

## AN ASSESSMENT OF CITIZEN CONTRIBUTED GROUND REFERENCE DATA FOR LAND COVER MAP ACCURACY ASSESSMENT

G. M. Foody<sup>a</sup>

<sup>a</sup> School of Geography, University of Nottingham, Nottingham, NG7 2RD, UK - giles.foody@nottingham.ac.uk

**KEY WORDS:** Citizen sensors, volunteered geographic information (VGI), neogeography, classification accuracy

### ABSTRACT:

It is now widely accepted that an accuracy assessment should be part of a thematic mapping programme. Authoritative good or best practices for accuracy assessment have been defined but are often impractical to implement. Key reasons for this situation are linked to the ground reference data used in the accuracy assessment. Typically, it is a challenge to acquire a large sample of high quality reference cases in accordance to desired sampling designs specified as conforming to good practice and the data collected are normally to some degree imperfect limiting their value to an accuracy assessment which implicitly assumes the use of a gold standard reference. Citizen sensors have great potential to aid aspects of accuracy assessment. In particular, they may be able to act as a source of ground reference data that may, for example, reduce sample size problems but concerns with data quality remain. The relative strengths and limitations of citizen contributed data for accuracy assessment are reviewed in the context of the authoritative good practices defined for studies of land cover by remote sensing. The article will highlight some of the ways that citizen contributed data have been used in accuracy assessment as well as some of the problems that require further attention, and indicate some of the potential ways forward in the future.

### 1. INTRODUCTION

The assessment of the accuracy of thematic maps such as those depicting land cover obtained via remote sensing has evolved considerably over the last four decades (e.g. Foody, 2002; Congalton and Green, 2009). It is now widely accepted that an accuracy assessment should be part of land cover mapping programmes. This is primarily because without an accuracy assessment each map produced is simply an untested hypothesis, one of many possible representations of the world which may or may not be fit for its intended purpose (Strahler et al., 2006). This is important as it is now very simple to produce thematic maps from remote sensing. Indeed there are, for example, numerous global land cover maps available but they do differ markedly in their representation and it is sometimes difficult to know which is the most suitable one to use in an application or how to best use the set without information on accuracy (Giri et al., 2005; Jung et al., 2006; McCallum et al. 2006; Fritz and See, 2008; Critically, a map is not suited for scientific inference without a rigorous assessment of its quality, leaving the map as little more than a pretty picture (M<sup>c</sup>Roberts, 2011).

An accuracy assessment may be used to do more than simply indicate the quality of a land cover map. Critically, an accuracy assessment may also be used to add value to the land cover map. By undertaking a rigorous accuracy assessment it may, for example be possible to refine estimates of the areal extent of land cover classes that occur within the mapped region. The latter can have a major impact on, amongst other things, estimates of the magnitude and direction of land cover changes (Olofsson et al., 2013) and ecosystem services valuations (Foody, 2015).

Good practices for land cover map accuracy assessment have been established (Strahler et al., 2006; Olofsson et al., 2013, 2014). Additionally adaptable resources have been made

available to the community to facilitate rigorous accuracy assessment (e.g. Olofsson et al., 2012). However, map accuracy assessment remains a challenging task. One fundamental problem is that an accuracy assessment requires ideally a gold standard reference data set to compare against the map(s) being evaluated (Foody, 2010, 2013). Frequently, however, the ground reference data are flawed in relation to their quantity and quality, which can impact negatively on accuracy assessment. There is also often a negative relationship between data quantity and quality, making it difficult to acquire a large high quality data set. It may, however, be possible for citizens to help reduce some of the problems by providing reference data.

Citizens have contributed to scientific research for centuries. However, recent technological developments, notably web2.0 which facilitates the collaboration and interaction of people with each other including two way data transfers and growth of user-generated content, combined with the proliferation of inexpensive location-aware devices have led to dramatic growth in citizen sensing and participation in collaborative volunteer projects (Arsanjani et al., 2015a). Citizen science research has grown enormously in recent years, revolutionizing parts of geography and forming a key component of future research priorities in the subject (CSDGSND, 2010).

It is now possible for almost anyone anywhere in the world to provide spatially located information that may be used to inform a diverse array of research and practical applications. This has seen the recent rise of citizen sensors and the provision of volunteered geographic information (VGI; Goodchild, 2007) to add to more conventional crowdsourcing activity. These various sources of citizen contributed data can differ greatly in detail, ranging from altruistic volunteering to paid crowdsourcing. In this article, attention is focused on data that originates from citizens who are typically, but not necessarily, amateurs and acting voluntarily. Moreover, the data contributed may have

been provided unintentionally and normally for little if any reward (e.g. data mined from social media *etc.*) or deliberately in response to a call for information from, perhaps, an authoritative mapping body.

