OPEN SOURCE SOFTWARE AND OPEN EDUCATIONAL MATERIAL ON LAND COVER MAPS INTERCOMPARISON AND VALIDATION

M. A. Brovelli [1], H. Wu [2], M. Minghini [1, *], M. E. Molinari [1], C. E. Kilsedar [1], X. Zheng [2], P. Shu [2], J. Chen [2]

[1] Department of Civil and Environmental Engineering, Politecnico di Milano, Piazza Leonardo da Vinci 32, 20133 Milano, Italy - (maria.brovelli, marco.minghini, moniaelisa.molinari, candaneylul.kilsedar)@polimi.it
[2] Department of Remote Sensing and Photogrammetry, National Geomatics Center of China, 28 Lianhuachi West Road, Haidian District, Beijing, 100830, China - (wuhao, xinyan_zheng, chenjun, pengshu)@ngcc.cn

Commission IV, WG IV/4

KEY WORDS: Capacity Building, Education, Land Cover, Open Access, Open Source, Validation

ABSTRACT:

Land Cover (LC) maps represent key resources to understand, model and address many global and local dynamics affecting our planet. They are usually derived from the classification of satellite imagery, after which a validation or intercomparison process is performed to assess their accuracy. This paper presents the project "Capacity Building for High-Resolution Land Cover Intercomparison and Validation", an educational initiative funded by the International Society for Photogrammetry and Remote Sensing (ISPRS) and mainly targeting developing countries. First, with the help of two open surveys, an analysis of the state of the art was performed which assessed the overall good awareness on LC maps and the needs and requirements for validating and comparing them, as well as the rich availability of educational material on this topic. The second task, currently under finalization, is the development of new educational material, based on open source software and released under an open access license, consisting of: an introduction to the GlobeLand30 (GL30) LC map and its online platform; a desktop GIS procedure showing two use cases on GL30 validation; and an application to collect LC data on the field to be used for validation. Finally, this educational material will be tested in practice in three workshops during the second half of the project, two of which held in developing countries: Dar es Salaam, Tanzania and Nairobi, Kenya.

1. INTRODUCTION

It is well known that we live in a rapidly changing world, where, compared to the past, it is much harder to model, control and forecast natural as well as anthropogenic dynamics. The ceaseless changes happening to the Earth environment at all scales, from local to global, such as those attributed to climate changes, point out the need to find new solutions in the geospatial domain. Providers of geospatial datasets should satisfy the demand for global, high-resolution (both spatial and temporal) and openly-licensed products, while the scientific community should provide robust, verified and replicable procedures for the assessment and validation of these products. This vision was fully shared within the Prague Declaration, approved in 2016 at the ISPRS XXIII Congress, which "calls on international communities to work together and promote multi-disciplinary collaboration towards providing reliable geospatial information to support societal transformations towards global sustainability" (ISPRS General Assembly, 2016). Land Cover (LC) maps represent a key class of global geospatial datasets, which are used as input for a variety of models (meteorological, ecological, biochemical, economic, hydrological, etc.) to monitor biodiversity and distribution of species, run climate models and assess environmental changes such as deforestation, desertification and urban expansion (see e.g. Foley et al., 2005; Feddema et al., 2005). Also, LC maps are one of the most powerful geospatial products able to measure the indicators for the 17 Sustainable Development Goals (SDGs), defined in the 2030 Agenda for Sustainable Development (United Nations, 2015) as well as to promote evidence-based policy-making on issues like soil consumption and deforestation.

LC maps are typically derived from the classification of satellite imagery, followed by a process of validation aimed at assessing their degree of adherence to the reality. The latter is performed through a spatial comparison of the classified LC dataset with a higher quality reference dataset representing the "ground truth", e.g. field surveys. Information on the accuracy of the classified map can be also obtained through the so-called intercomparison procedure, i.e. a benchmarking with another LC product. In both cases, the approach suggested by literature is the computation of a confusion matrix from which to extract many accuracy indexes, ranging from the most commonly used overall accuracy, user's and producer's accuracy to the more recently proposed allocation and quantity disagreements (Congalton and Green, 1999).

Currently, many countries and political organizations have their own high-resolution LC maps, but this is not the case for many developing countries. At the global level, traditionally course-resolution LC products are giving way to more detailed maps at a 30-m resolution such as the FROM-GLC (Finer Resolution Observation and Monitoring of Global Land Cover; Gong et al., 2013) and the GlobeLand30 (GL30; Chen et al., 2015). However, these high-resolution global LC maps still require validation at the international level to determine if they are useful for different applications. In parallel, open access to high-resolution satellite imagery such as those from Landsat and Sentinel has opened up the possibility for everyone to generate high-resolution LC maps. Not least, geospatial information generated from crowdsourcing initiatives such as OpenStreetMap (OSM; Mooney and Minghini, 2017) has a huge potential for the creation and update of LC maps (Fonte et al., 2017a). Therefore, it is crucial that both users and

---

* Corresponding author

producers of geospatial data in the domains of Geographic Information Systems (GIS) and Remote Sensing (RS) are able to apply scientific procedures for validating and comparing high-resolution LC maps, in order to validate one map against another (e.g. a local or self-produced map against a reference map) or to compare LC maps of the same area corresponding to different times in history to detect LC changes. The project "Capacity Building for High-Resolution Land Cover Intercomparison and Validation", described in this paper, addresses this need by developing outputs in the form of software solutions and educational material on the validation and intercomparison of high-resolution LC maps. The project outputs will be of wider utility, but will specifically target users in developing countries through the organization of some dedicated hands-on workshops.

