Radiant MLHub
EARTH IMAGERY FOR IMPACT

Dataset Documentation

Dataset Name: PlantVillage Kenya Ground Reference Crop Type Dataset

Location and boundaries

Overall Location Method

- ☐ Ground collection only
- ☒ Ground collection with boundary drawn using imagery
- ☐ Ground collection with spatial buffer added
- ☐ Boundary drawn from imagery
- ☐ Other _____
- ☐ Unknown

GeoLocation Device

- ☐ Industrial grade GPS (List model) _____
- ☐ Retail grade GPS
- ☐ Mobile Phone GPS
- ☐ N/A
- ☒ Unknown

Ground Boundary Method (Details explained in Appendix A)

- ☐ Live/Continuous point capture of walk-around
- ☒ Manual point capture of walk-around
- ☐ Manual point capture of polygon boundaries (not whole field)
- ☐ Manual point capture for later image annotation
- ☐ Manual point capture for spatial buffer within field
- ☐ Manual point capture while looking at but not in field, with heading recorded
- ☐ Other _____
- ☐ Unknown

Imagery used (Skip if no imagery used)

Sensor: Unknown

Date(s):

List scenes used in Appendix B

Imagery Annotation methods

- ☐ Boundaries drawn based on a single ground point captured
- ☒ Boundaries drawn/edited based on multiple ground points captured
- ☐ Buffer validated from ground point captured
- ☐ Boundary drawn without ground reference data (Include description of methods in Appendix C)
- ☐ Pixels annotated without ground reference data (Include description of methods in Appendix C)
- ☐ Unknown

Radiant MLHub
EARTH IMAGERY FOR IMPACT

Boundary inclusion

- ☒ Captured polygon includes the entire field/area
- ☐ Captured polygon includes only a sample of the field/area

Classification

Classification Type

- ☐ Land cover
- ☒ Crop type
- ☐ Other _____

Classes/fields used

Describe in Appendix D

Ground Referenced Classification

- ☒ Observation (Describe methods of determination in Appendix E)
- ☒ Survey/interview with land holder (Describe methods in Appendix E)
- ☐ Other (Describe methods in Appendix E)

Image Referenced Classification

Describe methods used in Appendix C

Data Properties

Property name	Property Description	Parameters/Allowed responses (optional)
Field ID	A 3 part field ID explained in the documentation	See Appendix F
Latitude	Latitude of the point where the surveyor collects data	
Longitude	Longitude of the point where the surveyor collects data	
Accuracy	Accuracy of the GPS Latitude and Longitude of the point where the surveyor collects data	
Survey Date	Date the survey was completed	YYYY-MM-DD
Water Resource	Rainfed vs Irrigated	Rainfed, Irrigated
Planting Date	Date the field was planted	YYYY-MM-DD
PlantingDate Method	Indicates if the planting date is "Recorded" or "Estimated"	Recorded, Estimated
Estimated Harvest Date	Harvest dates are not recorded for this dataset, and are estimated using the Planting Date and a common growing season length from FAO crop calendar	YYYY-MM-DD

Crop1	Major crop in the field	
Crop2	Other crops in the field	
Crop3	Other crops in the field	
Crop4	Other crops in the field	
Crop5	Other crops in the field	
Crop Density	Estimated density of the field	See Appendix F
Variety	Crop species variety	
CMD Rating	Cassava Mosaic Disease Rating, range [0, 10]	See Appendix F
CBSD Rating	Cassava Brown Streak Disease Rating, range [0, 10]	See Appendix F
CGM Rating	Cassava Green Mites Rating, range [0, 10]	See Appendix F
Disease Rating	Rating for any other disease that could be present, range [0, 10]	See Appendix F

Appendix A: Describe the method of geographic ground data collection

PlantVillage app is used to collect multiple points around the field and collectors have access to basemap imagery in the app during data collection. They use the basemap as a guide in collecting and verifying the points.

