

Using control data to determine the reliability of volunteered geographic information about land cover

Alexis Comber^{1*}, Linda See², Steffen Fritz², Marijn Van der Velde², Christoph Perger², and Giles Foody³

¹Department of Geography, University of Leicester, Leicester, LE1 7RH, UK

²International Institute of Applied Systems Analysis (IIASA), Laxenburg, A-2361
Austria

³School of Geography, University of Nottingham, Nottingham, NG7 2RD, UK

* corresponding author – email: ajc36@le.ac.uk

Abstract

There is much interest in using volunteered geographic information (VGI) in formal scientific analyses. This analysis uses VGI describing land cover that was captured using a web-based interface, linked to Google Earth. A number of control points, for which the land cover had been determined by experts allowed measures of the reliability of each volunteer in relation to each land cover class to be calculated. Geographically weighted kernels were used to estimate surfaces of volunteered land cover information accuracy and then to develop spatially distributed correspondences between the volunteer land cover class and land cover from 3 contemporary global datasets (GLC-2000, GlobCover and MODIS v.5). Specifically, a geographically weighted approach calculated local confusion matrices (correspondences) at each location in a central African study area and generated spatial distributions of user's, producer's, portmanteau, and partial portmanteau accuracies. These were used to evaluate the global datasets and to infer which of them was 'best' at describing Tree cover at each location in the study area. The resulting maps show where specific global datasets are recommended for analyses requiring Tree cover information. The methods presented in this research suggest that some of the concerns about the quality of VGI can be addressed through careful data collection, the use of control points to evaluate volunteer performance and spatially explicit analyses. A research agenda for the use and analysis of VGI about land cover is outlined.

Key words: VGI; accuracy; geographically weighted models; user's, producer's, Portmanteau and Partial Portmanteau accuracies;

1. Introduction

Volunteers have been observing and reporting information about environmental events and phenomena for a long time. One of the pioneers in this area was Robert Marsham, who in 1736, started to record the arrival of the first swallow at his home in Norfolk, England (The Guardian, 2011). The concept of *citizen science* emerged as information about a particular event but collected by many individuals was collated. In many cases the observation location was included, providing a geographical reference. Such geographically referenced observations have more recently been described as ‘volunteered geographical information’ (VGI) (Goodchild, 2007). This information is collected on a voluntary basis by interested individuals, frequently with no formal training. There are two broad strands of VGI available to the interested researcher. First, many historical datasets, often recording ‘natural’ phenomena such as phenological events, are being collated and made freely available. For example, data on the first leaf and first bloom dates of the common lilac in the USA as collated by Schwartz and Caprio (2003) are available to download (ftp://ftp.ncdc.noaa.gov/pub/data/paleo/phenology/north_america_lilac.txt). In the UK, spatio-temporal data describing different phenological events are held by The Woodland Trust, and can be viewed on their website (<http://www.naturescalendar.org.uk/>). The increased availability of such data is in part being driven by the need for public organisations to make their data holdings publicly available (Lister and the Climate Change Research Group, 2011). A second, more recent phenomenon is the availability of diverse information that is spatially referenced and can thus be considered as VGI. Information about all kinds of activities, in all kinds of formats, are contributed by members of the public to many

different hosting websites. The availability and collection of such data (see for example Haklay 2010; Coleman; 2010, Jones et al., 2012, van der Velde et al., 2012) is in part due to the many new ways in which the public can share information via the web, social networks, specific host sites (e.g. Flickr for photographs), and activities such as OpenStreetMap (Mooney and Corcoran, 2012). These are facilitated by the near-ubiquitous ability to capture data on location and to upload the information via many electronic devices (e.g. GPS enabled cameras, smartphones, electronic notebooks, etc). Thus, there is an increasing amount of spatially referenced or geo-located data available that could be used for formal scientific analyses.

The critical issue in the use of VGI relates to the quality of the information. In contrast to formal scientific experiments which include sampling design, training, data validation and some degree of scientific objectivity via the ‘designed experiment’ (e.g. Myers et al., 2010), in VGI there is no control over who records what, how they record it, or the quality of the information they provide (Hudson-Smith et al. 2009; Goodchild and Glennon, 2010; Haklay et al., 2010; Wiersma, 2010). As a result, recent research has sought to develop methods to assess the quality of VGI. For example, Brunsdon and Comber (2012) applied random coefficient modeling and bootstrapping approaches to overcome irregularities in the lilac data referred to above and suggested the need to consider the data collection methods before selecting the approach for model calibration. Haklay et al. (2010) and Tang and Lease (2011) suggested the use of multiple observations and observers to enable consensus-based data quality assessments and Foody and Boyd (2012) proposed a method for assessing the quality of VGI contributors using a latent class analysis of VGI in relation to land cover. The key point is that for VGI to be useful in scientific analyses there is a need

for some measure of its reliability, where ‘reliability’ here refers to the correctness or accuracy of the information. The need to quantify reliability in VGI is critical if this information is going to be used for scientific research. Without such measures there will always be a lack of trust or credibility in these data.

The Geo-Wiki project (www.geo-wiki.org) was developed at IIASA in collaboration with the University of Applied Sciences in Wiener Neustadt and the University of Freiburg (Fritz et al., 2012; Perger et al., 2012). Geo-Wiki is comprised of a web interface linked to Google Earth and different data collection campaigns have been launched with different aims. In these campaigns, volunteers are randomly provided with a series of predefined sample locations and asked to record what they observe at each location. The purpose of this paper is to develop a method for determining the reliability of the VGI about land cover collected during the first campaign, which was launched as the Human Impact Geo-Wiki (<http://humanimpact.geo-wiki.org>), and to explore how such information could be integrated into formal scientific analyses. A set of ‘control’ data points was used to generate measures of volunteer reliability. The accuracies of each volunteer in identifying each class were linked to the full dataset. Correspondences between volunteered land cover and 3 global land cover products (GLC-2000, GlobCover and MODIS v.5) were then generated. Reliability in this context refers to the per class correspondence measures (user’s, producer’s, portmanteau, and partial portmanteau accuracies) of each volunteer and their associated probabilities as outlined in Section 3, while control data are locations where experts have agreed on the land cover label or class. Spatially distributed measures of land cover correspondence (Comber et al., 2012; Comber, 2012),

weighted by volunteer reliabilities, were used to infer the most appropriate global dataset for describing tree cover at each location in a central African case study.

2. Background: Geo-Wiki and land cover accuracy

The initial Geo-Wiki system was established to encourage wide involvement in land cover validation (Perger et al., 2012). It incorporates a web-based interface using Google Earth and has become a modular system with different campaigns targeting specific objectives, such as validating cropland, urban areas and biomass (Fritz et al., 2009; 2011a; 2011b; 2012). Volunteers are recruited informally for different campaigns involving both the remote sensing community and the wider public. The opportunity to participate is completely open with no barriers to involvement. A campaign was undertaken to validate a map of land availability for biofuel production in the autumn of 2011 using the Human Impact Geo-Wiki. A random sample of locations was generated and 65 volunteers were recruited. They were a mixture of remote sensing experts, postgraduate students in related areas, other scientists and novices. On-going research is analysing the variation in reliability between the different volunteers, for example comparing experts with non-experts. Volunteers for the Human Impact study were asked to complete a short on-line tutorial to demonstrate the process (but this was not a requirement) and then to record the land cover at a series of locations. Based on their interpretation of the underlying satellite image / aerial photography they assigned each location to one of 10 predefined land cover classes. A critical feature of the Human Impact project was that the volunteers validated up to 299 control points – locations where experts had labelled the land cover. Control points were introduced randomly and 3 volunteers labelled all 299,

with each volunteer completing an average of 120 control points (s.d. 114). A full description of the Human Impact Geo-Wiki initiative is provided in Perger et al. (2012), but in brief, it was a targeted campaign that sought to gather data to validate a map of land available for biofuels. The control points were introduced to allow some assessment of the quality of information provided by contributors, and Perger et al. (2012) scored the participants on the number of points they contributed and the degree to which their evaluations matched the control points.

The capture and analysis of volunteered information including land cover through systems like Geo-Wiki, has the potential to provide valuable data for more formal scientific analysis, particularly in relation to land surface process such as land cover and land use. Land change is known to be a major variable, being for example a cause and a consequence of climate change, and presents the greatest threat to biodiversity (Feddema et al., 2005). However, there is considerable disagreement between different global land cover products regarding the amount and spatial distribution of land cover features particularly in relation to forest and cropland. For example, differences of as much as 20% have been found in the amount of land classified as arable or cropland when global land cover products have been compared (Fritz and See, 2005; See and Fritz, 2006; Fritz et al., 2011c). Thus the uncertainty in these products is so great that they cannot be used for global change detection. Formal approaches for analysing the reliability of land cover data have been developed (e.g. Strahler, 2006) but many land cover datasets are not validated using these protocols (Foody, 2002). The Geo-Wiki approach offers great potential to contribute to land cover validation (Iwao et al., 2006; Fritz et al., 2012).

