

Dataset Documentation

Dataset Name: Benin - Cashew plantation identification training data set

Location and boundaries

Overall Location Method

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

GeoLocation Device

- Industrial grade GPS (List model): _Garmin eTrex 10_
- 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: Airbus Pléiades, 50cm

Date(s): Captured between 15. and 29. March 2020

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

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
Class_name	Class identifier for pixels belonging to classified shape layers. Identifies position and area size of each shape layer.	Class 1 = Well-managed plantation Class 2 = Poorly-managed plantation Class 3 = Non-plantation Class 4 = Residential Class 5 = Background Class 6 = Uncertain

Appendix A: Describe the method of geographic ground data collection

1. Based on available geographical data for Benin, including 40cm resolution aerial imagery captured between 2014 and 2015, a 1000km² area in the centre of the country was chosen as a sample area.
2. Within this sample area, potential cashew plantations were first identified via available imagery, cartographic sources as well as farmer registries.
3. Then a ground mapping team was dispatched to collect cashew plantation polygons and geographic reference points, using a Garmin eTrex10 device.

4. Out of the 1000km² sample area, a 120km² refinement area was then chosen to create the final, refined data labels.
5. On this 120km² area, using 50cm Airbus Pléiades imagery (captured between 15. and 29. March 2020), the ground data labels were corrected and re-classified into the six classes listed above.
6. This step was necessary to include whole plantations and parts of plantations that featured well-managed and well-aligned cashew trees within a “well-managed plantations” class. Plantations only sparsely or randomly populated by cashews or plantations with a great variety of tree species next to cashew were included in a “poorly-managed plantation” class.

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

- Airbus Pléiades 50cm: captured between 15. and 29. March 2020

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

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

The six classes are defined as follows:

Class 1 - Well-managed plantation	Predominantly well-aligned lines of trees, with regular canopy spacing, only limited areas of closed canopies or random planting. Only little difference in tree / species appearance.
Class 2 - Poorly-managed plantation	Predominantly randomly aligned trees, with partially interlocked canopies and with differences in tree / species appearance noticeable.
Class 3 - Non-plantation	Tree-agglomerations that do not show cashew texture features and are usually characterized by dark green color and small crown size.
Class 4 - Uncertain	Trees that are difficult to assign to a class.
Class 5 - Residential	Urban features and at times large rock formations.
Class 6 - Background	All other regions.

Appendix F: Detailed description of the dataset properties

Area: 120km²

Location: Benin, West Africa

File format: GeoJSON

Data hosting: AWS

DOI: