A citizen science application for improving the spatial distribution of global forests

Dmitry Shchepashchenko1,2, Myroslava Lesiv1, Linda See1, Steffen Fritz1, Anatoly Shvidenko1, Christoph Perger1, Martina Dürauer1, Florian Kraxner1, Maria Schepaschenko3, Ian McCallum1

1 International Institute for Applied Systems Analysis, Laxenburg, Austria; schepd@iiasa.ac.at
2 Moscow State Forest University, Mytischi, Russia
3 Russian Institute of Continuous Education in Forestry, Pushkino, Russia; mariaschep@gmail.com

Abstract

Citizen science and crowdsourcing applications have witnessed explosive growth during recent years. Reasons include new technological innovations and the desire of citizens to engage with environmental issues and scientific research. IIASA has developed several tools relevant to forest cover monitoring and mapping, including the Geo-Wiki.org portal (Fritz et al., 2009, 2012), mobile applications (e.g. Geo-Wiki Pictures) and a gaming approach (i.e. Crop Capture – See et al., 2014). The team at IIASA runs crowdsourcing competitions to encourage people to participate; incentives include prizes and co-authorship on publications resulting from analysis of the data. The data collected through Geo-Wiki have been used for independent validation of existing land cover maps and production of hybrid land cover and forest maps (Fritz et al., 2013; See et al., 2013; 2014; Schepaschenko et al., 2015). A number of global forest cover products have been made available in the recent past, however, these datasets often contradict one other when compared spatially, and they are typically inconsistent with national statistics. In this paper we compare freely available remote sensing products with a sample collected via Geo-Wiki using high resolution imagery provided by Google Earth. This sample contains around 22K training points, where the percentage of forest cover was estimated. We applied the method of geographically weighted regression to calculate hybrid forest cover products both calibrated with FAO FRA statistics and not calibrated. Independent validation shows that the hybrid products are better than any of the individual products. These products are presented at http://geowiki.org both for exploration and download.

Keywords: forest cover, remote sensing, crowdsourcing.

Introduction, scope and main objectives

Forests cover around 30% of the global land surface (FAO 2010), comprise greater than 80% of the live biomass (Shvidenko et al 2005) and provide more than 90% of the carbon sink of terrestrial ecosystems (Le Quéré et al 2013, Pan et al 2011), playing a crucial role in the global carbon cycle. A number of global maps of forest cover and its change have been made available in the recent past, e.g. global land cover datasets (MODIS, GLOBCOVER, ESA/CCI LC, GLC-SHARE, GlobeLand30), tree cover (MODIS VCF, Hansen et al., 2013); lidar-based (e.g. Baccini et al 2012, Saatchi et al 2011) and radar-based products (Santoro et al., 2012, JAXA PALSAR). However, these datasets often contradict one other when compared spatially, they are typically inconsistent with national statistics and require independent validation. End-users want to know which map is better for a particular region and resolution as well as the accuracy.

We aimed to collect freely available maps of forest cover to validate them and produce a series of hybrid maps with better accuracy. Broad employ of crowdsourcing allowed us to collect necessary “ground truth” information.
Methodology/approach

We compare a number of existing forest cover maps with reference dataset collected with Geo-wiki web tool to produce a hybrid forest maps. The end products are two different types of hybrid maps of percentage forest cover. We refer to the first map as our “best guess” forest cover map. This reflects our best knowledge about forest cover based on remotely sensed products. The second type of product is the hybrid map, calibrated with FAO FRA national statistics (FAO 2010). The methodology to create these products is illustrated in Fig. 1).

Geographically Weighted Regression (GWR) was employed for integration different maps to a hybrid one (Brunsdon et al 1998, Comber et al 2013, Lesiv et al 2013).

The training data used in the GWR and the validation data for accuracy assessment were obtained through separate exercises via the Geo-Wiki tool for visualization, crowdsourcing and validation of land cover (Fritz et al., 2009; Fritz et al., 2012). This resulted in approximately 22K training data points 1 km size. 20K of them were used for the training GWR and others - for validation.
More details on the methodology can be found in Schepaschenko et al. (2015).

