

Article

Local Knowledge and Professional Background Have a Minimal Impact on Volunteer Citizen Science Performance in a Land-Cover Classification Task

Carl Salk ^{1,2,*}, Tobias Sturn ¹, Linda See ¹ and Steffen Fritz ¹

¹ International Institute for Applied Systems Analysis, Laxenburg A2361, Austria; sturn@iiasa.ac.at (T.S.); see@iiasa.ac.at (L.S.); fritz@iiasa.ac.at (S.F.)

² Southern Swedish Forest Research Centre, Swedish University of Agricultural Sciences, Alnarp SE-23053, Sweden

* Correspondence: salk@iiasa.ac.at; Tel.: +46-040-41-53-30

Academic Editors: Cidália Costa Fonte, Parth Sarathi Roy, Clement Atzberger and Prasad S. Thenkabail

Received: 29 July 2016; Accepted: 12 September 2016; Published: 20 September 2016

Abstract: The idea that closer things are more related than distant things, known as ‘Tobler’s first law of geography’, is fundamental to understanding many spatial processes. If this concept applies to volunteered geographic information (VGI), it could help to efficiently allocate tasks in citizen science campaigns and help to improve the overall quality of collected data. In this paper, we use classifications of satellite imagery by volunteers from around the world to test whether local familiarity with landscapes helps their performance. Our results show that volunteers identify cropland slightly better within their home country, and do slightly worse as a function of linear distance between their home and the location represented in an image. Volunteers with a professional background in remote sensing or land cover did no better than the general population at this task, but they did not show the decline with distance that was seen among other participants. Even in a landscape where pasture is easily confused for cropland, regional residents demonstrated no advantage. Where we did find evidence for local knowledge aiding classification performance, the realized impact of this effect was tiny. Rather, the inherent difficulty of a task is a much more important predictor of volunteer performance. These findings suggest that, at least for simple tasks, the geographical origin of VGI volunteers has little impact on their ability to complete image classifications.

Keywords: crowdsourcing; citizen science; data quality; Tobler’s Law; local knowledge; remote sensing; land cover; cropland; volunteered geographical information

1. Introduction

Crowdsourcing is a new term for an old, but increasingly important, concept: the completion of large projects by combining small distributed contributions from the public. Even though the term is not yet widely known, many crowdsourced products are widely used, such as Wikipedia, the online, user-contributed encyclopedia whose popularity—and perhaps even accuracy—rivals traditional reference materials [1]. Within the scientific community, the crowdsourced statistical platform R is displacing traditional closed-source for-profit statistical packages [2]. Crowdsourcing can refer to passive contribution of data—for instance the Google Maps traffic congestion layer that comes from location data sent by Android phones—or more active contributions.

When the goal of a crowdsourcing campaign is to promote and benefit from active public participation in research, the process is often called ‘citizen science’. Citizen science has a long tradition. The Audubon Society Christmas Bird Count has taken place every year since 1900, with an ever-growing geographical scope and number of participants [3]. In recent years, public participation

in the collection and analysis of data for scientific purposes has exploded [4]. This growth is due, in part, to the proliferation of smartphones and tablets that are always on, close at hand, and perpetually networked. Perhaps the best known internet-based citizen science campaign has been the Galaxy Zoo project [5] in which volunteers are tasked with the classification of galaxy shapes. This is the type of task for which crowdsourcing is ideally suited—difficult for computers to reliably perform, but relatively easy for humans, perhaps requiring only a simple training module to learn. While Galaxy Zoo is typical in that it takes a micro-task approach, more macro-task approaches are possible. The FoldIt project uses volunteer labor to determine the three dimensional structure of proteins based on their nucleotide compositions [6]. This task is more akin to a complicated game, like solving a Rubik's Cube, that can be so time consuming that most volunteers do not complete even a single task. Rather, the project is driven forward by a small cadre of participants with exceptional geometric visualization abilities.

When the scope of a crowdsourced project is explicitly geographical, it is often called 'volunteered geographical information' (VGI). Perhaps the best-known VGI project is OpenStreetMap (OSM) [7], an online, open-source mapping project in which the local knowledge of contributors is a key driver to its success. This work is somewhat intermediate between the micro-task and macro-task ends of the spectrum. Typical contributions involve mapping of geographical features such as street names, rivers, and footpath locations, among many others.

One promising area where VGI is so far underexploited is in applications of Earth observation, in particular for the collection of data for the training and validation of products derived from remote sensing [8]. Geo-Wiki is one example of a VGI application for gathering training and validation data for improving global land cover maps [9,10]. To be useful for training and validation, the data provided by volunteers must correctly identify relevant landscape features. This is true of all VGI that is being used in further applications and, hence, there is a considerable body of literature surrounding the development of quality control measures and methods for VGI [11,12]. A recent trend in VGI quality studies has been in the development of indicators related to the contributors and filtering systems driven by user needs [13–15]. Contributions to self-directed VGI campaigns, such as OSM, are often focused near volunteers' homes [16], but the relevance of local familiarity to campaigns centered on image classification micro-tasks is not well studied, although some distance-related effect may be expected based on the geographical literature. Tobler's first law of geography states: "everything is related to everything else, but near things are more related than distant things" [17]. This notion has proven relevant to studies across disciplines as diverse as botany [18], hydrology [19], politics [20], and urban planning [21]. As of August 2016, Tobler's original paper had over 1290 peer-reviewed citations on Web of Science, but only three of those contain the search terms 'crowdsourcing' or VGI [22–24]. These papers rely on Tobler's Law as a theoretical underpinning, but none tests it in the context of VGI. One conference paper has addressed distance and image rating ability [25], but we are unaware of any other study on this topic. Clearly, this is a fertile field for research.

In this paper, we investigate how factors like geographic distance, professional background, and familiarity of a landscape affect volunteer performance in a land-cover classification task. We also look at the impact of landscape familiarity on volunteer confidence in their ratings. These questions are answered using data from online land-cover classification games—developed as part of the Geo-Wiki set of tools—played by participants from around the world.

2. Materials and Methods

2.1. The Cropland Capture Game

The Cropland Capture game took place between November 2013 and May 2014. It has been described in detail elsewhere [26], so we only outline it briefly here. Cropland Capture was a gamified image classification campaign in which volunteers were asked to determine whether a satellite image or ground-based photograph contains cropland within a demarcated region. Volunteers had the option to answer 'yes', 'no', or 'maybe' if they were uncertain or unable to tell. Players were awarded

one point for answers that were correct, with correctness defined in terms of agreement with other volunteer ratings. Since the majority classification may be incorrect, particularly for difficult images, some forgiveness was built into the grading system. When between 20% and 80% of non-maybe volunteer ratings were in the ‘cropland’ category, answers of both ‘cropland’ and ‘no cropland’ were considered correct. In other words, points were deducted only when a volunteer disagreed with at least 80% of previous ratings of an image. When this criterion was met, one point was deducted from a volunteer’s score. Of the 108,367 images that received at least 10 volunteer ratings, only 8974, or 8.3% fell in the 20%–80% cropland rating range at the end of the competition. Responses of ‘maybe’ neither added to nor deducted from a volunteer’s score. While the competition period has ended, the game mechanics and interface can still be seen at the project website [27].

Volunteers taking part in Cropland Capture were asked to fill out a background questionnaire. Responses were voluntary, and we have no way to verify specific answers, but the overall response patterns were in line with demographics and our volunteer outreach efforts. Countries with large populations and places with widespread access to high-speed internet were well represented among the volunteers (Figure 1). All of our outreach and publicity for the game took place in the English- and German-language media, and countries where those languages are widely spoken made up seven of the top 10 countries of volunteer origin (Figure 1). However, a few responses were unlikely or completely impossible, and these were eliminated from the data analyzed. For instance, reported home countries included Antarctica, North Korea, and Bouvet Island, an uninhabited Norwegian territory in the south Atlantic.

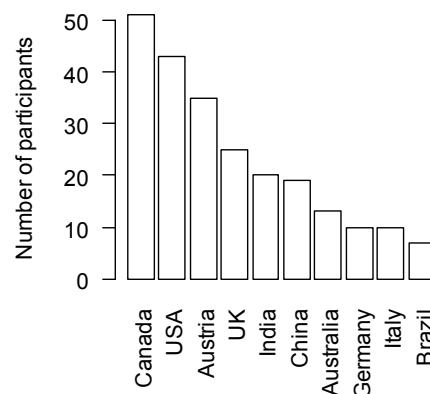


Figure 1. Distribution of players of the Cropland Capture game among the top 10 countries with the most participants. Only participants who rated at least 1000 images are included in this figure. An additional 105 players from 53 other countries were included in the analyses in this paper.

As a basis for evaluating the quality of responses we used a metric we refer to as ‘crowd accuracy’, or simply ‘accuracy’ for short. Crowd accuracy is simply the proportion of a player’s ratings that agree with the crowd’s majority vote for a particular image. For calculating this metric, responses of ‘maybe’ and all images with a tied vote were omitted. Ideally, we would like to use an externally-validated metric that relies on expert-derived information beyond the ratings only of other players of the game. Unfortunately, decisions taken to optimize use of player contributions inadvertently made expert validation of players’ performance more data hungry than was practical for all but the most active volunteers [28]. However, these difficulties are largely overcome by using the crowd accuracy metric as this measure shows strong positive correlations with other ways of measuring player performance [26].

2.2. The Land Cover Game

A second data source was a Geo-Wiki campaign in which participants were asked to identify the land cover classes present within a satellite image and estimate the percentage of the image occupied by each class. The potential classes were tree cover, shrub cover, herbaceous vegetation

or grassland, cropland, regularly flooded or wetland, urban, snow and ice, barren, and open water. Participants were also asked to rate how confident they were of their classifications on a four-point ordinal scale. This campaign differed from the Cropland Capture game in that it was implemented in the Google Earth API. This allowed users to pan and zoom to explore micro- and macro-scale features of the landscape to help make their determinations. During 2012, participants in an international summer school for PhD students were asked to list places in the world where they had lived and were familiar with the landscape. Each student was given images to evaluate that were roughly mixed between familiar and unfamiliar places. A total of 12 students took part in this exercise, and contributed 1516 ratings of 624 unique images. Since the students came from many parts of the world, what was a familiar location for one student was unfamiliar for most others.

2.3. Statistical Analysis—Cropland Capture

User accuracy on images from their home country or continent vs. other areas was compared using two-tailed binomial tests with the user as the level of replication. These tests essentially ask if individual users' image classification accuracy rates were higher at home than in other areas significantly more frequently than expected by chance alone. While this analysis only included participants who had rated at least 1000 images, sample size was still sometimes a consideration, particularly at the country level for players from countries with small land areas (the game did not take a user's country of origin into account when assigning particular images). To ensure that these results were not biased by small sample sizes or arbitrary cutoffs, we applied two different minimum thresholds, 10 and 50 images, for the number of home-country images rated in order to be included in this analysis. This and all subsequent analyses were performed using R version 3.0.2 (R Foundation for Statistical Computing, Vienna Austria).

We assessed user accuracy as a function of distance from home in three different ways. The first was great-circle distance between a volunteer's home and the image location. To compute this distance, we first determined the approximate latitude and longitude of the center of the city where the volunteer reported living. We then calculated the distance between this location and each point the volunteer had rated using the haversine formula [29]. The other two metrics were absolute latitude and longitude difference between a user's home and the point. Our logic for using these metrics was that they may reflect different components of the biophysical and cultural familiarity of a landscape. Latitude difference would reflect differences in user familiarity with boreal, temperate, or tropical landscapes, while longitude difference may have a bearing on the cultural similarities between familiar land-use practices and those depicted in the game images.

