# Improving smallholder representation in crop maps

Kendra Walker[1,2], Atahualpa Ayala[1], Lauren Sharwood[2], Robert Heilmayr[1,2]

[1]Bren School of Environmental Science & Management, University of California, Santa Barbara, Santa Barbara, CA 93106-5131
[2]Environmental Markets Lab, University of California, Santa Barbara, Santa Barbara, CA 93106-5131, USA

**Abstract**

Farms smaller than five hectares produce more than half of the world's food. However, many remotely sensed maps of croplands have been developed in ways that preferentially identify large, intensively managed fields. Here, we illustrate how remotely-sensed land cover maps could be improved to yield better information about smallholder fields, using Paraguay as a study system. We show that prior land cover maps dramatically underestimate smallholder cropping in Paraguay due to misclassification of these systems as a mixture of pastures, shrublands and forests. These maps fail to classify 55-99% of smallholder fields ($<$ 5 ha) as croplands, overlooking more than 50 to 120 thousand farmers (or 16-43% of the farming population) in the country. We develop national-scale land cover maps that more accurately identify smallholder croplands, conducting a series of experiments to identify specific methodological practices that enable the greatest improvements in smallholder cropland mapping. We show that modest adjustments to the training sample can improve smallholder crop recall in Paraguay by more than 35 percentage points compared to the best performing existing crop map, with other modeling adjustments improving smallholder crop recall even further. Our analysis underscores the need for remote sensing scientists to more carefully consider heterogeneity in cropping systems and to ensure that all farming systems are represented in training data for agricultural mapping.

***Keywords*** Paraguay · remote sensing · smallholder farming · crop mapping · training data · model sensitivity · algorithmic bias

## Highlights

- Prior maps misclassify 55-99% of smallholder croplands
- Our crop maps accurately represent small fields without sacrificing overall accuracy
- Common practices for training data collection fail to characterize small fields
- Resolution and segmentation improve smallholder recall, but training data is key

# 1 Introduction

Smallholder agriculture is critical for global food security as it supports the livelihoods and nutritional needs of a large portion of the planet's marginalized people. More than half of food calories produced globally come from areas where the average farm size is smaller than 5 ha (Samberg et al., 2016), while 30-35% of the global food supply comes from farms that are smaller than two hectares (Ricciardi et al., 2018). Such small farms directly affect the lives of two to three billion people (Woodhill et al., 2022). Both because so many people derive sustenance from smallholder farms and because such farms are particularly vulnerable to changing climate and markets, Sustainable Development Goal (SDG) target 2.3 seeks to support smallholders by increasing their productivity, incomes, and access to land. The ability to develop policies towards SDG 2.3 and to assess their success depends heavily on the ability to accurately locate smallholder farms and monitor changes (Fritz et al., 2024; See et al., 2015; You, Sun, 2022).

Despite the importance of smallholder farms, they are often overlooked in agricultural maps and statistics. Although more than 80% of farms worldwide are smaller than 2 ha, these small farms occupy just over one-tenth of agricultural land (Lowder et al., 2021). As a result, a data product can theoretically have over 90% accuracy in mapping agriculture while missing all smallholder farms. There are several reasons that smallholder farms are much harder to map than larger farms including heterogeneity, edge-effects, and lack of training data. Training data for crop mapping are more readily available from industrial farms, which, unlike smallholder farms, are often officially registered. The difficulty in obtaining training data for smallholder crops is compounded by the likely need for more training data for such crops compared to industrial crops because smallholder systems are more diverse and frequently mixed with other classes due to small field sizes (edge effects) (Weitkamp, Karimi, 2023).

Despite these difficulties, recent efforts have seen some success in mapping smallholder farming on local to global scales. Methods have been developed that can successfully segment very small fields (smaller than one hectare) using very high-resolution images, such as 0.5 m WorldView (McCarty et al., 2017; Neigh et al., 2018) and 2.5 m SPOT (Wang et al., 2022). Some success has been shown in segmenting fields in highly fragmented landscapes with 10 m Sentinel-2, but accuracy remains low for fields smaller than 2.56 ha (Song et al., 2023). Several projects have also reported mapping smallholder fields successfully with 3 m PlanetScope imagery, including mapping tillage practices (Liu et al., 2022) and crop type in India (Rao et al., 2021), field boundaries in Ghana (Estes et al., 2022), and a complex smallholder landscape across Mozambique (Rufin et al., 2022). However, since these commercial data sources can be expensive and have limited historical data, they pose challenges to standardized monitoring and comparison with historical land use. Maps developed with such sources are often one-off products, limited in their ability to assess systemic change, and may quickly become outdated given the rapid pace of agricultural change (Zeng et al., 2018; Bullock et al., 2021; You, Sun, 2022). For these reasons, we focus on methods that rely on publicly available data that can easily be replicated through time both to produce baseline analyses and to support future monitoring.

Efforts to map smallholder crops with publicly available data, such as Landsat and Sentinel, typically focus on specific crops or smaller regions. For example, Bey et al. (2020) map smallholder crops over a portion of Mozambique using Landsat and MODIS, Descals et al. (2024) map industrial and smallholder oil palm globally using Landsat and Sentinel data, and Jin et al. (2019) and Azzari et al. (2021) map smallholder corn at national scales in eastern Africa using Sentinel-1 and Sentinel-2. Lebourgeois et al. (2017) map smallholder crops in Madagsascar using Sentinel-2 and Pleides (0.5 m), but find that the very high resolution Pleides is only necessary if object-based methods are required. While these works provide insight into the mapping of specific smallholder cropping systems, they do so in areas where smallholder crops are the primary cropping system, thus limiting the ability to scale the methods up to larger areas where smallholder crops are a minority class. To our knowledge, our work is the first effort to use publicly available satellite data to map all smallholder and large-scale crops with high accuracy at a national scale in a landscape where smallholder crops are an important, yet minority, class.

The objective of this work is to advance methods to map smallholder systems within a diverse landscape containing both large-scale and smallholder farms, exploring how choices regarding training data and other modeling components affect accuracy in both smallholder and general crop maps. Paraguay presents an interesting landscape for such exploration, as it exhibits a similar farm-size distribution as seen on the global scale, while including many smallholder farming systems common to developing nations. Although well over 40% of farms in Paraguay are smaller than 5 ha, with the majority of those smaller than 2 ha, these farms occupy a very small fraction of the agricultural area. Farms larger than 100 ha, on the other hand, comprise six percent of farms but the vast majority of land area (Fig. 1b). Given this unequal distribution, a map of cropland in Paraguay that targets large-scale crops could have high accuracy in mapping cropland while missing a large proportion of farms and farmers. We explore whether it is possible in such skewed situations to increase the accuracy of mapping a rare minority class without sacrificing overall map accuracy. Given that smallholder crops occupy less than 5% of the landscape in Paraguay, our optimization task faces the challenges of any optimization problem aimed at improving the mapping of a rare class: a focus on overall accuracy will overlook improvements in the rare class, while too much focus on the rare class can overlook significant degradation in the overall product (Waldner et al., 2019b). Using a framework combining point-based and larger-area samples, we constructed measures to rate models on both overall accuracy in mapping crops and accuracy in mapping crops in areas with high smallholder activity.

We first describe existing crop maps for Paraguay in Section 2.2, then, in Section 3, outline our own general approach for modeling crops, which combines strengths of existing approaches with adjustments aimed at capturing more smallholder activity using publicly available data. We then describe in Section 3.5 the framework and set of tests we designed to systematically manipulate model components and select the best model to map smallholder crops in a situation where crop is a minority class compared to other land covers and smallholder crop is a minority class within crop. Our tests reveal that, despite good overall accuracy in mapping crops, existing maps capture very little of the smallholder crops in Paraguay. Based on our insights from our testing process, we make suggestions in Section 4.2 about elements of the modeling process that would be the most impactful for efforts to increase representation of smallholder farms in crop maps, and provide a general critique of algorithmic bias in crop mapping with common practices in section 5.

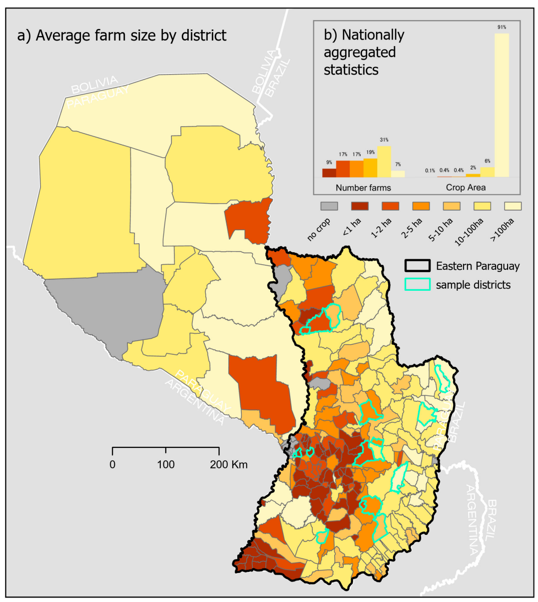



Figure 1: Farm size distribution in Paraguay according to the 2022 National Agricultural Census (MAG, 2023) (a) by district and (b) as aggregate statistics. Note that although farms are large in western Paraguay, the vast majority of crops are grown in eastern Paraguay. Our results suggest that the total crop area in smallholder fields (< 5 ha) shown in (b) is underestimated, but that the distribution is similar. Highlighted sample districts are those used for area-based optimization described in Section 3.4.3.

# 2 Background

## 2.1 Agriculture in Paraguay

Crops in Paraguay are primarily produced in the more fertile and populated eastern region, with agriculture in the more arid western region dominated by grazing lands. According to the most recent national agricultural census (MAG, 2023), about 98% of crops are produced in eastern Paraguay, although eastern Paraguay only comprises 40% of the country's land area (Fig 1). Soy is by far the dominant crop in Paraguay, with wheat and corn grown on the same plots in the off-season. The rapid expansion of soy in Paraguay in the beginning of the century has had potentially profound effects on the economy and livelihoods of smallholder farmers (Wesz, 2022). Soy accounts for over 40% of total exports in Paraguay, compared to 12% in 1995 (Avila Schmalko, Monroy Sarta, 2018). Between 2002 and 2014, the area occupied by agribusiness expanded by 139%, while peasant agriculture decreased by 50% (Ortega, 2016). As this agribusiness frontier has spread from the far eastern border with Brazil westward, smallholder farmers have been pushed to more marginalized lands (Wesz, 2022; Garcia-Lopez, Arizpe, 2010), experienced increased pressure to clear protected forests (Grossman, 2015), and have often chosen to abandon agriculture altogether (Soldi et al., 2019).

Despite this decline in area allocated to smallholder farming, the number and proportion of smallholder farmers has grown between 2008 and 2022 (MAG, 2023). Smallholder farms (defined in Paraguay as farms smaller than 50 ha) make up the majority of farms in Paraguay, with very small farms (smaller than 5 ha) comprising 43% (Fig. 1b). Because farms are typically made up of multiple fields, this means that an even greater majority of fields are smaller than 5 ha. However, as is true at the global level, the strong skew in field sizes in Paraguay means that crop maps can achieve reasonably accurate estimates by measuring only large fields while missing all the small fields and thus smallholder farmers. Overcoming these challenges to better map smallholder farming is necessary to fully understand linkages between agricultural industrialization and rural livelihoods and to promote policies to support family farms and improve food security, as pledged in SDG 3.2.

## 2.2 Existing crop-area maps and products for Paraguay

Many maps already detail croplands in Paraguay, including global and regional data products. We first explore the ability of existing products to accurately depict smallholder cropping systems. For our purposes here, we define smallholder crops as annual and perennial herbaceous crops (including sugar, cassava and horticulture; excluding forage, pasture, fallow, woody crops and bananas) in fields smaller than 5 ha. Given that many fields in Paraguay are much smaller than 1 ha, it is unlikely that they would be well represented in a product with a resolution of 100 m (1 ha equals one pixel). While 30 m resolution (1 ha equals $\tilde{1}1$ pixels) should be adequate to capture most smallholder fields, 10 m resolution (1 ha equals 100 pixels) would be ideal. Several global land cover products that include a crop category now exist at 10 m resolution, including the 10 m WorldCover from the Euopean Space Agency (ESA) (Zanaga et al., 2022), Dynamic World (Brown et al., 2022) and ESRI Land Cover (Karra et al., 2021). Through visual inspection, we found the ESA WorldCover map to have more detail for our study area than the other two (corroborated more generally by Xu et al. (2024)).

Global land cover maps have many competing objectives for optimization and accuracy assessment and might not represent crops as well as products specifically designed for crop mapping. Thus, we examine global crop maps, including the 30 m Landsat-derived Global Rainfed and Irrigated Cropland Product (LGRIP) from NASA MEaSUREs (Teluguntla et al., 2023), the 30 m Landsat-derived GLAD crop map from the University of Maryland (Potapov et al., 2022), and the 10 m WorldCereal Temporary Crops product from the ESA (Van Tricht et al., 2023b). We also included a 10 m global soy product (Song et al., 2021) as a comparison, although we did not expect it to capture most smallholder crops, which are often not soy.

Finally, we examined products more specific to the region, with the theory that more localized products would capture more diversity at that scale. These include the 30 m Atlantic Forest Trinational Collection 2 map by MapBiomas (MapBiomas, 2022), the organization with the most expertise in mapping the region, and a 30 m land cover map of the South American Southern Cone (Graesser et al., 2022), which is based on methods that provide the foundational structure for our own methodology. Table 1 provides a summary for all crop products explored, including the product's definition of crops, which varies from product to product but is not found in our analysis to have a large impact on the representation of smallholder cropland.

Table 1: characteristics of existing products providing crop estimates for Paraguay

| | Years | Res. | Region | Crop class | Training data | Source |
|---|---|---|---|---|---|---|
| **Ag.Census** | 2022 | district | national | >25 annual crops (+ forage, excluded here) & 9 perennial crops | Full census of 454,950 farm units (no model) | MAG (2023) |
| **WorldCover** | 2021 | 10 m | global | crop class (annual only(includes sugar)) 1 of 11 classes | stratified by class and "training location", but unclear what comprises a training location | Zanaga et al. (2022) |
| **LGRIP** | 2015[a] | 30 m | global | binary (irrigated & rainfed), all cultivated plants for food, feed & fiber. Includes perennial & tree crops. | >130,000 sample points with global distribution, from various sources. min mapping unit = .8 ha (90 m x 90m) | Teluguntla et al. (2023) |
| **WorldCereal** | 2021 | 10 m | global | binary crop (annual only (includes sugar)) | large but opportunistic. Points closest to Py. from industrial crops in Argentina (Boogaard et al., 2023) | Van Tricht et al. (2023a,b) |
| **World soy** | 2000 - 2021 | 30 m | global | binary, crop is soy only | extensive but focused only on soy | Song et al. (2021) |
| **MapBiomas** | 2021[b] | 30 m | Atlantic Forest & Chaco | two crop classes (annual & perennial) of 11 classes | stable samples (same LC for at least 30 of 35yrs), stratified by subregion (biome) | MapBiomas (2022) |
| **S.Cone** | 2000 - 2019[c] | 30 m | SA Southern Cone | crop class (annual only (includes sugar & forage)) 1 of 8 classes | ground samples (mostly industrial crops) in Argentina & Uruguay + systematic samples with FAO grid (3 polygons digitized within, focused on large crops) | Graesser et al. (2022) |
| **GLAD** | 2019[d] | 30 m | global | annual & perennial herbaceous crop, (includes forage & fallow up to 4 yrs., excludes woody crops & shifting ag) | Strategic selection of 924 Landsat tiles. Crop/no-crop samples selected within by visual interpretation in high-res imagery | Potapov et al. (2022) |
| **Fritz_hybrid** | 2019 | 500 m | global | uses GLAD definition | hybrid of GLAD & WorldCereal. Sampling only in some areas of disagreement. | Fritz et al. (2024) |
| **GeoWiki** | 2017 | 500 m | global | Field size map. Fields include crop and pasture. | random sampling within generous crop mask. 168 points in Py. | Lesiv et al. (2019) |
| **CELPy** | 2021 | 10 m | national | annual & perennial herbaceous crops (includes sugar, cassava and horticulture; excludes forage, pasture, fallow, banana & woody crops) | active iterative sampling throughout Py., both on ground and with high-res ($\leq .5m$) imagery | This |

[a] More recent years available for some locations, but only 2015 map available for Paraguay as of 12/2024.
[b] other versions available 1999-2022.
[c] 2019 map used here.
[d] available in 4-yr mosaics 1999-2019.

