Usage of VGI for Validation of Land Cover Maps

[1,2]Cidália C. Fonte, [3]Lucy Bastin, [4]Linda See, [5]Giles Foody, [6]Favio Lupia

[1] *Department of Mathematics, University of Coimbra, Coimbra, Portugal*

[2] *Institute for Systems Engineering and Computers at Coimbra (INESC Coimbra), Coimbra, Portugal*

[3] *School of Engineering and Applied Science, Aston University, Birmingham, B4 7ET, UK*

[4] *Ecosystems Services and Management Program, International Institute for Applied Systems Analysis (IIASA), Laxenburg, Austria*

[5] *School of Geography , University of Nottingham , Nottingham ,NG7 2RD, UK*

[6] *National Institute of Agricultural Economics (INEA), Rome, Italy*

Corresponding author: Cidália C. Fonte (e-mail: cfonte@mat.uc.pt)

Usage of VGI for Validation of Land Cover Maps

Volunteered Geographic Information (VGI) represents a growing source of potentially valuable data for land cover validation. However, the quality and credibility of the data remains a key concern, especially if VGI is to be adopted by more traditional map producers or integrated into authoritative sources in the future. This paper reviews different categories of spatial data quality and the main sources of VGI currently being used in the development and validation of land cover maps. The paper also proposes a framework for addressing different VGI quality assessment methodologies, which is used to identify gaps in approaches that could be used in VGI quality assessment in the future.

Keywords: VGI, land cover/land use, quality, validation, crowdsourcing

1. Introduction

Land cover maps are fundamental for a wide range of users and for many applications, such as planning, nature and biodiversity protection, environmental monitoring, management of natural resources, climate change and hydrological modelling (Feddema *et al.* 2005, Foley *et al.* 2005, Hassan *et al.* 2005, Nie *et al.* 2011, Verburg *et al.* 2011). They are often produced through the classification of remote sensing images, using automatic or semi-automatic approaches. However, due to the variability of maps generated with different methodologies (Lu and Weng 2007), a key issue for users is the accuracy of the produced maps, which determines their fitness-for-use for particular applications. Accuracy assessment is now widely regarded as an essential part of any land cover mapping programme, without which the map is simply an untested hypothesis, of potentially little, if any, value (Strahler *et al.* 2006).

The assessment of the accuracy of a land cover map is made by evaluating the degree to which the map agrees with a reference database of “ground truth”, which is meant to indicate the actual land cover observed for a sample of spatial units (e.g. pixels). The spatial units can be points, pixels or blocks of pixels and are identified using a suitable sampling strategy that generates an unbiased and representative subset of the population, so that the accuracy assessment of the sample may be used to estimate the population's accuracy (Steele 2005, Stehman 2009). The association of ground truth to each sample unit is usually done by experts using high resolution images, field visits or local knowledge of the area. The reference data are then compared to the land cover map, generally by building confusion or contingency matrices, from which one or more statistical accuracy indices can be extracted, which express different aspects of accuracy (e.g. Congalton and Green 1998; Foody 2002; Steele 2005; Pontius and Millones 2011).

A major challenge in this accuracy assessment approach is the creation of the reference data. The process of determining "ground truth" for all sample units may not be an easy task since it can be difficult to label the land cover of a site where even expert annotators can disagree with one another. In fact, "ground truth" never really exists, since error can be introduced at many points in the generation of a reference dataset, and even a small amount of error can propagate through the validation process to yield large errors in the accuracy assessment (Woodcock and Gopal 2000, Foody 2011, 2013). Another important requirement for reference data is that it be representative, and this is best achieved by using a properly-designed probabilistic sample design (Strahler *et al.* 2006, Stehman *et al.* 2012). Once a proper stratification of land cover classes has been performed, a large number of sample points is usually required in order to ensure statistical validity, and this is especially true when considering rare classes (Olofsson *et al.* 2012) or where a study aims to detect change between two points in time - effectively adding another dimension to the analysis. The construction of a high-quality authoritative reference dataset is therefore a time-consuming and expensive process. This is even more problematic when land cover maps covering large regions are to be assessed, when no high resolution images are available, or when the map producers are far from the area to be analysed. In these cases field visits are laborious, problematic or even impossible and thus no local knowledge of the area is available.

The difficulty of building reference databases, mainly for the validation of maps at a global scale, and the rising availability of data provided by volunteers worldwide, which can provide information on the type of land cover at different locations, has drawn the attention of scientists to the potential value of Volunteered Geographic Information (VGI) as a source of cheap, current and plentiful reference data.

The term VGI is used here to cover a wide range of data, to which a geographical location is associated, made available on the web by volunteers. Other terms, such as contributed geographic information, crowdsourcing, user-generated content or neogeography are also used to refer to this new type of data, even though they are not synonyms (Elwood *et al.* 2012, 2013, Harvey 2013, Poore and Wolf 2013, See, Mooney, *et al.* 2014). This information may be collected in many formats, ranging from text descriptions or photographs to complete maps created by the volunteers. A review on several of these sources of data may be found in Goodchild (2007), Haklay (2013) and Heipke (2010). On the face of it, VGI has huge potential to replace or complement authoritative data which are expensive or restricted, or to fill gaps in the available reference data, especially for global land cover monitoring. However, several questions are raised by this use; chief among these is how to guarantee the quality of VGI, given its patchy geographical distribution, its potential for contributor error or malicious misinformation, and its lack of homogeneity and representativity.

This paper aims to review the use of VGI as land cover reference data, evaluate its potential and identify the problems and challenges raised by the use of VGI as ground reference data. The review begins with a discussion on VGI quality, including methodologies currently or potentially used to ensure and assess quality, either quantitatively or qualitatively. We then identify and describe VGI platforms and projects whose outputs have already been used for land cover map validation, and others that may in the future be used for this aim, with particular attention to their quality assurance procedures. This is followed by a documentation of case studies where VGI has actually been used to validate land cover maps (including any quality assessments of the VGI which may have been carried out by the researchers in the course of their work), and some independent assessments of the resulting data quality. Finally, we discuss best practice, future potential and the challenges facing this field of analysis.

2. Quality of VGI

The assessment of spatial data quality is a very broad and diverse theme, since quality of a dataset has many aspects, such as internal consistency, completeness, precision, or closeness to reality (‘accuracy’) (Guptill and Morrison 1995, Devillers and Jeansoulin 2010). In combination, these aspects of quality define a dataset’s fitness-for-use for a certain purpose. When related to VGI, additional indicators of quality may also be considered, such as the credibility of the volunteers and the reliability of the information they supply. Flanagin and Metzger (2008) stress the differences between information credibility and accuracy. Credibility is indicated to have two dimensions: trustworthiness and expertise, and includes some subjective components which are complex to assess. The traditional meaning of accuracy is the degree of closeness to reality but it does not consider other factors such as the credibility of a contributor.

The issue of VGI data quality has been raised by many commentators and is one of the most important topics on the VGI research agenda (Elwood *et al.* 2013), since it ultimately determines the relevance and appropriateness of the data for use in real-world contexts. The key point is that for VGI to be useful in scientific analyses there is a need for some measure of its credibility and accuracy. Without such measures there will always be a lack of trust in these data. In particular, for the use of VGI as reference data to inform and validate land cover maps, a certain level of accuracy is fundamental to obtain credible results, since in this case VGI is supposed to represent ground truth.

As many aspects regarding data quality may be considered and many types of approaches have been proposed for its assessment, the most commonly used ones are reviewed here. Later attention focuses specifically on the individual quality aspects related to the use of VGI as ground reference data.

2.1 Approaches to address VGI quality

Several approaches and perspectives may be taken to respond to the challenge of VGI quality control. One perspective is to guide the volunteers and establish procedures to produce VGI with higher quality compared to the situation where that guidance was not present; another is to assess the quality of the obtained information, enabling users to focus attention upon only the most accurate data (Foody *et al.* 2014). These are not mutually exclusive, but complementary and interconnected.

In the first, *result-oriented* perspective, many approaches may be considered, such as determining which procedures are more likely to produce information with higher levels of quality. Some, more 'facilitated' applications use numerous rules to guide the contributors in what they are expected to do; others give the contributor freedom to supply information of different types without strict rules as to its content or format (Elwood *et al.* 2013). Also procedures that enable the identification and correction of erroneous data can be included in these.

The second, *evaluation-oriented* perspective, aims to identify methods that enable the assessment of the data quality, either qualitatively or quantitatively. These may be based, for example, on metadata, comparison with data of a known higher quality, or assessment of the contributor's accuracy of labelling.

Goodchild and Li (2012) categorize the different approaches to address the quality of VGI into three groups:

(1) 'Crowd-sourced', relying on consensus and agreement. As the number of contributors increases it is more likely that the results have higher quality, since errors are more easily identified and corrected (e.g. Haklay et al. 2010). Even though this principle may apply to populated regions, it does not apply correctly to more isolated regions where the number of possible contributors is small;

(2) 'Social', using contributors with a reliable history of quality contributions to review the work of other contributors;

(3) 'Geographical', identifying rules that enable the use of connections between the several types of information for each location, to assess the possibility that a certain attribute is correct at a certain location. In its simplest form, this approach is the most familiar to geoscience and land cover specialists, since it equates to traditional 'ground-truthing' against more credible data based on geographic context.

Goodchild and Li (2012) stress that these approaches aim at enhancing quality during the data acquisition and compilation phase. Therefore these approaches are result-oriented, but may depend on the development of evaluation strategies to assess data quality; for example, developing rules based on geographical context to assign a degree of credibility to the occurrence of a certain phenomenon at a given location.

Allahbakhsh et al. (2013) also provide a categorization for quality control measures of crowdsourced systems more generally but which has relevance to assessing the quality of VGI. They refer to these approaches as 'run-time approaches' and include among them: expert review, whereby the quality is checked by domain experts; output agreement and majority consensus, where multiple independent observations that agree on the same value are deemed to be correct or correctness is based on majority agreement; ground truth or comparison with a gold standard such as known answers; contributor evaluation, which assesses a current contribution based on past performance; real-time support, or processes for guiding contributors in real-time, and workflow management, whereby complicated tasks are broken down into workflows and monitored over time, which are then modified as necessary to improve the quality. The authors also outline what they refer to as 'design-time approaches', corresponding to our *result-oriented* category. These are divided into effective task preparation, in which the task is clearly defined along with evaluation criteria and compensation (more relevant to crowdsourcing approaches that are business-oriented); and the selection of workers, which is further divided into no selection (open to all); selection based on reputation; and selection based on credentials. The last two approaches are not often used in VGI.