Post ground data collection, Radiant Earth Foundation conducted a quality control of the polygons using Sentinel-2 imagery of the growing season as well as Google basemap imagery. Two actions were taken on the data 1) several polygons that had overlapping areas with different crop labels were removed, 2) invalid polygons where multiple points were collected in corner of the field (within a distance of less than 0.5m) and the overall shape was not convex, were corrected.

Appendix B: List imagery scenes used for annotation (ideally also included in metadata)

Appendix C: Describe how boundaries and classes were determined without ground reference data

Appendix D: List all top-level classes or the classification guidance used

Crops were identified using FAO list, and include:

Banana, Bean, Cabbage, Cassava, Cowpea, Fallowland, Groundnut, Maize, Millet, Sorghum, Soybean, Sugarcane, Sweetpotato, Tomato

Appendix E: Describe methods for determining classes based on direct/ground observation

Farmers were asked when available, otherwise observations of the field collector.

Radiant MLHub
EARTH IMAGERY FOR IMPACT

Appendix F: Detailed description of the dataset properties

Field ID: This has a 3 part number (#1.#2.#3):

#1 refers to the lead farmer working with PlantVillage. For example, Josephine is Lead Farmer 1. Her number would look like 1.#.#.

#2 refers to the number field the collector is at. For example, if the 4th field of the day is being surveyed with Josephine; the ID would be 1.04. If there is no third number that means all the information is known about this field.

#3 refers to if the field is an adjacent field. An adjacent field is constituted by being a field in the area but little to none of the information is known. Typically fields with a 3rd number just have a polygon and a crop.

Crop Density: Estimated density of the field, based off of the image below:

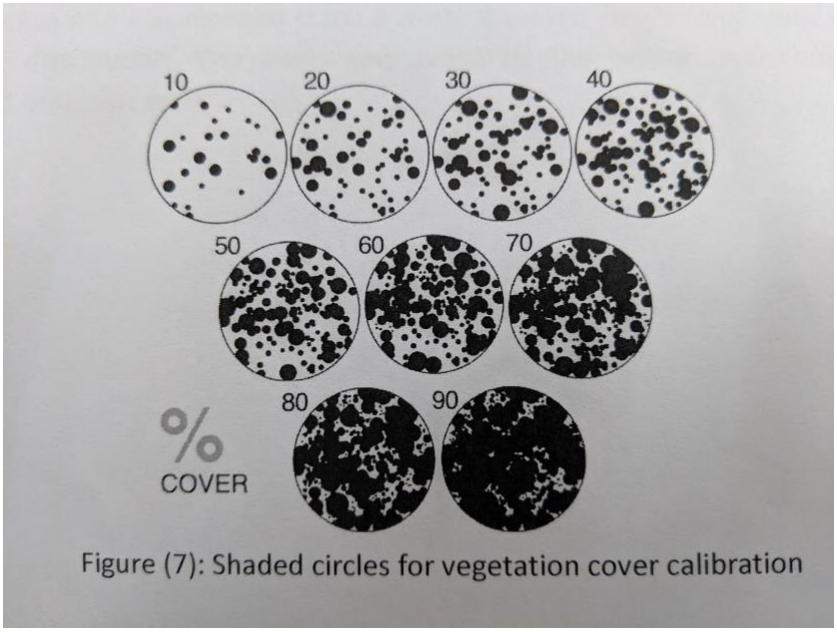

Figure (7): Shaded circles for vegetation cover calibration

Disease Ratings: There are different categories for collecting disease data. This measurement is a rating from 0-10. 0 is there is no disease present in the field, 10 being the disease is found in every plant. PlantVillage app allows collectors to scout the cassava or maize fields for the percentage infected for each disease and then input those results in the ODK categories

CMD(Cassava Mosaic Disease), CBSD(Cassava Brown Streak Disease), CGM(Cassava Green Mites) are all for cassava. These are ratings from 0(no disease presence) to 10 (every plant diseased)

Disease Prevalence is the last option which is meant for any other disease that could be present in fields other than cassava and maize. You will have to complete your own manual survey of the field and then mention what disease was prevalent in the comments.