Article

Crowd-Driven and Automated Mapping of Field Boundaries in Highly Fragmented Agricultural Landscapes of Ethiopia with Very High Spatial Resolution Imagery

Michael Marshall [1],*[ID]**, Sophie Crommelinck** [1][ID]**, Divyani Kohli** [1]**, Christoph Perger** [2][ID]**,
Michael Ying Yang** [1][ID]**, Aniruddha Ghosh** [3][ID]**, Steffen Fritz** [2]**, Kees de Bie** [1][ID] **and Andy Nelson** [1][ID]

[1]   Faculty of Geo-Information Science and Earth Observation (ITC), University of Twente, Hengelosestraat 99, 7514 AE Enschede, The Netherlands

[2]   Ecosystem Services and Management Program, International Institute for Applied Systems Analysis (IIASA), Schlossplatz 1, A-2361 Laxenburg, Austria

[3]   Environmental Science & Policy, University of California-Davis, Wickson Hall, 350 E Quad, Davis, CA 95616, USA

*   Correspondence: m.t.marshall@utwente.nl; Tel.: +31-534897193

check for updates

Abstract: Mapping the extent and location of field boundaries is critical to food security analysis but remains problematic in the Global South where such information is needed the most. The difficulty is due primarily to fragmentation in the landscape, small farm sizes, and irregular farm boundaries. Very high-resolution satellite imagery affords an opportunity to delineate such fields, but the challenge remains of determining such boundaries in a systematic and accurate way. In this paper, we compare a new crowd-driven manual digitization tool (Crop Land Extent) with two semi-automated methods (contour detection and multi-resolution segmentation) to determine farm boundaries from WorldView imagery in highly fragmented agricultural landscapes of Ethiopia. More than 7000 one square-kilometer image tiles were used for the analysis. The three methods were assessed using quantitative completeness and spatial correctness. Contour detection tended to under-segment when compared to manual digitization, resulting in better performance for larger (approaching 1 ha) sized fields. Multi-resolution segmentation on the other hand, tended to over-segment, resulting in better performance for small fields. Neither semi-automated method in their current realizations however are suitable for field boundary mapping in highly fragmented landscapes. Crowd-driven manual digitization is promising, but requires more oversight, quality control, and training than the current workflow could allow.

Keywords: agriculture; cropland; food security; image segmentation; object detection; crowdsourcing; remote sensing; WorldView

1. Introduction

Efforts to better target research and the extension of sustainable agriculture require basic information on the extent and location of fields, which are the fundamental land management units on which decisions are made regarding what, when, and how to grow crops [1]. More specifically, accurate information on individual field size or crop area enhances the estimation of current and potential productivity of a system and opportunities of yield gains. Crop area at the level of detail where individual fields can be discriminated is poorly estimated in regions of the world where field sizes are small (<2 ha) and agricultural landscapes are highly fragmented [2,3]. These regions are

primarily in the Global South where efforts are most needed to sustainably intensify agricultural systems to close prevalent yield gaps, boost productivity, and improve livelihoods [4]. Typically, information on crop area is collected through a national agricultural census or other ground survey. Unfortunately, the collection of these data is expensive and time consuming. Second, census reports often aggregate crop area information to different administrative units, making it impossible to relate individual fields with corresponding land management, agricultural practices, and production. Third, surveys based on statistical sampling schemes may not be representative of the population in complex landscapes. Earth observation (EO) imagery is often used to generate information about the physical aspects of agricultural production because the workflow can be largely automated and it can provide low-cost and consistent estimates of surface conditions over large areas through time [5,6]. Moderate (~30 m) resolution EO imagery has been used to map field boundaries for crop area estimation in large fields with regular geometries [7,8], though it is too coarse to delineate small fields with irregular geometries [9]. The emergence of very high spatial resolution (≤5 m) commercial satellites promises to overcome this obstacle but, to date, few studies have evaluated their effectiveness for wall-to-wall national coverage [10].

A major trade-off persists since the widespread use of EO imagery for crop area estimation at the national scale began with such programs as the Large Area Crop Inventory Experiment [11–13] and the Agriculture and Resources Inventory Surveys [14,15]. Costly ground surveys are required to calibrate remote sensing-based models to achieve acceptable levels of accuracy [16]. Husak et al. [17] and others [18–20] proposed a method that attempts to strike a balance between the cost and accuracy of field mapping in highly fragmented landscapes. It largely uses very high spatial resolution (VHR) EO imagery in lieu of ground data to train and test crop area models. The models are built on the concept of grid point sampling frames. A sample frame in this case, consists of a grid of "crop" and "no crop" manual interpretations from VHR imagery. The frames are scaled to a coarser resolution as proportions/probabilities of crop area using freely available geospatial data as model inputs. The success of this method varies considerably as cross-validated errors range from <2% to more than 36% [20]. The errors are largely attributed to the subjectivity of manual interpretation.

Crowdsourcing, first demonstrated for crop area estimation by Fritz et al. [21], has the potential to reduce the subjectivity of manual interpretation of VHR imagery by vastly increasing the number of interpreters and amount of data interpreted. The premise being, as the number of interpreters and data interpreted increases, the influence of outliers is reduced. Crowdsourcing in Fritz et al. [22] was facilitated by the Geo-Wiki platform (https://www.geo-wiki.org). As with other grid point sampling techniques, interpreters classify grid cells superimposed on EO images available through Google Maps®as "crop" or "not crop" [21]. The grids are used to calculate the probability of crop area. Unlike other manual techniques, interpreters perform the operation online and are assisted by online training materials. This enables the interpretation of hundreds or thousands of sample frames in the data cloud. The frames have been used for validation of other models or interpolated to provide a global surface of crop area. Recent campaigns have aimed to address the shortcomings of this technique, namely the low density of sample frames in both space and time, as well as insufficient quality control/assurance (see Lesiv et al. [23] for a recent example). New tools such as "do-it-yourself" (DIY) landcover can be used to map discrete field boundaries instead of gridded crop area probabilities [24]. This is advantageous, because decisions are typically made at the field level.

Yet another approach attempts to eliminate the impact of subjectivity in manual interpretation of VHR imagery by automating the interpretation process. A few automated methods, such as Debats et al. [25], estimate crop area probabilities from VHR imagery. In general, however, object-based image classification is preferred over other automated methods, because it structures image content spatially and semantically instead of spectrally [26]. This is advantageous, because VHR imagery tends to have lower spectral depth than other EO imagery. Pixels of similar characteristics in color, tone, texture, shadow, or semantics are grouped to high-level features. These high-level features incorporate model-driven knowledge and scene understanding. Object-based image classification typically derives

features on a pixel-by-pixel basis (i.e., edge and contour detection) to form spectrally homogeneous regions (image segmentation) [27]. Grouping pixels can contribute to image analysis once the object of interest is larger than the spatial resolution of the image [28]. The main limitation of these methods is that they tend to over-segment due to within-field spectral variability [29]. In addition, the degree of segmentation depends heavily on how similarity parameters are tuned a priori. Pixels can be aggregated to super-pixels to smooth out variations and reduce over-segmentation. Neigh et al. [30] applied smoothing kernels of varying window sizes before segmenting field boundaries from WorldView imagery in Ethiopia. Crommelinck et al. [31] used simple linear iterative clustering (SLIC) based on color similarity and pixel proximity to generate super-pixels from unmanned aerial system imagery. The super-pixels were combined with image contours to delineate cadastral boundaries. Garcia-Pedrero et al. [29] also used SLIC to delineate farm boundaries but extended it to consider spectral depth as well as color and space.

