**Survey data**

Outputs are mostly shared in shapefile (.shp) and data frame (.csv) formats.

**Uga\_study\_area\_poly** – this file is related to the actual study sites polygons. Please note they could vary in size. Typically, we were sampling 4 ~100x100m quadrants, however in some areas (esp. urban sites) we limited this to two or one ~100x100m quadrant. ‘field\_id’ is the unique key identifying study site number in this and following files. I have imported to those polygons information that we collected about the study site at the start of the survey.

**KLH\_Uga\_inhabited\_pts** – this file is related to a number of individual buildings that were inhabited within the survey site. Please note there are three categories: “Inhabited” is related to a 1-storey building and remaining to either 2 or 3 story buildings as seen in ‘Subcategor’ column.

**KLH\_Uga\_individual\_plants\_poly** – this is information related to scattered individual plants that we were mapping in our study areas. I assumed a 1-metre buffer around each plant and calculated areas of those buffers. You will find this area calculated in ‘area\_m2’ column. Link to study site is in ‘field\_id’ column.

**KLH\_Uga\_cassava\_fields\_poly** – the largest output containing information about cassava fields sizes and properties. There is one caveat here. Some polygons during the surveys were overlapping. This is mainly due to two reasons: sometimes two teams were running outside their designated study areas and in result could overlap two fields. The second reason is that the GPS accuracy could not be great in some locations. I solved it in the following way. I extracted those intersecting areas and created new polygons out of them. You will see columns with ‘\_Merged’ suffixes. This is where I concatenated information from overlapping polygons, and this will need to be probably averaged or split in half per polygon in the analysis. Some minor cleaning might be needed in this big file. I noticed that sometimes in the comments area there is information that should be placed in intercrop column.

**KLH\_Uga\_study\_area\_centre\_sample.shp/.csv -** They represent calculated centroids of the study areas (coordinates are found in the Centre\_Lon and Centre\_Lat columns). The **field\_id** column indicates study site numbers. The last four columns represent a value of the 4 raster layers in those centroid point locations. These are rasters that we would like to compare the collected data to:

* **Cass\_Prod** – cassava production per approx. sq km in tonnes for 2014 (my model). Max allowed value is 1000tonnes/1 sq km
* **Cass\_HA** - cassava harvested area per approx. sq km in tonnes for 2014 (my model). Max allowed model value is 50h/1 sq km
* **SPAM2010** – cassava production as predicted by the MapSPAM2010 model (tonnes per 100 sq km) for 2010
* **Population** – population number per 1 sq km as predicted by LandScan2014 (for 2014)

**Analyses**

**1\_alignData.R** – Code to align data inputs, create weightings and calculate the cassava production area.

**2\_alignRasterOutputs.R** – Aligns the data extracted to include additional buffer data layers. The data layers used are around the sampling points of the survey and they contain: 1. Cassava production; 2. Harvest Area; 3. Rural population in 2014; 4. settlements in 2018 extracted at 2km, 5km and 10km buffers.

**3\_modelVsSurvey.R** – This script compares modelled data in disaggregation models of cassava (CassavaMap and MapSPAM versus sampled data through the survey in Côte d’Ivoire and Uganda.

**4\_spatialTrends.R** – This script looks at large-scale spatial trends through the raster buffers.

**5\_surveyResults.R** – This script analyses the survey data.

**6\_modelVsSurvey\_baseline.R** – Makes a preliminary analysis of the data obtained through the disaggregation models and the survey.

**99\_figures.R** – This script summarises the results of the analyses by producing the figures for the manuscript.

**analysis\_aggregatingBuffers.R** and **analysis\_aggregatingModel.R** – These two scrips analyse whether aggregated predictions of cassava production correlate better using only the survey or the extracted buffer zones.

**country\_comparisons.R** – compares results between Côte d’Ivoire and Uganda