

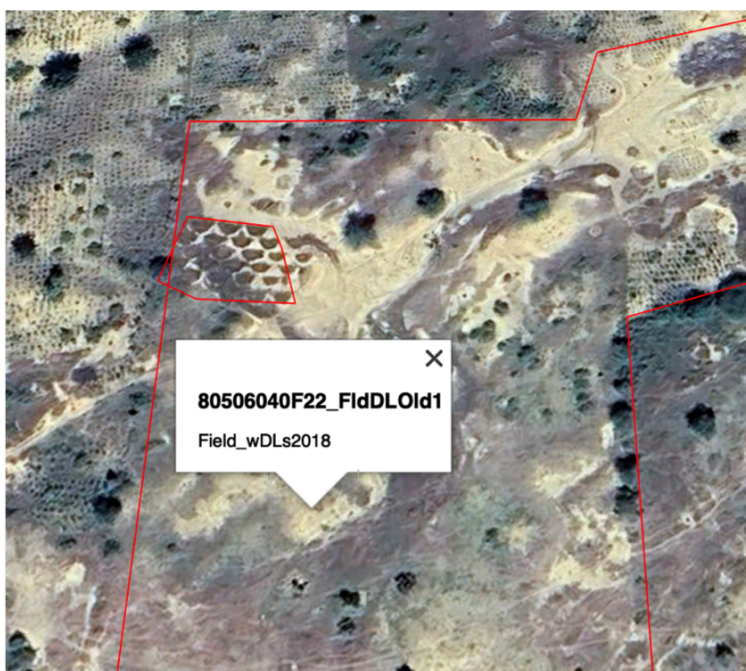
Appendix to
“How Many (Half) Moons?
Measuring Technology Adoption From the Sky and on the Ground”

Jenny C. Aker, Jennifer Burney, Alison Campion, B. Kelsey Jack and Chuan Liao

A Figures and Tables



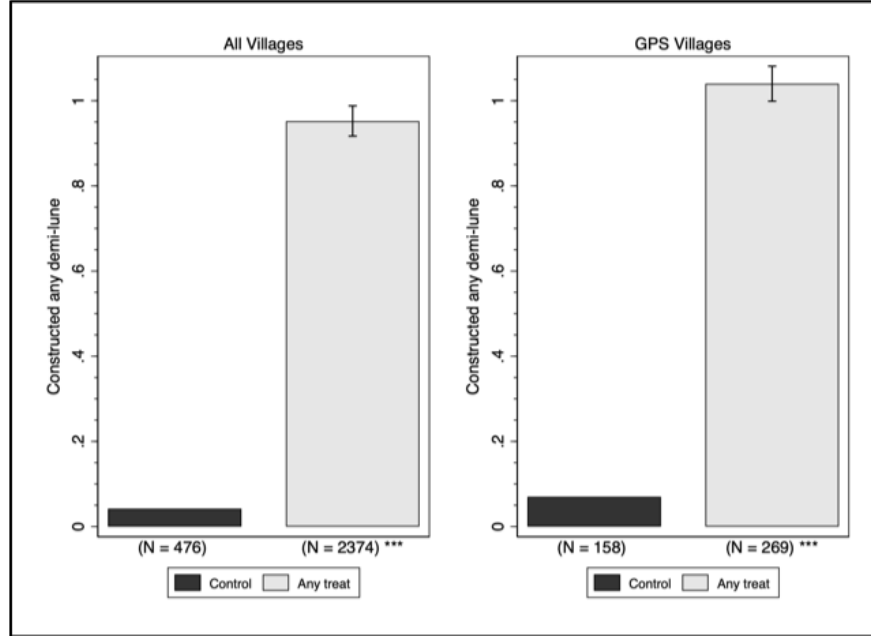
(a) Photograph of demi-lunes during growing season



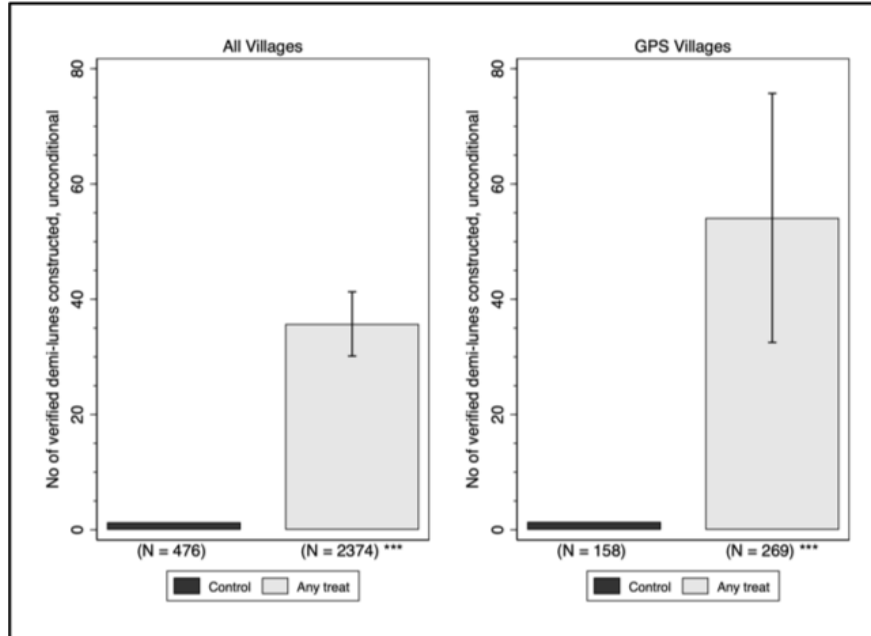
(b) Satellite image of demi-lunes from Google Earth

Figure A.1: Images of demi-lunes

Note: Figure (a) shows a photograph of demi-lunes planted with crops during the growing season. Figure (b) shows image of demi-lunes as observed in Planet/Google Earth satellite imagery. The outer red polygon defines the perimeter of the farmer's field. The inner red polygon demarcates the portion of the field where adoption took place between 2018 and 2021. We refer to the latter as the *subfield* adoption polygon, and the former as the field polygon.



(a) Extensive margin (any adoption)



(b) Intensive margin (number of demi-lunes)

Figure A.2: Treatment effects on adoption: Direct observation

Note: Figure (a) shows the extensive margin of adoption for the control and treatment groups in 2018. Figure (b) shows the intensive margin (continuous number of demi-lunes adopted) in 2018. In both (a) and (b), the full sample of 180 villages is shown on the left, and the subsample of 26 GPS villages is shown on the right.

