The Journal of Open Source Software

Semi-Automatic Classification Plugin: A Python tool for the download and processing of remote sensing images in QGIS

Luca Congedo[1]

1 Independent Researcher

DOI: 10.21105/joss.03172

Software

- Review
- Repository
- Archive

Editor: Kristen Thyng

Reviewers:

- @bstabler
- @joferkington

Submitted: 08 March 2021
Published: 12 July 2021

Summary

The `Semi-Automatic Classification Plugin` is a Python plugin for the software QGIS (QGIS Development Team, 2021) developed with the overall objective to facilitate land cover monitoring by people whose main field is not strictly remote sensing, but that could benefit from remote sensing analysis. The `Semi-Automatic Classification Plugin` provides a set of intertwined tools and a user interface for easing and automating the phases of land cover classification, from the download of remote sensing images, to the preprocessing (i.e. tools for preparing the data to the analysis or other calculations), the processing (i.e. tools for performing the classification of land cover or perform analysis), and postprocessing (i.e. tools for assessing the classification accuracy, refining the classification, or integrating additional data). The processing of remote sensing data can be computationally intensive, therefore most of the developed tools use Python multiprocessing to exploit the system CPU and RAM by splitting the work among multiple threads. The aim of this paper is to describe the main characteristics of the `Semi-Automatic Classification Plugin` for the land cover classification of remote sensing images.

Statement of need

Land cover is defined as the material at the Earth's surface, such as soil, vegetation, water, asphalt, etc. (Fisher & Unwin, 2005); during the last decades, remote sensing has been been widely used for land cover monitoring and assessing environmental changes (Rogan & Chen, 2004).

Remote sensing is the measurement, by a sensor, of the energy that is emanated from the Earth's surface. The source of the measured energy can be the Sun (i.e. passive remote sensing, when the sensors measure the energy reflected by the Earth's surface), or the sensor platform (i.e. active remote sensing, when the sensors emit the energy and measure the energy reflected by the Earth's surface) such as radar sensors which work in the microwave range (Richards & Jia, 2006).

Sensors can be installed on board of airplanes or satellites, measuring the electromagnetic radiation (i.e. radiance) at specific ranges of wavelength (also called bands). The result of the sensor measurements are converted into a digital image, having different characteristics (resolutions) depending on the sensor (Richards & Jia, 2006):

- Spatial resolution (also referred to as Instantaneous field of view), usually corresponding to pixel size measured in meters;

- Spectral resolution, is the number of the spectral bands located in the electromagnetic spectrum; usually, each band corresponds to an image raster;
- Radiometric resolution, is the number of binary digits representing the range of brightness values, usually measured as bit number.

During the last few years, the availability of satellite images has increased and become affordable, especially because of the free access to the Landsat data (a set of satellites developed by the National Aeronautics and Space Administration of USA, since the early 1970's) (United States Geological Survey, 2019), and the European satellites Sentinel-2 launched in 2015 (European Space Agency, 2018). The free availability of these data extended their use in various fields such as urban planning, agriculture, environmental monitoring, etc. (Nink et al., 2019; Pesaresi et al., 2016; Weiss et al., 2020).

For land cover analysis, it useful to define the reflectance as the ratio of reflected versus total incident energy (Richards & Jia, 2006); each material at the ground is characterized by a spectral signature, which is the reflectance as a function of wavelength. Therefore, it is possible to classify land cover using the spectral signatures of materials, through classification algorithms that label image pixels, allowing for the creation of thematic maps.

A supervised classification (also known as semi-automatic classification) is an image processing technique that aims at classifying land cover by training an algorithm with samples of material spectral signatures. There are several semi-automatic classification algorithms that in general require a priori knowledge of material spectral signatures, through the identification in the image of training pixels and calculation of the corresponding spectral signatures (Richards & Jia, 2006). Based on the training pixels (i.e. spectral signatures) the algorithm classifies the whole remote sensing image through a mathematical model.