The remainder of the paper is organized as follows. Section 2 provides an overview on capacity building on the validation and intercomparison of high-resolution LC maps by describing the aforementioned project in more detail. Sections 3 and 4 focus on two of the project main tasks, i.e. the survey on the state of the art and the development of educational material, respectively. The methodology designed and the results obtained are described in detail. Finally, Section 5 presents the future tasks of the projects and offers some directions for future research.

2. CAPACITY BUILDING ON LC MAPS VALIDATION

The project "Capacity Building for High-Resolution Land Cover Intercomparison and Validation" is funded by the International Society for Photogrammetry and Remote Sensing (ISPRS) within the series of Educational and Capacity Building Initiatives 2018 (http://www.isprs.org/society/ecbi/default.aspx). The main goal of the project, started in February 2018 and ending in January 2019, is to create new knowledge and skills on the validation and intercomparison of high-resolution LC maps. The project outputs will represent a useful capacity building tool for researchers and professionals in the fields of GIS and RS working in developing regions. The project workflow is outlined in Figure 1. The first task consists of an analysis of the current state of the art, which is meant as the combination of two separate analyses: on one side, the assessment of the awareness on the existence and importance of LC maps and the analysis of the needs, requirements and limiting factors in using and validating LC maps from a user's perspective, with special focus on developing countries; on the other side, the assessment and classification of the available educational material on the intercomparison and validation of high-resolution LC maps. Based on these results, the second task focuses on the development of new computer aided educational material on the intercomparison and validation of global high-resolution LC maps. This includes both teaching material (e.g. slide presentations and text documents), released under open access licenses, and software-based material (e.g. new modules for GIS software and code scripts/repositories), released under open source licenses to maximize its exploitation and impact. Finally, the third task consists in organizing three workshops on the intercomparison and validation of global high-resolution LC maps, where the developed material is tested in practice before its final release foreseen at the end of the project. Two of the three workshops are held in developing countries: Dar es Salaam, Tanzania and Nairobi, Kenya, both in September 2018, while the third one is held in Delft, The Netherlands, in October 2018. In accordance with the project timeline, at the time of writing (July 2018) the first project task has been completed, the second one is currently being finalized, while the third has been already setup and is ready for implementation. The next two sections describe the first and second tasks in more detail.
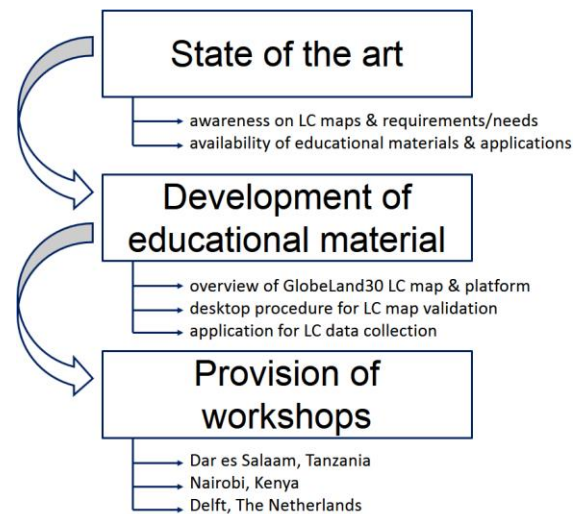

Figure 1. Workflow of the project "Capacity Building for High-Resolution Land Cover Intercomparison and Validation"

3. STATE OF THE ART

As mentioned before, the first task of the project consisted in the analysis of the state of the art about capacity building on LC maps and their validation and intercomparison, with special focus on developing countries. This initial step was essential to assess the current situation and to plan the following tasks accordingly, in particular in terms of the nature and content of the educational material to be developed. The following Subsections 3.1 and 3.2 describe the methodology adopted and the results achieved.

3.1 Methodology

The analysis of the state of the art was facilitated by the design of two surveys. The purpose of the first survey was to investigate the awareness about the importance and need of LC maps and to collect significant use cases for their production and validation. After asking for some personal information (name, surname, e-mail address, work affiliation and position), the survey included the following questions:
- Are you familiar with LC maps? (*yes/no answer*)
- Have you ever used LC maps in your research or professional work? (*yes/no answer*)
- If yes, which LC maps have you used? (*free text answer*)
- If yes, for which application(s) have you used the previous LC maps? (*free text answer*)
- Why do you think LC maps are useful, and thus their validation and intercomparison is important? Please provide at least two significant use cases/applications where the production, use or validation/intercomparison of Land Cover maps is important for your specific needs or interests. (*free text answer*)
- Which is the resolution of LC maps you believe is important for your applications? (*free text answer*)

The purpose of the second survey was to assess the current state of educational material and software applications focused on the intercomparison and validation of LC maps. In addition to the same personal information of the previous survey, the survey asked the following questions:
- Title/name of material - Provide the name or title of the training material/application. (*free text answer*)
- Type of material - Specify the type of material, e.g. text document, slides, code snippets, web application, etc. (*free text answer*)

- Software - Specify the GIS/Remote Sensing software, which is used by the learning material/application. (*free text answer*)
- License - Specify the license of training material/application or any other access restriction (answer NA if not applicable). (*free text answer*)
- Objective/Trainees - Describe the main objective of the learning material/application. Describe which trainees were trained by using the material (e.g.: professionals, university students, etc.). (*free text answer*)
- Land Cover maps - Describe which are the LC maps used in the learning material/application (answer NA if not applicable). (*free text answer*)
- Link to the training material - Enter the link to the training material, if it is available on the Internet. (*free text answer*)

To complete the analysis on the state of the art of the educational material, a separate review was performed by looking at both the available literature and the existing software packages in the field of GIS and RS.