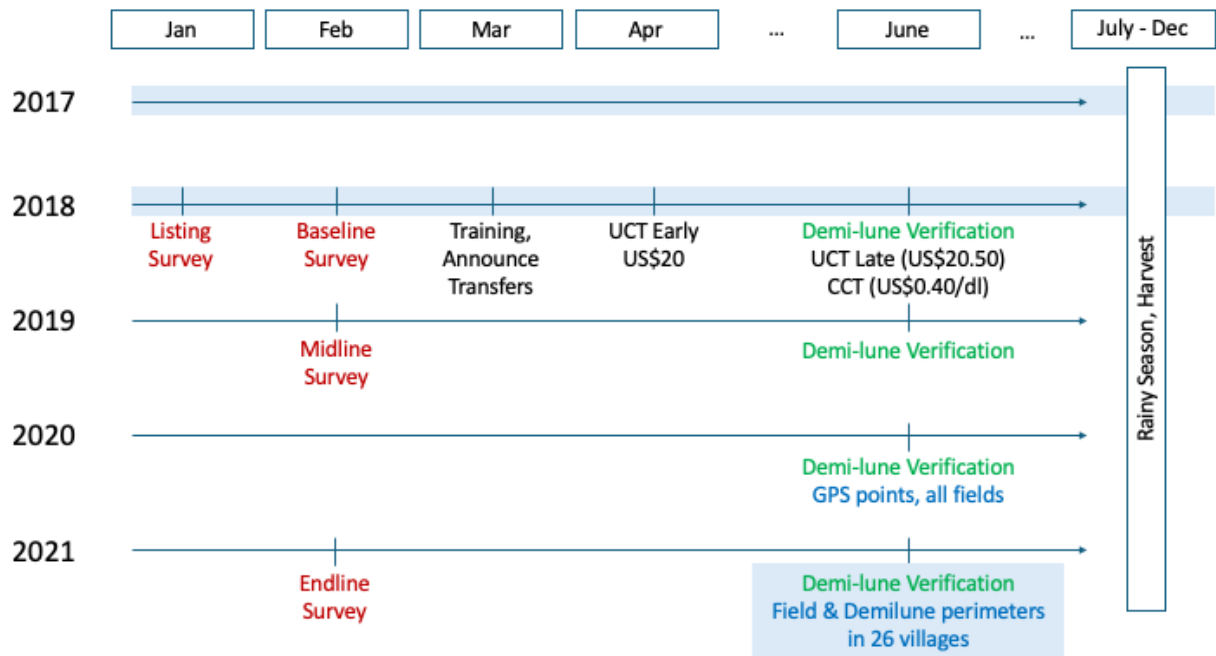
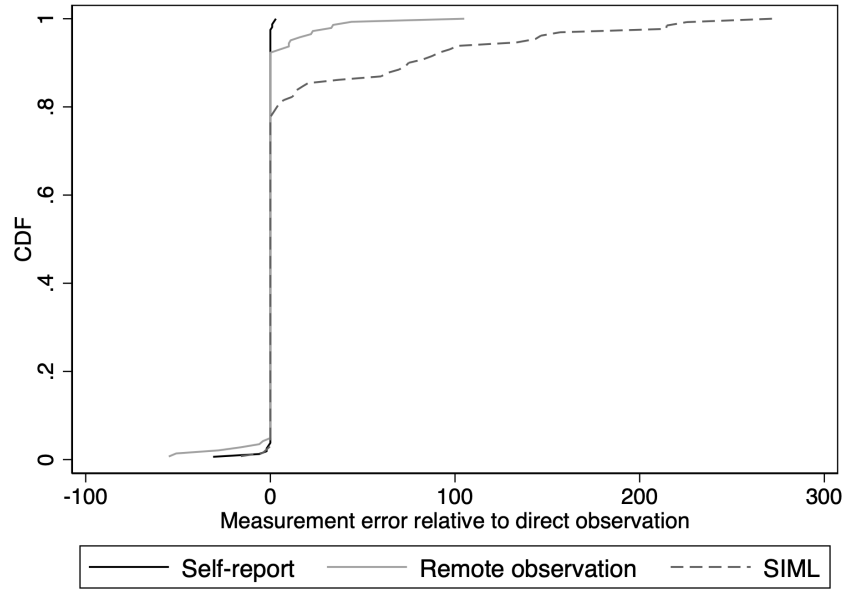


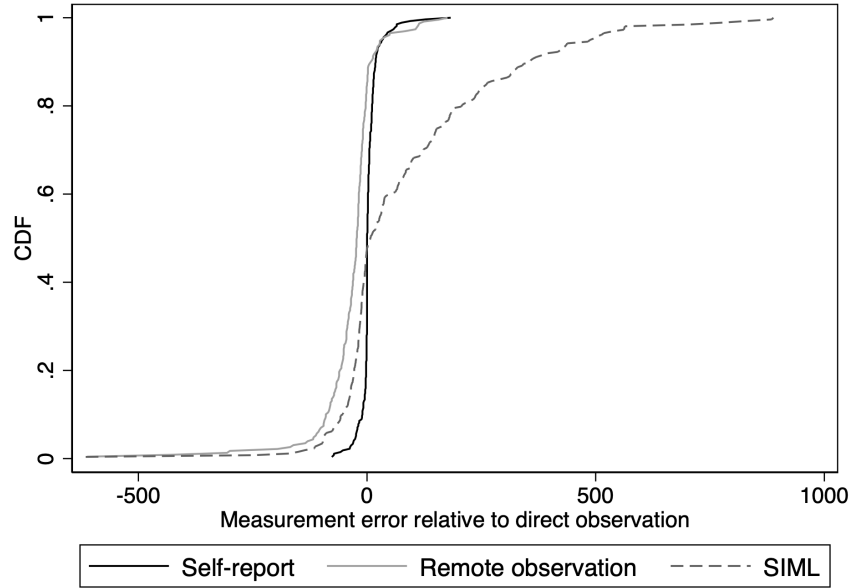
Figure A.3: Project and data collection timeline

Note: Timeline for project implementation and data collection activities between 2018 and 2021.

Implementation activities are in black, self-reported surveys are in red, enumerator observations are in green and GPS data collection points are in blue. During the enumerator observation visits, the enumerators also asked questions about self-reported adoption; we use these data for our self-reported measures. The data used in the *GPS sample* is highlighted in blue; data on field and subfield perimeters were collected in 2021, three years after initial interventions.



(a) Control group



(b) Treatment group

Figure A.4: Measurement error, intensive margin of adoption

Note: Measurement error ($Y_i^M - Y_i^*$) on the intensive margin, at the farmer level. The CDF of the difference is plotted separately for the control (a) and treatment (b) groups.

