Comparing Expert and Non-expert Conceptualisations of the Land: An Analysis of Crowdsourced Land Cover Data

Alexis Comber[1], Chris Brunsdon[2], Linda See[3], Steffen Fritz[3], and Ian McCallum[3]

[1] Department of Geography, University of Leicester, Leicester, LE1 7RH, UK
[2] Department of Geography, University of Liverpool, Liverpool, L69 3BX UK
[3] International Institute of Applied Systems Analysis (IIASA), A-2361, Laxenburg, Austria
ajc36@le.ac.uk, Christopher.Brunsdon@liverpool.ac.uk,
{see,fritz,mccallum}@iiasa.ac.at

Abstract. This research compares expert and non-expert conceptualisations of land cover data collected through a Google Earth web-based interface. In so doing it seeks to determine the impacts of varying landscape conceptualisations held by different groups of VGI contributors on decisions that may be made using crowdsourced data, in this case to select the best global land cover dataset in each location. Whilst much other work has considered the quality of VGI, as yet little research has considered the impact of varying semantics and conceptualisations on the use of VGI in formal scientific analyses. This study found that conceptualisation of cropland varies between experts and non-experts. A number of areas for further research are outlined.

Keywords: Volunteered Geographical Information (VGI), Land Cover, Geo-Wiki, Geographically Weighted Kernel.

1 Introduction

The increase in web-based technologies has resulted in many new forms of data creating and sharing. Individual citizens generate, upload and share a wide range of different types of data via online databases, much of which increasingly has a spatial reference. There is the potential of what is referred to as volunteered geographic information (VGI) [1] to change the nature of scientific investigation as one of its critical advantages of VGI is the potential increase in the volumes of data describing spatially referenced phenomena. Such citizen science activities are supported by the rise of digital, location enabled technologies which offer increased opportunities for the capture of spatial data and for citizens to share information about all kinds of processes and features that they observe in their daily lives.

One of the critical issues to be overcome for crowdsourced data to be included in scientific investigations relates to the quality of the information. In much scientific research data are collected under a formal experimental design which frequently includes consideration of sampling frameworks, quality assurance checks etc. Data are collected using well-established methods, by particular instruments or by people with appropriate training and expertise. Thus one of the critical issues in using crowdsourced data relates to the nature of crowd, their familiarity with the domain under investigation and consideration of any impacts of a lack of expertise on the quality of the data that are collected.

This paper generates measures of correspondence between crowdsourced data about land cover / land use and different global land cover datasets. The correspondences between locations identified as being ‘cropland’ by volunteers was statistically related to measures of the proportions of cropland at those locations from the global datasets. The largest correspondences were used to infer which global dataset best predicted the cropland identified by the crowd. The variation in these correspondences and inferences were examined by considering data contributed by remote sensing experts and by non-experts – ie with contrasting degrees of domain familiarity. In this way the paper explores the impacts of the differences between naïve and expert contributors of VGI and the conceptualisations of landscape features that they hold. This study uses data on land cover, collected as part of the Geo-Wiki project (www.geo-wiki.org), where the registration process captured self-reported measures of contributor expertise. The crowdsourced land cover data captured in this way has a number of potential applications and could, for example, be used to train or to validate statistical classification of remote sensed data. The results of this analysis will inform future uses of VGI by determining whether the differences in expertise about the domain under consideration (in this case remote sensing of cropland land cover) are important and should be considered in such future work.

2 Background

There are many examples of crowdsourced data being exploited that have resulted in novel scientific discoveries such as unravelling protein structures [2], discovering new galaxies [3], reporting of illegal deforestation [4]. The use of VGI is now commonplace in many areas of scientific investigation, from conservation [5] to urban planning [6], and VGI has been found to be particularly useful in endeavours to manage and understand important emerging problems such as ash dieback [7] and post-disaster damage assessments [8, 9]. The latter exemplify the critical advantage of crowdsourced data: the ability to rapidly collect and share large volumes of data describing many kinds of phenomena. These activities are facilitated by ubiquitous ability to capture and share data using many electronic devices (e.g. digital cameras, smartphones, tablets, etc) that are location-enabled, eg through in-built GPS capabilities. The result is an increasing amount of spatially referenced or geo-located data, captured through ubiquitous and low costs citizen owned sensors, that can directly and instantaneously capture and share data of the immediate environment, that are available for formal scientific analysis. This presents a number of challenges associated with use of VGI in formal analyses that relate to questions over data quality and reliability:

- Data are not collected as part of a controlled experiment [10] which may include experimental design, sampling frameworks, validation and error assessments, etc.
- There is no control over what or how information is recorded [11].
- The nature and ability of volunteers can vary greatly in their expertise in subject matter being recorded and may even be malicious [12].

For these reasons consideration of VGI or crowdsourced data quality has attracted increasing attention [11, 13-15] because the usefulness of crowdsourced data for incorporation into scientific analyses depends on its reliability and credibility. In comparison to the designed experiment there is no control over the information that is recorded [14, 16-18] and conflicting information make difficult for analysts to use the information [19] with confidence. One of the frequently employed approaches to overcome a lack of data quality reporting is to use information from a large number of contributors and a number of crowdsourcing projects show a positive relationship between the reliability of contributed data and the number of participants [20], supporting Linus' law [14]. However, other work has shown that providing more data from more contributors may be unhelpful if the volunteers are all similarly confused [21], or if the feature or process being described is associated with alternative conceptualisations, in which case additional contributions may dilute the usefulness of the data. For example, Pal and Foody [22] note that the curse of dimensionality or the Hughes effect [23] is commonly encountered in remote sensing, where the accuracy of a mapping project may decline as the volume of data increases. In order to bridge between these extremes, a number of studies have been undertaken that have sought to develop measures of contributor reliability. These include approaches based on rating data [24], quality assurances based on inputs from trusted individuals acting as gatekeepers [13], random coefficient modeling and bootstrapping approaches to address irregularities in phenology data [10] and the use of control data, where the features under investigation was known, to allow measures of contributor reliability to be generated [11]. More recent research has sought to characterise the reliability of individual volunteers for example through the application of latency measures [25, 26]. These provide an intrinsic measure of contributor quality derived entirely from the data itself and without any additional information such as reference data.