Crowdsourcing and automated object-based methods have been widely used to map field boundaries in the Global South with VHR imagery. Each method has its advantages and disadvantages. Crowdsourcing reduces subjectivity in manual classification and can provide wall-to-wall national coverage, but often suffers from under-sampling and poor-quality control/assurance. Automated object-based techniques are also less subjective, but typically require localized parameterization that hampers operationalization for wall-to-wall national coverage. Neither method has been compared to one another in a practical sense, i.e., from inception to completion of a large area assessment of field boundaries in the Global South. The purpose of this study therefore was to make such a practical comparison. The study used thousands of VHR (WorldView) imagery within a new online digitizing platform (CLE: Crop Land Extent; https://geo-wiki.org/cle) and trained campaign volunteers to manually digitize field boundaries in important agroecosystems of Ethiopia. Two common automated methods (contour detection and multi-resolution segmentation) were also employed for comparison purposes. The field boundaries were subsequently used as input for drought insurance pricing. The paper demonstrates the challenge of applying these methods to highly variable, smallholder farming systems in Africa where field sizes are small, prior information is poor, and where boundaries are unclear. Opportunities for these methods and their processing chains to be improved in future work were identified. The recommendations presented provide some points of departure for further research on the applicability of these methods before they can be considered for operational use.

2. Study Area

The comparison was performed across ten Woredas (districts) in the Amhara Highlands and Central Oromia region of Ethiopia (Figure 1). Both regions have experienced rapid population growth, land degradation from the expansion of farming and pastoralism, and drying/increased frequency of droughts due to climate change [32]. The Amhara Highlands are adjacent to Lake Tana, which is the main source of the Blue Nile [33]. Like other regions of the Ethiopian plateau, it resulted from rifting or spreading along the Central Rift Valley. The highlands have high soil fertility, mild temperatures (mean daily = 15 to 25 °C daily average), and ample rainfall (mean annual = 750 mm) (http://www.ethiopia.gov.et). These characteristics make it highly desirable for farming and animal husbandry, which together are the primary livelihood of more than 85% of the population. The human demand for arable land, complex topography, and unequitable land tenure, results in a highly fragmented landscape and small farm sizes (average < 1 ha) (Figure 2) [34]. Rainfall occurs primarily during the *Kiremt* (June–September) and secondarily during the *Belg* (February–May) season. Mixed cropping and intercropping are common. Teff is the main staple crop in the Amhara Highlands and it is one of the major teff-producing regions of Ethiopia. Other major staple crops grown in the region include barley, finger millet, maize, oats, pulses, sorghum, and wheat. The staple crops are typically cultivated during the *Kiremt* season to assure a long, but high-yielding growth period, while short duration crops, such as potatoes and yams, are cultivated during the *Belg* season. The central region of Oromia is located just southeast of the nation's capital (Addis Ababa). It is less mountainous

and climatically more variable than the Amhara Highlands due to its close proximity to the Central Rift Valley. As a result, farm sizes tend to be larger and can approach 10 ha in size. The climate transitions from hot (mean daily temperature = 27 to 39 °C) and dry (mean annual rainfall < 450 mm) in the valley to warm (mean daily temperature = 18 to 27 °C) and wet (mean annual rainfall = 450 to 820 mm) away from the valley. Similar to the Amhara Highlands, mixed cropping and intercropping is common in wetter areas. Farmers in these areas grow teff and other cereals, grains, and pulses during the *Kiremt* season. In drier areas, maize, and horse bean are grown during the *Kiremt* season instead.
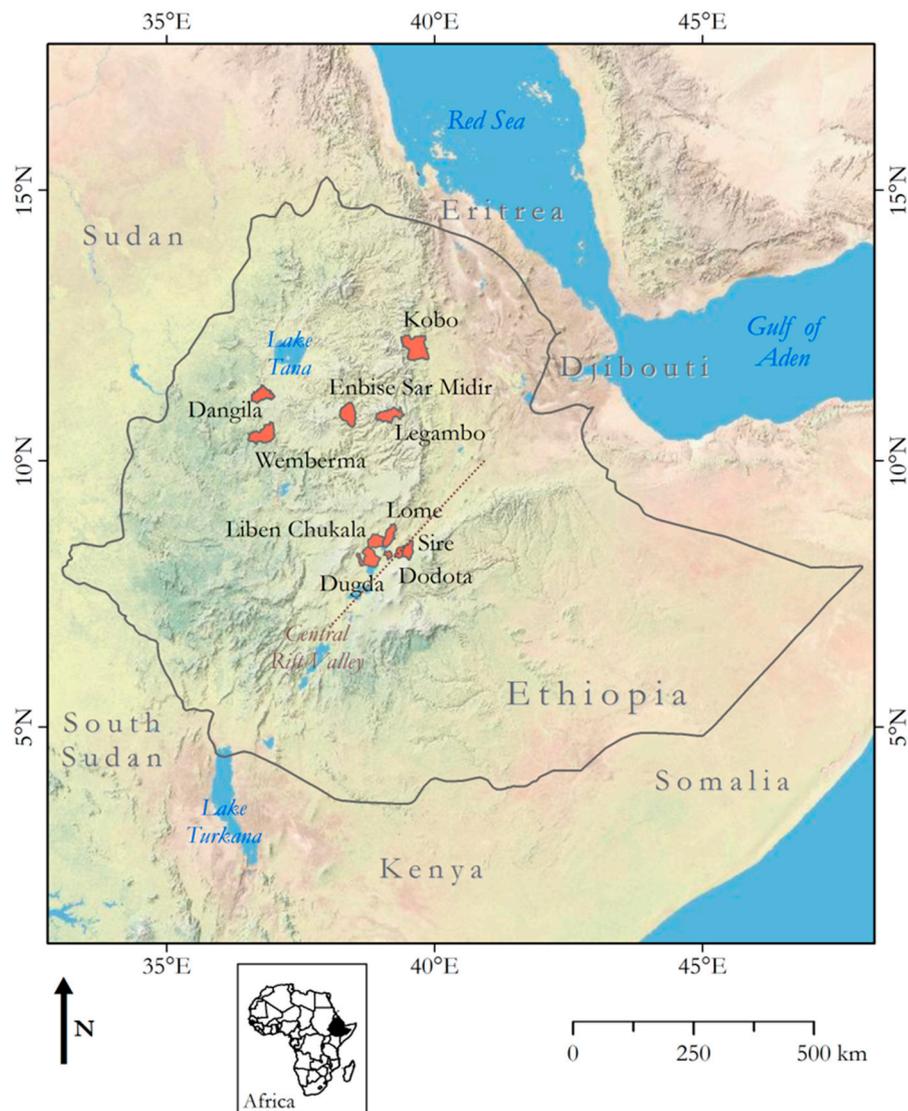
Figure 1. Ten woredas (sub-districts) in the Amhara Highlands (around Lake Tana) and Central Oromia region (adjacent to the Central Rift Valley) over which boundary mapping methods were evaluated using WorldView 2 and 3 imagery taken from 2009–2016.
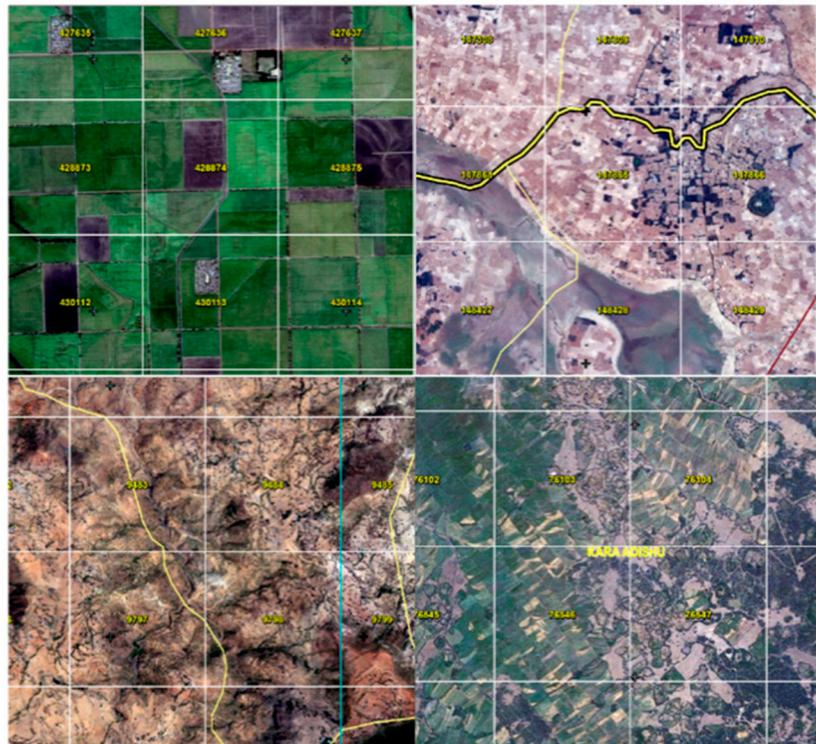
Figure 2. Examples of different cropland areas in Ethiopia illustrating the diversity in field size, shape, and delineation. Based on very high-resolution imagery (0.5 m/pixel) overlaid with 1 km × 1 km grids. Map data © 2016 Google.