Both the surveys were made available online to maximise their access. The first can be accessed at https://tinyurl.com/ydgg59ua, while the second can be accessed at https://tinyurl.com/yde9ykqg. To maximize the number of responses, the surveys were shared among several institutional, thematic, research and educational networks, including some specifically focused on developing countries. Examples are the Humanitarian OpenStreetMap Team (HOT, https://www.hotosm.org), the Missing Maps project (http://www.missingmaps.org), the YouthMappers global network (http://www.youthmappers.org) and the GeoForAll educational network (https://www.osgeo.org/initiatives/geo-for-all). Also, paper copies of the surveys were distributed at the last ISPRS TC III Symposium held in Beijing, China, on May 7-10, 2018.

3.2 Results

3.2.1 Survey on LC maps awareness and needs: The first survey was filled by 39 respondents from 24 different countries, including 5 countries from Europe, 8 from Africa, 9 from Asia, 1 from North America and 1 from Central America. In line with the project objective, most of the respondents were from developing countries (see Figure 2). The countries with most respondents were Kenya (7), Italy (4), United States (3), Nigeria (2), Bangladesh (2), Pakistan (2), and Tanzania (2). Respondents were mainly from universities (60%), followed by governmental organizations (e.g. the US National Aeronautics and Space Administration - NASA), inter-governmental organizations (e.g. the World Bank and the Food and Agriculture Organization - FAO), research institutes and non-governmental organizations (see Figure 3).

Figure 2. Distribution of the respondents to the first survey

Most of the respondents (92%) stated that they are familiar with LC maps and 85% of them have used LC maps in their research or professional work. In terms of the LC maps used, 1/3 of the respondents have used LC maps derived from Landsat imagery;

15% have used CORINE LC (a European product) and 13% have used GL30. Other LC maps have been used by about 5% of the respondents. These include LC maps derived from MODIS and, to a lesser extent, Sentinel imagery; the Global Land Cover maps from the United States Geological Survey (USGS); Urban Atlas (another European product); the Global Land Cover 2000 Project (GLC-2000); the GlobCover map; the products generated by the European Space Agency (ESA) Land Cover CCI project; the Global Urban Footprint (GUF) and the Global Human Settlement Layer (GHSL). Some respondents also pointed to local LC maps (e.g. COS, a LC map for Portugal), while some mentioned the use of OSM as a source of LC information.
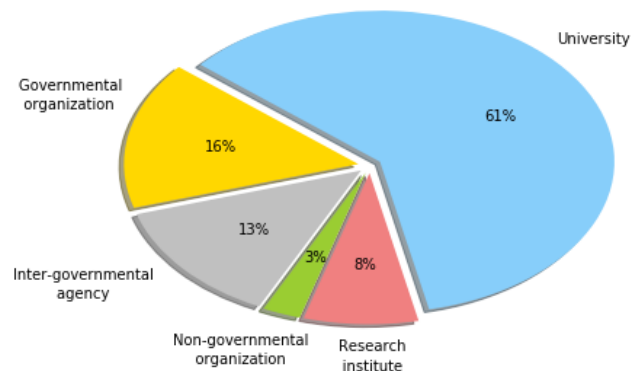

Figure 3. Affiliation types of the respondents to the first survey

Respondents highlighted the use of LC maps for a huge variety of applications: agricultural monitoring, disaster management, inventories of forests and greenhouse gases, resource mapping (e.g. forest and rangeland), LC change detection, environmental monitoring and assessment, modelling of LC projection, studies on climate change and food security, deforestation monitoring, census of LC categories, hydrological modelling of water quality, flood modelling, urban planning and urban growth analysis, study of disease incidence, and modelling of civil war impacts. LC maps were also used within several research applications, e.g. to compare and validate diverse LC maps, also with gamification approaches (see e.g. Fritz et al., 2009) as well as to create hybrid LC maps, including OSM-derived maps (Fonte et al., 2017b). Finally, the use of LC maps in RS and spatial analysis teaching at university was also reported. When asked about the reasons why the intercomparison and validation of LC maps is important, in addition to the applications listed above, respondents pointed out the need to count on accurate products to quantify and report environmental indicators, inform policy decision making towards sustainable development and societal well-being as well as guide error-free scientific research. Validation and intercomparison of LC maps can make users aware of the large disagreements existing between available LC maps, and can allow to identify the best datasets to be used at specific scales and for specific applications, thus reducing or mitigating the need to generate new products. Finally, respondents agreed that the best resolution of LC maps depends on the specific application, e.g. urban applications may need products at very high resolutions (i.e. <10 m) while maps at medium to high resolutions (i.e. between 30 and 100-500 m) may be sufficient for agricultural applications. However, respondents also agreed that high-resolution maps have a wider applicability than medium and low-resolution ones.

