Open Educational Resources for Validation of Global High-Resolution Land Cover Maps

Candan Eylül Kilsedar, Gorica Bratic, Monia Elisa Molinari, Marco Minghini, and Maria Antonia Brovelli

Politecnico di Milano, Department of Civil and Environmental Engineering, Piazza Leonardo da Vinci 32, 20133, Milan, Italy

Corresponding author:
Candan Eylül Kilsedar

Email address: candaneylul.kilsedar@polimi.it

ABSTRACT

Land cover (LC) maps are crucial to analyze and understand several phenomena, including urbanization, deforestation and climate change. This elevates the importance of their accuracy, which is assessed through a validation process. However, we observed that knowledge on the importance of LC maps and their validation is limited. Hence, a set of educational resources has been created to assist in the validation of LC maps. These resources, available under an open access license, focus on validation through open source and easy-to-use software. Moreover, addressing the lack of accurate and up-to-date reference LC data, an application has been developed that provides users a means to collect LC data.

INTRODUCTION

Land cover (LC) maps are key products for several applications, including natural resources management (Cui et al., 2011; Butt et al., 2015; Sambou et al., 2015; Hua, 2017), biodiversity monitoring (Kerr and Ostrovsky, 2003; Turner et al., 2003; Nagendra et al., 2013; Willis, 2015) and climate change modeling (Feddema et al., 2005; Verburg et al., 2011; Bontemps et al., 2013; Kim, 2016). Their production involves the classification of remotely-sensed imagery and its validation, which defines the degree of adherence of the LC map to the reality. The latter step is usually performed through a comparison of the classified dataset with a reference dataset representing the "ground truth" (e.g., field surveys) and the computation of a confusion matrix from which many accuracy indexes can be extracted (Stehman, 1997). These span from the most commonly used accuracy indexes such as overall accuracy, user's and producer's accuracy to the more recently proposed ones such as allocation and quantity disagreements (Koukoulas and Blackburn, 2001; Congalton and Green, 2009; Labatut and Cherifi, 2011; Pontius and Millones, 2011; Brovelli et al., 2015).

This work started within the project "Capacity Building for High-Resolution Land Cover Inter-comparison and Validation", one of the ISPRS 2018 Education and Capacity Building Initiatives (http://www.isprs.org/society/ecbi/default.aspx). The project focuses on global high-resolution LC maps, in particular, the GlobeLand30 (GL30). GL30 is the first global LC map at 30 m resolution, provided by National Geomatics Center of China, available for two reference years 2000 and 2010 (Chen et al., 2015). Based on an analysis of state of the art in educational resources on the validation of high-resolution LC maps (Brovelli et al., 2018a), the project creates new educational resources, which are entirely based on open source software. The resources are released under an open access license to maximize exploitation by the community and were used to provide three hands-on workshops in Dar es Salaam (Tanzania), Nairobi (Kenya) and Delft (The Netherlands).

MOTIVATIONS AND GOALS

The project "Capacity Building for High-Resolution Land Cover Inter-comparison and Validation" started in February 2018 with the aim of creating educational resources on the validation of global LC maps, using the results of the thorough review of the state of the art, which highlighted the need to increase the awareness of the LC validation process (Brovelli et al., 2018a).

The educational resources tackle the issues recognized during the review. Firstly, an effort was made during the workshops to emphasize the importance of validation as an integral part of the LC map production, as the review revealed that it is limited. Secondly, as the software available for validation are often complicated to use and not well-documented (Brovelli et al., 2018b), a set of user-friendly software to validate LC maps was developed, and detailed educational resources were created to address this issue. Lastly, as there is often a lack of accurate and up-to-date reference data because their collection is labor intensive and expensive (Stehman et al., 2000), an application was developed for collecting LC data in situ to assist in the creation of LC reference dataset that is of great importance for validation (See et al., 2015; Fritz et al., 2017; Stehman et al., 2018; McCallum et al., 2018). All the software developed was presented to the participants of the workshops with step-by-step instructions.

