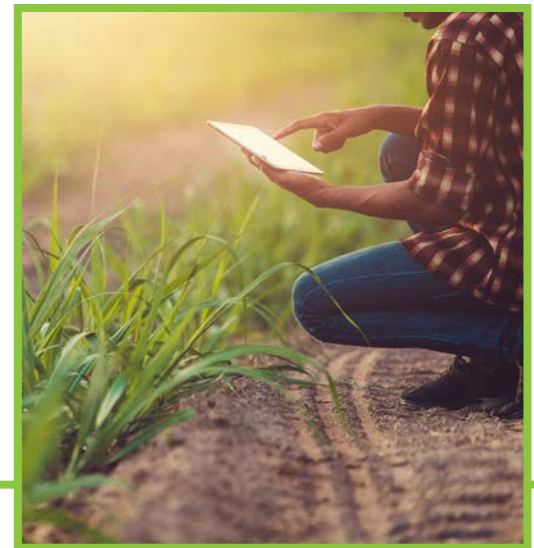




ENABLING CROP ANALYTICS AT SCALE (ECAAS)

Creating Open Agricultural Maps & Groundtruth Data to Better Deliver Farm Extension Services



Contents

1	Overview	3
2	Approach	5
	2.1 Cropland maps	7
	2.2 Crop type maps	9
3	Maps	13
4	Summary of Findings	16
	4.1 Crop type areas and distribution	17
	4.2 Sources of error	18
	4.3 Methods that improve map quality	19
	4.4 Key Findings	20
5	Next Steps	21
6	Data Availability	23
7	References	24
8	Appendix	25
	8.1.1 Derived image features	26
	8.1.2 Model Variables	27

1

Overview

The report provides a summary of the final map deliverables for the ECAAS project “[Creating Open Agricultural Maps and Groundtruth Data to Better Deliver Farm Extension Services](#)”. This project was developed to address the challenges of collecting ground truth data on crop types in a sustained manner and to create reliable maps of crop types that cover large areas and multiple seasons. The goal of developing such maps is to use them to improve subsequent extension services and develop value-added products for farmers and agribusinesses, in order to boost reach and revenues, and thereby support the expansion and continuation of ground-data collection and map development.

In this project, groundtruth data were collected by Farmerline field agents, who used the GPS-enabled Mergdata platform to delineate the boundaries of crop fields belonging to farmers with whom a relationship has already been established. These observations, which were collected in accordance with recommended best practices¹, were then used to create crop type maps from Sentinel-1, Sentinel-2, and PlanetScope imagery by leveraging recent advances in machine learning and Earth Observation, including deep learning models. The maps were developed for the Ejura-Tain region, the primary focus of the first phase of this project (**Figure 1.1**).

Figure 1:

Dynamic frames can be generated using rainfall or NDVI data. Rainfall data have greater predictive capabilities at larger geographic scales, while NDVI is more useful at sub-national scales.