Recent developments in land cover validation and accuracy reporting have proposed spatially explicit extensions to the standard method for reporting land cover accuracy based on the confusion matrix. Land cover errors are known to be spatially auto-correlated but the standard methods for describing them do not report their spatial distribution. Foody (2005) sought to address this and applied a kernel based approach to develop accuracy surfaces (e.g. user's accuracy for each class) using a fixed number of points under the kernel. Comber et al. (2012) extended Foody's approach and proposed geographically weighted models for describing spatial distributions of Boolean portmanteau and Fuzzy difference accuracies, where data points under the kernel were weighted by their distance from its centre. Portmanteau accuracy measures are described in more detail below. Comber (2012) further extended these methods and developed a geographically weighted confusion matrix model that generates spatial distributions of the error probabilities associated with user's and producer's accuracies. Such geographically weighted approaches for estimating local variations in accuracies provide a framework for the methods used in this research. In remote sensing validation, the confusion matrix is used to generate *global* correspondence measures between two datasets. In land cover validation, these are the land cover data being validated and some reference data considered to be of higher quality. In this research, *local* confusion matrices and correspondence measures were calculated using a geographically weighted kernel at regular intervals throughout the study area. The results are spatially distributed measures of correspondence and their associated probabilities, which can be mapped.

3. Methods

The approach taken was:

- 1) To use the control data to calculate measures of each volunteer's per class reliabilities, namely correspondences based on user's, producer's and portmanteau accuracies and a further partial portmanteau measure.
- 2) To apply these measures to the full volunteered land cover dataset, collected by the Human Impact Geo-Wiki initiative such that each data point had a measure of reliability depending on the volunteer and the class they indicated.
- 3) To compare volunteered land cover information with the GLC-2000, MODIS and Globcover data and to calculate geographically weighted measures of correspondence.
- 4) To infer the most appropriate global dataset in each location for a specific land cover class.

3.1 Data

The Human Impact project asked volunteers to classify the land cover into one of ten land cover classes: (1) Tree cover, (2) Shrub cover, (3) Herbaceous vegetation / Grassland, (4) Cultivated and managed, (5) Mosaic of cultivated and managed / natural vegetation, (6) Flooded / wetland, (7) Urban, (8) Snow and ice, (9) Barren and (10) Open Water. These class numbers are used in subsequent tables.

The project team at IIASA identified the land cover class at 299 control locations introduced randomly, against which the volunteer land cover classes were compared. In total, 7657 control records were provided by 65 volunteers. The control data were

filtered to exclude volunteers who had contributed less than 20 data points or where the land cover had not been recorded by the volunteer for some reason. This was to ensure sufficient data to enable reliable volunteer-specific accuracy measures to be calculated. The end result was a control dataset of 6906 records contributed by 47 volunteers, describing the land cover at 299 locations. This was used to characterise the reliabilities of each volunteer in identifying each land cover class. The locations of the control points are shown in Figure 1.

(insert figure 1 about here)

The calculation of the volunteer per class reliabilities is described and illustrated below. These were attached to the full Human Impact dataset based on the user that recorded the data point and the land cover class they allocated. The dataset contained 42,474 records after filtering for the 47 volunteers who contributed more than 20 validation points and for whom robust reliabilities could be calculated. Here, a central African case study was selected to illustrate the results. The case study area was defined by an arbitrary bounding box for the area shown in Figure 2 and contained 5,966 data points labelled by 47 volunteers.

(insert figure 2 about here)

3.2 Volunteer reliability

The correspondence between control and volunteer land cover classes was used to generate 4 measures describing the reliability of each volunteer, in relation to each class. Table 1 describes how the components of the different correspondence measures are calculated from a collapsed confusion matrix.

(insert Table 1 about here)

The 4 reliability measures were:

- 1) User's accuracy: $n_1 / (n_1 + n_2)$
- 2) Producer's accuracy: $n_1 / (n_1 + n_3)$
- 3) Portmanteau accuracy: $(n_1 + n_4) / (n_1 + n_2 + n_3 + n_4)$
- 4) Partial portmanteau accuracy: $(n_1) / (n_1 + n_2 + n_3)$

User's and producer's accuracies are described Congalton (1991). The portmanteau measure reflects whether a volunteer has correctly recorded the presence or absence of each class (Comber et al., 2012). It accommodates both the specificity and the sensitivity of the data. Sensitivity measures the proportion of the actual land cover class (i.e. positive identifications) that is correctly identified and specificity measures the proportion of negatives that is correctly identified. A second partial portmanteau measure of correspondence was generated from only those points where either the control or the contributor indicated the presence of the class under consideration.

3.3 Surfaces of volunteer reliability

Spatially distributed correspondence surfaces were generated from the user's, producer's, portmanteau and partial portmanteau measures described above using a geographically weighted kernel. This is a spatial interpolation process that uses a moving window to compute geographically weighted values at each point in a predefined set of locations, in this case spaced at 100km apart (see Figure 2), with data points that are further away from the specific location under consideration contributing less to the computation. The spacing was selected arbitrarily and reliability measures could have been computed over a finer or coarser distribution of locations. The weight, w_i , associated with each data point location (u_i, v_i) is a decreasing function of d_i , the distance from the centre of the window to (u_i, v_i) :

$$w_i = \begin{cases} (1 - \frac{d_i^2}{h^2})^2 & \text{if } d_i < h \\ 0 & \text{otherwise} \end{cases} \quad (\text{Eqn 1})$$

where h is known as the bandwidth and is specified in whatever map units are being used or as a proportion of data points. In this way, the weights associated with each location change depending on the location for which an accuracy probability is to be calculated. The bandwidth (or kernel size) may be varied to ensure that enough data points are used in calibration to minimise the cross validation prediction error (Comber et al. 2012). The number of data points is a trade-off between working with a dataset that is too small to calibrate the local model reliably, and too big to avoid averaging out local effects. In this research the bandwidth was optimised automatically using a leave-one-out cross validation procedure. This determines a bandwidth that optimises the prediction probability for each individual volunteer

value when it was removed from the dataset. Further details on bandwidth selection are in Fotheringham et al. (2002).

3.4 Linking to global land cover data

Land cover classes from three global datasets were compared with the volunteered land cover class: MODIS, GLC-2000 and GlobCover. The GLC-2000 map was developed with 14 months of satellite data from the VEGETATION instrument on board the SPOT 4 satellite. Regional maps were based on the Land Cover Classification System (LCCS) of FAO (Di Gregario and Jansen, 2000) and then harmonised to produce a global product with 22 classes (Fritz *et al.*, 2003). The project was coordinated by the Global Vegetation Monitoring Unit of the Joint Research Centre (JRC) of the European Union (Fritz *et al.*, 2003). The MODIS land cover product from Boston University is available at a resolution of 500m and uses the 17 classes of the IGBP (International Global Biosphere Project) legend (Loveland et al., 1998). The product was created using the Moderate Resolution Imaging Spectroradiometer instrument on the NASA Terra Platform using an automatic supervised classification method (Morisette et al., 2002; Friedl et al., 2010). Version 5 of the MODIS land cover data set is used in this paper. The recent release of GlobCover (Bicheron et al., 2008) for the reference year 2005 is the highest resolution global land cover product available (c. 300m×300m at the equator). This was developed by the European Space Agency and a number of partners including the Joint Research Centre of the European Union and the Catholic University of Louvain. The first version of the land cover product used in this paper was based on an

automatic processing chain using MERIS FR time series data from December 2004 to June 2006 (Bicheron et al., 2008).

The land cover class from each of the global datasets was extracted for each location at which volunteered information on land cover was described. For each of the global datasets the land cover class was aggregated into one of the 10 classes described above. This was to provide a common framework to evaluate the contributor land cover classes. The aggregation look up tables were devised by an expert who had worked with all of the datasets, had been involved in their creation and who was familiar with them, their nomenclatures, their underlying semantics and critically how they varied from each other in the way that they describe land cover. Table 2 describes the look up tables for these aggregations. The final dataset used for the analysis contained 4 land cover classes: 3 from the global land cover datasets and the volunteered land cover class. It also included measures of user's, producer's, partial portmanteau and portmanteau correspondences for that volunteer in identifying that class, which were outlined in Section 3.2.

