Improving Global Land Cover through Crowd-sourcing and Map Integration

L. See[1], S. Fritz[1], I. McCallum[1], C. Schill[2], C. Perger[3] and M. Obersteiner[1]

[1]International Institute for Applied Systems Analysis (IIASA), Schlossplatz 1, A-2361 Laxenburg, Austria
Telephone: (+43 2236) 807423
Fax: (+43 2236) 807599
Email: see@iiasa.ac.at

[2]University of Freiburg, Tennenbacherstr. 4, Freiburg, Germany
Telephone: (+49 761) 2038645
Fax: (+49 761) 2033701
Email: christian.schill@felis.uni-freiburg.de

[3]Fachhochschule Wiener Neustadt, Johannes Gutenberg-Strasse 3, A-2700 Wiener Neustadt, Austria
Telephone: (+43 2622) 89084243
Fax: (+43 2622) 8908499
Email: christoph.perger@fhwn.ac.at

1. Introduction

There are currently a large number of satellites orbiting the earth collecting vast amounts of Earth Observation (EO) data. With developments like Google Earth and Google Earth Engine, we are witnessing the democratization of EO through public access to high resolution satellite imagery via the internet. One important EO-derived product from satellites is global land cover. In the last decade, three global land cover products have been created: GLC-2000 (Fritz et al., 2003), MODIS (Friedl et al., 2002) and GlobCover (Bicheron et al., 2008). These datasets are currently used as inputs to a range of different global, regional and national scale applications, e.g. resource assessments of forest and agricultural land and inputs to global economic land use models.

There are, however, problems with land cover. A pixel-by-pixel comparison reveals areas of the world where these maps do not agree, in some cases by large amounts (Fritz and See, 2008). As a result we do not know precisely how much land is currently forested or under cultivation because the uncertainty in the estimates provided by these products is too high. This has clear implications for determining deforestation rates and how much land is available for. Users of these products also have a difficult choice, i.e. which is the best product to choose and what effect will this choice have on a particular application? For example, Fritz et al. (2010a) have shown that comparing global land cover (GLC-2000) with the equivalent MODIS product produces large areas of disagreement when assessing the amount of agricultural land available in parts of eastern Africa. The problem with these datasets lies in their validation, as at present, there are an insufficient number of in-situ validation points, which can serve both as input data for calibration algorithms of satellite data, and to validate land cover products. The Geo-Wiki application, developed by Fritz et al. (2009), has integrated Google Earth and crowd-sourcing as a way of increasing the amount of publically-supplied in-situ validation points. The ultimate goal is to use this crowd-sourced data to create a hybrid land cover product that is better than any currently available. The aim of this paper is discuss how validation from Geo-Wiki and a rule-based map integration algorithm could be used to develop such a hybrid product.

2. The Geo-Wiki Land Cover Crowd-sourcing Application

The Geo-Wiki Project (www.geo-wiki.org) was developed to encourage a global network of volunteers to help improve the quality of global land cover maps through crowd-sourcing (Fritz et al., 2009). Geo-Wiki overlays the GLC-2000, MODIS and GlobCover onto Google Earth as well as maps of where these different land cover products disagree. Volunteers can choose any area of land on the earth or an area of high disagreement. Geo-Wiki shows them where the pixels from each land cover product overlap and the land cover types as shown in fig. 1. The light blue rectangle is GLC-2000 and has the lowest resolution of 1km. The dark blue square is one pixel from the MODIS land cover product while the red square is GlobCover at the highest resolution of 300m. Volunteers are then asked to determine whether the land cover maps at that point agree with what they see based on Google Earth. Their input is recorded in a database, along with any photos they upload. At present there are 300 users registered on the system who have contributed more than 15,000 validation points.

3. Development of a Hybrid Product through Map Integration

Fritz et al. (2010b) have developed a methodology for combining five different land cover maps to create a cropland or forest extent using expert knowledge and national and sub-national statistics. However, to create a global land cover map is more complicated because the legends of the different land cover products do not match. An aggregated and simplified legend to which the different land cover products can be matched directly must first be created (e.g. Herold et al., 2008). This is already available on Geo-Wiki as a simplified legend with 11 classes.