The approaches described above all focus on the *quality* of volunteer and the data they contribute. As yet little research has considered how the background and characteristics of the contributor, citizen or volunteer influences their conceptualisation of landscape features and the impacts of vary conceptualisation on the data this is contributed. It is well know that many landscape features are conceptually uncertain, are conceptually vague or indeterininistic in nature, and that different individuals will conceptualise and describe the landscape in different ways depending on their background. Vague interpretations of objects and their boundaries have been considered extensively through the concepts of fiat and bone fide boundaries, corresponding to fiat and bone fide geographic objects [27-29]. More recent research has sought to formalize frameworks for describing and managing alternative conceptualisations of landscape features. These include the use of similarity measures to describe semantics and meanings associated with geographic information retrieval [30] and ontology frameworks [31] for understanding and grounding constructed or differently conceived objects. The latter include mechanisms for integrating different landscape conceptualisations and for describing the associated uncertainty. This research extends consideration of uncertainties associated with varying conceptualisations of the landscape as recorded in crowdsourced geographical information and the impacts on decision making.

3 Case Study: Conceptualisations of Land Cover

Land cover and land cover change have been found to be important variables in understanding land-atmosphere interactions and particularly the impacts of climate changes [32]. A number of different global datasets describe land cover but with considerable disagreement between them in the amount and distribution of different types of land cover features particularly in relation to forest and cropland. This is a long-standing problem, one that has been recognised since the emergence of different global datasets in the early 1990s: they describe significantly different amounts of land cover, for example with differences as great as 20% in the amount of land classified as arable or cropland [33]. The potential errors and uncertainties associated with these products mean that their input into applications such climate models is questionable. They certainly cannot be used to model land cover change. In other research the VGI on land cover as collected through the Geo-Wiki has been used to identify and to validate land cover [34] and to determine which global datasets best correspond to volunteered land cover class labels in different locations – an illustrative example is given in [11] – using spatially weighted kernels [35, 36].

The VGI on land cover can also be used to explore the varying conceptualisations of the land held by different groups or subsets of volunteers who contributed data to the Geo-Wiki project. A recent submission [37] provides a full analysis of the quality of the expert and non-expert data collected through the Human Impact Geo-Wiki campaign. Using control data points, where the actual land cover was known, they analysed the different degrees to which experts and non-expert volunteers correctly identified the land cover and the degree of human impact using a standard statistical data analysis. They found little difference between experts and non-experts in identifying human impact and noted that experts were better than non-experts at identifying the land cover class [37].

The study presented here extends this analysis. It uses VGI on cropland land cover to infer the best global land cover datasets at each location in two study areas. The suggested extension associated with this research, is that it seeks to explore the relative impacts of different levels of expertise rather than absolute ones which would require the use of control data (or ground truth) as in the previous work cited above. The aims of this study were:

- To determine whether the analysis of data contributed by expert and non-expert groups results in different optimal global datasets being selected in different locations.

- To examine the degree to which expert and non-expert groups of contributors hold different conceptualisations of the landscape.
- To explore whether this matters.

These were explored by comparing the data contributed by experts and non-experts and through analysis of cropland land cover and land use relating to agricultural activities as collected by the Geo-Wiki and as recorded in global land cover datasets.

4 Methods

VGI on land cover were collected through the Geo-Wiki project which incorporates a web-based interface that uses Google Earth [38]. Volunteers are invited to record the land cover at different locations. A screen grab is shown in Figure 1 by way of example. The Geo-Wiki project is a directed form of VGI creation in that volunteers are steered towards a structured reporting / scoring environment, they score the land cover in randomly selected global locations and their scores are stored in a databases. This is contrast to other activities such as OpenStreetMap, which provides a direct experiential link between the volunteer and the data being recorded such that a typical Open Street Map volunteer knows the place that is being updated.

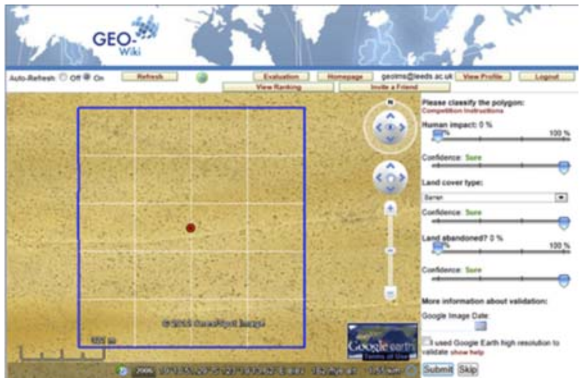

Fig. 1. An illustration of the Geo-Wiki interface

A number of Geo-Wiki campaigns have been initiated to gather different types of information related to specific objectives including agriculture (agriculture.geo-wiki.org), urban areas (urban.geo-wiki.org) and biomass (biomass.geo-wiki.org). The data used in this analysis were collected in the autumn of 2011 in a campaign to collect measures of Human Impact (http://humanimpact.geo-wiki.org). This was developed as a crowdsourcing competition with the aim of collecting sufficient data to validate a spatial dataset of land availability for biofuel production, which was released without any measures of data quality or error and was perceived to misrepresent land availability and thus would implicitly support land grabs in the Global South. As part of their registration, volunteers were asked to provide some background information. Critical for this analysis was, this information included questions about the volunteer's experience, profession and expertise.

Each volunteer was asked to complete an online tutorial in order to demonstrate the process and to provide basic training with explanations of the issues and concepts associated with land cover and land use mapping. The contributors completed as much or as little of the training that they wished to complete. As part of the competition, the top ten participants were offered co-authorship on a paper [39] and Amazon vouchers as incentives [38]. Volunteers were provided with a series of random locations and asked to record what they observed at each location and based on their interpretation of the landscape they assigned each location to one of 10 predefined land cover classes, including a cropland class of *Cultivated and managed*.