Citizens have the potential to become a major source of reference data for accuracy assessment. Issues of data quantity and quality may remain but given the existence of authoritative best practices it should be possible to gauge how suitable citizen contributed data are for some accuracy assessment tasks. This article first outlines the authoritative good practices for ground data collection in accuracy assessment. It will highlight the limitations of authoritative data sets before considering the relative merits of citizen contributed data for accuracy assessment. The focus is on only the use of the reference data in accuracy assessment but it should be noted that citizen derived data could, of course, be used in other parts of a mapping programme (e.g. for use in training a supervised classification of remotely sensed imagery). Additionally only conventional 'global' accuracy assessment is discussed although it should be noted that the broad geographic coverage that can be provided by citizens can be used to indicate spatial variation in map quality (e.g. Comber et al., 2013).

## 2. GOOD PRACTICES FOR REFERENCE DATA ACQUISITION

Remote sensing has considerable potential for the provision of environmental information for thematic mapping applications at a range of spatial and temporal scales (Foody and Curran, 1994; Cihlar, 2000; Wulder et al., 2008). Thematic maps are typically derived from remotely sensed imagery through a digital image classification (Mather and Koch, 2011). In this type of analysis it is typically assumed that the classes are discrete and mutually exclusive as well as exhaustively defined. These assumptions are not always satisfied, often leading to negative impacts on land cover map accuracy (e.g. Foody, 1996; Foody, 2004; Rocchini et al., 2013). However, in many cases the problems can be addressed and useful representations of land cover obtained.

Frequently supervised digital image classification analyses are used to obtain land cover maps from remotely sensed imagery. Beyond fundamental issues such as image pre-processing, such classification analyses comprise three stages: training, class allocation and testing. Issues connected with each stage can greatly impact upon the quality of the classification and hence the resulting map. Here, the focus is entirely on the reference data used in the final, testing, stage of the classification that seeks to indicate the quality of the classification, normally in terms of its accuracy.

Authoritative statements on good practice for land cover map accuracy assessment have been defined (Strahler et al., 2006; Olofsson et al., 2014). An accuracy assessment has three major components, namely the response design, the sampling design and the analysis (Stehman and Czaplewski, 1998). These apply to accuracy assessments using authoritative and/or citizen contributed data.

The response design sets out the protocol to determine if the class label depicted in the land cover map under evaluation is in agreement with the label contained in the ground reference data set. It includes issues such as the selection of the spatial unit (e.g. pixel, block or object) and the sources of information (e.g. reference data could come from field visits, inventories, aerial

photograph analysis *etc.*). The effect of error and uncertainty should also be considered. It may, for example, be useful to have each case labelled by multiple interpreters to give a guide to the quality of the reference data and to aid the definition of agreement (e.g. should only cases for which all interpreters agree on a label be used in an accuracy assessment, should secondary labels and certainty information be used *etc.*). The reference labelling protocol must also be defined which may be associated with challenges linked to the minimum mapping unit. Finally, while agreement may seem a simple concept there are many issues that require careful attention. These include problems linked to the ability to correctly locate a site geographically in both the land cover map and on the ground as well as the effects of inter-rater uncertainty in labelling and semantics. Further details on this, and the other, parts of an accuracy assessment are given in the literature (e.g. Stehman and Czaplewski, 1998; Strahler et al., 2006; Olofsson et al., 2014).

As the evaluation of classification accuracy cannot normally be undertaken for the entire map it is usual to base the assessment on a sample of cases. To ensure a statistically rigorous and credible accuracy assessment it is important that the sample used for the accuracy assessment is acquired following an appropriate design. Good practice recommendations call for the assessment to be based on the use of probability sampling. A range of designs are available, with choice between them often based on the accuracy objectives and key design criteria. Popular approaches include the use of simple random, stratified, systematic and cluster sampling. For each sampling design, recommendations on key factors such as suitable sample size may also be followed to ensure the sample meets the goals of a mapping project (e.g. Stehman, 1999, 2009, 2012).