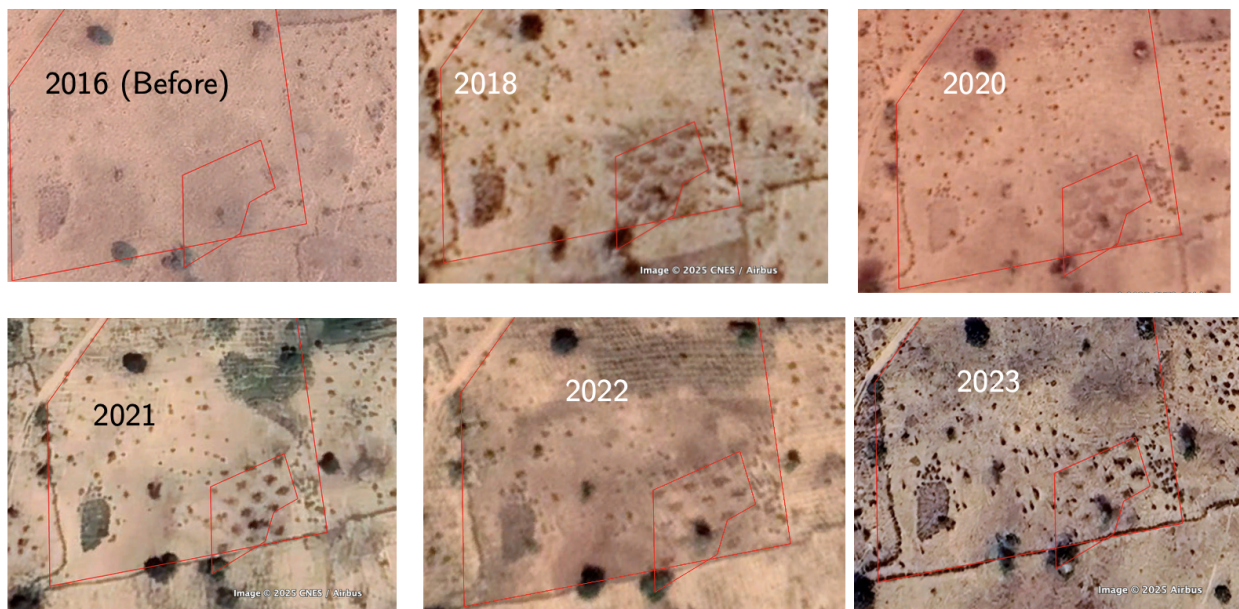


Figure A.5: Satellite images over time

Notes: These images show demi-lune adoption using Planet/Google Earth over time. These images were used for the remote observation technique.

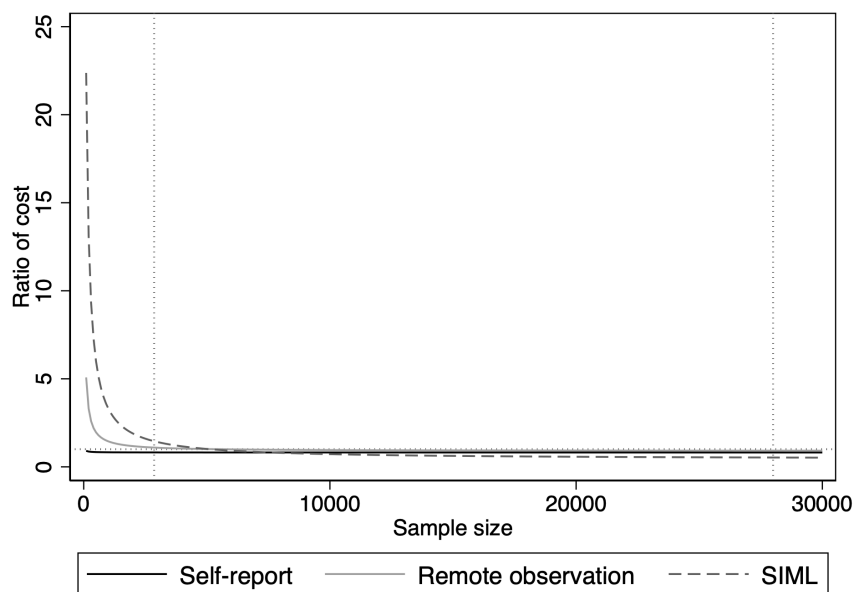


Figure A.6: Cost simulations ($t=3$), as sample size increases

Note: Simulation of the cost of each alternative measurement approach relative to direct observation, as the sample size varies along the horizontal axis. Three rounds of measurement are assumed. Vertical lines indicate the RCT sample and the scale up sample. The dashed horizontal line is at 1 (equal cost ratio between the approach and direct observation).

Table A.1: Pros and cons of each measurement method

Method	Pros	Cons
Direct observation	Accurate, easy analysis, no GPS required	Expensive at scale, potential access issues, potential experimenter demand effects
Survey self-report	Quick, can be bundled with other survey questions, easy analysis, no GPS required	Expensive at scale (unless phone calls), potential non-classical measurement error, potential experimenter demand effects
Remote observation	Can be done over different years and at different times of year Potentially easy analysis	Measures are sensitive to time of year (cloud cover) and data availability; potential Rosenthal effect; requires field perimeters;
SIML	Data available at high temporal frequency; multistreamal imagery can be used to predict other “downstream” outcomes	Measures are sensitive to cloud cover; resolution in public imagery limits methodology; high compute cost; high expertise burden; requires field perimeters

Note: Description of the pros and cons of different measurement methods for the detection of Demi-Lune adoption.

Table A.2: Treatment effects on measurement error

	(1)	(2)
	Extensive Margin	Intensive Margin
Panel A: Self-Report (Full Sample)		
Any treatment	0.00*	3.23***
	(0.00)	(0.51)
Mean of control	0.00	-0.20
Observations	2850	2850
R^2	0.01	0.01
Panel B: Self-Report (GPS Sample)		
Any treatment		2.77*
		(1.39)
Mean of control		-0.22
Observations		427
R^2		0.03
Panel C: Remote Observation (GPS Sample)		
Any treatment	-0.43***	-37.29***
	(0.07)	(12.10)
Mean of control	0.04	1.07
Observations	406	372
R^2	0.19	0.15
Panel D: SIML (GPS Sample)		
Any treatment	-0.41***	51.51*
	(0.10)	(25.16)
Mean of control	0.18	18.61
Observations	392	388
R^2	0.29	0.15

Note: This table shows the results from a regression of measurement error on a binary treatment variable and strata fixed effects, clustering the standard errors at the village level. For each panel, measurement error is calculated as $Y_i^M - Y_i^*$, where Y_i^M is either the measure of adoption from self-reports, remote observation or SIML, and Y_i^* is the measure of adoption from direct observation. Coefficient estimates are missing from Column (1), Panel B, as there was not enough variation in the dependent variable to estimate a regression. The sample sizes in Panel C varies slightly from those in Table 1, due to missing data from the direct observation measures.

Table A.3: Summary statistics and balance, GPS sample

	Full sample	GPS sample		
	Mean	Control mean	Treatment mean	Difference (p-val)
	(1)	(2)	(3)	(4)
Respondent is female	0.49	0.48	0.50	0.29
No. of adult household members	4.41	4.51	4.32	0.45
Owns a mobile phone	0.38	0.45	0.48	0.89
No. of hectares owned	3.90	3.68	4.54	0.66
No. of degraded hectares owned	1.81	1.64	2.00	0.49
Demi-lune experience	0.37	0.34	0.52	0.03
Number of obs	2,861	160	269	

Note: Means of household characteristics at baseline, in the full sample (column 1) and the GPS sample (Columns 2-4). The p-value on a test for equal means in the control and treatment groups is shown in Column 4. The table corresponds to Appendix Table A.3 in Aker and Jack (2025).