3. Materials and Methods

The Woredas were selected, because they are within the agricultural commercialization clusters (ACCs) of Ethiopia: http://www.ata.gov.et. The clusters are used to focus locally relevant investments for the development of high-value commodities. In this case, Kifiya Financial Technology PLC, a digital finance and payment services provider in Ethiopia, as well as business partner in an innovative drought peril insurance scheme (GIACIS: Geodata for Innovative Agricultural Credit Insurance Schemes; https://g4aw.spaceoffice.nl/en/projects/g4aw-projects/64/geodata-for-innovative-agricultural-credit-insurance-schemes-giacis-.html) was keen to understand the potential of VHR imagery to identify field boundaries in areas where they insure agricultural credit (small loans to smallholders farmers at the start of the season) against drought in the ACCs [35]. A random subset of the images was used to train three general approaches to extract field boundaries. A total of 7200 boundaries were delineated (Table 1). Two semi-automated techniques (contour detection and multi-resolution segmentation) were compared against CLE with two performance metrics (quantitative completeness and spatial correctness). Quantitative completeness answers, "How complete are the entire fields extracted?" and spatial correctness answers, "How correct is the extraction in a spatial sense?"

Table 1. The number of field boundaries (N) manually digitized using CLE for each Woreda.

Woreda	N
Wemberma	592
Dugda	1338
Dodota	253
Dangila	1505
Kobo	224
Legambo	198
Lome	1244
Liben Chukala	887
Sire	712
Enbise Sar Midir	247

3.1. Image Acquisition and Processing

Over 60,000 WorldView 2 and 3 image one square-kilometer tiles were acquired through a collective agreement between the University of California, Davis and the International Institute for Applied Systems Analysis (IIASA) with DigitalGlobe web services. The image tiles represented surface conditions of the Woredas in Ethiopia from 2009–2016. A random subset of 533 unique image tiles that intersected with the Woredas were selected for further analysis. Since multiple image tiles were available for the same location over the eight-year period, the subset consisted of the image tile closest to the *Kiremt* season with ≤20% cloud cover. Imagery within the *Kiremt* season were generally unavailable due to persistent cloud cover. In addition, recent images were selected preferentially over older images. For this reason, the semi-automated methods were performed after manual digitization to ensure that each approach was applied to the same images. The image tiles consisted of georectified clips of WorldView 2 or 3 true color composites (red, green, and blue). Other WorldView spectral bands were not used, because they were not available under the consortium agreement. The image tiles were provided at ~0.5 m spatial resolution in a Universal Transverse Mercator (zone 37N) projection (datum = WGS 84). They underwent radiometric (top-of-atmosphere) correction prior to the analysis.

3.2. Reference Data (REF) Creation with Manual Digitization

Field boundaries were delineated in each image tile with CLE by eleven trained campaign volunteers in Ethiopia; all were employees of Kifiya. Kifiya is an Ethiopian insurance company that operates a geodata-driven drought peril insurance product for smallholder farmers. The digitizers were all computer literate with some form of tertiary education. Some digitizers were more familiar with the geographic regions, which helped them interpret the images more quickly. This knowledge was shared during feedback sessions in the workshop. Like DIY landcover, CLE produces field boundaries instead of probabilities. As with Geo-Wiki, CLE is built on the open-source mapping framework OpenLayers. A key feature of the tool is that it can easily be updated to accommodate user needs. After the volunteers were registered with CLE and trained, they could begin digitizing the image tiles. CLE displays the image tiles at random and assures at least three volunteers digitize each image tile. Unlike Geo-Wiki, basic GUI tools are used in CLE to draw and digitize field boundaries over each image tile, instead of classifying superimposed grids as crop or not crop. Following the Joint Experiment for Crop Assessment and Monitoring guidelines for cropland definition and field data collection [36], field boundaries were defined as an enclosed area (≥0.3 ha) of annual crops. These areas consist primarily (>30%) of herbaceous vegetation cover but can also include some (<20%) tree or woody vegetation cover. They do not include fallow or pastureland.

The volunteers were trained on CLE in Addis Ababa on 18–19 May 2017. They were also provided with an eight-page user manual, which included a basic tutorial, digitizing rules, and example image tiles with clarifications. The training was designed to bring all volunteers to a common level of digitizing consistency and quality with the aim that only field boundaries for crops, adjacent fields or

fields containing more than one crop were accurately digitized. The training included three practical sessions to build up the expertise of the volunteers and expose them to increasingly complex digitizing tasks (see the Discussion section for more details). Among other training activities, the volunteers were given examples of image tiles that were too difficult to digitize due to cloud cover or other inconsistencies. If the image was skipped, it was removed from the analysis and an earlier image was selected in its place. Digitization attempts by subsequent participants were "snapped" to the first attempt to reduce minor geometric inconsistencies. Feedback sessions were held after each practical session to compare digitizing outputs and interpretation between volunteers. The sessions were very interactive and resulted in the following digitizing rules:

- Digitize crop field boundaries (not pastures or other delineated structures).
- Map two fields where there is a clear boundary visible between them.
- A field may have more than one crop (meaning different color) in it but it is still one field.
- Patterns within the field will help you to find the boundaries, if the pattern changes it is a new field.
- Lines of trees can often be boundaries, especially in hillside terraces where a field consists of several terraces.
- Approximate the boundaries, it does not have to be pixel-perfect.
- If it is not clear, then do not digitize boundaries; in other words, do not guess.

The digitizing took place between June and August 2017. Digitizing progress was monitored regularly on the CLE website (https://geo-wiki.org/Application/code/cle.html) where the progress per volunteer was recorded and where the digitized boundaries could be downloaded and viewed at any time during those three months. The project lead visually assessed the quality of the digitizing to date and referred to this in monthly calls between the project lead, the CLE developer and the local coordinator of the volunteer team. These calls focused on issues with regards to quality and adherence to the digitizing rules. At the same time, the local coordinator would report any problems faced by the volunteers or request clarification for cases that had not been considered during the training. The digitized field boundaries were stored and geo-tagged, so that they could easily be imported into a Geographic Information System (GIS) geodatabase.