As a crude summary, the required sample for an authoritative accuracy assessment can be defined following simple rules and recommendations (Stehman and Czaplewski, 1998; Strahler et al., 2006; Olofsson et al., 2014). The size of the sample, for example, may be estimated from sampling theory (Foody, 2009) or heuristics such as those that suggest at least 50 cases per-class acquired via an appropriate sample design (Congalton and Green, 2009). For example, if a simple random sampling design was to be used, the required sample size may be estimated from equation 1.

$$n = \frac{z_{\alpha/2}^2 P(1-P)}{h^2} \quad (1)$$

where  $P$  is a planning value for the population proportion of correctly allocated cases,  $h$  the half width of the desired confidence interval and  $z_{\alpha/2}$  the critical value of the normal distribution for the two-tailed significance level  $\alpha$  (Cochran, 1977).

The approach can be adapted to meet specific project needs. If, for example, the objective is to test the statistical significance of differences in map accuracy, perhaps in evaluating a set of different mapping approaches, the required sample size can be estimated using the same basic principles. For this, however, the probability of detecting a specified effect, which represents the minimum meaningful difference in classification accuracy, is represented by the power of the test,  $1-\beta$  (Fleiss et al. 2003). With  $\alpha$ ,  $1-\beta$  and the effect size selected, the required sample size from each of the populations being compared may be estimated using equation 2.

$$n = \frac{(z_{\alpha/2} \sqrt{2\bar{P}\bar{Q}} + z_{\beta} \sqrt{P_1Q_1 + P_2Q_2})^2}{(P_2 - P_1)^2} \quad (2)$$

in which  $\bar{P} = \frac{P_1 + P_2}{2}$  and  $\bar{Q} = 1 - \bar{P}$ . In this type of comparative study it is important to note that the sample size should be determined with care as sizes too small and too large can be problematic (Foody, 2009).

In many instances simple random sampling is not ideal. In such cases other designs may be used and a variety of probability designs are available, notably the use of stratified, systematic and cluster sampling designs. For each design, the sample size required may be calculated and this may be optimized to meet the specific objectives of a study (e.g. Stehman, 2012). Again the basis is straightforward with, for example, the size of the sample for stratum  $i$ ,  $n_i$ , in a stratified random sample of fixed size  $n$  estimated using equation 3.

$$n_i = \frac{nW_i \sqrt{P_i(1-P_i)}}{\sum_{i=1}^k W_i \sqrt{P_i(1-P_i)}} \quad (3)$$

in which  $W_i$  is the stratum weight determined from  $W_i = \frac{N_i}{N}$  where  $N$  is the population size and  $N_i$  the size of stratum  $i$ . This approach can be adapted to fit the specific circumstances of a study, such as variations in the cost per-stratum or project objectives (Cochran, 1977; Barnett, 1991; Som, 1996; Stehman, 2012).

In the analysis stage the aim is typically to obtain rigorous and credible accuracy information. This typically draws on analysis of the error matrix or confusion matrix that shows a cross-tabulation of the map and ground reference data labels for the sample of cases used. A range of quantitative measures of accuracy can be obtained from the matrix and it is important that the accuracy assessment takes into account the nature of the data used. Ideally, therefore, the error matrix, together with key information on issues such as the sample design used in its formation, should be reported in the output of an accuracy assessment. This allows other users to obtain information that they may need (e.g. for the calculation of standard errors and confidence intervals) but also because the matrix may be used to help refine estimates of key properties such as the areal extent of classes and so add value to the map (Olofsson et al., 2013). The formulae used to estimate accuracy values and their associated variances need to be selected in relation to the sample design used to acquire the data. Formulae for popular designs such as simple random, stratified random and cluster sampling are provided in Stehman and Foody (2009).

A rigorous accuracy assessment provides not only information on map quality but also means to enhance the value and usefulness of the map. The ability of an accuracy assessment to add value to a map can be illustrated with examples. In showing how a confusion matrix used for accuracy assessment can also aid accurate estimation of class areal extents Olofsson et al. (2013) provide an example focused on the estimation of the extent of deforestation in a region. In this example, a highly accurate map, its overall accuracy was ~94%, suggests that 22,353 ha of the study region had been deforested. However, adjusting for even the low levels of error present, the actual

areal extent was double what the map showed, at 45,651 ha. Not only is the difference large it has important implications to the carbon budget of the region as outlined by Olofsson et al. (2013). Similarly, Foody (2015) shows how errors in a land cover classification can have a large impact on valuations of ecosystem services. For example, using the National Land Cover Data (NLCD) set for the conterminous USA, which is ~84% accurate (Wickham et al., 2013), directly in a basic transfer function approach to ecosystem services valuation provides an estimate of US\$1118 billion yr<sup>-1</sup>. Adjusting the estimate for the pattern of mis-classification evident in the confusion matrix used in the assessment of map accuracy, however, shows that the value of the ecosystem services is markedly lower, at US\$600 billion yr<sup>-1</sup>. The pattern of error and the differential value of the classes will determine the size and direction of the change in value that arises when adjusting for the effects of mis-classification bias. For example, at a global scale the value of wetlands estimated from the IGBP DISCover land cover map rises from US\$1.92 trillion yr<sup>-1</sup> to US\$2.79 trillion yr<sup>-1</sup> when adjustment is made for classification error (Foody, 2015).

