

Clark University

Clark Digital Commons

Geography

Faculty Works by Department and/or School

6-1-2016

A platform for crowdsourcing the creation of representative, accurate landcover maps

Lyndon Estes

Clark University, lestes@clarku.edu

D. McRitchie

Princeton University

J. Choi

Princeton University

S. Debats

Princeton University

T. Evans

Indiana University Bloomington

See next page for additional authors

Follow this and additional works at: https://commons.clarku.edu/faculty_geography



Part of the [Geography Commons](#)

Repository Citation

Estes, Lyndon; McRitchie, D.; Choi, J.; Debats, S.; Evans, T.; Guthe, W.; Luo, D.; Ragazzo, G.; Zempleni, R.; and Caylor, K. K., "A platform for crowdsourcing the creation of representative, accurate landcover maps" (2016). *Geography*. 73.

https://commons.clarku.edu/faculty_geography/73

This Article is brought to you for free and open access by the Faculty Works by Department and/or School at Clark Digital Commons. It has been accepted for inclusion in Geography by an authorized administrator of Clark Digital Commons. For more information, please contact larobinson@clarku.edu, cstebbins@clarku.edu.

Authors

Lyndon Estes, D. McRitchie, J. Choi, S. Debats, T. Evans, W. Guthe, D. Luo, G. Ragazzo, R. Zempleni, and K. K. Caylor

A platform for crowdsourcing the creation of representative, accurate landcover maps

Estes, L.D.^{a,b,1,*}, McRitchie, D.^{c,1}, Choi, J.^a, Debats, S.^a, Evans, T.^d,
Guthe, W.^a, Luo, D.^a, Ragazzo, G.^a, Zempleni, R.^a, Caylor, K.K.^a

^a*Civil and Environmental Engineering, Princeton University, Princeton, NJ, 08544 USA*

^b*Woodrow Wilson School, Princeton University, Princeton, NJ, 08544 USA*

^c*Computational Science and Engineering Support, Office of Information Technology,
Princeton University, Princeton, NJ, 08544 USA*

^d*Department of Geography, Indiana University, Bloomington, IN 47405 USA*

Highlights

- DIYlandcover crowdsources the generation of landcover data, using human pattern recognition skill to create accurate maps with rich geometric detail.
- It incorporates representative sampling and worker-specific accuracy assessment protocols, and connects to a large online job market. This design addresses three problems with crowdsourced mapping: representativity; data reliability; product delivery speed.
- In a trial case, South African cropland was mapped with 91% accuracy by novice workers. A scaling up analysis found that an Africa-wide cropland map could potentially be developed using this software for \$2-3 million within 1.2-3.8 years.

Abstract

Accurate landcover maps are fundamental to understanding socio-economic and environmental patterns and processes, but existing datasets contain substantial errors. Crowdsourcing map creation may substantially improve accuracy, particularly for discrete cover types, but the quality and representa-

*Corresponding author

Email address: `lestes@princeton.edu` (Estes, L.D.)

¹Equal contributors

tiveness of crowdsourced data is hard to verify. We present an open-sourced platform, DIYlandcover, that serves representative samples of high resolution imagery to an online job market, where workers delineate individual landcover features of interest. Worker mapping skill is frequently assessed, providing estimates of overall map accuracy and a basis for performance-based payments. A trial of DIYlandcover showed that novice workers delineated South African cropland with 91% accuracy, exceeding the accuracy of current generation global landcover products, while capturing important geometric data. A scaling-up assessment suggests the possibility of developing an Africa-wide vector-based dataset of croplands for \$2-3 million within 1.2-3.8 years. DIYlandcover can be readily adapted to map other discrete cover types.

Keywords: remote sensing, landcover, crowd-sourcing, accuracy assessment, representative sampling, object extraction

1 Availability

DIYlandcover’s source code will be made available free of charge for suitable non-commercial purposes under a GPLv3 license, upon consultation with the authors. For those interested in commercial applications, the prospective licensee should contact Princeton University’s Office of Technology Licensing. The details of a specific application of the software for delineating crop fields in sub-Saharan Africa can be found at mappingafrica.princeton.edu, together with associated information about participating in the project, including digitizing rules and links for accessing the mapping interface.

1. Introduction

Regional maps of landcover provide critical information on food security estimates (e.g. [Monfreda et al., 2008](#); [Licker et al., 2010](#); [See et al., 2015](#); [Lobell, 2013](#)), models of land-atmosphere interactions (e.g. [Liang et al., 1994](#)), and calculations of carbon stocks (e.g. [Ruesch and Gibbs, 2008](#)), greenhouse gas emissions (e.g. [Searchinger et al., 2015](#)), and habitat change (e.g. [Gibbs et al., 2010](#)). These maps are particularly important in developing regions, such as sub-Saharan Africa, where government land use data are often lacking, error-prone, and inconsistent ([Ramankutty et al., 2008](#); [See et al., 2015](#)).

20 These developing regions are also experiencing rapid land use changes (Gibbs
21 et al., 2010; Rulli et al., 2013) that pose pressing development challenges (e.g.
22 how to feed people at substantially lower environmental cost Searchinger
23 et al., 2015).

24 Unfortunately, landcover datasets derived from medium to coarse reso-
25 lution satellite sensors are particularly inaccurate (Fritz et al., 2010; Fritz
26 and See, 2008). One major reason for poor accuracy is the fact that land use
27 patterns in these regions are dominated by smallholder farming. Smallholder
28 fields are typically smaller (≤ 2 ha) than the resolution (~ 6 ha) of the most
29 commonly used satellite imagery (Jain et al., 2013). Furthermore, smallhold-
30 ers often plant diverse mixtures of crops, which further increases within-pixel
31 heterogeneity (Jain et al., 2013), and their fields often contain remnant trees
32 and have irregular boundaries, which makes them spectrally harder to dis-
33 tinguish from the surrounding vegetation (See et al., 2015; Lobell, 2013).

34 New techniques for merging multiple landcover products are helping to
35 substantially improve map accuracy (Fritz et al., 2011, 2015). However, these
36 approaches cannot overcome the mismatch between sensor resolution and
37 smallholder field size. High resolution satellite imagery (< 5 m) is becom-
38 ing increasingly available—and presumably will become more affordable—so
39 the resolution problem should be solved in the near future (See et al., 2015;
40 Lobell, 2013). But high resolution comes at the expense of higher spectral
41 variability; centimeter-scale data require lower orbits, narrower swaths, and
42 greater communication bandwidth, which combine with clouds to greatly
43 limit the area that can be imaged under contemporaneous environmental
44 conditions, and from comparable viewing angles. This means that high res-
45 olution image mosaics covering large areas contain substantial and largely
46 uncorrectable spectral differences caused by variations in atmospheric con-
47 ditions, vegetation phenology, and bidirectional reflectance. This variability
48 propagates error in automated classifications over large regions, which can
49 already be substantial when there is high within-cover variability (Debats
50 et al., 2015), or high heterogeneity among cover types (Gross et al., 2013).

51 It remains a major challenge to develop algorithms that can accurately
52 classify landcover in the face of both increased image variability and substan-
53 tial spatial heterogeneity. Promising methods are emerging, however, which
54 draw on advances in computer vision and machine learning, such as semantic
55 segmentation (e.g. Schroff et al., 2008) and Randomized Quasi-Exhaustive
56 feature selection (Tokarczyk et al., 2015), to find optimal classifiers within
57 complex urban environments Frhlich et al. (2013) and highly variable small-

58 holder fields (e.g. [Debats et al., 2015](#)). However, these advances are primarily
59 in pixel-wise classification. Accurate, automated methods for extracting in-
60 dividual objects within a single cover type, particularly over wide areas, is
61 arguably even more difficult. Object delineation is an important goal of
62 landcover mapping, as cover geometries encode critical social and environ-
63 mental information ([Fritz et al., 2015](#)), and can play an important role in
64 improving environmental monitoring systems. For example, in agroecosys-
65 tems, field boundaries can provide a filter for extracting “pure”, crop-specific
66 time series of satellite-derived vegetation indices, which helps to improve the
67 accuracy of remotely sensed yield estimates ([Estes et al., 2013a,b](#)). Some
68 limited progress has been made with automated approaches, but these have
69 been demonstrated mainly for small areas where the cover objects have regu-
70 lar geometries and sharp boundaries (e.g. commercial agricultural fields [Yan
71 and Roy, 2014](#); [Ozdarici-Ok and Akyurek, 2014](#); [Ozdarici-Ok et al., 2015](#)).
72 Such methods are not yet proven over large areas with more complex, less
73 distinct cases.

74 An alternative approach is to employ humans, who are very adept at rec-
75 ognizing patterns in noisy images ([Biederman, 1987](#)). The superiority of hu-
76 man over machine pattern recognition provides the motivation for CAPTCHA
77 ([Ahn et al., 2003](#)), which secures websites by requiring human users to rec-
78 ognize fuzzy or irregular letters and numbers that are too difficult for auto-
79 mated algorithms to identify. Human-interpreted landcover maps are thus
80 likely to be consistently more accurate than automated classifiers. Unfor-
81 tunately, since humans are much slower at data processing than computers,
82 human-generated landcover maps covering large areas will require much more
83 time and expense to create. However, this problem is being alleviated by the
84 growth of the internet, which makes it increasingly feasible to turn pattern
85 recognition problems into many small tasks that are undertaken by a large
86 number of online workers—the human equivalent of parallel processing. This
87 ability to “crowdsource” ([Howe, 2006](#)) such work supports projects ranging
88 from galactic classification ([Lintott et al., 2008](#)) to ornithological surveys
89 ([Sullivan et al., 2009](#)). Crowdsourcing of landcover is already being used in
90 the Geo-wiki project, which uses online volunteers to correct landcover data
91 based on their own interpretations of high resolution satellite imagery ([Fritz
92 et al., 2009, 2012, 2015](#)). Recently, these data have been used to create the
93 most accurate (82%) global cropland map ([Fritz et al., 2011, 2015](#)).

94 While the use of crowdsourcing is an extremely promising development
95 for landcover mapping, and is being increasingly used for this and other en-

96 vironmental monitoring applications (Jacobson et al., 2015; Fraternali et al.,
97 2012; Schellekens et al., 2014), many existing projects (e.g. OpenStreetMap
98 (openstreetmap.org)) are geared towards users who create content accord-
99 ing to their personal interests, thus the resulting maps are unlikely to be
100 geographically representative (Fraternali et al., 2012). Furthermore, veri-
101 fying the accuracy of crowdsourced data is a challenge (Allahbakhsh and
102 Benatallah, 2013; Flanagin and Metzger, 2008; See et al., 2015) that remains
103 largely unaddressed by existing platforms. In terms of using crowdsourcing
104 to improve landcover data, prior efforts have focused primarily on validating
105 pixel-based classifications, and less on delineating individual cover objects,
106 which is arguably one of the greatest advantages that people have over ma-
107 chines. Indeed, recognizing and digitizing individual, discrete cover types
108 such as crop fields is considered fairly “straightforward” for humans (Yan
109 and Roy, 2014).

