Appendix

Data descriptors

Table 1: Master database column entries. We report the label used as provided in the dataset (csv file) and a brief description of the data represented in the table column.

label	description
report_id	unique image based identifier (assigned per season)
farmer	unique farmer id number
project_id	project id number (1 fand 2 for the first and second growing season respectively)
field	field id number (when multiple fields are acquired by a farmer)
site	unique site id
season_id	season id – defined as year_month-of-first-sowing
lat	decimal latitude
lon	decimal longitude
date	date of image acquisition (YYYY-MM-DD)
time	time of image acquisition (HH:MM:SS)
image	filename of the image as provided in the provided image dataset
r_dn	red digital number
g_dn	green digital number
b_dn	blue digital number
rcc_90	90th percentile of the red chromatic coordinate
gcc_90	90th percentile of the green chromatic coordinate
smooth_gcc_90	smoothed and normalized estimates of field based Gcc values
grvi_10	10th percentile of the green red vegetation index
spatial_location	the spatial location, either a town name or a grid cell index number
spatial_unit	the spatial aggregation used, either
growth_stage	wheat growth stage as assigned by human experts
lodging_labels	lodging labels as assigned by human experts
seed_variety	farmer reported seed variety
dam_rain	farmer reported rain damage
dam_hail	farmer reported hail damage
dam_high_temp	farmer reported temperature damage (heat)
dam_low_temp	farmer reported temperature damage (cold)

Table 1: Master database column entries. We report the label used as provided in the dataset (csv file) and a brief description of the data represented in the table column. *(continued)*

label	description
dam_pest	farmer reported pest
dam_lodging	farmer reported lodging
dam_wildlife	farmer reported wildlife damages
dam_fire	farmer reported fire damages
dam_unclassified	farmer reported damages, unclassified
man_harvest	farmer reported management, harvest
man_irrigate	farmer reported irrigation
man_till	farmer reported tilling
man_sow	farmer reported sowing
man_weed	farmer reported weeding
man_urea_kg_acre	farmer reported application of urea (kg / acre)
man_dap_kg_acre	farmer reported application of dap (kg / acre)
man_potash_kg_acre	farmer reported application of potash (kg / acre)
man_zinc_kg_acre	farmer reported application of zinc (kg / acre)
man_fungicide_kg_acre	farmer reported application of fungicide (kg / acre)
man_herbicide_kg_acre	farmer reported application of herbicide (kg / acre)
man_pesticide_kg_acre	farmer reported application of pesticide (kg / acre)
man_unclassified	unclassified management intervention
soil_type	soil type (loam,clay loam,loam sandy,clay or sandy)
drainage	soil drainage (good or poor)
sowing_date	date of sowing (YYYY-MM-DD)
harvest_quantity	harvest quantity from crop cutting ()
yield_expectation	farmer estimated yield at time of assessment ()
duration	duration of the growing season (based upon acquired image data)
nr_values	nr of processed images
spread	mean time difference between images
qa	ratio of the nr of values (images) and their spread

Table 2: Derivated summary database column entries. We report the label used as provided in the summary dataset (csv file) and a brief description of the data represented in the table column.

label	description
project_id	project id number (1 fand 2 for the first and second growing season respectively)
spatial_unit	the spatial aggregation used, either
spatial_location	the spatial location, either a town name or a grid cell index number
lat	decimal latitude
lon	decimal longitude
nr_fields	number of fields in a region
nr_farmers	number of farmers in a region
man_mean_date	mean manipulation date (across all interventions)
man_sd_date	standard deviation of the manipulation date (across all interventions)
nr_fields_irrigated	number of fields irrigated
nr_fields_weeded	number of fields weeded
nr_fields_tilled	number of fields tilled
nr_fields_sowed	number of fields sowed
nr_fields_harvested	number of fields harvested
nr_fields_urea	number of fields on which urea was applied
nr_fields_dap	number of fields on which dap was applied
nr_fields_potash	number of fields on which potash was applied
nr_fields_zinc	number of fields on which zinc was applied
nr_fields_fungicide	number of fields on which fungicide was applied
nr_fields_herbicide	number of fields on which herbicide was applied
nr_fields_pesticide	number of fields on which pesticide was applied
mean_urea_kg_acre	mean urea applied across the region (kg / acre)
sd_urea_kg_acre	standard deviation on the mean urea applied across the region (kg / acre)
mean_dap_kg_acre	mean urea applied across the region (kg / acre)
sd_dap_kg_acre	standard deviation on the mean urea applied across the region (kg / acre)
mean_potash_kg_acre	mean urea applied across the region (kg / acre)
sd_potash_kg_acre	standard deviation on the mean urea applied across the region (kg / acre)
mean_zinc_kg_acre	mean urea applied across the region (kg / acre)
sd_zinc_kg_acre	standard deviation on the mean urea applied across the region (kg / acre)
mean_fungicide_kg_acre	mean urea applied across the region (kg / acre)
sd_fungicide_kg_acre	standard deviation on the mean urea applied across the region (kg / acre)

