WP520 “Humanitarian Crisis”, Subscenario 1 - Comparison Exercise: Evaluation of accuracy

Guidelines:

- Detection of single dwellings
- Differentiation of dwelling size: e.g. exclusion of very small dwellings (< 6 sqm)

Differentiation of dwelling types:

- PLUS – camp facility/hut/hut_dark/hut_potential/rondavel
- DLR – camp facilities/dwelling_blue/dwelling_potential/dweeling_white
- JRC – no differentiation

A tool for assessing a spatially explicit accuracy has been developed in Python and is implemented in ArcGIS. The workflow of the tool is described in the upcoming section “Methodology: Polygon – Polygon”. For the site-specific as well as for the spatially-explicit accuracy, a buffer distance of 2m was considered reasonable as the automated approaches did not always delineate the complete objects and considering, that in camp planning a minimum distance of 2m between tents/dwellings is recommended [this buffer value can be changed in the analysis tool interactively]

A dwelling-type-specific accuracy could not be conducted due to missing class-descriptions in the reference data set and different dwelling types depending on analyses approaches used by the three partners (see above).

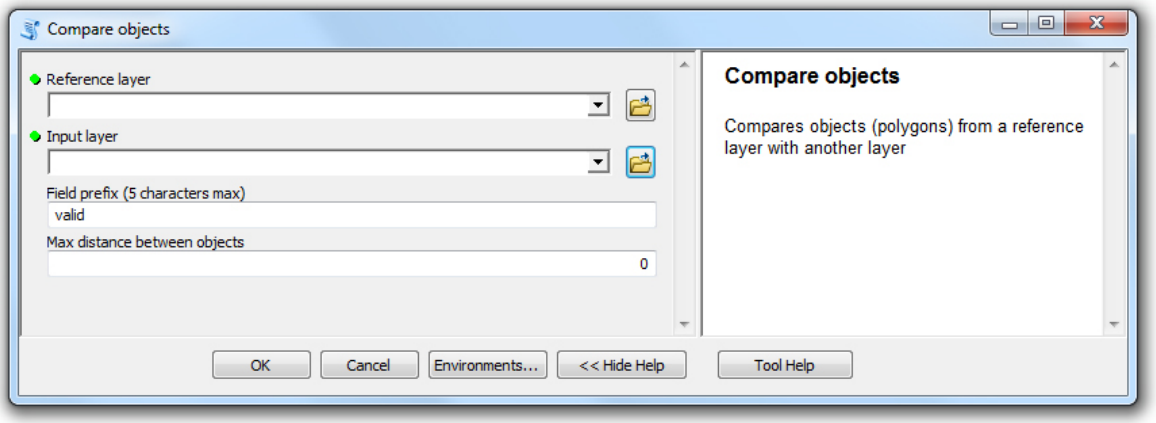
Aims of the tool:

- Count how many extracted dwelling-polygons do have an associated reference polygon
    - ⇨ Overall number of reference / overall number of classified dwellings * 100
- Avoid double counting (if e.g. a reference polygon is covered by more than one dwelling/object)
    - ⇨ Double-count is equal to false positive
- For reference polygons with more than one object, use the most appropriate object (i.e. the object with the largest overlap)
- Each reference-object is buffered on the fly accordingly to a certain user-defined distance. This ensures to also count associated dwellings, which are not directly overlapping the reference due to errors in the classification procedure. The selected small buffer size of 2m should ensure, that not neighboring dwellings are falsely taken into account
- One object can only be linked to one reference polygon
- Count how many dwellings do not have an associated reference polygon
    - ⇨ False positives
- Provide overall overlap of objects with reference polygons per block area
    - ⇨ Spatially explicit per block
- Provide overall overlap of objects with overall area of reference polygons
    - ⇨ Spatially explicit

Methodology – Polygon-Polygon

Input-Parameter:

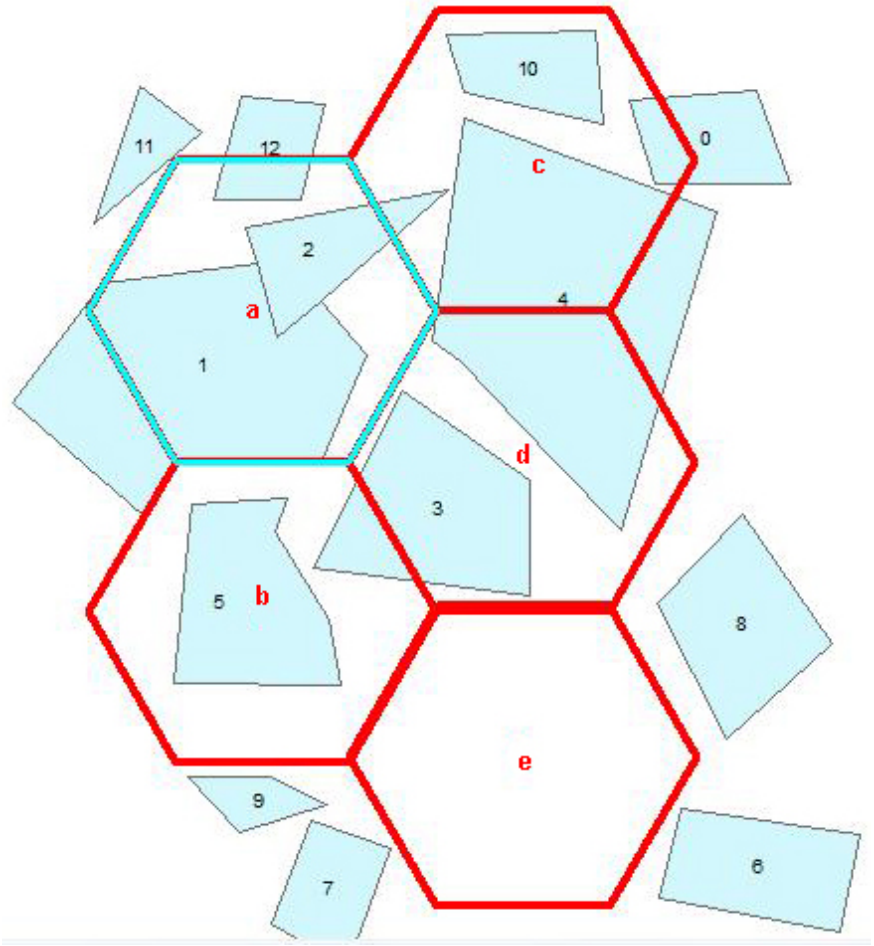
1. Reference layer with reference polygons.
2. Layer for comparison (henceforth called “Inputlayer”).
3. The prefix for the name of the columns in the attribute table. The attribute table of the reference layer is extended by 7 columns containing the results of the tool.
4. The distance of the buffer. The reference objects will be buffered with this distance.



Workflow:

1. The overlap of the Inputlayer object with the highest overlapping area is calculated and stored as *prefix_**big***[1] in the attribute table [list is virtually sorted to start from the highest overlapping polygons]
2. Objects from the Inputlayer with no overlap in the reference layer are marked with a zero in the column *prefix_**gd***, vice versa overlapping polygons are marked with a one.
3. The column *prefix_**gdid*** contains the ID of the matching Inputlayer-object. If there is no Inputlayer-object but a reference object available, the ID is set to 9999
4. Starting with the Inputlayer object that has the maximum overlap with a reference object, the single overlapping areas are calculated for each Inputlayer object with the reference layer objects, counting each Inputlayer object only once. The area is stored as *prefix_**gdar.***
5. The column *prefix_**bad*** contains the Inputlayer objects that are double-, triple- or more-counts and thus are additionally overlapping one reference object.
6. The column *prefix_**all*** lists the sum of all Inputlayer objects that match with a single reference object.
7. The overall overlap area of all Inputlayer objects matching with one reference object is given in the column *prefix_**alar*** without taking double- or more-counts into account.

[1] Remark: Prefixes of the field names can be defined by the user. Suffixes are a bit cryptic due to the field length limitations in shapefiles.



Examples of how the tool works:

Reference - Hexagon a

- Contains the objects 1,2,12. Object 1 has the largest overlap, thus *prefix_**big***, *prefix_**gdar*** and the ID *prefix_**gdid*** are interlinked to 1. The overall area *prefix_**alar*** is the sum of the overlapping area of the objects 1, 2 and 12. *Prefix_**bad*** is in this case two, since two objects are additionally overlapping with Hexagon a. *Prefix_**all*** is three, respectively. [If a buffer is selected, also object 11, 3 and 4 are taken into account]

Reference - Hexagon c

- Contains the objects 0,2,4,10. Object 4 has the largest overlap, thus *prefix_**big*** and *prefix_**gdar*** and the ID *prefix_**gd*** are interlinked to 4. The overall area *prefix_**alar*** is the sum of the overlapping area of the objects 0,2,4 and 10. *Prefix_**bad*** is in this case three, since three objects are additionally overlapping with Hexagon c. *Prefix_**all*** is four, respectively.

Reference - Hexagon d

- Contains the objects 3 and 4. Object 4 has the largest overlap but this object has been counted in Hexagon c already, since its overlap is larger there. The largest area in Hexagon d is nonetheless provided by 4, thus *prefix_**big*** is linked to 4. However, *prefix_**gdar*** and the ID *prefix_**gd*** are interlinked to 3, since 3 is the second-largest object and has not been counted yet. The overall area *prefix_**alar*** is the sum of the overlapping area of the objects 3 and 4. *Prefix_**bad*** is in this case one, *prefix_**all*** is two, accordingly.