We propose another perspective to classify the various approaches to VGI quality assessment, which are based on the level of automation and the type of data used for the quality assessment. Four levels of automation of quality assessment may be considered, illustrated in Table 1, which go from *fully automated* to *done manually by experts*. These levels of automation refer to the level of human intervention required once the process is implemented, but recognise that the more a process is executed within the crowd the more automated it is, since ultimately it will run by itself based on the intervention of the volunteers, requiring low or even no intervention by the project managers.

Four types of data required to assess the quality of a particular item may also be considered, illustrated in Table 2. At one end of the scale are methods using only the VGI itself (including any available metadata), which usually rely on consensus or on recognition of patterns in the behaviour of contributor to assess the trustworthiness of a volunteer or their contribution. At the other end of the scale are methods which require a full authoritative dataset, against which 'truth' the VGI can be assessed. The latter context raises interesting questions about whether VGI is sometimes, in fact, more 'true' than the authoritative data, particularly when up-to-date representations of the world are being sought.

The methods used to assess quality related to VGI may be classified by these two dimensions, i.e. the degree of automation involved and the data used, resulting in 16 types that appear as elements in the matrix shown in Table 3. These provide a categorization of existing approaches to VGI quality assessment, enabling a more structured analysis of the types of methodologies that have been developed and those that may be developed, and will be used to classify the methodologies presented in this paper. The categories 'crowdsourced' and 'social' proposed by Goodchild and Li (2012) correspond respectively to categories B and C in the framework proposed here, while the 'geographical' approach may correspond to categories A, 1 to 4, depending on the type of data used. From the categorization proposed by Allahbakhsh et al. (2013), the *expert view* corresponds to approaches of type D, *output agreement and majority consensus* to type B, and *ground truth* may correspond to approaches of type D3 or D4, or even to C3 or C4, if the selected volunteers are expected to give very good results. *Contributor evaluation* may be approaches of type A, C or D and most probably 1, 3 or 4, depending on how the contributor quality is assessed. *Real-time support* and *workflow management* are not directly considered in the proposed categorization, since they are result-oriented approaches which aim at generating better results through direct assistance to the volunteer. *Real-time support* involves expert intervention and training in real-time, which might be considered a variation of category D, while *workflow management* involves the development of workflows to facilitate complex tasks. These workflows could then be monitored with respect to quality and modified over time.

2.2 Assessment of VGI quality

Aspects relevant to the use of VGI for land cover map quality assessment are the contributor credibility as well as the classically-recognised aspects of spatial data quality, such as positional accuracy, thematic accuracy, completeness, currency, logical consistency and lineage. As indicated in the previous section, the quality assessment may require additional data, such as metadata, crowd knowledge or comparison with other sources of GI, volunteered or not. Some of these aspects are addressed in the following sections.

2.2.1 Credibility

When evaluating the credibility of a VGI dataset or contributor, it is possible to exploit metadata if this exists. In the absence of metadata, it is often necessary to seek patterns in the data or in the history of that contributor. Some of the facets of traditional metadata are of particular interest in assessing and using VGI. For example, the lineage of a record or dataset may include its edit history, information on who submitted it or on how it was measured, and can be especially important in the automated assessment of

VGI fitness-for-use. Examples of metadata potentially useful for VGI are equipment used in measurements; data on the volunteer, such as age, address, level of education or interests; date and time of data collection or atmospheric conditions at the time a particular observation was taken. Individual metadata about heterogeneous observations can be extremely useful in identifying bias and likely trustworthiness, as seen for example in the context of amateur weather monitoring (Bell *et al.* 2013, 2014) and digitised trails (Esmaili *et al.* 2013). There is also potential to automatically infer some information about the likely precision of a measurement, for example, by considering the nature of the device that captured a reading, the characteristics of the equipment used, such as a GNSS receiver, or the number of satellites used for the positioning and measurement duration. The likelihood of erroneous geotagging of a photograph may also be assessed by considering whether it was georeferenced using a GNSS receiver integrated in the camera, an external receiver, or if the photograph was geotagged through its positioning over satellite imagery or a digital map, and if so, whether high, medium or low spatial resolution images were used or the map scale.

It is also possible to construct metadata based on past behaviour of a user, or the number of times their contributions have been identified as erroneous by other volunteers, which requires the storing of all alterations and changes made to the system. This information may be used to associate a degree of credibility to the data, using approaches of types A1, B1, C1 or D1. From these the methods of type A1 are probably the most promising, since they enable, through the definition of a set of rules, the automatic extraction of quality information, which may be used as an initial indicator of credibility, enabling the exclusion of some VGI from an analysis based on the likelihood that it might be less trustworthy. An example of these procedures is the approach proposed by Lenders et al. (2008) where the contributor’s trustworthiness is assessed using the information about the volunteer's location and the time of the contribution. However, such simple 'rules-of-thumb' are limited in their distinguishing powers. For example, across a wide range of VGI applications, it is common that most of the data is provided by a few contributors, with a large proportion of contributors contributing only once. It might be assumed that the contributions of the prolific contributors have higher quality (Elwood *et al.* 2013), but the reality is often more complex. Even experienced users have their 'blind spots' and may lose their advantage when they move to unfamiliar regions or themes. Even though metadata may be very useful to infer several types of information related to data quality, only a minority of VGI applications require contributors to actively supply metadata about themselves or the data they record.

Credibility may also be assessed using volunteer or expert intervention (corresponding to approaches of types B, C or D). Bishr and Mantelas (2008) propose a 'trust and reputation model', where these two concepts together are proxies for data quality. Users rate each other's contributions on a score of 1 to 10, which makes up the reputation component. Users are also linked to one another through a social network, which can be used to measure the strength between two individuals. These two components are then combined to calculate a trust rating based on the ratings given to contributors and the strength of the relationship between them, which is divided by the log of the distance between a contributor's location and the observation. This trust model therefore takes both spatial context and reputation into account through user ratings and the relationships between contributors. The model remains theoretical, and was not applied in the paper cited above but an example of data collection for an urban growth scenario was outlined. This represents a type-B approach, although the inclusion of relationships via social networking could give greater weight to the ratings of certain individuals and may therefore be of category C. The data for the urban growth example would be crowdsourced, placing the method into category 2 in Table 2.

Some approaches have also been tested to assess volunteer credibility relating to the thematic information they provide. Since this aspect is related to thematic information, it is addressed in section 2.2.3, dedicated to thematic quality.

2.2.2 Positional accuracy

The positional accuracy of spatial data is usually associated with data georeferenced as points, lines or areas, such as road junctions or buildings. At present, most quality assessments for positional accuracy in VGI appear to use traditional 'geographic' approaches such as feature matching or control point checking against authoritative data (corresponding to approaches of type 4, either B, C or D). While there is potential for features contributed by multiple users to be simply 'geometrically averaged', this does not appear to be applied in practice. Positional accuracy may also be controlled either by the crowd, correcting the location of the data, or inserting comments within a particular VGI collection platform, if that option is available, when a wrong position is detected, usually using approaches of type B2 or B3, or by selected volunteers, corresponding to approaches C2 or C3.

Portable data collection technologies widely available to individual citizens are now capable of delivering a spatial precision exceeding ±10m (Coleman 2010). When combined with the increasing availability of Web-based maps and imagery (in some cases with very high spatial resolution) which can be used as digitising backdrops, it is not surprising that the positional accuracy of VGI increases, and is now appropriate for a wide range of applications. However, discrepancies and outright errors still arise, with diverse levels of magnitude and seriousness. For example, an analysis of positional accuracy of OSM in relation to Google Maps and Bing Maps (approach of type D3) was made by Ciepłuch et al. (2010) for sites in Ireland, which concluded that in some locations there were differences of up to 10m (in Google Maps) between these sources, although only for some types of features, which seemed to result from digitization over low resolution images. For a set of OpenStreetMap road features compared against the UK's Ordnance Survey data (approach of type D4), the average errors identified were 5.8m (Haklay 2010) - a distance unlikely to be seriously problematic for most land cover maps, but one which could cause small or narrow features (ponds, hedges, riparian habitats, etc.) to be missed or misplaced.

To correct and quantify these positional errors, conflation approaches which use a set of reference features are common (Coleman 2010, Girres and Touya 2010, Haklay 2010) for discrete data that fits an existing taxonomy. Canavosio-Zuzelski et al. (2013) also performed a positional accuracy assessment of OSM as part of a vector adjustment correction. However, in this case, rather than accepting official survey data as truth, both official data and OSM were assessed against independent stereo imagery, which means the technique can be applied to other national agency and topographic datasets and has the potential to identify areas where the VGI excels over the accepted dataset (approach of type D3). Thus the authors were able to assess OSM against USGS (United States Geological Survey) and TIGER (Topologically Integrated Geographic Encoding and Referencing) road data on a more-or-less equal footing - albeit for a very small area for which the aerial imagery was available. In general, the availability of such accurate benchmarking data is restricted, and this (or a requirement for very current information) may be the very reason why VGI is being elicited. The most successful examples of such quality control analyses are where feedback is given to the volunteers to enable them to improve their contributions - e.g., the OpenStreetMap Collaborative Project.

Positional accuracy of points representing geotagged photographs may also be considered and analysed. In Hochmair and Zielstra (2012) the location associated with the Flickr and Panoramio photographs was compared to the location of the photograph determined by the authors analysing what was represented in the photograph (approach of type D1 and D3). Several types of errors were identified, e.g., when the position assigned to some photographs was not the location from which the photograph was taken, but rather the position of what was represented in the photograph (potentially some distance away!). Another type of error was confusion between similar features that are present in the region (such as different bridges over a river close to each other), which became apparent when the location of the photographs was viewed on a satellite image or digital map.

The assessment of the positional accuracy or the extent mapping of patchy vegetation, highly-textured land use types and ecotones presents much more of a challenge. For land cover mapping, it is often the case that categorical labels (or degrees of similarity to those labels) are being elicited from contributors for attachment to user-supplied location points or to predefined polygon features. Absolute positional accuracy is still important, but more often relates to boundaries between mapped areas, or the location of single survey points, and the predominant source of inaccuracy is thematic misclassification (to which, of course, these positional inaccuracies can contribute).

2.2.3 Thematic quality

In mapping, thematic accuracy assesses the accuracy of classes or thematic tags associated with specific locations or objects placed in geographical space, such as classes assigned to pixels in a land cover map, a tag assigned to a linear entity or a polygon, as for example a highway, river, building or green area.