These data are considered here for two study areas: South America and Africa. Volunteered information on land cover was captured globally and 17,371 and 6888 data points were recorded in Africa and South America respectively as shown in Figure 2 and Figure 3. Of the data points in Africa, 4562 were recorded as Cultivated and managed and 1390 were similarly recorded in South America. At each location the proportions of cropland land cover as recorded in 5 global land cover datasets were extracted. The global land cover datasets were GlobCover, GLC, MODIS, GeoCover and Hansen and, although other datasets exist (e.g JRC), these datasets were selected to illustrate the analysis this variation in VGI contributed by different groups.

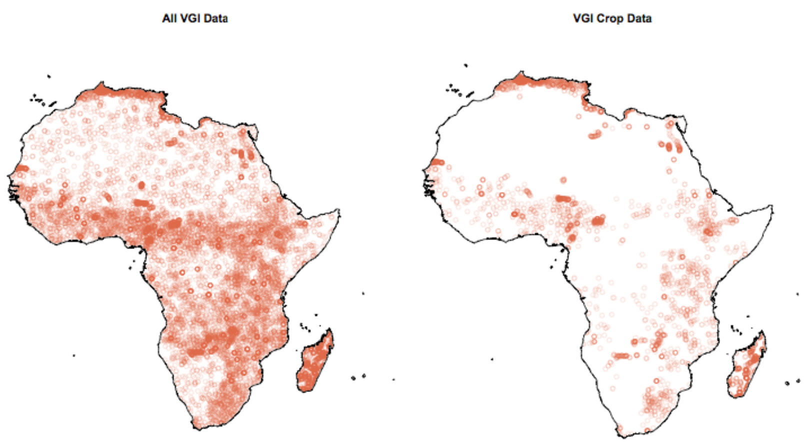

Fig. 2. VGI data in Africa and the *Cultivated and managed* class used in this study

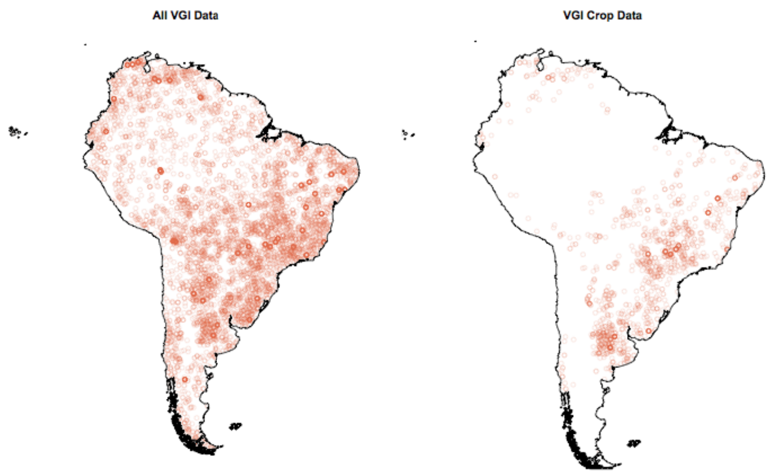

Fig. 3. VGI data in South America and the *Cultivated and managed* class used in this study

Measures of spatially distributed correspondence between VGI cropland and cropland data from the global datasets were generated using the methods described in detail elsewhere [11, 35, 36]. In brief, these approaches use a geographically weighted kernel to generate local correspondence measures calculated at discrete locations on grid covering the study area. This allows local measures of the spatial variations in relationships between the data inputs to be determined. In this case, the analysis considered the presence / absence of VGI cropland against the proportions of cropland cover in the global datasets, when considered simultaneously.

A logistic regression of the binary presence of the VGI cropland class against the proportions of cropland indicated by the global datasets was used to determine which of the different global datasets best predicted the VGI data, using the following calculations.

A *logit* function was defined by:

$$logit(Q) = \frac{exp(Q)}{1+exp(Q)} \tag{1}$$

The logistic geographically weighted regression was then calculated as follows the global land cover datasets:

$$P(y_i = 1) = \mathrm{logit}(b_{0(u_i,v_i)} + b_1 x_{1(u_i,v_i)} \ldots + b_n x_{n(u_i,v_i)}) \tag{2}$$

where $P(y_i = 1)$ is the probability that the VGI cropland cover class, y at location i is correctly predicted, b_0 is the intercept term, $x_{1..n}$ are the proportions of cropland indicated by the 6 global datasets under consideration, $b_{1..n}$ are the slopes and (u_i, v_i) is a vector of two dimensional co-ordinates describing the location of i over which the coefficient estimates are assumed to vary.

The basic idea here is that all of the global dataset proportions are considered as a series of independent variables in the logistic regression. The analyses returns a coefficient for each global dataset, the highest of which indicates the strongest effect in predicting the presence of the VGI data for cropland cover.

5 Results

5.1 VGI to Evaluate Global Datasets

As an initial investigation the relationship between crowdsourced data indicating the presence of the Managed and cultivated class were compared with the proportions of cropland as recorded at those locations in five global land cover datasets. The results of standard logistic regressions for the African and South American data are shown in Table 1 and Table 2.