**Results**

Both calibrated (using FAO FRA statistics) and non-calibrated (“best guess”) forest cover datasets were obtained based on the mentioned above method. “Best guess” map reflects our best knowledge about forest cover based on remotely sensed products. A hybrid map calibrated with FAO FRA country statistics at the national level suffers from country border effects in some regions (e.g. clearly visible in central Africa). This is a result of having different probability thresholds for each country, which reflects the quality of country-level statistics. The hybrid forest cover maps available at http://geo-wiki.org (Fig. 2).

![Geo-Wiki tool for crowdsourcing, demonstration, and dissemination](image)

**Fig. 2: Geo-Wiki tool for crowdsourcing, demonstration, and dissemination**

Independent validation of seven existing datasets on forest extent at the resolution of 1km together with hybrid maps was provided (Table 1).
Results from Table 1 demonstrate that the hybrid “best guess” product outperforms all of the input datasets and the other hybrid product. This “best guess” hybrid product has the highest overall accuracy (93%) and generally scores better in terms of sensitivity and specificity (93%). The second best product is Hansen’s TC.

**Discussion**

We demonstrated the method and produced the first map consistent with FAO FRA statistics. The map is freely available at Geo-Wiki.org portal. Hybrid map calibrated with FAO FRA country statistics at the national level suffers from country border effects in some regions (e.g. clearly visible in central Africa). This is a result of having different probability thresholds for each country, which reflects the quality of country-level statistics. This map shows less accuracy compare to “best guess” map (Table 1). However the FAO FRA calibrated map is required by global economic, biophysical and land use models, e.g. GTAP, IMAGE, GLOBIOM, IMPACT, etc. (Nelson et al., 2014). These statistics are also used, e.g.
in climate change negotiations, since FAO provides the only globally consistent estimates of forest, even though it is recognized that countries record forest statistics in different ways.

**Conclusions/outlook**

This paper outlines the development and validation of the first global hybrid forest cover map that integrates remote sensing products and data collected via crowdsourcing, which is also consistent with national forest statistics from FAO at the national level. An additional hybrid forest product is available as a “best guess” forest cover map that is independent of FAO. Validation of the individual and hybrid products confirmed that the integrated products represent forest cover better than any of the eight existing individual datasets at a resolution of 1km.

These hybrid products serve two purposes – first, they demonstrate that cost-effective improvement is possible through such a harmonized approach, combining existing products together with crowdsourced information and statistics. This type of approach could also be readily applied to improve the accuracy of other spatial biophysical products. Secondly, these products provide a dataset of increased confidence that can be used by both the monitoring and modelling communities. Such datasets are of extreme importance regarding the establishment of baselines for REDD and carbon cycle modelling, along with issues related to biofuel assessment. Furthermore, increased accuracy in the representation of forest cover will improve the accuracy of land use modelling more generally – particularly with regards to the interactions with the agricultural sector.

**Acknowledgements**

This study has been partly supported by the EC-funded 7th Framework Programme Projects SIGMA (603719), SIFCAS - FP7-PEOPLE-2013-IIF (627481); the GLOBBIO MASS project supported by ESA; the REDD-PAC project supported by the International Climate Initiative of the Federal Ministry for the Environment, Nature Conservation and Nuclear Safety (BMU), Germany.

The views expressed in this information product are those of the author(s) and do not necessarily reflect the views or policies of FAO.

**References**


DiMiceli CM, Carroll ML, Sohlberg RA, Huang C, Hansen MC, Townshend JRG. 2011. Annual Global Automated MODIS Vegetation Continuous Fields (MOD44B) at 250 m Spatial Resolution for Data Years Beginning Day 65, 2000 - 2010, Collection 5 Percent Tree Cover Available at: http://glcf.umd.edu/data/vcf/ [Accessed 04.05.2015]