The digitized field boundaries were imported into a geodatabase and merged into one master vectorized dataset (REF) for comparison with the semi-automated techniques. Despite the various quality control steps (training, feedback and monitoring) variations in boundary delineation skill across the volunteers were still observed (Figure 3) so additional steps were taken before the comparison.
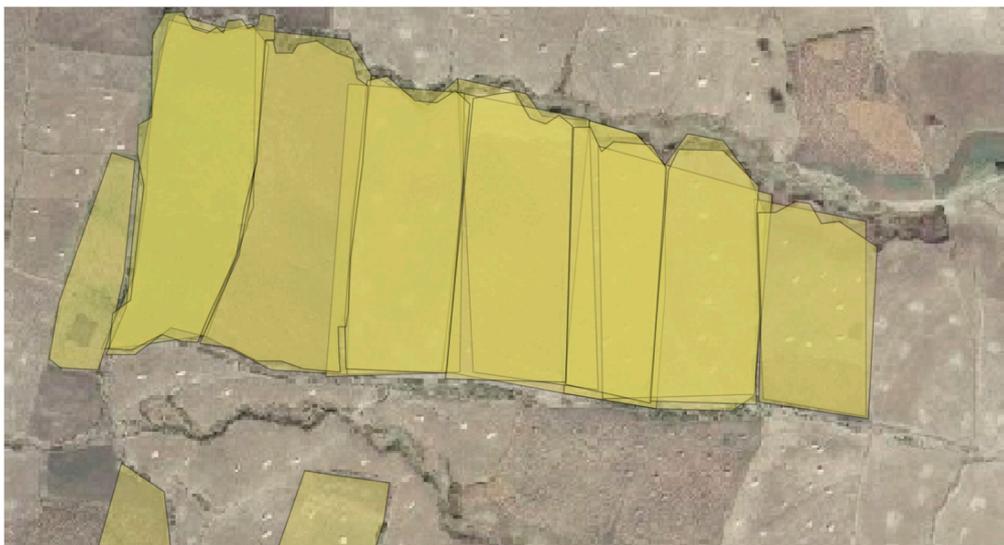
Figure 3. An example of multiple field boundaries digitized in CLE for the same image tile and displayed in a GIS. The digitized polygons are semi-transparent, with darker colors revealing where a field has been digitized by more than one Kifiya participant. Imagery © 2011 DigitalGlobe, Inc.

First, the boundaries were inspected visually to identify and remove any erroneous boundaries (Figure 4a). A boundary was considered erroneous if it was outside a 2 m (4 × spatial resolution) buffer of the nearest boundary. After, the boundaries were converted to a topology. A topology defines how polygons relate to one another geometrically. It is used to correct vectorized data automatically based on user-defined criteria. In this case, it was used to remove minor inconsistencies, such as gaps and overshoots (Figure 4b). Finally, the boundaries were averaged across volunteers as using the geometric center of available polygons of each side (Figure 4c). Averaging is the centerpiece of crowdsourcing: the average is more representative of a field than any single digitizer.

(a) Erroneous field boundary   (b) Boundary overshoot   (c) Boundary centroid
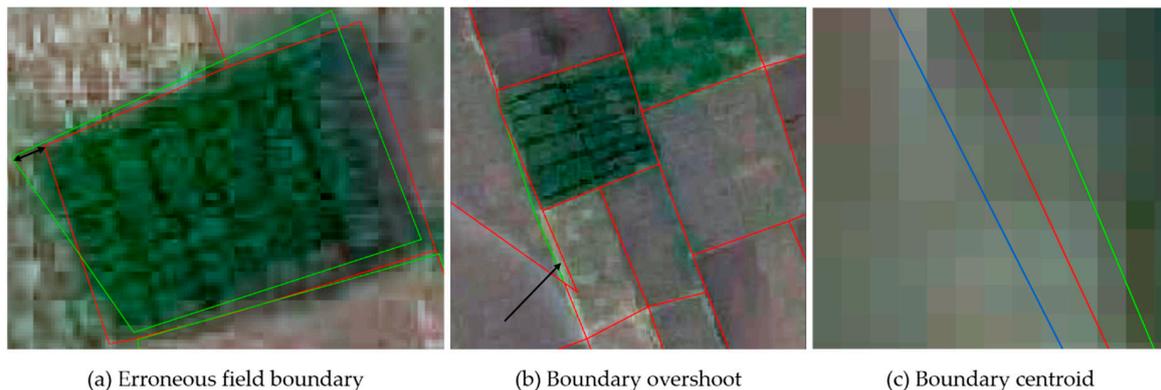
Figure 4. Major phases of processing to bring individual field boundaries into one master dataset (REF in red), which included averaging individual digitizers after elimination of erroneous field boundaries and gaps/slivers in green in (**a**) and (**b**). The green and blue lines in (**c**) represent boundaries defined by individual digitizers. Imagery © 2011 DigitalGlobe, Inc.

3.3. Multi-Resolution Segmentation (MRS)

The first method for boundary detection was implemented in Trimble eCognition®, which is an object-based image classification software. The first step in eCognition is segmentation, which is a process of dividing the image into regions or objects of homogeneous pixel values based on user-defined parameters. Multi-resolution segmentation (MRS) was used to segment the image into objects to coincide with the field boundaries [37]. For MRS, the size and constituents of the segments are controlled by assigning appropriate values to the key parameters: scale (SP), shape, and compactness to segment objects. The choice of these values can be determined a priori or through trial and error. To avoid a time-consuming and subjective selection of SP, an automated tool for parameterizing multi-scale image segmentation, referred to as the estimation of scale parameter (ESP2) [38], was used. ESP2 identifies scale parameters and segments the image based on the average local variance value of different layers. The shape parameter balances the spectral homogeneity and shape of the resulting objects as the sum of two should be equal to one. A default value of 0.1, as assigned in the ESP2 tool, was used to give more weight to spectral reflectance of the crop fields. The compactness parameter balances the compactness versus smoothness of the edges of objects [39]. A value of 0.5 was assigned to give equal weight to both compactness and smoothness. As implemented in eCognition, images were segmented at three hierarchical spatial levels (scales), i.e., level 1, 2, and 3. The choice of the output from the three spatial levels depends on the purpose of segmentation. Level 1 represents the most detailed segmentation and hence, produces the smallest segments, whereas level 3 is the coarsest and produces the largest segments. The output at spatial level 2 was found to be most suitable for the comparison, because it yielded field boundaries that tended to coincide with a subset of the manually digitized boundaries.

3.4. Contour Detection (CD)

Contour detection (CD) was implemented using the globalized probability of boundary (gPb) [40]. Contour detection refers to the process of finding boundaries between objects or segments in an image. gPb-based CD improves on early contour detection methods in two fundamental ways. Early approaches, such as Canny edge detection [41] extracted edges by calculating gradients of local brightness on a pixel-by-pixel (localized) basis, which are thereafter combined into contours. The approach typically detects irrelevant edges in textured regions, so more recent approaches, such as CD as implemented in gPb, include additional cues (texture and color). The cues on color, texture, and brightness are considered on a local pixel-level and a global image scale. This is done by combining the cues derived through edge detection and hierarchical image segmentation based on a k-threshold. As a result, gPb provides closed object outlines and eliminates irrelevant contours in textured regions. Crommelinck et al. [42] recently demonstrated gPb for cadastral mapping with images from an unmanned aerial system. It produced completeness and correctness rates of up to 80%. In this study, the same k-threshold was applied to all of the images. This parameter was selected to obtain comparable results in completeness and correctness, which inherits a balance between over- and under-segmentation.

3.5. Boundary Mapping Performance

MRS and CD were compared to REF using two measures: (i) quantitative completeness and (ii) spatial correctness. These measures are based on Heipke et al. [43] and are commonly reported in the literature for segmentation performance [44–46]. They were computed for a representative image tile from each of the Woredas given the large size of the dataset and computational demands of the two measures. One of the Woredas (Enbise Sar Midir) was removed from the comparison, because no image tile was available that was classified with all three methods. The completeness captured the percentage of field boundaries that were extracted by image segmentation. It was computed as the ratio of the number of field boundaries extracted from MRS or CD (N_{seg}) to the number of field boundaries defined by REF (N_{ref}). A field boundary was considered extracted when 70% of its outline was covered by the image segmentation. The counting of fields that were extracted and not extracted in MRS and CD was done by visual comparison to REF (Figure 5).

$$\text{completness } [\%] = \frac{N_{seg}}{N_{ref}} \times 100 \tag{1}$$

**(a)** extracted fields　　　　　　　　　　**(b)** not extracted fields
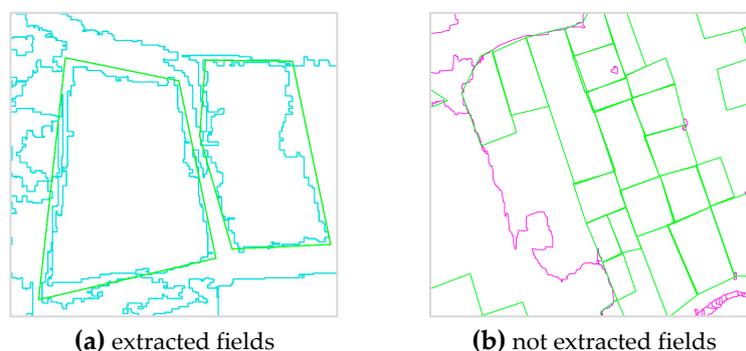
Figure 5. REF (green) overlaid with the image segmentation results from MRS (blue) and CD (pink). Image tiled subsets are representative of the general patterns of the two techniques.