(insert table 2 about here)

3.5 Comparing volunteered and global land cover

The next step in the analysis was to compare the volunteered land cover classes with global classes recorded by the GLC-2000, MODIS and GlobCover. The volunteer land cover data was used as the dependent variable in a binomial geographically

weighted regression (GWR) considering each land cover class in turn. The analysis regressed data indicating the presence (1) or absence (0) of a particular land cover class in both the global and volunteered datasets. The logit transform generated probability estimates from the regression coefficients of the degree to which the volunteered land cover class was predicted by the global dataset – i.e. the volunteer land cover was used as the reference data. Each land cover class considered in turn, but the results describe the analysis of the (1) Tree cover class.

A *logit* function was defined by:

$$\text{logit}(Q) = \frac{\exp(Q)}{1 + \exp(Q)} \quad (\text{Eqn 2})$$

The logistic geographically weighted regressions were then calculated as follows for each global land cover dataset:

$$\text{pr}(y_i = 1) = \text{logit}(b_{0(u_i, v_i)} + b_1 x_{1(u_i, v_i)}) \quad (\text{Eqn 3})$$

where $\text{pr}(y_i = 1)$ is the probability that the volunteer class of Tree cover is present, and x_1 is the explanatory or independent variable (the land cover class from the global dataset). The coefficient estimates are assumed to vary across the two-dimensional geographical space defined by the coordinates (u, v) and can be considered as functions of these coordinates, rather than constants as in a global regression. The coefficient estimates arising from Equation 3 were determined for each point in the predefined set of locations spaced at 100km apart (as described above and shown in

Figure 2). From these, the spatially distributed probabilities associated with the correspondence measures as described in Section 3.2 were calculated as follows:

User's accuracy:

$$pr(y=1|x=1) / (pr(y=1|x=1)+pr(y=0|x=1))$$

Producer's accuracy:

$$pr(y=1|x=1) / (pr(y=1|x=1)+pr(y=1|x=0))$$

Portmanteau:

$$pr(y=1|x=1) / (pr(y=1|x=1)+pr(y=0|x=1))+pr(y=1|x=0)+pr(y=0|x=0))$$

Partial portmanteau:

$$pr(y=1|x=1) / (pr(y=1|x=1)+pr(y=0|x=1))+pr(y=1|x=0))$$

where x is the predicted class in the global land cover data, y is the volunteered class and a value of 0 denotes absence of that class and 1 denotes presence, for both x and y . Note, that as the statistical relationships between x and y are generated from Equation 3, then the probabilities above will have values for each location over the geographic space (u, v) .

4. Results

The confusion matrix between control and volunteer land cover classes is shown in Table 3. The overall accuracy is 62% and it should be noted that classes 6 and 8 are absent in the control dataset.

(insert Table 3 about here)

For each volunteer and for each class, 4 correspondence measures were calculated. These have potential ranges of $[0,1]$ and 3 examples of volunteer reliability measures are shown in Table 4. It is evident that portmanteau is higher than both user's and producer's correspondences and partial portmanteau is lower. The correspondences for each volunteer were linked to the full Human Impact dataset by land cover class. So for example, in the VGI dataset, each record in the full dataset that was contributed by Volunteer A and scored as (1) Tree cover, had a user's value of 0.727, a producer's value of 0.833, a portmanteau value of 0.923 and partial portmanteau value of 0.635 attached to it.

(insert Table 4 about here)

The spatial variations in volunteer accuracies for all classes were estimated using a geographically weighted kernel and are shown in Figure 3. These are surfaces of the weighted means of the correspondence measures, whose weights use the same kernel as defined in Equation 1 in a similar manner to GWR – see for example Brunsdon et al. (2002). An optimal bandwidth was identified and it included 2.1% of the data

points nearest to each location. Figure 3 includes the location of the volunteer validation points with the size of the circles indicating the magnitude of their correspondence. Figure 4 shows the geographically weighted surfaces of mean correspondences for only the class of (1) Tree cover using a kernel bandwidth that included 0.8 % of the data points. For Figures 3 and 4, the geographically weighted kernel described in Equation 1 was used to interpolate the measures of volunteer reliability for all classes (Figure 3) and for just Tree cover (Figure 4). It is worth considering how the surfaces in Figures 3 and 4 might be used in relation to their derivation. If the control data are correct, then:

- User's accuracy or correspondence describes the degree to which the land cover data contributed by the volunteer is the same as the control land cover (how many times the volunteer got it right);
- Producer's accuracy or correspondence describes the probability of the control data being correctly identified by the volunteer (how much of the control land cover was correctly identified);
- Portmanteau accuracy or correspondence describes the combined probability that the volunteer correctly identified the control land cover when it was present and its absence when it was not;
- Partial portmanteau accuracy or correspondence describes the probability that the volunteer correctly identified the land cover when it was present in the control or their predictions.

(insert Figure 3 about here)

(insert Figure 4 about here)

Figures 5, 6, 7 and 8 show the different types of geographically weighted correspondence between volunteer land cover and land cover from each of the global datasets, for the class of (1) Tree cover. They are accompanied by a map describing which of the global datasets has the highest correspondence value in each location, along with contours describing the variation in the reliability of the volunteered information in that class. Land cover accuracy is known to vary spatially. The models used to generate the maps explicitly analyse spatial variations in accuracy and extend the work of Foody (2005) and Comber et al. (2012). These are in contrast to the usual measures for reporting correspondence which are aspatial, such as the standard confusion matrix. The results show that for the class of (1) Tree cover:

- 1) User's correspondences are higher for the GLC-2000 and GlobCover than for MODIS;
- 2) Correspondences with MODIS are stronger than the other global datasets when producer's accuracies are considered;
- 3) MODIS portmanteau correspondences are higher overall than GLC-2000 or GlobCover;
- 4) MODIS partial portmanteau correspondences are higher than GLC-2000 or GlobCover.

In each case, the spatial variations in these trends show where one dataset may be preferred over others. However, this preference will depend on the correspondence measure required for the particular analysis or hypothesis being tested. For example, in a situation where it is important to determine how much of the actual land cover was correctly identified, then correspondence described by the producer's accuracy may be preferred, and for a study requiring the best Tree cover information, MODIS may be selected as the preferred dataset. Alternatively, for a study in Cameroon, the

area in the Southwest of the study area, GLC-2000 data may be preferred. Or, for another study, a composite land cover dataset could be constructed from the 3 global datasets. Framing the study objectives in relation to the probabilistic logic of the different types of accuracy or correspondence (user's, producer's, portmanteau, partial portmanteau) allows the most suitable dataset, or dataset combinations, to be chosen. The differences between the correspondences with global land cover data could be tested for significance, but the objective here is to illustrate how geographically weighted methods for analysing volunteered land cover data, parameterised by volunteer reliabilities, could be used to explore the spatial variations in relationships and correspondences.

(insert Figure 5 about here)

(insert Figure 6 about here)

(insert Figure 7 about here)

(insert Figure 8 about here)

5. Discussion

This analysis evaluated the quality of the land cover information provided by volunteers through the incorporation of a set of control locations where the land cover was known. By comparing volunteered land cover with the control data, volunteer reliabilities were calculated, in this case using correspondence measures derived from the confusion matrix, the classic approach for reporting accuracy in remote sensing. The collection of control data – that is, locations where the land cover recorded by the volunteer could be compared with the land cover class determined by experts – via

the Geo-Wiki project allowed measures of volunteer reliability to be determined. Measures of the volunteer accuracies for each class calculated in this way were linked to the data points in the full global dataset. Geographically weighted models were used to analyse the spatial variation in the volunteer accuracies and to compare volunteer land cover classes with land cover from 3 global datasets. A geographically weighted kernel was used to construct surfaces of user's, producer's, portmanteau and partial portmanteau correspondences from the control data. The volunteered land cover classes were then compared to global land cover datasets in order to determine the correspondences between VGI land cover and the land cover as recorded in 3 global datasets. Measures of user's, producer's, portmanteau and partial portmanteau accuracies were calculated using a logistic geographically weighted regressions of the volunteered land cover and land cover classes from the global datasets. The spatial distributions of these correspondence measures were mapped. The highest correspondence was used to infer which global dataset may be most appropriate at each location in the study area, and contours of user reliabilities were overlaid. The results indicate how VGI on land cover, parameterised by some control measure, could be incorporated into formal scientific analyses.

The analysis could have included additional steps, for example exploring the thresholds of user reliability and combinatorial approaches to indicate compound probabilities, which may change the inferences made about the most appropriate global dataset (as in Figures 5 to 8). These will be explored in future work as the main purpose of this research was to develop generic approaches for evaluating VGI on land cover using the data collected under a semi-formal structure such as is afforded by the Geo-Wiki approach. This analysis has only described the class of (1) Tree

cover in detail. Similar data, reliabilities and spatial correspondences were generated for the other classes using the different probabilities associated with the correspondence matrix, but space limits their inclusion. The use of geographically weighted models to describe spatial variations in accuracy and correspondence between different sources of land cover information (in this case land cover from VGI and global datasets) provides more informative measures of correspondence than simple confusion or correspondence matrices: they generate probabilities which can be mapped whereas static correspondence measures do not (Foody, 2005; Comber et al., 2012; Comber 2012). Spatially distributed measures of accuracy are more informative because they show how and where error rates and correspondences vary.