# 3 Methods

## 3.1 Overview

Our general processing framework mimics procedures from the Harmonized Landsat and Sentinel-2 (HLS) project (Claverie et al., 2018) (as the HLS data for our region were not available at the onset of this project) and builds on methods from a prior project to map land cover across the South American Southern Cone (Graesser et al., 2022). We introduced modifications to this methodology to meet our specific goal of mapping crop types across Paraguay, including higher-resolution processing (10 m), different training data, a different feature set including semantic segmentation, and different modeling architecture. Our full methodology to map specific crop types across Paraguay is presented in Walker et al. (2025) and is more complex than might be necessary to simply map where crops are present, which is our objective here. This added complexity presents an opportunity to test the impact of various methodological decisions on the accuracy of resulting crop maps.

The primary classes of interest to us for this work are what we call "mixed crop", "homogeneous crop", and aggregated "non-crop". These primary classes are defined as:

**Mixed crop:** crop grown in conditions common to smallholder systems. A pixel is labeled as mixed crop if it contains herbaceous crops during the year and meets one or more of the following conditions: 1) the width of the field (the narrowest of the two primary dimensions) is less than 60 m, 2) the pixel contains a mixture of different crops, 3) the pixel contains a crop that we did not observe in large fields in eastern Paraguay (peanuts, beans, cassava, horticulture, tobacco, or sesame). (Note that we would theoretically like to have homogeneous crop classes for larger fields in the latter category, but did not have the training data to do so, as discussed in Section 4.2.3. Most of these crops are grown almost exclusively in smallholder or mixed-cropping systems in Paraguay, although sesame is observed in large fields in western Paraguay.)

**Homogeneous crop:** crop grown in conditions common to large fields, although actual field size, beyond the 60 m threshold for mixed crops, was not considered when labeling data. A pixel is considered homogeneous crop if it contains herbaceous crop during the year, does not meet the conditions of mixed crop, and is at least 10 m from the field edge. Pixels along the edge of fields containing homogeneous crops are not within the homogeneous-crop or mixed-crop classes, but rather assigned to a separate crop_edge class, as explained in Section 3.2.3.

**Med crop:** included as a separate crop class containing tall perennial or permanent crops (bananas, yerba-mate, vineyards, and orchards). This class was excluded from both crop and noncrop datasets for analyses here because it does not match the definition of crop for most of the other products we reviewed. Thus, although yerba-mate and bananas are often grown within smallholder systems, they are not included in our analysis of smallholder crops here.

**Non-crop:** all other land covers. For training and mapping purposes, these are divided into up to 22 classes (see Section 3.2.3 and SI Figs. A1 A2), however, for calibration and validation purposes here, they are grouped into a single "non-crop" class.

## 3.2 Land cover model

### 3.2.1 Pixel-level features

We use a pixel-level modeling approach – even when field-level segmentation is added (see Section 3.2.2), it is applied at the pixel level. Pixel-level features are generated from smoothed time series of all Landsat and Sentinel-2 images available for the time period (July 2021 to July 2022, plus a three-month padding on each side). The images are first processed into four spectral indices (kNDVI, GCVI, NDMI and NBR), which were determined to provide as much information as six indices (with EVI2 and WI included) during a preliminary phase of model optimization (Walker et al., 2025). Following this preliminary optimization, the pixel-based feature set determined to be most useful includes basic statistics for each index for the year, the wet season (Nov-Mar), and the dry season (May-Sept), as well as the phenological sequence represented by the smoothed kNDVI value from the 20th day of each month. More complex phenological variables, such as the rate of greenup, the rate of senescence, the length of season, and the peak of season, were explored but were not found to be important to map crop presence and are therefore excluded here. For the full list of pixel-level variables and more details on the methods, see Walker et al. (2025).

### 3.2.2 Polygon features (field segmentation)

For the model with segmentation, polygon-level features were added to the default model in the form of six additional bands that match the resolution of the pixel-level features. These bands include three pixel-level measures: crop probability, crop border probability, and distance to the border (if crop probability is greater than zero), as well as three polygon-level measures: field size, field homogeneity, and perimeter efficiency. The perimeter efficiency measure captures the deviation from a square and is helpful in identifying polygons that likely contain multiple smaller fields. Field homogeneity is the standard deviation of the average Nov/Dec GCVI value (after smoothing) for all pixels in a given polygon. The three pixel-level measures are based on work by Waldner, Diakogiannis (2020) and are the direct output of Cultionet (Graesser, 2023), the open-source semantic segmentation tool that we used. Cultionet was explicitly developed to map cropland extent and uses satellite image time series as inputs in a convolutional neural network model. We trained our segmentation model with 1026 1 km x 1 km sample chips comprising the same time series data for all four indices used in the pixel-level model along with a vector layer of manually digitized crop boundaries. Most of the training chips (726) were randomly sampled across eastern Paraguay, while 100 were randomly sampled across western Paraguay and another 200 were added in areas with more smallholder farming activity. The three inference outputs from the Cultionet model were used directly in our crop model and were also used to extract polygon vectors to estimate field areas and derive the three polygon-level features. A minimum mapping unit of 30 m x 30 m was required for polygon features due to the need for both interior crop extent and border pixels for this process. Smaller fields can have values in the first three pixel-level segmentation bands but not in the three polygon-level bands. See Walker et al. (2025) for more details on all aspects of the segmentation model.

### 3.2.3 Land cover schema

Although our analysis is based on the three primary classes, mixed crop, homogeneous crop, and non-crop, as defined in Section 3.1, the underlying maps were generated using many sub-classes to ensure representation of different land covers across Paraguay's heterogeneous landscape. Most pertinent to this analysis, a group of mixed classes beyond mixed crop were included to facilitate increased representation of other types of mixed pixels in experiments where we increased the representation of mixed crop in the training data, as discussed in Section 3.5. Early models without these classes were found to exaggerate crop area in other areas containing mixed vegetation and bare land, such as paths and grass edges, prompting these to be added as classes. Likewise, the crop_edge subclass was added to homogeneous crops to counter the ring of mixed crop pixels classified around the border of homogeneous fields in the absence of this class. This ring will not matter for general crop/no-crop mapping, however if a producer wants to quantify smallholder crops specifically, as we did here, mixed-crop borders around homogeneous fields need to be removed. This can be accomplished with a post-classification filter (which we also used, as discussed in Walker et al. (2025)), but we found that most could be removed preemptively by creating this crop_edge class and increasing its representation in the training data at the same proportion as mixed crop.

Our final classification scheme includes 36 land cover classes (summarized in SI Fig. A1 and documented in further detail in Walker et al. (2025)). While this many classes is unlikely necessary for simple crop/mixed-crop/non-crop mapping at a national scale, we included training data for all classes with the theory that greater representation of landscape diversity would improve the model as long as class size is small compared to the dominant classes and each non-crop class can be accurately discerned from crop in high-resolution imagery. Although we expected that including examples of as many classes as possible would improve landscape representation in a model, we did not necessarily expect that a model required to separate these extra classes would perform better than a simpler model. We conducted experiments to determine the ideal level of class complexity for this case as part of the optimization process, discussed in Section 3.5.

## 3.3 Model training

### 3.3.1 training dataset

Our pixel-level models are trained with pixel-level (point) data from various sources, including data from the Ministry of Agriculture, ground sampling, and interpretation of high-resolution imagery. The Ministry of Agriculture provided centroid coordinates for around 5000 fields growing industrial crops (soy, rice, and sugar) at a recent (but unknown) time, which we verified in Google Earth, removing any that did not appear to be crop in 2021/2022. To make our training data more representative of smaller agricultural systems in Paraguay, we collected additional samples, focusing on areas with more smallholder activity.

We conducted a ground sampling campaign focused on determining crop type for the 2021/2022 growing season. Although this level of detail is not likely necessary for mapping cropland in general, as is our objective here, the information from the ground provided important insights and feedback to calibrate image interpretation for the remaining training data. Ground samples were selected opportunistically in that they were confined to points observable from roads navigable in the wet season. As points were viewed from the roadside, they were also confined to locations where the land cover class 15 m away could be seen and interpreted unambiguously (i.e. away from class edges). Road samples have been found to be as effective in generating training data as random samples as long as the resulting sample is representative of the classes (Waldner et al., 2019a). However, we found it difficult to gather representative training data from the roadside, especially with regard to smallholder fields. This is because the properties of smallholder farmers are usually aligned with the houses along the road and the fields behind and thus inaccessible, as shown in Fig. 2).


Figure 2: Drone image of smallholder farms in Paraguay, showing typical configuration with houses and tree fences along the road and crops farther back.

To increase spatial representation of our training data to cover areas away from roads, increase representation in smallholder areas, and supplement data on non-crop land covers, we collected additional training data using image interpretation in WorldView or similar imagery in Google Earth. This sample included points randomly selected within the footprints of relevant images as well as points strategically selected to increase observations for under-represented classes and classes that were commonly misclassified as crop in initial model iterations. Sampling methods are described in more detail in Walker et al. (2025).

### 3.3.2 Training methods

We developed an efficient process using Google Earth and iterative active learning to expand our dataset beyond the samples already available from the ground campaign and the Ministry of Agriculture. For each iteration, we used our existing training data to classify a selection of 20 km by 20 km blocks representing the full landscape, then inspected misclassified areas in high-resolution imagery to create high-value data points and determine whether new classes existed that could be easily distinguished from crop in such imagery. If so, new training points for these classes were gathered opportunistically from the imagery to represent as broad a landscape as possible within the constraints of our available processed data. Others have found a similar iterative active learning framework to be effective (Tuia et al., 2009; Patel, Patel, 2023) when using probabilities and variance to select new sample areas. We found that direct selection of pixels in post-classified images can effectively direct efforts to difficult areas and thus dramatically reduce the time required to interpret and label new training data.

We expanded the class schema with iterations of this active sampling methodology adding classes as we discovered areas of confusion in our crop model. For example, wet grasslands were separated from the broader grassland category because they were frequently confused with rice in an initial model, and low-density palm forests were created as a category because they were often confused with bananas and smallholder crops. This is essentially opportunistic stratified sampling, where these classes would be captured with a very large sample size but are missed when sample sizes are lower because they occupy only a small proportion of the landscape. Although they comprise relatively small areas, classes frequently confused with crop can have a significant impact on accuracy in specific locations, especially where crop is also a minority class. By having many non-crop classes, we ensure that the training data represents the diverse landscape, although for the purpose of validation and analysis we aggregate all non-crop classes into this single category. The final list of classes can be found in SI Fig. A1.

The proportional representation of classes in the training data can have substantial impact on model performance (Foody et al., 1995; Weitkamp, Karimi, 2023; Gao et al., 2023). General practice suggests balancing the number of training samples in each class in accordance with its proportional representation in the landscape. However, when classes are highly imbalanced in the landscape, resulting models tend to be biased toward the majority class (Chawla, 2010; Waldner et al., 2019b). To reduce this bias, we first ensured that the majority classes are not overrepresented, by drawing a distribution from our full dataset to best match the proportions estimated from a preliminary map. We then compromised between proportional and equal allocation for minority classes by setting a minimum sample size for all classes. The total and minimum sample sizes were varied in a set of optimization experiments, as was the proportional representation of the mixed classes, as discussed in more detail in Section 3.5.

## 3.4 Model calibration and validation

### 3.4.1 Model performance metrics

As our primary interest is in the ability of the model to correctly find instances of smallholder crops, we focus on recall of the crop class (also called sensitivity, true-positive rate, or producer's accuracy) to evaluate model performance. Recall is the number of "hits", or labels correctly classified as belonging to a class, out of all observations for that class. An advantage of recall is that it is independent of prevalence as it only relies on the observations from a single class (Luque et al., 2019). Unlike with many other performance metrics (Guesne et al., 2024), recall for different models and maps can be compared directly without concern for how well the sample represents the actual class prevalence or how balanced the classes are. This is especially useful for cases such as this with high class imbalance, and for preliminary optimization tasks, where a holdout sample might be used instead of a more statistically rigorous random sample.

While our objective is to maximize crop recall, we do not want this to occur at the expense of non-crop recall (which is the true-negative rate, or the specificity, for the positive crop class). In other words, we do not want the model to improve crop recall by sloppily classifying a lot of non-crop pixels as crop as well. For this reason, we evaluate non-crop recall alongside crop recall. Because our focus here is on smallholder crops (largely represented by the mixed-crop class), we separate this mixed-crop class from the homogeneous-crop class and evaluate the recall for each. Without this split, homogeneous crops, which make up around 80% of the crop class, would dominate the recall metric and likely obscure any information about the minority mixed crops. Except for the purpose of area estimation, we are not interested in confusion between these two types of crop and in most cases consider this as a binary matrix, where a crop in either crop category is considered a "hit" as long as it is classified as crop.

Comparing models is easy in cases where recall for both crop and non-crop is higher for one model compared to another. However, it is often the case that improved recall for one class occurs at the expense of the recall for the other class. To choose the optimal model, a metric is thus needed that accounts for the performance for each class simultaneously. The ideal metric for this task is a topic of long-running debate (Pontius, Millones, 2011; Stehman, Foody, 2019) because the results and interpretation of overall performance metrics depend on sample composition, underlying prevalence, and the producer's valuation of tradeoffs in defining "optimal". For our task, for example, we prioritize high recall of mixed crop, but we do not want to allow too much error of commission to degrade the amount of non-crop correctly identified. However, the definition of "too much" is subjective. One way to obtain a more objective target for optimization is to acquire area estimates for each class prior to modeling them and selecting the model that is best able to replicate these areas. Obtaining such apriori area estimates can be difficult, however, especially in the case of minority classes. Furthermore, mixed-crop area (and resulting errors) vary substantially across the landscape, thus requiring spatially explicit area/error assessments for effective optimization.

We address this optimization dilemma with a two-step approach. we first use the F score as a preliminary means to compare models' ability to correctly identify crop while minimizing false instances. The actual F scores are rather meaningless (Foody, 2023), however, and optimizing on F score becomes a futile task when models are close to one another, especially when the actual area of mixed crop is unknown. We therefore include a second step, where the best candidates are promoted to a more rigorous and computationally expensive area-based evaluation for final model selection and validation. This area-based method involves wall-to-wall mapping of seventeen sample districts, as described in Section 3.4.3. Our final model is the one with the smallest absolute difference between the model- and sample-based crop-area estimates for all 17 districts.

### 3.4.2 Point-based calibration data

Our two-step model optimization process involves first point-based calibration data from a 20% holdout sample of the training data described in section 3.3.1 to evaluate class recall and inform initial model comparison, followed by area-based calibration data described in section 3.4.3 for more rigorous model selection. The holdout sample for the first step includes the three primary classes evaluated with the model performance metrics: mixed crop, homogeneous crop, and non-crop. The same holdout dataset was used to evaluate all models. We balanced subclasses (described in Section 3.2.3 and SI Fig. A1) within each main holdout class to reduce overrepresention of any single class, but designed the calibration set to drastically overrepresent the mixed-crop category overall. The mixed holdout sample comprises pixels assigned to the mixed class as defined in Section 3.1. The homogeneous crop class comprises pixels assigned to one of the homogeneous crop labels in a field larger than 2 ha. To simplify interpretation of results (for calibration only), we filtered the homogeneous crops set to exclude pixels from fields smaller than 2 ha. Field size was determined by overlaying our sample data with the polygon data described in section 3.2.2. Med-crops (defined in Section 3.1) were excluded from the calibration data.

### 3.4.3 Area-based calibration from district sample

While general accuracy metrics provide good initial insight into a model's performance, it is ultimately difficult to pinpoint the best model to produce a map focused on a rare class, as was our goal here, based on accuracy metrics alone. For a deeper assessment of performance for our top model candidates and other existing maps, we produced wall-to-wall maps for ten districts dominated by smallholder farming (average field size smaller than 5 ha), three districts dominated by industrial farming (average field size greater than 50 ha), and four districts with intermediate field size (Fig. 1). The districts were randomly selected within each category, with a maximum of two districts allowed per department to broaden the spatial representation. The average field size for each district was estimated from district-level crop data from the 2022 agricultural census (MAG, 2023), which provides the total area and number of farms in each district growing each crop for more than 35 individual crops. These average-field-size estimates align fairly well with estimates from the 2017 Geo-Wiki field-size map (Lesiv et al., 2019). While the census data can provide a good estimate of average field size, they do not provide reliable estimates of total crop area for each district because the number of crop cycles in each field is unknown; simply summing the area of all crops grown for each district (as we have done in Table 3) will overestimate the crop area if two or more crops are grown on the same field in a year. To get a better estimate of crop area for our 17 sample districts, we randomly sampled 300 points within each district and determined whether the point was crop or not by viewing it in high-resolution WorldVew imagery, using the same protocol as our training data. Information for the sample districts is provided in SI Table 6. A discussion on the error bounds on these estimates is also provided in the SI.