Spatial correctness investigated to what extent successfully extracted fields coincided with the reference data in a spatial sense. It was computed by first buffering the reference data. The buffer size should be chosen in accordance with the required accuracy. The International Association of Assessing

Officers has proposed accuracies for fields in rural areas of 2.4 m [47]. A conservative buffer with a 2 m radius was selected for REF. Next, the percentage of the segmented lines lying inside and outside of the buffer was calculated (Figure 6). This can be done either vector- or raster-based. For a raster-based approach, as was done in this study, the rasterized segments (REF, MRS, and CD) were resampled to the spatial resolution of the image tiles. Pixels within the buffer were classified as true positive (TP), while pixels outside the buffer were classified as false positive (FP). The pixels were summed for each category in a confusion matrix. The error of commission (2) and the correctness (3) are calculated from the confusion matrix.

$$\text{error of commission } [\%] = \frac{\text{FP}}{\text{FP} + \text{TP}} \times 100 \tag{2}$$

$$\text{correctness } [\%] = \frac{\text{TP}}{\text{FP} + \text{TP}} \times 100 = 1 - \text{error of commission} \tag{3}$$
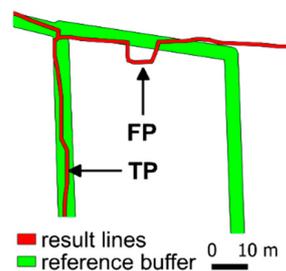
Figure 6. Correctness based on overlaying the buffered delineation and reference data to compute pixels being true positive (TP) or false positive (FP). These pixels are then summed to calculate (2) the error commission and (3) the correctness.

4. Results

Fields on average were less than 1 ha (Figure 7a) and had gentle slopes (<5%) (Figure 7b). The distributions were skewed to the right with fields as large as approximately 8 ha and very steeply sloping (nearly 45%). Since the distributions were highly skewed to the right, statistics are presented for the median, first quartile (25th percentile) and third quartile (75th percentile) range of values. Field boundaries in Dugda tended to be the largest (0.60 ha), followed by Dodota (0.57 ha), while in Enbise Sar Midir they were the smallest (0.33 ha, 0.23–0.49 ha). Field sizes in Dugda and Dodota were not only the largest, but also the most variable with interquartile ranges of 0.57 and 0.60 ha, respectively. Differences in the slopes among the Woredas were more mixed. Legambo was strongly sloping with a median slope of 12% and interquartile range of 10%. Liben Chukala was the most gently sloping with a median slope of <1% and an interquartile range of <1%. Other Woredas ranged between 1 and <10% slopes.

Figure 8 shows the REF boundaries and the segmentation results for CD and MRS for one image tile in the Amhara region. The comparison of MRS and CD with REF required the identification of one image tile per Woreda that contained representative results for all classes. Table 2 shows the results for image tiles from the nine Woredas that were suitable for comparison. On average, the completeness amounted to 46% for CD and 50% for MRS. Although these results appear similar, there was considerable variation among Woredas. CD results deviated more often from this average value compared to the MRS results. Neither technique appeared greatly impacted by the size of the fields. CD on average scored high for completeness in Dangila (96%), Lome (79%), and Wemberma (72%), while MRS scored high for completeness in Legambo (78%) and Lome (64%). Legambo was the most topographically complex, which may have contributed to the lower score of CD for completeness in this Woreda (39%). The correctness scores were considerably lower for both techniques, but CD on average scored higher (46%) than MRS (27%). CD scored particularly high for correctness in Woredas with the largest fields: Dangila (63%) and Dodota (63%). MRS scored poorly for correctness across the Woredas. The highest score for MRS was in Dangila (38%).
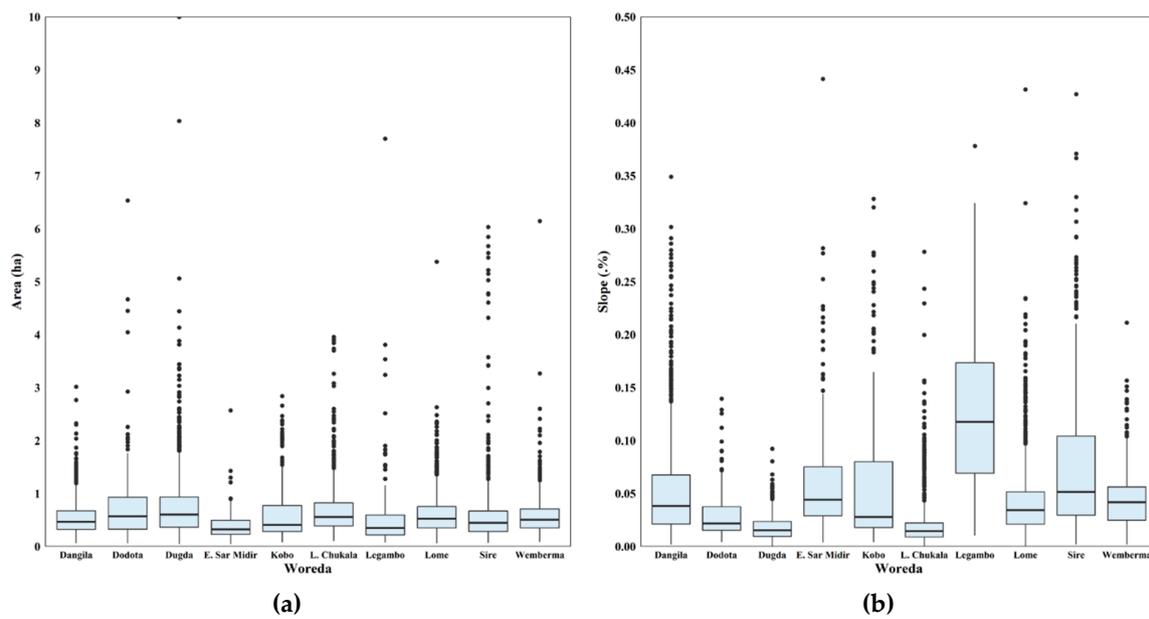
Figure 7. Boxplots showing the distribution of field boundary area (**a**) and average slope (**b**) as defined in REF for each Woreda. Slope was derived from the CGIAR CSI SRTM 90 m Digital Elevation Database v4.1 (https://cgiarcsi.community/data/srtm-90m-digital-elevation-database-v4-1/).
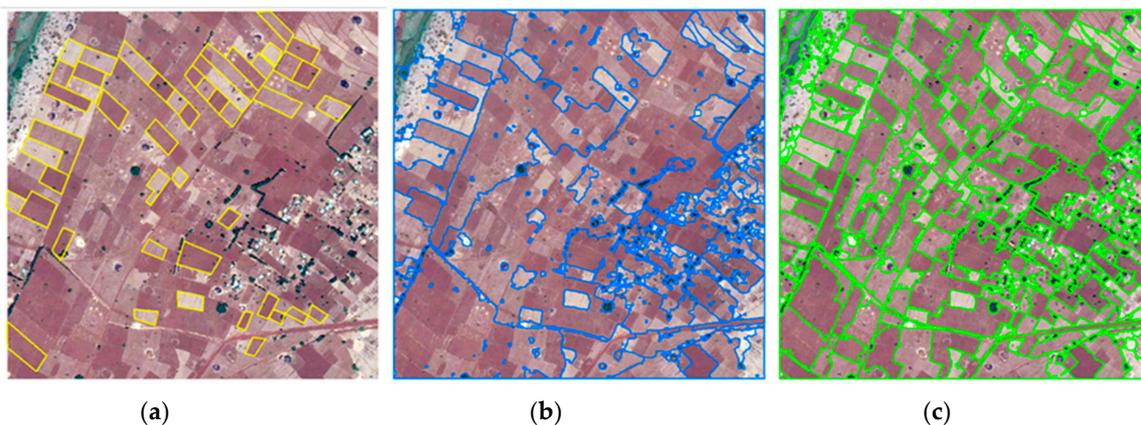
Figure 8. REF boundaries (**a**) and outputs from the two boundary detection methods, CD (**b**) and MRS (**c**), for one 1 km × 1 km tile in the Amhara Region. Image taken on 25 April, 2011. Imagery © 2011 DigitalGlobe, Inc.