Although the demands made by good practice documents may not seem onerous or problematic it is often difficult to acquire a ground reference data set in strict accordance to the authoritative good practices. Consequently, it is often impractical to follow the good practices. The sample used is often of inappropriate size and/or quality, impacting negatively on the accuracy assessment (Foody, 2009, 2010, 2013).

Even if the concerns with issues such as the sampling design have been addressed satisfactorily there are still other concerns, notably those linked to the quality of the reference data. Typically the ground reference data are used in an accuracy assessment as if perfect (i.e. that they are a gold standard or ground truth). Sometimes it is recognized that the ground reference data are flawed but the analysis proceeds as if it is perfect. This can be a dangerous situation in an accuracy assessment. It is possible for even small errors in the ground reference data set to be a source of substantial error and mis-interpretation in an accuracy assessment (Carlotto, 2009; Foody, 2013). For example, in a study of land cover change the effects of even very small reference data errors led to substantial mis-estimation of both classification accuracy and of the area of land undergoing change (Foody, 2013). As one example, for one simple scenario in which the area of a rare land cover change that actually occurs in 0.5% of the study area will be exaggerated by ~40 times if the ground data and land cover map used have an accuracy of 80% and 70% respectively. Fortunately, however, it is sometimes possible to address the effects of ground reference data error and obtain accurate estimates of map accuracy and class extent if the ground reference data error is well-known and characterized (Foody, 2010).

### 3. POTENTIAL OF CITIZEN CONTRIBUTED DATA

Some of the problems commonly encountered with ground reference data sets, even from highly authoritative sources, can potentially be addressed in a variety of way. At one extreme the effects of ground data error can, as noted above, be addressed directly if the error is well known and characterized. Alternatively, the problems of design-based accuracy assessment can, to some extent, be addressed by adopting model-based approaches. Standard components of design-based accuracy assessment, such as the confusion matrix and measures

of overall accuracy, are not encountered with model-based inference. However, the latter can be useful in relation to issues such as area estimation that can be used in some accuracy assessments (McRoberts, 2011). Additionally, as will be noted later in this section, modelling approaches can also provide a means to analyze imperfect data set such as those contributed by volunteers. Here, the main focus of attention is on how citizens could contribute to standard design-based accuracy assessment of land cover maps.

A key attraction of citizen sensing for ground reference data collection is their ability to contribute data at a range of spatial and temporal scales. Thus, citizen sensors could reduce or even possibly remove problems linked to ground data sample size, location and timing relative to image acquisition. Additionally, the data can arise in a range of different ways. Data could be contributed passively by exploiting information provided unintentionally, or actively in response to a request from a body that could steer the contributions to meet particular needs. An overview of the use of volunteered geographic information arising from citizen sensors in accuracy assessment is provided by Fonte et al. (2015a).

Although it may seem odd for citizens to contribute unknowingly to accuracy assessments this type of passive citizen sensing has occurred when members of the general public have uploaded photographs to sites such as Flickr or Panoramio. The photographs may have been added to the sites to share with friends and family, but they may have additional uses. Critically, the photographs also provide geolocated images that can be interpreted to yield land cover data that might be used as ground reference data in an accuracy assessment (Antoniou et al., 2010; Estima and Painho, 2013). Additionally, volunteers to projects such as OpenStreetMap may provide land cover data that could be used as reference data (Arjansani et al., 2015a; Estima and Painho, 2015). In a similar way, contributors to internet projects such as the Degree Confluence Project may provide unintentionally data that can be used in an accuracy assessment. These contributors visit the points of intersection of lines of latitude and longitude globally and take photographs of the site. The photographs acquired are available through the project website and may be interpreted to yield ground reference data for an accuracy assessment (Iwao et al., 2006; Foody and Boyd, 2013). Moreover, through the project the photographs of a site are up-dated enabling use through time. The systematic sampling design used is also compatible with best practice recommendations for accuracy assessment.