The measures of correspondence between VGI and global land cover data could be seen to be predicated on the assumption that the VGI is correct. However, this would be a naïve assumption as this research has generated correspondences commonly applied in remote sensing accuracy assessments, extended spatially using a geographically weighted kernel, as a way of generating comparison metrics. The spatially distributed measures of correspondence do however provide helpful information to aid the interpretation of both the VGI and the global land cover datasets, but the results of this research indicate the need for an informed ‘interpretation’ of the numerical correspondences to include allowances for variations in data quality.

The use of control points allowed VGI reliabilities to be determined and these in turn were used as confidence measures to accompany the inferences made about the suitability of different global land cover datasets. Such approaches can support user

choices about which data to use for a particular location and application. For example, having a composite dataset of forest, comprising the ‘best’ land cover data from various global datasets, may be important for global climate modelling. Additionally, the use of control points suggests that one possible way to overcome the reliability issues inherent in VGI is to set up formal sampling structures within geo-wikis.

There are some assumptions and limitations associated with the research. These include possible problems with the control data, where some of them may be less reliable and of poorer quality than others. Future work will explore the control data in more detail and the impact that variations in the quality and reliability of the control data have on the results. Additionally there are some limitations associated with the use of the global data in this way. They were collapsed into 10 classes, introducing further uncertainty into the results of the analyses. They were produced at different scales and the ‘scale’ at which the volunteer identified the land cover at each location did not use the same pixel frame as the global datasets. These issues are currently being investigated in on-going work. However, this research does present an approach for analysing the quality of volunteered land cover information and for integrating those into wider analyses using spatially distributed measures of accuracy and correspondence (Foody, 2005; Comber, 2012 and Comber et al., 2012).

This research has suggested a number of areas for future work, explicitly in relation to volunteered information on land. There is a need for research in the following broad areas:

- Comparison of different measures of user reliability using geographically weighted models, for instance the latent class model measures proposed by Foody and Boyd (2012) with the measures derived from the control points.
- Analysis of volunteer performance in relation to i) perceptions of scale, ii) their ability to discern different granularities of information; iii) how their reliability changes with the physical and experiential distances from the location being considered. Understanding these potential biases will allow more nuanced thresholds of user reliability to be determined and potentially may be part of a structured test that establishes a volunteer's characteristics before they participate in the program.
- The use of different formal methods for combining uncertain evidence (e.g. Fuzzy sets, Dempster-Shafer, Possibility and Endorsement theories) in order to develop measures of belief in volunteered information rather than just using a Linus Law approach. This is a strong and longstanding domain in informatics that has yet to be explored in relation to VGI.
- Investigation of the different types of decision about land cover data that are supported by the outputs of different uncertainty formalisms, particularly in relation to the acceptability of the information.

6. Conclusions

Many volunteer activities generate spatially referenced information. The key issue relating to the use of volunteered geographical information or data in scientific research concerns its unknown quality and the errors associated with using it in any specific analysis. This research has shown that VGI about land cover can be used in

formal analyses when it is linked to control data (locations where the land cover was known). The collection of control data via the Geo-Wiki project allowed the quality of volunteered land cover information to be parameterised, and geographically weighted models were used to generate surfaces of reliability, based on different kinds of accuracy measures. Generating geographically weighted models of the correspondences between VGI and 3 global land cover datasets allowed spatially explicit inferences regarding the most suitable global dataset to be made.

The key conclusions of this work are:

- Accuracies can be attached to VGI if some of it can be cross-referenced with control data;
- Spatially explicit methods for calculating correspondences, such as geographically weighted models, allow different surfaces of correspondence or accuracy to be generated (user's, producer's, portmanteau, partial portmanteau);
- VGI on land cover analysed in this way can be used to select the most appropriate datasets from a set of competing choices, in this case the GLC-2000, GlobCover and MODIS v.5 global land cover products, and to make inferences about their reliability in describing specific land cover classes at different locations.

On-going activities are extending this research area in a number of directions.

Acknowledgements

Alexis Comber would like to thank IIASA for hosting him during his visit to conduct this research, the Centre for Landscape and Climate Research and the University of

Leicester for a short period of study leave. All of the statistical analysis and mapping were implemented in R version 2.15.1, the open source statistical software <http://cran.r-project.org> , using the *spgwr* and *GISTools* libraries. The data and code used in this analysis will be provided to interested researchers on request.

References

- Bicheron, P., P. Defourny, C. Brockmann, L. Schouten, C. Vancutsem, M. Huc, S., Bontemps, M. Leroy, F. Achard, M. Herold, F. Ranera, and O. Arino (2008), *GlobCover: Products Description and Validation Report*, 18, Toulouse, France. URL: http://ionia1.esrin.esa.int/docs/GLOBCOVER_Products_Description_Validation_Report_I2.1.pdf
- Coleman D (2010). The potential and early limitations of volunteered geographic information. *Geomatica*, 64(2): 27–39
- Comber, A., Fisher, P.F., Brunsdon, C. and Khmag, A. (2012). Spatial analysis of remote sensing image classification accuracy. *Remote Sensing of Environment*, 127: 237–246.
- Comber A.J., (2012). Geographically weighted models for determining local confusion matrices: surfaces of user's and producer's accuracy. Paper accepted for publication in *Remote Sensing Letters*.
- Congalton, R.G. (1991). A review of assessing the accuracy of classifications of remotely sensed data, *Remote Sensing of Environment*, 37(1), 35-46.

- Feddema, J.J., Oleson, K.W., Bonan, G.B., Mearns, L.O., Buja, L.E., Meehl G A, Washington W M (2005) The importance of land-cover change in simulating future climates, *Science* 310:1674-1678.
- Foody, G. M. (2005). Local characterization of thematic classification accuracy through spatially constrained confusion matrices, *International Journal of Remote Sensing*, 26, 1217-1228
- Foody, G. M. and Boyd, D. S., (2012), 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, 203-208.
- Foody, G.M. (2002). Status of land cover classification accuracy assessment, *Remote Sensing of Environment*, 80, 185-201.
- Friedl M., Sulla-Menashe, D., Tan, B., Schneider, A., Ramankutty N., Sibley, A. and Huang, X. (2010). MODIS Collection 5 global land cover: algorithm refinements and characterization of new datasets. *Remote Sensing of Environment*, 114, 168–82
- Fritz, S. and See, L. (2005), Comparison of land cover maps using fuzzy agreement. *International Journal of Geographical Information Science*, 19(7), 787-807.
- Fritz, S., Bartholome , E., Belward, A., Hartley, A., Stibig, H.J., Eva, H., Mayaux, P., Bartalev, S., Latifovic, R., Kolmert, S., Roy, P., Agrawal, S., Bingfang, W., Wenting, X., Ledwith, M., Pekel, F.J., Giri, C., Mucher, S., De Badts, E., Tateishi, R., Champeaux, J-L. and Defourny, P., (2003). *Harmonisation, Mosaicing and Production of the Global Land Cover 2000 Database (Beta*

Version) (Luxembourg: Office for Official Publications of the European Communities).

- Fritz, S., McCallum, I., Schill, C., Perger, C., Grillmayer, R., Achard, F., Kraxner, F. and Obersteiner, M. (2009). Geo-Wiki.Org: The use of crowd-sourcing to improve global land cover. *Remote Sensing*, 1 (3), 345-354
- Fritz, S., See, L., McCallum, I., Schill, C., Perger, C. and Obersteiner, M. (2011a), Building a crowd-sourcing tool for the validation of urban extent and gridded population. *Lecture Notes in Computer Science*, 6783, 39-50.
- Fritz, S., Schepaschenko, D., McCallum, I., Perger, C., Schill, C., Obersteiner, M., Baccini, A., Gallaun, H., Kindermann, G., Kraxner, F., Saatchi, S., Santoro, M., See, L., Schmullius, C. and Shvidenko, A. (2011b), Observing terrestrial biomass globally: <http://Biomass.Geo-wiki.org>. AGU Fall Meeting 2011. San Francisco, 5-9 Dec 2011.
- Fritz, S., See, L., McCallum, I., Schill, C., Obersteiner, M., van der Velde, M., Boettcher, H., Havlik, P. and Achard, F. (2011c), Highlighting continued uncertainty in global land cover maps to the user community. *Environmental Research Letters*, 6, 044005.
- Fritz S, McCallum I, Schill C, Perger C, See L, Schepaschenko D, van der Velde M, Kraxner F and Obersteiner M., (2012). Geo-Wiki: An online platform for improving global land cover. *Environmental Modelling and Software*, 31: 110-123.
- Goodchild M.F. (2007). Citizens as sensors: the world of volunteered geography. *Geojournal* 69: 211-221.