Table 2. The quantitative completeness and spatial correctness of the manual versus object-based automated field boundary mapping techniques for a representative image tile in each Woreda.

Woreda	N_{fields}		Completeness [%] (N)		Correctness [%] (N)	
			CD	MRS	CD	MRS
Dangila	27		96 (26)	56 (15)	63	38
Wemberma	72		72 (52)	29 (21)	57	30
Legambo	18		39 (7)	78 (14)	59	35
Kobo	22		0 (0)	41 (9)	8	18
Dodota	29		17 (5)	59 (17)	63	31
Sire	16		19 (3)	31 (5)	33	17
Lome	28		79 (22)	64 (18)	54	28
Liben Chukala	8		50 (4)	38 (3)	45	27
Dugda	14		43 (6)	57 (8)	40	15
Total	234	*Average*	*46*	*50*	*46*	*27*

5. Discussion

The study generated crowd-driven maps of field boundaries from a number of VHR imagery across a diverse and representative set of topographically complex and highly fragmented agroecosystems in Ethiopia. Crowdsourcing was facilitated with a newly developed online tool. These field boundaries were compared against two semi-automated object-based techniques using standard performance metrics (quantitative completeness and spatial correctness). To the authors' knowledge, such a comprehensive comparison involving more than 7000 field boundaries derived from VHR imagery has never been performed. Three key findings of the study should be considered before the techniques are operationalized for practical use: (i) crowd-driven manual interpretation, especially when field boundaries are delineated, should involve considerable oversight and quality control/assurance with remedial training as needed; (ii) CD tended to capture larger field boundaries ($\rightarrow$1 ha), but not smaller field boundaries due to under-segmentation; and (iii) MRS as implemented with EPS tended to capture smaller field boundaries, but over-segmented larger field boundaries. Findings (ii) and (iii) indicate that neither semi-automated object-based approach is suitable for field boundary mapping over large areas in highly fragmented landscapes. CD and MRS are mainly governed by the k-threshold and SP, respectively. The former was too weak to differentiate between smaller field boundaries, while the latter gave too much weight to the spectral homogeneity compared to shape (compactness) of the field boundaries. Clearly, CD and MRS as implemented with EPS are location-specific and require user input unless a more sophisticated optimization procedure is employed.

There are several potential areas of improvement in the recruitment of trainees, training, the tool, and the remote supervision process. There was a large variation in quality across digitizers, in terms of number of vertices used to represent a boundary and in the position of the digitized boundary relative to the visible boundary on the image. Recruitment was restricted to Kifya employees. In the future, recruitment should be expanded to students in computer science and geography departments at local universities. These students most likely have greater computer skill and geographic information science/remote sensing experience. A pre-test could also be implemented to assure that volunteers have a certain level of computer skill and local knowledge of the study area, which was observed to improve the digitization skill. The digitizing training took place over two days and had three steps: (i) all trainees digitized the same five tiles and then conducted a peer review to compare performance and to reach agreement on an acceptable quality and level of detail in boundary representation; (ii) all trainees digitized the same 25 tiles, including tiles with obvious image quality issues, and conducted another round of peer review which included decision making on whether the image quality was sufficient to attempt digitizing or not, and; (iii) a longer session of digitizing randomly selected tiles to determine a realistic number of boundaries that could be mapped per hour per digitizer. Improvements in the training process could include a more interactive mode of digitizing training using a dashboard that shows the trainer how well the individually digitized boundaries of two users match. This could be used to drive discussion among the trainees and to create commonly agreed minimum levels of quality. Large discrepancies between digitizers may also represent challenging image tiles and again this would drive trainee discussion on when and when not to attempt digitizing. The digitizers could benefit from more contextual information during the digitizing process. Each tile was tagged with the name of the Woreda in which it was located and the Woreda name was displayed in the digitizing tool. The tool could be improved by including additional layers of information, such as a Google Maps background and allowing the digitizer to pan and zoom around in the vicinity of the tile to get a better feel for the landscape in the surroundings of the tile.

Follow up supervision was done remotely once per month to discuss any issues with the tool, to discuss consistency in interpretation across digitizers, and to assess overall progress. This remote oversight would have benefited from a dashboard to track the number of boundaries digitized per hour to detect instances where digitizers had mapped boundaries too rapidly and potentially with insufficient attention to detail. Additionally, diagnostic tools that could track the number of vertices per boundary would have also provided information on the level of detail of digitizing. The ability to

compare the same digitized boundaries from two digitizers would allow for specific interventions to re-train a digitizer where there were large discrepancies. While field boundary complexity will vary across the region, tile specific information could still be collated across digitizers as a way to assess consistency. Such information could have then been shared with the team of supervisors for remedial action and further training.

Selecting one optimal parameter for SP and CD for the entire region might be improved by considering the local variation in field size, shape, and contrast. While the field boundaries remain relatively stable across multiple years, it is often difficult to identify the edges between fields when crops are not grown. Both MRS as implemented with EPS and CD methods depend on the presence of edges to detect sharp transitions in color, texture or brightness. However, the availability of cloud free WorldView scenes during the growing seasons can be limited. Use of dry season images may have contributed to the poor performance of gPb-based CD in such cases. Although some studies, e.g., Neigh et al. [30], have reported the effect of seasonality on crop area estimates, further analysis is required to understand the effect of crop growth stages on field boundary delineation with semi-automatic methods. Smallholder plots in Africa may not always be homogeneous. Intercropping, intra-plot crop growth variation, unpaved dirt roads or trees create unique challenges for segmentation. In such cases, MRS as implemented with EPS tends to over-segment the fields capturing intra-plot variability, and it is almost impossible to find a single SP that would work across fields with such different physical characteristics. Data fusion is increasingly used in agronomy to take full advantage of the predictive power of high spatial (but low temporal) and low spatial (but high temporal) resolution satellite imagery. Alternative strategies to overcome these challenges therefore could involve the fusion of two different sources of data: (i) sub-meter resolution WorldView data for manual digitization and (ii) Planet data (https://www.planet.com/) with slightly lower spatial resolution (3 m) but much higher temporal frequency (~1–7 days) for semi-automatic field boundary detection. At 3 m resolution, some of the intra-plot variability would be smoothed out. Further, the integration of multiple observations within the season could reduce the confusion between croplands and pastures, and provide the opportunity to include time series information in the segmentation process [48]. The full spectral depth of WorldView imagery was not freely available to the consortium and therefore not used in this study. Increasing the spectral depth in combination with increasing the temporal resolution of imagery could greatly extend the advantages taken here to use high spatial resolution imagery. Finally, regarding the choice of gPb-based CD, it was observed that setting one k-threshold for all images might not be optimal: as the results revealed, the size and presence of visible field plots varied across different areas, which requires adjusting the level of over- or under- segmentation per area. CD could be considered as an initial workflow step to which further methods that refine the output to the actual field boundaries should be added. These can be methods of machine learning that learn which of the CD results are useful for field boundary extraction at a given location [29], or deep learning networks that directly detect boundaries from the remote sensing data [31].

