A generalized computer vision approach to mapping crop fields in heterogeneous agricultural landscapes

Stephanie R. Debats\textsuperscript{a,∗}, Dee Luo\textsuperscript{b}, Lyndon D. Estes\textsuperscript{a}, Thomas J. Fuchs\textsuperscript{c}, Kelly K. Caylor\textsuperscript{a}

\textsuperscript{a}Department of Civil \& Environmental Engineering, Princeton University, Princeton, NJ 08544, USA
\textsuperscript{b}Department of Operations Research \& Financial Engineering, Princeton University, Princeton, NJ 08544, USA
\textsuperscript{c}NASA Jet Propulsion Laboratory, Pasadena, CA 91109, USA

Abstract

Smallholder farms dominate in many parts of the world, particularly Sub-Saharan Africa. These systems are characterized by small, heterogeneous, and often indistinct field patterns, requiring a specialized methodology to map agricultural land cover. Using a variety of sites in South Africa, we present a new approach to mapping agricultural fields, based on efficient extraction of a vast set of simple, highly correlated, and interdependent features, followed by a random forest classifier. We achieved similar high performance across agricultural types, including the spectrally indistinct smallholder fields as well as the more easily distinguishable commercial fields, and demonstrated the ability to generalize performance across large geographic areas. In sensitivity analyses, we determined multi-temporal information provided greater gains in performance than the addition of multi-spectral bands available in DigitalGlobe Worldview-2 imagery.

Keywords: land cover, agriculture, Sub-Saharan Africa, computer vision, machine learning

1. Introduction

Improving the capacity to monitor the spatial distribution of agriculture, particularly among smallholder farmers, is critical to increasing agricultural productivity and food security in many parts of the world. Smallholder agriculture, which dominates in Sub-Saharan
Africa and other world regions, features rainfed production for household consumption and use of family labor and minimal technology (Altieri & Koohafkan, 2008; Eastwood et al., 2010; Gollin, 2014; Morton, 2007). These systems are also characterized by small, heterogeneous, and often indistinct field patterns (Estes et al., 2015; Fritz & See, 2008; Lobell, 2013; See et al., 2015). The prevalence of smallholder farms highlights the need for a specialized methodology to monitor agriculture across farming types, including both smallholder and commercial.

Of the regions where smallholder agriculture dominates, Sub-Saharan Africa is the most important, due to its geographic size and status as a potential center of agricultural growth in the coming decades. Of all farms in Sub-Saharan Africa, 80% are less than 2 hectares and the mean farm size of 1.6 hectares is significantly smaller than most world regions (Table 1) (von Braun, 2004; FAO, 1997; IFAD & UNEP, 2013; Lowder et al., 2014; Nagayets, 2005). Increasing agricultural productivity is crucial in Sub-Saharan Africa, because its population is expected to double by 2050 (Haub & Kaneda, 2013). The population remains predominantly rural despite recent urbanization (Masters et al., 2013), and 60% of the workforce is employed in agriculture (Bank et al., 2014). As the population grows, agricultural field sizes are driven down and farmers are pushed onto marginal lands. As a result, growing climate variability, characterized by less frequent and more intense rain events, increases farmers’ vulnerability to food insecurity (Davidson et al., 2003; Jayne et al., 2014; Masters et al., 2013; Oba et al., 2001; Thornton et al., 2009, 2011; World Bank, 2013). As smallholder field sizes decrease, large-scale international land acquisitions and government-sponsored agricultural growth corridors promote consolidation of remaining farmland into commercial enterprises (Cotula & Vermeulen, 2009; Davis et al., 2014; Gollin, 2014; Nogales, 2014; Rulli & D’Odorico, 2014; Rulli et al., 2013). Efforts to monitor agricultural change on the ground are confounded by a widespread shortage of agricultural data in Sub-Saharan Africa, in part due to limited government capacity (Alliance for a Green Revolution in Africa (AGRA), 2013; Carletto et al., 2013; Centre for Global Development & The African Population and
Health Research Centre, 2014; IFAD & UNEP, 2013; World Bank, 2013). Therefore, an accurate accounting of agricultural land cover, across both smallholder and commercial farming, is needed to track and promote food security in Sub-Saharan Africa.

Table 1: Estimated farm sizes of selected world regions
(von Braun, 2004; FAO, 1997; Nagayets, 2005)

<table>
<thead>
<tr>
<th>World region</th>
<th>Average farm size (hectares)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Africa</td>
<td>1.6</td>
</tr>
<tr>
<td>Asia</td>
<td>1.6</td>
</tr>
<tr>
<td>Western Europe</td>
<td>27.0</td>
</tr>
<tr>
<td>Latin America and Caribbean</td>
<td>67.0</td>
</tr>
<tr>
<td>North America</td>
<td>121.0</td>
</tr>
</tbody>
</table>

In data-sparse regions, remote sensing satellite imagery provides alternative means to monitor agriculture. However, land cover data sets derived from remote sensing contain large uncertainties regarding the total area as well as the spatial distribution of agriculture (Estes et al., 2015; Fritz & See, 2008; Fritz et al., 2011; Siebert et al., 2010). Moreover, readily available remote sensing data sets, such as MODIS (250 m) and Landsat (30 m), lack sufficient spatial resolution to study smallholder fields, which are often smaller than 1 hectare (100 m x 100 m) (Estes et al., 2015; Jain et al., 2013; Lobell, 2013). For example, a previous study using Landsat imagery to identify agricultural fields in the United States struggled with small, irregular fields of less than 1.5 hectares, though these conditions represented only a minority of studied fields (Yan & Roy, 2014).