110 In this paper, we describe *DIYlandcover* (or “Do-it-Yourself” land-
111 cover), a new platform for creating crowdsourced landcover data that ad-
112 dresses the three aforementioned limitations. DIYlandcover was designed for
113 mapping discrete, but “noisy”, cover types, where object extraction is of pri-
114 mary interest. Specifically, our platform provides online workers with tools to
115 1) delineate landcover objects within 2) representatively selected locations,
116 while the resulting maps are subjected to 3) periodic quality assessments
117 that provide estimates of individual worker and overall map accuracy. We
118 provide an overview of DIYlandcover’s design and mechanics, and report on
119 the results of a trial application mapping crop fields in South Africa, which
120 suggests that DIYlandcover allows inexperienced online workers to generate
121 high accuracy (>90%), geometrically rich, and geographically representative
122 landcover data at a much faster rate than is usually possible with human-
123 based mapping.

124 2. System design

125 The inspiration for DIYlandcover came from GeoTerraImage, a company
126 that mapped South Africa’s arable cropland by manually digitizing fields
127 visible in high resolution satellite imagery (GeoTerraImage, 2008). The re-
128 sulting map set is 97% accurate in distinguishing cropped from uncropped
129 areas at a 4 ha resolution (see detailed accuracy assessment in Appendix
130 S1), and provides rich detail on field type and geometry. However, making
131 these maps was an expensive and lengthy process; the estimated labor cost

132 for digitizing was $\$5 \text{ km}^{-2}$, and the project took approximately 2.5 years to
133 complete (Ferreira, pers. comm.).

134 We developed DIYlandcover to help overcome these constraints of cost
135 and production time, while retaining the advantages of human image in-
136 terpretation skill demonstrated by GeoTerraImage. Our platform connects
137 workers in an online job marketplace to a map application programming
138 interface (API) that hosts high resolution satellite imagery. DIYlandcover
139 currently works with Amazon’s Mechanical Turk (Services, 2012) and the
140 Google Maps API, but these could in principle be replaced by other services.
141 These two aspects of DIYlandcover substantially reduce both mapping costs
142 and completion times, because the imagery is free and the platform can access
143 a potentially large number of workers.

144 Given the distributed and anonymous nature of the online job market,
145 we cannot intensively train workers (as GeoTerraImage did), yet our map-
146 ping task is complex, requires significant image interpretation skill, and must
147 be completed in a systematic manner. Therefore, to ensure the scientific
148 quality of its maps, DIYlandcover incorporates site selection and accuracy
149 assessment protocols (Fig. 1). A sampling grid (SG in Fig. 1) over the
150 desired study region provides the basis for collecting stratified random sam-
151 ples. The first draw identifies sites where the researcher/administrator (the
152 “Requester”; Allahbakhsh and Benatallah, 2013) will provide landcover refer-
153 ence maps (black cells). Subsequent draws select sites where workers will
154 create new maps (grey cells). This sample of locations is then sent to the job
155 marketplace. All workers must pass an initial qualification test (Q1 in Fig.
156 1) that proves their ability to map a handful of sites with a minimum level
157 of skill. Once qualified, workers begin mapping. Each worker will map both
158 grey and black sites, which are respectively referred to as N (for normal) and
159 Q (for quality assessment) sites. Q sites are indistinguishable from N sites,
160 and are intermingled such that each worker has a Requester-defined proba-
161 bility of encountering a Q site. Completed maps from N sites are inserted
162 into DIYlandcover’s database (D), while maps from Q sites are first scored
163 according to their agreement with their reference maps (Q2 in Fig. 1). Maps
164 that fall below a minimum score are rejected. Map scores are incorporated
165 into a worker-specific quality score, which is used to assign confidence to all
166 maps generated by a worker, and to determine overall map accuracy. Work-
167 ers are paid (P in Fig. 1) for each site mapped, with the possibility of bonus
168 payments linked to quality scores.

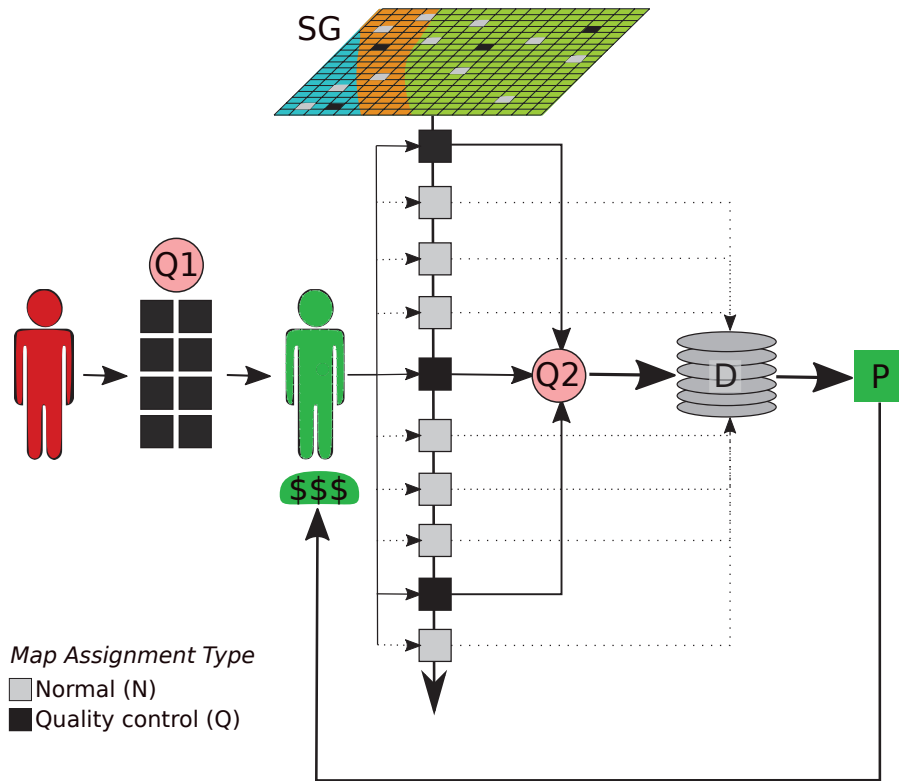


Figure 1: An overview of DIYlandcover’s design. A survey grid (SG) is overlaid on a geographic area, and then random samples (weighted proportionally to the probability of cover type presence, represented by green, orange, and blue) are drawn to specify where groundtruth maps will be generated (black cells) to support worker map quality (Q) assessment. Subsequent random draws (grey cells) select sites that are undertaken as normal (N) mapping assignments. N and Q sites are sent inter-mingled to the online job marketplace for mapping. A first time worker (red) must take an initial map qualification test (Q1), after which she or he is qualified (green) and begins mapping. Maps from N sites are stored in the database (D); Q site maps are first scored based on their agreement with groundtruth (Q2). This score contributes to a longer term worker quality score, which is used to assess overall map quality and allows performance-based bonuses to be paid on of fixed per site payments (P).

169 **3. The mechanics of DIYlandcover**

170 The basic structure of DIYlandcover consists of three elements (Fig. 2):
 171 the main server hosting DIYlandcover’s database, here a Linux virtual ma-

172 chine with PostgreSQL (9.4) with the PostGIS (2.1) spatial extension; a
173 map server hosting the satellite imagery, in this case the Google Maps API
174 (Developers, 2012); the online job market, Mechanical Turk (Services, 2012).
175 Within this structure several key processes govern the creation and manage-
176 ment of mapping tasks.

177 3.1. Site selection

178 A “master grid” covering the study area is first created as a PostGIS
179 table. Each cell provides a unique identifier, and the cell resolution defines
180 the area of an individual mapping task. This grid is intersected with a second
181 grid containing landcover occurrence probabilities, which are converted into
182 categorical weights. A third field is created that indicates whether each cell
183 is available to be mapped or not.

184 After the initial random draw (of a user specified size) is taken to identify
185 quality assessment (Q) sites (Section 2, Fig. 1), the selected cells’ status is
186 set to unavailable. The geometries are written to individual keyhole markup
187 language (KML) files, and their IDs are added to a “KML data” table, where
188 a field specifying cell type is set to “Q” to indicate that the corresponding
189 KMLs reference quality control sites. The user has to provide landcover refer-
190 ence maps for these sites, the geometries of which are stored in a “reference
191 maps” table.

192 The next draw collects sites that will form the normal (“N”) map produc-
193 tion process, where a worker (or workers) creates maps for locations where
194 the underlying landcover is unknown. This step is governed by *KMLGen-*
195 *erate*, an R process that connects to the database (via the RPostgreSQL
196 package; Conway et al., 2012), takes a weighted random draw of size X (a
197 parameter stored in the “configuration” table that holds all variables used
198 by DIYlandcover) from the master grid table, writes each cell geometry to a
199 separate KML file, adds the selected cell IDs to the KML data table, and sets
200 the field type value to “N”. The script changes the cell status in the master
201 grid to unavailable. As N type maps assignments are completed, their status
202 is set to mapped in the KML data table. *KMLGenerate* runs as a daemon,
203 selecting a new random draw as soon as the number of unmapped sites falls
204 below a specified number, ensuring that there is never a system delay in
205 sending mapping assignments to the job market (see 3.2).

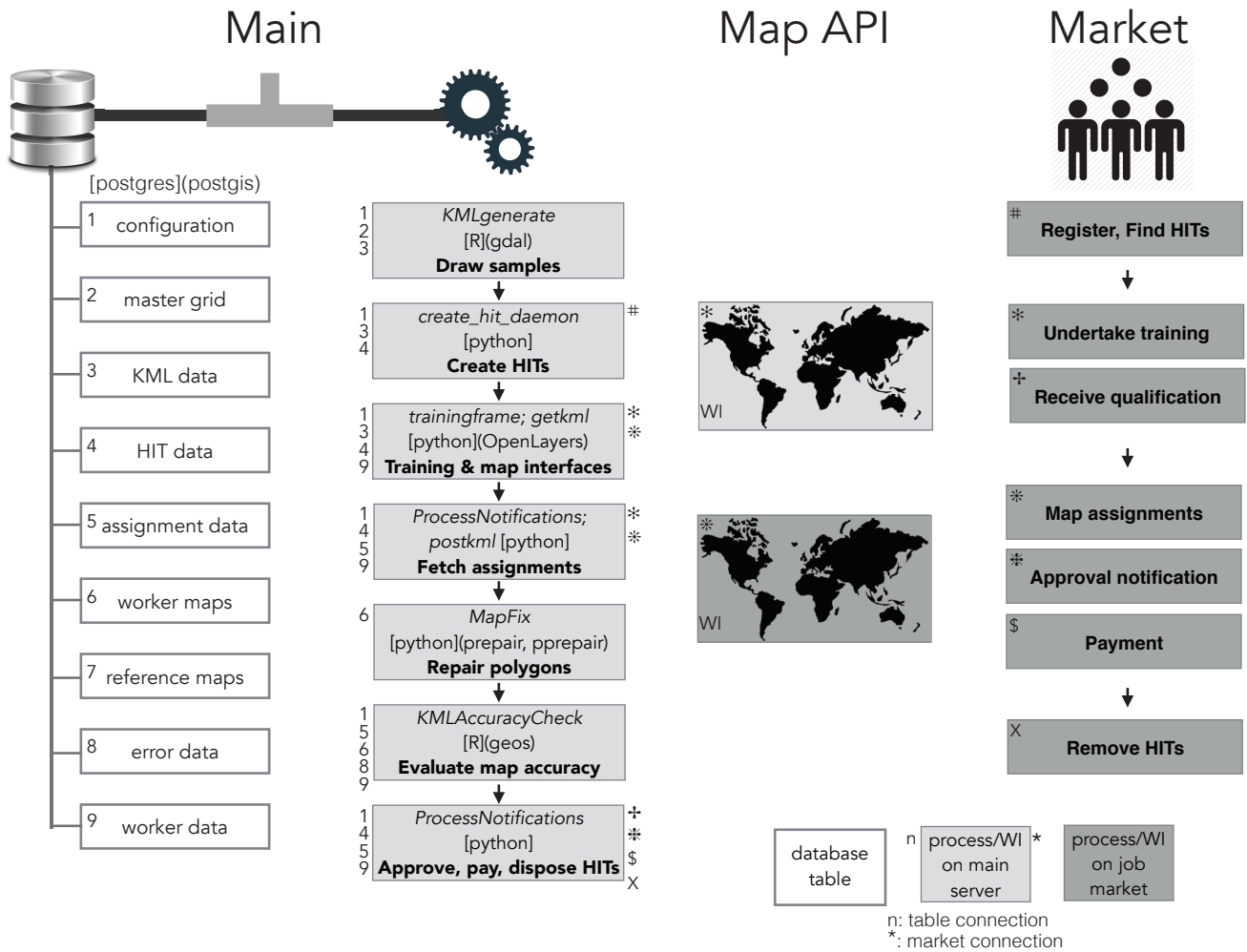


Figure 2: The components, and primary processes of DIYlandcover. The main server contains the system database and processes. Primary data tables are shown by the white boxes with grey borders. Primary processes are shown in light grey boxes (process names are italicized, primary software in brackets and its external dependencies in parenthesis, and description in bold). Server processes interact with specific data tables (indicated by the numbers to the left), and with processes that occur in the online job market (indicated by symbols to the right). The two versions (one for training, one for qualified workers) of the worker interface (WI) to the map API are shown, color-coded according to where they are hosted (on main server or online market).

