

Assuring the quality of VGI on land use and land cover: experiences and learnings from the LandSense project

Giles Foody, Gavin Long, Michael Schultz & Ana-Maria Olteanu-Raimond

To cite this article: Giles Foody, Gavin Long, Michael Schultz & Ana-Maria Olteanu-Raimond (2022): Assuring the quality of VGI on land use and land cover: experiences and learnings from the LandSense project, Geo-spatial Information Science, DOI: [10.1080/10095020.2022.2100285](https://doi.org/10.1080/10095020.2022.2100285)

To link to this article: <https://doi.org/10.1080/10095020.2022.2100285>



© 2022 Wuhan University. Published by Informa UK Limited, trading as Taylor & Francis Group.



Published online: 21 Jul 2022.



Submit your article to this journal [↗](#)



Article views: 120



View related articles [↗](#)



View Crossmark data [↗](#)

Assuring the quality of VGI on land use and land cover: experiences and learnings from the LandSense project

Giles Foody ^a, Gavin Long ^b, Michael Schultz ^c and Ana-Maria Olteanu-Raimond ^d

^aSchool of Geography, University of Nottingham, Nottingham, UK; ^bNottingham University Business School, Nottingham, UK; ^cGeographisches Institut, Universität Tübingen, Tübingen, Germany; ^dLASTIG, Univ Gustave Eiffel, IGN-ENSG, Paris, France

ABSTRACT

The potential of citizens as a source of geographical information has been recognized for many years. Such activity has grown recently due to the proliferation of inexpensive location aware devices and an ability to share data over the internet. Recently, a series of major projects, often cast as citizen observatories, have helped explore and develop this potential for a wide range of applications. Here, some of the experiences and learnings gained from part of one such project, which aimed to further the role of citizen science within Earth observation and help address environmental challenges, LandSense, are shared. The key focus is on quality assurance of citizen generated data on land use and land cover especially to support analyses of remotely sensed data and products. Particular focus is directed to quality assurance checks on photographic image quality, privacy, polygon overlap, positional accuracy and offset, contributor agreement, and categorical accuracy. The discussion aims to provide good practice advice to aid future studies and help fulfil the full potential of citizens as a source of volunteered geographical information (VGI).

ARTICLE HISTORY

Received 11 April 2021
Accepted 7 July 2022

KEYWORDS

Crowdsourcing; citizen sensing; volunteered geographical information (VGI); data quality assurance

1. Introduction

Citizen science for the provision of geographical data has a long history and is known by a variety of terms (See et al. 2016) but has developed rapidly since the advent of volunteered geography and volunteered geographical information (VGI) pioneered by Goodchild (2007). The term VGI is taken here to relate to geographical information provided by the public for little reward if any. Developments such as Web 2.0 and the proliferation of inexpensive location aware devices has greatly eased the acquisition of VGI and resulted in a considerable growth of the subject (Capineri et al. 2016; Foody et al. 2017a). A range of fundamental issues that are central to the acquisition, storage, management, distribution and use of VGI can be encountered in citizen science projects (Bastin, Schade, and Schill 2017; Demetriou et al. 2017). Here, the focus is on the quality of VGI on land use and land cover.

Citizen observatories offer a potential revolution in the field of land use and land cover monitoring. Citizens can greatly increase the capacity and frequency of data collection, enabling, for example, near real-time response to emerging environmental hazards (Ostermann, and Granell 2017) as well as facilitating conventional activities such as map updating (Olteanu Raimond et al. 2017a; Olteanu-Raimond et al. 2017b). The growing potential of VGI to augment or even replace authoritative geographical data on land cover and land use, which

can be expensive to acquire or have restrictions on access and use, has been recognized widely (Fonte et al. 2015a; Stehman et al. 2018). Indeed, citizens can be an attractive source of data on land cover and land use, addressing concerns such as the amount, spatial distribution and timeliness of authoritative data. For example, the use of citizen-based contributions is one effective way to increase the ground data available to support analyses of remotely sensed data (See et al. 2022). However, concerns have been raised about the quality of VGI when compared to the traditional land surveying techniques and methods employed by bodies such as national mapping agencies.

There are many concerns about the quality of citizen-derived data such as its accuracy, trustworthiness, distribution, and heterogeneity (Flanagin, and Metzger 2008; Elwood, Goodchild, and Sui 2012; Fonte et al. 2015a; Fogliaroni, D'Antonio, and Clementini 2018; Vahidi, Klinkenberg, and Yan 2018; Severinsen et al. 2019). Quality assurance (QA) and assessment are, therefore, important, and methods to ensure that the quality of VGI can be characterized so that they are effective for user needs are required if the full potential of VGI is to be realized. A range of approaches are available for the assessment of VGI quality (Goodchild, and Li 2012; Senaratne et al. 2017) and can be impacted by the source of reference data (Mocnik et al. 2018).

Issues of QA have been a major issue in citizen science projects such as LandSense. LandSense was a European Union (EU) funded project that sought to develop the use of citizen observations to enhance Earth observation studies (Moorthy et al. 2017; Wannemacher et al. 2018).

The underlying need for the LandSense project was to help achieve the full potential of remote sensing for Earth observation in environmental monitoring. The ground data requirements needed to support a remote sensing study can be demanding and difficult to satisfy, especially if the spatial and temporal dimensions of the study are large. However, citizen science offers a means to help address key challenges with ground data. The LandSense project linked Earth observation with contemporary citizen science to help deliver quality-assured ground data to complement and support environmental monitoring systems that use remotely sensed data (Moorthy et al. 2017). The core motivations of the project were to enhance the quality of land use and land cover products generated from remote sensing, to transform the approach to satellite monitoring and to increase the engagement of citizens in environmental monitoring activity (Moorthy 2020). In LandSense, citizens were encouraged to collect ground data which could be useful support to environmental monitoring systems and, if appropriate, integrated with authoritative data sets. The activity, thus, enhanced the role of citizen science in Earth observation, increasing the involvement of citizens in science as well as providing a means to generate accurate, timely and low cost ground data to support the use of remote sensing in addressing major environmental challenges.

A range of application themes were studied in the LandSense project, notably urban landscape dynamics, agricultural land use and forest and habitat monitoring. For each of the aforementioned themes, one or more pilot or demonstration studies were undertaken. Here, attention focuses on QA of data acquired from four pilot studies on urban landscape dynamics (OSMlanduse validation, City.Oases, MijnPark.NL and Paysages), one agricultural pilot (CropSupport) and one forest and habit monitoring (Natura.Alert) pilot. As with other studies focused on citizen science contributions to environmental research (e.g. Salk et al. 2016; Mobasher, Zipf, and Francis 2018; De Marchi, Ficorilli, and Biggeri 2022), attention was focused on lessons learnt from these studies. This adds to the growing literature on how experiences gained in crowdsourcing projects can enhance remote sensing research, notably as a source of ground reference data (See et al. 2022), and aid the design of future research programs.

There are many dimensions to data quality (Hickling Arthurs Low Corporation 2012; Fonte et al. 2015a). Here, the focus is entirely in relation to

the data acquired for the LandSense project and its objectives. LandSense QA processes were founded upon a review of the literature on quality assessment of VGI and the work of two previous EU funded projects: COST Action TD1202 (Foody et al. 2017a) and the COBWEB project (Higgins et al. 2016). Some of the key details are revisited here as they impacted the design and execution of QA checks undertaken in LandSense.

Building on the foundations provided by the literature and the two prior EU funded projects, a set of QA tools to meet the objectives of the LandSense project were developed and tested. In the course of this activity, good practices for QA of citizen-contributed data emerged and are reported here. Although the discussion is focused on LandSense applications, many of the practices and lessons learned should have broader relevance and could be applicable or adaptable to other studies. Good practice guidelines and protocols for VGI are emerging (e.g. Fonte et al. 2015b; Mooney et al. 2016; Minghini et al. 2017) and some are revisited and expanded upon based on the experiences gained from the LandSense project. The focus throughout is on good practices for projects using VGI recognizing that there is a desire to not alienate the citizen community by making the acquisition process too onerous or difficult (Fonte et al. 2015b; Foody et al. 2017b; Minghini et al. 2017).

Guidelines are provided for each of the QA tools developed and implemented during the LandSense project. The latter are checks on photographic image quality and privacy, polygon overlap, positional accuracy and offset, contributor agreement and categorical accuracy. Section 2 provides a brief review of the status of relevant QA checks. Section 3 reviews some key background issues including the LandSense pilot studies that generated VGI. Section 4 summarizes the QA checks undertaken and lessons learnt and Section 5 concludes the paper.

2. Quality assurance of VGI

Here, a brief overview of some key aspects of QA of VGI is provided to illustrate the current status of such activity before addressing LandSense specific issues. The discussion is focused on material relevant to the QA checks used in LandSense and structured under four general headings: photographic imagery, polygon overlap, positional accuracy and label quality.

2.1. Photographic imagery

Photographs, especially if geotagged, are a popular form of VGI (Brabyn, and Mark 2011; Elwood, Goodchild, and Sui 2012; Feick, and Robertson 2015; Chesnokova, and Purves 2018). Smartphones have, for example, made it effort-free to acquire

geotagged photographs, and these can be interpreted to yield useful land cover and land use information (Antoniou, Morley, and Haklay 2010; Xu et al. 2017). Photographs could be acquired for a variety of purposes and their use as VGI may sometimes have been unplanned. This is, for example, the case with the photography acquired in the Degree Confluence Project (<https://confluence.org>) where the photographs were acquired primarily as a leisure activity but which could be interpreted to yield land cover information and used as ground reference data to support analyses of remotely sensed data (Iwao et al. 2006). Critically, it means many photographs may have been obtained in an unconstrained manner. The photographer will not have been working to constraints of, for example, lighting, orientation and privacy. Consequently, photographs acquired by citizens may be degraded by concerns such as poor brightness levels and blur and hence be more difficult to use than constrained photographs upon which methods for information extraction and quality control are relatively advanced (Wolf, Hassner, and Maoz 2011; Nada et al. 2018).

Photograph quality can be analyzed and expressed in a variety of ways (Ke, Tang, and Jing 2006). It is often desirable to filter a set of photographs to eliminate or enhance those unsuitable because of problems such as inappropriate brightness levels or blur which can hinder interpretation (Lo et al. 2015; Griesbaum, Marx, and Höfle 2017; Wu et al. 2021). Indeed, studies using VGI have recognized that photographs provided by citizens can be improved by image enhancement operations (Elia, Balbo, and Boccardo 2018). In LandSense, attention was focused on three features: the degree of image blurriness, image brightness and the presence of privacy features. Other properties such as image resolution, bit or color depth, contrast and orientation (Szeliski 2010; Shima, Nakashima, and Yasuda 2017) could be added to the QA platform if required.

The usability of a photograph is in part a function of the degree of blurring present and hence, as in other studies (e.g. Havlik et al. 2013), blur checking was included in the QA checks. A blurred photograph could be acceptable if it is to be used for the estimation of something very general and easy to identify such as building height via the number of floors but of little or no value if identifying something very detailed and specific such as crop species type. Similarly, the brightness of a photograph affects its interpretability with extremely dark and bright photographs often problematic. The checks for image blur and brightness sought to identify potentially unsuitable photographs so that they could be excluded or subjected to enhancement operations to improve interpretability.

The aim of photograph privacy checks is to ensure the anonymity of people who appear, perhaps unintentionally, in photographs acquired as part of the pilot campaigns. While the photographs may have been acquired to act as a source of land cover and use information, they may, especially in urban areas, include people and individually identifiable features (e.g. vehicle license plates). This situation can be a major limitation especially in the context of ensuing compliance with the General Data Protection Regulation (GDPR), which may require that such features to be masked out in some way to ensure anonymity. Thus, features such as faces and vehicle license plates which can be linked to an individual must be identified and obscured.