- Goodchild M F and Glennon J A 2010 Crowdsourcing geographic information for disaster response: a research frontier. *International Journal of Digital Earth* 3: 231-241.
- Haklay M Basiouka S Antoniou V and Ather A 2010 How many volunteers does it take to map an area well? The validity of Linus' law to volunteered geographic information. *Cartographic Journal* 47: 315-322.
- Hudson-Smith A Batty M Crooks A and Milton R 2009 Mapping for the masses. Accessing Web 2.0 through crowdsourcing. *Social Science Computer Review* 27: 524-538.
- Jones, C. E., Mount, N. J. and Weber, P. (2012). The Rise of the GIS Volunteer. *Transactions in GIS*, 16:431-434
- Iwao K Nishida K Kinoshita T and Yamagata Y 2006 Validating land cover maps with Degree Confluence Project information. *Geophysical Research Letters* 33: L23404, doi:10.1029/2006GL027768.
- Lister M Adrian, the Climate Change Research Group (2011) Natural history collections as sources of long-term datasets. *Trends Ecol Evol* 26(4):153–154
- Loveland, T.R., Reed, B.C., Brown, J.F., Ohlen, D.O., Zhu, Z., Yang, L. and Merchant, J.W. (2000) Development of a global land cover characteristics database and IGBP DISCover from 1 km AVHRR data. *International Journal of Remote Sensing*, 21(6 & 7), 1303-1330.
- Mooney, P. and Corcoran, P. (2012). The Annotation Process in OpenStreetMap. *Transactions in GIS*, 16: 561-579.
- Morisette JT, Privette JL and Justice CO 2002. A framework for the validation of MODIS land products. *Remote Sensing of Environment*, 83: 77–96.

- Myers JL, Well A, Lorch RF (2010). *Research design and statistical analysis*, 3rd edn. Routledge, New York
- Perger C, Fritz S, See L, Schill C, Van der Velde M, McCallum I and Obersteiner M
2012 A campaign to collect volunteered geographic Information on land cover and human impact. In: Jekel T, Car A, Strobl J and Griesebner G (Eds.) *GI_Forum 2012: Geovizualisation, Society and Learning*. Herbert Wichmann Verlag, VDE VERLAG GMBH, Berlin/Offenbach, pp.83-91.
- Schwartz M, Caprio J (2003) *North American first leaf and first bloom lilac phenology data*. IGBP PAGES/World Data Center for Paleoclimatology Data; Contribution Series # 2003-078; NOAA/NGDC Paleoclimatology Program, Boulder CO, USA
- See, L. and Fritz, S. (2006), Towards a global hybrid land cover map for the year 2000. *IEEE Transactions on Geosciences and Remote Sensing*, 44: 1740-1746.
- Strahler A H, Boschetti L, Foody G M, Friedl M A, Hansen M C, Herold M, Mayaux P, Morisette J T, Stehman S V and Woodcock C E 2006 *Global Land Cover Validation: Recommendations for Evaluation and Accuracy Assessment of Global Land Cover Maps*, European Commission, Joint Research Centre, Ispra, Italy, EUR 22156 EN, 48pp.
- Tang W and Lease M 2011 Semi-supervised consensus labeling for crowdsourcing. In: Proceedings of the *SIGIR 2011 Workshop on Crowdsourcing for Information Retrieval*, July 28, 2011, Beijing, China.
- The Guardian (2011). *Weatherwatch: phenology in the UK*,
<http://www.guardian.co.uk/news/2011/apr/11/weatherwatch-phenology>

van der Velde, M., See, L., Fritz, S., Verheijen, F.G.A., Khabarov, N. and Obersteiner,

M. (2012). Generating crop calendars with Web search data. *Environmental Research Letters*, 7, 024022, doi:10.1088/1748-9326/7/2/024022

Wiersma Y F 2010 Birding 2.0: Citizen science and effective monitoring in the web

2.0 world. *Avian Conservation and Ecology* 5: 13 [http://www.ace-](http://www.ace-eco.org/vol5/iss2/art13/)
[eco.org/vol5/iss2/art13/](http://www.ace-eco.org/vol5/iss2/art13/)

<http://dx.doi.org/10.5751/ACE-00427-050213>

	Control true	Control false
Volunteered true	n_1	n_2
Volunteered false	n_3	n_4

Table 1. The 2-class error matrix used to calculate different accuracy measures.

Class	GLC-2000	MODIS	GlobCover
<i>(1) Tree cover</i>	1 to 10	1 to 5, 8, 9	40, 50, 60, 70, 90, 100, 110, 160, 170
<i>(2) Shrub cover</i>	11, 12	6, 7	130
<i>(3) Herbaceous / Grassland</i>	13	10	120, 140
<i>(4) Cultivated / Managed</i>	16	12	11, 14
<i>(5) Mosaic of cultivated & natural</i>	17, 18	14	20, 30
<i>(6) Flooded / wetland</i>	15	11	180
<i>(7) Urban</i>	22	13	190
<i>(8) Snow and ice</i>	21	15	220
<i>(9) Barren</i>	14, 19	16	150, 200
<i>(10) Open Water</i>	20	17	210

Table 2. The aggregation of global land cover classes.

		Control Class										User
		1	2	3	4	5	6	7	8	9	10	
Volunteered Class	1	800	69	47	8	223	0	0	0	0	0	0.697
	2	245	294	97	24	113	0	0	0	0	0	0.380
	3	112	99	279	61	154	0	0	0	1	0	0.395
	4	19	8	21	1689	226	0	0	0	0	0	0.860
	5	99	20	72	548	1141	0	0	0	0	0	0.607
	6	3	12	16	8	11	0	0	0	0	1	0.000
	7	4	0	0	30	14	0	26	0	0	0	0.351
	8	3	0	0	1	2	0	0	0	6	0	0.000
	9	19	38	98	21	43	0	0	0	51	0	0.189
	10	0	0	1	1	1	0	0	0	0	27	0.900
Producer		0.613	0.544	0.442	0.706	0.592	0	1	0	0.879	0.964	

Table 3. The correspondence matrix between control and volunteered land cover data, with User and Producer accuracies.