#### 3.4.4 Validation data

Unlike the calibration set used in optimization, the validation set used to assess our final products and compare with other products was designed to be comprehensive and fully representative of the landscape. Full validation of our final maps is provided in Walker et al. (2025). From the purpose here of evaluating performance in mapping crops in general and smallholder crops in particular, we used the data from the 5100 randomly sampled points within the 17 randomly sampled districts, as this dataset is independent of our training data. Because these districts were selected to highlight smallholder farming, this sample offers a large set of observations of smallholder crops while maintaining a random structure. This sample is helpful in assessing the strengths and weaknesses of products for different crop size categories. Assessing the overall accuracy for other products in Paraguay, on the other hand, would require a random sample across Paraguay and was not our objective here. To assign validation points to field-size classes, we overlaid our segmentation results on the sample points and extracted the field-area estimate for each point. Cases of points without polygons almost always indicate that the point is in a field smaller than 1 ha. Such cases were nonetheless reviewed with high-resolution imagery to ensure proper field-size assignment.

## 3.5 Model tests and optimization

We designed a series of tests to evaluate the contribution of various modeling components to the mapping smallholder agriculture. We hypothesized that the most important components driving the relative success in mapping smallholder areas that we observed in early models were feature inputs to the model, specifically the higher resolution of our inputs, as well as our use of segmented field features created with deep learning methods.

In addition to the feature set, the composition of the training set can affect the model outcomes by both the information that is included and how that information is balanced. It would be easiest to train our crop model on data already available from the Ministry of Agriculture. However, we expected that the lack of representation of smallholder cropland in the dataset provided would result in poor model performance in smallholder areas. Given that smallholder crops make up a very small percentage of Paraguay's cropland (3-4% in preliminary estimations), however, a model trained with too much information from these heterogeneous areas might perform worse overall. The tradeoff in providing the model with enough nuance in training data while maintaining adequate balance to not exaggerate the presence of small land covers applies both to smallholder crops and to crops in general. Paraguay's landscape is naturally quite imbalanced, with a crop to non-crop ratio around 25:75 in eastern Paraguay and far more skewed in the western Chaco region. Among our 36 land-cover classes, many represent very small percentages of total land area and would not be captured with typical random sampling protocols. For our base model, we balanced classes to maintain the general crop/no-crop balance seen on the ground while balancing subclasses to match proportions expected with random sampling, but with a minimum sample size for each class (varied from 50 to 200). This minimum sample size overrepresents small classes in the training data, but is necessary to provide the model with adequate information from these small classes. We then tested the effect of this balancing approach versus other sampling protocols, such as pure random, as well as the effect of varying parameters within the balancing framework. Under the theory that more heterogeneous classes need greater representation in the training data, we also explored the effect of upsampling the mixed-crop and other mixed classes.

The full set of models explored (with different mixtures of feature model, sample model, and thematic model components) are outlined below and summarized in Table 2.

- Feature model variations:
  - $\mathbf{F_{10}}$: Default feature model with inputs at 10 m resolution. To enable direct comparison with all models, this model only includes pixel-based features and not the six polygon features discussed in Section 3.2.2 (included in the $\mathrm{F_{SEG}}$ model below).
  - $\mathbf{F_{30}}$: Reduced spatial resolution, by downsampling time-series outputs to 30 m prior to feature creation

  - $\mathbf{F_{SEG}}$: All pixel-based features as well as the six polygon-based features derived from field segmentation (see Section 3.2.2) at 10 m resolution.

- Sample model variations:
  - $\mathbf{S_r}$: Simulated random sample. Includes training data from the sets described in Section 3.3.1, with class sizes proportional to the observed class proportion in preliminary maps (and no minimum class size). The objective is to have classes represented in the training data as similarly as possibly to that of a random sample.
  - $\mathbf{S_e}$: External crop data: includes only crop data from a dataset provided to us by the Ministry of Agriculture, consisting of points in soy, rice, and industrial sugar fields. This dataset was first checked against high-resolution imagery to remove stray points and points that no longer appeared to be crop in 2021/22. This is expected to be similar to a model with no mixed-class samples.
  - $\mathbf{S_x}$ (x=1to10): Re-sampling the mixed-crop class to x times the balanced mixed-class proportion (a conservative 1.65% was used for this purpose, which is about half the actual proportion). By default, other mixed classes (field_edge, grass_edge, and path) are also resampled to the same degree.

- Thematic model (classification schema) variations:
  - $\mathbf{C_{32}}$ default 32-class model
  - $\mathbf{C_2}$ to $\mathbf{C_{36}}$: alternative class models, from binary crop/no-crop model (C2), crop/no-crop with an additional mixed crop class (C3), crop/no-crop/med-crop/mixed-crop (C4), and other groupings with more detail within both the crop and non-crop classes (C8, C10, C12, C16, C18, c21, C23, C25, C36 – Fully described in SI Table A2).

Table 2: Model variation framework and tested models

| | | |
|---|---|---|
| **Feature Model (F)** | Spectral bands | $\mathbf{F_0}$: 4 indices from 6 bands (not varied) |
| Set of information provided for each instance | Spatial resolution | $\mathbf{F_0}$: 10 m |
| | | $\mathbf{F_{30}}$: 30 m |
| | Spatial context (texture, pixel vs. polygon) | $\mathbf{F_0}$: pixel only |
| | | $\mathbf{F_{SEG}}$: segmentation bands |
| | Variable selection | $\mathbf{F_0}$: (not varied here) |
| **Sample Model (S)** | | $\mathbf{S_0}$: $=S_6$ (after prelim testing) |
| Set of instances used to train the model | Training data source /Class balance | $\mathbf{Se}$: External crop data |
| | | $\mathbf{S_r}$: Random sample |
| | | $\mathbf{S_{2\text{-}8}}$: upsampling of rare class |
| **Schematic Model (C)** | Classification schema | $\mathbf{C_0}$: 32 classes |
| Classes used to represent the data | | $\mathbf{C_2}$: binary (crop /no-crop) |
| | | $\mathbf{C_3}$-$\mathbf{C_{36}}$:other schema |
| **Model architecture (A)** | Classifier | $\mathbf{A_0}$: Random Forest |
| Methods to generate model from given feature/instance set | | $\mathbf{A_{GB}}$: Gradient boosting |
| | Hyperparameters | $\mathbf{A_0}$: 100 trees |
| | | $\mathbf{A_{x200}}$- $\mathbf{A_{x500}}$: 200 to 500 estimators |

# 4 Results

## 4.1 Crop area and field size distribution in Paraguay

Our model is able to characterize small-scale agricultural activity in areas of very small field sizes and high landscape heterogeneity (Fig. 3). By joining our final model with our segmentation results, we estimate that smallholder fields smaller than 5 ha make up around 17% of total crop area in eastern Paraguay, with fields smaller than 2 ha making up 11-13% of total crop area (Fig. 4). A more robust area adjustment procedure based on the errors of the random sample in the sample districts, extrapolated to all of eastern Paraguay, yields estimates that 16.7% of fields are smaller than 5 ha and at least 10% are smaller than 2 ha (SI Table A4). Whichever estimation method is used, our results suggest that smallholder farms occupy considerably more than the <1% area detailed in Paraguay's agricultural census, but are still a minority class, comprising only 3-4% of the total land in eastern Paraguay. This is in line with our mixed-crop class, which we estimate covers 3.61 ±0.6% of eastern Paraguay (SI Table A2).

Our total crop area estimates, shown in Table 3 are only a few percentage points higher than the general agreement among prior remotely sensed products of 18-20% crop cover for eastern Paraguay. The products outside of this general agreement are LGRIP, which overestimates crop area for all district-level farm size categories, and S.Cone, GLAD, and the Fritz-hybrid map (which combines WorldCereal and GLAD), which underestimate crop area for all categories.

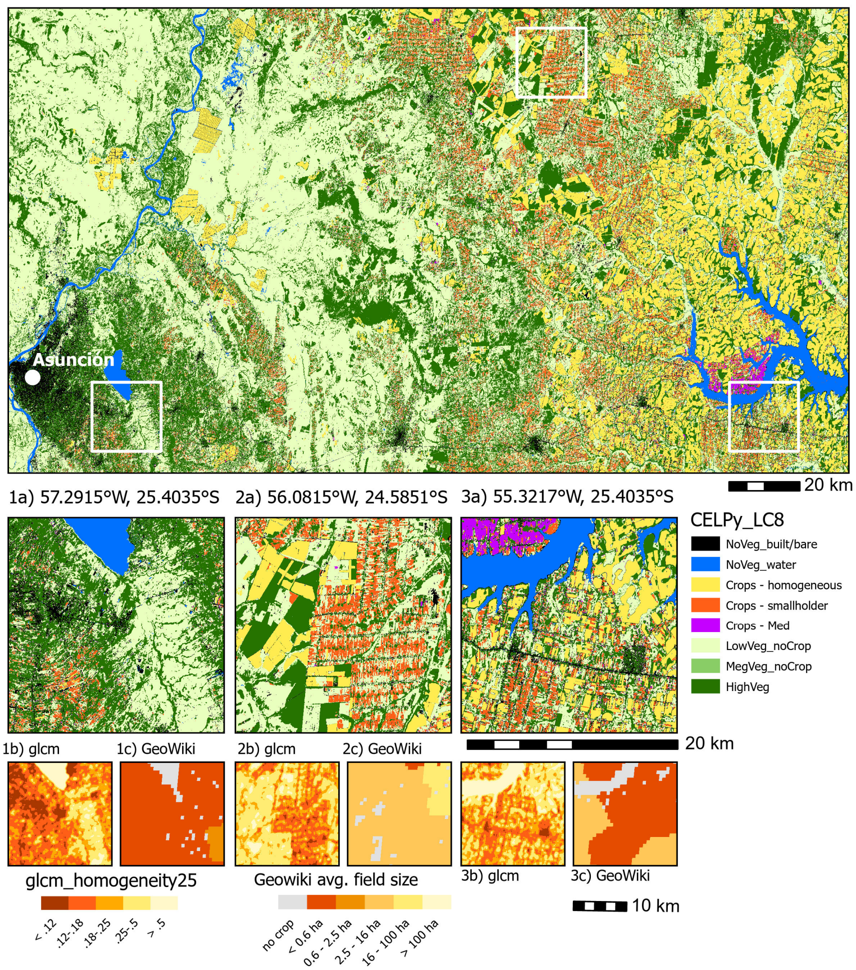



Figure 3: Three example 20 km x 20 km patch outputs from CELPy, 8-class model (classified with 32 classes and regrouped to 8). Corresponding homogeneity measures from a 25 cell gray-level co-occurrence matrix (glcm) window and GeoWiki average field size maps (Lesiv et al., 2019) are provided as examples of possible methods to increase sampling effort around smallholder crops.

Table 3: Estimates of % crop area in Paraguay in 2022

| | | | | District farm size bracket[a] | | | |
|---|---|---|---|---|---|---|---|
| | eastern | western | national | <2ha | 2-5ha | 5-16ha | >16ha |
| Ag.Census[b] | 28.8% | 0.4% | 11.7% | 5.0% | 19% | 33% | 58% |
| CELPy | 22.4% | 2.4% | 10.2% | 6.3% | 17% | 24% | 36% |
| MapBiomas | 20.2% | 1.0% | 8.6% | 3.1% | 12% | 21% | 36% |
| WorldCover | 20.8% | 2.4% | 9.7% | 3.1% | 13% | 22% | 37% |
| WorldCereal | 20.7% | 0.1% | 8.3% | 2.6% | 12% | 22% | 40% |
| WorldSoy | 16.0% | 0.1% | 6.4% | 0.7% | 7% | 16% | 30% |
| S.Cone(2018) | 12.8% | 0.3% | 5.3% | 0.5% | 5% | 13% | 28% |
| LGRIP(2015) | 36.8% | 14.8% | 23.3% | 20% | 32% | 41% | 49% |
| GLAD | 17.2% | 0.4% | 7.1% | 1.2% | 8% | 17% | 34% |
| Fritz_hybrid | 18.1% | 0.4% | 7.4% | 1.3% | 9% | 19% | 34% |

[a] Average farm size calculated from district-level census data MAG (2023). Districts with <1% crop in both Ag.Census and CELPy are excluded from calculations.
[b] Census data do not provide a strict area estimate as multiple crop cycles on the same land are summed for the year.

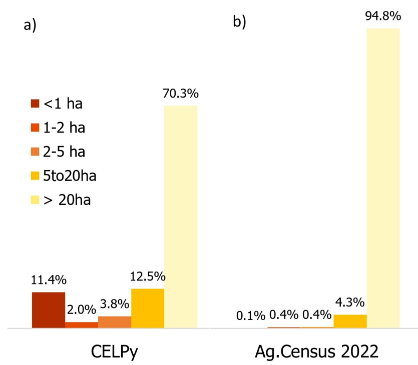



Figure 4: Active crop area distribution in 2022 by field size in eastern Paraguay, based on a) our CELPy model and b) the 2022 agricultural census data summary. Note that these data from the agricultural census are aggregated farm sizes rather than fields; as a farm can contain multiple fields, these figures are expected to be shifted a bit to the right. However, full smallholder land holdings are not usually larger than 5 ha.

Although our overall crop area estimates are similar to other maps, Table 3 shows that our district-level estimates are consistently higher than other products in districts where the average field size is small. Fig. 5 illustrates the clear relationship between the average field size and the percentage of cropland captured at the district level, with most products capturing almost no cropland in districts where average farm size is smaller than 2 ha. These observations that other products are missing crops grown in smallholder fields are corroborated by our recall assessment shown in Table 4. While most products have high recall rates for homogeneous crops, they have very low recall rates for the mixed crops typical of smallholder fields. (Although LGRIP does have a relatively high recall rate for smallholder crops, the low recall rate for no crop indicates a large overmapping of crops for this product. This can be seen in SI A3, which is Fig. 5 with LGRIP included.)

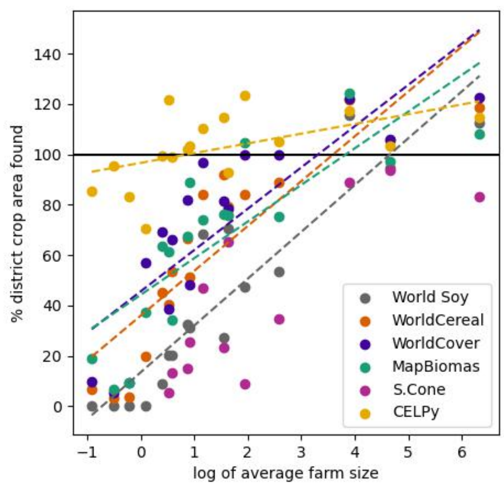



Figure 5: Relationship between average farm size in district and % of the crop area captured by different models. Percent crop area targets are estimated with 300 point random samples for each district. Percent captured is the comparison of that target with the crop area mapped by the product (e.g. if the target crop area is 30% and a map shows 6%, the percent of crop captured is 20%).

Table 4: Recall for crop (by field size) and no-crop, from independent sample

| | Smallholder crop recall | | | | Large crop recall | | Non-crop |
|---|---|---|---|---|---|---|---|
| | All | <1ha | 1-2ha | 2-5ha | 5-16ha | >16ha | recall |
| CELPy | 80% | 75% | 96% | 98% | 98% | 99.7% | 97% |
| WorldCover | 45% | 34% | 60% | 81% | 91% | 97% | 97% |
| WorldCereal | 32% | 23% | 46% | 77% | 88% | 97% | 98% |
| MapBiomas | 26% | 18% | 44% | 60% | 75% | 91% | 97% |
| WorldSoy | 8% | 2% | 9% | 39% | 65% | 87% | 99% |
| S.Cone | 1% | 1% | 0% | 0% | 57% | 71% | 99% |
| LGRIP | 79% | 78% | 78% | 87% | 87% | 94% | 67% |

Field size estimates based on our segmentation results or high resolution imagery in cases where our segmentation process did not produce polygons (common with very small fields). Recall scores based on sample of 5500 points randomly located in 17 sample districts.