Our first analyses of these response variables were generalized linear models (GLM) of correctness as a function of distance measured in the three ways described in the previous paragraph. Since the response variable was simply whether a classification was correct (1) or incorrect (0), all models (including those described below) used a binomial error model with a logit link. This simple regression approach is not a perfect fit to the data due to the non-independence among data points—each user performed many ratings and most points were evaluated by many volunteers. To ensure that this confounding factor did not lead to false results, we re-ran the distance models as linear mixed models with random effects specified for either volunteer or point. Generalized linear models with logit links were also applied to analyses of the interactions between distance of a point from a volunteer's home and other variables: professional background and whether a volunteer lived in a western (North America, Europe and Oceania) or non-western (Asia, Africa and Latin America) country. These and all other analyses were performed using the R package 'lme4' [30].

The impact of professional background on rating reliability was evaluated in two different ways. First, differences among volunteers as a function of self-reported career category were assessed using one-way ANOVA. The relative distance-based effects (see previous paragraph) among volunteers with a remote sensing or land cover (RS/LC) background were compared to other volunteers by adding an interaction term between distance and a dummy variable for RS/LC professionals to the GLM.

To test more specifically the effect of specialized local knowledge on image classification performance we took advantage of a subset of images within the Cropland Capture game that were extremely tricky for most participants to classify correctly. These images came from Ireland and western Great Britain where pastures are commonly bounded by rock walls or hedgerows. As seen in satellite imagery, this landscape is characterized by bright green geometrically-patterned patches. To most eyes, this looks like cropland in spite of not fitting the provided definition. We manually examined all satellite images in the game coming from the region bounded by longitudes 11°W and 1°W and latitudes 50°N and 55°N. This region covers the entire island of Ireland, all of Wales, western England, and Southwestern Scotland. Of these images, we selected 84 that many participants rated as containing cropland, but could be confirmed to contain no cropland. The confirmation was done with the aid of Google Earth and Street View. An example of such an image is shown in Figure 2. We tested the hypotheses that volunteers from Britain and Ireland would be better at correctly identifying these images as non-cropland than players from other places using a two-tailed exact binomial test. We also tested whether British and Irish players would be more likely to express uncertainty by rating these images as ‘maybe’ using the same statistical test.

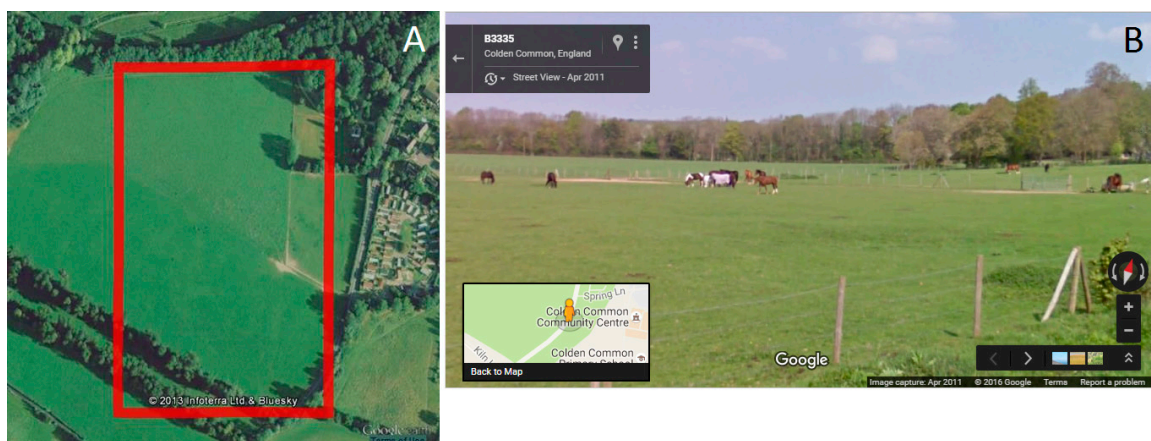


Figure 2. An example of pasture land in an image from the Cropland Capture game. (A) The image as seen by players of the game. Players overwhelmingly classified this as being cropland; and (B) a view of the same parcel from ground-level as seen in Google Street View, showing that this field is pasture, not cropland.

User response speed was assessed using timestamps associated with each rating. Since the server tended to cluster timestamps in approximately 10 s intervals, the individual differences among timestamps could not be directly analysed. However, when averaged, the differences between successive timestamps give an accurate indication of the mean response time of an individual volunteer. To avoid skewing means with large values that indicate a volunteer taking a break rather than answering slowly, we omitted all values >300 s, a value beyond which the response time distribution for similar tasks has been shown to break down [31]. Since individual volunteers’ mean response times were non-normally distributed, we tested for differences between remote sensing specialists and volunteers with other backgrounds using a Mann-Whitney-Wilcoxon U test.

2.4. Statistical Analysis—Land Cover Game

The land cover game involved answering several questions about an image, including the dominant land-cover types present and their cover percentages. Participants were also asked to evaluate their confidence about particular images using an ordinal scale, with four possible responses, ‘sure’, ‘quite sure’, ‘less sure’, and ‘unsure’. These responses were internally stored as the numbers 0, 10, 20, and 30, respectively. To test whether volunteers were more confident in their responses to tasks in familiar regions, we used a bootstrapping procedure known as a ‘permutation test.’

In this test, the confidence responses were randomly reassigned among the images. The mean of the newly-randomized uncertainty scores was then calculated for the familiar and unfamiliar locations. The randomization procedure was repeated 10,000 times. This process maintains the same distribution of scores while allowing calculation of the variability in the mean of the familiar and unfamiliar locations' confidence levels assuming they are drawn from the same distribution. The proportion of randomizations that resulted in group-level means with a bigger difference than the observed difference between the familiarity groups' means is an estimate of the probability of results more extreme than observed resulting from chance alone when the groups are drawn from identical distributions.