With active sensing, the citizens contributing data often do so to contribute to scientific research or practical applications. Critically a body seeking to assess the accuracy of a map can design key aspects of the accuracy assessment programme. For example, the sites to be visited for data collection could be specified following an appropriate probability sampling design for an accuracy assessment. The sampling approach can also be designed to fit with existing authoritatively defined data sets and resources, notably by blending the data sets and using explicitly adaptable resources such as the sample defined by Olofsson et al. (2012).

Moreover, given the recent growth of resources such as Google Earth that allow easy access to high quality and often fine spatial resolution imagery for the globe, the data collection need not involve fieldwork, although that can still be useful. A variety of internet based resources are available to help citizens label imagery that may be of anywhere on the planet from the

comfort of their own home (e.g. Fritz et al., 2012; Bastin et al., 2013).

Concerns with data quality can also be addressed to some degree. It is, for example, possible to have each site interpreted and labelled by multiple citizens which can aid some model-based analyses that can provide accuracy estimates (Foody et al., 2013). For example, latent class modelling allows estimates of the accuracy of the data contributed by citizens to be estimated from the data alone, without any reference data. The approach is based on the probability of observing the patterns of class allocation made by the set of citizens contributing to the task; each citizen need not label the exact same set of data as the approach can accommodate missing observations. The set of class labels provided by the citizens form the visible or manifest variables of the analysis and are used to provide information on the unobserved (latent) variable. In typical use, the set of citizens contributing,  $C$ , are each presented with a set of cases to label. The citizens may, for example, be presented with fine spatial resolution images for selected locations via an internet based system (e.g. Fritz et al., 2012). With,  $M_c$  representing one of the set of  $C$  manifest variables indexed  $1 \leq c \leq C$ , and its values are class labels represented by  $m_v$ , which lie in the range  $r(1-q)$  and using vector notation  $\mathbf{M}$  and  $\mathbf{m}$  to represent the complete response patterns (i.e.  $\mathbf{M}$  denotes  $(M_1, \dots, M_c)$  and  $\mathbf{m}$  denotes  $(m_1, \dots, m_q)$ ), the latent class model is that the probability of obtaining the response pattern  $\mathbf{m}$ , represented as  $Prob(\mathbf{M}=\mathbf{m})$ , is a weighted average of the  $q$  class-specific probabilities  $P(\mathbf{M}=\mathbf{m}|T=t)$  (Magidson and Vermunt, 2004). If the set of labels derived from each citizen can be assumed to be conditionally independent of those from all other citizens contributing labels, the latent class model may be written as equation 4.

$$Prob(\mathbf{M}=\mathbf{m}) = \sum_{t=1}^q Prob(T=t) \prod_{c=1}^C Prob(M_c = m_c | T=t) \quad (4)$$

in which  $Prob(T=t)$  is the proportion of cases belonging to latent class  $t$  (Yang and Becker, 1997; Vermunt and Magidson, 2003); the approach can often be readily adapted for situations in which there is dependence in the labelling. The quality of the model is generally illustrated by its fit to the data and this is commonly assessed with the likelihood ratio chi-squared statistic, with a model viewed as fitting the data if the calculated value of statistic obtained is sufficiently small to be attributable to the effect of chance (Magidson and Vermunt, 2004). Critically, this type of approach provides a means to assess the accuracy of maps without any reference data (Foody, 2012) and can also convey information on the quality of the citizens contributing data in terms of the accuracy of their labelling (Foody et al., 2013). The type of model may also allow the production of information on the confidence or certainty with which individual cases in the map have been classified that would be of value to some users. This could, for example, be used potentially to help illustrate the spatial variation in the uncertainty or quality of the labelling in a land cover map.

#### 4. CHALLENGES AND OPPORTUNITIES

Citizen contributed data has considerable potential for use in accuracy assessment but a range of challenges exist. The tension between the wisdom and power of the crowd versus mob rule are well-known (Roman, 2009). Before citizen contributed data become accepted widely for use in accuracy assessment a variety of concerns will need to be addressed. The latter extend well beyond the basic concerns with data quality and

trustworthiness, with problems connected with issues such as the location, timing and sustainability of data collection as well as a suite of legal and ethical concerns (Vandecasteele and Devillers, 2015; Arsanjani et al., 2015b). For example, the data sets obtained from citizens, especially that contributed unintentionally, may be acquired from highly unrepresentative samples.

A variety of approaches may be used to address the concerns with VGI. For example, markedly different approaches for assessing the quality of VGI are available (e.g. Goodchild and Li, 2012). Some approaches may simply follow a basic voting approach if there are multiple contributions on a particular case, others may have a hierarchy of contributors with some established and trusted people effectively acting as gatekeepers while others may make use of geographical contextual information to sense-check contributions or actually seek to infer quality from the data themselves.