Table 1. Coefficients of the logistic regression of African crowdsourced cropland data against measures derived from global land cover data

Term	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-1.745	0.025	-69.426	0
GLC	1.807	0.098	18.462	0
GlobCover	0.211	0.074	2.845	0.004
MODIS	1.593	0.103	15.444	0
GeoCover	2.139	0.223	9.578	0
Hansen	1.055	0.07	15.087	0

Table 2. Coefficients of the logistic regression of South American crowdsourced cropland data against measures derived from global land cover data

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-2.711	0.062	-43.501	0
GLC	1.536	0.161	9.561	0
GlobCover	0.69	0.13	5.294	0
MODIS	1.573	0.176	8.927	0
GeoCover	1.762	0.126	13.954	0
Hansen	0.635	0.152	4.172	0

These show that for both Africa and South America the proportions of cropland recorded by each of the different global land cover datasets are highly significant (<0.01) predictors of volunteered data on cropland (Cultivated and managed). Of interest are the varying degrees to which the different datasets correspond to the VGI: in both cases GeoCover has the strongest relationship. In Africa GlobCover has the weakest relationship and in South America is with the Hansen data and if we were to chose one dataset that best reflects the actual land cover on the ground, then GeoCover would be selected from this set of five global datasets.

However it is also possible to examine how these relationships vary spatially. A geographically weighted logistic regression was used to model the variation in the degree to which the individual global datasets best predicted the VGI data (ie had the highest coefficient) at different locations. In the geographically weighted approach [40], local regression models are estimated at discrete locations throughout the study area, in this case defined on 100km grid. At each location the local model is calculated from the data points under a kernel and their distance from the location under consideration weights their individual contribution to the model. The size of kernel can be predefined or can be automatically determined. In this case kernel sizes were specified to incorporate 0.01 (all cropland data) and 0.02 (expert and non-expert groups) of the data points. The application of these techniques in the context of the VGI on land cover is described in detail elsewhere [11].

The results of using a geographically weighted approach, i.e. extending the ordinary and stationary logistic regression, are shown in Figure 4. These show which of the global datasets best reflect the conceptualisation of the volunteers in each location (100km grid cell). By examining the maps in Figure 4, it is evident that the GLC data is best predicts cropland in areas with little cropland – the absence of cropland – and that the results in areas of cropland indicate that other datasets more closely correspond to the conceptualisations of the land held by volunteers.

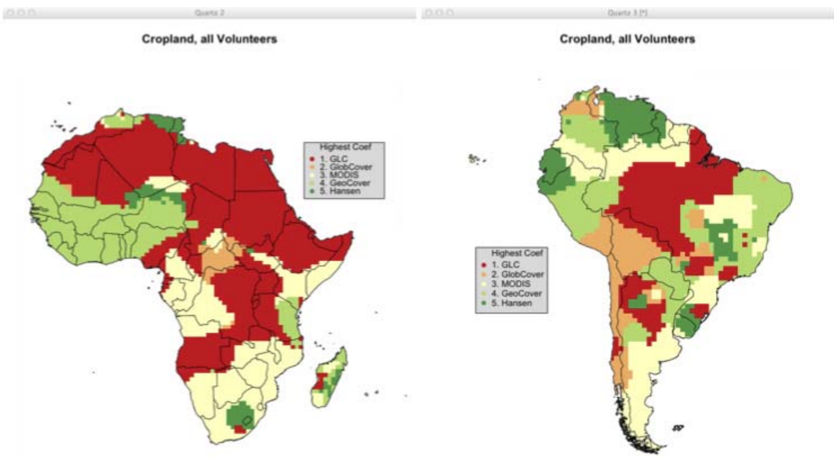

Fig. 4. The global datasets that most strongly correspond to the landscape conceptualisations held by all volunteers in different locations

5.2 Comparing Expert and Non-expert Evaluations of Global Datasets

The analyses described above were rerun, but this time after subsetting the data in order to compare Experts with Non-experts. The results are shown in Figures 5 and 6.

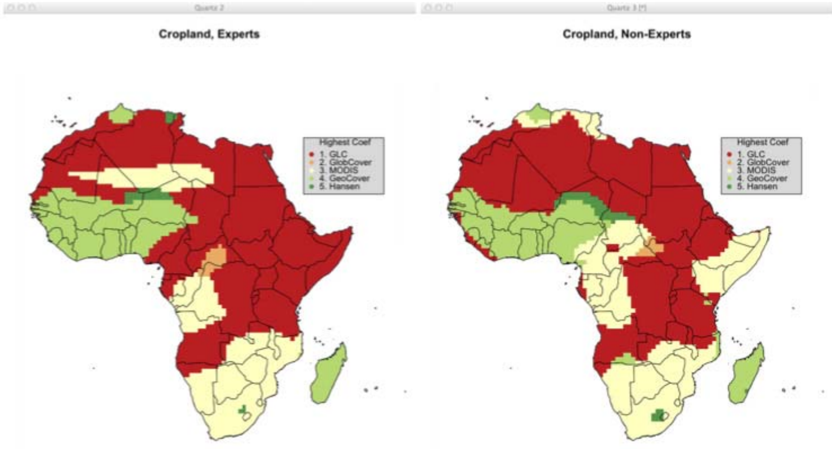

Fig. 5. The global datasets that most strongly correspond to the landscape conceptualisations held by experts and non-experts in Africa

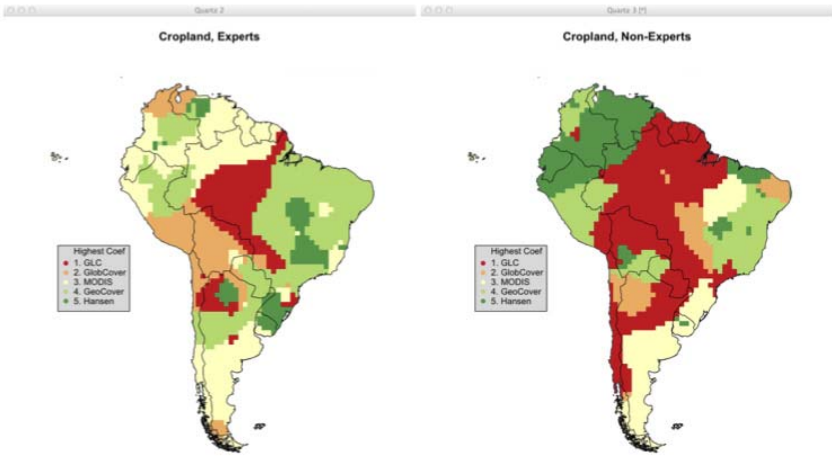

Fig. 6. The global datasets that most strongly correspond to the landscape conceptualisations held by experts and non-experts in South America

These show distinguishable differences in the spatial distribution of the global land cover datasets that most strongly corresponded with the crowdsourced data cropland / Cultivated and managed. The locations of difference are mapped in Figure 7 and it is noticeable there are much greater differences between expert and non-experts and the global land cover datasets they infer in South America than in Africa. This may be due to a number of factors:

- The nature of the land cover. The cropland and cultivated land cover may be more confusing or more difficult to discern in South America.
- The quality of the global land cover data in those areas. There are well known and large variations in the amounts of land cover features reported in different global datasets.