Since user-reported uncertainty is on an ordinal scale, there is no completely objective method of converting the responses into a numerical scale. To increase confidence in our results, we performed a sensitivity analysis by re-running the analysis described above with different numerical values assigned to the four responses on the sure-unsure scale. In these sensitivity analyses, the endpoints of the scale retained the same values, 0 for 'sure' and 30 for 'unsure'. The intervening two values were changed to represent different scenarios of dispersion and skew among the certainty levels. Under these scenarios, the values for the 'quite sure' and 'less sure' responses were 5 and 10, 20 and 25, 5 and 25, and 12.5 and 17.5.

3. Results

3.1. National and Continental Familiarity—Cropland Capture

Volunteers showed small, but significant, differences in classification accuracy between their home country and other countries. Of the 253 volunteers who rated at least 1000 points and at least 10 points in their home country, mean accuracy was 92.12% at home and 91.41% in other countries. Although small, this difference was statistically strong (two-tailed binomial test, 146 successes on 253 trials, $p = 0.017$). When more stringent cutoffs were used (a minimum of 50 points in the home country, limiting the sample to 145 volunteers), the small home country advantage persisted (92.51% at home vs. 92.05% away), however, its statistical strength disappeared (two-tailed binomial test, 77 successes on 145 trials, $p = 0.507$). This pattern was even weaker for ratings in volunteers' home continents vs. other continents. The 337 volunteers with at least 1000 total ratings averaged 91.35% agreement for images from their home continent and 91.30% on other continents (two-tailed binomial test, 175 successes on 337 trials, $p = 0.513$).

3.2. Distance-Based Familiarity—Cropland Capture

Classification accuracy decreased slightly with distance of the image location from a volunteer's home. This pattern persisted regardless of whether great circle distance, latitude difference, or longitude difference was used, but the statistical significance of the pattern depended on how it was measured and what random effects were included in the model. When measured along a great circle, accuracy decreased significantly with distance from home (GLM results in Table 1). Similar patterns were seen for both latitude (Table 1) and longitude difference between a player's home location and image locations (Table 1). Volunteers from western and non-western countries showed similar patterns of slightly reduced accuracy with distance from home (Figure 3A). Due to the binomial response model used in these analyses, standard measures of model fit, like R^2 , are not easily interpreted. However, the narrow range of fitted values seen in Figure 3 gives an indication of the relatively weak explanatory power of these models. Since most individual ratings agreed with the crowd majority vote, predicted agreement probability varied little (note the small ranges on the y-axes in Figure 3). When random effects were included for images, the results showed little change (Table 1). However, when random effects for volunteers were included, coefficients for distance measures shrank and model p-values increased; for latitude and longitude differences traditional cutoffs for significance were no longer met (Table 1).

Table 1. Key results from different models of the probability of correct answers on a game-based cropland-identification task. The negative coefficients indicate that user performance is worse with greater distance from their home. In the ‘model’ column, items in bold are predictor variables whose coefficients and p-values are given in the subsequent columns. Items in italics are random effects included to control for possible variation in individual volunteer skill (user) or task difficulty (image). Units for distance are in kilometers and Δ latitude and Δ longitude are in degrees. A total of $n = 2,065,714$ observations spread among 151,756 images and 301 users were used in these models.

Model	Coefficient	p-Value	AIC
correct ~ distance	-1.592×10^{-5}	$<2 \times 10^{-16}$	817,995
correct ~ distance + <i>user</i>	-2.483×10^{-6}	0.00194	797,140.3
correct ~ distance + <i>image</i>	-1.731×10^{-5}	$<2 \times 10^{-16}$	721,859.5
correct ~ Δlatitude	-0.0015033	$<2 \times 10^{-16}$	818,320
correct ~ Δlatitude + <i>user</i>	-0.0001469	0.252	797,148.7
correct ~ Δlatitude + <i>image</i>	-0.0020105	$<2 \times 10^{-16}$	722,088.3
correct ~ Δlongitude	-6.739×10^{-4}	$<2 \times 10^{-16}$	818,356
correct ~ Δlongitude + <i>user</i>	-9.353×10^{-5}	0.16	797,148.1
correct ~ Δlongitude + <i>image</i>	-7.512×10^{-4}	$<2 \times 10^{-16}$	722,142.8

3.3. Local Knowledge in a Confusing Landscape—Cropland Capture

A total of 3096 ratings of images showing pastures superficially similar to cropland in western Britain and Ireland were completed by players whose home country was known. Of these, 128 ratings were performed by players whose home was in the Republic of Ireland or the United Kingdom (Table 2). Players from these countries were no more likely than players from other countries to correctly rate these images as non-cropland (two-tailed binomial test, $p = 0.443$). Similarly, Irish and British players were no more likely to express uncertainty by using the ‘maybe’ rating than other players (two-tailed binomial test, $p = 0.505$; Table 2).

Table 2. Ratings of a subset of non-cropland images in Britain and Ireland as performed by Irish and British participants and participants from other countries. These images were particularly difficult as they contained subdivided pastures that looked like cropland to most volunteers. (A) The frequency of rating these images as cropland or non-cropland as a function of country of origin; and (B) the frequency of rating these images or expressing uncertainty by using the ‘maybe’ response as a function of country of origin. Note that all responses in Table 2A fall into the ‘certain’ line of Table 2B. The percentage values in parentheses are the proportion of values in a column that were rated in a certain way; for example, in (A), 96.1% of the responses by Irish/British participants were that the images were cropland. In both cases, differences between Irish/British players and the general population were not statistically discernible.