¹ <https://github.com/radianteearth/ground-referencing-guide>

2

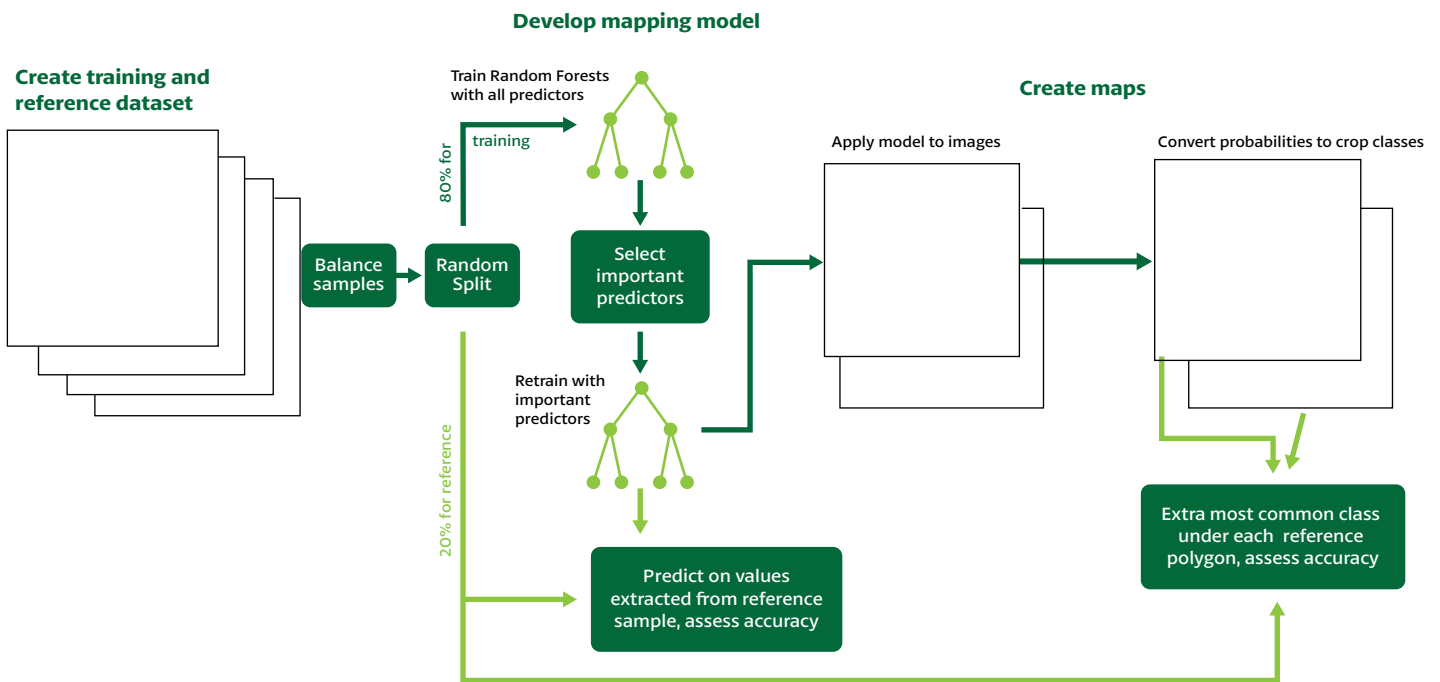
Approach

To develop these maps, we followed the approach described in the [report](#) from the first season and illustrated in Figure 2.1. This entails the following steps:

- Use field-collected crop type observations (so-called Class 1 labels) to develop a set of training and reference labels;
- Process time series of Sentinel-1, Sentinel-2, and PlanetScope data covering the relevant seasons;
- Extract image data underlying the labels;
- Train a Random Forests model to predict maize, rice, other crops, and noncrop classes, evaluating the model performance using a reserved subset of labels;
- Use the model to map the four classes throughout the study area
- Filter the predictions through high-resolution field boundary maps, developed using a deep learning model applied to PlanetScope data, in order to confine the predictions to those areas most likely to be crop fields.

Figure 2.1:

An overview of the approach used to develop the Random Forests-based mapping model.



In the next sections, we provide a summary of the model and approach used for developing the maps.

2.1 Cropland maps

To create the cropland maps for 2020 and 2021 that we used to filter the crop type predictions, we used a U-Net (a convolutional neural network; Ronneberger et al. 2015) model that was developed for the purpose of creating annual, country-scale maps of crop field boundaries. This model represents an evolution of a proof-of-concept method that was used to map Ghana's field boundaries for the year 2018 (Estes et al., 2021), and its application in this project demonstrates its capability to create yearly field boundary maps. We used a version of the model trained on 2018 PlanetScope imagery with a set of 4593 labels collected across Ghana, Tanzania, and the Republic of Congo. This "global" model was trained to classify both field boundaries and field interiors. We used the maps developed from the predictions of the interior class to identify and delineate individual fields or field clusters. The accuracy of this initial model was 86.1% for the field interior class, with corresponding true and false positive rates of 65.2 and 8.3%, respectively, indicating the model makes relatively little commission error but misses about 35% of fields.

We applied this model to the Ejura-Tain study region with the following steps:

1. We resampled PlanetScope basemap imagery from November 2020 to ~3 m resolution and retilled it to match the grid used for prior Ghana-wide mapping work (Estes et al, 2021).
2. We digitized a set of 120 field boundary labels on this imagery, from sites selected within the Ejura-Tain region.
3. We then refined the model using 108 of the collected labels. We froze the weights on the first 58 layers of the model, letting the model update parameters on the remaining layers over 15 epochs.

The resulting refined model, assessed against the 10% of labels reserved for validation, was 83.1% for the field interior class, with true and false positive rates of 61.9 and 11%. This performance is slightly lower than that of the global model, but this was expected given that 1) the imagery was processed differently (Planet uses a "best on top" approach to make basemaps) than the PlanetScope data used to train the original model (a temporal compositing procedure described in Estes et al., 2021), and 2) the Ejura-Tain region is particularly challenging to map, given its position at the transition between tropical forests and savannas, and corresponding high cloud cover (Estes et al., 2021). The primary implication of these error rates for map usage is that they will lead to underestimates of total cropped area, and, when used to filter crop type maps, to underestimates of total planted area for different crops. Such underestimates can be corrected using an appropriately designed map reference sample (Stehman et al, 2019).



We are also taking steps to further reduce error by introducing several refinements to the model architecture (e.g. more connections between model layers) and by increasing the number of training labels.

We used the refined model to map field interiors on imagery from both November, 2020 and November 2021, which is the middle of the short growing season that we focused on for mapping crop types. After predicting the field interiors for each region and each year, we used a simple approach to label clumps of pixels representing distinct fields or clusters of fields into individual *instances*. We converted these into vectors to provide field polygons. This approach in some cases results in multiple fields being grouped together, particularly where boundaries between adjacent fields are indistinct, but it nevertheless provides a much finer segmentation than the version 1 approach we developed for Ghana (Estes et al., 2021).