3.2.2 Survey on educational material on LC maps: Being more specific than the first survey, the second survey was filled by only 29 respondents from 21 different countries. Most of the respondents were the same as in the first survey; their geographic distribution is thus similar to the one shown in Figure 2, with the

exception of one new respondent for each of the following countries: Brazil, Canada, Nepal, The Netherlands and Uganda. However, the proportions of the respondents' affiliation types are pretty much the same as those shown in Figure 3.

Regarding the type of material, its license and the software it is based on, the emerging panorama is very heterogeneous. Half of the material reported by the respondents include standard documents, e.g. text/pdf documents or presentation slides; 29% of the material consist of specific websites or web applications; 13% include code snippets; and 8% feature multimedia content, i.e. videos and (in one case) an e-learning platform (see Figure 4). About 40% of this educational material on the intercomparison and validation of LC maps is based on the proprietary software ESRI ArcGIS (http://desktop.arcgis.com/en). This is followed by the open source software QGIS (https://qgis.org, 28%), which in many cases is used with ArcGIS within the same educational material. Other packages are the proprietary ERDAS IMAGINE (https://www.hexagongeospatial.com/products/power-portfolio/erdas-imagine, 17%) and the proprietary ENVI (https://www.harris.com/solution/envi, 21%) which again are often used in combination with other software within the same material. Less used packages reported include GRASS GIS (https://grass.osgeo.org), TerrSet (https://clarklabs.org/terrset), eCognition (http://www.ecognition.com), the IMPACT Toolbox (http://forobs.jrc.ec.europa.eu/products/software/impact.php), Google Earth for visualization, and specific mobile mapping apps such as LULC Mapper (https://mightysignal.com/a/google-play/com.servir.lulcmapper) and GEOVAL (Stylianidis et al., 2009). A software reported by the respondents is the LACO-Wiki web-based application (https://laco-wiki.net; See et al., 2017). It guides users in validating LC maps through a four-step process (upload of the dataset, creation and interpretation of the validation sample, and generation of validation results). Another notable web-based application reported by the respondents is the official platform of GL30 (http://www.globeland30.org), which offers a wealth of functions including visualization, download, extraction of statistics, and validation. This platform is described in detail in the following Subsection 4.1, since it is part of the educational material developed for the project. Finally, all the code snippets included in the educational material reported by the respondents were generated in Python and R.
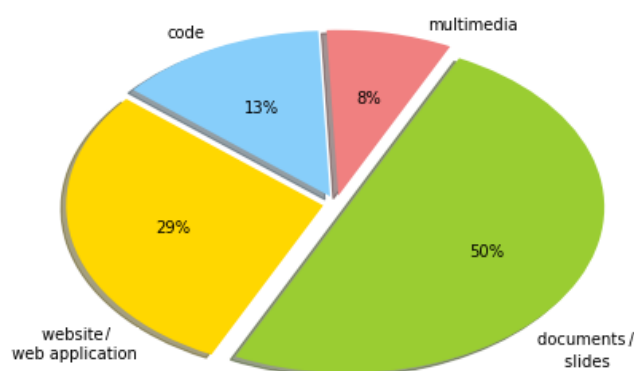

Figure 4. Types of educational material reported by the respondents to the second survey

Only about half of the respondents indicated the license of their educational material, while all the others simply answered with "NA". However, many among the respondents who indicated a license referred indeed to the license of the software instead of the material; therefore, these answers had to be discarded. Even the respondents who indicated a license for the material provided very generic answers such as "no license", "open" and "Creative Commons", which suggest that researchers and professionals are unfamiliar with licenses and/or they do not give them so much importance. Regarding the objective and the target users, most of the suggested educational material address (or were used to train) university students, followed by professionals (mainly experts in GIS and RS), researchers, governmental agencies and in one case also software developers (see Figure 5).
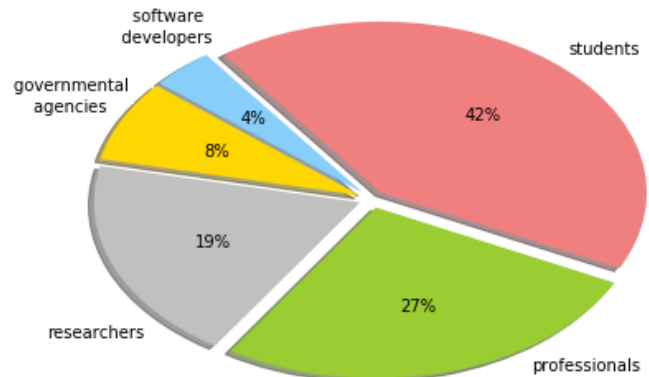

Figure 5. Target users for the educational material reported by the respondents to the second survey