Or, the observed differences may be due to alternative landscape conceptualisations held by the 2 groups. The differences between them can be more formally investigated by considering the sample locations for which land cover was recorded by an expert and a non-expert: 1165 data locations in Africa and 494 in South America.

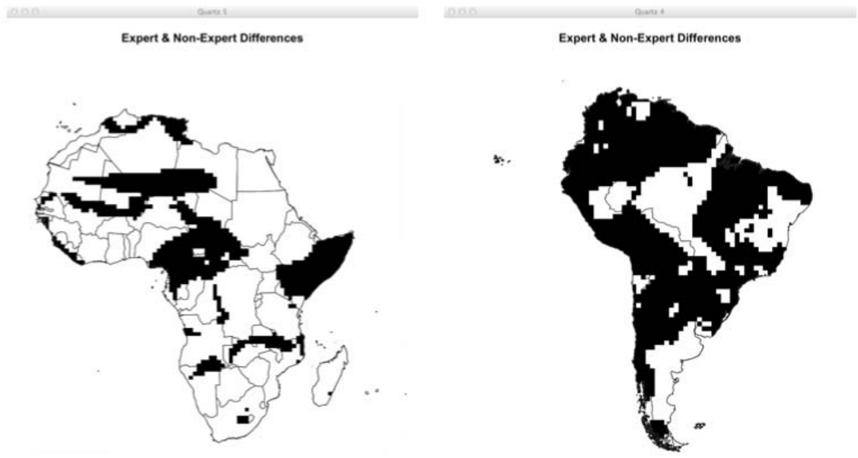

Fig. 7. Differences between the global datasets that best match crowdsourced data on cropland contributed by experts and non-experts

Tables 3 and 4 show correspondence matrices and the per class measures that can be derived from them for the case study areas. If the assumption is that the experts are correct then the equivalent of Type I and Type II error rates can be calculated on a per class basis.

In this context, Type I errors relate to errors of omission (exclusion) for cropland by the non-experts relative to the experts. These indicate the relative probabilities of areas being identified as cropland by experts and not being identified as such by non-experts. These are low (~25%) for Africa and high (~45%) for South America. Considering the Type II errors, the errors of commission (inclusion) by the experts, it is evident that there are similar levels for cropland in both the study areas (~25-30%) which indicates that the experts are including many false positives in their scoring. Of course, it is important to remember that these are *comparative* statements in the way that 'error' is being interpreted, rather than *absolute* statements about error.

Table 3. Correspondence between experts and non-experts in South Africa with the cropland class highlighted

Non-Expert \ Expert	1	2	3	4	5	6	7	8	9	10	Type II
1	111	50	26	2	29	2	0	0	0	0	0.505
2	25	75	22	6	14	1	0	0	10	0	0.490
3	6	9	28	3	9	3	0	0	9	0	0.418
4	0	4	4	215	71	0	0	0	3	0	0.724
5	4	8	5	69	168	0	0	1	2	1	0.651
6	0	0	0	0	0	0	0	0	1	2	0
7	0	0	0	0	0	0	1	0	0	0	1.000
8	1	0	0	0	0	0	0	0	3	0	0
9	1	6	11	4	1	0	0	1	138	0	0
10	0	0	0	0	0	0	0	0	0	0	0
Type I	0.750	0.493	0.292	0.719	0.575	0	1.000	0	0	0	

Table 4. Correspondence between experts and non-experts in South Africa with the cropland class highlighted

Non-Expert \ Expert	1	2	3	4	5	6	7	8	9	10	Type II
1	128	19	12	4	19	3	0	1	0	0	0.688
2	3	12	12	2	3	0	0	0	2	0	0.353
3	1	8	18	5	12	2	0	0	8	1	0.327
4	2	0	3	59	19	0	0	1	0	0	0.702
5	7	1	2	31	59	1	0	0	2	0	0.573
6	2	0	1	0	0	1	0	0	0	0	0.250
7	0	0	0	0	1	0	2	0	0	0	0.667
8	1	1	1	0	0	0	0	2	0	0	0.400
9	0	1	3	0	1	0	0	1	8	0	0
10	1	0	0	0	0	0	0	0	0	5	0
Type I	0.883	0.286	0.346	0.584	0.518	0.143	1	0.400	0	0	

5.3 Volunteer Confidence

The Geo-Wiki on Human Impact also asked volunteers to record their confidence in the land cover class that they allocated to each sample location. Four levels of confidence were available, *Sure*, *Quite sure*, *Less sure* and *Unsure*, and the analysis described below examines whether the differences between experts and non-experts observed persisted when their confidence was Sure. Although other work [37] has shown that experts to be better than non-experts in identifying land cover particularly herbaceous, cropland and mosaic classes, the confidence associated with those class allocations has not been considered. The expert and non-expert data described above were further subsetted to extract the data points that were allocated a Sure confidence attribute. These were then analysed in the same way as above to determine which global dataset best matched the certain VGI cropland data in the two study areas, and maps of difference were created.