A comparison of the mapped field boundaries for 2020 and 2021 over one of the 5x5 km tiles covering the study region is shown in **Figure 2.2**. Although some of the differences are due to model error, these maps to a substantial degree capture the significant year-to-year variations in field distributions that occur in these croplands.

2020

2021

Figure 2.2:

Cropland boundaries (yellow) mapped with a deep learning model for the month of November in 2020 (left) and 2021 (right). Field boundaries are shown over the corresponding PlanetScope basemaps for those two time periods, rendered in false color.

2.2 Crop type maps

We used the Class 1 labels collected by Farmerline's agents over two seasons to develop crop type models based on the Random Forests algorithm (Breiman, 2001). These included labels from both the first short season of 2020 (August–December) and the short season of 2021. We focused on these two seasons because there were insufficient observations from the long season of 2021 (~April–September) to train a model, and optical data are hard to obtain during this season because of frequent cloud cover. The number of observations from these two seasons (2020 and 2021) is shown in Table 2.1. The distribution of observations is shown in **Figure 2.3**.

Crop	Year	Count
Maize	2020	571
Maize	2021	434
Other	2020	18
Other	2021	618
Rice	2020	58
Rice	2021	305

Table 2.1:

The count of Class 1 labels collected by crop type by year.

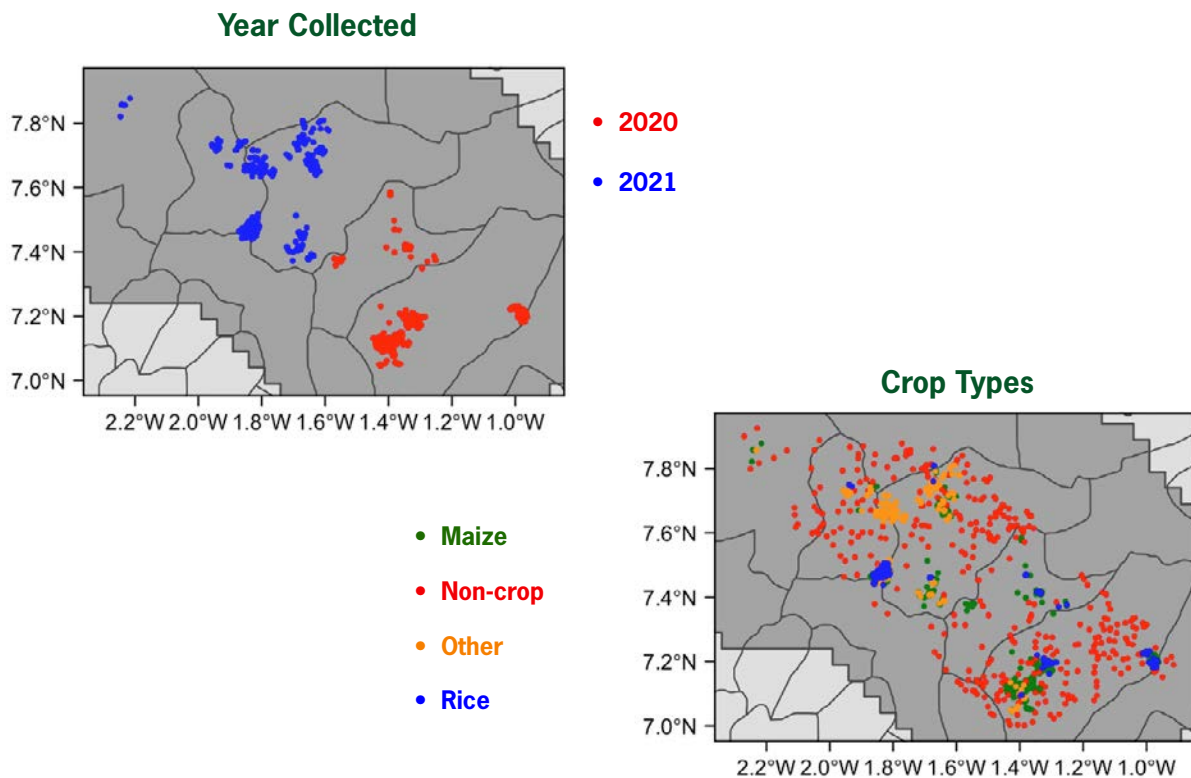


Figure 2.3:

Cropland boundaries (yellow) mapped with a deep learning model for the month of November in 2020 (left) and 2021 (right). Field boundaries are shown over the corresponding PlanetScope basemaps for those two time periods, rendered in false color.

To train the model, we developed stacks of Sentinel 1, Sentinel 2, and PlanetScope imagery covering each season, following the same approach we [described](#) for our initial model developed for the 2020 season, with the exception that the time periods used to cover season 2 were necessarily shorter because the imagery was only available through the end of November. The following images sets were used:

- Sentinel 1 imagery covering the full year through November:
- Sentinel 2 imagery processed into two seasonal composites, one covering a wide period over the major season prior (typically February–October, which is needed to have enough imagery to deal with the high cloud cover), and the second covering the short season following it;
- PlanetScope basemap composites for October and November 2021.