Table A.4: Implementation costs for each measurement approach

	Direct observation	Self-report	Remote observation	SIML
Fixed costs, one-time				
Staff training	300	350	100	20000
Data acquisition	0	0	450	450
Assumptions	Staff training	Staff training	Staff time	Staff time
Fixed costs, per-round				
Training data	0	0	1479	1479
Assumptions	N/A	N/A	Direct obs 427 fields	Direct obs 427 fields
Variable costs, one-time				
Geolocations	0	0	2.76	2.76
Assumptions	N/A	N/A	Staff time + transport	Staff time + transport
Variable costs, per round				
Farmer/field visits	2.76	2.5	0	0
Ground truth	0	0	1.56	0.25
Assumptions	Staff time + transport	Staff time + transport	Staff time	Compute cost

Note: Inputs to cost calculations and simulations. *One-time* refers to costs incurred once during the project. *Per round* refers to costs incurred during each round of data collection. Where possible, cost estimates are based on the RCT and/or measurement undertaken for this paper. In cases that no actual cost information is available, best approximations are used. Costs are in USD.

B Methods

B.1 Direct Observation

Data for the direct observation measure was collected by a team of highly-trained enumerators, who visited farmers' fields in May/June of each year of the RCT. They first asked the farmer whether he/she had adopted over a particular time frame. (For 2018, for example, the specific time frame referred to between January and June 2018). If the farmer had adopted, the enumerator visited the field and noted whether demi-lunes had been constructed and counted the total number of demi-lunes. The enumerator also noted whether the demi-lune adhered to quality norms (size, depth, positioning), and then noted a "quality-adjusted" number of demi-lunes. Those who did not report adopting did not receive a field visit, and hence the intensive margin of adoption for non-adopters was zero.

Direct observation visits were conducted for all 2,861 farmers across 180 villages for four years (2018-2021). Only 2018 direct observation data are used in this paper. There was very little attrition for these visits; attrition in 2018 was less than 0.5%, 1.5% in 2019, 0.9% in 2020, and 0.5% in 2021. 95% of households adopted on only one field.

B.2 Self-Reports

The survey self-reported measures were collected in a manner similar to that in the direct observation measures. A team of highly-trained enumerators visited farmers' fields in May/June of each year. They first asked the farmer whether he/she had adopted over a particular time frame, in our case, between January and June 2018. If the farmer reported adopting, the enumerator asked the number of demi-lunes adopted over the given time frame. If the farmer did not report adopting, they did not ask how many were adopted. As a result, the self-reported intensive margin of adoption for non-adopters was zero.

Since these self-reports were conducted at the same time, and during the same survey, as the direct observation visits, the surveys were conducted for all 2,861 farmers across 180 villages for four years. There was very little attrition for these visits; attrition in 2018 was less than 0.5%, 1.5% in 2019, 0.9% in 2020, and 0.5% in 2021.

B.3 Geocoordinate collection

Geocoordinates were collected from the RCT sample at several points during the research. The satellite-based measurement methods (remote observation and SIML) primarily rely upon the 2021 GPS sample, and the assumption that field perimeters did not meaningfully change between 2018 and 2021. We describe here all geocoordinate collection, including some measurements not used in this paper.

In 2019, during the direct observation visits, the enumeration team visited each field where households had reported adoption, regardless of treatment status. After noting adoption, the enumerator went to the center of the field and took the latitude and longitude of the location. This

yielded a sample of 2,297 field points, one for each household, and only for adopters.

In 2020, during the direct observation visits, the enumeration team visited each field where households had reported adoption, regardless of treatment status. After noting adoption, the enumerator went to the center of the field and took the latitude and longitude of the location. This yielded points on demi-lune fields for 2,377 households. 7% of households had adopted on more than one field, and so additional field points were taken. During the same 2020 field visit, enumerators also visited non-demi-lune fields, both degraded and non-degraded. Upon arriving at the field, the enumerator went to the center of the field and took the latitude and longitude of the location. This yielded points on non demi-lune fields. As a result, the 2020 dataset has field points for 2,817 households, both treatment and control, demi-lune and non-demi-lune.

In 2021, the research team identified 26 villages (from among the 180 villages), stratified by treatment status (known as the *GPS sample*). During the 2021 direct observation visits, the team collected three sets of coordinates in the GPS sample. First, they asked about any new demi-lune adoption (since January 2021) and visited those fields. On these fields, they collected field and adoption subfield perimeters (polygons). Almost all households only had one “new” demi-lune field. Second, enumerators then asked about demi-lune adoption between 2018 and 2021 (or “old” demi-lune adoption). They visited these “old” demi-lune fields and collected field and adoption subfield perimeters (polygons), for a total of 280 household observations. Approximately 90% of households had only adopted on one field, and hence had one observation. And third, enumerators then asked to visit non-demi-lune fields, first degraded and then non-degraded. They followed a similar protocol, collecting field perimeter data for non-demi-lune fields. Overall, this yielded a sample of geocoordinates for 429 households’ demi-lune and non-demi-lune fields, with no attrition. We primarily use the “old” demi-lune field perimeters for this paper, although the remote observation approach used other field perimeters to complete missing data.

B.4 Remote Observation

The remote observation methodology started with uploading the field perimeters for all households in the GPS sample. We used Google Earth Pro to visually inspect each field, starting in 2016 (before construction) and continuing each year, from 2018 to 2024. For each field, we noted the date that the image was available, the image quality (low, moderate or high), and whether or not demi-lunes were visible on the field for a given date (extensive margin). Most households’ fields had images available in October or November of 2018, approximately 5 months after the 2018 direct observation visit. A subset of villages only had images available in early 2019. Thus, if a farmer adopted in between June 2018 and June 2019, the remote observation approach would detect adoption, even though direct observation did not.

If the image quality was low, then no adoption could be detected and the observation was recorded as missing. This represented 2% of observations.

If the shapes of the demi-lunes were easily visible, we counted the number of observable demi-lunes

on that field on the entire field for a given date. If the shape of demi-lunes could be detected on a given field, but demi-lunes could not be accurately counted, demi-lunes were not counted and the intensive margin observation was recorded as missing. This was the case for 8% of observations.

A subset of households were visually inspected by two individuals to confirm the extensive and intensive margin. If there was a difference in the intensive margin numbers, an individual went back to the field to review and confirm the final number.

Since the data for the “old” demi-lune fields were only collected for those who had adopted since 2018, this meant that a number of households in control villages (or non-adopters in treatment villages) did not have field or adoption perimeters in the “old” demi-lune set of perimeters. In our analysis, we dealt with these in two ways. First, similar to the direct observation data, we imputed a zero value for household observation that had a missing field perimeter. This would therefore incorrectly attribute no adoption to a household who had adopted but had missing data, or no adoption to an image that may have been of low quality.

Since we had field perimeter data for all GPS households in 2021, we implemented a second approach, using all field perimeter data for the GPS households. If a given household did not have field perimeter data for “old” demi-lunes (as was the case for most of the control sample), we first inspected the field perimeters for “new demi-lune fields” (assuming that households would have adopted on such fields in 2018, if they had adopted), and then inspected images of farmers’ non-demi-lune fields. For each of these fields, we followed the same protocol as above: Observed the field perimeter in 2016 (before adoption) and again in 2018; noted the quality of the image in 2018; determined whether there was any adoption; and counted the number of demi-lunes. This process only yielded 12 “missings” in the GPS sample.