Table 2: Derivated summary database column entries. We report the label used as provided in the summary dataset (csv file) and a brief description of the data represented in the table column. *(continued)*

label	description
mean_herbicide_kg_acre	mean urea applied across the region (kg / acre)
sd_herbicide_kg_acre	standard deviation on the mean urea applied across the region (kg / acre)
dam_mean_date	mean date of sustained damage (across all classes)
dam_sd_date	standard deviation on the mean date of sustained damage (across all classes, in days)
nr_fields_dam_rain	number of fields damaged by rain
nr_fields_dam_hail	number of fields damaged by hail
nr_fields_dam_high_temp	number of fields damaged by high temperature
nr_fields_dam_low_temp	number of fields damaged by low temperature
nr_fields_dam_pest	number of fields damaged by pests
nr_fields_dam_lodging	number of fields damaged by lodging
nr_fields_dam_wildlife	number of fields damaged by wildlife
nr_fields_dam_fire	number of fields damaged by fire
nr_fields_dam_unclassified	number of fields with unclassified damage
mean_qa	mean quality assurance value
mean_spread	mean time difference between images across a region
mean_nr_values	mean nr of processed images per field across a region
total_nr_values	mean total number of images per field across a region
mean_duration	mean duration of the season per field across a region (in days)
mean_sowing_date	mean sowing date by field across a region (YYYY-MM-DD)
sd_sowing_date	standard deviation of the sowing date by field across a region (in days)
rising	phenophase of the rising part of greenness curve (73% threshold, tillering phase, in YYYY-MM-DD)
falling	phenophase of the falling part of greenness curve (83% threshold, heading phase, in YYYY-MM-DD)
rising_lower_ci	phenophase confidence intervals (YYYY-MM-DD)
rising_upper_ci	phenophase confidence intervals (YYYY-MM-DD)
falling_lower_ci	phenophase confidence intervals (YYYY-MM-DD)
falling_upper_ci	phenophase confidence intervals (YYYY-MM-DD)

Table 3: Derivated time series summary database column entries. We report the label used as provided in the time series summary dataset (csv file) and a brief description of the data represented in the table column. The time series summary data can be merged with the summary dataset for quality control.

label	description
project_id	project id number (1 fand 2 for the first and second growing season respectively)
spatial_unit	the spatial aggregation used, either GADM or WorldClim 2.5 or 5 degree units
spatial_location	the spatial location, either a town name or a grid cell index number
lat	decimal latitude
lon	decimal longitude
date	date of reported Gcc values (YYYY-MM-DD)
smooth_gcc_90	Normalized and smoothed Gcc for a given spatial unit and spatial location
smooth_gcc_90_lower_ci	Lower confidence interval of the normalized and smoothed Gcc
smooth_gcc_90_upper_ci	Upper confidence interval of the normalized and smoothed Gcc
rising	transition date (at 73% threshold) of the rising part of the normalized and smoothed Gcc curve (per spatial location)
falling	transition date (at 83% threshold) of the falling part of the normalized and smoothed Gcc curve (per spatial location)
rising_lower_ci	uncertainty intervals on the rising transition date, based upon the lower CI
rising_upper_ci	uncertainty intervals on the rising transition date, based upon the upper CI
falling_lower_ci	uncertainty intervals on the falling transition date, based upon the lower CI
falling_upper_ci	uncertainty intervals on the falling transition date, based upon the upper CI