Sentinel-1 data were processed further using harmonic regression to extract seasonality information from the time series, and a suite of vegetation indices was derived from the Sentinel-2 data (**see Table 8.1 in Appendix**).

All images were resampled to the 10 m resolution of Sentinel-1's visual and near-infrared bands, and the Class 1 labels were used to extract the values from the imagery from the corresponding season.

To train the model, we selected 80% of the observations from each crop type, reserving the remaining 20% for validation. Given the uneven numbers of labels across crop types, we balanced the sample by making several random draws from the rice and other crop type classes, randomly selecting subsets of pixels from each field, and averaging their values, in order to boost the final numbers in each class to be approximately equal (1000–1200 per class across training and validation samples). The training and validation samples were kept separate during this resampling process.

We used the resulting sample to train a multi-class Random Forests model, with a tree depth of 100 and 1000 trees. We fit two versions of the model. The first variant used the training set collected from both years to train the model (referred to here as the *two-year model*), and the second was trained using only those labels collected in 2021 (*the 2021 model*). We evaluated the variable importance of both models after training them the full set of predictors and then retained the 21 most important variables to train the final two-year model, and 28 to train the final 2021 model. The variable importance ranking and the cutoff used to train the final two-year model are illustrated in Figure 8.1 in the Appendix.

The two-year model was most effective, having an overall accuracy against the reserved test set of 77.4% when trained with the full set of predictors, and 76% when trained using the subset of most important variables. In contrast, the final 2021 model's accuracy was just 69.7%.

For the final two-year model, the accuracy per class varied. In terms of User's accuracy (the complement of commission, or false positive, error), Rice was classified least accurately (67%) and maize the most accurately (75%, **Table 2.2**), while the non-crop class was most accurately predicted (92%). The same pattern is seen in the Producer's accuracy (the complement of omission, or false negative, error). For the 2021 model (**Table 2.2**), which had a much smaller test sample, the User's accuracy was lowest for maize (56.4%) and highest for rice (65%), although this score is lower than the least accurate class of the two-year model. The producer's accuracy for maize was the highest (63%) of the three crop types. As with the two-year model, both accuracy measures were highest for the non-crop class. The higher accuracy for the two-year model presumably reflects its larger sample size (for both training and test sets) as well as its greater spatial coverage of the mapping region.

Table 2.2:

The Error matrix and accuracy measures for the four crop classes for the two-year Random Forest model (top) and the 2021 model (bottom). The overall (O), User's (U), and Producer's accuracies, as are total sample counts. The error matrix lists reference values in columns and predicted values in rows.

Two-year Model	Maize	Rice	Other	Non-crop	Total	U	O
Maize	146	28	17	3	194	75.3%	76
Rice	23	140	42	3	208	67.3%	
Other	18	44	178	14	254	70.1%	
Non-crop	2	4	11	197	214	92.1%	
Total	189	216	248	217	870		
P	77.2	64.8	71.8	90.8			

Two-year Model	Maize	Rice	Other	Non-crop	Total	U	O
Maize	44	22	10	2	78	56.4	69.7
Rice	13	58	18	0	89	65.2	
Other	11	11	49	5	76	64.5	
Non-crop	2	4	3	81	90	90	
Total	70	95	80	88	333		
P	62.9	61.1	61.3	92			

We used the two-year model to predict crop types for the full study area for both the 2020 and 2021 short seasons and used the 2021 model to make a second set of maps for the 2021 season. The total map extent was confined to a boundary drawn around the outermost locations of Class 1 labels collected during the combined 2020–2021 campaigns.



3

Maps

Each model produced four different probability layers, one for each crop class. We used two different approaches to create maps from these probability layers:

- 1. Thresholding:** Using a predefined threshold of 0.5, pixels were classified to a particular crop type when its predicted probability exceeded the threshold. If no crop exceeded that threshold, we assigned the class to "other". We excluded the non-crop class in this approach.
- 2. Max-class:** Each pixel was assigned the crop type having the highest predicted probability.

The first approach is more conservative and identifies the highest confidence predictions, but tends to substantially underestimate the area of each crop. The second approach can lead to more error, as it makes a classification even when there is high uncertainty due to close probability values between the four classes (e.g. Maize = 0.27, Rice = 0.26, Other = 0.24, non-crop = 0.23).

After classification, the resulting maps were filtered through the field boundary maps. The classified crop types corresponding to each field were extracted, and then the entire field was assigned the class of the most frequently occurring crop type.

The resulting filtered crop type maps are shown in **Figure 3.1**, and the areas for each crop derived from the max-class maps are shown in **Figure 3.2**.