206 *3.2. Creating mapping assignments*

207 Following selection, each site is converted into a mapping task for online
208 workers. These tasks are referred to as Human Intelligence Tasks (HITs),
209 in Mechanical Turk’s parlance. HITs are created by (*create_hit_daemon*), a
210 python daemon that uses the boto library to interface with Mechanical Turk
211 (MT). The daemon polls MT (at regular intervals) to see how many DIY-
212 landcover HITs of types Q and N exist on MT (zero at start of production),
213 and whether they fall below their minimum required numbers. These num-
214 bers are calculated from two configuration parameters: the minimum total
215 number of HITs that should be available on MT, and the percentage of these
216 that should be of Q type. If the actual numbers of each type fall below their
217 target numbers, *create_hit_daemon* selects the IDs of available KMLs from
218 the KML data table, and sends these together with associated HIT metadata,
219 which includes the pay rate, the number of times the HIT should be mapped,
220 the qualifications required to undertake the HIT (see 3.5), and a definition of
221 the task. MT then registers each HIT and provides it with a unique HIT ID
222 and registration time, which is logged into a “HIT data” table on the main
223 server.

224 *3.3. Undertaking the mapping assignment*

225 Once a HIT is registered on MT, it is visible to all workers in the mar-
226 ketplace, but can only be undertaken by qualified workers (see 3.5). Quali-
227 fied workers who choose to undertake DIYlandcover-generated HITs are first
228 shown a default HIT preview, and they must choose to accept it before they
229 can see the actual location to map. This step helps prevent workers from
230 declining more challenging sites, which bias the sample towards simpler land-
231 covers.

232 To enable workers to perform a mapping HIT, DIYlandcover uses an
233 OpenLayers interface to the image server, which sits within MT’s user screen,
234 centers the map view on the site of the HIT location, and provides a set
235 of digitization tools (Fig. 3). As soon as the worker accepts the HIT, it
236 becomes a mapping assignment that is issued a unique assignment ID. A Web
237 Server Gateway Interface (wsgi) script, *getKML*, retrieves the OpenLayers
238 javascript, the frame size parameters for the MT interface, the url for the
239 KML demarcating the sample site, and user instructions (e.g. tool use tips),
240 and passes these to MT, and collects the worker, assignment, and HIT IDs
241 and acceptance time, and records these into the “assignment data” table.

242 The worker then draws polygons around the landcover type(s) of interest
 243 that intersect the KML sample square (Fig. 3), and has the option to edit
 244 or delete individual geometries and provide comments. On completion, the
 245 worker saves the map, and is then taken to the next HIT preview screen.
 246 Alternatively, the worker may choose to return the assignment uncompleted.
 247 If this happens more than a specified number of times, the worker’s qual-
 248 ification can be revoked (see 3.4), which is another check against sample
 249 selection bias. The assignment is automatically abandoned if it is not com-
 250 pleted within a defined time. We impose this last restriction to minimize
 251 bias in the estimation of wage rates (see 4.2); if workers leave the assignment
 252 unfinished on their computer for long periods, the amount of time required
 253 to complete assignments will be inflated.

254 When the assignment is completed, returned, or abandoned, MT sends
 255 an email notification to the main server, where it is retrieved by *ProcessNo-*
 256 *tifications*, a python process. If the assignment is returned or abandoned,
 257 it is marked as unprocessed and returned to the pool of available HITs on
 258 MT, and the worker receives no pay. If the assignment was completed, post-
 259 processing routines are triggered.

260 3.4. Processing completed assignments

261 Several processing steps must be performed before the worker is paid for
 262 the completed assignment, which depend on whether the worker created any
 263 polygons during the assignment, and whether it was of Q or N type. If
 264 the worker created polygons, then the geometries, KML ID, assignment ID,
 265 and completion time are stored in the “user maps” data table by process
 266 *postKML*, which then triggers *mapFix*, a python script that invokes *prepair*
 267 and *pprepair* (Ohori et al., 2012), which repair the topologies of single and
 268 multi-polygons, respectively. This step is essential because hand-digitized
 269 polygon data often contain errors, such as self-intersections and unintended
 270 overlaps, which can render topologies invalid and cause subsequent spatial
 271 analyses (per 3.4) to fail. The repaired geometries are then inserted into the
 272 user maps table.

273 Next, the assignment is given a score, which is recorded in the assignment
 274 data table. If the assignment was of N type, this score is null; for Q type,
 275 *KMLAccuracyCheck*, an R process, is called to compare the worker’s and
 276 reference maps, with the score determined by:

$$S = \beta_1 C + \beta_2 O + \beta_3 I \quad (1)$$

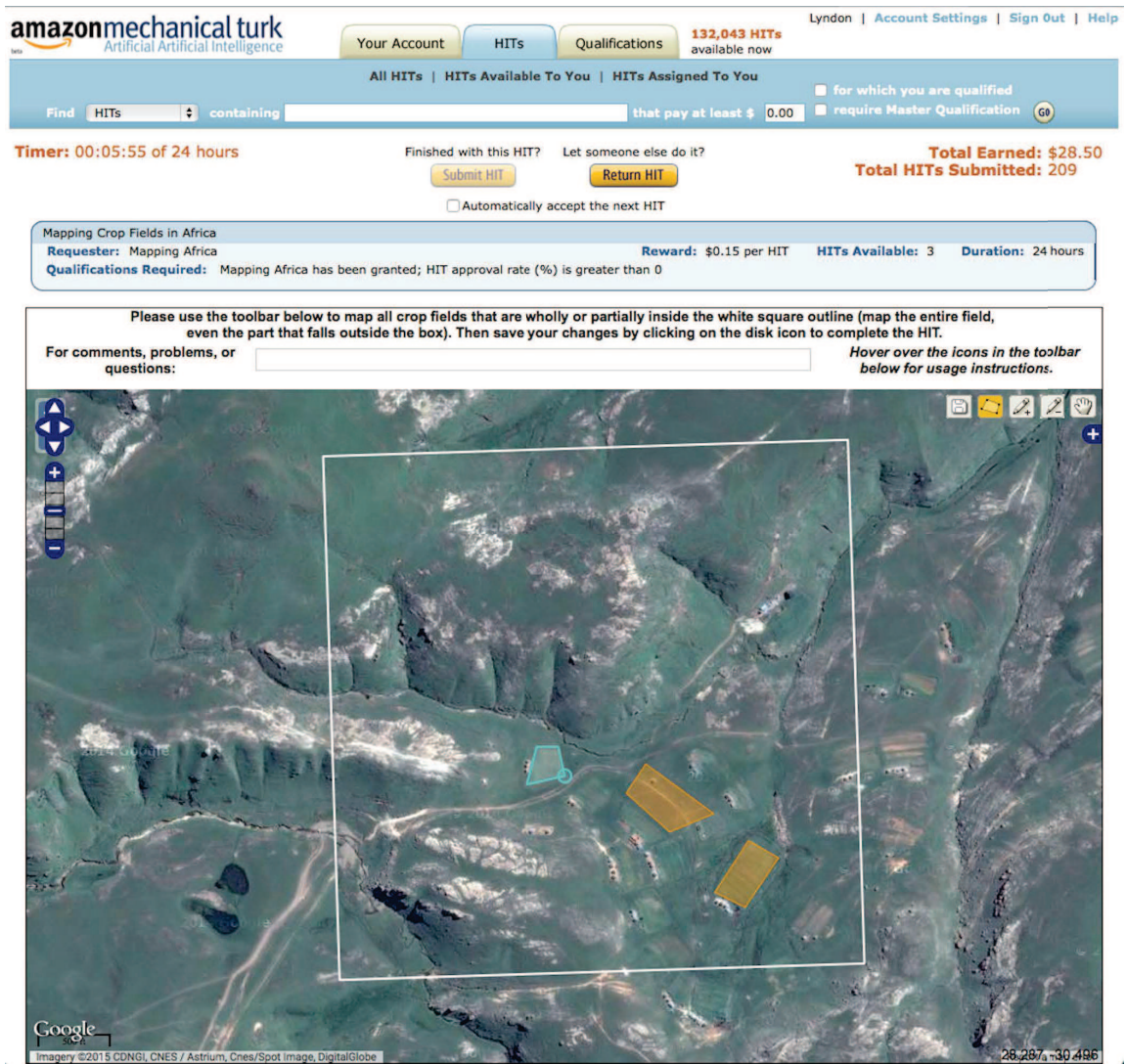


Figure 3: The DIYlandcover mapping interface within Amazon.com’s Mechanical Turk job marketplace. The white square is the KML sampling frame, gold polygons are completed crop field polygons, the blue polygon is a field in the process of being mapped. Mapping controls are in upper right corner of the image frame.

277 Where S is overall mapping accuracy, β_1 - β_3 are user-defined weights, and:

$$C = 1 - \frac{\text{abs}(n - N)}{\max(n, N)} \quad (2)$$

$$O = \frac{a}{a + c} \quad (3)$$

$$I = \frac{A + D}{A + B + C + D} \quad (4)$$

278 Or:

$$I = \left(\frac{A}{A + C} + \frac{D}{B + D} \right) 0.5 \quad (5)$$

279 With C being count error, or the agreement between the number of landcover
 280 polygons in the worker’s maps (n) and in the reference data (N). O measures
 281 map agreement for those parts of the worker’s and reference polygons that
 282 fall *outside* of the KML grid, where a is the area of overlap, and c is the false
 283 negative error (i.e. the area of reference field polygons falling outside the grid
 284 that the worker failed to map). I measures map accuracy *inside* the KML
 285 grid, with A being the grid interior equivalent of a, B the false positive error
 286 (i.e. landcover incorrectly labelled by the worker), C the false negative error
 287 (landcover area missed by the worker), and D the true negative area (area
 288 correctly left unmapped). I can be calculated using standard classification
 289 accuracy (Eq. 4), or a variant of the True Skill Statistic (Eq. 5 [Allouche](#)
 290 [et al., 2006](#)), a more stringent measure that corrects for class prevalence,
 291 which we compressed to fall between 0 and 1 rather than -1 to 1. The areas
 292 of a, c, A, B, C, D are calculated using intersection and difference operations
 293 provided by the rgeos library ([Bivand and Rundel, 2013](#)), after transforming
 294 maps to a projected coordinate system.