The results in this paper use data from the second approach, as this is more complete and does not require imputation. Nevertheless, both approaches yield similar results.

B.5 Satellite Imagery & Machine Learning

B.5.1 Data

For SIML, we downloaded the Sentinel-2 L1C Top-of-Atmosphere reflectance product at 10m resolution. We then applied an atmospheric correction¹ to create the equivalent of the L2A (Bottom-of-Atmosphere) imagery. In processing the imagery, we also applied a Bidirectional Reflectance Distribution Function (BRDF) correction and cloud mask. We downloaded scenes for every 10 day period for the years 2017 and 2018, then took monthly averages for each band. We then calculated the vegetation and water indices for each month in each year, listed below in Table B.1. Lastly we calculated the difference for each raw band and index for each month between 2018 and 2017. These pixel-level annual differences were included as input data to the random forest model.

¹Using the Sen2Cor command line program.

Individual pixels were extracted from the Sentinel imagery in one of three classes: (1) full demi-lune coverage, (2) partial demi-lune coverage, or (3) no demi-lune coverage. Pixels that were more than 97.5% contained by the enumerator-collected subfield polygons where demi-lunes were adopted are classified as (1) full coverage, while pixels in which some portion less than 97.5% is contained by the polygon are considered (2) partial coverage. Pixels that lie outside of the demilune subfield polygon but inside a farmer’s field are classified as (3) no coverage. The distribution of partial vs. full pixels within the subfield polygons containing demi-lunes is shown in Figure B.1.

For training data, we select only fields that had demi-lunes in 2018. To ensure we limit the impact of any error in the boundaries of the subfield polygons, pixels classified as (3) no demi-lune coverage were extracted from fields that reported to have no demi-lunes anywhere. Lastly, to construct the training dataset we sampled up to 60 pixels from each subfield demi-lune polygon and up to 15 pixels from the non-demi-lune field polygons. This sampling ensured a balanced dataset while still including pixels from every polygon, as there are more, larger non-demi-lune field polygons relative to the smaller subfield demi-lune polygons.

Of the 429 households originally included in the 2021 GPS sample, 14 were dropped due to missing or inadequate GPS data points; 17 were dropped because the household had built demi-lunes after 2018 and had no non-demi-lune fields; 5 were dropped because they did not have a subfield polygon; and 1 outlier field was dropped (a subfield polygon 5 times larger than any other). Thus the sample for the SIML analysis was 392 households.

B.5.2 Random Forest Model

We first selected the hyperparameters for the random forest model using a 10-fold cross validation. The number of trees, maximum tree depth, and minimum number of samples per leaf were determined by the best-performing model from this cross validation. We then partitioned the data into five 80-20 training-testing sets, in which each pixel was included in the test set in exactly one of the five partitions. We stratified these splits by farmer to ensure that all pixels associated with fields held by a given farmer would be grouped together as either training or testing in each partition. For each of these five partitions, we trained the model on the 80% training sample and tested it on the remaining 20% sample. For each farmer in the test set, we also made out-of-sample predictions for every pixel in the full field that contained the demi-lunes, as we had only included the pixels within the subfield demi-lune polygon in the original training dataset. This approach allowed us to make an out-of-sample prediction for every pixel in our dataset even with our relatively small sample size.

As our model predicts full, partial, or no coverage at the pixel level, we have to make assumptions to estimate the intensive margin (i.e., the number of demi-lunes a farmer built). To do so, we assume partial pixels to be 48% of a full pixel, as that is the mean of the distribution of partial pixels. We then summed the total number of pixels predicted to have demi-lunes across full and partial pixels for each

farmer. By assuming that 2.5 demi-lunes are built on each $100m^2$ pixel, following guidelines for proper demi-lune construction (250/ha), we calculated a predicted number of demi-lunes for each farmer.

B.5.3 Results

Table B.2 lists all the variables input into the model along with a t-test showing if the mean values of demi-lune pixels are statistically significantly different from the mean values of non-demi-lune pixels for each variable.² The table indicates significant differences between pixels with and without demi-lunes, meaning a model may be able to use these variables to classify pixels with and without demi-lunes. Though even with mean differences, there is still significant overlap in the distributions between the two groups, making the classification a non-trivial task.

The variation that the random forest model leverages in classification is shown in Figure B.2, which plots the difference-in-differences timeseries for two example indices. To construct this timeseries, we take the monthly average of all demi-lune pixels and all non-demi-lune pixels for each index. We then take the difference between these two sets for each month, then subtract the monthly 2017 differences from the 2018 differences. If the resulting value was zero, there would be no variation for the model to exploit in classification. However, as Figure B.2 demonstrates, the difference-in-differences timeseries points to variability between the demi-lune and non-demi-lune pixels and across years that the random forest model can leverage in classification.

Our main random forest model uses the categorical class labels, though we also run the model using a simpler binary classification without the partial pixel category (meaning partial pixels would be considered as having demi-lunes). Table B.3 reports the results of both models, at the pixel-level as well as aggregated to the farmer-level. To aggregate our results to the farmer level, we considered a farmer as having demi-lunes if any pixel in one of their fields was predicted to have any demi-lunes, either full or partially.

Our accuracy is much higher when we aggregate to the farmer level. This is largely because our pixel-level model has a high level of false negatives, as can be seen in the confusion matrix in Figure B.3. By aggregating up to the farmer-level and having the most generous rule for labeling a farmer as having any demi-lunes (i.e. any pixel with any demi-lunes), we reduce the number of false negatives, increasing overall accuracy.

B.6 Tables and Figures

²Only pixels that were fully contained by their respective polygons were included in this table (i.e. partial pixels were excluded)

Table B.1: Indices calculated from Sentinel data

Index	Full Name	Equation
EVI	Enhanced Vegetation Index (2-band)	$2.5 \cdot \frac{NIR-RED}{NIR+1+2.4RED}$
NDVI	Normalized Difference Vegetation Index	$\frac{NIR-RED}{NIR+RED}$
SAVI	Soil-Adjusted Vegetation Index	$1.5 \cdot \frac{NIR-RED}{NIR+RED+5}$
MSAVI	Modified Soil-Adjusted Vegetation Index	$\frac{2 \cdot NIR + 1 - \sqrt{(2 \cdot NIR + 1)^2 - 8 \cdot (NIR - RED)}}{2}$
SWIR1	Shortwave Infrared 1	$SWIR1$ (raw band value)
SWIR2	Shortwave Infrared 2	$SWIR2$ (raw band value)
NDWI	Normalized Difference Water Index	$\frac{GREEN-NIR}{GREEN+NIR}$
NDMI	Normalized Difference Moisture Index	$\frac{NIR-SWIR1}{NIR+SWIR1}$

Note: This table lists the indices calculated for each pixel from the Sentinel-2 data. The index, its full name, and its equation are included. We calculate monthly indices for 2017 and 2018, then the difference between the 2018 and 2017 monthly values.

