

Open educational resources for the validation of global high-resolution land cover maps

Land cover (LC) maps are crucial to understand and analyze several phenomena, including urbanization, deforestation and climate change. This elevates the importance of their accuracy, which is assessed through a validation process. However, it has been 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 the validation procedure through open source and easy-to-use tools. 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.

Open Educational Resources for the 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: candaneyul.kilsedar@polimi.it

ABSTRACT

Land cover (LC) maps are crucial to understand and analyze several phenomena, including urbanization, deforestation and climate change. This elevates the importance of their accuracy, which is assessed through a validation process. However, it has been 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 the validation procedure through open source and easy-to-use tools. 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, biodiversity monitoring and climate change modeling. Their production involves the classification of remotely-sensed imagery and its accuracy assessment or 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. 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 (Congalton and Green, 2009; Koukoulas and Blackburn, 2001; Labatut and Cherifi, 2011; Pontius and Millones, 2011).

This work is born within the project “Capacity Building for High-Resolution Land Cover Intercomparison and Validation”, one of the funded ISPRS Education and Capacity Building 2018 Initiatives (<http://www.isprs.org/society/ecbi/default.aspx>). The project focuses on global high-resolution LC maps, in particular the GlobeLand30 (GL30), which is the first global LC map at 30 m resolution, provided by National Geomatics Center of China and available for the two reference years 2000 and 2010. Based on an analysis of the state of the art in educational resources on the validation of high-resolution LC maps, the project creates new educational resources, which are fully based on open source geospatial 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 Intercomparison and Validation” started in February 2018 with the aim of creating educational resources on the intercomparison and validation

of global LC maps (Brovelli et al., in press). The thorough review of the state of the art in this field highlighted the need to increase the awareness on the LC intercomparison/validation process.

The educational resources were created to tackle the issues recognized during the review. Firstly, an effort was made to emphasize the importance of validation as an integral part of the LC map production. Secondly, software tools available for validation are often complex to use and not well-documented (Brovelli et al., 2018). Thus, a set of user-friendly software tools to validate LC maps were developed and detailed educational resources were created to address this issue. Lastly, there is often a lack of accurate and up-to-date reference data, because their collection is labor intensive and expensive. As a result, 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 the validation. These tools were presented to the participants of the workshops through hands-on sessions with step-by-step instructions.

CONTENT OF THE EDUCATIONAL RESOURCES

The educational resources (http://geomobile.como.polimi.it/website/docs/training/ISPRS2018/Validation_QGIS.zip, http://geomobile.como.polimi.it/website/docs/training/ISPRS2018/Land_Cover_Collector.zip) include practical guides for the intercomparison and validation of high-resolution LC maps using QGIS 2.18 open source 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 tools developed for validation require datasets with the same LC classification system, the same coordinate reference system, and in the case of raster map, the same resolution as well. Therefore, some preprocessing steps on the datasets are required to successfully perform validation. 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. (2018) and Bratic et al. (in press).

Validation with Vector Points Reference Dataset

The first case study is related to the validation of GL30 (2010 version) through a comparison against a reference vector points dataset of 2009 obtained from the Land Use and Coverage Area frame Survey (LUCAS) (<http://ec.europa.eu/eurostat/web/lucas/overview>) programme promoted by Eurostat. The area of interest is the Lombardy Region (Northern Italy). The guide shows how to use the tools available in QGIS 2.18 to aid in preprocessing and how to use the tool developed to perform the validation.

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. On the other hand, 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 a 0 value to a nodata value; and iii) reclassifying two LC classes (grassland and tundra) to merge them, since they are grouped into a unique class in LUCAS.

The steps described above are performed by means of standard tools provided by QGIS. Conversely, the last step, i.e. validation, is performed through an ad hoc PyQGIS script named *pts_lcval* (<https://github.com/GoricaB/Land-cover-validation>). Given as input the two preprocessed datasets, the tool 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. After that, the tool computes the confusion matrix between the two preprocessed datasets and its derived global and per-class statistics, which are returned to the user as CSV files together with the updated dataset of LUCAS points.

Validation with Raster Reference Dataset

The second case study proposes the validation of GL30 (2010 version) by means of a comparison with a local reference raster dataset available for 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 Lombardy Region.

As in the previous case, the guide explains the procedure to perform the comparison and shows how to carry out each step using QGIS 2.18 software. Since DUSAF is provided in vector format, a preprocessing step is required to rasterize it. Rasterization is performed according to its first level LC classification.

95 The resulting raster map has the same resolution as GL30. To harmonize the classification system of the
96 two datasets, GL30 is reclassified according to the DUSAF first level classification. After that, an ad hoc
97 PyQGIS script named *raster_lcval* (<https://github.com/GoricaB/Land-cover-validation>) is used to perform
98 the comparison between the two datasets. The tool developed requires the two preprocessed datasets as
99 inputs and performs a pixel-by-pixel comparison that results in the confusion matrix computation and the
100 global and per-class indexes.