295 We include the O metric to encourage workers to completely map features
 296 intersecting the sampling grid (i.e. either falling entirely within or both
 297 within and outside of it), in order to have unbiased estimates of landcover
 298 size classes. However, we can only partially assess the accuracy of exterior
 299 features because it is impossible to correctly define negative space outside
 300 the sample grid, since it is both unbounded and may contain target features
 301 that will not be mapped because they do not intersect the grid. An error
 302 map showing each of the accuracy assessment components is illustrated in
 303 Figure 4.

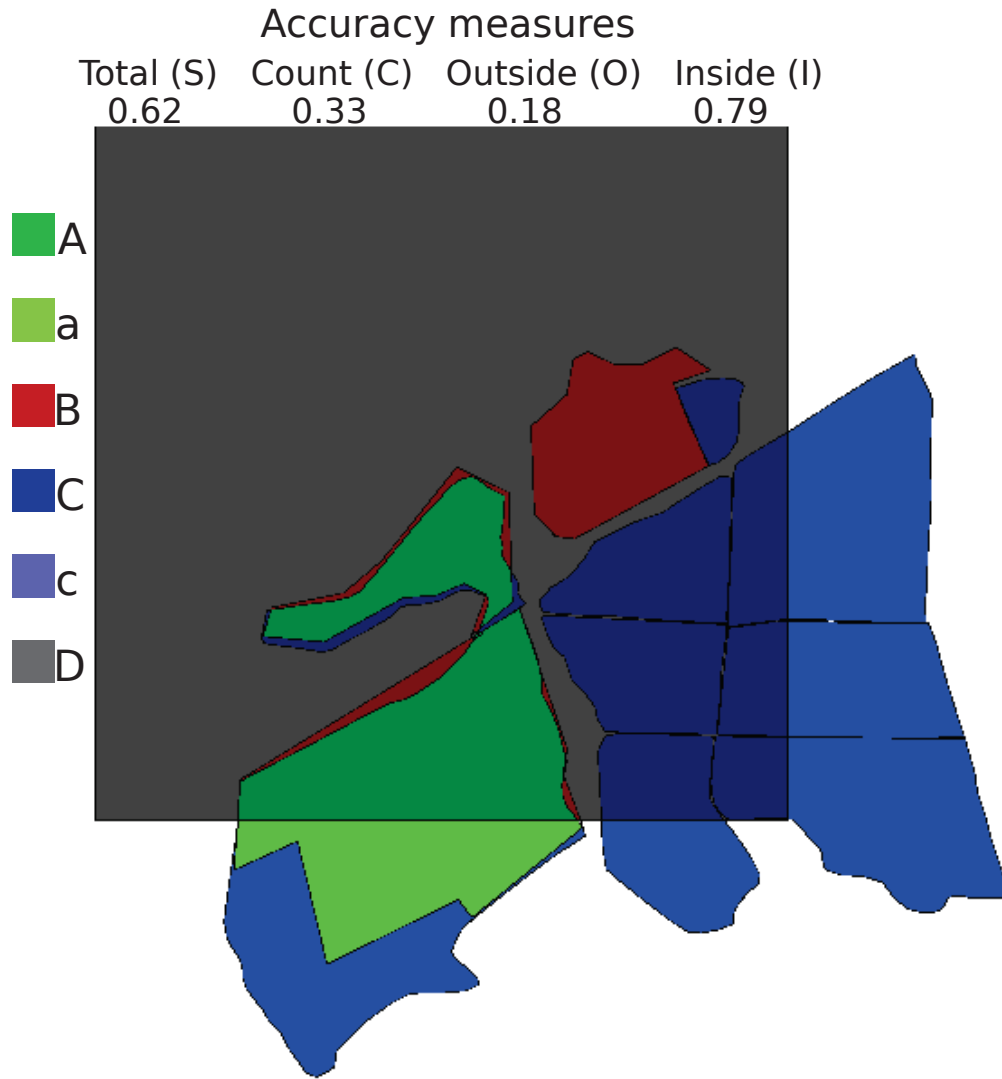


Figure 4: A graphical illustration of the accuracy assessment algorithm (as applied to cropland maps), providing the resulting scores for overall accuracy (Eq. 1) and count, outside, and inside error (Eqs. 2-5), where each component ranges between 0 (most error) and 1 (no error). The area of each error component is color-coded: A (agreement inside the grid), a (agreement outside), B (false positive error inside the grid), C (false negative inside), c (false negative outside), and D (true negative inside).

304 Once the algorithm has run, all accuracy measures (S, C, O, I) are stored
 305 in the “error data” table, while S is stored in the assignment data table. S
 306 is also added to a vector of Q scores for the specific worker (stored in the
 307 “worker data” table), which is used to calculate a moving average of the
 308 worker’s recent performance. If S is above a minimum accuracy threshold,
 309 then the assignment is approved. If rejected, then payment is withheld,
 310 and a notice is sent to MT where it is added to the worker’s system-wide
 311 rejection rate. Successive rejections can result in the revocation of mapping
 312 qualifications if a worker’s *quality* score drops below the accuracy threshold.
 313 The quality score is:

$$\text{quality} = \frac{S_i + S_{i-1} + \dots + S_{i-(j-1)}}{j} - \beta_4 \frac{R_i + R_{i-1} + \dots + R_{i-(j-1)}}{j} \quad (6)$$

314 Where i is the most recent S value calculated, and j the total number
 315 of recent S scores to use in calculating a mean S. To minimize assignment
 316 selection bias (see 3.3), an additional penalty, the worker’s rate of assignments
 317 accepted but returned without completing (R , which equals 1 for a return, 0
 318 for a completion), is multiplied by a weight β_4 and deducted.

319 In cases where the worker returns no maps for a Q type assignment, map
 320 storage and cleaning does not occur before *KMLAccuracyCheck* is run. In
 321 these cases, the C and O scores (Eq. 2 & 3) reduce to 1 where the reference
 322 map has no landcover polygons, or 0 if it does. If the assignment is of N
 323 type, it is scored as NULL and added to the assignment data table.

324 Unlike the Q type, N assignments are automatically approved, under the
 325 logic that the worker’s quality score at the time of map creation is indicative
 326 of that map’s accuracy. The exception to this is N assignments created by
 327 a newly qualified worker (see 3.5), which are marked as “untrusted” in the
 328 assignment data table until that worker completes as many Q assignments as
 329 are needed to calculate the moving average accuracy score. Upon assignment
 330 approval, *ProcessNotifications* relays a message to MT and the worker is paid
 331 (see 3.6) from the Requester’s account, and then removes the corresponding
 332 HIT from MT. Q sites will be re-created as HITs multiple times, while N
 333 sites are mapped just one time.

334 3.5. Worker qualification and payments

335 All workers performing mapping assignments must first be qualified, which
 336 is treated as a special case of Q type assignments. MT evaluates the quali-
 337 fication status of each worker attempting to access a DIYlandcover HIT. If

338 the worker is not qualified, a link to a training module is presented on the
339 MT interface. The module, which is hosted on the main server, is managed
340 by *trainingframe*, a python process, which issues each new trainee a unique
341 training ID. The trainee first watches video tutorials explaining the project
342 and its mapping rules, and is then required to map several training sites,
343 the accuracy of which is assessed by *KMLAccuracyCheck*. Trainees must
344 map each site to the minimum accuracy standard, but are given unlimited
345 chances to do so. A separate set of tables mirroring those used for collect-
346 ing map, assignment, worker, and error data is used to record training data.
347 Once a worker successfully completes all training sites, a qualification re-
348 quest is posted on MT. A daemon, *process_qualifications_requests*, polls MT
349 at specified intervals, collects these requests together with associated worker
350 and training IDs, examines for each worker whether all training sites were
351 completed successfully, and, if so, adds the trainee’s worker ID to the worker
352 data table, sets the qualification status to true, then sends a notice to MT
353 that the worker is now qualified. Candidate workers who fail to pass all train-
354 ing sites, or workers whose qualifications are revoked due to poor accuracy
355 (see 3.4), can repeat the training to qualify/re-qualify.

356 Upon qualification, workers are paid a small bonus, and can begin map-
357 ping assignments. Workers are paid a flat rate for approved assignments. To
358 incentivize worker performance, DIYlandcover also allows bonus payments
359 to be made based on the worker’s accuracy score. If implemented, the bonus
360 algorithm, managed by *ProcessNotifications*, pays an extra per assignment
361 amount if the worker’s quality score exceeds certain thresholds.

362 4. Applying DIYlandcover to map South African crop fields

363 We examined the capabilities of DIYlandcover by applying it to map
364 crop field boundaries in South Africa. South Africa was a convenient test
365 case because its cropland was already mapped (see section 2; [GeoTerraImage,](#)
366 [2008](#)) using similar methods, providing both an objective means for evaluat-
367 ing DIYlandcover’s performance, and a readily adaptable source of reference
368 maps. Furthermore, South Africa’s diversity of agricultural systems are rep-
369 resentative of the image interpretation challenges facing workers. This mix
370 ranges from hard to detect communal and smallholder agriculture, to more
371 easily discerned industrial fields ([Hardy et al., 2011](#)). South Africa also pro-
372 vides the test site for the [Mapping Africa](#) project, which aims to create high
373 quality cropland maps for sub-Saharan Africa.

374 *4.1. Mapping set-up*

375 We created a 1X1 km, Albers Equal Area Conic-projected sampling grid
376 for South Africa, and used logistic regression to model the probability of
377 cropland presence throughout the country. Equally sized random draws of
378 points selected inside and outside GTI field boundary polygons provided the
379 positive and negative responses of the dependent variable, while predictors
380 were derived from gridded rainfall and elevation data and a map of protected
381 areas (for further details on these variables see [Estes et al., 2013b, 2014](#)). The
382 resulting probability was divided into quartiles, which provided the weights
383 used by *KMLGenerate*.

384 For Q sites, we used the modeled cropland probability categories to draw
385 a random select 609 grid cells (0.05% of South Africa’s area), providing a
386 representative sample of South Africa weighted towards agricultural areas.
387 We intersected these with the GTI polygons to create the associated Q data
388 tables (3.1). These polygons were then further edited to make the Q maps
389 consistent with imagery in the Google Maps API, and to conform with the
390 specific mapping rules that we set for workers (Table 1). Workers were asked
391 to map sites where crop fields were actively or very recently (i.e. within the
392 past 2-3 years) used for arable agriculture. This category of agriculture takes
393 many forms in South Africa (see Appendix S1 for an illustration), ranging
394 from large, clearly defined, commercial fields to less geometrically distinct
395 smallholder fields, which often contain trees and mixed crops. Long term
396 fallows, tree crops (orchards, commercial afforestation), and non-agricultural
397 areas were left unmapped. In cases of uncertainty (e.g. the worker had
398 trouble telling whether the field was active or abandoned), workers were
399 asked to map every second field. On top of the high variability in arable
400 fields, narrowing the mappers’ focus actually made the task more challenging,
401 because the agricultural types described in Table 1 often look similar, which
402 increases the risk of both false positive and false negative errors. For instance,
403 it is often difficult to tell whether a field is active or abandoned, while young
404 orchards or recently cleared forest compartments can be mistaken for arable
405 fields. In all these examples, field boundaries tend to be clearly visible, thus
406 more inclusive mapping rules would likely reduce both types of error.