High-resolution satellite imagery (≤ 2 m) provides the necessary detail to observe smallholder agriculture and is becoming increasingly available and affordable (Estes et al., 2015; Fritz et al., 2015; Hayes et al., 2014; Jain et al., 2013; Lobell, 2013; See et al., 2015). As coverage improves, automated classification algorithms are used to extract actionable data from images. Prior efforts to automatically classify agricultural fields using high-resolution imagery have struggled due to (1) the nature and appearance of smallholder agriculture, (2) the properties of high-resolution imagery itself, and (3) the design of classification algorithms.

First, the high spatial variability in land cover poses a classification challenge, particularly
among smallholder agricultural fields, which are small and irregularly shaped (Palm et al., 2010). Smallholder fields are less visually defined, exhibiting indistinct boundaries between neighboring fields as well as ambiguity between fields and natural vegetation (Estes et al., 2015; Fritz & See, 2008; Lobell, 2013; See et al., 2015). They also exhibit more variability in spectral and phenological signatures, due to rainfed farming, sub-optimal management, low cropping intensity, fallowing, abandonment, and the inclusion of large trees within fields (Jain et al., 2013; Mayes et al., 2015; Siebert et al., 2010; Vintrou et al., 2012) (Figure 1). Thus, these conditions highlight the need for a specialized methodology for smallholder agriculture.

Figure 1: Examples of false color (NIR1, red, green) images of smallholder (a, b) and commercial (c, d) agriculture in Sub-Saharan Africa, highlighting the highly variable spectral characteristics of smallholder agriculture. Example images are 500 x 500 pixels of 2 m resolution, centered at: (a) 25.0383° S, 27.9955° E, (b) 25.0234° S, 29.5551° E, (c) 26.4228° S, 25.9284° E, and (d) 25.9859° S, 25.9573° E.

Second, the level of detail increases in high-resolution imagery, which raw spectral values or simple features have difficulty describing. In high-resolution imagery, land cover classes have lower inter-class and higher intra-class spectral variability, creating ambiguities in classification (Lu & Weng, 2007; Tokarczyk et al., 2013, 2015). Variability also increases with image mosaicking, which is often necessary due to lower collection frequency of high-resolution imagery (Estes et al., 2015; Hayes et al., 2014). Previous efforts to handle increased detail in high-resolution imagery have focused on expanding the feature space by manually handcrafting higher order features suitable to a specific application. Features that capture textural and contextual information (e.g. grey-level co-occurrence matrices (GLCM), filter bank responses, and textons) have been found to improve classification accuracy over
spectral information alone, but their use has been limited by their complexity and high computational cost (Butusov, 2003; Kurosu et al., 2001; Leung & Malik, 2001; Lu & Weng, 2007; Podest & Saatchi, 2002; Rao et al., 2002; Schmid, 2001; Shaban & Dikshit, 2001; Shotton et al., 2008; Tokarczyk et al., 2013, 2015). Incorporating additional multi-spectral bands in features, either as raw inputs or calculated image transformations like vegetation indices, may also increase accuracy, but without selection of the most discriminative bands, this approach is limited by high band correlation (Lu & Weng, 2007; Mausel et al., 1990; Thenkabail et al., 2004). Multi-temporal imagery has also been found to increase accuracy, especially for agriculture, by capturing details about phenological profiles and filling in missing data (Duveiller & Defourny, 2010; Guerschman et al., 2003; Liu et al., 2002; Lu & Weng, 2007; Oetter et al., 2001). Overall, extracting more features derivates greater information from high-resolution imagery. However, classification algorithms struggle with the expanded feature space, necessitating the use of feature selection to determine the most useful ones, often in a separate stage prior to classification (Hughes, 1968; Lu & Weng, 2007; Price et al., 2002).

Third, a classification algorithm must efficiently handle the vast feature spaces of highly correlated and interdependent features required to adequately describe smallholder agriculture in high-resolution imagery (Hughes, 1968; Lu & Weng, 2007; Price et al., 2002). An algorithm must also resist overfitting to training data and have high generalization performance to classify large expanses of new imagery (Mascaro et al., 2014). Previous efforts have found that tracking land cover changes over time lends itself to a supervised classification approach, in which a classifier is initially trained to identify prescribed classes with labeled data and then repeatedly deployed on a time series of images (McIver & Friedl, 2001). Yet supervised classification of remote sensing imagery is often deemed a local approach for small areas, with a reputation of being difficult to repeat over large areas (Mayes et al., 2015). Within supervised classification, non-parametric classifiers are increasingly preferred, due to their ability to handle extremely large feature spaces and data sets, robustness to outliers and noise, and outperformance of parametric classifiers in complex landscapes (Adam et al., 2014;
De Fries et al., 1998; Friedl et al., 1999; Gislason et al., 2006; Gopal et al., 1999; Ham et al., 2005; Hayes et al., 2014; Lu & Weng, 2007; McIver & Friedl, 2001; Paola & Schowengerdt, 1995; Rodriguez-Galiano et al., 2011). Furthermore, classifiers with probabilistic output, as opposed to hard classifications, are gaining popularity in remote sensing for highlighting the spatial variation in classification quality and confidence, which is crucial for utilizing results in decision making (Liu et al., 2004; Lu & Weng, 2007; McIver & Friedl, 2001). Probabilistic output can also be post-processed in a variety of ways, ranging from simple thresholding for obtaining hard classifications to image segmentation and object detection.