As already mentioned, a separate review on available educational material on the validation and intercomparison of LC maps was also performed. In particular, an overview was made on the most relevant free and open source software packages for LC mapping, including both classification and validation/intercomparison, and the related educational material, when existing (Brovelli et al., 2018). In addition to the abovementioned QGIS and GRASS GIS, these software packages include SNAP (http://step.esa.int/main), Orfeo ToolBox (OTB, https://www.orfeo-toolbox.org) and SAGA (http://www.saga-gis.org). All the tutorials and training materials based on these software packages are available under open access licenses. Another review was performed on ready-to-use software applications allowing users to compute accuracy measures from a confusion matrix. A notable application is the Map Accuracy Tools website (http://tool.laco-wiki.net; Salk et al., 2018), which computes multiple accuracy metrics, both global and per-class. The underlying code is written in R and is available on the same website, although unlicensed. Another application is an Excel file downloadable from http://www2.clarku.edu/~rpontius, which given a confusion matrix allows again to compute a number of summary statistics (Pontius and Millones, 2011). It is also worth mentioning the dtwSat package, written in R and licensed under the open source General Public License (GPL; Free Software Foundation, 2007), which focuses on satellite image time series analysis and implements functions for LC map validation.

4. DEVELOPMENT OF EDUCATIONAL MATERIAL

As mentioned above, the following task of the project (currently under finalization) consists in the development of new educational material. This is composed of a combination of slides and code scripts, which will make full use of open source software and will be released under the open access Creative Commons Attribution 4.0 License (CC BY 4.0; Creative Commons, 2018) to maximize its diffusion and reuse. In addition, the educational material will be tested in practice during the three workshops to be carried out in the second half of the project duration. The educational material is composed of three parts, which are separately described in the following Subsections 4.1, 4.2 and 4.3: an introduction to GL30, which is the selected global, high-resolution LC map, and its

online platform; a desktop procedure for GL30 validation and intercomparison; and an application for collecting LC map data.

4.1 GL30 and its online platform

GL30 is the output of a global LC mapping programme, launched in China in 2010 with the aim of producing a 30 m LC product for the years 2000 and 2010 and within a four-year period. GL30 uses a classification scheme composed of 10 first-level classes, namely water bodies, wetland, artificial surfaces, cultivated land, permanent snow/ice, forest, shrubland, grassland, bareland and tundra. Satellite images used for generating GL30 are mainly 30 m multispectral images from Landsat TM and ETM+ and from the Chinese Environmental Disaster Alleviation Satellite (HJ-1). The classification approach adopted for GL30 was pixel-object-knowledge (POK) based, i.e. it consisted of the three consecutive steps of pixel-based classification, object-based labeling, and knowledge-based verification (Chen et al., 2015). The overall accuracy of the 2010 version of GL30 reaches 80.33%, however this global LC dataset is still undergoing validation to assess its usability for different applications at different scales (see e.g. Jokar Arsanjani et al., 2016).

After an introduction to GL30, the educational material focuses on its official web-based platform (http://www.globeland30.org). This platform provides several functions related to GL30. The first one is data visualization and browsing, which allows users to navigate the 2000 and 2010 versions of GL30, Landsat imagery and other LC map layers (e.g. CORINE LC and GlobCover) up to the pixel level. An ad hoc function allows to split the screen to visually compare any two of these layers (see Figure 6). For each of the 10 first-level classes of GL30, the platform provides a number of links to display the web map centered on significant areas of the world corresponding to that class (e.g. a list of deserts for the bareland class). A 3D globe visualization of GL30 with a number of ancillary layers, e.g. base maps and Digital Elevation Models (DEMs), is also available.
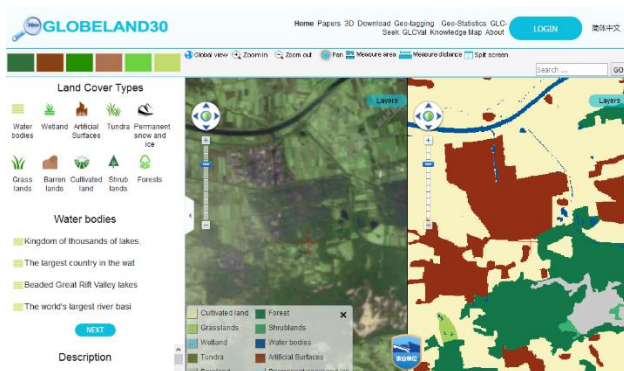

Figure 6. GL30 online platform showing a comparison between a satellite imagery (left) and the 2010 version of GL30 (right)

The second function provided is GL30 data download, which is only available after creating an account and logging in to the platform. Users can select the area for download by entering the GL30 map sheet number, the coordinates of the desired bounding box or by drawing a polygon on the map. Users are then presented the list of the GL30 map sheets corresponding to their selection, that they can download after visualizing and accepting the GL30 product utilization agreement. Another available function is the extraction of GL30-related statistics. Users can select an area of interest (by choosing a country from a drop-down menu or from the map, or by drawing a polygon on the map) and statistics and graphs are displayed on the GL30 class distribution in that area;

statistics on the comparison between the 2000 and 2010 versions of GL30 are also provided. Finally, the GL30 online platform offers a validation function. Users have first to select the region where they want to validate GL30: by selecting the country and the administrative region of interest from a drop-down menu, or by uploading a file defining the bounding box. Users have then to create a validation sample by defining the sampling method (landscape shape index or spatial autocorrelation), the selection method (stratified or random) and the confidence level. Once the sample is visualized on the map, users can perform manual validation using several satellite images or by connecting to an external Web Map Service (WMS) layer. After validating each sample point (i.e. indicating whether the GL30 classification of that point is correct or not), users can visualize the confusion matrix and the related statistics.