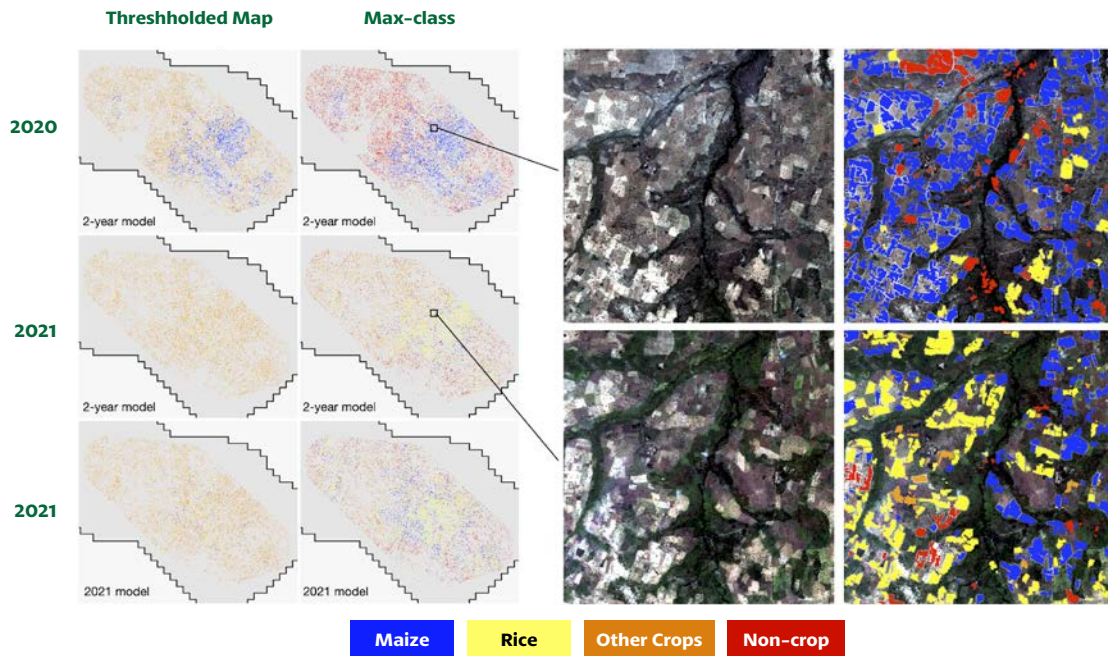


Figure 3.1:

Crop type maps for 2020 and 2021 over the Ejura Tain region, made using two different Random Forests models and two different techniques for converting model predictions into maps. The top two rows show maps made with a model trained using Class 1 labels collected in 2020 and 2021, while the bottom row shows maps made for 2021 using a model developed using 2021 labels only. The two columns on the main maps show, on the left, maps made by thresholding the predicted probabilities to classify the crop in the pixel, and (right), maps in which each pixel was assigned to the class with the highest predicted probability. The four panels on the right show the PlanetScope imagery from November of 2020 (top row) and 2021 (bottom row), with the crop types in each field predicted by the two-year model.

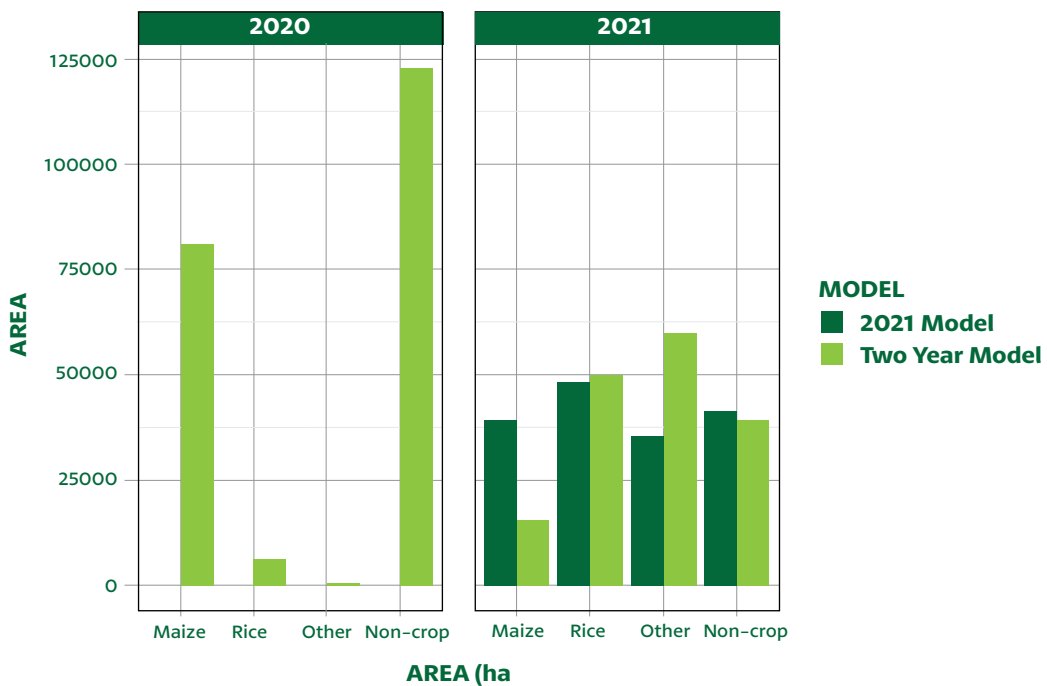


Figure 3.2:

Areas of each crop type by year, as mapped using the max-class versions of the two-year and 2021 models.