407 The system was set to make random draws of 500 N sites from the master
408 grid each time the number of N sites available for mapping fell below 500
409 (3.1), in order to ensure that no system latency occurred as the system
410 selected new mapping locations. At least 10 HITs, 80% N and 20% Q, were
411 maintained on MT at any given time, with the system polling MT every 10

Table 1: Rules for mapping crop fields in the South Africa-focused application of DIY-landcover. Workers were asked to map only currently active (i.e. farmed within the past 2-3 years) annual crop fields, and to not delineate other agricultural types.

Feature type	Action
No cropland visible	Don't map
Active annual crop field	Map
Fallow crop field	Don't map
Unsure if active crop fields	Map every second feature
Permanent tree crops (orchards/plantations)	Don't map
Improved pastures	Don't map

412 seconds to see if new HITs were needed (3.2). This relatively low number
 413 allowed rapid cycling of HITs through MT, while the ratio of N:Q HITs
 414 ensured that worker accuracy was assessed (3.4) frequently during the trial
 415 (once in every five assignments).

416 The accuracy algorithms (Eq. 1) β terms were set as 0.1, 0.2, and 0.7
 417 for the C (Eq. 2), O (Eq. 3), and I terms (here Eq. 4). We selected a low
 418 weight for C because determining the boundaries of individual fields from
 419 overhead imagery is fairly subjective, even for expert observers, and we did
 420 not want to unduly penalize workers for a difference in judgement, yet we also
 421 wanted to discourage rapid mapping that erased boundaries between clearly
 422 distinct fields. We gave O a slightly larger weight to stress the importance of
 423 completed fields that extended outside the sample grid, but a larger weight
 424 would give the worker too much credit for cases where no fields intersected
 425 the grid. The I term was weighted most heavily because it is the only place
 426 where workers' abilities to correctly distinguish null space can be assessed.
 427 We used the same weights to assess assignments within the 8-site training
 428 module.

429 Payment was set at \$0.15 per assignment. A four-tier bonus payment
 430 algorithm was also written into logic. We did not implement this logic in
 431 our initial trial, in order to first assess whether the base rate would allow
 432 workers to achieve our target wage of \$8-10 hour⁻¹, but we evaluated the
 433 cost implications of bonus payments set to \$0.01, \$0.02, \$0.03, and \$0.05 for
 434 worker quality scores exceeding 0.85, 0.95, 0.975, or 0.99, respectively.

435 *4.2. Trial results*

436 The Mapping Africa trial ran on Mechanical Turk for 26.4 hours between
437 October 2-3, 2013, resulting in 945 mapping assignments, of which 882 were
438 approved, 10 were rejected (due to failing accuracy scores), and 53 were not
439 completed (i.e. returned or abandoned). A total of 707 N sites with 216
440 (31%) containing worker-delineated polygons were mapped, as well as 185 Q
441 sites, with 65 (35%) having fields (Fig. 5).

442 These sites were mapped by 15 different workers, from a pool of 18 who
443 passed the initial qualification test. A further 18 took the qualification test
444 but failed to pass. The distribution of mapping effort was highly skewed,
445 with three workers completing 65% of the total assignments (Fig. 6A). The
446 average Q:N assignment ratio for each worker was 18%, but there was high
447 variability among workers who completed less than 50 assignments (Fig. 6A).
448 The mean accuracy assessed across all Q sites (using Eq. 1 with Eq. 4) was
449 0.91 (out of 1), but Q sites containing fields were mapped with lower overall
450 accuracy (0.79) than sites without fields (0.97; Fig. 6B). Using just the inside
451 component of the score (Equation 4), accuracy was higher for sites with
452 fields (0.89 with fields versus 0.99 without). To understand these accuracy
453 discrepancies more fully, the number of polygon vertices in the reference
454 polygons can be used as proxy for cropland complexity, and thus assignment
455 difficulty. Worker accuracy declined significantly, albeit weakly ($p < 0.048$),
456 in relation to this complexity (Fig. 6C). Worker effort also declined strongly
457 as a function of map complexity (Fig. 6D); the more fields there were to
458 map—or the more intricate their boundaries—the fewer vertices placed by
459 workers, presumably to minimize mapping time. This reduction in effort may
460 partially explain the increased error.

461 Replacing Equation 4 with Equation 5 (the True Skill Statistic; TSS),
462 which corrects for class prevalence (Allouche et al., 2006), to calculate map
463 score (Eq. 1) removed the significant negative relationship between map
464 score and complexity (F-statistic: 0.98; $p < 0.32$). At sites with only a few
465 fields, which are both less complex and typically having a much higher share
466 of non-cropped than cropped area, Eq. 4 was more lenient than at more
467 complex sites, because the worker received proportionally more credit for
468 “mapping” the uncropped space. This tendency is seen in Fig. 6E, which
469 shows that map scores calculated using Equation 4 were generally higher
470 than those assessed using Equation 5; for sites with fields, scores were on
471 average 0.1 higher, and up to 0.14 greater where fields were relatively simple

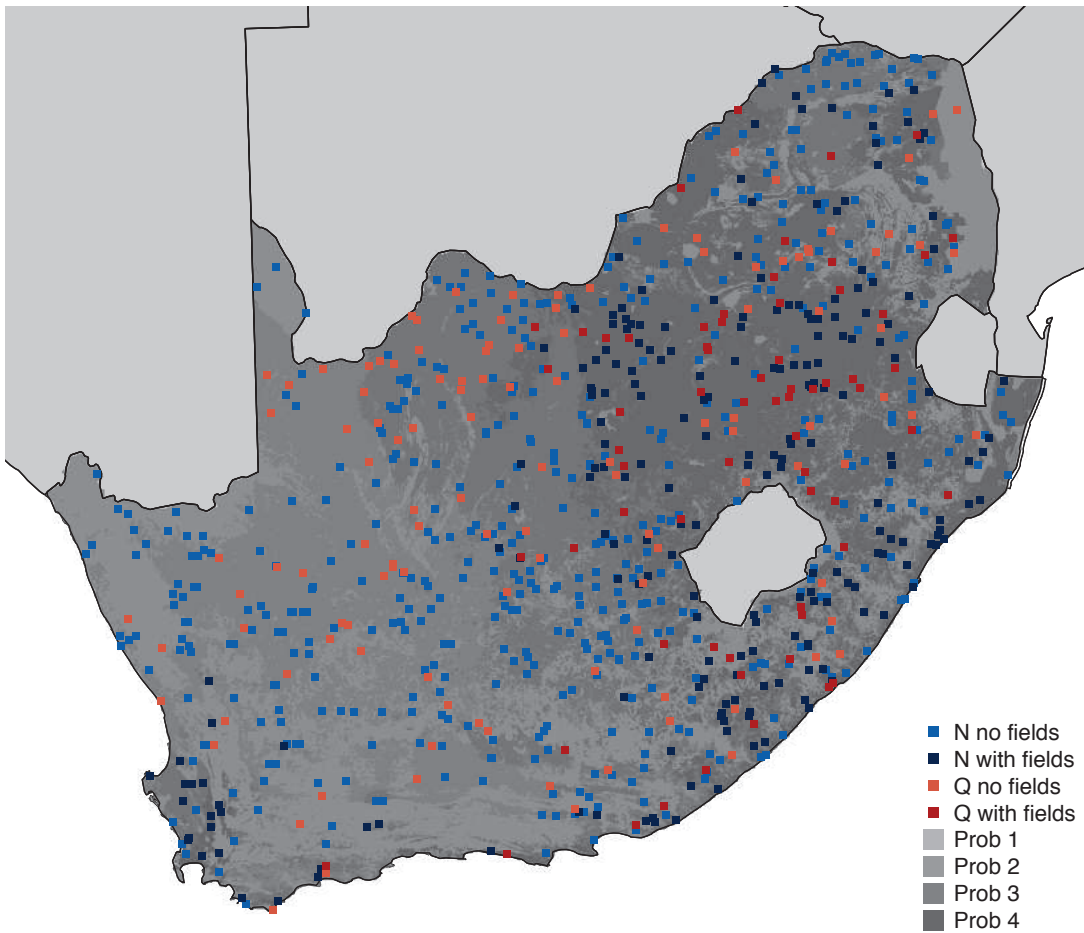


Figure 5: A map of DIYlandcover trial results over South Africa, showing the distribution of mapped sites, color-coded according to their assignment type (Q or N) and whether they contained worker-mapped crop field boundary polygons. The grey shading indicates the four-category weighting derived from a logistic regression model of cropland occurrence.

472 (i.e. where truth maps had <25-50 vertices, indicating both low complexity
 473 and small areas).

474 Accuracy appears to improve with experience, as workers' average accu-
 475 racy scores increased in proportion to the number of Q assignments com-
 476 pleted. Accuracy gains increased rapidly below 20-25 completed Q assign-
 477 ments, after which they leveled out between 0.9 and 1 (Fig. 6F).

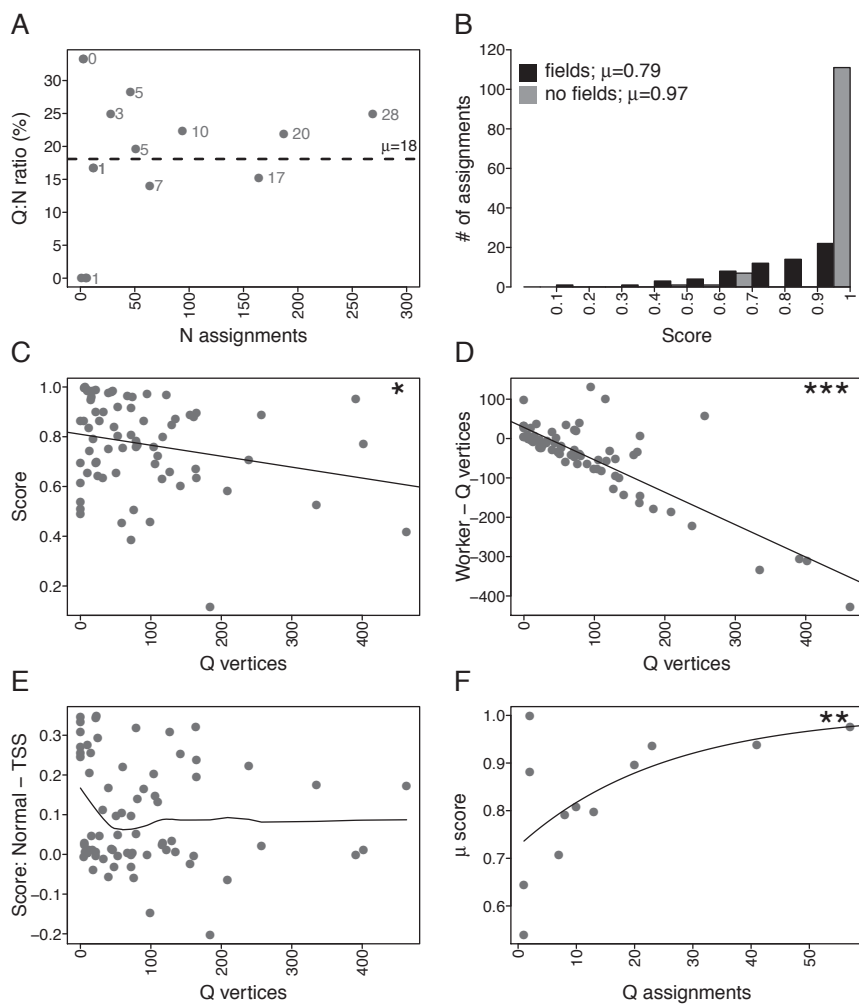


Figure 6: Results from initial DIYlandcover trial, including A) the total number of completed assignments per worker versus the ratio of Q to N type assignments (values in grey next to points represent the percent of total assignments); B) the distribution of accuracy scores segregated by assignment type (black bars = Q type, grey bars = N type); the number of vertices in reference map polygons versus C) accuracy score, D) the difference between the number of vertices in workers' and reference map polygons, and E) the difference between accuracy scores calculated using Equation 4 or Equation 5 (the True Skill Statistic); F) the number of Q assignments completed by each worker versus worker mean accuracy score. Significance of regression fits in C, D, F are: * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$. C and D are linear models, F is asymptotic regression.