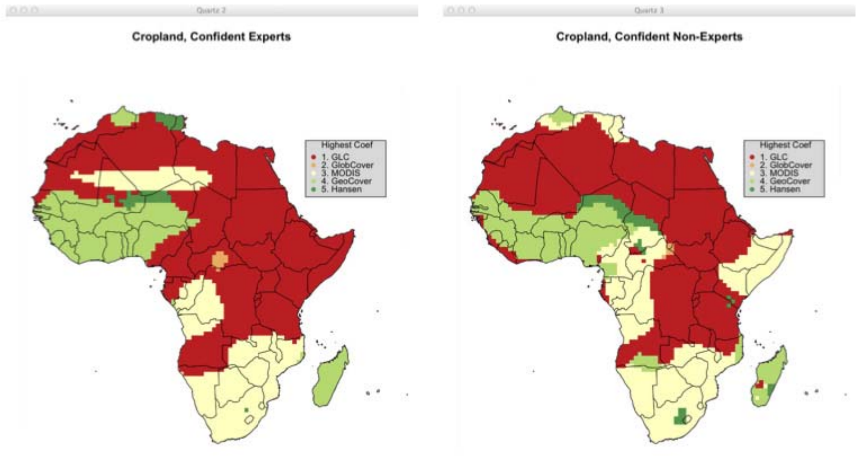

Fig. 8. The global datasets that most strongly correspond to the landscape conceptualisations held by *Confident* experts and non-experts in Africa

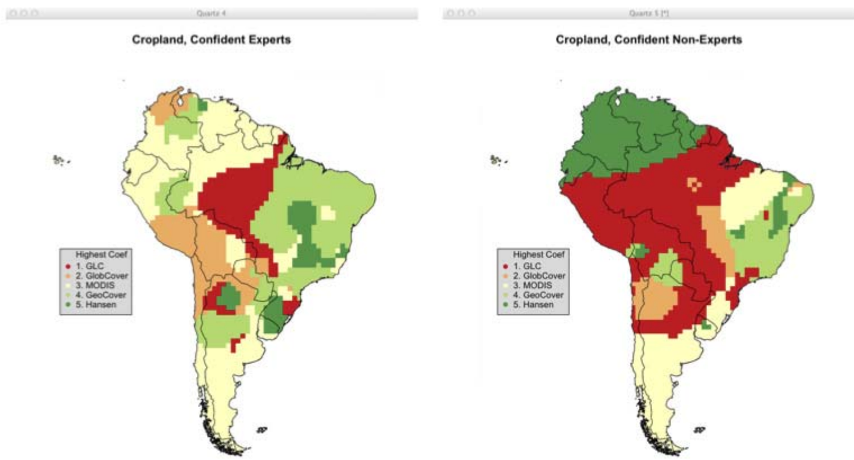

Fig. 9. The global datasets that most strongly correspond to the landscape conceptualisations held by *Confident* experts and non-experts in South America

Figures 8 and 9 show the mappings of the global datasets that most strongly corresponded to the different subsets of VGI data and Figure 10 shows the differences between the expert and non-expert groups. Different levels of confidence were not considered. Interestingly, when these results and figures are compared with Figures 5-7 it is evident that, for Africa, extracting the crowdsourced data where the volunteers were confident (sure) resulted different global datasets being selected by experts and non-experts. In South America it did not.

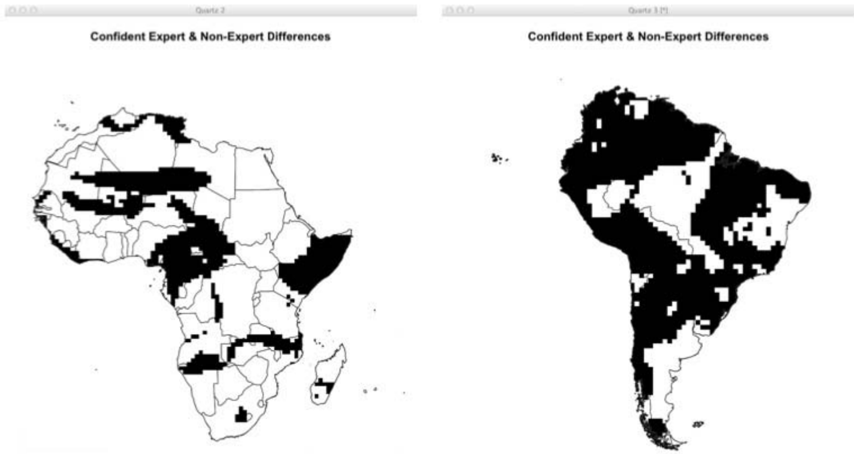

Fig. 10. Differences between the global datasets that best match crowdsourced data on cropland contributed by *Confident* experts and non-experts

5.4 Summary of Results

The aim of this analysis was to explore the varying conceptualisations of cropland land cover held volunteers. The example application was the use of VGI to select from a range of different global land cover datasets, whose quality is known to vary.

Evaluating Global Datasets
For both study areas, the proportions of cropland recorded by each of the global land cover datasets were found to be significant predictors of the presence of crowdsourced data of Cultivated and managed.

GeoCover was found to have the strongest relationship with volunteer conceptualisation of cropland with the GlobCover and Hansen datasets having the weakest relationships in Africa and South America respectively.

The GLC data was found to best predict the absence of cropland being recorded by volunteers.

Comparing Expert and Non-Expert Conceptualisation of Global Datasets
The differences between expert and non-experts conceptualisations of land cover and global datasets measures of cropland were much greater in South America than in

Africa, potentially indicating the impacts of different conceptualisations (eg for data selection) of this feature in these areas.

If the expert data are considered as the referent then low (~25%) errors omission (exclusion) occurred in Africa with high levels (~45%) for South America. Errors of commission (inclusion) were similar in both the study areas (~25-30%).

Volunteer Confidence

The impacts of considering only Sure / certain crowdsourced data on cropland were much greater for the selection of the best global dataset in Africa than in South America.