This study develops a methodology to differentiate heterogeneous agricultural land cover in high-resolution imagery of Sub-Saharan Africa that is effective across a range of agricultural types, including the small, irregular fields of the dominant smallholder class. In a supervised classification approach, we utilize techniques from computer vision and machine learning, two fields that are increasingly used in remote sensing analyses as high-resolution imagery becomes readily available. Computer vision provides the means to detect, recognize, and track complex objects in images, while machine learning enables assimilation of vast quantities of data (Szeliski, 2010), a powerful combination for deriving actionable data from high-resolution imagery that generalizes across large areas.

First, to address the problems of adequately describing the small, irregular fields of smallholder agriculture and handling the great level of detail in high-resolution imagery, we efficiently extract a unique set of many simple, highly correlated, and interdependent features. Second, we utilize a supervised, non-parametric machine learning classifier to handle the vast feature space and provide probabilistic output of the likelihood that pixels belong to agricultural fields. We assess our classification accuracy using the metrics of receiver operating characteristic (ROC) curves and the related area under the ROC curve (AUC), finding similar high performance (AUC > 0.90) across agriculture types, including smallholders. We also investigate the relative roles of multi-temporal and multi-spectral information, given the data limitations of high-resolution satellite imagery in Sub-Saharan Africa, and find
that multi-temporal imagery contributes to greater performance gains. Finally, we propose further work to post-process the probabilistic output into maps of individual agricultural fields.

2. Data

2.1. Study area

South Africa was selected as our study area, due to greater coverage of high-resolution imagery than other areas of Sub-Saharan Africa, as well as availability of official government-sponsored agricultural land cover maps (GeoTerraImage, 2008). In addition, South Africa’s agricultural sector is divided between commercial agriculture and smallholder subsistence households in the former homeland areas of the apartheid era (Pienaar, 2013). Maize is the most important crop for commercial agriculture as well as smallholders with a planted area over four times greater than the land devoted to all other summer crops (Hardy et al., 2011). South African smallholders typically grow maize as their main staple food in subsistence mixed crop-livestock systems, which is representative of typical agriculture in Sub-Saharan Africa as a whole (Jones & Thornton, 2003).

Eight study sites were selected with an emphasis on smallholder rainfed subsistence fields, as well as a mix of rainfed commercial and center pivot irrigated fields from the primary agricultural regions of the country, based on availability of imagery (Figure 2). Each of these three agricultural types were present in at least three of the eight sites. The range of sites enabled testing of the generalizability of our algorithm across various landscapes and farming types that are representative of much of Sub-Saharan Africa.
2.2. Satellite imagery

For this study, we acquired DigitalGlobe Worldview-2 imagery, which typifies the spatial resolution of future sensors and provides a good compromise between cost, resolution (spatial, spectral, and temporal), and geographic coverage for our application. For each site, we obtained pairs of orthorectified images, each 25 km$^2$, which include a growing season image (December-April) and an off season image (July-November) within the same year or one year apart. Each image contains a panchromatic band at 0.5 meter resolution (aggregated using mean values to 2 meter resolution) and eight multi-spectral bands at 2 meter resolution (Table 2).

Table 2: DigitalGlobe Worldview-2 multi-spectral bands

<table>
<thead>
<tr>
<th>Band</th>
<th>Spectral range (µm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coastal Blue</td>
<td>0.40 - 0.45</td>
</tr>
<tr>
<td>Blue</td>
<td>0.45 - 0.51</td>
</tr>
<tr>
<td>Green</td>
<td>0.51 - 0.58</td>
</tr>
<tr>
<td>Yellow</td>
<td>0.585 - 0.625</td>
</tr>
<tr>
<td>Red</td>
<td>0.63 - 0.69</td>
</tr>
<tr>
<td>Red Edge</td>
<td>0.705 - 0.745</td>
</tr>
<tr>
<td>Near Infrared (NIR1)</td>
<td>0.77 - 0.895</td>
</tr>
<tr>
<td>Near Infrared (NIR2)</td>
<td>0.86 - 1.04</td>
</tr>
</tbody>
</table>
3. Methodology

3.1. Overview

We frame our agricultural field classification task as a supervised binary classification problem to determine the unknown class $y$ of pixels as either field or non-field. In supervised classification, a classifier is initially trained with labelled data and then deployed on previously unseen data. Specifically, using an input feature vector, $x$, extracted for a pixel of interest, a trained classifier determines the posterior class probability of the pixel, $P(y|x)$, expressed as the likelihood of the pixel belonging to a field.

We define agriculture as fields that have been recently active. This approach is preferred over classifying currently active fields, due to the frequent fallowing and rotation of fields, variety of planted crops, and range of planting dates that is characteristic of Sub-Saharan African agriculture. Fields that are obviously grown over and abandoned are labelled as non-field areas. Given the complex, small-scale nature of Sub-Saharan African agriculture, some disagreement in labeling is expected. As is the case with most machine learning algorithms, we expect the volume and variety of training data to enable the algorithm to handle some level of noise.

3.2. Hand-digitization of labelled data

To supply labelled data for supervised classification, each agricultural field was hand-digitized and additional metadata on agricultural type (rainfed subsistence, rainfed commercial, or center pivot) was stored. The hand-digitization was performed in QGIS (QGIS Development Team, 2015) using the 2 meter resolution data as well as the panchromatic band at the original 0.5 meter resolution for reference. Hand-digitized field boundaries were also cross-referenced with official agricultural land cover maps from the South African government to increase ground-truth accuracy. Hand-digitization of entire fields was preferred over classifying randomly selected points, because it provided a wider variety of examples in each class that captured within-field spectral variability. The resulting field boundaries were
then rasterized into labeled images to be included in the training data (Figure 3).

Figure 3: Example of Worldview-2 imagery (a) and hand-digitized field boundaries (b). Example image is 2400 x 2400 pixels of 2 m resolution, centered at 25.0208° S, 29.5539° E.