There is a large literature on automatic face and license plate detection and considerable success has been achieved with constrained photographs. Photographs acquired by citizens are, however, typically unconstrained images, which present greater challenge and consequently are associated with less accurate detection of privacy features (Yang et al. 2016; Silva, and Jung 2018). Thus, while very high (90 +%) detection rates may be obtained from constrained images lower rates are expected from unconstrained images. For example, Yang et al. (2016) observe an accuracy of approximately 70% for detecting faces in images and this might be considered a reasonable performance target for a nonspecialist VGI project. Full GDPR compliance may, therefore, require some further action such as manual interpretation in addition to the use of an automated feature detector.

2.2. Polygon overlap

Some VGI takes the form of polygons (e.g. field boundaries) which can be especially useful in, for example, object-based analyses of remotely sensed data where the object may be a feature such as an agricultural field. A common concern with such data are features such as overlapping polygons. This type of problem can be readily addressed using conventional intersect functions in basic geographical information systems (Burrough et al. 2015). Consequently, VGI projects may make use of basic tools to identify overlaps and potentially remove these errors through operations such as the clipping of overlapping polygons (Wadembere, and Ogao 2010). There are also a range of open-source resources with functions to allow polygon topology correction that could be implemented to correct overlaps (Obe, and Hsu 2011).

2.3. Positional accuracy

Positional accuracy is a fundamental issue with geographic information. The topic is well established with methods of measurement and standards available

(Congalton, and Green 2009). Positional accuracy will impact on VGI in many ways. Positional errors could, for example, be a source of error with polygon data. Positional error is also a major issue with geotagged photographs. Indeed, popular sources of geotagged photographs may differ substantially in terms of the magnitude of positional error present (Zielstra, and Hochmair 2013). Such error is, however, often a major concern, especially, for example, in site-specific approaches to accuracy assessment used in remote sensing (Congalton and Green 2009). Studies on location tracking using mobile devices of the sort citizens may use in acquiring VGI have shown horizontal accuracies typically in the range 5–15 m (Merry, and Bettinger 2019; Menard et al. 2011). Newer devices tend to demonstrate higher positional accuracy. For example, Zandbergen (2009) found an average horizontal position error of 10 m for the iPhone 3 G. This was reduced to 6.5 m for the iPhone 4S according to Garnett and Stewart (2015). The technical capability of typical mobile devices that will be used in data collection should be considered when assessing the degree of positional accuracy that can be achieved. This may influence the design of any data collection exercise.

2.4. Label quality

In many studies, the VGI acquisition involves labeling cases (e.g. allocating a land cover class to a field). The quality of the labels may be assessed in a variety of ways. Of relevance to LandSense is an interest in the quality of the labeling by different citizens, thus determining the extent to which they agree in their labeling, but also in the accuracy of labeling, thus determining the amount of error in the labeling by comparison to reality; note it is possible for citizens to agree in labeling but be incorrect. In both the assessment of contributor agreement and of accuracy, the core QA task involves comparing a set of labels. A basis for both types of assessment is similar to that suggested for accuracy assessment in remote sensing (Congalton, and Green 2009) and for which good practices exist (Olofsson et al. 2014). Recently, methods that account for spatial autocorrelation commonly encountered with geographic information have been promoted for accuracy assessment (Ploton et al. 2020). Such approaches are, however, unsuitable for the task and the conventional approach based on design-based inference upon which established good practice guidance (Olofsson et al. 2014) is based should still be used (Wadoux et al. 2021; Meyer, and Pebesma 2022). The latter has three main components: response design, sampling design and analysis (Stehman, and Czaplewski 1998). Some of the main issues in this good practice advice presented by (Olofsson et al. 2014) are restated and revisited in this Section to encourage their use in other VGI studies.

The steps that lead to a decision regarding agreement between the two or more sets of labeled data lie within the response design. This includes a labeling protocol that provides specification on exactly what is being labeled and the definition of agreement to be used (Stehman, and Czaplewski 1998; Olofsson et al. 2014). Thus, for example, it is important that there is clear and unambiguous guidance on the task so that the contributors understand exactly what the task is, the options open to them, and that a suitable method is used to assess their agreement. These various issues can be non-trivial and hence clear guidance and supporting documentation may be required. For example, in the seemingly simple task of allocating a land use class label to a building it must be clear what an individual building is. For example, is it a discrete detached feature or can there be spatially joined buildings? The latter distinction can be significant, for instance, in residential areas where the distinction between detached and terraced housing may be important. The classes to be used should also normally be clear, discrete, mutually exclusive and exhaustively defined. Although the latter may seem straightforward, it can be easy to fail to fully satisfy the assumed condition. For example, in a study focused on agricultural crops it may be possible to exhaustively define every crop type that could occur but fail to include classes such as water, urban or woodland that could form part of the entire landscape. There should also be means to address common problems such as what to do for cases that may involve a mixture of classes (e.g. contributors could be instructed to label to the dominant class). Finally, an appropriate analysis is required to provide meaningful information on the level of agreement that exists between the data provided by the contributors. Issues such as sampling design which determines the locations where labels are acquired for should be considered especially if seeking to generalize the results which is often the case in the assessment of the accuracy of land use and land cover maps. In some studies, especially those focused on contributor agreement, sampling issues may not be critical as attention may simply be directed at the degree of agreement in labeling of a specific set of data.

A key part of the response design for the assessment of both contributor agreement and accuracy is the labeling protocol, and this can be challenging as many class definitions often exist (e.g. Comber, Fisher, and Wadsworth 2005; Comber, Wadsworth, and Fisher 2008; Ahlqvist 2008). In LandSense, this challenge was also sometimes magnified as highly subjective phenomena related to perceptions of place were studied. It is critical that a clear and unambiguous set of classes are defined and that all of the contributors can use these labels effectively. To help achieve this situation, the response design should

include means to enhance the consistency in labeling (Olofsson et al. 2014) perhaps by provision of supporting documentation (e.g. classification keys and examples) and training in labeling. It may also be useful to acquire self-confidence ratings in labeling (e.g. Comber, See, and Fritz 2014) to help filter a data set, although these can be problematic as poor contributors may over-estimate their abilities (Ehrlinger et al. 2008). It is also important to ensure that the contributors are labeling the same geographical feature. In LandSense, a key aid to this is to ensure geolocation and hence checks on positional accuracy can usefully aid the QA of labeled data.

A fundamental issue is the need to define clearly and unambiguously what is being labeled. In some LandSense applications, the focus was on the labeling of remotely sensed imagery, often obtained via an image classification analysis to yield a thematic map. In such circumstances, there is a need to define the minimum mapping unit and a spatial unit for the accuracy assessment (Olofsson et al. 2014). The former may vary from application to application but can have important impacts on a map and its accuracy which are often influenced by the size and spatial distribution of the land cover patches (Saura 2002; Knight, and Lunetta 2003). The spatial unit for accuracy assessment is a central feature in popular site-specific approach to accuracy assessment (Congalton, and Green 2009). In this approach to accuracy assessment, the analysis is based upon a comparison of the contributed class label and that which exists in reality and shown in the reference data set. It may be helpful, but is not essential, that the same spatial unit is used in both data sets. The latter highlights the importance of ensuring accurate geolocation and hence the potential value of the positional accuracy checks.

A commonly encountered problem in analyses of remotely sensed data is that the pixel is an arbitrary spatial unit and it may sometimes be preferable to use something else such as a land parcel or field which can be easier to locate. Differences in the spatial unit can be accommodated, but the analysis should recognize the issue. Land parcels are typically polygonal features and may be obtained in some practical applications via an image segmentation analysis. There are many methods and challenges in segmenting an image optimally (Costa, Foody, and Boyd 2018) and it may be helpful to ensure that the reference data and polygon of interest spatially overlap using a polygon overlap check. Regardless of the specific spatial unit used, it is common to find that class mixing occurs. For example, mixed pixels may be common as it is an artificial unit that may not provide a realistic representation of the land mosaic but mixed objects are also commonly encountered (Costa, Foody, and Boyd 2017). A way to accommodate such mixed cases in the analysis is required. For example, a mixed unit could be allocated

the dominant class label and the accuracy assessment proceeds as normal; although this is an imperfect approach as it essentially degrades the data. Alternatively, a fuzzy approach to accuracy assessment could be adopted when mixed units occur (Gopal, and Woodcock 1994; Stehman, and Foody 2019).

A fundamental assumption made in an accuracy assessment is that the reference data set represents reality or ground truth (i.e. a “gold standard” reference data set). However, this is rarely the case and the source of reference data has implications to the assessment of data quality (Mocnik et al. 2018). Error inevitably exists, and it is important to recognize that error in the reference data set, even at small amounts, can substantially degrade an analysis (Foody 2013). A common error source is spatial mis-registration (Pontius 2000), and this could be reduced using positional accuracy tools. Sometimes the reference data come from multiple contributors (Wulder et al. 2007; Wickham et al. 2013). In such situations, it is common to focus on only a subset of the cases. For example, sometimes it may be appropriate to use only the cases upon which all of the contributors agree on labels (Scepan 1999). In other instances, it may be appropriate to use a consensus label for each case. Some VGI-based projects have suggested that a set of between 3 and 15 contributors is often sufficient for some common applications (Haklay et al. 2010; Foody et al. 2015). Care must be taken to ensure an appropriate approach is adopted. For example, by focusing on only cases of complete agreement, a data set may be limited to relatively unrepresentative sample locations of simple homogeneous composition. It is also common for disagreements to be largest and most important for rare classes (Wulder et al. 2007; Stehman, and Foody 2019; Xing et al. 2021) and so by ignoring such cases a study may end up excluding rare classes by accident. Furthermore, in projects such as LandSense, the reference data may often arise from volunteers (Laso Bayas et al. 2017; Waldner et al. 2019) as well as experts (Xing et al. 2021) and often involves either fieldwork or the interpretation of aerial photography or satellite imagery of the region mapped. Irrespective of the source of the reference data, some variation in labeling is often observed (e.g. Xing et al. 2021). Good practice advice (Olofsson et al. 2014) urges that the reference data be of higher quality than the map being evaluated.

In many scenarios, the sampling design used to acquire the data for an accuracy assessment is of critical importance. If the aim is to not merely consider the accuracy with which a particular testing set has been classified and there is a desire to generalize, such as to an entire map, then good practice guidelines call for the use of a probability sample design (Olofsson et al. 2014). The latter include popular

approaches such as simple random, stratified, cluster and systematic sampling (Stehman 1999) and guidance on the selection for different study objectives is provided in the literature (Stehman 2009). Of key concern for typical LandSense type application is that if there is a desire to assess the accuracy of a land use or land cover map a probability sample should be used and its properties considered in the estimation of accuracy (Olofsson et al. 2014). Thus, for example, if a stratified sample design is used, the size of the strata should be accommodated in the estimation of map accuracy or phenomena such as class areal extent (Olofsson et al. 2013). The required sample size can be estimated from sampling theory and is important in influencing the width of the confidence interval that may be fitted to estimates. Alternatively, simple heuristics, such as a minimum of 50 cases per class (Hay 1979), may be adopted if appropriate. Note that data from a non-probability sample, as may occur with some VGI, can be usefully integrated with such data to enhance analyses (Stehman et al. 2018).

Finally, the cross-tabulation of the labels from two sources being evaluated yields a confusion matrix from which a set of quantitative measures of labeling quality can be estimated (Olofsson et al. 2014). In the case of an accuracy assessment the matrix is often termed an error matrix as the reference data set represents reality. In the matrix, the cases lying on the main diagonal are those on which both data sources agree while the cases in off-diagonal elements represent disagreements. There are many measures of agreement and accuracy that can be calculated from the confusion or error matrix (Card 1982; Fielding, and Bell 1997; Olofsson et al. 2014). Some of the measures may be used in assessments of both contributor agreement and accuracy, but care may be needed in interpretation. Note, for example, that the kappa coefficient which is widely used in QA is suited for use in measuring contributor agreement but not in accuracy assessment (Pontius, and Millones 2011; Foody 2020). Other widely used quality measures include the proportion of all cases that were correctly classified based on the reference data as well as measures for individual classes (i.e. measures the proportion of correct classifications of a given class based on the reference data). As the assessment is typically based upon a sample of cases, good practice would be to fit confidence limits to the estimate to indicate the degree of uncertainty present (Olofsson et al. 2014).