Table B.2: Difference between demi-lune and no-demi-lune pixels

variable	mean_0	mean_1	mean_diff_1_0	t_stat	p_value
ndwi_05	0.105	0.325	0.220	-13.600	0.000
ndwi_07	-0.846	-0.676	0.170	-18.310	0.000
green_05	0.573	0.707	0.135	-7.029	0.000
ndwi_01	-0.142	-0.021	0.121	-41.151	0.000
red_05	0.360	0.472	0.112	-8.308	0.000
ndvi_10	0.145	0.246	0.101	-48.564	0.000
ndvi_09	-0.022	0.058	0.080	-27.869	0.000
ndvi_06	-0.160	-0.082	0.077	-27.525	0.000
msavi_06	-0.172	-0.097	0.075	-20.081	0.000
savi_06	-0.161	-0.088	0.074	-23.167	0.000
ndwi_08	-0.303	-0.230	0.073	-19.176	0.000
red_08	-0.186	-0.117	0.070	-18.991	0.000
savi_10	0.084	0.149	0.065	-49.482	0.000
evi2_10	0.087	0.151	0.064	-49.311	0.000
evi2_06	-0.140	-0.076	0.064	-23.724	0.000
green_08	-0.258	-0.194	0.064	-14.457	0.000
msavi_10	0.080	0.139	0.059	-48.726	0.000
savi_09	-0.000	0.057	0.057	-28.268	0.000
evi2_09	0.006	0.059	0.053	-27.745	0.000
msavi_09	0.006	0.057	0.051	-27.521	0.000
ndvi_12	-0.101	-0.062	0.039	-42.572	0.000
ndvi_02	-0.029	0.002	0.030	-16.315	0.000
red_01	-0.059	-0.029	0.030	-30.325	0.000
swir1_08	0.028	0.058	0.029	-29.699	0.000
swir2_08	0.034	0.063	0.029	-27.821	0.000
savi_12	-0.078	-0.053	0.025	-38.986	0.000
evi2_12	-0.075	-0.051	0.024	-39.152	0.000
swir1_12	-0.005	0.019	0.024	-38.839	0.000
ndwi_04	0.019	0.043	0.024	-10.454	0.000
ndwi_11	-0.006	0.017	0.023	-16.607	0.000
msavi_12	-0.073	-0.051	0.023	-37.909	0.000
green_01	-0.041	-0.019	0.021	-28.075	0.000
swir1_06	0.025	0.046	0.021	-17.591	0.000
swir2_12	-0.006	0.013	0.020	-30.422	0.000
swir2_06	0.019	0.037	0.018	-16.054	0.000
nir_12	0.005	0.022	0.017	-31.199	0.000
savi_02	-0.006	0.010	0.017	-26.039	0.000
evi2_02	-0.006	0.010	0.016	-26.167	0.000
nir_06	0.038	0.053	0.016	-14.355	0.000
msavi_02	-0.004	0.011	0.015	-26.727	0.000
nir_08	0.017	0.032	0.015	-19.043	0.000
ndmi_07	-0.004	0.005	0.008	-7.518	0.000
ndmi_05	-0.002	0.006	0.008	-10.167	0.000
nir_09	0.010	0.018	0.008	-9.027	0.000
swir2_01	0.034	0.041	0.007	-12.020	0.000
red_11	0.019	0.026	0.007	-9.859	0.000
green_11	0.030	0.036	0.006	-10.030	0.000
green_12	0.059	0.064	0.005	-10.675	0.000
ndmi_09	0.038	0.042	0.004	-6.348	0.000
swir1_01	0.007	0.010	0.003	-6.400	0.000
ndmi_10	0.004	0.007	0.003	-6.889	0.000
red_12	0.061	0.063	0.003	-5.691	0.000
ndvi_03	-0.025	-0.022	0.002	-6.146	0.000
evi2_03	-0.021	-0.020	0.002	-4.181	0.000
savi_03	-0.022	-0.020	0.002	-4.100	0.000
msavi_03	-0.021	-0.019	0.001	-3.858	0.000
swir1_03	0.036	0.034	-0.001	2.318	0.021
green_03	0.035	0.034	-0.001	3.764	0.000
nir_01	-0.034	-0.036	-0.002	4.364	0.000
swir1_02	0.026	0.025	-0.002	2.565	0.010
ndmi_06	0.017	0.015	-0.002	5.502	0.000
swir2_11	0.019	0.017	-0.002	3.192	0.001
ndmi_01	-0.039	-0.041	-0.002	9.459	0.000
ndmi_03	-0.002	-0.004	-0.002	12.579	0.000
ndmi_02	-0.004	-0.006	-0.002	10.345	0.000
nir_11	0.033	0.030	-0.003	5.002	0.000
nir_03	0.027	0.024	-0.003	7.963	0.000
swir1_11	0.026	0.023	-0.003	4.891	0.000
nir_02	0.017	0.014	-0.003	5.426	0.000
ndmi_11	0.014	0.010	-0.003	10.837	0.000
ndmi_12	0.009	0.006	-0.004	15.162	0.000