	A			B	
	Irish or British	Other Country		Irish or British	Other Country
Cropland	122 (96.1%)	2866 (97.1%)	Certain	127 (99.2%)	2952 (99.5%)
Not Cropland	5 (3.9%)	86 (2.9%)	Uncertain	1 (0.8%)	16 (0.5%)

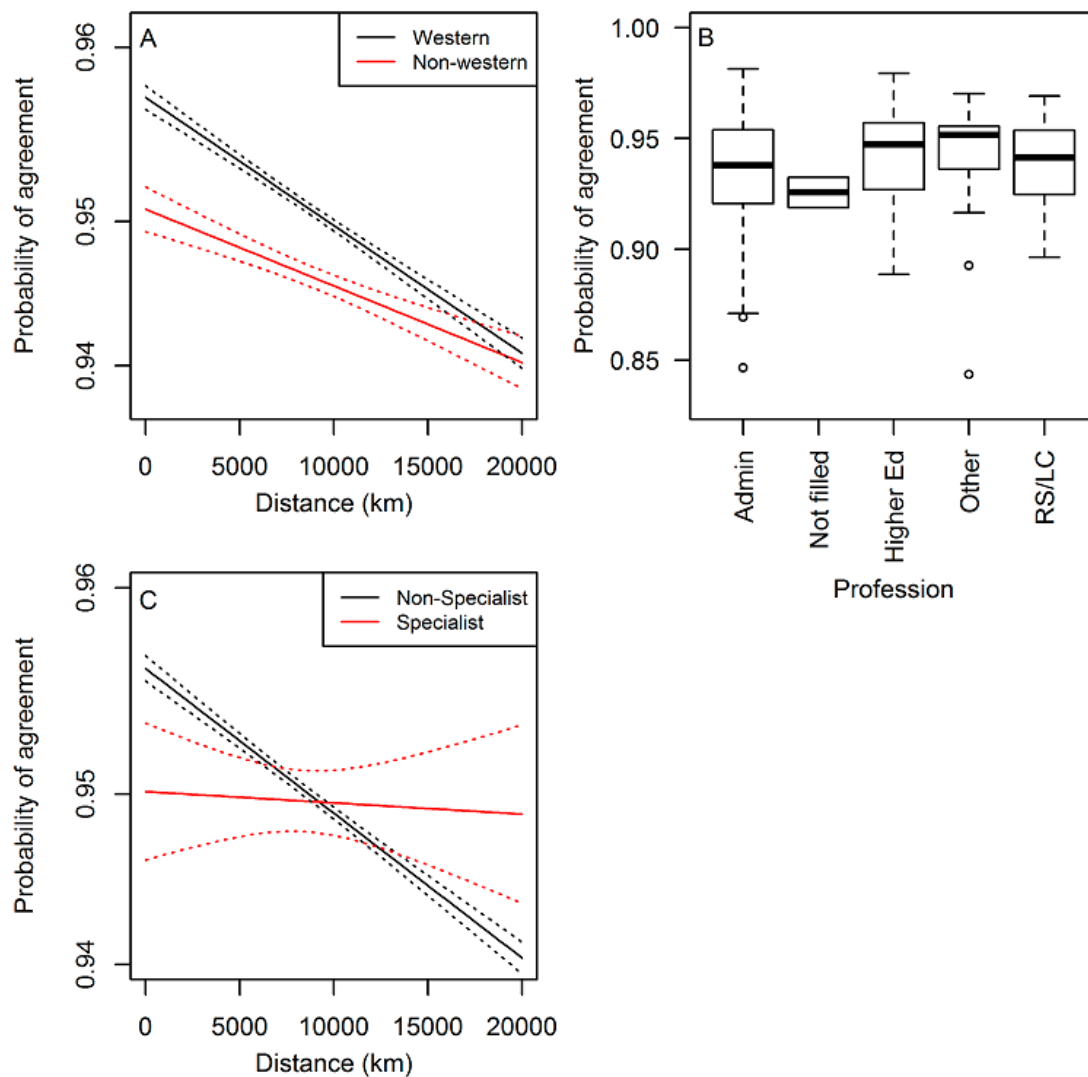


Figure 3. Predicted probability of agreement with the majority of ratings for an image in the Cropland Capture game. (A) As a function of the image location's great circle distance from home and a volunteer's home region ('western' includes North America, Europe, and Oceania; 'non-western' is Asia, Africa, and Latin America); (B) As a function of self-reported profession. RS/LC stands for 'Remote sensing/land cover'. Differences among groups are non-significant (see main text); (C) As a function of image distance from home and professional background. The 'specialist' professional background applies to users who reported their profession as 'remote sensing or land cover', while the 'non-specialist category applies to everyone else. The dotted lines show ± 2 standard error confidence intervals. The y-axes in panels A and C are on a logit scale due to the binomial probability model used for the response variable. Note the small range of values on these y-axes—this indicates low explanatory power of the models, in spite of their statistical significance (see main text).

3.4. Professional Background—Cropland Capture

Volunteers in the Cropland Capture game showed no direct pattern of work quality as a function of professional background. Among users with >1000 points rated, there was no significant difference among professions (ANOVA, $F(4) = 1.1151$, $p = 0.349$). Users with a background in remote sensing or land cover had an average rate of agreeing with the crowd (Figure 3B). Similarly, a logit-linked GLM assessing how well a professional background in remote sensing or land cover predicted agreement with the volunteer majority vote showed no relationship (full results not shown). A separate model with an interaction term between professional background (non-specialist vs. specialist in

remote sensing or land cover background) and distance from home to the image location showed that the non-specialists actually performed slightly better than specialists near home (Table 3; Figure 3C). This advantage decreased with distance from home; non-specialist performance decreased with distance, but specialists maintained similar performance regardless of distance from home (Table 3; Figure 3C). As in the analysis combining all professions (Figure 3C), the magnitude of this distance-based trend was very small. Volunteers with a remote sensing or land cover background worked faster than other volunteers. Average response time was 2.248 s among 10 RS/LC specialists and 2.712 s for 52 other volunteers. However, these numbers were quite variable among participants and the difference between the two groups was statistically weak (unpaired two-tailed Mann-Whitney test, $U = 212$, $p = 0.363$).