478 We paid \$132.30 to workers for the 882 approved assignments, with a
479 total cost to the project of \$145.53 after accounting for Amazon.com’s 10%
480 Requester surcharge. Of this, \$28.88 was paid for the 175 approved Q as-
481 signments and \$116.66 for the 707 N assignments. Our post-hoc application
482 of the bonus algorithm, which requires workers to complete at least five Q
483 assignments (8 of 15 met this requirement), would have added \$21.89 (15%)
484 to the trial’s cost.

485 To examine the effective worker wage (i.e. the amount the worker would
486 expect at these rates assuming constant, uninterrupted work), we divided
487 total pay by the mean assignment duration, calculated as the difference be-
488 tween assignment acceptance and completion times. Since workers could
489 accept assignments without immediately completing them (maximum assign-
490 ment duration was 24 hours), we could not precisely measure mapping time.
491 However, our experience suggests that the most complicated sites require
492 <30 minutes of mapping effort, thus we excluded any assignments taking
493 longer than this. The resulting average effective wage was \$10.80 hr⁻¹ across
494 all sites, but just \$3.26 hr⁻¹ for sites having fields compared to \$13.40 hr⁻¹
495 for sites without fields. Factoring in bonus payments, these would have been
496 \$11.65 overall and \$3.55 and \$14.53 for sites with and without fields (Table
497 2).

498 The flat rate cost to map a single square kilometer was \$0.165, including
499 the cost of accuracy assessment and Amazon.com’s fees, or \$0.19 had we
500 included bonus payments.

501 *4.3. Estimating the costs of scaling up*

502 We used the time and cost results from the trial to estimate the potential
503 costs of mapping larger regions, in terms of worker payments and total map-
504 ping time, using two different payment models. One models used fixed base
505 rates (as in our trial), the other variable rates linked to potential mapping
506 effort, and in each case we tested two different levels of payments: for the
507 fixed case we used rates of \$0.15 and \$0.05²; for the variable rate, payments
508 rates were set using the following formula:

²Estimated as the approximate difference between US and South African minimum wages, <http://businesstech.co.za/news/international/87614/minimum-wage-in-south-africa-vs-the-world/>

$$R = \$0.01 + \left(\sum_{w=1}^n -1 \right) I \quad (7)$$

509 Where R is the rate, w is a categorical weight derived from a map of
 510 cropland probability, and I is an increment, set here to \$0.07 and \$0.023 for
 511 the higher and lower pay models, respectively (see Appendix S3 for meth-
 512 ods). For the cropland weights, we converted the GeoWiki 1 km² cropland
 513 percentage map (Fritz et al., 2015) into a 10 category map (where 1 = 0-10%
 514 cropland and 10 = 90-100% cropland). This map provided a more finely
 515 resolved set of weights than our four category map, and covered the entire
 516 continent. Since cropland percentages correlate positively with field area
 517 and number (albeit with area), these weights also provided a proxy mea-
 518 sure for likely mapping effort. We confirmed this assumption by extracting
 519 the new weights corresponding to the areas mapped, and used them in a
 520 least squares regression to model workers' observed mapping times ($R^2=0.1$,
 521 $p<0.0001$; Appendix S3). The map weights were extracted into a reordered
 522 vector using DIYlandcover's weighted random sampling protocol (see 3.1),
 523 and then used with Equation 7 to assign payments for each site. We added
 524 the trial mean bonus rate (\$0.023) onto these payments (and to the fixed pay
 525 rates), then calculated the cumulative cost for mapping all 29,924,000 km²
 526 of Africa for each pay model, multiplied by 1.4 to represent 1) an additional
 527 10% of mapping effort for quality assessment, and 2) administrative costs of
 528 30%.

529 To estimate the total time required to map the continent, we used the
 530 predicted mapping times resulting from the regression model. The model was
 531 run 1000 times on random subsets of the data to obtain prediction uncertain-
 532 ties for each weight level, from which the mean, 2.5th, and 97.5th percentile
 533 values were extracted. These predicted time values were assigned to their cor-
 534 responding weights in the reordered weights vector, from which mean, upper,
 535 and lower estimates of cumulative mapping hours were calculated (Appendix
 536 S3). We then created three hypothetical models of worker involvement, in
 537 which either 100, 250, or 500 workers, each mapping one hour per day, under-
 538 took the work, and used the resulting daily total mapping hours to convert
 539 the cumulative mapping time into estimates of how long it would take to
 540 map the entire continent (in years).

541 The cost model results show that variable pay rates would be considerably
 542 more efficient than fixed rate methods (Figure 7, left panel). At the trial pay

543 rates, it would cost \$7.23 million to map the entire continent, whereas it
 544 would cost just \$3.47 million using a variable pay rate, which is not much
 545 higher than \$3.04 million that would be required if pay was at the lower
 546 fixed rate. Applying the cheaper variable rate scheme, the cost would be
 547 just \$2.07 million to map all of Africa. Applying these alternate payments
 548 rates for the sites mapped during the trial reveals that variable rates would
 549 produce overall effective wages comparable to fixed rates, while paying nearly
 550 50% higher, on average, for mapping sites with fields (Table 3).

551 Total mapping time estimates vary widely according to the number of
 552 workers involved, ranging from more than 19 years to map the whole con-
 553 tinent with just 100 workers involved (i.e. 100 mapping hours per day) at
 554 the upper confidence limits of mapping time, to 1.2 - 3.8 years (mean = 1.9
 555 years) if 500 workers map.

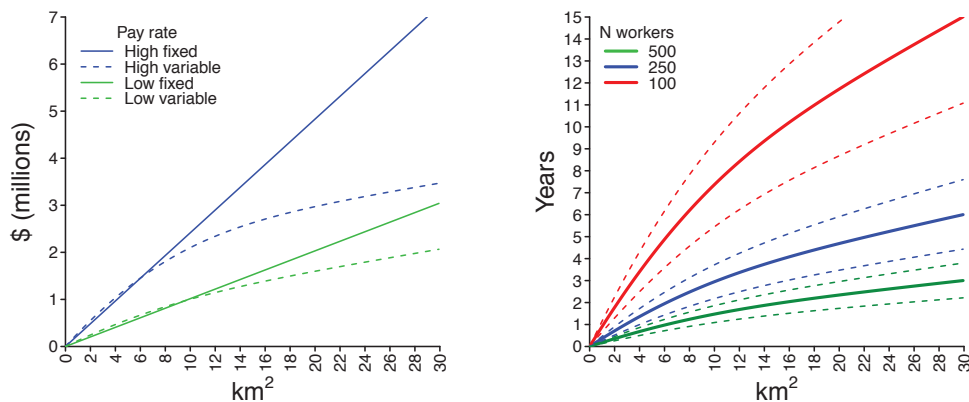


Figure 7: Several estimates of the cumulative cost (left) and time (right) of using DIYlandcover to map cropland throughout Africa. Cost estimates are based on either fixed (solid line) or variable (dashed line) rates, using higher (blue) and lower (green) cost models for each case. The cumulative mapping time is calculated in years, based on three different levels of worker involvement (100 [red], 250 [blue], and 100 [green] workers, each mapping for 1 hour/day). Solid lines indicate regression-predicted means, dashed lines the upper and lower confidence bounds for each model.

556 5. Discussion

557 The initial trial demonstrates that DIYlandcover can be an effective plat-
 558 form for generating high quality maps of discrete landcover types. This qual-
 559 ity is attributable to humans' superior ability to recognize objects in patterns

Table 2: Effective wages for workers (in $\$ \text{hr}^{-1}$) under two fixed and two variable payment schemes, calculated as overall averages and for mapping sites with and without fields, and including mean bonus rates.

Payment method	Overall	With fields	Without fields
Fixed high	11.65	3.55	14.53
Variable high	10.16	5.19	12.16
Fixed low	4.44	1.38	5.59
Variable low	4.43	2.07	5.39

560 in noisy backgrounds, and is the reason why expert image interpretation is
 561 a key component of training and assessing existing landcover mapping al-
 562 gorithms (e.g. [Fritz et al., 2011, 2012](#); [Hansen et al., 2013](#)). Here we found
 563 that workers with less than 24 hours of mapping experience were able to map
 564 cropland with 91% accuracy. Although accuracy and mapping precision de-
 565 creased when sites contained crop fields, and in proportion to the complexity
 566 of those fields (Fig. 6), the overall accuracy was higher than the latest gen-
 567 eration landcover dataset of comparable resolution (82%; [Fritz et al., 2015](#)),
 568 and was close to that achieved by GTI’s trained workers. Compared to GTI’s
 569 performance at sites with fields, using the most comparable accuracy met-
 570 ric (Equation 4), DIYlandcover showed similar performance—even though the
 571 score was 6% lower than GTI’s 95% (see Appendix S1)—because GTI mapped
 572 using a more inclusive set of rules, thereby reducing error rates, and DIY-
 573 landcover’s accuracy algorithms are more precise than the one used to assess
 574 GTI performance (see Fig. 6B and Appendix S1). The positive relationship
 575 we see between worker experience and score (Fig. 6F) also suggests that
 576 DIYlandcover’s accuracy improves with time, and we expect that the im-
 577 plementation of bonus payments for performance will also improve mapping
 578 skill. These latter two points will need to be evaluated after a lengthier period
 579 of production, as does the affect of the different accuracy component weights
 580 (Eq. 1) in terms of influencing worker—and thus system—performance.

581 The trial also suggested that DIYlandcover has the potential to generate
 582 map data relatively rapidly, given an adequate number of workers (Figure 7).
 583 With 500 workers each contributing one hour of work per day, we estimate
 584 that all of Africa could potentially be mapped in approximately 1.9 years,
 585 which is roughly six month faster than the time needed to create GTI’s map
 586 for South Africa. It is not unreasonable to think that this level of worker in-
 587 volvement is feasible on a for-pay crowdsourcing platform, particularly given

588 the payment rates we applied, which were substantially greater than the \$2
589 hr^{-1} received by Mechanical Turk workers (Marvit, 2014).

