Workshop: Satellite data analysis and Machine Learning classification with QGIS

Gorica Bratic and Maria Antonia Brovelli
Politecnico di Milano Department of Civil and Environmental Engineering - Milano (Italy)

The scope of the workshop is to introduce how classification of satellite imagery can be done with QGIS (https://www.qgis.org/en/site/) by showing how to retrieve, process and classify satellite imagery, as well as how to assess performance of machine learning algorithms through error matrix and accuracy indexes.
The workshop involves two QGIS plugins: Semi-automatic Classification Plugin (SCP) and dzetsaka. SCP is used for majority of preprocessing operations such as retrieval of the Sentinel 2 imagery for an area of interest, DOS (Dark object subtraction) atmospheric correction, selection of specific bands for classification, creation of composite and computation of band algebra (i.e., Normalized Difference Vegetation Index (NDVI). The dzetsaka plugin is used to detect and classify built-up areas starting from preprocessed satellite imagery with Gaussian Mixture Model, Random Forest and K-Nearest Neighbors machine learning algorithms.
Besides the two plugins, some core QGIS functionalities and are included in the workshop for clipping satellite imagery and creating vector file of training data. Lastly, outcomes of the machine learning algorithm are compared with the global map of human settlements – GHS-BUILT (Sentinel-1) produced by Joint Research Center (JRC) of European Commission to assess their performance. Before being used for assessing algorithms' performances, GHS-BUILT (Sentinel-1) is adapted to coordinate reference system, resolution, and classes of classification outcomes. Adaptation of GHS-BUILT (Sentinel-1) involves many isolated operations (reprojection, tile merging, resampling, and reclassification). For this reason, the QGIS Graphical Modeler is introduced in the exercise because it allows automation of chain of operations. Besides the adaptation of GHS-BUILT (Sentinel-1), the computation of error matrix and accuracy indexes for each classification outcome are integrated with the Graphical Modeler too.

Requirements for the exercise (guides for installation will be provided):

- QGIS Desktop 3.10.x with GRASS
- scikit-learn library for Python3 in QGIS
- Account on Copernicus Open Access Hub (for retrieving Sentinel-2 imagery)

Dataset for the exercise:

- Vector of bounding box of region of interest
- Training dataset
- GHS-BUILT (Sentinel-1) data in the region of interest
- Satellite imagery (Since download of imagery takes a lot of time, we will provide satellite imagery to the participants, but however the procedure of download will be explained)

Schedule. The workshop is organized in two parts of about 2 hours each (the pace of the workshop is slow in such a way to give time to "average" attendees the time of following and doing the exercise.

- introduction and pre-processing
- processing/assessment