Table 3. Regression table for the model of agreement with the majority classification of an image as a function of distance of that image from a participant’s home and professional background. The coefficient is either an absolute difference in probability correct or a change in probability correct/km where distance is involved in a term. The asterisk symbol (*) denotes an interaction between two variables.

Variable	Coefficient	Std. Err	<i>p</i> -Value
(Intercept)	2.947	0.0387	<0.00001
Non-specialist	0.1394	0.0394	0.00040
Distance	−0.00000127	0.0000041	0.7563
Non-specialist*distance	−0.0000152	0.0000042	0.00029

3.5. Volunteer Confidence—Land Cover Game

Volunteers were more confident in their ratings in familiar regions. The mean uncertainty score was 6.24 for images from familiar areas, and 8.89 in unfamiliar areas. The permutation test revealed strong statistical support for this difference; all 10,000 randomizations of the uncertainty scores resulted in differences between the group-level means that were smaller than observed in the un-randomized data. The sensitivity analysis showed these results to be robust to changes in the numbers assigned to the ordinal uncertainty scores; in all scalings of the data, all 10,000 randomizations resulted in mean differences smaller than the observed difference.

4. Discussion

This paper has tested an idea analogous to Tobler’s Law [17]—that land-cover classification tasks are easier when they are from closer to the home of the person doing the classification. Our results have found support for this notion in some circumstances, but the magnitude of the effect is so small that, at least for cropland detection, it is unlikely to be of much practical importance for future design of global classification tasks. Similar trends toward decreasing accuracy with distance were seen whether the response variable was the home country vs. other countries or simply the linear distance from one’s home. Even when particularly tricky Irish and British pasturelands were singled out for evaluation, residents of those areas showed no advantage. When home vs. other continent, latitudinal or longitudinal distance were used as response variables, trends were all in the same direction, although non-significantly so. These findings are congruent with the results of the only other study we are aware of looking at distance effects in VGI [25], which analysed Geo-Wiki data from the first human impact campaign that was run in 2011 [32]. They found that the relative odds of participants correctly choosing the right land cover decreased by 2% per 1000 km distance from the participant’s home location, with little difference between land cover types. Yet location and local knowledge have been found to be relevant in studies of other types of tasks. A comparison of the contributions of local volunteers in classifying road types from imagery with professional surveyors found that the accuracy increased from 68% to 92% when the data were collected by the local volunteers [33]. OpenStreetMap is built on the idea that local knowledge is of fundamental

importance in contributing and tagging features [34] yet many humanitarian mapping exercises use remote mappers to provide rapid information for response teams. A recent study compared the results from remote mappers with local mappers in terms of how well they digitized buildings in Kathmandu [35]. The results showed that the remote mappers missed around 10% of buildings and that there were issues with positional accuracy. These two latter examples were focused on data collection in a small area where the effect of local knowledge may simply have much more influence than on the worldwide classification data presented here. From a design perspective, there appears to be little need to geographically target the tasks to individuals in closer proximity to the task location for these types of global data collection exercises.

Professional background has little first-order relationship with task accuracy, although it apparently interacts with distance, suggesting that the distance effects described above may be common among the general population, but do not apply to remote sensing specialists. In contrast, a study of the Geo-Wiki database found relative odds of correctly predicting land cover around 1.7 times larger for experts when compared to non-experts in remote sensing [25]. This may be because the Geo-Wiki human impact campaign involved more detailed identification of all land cover types while Cropland Capture focussed on only one type, so it was a much simpler task. This points toward a widely accepted axiom of game design, that simpler tasks require less training.

Interestingly, self-rated confidence of responses in a small experiment involving students increased in familiar landscapes. While this was not an idea we could test on the large dataset used for the rest of our analyses, it suggests that confidence (or lack thereof in unfamiliar regions) has little to do with actual performance—confidence varied much more than did actual task performance. In contrast, the relative odds of correctly classifying the land cover was twice as high for confident volunteers compared to non-confident ones in the same Geo-Wiki study cited above [25]. Again, this may relate to the increased task difficulty, and should motivate keeping tasks as simple as possible while preserving meaningful scientific use.

The volunteer participants in this study somewhat over-represent western industrialized countries of North America and Europe (Table 1). This is a common pattern found in many citizen science initiatives [36] and likely has several explanations, including the targeting of our media outreach, English as the language of the application and the availability of high-speed internet connections. Regional mis-representation could potentially explain some of our results. Smallholder and subsistence agriculture are more common in non-western countries, so participants from those areas might enjoy an advantage in recognizing these types of farming. Conversely, the large-scale mechanized agriculture typical of industrialized countries and its geometrically-patterned landscapes should be relatively easy for anyone to identify. This possibility is congruent with the slightly stronger distance-based effects we observe among volunteers from industrialized nations (Figure 3A). However, Asia and Latin America, but not Africa, still had substantial representation among the participants. This, combined with the quite small impact of distance-based effects, supports the conclusion that the overall impact of regional effects on our results was minimal.

In spite of the small magnitude of the location-based effects we observed, finer-grained local knowledge may still play a role in volunteer task performance. For instance, the result that British and Irish volunteers are no better than others at recognizing the unique pasture landscapes in their region may be due, in part, to combining urban and rural residents for this analysis. It is possible that rural residents of these countries, particularly those in areas where grazing is commonplace, would be better at this task than urban residents. Unfortunately, our data was insufficient to parse this possibility. Similarly, rural residents of non-western countries may be quite adept at identifying local agricultural patterns, a phenomenon difficult to detect in our worldwide database. Bias in participation in urban areas is a pattern strongly identified in other geographical crowdsourcing applications such as OpenStreetMap [37]. It would be interesting to test whether differences in urban-rural knowledge have an effect on the accuracy of classification tasks related to cropland if targeted recruitment strategies can be incorporated into future campaigns.