This assessment may be performed using a traditional approach, where the information is compared by experts with reference data (satellite imagery or authoritative data), using approaches of type D3 and D4. A comparison to information with comparable semantics available from different VGI initiatives for the same location can also be done, although in this case it is more an assessment of consistency, since none of the data is considered to represent the truth. As stressed by Sui et al. (2013) these geographic approaches have not yet been developed enough. They correspond to approaches of type D2 (and possibly in the future, when automated, of type A2) and may be used to control data quality but not to assess its accuracy. As for the positional accuracy, the crowd or selected volunteers may also control the accuracy of this type of information, correcting erroneous contributions or tagging them, usually corresponding to methods of type B2, B3, C2 and C3.

Methods for the automatic computation of contributor reliability regarding thematic information in VGI have been proposed by several authors. Haklay et al. (2010) and Tang and Lease (2011) stress the need for multiple observations and observers to enable consensus-based data quality assessments. Foody and Boyd (2012) and Foody et al. (2013) proposed a method for using these repeated observations to concretely assess the quality of VGI contributors using a latent class analysis of VGI in relation to land cover.

When considering thematic quality, the issue of contributor reliability can be more complicated than a single ranking. Some contributors excel at labelling particular types of objects or habitats, but perform poorly elsewhere in the problem domain.

Knowledge of the strengths and weaknesses of the volunteers allows a more nuanced consideration of the trustworthiness of their contributions, but often requires independent reference data to compute. To this end, Comber et al. (2013) calculated the consistency and skill of each volunteer in relation to each land cover class, using a number of control points for which the land cover had been independently determined by experts, and demonstrated that at least some concerns about the quality of VGI can be addressed through careful data collection, the use of control points to evaluate volunteer performance and spatially explicit analyses.

The assignment of thematic information in VGI has many similarities to the extensive tagging and relevance assessment of documents by volunteers or paid contractors working via systems such as Amazon’s Mechanical Turk, and we have therefore paid attention to the methods used in those fields to validate the assigned labels. Many land cover mapping challenges are effectively labelling problems, where predefined pixels or spatial features must be assigned to particular classes (Fritz *et al.* 2012, Lindquist *et al.* 2012).

Currently, the majority of VGI is contributed for free, by volunteers, but there is an increasing interest in contracting out classification tasks such as land cover labelling to paid workers in the cloud. In such contexts, spam and errors are common, whether these stem from a lack of skill or from deliberate attempts to mislead (including attempts to cheat the system in a way that cannot be easily detected). A number of strategies have been proposed and evaluated for getting the best value out of contracted labellers, and in particular for trading off the value of new information about unlabelled entities against the value of reinforcing or correcting information about entities which have been labelled repeatedly (Ipeirotis *et al.* 2014). One consideration when deciding between accuracy improvement and new data acquisition must be the possible impact of errors when a dataset is used in the real world – a similar balancing act to the calculation of ROC (Receiver Operating Characteristic) curves or sensitivity/specificity calculations for classifiers and prediction algorithms. The problem of risk and liability, where it is considered in the VGI world, is usually sidestepped by the use of disclaimers but if VGI data begins to seriously underpin Spatial Data Infrastructures (SDIs) and commercial products, the issue will become more pressing.

Many of the non-VGI labelling tasks described have marked parallels to VGI problems: for example, data points are often being collected, like 'ground truth' in order to carry out a supervised classification, and in many cases the labelling is not simply binary or categorical. In such cases, the variation between labellers is not simply noise; often, the uncertainty and disagreement, if recorded and analysed, can yield important information about the real world. In the case of VGI this could include conditions on the ground such as vegetation succession, change of ownership or mixing of land covers. Many papers in the field also note the importance of training for labellers as well as for models (e.g. Clark and Aide 2011, Fritz *et al.* 2012), and show the sorts of learning curves which are possible with varying quantities and qualities of reference data.

Of course, even well-trained users vary in their accuracy, and differences between experts and non-experts are also likely to exist. A comparison of the quality results of expert and non-expert volunteers for tag assignment was done by See et al. (2013). The results showed that in some types of tags (in this particular case "human impact") non-expert volunteers produced results as good as the experts, probably because the concept was new to both non-experts and experts alike so both had the same learning curves. However, for some land cover classes the experts (some of whom had considerable experience in image classification) performed better but the non-experts demonstrated improvements over time, especially when feedback on the quality of their results was given to them. Differences between volunteers are always likely to exist, and therefore in the examples of ‘social’ quality assessment described above (corresponding to types B2), known individuals could be identified and given a more trusted status, and these individuals could then be actively responsible for reviewing the work of others (approaches of type C2). However, in the context of labelling for commercial gain, the workers do not see the submissions of others, and it is necessary to automate the process of identifying trustworthy experts against whom the work of others can be benchmarked (Raykar and Yu 2012). Vuurens and de Vries (2012) tackle this issue by deriving patterns from the behaviour of different worker types, and attempt to diagnose the nature, and thus the likely error rate, of particular workers (approaches of type D1). For example, they note that ‘diligent’ workers are less likely to differ in their votes by more than one step on an ordinal scale, and exploit this fact to interpret the difference between contributors’ judgements to identify their trustworthiness. Once automated, this approach equates to the ‘crowdsourced’ approach - A1 in Table 3 - though by identifying more trustworthy individuals it mimics the ‘social’ approach. Such a strategy may potentially be adapted for land cover contexts where there is a logical continuity to the categorical classes being labelled, or where continuous judgements (e.g., ‘percentage of vegetation cover’) are being solicited. However, there are many contexts where no natural ordering is present in the labels from which a contributor is to choose.

2.2.4 Completeness

Completeness is a hot topic in VGI, since many volunteered datasets are demonstrably biased towards particular spatial regions (e.g. Haklay 2010), but also towards certain features which are easier to measure or towards themes or 'pet features' (Bégin *et al.* 2013) which are of particular interest to the contributing individual, or even motivated by accessibility or digital inclusion (Zielstra and Zipf 2010). This reliance on the motivation of individual volunteers will determine the resolution, homogeneity, representativity and domain consistency of the resulting data. Where a principled sampling strategy can be imposed on volunteers (for example, a probabilistic schema, or the systematic, even grid of the Degrees of Confluence Project - see section 3) the volunteered data have the potential to be more broadly applicable - but its value will depend on the coverage of the design by volunteers, meaning that many platforms must actively direct users to the desired locations, trading off potentially rich information elsewhere against an even placement of observations.

In many areas, the number of digitised features may exceed that in an authoritative dataset (Neis *et al.* 2011), making a simple comparison of feature counts inappropriate, and requiring a subtler consideration of commission and omission (Jackson *et al.* 2013). Koukoletsos et al. (2012) present a method which does have promise for such contexts, combining geometric and attribute constraints to match road segments in OSM with an authoritative dataset, and achieve a tile-by-tile completeness assessment. Haklay (2010) identifies a bias in UK OpenStreetMap data coverage towards more affluent areas, and relates this to the fact that socially marginal (and less-mapped) areas may be the very locations where charities and agencies requiring free data are operating. Cipełuch et al. (2010) also compared the spatial coverage of OSM to Google Maps and Bing Maps (approach of type D3), and also identified regions with different levels of coverage in the three data sets. Globally, this bias is being somewhat redressed by the volunteers' own efforts to improve coverage, and by focussed initiatives such as KompetisiOSM in Indonesia (http://openstreetmap.or.id/) but it remains the case that coverage is extremely heterogeneous, both spatially and thematically, and that the absence of information in an area is often a particularly shaky basis for drawing any conclusion about the state or rate of change of land cover / land use. However, careful post-processing of VGI can increase its value for a robust analysis, provided that the model calibration is informed by a consideration of the way in which the data were collected, and its likely biases. Brunsdon and Comber (2012) specifically addressed the lack of experimental design in a volunteered dataset recording the first flowering date of lilacs in the US, by applying random coefficient modelling and bootstrapping approaches to tease out more reliable information on phenological trends.

2.2.5 Currency

Currency is one aspect of traditional data quality where VGI can be expected to excel over authoritative data, especially in dynamically changing environments, given the large numbers of citizens who are acting as sensors at any one time. However, this is often a tradeoff against other facets of data quality. The issue of representativity becomes even more vexed when the spatial domain is extended to the spatio-temporal domain, and unless a temporal sampling scheme is also imposed upon contributors, the density and coverage of a VGI dataset over a small time range can be very limited. For citizen sensor networks which are largely made up of automated instruments, such as the Weather Underground, the observation pattern across time is fairly consistent. However, in other contexts (for example, presence-only species observations and the mapping of urban infrastructure), a user will need to carefully consider the ranges of data which are appropriate for their purpose, and whether cumulative observations are valuable. In making this decision, they will probably require metadata on the individual features - for example, date stamps. An important consideration here is that the date stamp should reflect the time at which the measurement or observation was made, rather than the time at which it was uploaded or digitised, as occurs in photographs uploaded to Panoramio, where only the time of upload is recorded (Fritz *et al.* 2009).

2.2.6 Logical consistency

The logical consistency of an observation makes little sense in isolation: it must usually be assessed with reference to other data from the same source, or from independent (and sometimes authoritative) data, and lends itself to automated quality assessment - for example, the use of rules such as ‘forest fires are highly unlikely in dense urban areas’. Bonter and Cooper (2012) discuss the use of a *smart filter system* in the context of species identification in Project FeederWatch. When participants enter counts of species that are too high or species that do not normally appear on standard lists, then the filter is activated and users are informed of unusual observations, thereby correcting potential errors in real-time. Similar smart filters could be devised and put into place in VGI projects thereby addressing some aspects of logical consistency.

3. VGI as Reference Data

3.1 Types of VGI used for land cover map creation and validation

Several sources of VGI with different characteristics have and may be used to assist in the creation of land cover maps and assess their quality. The main sources of data that have been used for this purpose include:

(1) photographs and descriptions collected by the Degrees of Confluence Project (*eg.* Iwao et al., 2006; Foody and Boyd, 2013; Iwao et al., 2011);

(2) photographs posted by volunteers at sites, namely Panoramio, Flickr and Geograph (Wang *et al.* 2012, Estima and Painho 2013a);

(3) volunteer initiatives to map the world, such as OpenStreetMap (Estima and Painho 2013b, Jokar Arsanjani *et al.* 2013);

(4) land cover data collected by projects such as Geo-Wiki (e.g. Fritz et al., 2013; Comber et al., 2013) and VIEW-IT (Clark and Aide, 2011; Aide et al., 2013; Redo et al., 2012).