There are many challenges in accuracy assessment, and other approaches may sometimes be suitable. For example, if the map and reference data legends do not match or perhaps contain a different number of classes then alternative methods, such as those based on entropy, may be undertaken (Finn 1993; Stehman, and Foody 2019). Similarly, there are also measures to address deviations from the standard scenario based on a comparison of a pair of labels for each case. For

example, in situations when agreement by more than two contributors is being assessed it may be possible to use a measure such as Fleiss's kappa (Fleiss 1971).

3. Background details and the LandSense pilot studies

The LandSense pilot studies differed in a number of ways, which impacted on QA tasks. For example, pilot studies used dissimilar scripting languages, version control systems and deployment platforms. However, a federated approach was used to accommodate for these differences in terms of accessibility and use. The latter included a common quality control entity to identify, assess and potentially correct data quality concerns across these multiple applications, themes and pilot studies.

Data standards for LandSense, and how these interact with QA procedures, included essential and recommended Data Collection Requirements (DCRs). DCRs define the nature of the data collected and how it relates to the QA functions. Essential DCRs relate to mandatory data requirements, such as privacy checks on photographic imagery and time stamps on data. Recommended DCRs are not mandatory requirements but relate to data protocols aimed at promoting good data quality. For example, the latter include checking photographs for blurring and appropriate brightness levels as well as the level of positional accuracy for contributed data.

Throughout the evolution of the LandSense project, all pilot studies were open to change and adaptation throughout their development regarding their concept, audience, content, and technicalities. This enabled them to focus better on user needs and to take into account lessons learnt during initial data collection campaigns. As a result, data protocols evolved and were subjected to reconsideration and reframing. Hence, a key aspect for good practice of the LandSense data protocols was the ability to accommodate these dynamics to ensure that ongoing operational use and development could occur simultaneously.

The pilot studies focused on the urban landscape dynamics theme witnessed changes to the data collection protocol and data schema to take better account of user's needs as well as development of more citizen-friendly tasks such as validation to take into account the lessons learned from initial data collection campaigns. This resulted in the simplification of the data collection protocol and data model. In contrast, the pilot studies for of the agricultural (CropSupport) and Forest Monitoring (Natura.Alert) themes evolved more gradually without any fundamental changes within their foci, content, or data schema.

Balancing the need for each pilot study to develop its own bespoke methods whilst maintaining some standardization across the LandSense project was

a challenge throughout the project. This is not surprising given the diversity of the pilot studies. In addition, the differing requirements of the various stakeholders to the LandSense project (e.g. users, government, commercial and academic) should also be considered. For example, commercial partners may be reluctant to release their intellectual property under open access conditions whereas funding organizations (e.g. the EU, national funding bodies, etc.) may require open access as part of the funding conditions. The concepts of cohesion, adaptability, and accessibility governed the design of data structures for the LandSense pilot studies.

Cohesion is the capacity to use the pilot studies in a combined manner such that they each could benefit from the others and may reuse parts where appropriate. Cohesion was aided by agreeing on shared standards. The adoption of established geodata standards acted to simplify access and reduce friction for data access, particularly with an interlinked federation environment. For LandSense, geojson was used for vector data since it is robust, adaptable, and suited for web applications. In addition, the World Geodetic System, which is the default projection of geojson, was used as it supports deployment scalability which is important for integrating data collection over the world. Furthermore, coding in R and Python was encouraged as they can be installed at no cost, are human readable and understandable by a broader academic and non-academic public. The code was maintained within a project managed git (gitLab, gitHub) environment allowing multiple access.

Adaptability is the capacity to keep a citizen science project flexible in terms of the requirements dictated by commercial, technical, scientific or policy constraints. Advances in technology may offer new, unanticipated, opportunities to which a project should be able to adapt to. Similarly, the aims and scope of a study may change and hence the ability to adapt can be important. For example, the app data protocol should support a rapid upscale or down-scaling capacity depending on the intensity of use. Containerized deployment chunks through Docker containers (Merkel 2014) and eventually Docker

swarms (Freeman 2017) or similar were recommended for use in parts of LandSense such as in the photograph privacy checks. These can be extended as needed and can be deployed within cloud platforms and as such adapt elastically to usage intensity.

Accessibility, particularly in relation to privacy and openness is important in the collection and use of VGI (Mooney et al. 2017). Privacy is an important part of the broader issue of ensuring compliance with the GDPR. In some instances, privacy may be waived by contributors at the outset of a project when giving free, informed and willing consent for their contributions to be linked to their identity. Additionally, a key aspect of accessibility is the degree to which both data and any associated code and/or applications is openly available for access for both contributors and wider public access. Complete openness may not always be possible. Privacy concerns will limit some access, and open availability of code developed may sometimes be difficult due to licensing issues. However, good practice should be to reduce such complications wherever possible by avoiding where possible the use of licensed code and/or data.

4. QA checks and suggested good practices

The experiences and lessons learned from undertaking the QA checks on VGI acquired during LandSense are briefly summarized below. The QA checks focused on photographic quality and privacy, polygon overlap, positional accuracy and offset, contributor agreement and categorical accuracy. Table 1 shows the various LandSense QA checks undertaken and the number of checks performed across the various LandSense themes and pilot studies.

4.1. Photograph quality checks

The checks aim to ensure that the photographs used and stored are of appropriate quality for the application in-hand. The checks on blur and

Table 1. The data quality analyses applied to different LandSense pilots. The numbers in each cell indicate the number of checks carried out for that QA service and pilot. Gray cells indicate that that particular QA service was not applicable to the pilot study.

QA check	Urban landscape dynamics				Agricultural land use	Forest and habitat monitoring
	OSMlanduse validation	City. Oases	MijnPark.NL	Paysages	CropSupport	Natura.Alert
Photograph brightness & blur	-	197	372	260	210	512
Photograph privacy	-	197	372	260	210	512
Polygon topology	-	-	-	-	202	-
Positional accuracy	-	878	377	-	207	-
Positional offset	-	443	361	-	211	-
Categorical accuracy	10,806	-	-	467	-	571
Contributor agreement	150 sites/points	-	30 sites/points	650 sites/points	-	-

brightness aimed mainly to filter out low-quality images. In relation to privacy, a key concern was to detect features that could identify an individual, such as faces or vehicle license plates, so that they could be masked out.

General good practice for photograph blur and brightness would require there to be a quality scale with a predefined threshold value (or values) against which the check is performed. If the result falls outside the set of acceptable values for an image, then the check should return a failure result for it. In addition to recording the pass/failure state, it was found that the quantified result from the quality check (e.g. brightness level and blur level) should also be recorded. Retaining the result enables users of the QA system to rerun the quality checks as a post process with a range of differing thresholds to examine the impacts of variations to the quality procedures and act to meet the particular needs of the study in-hand. This is important as users may differ greatly in needs and hence a photograph acceptable to one may be unsuitable for another.

4.1.1. Blur check

The blur check QA analysis was based on an assessment of the degree of blurriness obtained through the application of a Laplace filter to the image (Pech-Pacheco et al. 2000). Specifically, blurriness was determined through convolving the image with a 3×3 Laplace kernel and calculating the variance on the resulting image. The LandSense QA platform implementation of blur checking calculated an image's blurriness on an eight-bit radiometric scale producing a result in the 0–255 range (where 0 indicates fully blurred and 255 is no blur present).

The photographs acquired in the LandSense pilot studies were generally of a high quality. Less than 3% of the 1549 images processed by the LandSense QA service (Table 1) had blur levels below 251 and only these appeared problematic. A blur threshold of 250 was used in LandSense. Examples of photographs over a range of blur levels are shown in Figure 1 and a summary of the degree of blurring from each pilot study is provided in Table 2. Note that a photograph may pass this check but still be of limited value because of other concerns such as brightness.



Figure 1. Examples of 4 photographic images acquired during the LandSense project and the calculated blur level. (a) paysages pilot, blur level 162, (b) paysages pilot, blur level 216, (c) natura pilot blur level 248 and (d) MijnPark.NI pilot blur level 249. Note these all fail to meet the blur threshold used in LandSense but could still be useful to other studies.

Table 2. Summary of the photograph blur and brightness check results for the pilot studies. Numbers indicate the actual number of photographs (and as a % of the relevant total).

Statistic	All Pilots	MijnPark.NL	Paysages	CropSupport	Natura Alert	City.Oases
Number of blurred images (%)	41 (2.6%)	5 (1.3%)	11 (4.2%)	0 (0%)	0 (0%)	25 (12.7%)
Average blur level	249.2	251.7	252.1	254.0	253.9	223
Number of dark images (%)	431 (27.7%)	187 (50.3%)	84 (32.3%)	7 (3.3%)	72 (14.1%)	81 (41%)
Average brightness	111.4	99.3	108.5	122.7	123.0	95.5

In addition to identifying good practice for ascertaining whether an image is blurred, it is also important to consider what options are available once a blurred image has been detected. For the purposes of LandSense, and given the low number of blurred images that were found, it was determined either to exclude blurred images from the pilot study datasets or to include them with metadata highlighting their quality issues. Another option that could be considered is to apply a sharpening algorithm to reduce the blurriness of an image. To illustrate the potential of this approach, examples of blurred images were sharpened manually using Photoshop (Figure 2). As shown, minor levels of blurring can be addressed using standard sharpening functions in image processing software but significant blurring, especially when combined with an already dark image, was difficult to correct.

4.1.2. Brightness check

For the brightness QA check, the relative luminance of contributed photographs was calculated and expressed on an 8-bit scale. In the latter, the pixel DN values ranged from 0 (completely dark) to 255 (completely white). The choice of an appropriate threshold to identify unsuitable images is highly dependent on the

data collection context, and the specific image quality requirements of the project but from experience with the photographs acquired a threshold of 100 was selected for general applications in LandSense. However, a threshold may need to be adapted for particular needs and hence the desirability of storing the brightness check score for each photograph along with the photograph. For example, in terms of context, if images are likely to be taken in poorly lit environments (e.g. at night or indoors) then it may be wise to lower the quality threshold for brightness to increase the likelihood of image submissions passing the check. Conversely, if the project requirements specify that only high-quality image submissions are acceptable then the threshold may need to be increased. Although brightness checking may seem rather basic, a large proportion of photographs acquired were dark. Indeed, the proportion of photographs failing the brightness check varied greatly in the pilots with up to half of the acquired images acquired in a pilot study being viewed as too dark (Table 2). Figure 3 illustrates the quality of images that fail to reach the threshold of 100 used by LandSense. It is evident that images at the lower end of the spectrum convey little visual information.



(a)



(b)

Figure 2. Examples of images before (left) and after (right) correction. (a) paygages pilot, successful image sharpening to correct for blur and (b) paygages pilot unsuccessful correction for blur and brightness.



Figure 3. Examples of photographic images over a range of brightness levels below the threshold used in LandSense. Brightness level of images shown (from left to right) are 2,8,36,46,56,68,81 and 94.

Unlike the blur check function, where extreme image sharpness does not normally cause quality concerns, there is potential for reduced quality at each end of the brightness scale. Both extremely low and high levels of brightness can lead to images where detail is lost. This is evident in [Figure 4](#) which shows a spectrum of photographs acquired from the LandSense pilot studies that passed the basic brightness check (i.e. had a score of >100). The majority of these images are of good quality, but it is evident that the brightest image was of relatively low quality. Consequently, for good practice, it may be necessary to have both a minimum and maximum threshold value set for brightness checks in future work. Across the set of pilot studies, only 3 of 1549 photographs assessed had a value >200 and this value was used as the upper brightness threshold in LandSense

Photographs that fail to pass the brightness check could be excluded, marked as being of low quality or automatically brightened. The latter option may seem an obvious choice, but care should be taken when modifying images collected by users. Depending on the nature of the intellectual property rights for a VGI campaign, it may not be possible to make modifications to user collected images without infringing their rights. It may also be ethically questionable to make changes to data collected by users. If the need to modify user-collected imagery to improve quality is a desired goal of a citizen science campaign,

permission for image modification should be included in the data collection agreement with users as part of good practice.

