSL built-up area recognition system was ported in the decametric resolution, open remote sensing data domain and tested with global collections of image data records collected by the Landsat satellite platform in the past 40 years (Pesaresi et al., 2016a). In the first edition, the GHSL was made available as seamless global mosaic at high spatial resolution (approx. 38 m) and for various epochs (1975, 1990, 2000, 2014, see Fig. 1). The GHSL satellite-derived information production technology is based on symbolic machine learning (SML), a supervised data classification inspired by methods for DNA microarrays data analysis used in biomedical informatics for the clustering of gene expressions (Pesaresi et al., 2016b) in large datasets. Current efforts at the Joint Research Center focus on the production of GHSL versions with improved temporal and spatial resolution using Sentinel1 and Sentinel2 satellite imagery.

2.2. Parcel data

Open cadastral and tax assessment data have become increasingly available to the public – often in GIS-compatible format – for several regions in the U.S. (von Meyer and Jones, 2013) and other countries. For this study, all publicly available parcel data that could be obtained through open-sources were used. Cadastral parcel boundaries are