The lack of capture of smallholder crops may have a small effect on crop area estimates at the national level, yet the effect on crop distribution is profound. Figure 6 shows how crop areas are underestimated by other products at the local level, while Figure 3 shows the impact across a larger landscape.

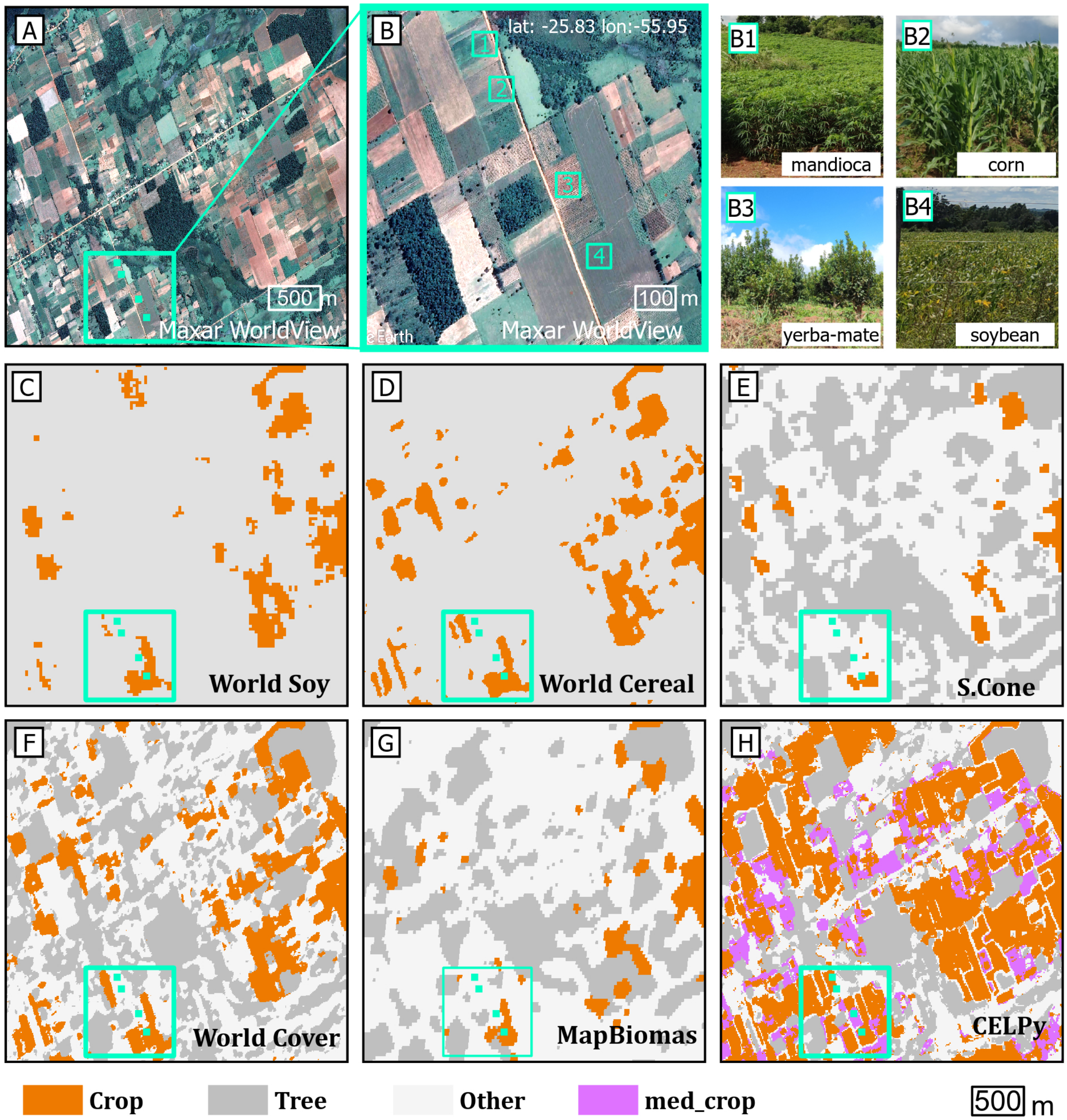



Figure 6: Comparison of cropland mapping by existing products and CELPy in a sample smallholder area. Trees are included to facilitate orientation, but note that C) and D) are binary crop/no-crop products

## 4.2 Methodological experiments to improve smallholder mapping

Here we conduct a variety of tests to isolate the individual contributions of different modeling decisions towards the objective of mapping smallholder crops without degrading overall model performance. Our main model variations are summarized in Table 5, and reveal that the most important component in improving the recall of smallholder crops is the sample model (training data), while the feature model is the most important for the overall results. The details of the results for each component are discussed in the following sections.

Table 5: Effect of model structure on performance

| | default | $F_{30}$ | $S_e$ | $S_r$ | $F_{SEG}$ |
|---|---|---|---|---|---|
| Recall mixed crop (as any crop) | 0.77 | 0.72 | 0.13 | 0.33 | 0.80 |
| Recall homogeneous crop (as any crop) | 0.93 | 0.91 | 0.90 | 0.90 | 0.97 |
| Recall no_crop | 0.87 | 0.84 | 0.97 | 0.95 | 0.91 |
| Kappa (crop/no_crop) | 0.70 | 0.63 | 0.69 | 0.69 | 0.79 |
| F1_(crop/no_crop) | 0.78 | 0.74 | 0.48 | 0.76 | 0.88 |
| OA_(crop/no_crop) | 0.89 | 0.86 | 0.91 | 0.89 | 0.89 |
| avg % of district crop % predicted | | | | | |
| for smallholder districts | 170 | 177 | 33 | 74 | 99 |
| for industrial ag. districts | 116 | 116 | 113 | 112 | 112 |

All model components are default unless specified: $S_6$, $F_{10}$, $C_{32}$, $A_{RF}$. See Section 3.5 & Table 2 for description of models. The final CELPy model is $F_{SEG}$ (or $F_{SEG}S_6C_{32}A_{RF}$).

### 4.2.1 Features included in model

The model including segmentation performs significantly better than the default model, while both perform significantly better than a model downsampled to 30 m resolution (see Table 5). This ranking of feature model performance holds true for different variations in the sample model, although the sample model itself has the most influence on the ability to map smallholder fields.

### 4.2.2 Sample used to train model

The biggest increase in model accuracy in mapping smallholder crops occurs with the inclusion of mixed-crop samples in the training data. Figure 7 shows that the lowest F scores occur when there are no mixed-crop samples represented in the training data. The increase in recall for mixed crop and the overall F score for all models as mixed-crop samples are added is larger than the difference in those values between models at any fixed sample size. (Note that the models with segmentation applied cannot be separated fully from mixed crop training data to test the effect without it because smallholder fields were included in the training data for the segmentation process.)

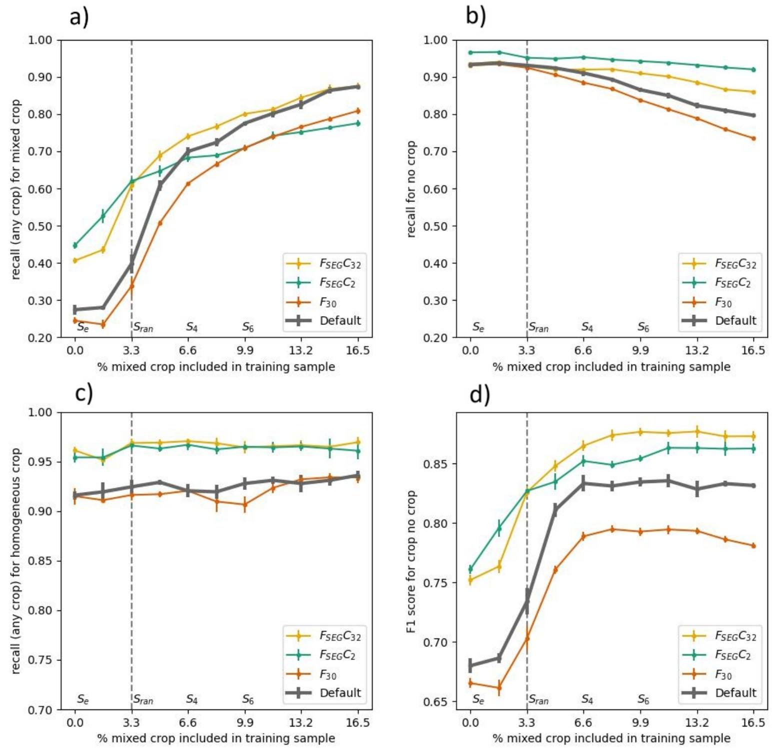



Figure 7: Effect of sample and feature model adjustments on crop mapping. A) Recall for mixed crop improves with representation in training data, while B) recall for no crop decreases with increased mixed crop in training data. C) Recall for homogeneous crops is relatively high and stable for all models. D) The biggest influence on overall model performance is the inclusion of mixed crop in the training sample, although the benefit plateaus around 8% due to the tradeoffs seen in A & B. The model with segmentation performs significantly better than the default model, while the models at 10 m resolution perform significantly better than a model at 30 m resolution. Vertical bars are standard deviations for ten model generations tested with the same holdout set. (Note: the default model is $F_{10}C_{32}$ without segmentation.)

The top two panels in Fig. 7 show the trade-off between improvement in recall for mixed crops and degradation of recall for non-crop as the number of mixed-crop samples increases in the training data. The recall for homogeneous crops, on the other hand, is little affected. This is because mixed crops resemble non-crop land covers more than they resemble homogeneous crops, as revealed in the principal component analysis in Fig. 8. Because non-crop landcovers occupy at least 75% of the landscape in eastern Paraguay, relatively small declines in performance in mapping non-crop can result in large overestimation of crop area. The model with segmentation offers little improvement in mapping the smallholder class specifically, but is an important improvement to the default model because it reduces errors of commission. The F scores shown here are useful for optimization relative to each other, but may not reflect true optimums for mapping crop area because the sample size of each class is not proportional to its expected representation. To ensure that the model with the optimum F score here is really performing the best, we produced wall-to-wall maps for the sample districts and selected the model the minimized the residuals (SI Fig.A4). The results of both tests show that the optimal model occurs at $S_6$, which has a mixed-class proportion of about 10%. This is about three times the proportion that would be expected from pure random sampling. After twice the proportion of random sampling, however, the marginal gains are small.

The gains from including about 10% mixed crop in the training sample occur regardless of sample size (after a class minimum of 50 samples is enforced). Fig 9 shows that the largest gains in mixed crop recall and the F score for crop/no-crop occur as the proportion of mixed crop is increased in the training data from zero to ten percent, regardless of the total sample size. As the total sample size is increased from 1000 to 5000 (and the mixed crop sample is increased proportionally from 100 to 500), slight gains are observed in the mixed crop recall and overall score, but these gains from the sample size are small compared to the gains from the proportional representation. These trends are true regardless of whether the segmentation features (which already include mixed and homogeneous crop training data in their processing) are included in the model.

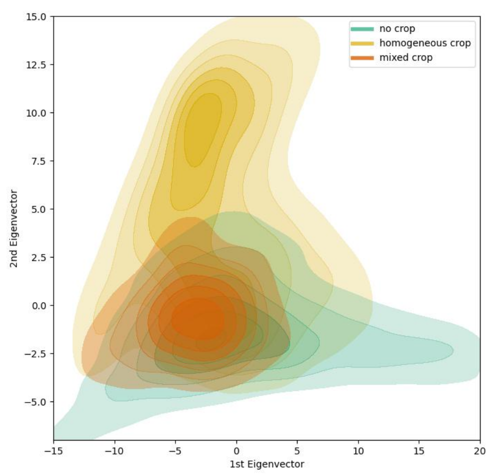



Figure 8: Principal Component Analysis reveals that crops on smallholder fields are more similar to non-crop landscapes than to the homogeneous crops on larger fields.

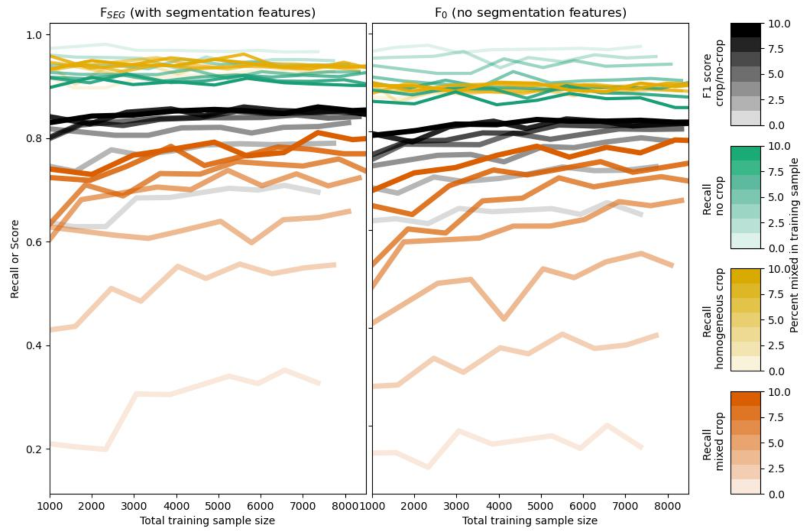



Figure 9: Effect of sample size and mixed composition on recall and F1 score of crops and non crop. Homogeneous crop recall is unaffected as total sample size increases from 1000 to 8000, whereas recall for mixed crop formations typical of smallholder fields increases mildly with sample size. However, the most impact on mixed crop recall occurs as its proportion in the training data is increased from 0-10%. These trends are similar whether segmentation features (which incorporate mixed-crop training data) are included in the model (left panel) or not (right panel).

**Influence of common sampling strategies**

Given that training data collection can be expensive, existing maps are often used to either directly supply training data or to limit the sampling space to areas most likely to provide productive data. However, the use of other maps may exclude important areas from the sample space. Fig 10 shows the breakdown by average field size of agreed crop area for five crop maps covering Paraguay. A sample drawn from agreed crop area for two of these maps would under-represent smallholder fields smaller than one hectare by 75% on average and 40% at best. Likewise, fields of one to two hectares would be underrepresented by 50%. The more maps combined to ensure that the crop area sampled is actually crop area, the more small fields will be excluded from this agreement sample. A very large portion of the agreement of these maps is also seen as crop in the World soy map (SI Table 6), further illustrating that the agreement occurs mostly in areas with homogeneous crops.

Rather than using other crop area maps to determine the location to sample for a new crop map, more information may be gained by using maps that provide different types of information, such as field size. We found that sampling from the smaller field strata of the Geo-Wiki field-size map (Lesiv et al., 2019) (see Table 1) would result in a sample with around 20% smallholder fields in Paraguay (Si Table 6), which would decrease effort by a factor of five compared to random sampling for a project trying to increase representation of smallholder fields.

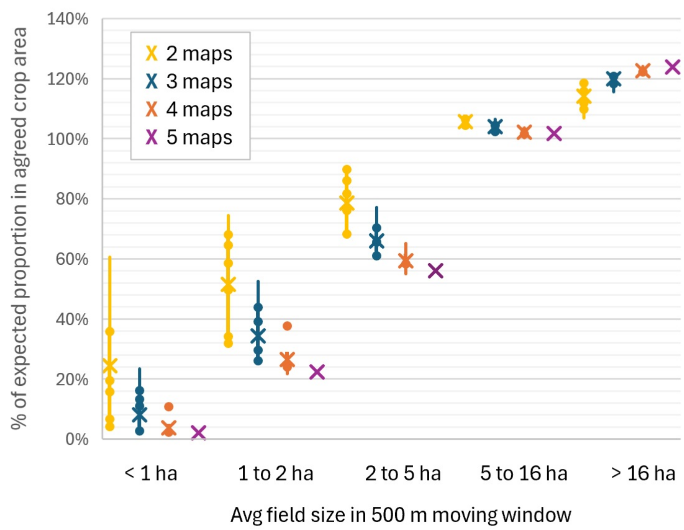



Figure 10: Agreement samples based on convergence of the evidence of existing maps vastly under-represent smallholder crops. Five maps used in this analysis are: World Cover, World Cereal, LGRIP, MapBiomas and S.Cone (see Table 1) y-axis is 1 - ($p_{ex}$-$p_{obs}$) /$p_{ex}$, with $p_{ex}$ being the expected proportion of crop for each category based on a 500 m moving window of the final CELPy map (SI Table 6) and $p_{obs}$ being the proportion of the agreed crop pixels belonging to that category

.