3.3. Feature extraction

We overcame the problems of manual feature design and selection by extracting a comprehensive set of several thousand simple, highly correlated, and interdependent features, using an expanded version of Randomized Quasi-Exhaustive (RQE) features (Tokarczyk et al., 2013, 2015). These features were an alternative to calculating specific features tailored to one data set, increasing the generalization performance of the algorithm. A comparison between RQE and standard feature sets used in remote sensing applications is included in Tokarczyk et al. (2013, 2015).

The features were computed efficiently based on integral images, a pioneering image representation from computer graphics and computer vision (Crow, 1984; Viola & Jones, 2001). Integral images, also called summed area tables, are both a data structure and an algorithm, enabling fast, efficient calculation of the sum of values in a rectangular subset of an image, which forms the basis of each of the calculated features. An image is pre-processed into a summed area table in which a pixel, $s$, is defined as the sum of all original image pixels, $i$, that are located above and to the left:

$$s^\sum(x, y) = \sum_{x' \leq x, y' \leq y} i(x', y') \quad (1)$$

Each pixel in the summed area table can be efficiently calculated in a single pass using...
image pixels and previously calculated values, thus building up the table beginning in the top left corner:

\[ s(x, y) = i(x, y) + s(x - 1, y) + s(x, y - 1) - s(x - 1, y - 1) \]  
\[ (2) \]

Therefore, the sum of any rectangular region of the image can be computed in constant time, using the four pixels of the summed area table that define the corners of the rectangle (top left: A, top right: B, bottom left: C, and bottom right: D):

\[ \sum_{x_0 < x \leq x_1} \sum_{y_0 < y \leq y_1} i(x, y) = s(D) + s(A) - s(B) - s(C) \]  
\[ (3) \]

Using integral images, the features were generated within a sliding window, or small image patch surrounding the pixel of interest, thereby extracting contextual and textural data and acting as an intermediate between pixel-based and object-based classification (Bekker et al., 2014; Shotton et al., 2008; Tokarczyk et al., 2013, 2015). A sliding window of 5 x 5 pixels provided the best performance for our application. The features were extracted in an identical process for each pixel of interest by applying a set of feature types to a set of feature channels.

### 3.3.1. Feature channels

To extract features, we first constructed a stack of feature channels, consisting of the original imagery bands (Figure 4a) and several intermediate image products (Figure 4b-d) that drew additional information from the image pairs.

The first group of intermediate image products (Figure 4b) were calculated with mathematical morphological operations, which “simplify image data preserving their essential shape characteristics and eliminating irrelevancies” (Haralick et al., 1987). We applied morphological image transformations at two scales to smooth natural vegetation and highlight structural information in the image (Benediktsson et al., 2003). We used a series of basic morphological operators, including erosion, dilation, opening, and closing. We also use filter-
ing by reconstruction, which has been shown to better preserve shapes in the original image (Crespo et al., 1995).

The second group of intermediate imagery products (Figure 4c) consisted of probability maps of where circular objects are likely to occur, which highlighted center pivot agriculture. Without these added feature channels, center pivot agriculture can complicate classification, due to differences in the phenology of irrigated and rainfed crops. The probability maps were calculated by passing a simple circle detector, a common tool in computer vision, over the imagery and ranking potential circles based on the metric of circle strength.

After combining the original image bands and the intermediate image products, mean filtering was applied to each channel (Figure 4d). Mean filtering smoothed and removed noise from the images by replacing each pixel with the mean of surrounding pixel values within sliding windows of 3 x 3 and 5 x 5 pixels, which did not result in a significant loss of detail. Thus, with the addition of the intermediate image products and mean filtering, the training data set expanded from the original nine imagery bands to 72 feature channels.

Figure 4: Feature channels include (a) growing and off season images with panchromatic and eight multispectral bands (18 channels), (b) two morphological image transformations of panchromatic bands (4 channels), (c) circle detection in growing and off season images (2 channels), and (d) 3 x 3 and 5 x 5 mean filtering applied to all previous channels (48 channels). A set of feature types was applied to the resulting set of 72 feature channels during feature extraction. Example image is 2400 x 2400 pixels of 2 m resolution, centered at 26.4216° S, 25.9194° E.
3.3.2. Feature types

We defined a range of feature types, which served as templates for extracting features from the feature channels for a pixel of interest. The feature types (Figure 5) were calculated based on the pixel values or the mean value of a patch of pixels within a sliding window centered around a pixel of interest.

The feature types increased in complexity, yet remained efficient to calculate using integral images. Within a single feature channel, the first feature type (Figure 5a) extracted every pixel value within the sliding window, the second (Figure 5b) calculated the mean value of a patch around the pixel of interest, and the third (Figure 5c) compared the mean values of patches of differing sizes around the pixel of interest.

The next group of feature types utilized comprehensive pairs of feature channels to increase descriptive power. To compare the mean values of a patch in two feature channels, the first feature type (Figure 5d) took the difference, while the second (Figure 5e) introduced a nonlinear comparison similar to the calculation of the normalized difference vegetation index (NDVI).

To more efficiently describe a wider range of textures and extract positional information, an additional feature type (Figure 5f) included randomized elements. Several hundred patches of randomized size and position was selected within the sliding window, mirrored over the central pixel of interest, and the difference was calculated between the mean pixel value of the two patches.
3.4. Classifier selection

We used a random forest classifier from the field of machine learning, trained on the class labels from the hand-digitized field boundaries and the extracted feature vectors for the pixels in the training data set, to output the posterior probability of each pixel belonging to an agricultural field. The algorithm is a custom implementation of the random forest classifier that is fully parallelized and optimized for extracting and handling features from imagery. The algorithm code was written in Matlab/C++ and based on the work of Schroff et al. (2008) and Tokarczyk et al. (2013, 2015).