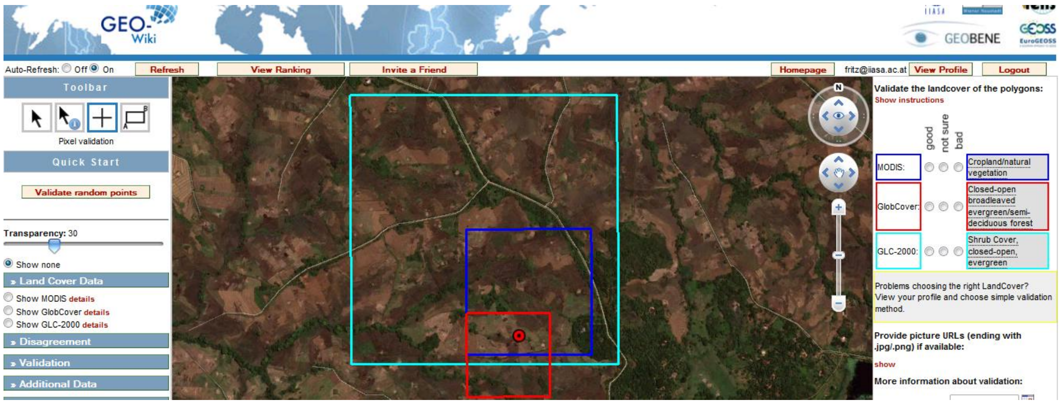

Figure 1. Illustration of Geo-wiki.org for improving land cover information

To use the validation data for creating a hybrid map, there must be sufficient confidence in the data before they are used in the new hybrid product. To gain confidence, people are directed to the same validation site so that a frequency distribution can be derived and agreement can be reached on which product is better. Once a certain threshold had been achieved, the point qualifies to be a validation point to be used in the hybrid map generation. Those validation points where the confidence is low will not be used in the hybrid map production

To create a hybrid map, a rule-based system is currently under development. For each pixel, the system will query whether there is disagreement between the land cover products as follows:

a. Where no disagreement exists, the hybrid land cover map will be assigned the class from the aggregated legend unless validation data exist at that pixel which disagrees with the land cover products. If there is sufficient confidence in the validation data, then these data will be used to correct the information from the land cover products in the new hybrid product.
b. Where disagreement between land cover products exists, the validation data from Geo-Wiki will be used. Where validation data of a sufficient confidence are available at that pixel, this will be used to assign the land cover class. Where no validation data exist, we will employ a search algorithm to determine if the same corrections have been applied using validation points within a certain radius,

which will be used to assign the land cover class. This will also increase our confidence in the validation sites.

This rule-based system will be implemented in Geo-Wiki in the next few months and an example of a hybrid map and the issues that have arisen from the implementation will be presented at the conference.

4. Ongoing Developments and Further Research

There are several ongoing developments with Geo-Wiki to improve the volume and coverage of data collected through the website. The first addresses the problem that there is currently little incentive for volunteers to willingly validate global land cover. One method of providing this incentive will be to develop a game that encourages users to play while simultaneously providing land cover validation information. An Austrian Funding Agency project called LandSpotting, which will begin in Feb 2011, addresses the creation and implementation of such a game. A second development is from the research community. A workshop on land cover validation (sponsored by the International Livestock Research Institute (ILRI) and to be held in June 2011 at IIASA) will bring together land cover and validation experts to discuss methods for creating hybrid products and the sharing of land cover products and validation data via Geo-Wiki. This will increase the size of the validation database and particularly the coverage across Africa and south-east Asia. The progress on these advances will also be presented at the conference.

5. Acknowledgements

This research was supported by the European Community's Framework Programme via the Project EuroGEOSS (No. 226487) and by the Austrian Research Funding Agency (FFG) via the Project LandSpotting. (No. 828332).

6. References

Bicheron P, Defourny P, Brockman C, Schouten L, Vancutsem C, Huc M, Bontemps S, Leroy M, Achard F, Herold M, Ranera F and Arino O, 2008, GLOBCOVER, http://ionia1.esrin.esa.int/docs/GLOBCOVER_Products_Description_Validation_Report_I2.1.pdf

Friedl MA, McIver DK, Hodges JCF, Zhang XY, Muchoney D, Strahler AH, Woodcock CE, Gopal S, Schneider A, Cooper A, Baccini A, Gao F, Schaaf C, 2002, Global land cover mapping from MODIS: algorithms and early results. *Remote Sensing of Environment*, 83:287-302.

Fritz, S., Bartholomé, E., Belward, A., Hartley, A., Stibig H.J., Eva, H., Mayaux, P., Bartalev, S., Latifovic, R., Kolmert, S., Roy, P., Agrawal, S., Bingfang, W., Wenting, X., Ledwith, M., Pekel, F.J., Giri, C., Mücher, S., de Badts, E., Tateishi, R., Champeaux, J-L., Defourny, P., 2003.

Harmonisation, mosaicing and production of the Global Land Cover 2000 database (Beta Version), Luxembourg: Office for Official Publications of the European Communities, EUR 20849 EN, 41 pp., ISBN 92-894-6332-5*.

Fritz S, and See L, 2008, Quantifying uncertainty and spatial disagreement in the comparison of Global Land Cover for different applications, *Global Change Biology*, 14:1-23.

Fritz S, McCallum I, Schill C, Perger C, Grillmayer R, Achard F, Kraxner F and Obersteiner M, 2009, Geo-Wiki.Org: The use of crowd-sourcing to improve global land cover. *Remote Sensing*, 1(3):345-354.

Fritz S, See LM and Rembold F, 2010a, Comparison of global and regional land cover maps with statistical information for the agricultural domain in Africa. *International Journal of Remote Sensing*, 25(7-8):1527-1532.

Fritz S, You L, Bun A, See LM, McCallum I, Liu J, Hansen M and Obersteiner M, 2010b, Cropland for Sub-Saharan Africa: A synergistic approach using five land cover datasets. *Geophysical Research Letters*, doi:10.1029/2010GL046231, in press.

Herold M, Mayaux P, Woodcock CE, Baccini A and Schmullius C (2008) Some challenges in global land cover mapping: an assessment of agreement and accuracy in existing 1 km datasets. *Remote Sensing of Environment*, 112: 2538–2556.