4.2 Desktop procedure for GL30 validation

The second part of the educational material consists of a desktop procedure for LC map validation, implemented in QGIS and using again GL30 as sample dataset. The material shows two different ways to validate GL30, i.e. using first a vector layer of points and then a raster map as the reference dataset. In both cases, the GL30 and the reference dataset are required to have the same LC classification scheme and the same coordinate/reference system; in the case of raster data, they must also have the same resolution. This requires that some pre-processing is made on the datasets involved. The test area chosen for the validation exercises is Lombardy Region (northern Italy).

The first use case shows how to validate the 2010 version of GL30 through comparison against data from the Land Use Cover Area frame Sample (LUCAS, http://ec.europa.eu/eurostat/web/lucas), which is the European official LC in-situ dataset, derived from rigorous surveys carried out every 3 years at points located across EU member states (Eurostat, 2015). In detail, the 2009 LUCAS dataset is used for the validation. The educational material shows the procedure to perform validation using the tools available in QGIS. First, pre-processing of the LUCAS dataset is required to: a) extract the LUCAS points in the region of interest; b) reproject this dataset into the GL30 reference system (WGS84/UTM 32N); and c) reclassify the dataset according to the GL30 nomenclature. GL30 also needs some pre-processing. First, since it is provided in tiles, a merging of two tiles is required to obtain a single dataset for Lombardy Region. Moreover, the cells having a value equal to 0 must be set as NoData. Finally, a reclassification is required to merge two LC classes (grass and tundra) which are considered as a single class in LUCAS. While all these steps are performed using the standard QGIS tools, validation is performed through a custom script for PyQGIS, named *pts_lcval.py* and available at https://github.com/GoricaB/Land-cover-validation. PyQGIS offers scripting support in QGIS using the Python language (https://docs.qgis.org/testing/en/docs/pyqgis_developer_cookbook). Given in input the pre-processed LUCAS and GL30 datasets (shown in Figure 7), the script adds a new attribute to the LUCAS dataset and updates it with the GL30 LC class at the location of each point. Then, the script computes the confusion matrix and the related global and per-class statistics, which are returned to the user in CSV format together with the updated LUCAS dataset.

In the second use case, the 2010 version of GL30 is validated through comparison against a 2012 reference raster map derived from DUSAF (Destinazione d'Uso dei Suoli Agricoli e Forestali, *Use Destination of Agricultural and Forest Land*), i.e. a land use/ land cover dataset available for Lombardy Region at the scale 1:10000. The area chosen for validation is limited to the province of Como, located in the northern part of Lombardy Region. Again,

the educational material shows how to perform validation using the available QGIS tools. Since DUSAF is originally provided in vector format, a pre-processing step is first required to rasterize it (in particular its first-level classification) according to the same resolution of GL30. GL30 is then reclassified to match the LC nomenclature of DUSAF. Finally, the comparison between GL30 and DUSAF is performed by another PyQGIS script, named *raster_lcval.py* and hosted at https://github.com/GoricaB/Land-cover-validation. The script requires the two raster datasets in input and performs a pixel-by-pixel comparison to generate the confusion matrix and the global and per-classes statistics.
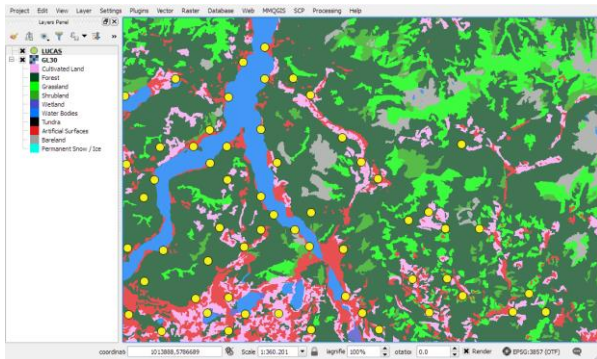

Figure 7. Screenshot of QGIS showing the pre-processed GL30 and LUCAS datasets in Lombardy Region, Italy.

4.3 Application for LC data collection

As the last part of the educational material, an application was also developed to allow users to collect field data according to the same LC nomenclature of GL30. This data can then be used to validate GL30, e.g. using the same methodology described in Subsection 4.2. The application, named Land Cover Collector, is released under the GPL 3.0 (Free Software Foundation, 2007); its source code is available at https://github.com/kilsedar/land-cover-collector. The architecture of the application, depicted in Figure 8, is based on the Apache Cordova mobile application development framework (https://cordova.apache.org), thus it is available as a mobile application for Android and iOS as well as a web application at https://landcover.como.polimi.it/collector. Being responsive, the latter can be accessed from both desktop and mobile browsers. Leaflet (https://leafletjs.com) application programming interface (API) is used for web mapping. The data is initially stored locally using PouchDB (https://pouchdb.com) and then replicated to CouchDB (http://couchdb.apache.org) on the server when an Internet connection is available. This method enables offline data collection, which is vital for collecting LC data in remote places.
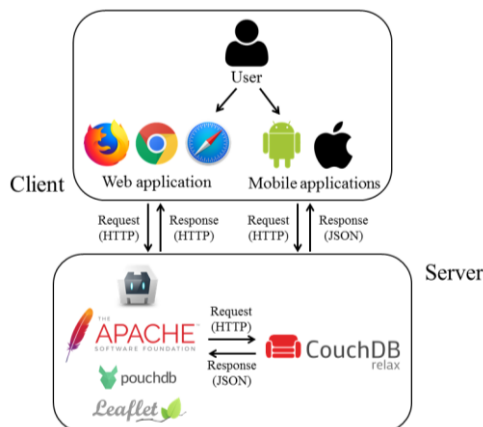

Figure 8. Architecture of the Land Cover Collector application.

