

Using Student Volunteers to Crowdfund Land Cover Information

Franziska ALBRECHT^{1*}, Mario ZUSSNER¹, Christoph PERGER², Martina DÜRAUER²,
Linda SEE², Ian MCCALLUM², Steffen FRITZ² and Wolfgang KAINZ¹

¹Dept. of Geography and Regional Research,
University of Vienna / Austria · franziska.albrecht@univie.ac.at
²IIASA, Ecosystem Services and Management Program, Laxenburg / Austria

This contribution was double-blind reviewed as extended abstract.

Abstract

Accurate geospatial cropland information is one of several critical input parameters needed for the assessment of food security and for achieving sustainable development. However, current land cover products are either not accurate enough or lacking in many food insecure countries. Here we demonstrate how we can use the Geo-Wiki crowdsourcing tool and student volunteers to generate a cropland map of Myanmar, and validate a global hybrid cropland map. The preliminary results clearly indicate that the method could be applied in a similar manner to other countries.

1 Introduction

Accurate geospatial cropland information is one of several critical input parameters needed for the assessment of food security by international organizations and aid agencies, which can lead to international response. However, none of the main global land cover products is able to capture cropland extents at a sufficient accuracy (FRITZ et al. 2013b, FRITZ et al. 2011). When these products are compared, large areas of spatial disagreement are apparent. The use of crowdsourcing as a means of collecting information using citizens has grown rapidly over the last decade, and many applications have evolved e.g. eBird project (WOOD et al. 2011) or GalaxyZoo (CLERY 2011). Land cover is one type of information that is also being crowdsourced (e.g. FRITZ et al. 2013a, SEE et al. 2013, SEE et al. 2014). The aim of this study is to demonstrate how land cover information can be generated using the Geo-Wiki crowdsourcing tool (FRITZ et al. 2009) and student volunteers. We use Myanmar as an example country, since it lacks up-to-date land cover information. Furthermore, the country is politically unstable, and mapping on the ground is not possible. The paper is intended to give an insight into the possibilities of this innovative data collection tool.

2 Data and Methods

Involving 97 geography students enrolled at the University of Vienna, the crowdsourcing campaign took place as a science project in autumn 2013. A modified branch of Geo-Wiki (accessible from hackathon.geo-wiki.org) was used to collect settlement and cropland

information in Myanmar for a sample of 1km² pixels. The sample was taken at regular intervals of 10km across the country, resulting in 6557 unique sample points or roughly 10% of the area of Myanmar. The students were provided with a random set of pixels from this sample and were asked to identify the degrees of settlement and cultivation based on what they could see on the underlying Google Earth imagery. The degrees of both settlement and cultivation were chosen from a set of four classes: (1) none, (2) low, (3) medium or (4) high. Students were also asked to indicate their level of confidence in their choices, ranging from low to high. None of the participants had experience in land cover interpretation from satellite imagery, but all were trained by an expert to guarantee a certain level of quality in their interpretations. Additionally, they were provided with simple instructions for further reference. In total, around 29,000 pixels were collected, where each student collected 151 pixels on average, and every pixel was visited between 3 to 11 times, respectively. Data collection of multiple pixels at the same location allowed us to determine the quality of the crowdsourced geospatial data.

3 Results

3.1 Degree of cultivation and settlement

With respect to settlement, the majority of the sample pixels did not show any evidence of urbanisation (75.5%), roughly one fifth is sparsely settled (21.5%), and less than four percent show medium to high degrees of settlement. As settlement and cultivated areas in developing countries are closely connected to each other, the same pattern is apparent for cultivated area. The majority of the sample pixels indicated uncultivated areas (56.5%). However, almost one third of the analysed area is covered by cropland, showing medium (9.5%) and high (20.1%) degrees of cultivation (see Fig. 1).

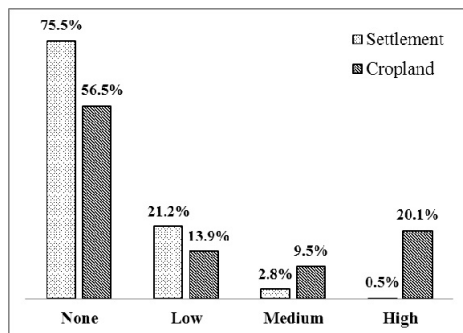


Fig. 1:
Percentage of cropland and settlement
for the 6557 sample pixels

Using ordinary kriging interpolation methods, we generated a cropland map for Myanmar based on the sample points. We excluded all pixel values below 10% cropland, as we assumed them to be highly uncertain. The map illustrates that large parts of the country are not or sparsely cultivated, confirming previous results. The final map is shown in Fig. 2.

