Exploring the potential role of volunteer citizen sensors in land cover map accuracy assessment

Giles M. Foody[1] *and Doreen S. Boyd*[1]

[1] School of Geography, University of Nottingham, Nottingham, NG7 2RD, UK
giles.foody@nottingham.ac.uk, doreen.boyd@nottingham.ac.uk

Abstract

Two sources of volunteered geographical information are used to provide reference data on forest to inform the evaluation of the accuracy of the Globcover land cover map for West Africa. Ground based photographs acquired through an internet-based collaborative project were interpreted by a further set of four volunteers to provide reference data on forest. Little agreement was found between a set of four volunteer interpreters but the set of labels they generated enabled a latent class model to be used to estimate forest cover. The latter is a key environmental variable as well as a basic indicator of accuracy on a non site-specific basis. The lack of gold standard reference makes detailed assessment difficult but the close correspondence between the model derived estimate of forest cover and that depicted in the map suggest the approach, based on volunteered data, may have value in accuracy assessment.

Keywords: remote sensing, citizen sensing, neogeography, crowd-sourcing, accuracy, latent class model, forest, Globcover.

1. Introduction

Land cover is a critical environmental variable. It has, for example, direct effect on numerous environmental processes and strongly impacts on human health and well-being. Land cover change is also a major variable of interest being amongst other things, the single most important variable affecting ecological systems and the greatest threat to biodiversity (Vitousek, 1994; Chapin *et al.*, 2000). There is, therefore, considerable need for accurate and timely land cover data and commonly the only practical means to derive this information is via satellite remote sensing.

Although remote sensing has considerable potential as a source of information on land cover there are many challenges to address in production of accurate land cover maps which allow the accurate estimation of variables such as class extent (Cihlar, 2000; Foody, 2002; Gallego, 2004; Rindfuss *et al.*, 2004). It is now widely accepted that any land cover map derived from remotely sensed imagery should be accompanied by a rigorously derived statement of its accuracy otherwise it remains an untested hypothesis of limited utility (Strahler *et al.*, 2006). The evaluation of map accuracy is, however, a non-trivial task. Many problems are encountered, often including issues connected to the quality and quantity of the reference data used to evaluate the map (Carlotto, 2009; Foody, 2010; Pontius and Petrova, 2010). The main recommended approach to accuracy assessment (Strahler *et al.*, 2006) requires the collection of a potentially large sample of high quality ground reference data at sites selected in accordance to carefully designed probability sampling design. The acquisition of an appropriate reference data set is, however, challenging and alternative approaches may sometimes be needed.

Some of the problems associated with the ground reference data may be tackled through the use of volunteered information from citizen sensors (Fritz *et al.*, 2012). Citizen sensing has emerged recently and offers immense opportunities and challenges to contemporary geography (Goodchild, 2007; Hudson-Smith *et al.*, 2009; CSDGSND, 2010; Voigt *et al.*, 2011). It offers, for example, the ability to acquire large volumes of data inexpensively over large areas which can be of great value to tasks such as map production and validation. There are, however, many concerns, including the variable, and typically unknown, quality of the data provided by volunteers together with a set of ethical and legal concerns associated with the collection and use of the data (Goodchild and Glennon 2010; CSDGSND, 2010).

In this article, the potential role of volunteered information in accuracy assessment is explored in the context of mapping forests in West Africa. Information on forest extent is required for numerous applications. For example, changes in forest extent caused by deforestation have a major impact on the global carbon budget and so accurate information is required for major international programmes such as the UN REDD+ (Woodcock *et al.*, 2010). Two sources of volunteered information are used here to support an evaluation of citizen sensor-based reference data with regard to the Globcover global land cover map: photographic data from a web-based collaborative project and interpretations by a further set of volunteers. Recognizing that this labeling process may contain potentially large amounts of error, the value of using multiple volunteers to interpret and label each site is explored. From the resulting set of labels, a latent class analysis (Magidson and Vermunt, 2004; Foody, 2012) is used to estimate the extent of forest cover over the region of study. Forest cover is a key environmental variable and also be used as a non site-specific measure of classification accuracy.

2. Data and methods

Attention focused on mapping forest in West Africa, specifically the accuracy with which forest was depicted in the European Space Agency's Globcover map (Bicheron *et al*., 2008). There are many definitions of forest (Comber *et al*., 2005; Alqvist, 2008) and here a site was considered to be forested if it had at least 15% canopy cover. Thus, forest comprised a number of classes depicted in the Globcover map; representing an aggregation of mosaic, closed and open forest classes.