Originally developed by Breiman (2001), a random forest is a collection of decision trees, each trained on a random subset of the training data, which allows for full parallelization of the algorithm. Each decision tree is composed of nodes, which perform binary splits on the training data based on a threshold value for a selected feature and pass the training data to the next level of nodes. Each node is constructed by evaluating a random subset of features to maximize information gain. The training data passes through the decision tree, resulting in a pixel classification at the terminal node. Results are averaged across
the trees to obtain a posterior class probability for the pixel. The two main sources of random behavior, random subsets of data for training individual trees and random subsets of features for node selection, increase generalization performance by creating independent decision trees that do not overfit to the training data set (Bekker et al., 2014; Fan et al., 2005; Shotton et al., 2008; Wagstaff et al., 2013).

Variations of the random forest algorithm achieved the highest accuracy in an evaluation of 179 classifiers on 121 commonly used machine learning benchmarking data sets (Fernández-Delgado et al., 2014). Random forests are also gaining popularity outside the machine learning community, particularly in the natural sciences, because the algorithm excels in handling complex, nonlinear, high dimensional data sets (Cutler et al., 2007; Touw et al., 2013). Random forests have been used in a variety of applications, including ecological modeling (Cutler et al., 2007; Evans et al., 2010; Prasad et al., 2006), global tropical forest carbon mapping (Baccini et al., 2012; Mascaro et al., 2014), classification of geologic surfaces for space exploration missions (Bekker et al., 2014; Wagstaff et al., 2013), and medical image processing (Kaynig et al., 2010). They are also increasingly used for remote sensing land cover classification (Gislason et al., 2006; Ham et al., 2005; Jain et al., 2013; Mascaro et al., 2014; Pal, 2005; Rodriguez-Galiano & Chica-Olmo, 2012; Rodriguez-Galiano et al., 2011, 2012a,b; Shao & Lunetta, 2012; Vintrou et al., 2012), including those using high-resolution imagery (Adam et al., 2014; Hayes et al., 2014).

Several parameters of the random forest classifier were optimized for our application. Each tree in the random forest was trained using a different random selection of 5,000,000 training data points (about 12%) split evenly between field and non-field pixels, which was determined to be sufficient due to the high correlation between neighboring pixels. During the training phase, 100 trees were grown with a maximum tree depth of 10 levels of nodes.

3.5. Cross-validation scheme

To assess the algorithm's accuracy, we used an image-based, leave-one-out (LOO) cross-validation scheme, in which training data was drawn from seven of the eight images and
the remaining image was used for testing, rotating through all combinations. A LOO cross-validation scheme ensured that each agricultural type (rainfed subsistence, rainfed commercial, or center pivot) occurred in at least two of the training images, thus maintaining a representative array of examples for a machine learning classifier to learn from and capturing the agricultural variability of the targeted geographic area of application. Our cross-validation scheme simulates classifying a new image based on a curated database of hand-digitized field boundaries for future, large-scale deployments of the algorithm.

Using an image-based cross-validation scheme also ensured that the algorithm only classified previously unseen data and that highly correlated pixels within an image were not divided between training and test sets, which could inflate performance. Other cross-validation approaches that are based on pixels instead of images, such as randomly splitting pixels into training and testing sets, dividing an image into training and testing regions, or the “out-of-bag” validation scheme built into random forests, may cause the algorithm to overfit to the data set and produce erroneously high accuracy scores due to the high spatial correlation within images (Friedl et al., 2010; McIver & Friedl, 2001).

3.6. Performance metrics

To assess performance, we used two common metrics for binary classification in computer vision and machine learning: the receiver operating characteristic (ROC) curve and the related area under the ROC curve (AUC). ROC curves were first used to compare the performance of multiple machine learning classifiers by Spackman (1989).

Probabilistic output from a classifier, which defines the likelihood of pixels belonging to a class, can be converted to hard class assignments through thresholding. The accuracy of the class assignments for a particular threshold can be described in a confusion matrix, which produces a single point in ROC space (Figure 6). Numerous points from varying the threshold produce an ROC curve, which can be summarized by the AUC metric (Fawcett, 2006). For reference, AUC values of 1.0 correspond to perfect discrimination, while models with AUC values of less than 0.5 are considered to have no predictive capability. An AUC
value of 0.7 is usually determined as the threshold for usefulness in an application, while scores greater than 0.9 denote high performance and are the most useful for classification (Swets, 1988).

<table>
<thead>
<tr>
<th>Actual class</th>
<th>Field</th>
<th>Non-field</th>
</tr>
</thead>
<tbody>
<tr>
<td>Field</td>
<td>True positive</td>
<td>False negative</td>
</tr>
<tr>
<td>Non-field</td>
<td>False positive</td>
<td>True negative</td>
</tr>
</tbody>
</table>

**True positive rate:** \( TPR = \frac{TP}{TP + FN} \)

**False positive rate:** \( FPR = \frac{FP}{FP + TN} \)

Figure 6: Applying a threshold to the probabilistic output of a classifier results in class assignments (i.e. field or non-field pixel). A confusion matrix (a) compares the predicted classes to the actual classes. The true positive rate and false positive rate can be calculated from the confusion matrix, constituting a single point in ROC space (b). By varying the threshold, an ROC curve can be constructed to more fully describe a classifier’s performance, regardless of any particular threshold.