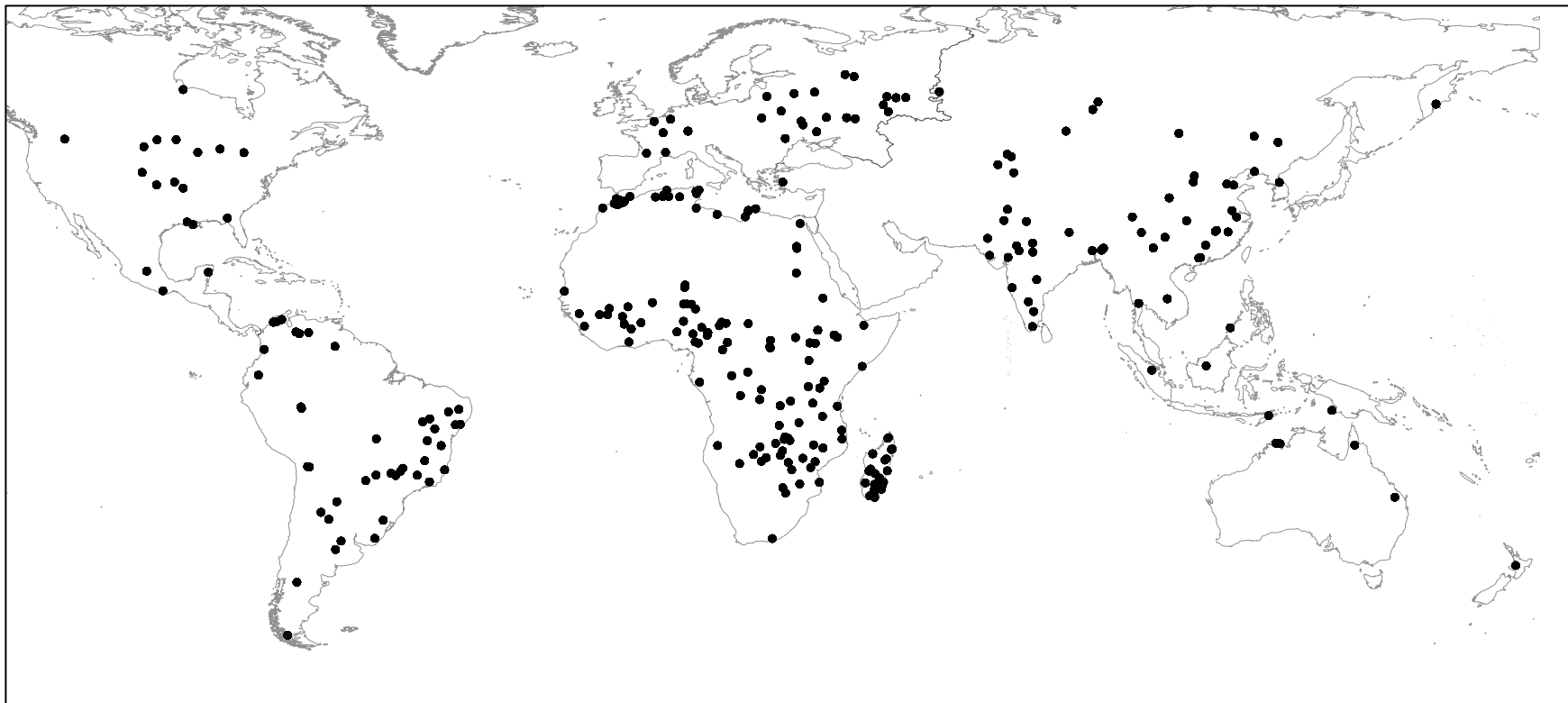


Figure 1. The location of the 299 control points.



Figure 2. The case study area and the location of the 5966 data points, with the locations at which the geographically weighted measures were computed.

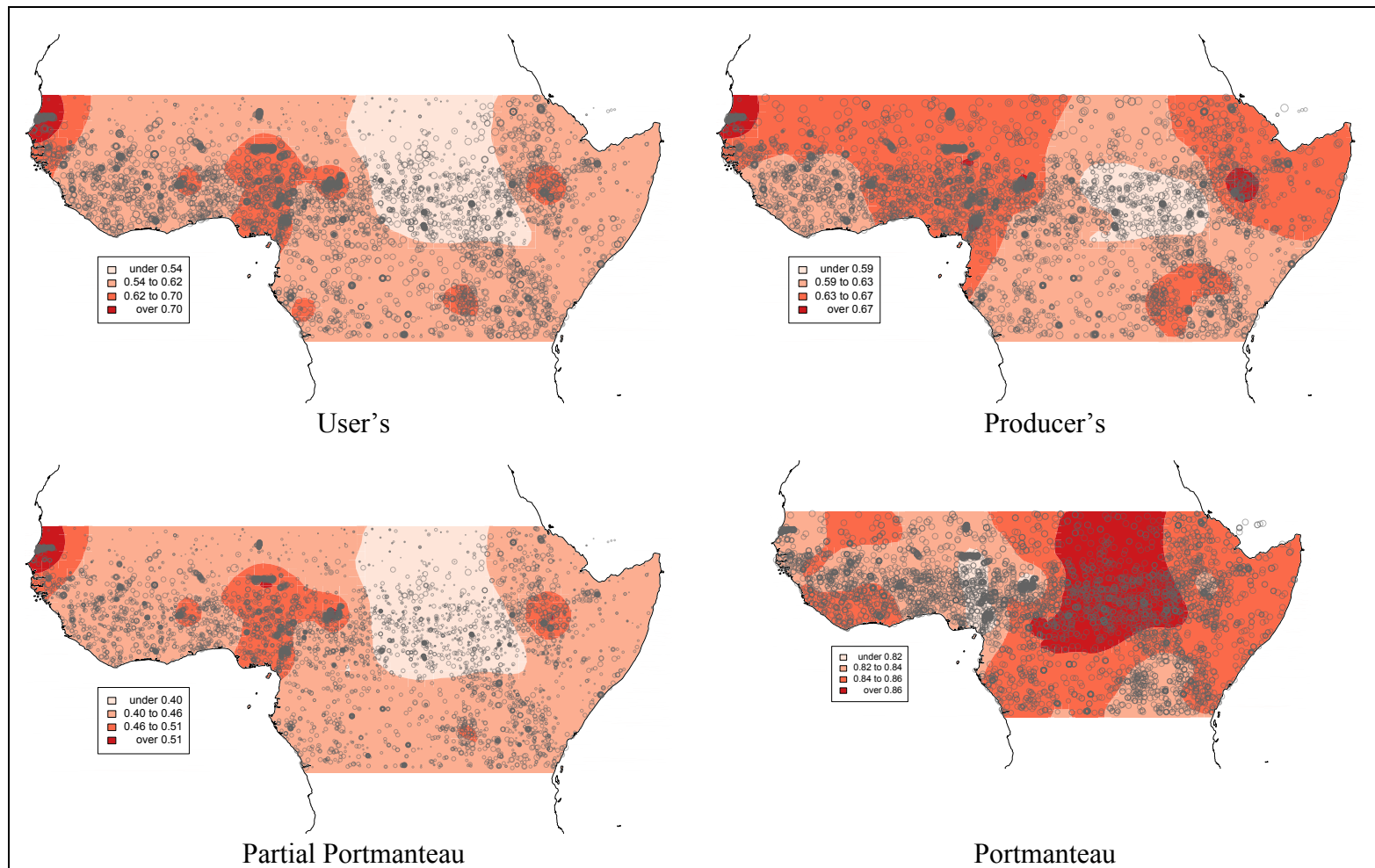


Figure 3. Surfaces of mean volunteer reliabilities estimated from the correspondences of all classes. The circles show the location of the volunteer validation points and their size indicates volunteer accuracies.

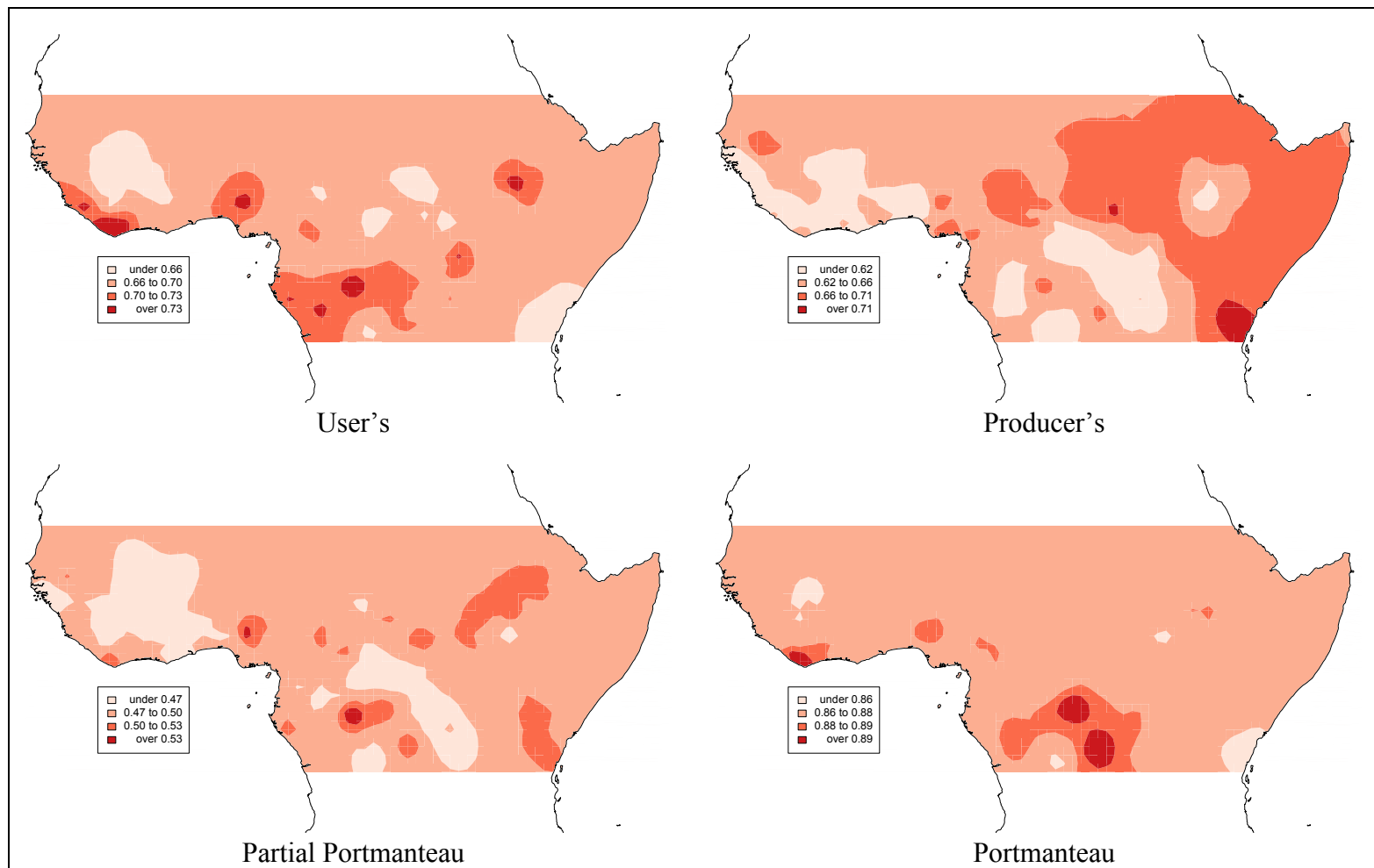


Figure 4. Surfaces of volunteer mean reliabilities estimated for the class of (1) Tree cover.