CONTENT OF THE EDUCATIONAL RESOURCES

The educational resources (http://www.gisgeolab.polimi.it/docs/training/ISPRS2018/Validation_QGIS.zip, http://www.gisgeolab.polimi.it/docs/training/ISPRS2018/Land_Cover_Collector.zip) include practical guides for the validation of high-resolution LC maps using QGIS 2.18 open source geospatial software and an application for collecting ground truth LC data.

Validation Procedure

The LC dataset selected for validation is GL30. In the practical guide based on QGIS 2.18, validation is performed using two types of reference datasets: vector points and raster map. The software developed for validation require datasets with the same LC classification system, the same coordinate reference system, and in the case of a raster map, the same resolution as well. Therefore, some preprocessing steps on the datasets are required to perform validation successfully. Hereafter, two case studies included in the educational resources, which make use of these two types of reference datasets are described in detail. More information can be found in the work of Bratic et al. (2018a) and Bratic et al. (2018b). The educational resources show how to use the tools available in QGIS 2.18 to aid in preprocessing and how to use the software developed to perform the validation.
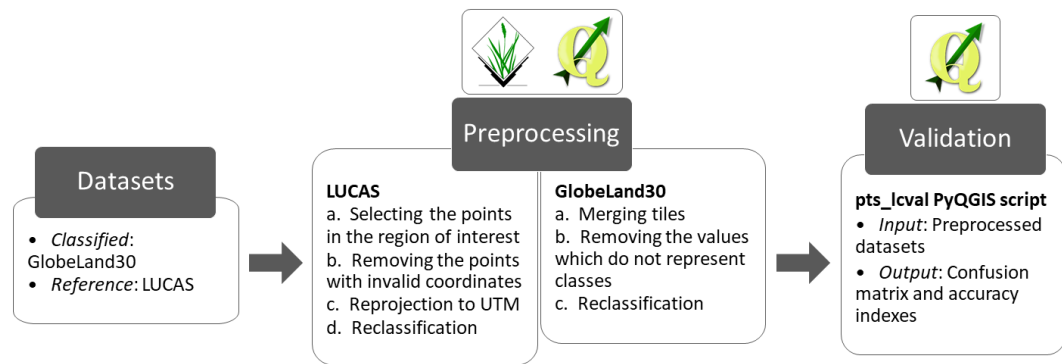
Validation with Vector Points Reference Dataset

The first case study validates GL30 of 2010 using a vector points reference dataset of 2009 obtained from the Land Use and Coverage Area frame Survey (LUCAS) programme (http://ec.europa.eu/eurostat/web/lucas/overview) promoted by Eurostat. The area of interest is the Lombardy region in northern Italy.

Preprocessing of the LUCAS dataset is required to i) select the LUCAS points located in the region of interest, ii) remove the points having null coordinates, iii) reproject the dataset according to the GL30 reference system (WGS84/UTM32N), and iv) reclassify the dataset according to the GL30 LC classification. Besides, the preprocessing of GL30 consists of three steps: i) merging tiles to obtain a unique dataset for Lombardy region, ii) setting the value of cells having zero value to no data value, and iii) reclassifying two LC classes (grassland and tundra) to merge them, since they are one class in LUCAS. Later, validation is performed through an ad hoc PyQGIS script named *pts_lcval* (https://github.com/GoricaB/Land-cover-validation). The script creates a new column in the attribute table of LUCAS points and fills it with the values of the GL30 dataset at the positions of LUCAS points using the two preprocessed datasets. After that, the script performs a pixel-by-pixel comparison that results in the confusion matrix computation and the global and per-class indexes, which are saved as CSV files. The whole procedure is summarized in Figure 1.
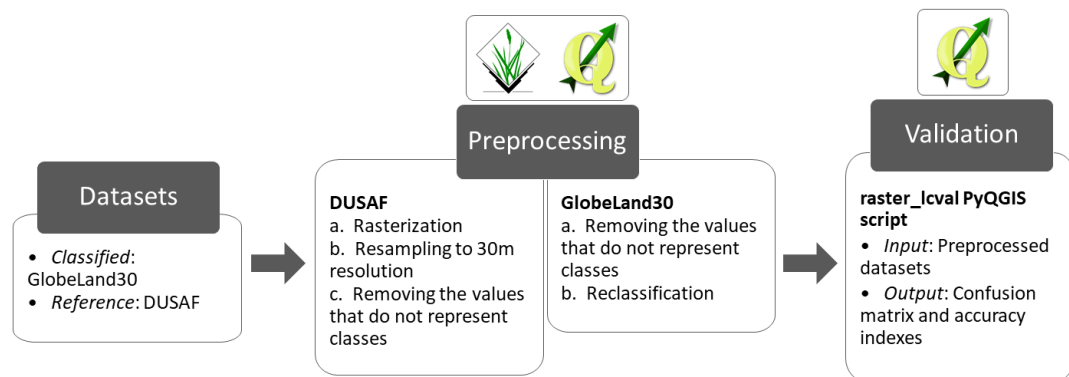
Validation with Raster Reference Dataset