The application allows for point-based collection of LC data. When collecting a LC observation, in addition to enabling their GPS, users are required to indicate the LC class according to the GL30 nomenclature (see Figure 9a) and the degree of certainty; to take four photos in the north, east, south and west directions; and optionally to add a comment. A button available on the application explains the guidelines for data collection. The collected LC data can be visualized as clusters, aggregated based on the declared LC class (see Figure 9b). The details of each single point can be queried by clicking it on the map (see Figure 9c). The application is currently available in eight languages: English, Italian, Arabic, Russian, Chinese, Portuguese, French and Spanish. Collected data are released under the Open Database License (ODbL; Open Data Commons, 2018) and can be downloaded in JavaScript Object Notation (JSON) format.
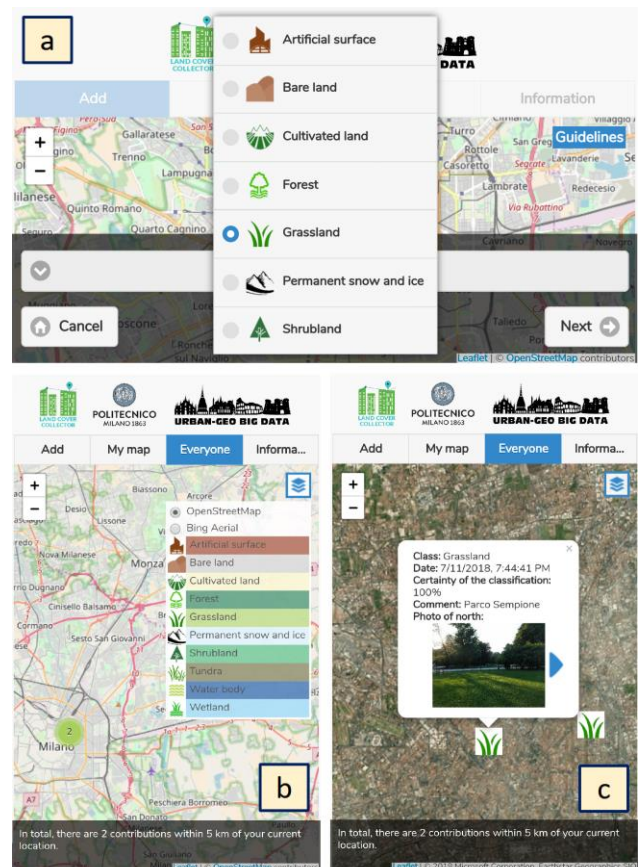

Figure 9. Screenshots of the Land Cover Collector application showing LC classification (a), visualization of data aggregated in clusters (b) and query of a collected point (c)

5. CONCLUSIONS

The availability of accurate and up-to-date geospatial datasets is crucial to understand, model and address the global dynamics affecting our planet. LC maps represent a key resource to study many of these phenomena, which are both causes and effects of global climate changes. The purpose of the project "Capacity Building for High-Resolution Land Cover Inter-comparison and Validation", presented in this paper, is exactly to raise awareness about the importance and use of LC maps, in particular on their validation and intercomparison which is a fundamental step for their exploitation. Developing countries are expected to benefit the most, since in these countries global maps are often the only LC source available. As already mentioned, in addition to the

development and release of openly-licensed educational material, the project will specifically target developing countries through the provision of workshops in Tanzania and Kenya. These will also help to adapt and finalize the educational material before its final release at the end of the project.

In addition, the capacity to validate and compare LC products has become increasingly important thanks to at least two new sources of LC information. On one side, several countries and political organizations have established their own Earth Observation programmes to provide users with free high-resolution imagery on a daily basis. The most ambitious among these is Copernicus (http://copernicus.eu), which exploits imagery delivered from the Sentinel satellites of the European Space Agency (ESA) with the mission to establish a European capacity for Earth Observation, where LC monitoring is one of the core services. On the other side different sources of crowdsourced geographic information, which provide a continuous flow of free geospatial data, have been recognized useful for LC purposes. In addition to OSM, geotagged photographs represent a notable example (Antoniou et al., 2016). Topics such as the validation of these new types of in situ observations and their integration with remotely-sensed data to derive new and advanced LC products will offer fertile ground for the research of the years to come.

REFERENCES

Antoniou, V., Fonte, C.C., See, L., Estima, J., Jokar Arsanjani, J., Lupia, F., Minghini, M., Foody, G.M., and Fritz, S., 2016. Investigating the feasibility of geo-tagged photographs as sources of land cover input data. *ISPRS International Journal of Geo-Information*, 5(5), 64.