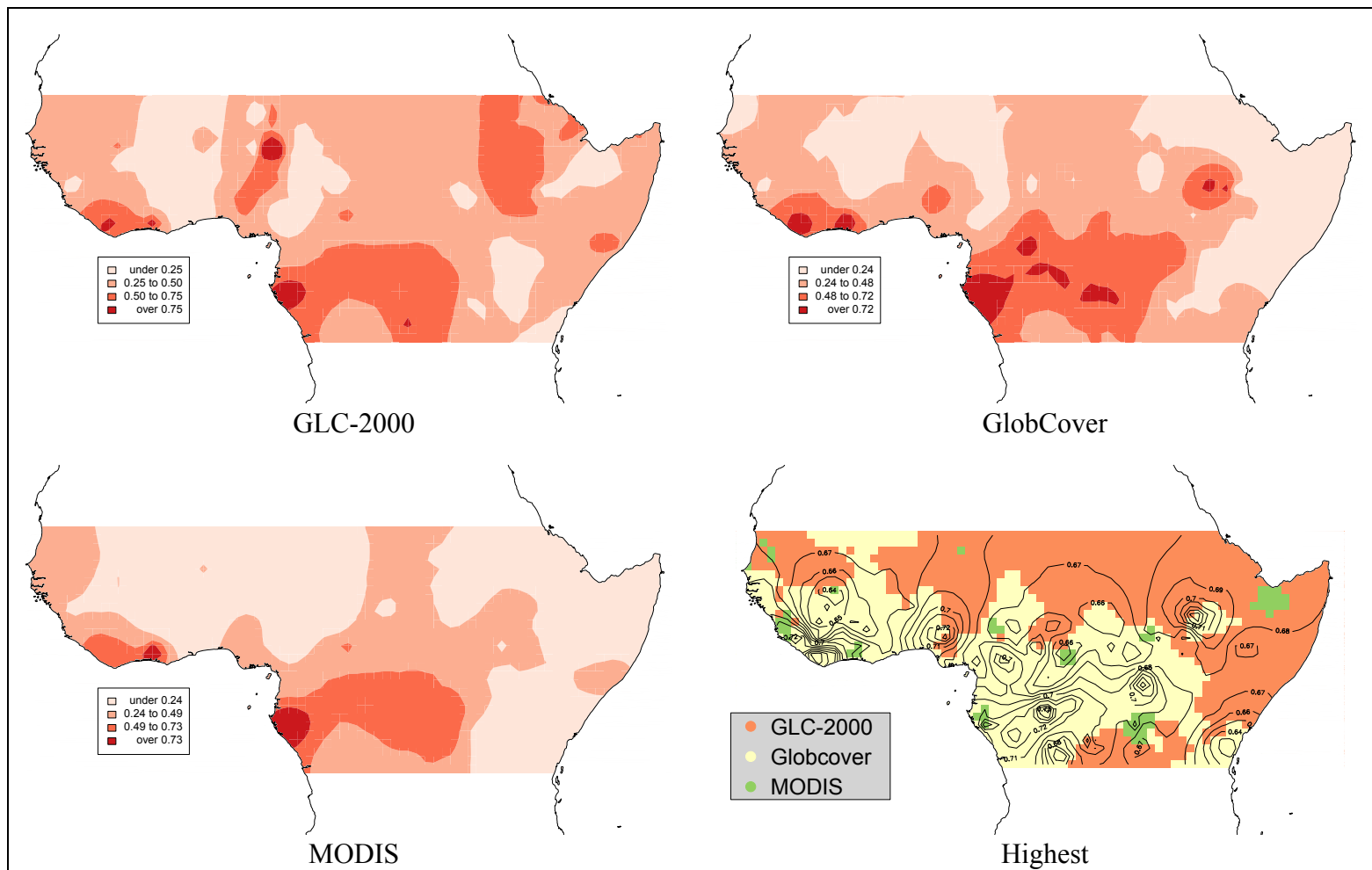


Figure 6. User's correspondences for Tree cover derived from comparisons between volunteered land cover and global datasets, with a map of the global datasets with the highest user's value at each location, overlaid with contours of the geographically weighted variation in the user's accuracy for Tree cover.

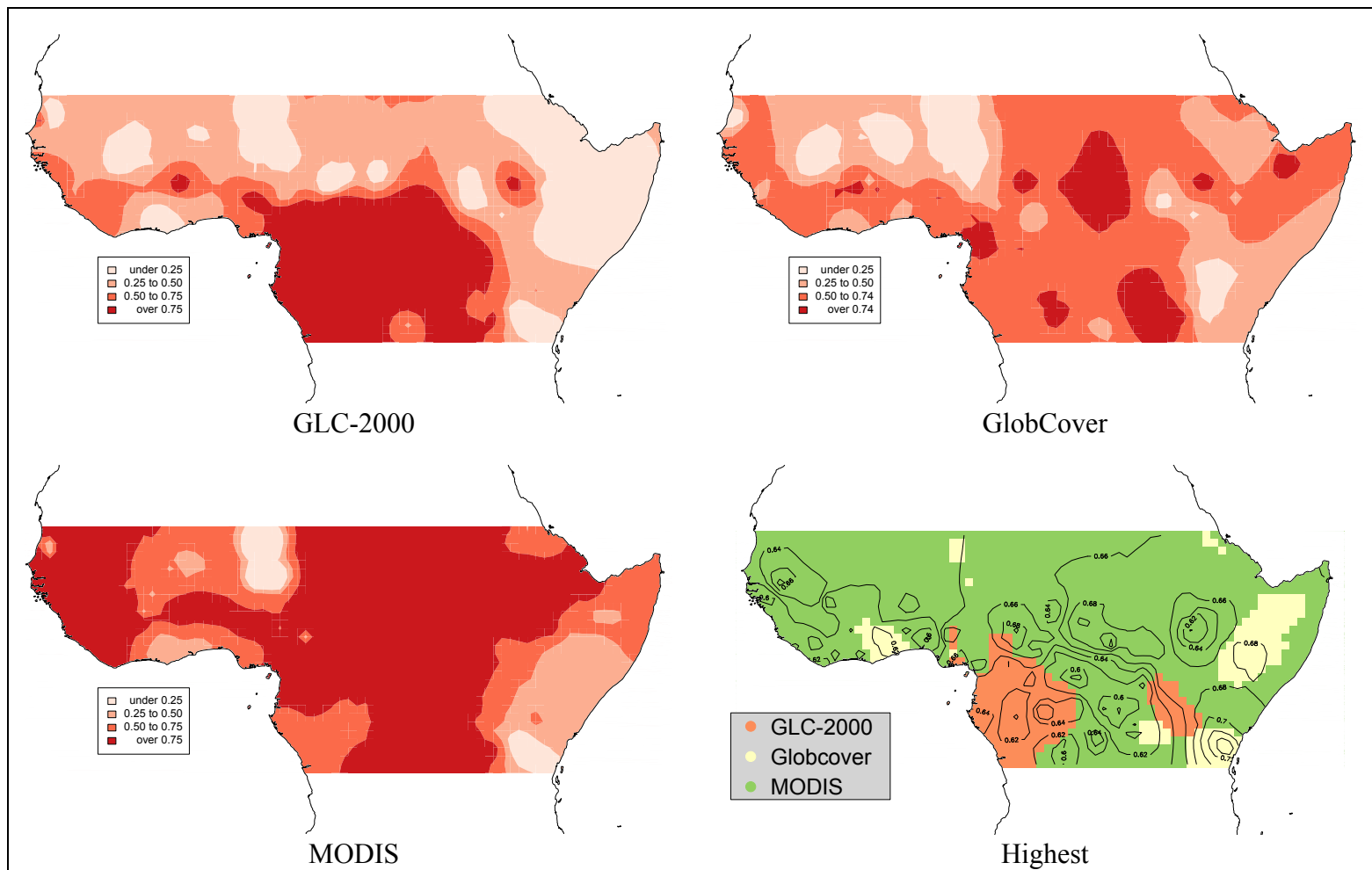


Figure 6. Producer's correspondences for Tree cover derived from comparisons between volunteered land cover and global datasets, with a map of the global datasets with the highest producer's value at each location, overlaid with contours of the geographically weighted variation in the producer's accuracy for Tree cover.