The second case study proposes the validation of GL30 of 2010 employing a comparison with a local reference dataset available for the Lombardy region for the reference year 2012 (DUSAF, *Destinazione d'Uso dei Suoli Agricoli e Forestali*). The area of interest is the province of Como, located in the northern part of the Lombardy region.

**Figure 1.** Validation procedure with vector points reference dataset

Since DUSAF is in vector format, a preprocessing step is required to rasterize it. Rasterization is performed according to its first level classification. The resulting raster map has the same resolution as GL30. Besides, the values that do not represent the classes, i.e., no data values, in both DUSAF and GL30 are removed. To harmonize the classification system of the two datasets, GL30 is reclassified according to the DUSAF first level classification. After that, an ad hoc PyQGIS script named *raster_lcval* (https://github.com/GoricaB/Land-cover-validation) is used to perform the validation of GL30. The script uses the two preprocessed datasets and computes the confusion matrix and the global and per-class indexes, which are saved as CSV files. The whole procedure is summarized in Figure 2.

Figure 2. Validation procedure with raster reference dataset

Reference Data Collection
Land Cover Collector (https://github.com/kilsedar/land-cover-collector) is a free and open source application, which allows the collection of LC reference data for GL30 validation. It relies on Apache Cordova mobile application development framework (https://cordova.apache.org), and it is available on the Web ((https://landcover.como.polimi.it/collector)) and Android and iOS mobile devices. Leaflet (https://leafletjs.com) is used for Web mapping. The LC points collected are stored using both PouchDB (https://pouchdb.com) and CouchDB (http://couchdb.apache.org), which enables both offline and online data collection. The LC class of each collected point, the date of collection, the user's degree of certainty on the correctness of the stated classification, the photos in north, east, south and west directions, and the user's comment (optional) are stored. Collected data can be visualized and queried on a map within the application, as illustrated in Figure 3. Collected points are licensed under the Open Database License (ODbL) v1.0 (https://opendatacommons.org/licenses/odbl/1.0/), and can be downloaded within the application in JSON format. The application is currently available in eight languages: English, Italian, Arabic, Russian, Chinese, Portuguese, French and Spanish.
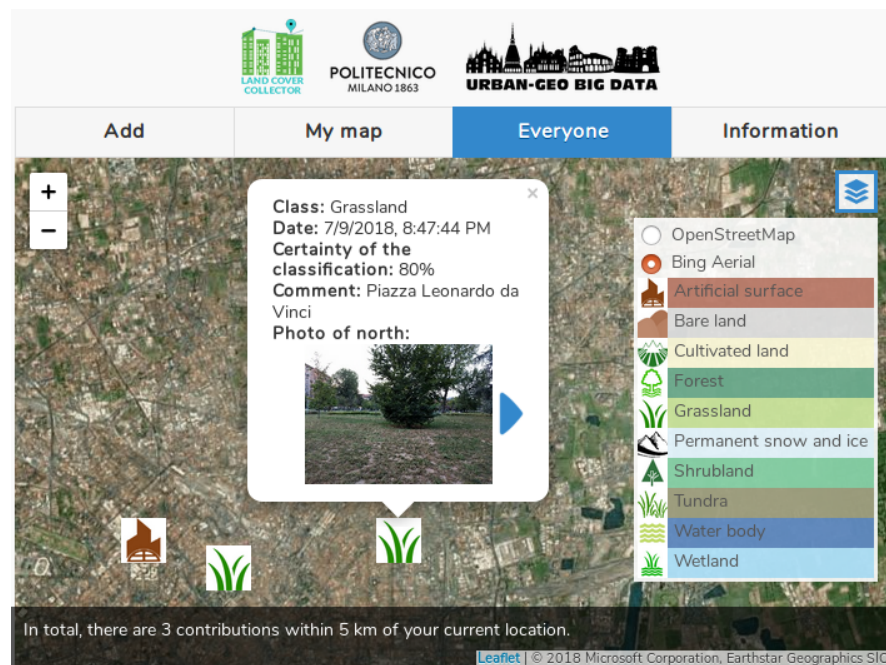
Figure 3. Visualization and query of collected data in Land Cover Collector