Two sources of volunteered data were used. First, freely available photographic data on ground conditions at the point of intersection of lines of latitude and longitude provided by the Degrees of Confluence project (www.confluence.org) were used as a source of spatially extensive information on ground cover. In brief, volunteers visit the confluence points and at each acquire a set of photographs; usually four photographs viewing north, south, east and west from the confluence point. Additional data is often obtained, including evidence of the photographs having been acquired at the correct location and text that could also reveal a lot of useful information about the site. Here, attention is focused solely on the set of photographs acquired at or near the point of confluence (Figure 1). The photographs for all 99 successfully visited and documented confluence points available (winter 2011) for the Ivory Coast, Ghana, Togo, Benin and Nigeria were used. A second set of data was derived from volunteers at Nottingham University who labelled each

Figure 1. Examples of photographs for confluence 11°N 1°W (obtained from www.confluence.org and reproduced with permission of R. Mautz).

confluence point based on the set of photographs available for it. Here, attention focused solely on whether or not the site was forest, using the above definition of forest. Each site was labelled independently by a total of four people (A-D).

For each point of confluence, a set of four labels indicating forest presence (1) or absence (0) was, therefore, generated. These labels supplemented the label indicated on the Globcover map (G). The labels derived by each volunteer were compared against those in the map to provide a guide to the degree of agreement between the volunteer derived reference data and the mapped depiction. It must be noted that this assessment is one of only the degree of agreement as the lack of an true gold-standard reference data set means accuracy cannot be assessed. It should be noted, however, that the Globcover map has an estimated accuracy of ~73% (when assessed using relatively certain reference data; Bicheron *et al*., 2008). Given the focus on agreement between different sources of class labels, the proportion of cases upon which labelling agrees and the kappa coefficient of agreement were used to quantitatively summarise agreement. The latter, despite its unsuitability as an index of accuracy in remote sensing (Pontius and Millones, 2011) is suitable as a measure of inter-rater agreement and was derived from

$$\hat{\kappa} = \frac{p_o - p_c}{1 - p_c}$$

where p_o is the observed proportion of agreement and p_c the proportion of agreement expected by chance.

In addition to assessing the degree of agreement between each volunteer's labels and the mapped labels on a pairwise basis, the set of four labels derived from all volunteers were combined to yield information on forest cover and the accuracy of its representation. A key focus was on the estimation of forest extent in the absence of reference data while recognising that the labels provided by each volunteer are likely to be imperfect. The potential to use data volunteered from a variety of sources to provide useful information on forest extent and the accuracy of the Globcover map was explored with the use of a latent class model (Magidon and Vermunt, 2004). A standard latent class model for a scenario in which the true status on forest occurrence is represented by a single latent (unobserved) variable, F, in which there are two latent classes, forest (1) and non-forest (0) and based on the outputs of four independent labellers A-D, may be written as

$$\pi_{abcdt} = \pi_t^F \pi_{abcdt}^{ABCD|F}$$

where the outputs from volinteers A-D are labels $a\text{-}d = 0{,}1$, $\pi_{abcdt}^{ABCD|F}$ is the conditional probability that the pattern of class labels derived is (a,b,c,d) given that the case has a status t (1 or 0) and π_t^F is the probability that a case has the status t (Vermunt, 1997; Yang and Becker, 1997; Magidson and Vermunt, 2004). The conditional probabilities that represent the producer's accuracy for each volunteer are parameters of the latent class model yielding an ability to estimate the accuracy of the individual volunteered data sources. This latter issue has been demonstrated for situations in which a high quality reference data set is available for validation (Foody, 2012). More critically for this article, the probability of a point being forested, is also a parameter of the model. Thus, if a model can be constructed that fits with the observed data one of its parameters is an estimate of the extent of the class of interest which is also a non site-specific measure of map accuracy. This variable may, therefore, be estimated without reference data and from a set of imperfect labels derived from volunteers. Since, it would be expected that the volunteers would tend to correctly label the cases that were clearly non-forest and struggle most with the labeling of points that had some tree cover, probably close to the 15% threshold value, a degree of correlation in their labeling was expected. To accommodate for this, the standard latent class model was adjusted to allow for the expected conditional dependence between the volunteers.

3. Preliminary results and discussion

The set of photographs for each confluence point were used to generate four independently-derived sets of labels. The four volunteers varied greatly in terms of labelling. The degree of agreement between pairs of labellers varied between 62.66 and 79.79% or, in terms of the kappa coefficient from 0.282 to 0.556 (Table 1). These relatively low levels of agreement highlight some concerns for the use of volunteered data in accuracy assessment. Although interpretation is limited by the absence of a gold-standard reference, the low levels of agreement may suggest that the individual volunteers differ greatly in terms of their perception of forest.