The first three correspond to data gathered for other purposes that may be useful for the aim of land cover map creation and validation, while Geo-Wiki and VIEW-IT correspond to projects created with the aim of collecting data on land cover for training and land cover map validation. For each of the projects indicated above, we present a description of the platform, the quality control procedures and the sampling strategies available, if any.

Several authors and projects use one or more of these sources of VGI, sometimes for validation but also in combination with additional non-VGI data and using other approaches to assess accuracy, including interpretation of satellite images and field visits (Iwao *et al.* 2006, Lindquist *et al.* 2012).

Other projects are available that may provide useful information for land cover mapping, such as Wikimapia (http://wikimapia.org), which largely collects ‘points of interest’ for human activity, and Wikiloc (http://www.wikiloc.com), which collects digitised trails relating to outdoor activity with associated photographs. Both have potential (albeit limited) to yield contextual information about land cover. However, they are not described in this paper, since we found no instances where the data had been used to derive or validate land cover maps.

An untapped source of information is descriptions of habitats from species identification sites such as iSpot and iNaturalist. When users identify species, they can also indicate the type of habitat, which if mapped into land cover classes, could be a valuable source of information for land cover map creation or validation.

3.2 Sources of VGI used as reference data

3.2.1 Degrees of Confluence Project

The Degrees of Confluence Project (DCP) (http://confluence.org/) was created in 1996. The aim of the project is for participants to collectively visit every latitude / longitude intersection point and then collect photographs oriented in the four cardinal directions (north, south, east and west), as well as descriptions of the landscape, to create an organized sample of the world. Many photographs and descriptions can be submitted for each confluence, which results in a multi-temporal collection of information. By April 2014 the website statistics report 6,278 confluence points successfully visited, corresponding to 38% of the 16,345 total confluence points, and 105,682 photographs collected across 186 countries.

The DCP allows the collection of several photographs per point, as well as the inclusion of textual descriptions of the site. In order for the photographs and textual descriptions to become usable ground data for land cover mapping, they must be labelled, by volunteers or experts, as belonging to certain land cover classes (Iwao *et al.* 2006, Foody and Boyd 2013).

The submission policy to the DCP involves a preliminary check for errors by regional coordinators, which identify obvious mistakes and malicious submissions. This verification of obvious errors is done through the comparison of what is shown in the photographs with maps, or the assessment of correspondence between the description and what is shown in the photographs. Actual communication with the volunteers may be done to clarify any doubts.

These quality control approaches are mainly result oriented, social and are of types C1, C2, C3 and C4, depending on the information used. For example, the coordinators may use the VGI data itself through check of self-consistencies (C1), VGI may be used from other VGI initiatives such as OSM (C2), and satellite images may be used (C3) or even authoritative maps (C4).

Additional aspects related to the data quality may be evaluated by users, such as assessing the positional accuracy of the photographs by comparing what the photographs show and inferring a likely location from which they were taken (type D3) (Hochmair and Zielstra 2012), or assessing the thematic accuracy by a comparison of descriptions (type D1, or eventually A1, if automated). The latter process can give some insight into the difficulties involved in assigning a land cover class to that location, and this in turn may give some indication of thematic accuracy.

3.2.2 Geograph

Geograph (http://www.geograph.org/) is an initiative that encourages people to collect and submit photographs that are representative of every square kilometre of Great Britain, where the project was first started by Gary Rogers. Geograph has now been extended to Ireland, Germany and the Channel Islands. To date, there are 12,050 contributors who have submitted just under 4 million photographs that cover 81.7% of the total area of Great Britain and Ireland. Geograph Germany is a much newer initiative with only 150 users so far who have covered just over 4% of the country while

20 users have already covered more than 50% of the Channel Islands. There are plans to extend the project to Corsica.

Anyone accessing the site can view information about the photographs, including tags assigned to each one and start discussions on individual photographs (although this latter feature requires logging into the system). Users can also view the location of the photograph on Google Earth, Google Maps, the corresponding OS map sheet and the Geograph map interface as well as viewing additional links related to this site.

Once logged in, users can upload geo-tagged photographs, manually enter the location or do a bulk upload of the photographs, either using their bespoke facilities or from Picasa. There does not appear to be any automated checking by the system itself but there is a team of moderators who review the photographs that are uploaded (i.e. approach C or D, probably using all sources of information available to them from the automatically generated information in the photograph to authoritative data, e.g. OS maps, covering 1 to 4). Moreover, users can disagree with the location or title of a photograph and make suggestions for changes, i.e. approach B, using local knowledge or other data, i.e. 2 to 4.

3.2.3 Panoramio

Panoramio (http://www.panoramio.com/) is a web site created in 2005 to collect photographs of the world, to which a geo-location and a date of upload is associated. The main aim of the website is to document the world with photographs, so most of the photographs illustrate places. The volunteers may assign small descriptions to the uploaded photographs, as well as tags, which can be used to group them into categories. A selection of photographs can be accessed as a layer in Google Earth and Google Maps by their location. According to the website Panorank (http://www.panorank.com) by May 2014 the Panoramio users were around 8 million with a total number of uploaded photographs of approximately 83 million.

The inclusion of photographs in Panoramio requires a prior review, which allows a control of the type of photographs that may be included on the site (C1 or D1). Panoramio also allows the correction of the position of the photographs by the volunteers (type B). However, no change information or versioning control is made available.

Panoramio allows the insertion of comments on the images, which may allow the inclusion of comments on problems related to the spatial location of the image, but these are usually used to comment on the photograph itself.

There is a selection of photographs that are displayed on Google Earth on a monthly basis. This may motivate the users to upload good images with accurate geographic positioning.

The positional information regarding Panoramio photographs may be entered automatically if an exchangeable image file (EXIF) is used and the camera has a built-in GPS receiver. Alternatively, the photographs position can be uploaded manually, obtained from an external GPS receiver, or the photographs may be placed by the volunteer over a satellite image. Even though positional error may exist in all of these cases, they may have different degrees of accuracy.

Photographs at Panoramio have a date of upload but not the date when the photograph was taken, so there is no temporal information associated with the photographs, unless an EXIF file is used. This can be a limitation to the use of these photographs for land cover map validation purposes (Fritz *et al.* 2009).

Since there is no procedure available to ask volunteers to document a specific location, the use of Panoramio photographs alone to assess the accuracy of land cover maps requires the use of the available photographs, whose location is predefined. Moreover, the distribution of photographs is uneven in space, i.e. there are regions with many photographs and regions with no photographs. This may lead to bias in the accuracy assessment, and therefore sampling strategies need to be carefully considered in order to choose the photographs that are used.

3.2.4 Flickr

The Flickr initiative (https://www.flickr.com/) started in 2004 to help people share their photographs and videos with others. The application is not targeted to a particular kind of photograph or video, but gives freedom to the volunteers to submit all types of images. Some metadata are automatically associated with the photographs, such as date of upload, date of the photograph, camera used, and location obtained by an inbuilt GPS receiver, if available. Geographical location may also be associated with the photographs by locating them on a map or a satellite image. The user can add tags and descriptions to the photographs and include them into thematic groups, which may help find photographs by themes. There are a set of community guidelines and the content of the site is subject to moderation, but just to prevent abusive use. More than 150 million geotagged photographs were available in Flickr in April 2014, including all types of photographs. The photographs in Flickr have the same issues with spatial bias, i.e. they would not be a representative sample if they were to be used for land cover map accuracy assessment, requiring the use of an appropriate sampling strategy.

3.2.5 The Geo-Wiki Project

The Geo-Wiki Project (http://www.geo-wiki.org/) was started in 2009 at the International Institute for Applied Systems Analysis (IIASA) in collaboration with the University of Applied Sciences in Wiener Neustadt and the University of Freiburg (Fritz *et al.* 2009, 2012, Perger *et al.* 2012). The main objective of the project is to facilitate the collection of in-situ land cover data, and to assist in the training and validation of global land cover maps using high resolution satellite imagery available on Google Earth. One of the main motivations for the creation of this application was the large spatial disagreements between the three main global land cover maps (GLC-2000, MODIS and GlobCover) when compared with one another. It is also possible to upload pictures of locations visited (either manually or through the Pictures Geo-Wiki mobile application), load additional data such as the photographs and descriptions available at the DCP website or Panoramio, statistical data on the percentage of land cover for some classes, such as cropland and forest as well as five year NDVI (Normalized Difference Vegetation Index) averages at 10 day intervals across the year, to help differentiate between, e.g. evergreen and deciduous vegetation.

From the main Geo-Wiki application, volunteers can go to any location and indicate whether the three main global land cover products are good or bad at representing a given location as visible from Google Earth. However, this has produced very little VGI. Instead, a competition branch of Geo-Wiki is used in concentrated campaigns where volunteers are given random locations on the Earth's surface and are asked to identify the land cover types visible using a simplified legend of ten land cover types, similar to the ones proposed by Herold et al. (2008). Although there are no restrictions in who can participate, the main contributors have been experts in remote sensing and geospatial sciences or students in a related field (Fritz *et al.* 2012). The actual crowd has been engaged more recently through the Cropland Capture game (See, Sturn, *et al.* 2014), which asks users to determine if any cropland is visible in a given pixel or photograph viewed in the game, which is essentially a simplified, game version of the competition site. The game uses a scoring system in which correct answers are assigned a point and incorrect answers result in point loss as well as a prize draw to be held at the end of the six month period over which the game is running.

Geo-Wiki provides some tools that aim to control the quality of the data provided by the volunteers. In addition to a manual on how to use the platform, it also provides on-line instructions and videos to help volunteers to classify the land cover (result-oriented approaches), along with some classification of users according to their skills in identifying land cover correctly. In the past this has been done after the competition (using comparison with control points from experts, hence D3 in terms of the generation of these control points, but then applied in an automated way, i.e. A3). Geo-Wiki also allows contributors to associate a degree of confidence (from high to unsure) to the class assignment at each location (type B2) and asks the volunteer to indicate whether their classification was done over a high resolution satellite image or not (corresponding to different levels of B3), which may be used as an indicator of data accuracy. For the data where control points are not available, some of the validation data have been consolidated, e.g. where multiple contributions have been made at the same location. If the data have been used in subsequent validation exercises, only those contributions where agreement is higher than 65% have been used. This corresponds to approaches A2 and B2.

The most recent Cropland Capture game uses a combination of methods including:

(1) the use of pixels where the answers have been agreed upon by experts or ‘control pixels’ (A3 + D3);

(2) where no controls exist, a majority rule is implemented whereby initially players are correct until sufficient data have been collected at a single point to use the majority rule – this determines if players receive a point or not (A2); and

(3) players can challenge the answer determined by the majority rule – experts then intervene (D3), awarding the player who challenged the answer multiple points or subtracting multiple points if they were incorrect (C2).