As more complex post-processing of probabilistic output replaces simple thresholding, ROC curves become a more useful indicator of performance than single confusion matrices. Furthermore, AUC has been shown to provide a superior metric of performance than accuracy, which is traditionally used, in part because accuracy disregards the confidence of class predictions provided by the probabilistic output of many machine learning classifiers (Huang & Ling, 2005).

For this work, metrics were calculated on a per-image basis, as opposed to a per-pixel basis, to assess generalization performance as the algorithm is deployed on previously unseen satellite images, which may vary widely in appearance. To provide further detail on the binary classification accuracy, the pixels belonging to agricultural fields in each image were divided into three sub-classes (rainfed subsistence, rainfed commercial, and center pivot) and ROC curves and AUC values were calculated for each.
3.7. Experiments

We carried out a series of experiments to evaluate the algorithm’s performance under varying conditions of data availability, recognizing that Sub-Saharan Africa does not yet have the same remote sensing coverage as other regions of the world. Specifically, we investigated the relative importance of temporal versus spectral depth by varying the number of images per site as well as the number of multi-spectral bands used out of the eight bands available in Worldview-2 imagery (Table 3).

Table 3: Combinations of multi-temporal imagery and multi-spectral bands used in sensitivity analyses.

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Multi-temporal images</th>
<th>Multi-spectral bands</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary analysis</td>
<td>Growing season, off season</td>
<td>Panchromatic, RGB, NIR1, coastal blue, yellow, red edge, NIR2</td>
</tr>
<tr>
<td>Multi-temporal sensitivity analysis</td>
<td>(a) Growing season</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(b) Off season</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(c) Growing season, off season</td>
<td></td>
</tr>
<tr>
<td>Multi-spectral sensitivity analysis</td>
<td>Growing season, off season</td>
<td>(a) Panchromatic, RGB</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(b) Panchromatic, RGB, NIR1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(c) Panchromatic, RGB, NIR1, coastal blue, yellow, red edge, NIR2</td>
</tr>
</tbody>
</table>

4. Results

4.1. Primary analysis

In the primary analysis, where the algorithm was trained using both growing season and off season images, as well as the panchromatic band and all eight multi-spectral bands, the mean AUC value across images (n=8) was 0.91 ± 0.02 (Figure 7). For individual agricultural sub-classes, mean AUC values were greater than 0.90 for center pivot, rainfed commercial, and rainfed subsistence, signifying high performance (Figure 8).
Figure 7: Receiver operating characteristic (ROC) curves for each image (n = 8, M = 0.91, SE = 0.02). Classifiers were trained on both growing season and off season images as well as the panchromatic band and all eight multi-spectral bands in a leave-one-out cross-validation scheme (trained on seven images and tested on remaining one image).

Figure 8: Mean and standard error of AUC values of each agricultural sub-class across images: center pivot (M = 0.93, SE = 0.04), rainfed commercial (M = 0.92, SE = 0.03), and rainfed subsistence (M = 0.90, SE = 0.02).

In addition to overall performance, the algorithm’s output was assessed qualitatively by examining how well areas of high probability of belonging to an agricultural field corresponded to hand-digitized fields. We aimed to achieve not only high overall accuracy, but also show coherent regions of similar posterior probability corresponding to fields, which are
important for further work on segmentation of individual fields.

Figure 9 depicts examples of the algorithm’s performance. Fields vary in appearance (color, texture, brightness, vegetation type, phenology) between the examples, yet generally correspond to areas of increased probability of belonging to a field over background pixels in Figures 9a - 9c. Figure 9a shows discretization between individual smallholder fields and accurately distinguishes between uncultivated or abandoned fields and active fields with a high degree of confidence. Figure 9b depicts differentiation of smallholder fields, though they contain trees and shrubs, from natural background vegetation. Figure 9c displays differentiation of commercial agriculture, both rainfed and center pivot, from background vegetation into cohesive regions of similar confidence, regardless of the fact that the half of the center pivot has the opposite growing season of the rainfed fields. Figure 9d, however, depicts a case of lower algorithm performance, in which degraded soil areas and abandoned fields are assigned similar posterior probability values as pixels within active fields.
4.2. Sensitivity analyses: Importance of multi-temporal images and multi-spectral bands

Training the classifier with multi-temporal imagery (both growing season and off season images) increased AUC values and decreased standard error, compared to training with a single image (Figure 10a). However, training the classifier with an increasing number of the eight available multi-spectral bands in Worldview-2 imagery had little effect on the mean
AUC values or standard error of the classifier (Figure 10b).

Figure 10: Mean and standard error of AUC values across images (n=8) for (a) multi-temporal sensitivity analysis, which trained the algorithm by varying the number of images per site, but using the panchromatic band and all eight multi-spectral bands: Off-season image (M = 0.87, SE = 0.03), growing season image (M = 0.88, SE = 0.03), and both images (M = 0.91, SE = 0.03); and (b) multi-spectral sensitivity analysis, which trained the algorithm by varying the number of multi-spectral bands, but using both images (growing season and off season): Pan + RGB (M = 0.90, SE = 0.02), Pan + RGB + NIR1 (M = 0.90, SE = 0.02), and Pan + 8 bands (M = 0.91, SE = 0.02).