As awareness of the challenges in using VGI grows there is increasing effort on methods to reduce problems and tentative steps to the definition of good practices for VGI collection are emerging (Fonte et al., 2015b). The issues are also not always straightforward. For example, some citizen science projects allow multiple contributions for same case while others actively discourage it. The former allows multiple labels to be available for each case which can aid some analyses but the latter would act to reduce duplication of effort and encourage a larger sample of cases to be labelled, albeit individually. The relative value of these approaches may differ between applications.

Finally, there is, of course, considerable scope for blending VGI with authoritative data sets although some users may wish to ensure that the data sources used can be identified so that attention may focus on cases from just one source independent of the other. For example, the design used by Olofsson et al. (2012) is adaptable and it would be possible to direct citizens to sites to collect data in order to meet specific research priorities (e.g. to increase the precision of estimates in a stratum of interest).

## 5. CONCLUSIONS

Good practices for authoritative accuracy assessment have been defined but may sometimes be impractical to implement. One key problem encountered commonly is the acquisition of a suitable ground reference data set on which to base the accuracy assessment.

Citizen sensing provides the potential to help address some of the problems encountered in the assessment of land cover map accuracy. It is not a panacea but does have the ability to provide reference data over a range of spatial and temporal scales. Although numerous concerns exist with citizen contributed data, and especially their quality, these are also research opportunities. Means to work effectively with citizen sensor data and to enhance future data acquisitions by defining good practices are emerging and it is anticipated that such data will increasingly be used to inform assessments of land cover map accuracy.

## ACKNOWLEDGEMENTS

I am grateful to the three referees for their helpful comments on the original version of this article and the conference organizers for the invitation to present this article at the conference.

## REFERENCES

- Antoniou, V., Morley, J. and Haklay, M., 2010. Web 2.0 geotagged photos: assessing the spatial dimension of the phenomenon. *Geomatica*, 64 (1), 99-110.
- Arsanjani, J. J., Mooney, P., Zipf, A., and Schauss, A., 2015a. Quality assessment of the contributed land use information from OpenStreetMap versus authoritative datasets. In Arsanjani, J. J., Zipf, A., Mooney, P., and Helbich, M. (editors), *OpenStreetMap in GIS*, Springer, Berlin, pp. 37-58.
- Arsanjani, J. J., Zipf, A., Mooney, P., and Helbich, M., 2015b. An introduction to OpenStreetMap in geographic information science: experiences, research and applications. In Arsanjani, J. J., Zipf, A., Mooney, P., and Helbich, M. (editors), *OpenStreetMap in GIS*, Springer, Berlin, pp. 1-15.
- Barnett, V., 1991, *Sample Survey Principles and Methods*, Arnold, London.
- Bastin, L., Buchanan, G., Beresford, A., Pekel, J. F. and Dubois, G., 2013. Open-source mapping and services for Web-based land-cover validation. *Ecological Informatics* 14, pp. 9-16.
- Carlotto, M. J., 2009. Effect of errors in ground truth on classification accuracy. *International Journal of Remote Sensing* 30 (18), pp. 4831-4849.
- Cihlar, J., 2000. Land cover mapping of large areas from satellites: status and research priorities, *International Journal of Remote Sensing*, 21, pp. 1093-1114.
- Cochran, W. G., 1977. *Sampling Techniques*, third edition, John Wiley, New York.
- Comber, A., See, L., Fritz, S., Van der Velde, M., Perger, C., and Foody, G., 2013. Using control data to determine the reliability of volunteered geographic information about land cover. *International Journal of Applied Earth Observation and Geoinformation*, 23, pp. 37-48.
- Congalton, R. G. and Green, K., 2009. *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices*. Second edition, CRC press.
- CSDGSND (Committee on Strategic Directions for the Geographical Sciences in the Next Decade), 2010. *Understanding the Changing Planet: Strategic Directions for the Geographical Sciences*, National Academies Press, USA.
- Estima, J. and Painho, M., 2013. Flickr geotagged and publicly available photos: preliminary study of its adequacy for helping quality control of Corine land cover In *ICCSA 2013* Springer Berlin Heidelberg, pp. 205-220.
- Estima, J. and Painho, M., 2015. Investigating the potential of OpenStreetMap for land use/land cover production: A case study for continental Portugal. In Arsanjani, J. J., Zipf, A., Mooney, P., and Helbich, M. (editors), *OpenStreetMap in GIS*, Springer, Berlin, pp. 273-293.
- Fleiss, J. L., Levin, B. and Paik, M. C., 2003. *Statistical Methods for Rates and Proportions*, 3rd edition, Wiley, New Jersey.