The results of this analysis are of little practical use in countries in Europe and North America where parcel information is available and frequently used to monitor crops. However, in the Global South, information on the location and extent of field boundaries is scarce, due to weak institutions, poor infrastructure, and a lack of incentives. At the same time, such information is desperately needed for agroecosystems in the region to develop sustainably. The workflow demonstrates that field boundary mapping over large areas in the Global South is possible if donors and/or national governments are willing to invest in limited computer infrastructure for cloud-computing and crowdsourcing, as well as very high spatial resolution imagery. The results lean toward crowdsourcing for field boundary mapping in Ethiopia, but also point to new avenues for improvement that researchers and practitioners can take to operationalize an object-oriented approach. The large dataset could be used to test these new alternatives. Already, the data are being used to evaluate a blended census and multi-scale remote sensing approach to map the probability of crop fields in the Oromia region.

6. Conclusions

Field boundaries can provide important structural information about individual plots, including size. Farm size plays an important role not only for farmers, but also for policy makers. There is an ongoing debate about the relationship between farm size and productivity across the Global South. Studies using household surveys with farmer self-reporting data have reported an inverse relationship between farm size and productivity [49]. Such a relationship would trigger different policies for farm management (e.g., subdivision of larger farms to smaller ones) or agriculture input supply (e.g., fertilizer recommendation based on farm size) to improve agricultural production at the national level [50]. However, studies using GPS-based plot area measurements did not find small farms to be more productive than larger farms [51,52]. These studies concluded that the inverse relationship is attributed to systematic bias in farmer-reported estimates. While GPS-based plot areas are unbiased, it comes at a significantly higher cost, and it is not a viable option across large areas over repeated times. The crowd-driven method presented here for the first time and object-oriented methods compared from the literature can provide practical alternatives to analyze the farm size-productivity relationship at larger scale at much cheaper cost, and thus provide novel insights with unbiased quality data. Given the observation that CD tended to capture larger field boundaries and MRS tended to capture smaller field boundaries, one prominent direction is combining these two methods for mapping complex field boundaries. In the future, a crowd-driven framework can be constructed where the crowd picks either MRS or CD for mapping depending on the nature of the field boundaries.

References

1. See, L.; Fritz, S.; You, L.; Ramankutty, N.; Herrero, M.; Justice, C.; Becker-Reshef, I.; Thornton, P.; Erb, K.; Gong, P.; et al. Improved global cropland data as an essential ingredient for food security. *Glob. Food Secur.* **2015**, *4*, 37–45. [CrossRef]
2. Fritz, S.; See, L. Identifying and quantifying uncertainty and spatial disagreement in the comparison of Global Land Cover for different applications. *Glob. Chang. Biol.* **2008**, *14*, 1057–1075. [CrossRef]
3. Ricciardi, V.; Ramankutty, N.; Mehrabi, Z.; Jarvis, L.; Chookolingo, B. How much of the world's food do smallholders produce? *Glob. Food Secur.* **2018**, *17*, 64–72. [CrossRef]

4. Herrero, M.; Thornton, P.K.; Power, B.; Bogard, J.R.; Remans, R.; Fritz, S.; Gerber, J.S.; Nelson, G.; See, L.; Waha, K.; et al. Farming and the geography of nutrient production for human use: A transdisciplinary analysis. *Lancet Planet. Health* **2017**, *1*, e33–e42. [CrossRef]

5. Yan, L.; Roy, D.P. Automated crop field extraction from multi-temporal Web Enabled Landsat Data—ScienceDirect. *Remote Sens. Environ.* **2014**, *144*, 42–64. [CrossRef]

6. Yan, L.; Roy, D.P. Roy Conterminous United States crop field size quantification from multi-temporal Landsat data—ScienceDirect. *Remote Sens. Environ.* **2016**, *172*, 67–86. [CrossRef]

7. Graesser, J.; Ramankutty, N. Detection of cropland field parcels from Landsat imagery. *Remote Sens. Environ.* **2017**, *201*, 165–180. [CrossRef]

8. White, E.V.; Roy, D.P. A contemporary decennial examination of changing agricultural field sizes using Landsat time series data. *GEO Geogr. Environ.* **2015**, *2*, 33–54. [CrossRef]

9. Marshall, M.T.; Husak, G.J.; Michaelsen, J.; Funk, C.; Pedreros, D.; Adoum, A. Testing a high-resolution satellite interpretation technique for crop area monitoring in developing countries. *Int. J. Remote Sens.* **2011**, *32*, 7997–8012. [CrossRef]

10. Crommelinck, S.; Bennett, R.; Gerke, M.; Nex, F.; Yang, M.Y.; Vosselman, G. Review of Automatic Feature Extraction from High-Resolution Optical Sensor Data for UAV-Based Cadastral Mapping. *Remote Sens.* **2016**, *8*, 689. [CrossRef]

11. Hammond, A.L. Crop Forecasting from Space: Toward a Global Food Watch. *Science* **1975**, *188*, 434–436. [CrossRef]

12. MacDonald, R.B.; Hall, F.G. Global Crop Forecasting. *Science* **1980**, *208*, 670–679. [CrossRef]

13. MacDonald, R.B.; Hall, F.G.; Erb, R.B. The Use of LANDSAT Data in a Large Area Crop Inventory Experiment (LACIE). In Proceedings of the LARS Symposia, West Lafayette, Indiana, 3–5 June 1975; Institute of Electrical and Electronics Engineers, Inc.: New York, NY, USA, 1975; p. 25.

14. Hixson, M.M.; David, B.J.; Bauer, M.E. Sampling Landsat classifications for crop area estimation. *Photogram. Eng. Remote Sens.* **1981**, *47*, 1343–1348.

15. Hixson, M.M.; Davis, S.M.; Bauer, M.E. Evaluation of a Segment-Based Landsat Full-Frame Approach to Crop Area Estimation. In Proceedings of the LARS Symposia, West Lafayette, Indiana, 23–26 June 1981; Institute of Electrical and Electronics Engineers, Inc.: New York, NY, USA, 1981; p. 11.

16. Gallego, F.J. *Crop Area Estimation in the MARS Project*; Space Applications Institute: Brussels, Belgium, 1999; p. 11.

17. Husak, G.J.; Marshall, M.T.; Michaelsen, J.; Pedreros, D.; Funk, C.; Galu, G. Crop area estimation using high and medium resolution satellite imagery in areas with complex topography. *J. Geophys. Res.* **2008**. [CrossRef]

18. Grace, K.; Husak, G.J.; Harrison, L.; Pedreros, D.; Michaelsen, J. Using high resolution satellite imagery to estimate cropped area in Guatemala and Haiti. *Appl. Geogr.* **2012**, *32*, 433–440. [CrossRef]

19. Grace, K.; Husak, G.; Bogle, S. Estimating agricultural production in marginal and food insecure areas in Kenya using very high resolution remotely sensed imagery. *Appl. Geogr.* **2014**, *55*, 257–265. [CrossRef]

20. Husak, G.; Grace, K. In search of a global model of cultivation: Using remote sensing to examine the characteristics and constraints of agricultural production in the developing world. *Food Secur.* **2016**, *8*, 167–177. [CrossRef]

21. Fritz, S.; See, L.; McCallum, I.; You, L.; Bun, A.; Moltchanova, E.; Duerauer, M.; Albrecht, F.; Schill, C.; Perger, C.; et al. Mapping global cropland and field size. *Glob. Chang. Biol* **2015**, *21*, 1980–1992. [CrossRef]