590 Our cost assessment (Fig. 7) indicates that linking pay to cover occur-
591 rence probabilities—and site selection to a finer gradation of weights—could
592 greatly reduce the overall costs of “wall-to-wall” mapping, while maintaining
593 fair worker wages. Although \$3 million is a significant amount of money, we
594 argue that it would be a fairly cheap price to pay for a vector-based map of
595 individual crop fields covering all of Africa. The overall mapping costs could
596 be reduced to \$2 million if payments are made at the lower rates we assessed.
597 We do not advise this approach when running the software on MT, as the
598 worker base is primarily in the US given Amazon’s fairly strict payment
599 rules³, and because there is growing concern about the exploitative nature of
600 crowdsourcing (Marvit, 2014). However, it may be possible to pay *fairly* at
601 lower rates if DIYlandcover is ported to job sites where workers can access
602 it from countries with lower prevailing wages. For instance, in our example,
603 had we been able to enlist workers in South Africa, where the exchange rate
604 favors the dollar, we could have paid less and had mapping undertaken by
605 workers who were familiar with those landscapes.

606 Other costs related to system development time could also be reduced,
607 particularly with respect to generating reference maps. Our trial reference
608 maps took several weeks to digitize, and were based on the judgement of a few
609 people. A third type of mapping HIT, one that allows repeated mapping by
610 multiple workers, would help mitigate these problems of cost and subjectivity.
611 The resulting maps could be combined to create a more robust “truth” based
612 on between-worker agreement, as illustrated by the combined maps from the
613 eight qualification sites used in the trial (Fig. 7). This approach could greatly
614 minimize the time required to develop reference data, and we suspect that the
615 consensus maps of many workers (which could be weighted based on worker
616 quality scores) will be more accurate than those of one or two experts (the last
617 assumption must be verified against field-collected boundary data). Another
618 advantage of this approach, which will be incorporated in the next update
619 of DIYlandcover, is that the quality assessment protocol would essentially
620 become a form of peer review.

621 Of course, the costs and necessary mapping time assessed here may still be
622 too much for many users who need spatially comprehensive, large area land-

³www.mturk.com/mturk/help?helpPage=worker

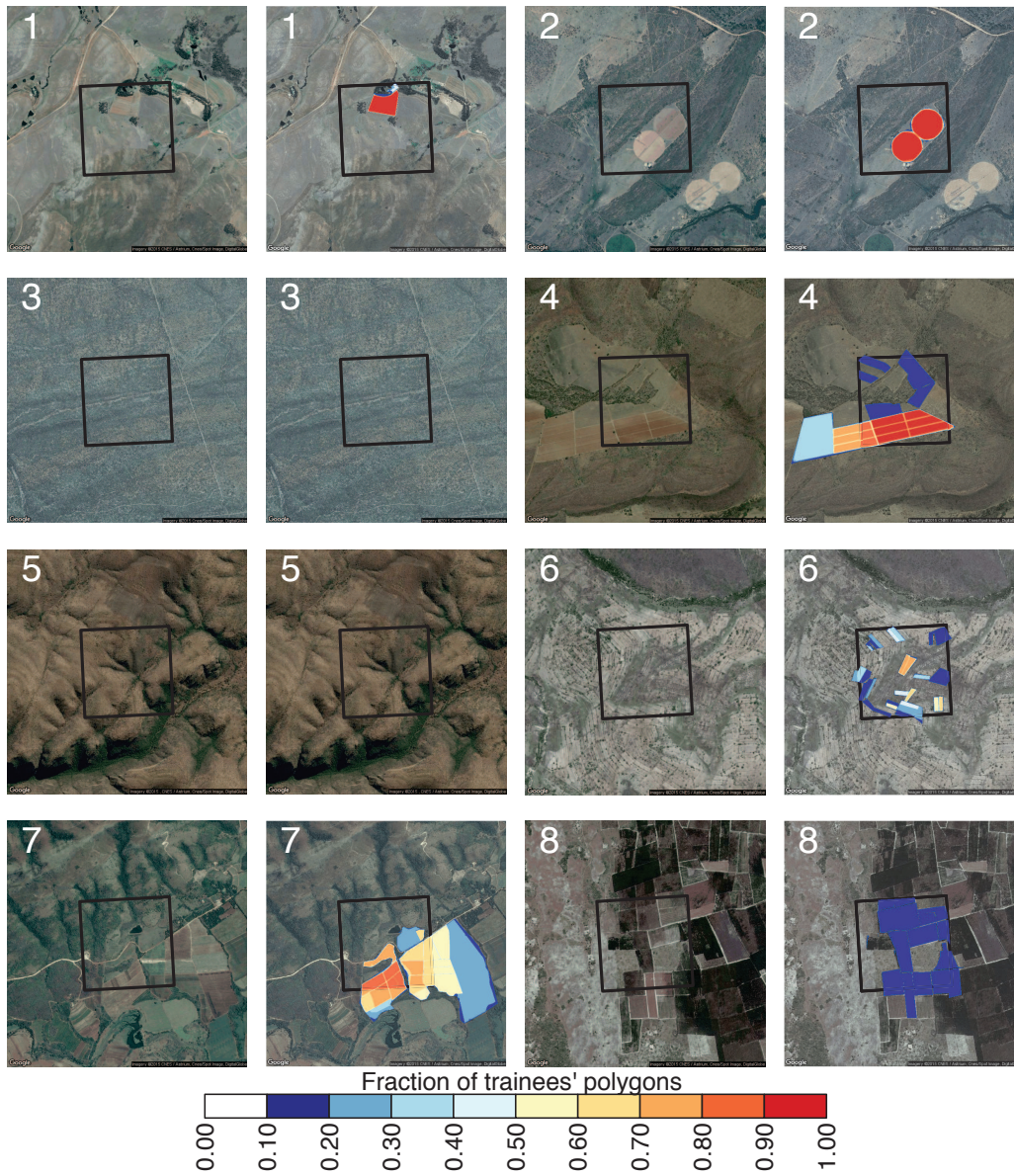


Figure 8: The eight sites used in the South Africa trial qualification test. Columns 1 and 3 show the unmapped imagery; columns 2 and 4 display the combined maps of all 19 trainees. Map colors indicate the fraction of trainee maps that overlap.

623 cover maps. To reduce costs, DIYlandcover could be ported to a server that
624 supports voluntary crowdsourcing efforts, similar to the Geo-Wiki project
625 (Fritz et al., 2012), but this would not address the problem of longer map-
626 ping times. An alternative, more advanced, application would be to use
627 DIYlandcover to train and test newer computer vision approaches for map-
628 ping noisy landcover types, such as smallholder crop fields (e.g. Debats et al.,
629 2015). In this case, DIYlandcover would work iteratively with the algorithm
630 until an acceptable level of accuracy is achieved, with site selection weighted
631 towards areas of greatest classification error after each step. This approach
632 could strike the best balance between cost, mapping speed, and accuracy, as
633 it would harness the complementary strengths of human (more effective at
634 recognizing patterns in noisy RGB or black and white images) and computer
635 image classifiers (able to extract patterns in high dimensional data, such as
636 multispectral imagery, which are hard for humans to interpret). An alterna-
637 tive possibility for this use case—where DIYlandcover validates broader-scale
638 methods—would be test and refine the judgement-based size class estimates
639 created under the Geo-Wiki project (Fritz et al., 2015). DIYlandcover is
640 highly complementary to this methodology, given its emphasis on precisely
641 measuring landcover geometries.

642 Beyond the questions of accuracy, cost, and time, the geometric de-
643 tail captured by vector boundaries is a key data dimension that is lacking
644 from current landcover products, and difficult to obtain from automated ap-
645 proaches. It is this capability to map individual features that may appeal to
646 the broadest range of potential users, as geometry data provide information
647 on a number of important social, economic, and environmental processes. For
648 instance, crop field sizes can be effective predictors of agricultural economic
649 metrics, and as such development-oriented agencies may be significantly in-
650 terested in using this tool to generate these data (Fritz et al., 2015). Al-
651 ternatively, conservationists wanting to identify habitat fragments to protect
652 may be interested in more precise boundary data, as patch geometry cor-
653 relate with extinction (Laurance et al., 2011) and thus conservation values.
654 Other potential uses include mapping buildings, burn scars, water holes, and
655 termite mounds, to name a few.

656 **Acknowledgements**

657 This work was supported by funds from the Princeton Environmental
658 Institute Grand Challenges program, the NASA New Investigator Program

659 (NNX15AC64G), and the National Science Foundation (SES-1360463 and
660 BCS-1026776).

661 **Supplementary materials**

662 Supplementary methods are presented in Appendix S1; Supporting figures
663 S1 and S2 are found in AppendixS2; Appendix S2 contains the R code used
664 to analyze trial results and create related figures.

665 **References**

666 Ahn, L. v., Blum, M., Hopper, N. J., Langford, J., 2003. CAPTCHA: Using
667 Hard AI Problems for Security. In: Biham, E. (Ed.), *Advances in Cryptol-*
668 *ogy EUROCRYPT 2003*. No. 2656 in *Lecture Notes in Computer Science*.
669 Springer Berlin Heidelberg, pp. 294–311.

670 Allahbakhsh, M., Benatallah, B., 2013. Quality control in crowdsourcing
671 systems. *IEEE Internet Computing* 17 (2), 76–81.

672 Allouche, O., Tsoar, A., Kadmon, R., Dec. 2006. Assessing the accuracy of
673 species distribution models: prevalence, kappa and the true skill statistic
674 (TSS). *Journal of Applied Ecology* 43 (6), 1223–1232.

675 Biederman, I., 1987. Recognition-by-components: a theory of human image
676 understanding. *Psychological review* 94 (2), 115.

677 Bivand, R., Rundel, C., 2013. rgeos: interface to geometry engine-
678 open source (GEOS). R package ver. 0.33.< [http://cran.r-project.](http://cran.r-project.org/web/packages/rgeos/index.html)
679 [org/web/packages/rgeos/index.html](http://cran.r-project.org/web/packages/rgeos/index.html).

680 Conway, J., Eddelbuettel, D., Nishiyama, T., Prayaga, S. K., Tiffin, N., 2012.
681 RPostgreSQL: R interface to the PostgreSQL database system (2010). R
682 package version 0.1-7.

683 Debats, S., Luo, D., Estes, L., Fuchs, T., Caylor, K., 2015. A generalized
684 computer vision approach to mapping agricultural fields in Sub-Saharan
685 Africa. *PeerJ PrePrints* 3, e1688.