### 4.2.3 Classes included in the model (thematic model)

The thematic model composition had little effect on overall performance in mapping crops in general, but a perceptible effect on mapping smallholder crops. This is especially true for binary models or models with only a few grouped classes representing crop and non-crop. Fig.11 shows that recall of the smallholder class increases by 10-20% as model complexity increases from around five to around ten classes, although the recall for no-crop decreases slightly in turn, resulting in similar overall performance (measured by the F1 score). Our best model occurs with a thematic schema comprising 32 classes, although the gains for mapping the presence/absence of crops are insignificant after 23 classes. The additional classes were included for other mapping purposes and are retained here because they do not hurt the model. It is important to note that the results of our thematic model are context-dependent and likely change with the quality and composition of the training data available for each class. The most complex model we tested resulted in decreased accuracy, likely due to poor training data for these additional classes rather than a limit on complexity for a random forest model. Our final four classes (cassava, horticulture, sesame, and tobacco) are very heterogeneous and likely require more training data than more homogeneous classes. However, we had very small and uncertain samples for these classes and likely a high degree of contamination with the general mixed-crop class. In this case, grouping these classes with the mixed-crop class provided better results. This makes sense as these crops are almost always grown on smallholder fields in eastern Paraguay.

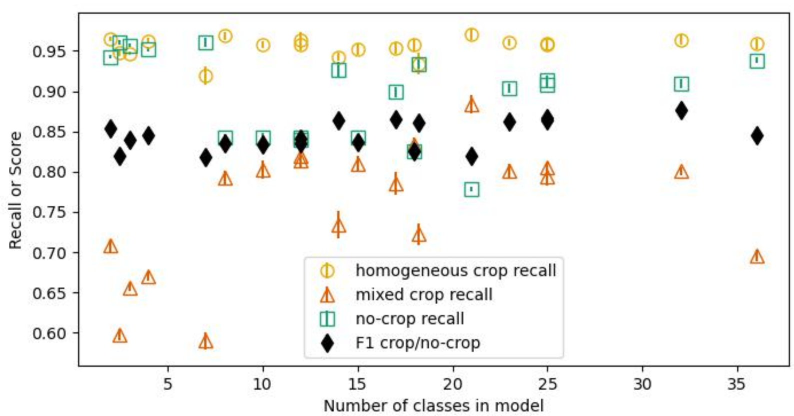



Figure 11: Accuracy metrics for random forest models with different class complexity. The classes included for each model are provided in SI Fig A2. Class complexity has little impact on overall model performance even when more than 30 classes are included. Small changes can have bigger impacts on minority classes, however. The order in which classes are included is context-dependent and not indicative of an ideal number or order. This does indicate, however, that a two step binary-first model would result in lower mixed class recall than a single, more complex model.

#### 4.2.4 Model architecture

Random Forest performed slightly better than Gradient Boosting for all model versions, but the increase in the F score for the best model was only 0.02 (SI Fig. A5). Increasing the number of estimators from 100 to 500 did not improve the random forest model (SI Fig. A6). A midpoint of 300 estimators was found to be optimal with gradient boosting, but this only alters the F-score by around 0.01. Changes in other model components were found to have much greater effects than changes in architecture or hyperparameter tuning.

# 5 Discussion

## 5.1 Algorithmic bias in crop maps

We estimate that 17% of crop area in Paraguay is located on smallholder farms. While this only represents 1.5% of Paraguay's total land (3.6% of land in eastern Paraguay), landscapes dominated by smallholder farms smaller than 5 ha cover an important part of the landscape, particularly in eastern Paraguay. The cropland on these farms has largely gone unseen in prior global and regional crop maps. We estimate that, in missing 55-99% of smallholder fields, these prior data products overlook the croplands managed by 16-43% of the country's farming population. This bias poses important challenges for the effective use of crop maps for policy implementation, especially when policies seek to support or influence smallholder farming practices.

The tendency of crop maps to overlook smallholder cropping systems is likely to be a more general limitation underpinning existing global and regional crop maps outside of Paraguay. Previous studies have described the importance of tailoring training data collection and remote sensing methods in landscapes dominated by smallholder producers (Azzari et al., 2021; Bey et al., 2020; Thomas et al., 2020). These methods must overcome several challenges that are unique to mapping smallholder cropping systems including spectral heterogeneity, fine-scale spatial variation and edge effects, and an absence of prior high-quality training data (Lark et al., 2021; Jin et al., 2019). Since smallholder crops often represent a relatively small proportion of the total area of crops, regional and global land cover data products can achieve high aggregate crop accuracies without addressing these challenges. Even in the rare cases when regional and global crop mapping initiatives explicitly take on the challenge of mapping smallholder croplands, the resulting map accuracies for smallholder crops tend to be much lower than for industrial crops. For example, in their analysis of global oil palm plantations, Descals et al. (2024) successfully map both industrial and smallholder plantations, although the resulting recall differs dramatically across the two classes (91% for industrial plantations vs 71% for smallholders).

The systematic under-representation of smallholder croplands in remotely sensed crop maps represents a further example of data bias that could disadvantage smallholder producers (Bronson et al., 2021; Mayuravaani et al., 2024). For example, prior studies have raised concerns that emerging precision agriculture technologies may be developed in ways that advantage large-scale commercial producers (Rotz et al., 2019; Visser et al., 2021). In our case, the absence of smallholder fields in crop maps may influence critical policy decisions by, for example, affecting where extension services are targeted or underestimating the impact of weather shocks on crops that vulnerable populations depend upon. In turn, these misinformed decisions risk negatively affecting the billions of people who depend upon small farms for their nutrition or livelihoods. Given these risks, researchers creating crop maps should prioritize the development and use of methods that accurately represent all agricultural systems.

## 5.2 Methods for more accurate maps of smallholder crop systems

### 5.2.1 Training data

Our tests highlight the crucial role that training data sample design plays in creating land cover maps that accurately reflect smallholder cropping. Similar to other studies that find that the size of the training set is less important than its composition when classes are imbalanced (Weitkamp, Karimi, 2023; Waldner et al., 2019b), we find that the representation of the mixed crop class is the most influential component in mapping the minority smallholder class. We show that a pure random sample that includes smallholder crops at the same proportion as they are represented in the landscape can increase smallholder crop mapping by 20-40 percentage points compared to a training dataset whose observations of croplands are taken entirely from homogeneous crops grown in larger fields. In our study context, we show that this proportional random sample is still insufficient – the best-performing models are trained with datasets that over-sample smallholder crop observations by a factor of two to three with respect to a fully random sample. Importantly, smallholder crops often exhibit spectral and phenological characteristics that are more similar to non-crop classes such as grass (Neigh et al., 2018) or trees (Karlson et al., 2020) than to more homogeneous crops, as a general separability test demonstrates for our data (Fig. 8). In addition, smallholder crops exhibit a great deal of intra-class variability (Ibrahim et al., 2021; Rufin et al., 2022; Mohammed et al., 2020), due to high diversity within smallholder systems Bégué et al. (2018). Due to these characteristics, large samples of smallholder crops are needed to train accurate classifiers.

Although our results highlight the importance of *oversampling* smallholder observations, common methods for training data collection typically lead to dramatic *undersampling* of smallholder observations. Given this contrast, we posit that typical approaches to training data collection are the primary explanation for prior maps' inability to accurately represent smallholder crops. Reviewing the prior literature, we provide four primary explanations for why researchers tend to undersample smallholder crops.

First, smallholder crops are often explicitly or implicitly removed from training data since they are a heterogeneous, or "messy," class. Mixed-class and edge pixels are often excluded from training data by explicitly dropping such pixels (e.g. Stanimirova et al., 2023; Gumma et al., 2020; Thenkabail et al., 2021; Zhang, Roy, 2017), enforcing a larger minimum mapping unit in polygon data (e.g. Brown et al., 2022; Büttner, 2014), performing polygon erosion prior to sampling (e.g. Radoux et al., 2014; Leinenkugel et al., 2019) or allowing enumerators discretion in choosing crop samples without specific targeting or training regarding smallholder crops (e.g. Graesser et al., 2022). Because many smallholder crops would fall under the definition of a mixed class at 10 m resolution, and most would be considered mixed at 30 m resolution or lower, they will be excluded by such practices. This consideration of heterogeneity applies not only to the spatial but also to the temporal composition of smallholder crops. To stabilize samples for time-series mapping, samples may be excluded from training if they do not remain the same land cover for several years (Zhang et al., 2024). For most of their products, for example, MapBiomas requires training samples to be the same land cover for 30 of 35 years (MapBiomas, 2022). As smallholder fields are much more dynamic than large industrial fields, such fields will almost certainly be excluded by this rule. Rather than avoid mixed pixels, we find that it is important to actively include mixed pixels in training data to produce a truly representative map.

Even when mixed pixels are not actively avoided or removed, mixed crops are often implicitly left out of the training data due to quality control measures. To reduce errors of commission in crop training data, producers will often create rules that result in high rates of omission of smallholder crops. Quality control rules to reduce uncertainty in manually labeled data, such as dropping pixels where enumerators do not agree (e.g. Estes et al., 2022; Stanimirova et al., 2023) will likely bias samples toward homogeneous crops and away from messier smallholder crops. Likewise, using existing maps to control quality in training data, for example, with the use of WorldCover by the recent Global Land Cover Training dataset (GLANCE) to remove potential training points that do not agree (Stanimirova et al., 2023), will bias samples toward homogeneous crops as we have shown in Fig. 10.

A third reason smallholder crops are often excluded from training data is that they are more difficult or more costly to sample. Some products rely on opportunistic sampling to acquire the necessary quantity of training data. Crop data from external sources tend to overrepresent crops in industrial farms because these data are more readily available. For WorldCereal, for example, the nearest training samples to Paraguay come from agencies interested in commercial agriculture in Argentina (Boogaard et al., 2023). S.Cone is also trained largely with data from industrial crops (Graesser et al., 2022). Because collecting training data is costly, many suggest the potential for "reverse engineering of land cover maps" (Radoux et al., 2014), or using existing products as training data for future products without creating any new training data. For example, Cherif et al. (2022) propose using MapBiomas maps to produce automatic training data for the Amazon Basin. We have shown here why relying only on information from existing maps would certainly under-represent smallholder farms in downstream products.

When new training data are collected, producers often use existing products to reduce costs by limiting the sample space to areas more likely to generate informative data. In fact, it is generally accepted as best practice to use existing maps to exclude areas where no crop is expected to occur to reduce noise (Azzari et al., 2021) or inform stratified sampling. A producer can maximize the likelihood that crop samples are actually crop using "convergence of the evidence" (Jung et al., 2006), or "synergy mapping" (Fritz et al., 2003), to reduce the sampling space to the area of agreement of other maps (Hermosilla et al., 2022; Li et al., 2021; Zhang et al., 2022; Bratic et al., 2023). However, we show that such an approach will reduce the representation of smallholder crops and bias the training data and resulting product toward homogeneous crops, especially as more products are added for consideration (see Fig. 10). In general, agreement-based sampling reduces the variance of the training sample, a characteristic that generally makes the model less robust (He et al., 2024; Zhang et al., 2018). As an alternative, Fritz et al. (2024) suggest sampling only in "hotspots of disagreement" between maps to adjust those areas. This could increase variance, but would not result in the recall of much additional smallholder area in the case presented here, as we find that no individual map captures this area well. Indeed, the resulting map by Fritz et. al. (a hybrid of WorldCereal and GLAD) looks much like its input WorldCereal at 500 m resolution in Paraguay.

While we find that using existing discrete crop or land use products as masks for future products will result in an underestimation of smallholder cropland, we understand the need to reduce the sampling space for training efforts, especially for large-scale mapping. We suggest exploring other products or tools to direct sampling efforts towards the most informative samples. For example, we found that using the Geo-Wiki (Lesiv et al., 2019) field-size map would reduce the search area for smallholder crops by a factor of five in Paraguay. Another sampling approach to increase the representation of smallholder crops would be to sample in areas with higher heterogeneity, which can be located by computing homogeneity from a gray-level co-occurrence matrix (glcm) based on the input imagery. Preliminary observations suggest that results using a 25-pixel glcm correspond well to smallholder farming area (see Fig. 3) and would reduce the sampling space considerably. This approach has been used to enhance sampling for urban mapping (Kuffer et al., 2016; Zhang et al., 2014), as homogeneity metrics are generally lowest in urban areas. Sampling from locations that are slightly more homogeneous than urban landscapes could prove fruitful for locating smallholder areas, as illustrated in Fig 3. Combining a field size or heterogeneity filter with sampling methods such as ours that use Google Earth for rapid pixel interpretation can provide large gains to smallholdher mapping for modest costs, even for large-scale mapping efforts.

### 5.2.2 Modeling choices beyond training composition

While representation of smallholder crops in the training data has the most effect on mapping smallholder crops of the modeling components we tested, other components such as contextual information (from segmentation), resolution, and class complexity are important.

**Pixel-based vs. object-based classification:** We find that segmentation improves the pixel-based model, but object-based classification alone would not adequately capture smallholder agriculture; 38% of smallholder crops ¡ 5 ha (73% ¡ 1 ha) were not captured with our segmentation methods, even though smallholder crops were included in the training data for the segmentation model. This is partially due to the minimum mapping requirement of nine pixels (an interior pixel and border pixels) for our segmentation methods, but this is less than 0.1 ha with the Sentinel imagery we used, and smaller than most fields we observed. We think that the missed segmentation is more often due to the frequent configurational shifting of many smallholder fields throughout the year, making them difficult to segment in an annual product. Although we were unable to represent these fields for an object-based analysis, the process did create informative features for these smallholder areas that improved their classification in the pixel-based model. This improvement is likely due to the higher-quality training data and the deep learning process involved in generating these features. Other approaches of incorporating textural attributes into pixel-based features, such as those involving glcms or other landscape metrics, have been useful for other crop mapping efforts (Chen et al., 2024; Dimitrov et al., 2024) and warrant more exploration as an alternative to our intensive segmentation methods to enhance smallholder mapping.

**Resolution:** We find that an enhanced resolution from 30 m to 10 m does result in better representation of smallholder cropland by increasing the smallholder recall by about 10% and increasing no-crop recall (decreasing errors of commission for crops) by about 4%. By recording the neighborhood heterogeneity (Sweeney, Evans, 2012), or the number of surrounding pixels that would receive the same classification, when labeling sample data, we were able to estimate the extra omission error that would be expected when downsampling from 10 m to 30 m resolution. We estimated this extra omission to be 12%, which is in line with the difference we see between the 30 m and 10 m model results. Based on a rough Pareto boundary test (Waldner, 2017; Boschetti et al., 2004) using crop perimeters from our segmentation training data, we do not expect that using 3 m resolution data (e.g. PlanetScope) would improve the model significantly for the smallholder systems in our study area. Higher resolutions might be beneficial, especially if a study is interested in capturing fields smaller than 0.1 ha, but appropriate training data is likely still the most important component.

**Class complexity:** We find that a random forest classification model can perform well with a fairly complex classification schema comprising 32 classes. Our 32-class model performed only slightly better than a binary crop/no_crop model overall, but it is important to note that all 32 classes were still represented in the training sample for the crop/no_crop model. A binary model that does not represent edge cases where confusion most often occurs will likely not perform as well Foody, Mathur (2006). While some authors advocate for first separating crop from non-crop before performing more nuanced classifications, (Lebourgeois et al., 2017), we found that such hierarchical classification would reduce the recall for mixed crops by around 20%, likely because they resemble non-crop more than they resemble other homogeneous crops. Overall, we find that the schematic model can help ensure collection of high-quality training and calibration data, but that the final set of classes used for classification is less important (Waldner et al., 2019b; Rebbapragada et al., 2008).

**Model architecture and hyperparameter tuning:** While we found random forest to perform slightly better than gradient boosting for almost all versions of our model, we find that the choice of architecture to be fairly unimportant compared to the other modeling components we tested. Likewise, the number of estimators or trees used to build the model made little difference to the output. While some cases may warrant this fine tuned optimization, our work supports criticisms by other authors (e.g. Waldner et al., 2019b; Weitkamp, Karimi, 2023) that the importance of modeling architecture and hyperparameter tuning seems inflated in the literature compared to potentially more important modeling choices such as training data composition and feature selection.