Even where permission is given to modify user-collected data, automatic brightening of images can lead to unforeseen quality issues. For example, parts of photograph may become clearer but in others interpretability may decline due to issues such as over-saturation. It is important that adjustment of photograph properties be undertaken in a way that is tailored to the specific needs of a study.

Brightening images may also be useful in improving the accuracy of privacy checks, discussed below. It may be expected that dark images are more likely to generate errors in privacy checks as the privacy features are difficult to detect (e.g. faces in a dark part of a photograph). However, such features may become detectable after brightening. Thus, good practice would be to undertake face detection after brightness correction.

4.1.3. Privacy checks on photographs

The photographs acquired as part of the LandSense project, as with other citizen science observatories, vary greatly in part because of the nature of the pilot studies. For example, the majority of CropSupport imagery was of agricultural land and fields of growing crops, whereas images acquired in the City.Oases pilot were primarily of urban scenes. The likelihood of privacy concerns arising in the latter is much greater



Figure 4. Examples of photographic images acquired that had high brightness levels. Brightness level of images shown (from left to right) are 110, 139, 151, 171, and 185.

than in the former. However, any privacy concerns such as faces and/or license plates included in photographs even if in the background and located some distance from the camera may need masking.

The definition of success and failure in the context of privacy checks is also not as simple and discrete as it may initially appear. For example, it may seem straightforward to define failure as being the presence of any undetected privacy feature after the privacy check has been performed. However, if the feature was not clearly visible in the original image, due to its distance from the camera or being partially obscured, then some judgment may be required to determine whether the feature needed to be removed to maintain the subject's anonymity. A similar, but less critical, issue also applies to errors of commission (i.e. photographs where a privacy feature was detected but either did not exist or was not clearly visible in the original image). For example, text on the side of a vehicle could be erroneously detected as a license plate that requires masking. While commission errors indicate a failure of the privacy check, some level of commission error is likely to be acceptable in order to minimize the likelihood of more serious omission errors where a privacy feature that should be masked out exists but was not detected. However, excessive levels of commission error could substantially reduce the quantity of the photographs and may potentially remove information from photographs. This was found to be an issue with the face detection service in the CropSupport pilot study and is further described below.

It is unlikely that any current automated privacy check can consistently achieve 100% accuracy (Liao, Jain, and Stan 2014). Therefore, if it is critical (e.g. for GDPR requirements) that all privacy features are identified then some further processing, such as manual intervention, may be required. For citizen observatories like LandSense, one option could be to use citizen contributors to provide manual feedback on the presence of privacy features in an image as a post-processing step in any QA privacy check procedure. A potential model for such manual interventions was explored during the LandSense project. Using IIASA's Picture Pile software (Danylo et al. 2018), photographs acquired in the City.Oases pilot study (>1700 images) were manually filtered to exclude all images that did not require privacy checks. The same process could be employed post-privacy check to ensure that all privacy features have been successfully identified and anonymized.

Another option explored by the City.Oases pilot was for contributors to identify manually in the pilot app any privacy features during the image upload process. In the upload page, the users were asked to blur out parts of the images which showed a face or vehicle license plate before being uploaded. When the users click the edit

button on the upload page, they get to the edit page on which they can select a photograph. When the user has selected an image, s/he can blur out parts of the image by touching the points on the image that should be blurred. It should be noted that this manual privacy check, as with the automated detection service, is unlikely to be 100% accurate. QA checks on the City.Oases image library found some of the images where manual blurring had failed to identify all privacy features.

4.1.3.1. Face detection service (FDS). The FDS model was created using TensorFlow and Google's TensorFlow Object Detection API. The model used was mobilenet SSD (single-shot multibox detector) included in the TensorFlow API. The performance of the FDS in terms of detection accuracy was relatively good for clearly visible faces but much poorer when faces were obscured, poorly lit or were in the background of the image.

Overall performance for the FDS seems to match the level of accuracy quoted in the literature (e.g. Yang et al. 2016) for the types of unconstrained imagery collected by LandSense with a detection accuracy of around 70%. Almost half of the 49 detection failures observed could be categorized as borderline cases where it is arguable whether any faces in the image were clearly recognizable and needed to be blurred.

Performance across the different pilot studies varied with the urban-based pilots showing the highest level of omission errors, mainly due to the increased presence of people in the images collected (Figure 5). Commission errors were more common in the agricultural and habitat-monitoring pilots, with the CropSupport pilot in particular showing a high degree of commission errors (Figure 6). One potential method to improve overall FDS performance would be to use differentiated detection thresholds based on the probability and prevalence of faces in the photographs collected.

Retraining the FDS using the images collected by the LandSense pilots was an option investigated, but it was determined that it may not be expected to substantially increase accuracy since it is already at a level matching the state of the art in face detection for unconstrained photographs. Additionally, as with

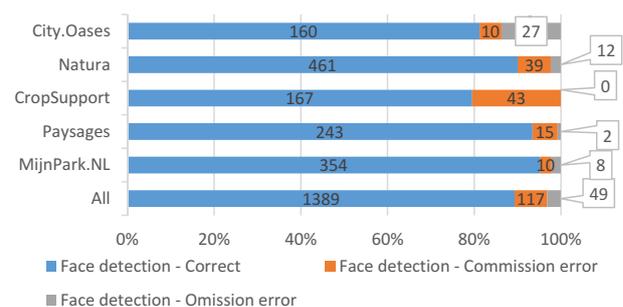


Figure 5. Accuracy results of face detection service for all pilots.

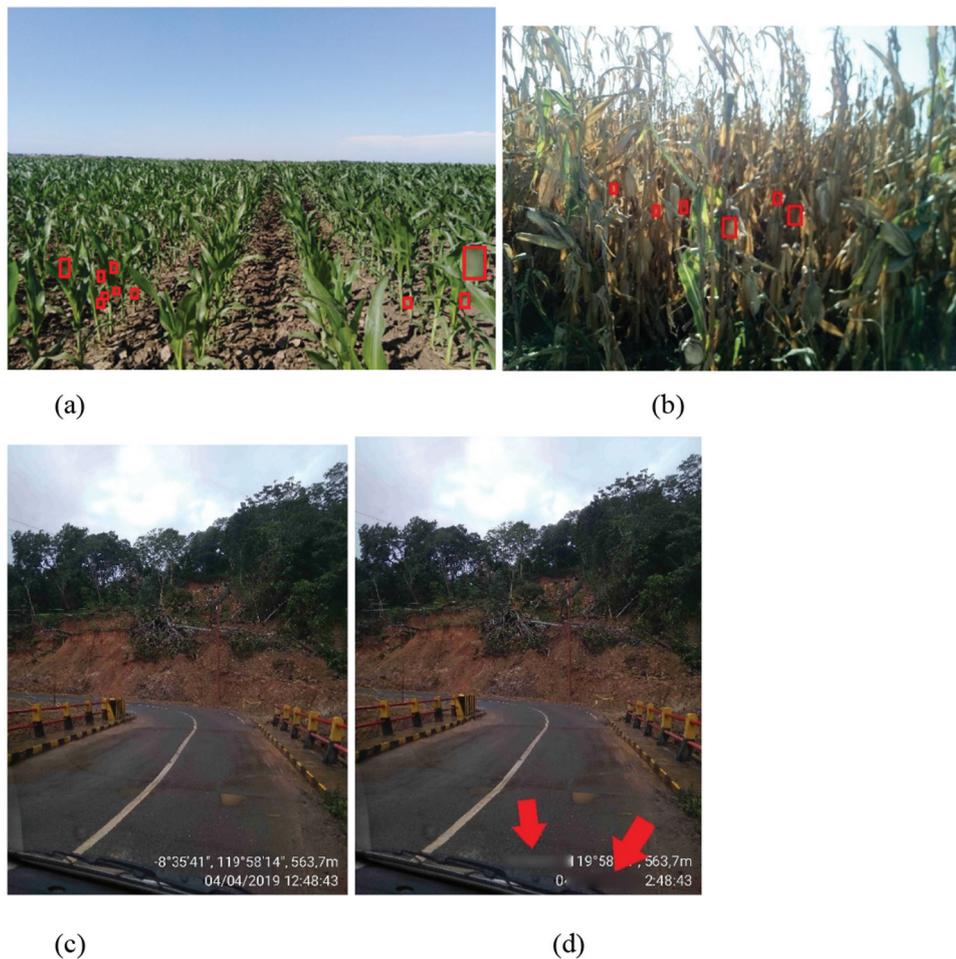


Figure 6. Examples of commission errors. (a) and (b) show commission errors for faces detected in crops, with red boxed areas supposedly containing a face, and (c) and (d) show a photograph before and after application of the licence plate detection software; note the metadata text in the bottom right of (c) is incorrectly masked out in (d).

other analysis based on machine learning algorithms, retraining must be carefully performed to ensure that the retrained model shows genuine improvement. There is also a danger that an inductively trained tool may work successfully for the site it was constructed on but generalize poorly to other sites.

As indicated above, no automated privacy check is likely to be perfect and some form of manual checking may need to be integrated into an FDS. Manually pre-filtering the images using IIASA's Picture Pile software greatly reduced the number of images for processing and hence reduced processing time considerably. This type of filtering process could be carried out by contributors after the initial data collection phase.

If 100% accuracy is needed, then some form of post-process manual verification check appears to be necessary. This could be performed by the QA team, pilot data manager or by the contributors themselves. If GDPR regulations are the primary motivation for privacy checks, care must be taken to ensure the regulations are not breached as part of the manual verification process itself. Any images that potentially

could be in breach of GDPR should not be located on a publicly accessible server until the QA FDS process is fully complete.

4.1.3.2. License plate detection. LandSense's implementation of the License Plate Detection Service (LPDS) used the Open Source version of the OpenALPR algorithm as described by Masood et al. (2017). Its overall performance for the photographs collected appears on initial examination to be very high (97%, Figure 7) but the accuracy is inflated due to imbalance in class size, with a very low number of license plates visible in the images. If its accuracy is based solely on the images with visible license plates, then it only recognizes only one license plate out of the 38 in total. As with the FDS, many of the license plate failures are borderline cases where the vehicle's license plate is not clearly visible and it is debatable whether any LPDS would have been able to detect it accurately. To examine this, images from the City.Oases pilot, which had the largest number of license plates present, were evaluated using other detection services.

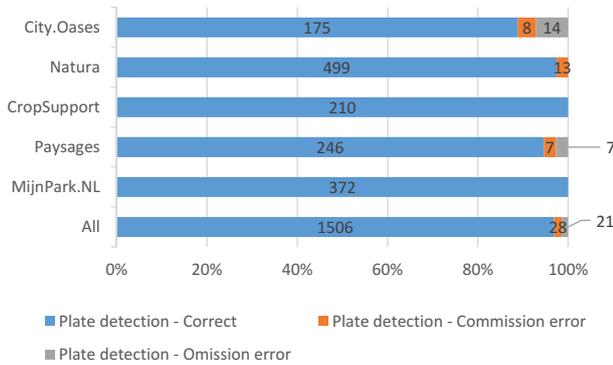


Figure 7. Accuracy results of license plate detection service for all pilots.

The City.Oases pilot had 14 images with license plates present that were not detected by the LandSense LPDS. These images were reprocessed using the commercial version of OpenALPR, which offers improved performance over the open source version used by LandSense. Even using this version, only license plates in 5 of the 14 images were detected.

Another commercial LPDS, PlateRecogniser (<https://platerecognizer.com>) was also tested. The latter performed identically to the commercial version of OpenALPR. Photographs for which PlateRecogniser detected license plates not identified by LandSense LPDS or the commercial version of OpenALPR may still include errors. For example, Figure 8 notes the presence of a commission error related to the construction signage contained within the image.

The performance of the LandSense LPDS used could be improved through further training. However, it is also clear that even with training, some license plates may not be identifiable due to the unconstrained nature of the images collected. The low number of images with license plates in the LandSense dataset may not be sufficient for retraining. Suitable images would need to be sourced for retraining the algorithm. In addition, it is debatable whether the cost of retraining the LPDS would be a worthwhile use of

time and resources given the low prevalence of license plates in the data collected and there would be no guarantee that any improvement would be substantial given the performance of the commercial LPDS.