4

Summary of Findings

4.1 Crop type areas and distributions

The maps show similar patterns in the spatial distribution of maize and rice, which are most concentrated in the center to the eastern half of the region. However, there were substantial differences in the areas of crop types between 2020 and 2021, with maize having a substantially larger mapped area in 2020 than in 2021, while the rice was predicted to be more abundant than maize in 2021 (**Figures 3.1**).

The two-year max class map found substantially more area of other crops in 2021 than in 2020, detecting almost no areas of “other” crops in 2020. In 2020, the non-crop class occupied the largest total area, even though these predictions were confined to areas mapped as croplands.

The 2021 max-class model showed more balance between the four classes than the two-year model's predictions for 2021, which predicted substantially greater areas of rice and other crops than maize.

The total area of cropland mapped was 210373 ha in 2020 and 164827 ha in 2021. This difference indicates that the 2021 cropland map had a larger omission error than the 2020 map.

4.2 Sources of error

These findings more likely reflect underlying issues with the input data than actual differences in crop type distributions. For example, it is unlikely that the area of maize planted in 2021 was smaller than that of rice. However, the exact nature of these discrepancies cannot be resolved in the absence of a statistically robust reference sample, which is currently under development (**see Next Steps**).

One of the primary issues impacting results is the spatially distribution of the samples. The collection of Class 1 labels was geographically clustered by both season and crop type (**Figure 2.3**). This clustering surely introduced biases into the models and may help explain the substantial variability in detected crop types between years. It also explains why the map accuracy assessment, even though it was based on independent samples, shows reasonable accuracy while the maps themselves show large discrepancies—a clustered sample produces misleading performance measures, just as a clustered training sample introduces model bias.

The second key data issue was the incomplete image time series for 2021. We were only able to collect imagery for part of the 2021 short season (through the end of November), and thus the input data did not capture the full crop phenology. This surely increased the classification uncertainty in 2021. This higher uncertainty is reflected in the smaller number of maize and rice fields that were detected using the thresholding approach in 2021 relative to 2020, which reflects the fact that the predicted probabilities for each crop type were low.

Beyond these data issues, the local environment complicates mapping. The high cloud cover in this region makes it difficult to construct time series of Sentinel-2 data. Sentinel-2 has proven to be the most useful source of data for crop type classification in similar environments (Azzari et al, 2021), and it also provided several of the most important predictors for our models. However, here we were forced to rely on temporal composites of Sentinel-2 constructed over broad time periods, in order to overcome the high cloud cover. In doing so, we lost valuable phenological information. This information loss was mitigated by the availability of PlanetScope imagery, which provided data during two of the key growing months (October and November), and was also among the most important predictors (**Figure 8.1**). Another factor is the nature of the local cropping systems, in which crops are often planted inter-mixed with trees and other vegetation that confuse the classification algorithms.



4.3 Methods that improve map quality

Several aspects of our methodology proved useful to help improve results in this hard-to-map environment. In particular, the deep learning-based cropland mapping approach is effective in delineating most individual fields or small groupings of adjacent fields. Although this method still misses a number of fields and does commit some false positive (albeit at a much lower rate), both types of errors can be reduced by refining model architecture and collecting a larger number of labels to refine the model (e.g. 500–1000 within the existing mapping region), divided between the 2020 and 2021 imagery. Despite these errors, the approach appears capable of detecting the substantial variation in field distributions that occur between years. The existing pre-trained model can be readily updated and applied to subsequent years, and transferred to different locations. This approach, therefore, shows great promise for tracking cropland change and provides an effective and scalable tool for filtering crop type maps.

With respect to crop type models, the thresholding approach applied to the Random Forests model holds promise as a means for improving model quality. Although we see here that thresholding substantially underestimated maize and rice abundance, particularly in 2021, it is nevertheless useful because it highlights high confidence predictions. These predictions can be converted to synthetic labels (we refer to these as Class 3 labels) to boost model training samples. Synthetic labels for crop type mapping is an approach that is also being advanced by the Radiant Earth Foundation.

In terms of data sources, using PlanetScope imagery, which has a higher revisit rate enabling composites that are less impacted by cloud, helped to overcome the loss of phenological information provided by Sentinel-2 data. Combined with Sentinel-1 data, this source of imagery helped to improve model accuracy. This finding is in accord with a recent crop type mapping study over small-scale farming systems in India, which found that models based on Sentinel-1 and PlanetScope predictors slightly outperformed models that relied on Sentinel-1 and Sentinel-2 predictors, due to the improved spatial resolution of PlanetScope (Rao et al, 2021). This benefit of using PlanetScope imagery should continue to increase (if budgets allow for imagery purchase) as the recently launch Super Dove satellites have improved spectral resolution, which will further improve classification accuracy.