22. Fritz, S.; McCallum, I.; Schill, C.; Perger, C.; See, L.; Schepaschenko, D.; van der Velde, M.; Kraxner, F.; Obersteiner, M. Geo-Wiki: An online platform for improving global land cover. *Environ. Model. Softw.* **2012**, *31*, 110–123. [CrossRef]

23. Lesiv, M.; Bayas, J.C.L.; See, L.; Duerauer, M.; Dahlia, D.; Durando, N.; Hazarika, R.; Sahariah, P.K.; Vakolyuk, M.; Blyshchyk, V.; et al. Estimating the global distribution of field size using crowdsourcing. *Glob. Chang. Biol.* **2019**, *25*, 174–186. [CrossRef]

24. Estes, L.D.; McRitchie, D.; Choi, J.; Debats, S.; Evans, T.; Guthe, W.; Luo, D.; Ragazzo, G.; Zempleni, R.; Caylor, K.K. A platform for crowdsourcing the creation of representative, accurate landcover maps. *Environ. Model. Softw.* **2016**, *80*, 41–53. [CrossRef]

25. Debats, S.R.; Luo, D.; Estes, L.D.; Fuchs, T.J.; Caylor, K.K. A generalized computer vision approach to mapping crop fields in heterogeneous agricultural landscapes. *Remote Sens. Environ.* **2016**, *179*, 210–221. [CrossRef]

26. Mueller, M.; Segl, K.; Kaufmann, H. Edge- and region-based segmentation technique for the extraction of large, man-made objects in high-resolution satellite imagery. *Pattern Recognit.* **2004**, *37*, 1619–1628. [CrossRef]

27. Turker, M.; Kok, E.H. Field-based sub-boundary extraction from remote sensing imagery using perceptual grouping. *ISPRS J. Photogramm. Remote Sens.* **2013**, *79*, 106–121. [CrossRef]

28. Blaschke, T. Object based image analysis for remote sensing. *ISPRS J. Photogramm. Remote Sens.* **2010**, *65*, 2–16. [CrossRef]

29. Garcia-Pedrero, A.; Gonzalo-Martín, C.; Lillo-Saavedra, M.; Rodríguez-Esparragón, D. The Outlining of Agricultural Plots Based on Spatiotemporal Consensus Segmentation. *Remote Sens.* **2018**, *10*, 1991. [CrossRef]

30. Neigh, C.S.R.; Carroll, M.L.; Wooten, M.R.; McCarty, J.L.; Powell, B.F.; Husak, G.J.; Enenkel, M.; Hain, C.R. Smallholder crop area mapped with wall-to-wall WorldView sub-meter panchromatic image texture: A test case for Tigray, Ethiopia. *Remote Sens. Environ.* **2018**, *212*, 8–20. [CrossRef]

31. Crommelinck, S.; Höfle, B.; Koeva, M.N.; Yang, M.Y.; Vosselman, G. Interactive cadastral boundary delineation from UAV data. In Proceedings of the ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Riva del Garda, Italy, 4–7 June 2018; Volume IV-2, pp. 81–88.

32. Funk, C.; Rowland, J.; Eilerts, G.; Kebebe, E.; Biru, N.; White, L.; Galu, G. *A Climate Trend Analysis of Ethiopia*; U.S. Geological Survey: Sioux Falls, SD, USA, 2012; p. 6.

33. DEWA. *Africa: Atlas of Our Changing Environment*; United Nations Environment Programme (UNEP): Nairobi, Kenya, 2008.

34. Place, F. Land Tenure and Agricultural Productivity in Africa: A Comparative Analysis of the Economics Literature and Recent Policy Strategies and Reforms. *World Dev.* **2009**, *37*, 1326–1336. [CrossRef]

35. De Leeuw, J.; Vrieling, A.; Shee, A.; Atzberger, C.; Hadgu, K.; Biradar, C.; Keah, H.; Turvey, C. The Potential and Uptake of Remote Sensing in Insurance: A Review. *Remote Sens.* **2014**, *6*, 10888–10912. [CrossRef]

36. JECAM. *Guidelines for Cropland and Crop Type Definition and Field Data Collection*; Group on Earth Observations: Ottawa, ON, Canada, 2014; p. 7.

37. Baatz, M.; Schäpe, A. Multiresolution Segmentation—An optimization approach for high quality multi-scale image segmentation. In *Angewandte Geographische Informationsverarbeitung*; Wichmann: Heidelberg, Germany, 2000; Volume XII, pp. 12–23.

38. Drǎguţ, L.; Tiede, D.; Levick, S.R. ESP: A tool to estimate scale parameter for multiresolution image segmentation of remotely sensed data. *Int. J. Geogr. Inf. Sci.* **2010**, *24*, 859–871. [CrossRef]

39. Drǎguţ, L.; Csillik, O.; Eisank, C.; Tiede, D. Automated parameterisation for multi-scale image segmentation on multiple layers. *ISPRS J. Photogramm. Remote Sens.* **2014**, *88*, 119–127. [CrossRef]

40. Arbelaez, P.; Maire, M.; Fowlkes, C.; Malik, J. Contour Detection and Hierarchical Image Segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2011**, *33*, 898–916. [CrossRef]

41. Canny, J. A Computational Approach to Edge Detection. In *Readings in Computer Vision*; Fischler, M.A., Firschein, O., Eds.; Morgan Kaufmann: San Francisco, CA, USA, 1987; pp. 184–203. ISBN 978-0-08-051581-6.

42. Crommelinck, S.; Bennett, R.; Gerke, M.; Yang, M.Y.; Vosselman, G. Contour Detection for UAV-Based Cadastral Mapping. *Remote Sens.* **2017**, *9*, 171. [CrossRef]

43. Heipke, C.; Mayer, H.; Wiedemann, C.; Jamet, O. Evaluation of Automatic Road Extraction. In Proceedings of the International Archives of Photogrammetry and Remote Sensing, Munich, Germany, 19–21 September 2007; pp. 47–56.

44. Quackenbush, L.J. A Review of Techniques for Extracting Linear Features from Imagery. Available online: https://www.ingentaconnect.com/content/asprs/pers/2004/00000070/00000012/art00004 (accessed on 14 April 2019).

45. Rottensteiner, F.; Sohn, G.; Gerke, M.; Wegner, J.D.; Breitkopf, U.; Jung, J. Results of the ISPRS benchmark on urban object detection and 3D building reconstruction. *ISPRS J. Photogramm. Remote Sens.* **2014**, *93*, 256–271. [CrossRef]

46. Rutzinger, M.; Rottensteiner, F.; Pfeifer, N. A Comparison of Evaluation Techniques for Building Extraction from Airborne Laser Scanning. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2009**, *2*, 11–20. [CrossRef]

47. *IAAO Standard on Digital Cadastral Maps and Parcel Identifiers*; International Association of Assessing Officers: Kansas City, MO, USA, 2015; p. 24.

48. Belgiu, M.; Csillik, O. Sentinel-2 cropland mapping using pixel-based and object-based time-weighted dynamic time warping analysis. *Remote Sens. Environ.* **2018**, *204*, 509–523. [CrossRef]

49. Gautam, M.; Ahmed, M. Too small to be beautiful? The farm size and productivity relationship in Bangladesh. *Food Policy* **2019**, *84*, 165–175. [CrossRef]

50. Barrett, C.B.; Bellemare, M.F.; Hou, J.Y. Reconsidering Conventional Explanations of the Inverse Productivity–Size Relationship. *World Dev.* **2010**, *38*, 88–97. [CrossRef]

51. Carletto, C.; Gourlay, S.; Winters, P. From Guesstimates to GPStimates: Land Area Measurement and Implications for Agricultural Analysis. *J. Afr. Econ.* **2015**, *24*, 593–628. [CrossRef]

52. Carletto, C.; Savastano, S.; Zezza, A. Fact or artifact: The impact of measurement errors on the farm size–productivity relationship. *J. Dev. Econ.* **2013**, *103*, 254–261. [CrossRef]