CROPYIELD

Task: Sentinel-2 Level 2A product parcel cropping and processing to an Analysis Ready Format.

Tentative algorithm

1. Set path to local data Directory (Sentinel-2 L2A tiles)
2. Read the parcel shapefile (columns: parcel_ID, farm_ID, plant_ID, startDate, endDate, geometry (Polygon)) into dataSet (geopandas). parcel_ID is a unique key.
3. Read the kml file for Sentinel-2 tiles over Finland
4. Create input set: EITHER (for each parcel_ID (row) in dataSet) OR (for each farm_ID where plant_ID == {value}, -> multiple polygons: a subset or MultiPolygon object)
    a. check within which tiles it goes
    b. Select all those tiles from Directory
    c. For each image file under a tile between startDate - endDate,
        Crop [‘geometry’] pixels all_touched=FALSE.
        For each band b (parameter list b, e.g. b = [2, 3, 4, 5, 6, 7, 8, 8A, 11, 12]),
            From pixel values, remove extreme values (close to zero or close to maximum),
            If CLOUDMASKING == TRUE, mask pixels having a cloud value(s), Save the length of the list of remaining values (count),
            Make an p bin histogram (percentile). Parameter p (e.g. p = 4).
    d. Save vectors to a numpy array (e.g. if p = 4 -> band, parcel_ID OR farm_ID+plant_ID, day-of-year (datatake sensing time), value1, value2, value3, value4, count)
    e. Save metadata to e.g. data frame (parcel_ID OR farm_ID+plant_ID, year, day-of-year, name of the file (tile), mission ID (SA|SB), count)
5. Save numpy array to a file .npy
6. Save metadata to a file .csv

- dstack numpy array
- First try saving to file an array per tile. If the size of the array is moderate, let’s consider packing all tiles to one array and one file.

Deliverables

First version -> testing -> finetune -> testing -> final version

- Source codes including Dockerfile.
- The structure of files and codes should follow the best practices.
- Delivered as an upload/clone to Luke's GitHub repository (github.com/lukefi).
- Author and license (CC BY 4.0) of the code should be written in the header of the code file.

Preferred timetable

- First version: 9/2019
- Final version 4/2020