101 Reference Data Collection

102 Land Cover Collector (<https://github.com/kilsedar/land-cover-collector>) is a free and open source applica-
103 tion, which allows the collection of LC field data for GL30 validation. It relies on the Apache Cordova
104 mobile application development framework (<https://cordova.apache.org>), and it is available for Android
105 and iOS devices, also on Web (<https://landcover.como.polimi.it/collector>). Leaflet (<https://leafletjs.com>) is
106 used for Web mapping. The LC points collected are stored using both PouchDB (<https://pouchdb.com>)
107 and CouchDB (<http://couchdb.apache.org>) to enable offline data collection. The LC class of each collected
108 point, the date of collection, the user's degree of certainty on the correctness of the stated classifica-
109 tion, the photos in north, east, south and west directions, and the user's comment (optional) are stored
110 in the database. Collected data can be visualized and queried on a map within the application, as il-
111 lustrated in Figure 1. Collected points are licensed under the Open Database License (ODbL) v1.0
112 (<https://opendatacommons.org/licenses/odbl/1.0/>), and can be downloaded within the application in JSON
113 format. The application is currently available in eight languages: English, Italian, Arabic, Russian,
114 Chinese, Portuguese, French and Spanish.

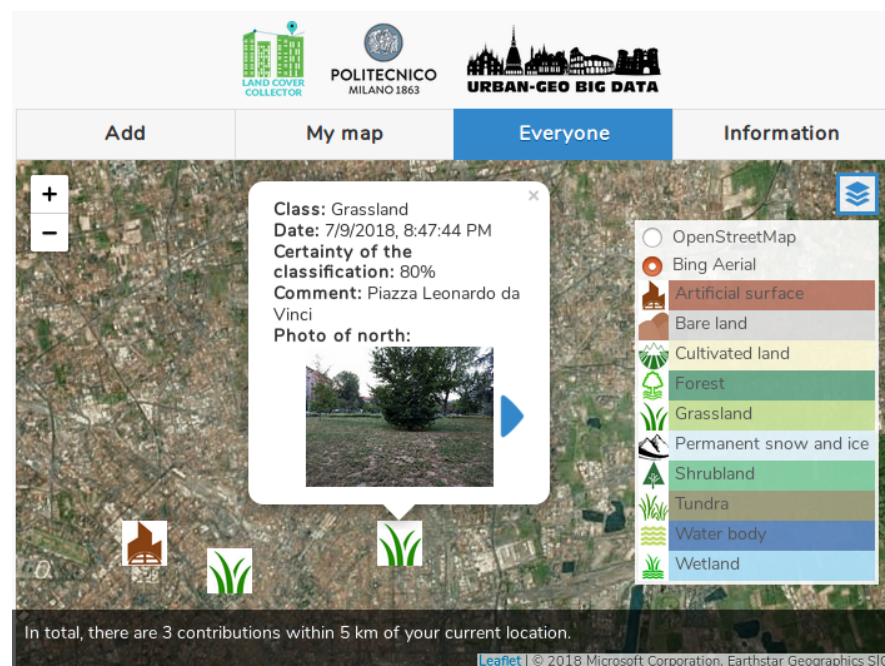


Figure 1. Visualization and query of collected data in Land Cover Collector.

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

119 CONCLUSIONS

120 Educational resources on the intercomparison and validation of LC maps were created based on the
121 needs identified by a literature review. The resources highlight the importance of validation and provide
122 easy-to-use tools and instructions for validation and reference data collection. Education resources are

prepared in the form of slide presentations, which were presented in three workshops (two of which held in developing countries) to fulfill the goal of the project as well as to promote their use within a wide community. Being the resources fully 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 all around the world. Future work includes using the data collected through the Land Cover Collector application to validate GL30.

ACKNOWLEDGEMENTS

This work is partially funded by the ISPRS 2018 Education and Capacity Building Initiative named “Capacity Building for High-Resolution Land Cover Inter-comparison and Validation” and the project URBAN GEO BIG DATA, a Project of National Interest (PRIN) funded by the Italian Ministry of Education, University and Research (MIUR) – id. 20159CNLW8.

REFERENCES

- Bratic, G., Brovelli, M. A., and Molinari, M. E. (2018). 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. (in press). 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*.
- Brovelli, M. A., Minghini, M., Molinari, M. E., Kilsedar, C. E., Wu, H., Zheng, X., Chen, J., and Shu, P. (in press). 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*.
- Brovelli, M. A., Minghini, M., Molinari, M. E., Wu, H., Zheng, X., and Chen, J. (2018). 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.
- Congalton, R. and Green, K. (2009). *Assessing the Accuracy of Remotely Sensed Data*. Boca Raton: CRC Press, 2 edition.
- 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.
- 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.