For individual agricultural sub-classes (Figure 11), rainfed types of agriculture, both commercial and subsistence, responded more positively to training with multi-temporal imagery than center pivot irrigated agriculture, as shown by increasing AUC values and decreasing standard error (Figure 11a). However, center pivot agriculture showed slightly increased AUC values when trained with all eight multi-spectral bands (Figure 11b).
5. Discussion and conclusions

We have presented here a new approach to mapping agricultural fields, based on efficient
extraction of a vast set of features followed by a random forest classifier. The generalizability
of this approach is demonstrated by high and comparable performance across spectrally
indistinct smallholder fields and more easily distinguishable commercial fields. Though we
were limited to leave-one-out cross-validation to maintain a representative array of training
examples, our image-based cross-validation scheme, our use of the ROC/AUC metrics, and
the number of images in our data set provide a significant improvement over existing work
and further boost confidence in the generalization performance.

In our primary analysis using multi-temporal imagery (growing season and off season
images) as well as the panchromatic band and all eight available multi-spectral bands, our
algorithm achieved similar high performance (AUC > 0.90) across agricultural types. These
results are consistent with the hypothesis that the set of features was capable of capturing
subtle textural and contextual details, enabling the classifier to perform well in difficult cases.
of smallholder agriculture, such as small, irregular rainfed fields with indistinct boundaries and low differentiation between cultivated areas, abandoned fields, and natural vegetation. Incorporating circle detection in the features also allowed center pivot irrigated fields to be accurately classified, despite some fields having a growing season opposite of rainfed fields.

Our findings related to expanded feature spaces are supported by the work of Mukashema et al. (2014), who found that spectral values alone could not adequately distinguish Rwandan coffee fields in high resolution imagery, which also exhibit similarly high variability. The authors increased classification accuracy by extracting features from ancillary data, including a digital elevation map (DEM) and a forest map. Expanded feature spaces were further explored in Hayes et al. (2014), who overcame low spectral resolution (red, blue, green, and NIR bands) by using a random forest classifier and ancillary data to map agricultural fields in high resolution imagery of the United States. In comparison to the smallholder fields in our study area, the U.S. fields were more spectrally and geometrically distinct, necessitating only the addition of features from one texture measure, an NDVI channel, and a DEM to achieve the required accuracy.

In sensitivity analyses, we determined that multi-temporal imagery increased performance more than adding multi-spectral bands up to the eight bands available in Worldview-2 imagery, particularly for the rainfed subsistence and rainfed commercial sub-classes. These results suggest that seasonal changes in texture and patterns are more descriptive than the spectral signatures themselves, which can be highly variable within the dominant smallholder class. Moreover, multi-temporal data reduced variability in the results, as evidenced by decreased standard error in mean AUC values. The greater discriminative power of multi-temporal data shown in our results suggests that the rapidly increasing number of small, low cost satellites may provide a substantial benefit by improving return intervals. These “nanosats”, pioneered by such companies as Planet Labs and Skybox, place greater emphasis on timely, high-resolution imaging, as opposed to greater spectral depth. The degree to which higher temporal resolution may be more advantageous than greater spectral depth
would need to be further explored in tests against broader range multi-spectral or hyper-
spectral datasets, which at this level of spatial resolution are primarily confined to aerial
imaging systems.

Similar to our results, Löw et al. (2015) found that several multi-temporal images gen-
erally resulted in higher accuracy and lower uncertainty for high-resolution classification of
Central Asian irrigated agriculture compared to single date images. Also, Yan & Roy (2014)
employed multiple years of weekly Landsat data to extract crop fields in the United States,
highlighting the ability of multi-temporal data to handle the problems of crop rotations and
inter-annual variability, which are also relevant to Sub-Saharan African agriculture.

As temporal resolution of hyperspatial imagery increases in Sub-Saharan Africa, future
work can explore whether a single pair of growing season and off season images captures the
most crucial temporal patterns in the simplest form for the classifier, or if the use of expanded
time series of images is warranted. Additional multi-temporal imagery could highlight more
subtle seasonal pattern changes that would help differentiate between cultivated fields and
abandoned fields or degraded soils. These cases had lower accuracy than differentiating
between cultivated fields and natural vegetation, due to similarity of visual patterns. Finer
temporal resolution could also improve accuracy in cases where the temporal gap in image
pairs created disagreement (e.g. a cultivated field in the growing season image was unplowed
natural vegetation in the off season image).

As shown in Figure 9, the classification algorithm produced coherent areas of similar pos-
terior probability, which are critical for further post-processing to extract individual fields as
well as assessing the confidence of the output for decision making applications. A study by
Ozdarici-Ok et al. (2015) achieved promising results in segmenting individual agricultural
fields in high-resolution images of Turkey, using global conditional random field (CRF) label-
ing with graph cuts, a segmentation technique from image processing and computer vision
that smooths the posterior class probabilities from a classifier (Boykov & Kolmogorov, 2004;
Boykov et al., 2001). Though most of the studied fields were clearly defined, rectangular in
shape, and single-cropped, their approach could be expanded to include larger feature spaces that are crucial for distinguishing smallholder agriculture in Sub-Saharan Africa.

Overall, our algorithm provides a flexible and parallelized framework that can be expanded for growing data demands, such as additional training data as well as increases in temporal and spectral resolution of imagery. While the algorithm has shown high generalization performance, adding training data that is representative of the targeted application areas, especially those with unique agricultural characteristics, can aid in increasing the geographic range. To maximize computational efficiency, future algorithm development will focus on optimizing training data using active learning (Crawford et al., 2013; Egorov et al., 2015; Foody & Mathur, 2004). This iterative approach would interactively select the most informative training images for hand-digitization, based on image similarity metrics and changes in algorithm accuracy. As opposed to random selection of images for hand-digitization, our goal is to minimize the training data while maximizing algorithm accuracy as well as select new images to expand the geographic range of our algorithm. In addition, this approach could be combined with crowd-sourcing to expedite the process of hand-digitization (Estes et al., 2015).