4.2. QA checks for polygon overlap

For the CropSupport pilot, QA checks applicable to polygonal data were developed to check for overlapping polygons. Checking for overlapping polygons is a relatively simple spatial problem, based on standard GIS intersect functions (Burrough et al. 2015). We implemented a post-processing approach, the responsibility for correcting polygon overlap errors passed from the citizen contributing data to the data manager for the pilot. This process may be difficult without local knowledge or other ancillary data (e.g. detailed high resolution remotely sensed imagery). It also added additional manual QA processes into the workflow; however, it was deemed appropriate to safeguard quality of data.

With the LandSense QA service, each instance of polygon overlap counted individually. That is, if two polygons were overlapping this counted as two overlapping polygons. However, with other GIS tools, the same example may be reported as a single instance of overlapping. The difference between the number of overlaps and the number of overlapping polygons can lead to confusion when comparing results from different tools. Good practice requires that documentation, therefore, is clear, and transparent to ensure meaning is apparent.

Although amending overlapping polygons was not included in the LandSense QA tool, there are a range of approaches that could be used. Establishing good practice guidelines requires the consideration of the various options within the context of the specific citizen science project being undertaken. For example, the CropSupport pilot focused on adding polygons representing crop boundaries, therefore polygons do

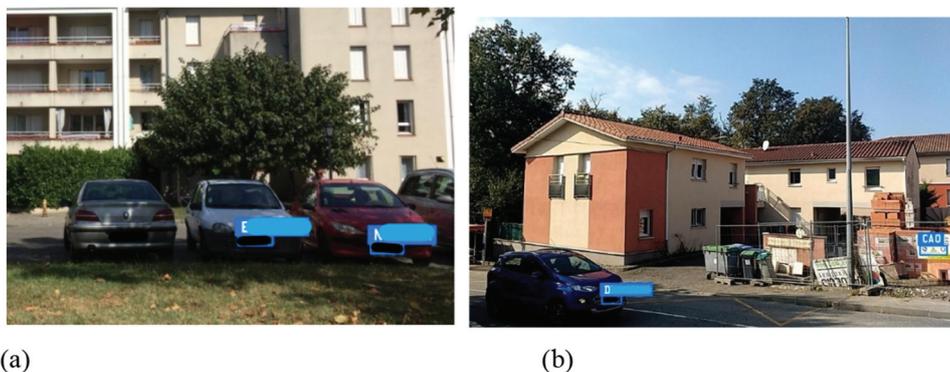


Figure 8. Examples of outputs from the licence plate detector. (a) output from PlateRecogniser LPDS which detected one more license plate (shown in blue) than the OpenALPR and (b) output from PlateRecogniser which shows detection of a licence plate that was identified by LandSense LPDS or the commercial version of OpenALPR. Note the presence of a commission error related to the construction signage on the right hand side of the image.

not need to be contiguous, but they should not overlap. Conversely, a project using VGI to map land use may need to ensure that polygons are contiguous in addition to not overlapping. Other models might allow polygons to overlap but could allow a polygon to be within another polygon to enable hierarchical structuring of land use categories. In each case, the process of identifying overlaps does not alter but guidelines for dealing with any overlaps identified may vary according to the specific use case. For LandSense, overlapping polygons were not allowed and in the CropSupport pilot study users would be required to redraw their contributed polygon if it overlapped an existing polygon. A concern with this approach is that it assumes that the original polygon is correct. Other means to address the problem of overlapping polygons could be adopted. For example, the option to redraw a polygon or the boundaries of the existing polygon it overlaps with could be provided. A potential problem with this option is that it may allow a contributor to edit another user's data. There would be no way to guarantee that any edits made to polygons were correct and the original contributor's data could be lost unless previous versions stored. To address this concern, it could be possible to flag overlapping polygons for review. Users could also be asked to add comments on the overlap that could assist in solving the problem. The overlapping polygons could then be reviewed by the original contributors, or other users, and a consensus could be reached as to the correct polygon boundaries. This option would require the tools available to users to include snapping and topological functionalities. Finally, automatic polygon correction methods such as existing in some GIS packages could be adopted. However, caution should be taken when employing automated solutions, as these may simply redraw the polygon boundaries to remove any intersecting areas between them and there is no guarantee that the new polygons are an accurate reflection of reality.

4.3. QA checks for positional accuracy and offset

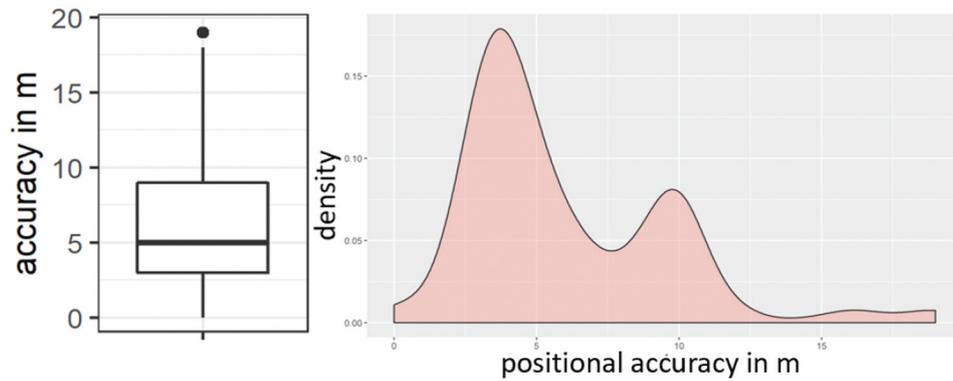
Positional accuracy is an important issue in the collection of data on land cover and land use (Congalton, and Green 2009). Many analyses of data, for instance, assume high positional accuracy and hence it is important to establish the level of accuracy required and that achieved. Positional accuracy is, for example, important in site-specific accuracy assessment discussed in Section 4.4 below.

The measurement of positional accuracy and offset are relatively straightforward and are widely discussed in the literature. In terms of good practice, the

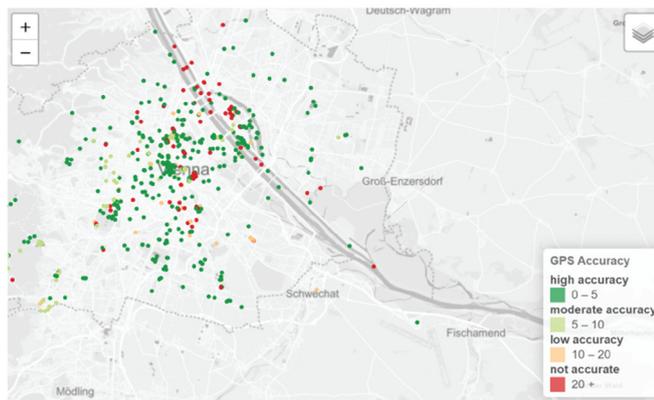
definition of accuracy and offset requirements is the key issue to be discussed here. In line with the other QA services discussed, the context of the data collection exercise being undertaken is a critical element in the definition of positional accuracy requirements. The scale of the land use feature being surveyed will substantially impact on the level of positional accuracy required. The density of observation points within a given area will also affect the level of positional accuracy required. To illustrate this, where areas being observed are large, open land use features, such as fields of crops for the CropSupport pilot, the degree of positional accuracy for the observation point often does not need to be high to be of sufficient quality. In contrast, where the features being observed are small and densely packed, as in urban green spaces, such as those observed in the MijnPark.NL pilot, a higher level of positional accuracy may be required to ensure that the contributor has observed the correct feature. As a guide to the achieved positional accuracy, Figure 9 shows a summary of results for the City.Oases pilot which shows the variability in accuracy in terms of magnitude but also in space.

The local environment will influence the level of positional accuracy that can be achieved. In urban areas, the presence of tall buildings and other obstructions to a clear view of the sky will reduce positional accuracy. Observations taken in more open, rural or suburban locations, such as the CropSupport pilot, will not suffer from similar GPS signal quality issues. In addition to climatic and locational factors, the technical specifications of the mobile device used to make the observations could affect the degree of positional accuracy obtainable. For many guided campaigns, where both positional accuracy of the observation point and its positional offset from the reference point are important, the project's design may need careful attention. The combination of errors of positional accuracy and offset can lead to large disparities in spatial accuracy of observations. For this reason, guided campaigns, in particular, should integrate the likely spatial accuracy of the mobile devices used to collect data in their design. A guided campaign should also consider the likelihood of obstructions to GPS signals and their potential impact on positional accuracy when choosing the location of observation points.

One potential method for reducing spatial accuracy issues in a guided campaign might be to consider the use of technological solutions to reduce or eliminate the positional offset problem. For example, QR codes could be installed at the observation points and users could be required to scan the code to indicate that they are at the correct location. This option would require



(a)



(b)

Figure 9. Example of positional errors for the city.Oases pilot. (a) statistical summary of the positional accuracy results and (b) a map showing the spatial distribution of errors of differing magnitude (base map provided by OpenStreetmap, © OpenStreetmap contributors).

the presence of some physical infrastructure to which the QR code can be attached (e.g. post, bench, tree *etc.*). Another technological solution could be the installation of Bluetooth beacons at the observation points that could then log the presence of a Bluetooth enabled phone at the location and inform the user via the pilot app that a contributor was in the correct location.

When considering an appropriate level of positional offset for guided campaigns, it may be considered good practice to allow variable offset thresholds. For the LandSense guided campaigns, MijPark.NL and City.Oases, a fixed value of 20 m for positional offset was used as both pilots were collecting data in similar urban environments. However, it would be possible to vary the offset based on the observability of each observation point (i.e. positional offset threshold could be relaxed for large features and tightened for small features or ones that need to be viewed from a specific geographical location). As elsewhere, the good practice advice is that settings such as offset thresholds be selected for the requirements of a particular application.

4.4. QA checks on label quality

The QA checks on label quality focused on two commonly encountered issues with VGI: contributor agreement and thematic accuracy. The QA checks for contributor agreement sought to determine the extent to which two or more citizens participating to a project agree in the labeling of cases. The core focus in LandSense applications was typically either upon agreement on labeling to conventional LULC classes (e.g. OSMLanduse and Paysages) or to classes that indicate the perception of places (e.g. City.Oases and MijPark.NL). If a reference set representing reality was available, the focus was typically on the accuracy of the labels relative to the reference.

In the course of the LandSense project, some challenges and issues were encountered and two are highlighted here. First, there were instances in which only one contributor obtained data. In such circumstances, agreement between contributors cannot be assessed. It may, however, be possible to evaluate other aspects of the quality of the data. For example, it may be possible to determine how the contributed data fits within its geographical context (Goodchild, and Li 2012) or to

compare the labels against some other source of data. Second, in many instances, a large number of contributors may participate. For example, in LandSense between 2 and 40, contributors provided data for Mijnpark.NL and from 3 to 6 for contributors for Paysages. Sometimes the number of participants may be so large that it actually acts to degrade a study (Foody et al. 2015). In such circumstances, a range of possible approaches can be used to seek to maximize the value of the data. For example, attention could focus on cases for which all or a majority of contributors agree. Information on contributor performance may also be used to weight contributed data to ultimately increase classification accuracy (Foody et al. 2018). Thus, having multiple contributors labeling can be useful and although the required number of contributors is likely to be application dependent it would be good practice to strive for such a community of contributors in future work. However, in doing so the composition of the crowd needs attention (Comber et al. 2016). Within LandSense, for example, contributors with different personal profiles performed differently (Olteanu-Raimond et al. 2020) and this may have implications to future studies, especially in relation to issues such as the selection and training of potential contributors.

In special circumstances, it can be necessary to deviate from good practices, and if this is the case, open and honest reporting of the accuracy assessment is required to aid interpretation (Stehman, and Foody 2019). For example, it may sometimes not be possible for a citizen to visit a randomly selected site to acquire VGI. In such circumstances, the problem should be noted and reported as it may aid interpretation. Other approaches to indicate accuracy, such as those based on social status and geographical context may also be used to help support an accuracy assessment (Goodchild, and Li 2012).