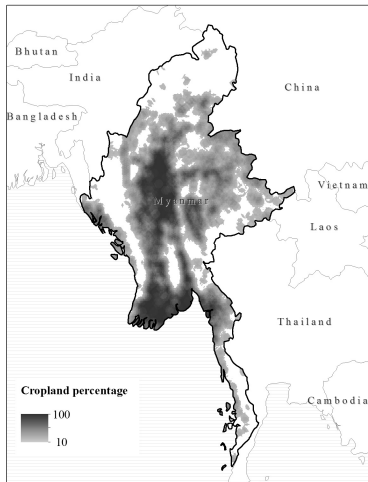


Fig. 2:
Cropland map of Myanmar based on interpolation of the crowdsourced data

3.2 Quality assurance of the crowdsourced data

With respect to the user-indicated confidence, our analysis revealed that the majority of data was collected with high confidence (i.e. 80%), where only 4% had low confidence attached. We then calculated a weighted average based on the multiple classifications and confidences at each unique location, for each class (none to high), and land cover type (settlement or cultivation), in order to estimate the reliability of the data at each location. For example, if all users agreed on one class at a given location, it is most likely that it represents the actual land cover situation. In contrast, if different classes were chosen for one point, the land cover type is less likely, showing a lower weighted average confidence. For cropland, the mean weighted average confidence was estimated at roughly 80%. Results for settlement are higher at 85%. Based on these results we are confident that cropland and settlement are classified correctly at 3634 and 4141 locations respectively.

3.3 Accuracy assessment of cropland

We performed an accuracy assessment to compare our crowdsourced cropland information with a recently developed global hybrid cropland map (hereafter: hybrid map, FRITZ et al. 2013b). The hybrid map was considered as ground truth or reference data (Producer), while the crowdsourced dataset represented the classification data (User). The analysis revealed an overall accuracy of 77.1%, having users and producer's accuracies ranging from 72 to 82%. We further calculated Cohen's Kappa (κ -value), a statistical estimate for inter-rater-agreement, which takes the agreement by chance into account. Comparing the hybrid map and the crowdsourced dataset we derived a κ -value of 0.54, indicating moderate agreement between the datasets. This is higher than the κ -value derived for other global land cover products like GLC2000 ($\kappa=0.38$), GlobCover2005 ($\kappa=0.45$) and MODIS2010 ($\kappa=0.47$), all showing lower agreement with our crowdsourced data.

4 Discussion and Conclusion

By running a crowdsourced campaign as a research project at the University of Vienna, we were able to capture roughly 10% of the area of Myanmar. Due to our methodological design and the participation of nearly 100 students, we were able to determine the quality of the crowdsourced data. This is usually hard to implement due to a lack of participants (e.g. see ALBRECHT et al. 2013). Based on our analysis, we are confident that more than 50% of the points were classified correctly. When comparing our dataset with a recently developed hybrid cropland map (FRITZ et al. 2013b), we revealed moderate agreement. In contrast, the κ -values of GLC2000, GlobCover2005 and MODIS2010 revealed low agreement, which is most likely due to the disagreement in these global land cover products as outlined in e.g. FRITZ et al. (2011) and FRITZ et al. (2013b). Our preliminary results give us confidence that crowdsourcing is an appropriate tool for gathering up-to-date land cover information of areas that are, for example, not easily accessible or restricted. Further, we are confident that we could use this information to try to understand where global land cover products are problematic, so that these areas can be improved in the future.

References

- ALBRECHT, F., PERGER, C., SCHILL, C., SEE, L., MCCALLUM, I., COLLINS, R. & FRITZ, S. (2013), Using Crowdsourcing to Examine Land Acquisitions in Ethiopia. In: JEKEL, T., CAR, A., STROBL, J. & GRIESEBNER, G. (Eds.), *GI_Form 2013: Creating the GISociety. Conference Proceedings*. Berlin/Offenbach, Wichmann, 100-104.
- CLERY, D. (2011), Galaxy Zoo Volunteers Share Pain and Glory of Research. *Science*, 333, 173-175.
- FRITZ, S., MCCALLUM, I., SCHILL, C., PERGER, C., GRILLMAYER, R., ACHARD, F., KRAXNER, F. & OBERSTEINER, M. (2009), Geo-Wiki.Org: The Use of Crowdsourcing to Improve Global Land Cover. *Remote Sensing*, 1, 345-354.
- FRITZ, S. et al. (2013a), Downgrading Recent Estimates of Land Available for Biofuel Production. *Environmental Science & Technology*, 47, 1688-1694.
- FRITZ, S. et al. (2013b), The Need for Improved Maps of Global Cropland. *Eos, Transactions American Geophysical Union*, 94, 31-32.
- FRITZ, S., YOU, L., BUN, A., SEE, L., MCCALLUM, I., SCHILL, C., PERGER, C., LIU, J., HANSEN, M. & OBERSTEINER, M. (2011), Cropland for sub-Saharan Africa: A synergistic approach using five land cover data sets. *Geophysical Research Letters*, 38, L04404.
- SEE, L. et al. (2014), *Mapping human impact using crowdsourcing*. Dordrecht/Heidelberg, Springer.
- SEE, L., MCCALLUM, I., FRITZ, S., PERGER, C., KRAXNER, F., OBERSTEINER, M., DEKA BARUAH, U., MILI, N. & RAM KALITA, N. (2013), Mapping cropland in Ethiopia using crowdsourcing. *International Journal of Geosciences*, 4, 6-13.
- WOOD, C., SULLIVAN, B., ILIFF, M., FINK, D. & KELLING, S. (2011), eBird: Engaging Birders in Science and Conservation. *PLoS Biol* 9, e1001220.