Building a Crowd-Sourcing Tool for the Validation of Urban Extent and Gridded Population

Steffen Fritz[1], Linda See[1], Ian McCallum[1], Christian Schill[1,2], Christoph Perger[3], and Michael Obersteiner[1]

[1] International Institute for Applied Systems Analysis (IIASA), Schlossplatz 1, A-2361 Laxenburg, Austria
[2] University of Freiburg, Tennenbacherstr. 4, Freiburg, Germany
[3] Fachhochschule Wiener Neustadt, Johannes Gutenberg-Strasse 3, A-2700 Wiener Neustadt, Austria
{fritz,see,mccallum,obersteiner}@iiasa.ac.at, christian.schill@felis.uni-freiburg.de, christoph.perger@fhwn.ac.at

Abstract. This paper provides an overview of the crowd-sourcing tool Geo-Wiki, which is used to collect in-situ land cover validation data from the public. This tool is now being modularized in order to allow for domain specific land cover validation. Agriculture and biomass versions of Geo-Wiki are already operational. The next module, which is called urban.geo-wiki.org, is aimed at the validation of urban extent and gridded population data. The aim of this paper is to outline the structure of this module and the datasets that will aid in the validation of this land cover type and gridded population data. The ultimate aim of Geo-Wiki is to produce a hybrid land cover map that is better than existing products, including estimates of urban extent.

Keywords: Crowd-sourcing, urban areas, land cover validation, Google Earth.

1 Introduction

Cities are one of the greatest challenges of the modern age, where the rapid urbanization of the last century has led to many problems such as pollution, climate change, food security, global health issues and increases in total water demand [1]. More than half of the world's population currently lives in cities, but by 2050, this will increase to 80% as continued movement to cities takes place in China, India, Africa and Asia [2]. Urban areas are captured in global land cover datasets such as the GLC-2000 [3], MODIS [4] and GlobCover [5], which have been created as baseline terrestrial products in the last decade. These products are used in a number of applications such as global and regional land use modeling and land use change [6, 7, 8].

The problem with global land cover products has been highlighted in a number of comparison studies, which have revealed significant amounts of spatial disagreement in land cover types [9] [10] [11] [12] with a particular focus on the forest and cropland domains. The need for a more accurate global map of urban extent has already

B. Murgante et al. (Eds.): ICCSA 2011, Part II, LNCS 6783, pp. 39–50, 2011.
© Springer-Verlag Berlin Heidelberg 2011

been highlighted in [13] in the development of a MODIS urban extent product and gridded population. Information on urban extent, in particular changes in urbanization, and population forecasts are necessary for the estimation of future carbon emissions [14] [15]. They are also important in trying to define world urban areas and population [16]. However, future projections of urbanization and future population will only be as good as the current input data, yet there are obvious discrepancies and uncertainties with the global datasets used to create these inputs.

One reason why the various global land cover maps disagree is because of insufficient ground-truth or in-situ data for the calibration and validation of these maps. Google Earth and Google Earth Engine provide ideal mechanisms for improving the validation of global land cover, including urban extent. By superimposing urban land cover onto Google Earth, the accuracy of urban extent from any land cover product can be assessed. Other datasets such as gridded population can also be superimposed on Google Earth and validated in the same way. However, the task of validation is an enormous one, and is therefore an ideal candidate for crowd-sourcing, in particular because Google Earth is freely available to the public via the internet. Crowd-sourcing involves the public provision of information to different websites on the internet [17], made possible because of advances in Web2.0 technology and an active and willing public. Examples of successful crowd-sourcing applications include the eBird project [18] containing greater than 48 million bird sightings from the public, Galaxy Zoo [19], where the public classify galaxies and work with scientists in making new discoveries, and FoldIt, which is a serious game in which the public creates protein structures that may lead to the treatment of a disease [20]. There are also many successful examples of spatial crowd-sourcing applications, which provide volunteered geographic information (VGI) [21]. Openstreetmap (openstreetmap.org) is a street map wiki in which members of the public can edit and update street information for any location on the Earth's land surface. Other examples of VGI include wikimapia (wikimapia.org), the Degree Confluence project (confluence.org), MapAction (mapaction.org) and the European Environment Agency's 'Eye on Earth', which involves the wider public in monitoring air and water pollution in the environment.

Google Earth and crowd-sourcing have been combined together in the Geo-Wiki application, developed in [22], as a way of increasing the database of in-situ calibration and validation points for improving land cover. Geo-Wiki also represents one initiative in the growing trend of 'community remote sensing' in which remote sensing data are combined with local information to directly engage and empower communities in understanding and arguing environmental issues [23]. Geo-Wiki could be used by the public for a range of environmental applications. However, from a land cover perspective, the ultimate goal of Geo-Wiki is to use the crowd-sourced calibration/validation database to create a hybrid land cover product that is better than any of the individual products that are currently available.

The aim of this paper is to provide an overview of the Geo-Wiki application and the latest module to be developed: urban.geo-wiki.org, which covers validation of urban areas and gridded population. The need for such a module becomes evident through an examination of the disagreement in urban extent based on the GLC-2000, MODIS and GlobCover. Plans for how Geo-Wiki will be developed in the short to long term are also presented.