Remote sensing images are spatial data and require the use of a Geographic Information System (GIS) for visualization and processing. The development of open source GIS and processing software can foster environmental monitoring, which can be performed with no cost for software and remote sensing images considering the free data availability. QGIS is an open source GIS software which provides several tools for data visualization and analysis; it is mainly written in C++ code, but it also allows to extend the functions thereof through API and plugins written in Python (QGIS Development Team, 2021). QGIS has a large repository of plugins that improve data analysis and also provide access to several image processing tools included in other open source programs such as GRASS (GRASS Development Team, 2017), GDAL (GDAL/OGR contributors, 2021), and Orfeo Toolbox (OTB Development Team, 2020). These programs allow for the processing of satellite images, nevertheless their use can be difficult for people interested in land cover classifications but specialized in fields other than remote sensing, because of the several steps required for image processing.

The `Semi-Automatic Classification Plugin` aims to provide a complete suite of tools for processing remote sensing data, easing the phases related to the download, the preprocessing of images, and the postprocessing of classifications.

Several tutorials about the `Semi-Automatic Classification Plugin` are freely available on the official website, and the user manual is being translated into several languages by the user community.

The `Semi-Automatic Classification Plugin` has been cited in several studies about land cover, urban growth, and image processing (Arekhi et al., 2019; Damtew et al., 2021; Furukawa et al., 2020; Garilli et al., 2021; Huq et al., 2019; Leroux et al., 2018; Palafox-Juárez et al., 2021; Pelage et al., 2019; Teodoro & Amaral, 2019; Zhang et al., 2018).

The Journal of Open Source Software

Overview of the Semi-Automatic Classification Plugin

The `Semi-Automatic Classification Plugin` interfaces is composed of several modules (as illustrated in Figure 1). A module allows for searching and downloading freely available images (in particular ASTER, GOES, Landsat, MODIS, Sentinel-1, Sentinel-2, and Sentinel-3). It is possible to perform the preprocessing and raster calculations automatically after the download, by setting a few parameters in the user interface.

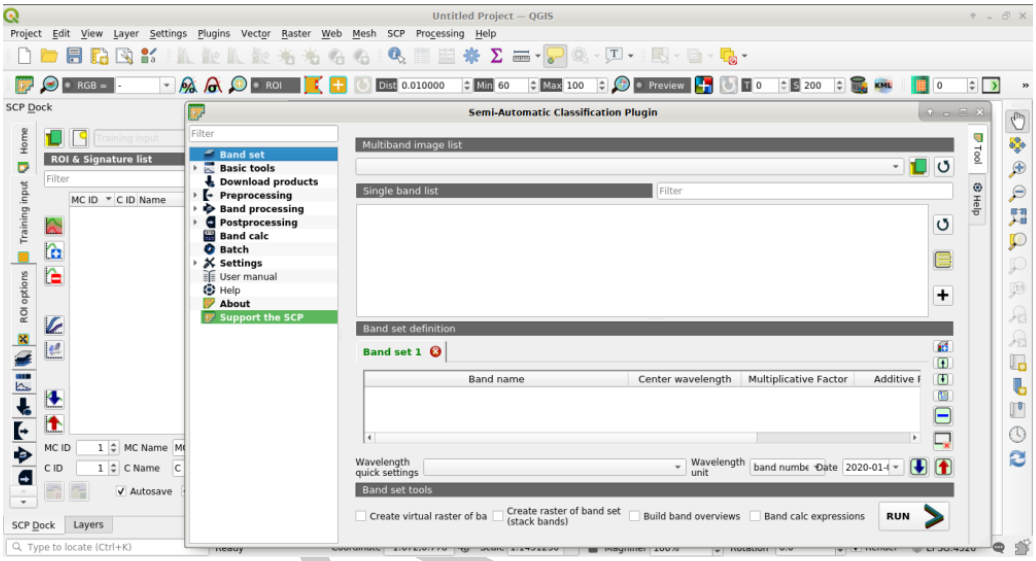

Figure 1: User interface of the `Semi-Automatic Classification Plugin` for QGIS.

The interface allows to define an image input (named `band set`) which is the set of raster bands to be processed.