- Fonte, C. C., Bastin, L., See, L., Foody, G. and Lupia, F., 2015a. Usability of VGI for validation of land cover maps. *International Journal of Geographical Information Science*, (DOI:10.1080/13658816.2015.1018266, in press).
- Fonte, C. C., Bastin, L., See, L., Foody, G. and Estima, J., 2015b. Good practice guidelines for assessing VGI data quality. Proceedings of AGILE 2015, 9-12 June 2015, Lisbon, Portugal (in press)
- Foody, G. M., 1996. Approaches for the production and evaluation of fuzzy land cover classifications from remotely-sensed data. *International Journal of Remote Sensing*, 17(7), pp. 1317-1340.
- Foody, G. M., 2002. Status of land cover classification accuracy assessment. *Remote Sensing of Environment*, 80 (1), pp. 185-201.
- Foody, G. M., 2004. Supervised image classification by MLP and RBF neural networks with and without an exhaustively defined set of classes. *International Journal of Remote Sensing*, 25 (15), pp. 3091-3104.
- Foody, G. M., 2009. Sample size determination for image classification accuracy assessment and comparison. *International Journal of Remote Sensing*, 30 (20), pp. 5273-5291.
- Foody, G. M., 2010. Assessing the accuracy of land cover change with imperfect ground reference data. *Remote Sensing of Environment*, 114 (10), pp. 2271-2285.
- Foody, G. M. 2012. Latent class modelling for site and non-site specific classification accuracy assessment without ground data *IEEE Transactions on Geoscience and Remote Sensing*, 50 (7), pp. 2827-2838.
- Foody, G. M., 2013. Ground reference data error and the mis-estimation of the area of land cover change as a function of its abundance. *Remote Sensing Letters*, 4 (8), pp. 783-792.
- Foody, G. M., 2015. Valuing map validation: the need for rigorous land cover map accuracy assessment in economic valuations of ecosystem services. *Ecological Economics*, 111, pp. 23-28.
- Foody, G. M. and Boyd, D. S., 2013. Using volunteered data in land cover map validation: mapping West African forests. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 6 (3), pp. 1305-1312.
- Foody, G. M., and Curran, P. J., 1994. *Environmental Remote Sensing from Regional to Global Scales*. John Wiley and Sons Ltd, Chichester.
- Foody, G. M., See, L., Fritz, S., Van der Velde, M, Perger, C., Schill, C., and Boyd, D. S., 2013. Assessing the accuracy of volunteered geographic information arising from multiple contributors to an internet based collaborative project. *Transactions in GIS* 17 (6), pp. 847-860.
- Fritz, S. and See, L., 2008. Identifying and quantifying uncertainty and spatial disagreement in the comparison of Global Land Cover for different applications. *Global Change Biology*, 14 (5), pp. 1057-1075.
- Fritz, S., McCallum, I., Schill, C., Perger, C., See, L., Schepaschenko, D., van der Velde, M., Kraxner, F., and Obersteiner, M., 2012. Geo-Wiki: An online platform for improving global land cover. *Environmental Modelling and Software*, 31, pp. 110-123.
- Giri, C., Zhu, Z., and Reed, B., 2005. A comparative analysis of the Global Land Cover 2000 and MODIS land cover data sets. *Remote Sensing of Environment*, 94 (1), pp. 123-132.
- Goodchild, M. F., 2007. Citizens as sensors: The world of volunteered geography. *GeoJournal*, 69 (4), pp. 211-221.
- Goodchild, M. F., and Li, L., 2012. Assuring the quality of volunteered geographic information. *Spatial Statistics*, 1, pp. 110-120.
- Iwao, K., Nishida, K., Kinoshita T. and Yamagata, Y. 2006. Validating land cover maps with Degree Confluence Project information. *Geophysical Research Letters* 33, p. L23404.
- Jung, M., Henkel, K., Herold, M., and Churkina, G., 2006. Exploiting synergies of global land cover products for carbon cycle modeling. *Remote Sensing of Environment*, 101 (4), pp. 534-553.
- Magidson, J. and Vermunt, J. K. 2004. Latent class models. In Kaplan, D (editor) *The SAGE Handbook of Quantitative Methodology for the Social Sciences*, Sage, Thousand Oaks, pp. 175-198.
- Mather, P. and Koch, M., 2011. *Computer Processing of Remotely-Sensed Images: An Introduction*, 4<sup>th</sup> edition, John Wiley and Sons Ltd, Chichester.
- McCallum, I., Obersteiner, M., Nilsson, S. and Shvidenko, A., 2006. A spatial comparison of four satellite derived 1km global land cover datasets. *International Journal of Applied Earth Observation and Geoinformation*, 8, pp. 246-255.
- McRoberts, R. E., 2011. Satellite image-based maps: Scientific inference or pretty pictures? *Remote Sensing of Environment*, 115 (2), 715-724.
- Olofsson, P., Stehman, S. V., Woodcock, C. E., Sulla-Menashe, D., Sibley, A. M., Newell, J. D., Friedl, M. A. and Herold, M., 2012. A global land-cover validation data set, part I: fundamental design principles *International Journal of Remote Sensing*, 33 (18), pp. 5768-5788.
- Olofsson, P., Foody, G. M., Stehman, S. V., and Woodcock, C. E., 2013. Making better use of accuracy data in land change studies: Estimating accuracy and area and quantifying uncertainty using stratified estimation. *Remote Sensing of Environment*, 129, pp. 122-131.
- Olofsson, P., Foody, G. M., Herold, M., Stehman, S.V., Woodcock, C. E. and Wulder, M. A. 2014. Good practices for assessing accuracy and estimating area of land change. *Remote Sensing of Environment*, 148, pp. 42-57.
- Rocchini, D., Foody, G. M., Nagendra, H., Ricotta, C., Anand, M., He, K. S. Amicic, V., Kleinschmit, B., Forster, M., Schmidtlein, S., Feilhauer, H., Ghisla, A., Metz, M. and Neteler, M., 2013. Uncertainty in ecosystem mapping by remote sensing. *Computers and Geosciences*, 50, pp. 128-135.