Although untested in this study, data fusion approaches that combine multiple image sources, including PlanetScope, Sentinel-1, Sentinel-2, Landsat, and even MODIS (e.g. Houbourg et al, 2018; Kpienbaareh et al, 2021; Orynbaikyzy et al., 2020), may further improve crop type, classification models. Fusion methods that incorporate PlanetScope, thereby increasing the spectral content of predictor data while achieving high spatial and temporal resolution (e.g. Houbourg et al, 2018), also hold promise for helping to improve classification accuracy where crop types are mixed or planted in very small plots.



4.4 Key Findings

The following key findings emerged from this project:

- A convolutional neural network (CNN) applied to PlanetScope imagery showed promise for mapping annual changes in crop field boundaries in highly dynamic agricultural landscapes. These field boundary maps provided an effective filter for crop type mapping models, although further improvements are still needed to reduce the CNN's omission error;
- Geographic clustering in the groundtruth data led to biased models and made it difficult to reliably assess model performance. Both of these shortcomings may be remedied by using Class 2 labels created from drone imagery that was collected following a probability-based sample design;
- The number and geographic distribution of training labels may be further improved through the use of synthetic Class 3 labels, which are drawn from the highest confidence predictions resulting from the initial crop type maps;
- Using PlanetScope imagery helped to mitigate the loss of information from Sentinel-2 due to high cloud cover.



5

Next Steps

During the next months we will take several steps to improve these results and use these to develop reliable estimates of crop type distributions in the study region.

First, we will complete the full-time series of imagery for the current season after January, and retrain the model, which should improve with more phenological data.

Second, we will develop Class 2 labels from visual interpretation of UAS imagery that was collected during a drone collection campaign conducted in November/December in this region, following a probability design. These data will allow us to improve the spatial and class distribution of training samples, and provide a statistically independent reference sample that can be used to reliably assess map accuracy.

Third, we will use the thresholding approach to develop synthetic Class 3 labels. We will add these and a portion of Class 2 labels to the training sample and use to retrain the model and assess performance gains.

Fourth, we will replace the Random Forests model with two neural network approaches that we are developing under Phase 2 of this project.

Fifth, we will continue to refine the deep learning-based cropland mapping approach, in order to further improve maps used for filtering crop types.



6

Data Availability

The crop type map and labels are hosted on a public bucket on AWS S3. We created a **GitHub repository** that provides instructions for viewing and downloading these data. They are also available on a public **Box folder** for download by users who cannot access AWS resources.

This repository is currently private. Upon approval of this deliverable, it will be made open so that the data will be publicly available.

The labels will also be submitted to Radiant MLHub following final validation by the Farmerline team.

7

References

1. Azzari, G., M. Jain, and D. B. Lobell. 2017. Towards fine-resolution global maps of crop yields: Testing multiple methods and satellites in three countries. *Remote Sensing of Environment* 202:129–141.
2. Breiman, L. 2001. Random Forests. *Machine Learning* 45:5–32.
3. Estes, L. D., S. Ye, L. Song, B. Luo, J. R. Eastman, Z. Meng, Q. Zhang, D. McRitchie, S. R. Debats, J. Muhando, A. H. Amukoa, B. W. Kaloo, J. Makuru, B. K. Mbatia, I. M. Muasa, J. Mucha, A. M. Mugami, J. M. Mugami, F. W. Muinde, F. M. Mwawaza, J. Ochieng, C. J. Oduol, P. Oduor, T. Wanjiku, J. G. Wanyoike, R. Avery, and K. Caylor. 2021. High resolution, annual maps of field boundaries for smallholder-dominated croplands at national scales. *Frontiers in Artificial Intelligence*. DOI:10.3389/frai.2021.744863
4. Houborg, R., and M. F. McCabe. 2018. A Cubesat enabled Spatio-Temporal Enhancement Method (CESTEM) utilizing Planet, Landsat and MODIS data. *Remote Sensing of Environment* 209:211–226.
5. Jin, Z., G. Azzari, C. You, S. Di Tommaso, S. Aston, M. Burke, and D. B. Lobell. 2019. Smallholder maize area and yield mapping at national scales with Google Earth Engine. *Remote Sensing of Environment* 228:115–128.
6. Kpienbaareh, D., X. Sun, J. Wang, I. Luginaah, R. Bezner Kerr, E. Lupafya, and L. Dakishoni. 2021. Crop Type and Land Cover Mapping in Northern Malawi Using the Integration of Sentinel-1, Sentinel-2, and PlanetScope Satellite Data. *Remote Sensing* 13:700.
7. Orynbaikyzy, A., U. Gessner, B. Mack, and C. Conrad. 2020. Crop Type Classification Using Fusion of Sentinel-1 and Sentinel-2 Data: Assessing the Impact of Feature Selection, Optical Data Availability, and Parcel Sizes on the Accuracies. *Remote Sensing* 12:2779.
8. Rao, P., W. Zhou, N. Bhattarai, A. K. Srivastava, B. Singh, S. Poonia, D. B. Lobell, and M. Jain. 2021. Using Sentinel-1, Sentinel-2, and Planet Imagery to Map Crop Type of Smallholder Farms. *Remote Sensing* 13:1870.
9. Ronneberger, O., P. Fischer, and T. Brox. 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation. Pages 234–241 in N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, editors. *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*. Springer International Publishing, Cham.