686 Developers, G., 2012. Google Maps API.

- 687 Estes, L. D., Beukes, H., Bradley, B. A., Debats, S. R., Oppenheimer, M.,
688 Ruane, A. C., Schulze, R., Tadross, M., Dec. 2013a. Projected climate im-
689 pacts to South African maize and wheat production in 2055: a comparison
690 of empirical and mechanistic modeling approaches. *Global Change Biology*
691 19 (12), 3762–3774.
- 692 Estes, L. D., Bradley, B. A., Beukes, H., Hole, D. G., Lau, M., Oppenheimer,
693 M. G., Schulze, R., Tadross, M. A., Turner, W. R., Aug. 2013b. Compar-
694 ing mechanistic and empirical model projections of crop suitability and
695 productivity: implications for ecological forecasting. *Global Ecology and*
696 *Biogeography* 22 (8), 1007–1018.
- 697 Estes, L. D., Paroz, L.-L., Bradley, B. A., Green, J. M., Hole, D. G., Hol-
698 ness, S., Ziv, G., Oppenheimer, M. G., Wilcove, D. S., Apr. 2014. Using
699 changes in agricultural utility to quantify future climate-induced risk to
700 conservation. *Conservation Biology* 28 (2), 427–437.
- 701 Flanagan, A. J., Metzger, M. J., Jul. 2008. The credibility of volunteered
702 geographic information. *GeoJournal* 72 (3-4), 137–148.
- 703 Fraternali, P., Castelletti, A., Soncini-Sessa, R., Ruiz, C. V., Rizzoli, A. E.,
704 2012. Putting humans in the loop: Social computing for Water Resources
705 Management. *Environmental Modelling & Software* 37, 68–77.
- 706 Fritz, S., McCallum, I., Schill, C., Perger, C., Grillmayer, R., Achard, F.,
707 Kraxner, F., Obersteiner, M., Aug. 2009. Geo-Wiki.Org: The Use of
708 Crowdsourcing to Improve Global Land Cover. *Remote Sensing* 1 (3), 345–
709 354.
- 710 Fritz, S., McCallum, I., Schill, C., Perger, C., See, L., Schepaschenko, D.,
711 van der Velde, M., Kraxner, F., Obersteiner, M., May 2012. Geo-Wiki: An
712 online platform for improving global land cover. *Environmental Modelling*
713 *& Software* 31, 110–123.
- 714 Fritz, S., See, L., 2008. Identifying and quantifying uncertainty and spatial
715 disagreement in the comparison of Global Land Cover for different appli-
716 cations. *Global Change Biology* 14 (5), 1057–1075.
- 717 Fritz, S., See, L., McCallum, I., You, L., Bun, A., Moltchanova, E., Duer-
718 auer, M., Albrecht, F., Schill, C., Perger, C., Havlik, P., Mosnier, A.,

- 719 Thornton, P., Wood-Sichra, U., Herrero, M., Becker-Reshef, I., Justice,
720 C., Hansen, M., Gong, P., Abdel Aziz, S., Cipriani, A., Cumani, R., Cec-
721 chi, G., Conchedda, G., Ferreira, S., Gomez, A., Haffani, M., Kayitakire,
722 F., Malanding, J., Mueller, R., Newby, T., Nonguierma, A., Olusegun, A.,
723 Ortner, S., Rajak, D. R., Rocha, J., Schepaschenko, D., Schepaschenko,
724 M., Terekhov, A., Tiangwa, A., Vancutsem, C., Vintrou, E., Wenbin, W.,
725 van der Velde, M., Dunwoody, A., Kraxner, F., Obersteiner, M., 2015.
726 Mapping global cropland and field size. *Global Change Biology* 21 (5),
727 1980–1992.
- 728 Fritz, S., See, L., Rembold, F., 2010. Comparison of global and regional
729 land cover maps with statistical information for the agricultural domain
730 in Africa. *International Journal of Remote Sensing* 31 (9), 2237–2256.
- 731 Fritz, S., You, L., Bun, A., See, L., McCallum, I., Schill, C., Perger, C., Liu,
732 J., Hansen, M., Obersteiner, M., 2011. Cropland for sub-Saharan Africa: A
733 synergistic approach using five land cover data sets. *Geophysical Research*
734 *Letters* 38, L04404.
- 735 Frhlich, B., Bach, E., Walde, I., Hese, S., Schmullius, C., Denzler, J., 2013.
736 Land cover classification of satellite images using contextual information.
737 *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Infor-*
738 *mation Sciences* 3, W1.
- 739 GeoTerraImage, 2008. South African crop field boundaries. Tech. rep.,
740 <http://www.geoterraimage.com>.
- 741 Gibbs, H. K., Ruesch, A. S., Achard, F., Clayton, M. K., Holmgren, P.,
742 Ramankutty, N., Foley, J. A., Sep. 2010. Tropical forests were the primary
743 sources of new agricultural land in the 1980s and 1990s. *Proceedings of the*
744 *National Academy of Sciences* 107 (38), 16732–16737.
- 745 Gross, D., Dubois, G., Pekel, J.-F., Mayaux, P., Holmgren, M., Prins, H.,
746 Rondinini, C., Boitani, L., Mar. 2013. Monitoring land cover changes in
747 African protected areas in the 21st century. *Ecological Informatics* 14, 31–
748 37.
- 749 Hansen, M. C., Potapov, P. V., Moore, R., Hancher, M., Turubanova, S. A.,
750 Tyukavina, A., Thau, D., Stehman, S. V., Goetz, S. J., Loveland, T. R.,
751 Kommareddy, A., Egorov, A., Chini, L., Justice, C. O., Townshend, J.

- 752 R. G., Nov. 2013. High-Resolution Global Maps of 21st-Century Forest
753 Cover Change. *Science* 342 (6160), 850–853.
- 754 Hardy, M., Dziba, L., Kilian, W., Tolmay, J., 2011. Rainfed Farming Systems
755 in South Africa. In: Tow, P., Cooper, I., Partridge, I., Birch, C. (Eds.),
756 Rainfed Farming Systems. Springer Netherlands, pp. 395–432.
- 757 Howe, J., 2006. The rise of crowdsourcing. *Wired magazine* 14 (6), 1–4.
- 758 Jacobson, A., Dhanota, J., Godfrey, J., Jacobson, H., Rossman, Z., Stan-
759 ish, A., Walker, H., Riggio, J., Oct. 2015. A novel approach to mapping
760 land conversion using Google Earth with an application to East Africa.
761 *Environmental Modelling & Software* 72, 1–9.
- 762 Jain, M., Mondal, P., DeFries, R. S., Small, C., Galford, G. L., Jul. 2013.
763 Mapping cropping intensity of smallholder farms: A comparison of meth-
764 ods using multiple sensors. *Remote Sensing of Environment* 134, 210–223.
- 765 Laurance, W. F., Camargo, J. L., Luizao, R. C., Laurance, S. G., Pimm, S. L.,
766 Bruna, E. M., Stouffer, P. C., Bruce Williamson, G., Bentez-Malvido, J.,
767 Vasconcelos, H. L., Van Houtan, K. S., Zartman, C. E., Boyle, S. A., Did-
768 ham, R. K., Andrade, A., Lovejoy, T. E., Jan. 2011. The fate of Amazonian
769 forest fragments: A 32-year investigation. *Biological Conservation* 144 (1),
770 56–67.
- 771 Liang, X., Lettenmaier, D. P., Wood, E. F., Burges, S. J., 1994. A simple hy-
772 drologically based model of land surface water and energy fluxes for general
773 circulation models. *Journal of Geophysical Research* 99 (D7), 14415.
- 774 Licker, R., Johnston, M., Foley, J. A., Barford, C., Kucharik, C. J., Mon-
775 freda, C., Ramankutty, N., Nov. 2010. Mind the gap: how do climate and
776 agricultural management explain the yield gap of croplands around the
777 world? *Global Ecology and Biogeography* 19 (6), 769–782.
- 778 Lintott, C. J., Schawinski, K., Slosar, A., Land, K., Bamford, S., Thomas,
779 D., Raddick, M. J., Nichol, R. C., Szalay, A., Andreescu, D., Murray, P.,
780 Vandenberg, J., Sep. 2008. Galaxy Zoo: morphologies derived from visual
781 inspection of galaxies from the Sloan Digital Sky Survey. *Monthly Notices*
782 of the Royal Astronomical Society 389 (3), 1179–1189.

- 783 Lobell, D. B., Mar. 2013. The use of satellite data for crop yield gap analysis.
784 Field Crops Research 143, 56–64.
- 785 Marvit, M. Z., Feb. 2014. How Crowdworkers Became the Ghosts in the
786 Digital Machine. The Nation.
- 787 Monfreda, C., Ramankutty, N., Foley, J. A., 2008. Farming the planet: 2.
788 Geographic distribution of crop areas, yields, physiological types, and net
789 primary production in the year 2000. Global Biogeochemical Cycles 22,
790 GB1022.
- 791 Ohori, K. A., Ledoux, H., Meijers, M., Oct. 2012. Validation and automatic
792 repair of planar partitions using a constrained triangulation. Photogram-
793 metrie - Fernerkundung - Geoinformation 2012 (5), 613–630.
- 794 Ozdarici-Ok, A., Akyurek, Z., Sep. 2014. Object-Based Classification of
795 Multi-temporal Images for Agricultural Crop Mapping in Karacabey Plain,
796 Turkey. ISPRS - International Archives of the Photogrammetry, Remote
797 Sensing and Spatial Information Sciences XL-7, 127–132.
- 798 Ozdarici-Ok, A., Ok, A. O., Schindler, K., May 2015. Mapping of Agricul-
799 tural Crops from Single High-Resolution Multispectral ImagesData-Driven
800 Smoothing vs. Parcel-Based Smoothing. Remote Sensing 7 (5), 5611–5638.
- 801 Ramankutty, N., Evan, A. T., Monfreda, C., Foley, J. A., Jan. 2008. Farming
802 the planet: 1. Geographic distribution of global agricultural lands in the
803 year 2000. Global Biogeochemical Cycles 22, 19 PP.
- 804 Ruesch, A., Gibbs, H. K., 2008. New IPCC Tier-1 global biomass
805 carbon map for the year 2000. Carbon Dioxide Information
806 Analysis Center (CDIAC), Oak Ridge National Laboratory,
807 Oak Ridge, Tennessee. Available online at: [http://cdiac.ornl.
808 gov/epubs/ndp/global_carbon/carbon_documentation.html](http://cdiac.ornl.gov/epubs/ndp/global_carbon/carbon_documentation.html).
- 809 Rulli, M. C., Savioli, A., D’Odorico, P., Jan. 2013. Global land and water
810 grabbing. Proceedings of the National Academy of Sciences 110 (3), 892–
811 897.
- 812 Schellekens, J., Brolsma, R., Dahm, R., Donchyts, G., Winsemius, H.,
813 Nov. 2014. Rapid setup of hydrological and hydraulic models using Open-
814 StreetMap and the SRTM derived digital elevation model. Environmental
815 Modelling & Software 61, 98–105.

- 816 Schroff, F., Criminisi, A., Zisserman, A., 2008. Object class segmentation
817 using random forests. *British Machine Vision Association*, pp. 54.1–54.10.
- 818 Searchinger, T. D., Estes, L., Thornton, P. K., Beringer, T., Notenbaert, A.,
819 Rubenstein, D., Heimlich, R., Licker, R., Herrero, M., May 2015. High
820 carbon and biodiversity costs from converting Africa’s wet savannahs to
821 cropland. *Nature Climate Change* 5 (5), 481–486.
- 822 See, L., Fritz, S., You, L., Ramankutty, N., Herrero, M., Justice, C., Becker-
823 Reshef, I., Thornton, P., Erb, K., Gong, P., Tang, H., van der Velde,
824 M., Ericksen, P., McCallum, I., Kraxner, F., Obersteiner, M., Mar. 2015.
825 Improved global cropland data as an essential ingredient for food security.
826 *Global Food Security* 4, 37–45.
- 827 Services, A. W., 2012. Amazon Mechanical Turk.
- 828 Sullivan, B. L., Wood, C. L., Iliff, M. J., Bonney, R. E., Fink, D., Kelling, S.,
829 Oct. 2009. eBird: A citizen-based bird observation network in the biological
830 sciences. *Biological Conservation* 142 (10), 2282–2292.
- 831 Tokarczyk, P., Wegner, J., Walk, S., Schindler, K., Jan. 2015. Features, Color
832 Spaces, and Boosting: New Insights on Semantic Classification of Remote
833 Sensing Images. *IEEE Transactions on Geoscience and Remote Sensing*
834 53 (1), 280–295.
- 835 Yan, L., Roy, D. P., Mar. 2014. Automated crop field extraction from multi-
836 temporal Web Enabled Landsat Data. *Remote Sensing of Environment*
837 144, 42–64.