AssesSeg for Windows 64bit is a command line executable software that calculate a modified version of Euclidian Distance 2 index.

AssesSeg deals only with ESRI polygon shapefile and its source code is written with open source libraries (python 2.7, gdal).

AssesSeg needs two different typologies of input: a reference shapefile and output segmentation shapefiles. Actually, they are called output because they are supposed to be produced with segmentation software.

The output of AssesSeg is an .xlsx file that ha the following structure:

- n spreadsheet, one for each group of output segmentation files (see Use of AssesSeg.exe);
- for each n-th spreadsheet there are the following columns:
  - name: name of the i-th output segmentation shapefile in the n-th output segmentation shapefile output.
  - Scale, shape, compactness: if the i-th output segmentation shapefiles comes from a multiresolution algorithm and has a specific name (see Use of AssesSeg.exe) this three column will record the input scale, the shape and the compactness values of the i-th output shapefile. If the name of the selected output segmentation shapefile does not respect the required syntax ruled (see Use of AssesSeg.exe) then the scale, the shape and the compactness values will be set to 0.
  - n. gt geometry: number of selected reference geometries that respect the selection criteria (e.g overlapping condition);
  - n. seg geometry: number of selected output segmentation shapefile geometries that respect the selection criteria (e.g overlapping condition);
  - area_gt: total area of selected groud truth geometries expressed in square of the same unit of the internal reference system of the reference shapefile (if reference system is utm the unit is meters).
  - under seg area: see reference section;
  - nsr: number-of-segmentation ratio , see reference section;
  - pse: potential segmentation error, see reference section;
  - ED2_index: modified ED_index.

Use of AssesSeg.exe

AssesSeg to run needs a working folder:

The working folder has the following subfolders:

- /gt: you have to past your reference shapefile here.
- /seg_test:
  - in the seg_test folder you have to create n folders (one for every excel spread sheet in the output excel file).
  - The name of the folders in /seg_test is the name of the n-th excel spreadsheet in the output excel file.
  - In the n-th excel spreadsheet you will have the result related to the shapefiles inside the n-th folder.
  - If the shapefiles in the n-th folder are named SclXX_ShpY.Y_CompZ.Z.shp or SclXXX_ShpY.Y_CompZ.Z.shp then AssesSeg is able to recognize scale (XX or XXX),

shape (Y.Y) and compactness (Z.Z): this could be useful if segmentation files were produced with a multiresolution segmentation algorithm.

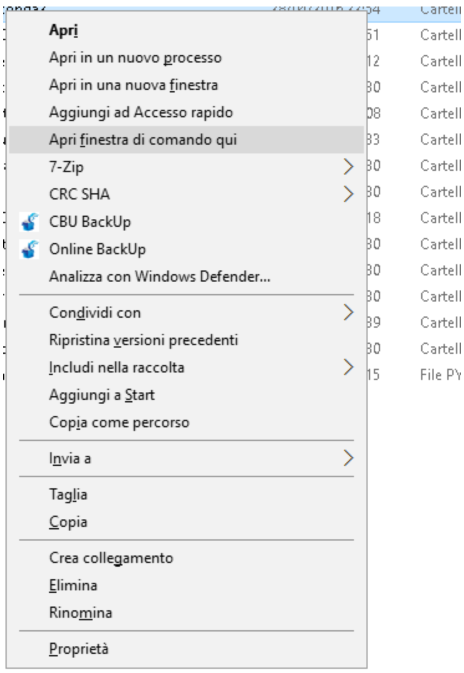
The syntax to run the program is:

AssesSeg -w <working folder> -o <outputfile> -p <overlapping percentage>

- “working folder” : path of the working folder;
- “outputfile”: output name of the excel file (this will be saved in the working folder)
- p: overlapping percentage (see reference section).
- the inverted commas are always necessary for the –w and –o options

In order to easily open a command line in the folder in which is located AssesSeg.exe you need to hold the shift key when you right click with the mouse on the selected folder.

The select “open command window here”



Example:

- If the working folder is located in : E:\documenti\Python Scripts\AssesSeg;

- If the output excel file name is : test
- If the overlapping percentage is 50

You simply need to type:

AssesSeg –w “E:\documenti\Python Scripts\AssesSeg” –o “test” –p 50

the inverted commas are always necessary for the –w and –o options.

Theoretical framework:

ED2, in its original formulation, starts the computations with the definition of the corresponding segment dataset. For each considered image segmentation output, the dataset owned the segments that spatially overlap the reference polygons. A further constrain is imposed over the corresponding segment dataset Liu et al. (2012): a considered segment can be labelled as a corresponding segment if the area of intersection between a reference polygon and the candidate segment is more than half the area of either the reference polygon or the candidate segment (overlapping criteria). After defying the corresponding segments dataset, the ED2 index (1) evaluates the segmentation quality in a two dimensional Euclidean space by means of the Potential Segmentation Error (PSE) and of the Number-of-Segments Ratio (NSR). The PSE (2) measures the geometric discrepancy as the ratio between the total area of under-segments and the total area of reference polygons. The under-segmentation is present when the contour of reference polygons r_k divides a corresponding segment s_i into at least two parts. Parts that falls outside are the under-segmented area (PSE = 0 means no under-segmentation). The NRS (3) measures the arithmetic discrepancy between the reference geometries and the candidate segments and is defined as the absolute difference between the number of reference polygons (m) and the number of corresponding segments (v) divided by the number of reference polygons.

$$\mathrm{ED2} = \sqrt{(\mathrm{PSE})^2 + (\mathrm{NSR})^2} \tag{1}$$

$$\mathrm{PSE} = \frac{\sum |s_i - r_k|}{\sum |r_k|} \tag{2}$$

$$\mathrm{NSR} = \frac{|m - v|}{m} \tag{3}$$

A high ED2 value indicates a significant geometric discrepancy, otherwise a significant arithmetic discrepancy, or both. The implemented modification of the ED2 index was introduced to consider the side effect of the overlapping criteria that act as a filter both on candidate corresponding segments and on reference geometries. When the number of reference geometries rises it is not infrequent the case of reference geometries without any candidate segments. In cases or reference geometries without any candidate segments, the true number of employed reference geometries will be lower than the original one. The ED2 index should take this into account.

AssesSeg corrects the overlapping criteria side effect by increasing both the PSE and NSR when not all reference geometries meet the overlapping criteria. If n is the number of excluded reference geometries the new computed PSE (4) and NSR (5) will be:

$$\mathrm{PSE_{new}} = \frac{\sum|s_i - r_k| + n \times \max(|s_i - r_k|)}{\sum |r_k|} \tag{4}$$

$$\mathrm{NSR_{new}} = \frac{|m - v - n \times v_{max}|}{m - n} \tag{5}$$

Where $\max(|s_i - r_k|)$ is the maximum over segmented area found for a single reference geometry; v_{max} is the maximum number of corresponding segments found for one single reference geometry; $\sum |r_k|$ is the total area of the m – n reference geometries.

The figure shows the logic process followed by the software for a single segmentation output shapefile (in input for the software)

References

Liu, Y., Bian, L., Meng, Y., Wang, H., Zhang, S., Yang, Y., Shao, X., Wang, B., 2012. Discrepancy measures for selecting optimal combination of parameter values in object-based image analysis. ISPRS Journal of Photogrammetry and Remote Sensing 68, 144-156.