Table 1: Kappa coefficients indicating degree of inter-rater agreement.

Source	B	C	D	G
A	0.367	0.495	0.553	0.235
B	-	0.556	0.282	0.170
C		-	0.458	0.556
D			-	0.154

A similar wide range of agreement to the labels depicted on the Globcover map was also evident, with agreement varying from 58.58% to 77.77% or 0.170 to 0.556 in terms of the kappa coefficient. These results suggest only relatively low degrees of agreement in the labelling provided by the different sources. Indeed, the low levels of agreement may suggest that none of the individual volunteers was able to provide data that could be confidently used in accuracy assessment. There are many potential sources of uncertainty and error in the analysis including problems with the photographic reference data, its labelling, the time period between the mapping and photography as well as error in the Globcover map.

The estimate of the extent of forest that may be derived from the the volunteered labels varied from 32.33% to 57.57% of the region. Although Globcover has an estimated accuracy of ~73%, the extent of forest based on its representation is 50.50%. Thus, the volunteered data provides estimates around a value that seems reasonable but the range of estimates was wide and there is no obvious source of information to distinguish between the volunteers in terms of labelling quality.

A variety of approaches might be used to try and derive an enhanced estimate of forest extent and map accuracy. For example, if there was confidence in the reference data set an estimate may be derived from each confusion matrix using a method such as that discussed by Card (1982). Alternatively, a simple approach such as taking the average of the 4 labellers could be adopted. This approach yielded an estimate of 42.67%. Similarly, it would be possible to combine the individual classifications in a basic ensemble method. For example, by allocating each point to the class associated with the most frequently allocated label, and making a random allocation in the case of tie, yielded an estimate of 40.40%. Here, however, attention is focused on the potential of the latent class modelling approach. Using the set of four class labels for each point derived by the volunteers (A-D), a latent class model based on equation 2 yielded an estimate of π_t^F that suggests 50.15% of the region was covered with forest. In the absence of gold-standard reference data it is difficult to evaluate the quality of this estimate but its closeness to the value derived from Globcover (50.50%) gives some reassurance to the quality of the mapped representation, although the asymmetry in the omission and commission errors (Bicheron *et al*., 2008) suggest that Globcover's value may be biased.

The results indicate that imperfect data arising from volunteered sources has the potential to provide useful information on map properties and accuracy. It should also be noted that the volunteered data have other advantages. For example, the systematic sample design used ensures that data are derived over a widespread area and could be used to aid map production (e.g. by provision of prior information on class abundance) and, as variance terms may be defined for data acquired by this design, the data are suitable for many map accuracy and comparison objectives.

4. Conclusions

Recent advances in geoinformation technology, notably remote and citizen sensing give immense opportunities to map and monitor the Earth's environment and address recognised international research priorities. The potential of volunteered data to aid the evaluation of land cover maps was explored and it was noted that:

1. Web based collaborative activities such as the Degrees of Confluence project can provide useful spatially extensive data to aid map evaluation.
2. The photographs acquired at each confluence point may be interpreted by further volunteers to provide class labels of value to the map evaluation in-hand. Here, volunteers were used to label each point as forest or non-forest.
3. Recognising that the volunteered data is likely to be imperfect, latent class analysis was used to derive estimate of forest cover from a set of labels derived from multiple observers.
4. While there was relatively low levels of inter-rater agreement and each volunteer's labels showed poor agreement with the Globcover map, it was evident that the estimate of forest cover derived from the latent class analysis was close to that depicted in the map. Although there are many uncertainties, this suggests that the potential of using volunteered data together with methods such as latent class analysis are worthy of further exploration.

Acknowledgments

We are grateful to the volunteers, notably Rainer Mautz and other contributors to the Degrees of Confluence and students at Nottingham University as well as ESA and the ESA GlobCover Project, led by MEDIAS-France, for GlobCover products.

References

Ahlqvist, O. (2008) "Extending post-classification change detection using semantic similarity metrics to overcome class heterogeneity: a study of 1992 and 2001 US National Land Cover Database changes", *Remote Sensing of Environment*, vol. 112: 1226-1241.