The response variable used in part of our study (the rate classification agreement with the majority of other users) may reasonably be questioned as a metric of use performance [26]. However, we believe that it still provides a useful assessment of the impact of local knowledge on response quality. A previous study using this data has shown that for the vast majority of tasks, there is little disagreement between volunteers and experts. Even for the tiny proportion of the hardest points, the crowd agreed with experts nearly 80% of the time [28]. Thus, the volunteer majority is correct in the vast majority of cases, providing a robust comparison for the distance- and region-based analyses in this study. This previously-observed difference in agreement with experts between easy and hard points is of a magnitude that dwarfs the variation in performance with distance or country demonstrated in this paper. Thus, the inherent difficulty of a task appears much more important for predicting classification correctness than any trait of individual volunteers that we assessed. Distance, country, and profession have truly tiny effects in comparison. However, it would be interesting to evaluate volunteer performance in a smaller geographical region where the effect of local knowledge may be much more relevant.

5. Conclusions

These results have clear consequences for VGI campaign implementation. From these findings, it appears that the geographic origin of the participants has little impact on their ability to identify cropland. While encouraging participation from around the world is certainly important for broadening the impact of science and distributing the benefits of citizen science as widely as possible, it appears that these factors do not greatly affect the scientific outcome of projects. It is important to caution, however, that this result derives from a simple binary classification task. Tasks with more complex response possibilities may prove not to follow the patterns we have shown here, particularly if they involve searching for or classifying geographical features that are highly localized or culturally specific. However, even when we focused on pastures in Ireland and Britain that were widely misclassified as cropland, residents of those islands misinterpreted these unique landscapes just as much as everyone else. Overall, our results show little cause for worry about where volunteers come from for more general types of VGI tasks.

Acknowledgments: We would like to thank Elena Moltchanova, Glenn Wright and Juan-Carlos Laso-Bayas for statistical advice. This research was funded by the European Research Council CrowdLand project (Grant No. 617754).

Author Contributions: Tobias Sturn, Linda See and Steffen Fritz conceived the Cropland Capture game; Tobias Sturn designed and programmed the game; Carl Salk analyzed the data; Carl Salk and Linda See wrote the paper.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Giles, J. Internet encyclopaedias go head to head. *Nature* **2005**, *438*, 900–901. [[CrossRef](#)] [[PubMed](#)]
2. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2014.
3. Silvertown, J. A new dawn for citizen science. *Trends Ecol. Evol.* **2009**, *24*, 467–471. [[CrossRef](#)] [[PubMed](#)]
4. Bonney, R.; Shirk, J.L.; Phillips, T.B.; Wiggins, A.; Ballard, H.L.; Miller-Rushing, A.J.; Parrish, J.K. Next steps for citizen science. *Science* **2014**, *343*, 1436–1437. [[CrossRef](#)] [[PubMed](#)]
5. Raddick, M.J.; Bracey, G.; Gay, P.L.; Lintott, C.J.; Cardamone, C.; Murray, P.; Schawinski, K.; Szalay, A.S.; Vandenberg, J. Galaxy Zoo: Motivations of citizen scientists. *Astron. Educ. Rev.* **2013**, *12*, 010106.
6. Good, B.M.; Su, A.I. Games with a scientific purpose. *Genome Boil.* **2011**, *12*. [[CrossRef](#)] [[PubMed](#)]
7. Jokar Arsanjani, J.; Zipf, A.; Mooney, P.; Helbich, M. *OpenStreetMap in GIScience*; Springer International Publishing: Cham, Switzerland, 2015.