6 Discussion

There is much interest in methods and applications to take advantage of the large volumes of crowdsourced data that are potentially available to scientific research. The ubiquity of location-enabled devices that attach a geographic reference or geo-tag to such information that is easily shared via various portals and networks provides spatial information scientists with a number of challenges. Not least of these is the need to understand how the concepts embedded in the data that are contributed relate to the intended use of that data. This research has shown that conceptualisations between expert and non-expert groups vary and that decisions made using crowdsourced data (in this global dataset selection) will also vary.

There has been much recent research to understand the quality and reliability of crowdsourced geographical data particularly in relation to the data collected by the Geo-Wick project and methods have been developed for generating error and uncertainty measures [11, 14, 15, 25, 26]. However, as yet little consideration has been given to different conceptualisations of landscape features held by different contributors and the impacts of those. This research has highlighted those differences and their impacts on decisions and the need for further investigation into formal structures to allow such differences to be modelled and reasoned with.

Participation in VGI-related activities may empower citizens to create spatial information and to represent *their* world [41, 42], but may also enhance social disparities across the digital divide as “the favored few “(with digital access) are allowed to exploit “the mediocre many” (without) (quotes from [43]). As spatial data creating and mapping is an inherently politically and socially mediated activity [44], the digital divide results in disparities in the opportunity to create information. Globally, a narrow subset of the world’s population has the opportunity to describe what *is there*. For instance 6.2% of total internet use in 2011 was in Africa despite it accounting for 14.7% of the global population[1], and within developed countries the digital divide reflect an urban bias in access to broadband technologies [45].

[1] http://www.newmediatrendwatch.com/world-overview/

This suggests a number of areas appropriate for further research that are being developed in on-going research by the authors:

- Analysis of the impacts of volunteer characteristics such as their background, training and socio-economic contexts on their perceptions and what they record.
- Consideration of the digital context within which information is volunteered, the impacts of the digital divide on participation, both in terms of contribution and the extent to which participation has the unintended effect of increasing the digital divide.
- Analysis of the degree to which spatial knowledge is shaped by identity, power, and socioeconomic status and the extent to which is spatial data handling through VGI is socially and politically mediated.

References

1. Goodchild, M.F.: Citizens as sensors: the world of volunteered geography. GeoJournal 69, 211–221 (2007)
2. Khatib, F., DiMaio, F., Group, F.C., Group, F.V.C., Cooper, S., et al.: Crystal structure of a monomeric retroviral protease solved by protein folding game players. Nature Structural & Molecular Biology 18, 1175–1177 (2011), doi:10.1038/nsmb.2119
3. Timmer, J.: Galaxy Zoo shows how well crowdsourced citizen science works (2010), `http://arstechnica.com/science/news/2010/10/galaxy-zoo-shows-how-well-crowdsourced-citizen-science-works.ars`
4. Nayar, A. (2009), Model predicts future deforestation. Nature News, `http://www.nature.com/news/2009/091120/full/news.2009.1100.html`
5. Newell, D.A., Pembroke, M.M., Boyd, W.E.: Crowd Sourcing for Conservation: Web 2.0 a powerful tool for biologists. Future Internet 4, 551–562 (2012)
6. Brabham, D.C.: Crowdsourcing the public participation process for planning projects. Planning Theory 8, 242–262 (2009)
7. MacLean, D., Yoshida, K., Edwards, A., Crossman, L., Clavijo, B., Clark, M., Swarbreck, D., Bashton, M., Chapman, P., Gijzen, M., Caccamo, M., Downie, A., Kamoun, S., Saunders, D.G.O.: Crowdsourcing genomic analyses of ash and ash dieback – power to the people. GigaScience 2(2) (2013), doi:10.1186/2047-217X-2-2
8. Goodchild, M.F., Glennon, J.A.: Crowdsourcing geographic information for disaster response: a research frontier. International Journal of Digital Earth 3, 231–241 (2010)
9. Zook, M., et al.: Volunteered Geographic Information and Crowdsourcing Disaster Relief: A Case Study of the Haitian Earthquake. World Medical & Health Policy 2(2), 7–33 (2010)
10. Brunsdon, C., Comber, A.J.: Experiences with Citizen Science: Assessing Changes in the North American Spring. Geoinformatica (2012), doi:10.1007/s10707-012-0159-6
11. Comber, A., See, L., Fritz, S., Van der Velde, M., Perger, C., et al.: Using control data to determine the reliability of volunteered geographic information about land cover. International Journal of Applied Earth Observation and Geoinformation 23, 37–48 (2013), doi:10.1016/j.jag.2012.11.002