Continued on next page

Table B.2: Difference between demi-lune and no-demi-lune pixels

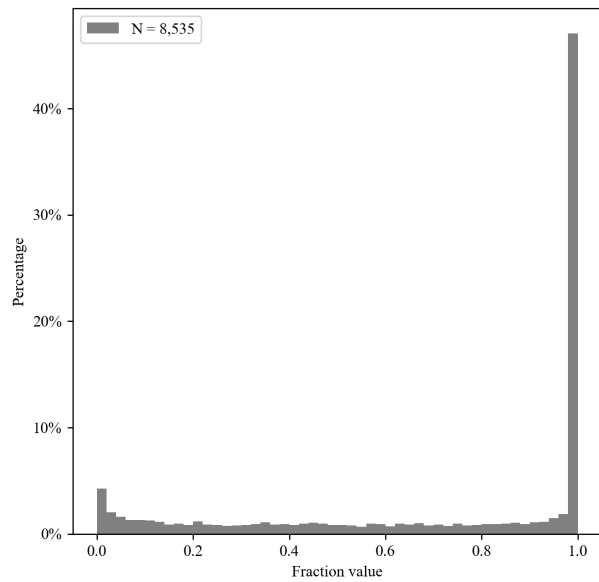
variable	mean_0	mean_1	mean_diff_1_0	t_stat	p_value
swir2_09	-0.012	-0.016	-0.004	4.028	0.000
red_03	0.044	0.040	-0.004	10.900	0.000
ndmi_04	-0.015	-0.019	-0.005	15.445	0.000
ndmi_08	-0.011	-0.020	-0.009	12.787	0.000
green_02	0.011	0.001	-0.010	18.274	0.000
swir2_07	0.059	0.046	-0.013	4.012	0.000
msavi_11	0.016	0.002	-0.014	16.545	0.000
evi2_11	0.016	0.002	-0.014	16.512	0.000
red_02	0.016	0.001	-0.015	22.441	0.000
nir_05	0.061	0.045	-0.016	9.280	0.000
savi_11	0.019	0.002	-0.017	17.691	0.000
evi2_07	0.533	0.509	-0.025	4.951	0.000
ndwi_02	0.023	-0.003	-0.026	15.454	0.000
ndvi_11	0.032	0.004	-0.028	19.600	0.000
msavi_04	-0.005	-0.033	-0.028	18.766	0.000
evi2_04	-0.002	-0.030	-0.029	19.560	0.000
swir2_05	0.082	0.053	-0.029	15.266	0.000
swir1_05	0.078	0.048	-0.030	14.023	0.000
red_09	-0.001	-0.032	-0.031	21.777	0.000
savi_04	0.001	-0.031	-0.032	20.977	0.000
ndwi_12	0.115	0.080	-0.035	35.350	0.000
green_09	-0.020	-0.058	-0.038	25.146	0.000
ndvi_04	0.015	-0.027	-0.042	24.582	0.000
evi2_08	0.182	0.131	-0.051	16.208	0.000
red_04	0.102	0.049	-0.053	29.962	0.000
msavi_08	0.198	0.143	-0.055	15.498	0.000
savi_08	0.196	0.140	-0.056	16.403	0.000
savi_07	0.649	0.592	-0.057	9.287	0.000
nir_10	-0.056	-0.113	-0.057	43.842	0.000
swir2_10	-0.054	-0.112	-0.057	44.659	0.000
green_04	0.121	0.061	-0.060	27.781	0.000
red_06	0.237	0.172	-0.065	12.645	0.000
swir1_10	-0.065	-0.134	-0.068	45.252	0.000
savi_01	0.075	0.007	-0.069	41.638	0.000
evi2_01	0.078	0.007	-0.070	42.089	0.000
msavi_01	0.079	0.007	-0.072	42.733	0.000
green_10	-0.075	-0.148	-0.073	52.090	0.000
red_07	-0.518	-0.592	-0.074	12.143	0.000
green_07	-0.566	-0.640	-0.074	11.418	0.000
nir_04	0.087	0.012	-0.075	52.719	0.000
ndvi_08	0.220	0.143	-0.077	22.434	0.000
swir2_04	0.118	0.038	-0.080	53.806	0.000
red_10	-0.090	-0.176	-0.086	50.824	0.000
swir1_04	0.128	0.041	-0.087	51.544	0.000
ndwi_09	-0.003	-0.095	-0.092	29.740	0.000
ndwi_10	-0.138	-0.239	-0.101	49.379	0.000
green_06	0.334	0.224	-0.109	13.593	0.000
ndwi_06	0.246	0.123	-0.123	29.979	0.000
ndvi_07	0.732	0.593	-0.139	16.832	0.000
evi2_05	-0.073	-0.218	-0.145	13.521	0.000
ndvi_01	0.179	0.031	-0.148	42.106	0.000
savi_05	-0.102	-0.258	-0.156	13.024	0.000
msavi_05	-0.146	-0.305	-0.159	12.217	0.000
ndvi_05	0.037	-0.178	-0.216	14.849	0.000
swir2_02	0.050	0.051	0.001	-1.245	0.213
ndwi_03	0.025	0.025	0.000	-1.075	0.283
swir2_03	0.054	0.054	-0.000	0.759	0.448
swir1_09	-0.024	-0.026	-0.002	1.812	0.070
nir_07	0.078	0.072	-0.006	1.772	0.077
swir1_07	0.082	0.075	-0.007	1.714	0.087
msavi_07	0.622	0.614	-0.008	1.300	0.194

Note: The difference in variables between pixels that entirely contain demi-lunes (1) and those that do not (0). Each variable used in the random forest model is listed in the first column, where the last two digits indicate the month. For example, ndvi_11 is the difference in the NDVI measure between 2018 and 2017 in the month of November. The second and third columns report the mean values of all pixels without (mean_0) and with (mean_1) demi-lunes. The fourth column reports the difference between columns two and three. The last three columns summarize the t-test that reports if the mean values of each group are statistically significantly different from each other.

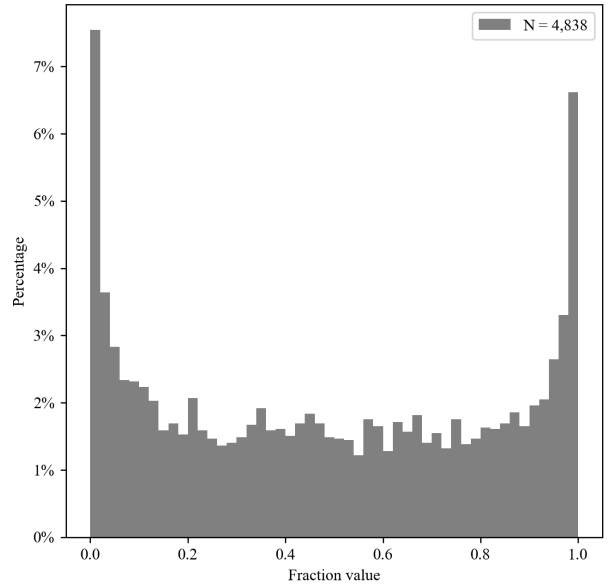
Table B.3: Summary of training and testing accuracies for DL subsets

	Training Classes	Accuracy Type	Accuracy
1.a	Categorical	Training accuracy	62.4%
1.b	Categorical	Training accuracy aggregated to binary classes	73.9%
1.c	Categorical	Out-of-sample accuracy	72.7%
1.d	Categorical	Out-of-sample accuracy aggregated to binary classes	74.2%
1.e	Categorical	Out-of-sample accuracy aggregated to farmer-level	77.9 %
2.a	Binary	Training accuracy	76.8%
2.b	Binary	Out-of-sample accuracy	68.7%
2.c	Binary	Out-of-sample accuracy aggregated to farmer-level	79.1%

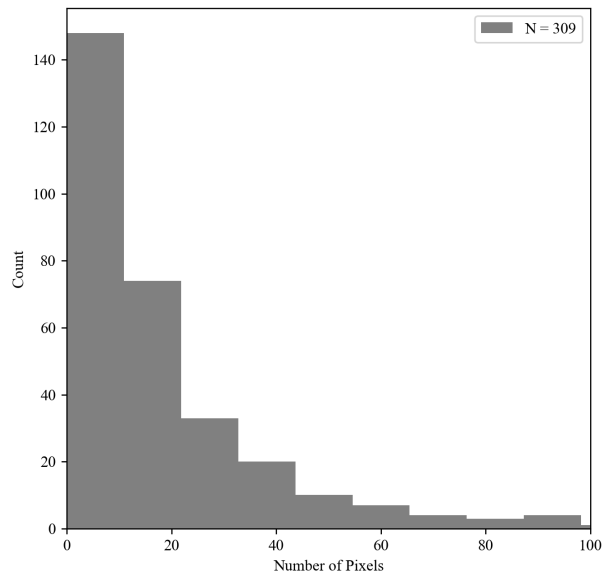
Note: This table summarizes the accuracies of the random forest model. Model (1) used categorical classes of full, partial, and no demi-lune coverage while model (2) used binary categories of simply any demi-lune coverage or no demi-unes. For the categorical (main) model results, we report the pixel-level training accuracy (1.a) and also aggregate the categories up to the simple binary classification (1.b), where partial pixels are classified as having demi-lunes. We report the pixel-level out-of-sample accuracy for all categories (1.c) as well as the aggregated binary classes (1.d). We then aggregate the out-of-sample results to the farmer-level (1.e), where a farmer is considered to have demi-lunes if any pixel in their fields is predicted to have any demi-lunes. Analogous accuracies are reported of the model trained on the simple binary categories (2.a, 2.b, 2.c).