# 6 Conclusion

With the rapid advancement of machine learning methods, it is increasingly possible to generate detailed large-scale crop maps to support decision-making. For example, crop maps increasingly serve as important inputs for the targeting of agricultural support, early warning systems for food insecurity, and the identification of agriculture's environmental impacts. However, as the influence of machine learning on decision making proliferates in many fields, so do warnings that issues of inequality and unintentional algorithmic bias (Pessach, Shmueli, 2023) can lead to decisions that further marginalize already marginalized populations. A common source of bias in machine learning is the underrepresentation of marginalized groups compared to privileged groups in training data (Pessach, Shmueli, 2021). We have shown here how poor representation of smallholder farms in training data leads to crop maps that miss large areas of smallholder farming. This bias is intensified by the increasing trend to use existing land cover maps for subsequent analyses, which can create substantial downstream errors. If used to make decisions on agricultural support or climate adaptation, crop maps that miss a large portion of smallholder farms risk further disadvantaging these marginalized communities. As with other domains where machine learning is becoming prominent in decision-making, the current trend of large-scale, data-intense mapping that makes data more available to all may unintentionally hide important information from policymakers and practitioners about underserved communities whose lives they intend to transform (Gaikwad et al., 2022). By integrating local data in large-scale mapping efforts, machine learning can better support marginalized populations such as smallholder farmers, rather than further disadvantaging them (You, Sun, 2022). Here we show how an active learning approach, with sensitivity to local landscapes and systems, can be used to efficiently integrate local data to better map smallholder farming.

**Acknowledgements** The staff at Paraguay's Ministerio de Agricultura y Ganadería generously provided access to district-level agricultural census data and industrial crop points. Jordan Graesser set up the time series processing pipeline and provided much of the foundational code. Keyla Morales and Conner Edwards helped create training and validation data, Ryan Ashraf helped with model building and optimization, and Camila Berger helped to collect field data. The GRIT team at UCSB provided technical assistance to develop and maintain our pipeline on Earth Research Institute's (ERI) cluster computing network. This work was supported by the National Aeronautics and Space Administration under Grant No. 80NSSC20K1489 issued through the Land Cover and Land Use Change Program.

**Author contributions: CRediT KW**: Conceptualization, Methodology, Investigation, Formal analysis, Data Curation, Software, Supervision, Writing - Original Draft **AA**: Investigation, Data Curation, Resources **LS**: Methodology, Data Curation **RH**: Conceptualization, Funding acquisition, Resources, Supervision, Writing - Original Draft

**Funding Sources** This work was supported by the National Aeronautics and Space Administration under Grant No. 80NSSC20K1489 issued through the Land Cover and Land Use Change Program.

# References

*Avila Schmalko Claudia, Monroy Sarta Ariym.* Mapeando el Agronegocio en Paraguay. Asuncion, Paraguay: Base Investiagciones Sociales, 2018.

*Azzari George, Jain Shruti, Jeffries Graham, Kilic Talip, Murray Siobhan.* Understanding the Requirements for Surveys to Support Satellite-Based Crop Type Mapping: Evidence from Sub-Saharan Africa // Remote Sensing. 2021.

*Bey Adia, Jetimane Julieta, Lisboa Sá Nogueira, Ribeiro Natasha, Sitoe Almeida, Meyfroidt Patrick.* Mapping smallholder and large-scale cropland dynamics with a flexible classification system and pixel-based composites in an emerging frontier of Mozambique // Remote Sensing of Environment. 2020. 239. 111611.

*Boogaard H., Pratihast A. K., Bayas J. C. L., Karanam S., Fritz S., Van Tricht K., Degerickx J., Gilliams S.* Building a community-based open harmonised reference data repository for global crop mapping // Plos One. 2023. 18, 7.

*Boschetti L., Brivio P. A., Flasse S., ieee .* Pareto Boundary: a useful tool in the accuracy assessment of low spatial resolution thematic products //

IEEE International Geoscience and Remote Sensing Symposium. 2004. 3759–3762. (IEEE International Symposium on Geoscience and Remote Sensing IGARSS).

*Bratic G., Oxoli D., Brovelli M. A.* Map of Land Cover Agreement: Ensambling Existing Datasets for Large-Scale Training Data Provision // Remote Sensing. 2023. 15, 15.

*Bronson Kelly, Rotz Sarah, D'Alessandro Adrian.* The human impact of data bias and the digital agricultural revolution // Handbook on the human impact of agriculture. 2021. 119–137.

*Brown Christopher F., Brumby Steven P., Guzder-Williams Brookie, Birch Tanya, Hyde Samantha Brooks, Mazzariello Joseph, Czerwinski Wanda, Pasquarella Valerie J., Haertel Robert, Ilyushchenko Simon, Schwehr Kurt, Weisse Mikaela, Stolle Fred, Hanson Craig, Guinan Oliver, Moore Rebecca, Tait Alexander M.* Dynamic World, Near real-time global 10m land use land cover mapping // Scientific Data. 2022. 9, 1. 251.

*Bullock E. L., Healey S. P., Yang Z. Q., Oduor P., Gorelick N., Omondi S., Ouko E., Cohen W. B.* Three Decades of Land Cover Change in East Africa // Land. 2021. 10, 2.

*Büttner György.* CORINE land cover and land cover change products // Land use and land cover mapping in Europe: practices & trends. 2014. 55–74.

*Bégué A., Arvor D., Bellon B., Betbeder J., Abelleyra D. de, Ferraz R. P. D., Lebourgeois V., Lelong C., Simoes M., Verón S. R.* Remote Sensing and Cropping Practices: A Review // Remote Sensing. 2018. 10, 1.

*Chawla N. V.* Data Mining for Imbalanced Datasets: An Overview. 2010. 875–886. (Data Mining and Knowledge Discovery Handbook, Second Edition).

*Chen Chen, Dong Taifeng, Wang Zhaohai, Wang Chen, Song Wenyao, Zhang Huanxue.* Exploring optimal features and image analysis methods for crop type classification from the perspective of crop landscape heterogeneity // Remote Sensing Applications: Society and Environment. 2024. 36. 101308.

*Cherif E., Hell M., Brandmeier M.* DeepForest: Novel Deep Learning Models for Land Use and Land Cover Classification Using Multi-Temporal and -Modal Sentinel Data of the Amazon Basin // Remote Sensing. 2022. 14, 19.

*Claverie Martin, Ju Junchang, Masek Jeffrey G., Dungan Jennifer L., Vermote Eric F., Roger Jean-Claude, Skakun Sergii V., Justice Christopher O.* The Harmonized Landsat and Sentinel-2 surface reflectance data set // Remote Sensing of Environment. 2018. 219. 145–161.

*Descals A., Gaveau D. L. A., Wich S., Szantoi Z., Meijaard E.* Global mapping of oil palm planting year from 1990 to 2021 // Earth System Science Data. 2024. 16, 11. 5111–5129.

*Dimitrov Petar, Roumenina Eugenia, Ganeva Dessislava, Gikov Alexander, Kamenova Ilina, Bozhanova Violeta.* Enhancing Pléiades-based crop mapping with multi-temporal and texture information // Remote Sensing Applications: Society and Environment. 2024.

*Estes L. D., Ye S., Song L., Luo B. K., Eastman J. R., Meng Z. H., Zhang Q., McRitchie D., Debats S. R., Muhando J., Amukoa A. H., Kaloo B. W., Makuru J., Mbatia B. K., Muasa I. M., Mucha J., Mugami A. M., Mugami J. M., Muinde F. W., Mwawaza F. M., Ochieng J., Oduol C. J., Oduor P., Wanjiku T., Wanyoike J. G., Avery R. B., Caylor K. K.* High Resolution, Annual Maps of Field Boundaries for Smallholder-Dominated Croplands at National Scales // Frontiers in Artificial Intelligence. 2022. 4.

*Foody G. M.* Challenges in the real world use of classification accuracy metrics: From recall and precision to the Matthews correlation coefficient // Plos One. 2023. 18, 10.

*Foody G. M., Mathur A.* The use of small training sets containing mixed pixels for accurate hard image classification: Training on mixed spectral responses for classification by a SVM // Remote Sensing of Environment. 2006. 103, 2. 179–189.

*Foody G. M., McCulloch M. B., Yates W. B.* The effect of training set size and composition on artificial neural-network classification // International Journal of Remote Sensing. 1995. 16, 9. 1707–1723.

*Fritz S., Lesiv M., See L., Schepaschenko D., Guzmán K. P., Bayas J. C. L., Schepaschenko M., Georgieva I., Collivignarelli F., Meroni M., Kerdiles H., Rembold F., McCallum I.* A new global hybrid map of annual herbaceous cropland at a 500 m resolution for the year 2019 // Environmental Research Letters. 2024. 19, 8.

Harmonisation, mosaicing and production of the Global Land Cover 2000 database (Beta Version). // . 2003.

*Gaikwad S., Iyer S., Lunga D., Yabe T., Liang X. F., Ananthabhotla B., Behari N., Guggilam S., Chi G. H., Acm* . Data-driven Humanitarian Mapping and Policymaking: Toward Planetary-Scale Resilience, Equity, and Sustainability // 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KKD). 2022. 4872–4873.

*Gao Z. T., Guo D. L., Ryu D., Western A. W.* Training sample selection for robust multi-year within-season crop classification using machine learning // Computers and Electronics in Agriculture. 2023. 210.

*Garcia-Lopez G. A., Arizpe N.* Participatory processes in the soy conflicts in Paraguay and Argentina // Ecological Economics. 2010. 70, 2. 196–206.

*Graesser J., Stanimirova R., Tarrio K., Copati E. J., Volante J. N., Veron S. R., Banchero S., Elena H., Abelleyra D. de, Friedl M. A.* Temporally-Consistent Annual Land Cover from Landsat Time Series in the Southern Cone of South America // Remote Sensing. 2022. 14, 16.

*Graesser Jordan.* Cultionet. 2023. Accessed: 2023-11-20.

*Grossman J. J.* Ecosystem service trade-offs and land use among smallholder farmers in eastern Paraguay // Ecology and Society. 2015. 20, 1. Grossman, Jake J. Grossman, Jake/0000-0001-6468-8551.

*Guesne S. J. J., Hanser T., Werner S., Boobier S., Scott S.* Mind your prevalence! // Journal of Cheminformatics. 2024. 16, 1.

*Gumma M. K., Thenkabail P. S., Teluguntla P. G., Oliphant A., Xiong J., Giri C., Pyla V., Dixit S., Whitbread A. M.* Agricultural cropland extent and areas of South Asia derived using Landsat satellite 30-m time-series big-data using random forest machine learning algorithms on the Google Earth Engine cloud // Giscience Remote Sensing. 2020. 57, 3. 302–322.

*He Wei, Li Lianfa, Gao Xilin.* Geocomplexity Statistical Indicator to Enhance Multiclass Semantic Segmentation of Remotely Sensed Data with Less Sampling Bias // Remote Sensing. 2024. 16, 11.

*Hermosilla Txomin, Wulder Michael A., White Joanne C., Coops Nicholas C.* Land cover classification in an era of big and open data: Optimizing localized implementation and training data selection to improve mapping outcomes // Remote Sensing of Environment. 2022. 268. 112780.

*Ibrahim E. S., Rufin P., Nill L., Kamali B., Nendel C., Hostert P.* Mapping Crop Types and Cropping Systems in Nigeria with Sentinel-2 Imagery // Remote Sensing. 2021. 13, 17.

*Jin Zhenong, Azzari George, You Calum, Di Tommaso Stefania, Aston Stephen, Burke Marshall, Lobell David B.* Smallholder maize area and yield mapping at national scales with Google Earth Engine // Remote Sensing of Environment. 2019. 228. 115–128.

*Jung Martin, Henkel Kathrin, Herold Martin, Churkina Galina.* Exploiting synergies of global land cover products for carbon cycle modeling // Remote Sensing of Environment. 2006. 101. 534–553.

*Karlson M., Ostwald M., Bayala J., Bazié H. R., Ouedraogo A. S., Soro B., Sanou J., Reese H.* The Potential of Sentinel-2 for Crop Production Estimation in a Smallholder Agroforestry Landscape, Burkina Faso // Frontiers in Environmental Science. 2020. 8.

*Karra Krishna, Kontgis Caitlin, Statman-Weil Zoe, Mazzariello Joseph C, Mathis Mark, Brumby Steven P.* Global land use/land cover with Sentinel 2 and deep learning // 2021 IEEE international geoscience and remote sensing symposium IGARSS. 2021. 4704–4707.

*Kuffer Monika, Pfeffer Karin, Sliuzas Richard, Baud Isa.* Extraction of Slum Areas From VHR Imagery Using GLCM Variance // Ieee Journal of Selected Topics in Applied Earth Observations and Remote Sensing. 2016. 9, 5. 1830–1840.

*Lark Tyler J., Schelly Ian H., Gibbs Holly K.* Accuracy, Bias, and Improvements in Mapping Crops and Cropland across the United States Using the USDA Cropland Data Layer // Remote Sensing. 2021. 13, 5. 968.

*Lebourgeois V., Dupuy S., Vintrou E., Ameline M., Butler S., Begue A.* A Combined Random Forest and OBIA Classification Scheme for Mapping Smallholder Agriculture at Different Nomenclature Levels Using Multisource Data (Simulated Sentinel-2 Time Series, VHRS and DEM) // Remote Sensing. 2017. 9, 3.

*Leinenkugel P., Deck R., Huth J., Ottinger M., Mack B.* The Potential of Open Geodata for Automated Large-Scale Land Use and Land Cover Classification // Remote Sensing. 2019. 11, 19.

*Lesiv M., Bayas J. C. L., See L., Duerauer M., Dahlia D., Durando N., Hazarika R., Sahariah P. K., Vakolyuk M., Blyshchyk V., Bilous A., Perez-Hoyos A., Gengler S., Prestele R., Bilous S., Akhtar I. U., Singha K., Choudhury S. B., Chetri T., Malek Z., Bungnamei K., Saikia A., Sahariah D., Narzary W., Danylo O., Sturn T., Karner M., McCallum I., Schepaschenko D., Moltchanova E., Fraisl D., Moorthy I., Fritz S.* Estimating the global distribution of field size using crowdsourcing // Global Change Biology. 2019. 25, 1. 174–186.

*Li Congcong, Xian George, Zhou Qiang, Pengra Bruce W.* A novel automatic phenology learning (APL) method of training sample selection using multiple datasets for time-series land cover mapping // Remote Sensing of Environment. 2021. 266. 112670.

*Liu Y., Rao P., Zhou W. Q., Singh B., Srivastava A. K., Poonia S. P., Van Berkel D., Jain M.* Using Sentinel-1, Sentinel-2, and Planet satellite data to map field-level tillage practices in smallholder systems // Plos One. 2022. 17, 11.

*Lowder S. K., Sanchez M. V., Bertini R.* Which farms feed the world and has farmland become more concentrated? // World Development. 2021. 142.

*Luque A., Carrasco A., Martín A., Heras A. de las.* The impact of class imbalance in classification performance metrics based on the binary confusion matrix // Pattern Recognition. 2019. 91. 216–231.

*MAG Dirección de Censos y Estadísticas Agropecuarias Ministerio de Agricultura y Ganadería.* Censo Agropecuario Nacional 2022. 2023. Public depertment-level data accessed at: https://www.datos.gov.py/dataset/censo-agropecuario-nacional-can-2022. District-level data accessed by invitation at:

https://geoportal.mag.gov.py on 22-Nov-2024. Confidentiality of these data is dictated by Ley N.º 6670/2020 de la República del Paraguay.

*MapBiomas* . MapBiomas Trinational Atlantic Forest Project – Collection 2, Version 1 of the Annual Coverage and Land Use Series. 2022. Accessed at: https://gee-community-catalog.org/projects/mapbiomas.

*Mayuravaani Mathuranathan, Ramanan Amirthalingam, Perera Maneesha, Senanayake Damith Asanka, Vidanaarachchi Rajith.* Insights into Artificial Intelligence Bias: Implications for Agriculture // Digital Society. 2024. 3, 3. 48.

*McCarty J. L., Neigh C. S. R., Carroll M. L., Wooten M. R.* Extracting smallholder cropped area in Tigray, Ethiopia with wall-to-wall sub-meter WorldView and moderate resolution Landsat 8 imagery // Remote Sensing of Environment. 2017. 202. 142–151.

