

Dataset Documentation

Dataset Name:

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: Landsat 5-8

Date(s): 2008 - 2018

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 (optional)
sampler	Institution leading the field campaign	
country	Country of sampling	
region	Geographic region of sampling	
date	Date of sampling campaign	yyyy-mm-dd
year	Year of sampling campaign	
label_1	Crop type; double cropping is labeled as CROP1-CROP2 (e.g. "wheat-other")	
label_2	Phenological class	
area	Polygon area in m2	

Appendix A: Describe the method of geographic ground data collection

Polygons were drawn from the GPS points collected on the field through image interpretation, relying on multi-temporal, very high-resolution satellite imagery from Google Earth (GE). The drawn polygons avoid non-vegetated areas, and samples with an unclear land cover (e.g. mixed crops), were excluded, as well as those with no valid GE Data due to the lack of observations or heavy cloud-cover.

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

All available collection 1 Landsat Data (i.e. Landsat 4, 5, 7 and 8) within a buffer of 2 months of each sampling year (i.e. 2008, 2011, 2015, 2016, 2017 and 2018) were used in the evaluation of crop type labels.

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

The top level class scheme (referred to as "label_2" in the present dataset) lists phenological classes, highlighting differences in the growth patterns of different crops and certain management practices:

- "winter"
- "summer"
- "double" (cropping)
- "permanent"
- "fallow"
- "unclear" (i.e. no distinguishable phenological cycle)

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

The classification of samples into different crop types was performed through visual interpretation on the field. Then, the validity of the assigned classes was tested against per-field, high-resolution, equidistant time-series of Normalized Difference Vegetation Index (NDVI). This step helped identify inconsistencies in class assignments in light of expert and data-driven knowledge on the phenological behavior of individual crop types.

Include any additional information/extra space as Appendix F+

For details on the methodology, please consult the publication dedicated to this dataset (<https://www.nature.com/articles/s41597-020-00591-2>).