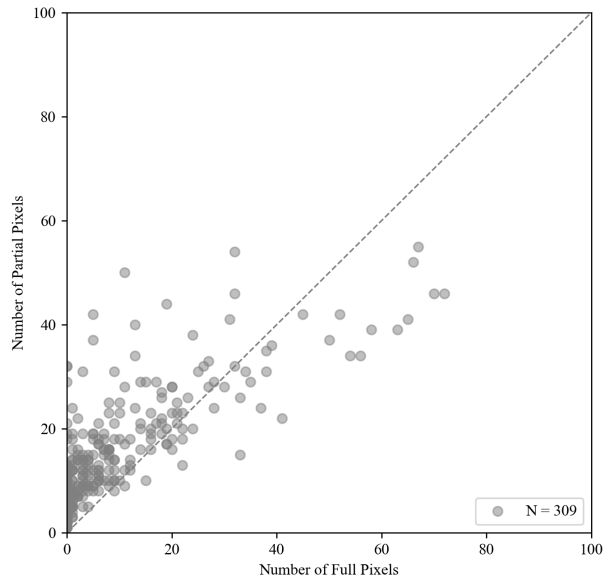
(a) Pixels in demi-lune subfield polygons



(b) Partial pixels in demi-lune subfield polygons



(c) Number of pixels per demi-lune subfield polygon



(d) Pixel types in each demi-lune subfield polygon

Figure B.1: Distribution of pixel types within the demi-lune subfield polygons.

Note: Panel (a) shows that among the 8,522 pixels contained by the 309 polygons included in the analysis, roughly 50% are entirely contained by their respective polygon while the remainder are only partially contained. Panel (b) shows that the distribution of the fractional pixels is fairly continuous between 0% and 100%, with mass clustered at the two end members. Panel (c) shows that most polygons are fairly small, with the majority containing < 50 pixels. Lastly, panel (d) shows that most polygons have more partial pixels than full pixels, particularly among smaller polygons.

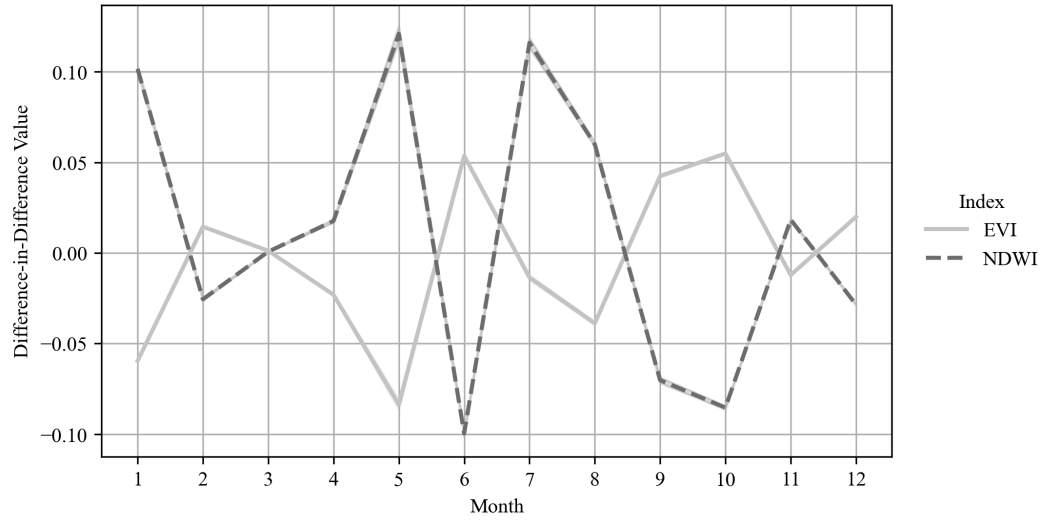
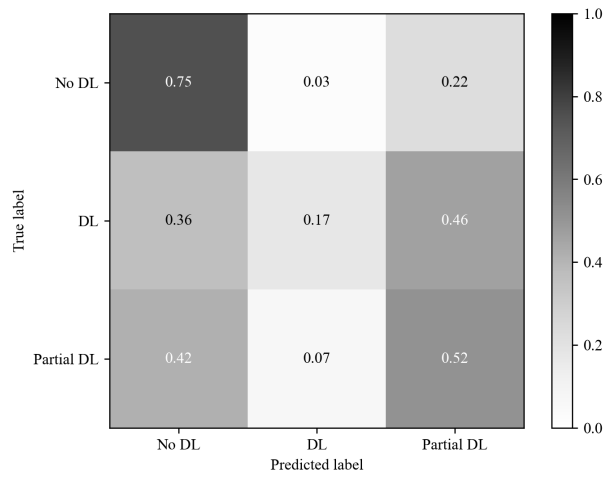
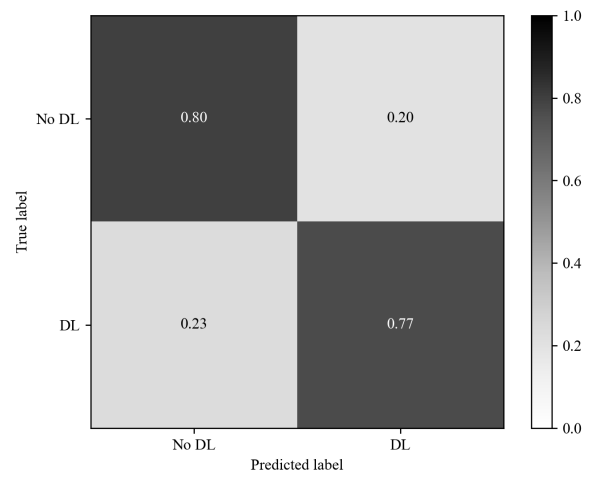


Figure B.2: Difference-in-Differences Timeseries

Note: This figure shows the difference-in-differences of pixels with demi-lune and those without non-demi-lunes between 2018 and 2017 for two example indices. For each month, the mean value of pixels from all non-demi-lune polygons was subtracted from the mean value of all pixels in demi-lune polygons. The 2017 monthly differences were then subtracted from the 2018 differences to get the difference-in-differences value plotted here. These monthly difference-in-differences values are the input data for the random forest model.



(a) Pixel-level model result



(b) Farmer-level model result, aggregated classes

Figure B.3: Confusion matrices summarizing out-of-sample result

Note: Panel (a) shows the confusion matrix of the pixel-level random forest model for the out-of-sample prediction, which had an accuracy of 48%. There are a large number of false negatives at the pixel level. Panel (b) shows the results aggregated to the farmer-level, and labels aggregated to simple “DL” (having demi-lunes) and “No DL” (not having demi-lunes). Farmers are classified as having demi-lunes if any pixel in their fields is classified as either fully or partially having demi-lunes.