5. Conclusions

This article summarizes some of the key experiences and lessons learnt in the LandSense project in relation to QA of VGI. Universal guidelines cannot be offered as the needs of studies may vary greatly but some general conclusions may be drawn. For example, it was stressed that a compromise should be reached between maintaining data protocols and schemas rigidly (to simplify and streamline QA processes) whilst still allowing flexibility for each specific pilot study to identify a data model that fits their subject area and requirements as well as allow projects to evolve. Some common data elements (e.g. GPS accuracy) and formats (e.g. geojson) may be necessary to

enable effective QA processes to be performed, but these need to be defined in a consensual manner to encourage buy-in from those involved in each pilot study.

Examples of good practice identified throughout the project were outlined for each QA check of relevance to LandSense. The checks covered common issues such as photographic image quality and privacy checks, polygon overlap, positional accuracy and offset, and labeling quality in assessments of contributor agreement and categorical accuracy. As visual interpretation of images, both from ground-based photography and remote sensing, are central to much of VGI, the basic issues of image quality are of fundamental importance. Being able to satisfactorily interpret an image and know the location it relates to are critical to the successful use of VGI as ground data. While citizen science has much to offer a key experience gained, however, was that there are important challenges in the use of citizen-based VGI. If designing a crowdsourcing project, for example, do not assume that citizens will provide photographs of appropriate illumination and blur. We also do not assume that current image analysis software can fully complete necessary privacy checks. Factoring in some human interpretation into the data processing system may be wise.

Photographs were acquired mainly as a source of information on land use and land cover but before this could be extracted they were subjected to three QA checks: blur, brightness and privacy. Checks on image blur and brightness were based on the use of a numerical quality scales. Threshold values on these scales may then be used to identify the photographs of suitable quality and those which could be excluded or subjected to enhancement operations. From the experience gained in LandSense, it is suggested that the quantitative score for each photograph generated in the QA checking should be retained with it so that future, possibly unplanned, work could easily use different quality thresholds if appropriate for the task. The effect of different thresholds on the relative size of commission and omission errors should also be considered within the context of a study. For example, with privacy checks for features such as faces, different thresholds may be required in urban and rural areas. Finally, in some analyses, such as in face detection as part of a privacy check, it was highlighted that current automated approaches may be inadequate alone. Such tasks may require inclusion of additional, perhaps manual, checking. The accuracy of privacy checks was less than might be expected from the literature on topics such as face detection from constrained photographs. The experience gained from LandSense

was that some form of manual check was required if all privacy features were to be detected and masked out. Experience also showed that if images were to be enhanced this needs some care. For example, a basic brightness correction may enhance the interpretability of a photograph but this can cause problems. First, the change may make privacy features more apparent and hence it is recommended that privacy checks are undertaken after other QA checks. Second, permission may be required from the contributor to alter the image and hence if such changes are anticipated permission for such activity should be acquired when gaining consent from a contributor to participate in a project.

The positional accuracy of contributed data was important, and requirements may vary depending on the features under study. Thus, thresholds for positional accuracy may vary between urban and rural sites and depending on the nature of the features under study. The ability to use physical infrastructure to enhance confidence in geospatial data was encouraged if appropriate. The quality of the devices likely to be used by contributors also needs consideration, especially as newer more expensive devices often provide high positional accuracy.

Assessments of labeling quality were encountered when seeking to evaluate contributor agreement or thematic accuracy. Such assessments were based on comparisons of multiple labels applied to a set of geographical features (e.g. crop type by a set of contributors). How the multiple labels are used requires careful consideration. For example, a basic consensus approach could be used to generate a label or only those cases labeled the same by a majority or indeed all labelers could be used. Differences between groups of labelers were also observed in LandSense and hence the personal profile of contributors may need careful consideration. In LandSense, an additional observation was that it is possible to weight contributed data to obtain an enhanced ensemble label. A problem sometimes observed was that a site may get labeled by only one contributor. In such a case, the standard approach to evaluating labeling quality cannot be used and a project should be adaptable and consider other QA approaches.

Adaptability was also required in the assessment of polygon overlap. The approach adopted evolved such that the check was moved from a near real-time task for the contributor to a post-processing operation to be undertaken by the project staff. This change can be important, especially if detailed local knowledge is desired as the project staff, unlike some contributors, may not have it. In LandSense, it also required the incorporation of a manual assessment into the process. Finally, it was stressed that many seemingly basic

issues need to be carefully considered. For example, terminology must be clear. Thus, thematic classes used need clear and careful definition as does the spatial unit used and deviation from common assumed conditions explained (e.g. there needs to be a way to deal with problematic cases such as mixed units in a classification). Clarity was also important in relation to comparison of results between methods. For example, the LandSense polygon overlap QA check would report an overlap between two polygons as a single case but other systems and researchers might count this as two cases.

It is hoped that these lessons from LandSense help other projects. Note that the outcomes are explicitly proposed as good practice and should not be treated as the only or “best” solution. The specific requirements and context of a VGI exercise will be a key driver in the applicability and utility of the various good practices shown. Many of the practices discussed represent a compromise between maximizing data quality whilst maintaining important features such as contributor engagement. While technological advancements may be expected to enhance aspects such as automated privacy checks, the experiences gained in LandSense may help others develop citizen science projects and help fulfil the potential of VGI. LandSense provided experiences and learnings of value to other projects for the effective acquisition and use VGI. Critically, our experiences urge caution against assuming that basic issues such as photograph quality can be downplayed or that automated methods for common applications will be sufficient. Including basic checks (e.g. image brightness and blur) and planning for some manual input to QA of VGI is important until relevant technological advancements occur.

Acknowledgments

We are grateful for the input of all LandSense collaborators especially those working on the demonstration projects and key colleagues notably Julian Rosser, Sam Meek, Inian Moorthy, Linda See and Steffen Fritz. We also thank the three referees and the editors for their comments on the article. The work reported in this article was undertaken while MS was at Heidelberg University.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This research was funded by the European Commission’s Horizon 2020 program as part of the LandSense project [grant number 689812]; Horizon 2020 [LandSense, 689812]

Notes on contributors

Giles Foody is currently Professor of Geographical Information Science at the University of Nottingham. He completed BSc and PhD degrees in geography at the University of Sheffield in 1983 and 1986, respectively. His main research interests focus on remote sensing for environmental applications at scales ranging from sub-pixel to global.

Gavin Long is a research fellow at the University of Nottingham. A computer science graduate, he received his PhD from the University's School of Engineering in 2010. His research interests cover a broad range of disciplines with a particular focus on applied research in the fields of data science, computer simulation and machine learning. He is currently a member of Nottingham University Business School's N/Lab investigating the use of 'big data' and machine learning techniques to augment social policy and predict demographics.

Michael Schultz is currently pursuing his habilitation at the University of Tübingen on the topic of wealth and poverty prediction using remote sensing. He completed his BSc and MSc degrees in geography and geoinformatics at the University of Jena in 2007 and 2010, respectively. He was awarded a PhD title for time series analysis by the University of Wageningen. His activities range from operational application of remote sensing and creation of novel data products through the facilitation of AI and data fusion across sciences.

Ana-Maria Olteanu-Raimond is a senior researcher in GIS and co-director of LASTIG laboratory. Her research interests include the integration of heterogeneous spatial data, imperfect information fusion, data matching, and collaborative mechanisms and qualification of VGI for joint use with authoritative spatial data. She is also conducting research on citizen science: data collection, platforms, tasks, motivation and sustainability.

ORCID

Giles Foody  <http://orcid.org/0000-0001-6464-3054>
 Gavin Long  <http://orcid.org/0000-0003-4947-5712>
 Michael Schultz  <http://orcid.org/0000-0002-4813-090X>
 Ana-Maria Olteanu-Raimond  <http://orcid.org/0000-0002-1101-1333>

Data availability

Key LandSense materials are freely available. Key data sets are available at <https://doi.org/10.1080/10095020.2022.2100285> and tools at <https://doi.org/10.1080/10095020.2022.2100285>. The authors would be pleased to consider reasonable requests to share other resources.

Ethical approval

Ethical approval for the work reported was gained from the Ethics Committee of the School of Geography, University of Nottingham on 12 March 2020. The file name for this approval is Giles Foody 063v2. The data used were acquired

by LandSense partners who gained informed consent to ensure GDPR compliance.

References

- Ahlqvist, O. 2008. "Extending Post-Classification Change Detection Using Semantic Similarity Metrics to Overcome Class Heterogeneity: A Study of 1992 and 2001 US National Land Cover Database Changes." *Remote Sensing of Environment* 112 (3): 1226–1241.
- Antoniou, V., J. Morley, and M. Haklay. 2010. "Web 2.0 Geotagged Photos: Assessing the Spatial Dimension of the Phenomenon." *Geomatica* 64 (1): 99–110.
- Bastin, L., S. Schade, and C. Schill. 2017. "Data and Metadata Management for Better VGI Reusability." In *Mapping and the Citizen Sensor*, edited by G. Foody, L. See, S. Fritz, P. Mooney, C. C. Fonte, A. M. Olteanu-Raimond, and V. Antoniou, 249–272. London: Ubiquity Press.
- Brabyn, L., and D. M. Mark. 2011. "Using Viewsheds, GIS, and a Landscape Classification to Tag Landscape Photographs." *Applied Geography* 31 (3): 1115–1122. doi:10.1016/j.apgeog.2011.03.003.
- Burrough, P. A., R. McDonnell, R. A. McDonnell, and C. D. Lloyd. 2015. *Principles of Geographical Information Systems*. Oxford: Oxford University Press.
- Capineri, C., M. Haklay, H. Huang, V. Antoniou, J. Kettunen, F. Ostermann, and R. Purves. 2016. *European Handbook of Crowdsourced Geographic Information*. London: Ubiquity Press.
- Card, D. H. 1982. "Using Known Map Category Marginal Frequencies to Improve Estimates of Thematic Map Accuracy." *Photogrammetric Engineering and Remote Sensing* 48: 431–439.
- Chesnokova, O., and R. S. Purves. 2018. "From Image Descriptions to Perceived Sounds and Sources in Landscape: Analyzing Aural Experience Through Text." *Applied Geography* 93: 103–111. doi:10.1016/j.apgeog.2018.02.014.
- Comber, A., P. Fisher, and R. Wadsworth. 2005. "What is Land Cover?" *Environment and Planning B, Planning & Design* 32 (2): 199–209. doi:10.1068/b31135.
- Comber, A. J., R. A. Wadsworth, and P. F. Fisher. 2008. "Using Semantics to Clarify the Conceptual Confusion Between Land Cover and Land Use: The Example of 'Forest'." *Journal of Land Use Science* 3 (2–3): 185–198.
- Comber, A., L. See, and S. Fritz. 2014. "The Impact of Contributor Confidence, Expertise and Distance on the Crowdsourced Land Cover Data Quality." *GI Forum 2014-Geospatial Innovation for Society* 309–313. doi:10.1553/giscience2014s309.
- Comber, A., P. Mooney, R. S. Purves, D. Rocchini, and A. Walz. 2016. "Crowdsourcing: It Matters Who the Crowd Are. The Impacts of Between Group Variations in Recording Land Cover." *PLoS One* 11 (7): e0158329.
- Congalton, R. G., and K. Green. 2009. *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices*. 2nd ed. Boca Raton: CRC Press.
- Congalton, R. G., R. G. Oderwald, and R. A. Mead. 2009. "Assessing Landsat Classification Accuracy Using Discrete Multivariate Analysis Statistical Techniques." *Photogrammetric Engineering and Remote Sensing* 49 (12): 1671–1678.

- Costa, H., G. M. Foody, and D. S. Boyd. 2017. "Using Mixed Objects in the Training of Object-Based Image Classifications." *Remote Sensing of Environment* 190: 188–197. doi:10.1016/j.rse.2016.12.017.
- Costa, H., G. M. Foody, and D. S. Boyd. 2018. "Supervised Methods of Image Segmentation Accuracy Assessment in Land Cover Mapping." *Remote Sensing of Environment* 205: 338–351. doi:10.1016/j.rse.2017.11.024.
- Danylo, O., I. Moorthy, T. Sturn, L. See, J. C. Laso Bayas, D. Domian, D. Fraisl, C. Giovando, et al. 2018. "The Picture Pile Tool for Rapid Image Assessment: A Demonstration Using Hurricane Matthew." *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences IV-4*: 27–32. doi:10.5194/isprs-annals-IV-4-27-2018.
- De Marchi, B., A. Ficorilli, and A. Biggeri. 2022. "Research is in the Air in Valle Del Serchio." *Futures* 137: 102906. doi:10.1016/j.futures.2022.102906.
- Demetriou, D., M. Campagna, I. Racetin, and M. Konecny. 2017. "Integrating Spatial Data Infrastructures (SDIs) with Volunteered Geographic Information (VGI) Creating a Global GIS Platform." In *Mapping and the Citizen Sensor*, edited by G. Foody, L. See, S. Fritz, P. Mooney, C. C. Fonte, A. M. Olteanu-Raimond, and V. Antoniou, 273–297. London: Ubiquity Press.
- Ehrlinger, J., K. Johnson, M. Banner, D. Dunning, and J. Kruger. 2008. "Why the Unskilled are Unaware: Further Explorations of (Absent) Self-Insight Among the Incompetent." *Organizational Behavior and Human Decision Processes* 105 (1): 98–121. doi:10.1016/j.obhdp.2007.05.002.
- Elia, A., S. Balbo, and P. Boccardo. 2018. "A Quality Comparison Between Professional and Crowdsourced Data in Emergency Mapping for Potential Cooperation of the Services." *European Journal of Remote Sensing* 51 (1): 572–586. doi:10.1080/22797254.2018.1460567.
- Elwood, S., M. F. Goodchild, and D. Z. Sui. 2012. "Researching Volunteered Geographic Information: Spatial Data, Geographic Research, and New Social Practice." *Annals of the Association of American Geographers* 102: 571–590.
- Feick, R., and C. Robertson. 2015. "A Multi-Scale Approach to Exploring Urban Places in Geotagged Photographs." *Computers, Environment and Urban Systems* 53: 96–109. doi:10.1016/j.compenvurbysys.2013.11.006.
- Fielding, A. H., and J. F. Bell. 1997. "A Review of Methods for the Assessment of Prediction Errors in Conservation Presence/absence Models." *Environmental Conservation* 24: 38–49.
- Finn, J. T. 1993. "Use of the Average Mutual Information Index in Evaluating Classification Error and Consistency." *International Journal of Geographical Information Science* 7 (4): 349–366.
- Flanagin, A., and M. Metzger. 2008. "The Credibility of Volunteered Geographic Information." *GeoJournal* 72 (3–4): 137–148.
- Fleiss, J. L. 1971. "Measuring Nominal Scale Agreement Among Many Raters." *Psychological Bulletin* 76: 378–382.
- Fogliaroni, P., F. D'-Antonio, and E. Clementini. 2018. "Data Trustworthiness and User Reputation as Indicators of VGI Quality." *Geo-Spatial Information Science* 21 (3): 213–233. doi:10.1080/10095020.2018.1496556.
- Fonte, C. C., L. Bastin, L. See, G. Foody, and F. Lupia. 2015a. "Usability of VGI for Validation of Land Cover Maps." *International Journal of Geographical Information Science* 29 (7): 1269–1291. doi:10.1080/13658816.2015.1018266.
- Fonte, C. C., L. Bastin, L. See, G. Foody, and J. Estima. 2015b. "Good Practice Guidelines for Assessing VGI Data Quality." In *The 18th AGILE International Conference on Geographic Information Science*. Lund, Sweden.
- 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: 783–792.
- Foody, G. M., L. See, S. Fritz, M. Van der Velde, C. Perger, C. Schill, D. S. Boyd, and A. Comber. 2015. "Accurate Attribute Mapping from Volunteered Geographic Information: Issues of Volunteer Quantity and Quality." *The Cartographic Journal* 52 (4): 336–344. doi:10.1080/00087041.2015.1108658.
- Foody, G., L. See, S. Fritz, P. Mooney, C. Fonte, A. M. Olteanu-Raimond, and V. Antoniou, edited by. 2017a. *Mapping and the Citizen Sensor* 398pp. London: Ubiquity Press. doi:10.5334/bbf.
- Foody, G., S. Fritz, C. C. Fonte, L. Bastin, A. M. Olteanu-Raimond, P. Mooney, L. See, et al. 2017b. In *Mapping and the Citizen Sensor*, In edited by G. Foody, L. See, S. Fritz, P. Mooney, C. C. Fonte, A. M. Olteanu-Raimond, and V. Antoniou, 1–12. London: Ubiquity Press.
- Foody, G., L. See, S. Fritz, I. Moorthy, C. Perger, C. Schill, and D. Boyd. 2018. "Increasing the Accuracy of Crowdsourced Information on Land Cover via a Voting Procedure Weighted by Information Inferred from the Contributed Data." *ISPRS International Journal of Geo-Information* 7 (3): 80. doi:10.3390/ijgi7030080.
- Foody, G. M. 2020. "Explaining the Unsuitability of the Kappa Coefficient in the Assessment and Comparison of the Accuracy of Thematic Maps Obtained by Image Classification." *Remote Sensing of Environment* 239: 111630. doi:10.1016/j.rse.2019.111630.
- Freeman, A. 2017. *Essential Docker for ASP.NET Core MVC*. Berkeley, CA: Apress.
- Garnett, R., and R. Stewart. 2015. "Comparison of GPS Units and Mobile Apple GPS Capabilities in an Urban Landscape." *Cartography and Geographic Information Science* 42: 1–8.
- Goodchild, M. F. 2007. "Citizens as Sensors: The World of Volunteered Geography." *GeoJournal* 69 (4): 211–221. doi:10.1007/s10708-007-9111-y.
- Goodchild, M. F., and L. Li. 2012. "Assuring the Quality of Volunteered Geographic Information." *Spatial Statistics* 1: 110–120. doi:10.1016/j.spasta.2012.03.002.
- Gopal, S., and C. Woodcock. 1994. "Theory and Methods for Accuracy Assessment of Thematic Maps Using Fuzzy Sets." *Photogrammetric Engineering and Remote Sensing* 60: 181–188.
- Griesbaum, L., S. Marx, and B. Höfle. 2017. "Direct Local Building Inundation Depth Determination in 3-D Point Clouds Generated from User-Generated Flood Images." *Natural Hazards and Earth System Sciences* 17 (7): 1191–1201. doi:10.5194/nhess-17-1191-2017.
- Haklay, M., S. Basiouka, V. Antoniou, and A. Ather. 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. doi:10.1179/000870410X12911304958827.
- Havlik, D., J. Soriano, C. Granell, S. E. Middleton, H. van der Schaaf, A. Berre, and J. Pielorz. 2013. “Future Internet Enablers for VGI Applications.” *Proceedings of the 27th Conference on Environmental Informatics (EnviroInfo 2013)*, Aachen, Hamburg, Germany: Shaker Verlag, pp. 622–630.
- Hay, A. M. 1979. “Sampling Designs to Test Land-Use Map Accuracy.” *Photogrammetric Engineering and Remote Sensing* 45: 529–533.
- Hickling Arthurs Low Corporation. 2012. *Volunteered Geographic Information (VGI) Primer*, Canadian Geospatial Data Infrastructure Product 21e. Retrieved from: <https://geoscan.nrcan.gc.ca/starweb/geoscan/servelet.starweb?path=geoscan/download.web&search1=R=291948>
- Higgins, C.I., J. Williams, D. G. Leibovici, I. Simonis, M. J. Davis, C. Muldoon, P. van Genuchten, G. O’Hare, and S. Wiemann. 2016. “Citizen Observatory WEB (COBWEB): A Generic Infrastructure Platform to Facilitate the Collection of Citizen Science Data for Environmental Monitoring.” *International Journal of Spatial Data Infrastructures Research* 11: 20–48.
- Iwao, K., K. Nishida, T. Kinoshita, and Y. Yamagata. 2006. “Validating Land Cover Maps with Degree Confluence Project Information.” *Geophysical Research Letters* 33 (23): 123404. doi:10.1029/2006GL027768.
- Ke, Y., X. Tang, and F. Jing 2006. “The Design of High-Level Features for Photo Quality Assessment.” *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’06)*, New York, NY, USA, 2006, 419–426. doi:10.1109/CVPR.2006.303.
- Knight, J. F., and R. S. Lunetta. 2003. “An Experimental Assessment of Minimum Mapping Unit Size.” *IEEE Transactions on Geoscience and Remote Sensing* 41 (9): 2132–2134. doi:10.1109/TGRS.2003.816587.
- Laso Bayas, J. C., M. Lesiv, F. Waldner, A. Schucknecht, M. Duerauer, L. See, S. Fritz, D. Fraisl, et al. 2017. “A Global Reference Database of Crowdsourced Cropland Data Collected Using the Geo-Wiki Platform.” *Scientific Data* 4 :e170136. doi:10.1038/sdata.2017.136.
- Liao, S., A. Jain, and L. Stan. 2014. “A Fast and Accurate Unconstrained Face Detector.” *IEEE Transactions on Pattern Analysis and Machine Intelligence* 38: 211–223.
- Lo, S. W., J. H. Wu, F. P. Lin, and C. H. Hsu. 2015. “Visual Sensing for Urban Flood Monitoring.” *Sensors* 15: 20006–20029.
- Masood, S. Z., G. Shu, A. Dehghan, and E. G. Ortiz. 2017. “License Plate Detection and Recognition Using Deeply Learned Convolutional Neural Networks.” arXiv preprint arXiv:1703.07330.
- Menard, T., J. Miller, M. Mowak, and D. Norris 2011. “Comparing the GPS Capabilities of the Samsung Galaxy S, Motorola Droid X, and the Apple iPhone for Vehicle Tracking Using FreeSim Mobile.” *14th International IEEE Conference of Intelligent Transportation Systems*, Washington, DC.
- Merkel, D. 2014. “Docker: Lightweight Linux Containers for Consistent Development and Deployment.” *Linux Journal* 239: 2.
- Merry, K., and P. Bettinger. 2019. “Smartphone GPS Accuracy Study in an Urban Environment.” *PLoS One* 14 (7): e0219890.
- Meyer, H., and E. Pebesma. 2022. “Machine Learning-Based Global Maps of Ecological Variables and the Challenge of Assessing Them.” *Nature Communications* 13 (1): 1–4. doi:10.1038/s41467-022-29838-9.
- Minghini, M., V. Antoniou, C. C. Fonte, J. Estima, A. M. Olteanu-Raimond, L. See, and M. Laakso, et al. 2017. “The Relevance of Protocols for VGI Collection.” In *Mapping and the Citizen Sensor*, edited by G. Foody, L. See, S. Fritz, P. Mooney, C. C. Fonte, A. M. Olteanu-Raimond, and V. Antoniou, 223–247. London: Ubiquity Press.
- Mobasheri, A., A. Zipf, and L. Francis. 2018. “OpenStreetmap Data Quality Enrichment Through Awareness Raising and Collective Action Tools—experiences from a European Project.” *Geo-Spatial Information Science* 21 (3): 234–246. doi:10.1080/10095020.2018.1493817.
- Mocnik, F. B., A. Mobasheri, L. Griesbaum, M. Eckle, C. Jacobs, and C. Klöner. 2018. “A Grounding-Based Ontology of Data Quality Measures.” *Journal of Spatial Information Science* 2018 (16): 1–25.
- Mooney, P., M. Minghini, M. Laakso, V. Antoniou, A. M. Olteanu-Raimond, and A. Skopeliti. 2016. “Towards a Protocol for the Collection of VGI Vector Data.” *ISPRS International Journal of Geo-Information* 5 (11): 217.
- Mooney, P., A. M. Olteanu-Raimond, G. Touya, N. Juul, S. Alvanides, and N. Kerle. 2017. “Considerations of Privacy, Ethics and Legal Issues in Volunteered Geographic Information.” In *Mapping and the Citizen Sensor*, edited by G. Foody, L. See, S. Fritz, P. Mooney, C. C. Fonte, A. M. Olteanu-Raimond, and V. Antoniou, 119–135. London: Ubiquity Press.
- Moorthy, I., S. Fritz, L. See, and I. McCallum. 2017. “LandSense: A Citizen Observatory and Innovation Marketplace for Land Use and Land Cover Monitoring.” In *EGU General Assembly Conference Abstracts*, (pp. 8562). Vienna, Austria.
- Moorthy, I. 2020. “LandSense: A Citizen Observatory and Innovation Marketplace for Land Use and Land Cover Monitoring.” *18th European Week of Regions and Cities*, Zenodo. doi:10.5281/zenodo.4146821
- Nada, H., V. A. Sindagi, H. Zhang, and V. M. Patel. 2018. “Pushing the Limits of Unconstrained Face Detection: A Challenge Dataset and Baseline Results.” In *2018 IEEE 9th International Conference on Biometrics Theory, Applications and Systems (BTAS)*, Redondo Beach, CA, USA: IEEE.
- Obe, R., and L. Hsu. 2011. “PostGis in Action.” *GEOInformatics* 14 (8): 30.
- Olofsson, P., G. M. Foody, S. V. Stehman, and C. E. Woodcock. 2013. “Making Better Use of Accuracy Data in Land Change Studies: Estimating Accuracy and Area and Quantifying Uncertainty Using Stratified Estimation.” *Remote Sensing of Environment* 129: 122–131. doi:10.1016/j.rse.2012.10.031.
- Olofsson, P., G. M. Foody, M. Herold, S. V. Stehman, C. E. Woodcock, and M. A. Wulder. 2014. “Good Practices for Estimating Area and Assessing Accuracy of Land Change.” *Remote Sensing of Environment* 148: 42–57. doi:10.1016/j.rse.2014.02.015.
- Olteanu Raimond, A. M., G. Hart, G. M. Foody, G. Touya, T. Kellenberger, and D. Demetriou. 2017a. “The Scale of VGI in Map Production: A Perspective on European National Mapping Agencies.” *Transactions in GIS* 21 (1): 74–90.