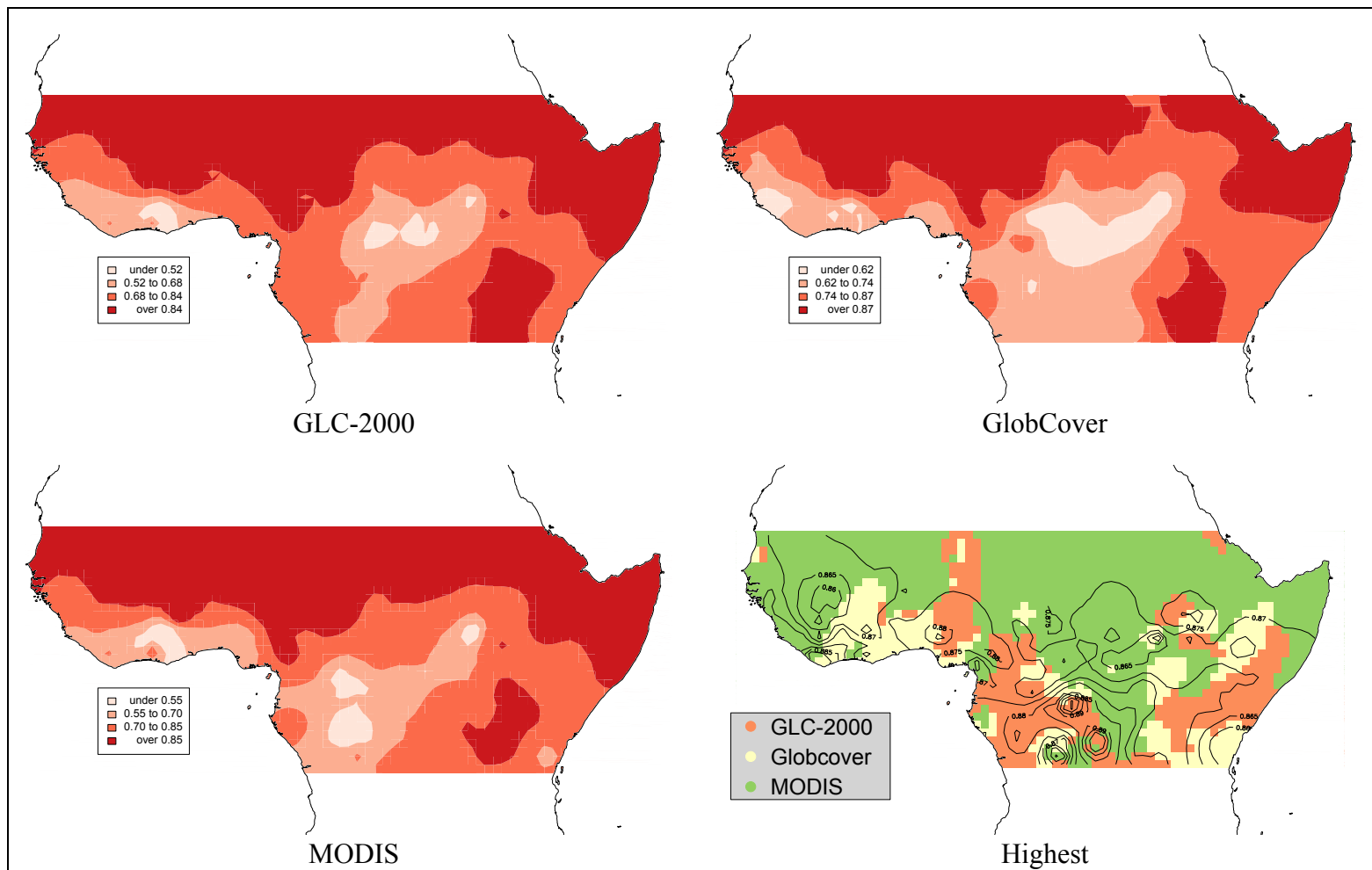


Figure 7. Portmanteau correspondences for Tree cover derived from comparisons between volunteered land cover and global datasets, with a map of the global datasets with the highest portmanteau value at each location, overlaid with contours of the geographically weighted variation in the portmanteau accuracy for Tree cover.

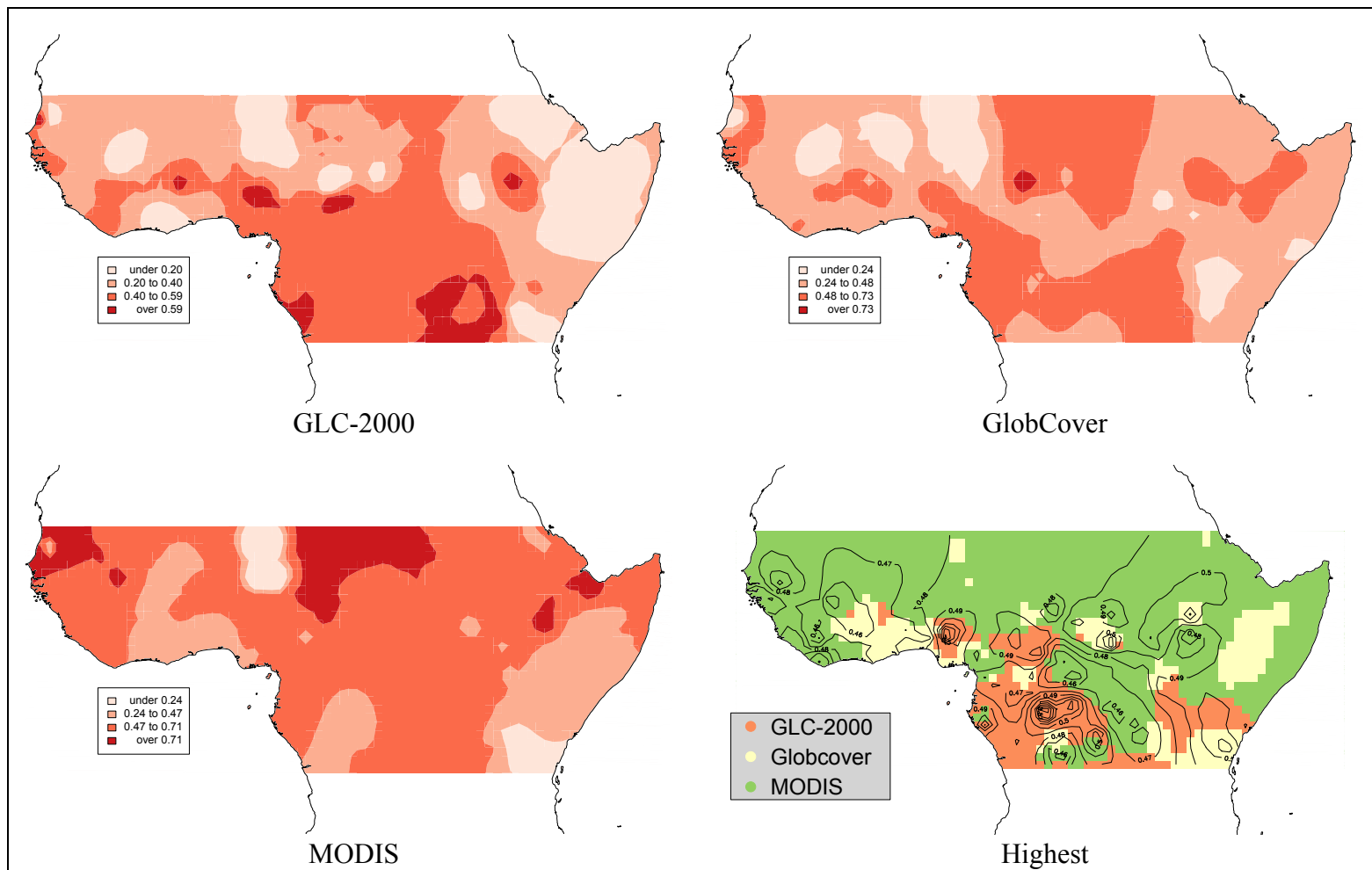


Figure 8. Partial portmanteau correspondences for Tree cover derived from comparisons between volunteered land cover and global datasets, with a map of the global datasets with the highest partial portmanteau value at each location, overlaid with contours of the geographically weighted variation in the partial portmanteau accuracy for Tree cover.