Future competitions are planned and additional mechanisms to control for quality will be introduced, including checking by experts (D3) interactively and an internal rating system developed from a combination of factors such as performance, spatial proximity, peer review, etc. (A1, A2, B2, C2).

3.2.5 VIEW-IT Project

The Virtual Interpretation of Earth Web-Interface Tool (VIEW-IT) described by Clark and Aide (2011) is a collaborative Web-based system for automating the collection of reference data for producing and verifying the accuracy of land use/land cover maps derived from MODIS imagery. The browser-based tool aims to collect crowdsourcing interpretation of reference data from high resolution imagery available on Google Earth and allow users to visually estimate the percent cover of seven basic land cover/land use classes within a sample grid. The tool builds on the approach developed by the Geo-Wiki Project, and is described as a prototype aimed at building a global community of volunteer interpreters, especially in developing countries, where land change occurs very frequently. No link was found to this platform.

The VIEW-IT application allows the use of historical images from Google Earth, as well as Panoramio photographs, Google Charts for viewing temporal

Enhanced Vegetation Index (EVI) data and an administration data summary, as well as ArcGIS Server for displaying biome and ecoregion polygons (B3, B4). This allows the use of several types of data to perform the classification of the sample points used for the accuracy assessment.

There are two sample approaches which can be used in this application, namely, samples created by the administrator (which may use any sampling protocol) and the user can select the sample locations manually. In this latter approach it is possible to analyse the characteristics of the additional information available at that location to decide if that sample location is a good one or not (Clark and Aide, 2011).

Each VIEW-IT sample unit is a 250 × 250 m square corresponding to a MODIS pixel. This square is further decomposed into a 4 × 5-cell grid, each covering 5% of the 250 m square.

To improve the quality of the reference data, interpretations follow a protocol which provides instructions to either expert or volunteer users on how to assign the classes to the samples, enabling a decrease in thematic errors. The system allows an estimate of the percentage of land use/land cover classes at each sample unit and records the year of the image used to make the classification. If the first interpretation is from a user, the application allows the inclusion of additional interpretations made by other users (approach of type B2) without knowing the results produced by each user. The system assigns the class corresponding to the larger percentage to the sample units and if different percentages were assigned to it by the users the average of the percentage indicated by them is considered, but the original percentage information is kept in the system. Where discrepancies are found or if the image year used for the classification is different, then the answers are analysed by an expert (D3, D4). In this case the expert has access to the information about the identity of the users, so that it is possible to identify their credibility (probably an approach of type D1 or D2). If the classification is made by an expert, the classification process is closed.

The volunteers are chosen by the system administrators and have prior training using a sample dataset and their interpretation results are verified before using the system, so it is not yet a system open to the crowd. The developers however express the will to expand it to the global scale and to a larger community of users. This is similar to that of the Web-based validation tools described in Bastin et al. (2013) which were first evaluated by a limited set of trusted experts and volunteers, but then expanded into a platform suitable for citizen labelling of multi-temporal land cover across a carefully designed set of sample points.

3.2.7 OpenStreetMap

OpenStreetMap (http://www.openstreetmap.org/) is a global initiative whereby volunteers digitise detailed information on features and infrastructure, according to a model rather similar to many topographic maps. Point, line and polygon data are collaboratively submitted and edited to generate a plane-view representation of the Earth. A detailed taxonomy of tags allows buildings, for example, to be annotated with information defining their purpose and nature. In many areas (particularly developed urban zones), OpenStreetMap is more complete and informative than authoritative alternatives (Neis *et al.* 2011). However, its density and currency depends on local survey effort. Some well-defined projects exist to map regions, for example, where an urgent humanitarian response is need, such as the case of the Haiti 2010 earthquake or the more recent Guinea Ebola epidemic (http://hot.openstreetmap.org/projects). For assessments of land use that rely on the density of buildings, the hard surface and tags denoting human activity, or which map to the existing tags (http://wiki.openstreetmap.org/wiki/Key:landuse), Open Street Map can be a valuable source of information in areas where it is relatively complete: for example, urban land-use maps of impressive quality have been derived using automated decision rules and computation of coverage proportions (Jokar Arsanjani *et al.* 2013). For natural land cover types, Open Street Map has a set of agreed tags (http://wiki.openstreetmap.org/wiki/Key:natural), but these are used far less frequently, and the data may be much more heterogeneous in its detail and quality. An analysis of class coverage on a national scale, when compared to the Corine Land Cover map, showed promising results mainly for water and urban classes (Estima and Painho 2013b). There is an ongoing debate among the OSM community as to the detail with which land use and land cover should be represented in the accepted tags. Suggested conformance to official schemes such as the Land Cover Classification System (LCCS) (Di Gregorio 2005) of the Food and Agriculture Organization (FAO) have been generally seen as potentially too complex for general contributors to supply. In brief, OSM data are potentially of value for land cover validation, but present some problems due to their spatial and semantic patchiness.

4. Uses of VGI for Quality Assessment of Land Cover Maps

In this section, projects where VGI was used to assess the accuracy of land cover maps are described, indicating the data used by the authors, the procedures applied to assess the quality of the VGI and additional approaches to improve the quality of the accuracy assessment results.

VGI has been used to validate land cover maps based on two main approaches, consisting of using data such as photographs and descriptions provided by volunteers in platforms such as DCP and Panoramio, which have then to be interpreted and classified for that purpose either by other volunteers, usually selected, or experts; or by using classifications directly provided by the crowd, which may have been given access to several types of data, such as satellite imagery, photographs or NDVI values to perform the classification.

4.1 Using photographs and descriptions

For this type of approach the DCP was already tested, with promising results. Iwao et al. (2006) used 749 photographs extracted from the DCP, and the descriptions associated with them to assess the accuracy of different land cover maps of Eurasia, namely GLC2000, MOD12, UMD and GLCC. To assess the accuracy of the descriptions provided by the volunteers, three individuals with different backgrounds confirmed that the descriptions were appropriate and did not depend on expertise. These three individuals then assigned classes to all 749 sites and the land cover class assigned to the sites was the most frequent class assigned by the three volunteers, which multiple interpretations are often used in the development of ground reference datasets for land cover validation (Bontemps *et al.* 2011).

Additional procedures were used to assess the quality of these data. To assess the positional accuracy of the photographs, a set of eight confluence points was selected corresponding to sites visited more than four times. If the descriptions given by the different volunteers did not change much, it was then considered that positional accuracy could be trusted. An evaluation was also made as to whether the descriptions had changed over time to identify changes in land cover.

To assess the accuracy of the thematic information extracted from the photographs and descriptions, the classification was compared to the classification of Landsat false-colour images for thirty sites. Iwao et al. (2006) still made field visits to some of the sites. According to the authors, the results showed that the validation made using the DCP data presented the same or even higher accuracy than the one obtained with visual interpretation of Landsat images.

Several types of quality assessment were used by Iwao et al. (2006) in this project. To check the variability of the volunteer outputs, an approach of type B2 was used. Positional accuracy of the VGI was also determined using an approach of type B2. Additional checks of type D3 were still used to confirm the suitability of the data used for the assessment.

Iwao et al. (2011) also used the approach described in Iwao et al. (2006) to assess the accuracy of a land cover map generated by combining three existing land cover maps. The photographs and descriptions available at the DCP for 4,211 sites were used for the validation. No further details are given on additional quality control methodologies used.

Biradar et al. (2009) used 3,982 DCP sites along with field data and Google Earth interpretations to help label the classes of their global map of rainfed cropland areas. Of the original 6,000 DCP sites for which descriptions and photographs were downloaded, a large number did not have sufficient information to determine the land use/land cover and therefore had to be discarded. Only a sample of the field data and the Google Earth interpretations were then used for the accuracy assessment of the map. This may be due to the fact that the data from the DCP sites were not verified by Biradar et al. (2009), although as described in section 3.2.1 above, the DCP data are subject to some quality control procedures on the DCP side.

Foody and Boyd (2013) tested the use of photographs available at the DCP to assess the accuracy of the Globcover map of tropical forests in West Africa. Photographs acquired at ninety nine confluence points were used. The photographs were then interpreted independently by four volunteers, who labeled them as representing either forest or non-forest. Since errors were expected to occur during the labelling process, a latent class model was used to estimate the user's and producer's accuracy of the classification as forest or non-forest. The descriptions associated with the photographs were not used in this study. The results showed that the labelling of the photographs varied greatly between volunteers, which raises some concerns about the possible use of VGI for accuracy assessment, especially if no means to select volunteers based on the quality of their work is used. Also low levels of agreement were observed between the reference data and the Globcover map, even though many sources of uncertainty may contribute to the observed disagreements. However, the use of latent class analysis was shown to produce useful information.

Kinley (2013) compared land cover data from an area in Hampshire, UK, with tags from Geograph photographs and OSM data. The results showed a poor match between OSM and the authoritative data but a higher match between the Geograph photographs and the land cover map. The advantage of Geograph compared to Flickr and OSM for the study area considered was a much higher spatial coverage. Temporal coverage of Geograph was also shown to be good, which indicates that this source of VGI could be useful in areas where ground truth information is not available.

To the authors' knowledge no accuracy assessment of land cover has been done so far using photographs exclusively from Panoramio or Flickr. However, Wang et al. (2012) used these data to assist in the training of the classifiers, Hochmair and Zielstra (2012) assessed the positional accuracy of Panoramio and Flickr photographs, and Estima and Painho (2013a) assessed the availability of Flickr photographs on a country level to determine if they could be used for land cover map accuracy assessment. These preliminary developments might indicate that the adequacy of this information for land cover map validation will also be tested in the future, with the possible advantage of also providing a useful source of data to assess not only land cover but also land use (Newsam 2010).

In some cases this type of data was used as additional data to validate land cover map. This is the case in the validation of the Global Forest Resources Assessment (FRA) Remote Sensing Survey. This project, performed by a partnership between the Food and Agriculture Organization of the United Nations and the European Commission Joint Research Centre (Lindquist *et al.* 2012), enabled the estimation of the global forest land use and change between 1990 and 2005. The survey was made through the classification of a sample of Landsat satellite imagery at the intersection of each degree of latitude and longitude. The validation of the classification was made using Google Earth and images from both DCP and Panoramio (Lindquist *et al.* 2012), but no formal accuracy assessment was done.