8. See, L.; Fritz, S.; Dias, E.; Hendriks, E.; Mijling, B.; Snik, F.; Stammes, P.; Vescovi, F.; Zeug, G.; Mathieu, P.-P.; et al. A new generation of tools for crowdsourcing and citizen science to support Earth Observation calibration and validation. *IEEE Geosci. Remote Sens. Mag.* **2016**, in press. [[CrossRef](#)]
9. Fritz, S.; McCallum, I.; Schill, C.; Perger, C.; See, L.; Schepaschenko, D.; van der Velde, M.; Kraxner, F.; Obersteiner, M. Geo-Wiki: An online platform for land cover validation and the improvement of global land cover. *Environ. Model. Softw.* **2012**, *31*, 110–123. [[CrossRef](#)]
10. See, L.; Fritz, S.; Perger, C.; Schill, C.; McCallum, I.; Schepaschenko, D.; Karner, M.; Kraxner, F.; Obersteiner, M. Harnessing the power of volunteers, the Internet and Google Earth to collect and validate global spatial information using Geo-Wiki. *Technol. Forecast. Soc. Chang.* **2015**, *98*, 324–335. [[CrossRef](#)]
11. Antoniou, V.; Skopeliti, A. Measures and indicators of VGI quality: An overview. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **2015**, II-3/W5. [[CrossRef](#)]
12. Senaratne, H.; Mobasher, A.; Ali, A.L.; Capineri, C.; Haklay, M. A review of volunteered geographic information quality assessment methods. *Int. J. Geogr. Inf. Sci.* **2016**, 1–29. [[CrossRef](#)]
13. Bordogna, G.; Carrara, P.; Criscuolo, L.; Pepe, M.; Rampini, A. On predicting and improving the quality of Volunteer Geographic Information projects. *Int. J. Digit. Earth* **2014**, *9*, 134–155. [[CrossRef](#)]
14. Meek, S.; Jackson, M.J.; Leibovici, D.G. A flexible framework for assessing the quality of crowdsourced data. In Proceedings of the AGILE'2014 International Conference on Geographic Information Science, Castellón, Spain, 3–6 June 2014.
15. Meek, S.; Jackson, M.; Leibovici, D.G. A BPMN solution for chaining OGC services to quality assure location-based crowdsourced data. *Comput. Geosci.* **2016**, *87*, 76–83. [[CrossRef](#)]
16. Zielstra, D.; Zipf, A. A comparative study of proprietary geodata and volunteered geographic information for Germany. In Proceedings of the 13th AGILE International Conference on Geographic Information Science, Guimarães, Portugal, 11–14 May 2010.
17. Tobler, W.R. A computer movie simulating urban growth in the Detroit region. *Econ. Geogr.* **1970**, *46*, 234–240. [[CrossRef](#)]
18. Bjorholm, S.; Svenning, J.C.; Skov, F.; Balslev, H. To what extent does Tobler's 1st law of geography apply to macroecology? A case study using American palms (Arecaceae). *BMC Ecol.* **2008**, *8*. [[CrossRef](#)] [[PubMed](#)]
19. Wechsler, S.P. Uncertainties associated with digital elevation models for hydrologic applications: A review. *Hydrol. Earth Syst. Sci.* **2007**, *11*, 1481–1500. [[CrossRef](#)]
20. Franzese, R.J.; Hays, J.C. Interdependence in Comparative Politics Substance, Theory, Empirics, Substance. *Comp. Political Stud.* **2008**, *41*, 742–780. [[CrossRef](#)]
21. Miller, H.J. Potential contributions of spatial analysis to geographic information systems for transportation (GIS-T). *Geogr. Anal.* **1999**, *31*, 373–399. [[CrossRef](#)]
22. Goodchild, M.F.; Li, L.N. Assuring the quality of volunteered geographic information. *Spat. Stat.* **2012**, *1*, 110–120. [[CrossRef](#)]
23. Herfort, B.; de Albuquerque, J.P.; Schelhorn, S.J.; Zipf, A. Exploring the Geographical Relations between Social Media and Flood Phenomena to Improve Situational Awareness. In *Connecting a Digital Europe through Location and Place*; Springer International Publishing: Cham, Switzerland, 2014; pp. 55–71.
24. Comber, A.; Fonte, C.; Foody, G.; Fritz, S.; Harris, P.; Olteanu-Raimond, A.M.; See, L. Geographically weighted evidence combination approaches for combining discordant and inconsistent volunteered geographical information. *Geoinformatica* **2016**, *20*, 503–527. [[CrossRef](#)]
25. Comber, A.; See, L.; Fritz, S. The Impact of Contributor Confidence, Expertise and Distance on the Crowdsourced Land Cover Data Quality. In *GI_Forum 2014: Geospatial Innovation for Society*; Vogler, R., Car, A., Strobl, J., Griesebner, G., Eds.; Herbert Wichmann Verlag: Berlin, Germany; Offenbach, Germany, 2014.
26. Salk, C.F.; Sturn, T.; See, L.; Fritz, S.; Perger, C. Assessing quality of volunteer crowdsourcing contributions: Lessons from the Cropland Capture game. *Int. J. Digit. Earth* **2016**, *9*, 410–426. [[CrossRef](#)]
27. Cropland Capture. Available online: www.geo-wiki.org/games/croplandcapture (accessed on 17 September 2016).
28. Salk, C.F.; Sturn, T.; See, L.; Fritz, S. Limitations of majority agreement in crowdsourced image interpretation. *Trans. GIS* **2016**. [[CrossRef](#)]
29. Robusto, C.C. The cosine-haversine formula. *Am. Math. Mon.* **1957**, *64*, 38–40. [[CrossRef](#)]
30. Bates, D.; Maechler, M.; Bolker, B.; Walker, S. Fitting Linear Mixed-Effects Models Using lme4. *J. Stat. Softw.* **2015**, *67*, 1–48. [[CrossRef](#)]

31. See, L.; Comber, A.; 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. *PLoS ONE* **2013**, *8*, e69958. [[CrossRef](#)] [[PubMed](#)]
32. Perger, C.; Fritz, S.; See, L.; Schill, C.; van der Velde, M.; McCallum, I.; Obersteiner, M. A campaign to collect volunteered geographic information on land cover and human impact. In *GI_Forum 2012: Geovizualisation, Society and Learning*; Jekel, T., Car, A., Strobl, J., Griesebner, G., Eds.; Herbert Wichmann Verlag, VDE VERLAG GMBH: Berlin, Germany; Offenbach, Germany, 2012; pp. 83–91.
33. De Leeuw, J.; Said, M.; Ortegah, L.; Nagda, S.; Georgiadou, Y.; DeBlois, M. An Assessment of the Accuracy of Volunteered Road Map Production in Western Kenya. *Remote Sens.* **2011**, *3*, 247–256. [[CrossRef](#)]
34. Budhathoki, N.R.; Haythornthwaite, C. Motivation for open collaboration: Crowd and community models and the case of OpenStreetMap. *Am. Behav. Sci.* **2012**, *57*, 548–575. [[CrossRef](#)]
35. Eckle, M.; Porto de Albuquerque, J. Quality Assessment of Remote Mapping in OpenStreetMap for Disaster Management Purposes. In Proceedings of the ISCRAM 2015 Conference, Kristiansand, Norway, 24–27 May 2015.
36. Chandler, M.; See, L.; Copas, K.; Schmidt, A.; Claramunt, B.; Danielsen, F.; Legrind, J.; Masinde, S.; Miller Rushing, A.; Newman, G.; et al. Contribution of citizen science towards international biodiversity monitoring. *Biol. Conserv.* **2016**, in press.
37. Neis, P.; Zielstra, D. Recent developments and future trends in volunteered geographic information research: The case of OpenStreetMap. *Future Internet* **2014**, *6*, 76–106. [[CrossRef](#)]



© 2016 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC-BY) license (<http://creativecommons.org/licenses/by/4.0/>).