Finally, our approach has shown the ability to generalize well across large areas and disparate agricultural land cover types, paving the way for its large-scale deployment in which new images are classified using a curated database of hand-digitized field boundaries. In addition to providing high-resolution maps of agricultural land cover and reducing uncertainties in existing data sets, applying this algorithm over Sub-Saharan Africa can provide a critical and improved constraint for regional crop productivity (Estes et al., 2013; Rojas et al., 2011) and agro-climatological assessments (Estes et al., 2014), as well as evaluations of land cover change.
Acknowledgements

This work was supported by funds from the Princeton Environmental Institute, including the Walbridge Fund, the Mary and Randall Hack ’69 Research Fund, and the Program in Science, Technology, and Environmental Policy (PEI-STEP) Fellowship; the NASA Jet Propulsion Laboratory Strategic University Partnerships (JPL SURP) Graduate Research Program (1524338); and the National Science Foundation (SES-1360463 and BCS-1026776).


scale Farmers in a Liberalized Trade Environment* (pp. 21–52).


Butusov, O. (2003). Textural Classification of Forest Types from Landsat 7 Imagery. *Map-
ing Sciences & Remote Sensing, 40*, 91–104.


Centre for Global Development, & The African Population and Health Research Centre
(2014). *Delivering on the Data Revolution in Sub-Saharan Africa: Final Report onf the
Data for African Development Working Group*. Centre for Global Development.

Cotula, L., & Vermeulen, S. (2009). Deal or No Deal: The Outlook for Agricultural Land
Investment in Africa. *International Affairs, 85*, 1233–1247.


by reconstruction. *Signal Processing, 47*, 201–225.

graphics, 18*, 207–212.

Cutler, D. R., Edwards, T. C., Jr, Beard, K. H., Cutler, A., Hess, K. T., Gibson, J., &


List of figure captions

- Figure 1: Examples of false color (NIR1, red, green) images of smallholder (a, b) and commercial (c, d) agriculture in Sub-Saharan Africa, highlighting the highly variable spectral characteristics of smallholder agriculture. Example images are 500 x 500 pixels of 2 m resolution, centered at: (a) 25.0383° S, 27.9955° E, (b) 25.0234° S, 29.5551° E, (c) 26.4228° S, 25.9284° E, and (d) 25.9859° S, 25.9573° E.

- Figure 2: Worldview-2 imagery sites in South Africa (n=8)

- Figure 3: Example of Worldview-2 imagery (a) and hand-digitized field boundaries (b). Example image is 2400 x 2400 pixels of 2 m resolution, centered at 25.0208° S, 29.5539° E.

- Figure 4: Feature channels include (a) growing and off season images with panchromatic and eight multi-spectral bands (18 channels), (b) two morphological image transformations of panchromatic bands (4 channels), (c) circle detection in growing and off season images (2 channels), and (d) 3 x 3 and 5 x 5 mean filtering applied to all previous channels (48 channels). A set of feature types was applied to the resulting set of 72 feature channels during feature extraction. Example image is 2400 x 2400 pixels of 2 m resolution, centered at 26.4216° S, 25.9194° E.

- Figure 5: Descriptions of feature types that were applied to the set of feature channels during feature extraction.

- Figure 6: Applying a threshold to the probabilistic output of a classifier results in class assignments (i.e. field or non-field pixel). A confusion matrix (a) compares the predicted classes to the actual classes. The true positive rate and false positive rate can be calculated from the confusion matrix, constituting a single point in ROC space (b). By varying the threshold, an ROC curve can be constructed to more fully describe a classifier’s performance, regardless of any particular threshold.
Figure 7: Receiver operating characteristic (ROC) curves for each image (n = 8, M = 0.91, SE = 0.02). Classifiers were trained on both growing season and off season images as well as the panchromatic band and all eight multi-spectral bands in a leave-one-out cross-validation scheme (trained on seven images and tested on remaining one image).

Figure 8: Mean and standard error of AUC values of each agricultural sub-class across images: center pivot (M = 0.93, SE = 0.04), rainfed commercial (M = 0.92, SE = 0.03), and rainfed subsistence (M = 0.90, SE = 0.02).

Figure 9: Examples of algorithm’s performance in image subsets of 250 x 250 pixels (2 m). From left to right, the columns display the growing season and off season images, the hand-digitized field boundaries, and the classifier output of the posterior probability that pixels belong to an agricultural field. Centers of image subsets are located at: (a) 25.0224° S, 29.5542° E, (b) 25.3052° S, 31.2272° E, (c) 26.4242° S, 25.9050° E, and (d) 23.2124° S, 30.5684° E.

Figure 10: Mean and standard error of AUC values across images (n=8) for (a) multi-temporal sensitivity analysis, which trained the algorithm by varying the number of images per site, but using the panchromatic band and all eight multi-spectral bands: Off-season image (M = 0.87, SE = 0.03), growing season image (M = 0.88, SE = 0.03), and both images (M = 0.91, SE = 0.03); and (b) multi-spectral sensitivity analysis, which trained the algorithm by varying the number of multi-spectral bands, but using both images (growing season and off season): Pan + RGB (M = 0.90, SE = 0.02), Pan + RGB + NIR1 (M = 0.90, SE = 0.02), and Pan + 8 bands (M = 0.91, SE = 0.02).

Figure 11: Mean and standard error of AUC values of each agricultural sub-class across images (n=8) for: (a) multi-temporal sensitivity analysis, which trained the algorithm by varying the number of images per site, but using the panchromatic band and all eight multi-spectral bands; and (b) multi-spectral sensitivity analysis, which trained
the algorithm by varying the number of multi-spectral bands, but using both images (growing season and off season).