Protocol to assess spatial accuracy in case of data type FieldObservationSurvey focussing on crop land and arable crop types[1]

Level 1: Assess spatial accuracy of datasets (point or polygon) by benchmarking against non-arable spatial context features such as infrastructure

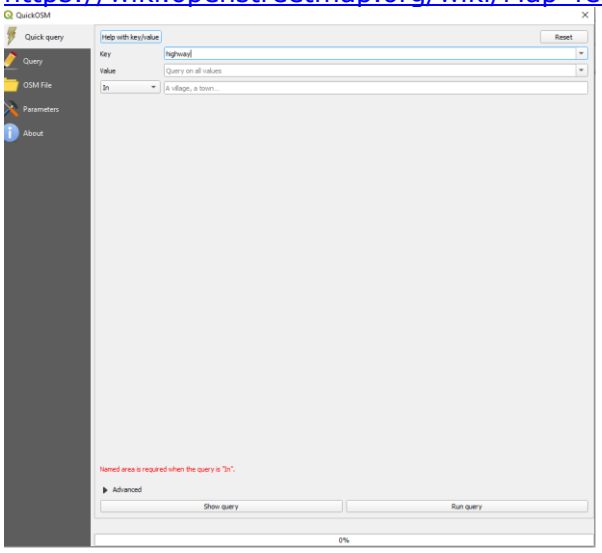
Step 1: Search and Download the country specific datasets (e.g., roads, water bodies, railway, buildings, nature areas etc.)

Step 1.1: Following are useful links to download country specific datasets:

- Country specific data: https://www.diva-gis.org/gdata
- Planet monthly base maps tropical 64 countries https://www.planet.com/nicfi/
- Freely available geographic datasets https://freegisdata.rtwilson.com/

Step 1.2: If country specific data is not present then extract required layers from openstreetmap (OSM). Here are the steps to get OSM layers.

- Install QuickOSM plugin
- Perform QuickOSM query to get the features (such as buildings, roads, rivers, railways, nature etc.). Detailed about OSM features can be explored on this link https://wiki.openstreetmap.org/wiki/Map_features#Natural



Step 1.3: Define the correct WGS UTM projection (e.g. WGS_1984_UTM_Zone_37S for Tanzania). The UTM system allows the coordinate numbering system to be tied directly to a distance measuring system.

[1] Thus, non-agricultural lands are excluded from these checks and actually are used as reference data to check the correctness of data

In case of point observations:

Step 1.4a: Perform a spatial distance analysis with point or line features (buildings, roads, railways, rivers) and reject the point observations which are less than 20 m from these features.

Step 1.4b: Perform an intersect analysis with polygon features (nature areas, water bodies etc.) and reject point observations which falls in this intersect.

In case of polygon observations:

Step 1.4: Perform an intersect analysis with point, line and polygon features (nature areas, water bodies, houses etc.) and reject polygon observations which intersect.

Step 1.5: Save cleaned data and register the number of rejected cases.

Level 2: Visual interpretation of selected samples using high-resolution satellite

Step 2.1: Select 5% of the cleaned input dataset from level 1 (use 10% if datasets have less than 200 features); use the "Subset Features" tool to select the features randomly.

Step 2.2: Load the high resolutions base maps (tutorial: add the base layer Map)

Step 2.3: Check all sampled observations against non-arable land covers (by interactively checking other land covers in the same area).

Step 2.4: Reject suspicious cases

Step 2.5: Register the number of suspicious cases. Based on your observation, select one of the best fit case scenario for average confidence score calculation

Case 0: Evaluated samples of cleaned data show no issues

Case 1: Evaluated samples of cleaned data show issues (between 1-10%)

Case 2: Evaluated samples of cleaned data show issues (between 10-25%)

Case 3: Evaluated samples of cleaned data show issues (between 25-50%)

Case 4: Evaluated samples of cleaned data show many issues (>50%)