Roman, D. 2009. Crowdsourcing and the question of expertise. *Communications of the ACM* 52: pp. 12.

Som, R. K., 1996, *Practical Sampling Techniques*, second edition, Marcel Dekker, NewYork.

Stehman, S. V., 1999. Basic probability sampling designs for thematic map accuracy assessment. *International Journal of Remote Sensing*, 20 (12), pp. 2423-2441.

Stehman, S. V., 2009. Sampling designs for accuracy assessment of land cover. *International Journal of Remote Sensing*, 30 (20), pp. 5243-5272.

Stehman, S. V., 2012. Impact of sample size allocation when using stratified random sampling to estimate accuracy and area of land-cover change. *Remote Sensing Letters*, 3 (2), pp. 111-120.

Stehman, S. V. and Czaplewski, R. L., 1998. Design and analysis for thematic map accuracy assessment: fundamental principles. *Remote Sensing of Environment*, 64 (3), pp. 331-344.

Stehman, S. V. and Foody, G. M., 2009. Accuracy assessment, In Warner, T. A., Nellis, M. D. and Foody, G. M. (editors) *The SAGE Handbook of Remote Sensing*, Sage, London, pp. 297-309.

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*. European Commission, EUR 22156 EN, 2006.

Vandecasteele, A. and Devillers, R., 2015. Improving volunteered geographic information quality using a tag recommender system: The case of OpenStreetMap, In Arsanjani, J. J., Zipf, A., Mooney, P., and Helbich, M. (editors), *OpenStreetMap in GIS*, Springer, Berlin, pp. 59-80.

Vermunt, J. K. and Magidson, J., 2003. Latent class analysis. In M. Lewis-Beck, A. E. Bryman and T. F. Liao (editors), *The Sage Encyclopaedia of Social Science Research Methods*, volume 2, Sage, Publications, Thousand Oaks, pp. 549-553.

Wickham, J.D., Stehman, S.V., Gass, L., Dewitz, J., Fry, J.A., Wade, T.G., 2013. Accuracy assessment of NLCD 2006 land cover and impervious surface. *Remote Sensing of Environment*, 130, pp. 294-304.

Wulder, M. A., White, J. C., Goward, S. N., Masek, J. G., Irons, J. R., Herold, M., Cohen, W. B., Loveland, T. R., and Woodcock, C. E., 2008. Landsat continuity: Issues and opportunities for land cover monitoring, *Remote Sensing of Environment*, 112 (3), pp. 955-969.

Yang, I. and Becker, M. P. 1997. Latent variable modelling of diagnostic accuracy. *Biometrics*, 53, pp. 948-958.