12. Vuurens, J., de Vries, A.P., Eickhoff, C.: How Much Spam Can You Take? An Analysis of Crowdsourcing Results to Increase Accuracy. In: Proceedings of the SIGIR 2011 Workshop on Crowdsourcing for Information Retrieval (CIR), Beijing, China (2011)
13. Goodchild, M.F., Li, L.: Assuring the quality of volunteered geographic information. Spatial Statistics 1, 110–120 (2012), doi:10.1016/j.spasta.2012.03.002
14. Haklay, M., Basiouka, S., Antoniou, V., Ather, A.: How Many Volunteers Does it Take to Map an Area Well? The Validity of Linus' Law to Volunteered Geographic Information. The Cartographic Journal 47, 315–322 (2010)
15. Foody, G.M., Boyd, D.S.: Using volunteered data in land cover map validation: Mapping tropical forests across West Africa. In: 2012 IEEE International on Geoscience and Remote Sensing Symposium (IGARSS), pp. 6207–6208 (2012)
16. Hudson-Smith, A., Batty, M., Crooks, A., Milton, R.: Mapping for the masses. Accessing Web 2.0 through crowdsoucring. Social Science Computer Review 27, 524–538 (2009)
17. Li, L., Goodchild, M.F.: The Role of Social Networks in Emergency Management: A Research Agenda. In: International Journal of Information Systems for Crisis Response and Management, vol. 2(4), 11 pages (2010), doi:10.4018/jiscrm.2010100104
18. Wiersma, Y.F.: Birding 2.0: Citizen science and effective monitoring in the web 2.0 world. Avian Conservation and Ecology 5, 13 (2010), `http://www.ace-eco.org/vol5/iss2/art13/`, `http://dx.doi.org/10.5751/ACE-00427-050213`
19. Voigt, S., et al.: Rapid damage assessment and situation mapping: learning from the 2010 Haiti earthquake. Photogram. Eng. and Rem. Sen. 77, 923–931 (2011)
20. Snow, R., O'Connor, B., Jurafsky, D., Ng, A.Y.: Cheap and fast—but is it good?: evaluating non-expert annotations for natural language tasks. In: 2008 Proceedings of the Conference on Empirical Methods in Natural Language Processing, pp. 254–263 (2008)
21. Flanagin, A., Metzger, M.: The credibility of Volunteered Geographic Information. GeoJournal 72, 137–148 (2008)
22. Pal, M., Foody, G.M.: Feature Selection for Classification of Hyperspectral Data by SVM. IEEE Transactions on Geoscience and Remote Sensing 48(5), 2297–2307 (2010)
23. Hughes, G.F.: On the mean accuracy of statistical pattern recognizers. IEEE Transactions on Information Theory 14(1), 55–63 (1968), doi:10.1109/TIT.1968.1054102
24. Raykar, V.C., Yu, S.: Ranking annotators for crowdsourced labeling tasks. Advances in Neural Information Processing 24 (2011)
25. Foody, G.M., Boyd, D.S.: Exploring the potential role of volunteer citizen sensors in land cover map accuracy assessment. In: Proceedings of the 10th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Science (Accuracy 2012), Florianopolis, Brazil, pp. 203–208 (2012)
26. Foody, G.M., See, L., Fritz, S., Van der Velde, M., Perger, C., Schill, C., Boyd, D.S., Comber, A.: Accurate attribute mapping from volunteered geographic information: issues of volunteer quantity and quality. Paper to be submitted to Cartographic Journal (in preparation)
27. Smith, B.: On drawing lines on a map. Spatial Information Theory A Theoretical Basis for GIS, 475–484 (1995)
28. Smith, B.: Fiat Objects. Topoi 20, 131–148 (2001)
29. Smith, B., Mark, D.: Geographical categories: an ontological investigation. International Journal of Geographical Information Science 15, 591–612 (2001)
30. Janowicz, K., Kessler, C.: The Role of Ontology in Improving Gazetteer Interaction. International Journal of Geographical Information Science (IJGIS) 22(10), 1129–1157 (2008)

31. Kuhn, W.: Semantic engineering. Research Trends in Geographic Information Science, 63–76 (2009)
32. Feddema, J.J., Oleson, K.W., Bonan, G.B., Mearns, L.O., Buja, L.E., Meehl, G.A., Washington, W.M.: The importance of land-cover change in simulating future climates. Science 310, 1674–1678 (2005)
33. Fritz, S., See, L., McCallum, I., Schill, C., Obersteiner, M., et al.: Highlighting continued uncertainty in global land cover maps for the user community. Environmental Research Letters 6, 044005 (2011), doi:10.1088/1748-9326/6/4/044005
34. Fritz, S., McCallum, I., Schill, C., Perger, C., See, L., et al.: Geo-Wiki: An online platform for improving global land cover. Environmental Modelling & Software 31, 110–123 (2012), doi:10.1016/j.envsoft.2011.11.015
35. Comber, A.J.: Geographically weighted methods for estimating local surfaces of overall, user and producer accuracies. Remote Sensing Letters 4(4), 373–380 (2013), doi:10.1080/2150704X.2012.736694
36. Comber, A., Fisher, P.F., Brunsdon, C., Khmag, A.: Spatial analysis of remote sensing image classification accuracy. Remote Sensing of Environment 127, 237–246 (2012)
37. See, L., Comber, A.J., Salk, C., Fritz, S., Van der Velde, M., Perger, C., Schill, C., McCallum, I., Kraxner, F., Obersteiner, M.: Comparing the quality of crowdsourced data contributed by expert and non-experts. Paper submitted to PLOSOne (submitted, February 2013)
38. Perger, C., Fritz, S., See, L., Schill, C., Van der Velde, M., et al.: A campaign to collect volunteered geographic Information on land cover and human impact. In: Jekel, T., Car, A., Strobl, J., Griesebner, G. (eds.) GI_Forum 2012: Geovisualisation, Society and Learning, pp. 83–91. Herbert Wichmann Verlag, Berlin (2012)
39. Fritz, S., See, L., Van der Velde, M., Nalepa, R.A., Perger, C., et al.: Downgrading recent estimates of land available for biofuel production. Environ. Sci. Technol. 47, 1688–1694 (2013), doi:10.1021/es303141h
40. Brunsdon, C.F., Fotheringham, A.S., Charlton, M.: Geographically Weighted Regression - A Method for Exploring Spatial Non-Stationarity. Geographical Analysis 28, 281–298 (1996)
41. Tulloch, D.A.: Is VGI participation? From vernal pools to video games. Geojournal 72(3-4), 161–171 (2008)
42. Coleman, D.J., Georgiadou, Y., Labonté, J.: Volunteered geographic information: The nature and motivation of producers. International Journal of Spatial Data Infrastructures Research 4, 332–358 (2009)
43. Sui, D.Z.: The wikification of GIS and its consequences: Or Angelina Jolie's new tattoo and the future of GIS. Environment and Urban Systems 32, 1–5 (2008)
44. Harvey, F., Chrisman, N.: Boundary objects and the social construction of GIS technology. Environment and Planning A 30, 1683–1694 (1998)
45. ONS, Internet Access - Households and Individuals (2012), http://www.ons.gov.uk/ons/dcp171778_275775.pdf