4.2 Using classifications made by the crowd

As referred to in section 3, the Geo-Wiki and VIEW-IT projects were developed with the overall aim of land cover map validation, as well as integrating a variety of potentially useful data into the validation process. The Geo-Wiki project has hosted and provided data for several projects related to land cover map production and validation (Perger *et al.* 2012, Comber *et al.* 2013, Foody *et al.* 2013, Fritz *et al.* 2013, See, Comber, *et al.* 2013, See, McCallum, *et al.* 2013).

Fritz et al. (2013) and Perger et al. (2012) outline how VGI collected by the Geo-Wiki project from the human impact branch (http://humanimpact.geo-wiki.org) has been used to validate a map of land availability for the production of biofuel. This project was organized as a competition, and the contributors were scored based on the number of 1 km^2 pixels validated and the accuracy of the classifications. The project enabled the collection of a large number of points (around 55,000 from approximately 36,000 unique locations, from which around 18,000 were used in the map validation exercise). Some of these pixels were control points, which were also classified by experts, and enabled the assessment of the quality of the volunteer contributions. The overall accuracy of the classifications made by the crowd was between 66% and 76% and the agreement between the volunteer classifications was 83%. Quality was further assured by correcting for biases based on the number of classifications provided and for specific land cover types. See et al. (2013) used the results of this same project to assess the variability of class assignment between experts and non-experts, and concluded that for assessing human impact, experts and non-experts were shown to have similar performances, while for some land cover classes, the experts performed better. Although the project was considered to be successful, several aspects were identified that could further improve the results, such as allowing for indication of the percentage of land cover types, use of additional auxiliary data, such as geological maps, and implementation of a mechanism for enabling communication between the volunteers, allowing for users to learn through this channel.

Foody et al. (2013) used data collected by this project to assess the accuracy of the VGI provided by multiple volunteers, which showed considerable variation between volunteers. They then used latent class analysis to extract information on the quality of the resulting data, including the producer’s accuracy without using reference data.

See et al. (2013) also showed that LC data collected by the Geo-Wiki project could be used to map cropland using interpolation, whereby the map produced for Ethiopia had higher accuracy than existing global cropland maps.

The Geo-Wiki project has also developed several branches with different aims, including the assessment of different biomass datasets (http://biomass.geo-wiki.org), classification of urban areas into local climate zones (http://cities.geo-wiki.org), a repository for global maps of livestock (http://livestock.geo-wiki.org), a validation tool for regional-scale land cover and land cover change (http://lacoval.geo-wiki.org) and the validation of Australian maps of land cover and biophysical variables (http://auscover.geo-wiki.org). All of these branches use Geo-Wiki capabilities to validate different types of LC data.

The VIEW-IT project was used to acquire reference data to train classifiers and validate the classification results of several projects, such as the production and validation of a land use/land cover map for Latin America and the Caribbean (Clark and Aide, 2011), assessment of deforestation and reforestation of Latin America and the Caribbean (Aide et al., 2013) and identification of forest transitions in central America (Redo et al., 2012). Since VIEW-IT uses selected volunteers to perform the classifications in the described applications, they receive initial training using an example dataset, which enables the assessment of their performance before using the system. Therefore, some problems that may occur in projects that are open to all volunteers are not likely to occur, such as malicious contributions and incorrect classifications due to lack of knowledge.

Using several types of volunteers and only high resolution satellite imagery, De Leeuw et al. (2011) undertook an interesting experiment to assess the thematic accuracy in Kenya (in this case the classification of road types from imagery) using individuals with no surveying experience but local knowledge, professional surveyors with local knowledge, and professional surveyors without local knowledge. The results showed that overall, local knowledge resulted in higher accuracy, regardless of whether the individuals had surveying experience or not. Those with surveying experience but no local knowledge did considerably worse in terms of accuracy, i.e. 68% compared to 92%. There was also a difference in accuracy based on the types of roads classified, where local knowledge helped identify smaller roads and tracks more accurately than tarmacked roads (or roads which could be more easily identified from the images). The conclusions were that communities with local knowledge should be involved in the co-production of spatial information. Not only would this reduce costs and be more accurate, but the maps could be updated more frequently. The quality control was ensured by experts who visited the roads on the ground (i.e. approach of type D4).

5. Discussion and Conclusions

Amongst the wide variety of VGI currently available, some have been used as sources of data to assist in the validation of land cover maps. Two projects were developed for this aim, namely the Geo-Wiki project and the VIEW-IT project. Both use images made available by Google Earth and enable the inclusion of several other types of data to assist the volunteers, such as photographs from the DCP and Panoramio, and environmental information for the generation of more reliable information. Both projects have some training procedures and the data have subsequently been used in research. Therefore, they can be considered as promising tools. The developers of Geo-Wiki have developed several approaches to the assessment of data quality, and plan to continue to develop more approaches in this area in the future, since this is crucial for the appropriate use of VGI for these types of applications. The VIEW-IT project presents characteristics similar to the Geo-Wiki project, but is not openly available to all volunteers. Rather it relies on the use of selected volunteers for particular projects and has therefore implemented some preliminary control over the volunteer performance.

Several experiments have also been undertaken in which photographs simply collected by volunteers have been used to validate land cover maps. Photographs from the DCP have been used for this process (Iwao *et al.* 2006), although in some cases when their descriptions were not used, more divergence in the classification of the photographs was observed (Foody and Boyd 2013). This may be due to the information provided in these descriptions, to the different information used, or to other factors related to the classification of the photographs. Two aspects make the data collected by the DCP particularly useful for land cover map validation. Firstly, the photographs are collected using a systematic approach (at every integer degree of latitude and longitude), which results in a collection of "ground truth" appropriate for accuracy assessment (Stehman 2009). Secondly, at each location, photographs are collected in the four cardinal directions, which is useful to have a better understanding of the region in which the point is located. Photograph descriptions can also be of use to improve the classification of LC at these sites.

Some preliminary studies have been done regarding the use of photographs from Panoramio and Flickr, but further studies are still needed to assess the applicability of using only this source of data for the validation of land cover. One difficulty may be the uneven spatial distribution of the photographs, either geographically, temporally and by LC class, and the generation of samples may not be representative of the population (Estima and Painho 2013a). In addition, their positional accuracy may vary considerably, as shown by Hochmair and Zielstra (2012), and it can be difficult to extract information on LC classes from the photographs, since the photographs are not taken with this original purpose in mind (Estima *et al.* 2014).

To the authors' knowledge, OSM has not yet been used at any significant scale to extract reference data for validating land cover maps. However, Jokar Arsanjani et al.

(2013) showed that it is possible to produce a land cover map of urban areas using data from OSM, which suggests that OSM may eventually be useful as a source for producing reference data to assess the accuracy of ground truth, particularly in regions with high coverage of data, such as urban areas. Preliminary work has been undertaken by Estima and Painho (2013b) to establish a relation between OSM and the Corine Land Cover level 1 classes with good correspondence between the two. It is therefore expected that further developments will proceed with using OSM data for land cover map generation and validation. However, the use of this data at a much finer resolution may not be possible due to the availability of the data to date, since there may not be enough information to assess the accuracy of some classes (Estima and Painho 2013b).

One key point in considering the use of VGI to validate land cover maps is the data quality. Foody (2009, 2010, 2011, 2013) has repeatedly shown the large impacts that imperfect ground reference data may have on the results of the accuracy assessment, which demonstrates that the use of VGI for this purpose needs to be carefully controlled if reliable results are to be achieved. Therefore it is necessary to develop methods to assess the quality of VGI, so that only data with high levels of quality are used. Some of the aspects related to the quality of VGI have common features to other areas, not related to GI, such as the assessment of contributors' credibility and the labelling accuracy, and may be studied in a more general framework, benefiting from work already done in other areas. Therefore, a more focused review needs to be done regarding the methodologies already developed, for example, for the assessment of credibility of contributors for other applications. However, other aspects are specific to GI, such as positional accuracy, completeness and currency of the data, and these should be addressed within the context of GI requirements (Elwood *et al.* 2013). Even though VGI may have different levels of quality, as Foody et al. (2013)

have shown, when enough data are available, it is possible to apply methodologies that enable the extraction of useful information. Moreover, See et al. (2013), Iwao (2006) and De Leeuw et al. (2011) have shown that the contributions of volunteers may, in some cases, be as good as experts or even better, since locals with some training are more likely to produce better results than experts with no local knowledge, and therefore this source of information may be valuable for many applications.

Examining the types of approaches (Table 3) that have been used for quality assessment, it can be seen that for thematic and positional accuracy in particular, not all types of approaches have been used. For example, no methods of type A (automated) were identified for the applications reviewed above, indicating that the development of methods for these types of quality measures are still lacking. In fact, only a few automated methods have been developed so far, and most of them are to assess contributor credibility. Further automation, however, seems possible and desirable, developing more methods particularly of types A1 and A2, including, for example, several sources of VGI, as suggested by Goodchild and Li (2012) for the geographic approaches, and the assignment of a credibility or uncertainty tag to the information, so that their fitness for use could be easily assessed. This would enable the user to have information on the potential usefulness of VGI without performing expert checks (type D) to assess the information quality. It can also be noted that methods using additional data of type 1 (including metadata) are also very few. Several authors even recommend the blending of metadata and data quality measures, integrating conventional and new approaches oriented towards VGI (Coleman 2013, Johnson and Sieber 2013, Poore and Wolf 2013).

Another suggestion made by Dobson (2013) is that the construction of hybrid datasets, containing both authoritative data and crowdsourced data, as made by Google to create Google Maps, is probably the best way to increase the quality of datasets. This approach enables the use of crowdsourced knowledge to identify changes and additional data not easily collected by authorities, and maintains some of the advantages of authoritative data, mainly regarding data quality. If the lineage of the data is kept and provided to the user of the information, it will be possible to choose which information may be used for each application, based on the credibility of its source (e.g. crowdsourced by only one volunteer, by several volunteers or authoritative), as made by Tom Tom (Coleman 2010). Datasets of this type would also be valuable sources of data to both train and validate land cover maps, or even produce these maps, as proposed by Jokar Arsanjani et al. (2013) using data from OSM.

Another important aspect regarding volunteered information is to keep the volunteers interested in contributing. Very good platforms may be built for the crowd to provide relevant information, but if the contributors are not motivated and interested, then no information will be produced. For example, the Geo-Wiki team developed some branches of the application into competitions and more recently a game, which enabled them to collect more data than their initial approach. It is therefore important to develop methodologies to attract more volunteers, so that enough data are generated, enabling the use of strategies based on the crowd-sourced approach (Goodchild and Li 2012) and to provide results based on methodologies that incorporate data provided by many volunteers, such as the one proposed by Foody and Boyd (2013) and Foody et al. (2013), based on latent class analysis. Another aspect that may enable an improvement of the results is the possibility that the volunteers are able to communicate with each other, enabling a discussion when difficulties arise (Haklay *et al.* 2010, Perger *et al.* 2012).