The following tools are available for `preprocessing`:

- Conversion to reflectance for ASTER, GOES, Landsat, MODIS, Sentinel-1, Sentinel-2, and Sentinel-3 images;
- Clip multiple rasters at once;
- Cloud masking based on the values of a raster mask;
- Mosaic band sets, merging the corresponding bands of two or more band sets;
- Statistic calculation for neighbor pixels;
- Reprojection of the coordinates of raster bands;
- Splitting or stacking raster bands in a unique file;
- Vector to raster conversion.

The following `processing` tools are available:

- Band combination to get a raster where each value corresponds to a combination of input values;
- Clustering for unsupervised classification (i.e. without training input);
- Principal Component Analysis of band set;
- Calculation of the spectral distance between every corresponding pixel of two band sets;
- Classification using one of the available algorithms, such as Minimum Distance (Richards & Jia, 2006), Maximum Likelihood (Richards & Jia, 2006), Spectral Angle Mapping (Kruse et al., 1993), and Random Forest (Breiman, 2001).

Considering that semi-automatic classification algorithms require training pixels (i.e. spectral signatures), a specific dock interface for training input creation and spectral signature calculation is available, which allows to create polygons interactively (manually or through region growing) and import spectral libraries such as the USGS Spectral Library (Kokaly et al., 2017). The interface allows to create classification previews on small portion of the image to assess the results before launching the classification process for the whole image. In addition, the plot of spectral signatures can be visualized in order to assess the spectral separability of signatures.

A band calculator is available for performing mathematical and conditional calculation using input rasters, for instance for the calculation of spectral indices.

The following `postprocessing` tools are available for refining the classification output and further analyses:

- Accuracy assessment of the classification;
- Classification dilation, erosion, and sieve for refining the patches of the classes;
- Classification report for calculation of class statistics such as number of pixels, percentage and area;
- Conversion of classification from raster to vector;
- Manual editing of raster values;
- Assessment of land cover change comparing two classifications;
- Statistics related to an input raster for every unique value of a reference vector or raster;
- Reclassification of raster values.

Finally, the tool `Batch` allows for performing a series of functions consecutively and automatically through the definition of a script.

The `Semi-Automatic Classification Plugin` for QGIS is developed with Python 3 and requires the installation of GDAL, NumPy, SciPy, and Matplotlib (as illustrated in Figure 2). The user interface is developed using the Qt framework. The additional installation of SNAP (European Space Agency, 2021) is required for the Sentinel-1 preprocessing and random forest classification tools.

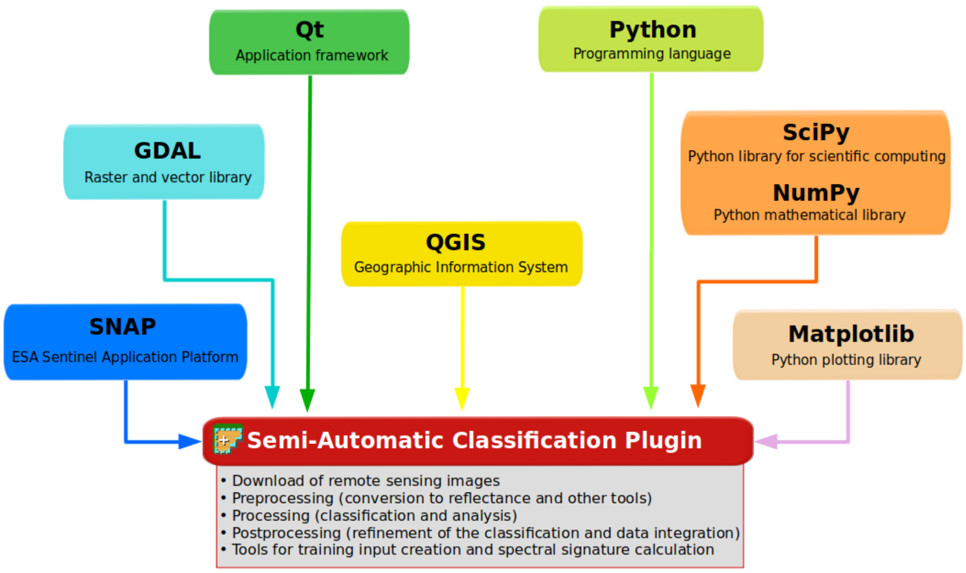

Figure 2: Framework of the `Semi-Automatic Classification Plugin` main dependencies.