Bicheron, P., Defourny, P., Brockmann, C., Schouten, L., Vancustem, C., Huc, M., Bontemps, S., Leroy, M., Achard, F., Herold, M., Ranera, F, Arino, O.(2008) *Globcover: Products Description and Validation Report*, Medias France, Toulouse, 47pp.

Card, D. H. (1982) "Using map category marginal frequencies to improve estimates of thematic map accuracy". *Photogram. Eng. Remote Sens*, vol. 49: 431–439.

Chapin, F.S., *et al*.(2000) "Consequences of changing biodiversity", *Nature* vol 405:234–42.

Carlotto, M. J. (2009) "Effect of errors in ground truth on classification accuracy", *International Journal of Remote Sensing*, vol. 30: 4831-4849.

Comber, A., Fisher, P., Wadsworth, R (2005) "What is land cover?" *Environment and Planning B*, vol. 32: 199-209.

CSDGSND (Committee on Strategic Directions for the Geographical Sciences in the Next Decade), (2010) *Understanding the Changing Planet: Strategic Directions for Geog. Sciences*, National Academies Press.

Foody, G. M. (2002) "Status of land cover classification accuracy assessment", *Remote Sensing of Environment*, vol. 80: 185-201.

Foody, G. M. (2010), "Assessing the accuracy of land cover change with imperfect ground reference data", *Remote Sensing of Environment*, vol. 114: 2271-2285.

Foody, G.M.(2012) "Latent class modelling for site and non-site specific classification accuracy assessment without ground data", *IEEE Trans. on Geosci. & Rem. Sen.* (in press).

Fritz S., McCallum, I., Schill, C., Perger, C., See, L., Schepaschenko, D., van der Velde, M. Kraxner, F., Obsersteiner, M. (2012) "Geo-Wiki: An online platform for improving global land cover", *Environmental Modelling and Software*, vol. 31: 110-123.

Gallego, F. J. (2004) "Remote sensing and land cover area estimation", *International Journal of Remote Sensing*, vol. 25: 3019-3047.

Goodchild, M. F. (2007) "Citizens as sensors: the world of volunteered geography", *GeoJournal*, vol. 69: 211-221.

Goodchild, M. F., Glennon, J. A. (2010) "Crowdsourcing geographic information for disaster response: a research frontier", *International Journal of Digital Earth*, vol. 3: 231-241.

Hudson-Smith, A., Batty, M., Crooks, A., Milton, R. (2009) "Mapping for the masses. Accessing Web 2.0 through crowdsoucring",*Social Sci Comp. Review*, vol.27: 524-538.

Magidson, J., Vermunt, J.K. (2004) "Latent class models", In Kaplan, D (ed) *The SAGE Handbook of Quantitative Methodology for the Social Sciences*, Sage, 175-198.

Pontius, R. G., Millones, M. (2011) "Death to kappa: birth of quantity disagreement and allocation disagreement for accuracy assessment", *International Journal of Remote Sensing*, vol. 32: 4407-4429.

Pontius, R. G., Petrova, S. H. (2010) "Assessing a predictive model of land change using uncertain data", *Environmental Modelling and Software*, vol. 25: 299-309.

Rindfuss, R. R., Walsh, S. J., Turner Ii, B. L., Fox, J. and Mishra, V. (2004) "Developing a science of land change: challenges and methodological issues", *Proceedings of the National Academy of Sciences USA*, vol. 101: 13976-13981.

Strahler, A. H., Boschetti, L., Foody, G. M., Friedl, M. A., Hansen, M. C., Herold, M., Mayaux, P., Morisette, J. T., Stehman, S. V. and Woodcock, C. E. (2006) *Global Land Cover Validation: Recommendations for Evaluation and Accuracy Assessment of Global Land Cover Maps*, Technical Report, Joint Research Centre, Ispra, EUR 22156 EN, 48pp.

Vermunt, J. K. (1997) *LEM: A General Program for the Analysis of Categorical Data, Department of Methodology and Statistics, Tilburg University, The Netherlands.*

Vitousek, P.M. (1994) Beyond global warming–ecology and global change,*Ecol*,75,1861-76.

Voigt, S. *et al.* (2011) "Rapid damage assessment and situation mapping: learning from the 2010 Haiti earthquake", *Photogram. Eng. and Rem. Sen.*, vol. 77: 923-931.

Woodcock CE *et al.* (2010) "The effect of map accuracy on estimates of terrestrial carbon budget", *Proceedings of Accuracy 2010*, Leicester, 9-11.

Yang, I., Becker, M. P. (1997) "Latent variable modelling of diagnostic accuracy", *Biometrics*, vol. 53: 948-958.