VGI is a rich source of data that may be valuable for many applications, including land cover map validation but also map production. However, there are only a few applications in the literature that demonstrate this potential. As approaches to data quality become more mature and VGI becomes a more accepted source of information, land cover map generation may be radically improved by this new and growing source of volunteer data.

Acknowledgements

This research was partly supported by the EU COST Action TD1202 Mapping and Citizen Sensor. C. Fonte work was partially supported by the Portuguese Foundation for Science and Technology under project grant PEst-OE/ EEI/UI308/2014

References

Allahbakhsh, M., Benatallah, B., Ignjatovic, A., Motahari-Nezhad, H.R., Bertino, E., and Dustdar, S., 2013. Quality control in crowdsourcing systems: Issues and directions. *IEEE Internet Computing*, 17 (2), 76–81.

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, 9–16.

Bégin, D., Devillers, R., and Roche, S., 2013. Assessing volunteered geographica information (VGI) quality based on contributors' mapping behaviours. *In*: *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. Presented at the 8th International Symposium on Spatial Data Quality, Hong Kong, China, 149–154.

Bell, S., Cornford, D., and Bastin, L., 2013. The state of automated amateur weather observations. *Weather*, 68 (2), 36–41.

Bell, S., Cornford, D., and Bastin, L., 2014. How good are citizen weather stations? Addressing a biased opinion. *Weather*, In press.

Biradar, C.M., Thenkabail, P.S., Noojipady, P., Li, Y., Dheeravath, V., Turral, H., Velpuri, M., Gumma, M.K., Gangalakunta, O.R.P., Cai, X.L., Xiao, X., Schull, M.A., Alankara, R.D., Gunasinghe, S., and Mohideen, S., 2009. A global map of rainfed cropland areas (GMRCA) at the end of last millennium using remote sensing. *International Journal of Applied Earth Observation and Geoinformation*, 11 (2), 114–129.

Bishr, M. and Mantelas, L., 2008. A trust and reputation model for filtering and classifying knowledge about urban growth. *GeoJournal*, 72 (3-4), 229–237.

Bontemps, S., Defourny, P., van Bogaert, E., Arino, O., Kalogirou, V., and Perez, J.R., 2011. GLOBCOVER 2009: Products Description and Validation Report.
Bonter, D.N. and Cooper, C.B., 2012. Data validation in citizen science: a case study from Project FeederWatch. *Frontiers in Ecology and the Environment,* 10 (6), 305–307.
Brunsdon, C. and Comber, L., 2012. Assessing the changing flowering date of the common lilac in North America: a random coefficient model approach. *GeoInformatica,* 16 (4), 675–690.
Canavosio-Zuzelski, R., Agouris, P., and Doucette, P., 2013. A photogrammetric approach for assessing positional accuracy of OpenStreetMap© roads. *ISPRS International Journal of Geo-Information,* 2 (2), 276–301.
Ciepłuch, B., Jacob, R., Mooney, P., and Winstanley, A., 2010. Comparison of the accuracy of OpenStreetMap for Ireland with Google Maps and Bing Maps. *Proceedings of the Ninth International Symposium on Spatial Accuracy Assessment in Natural Resuorces and Enviromental Sciences 20-23rd July 2010,* 337.
Clark, M.L. and Aide, T.M., 2011. Virtual Interpretation of Earth Web-Interface Tool (VIEW-IT) for collecting land-use/land-cover reference data. *Remote Sensing*, 3 (3), 601–620.
Coleman, D., 2010. Volunteered geographic information in spatial data infrastructure: An early look at opportunities and constraints. *In*: A. Rajabifard, J. Crompvoets, M. Kanantari, and B. Kok, eds. *Spatially Enabling Society: Research, Emerging Trends and Critical Assessment*. Leuven, Belgium: Leuven University Press, 1–18.
Coleman, D., 2013. Potential contributions and challenges of VGI for conventional topographic base-mapping programs. *In*: D. Sui, S. Elwood, and M. Goodchild, eds. *Crowdsourcing Geographic Knowledge*. Springer Netherlands, 245–263.
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, 37–48.
Congalton, R.G. and Green, K., 1998. *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices*. CRC Press.
Devillers, R. and Jeansoulin, R., 2010. *Fundamentals of Spatial Data Quality*. John Wiley & Sons.
Dobson, S., 2013. VGI as a compilation tool for navigation map databases. *In*: D.Z. Sui, S. Elwood, and M. Goodchild, eds. *Crowdsourcing Geographic Knowledge*. Netherlands: Springer, 307–327.
Elwood, S., Goodchild, M.F., and Sui, D.Z., 2012. Researching volunteered geographic information: Spatial data, geographic research, and new social practice. *Annals of the Association of American Geographers*, 102 (3), 571–590.
Elwood, S., Goodchild, M.F., and Sui, D.Z., 2013. Prospects for VGI research and the emerging fourth paradigm. *In*: D.Z. Sui, S. Elwood, and M. Goodchild, eds. *Crowdsourcing Geographic Knowledge*. Netherlands: Springer, 261–275.
Esmaili, R., Naseri, F., and Esmaili, A., 2013. Quality assessment of volunteered geographic information. *American Journal of Geographic Information System*, 2 (2), 19–26.
Estima, J., Fonte, C.C., and Painho, M., 2014. Comparative study of Land Use/Cover classification using Flickr photos, satellite imagery and Corine Land Cover

database. *In*: *17th AGILE International Conference on Geographic Information Science*. Castellon, Spain.
Estima, J. and Painho, M., 2013a. Flickr geotagged and publicly available photos: Preliminary study of its adequacy for helping quality control of Corine land cover. *In*: B. Murgante, S. Misra, M. Carlini, C.M. Torre, H.-Q. Nguyen, D. Taniar, B.O. Apduhan, and O. Gervasi, eds. *Computational Science and Its Applications – ICCSA 2013*. Springer Berlin Heidelberg, 205–220.
Estima, J. and Painho, M., 2013b. Exploratory analysis of OpenStreetMap for land use classification. *In*: *Proceedings of the Second ACM SIGSPATIAL International Workshop on Crowdsourced and Volunteered Geographic Information*. New York, NY, USA: ACM, 39–46.
Feddema, J.J., Oleson, K.W., Bonan, G.B., Mearns, L.O., Buja, L.E., Meehl, G.A., and Washington, W.M., 2005. The Importance of Land-Cover Change in Simulating Future Climates. *Science*, 310 (5754), 1674 –1678.
Flanagin, A. and Metzger, M., 2008. The credibility of volunteered geographic information. *GeoJournal*, 72, 137–148.
Foley, J.A., DeFries, R., Asner, G.P., Barford, C., Bonan, G., Carpenter, S.R., Chapin, F.S., Coe, M.T., Daily, G.C., Gibbs, H.K., Helkowski, J.H., Holloway, T., Howard, E.A., Kucharik, C.J., Monfreda, C., Patz, J.A., Prentice, I.C., Ramankutty, N., and Snyder, P.K., 2005. Global Consequences of Land Use. *Science*, 309 (5734), 570 –574.
Foody, G.M., 2002. Status of land cover classification accuracy assessment. *Remote Sensing of Environment*, 80 (1), 185–201.
Foody, G.M., 2009. The impact of imperfect ground reference data on the accuracy of land cover change estimation. *International Journal of Remote Sensing*, 30 (12), 3275–3281.
Foody, G.M., 2010. Assessing the accuracy of land cover change with imperfect ground reference data. *Remote Sensing of Environment*, 114 (10), 2271–2285.
Foody, G.M., 2011. Impacts of imperfect reference data on the apparent accuracy of species presence-absence models and their predictions: Imperfect reference data. *Global Ecology and Biogeography*, 20 (3), 498–508.
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), 783–792.
Foody, G.M. and Boyd, D.S., 2012. Using volunteered data in land cover map validation: Mapping tropical forests across West Africa. *In*: *Geoscience and Remote Sensing Symposium (IGARSS), 2012 IEEE International*. Presented at the Geoscience and Remote Sensing Symposium (IGARSS), 2012 IEEE International, 6207–6208.
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), 1305–1312.
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), 847–860.
Foody, G.M., See, L., Fritz, S., van der Velde, M., Perger, C., Schill, C., Boyd, D.S., and Comber, A., 2014. Accurate attribute mapping from Volunteered Geographic Information: Issues of volunteer quantity and quality. *The Cartographic Journal*, 1743277413Y.000.

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

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 & Software*, 31, 110–123.

Fritz, S., See, L., van der Velde, M., Nalepa, R.A., Perger, C., Schill, C., McCallum, I., Schepaschenko, D., Kraxner, F., Cai, X., Zhang, X., Ortner, S., Hazarika, R., Cipriani, A., Di Bella, C., Rabia, A.H., Garcia, A., Vakolyuk, M., Singha, K., Beget, M.E., Erasmi, S., Albrecht, F., Shaw, B., and Obersteiner, M., 2013. Downgrading recent estimates of land available for biofuel production. *Environmental Science & Technology*, 47 (3), 1688–1694.

Girres, J.-F. and Touya, G., 2010. Quality assessment of the French OpenStreetMap dataset. *Transactions in GIS*, 14 (4), 435–459.

Goodchild, M.F., 2007. Citizens as sensors: the world of volunteered geography. *GeoJournal*, 69 (4), 211–221.

Goodchild, M.F. and Li, L., 2012. Assuring the quality of volunteered geographic information. *Spatial Statistics*, 1, 110–120.

Di Gregorio, A., 2005. *Land Cover Classification System. Version 2*. Rome Italy: Food and Agriculture Organization of the United Nations.

Guptill, S.C. and Morrison, J.L., 1995. *Elements of Spatial Data Quality*. Elsevier Science Limited.

Haklay, M., 2010. How good is volunteered geographical information? A comparative study of OpenStreetMap and Ordnance Survey datasets. *Environment and Planning B: Planning and Design*, 37, 682–703.

Haklay, M., 2013. Citizen science and volunteered geographic information: Overview and typology of participation. *In*: D. Sui, S. Elwood, and M. Goodchild, eds. *Crowdsourcing Geographic Knowledge*. Springer Netherlands, 105–122.

Haklay, M., Basiouka, S., Antoniou, V., and Ather, A., 2010. How many volunteers does it take to map an area well? The validity of Linus' Law to volunteered geographic information. *The Cartographic Journal*, 47 (4), 315–322.