A testing tool is available for verifying the correct installation of the required dependencies and check the main functions of the `Semi-Automatic Classification Plugin`.

Acknowledgements

The first version of the Semi-Automatic Classification Plugin was developed by Luca Congedo in 2012 for the ACC Dar Project in order to create a tool for the classification of land cover in an affordable and automatic fashion; following versions 2, 3, 4, and 5 of Semi-Automatic Classification Plugin were developed by Luca Congedo as personal commitment to the remote sensing field and open source philosophy. Semi-Automatic Classification Plugin version 6 was developed in the frame of Luca Congedo's PhD in Landscape and Environment at Sapienza University of Rome. `Semi-Automatic Classification Plugin` version 7 was developed by Luca Congedo as personal commitment to the remote sensing field and open source philosophy.

Special thanks to all the users for contributing to the Semi-Automatic Classification Plugin and translating the interface and user manual to other languages.

References

Arekhi, M., Goksel, C., Balik Sanli, F., & Senel, G. (2019). Comparative Evaluation of the Spectral and Spatial Consistency of Sentinel-2 and Landsat-8 OLI Data for Igneada Longos Forest. *ISPRS International Journal of Geo-Information*, *8*(2). https://doi.org/10.3390/ijgi8020056

Breiman, L. (2001). Random Forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Damtew, Y. T., Verbeiren, B., Awoke, A., & Triest, L. (2021). Satellite Imageries and Field Data of Macrophytes Reveal a Regime Shift of a Tropical Lake (Lake Ziway, Ethiopia). *Water*, *13*. https://doi.org/10.3390/w13040396

European Space Agency. (2018). *Sentinel-2 Products Specification Document*. European Space Agency. https://sentinels.copernicus.eu/web/sentinel/user-guides/sentinel-2-msi/document-library

European Space Agency. (2021). *SNAP - ESA sentinel application platform* (Version 8.0.0) [Computer software]. http://step.esa.int

Fisher, P. F., & Unwin, D. J. (2005). *Re-Presenting GIS* (J. W. &. Sons, Ed.). John Wiley & Sons. ISBN: 978-0-470-84847-0

Furukawa, F., Morimoto, J., Yoshimura, N., & Kaneko, M. (2020). Comparison of Conventional Change Detection Methodologies Using High-Resolution Imagery to Find Forest Damage Caused by Typhoons. *Remote Sensing*, *12*(19). https://doi.org/10.3390/rs12193242

Garilli, E., Bruno, N., Autelitano, F., Roncella, R., & Giuliani, F. (2021). Automatic detection of stone pavement's pattern based on UAV photogrammetry. *Automation in Construction*, *122*, 14. https://doi.org/https://doi.org/10.1016/j.autcon.2020.103477

GDAL/OGR contributors. (2021). *GDAL/OGR geospatial data abstraction software library*. Open Source Geospatial Foundation. https://gdal.org

GRASS Development Team. (2017). *Geographic resources analysis support system (GRASS GIS) software, version 7.2*. Open Source Geospatial Foundation. http://grass.osgeo.org

Huq, N., Bruns, A., & Ribbe, L. (2019). Interactions between freshwater ecosystem services and land cover changes in southern Bangladesh: A perspective from short-term (seasonal) and long-term (1973–2014) scale. *Science of The Total Environment*, *650*, 132–143. https://doi.org/https://doi.org/10.1016/j.scitotenv.2018.08.430

Kokaly, R. F., Clark, R. N., Swayze, G. A., Livo, K. E., Hoefen, T. M., Pearson, N. C., Wise, R. A., Benzel, W. M., Lowers, H. A., Driscoll, R. L., & Klein, A. J. (2017). USGS spectral library version 7. In U. S. G. Survey (Ed.), *Data Series* (p. 68). http://pubs.er.usgs.gov/publication/ds1035

Kruse, F. A., Lefkoff, A. B., Boardman, J. B., Heidebrecht, K. B., Shapiro, A. T., Barloon, P. J., & Goetz, A. F. H. (1993). The Spectral Image Processing System (SIPS) - Interactive Visualization and Analysis of Imaging spectrometer Data. *Remote Sensing of Environment*, *44*(/), 145–163.