Brovelli, M.A., Minghini, M., Molinari, M.E., Wu, H., Zheng, X., and Chen, J., 2018. Capacity building for high-resolution land cover itercomparison and validation: what is available and what is needed. In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Dar es Salaam, Tanzania, Vol. XLII-4, Part W8, pp. 15-22.

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, pp. 7-27.

Congalton, R.G., and Green, K., 1999. *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices*. CRC Press, Boca Raton.

Creative Commons, 2018. Attribution 4.0 International (CC BY 4.0) https://creativecommons.org/licenses/by/4.0 (13 July 2018).

Eurostat, 2015. LUCAS 2015 (Land Use / Cover Area Frame Survey). Technical reference document C1 Instructions for Surveyors. http://ec.europa.eu/eurostat/documents/205002/6786255/LUCAS2015-C1-Instructions-20150227.pdf (14 July 2018).

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), pp. 1674-1678.

Foley, J.A., DeFries, R., Asner, G.P., Barford, C., Bonan, G., Carpenter, S.R., Chapin, F.S., Coe, M.T., Daily, G.C., Gibbs, H.K., Helkowski, J.H., Holloway, T., Howard, E.A., Kucharik, C.J., Monfreda, C., Patz, J.A., Prentice, I.C., Ramankutty, N., and Snyder, P.K., 2005. Global Consequences of Land Use. *Science*, 309(5734), pp. 570-574.

Fonte, C.C., Patriarca, J.A., Minghini, M., Antoniou, V., See, L., and Brovelli, M.A., 2017a. Using OpenStreetMap to Create Land Use and Land Cover Maps: Development of an Application. In: *Volunteered Geographic Information and the Future of Geospatial Data*. IGI Global, Hershey, pp. 113-137.

Fonte, C.C., Minghini, M., Patriarca, J., Antoniou, V., See, L., and Skopeliti, A., 2017b. Generating up-to-date and detailed land use and land cover maps using OpenStreetMap and GlobeLand30. *ISPRS International Journal of Geo-Information*, 6(4), 125.

Free Software Foundation, 2007. GNU General Public License, https://www.gnu.org/licenses/gpl-3.0.en.html (14 July 2018).

Fritz, S., McCallum, I., Schill, C., Perger, C., and Grillmayer, R., 2009. Geo-Wiki.Org: The use of crowdsourcing to improve global land cover. *Remote Sensing*, 1(3), pp. 345-354.

Gong, P., Wang, J., Yu, L., Zhao, Y.C., Zhao, Y.Y., Liang, L., Niu, Z.G., Huang, X.M., Fu, H.H., Liu, S., Li, C.C., Li, X.Y., Fu, W., Liu, C.X., Xu, Y., Wang, X.Y., Cheng, Q., Hu, L.Y., Yao, W.B., Zhang, H., Zhu, P., Zhao, Z.Y., Zhang, H.Y., Zheng, Y.M., Ji, L.Y., Zhang, Y.W., Chen, H., Yan, A., Guo, J.H., Yu, L., Wang, L., Liu, X.J., Shi, T.T., Zhu, M.H., Chen, Y.L., Yang, G.W., Tang, P., Xu, B., Ciri, C., Clinton, N., Zhu, Z.L., Chen, J., and Chen, J., 2013. Finer resolution observation and monitoring of global land cover: first mapping results with Landsat TM and ETM+ data. *International Journal of Remote Sensing*, 34(7), pp. 2607-2654.

ISPRS General Assembly, 2016. Prague Declaration. http://www.isprs.org/congresses/prague2016/The_Prague_Declaration_final.pdf (10 July 2018).

Jokar Arsanjani, J., See, L., and Tayyebi, A., 2016. Assessing the suitability of GlobeLand30 for mapping land cover in Germany. *International Journal of Digital Earth*, 9(9), pp. 873-891.

Mooney, P., and Minghini, M., 2017. A review of OpenStreetMap data. In: *Mapping and the Citizen Sensor*. Ubiquity Press, London, pp. 37-59.

Open Data Commons, 2018. Open Database License (ODbL) v1.0, https://opendatacommons.org/licenses/odbl/1-0 (14 July 2018).

Pontius Jr, R.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), pp. 4407-4429.

Salk, C., Fritz, S., See, L., Dresel, C., and McCallum, I., 2018. An Exploration of Some Pitfalls of Thematic Map Assessment Using the New Map Tools Resource. *Remote Sensing*, 10(3), 376.

See, L., Laso Bayas, J.C., Schepaschenko, D., Perger, C., Dresel, C., Maus, V., Salk, C., Weichselbaum, J., Lesiv, M., McCallum, I., Moorthy, I., and Fritz, S., 2017. LACO-Wiki: A new online land cover validation tool demonstrated using GlobeLand30 for Kenya. *Remote Sensing*, 9(7), 754.

Stylianidis, E., Roustanis, T., and Karanikolas, N., 2009. A Geographical Information System for real estate (GEOVAL). In: *Location Based Services and TeleCartography II*. Springer, Berlin, pp. 317-329.

United Nations, 2015. Transforming our world: The 2030 agenda for sustainable development. A/RES/70/1. https://sustainabledevelopment.un.org/content/documents/21252 030%20Agenda%20for%20Sustainable%20Development%20w eb.pdf (14 July 2018).