The corresponding educational resources explain how to collect LC reference data by providing detailed instructions on how to use the Land Cover Collector application. Furthermore, guidelines on how to prevent the violation of privacy rights while taking photos and breaking the law by entering forbidden places are provided.

CONCLUSIONS

Educational resources on the validation of LC maps and reference data collection were created based on the needs identified by a literature review. The resources highlight the importance of validation and provide easy-to-use software and instructions for validation and reference data collection. Educational resources are prepared in the form of slide presentations, which were presented in three workshops (two of which were held in developing countries) to fulfill the goal of the project as well as to promote their use within a broad community. As the resources are entirely based on open source software and available under an open access license, it is expected that the awareness on the importance of validation of LC maps, and on how this process can be carried out in practice, will be raised extensively. Future work includes using the data collected through the Land Cover Collector application to validate GL30.

ACKNOWLEDGEMENTS

REFERENCES

Bontemps, S., Defourny, P., Radoux, J., Van Bogaert, E., Lamarche, C., Achard, F., Mayaux, P., Boettcher, M., Brockmann, C., Kirches, G., Zulkhe, M., Kalogirou, V., Seifert, F. M., and Arino, O. (2013). Consistent global land cover maps for climate modelling communities: Current achievements of the ESA' Land Cover CCI. In *ESA Living Planet Symposium*, volume 722 of *ESA Special Publication*, page 62.

Bratic, G., Brovelli, M. A., and Molinari, M. E. (2018a). A free and open source tool to assess the accuracy of land cover maps: Implementation and application to Lombardy region (Italy). *ISPRS*

- *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-3:87–92.

Bratic, G., Molinari, M. E., and Brovelli, M. A. (2018b). Validation of the global high-resolution GlobeLand30 land cover map in Europe using the LUCAS database. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-4:51–59.

Brovelli, M. A., Minghini, M., Molinari, M. E., Kilsedar, C. E., Wu, H., Zheng, X., Chen, J., and Shu, P. (2018a). Open source software and open educational material on land cover maps intercomparison and validation. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-4:61–68.

Brovelli, M. A., Minghini, M., Molinari, M. E., Wu, H., Zheng, X., and Chen, J. (2018b). Capacity building for high-resolution land cover intercomparison and validation: What is available and what is needed. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-4/W8:15–22.

Brovelli, M. A., Molinari, M. E., Hussein, E., Chen, J., and Li, R. (2015). The first comprehensive accuracy assessment of GlobeLand30 at a national level: Methodology and results. *Remote Sensing*, 7(4):4191–4212.

Butt, A., Shabbir, R., Ahmad, S. S., and Aziz, N. (2015). Land use change mapping and analysis using remote sensing and GIS: A case study of Simly watershed, Islamabad, Pakistan. *The Egyptian Journal of Remote Sensing and Space Science*, 18(2):251–259.

Chen, J., Chen, J., Liao, A., Cao, X., Chen, L., Chen, X., He, C., Han, G., Peng, S., Lu, M., Zhang, W., Tong, X., and Mills, J. (2015). Global land cover mapping at 30 m resolution: A POK-based operational approach. *ISPRS Journal of Photogrammetry and Remote Sensing*, 103:7–27.

Congalton, R. and Green, K. (2009). *Assessing the Accuracy of Remotely Sensed Data*. Boca Raton: CRC Press, 2 edition.

Cui, G., Lee, W.-K., Kwak, D.-A., Choi, S., Park, T., and Lee, J. (2011). Desertification monitoring by LANDSAT TM satellite imagery. *Forest Science and Technology*, 7(3):110–116.