*Mohammed Issamaldin, Marshall Michael, Bie Kees de, Estes Lyndon, Nelson Andy.* A blended census and multiscale remote sensing approach to probabilistic cropland mapping in complex landscapes // ISPRS Journal of Photogrammetry and Remote Sensing. 2020. 161. 233–245.

*Neigh Christopher S. R., Carroll Mark L., Wooten Margaret R., McCarty Jessica L., Powell Bristol F., Husak Gregory J., Enenkel Markus, Hain Christopher R.* Smallholder crop area mapped with wall-to-wall WorldView submeter panchromatic image texture: A test case for Tigray, Ethiopia // Remote Sensing of Environment. 2018. 212. 8–20.

*Ortega Guillermo.* Mapeamiento del Extractivismo. Asunción, Paraguay: Base Investigaciones Sociales, 2016.

*Patel U., Patel V.* A comprehensive review: active learning for hyperspectral image classifications // Earth Science Informatics. 2023.

*Pessach D., Shmueli E.* Improving fairness of artificial intelligence algorithms in Privileged-Group Selection Bias data settings // Expert Systems with Applications. 2021. 185.

*Pessach D., Shmueli E.* A Review on Fairness in Machine Learning // Acm Computing Surveys. 2023. 55, 3.

*Pontius Robert Gilmore Jr., Millones Marco.* Death to Kappa: birth of quantity disagreement and allocation disagreement for accuracy assessment // International Journal of Remote Sensing. 2011. 32, 15. 4407–4429.

*Potapov Peter V., Turubanova Svetlana, Hansen Matthew C., Tyukavina Alexandra, Zalles Viviana, Khan Ahmad, Song Xiao-Peng, Pickens Amy, Shen Quan, Cortez Jocelyn.* Global maps of cropland extent and change show accelerated cropland expansion in the twenty-first century // Nature Food. 2022. 3, 1. 19–+.

*Radoux Julien, Lamarche Céline, Van Bogaert Eric, Bontemps Sophie, Brockmann Carsten, Defourny Pierre.* Automated training sample extraction for global land cover mapping // Remote Sensing. 2014. 6, 5. 3965–3987.

*Rao Preeti, Zhou Weiqi, Bhattarai Nishan, Srivastava Amit K., Singh Balwinder, Poonia Shishpal, Lobell David B., Jain Meha.* Using Sentinel-1, Sentinel-2, and Planet Imagery to Map Crop Type of Smallholder Farms // Remote Sensing. 2021. 13, 10.

*Rebbapragada Umaa, Lomasky Rachel, Brodley Carla E, Friedl Mark A.* Generating high-quality training data for automated land-cover mapping // IGARSS 2008-2008 IEEE International Geoscience and Remote Sensing Symposium. 4. 2008. IV–546–IV–548.

*Ricciardi V., Ramankutty N., Mehrabi L. Z. andJarvis, Chookolingo B.* How much of the world's food do smallholders produce? // Global Food Security. 2018. 17. 64–72.

*Rotz S., Gravely E., Mosby I., Duncan E., Finnis E., Horgan M., LeBlanc J., Martin R., Neufeld H. T., Nixon A., Pant L., Shalla V., Fraser E.* Automated pastures and the digital divide: How agricultural technologies are shaping labour and rural communities // Journal of Rural Studies. 2019. 68. 112–122.

*Rufin Philippe, Bey Adia, Picoli Michelle, Meyfroidt Patrick.* Large-area mapping of active cropland and short-term fallows in smallholder landscapes using PlanetScope data // International Journal of Applied Earth Observation and Geoinformation. 2022. 112. 102937.

*Samberg L. H., Gerber J. S., Ramankutty N., Herrero M., West P. C.* Subnational distribution of average farm size and smallholder contributions to global food production // Environmental Research Letters. 2016. 11, 12.

*See L. D., Fritz S., You L. Z., Ramankutty N., Herrero M., Justice C., Becker-Reshef I., Thornton P., Erb K., Gong P., Tang H. J., Velde M. van der, Ericksen P., McCallum I., Kraxner F., Obersteiner M.* Improved global cropland data as an essential ingredient for food security // Global Food Security-Agriculture Policy Economics and Environment. 2015. 4. 37–45.

*Soldi A., Meza M. J. A., Guareschi M., Donati M., Ortiz A. I.* Sustainability Assessment of Agricultural Systems in Paraguay: A Comparative Study Using FAO's SAFA Framework // Sustainability. 2019. 11, 13.

*Song W. Y., Wang C., Dong T. F., Wang Z. H., Wang C. X., Mu X. D., Zhang H. X.* Hierarchical extraction of cropland boundaries using Sentinel-2 time-series data in fragmented agricultural landscapes // Computers and Electronics in Agriculture. 2023. 212.

*Song Xiao-Peng, Hansen Matthew C., Potapov Peter V., Adusei Bernard, Pickering Jeffrey, Adami Marcos, Lima Andre, Zalles Viviana, Stehman*

*Stephen V., Di Bella Carlos M., Conde Maria C., Copati Esteban J., Fernandes Lucas B., Hernandez-Serna Andres, Jantz Samuel M., Pickens Amy H., Turubanova Svetlana, Tyukavina Alexandra.* Massive soybean expansion in South America since 2000 and implications for conservation // Nature Sustainability. 2021. 4, 9. 784–792.

*Stanimirova R., Tarrio K., Turlej K., McAvoy K., Stonebrook S., Hu K. T., Arévalo P., Bullock E. L., Zhang Y. T., Woodcock C. E., Olofsson P., Zhu Z., Barber C. P., Souza Jr. C. M., Chen S. J., Wang J. A., Mensah F., Calderón-Loor M., Hadjikakou M., Bryan B. A., Graesser J., Beyene D. L., Mutasha B., Siame S., Siampale A., Friedl M. A.* A global land cover training dataset from 1984 to 2020 // Scientific Data. 2023. 10, 1.

*Stehman S. V., Foody G. M.* Key issues in rigorous accuracy assessment of land cover products // Remote Sensing of Environment. 2019. 231.

*Sweeney S. P., Evans T. P.* An edge-oriented approach to thematic map error assessment // Geocarto International. 2012. 27, 1. 31–56.

*Teluguntla P., Thenkabail P., Oliphant A., Gumma M., Aneece I., Foley D., McCormick R.* Landsat-Derived Global Rainfed and Irrigated-Cropland Product 30 m V001. 2023. accessed at: https://doi.org/10.5067/Community/LGRIP/LGRIP30.001.

*Thenkabail P. S., Teluguntla P. G., Xiong J., Oliphant A., Congalton R. G., Ozdogan M., Gumma M. K., Tilton J. C., Giri C., Milesi C., Phalke A., Massey R., Yadav K., Sankey T., Zhong Y., Aneece I., Foley D.* Global Cropland-Extent Product at 30-m Resolution (GCEP30) Derived from Landsat Satellite Time-Series Data for the Year 2015 Using Multiple Machine-Learning Algorithms on Google Earth Engine Cloud. 2021.

*Thomas N., Neigh C. S. R., Carroll M. L., McCarty J. L., Bunting P.* Fusion Approach for Remotely-Sensed Mapping of Agriculture (FARMA): A Scalable Open Source Method for Land Cover Monitoring Using Data Fusion // Remote Sensing. 2020. 12, 20.

*Tuia D., Ratle F., Pacifici F., Kanevski M. F., Emery W. J.* Active Learning Methods for Remote Sensing Image Classification // Geoscience and Remote Sensing, IEEE Transactions on. 2009. 47, 7. 2218–2232.

*Van Tricht K., Degerickx J., Gilliams S., Zanaga D., Battude M., Grosu A., Brombacher J., Lesiv M., Bayas J. C. L., Karanam S., Fritz S., Becker-Reshef I., Franch B., Mollà-Bononad B., Boogaard H., Pratihast A. K., Koetz B., Szantoi Z.* WorldCereal: a dynamic open-source system for global-scale, seasonal, and reproducible crop and irrigation mapping // Earth System Science Data. 2023a. 15, 12. 5491–5515.

*Van Tricht Kristof, Degerickx Jeroen, Gilliams Sven, Zanaga Daniele, Battude Marjorie, Grosu Alex, Brombacher Joost, Lesiv Myroslava, Bayas Juan Carlos Laso, Karanam Santosh, others* . WorldCereal: a dynamic open-source system for global-scale, seasonal, and reproducible crop and irrigation mapping // Earth System Science Data. 2023b. 15, 12. 5491–5515.

*Visser O., Sippel S. R., Thiemann L.* Imprecision farming? Examining the (in)accuracy and risks of digital agriculture // Journal of Rural Studies. 2021. 86. 623–632.

*Waldner François, Diakogiannis Foivos I.* Deep learning on edge: Extracting field boundaries from satellite images with a convolutional neural network // Remote Sensing of Environment. 2020. 245. 111741.

*Waldner Nicolas FrançoisBellemans, Hochman Zvi, Newby Terence, Abelleyra Diego de, Verón Santiago R., Bartalev Sergey, Lavreniuk Mykola, Kussul Nataliia, Maire Guerric Le, Simoes Margareth, Skakun Sergii, Defourny Pierre.* Roadside collection of training data for cropland mapping is viable when environmental and management gradients are surveyed // International Journal of Applied Earth Observation and Geoinformation. 2019a. 80. 82–93.

*Waldner Pierre FrançoisDefourny.* Where can pixel counting area estimates meet user-defined accuracy requirements? // International Journal of Applied Earth Observation and Geoinformation. 2017. 60. 1–10.

*Waldner Yang FrançoisChen, Lawes Roger, Hochman Zvi.* Needle in a haystack: Mapping rare and infrequent crops using satellite imagery and data balancing methods // Remote Sensing of Environment. 2019b. 233. 111375.

*Walker K., Sharwood L., Ayala A., Graesser J., Heilmayr R.* CELPy: Conservation and Economics Lab Paraguay landcover maps 2017-2024 // in prep. 2025. to be published prior to acceptance of this article, as preprint ArXiv document if not yet formally published. Current draft is available at: https://klwalker-sb.github.io/klwalker-sb/files/py$_{m}apping_{d}ata_{p}aper.pdf$.

*Wang S., Waldner F., Lobell D. B.* Unlocking Large-Scale Crop Field Delineation in Smallholder Farming Systems with Transfer Learning and Weak Supervision // Remote Sensing. 2022. 14, 22.

*Weitkamp T., Karimi P.* Evaluating the Effect of Training Data Size and Composition on the Accuracy of Smallholder Irrigated Agriculture Mapping in Mozambique Using Remote Sensing and Machine Learning Algorithms // Remote Sensing. 2023. 15, 12.

*Wesz V. J.* Soybean production in Paraguay: Agribusiness, economic change and agrarian transformations // Journal of Agrarian Change. 2022. 22, 2. 317–340.

*Woodhill Jim, Kishore Avinash, Njuki Jemimah, Jones Kristal, Hasnain Saher.* Food systems and rural wellbeing: challenges and opportunities // Food Security. 2022. 14. 1099–1121.

*Xu P. P., Tsendbazar N. E., Herold M., Bruin S. de, Koopmans M., Birch T., Carter S., Fritz S., Lesiv M., Mazur E., Pickens A., Potapov P., Stolle F., Tyukavina A., Kerchove R. van de, Zanaga D.* Comparative validation of recent 10 m-resolution global land cover maps // Remote Sensing of Environment. 2024. 311.

*You L. Z., Sun Z. L.* Mapping global cropping system: Challenges, opportunities, and future perspectives // Crop and Environment. 2022. 1, 1. 68–73.

*Zanaga D., Van De Kerchove R., Daems D., De Keersmaecker W., Brockmann C., Kirches G., Wevers J., Cartus O., Santoro M., Fritz S., Lesiv M., Herold M., Tsendbazar N.-E., Xu P., Ramoino F., Arino O.* ESA WorldCover 10 m 2021 v200. 2022. Accessed at: https://doi.org/10.5281/zenodo.7254221.

*Zeng Z. Z., Estes L., Ziegler A. D., Chen A. P., Searchinger T., Hua F. Y., Guan K. Y., Jintrawet A., Wood E. F.* Highland cropland expansion and forest loss in Southeast Asia in the twenty-first century // Nature Geoscience. 2018. 11, 8. 556–+.

*Zhang Hankui K., Roy David P.* Using the 500m MODIS land cover product to derive a consistent continental scale 30m Landsat land cover classification // Remote Sensing of Environment. 2017. 197. 15–34.

*Zhang J. X., Yang W. J., Zhang W. L., Wang Y., Liu D., Xiu Y. C.* An Explorative Study on Estimating Local Accuracies in Land-Cover Information Using Logistic Regression and Class-Heterogeneity-Stratified Data // Remote Sensing. 2018. 10, 10.

*Zhang Jun, Li Peijun, Wang Jinfei.* Urban Built-Up Area Extraction from Landsat TM/ETM plus Images Using Spectral Information and Multivariate Texture // Remote Sensing. 2014. 6, 8. 7339–7359.

*Zhang X., Zhao T. T., Xu H., Liu W. D., Wang J. Q., Chen X. D., Liu L. Y.* GLC FCS30D: the first global 30 m land-cover dynamics monitoring product with a fine classification system for the period from 1985 to 2022 generated using dense-time-series Landsat imagery and the continuous change-detection method // Earth System Science Data. 2024. 16, 3. 1353–1381.

*Zhang Z., Tang P., Hu C. M., Liu Z. Q., Zhang W. X., Tang L.* Seeded Classification of Satellite Image Time Series with Lower-Bounded Dynamic Time Warping // Remote Sensing. 2022. 14, 12.

# Supplementary material

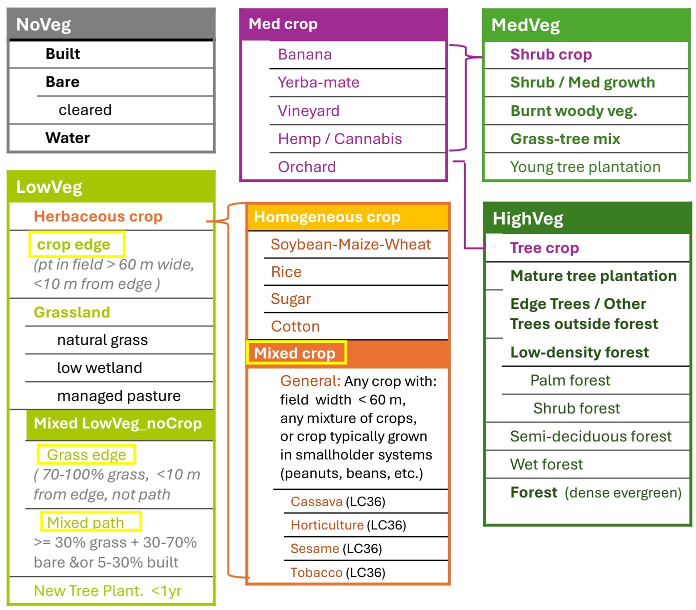
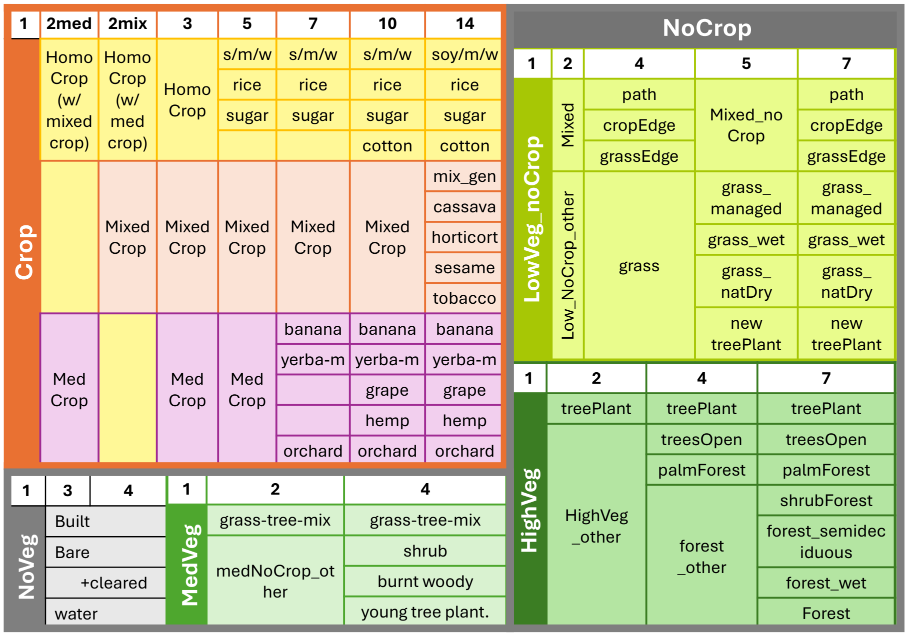



Figure A1: CELPy classes for full model. Orange and purple labels are within crop class and other labels are within non-crop class in binary model. Labels with white text are the eight classes in LC8. Yellow rectangles highlight the mixed classes that are increased proportionately for mixed-class representation tests




**Class Schema (LC_X)**

| | 2 | 3mix | 3med | 4 | 7 | 8 | 10 | 12 | 14 | 15 | 17 | 18b | 18 | 21 | 23 | 25 | 32 | 36 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Crop | 1 | 2mix | 2med | 3 | 3 | 3 | 5 | 5 | 5 | 7 | 5 | 7 | 7 | 7 | 7 | 7 | 10 | 14 |
| NoVeg | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 4 | 4 | 4 |
| LowVeg_noCrop | | | | | 1 | 2 | 2 | 2 | 4 | 2 | 7 | 4 | 2 | 4 | 7 | 7 | 7 | 7 |
| MedVeg_noCrop | | | | | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 4 | 4 |
| HighVeg | | | | | 1 | 1 | 1 | 1 | 1 | 2 | 1 | 2 | 4 | 4 | 4 | 4 | 7 | 7 |

Figure A2: Hierarchal breakdown of classes with schematic complexity chart

Table A1: Confusion Matrix for combined district sample

| | homo crop | mixed crop | no-crop | UA | UAcnc[a] | PA | PAcnc[a] |
|---|---|---|---|---|---|---|---|
| homo crop | 851 | 60 | 36 | 90.0% | 96.2% | 92.4% | 93.4% |
| mixed crop | 9 | 265 | 91 | 72.6% | 75.0% | 65.4% | 80.0% |
| no-crop | 61 | 80 | 3794 | 96.4% | 96.4% | 96.8% | 96.8% |

[a] cnc = crop/no-crop. Mixed crop and homo crop are grouped.