Harvey, F., 2013. To volunteer or to contribute locational information? Towards truth in labeling for crowdsourced geographic information. *In*: D. Sui, S. Elwood, and M. Goodchild, eds. *Crowdsourcing Geographic Knowledge*. Springer Netherlands, 31–42.

Hassan, R., Scholes, R., and Ash, N., 2005. *Ecosystems and Human Well-Being - Current State and Trends: volume I*. Island Press.

Heipke, C., 2010. Crowdsourcing geospatial data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 65 (6), 550–557.

Herold, M., Mayaux, P., Woodcock, C.E., 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 (5), 2538–2556.

Hochmair, H.H. and Zielstra, D., 2012. Positional accuracy of Flickr and Panoramio images in Europe. *In*: A. Car, G. Griesebner, and J. Strobl, eds. *Proceedings of the Geoinformatics Forum*. Presented at the Geospatial Crossroads @ GI Forum '12. Proceedings of the Geoinformatics Forum, Heidelberg: Wichman, 14–23.

Ipeirotis, P.G., Provost, F., Sheng, V.S., and Wang, J., 2014. Repeated labeling using multiple noisy labelers. *Data Mining and Knowledge Discovery*, 28 (2), 402–441.

Iwao, K., Nasahara, K.N., Kinoshita, T., Yamagata, Y., Patton, D., and Tsuchida, S., 2011. Creation of new global land cover map with map integration. *Journal of Geographic Information System*, 3 (2), 160–165.

Iwao, K., Nishida, K., Kinoshita, T., and Yamagata, Y., 2006. Validating land cover maps with Degree Confluence Project information. *Geophysical Research Letters*, 33 (23), n/a–n/a.

Jackson, S., Mullen, W., Agouris, P., Crooks, A., Croitoru, A., and Stefanidis, A., 2013. Assessing completeness and spatial error of features in volunteered geographic information. *ISPRS International Journal of Geo-Information*, 2 (2), 507–530.

Johnson, P. and Sieber, R., 2013. Situating the adoption of VGI by government. *In*: D. Sui, S. Elwood, and M. Goodchild, eds. *Crowdsourcing Geographic Knowledge*. Springer Netherlands, 65–81.

Jokar Arsanjani, J., Helbich, M., Bakillah, M., Hagenauer, J., and Zipf, A., 2013. Toward mapping land-use patterns from volunteered geographic information. *International Journal of Geographical Information Science*, 27 (12), 2264–2278.

Kinley, L., 2013. Assessing the potential for crowdsourced geospatial content to enhance the quality of authoritative land cover mapping. Presented at the AGI GeoCommunity'13 Open for Business, Nottingham, UK: AGI.

Koukoletsos, T., Haklay, M., and Ellul, C., 2012. Assessing data completeness of VGI through an automated matching procedure for linear data. *Transactions in GIS*, 16 (4), 477–498.

De Leeuw, J., Said, M., Ortegah, L., Nagda, S., Georgiadou, Y., and DeBlois, M., 2011. An assessment of the accuracy of volunteered road map production in Western Kenya. *Remote Sensing*, 3 (12), 247–256.

Lenders, V., Koukoumidis, E., Zhang, P., and Martonosi, M., 2008. Location-based trust for mobile user-generated content: Applications, challenges and implementations. *In*: *Proceedings of the 9th Workshop on Mobile Computing Systems and Applications*. New York, NY, USA: ACM, 60–64.

Lindquist, E.J., D'Annunzio, R., Gerrand, A., MacDicken, K., Achard, F., Beuchle, R., Brink, A., Eva, H.D., Mayaux, P., San-Miguel-Ayanz, J., and Stibig, H.-J., 2012. *Global Forest Land-Use Change 1990–2005*. Rome, Italy: Food and Agriculture Organization of the United Nations and European Commission Joint Research Centre, FAO Forestry Paper No. No. 169.

Lu, D. and Weng, Q., 2007. A survey of image classification methods and techniques for improving classification performance. *International Journal of Remote Sensing*, 28 (5), 823–870.

Neis, P., Zielstra, D., and Zipf, A., 2011. The street network evolution of crowdsourced maps: OpenStreetMap in Germany 2007–2011. *Future Internet*, 4 (4), 1–21.

Newsam, S., 2010. Crowdsourcing what is where: Community-contributed photos as Volunteered Geographic Information. *IEEE MultiMedia*, 17 (4), 36–45.

Nie, W., Yuan, Y., Kepner, W., Nash, M.S., Jackson, M., and Erickson, C., 2011. Assessing impacts of Landuse and Landcover changes on hydrology for the upper San Pedro watershed. *Journal of Hydrology*, 407 (1–4), 105–114.

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), 5768–5788.
Perger, C., Fritz, S., See, L., Schill, C., van der Velde, M., McCallum, I., and Obersteiner, M., 2012. A campaign to collect volunteered geographic information on land cover and human impact. *In*: T. Jekel, A. Car, J. Strobl, and G. Griesebner, eds. *GI_Forum 2012: Geovisualisation, Society and Learning*. Berlin / Offenbach: Herbert Wichmann Verlag, 83–91.
Pontius, R.G. and Millones, M., 2011. Death to Kappa: birth of quantity disagreement and allocation disagreement for accuracy assessment. *International Journal of Remote Sensing*, 32 (15), 4407–4429.
Poore, B.S. and Wolf, E.B., 2013. Metadata squared: Enhancing its usability for Volunteered Geographic Information and the GeoWeb. *In*: D. Sui, S. Elwood, and M. Goodchild, eds. *Crowdsourcing Geographic Knowledge*. Springer Netherlands, 43–64.
Raykar, V.C. and Yu, S., 2012. Eliminating spammers and ranking annotators for crowdsourced labeling tasks. *J. Mach. Learn. Res.*, 13, 491–518.
See, L., Comber, A., Salk, C., Fritz, S., van der Velde, M., Perger, C., Schill, C., McCallum, I., Kraxner, F., and Obersteiner, M., 2013. Comparing the quality of crowdsourced data contributed by expert and non-experts. *PLoS ONE*, 8 (7), e69958.
See, L., McCallum, I., Fritz, S., Perger, C., Kraxner, F., Obersteiner, M., Deka Baruah, U., Mili, N., and Ram Kalita, N., 2013. Mapping cropland in Ethiopia using crowdsourcing. *International Journal of Geosciences*, 4 (6A1), 6–13.
See, L., Mooney, P., and Foody, G.M., 2014. *Working Group 1 Acquiring and Managing VGI*. COST Action Mapping and the Citizen Sensor Working Paper.
See, L., Sturn, T., Perger, C., Fritz, S., McCallum, I., and Salk, C., 2014. Cropland Capture: A gaming approach to improve global land cover. *In*: *17th AGILE International Conference on Geographic Information Science*. Castellon, Spain.
Steele, B.M., 2005. Maximum posterior probability estimators of map accuracy. *Remote Sensing of Environment*, 99 (3), 254–270.
Stehman, S.V., 2009. Sampling designs for accuracy assessment of land cover. *International Journal of Remote Sensing*, 30 (20), 5243–5272.
Stehman, S.V., Olofsson, P., Woodcock, C.E., Herold, M., and Friedl, M.A., 2012. A global land-cover validation data set, II: augmenting a stratified sampling design to estimate accuracy by region and land-cover class. *International Journal of Remote Sensing*, 33 (22), 6975–6993.
Strahler, A., Boschetti, L., Foody, G.M., Friedl, M.A., Hansen, M.C., Herold, M., Mayaux, P., Morisette, J.T., Stehman, S.V., and Woodcock, C., 2006. *Global Land Cover Validation: Recommendations for Evaluation and Accuracy Assessment of Global Land Cover Maps*. Luxembourg: Office for Official Publications of the European Communities.
Tang, W. and Lease, M., 2011. Semi-supervised consensus labeling for crowdsourcing. Presented at the SIGIR 2011 Workshop on Crowdsourcing for Information Retri eval, Beijing, China, 36–41.
Verburg, P.H., Neumann, K., and Nol, L., 2011. Challenges in using land use and land cover data for global change studies. *Global Change Biology*, 17 (2), 974–989.
Vuurens, J.B.P. and de Vries, A.P., 2012. Obtaining high-quality relevance judgments using crowdsourcing. *IEEE Internet Computing*, 16 (5), 20–27.

Wang, S.-Y., Liao, W.-S., Hsieh, L.-C., Chen, Y.-Y., and Hsu, W.H., 2012. Learning by expansion: Exploiting social media for image classification with few training examples. *Neurocomputing*, 95, 117–125.
Woodcock, C.E. and Gopal, S., 2000. Fuzzy set theory and thematic maps: accuracy assessment and area estimation. *International Journal of Geographical Information Science*, 14 (2), 153–172.
Zielstra, D. and Zipf, A., 2010. A comparative study of proprietary geodata and volunteered geographic information for Germany. Presented at the 13 th AGILE International Conference on Geographi c Information Science 2010, Guimarães, Portugal, 1–15.

Table 1. Degree of automation of the VGI quality assessment

Fully automated

Type	Description
A	Methods are fully automated and use available information in the application
B	Corrections or assessments made by the crowd - they do require human intervention, but it is made by the crowd, not by the system managers or other experts
C	Corrections or assessments made by selected volunteers, which are supposed to create more reliable information. This requires the intervention of the system managers or other experts, in the choice of the reliable volunteers
D	Corrections or assessments made by experts

Manual

Decreasing degree of automation

Increasing degree of quality and costs

Table 2. Types of data used for quality assessment

Type	Description
1	Data which can be automatically collected by the system, either about the volunteer or the contributed VGI, such as some types of metadata (e.g. number of contributions, geographic location, date of contribution)
2	Data provided by volunteers in colaborative systems
3	Proprietary data, collected and owned by firms or institutions, to which access is granted and can be used as base data to assess some types of accuracy, such as positioning over satellite images (these may have several degrees of quality, related for example with resolution or scale)
4	Authoritative data, ground truth

Increasing degree of outcomes quality

Table 3. Relation between degree of automation and data used to assess accuracy

		Degree of automation and involvement of the professionals			
		Automated (A)	Made by the crowd (B)	Selected volunteers (C)	Experts (D)
Types of additional data used for quality assessment	Available data / metadata (1)	A1	B1	C1	D1
	Crowdsourced data (2)	A2	B2	C2	D2
	Base proprietary data freely availabe (3)	A3	B3	C3	D3
	Authoritative data (4)	A4	B4	C4	D4