- Olteanu-Raimond, A. M., M. Laakso, V. Antoniou, C. C. Fonte, A. Fonseca, M. Grus, J. Harding, T. Kellenberger, M. Minghini, and A. Skopeliti. 2017b. "VGI in National Mapping Agencies: Experiences and Recommendations." In *Mapping and the Citizen Sensor*, edited by G. Foody, L. See, S. Fritz, P. Mooney, C. C. Fonte, A. M. Olteanu-Raimond, and V. Antoniou, 299–326. London: Ubiquity Press.
- Olteanu-Raimond, A. M., L. See, M. Schultz, G. Foody, M. Riffler, T. Gasber, L. Jolivet, et al. 2020. "Use of Automated Change Detection and VGI Sources for Identifying and Validating Urban Land Use Change." *Remote Sensing* 12 (7): 1186.
- Ostermann, F. O., and C. Granell. 2017. "Advancing Science with VGI: Reproducibility and Replicability of Recent Studies Using VGI." *Transactions in GIS* 21 (2): 224–237. doi:10.1111/tgis.12195.
- Pech-Pacheco, J., L. G. Cristobal, J. Chamorro-Martinez, and J. Fernandez-Valdivia. 2000. "Diatom Autofocusing in Brightfield Microscopy: A Comparative Study." *Proceedings 15th International Conference on Pattern Recognition* Barcelona, Spain. ICPR-2000, 2000, pp. 314–317.
- Ploton, P., F. Mortier, M. Réjou-Méchain, N. Barbier, N. Picard, V. Rossi, C. Dormann, et al. 2020. "Spatial Validation Reveals Poor Predictive Performance of Large-Scale Ecological Mapping Models." *Nature Communications* 11 (1): 1–11. doi:10.1038/s41467-020-18321-y.
- Pontius, R. G. 2000. "Quantification Error versus Location Error in Comparison of Categorical Maps." *Photogrammetric Engineering and Remote Sensing* 66: 1011–1016.
- Pontius, R. G., Jr, and M. Millones. 2011. "Death to Kappa: Birth of Quantity Disagreement and Allocation Disagreement for Accuracy Assessment." *International Journal of Remote Sensing* 32 (15): 4407–4429. doi:10.1080/01431161.2011.552923.
- Salk, C. F., T. Sturn, L. See, S. Fritz, and C. Perger. 2016. "Assessing Quality of Volunteer Crowdsourcing Contributions: Lessons from the Cropland Capture Game." *International Journal of Digital Earth* 9 (4): 410–426. doi:10.1080/17538947.2015.1039609.
- Saura, S. 2002. "Effects of Minimum Mapping Unit on Land Cover Data Spatial Configuration and Composition." *International Journal of Remote Sensing* 23 (22): 4853–4880. doi:10.1080/01431160110114493.
- Scepan, J. 1999. "Thematic Validation of High-Resolution Global Land-Cover Data Sets." *Photogrammetric Engineering and Remote Sensing* 65: 1051–1060.
- See, L., P. Mooney, G. Foody, L. Bastin, A. Comber, J. Estima, S. Fritz, et al. 2016. "Crowdsourcing, Citizen Science or Volunteered Geographic Information? the Current State of Crowdsourced Geographic Information." *ISPRS International Journal of Geo-Information* 5 (5): 55.
- See, L., and J. C. L. Bayas, M. Lesiv, D. Schepaschenko, O. Danylo, I. McCallum, M. Dürauer, I. Georgieva, D. Domian, D. Fraisl, and G. Hager. 2022. "Lessons Learned in Developing Reference Data Sets with the Contribution of Citizens: The Geo-Wiki Experience." *Environmental Research Letters* 17 (6): 065003.
- Senaratne, H., A. Mobasher, A. L. Ali, C. Capineri, and M. Haklay. 2017. "A Review of Volunteered Geographic Information Quality Assessment Methods." *International Journal of Geographical Information Science* 31 (1): 139–167. doi:10.1080/13658816.2016.1189556.
- Severinsen, J., M. de Roiste, F. Reitsma, and E. Hartato. 2019. "Vgtrust: Measuring Trust for Volunteered Geographic Information." *International Journal of Geographical Information Science* 33 (8): 1683–1701. doi:10.1080/13658816.2019.1572893.
- Shima, Y., Y. Nakashima, and M. Yasuda. 2017. "Detecting Orientation of In-Plain Rotated Face Images Based on Category Classification by Deep Learning." In *TENCON 2017–2017 IEEE Region 10 Conference* Penang, Malaysia, IEEE, 127–132.
- Silva, S. M., and C. R. Jung. 2018. "License Plate Detection and Recognition in Unconstrained Scenarios." In *Proceedings of the European Conference on Computer Vision (ECCV)*, Munich, Germany, 580–596.
- Stehman, S. V., and R. L. Czaplewski. 1998. "Design and Analysis for Thematic Map Accuracy Assessment: Fundamental Principles." *Remote Sensing of Environment* 64 (3): 331–344. doi:10.1016/S0034-4257(98)00010-8.
- Stehman, S. V. 1999. "Basic Probability Sampling Designs for Thematic Map Accuracy Assessment." *International Journal of Remote Sensing* 20 (12): 2423–2441. doi:10.1080/014311699212100.
- Stehman, S. V. 2009. "Sampling Designs for Accuracy Assessment of Land Cover." *International Journal of Remote Sensing* 30: 5243–5272.
- Stehman, S. V., C. C. Fonte, G. M. Foody, and L. See. 2018. "Using Volunteered Geographic Information (VGI) in Design-Based Statistical Inference for Area Estimation and Accuracy Assessment of Land Cover." *Remote Sensing of Environment* 212: 47–59. doi:10.1016/j.rse.2018.04.014.
- Stehman, S. V., and G. M. Foody. 2019. "Key Issues in Rigorous Accuracy Assessment of Land Cover Products." *Remote Sensing of Environment* 231: 111199. doi:10.1016/j.rse.2019.05.018.
- Szeliski, R. 2010. *Computer Vision: Algorithms and Applications*. London: Springer Science & Business Media.
- Vahidi, H., B. Klinkenberg, and W. Yan. 2018. "Trust as a Proxy Indicator for Intrinsic Quality of Volunteered Geographic Information in Biodiversity Monitoring Programs." *GIScience & Remote Sensing* 55 (4): 502–538. doi:10.1080/15481603.2017.1413794.
- Wadembere, I., and P. Ogao. 2010. "Geometry Updating for Geospatial Data Integration." *International Archives of Photogrammetry and Remote Sensing* 38: 52–57.
- Wadoux, A. M. C., G. B. M. Heuvelink, S. de Bruin, and D. J. Brus. 2021. "Spatial Cross-Validation is Not the Right Way to Evaluate Map Accuracy." *Ecological Modelling* 457: 109692.
- Waldner, F., A. Schucknecht, M. Lesiv, J. Gallego, L. See, A. Pérez-Hoyos, R. D'Andrimont, T. de Maet, et al. 2019. "Conflation of Expert and Crowd Reference Data to Validate Global Binary Thematic Maps." *Remote Sensing of Environment* 221: 235–246. doi:10.1016/j.rse.2018.10.039.
- Wannemacher, K., B. Birli, T. Sturn, R. Stiles, I. Moorthy, L. See, and S. Fritz. 2018. "Using Citizen Science to Help Monitor Urban Landscape Changes and Drive Improvements." *Journal for Geographical Information Science-Gi_forum* 1: 336–343.
- Wickham, J. D., S. V. Stehman, L. Gass, J. Dewitz, J. A. Fry, and T. G. Wade. 2013. "Accuracy Assessment of NLCD 2006 Land Cover and Impervious Surface." *Remote Sensing of Environment* 130: 294–304. doi:10.1016/j.rse.2012.12.001.

- Wolf, L., T. Hassner, and I. Maoz. 2011 June. "Face Recognition in Unconstrained Videos with Matched Background Similarity." In *CVPR 2011*, Colorado Springs, CO, USA: IEEE, 529–534.
- Wu, F., B. Wu, M. Zhang, H. Zeng, and F. Tian. 2021. "Identification of Crop Type in Crowdsourced Road View Photos with Deep Convolutional Neural Network." *Sensors* 21 (4): 1165. doi:10.3390/s21041165.
- Wulder, M. A., J. C. White, S. Magnussen, and S. McDonald. 2007. "Validation of a Large Area Land Cover Product Using Purpose-Acquired Airborne Video." *Remote Sensing of Environment* 106: 480–491.
- Xing, D., S. V. Stehman, G. M. Foody, and B. W. Pengra. 2021. "Comparison of Simple Averaging and Latent Class Modeling to Estimate the Area of Land Cover in the Presence of Reference Data Variability." *Land* 10 (1): 35. doi:10.3390/land10010035.
- Xu, G., X. Zhu, D. Fu, J. Dong, and X. Xiao. 2017. "Automatic Land Cover Classification of Geo-Tagged Field Photos by Deep Learning." *Environmental Modelling and Software* 91: 127–134. doi:10.1016/j.envsoft.2017.02.004.
- Yang, S., P. Luo, C. C. Loy, and X. Tang. 2016. "Wider Face: A Face Detection Benchmark." In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (5525–5533). Las Vegas, NV, USA.
- Zandbergen, P. A. 2009. "Accuracy of iPhone Locations: A Comparison of Assisted GPS, WiFi and Cellular Positioning." *Transactions in GIS* 13 (S1): 5–25. doi:10.1111/j.1467-9671.2009.01152.x.
- Zielstra, D., and H. H. Hochmair. 2013. "Positional Accuracy Analysis of Flickr and Panoramio Images for Selected World Regions." *Journal of Spatial Science* 58 (2): 251–273. doi:10.1080/14498596.2013.801331.