Table A2: Adjusted Area Estimates for homo crop, mixed crop and no crop based on combined district sample

| | homo crop | mixed crop | no-crop |
|---|---|---|---|
| p. homo crop | 17.78% | 0.53% | 0.70% |
| p. mixed crop | 0.19% | 2.36% | 1.78% |
| p. no crop | 1.28% | 0.71% | 74.21% |
| Map area | 19.79% | 3.25% | 76.96% |
| **Adjusted area** | **19.24%** | **3.61%** | **76.69%** |
| 95% CI | 0.51% | 0.24% | 0.62% |

Table A3: Confusion Matrix for combined district sample - by field size tier

| | | | | |
|---|---|---|---|---|
| **< 2ha** | | | | |
| | crop | no-crop | UA | PA |
| crop | 288 | 86 | 77% | 78% |
| no-crop | 83 | 1786 | 96% | 95% |
| **2-5 ha** | | | | |
| | crop | no-crop | UA | PA |
| crop | 56 | 8 | 88% | 69% |
| no-crop | 25 | 1653 | 98.5% | 99.5% |
| **> 5 ha** | | | | |
| | crop | no-crop | UA | PA |
| crop | 835 | 6 | 99.3% | 98% |
| no-crop | 33 | 355 | 92% | 98% |

Table A4: Adjusted crop area estimates by average field size

| field size[a] | %crop | %no-crop | %tier[b] | area estimate | % of cropland |
|---|---|---|---|---|---|
| $<$ 2 ha | 13.5 ±0.2% | 85.5±0.3% | 16.8% | 2.27 ±0.03% | 10.2% |
| 2-5 ha | 32.6±3.6% | 72.1.5±0.06% | 4.4% | 1.44 ±0.2% | 6.5% |
| $>$ 5ha | 66.9 ±0.2% | 32.4±0.3% | 27.8% | 18.6 ±0.04% | 83.3% |

± are 95% confidence intervals

[a] Pixels are assigned field size based on average crop field size (calculated from segmentation results) in a 500 m window. (Crops mapped without segmentation polygons are assigned 0.5ha)). Many of the fields in the 2-5 ha group may actually by smaller in areas that border very large fields (that bring up the average).

[b] Percent of eastern Paraguay covered by this tier. For the $<$ 2 ha tier, this does not include areas without crop.

Table A5: Percent crop also identified as crop in WorldSoy

| | |
|---|---|
| **crop in single map** | |
| WorldCereal | 74% |
| WorldCover | 62% |
| LGRIP | 26% |
| Mapbiomas | 68% |
| S.Cone | 87% |
| **crop in two maps (both)** | |
| WorldCereal & WorldCover | 79% |
| WorldCereal & LGRIP | 81% |
| WorldCover & LGRIP | 69% |
| WorldCereal & Mapbiomas | 83% |
| WorldCover & Mapbiomas | 80% |
| LGRIP & Mapbiomas | 78% |
| **crop in three maps (all)** | |
| WorldCereal, WorldCover & LGRIP | 84% |
| WorldCereal, Mapbiomas & LGRIP | 87% |
| WorldCover, Mapbiomas & LGRIP | 84% |
| WorldCereal, WorldCover & Mapbiomas | 85% |
| **crop in four maps (all)** | |
| WorldCereal, WorldCover, LGRIP & Mapbiomas | 88% |

Table A6: Characteristics of sampled districts

| District | Ag. census 2022 %Crop[a] | Ag. census 2022 avgFarm[b] (ha) | GeoWiki avgField (ha) | random sample %Crop | WSoy | WCover | WCereal | MapBiomas | CELPy low crop | CELPy med crop | adjusted area[c] low crop |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Smallholder** | | | | | | | | | | | |
| 1119_JAS | 5% | 0.4 | <0.64 | 9% | 0% | 1% | 1% | 2% | 8% | 2.9% | $5.8 \pm 1.5\%$ |
| 1106_Itaugua | 1.6% | 0.6 | <0.64 | 7% | 0% | 0.5% | 0.3% | 0.6% | 7% | 1.7% | $5.6 \pm 2.0\%$ |
| 0301_Caacupe | 5% | 0.8 | <0.64 | 5% | 0% | 0.5% | 0.2% | 0.5% | 4% | 0.5% | $4.6 \pm 1.6\%$ |
| 0102_Belen | 14% | 1.1 | 0.64 - 2.5 | 15% | 0% | 9% | 3% | 6% | 11% | 0.1% | $9.7 \pm 1.7\%$ |
| 0103_Horqueta | 9% | 1.5 | 0.64 - 2.5 | 11% | 1% | 9% | 6% | 8% | 12% | 0.3% | $13.1 \pm 3.4\%$ |
| 0603_BuenaVista | 17% | 1.7 | 0.64 - 2.5 | 13% | 3% | 5% | 5% | 8% | 14% | 2.1% | $14.9 \pm 2.1\%$ |
| 0502_Caaguazu | 18% | 1.8 | 0.64 - 2.5 | 24% | 5% | 16% | 13% | 8% | 23% | 1.5% | $22.5 \pm 2.4\%$ |
| 0217_Capibary | 21% | 2.4 | 2.5 - 16 | 26% | 9% | 21% | 17% | 18% | 25% | 1.2% | $22.6 \pm 2.1\%$ |
| 0607_SJN | 13% | 2.5 | 0.64 - 2.5 | 13% | 4% | 6% | 6% | 11% | 11% | 1.6% | $12.7 \pm 2.1\%$ |
| 0518_3deFeb | 45% | 3.2 | 0.64 - 2.5 | 33% | 23% | 28% | 28% | 24% | 35% | 1.4% | $35.5 \pm 2.3\%$ |
| **Medium** | | | | | | | | | | | |
| 0806_SantaMaria | 20% | 5.0 | 0.64 - 2.5 | 20% | 5% | 16% | 18% | 15% | 22% | 1.0% | $26.3 \pm 2.6\%$ |
| 0725_AltoVera | 27% | 5.1 | 16 - 100 | 21% | 15% | 17% | 17% | 16% | 18% | 1.3% | $17.3 \pm 2.1\%$ |
| 0112_Arroyito | 7% | 6.9 | 0.64 - 2.5 | 9% | 4% | 8% | 9% | 9% | 11% | 0.1% | $18.2 \pm 3.0\%$ |
| 0506_DrJuan | 69% | 13.2 | 2.5 - 16 | 40% | 20% | 40% | 36% | 30% | 41% | 1.4% | $41.9 \pm 2.7\%$ |
| **Industrial** | | | | | | | | | | | |
| 1016_MingPora | 84% | 50 | 16 - 100 | 58% | 67% | 70% | 70% | 72% | 67% | 0.3% | $68 \pm 1.5\%$ |
| 1012_SanCristobal | 119% | 105 | 16 - 100 | 54% | 51% | 57% | 53% | 57% | 56% | 0.4% | $57.0 \pm 1.7\%$ |
| 1409_LaPaloma | 113% | 562 | 16 - 100 | 59% | 67% | 73% | 70% | 63% | 68% | 0.1% | $69.0 \pm 1.3\%$ |

[a] %Crop = Total crop area / district area. As the same area can be planted with different crops in a year, this number can over represent actual agricultural area.
[b] avgFarm is the average of the total area of each single crop planted by a single farming entity. As a crop can be divided over multiple fields, field size can be smaller than farm size.
[c] Direct stratified estimator combining CELPy with reference sample. Error is 95% confidence intervals. (calculations based on Olofsson 2014)

Table A7: Estimated proportion of crop in eastern Paraguay in each field size bracket

| field size | pixel count | windowed average |
|---|---|---|
| < 1 ha | 11.4% | 9.2 % |
| 1 to 2 ha | 2.0% | 2.2% |
| 2 to 5 ha | 3.8% | 5.2 % |
| 5 to 16 ha | 10.5% | 30.9% |
| > 16 ha | 72.3% | 52.5% |
| smallholder <2ha | 13.4% | 11.4% |
| smallholder <5ha | 17.2% | 16.6% |

Pixel count is the direct count of crop pixels in the final CELPy model, overlaid with field area from the segmentation results. Crop without segmentation area results are assigned to the < 1 ha category. Windowed average applies a 500 m moving window applied directly to the segmentation data to produce a gridded estimate of average field size to enable comparison with other products. Resulting 500 m pixels with <5% crop area in the final CELPy map are assigned 0 and considered non-crop areas, while pixels with 0 in the segmentation data but with >5% crop area are assigned to the <1 ha field-size category.

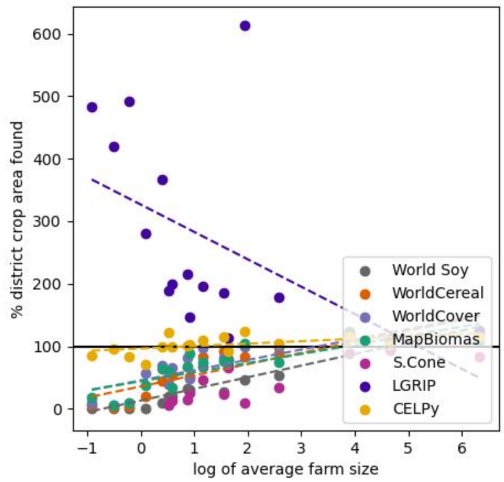



Figure A3: Relationship between average farm size in district and % of the crop area captured by different models. Percent crop area targets are estimated with 300 point random samples for each district. Percent captured is the comparison of that target with the crop area mapped by the product (e.g. if the target crop area is 30% and a map shows 6%, the percent of crop captured is 20%).

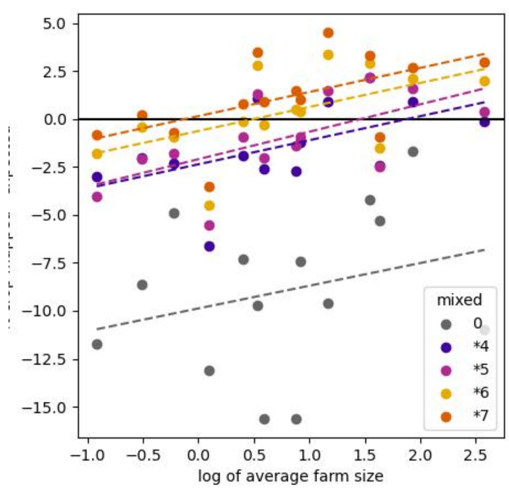



Figure A4: Model optimization based on sample districts. The amount of mixed crop in the sample was selected based on the model with the minimum absolute difference between the expected (sample-based) crop area estimate and the observed (model-based) estimate, summed for all sample districts. The horizontal black line at y=0 illustrates this target.

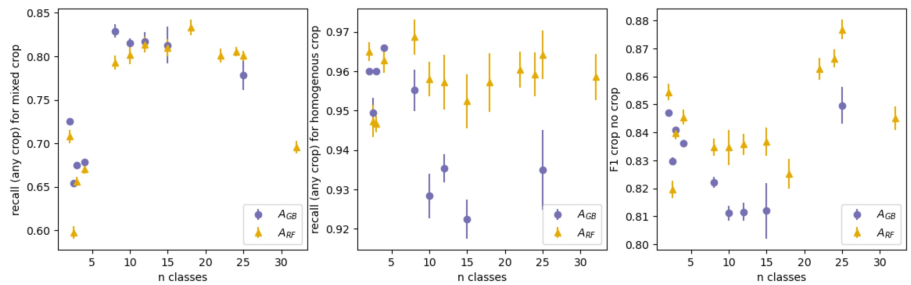



Figure A5: Comparison of performance of model using Gradient Boosting as architecture vs. Random Forest

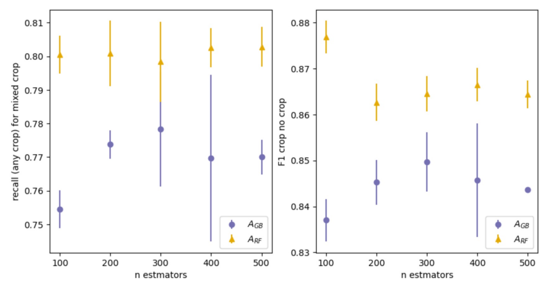



Figure A6: hyperparameter tuning does not have significant effect on performance

Table A8: CELPy breakdown for each GeoWiki strata

| | GeoWiki average field size (ha) | | | | |
|---|---|---|---|---|---|
| | <0.6 | 0.6 - 2.6 | 2.6 - 16 | 16 - 100 | >100 |
| not crop area | 71.6% | 66.0% | 48.8% | 44.3% | 28.8% |
| < 1ha | 10.7% | 16.3% | 10.3% | 4.0% | 1.0% |
| 1to2ha | 1.6% | 2.3% | 2.9% | 1.3% | 0.6% |
| 2to5ha | 3.0% | 3.4% | 6.7% | 3.0% | 3.0% |
| 5to16ha | 8.0% | 7.6% | 20.5% | 2 1.1% | 27.2% |
| >16ha | 5.1% | 4.4% | 10.9% | 26.4% | 39.4% |

CELPy class is determined with 500 m moving window over segmentation data. GeoWiki field size is from 2017 field size map by Lesiv et. al. Lesiv et al. (2019). Note: zero is excluded from all GeoWiki calculations

Table A9: Potential of using GeoWiki field size map to target smallholder areas: Expected composition of sample within GeoWiki strata based on CELPy observations

| | GeoWiki field size[a] | | |
|---|---|---|---|
| CELPy[b] | < 0.6 ha | < 2.5 ha | < 16 ha |
| not crop area | 71.57% | 68.58% | 58.23% |
| smallholder fields (< 5 ha) | 15.31% | 18.88% | 19.37% |
| medium fields(5 - 16ha) | 8.0% | 7.8% | 14.4% |
| large fields (>16 ha) | 5.1% | 4.7%% | 8.0% |

[a] GeoWiki field size is from 2017 field size map by Lesiv et. al. Lesiv et al. (2019). Each column class incudes nominal class and all smaller classes excluding zero.
[b] CELPy field size class is determined with 500 m moving window over segmentation data (unsegmented crop pixels are labeled as <1 ha).