Feddema, J. J., Oleson, K. W., Bonan, G. B., Mearns, L. O., Buja, L. E., Meehl, G. A., and Washington, W. M. (2005). The importance of land-cover change in simulating future climates. *Science*, 310(5754):1674–1678.

Fritz, S., Fonte, C. C., and See, L. (2017). The role of citizen science in Earth observation. *Remote Sensing*, 9(4).

Hua, A. K. (2017). Land use land cover changes in detection of water quality: A study based on remote sensing and multivariate statistics. *Journal of Environmental Research and Public Health*, 2017:1–12.

Kerr, J. T. and Ostrovsky, M. (2003). From space to species: ecological applications for remote sensing. *Trends in Ecology & Evolution*, 18(6):299–305.

Kim, C. (2016). Land use classification and land use change analysis using satellite images in Lombok island, Indonesia. *Forest Science and Technology*, 12(4):183–191.

Koukoulas, S. and Blackburn, G. A. (2001). Introducing new indices for accuracy evaluation of classified images representing semi-natural woodland environments. *American Society for Photogrammetry and Remote Sensing*, 67(4):499–510.

Labatut, V. and Cherifi, H. (2011). Accuracy measures for the comparison of classifiers. In Ali, A.-D., editor, *The 5th International Conference on Information Technology*, pages 1–5, Amman, Jordan. Al-Zaytoonah University of Jordan.

McCallum, I., See, L., Sturn, T., Salk, C., Perger, C., Duerauer, M., Karner, M., Moorthy, I., Domian, D., Schepaschenko, D., and Fritz, S. (2018). Engaging citizens in environmental monitoring via gaming. *International Journal of Spatial Data Infrastructures Research*, 13:15–23.

Nagendra, H., Lucas, R., Honrado, J. P., Jongman, R. H., Tarantino, C., Adamo, M., and Mairota, P. (2013). Remote sensing for conservation monitoring: Assessing protected areas, habitat extent, habitat condition, species diversity, and threats. *Ecological Indicators*, 33:45–59.

Pontius, R. J. G. and Millones, M. (2011). Death to Kappa: Birth of quantity disagreement and allocation disagreement for accuracy assessment. *International Journal of Remote Sensing*, 32(15):4407–4429.

Sambou, S., Lykke, A. M., Sambou, H., Guiro, I., Sambou, B., and Mbow, C. (2015). Land use-land cover change and drivers of deforestation in the Patako protected area (Center-West of Senegal). *American Journal of Environmental Protection*, 4(6):306–317.

See, L., Fritz, S., Perger, C., Schill, C., McCallum, I., Schepaschenko, D., Duerauer, M., Sturn, T.,

Karner, M., Kraxner, F., and Obersteiner, M. (2015). Harnessing the power of volunteers, the Internet and Google Earth to collect and validate global spatial information using Geo-Wiki. *Technological Forecasting and Social Change*, 98:324–335.

Stehman, S. V. (1997). Selecting and interpreting measures of thematic classification accuracy. *Remote Sensing of Environment*, 62(1):77–89.

Stehman, S. V., Czaplewski, R. L., Nusser, S. M., Yang, L., and Zhu, Z. (2000). Combining accuracy assessment of land-cover maps with environmental monitoring programs. *Environmental Monitoring and Assessment*, 64(1):115–126.

Stehman, S. V., Fonte, C. C., Foody, G. M., and See, L. (2018). Using volunteered geographic information (VGI) in design-based statistical inference for area estimation and accuracy assessment of land cover. *Remote Sensing of Environment*, 212:47–59.

Turner, W., Spector, S., Gardiner, N., Fladeland, M., Sterling, E., and Steininger, M. (2003). Remote sensing for biodiversity science and conservation. *Trends in Ecology & Evolution*, 18(6):306–314.

Verburg, P. H., Neumann, K., and Nol, L. (2011). Challenges in using land use and land cover data for global change studies. *Global Change Biology*, 17(2):974–989.

Willis, K. S. (2015). Remote sensing change detection for ecological monitoring in United States protected areas. *Biological Conservation*, 182:233–242.