8

Appendix

This appendix contains further details on methods and results.

8.1.1 Derived image features

From the processed imagery, we derived a number of additional features to be used as predictors. We fit a “Least Absolute Shrinkage and Selection Operator” (LASSO) regression to the Sentinel-1 time series, resulting in 6 coefficients extracted from the full annual time series, which are informative about vegetation phenology. We also derived the following vegetation indices from Sentinel-2 data (following Jin et al; 2019):

Table 8.1:

Vegetation indices derived from the bands of Sentinel-2. All indices were derived for each seasonal composite.

Index	Formula
NDVI	$(\text{NIR} - \text{Red}) / (\text{NIR} + \text{Red})$
GCVI	$(\text{NIR}/\text{Green}) - 1$
RG1_GCVI	$(\text{NIR}/\text{RedEdge1}) - 1$
RG1_GCVI	$(\text{NIR}/\text{RedEdge2}) - 1$
MTCI	$(\text{NIR} - \text{RedEdge1}) / (\text{RedEdge1} - \text{Red})$
MTCI2	$(\text{RedEdge2} - \text{RedEdge1}) / (\text{RedEdge1} - \text{Red})$
REIP	$700 + 40 * ((\text{Red} + \text{RedEdge3}) / 2 - \text{RedEdge1}) / (\text{RedEdge3} - \text{RedEdge1})$
NBR1	$(\text{NIR} - \text{SWIR1}) / (\text{NIR} + \text{SWIR1})$
NBR2	$(\text{NIR} - \text{SWIR2}) / (\text{NIR} + \text{SWIR2})$

Index	Formula
NDTI	$(\text{SWIR1} - \text{SWIR2}) / (\text{SWIR1} + \text{SWIR2})$
CRC	$(\text{SWIR1} - \text{Green}) / (\text{SWIR1} + \text{Green})$
STI	$\text{SWIR1} / \text{SWIR2}$



8.1.2 Model variables

The ranking of the importance of variables in the two-year Random Forests models is shown in **Figure 8.1**.



Figure 8.1:

Variable importance is assessed in terms of the mean decrease in model accuracy when the variable is removed. Variables named 'PLA_Oct' and 'PLA_Nov' refer to PlanetScope basemaps from October or November of either year (depending on which season the corresponding label was collected in), followed by the band number (indicated after the underscore). Sentinel-2 bands are those beginning with 'B', followed by a channel number and interval number (1 = February – October 2020; 2: October 2020 – January 2021). Vegetation indices are described in Table 2.2. Sentinel-1 harmonic coefficients for the VV or VH polarizations are indicated by their coefficient number. The red line indicates the threshold used for retaining variables to use in the final model.



**Enabling Crop
Analytics At Scale**

Creating Open Agricultural Maps and Groundtruth Data to Better Deliver Farm Extension Services

The Enabling Satellite-based Crop Analytics at Scale (ECAAS) Initiative is a multi-phase project that aims to catalyze the development, availability, and uptake of agricultural remote-sensing data and subsequent applications in smallholder farming systems. The initiative is funded by The Bill & Melinda Gates Foundation and implemented by Tetra Tech.

**info.ecaas@tetratech.com
cropanalytics.net**

Deliverables:

D1.10: Improved cropland maps (all districts)
D1.11: Year 2 crop models, crop type maps, and model release
D2.4: Hosted platform with ground truth and non-value added map data for public access

Authored & edited by:

1. Lyndon Estes, Clark University
2. Amos Olerterey Wussah, Farmerline Limited
3. Mary Dziedzorm Asipunu, Farmerline Ltd

Contributions from:

1. Zexing Zheng, Clark University
2. Lei Song, Clark University
3. Sitian Xiong, Clark University
4. Aandishah Samara, Clark University
5. Sai Vishal Muda, Clark University
6. Ismail Alatise, Clark University
7. Sam Khallaghi, Clark University
8. Worlali Senyo, Farmerline Ltd
9. Mami Araba Kwenua Amuah, Farmerline Ltd
10. Dzidzor Boni, Farmerline Ltd
11. Richmond Aryee Lakoussan, Farmerline Ltd
12. Marvin Armah, Farmerline Ltd