Leroux, L., Congedo, L., Bellón, B., Gaetano, R., & Bégué, A. (2018). Land Cover Mapping Using Sentinel-2 Images and the Semi-Automatic Classification Plugin: A Northern Burkina Faso Case Study. In *QGIS and Applications in Agriculture and Forest* (pp. 119–151). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781119457107.ch4

Nink, S., Hill, J., Stoffels, J., Buddenbaum, H., Frantz, D., & Langshausen, J. (2019). Using Landsat and Sentinel-2 Data for the Generation of Continuously Updated Forest Type Information Layers in a Cross-Border Region. *Remote Sensing*, *11*(20). https://doi.org/10.3390/rs11202337

OTB Development Team. (2020). *Orfeo ToolBox 7.2* (Version 7.2.0) [Computer software]. Zenodo. https://doi.org/10.5281/zenodo.4066373

Palafox-Juárez, E. B., López-Martínez, J. O., Hernández-Stefanoni, J. L., & Hernández-Nuñez, H. (2021). Impact of Urban Land-Cover Changes on the Spatial-Temporal Land Surface Temperature in a Tropical City of Mexico. *ISPRS International Journal of Geo-Information*, *10*(2). https://doi.org/10.3390/ijgi10020076

Pelage, L., Domalain, G., Lira, A. S., Travassos, P., & Frédou, T. (2019). Coastal Land Use in Northeast Brazil: Mangrove Coverage Evolution Over Three Decades. *Tropical Conservation Science*, *12*, 15. https://doi.org/10.1177/1940082918822411

Pesaresi, M., Corbane, C., Julea, A., Florczyk, A. J., Syrris, V., & Soille, P. (2016). Assessment of the Added-Value of Sentinel-2 for Detecting Built-up Areas. *Remote Sensing*, *8*(4), 299. https://doi.org/10.3390/rs8040299

QGIS Development Team. (2021). *QGIS geographic information system*. QGIS Association. https://www.qgis.org

Richards, J. A., & Jia, X. (2006). *Remote Sensing Digital Image Analysis: An Introduction* (Springer, Ed.). Springer.

Rogan, J., & Chen, D. (2004). Remote sensing technology for mapping and monitoring land-cover and land-use change. *Progress in Planning*, *61*(4), 301–325. https://doi.org/10.1016/S0305-9006(03)00066-7

Teodoro, A., & Amaral, A. (2019). A Statistical and Spatial Analysis of Portuguese Forest Fires in Summer 2016 Considering Landsat 8 and Sentinel 2A Data. *Environments*, *6*(3). https://doi.org/10.3390/environments6030036

United States Geological Survey. (2019). *Landsat 8 (L8) Data Users Handbook. Version 5.0*. Department of the Interior U.S. Geological Survey. https://prd-wret.s3.us-west-2.amazonaws.com/assets/palladium/production/atoms/files/LSDS-1574_L8_Data_Users_Handbook-v5.0.pdf

Weiss, M., Jacob, F., & Duveiller, G. (2020). *Remote sensing for agricultural applications: A meta-review*. *236*, 111402. https://doi.org/https://doi.org/10.1016/j.rse.2019.111402

Zhang, Y., Zhang, H., & Sun, Z. (2018). Effects of Urban Growth on Architectural Heritage: The Case of Buddhist Monasteries in the Qinghai-Tibet Plateau. *Sustainability*, *10*(5). https://doi.org/10.3390/su10051593