VALIDATION OF THE GLOBAL HIGH-RESOLUTION GLOBELAND30 LAND COVER MAP IN EUROPE USING LAND COVER FIELD SURVEY DATABASE - LUCAS

G. Bratic[1], M. E. Molinari[1], M. A. Brovelli[1,*]

[1] Politecnico di Milano, Department of Civil and Environmental Engineering, Piazza Leonardo da Vinci 32, 20133 Milan, Italy -
(gorica.bratic, moniaelisa.molinari, maria.brovelli)@polimi.it

Commission IV, WG IV/4

KEY WORDS: Global Land Cover, Thematic Accuracy Assessment, Free and Open Source, Error Matrix, Inexpensive validation, LUCAS

ABSTRACT:

High-resolution land cover maps are one of the technological innovations driving improvements in many fields influenced by Geographic Information Systems (GIS) and Remote Sensing. In particular, the GlobeLand30 (GL30), global LC map with spatial resolution of 30m, is thought to be one of the highest quality high-resolution products. However, these LC maps require validation to determine their suitability for a particular purpose. One of the best ways to provide useful validation reference data is to do a high-level accuracy field survey, but this is time consuming and expensive. Another option is to exploit already available datasets. This study assesses thematic accuracy of GL30 in Europe using LUCAS as a validation reference, because it is a free and open field survey database. The results were generally not good, and very bad for some classes. Analysis was then restricted to a small region (Lombardy, Italy) where LC data of higher resolution than those of GL30 were available. LUCAS was also found to be incoherent with this product. Further comparisons of LUCAS with other independent sources confirmed that the LC attributes of LUCAS are inconsistent with expectations. Although these findings may not be generalized to other regions, the results warn against the suitability of LUCAS as ground truth for LC validation. The paper discusses the process of thematic accuracy assessment of the GL30 and the applicability of LUCAS for high-resolution global LC validation.

1. INTRODUCTION

Land cover (LC) maps are categorical-type models of the Earth's surface that represent materials covering it, like water, forest, urban areas, etc. They are the products of classification procedures, done on satellite/aerial imagery by some classification algorithm. The maps can be employed for a variety of applications, including understanding climate change, natural resource management and conserving biodiversity. LC often provides valuable information for remote, hardly accessible areas and in developing countries where official data are frequently missing.

Some existing LC maps (Table 1) depict only one specific theme and they are characterized only by two classes like Global Human Settlement Layer (Pesaresi et al., 2016), Global Urban Footprint (Esch et al., 2013), Global Surface Water (Pekel et al., 2016), Forest/Non-forest map (Shimada et al., 2014), Tree Canopy Cover (Hansen et al., 2013), Global Forest Cover Gain (Hansen et al., 2013), and Global Forest Cover Loss (Hansen et al., 2013)); others are multi-class like GlobeLand30 (Chen et al., 2015), FROM-GLC (Gong et al., 2013)). Some have multi-temporal coverage. Among all global high resolution LC maps, the study focuses on GlobeLand30 (GL30) because it is multi-class and have more temporal repetition than FROM-GLC, and therefore it can potentially meet the needs of many users.

The aim of this study was to compute the accuracy indexes of GL30 for Europe, providing useful indicators that can help users to decide if the quality of the map is suitable for a specific target objective (e.g. monitoring climate change, capturing perma-

Global map name	Resolution	Year
GlobeLand30 (GL30)	30m	2000, 2010, 2015[1]
FROM-GLC	30m	2010, 2015
Global Urban Footprint	12m	2011
Global Human Settlement Layer (GHS BUILT-UP GRID S1)	20m	2016
Global Surface Water	30m	1984 - 2015
Forest / Non-Forest map	25m	2007 - 2010 2015 - 2016
Tree canopy cover	30m	2000
Global forest cover gain	30m	2000-2012
Global forest cover loss	30m	2000-2015

Table 1. Existing high-resolution global land cover maps

nent snow and ice melting rates, deforestation rates, urban planning, etc.). Accuracy indexes are computed from an error matrix, which is derived by the comparison of GL30 against a reference dataset. In this study, the LUCAS (Land Use/Cover Area frame Statistical survey) dataset for the year 2009 was chosen as the reference dataset for validation of GL30 2010, because it contains authoritative in-situ observations available free of charge.

For its characteristics, LUCAS in theory could be an optimal dataset for assessing land cover maps. Despite that, as far as the authors have been able to find, very few studies are available in literature about land cover assessment using LUCAS. Namely, in one study (Gallego, 2011), CORINE Land Cover 2000 (CLC2000) was validated. The author recognized that the different minimum mapping units (MMU) of LUCAS (7m2) and

---

*Corresponding author
[1] The GL30 map for 2015 is announced, but not published yet.

CLC2000 (25ha) was partially affecting validation results due to the inability of CLC2000 to capture features in as much detail as LUCAS. Another contribution to error was attributed to photo-interpretation mistakes in LUCAS. Moreover, certain classes did not contain a sufficient number of sample points, so per-class accuracy indexes (Omission and Commission error) could not be confidently estimated. In another study, Karydas et al. (2015) used LUCAS 2009 to validate the 2007 WWF Hellas LC map of Greece, with 30m resolution. A two-tier methodology was employed: first, a so called, "automated" process based validation directly on exploiting information about LC of LUCAS data; and second, a "supervised process" reinterpreted photos collected in the LUCAS campaign. The accuracy of the "supervised" procedure was lower than of the "automated" procedure, which the authors ascribed to misclassification, probably due to mistakes the surveyors made in the implementation of the LUCAS protocol in Greece. Both of the studies used non-squared matrices due to having different classes of reference and classified datasets.

This paper started with the idea of assessing GL30 2010 for all of Europe using LUCAS 2009 (the two datasets are more or less contemporary), considering that the difference between their MMUs is definitely smaller than the one between LUCAS and CLC 2000. Also, it is possible to directly link the classes of GL30 and LUCAS using a standard squared error matrix, which is an advantage since non comparable classes could give misleading results. What was found is that GL30 is not coherent with LUCAS, especially in some classes. Taking into account both, these results and the outcomes of the previously mentioned papers, the study included a comparison of LUCAS to other independent sources of information (authoritative and not) to determine if there is a systematic error in the data that can be removed so that the potential of the LUCAS database can be exploited.

In the following sections, the steps of the work and the results obtained are reported. Namely, section 2 describes the data used, section 3 presents the processing of the data, and sections 4 and 5 are related to the assessment results in Europe and in a small region where the LUCAS dataset was more deeply analyzed. At the end, in section 6, some conclusions are drawn.

2. DATASETS

In this section the different datasets used in the work are described. GlobeLand30 is the dataset that was chosen to be assessed and LUCAS was chosen as a ground truth reference. The last dataset, DUSAF, was considered when some doubtful and unexpected results were found.

2.1 GLOBELAND30

Two global maps for the baseline years of 2000 and 2010 are produced by the National Geomatics Center of China (NGCC) as a result of the project "Global Land Cover Mapping at Finer Resolution" (Chen, 2010).

GL30 (http://www.globallandcover.com) is derived by classification of TM5 and ETM + of the United States' Land Resources Satellite (Landsat) imagery and the multispectral images of China's Environmental Disaster Alleviation Satellite (HJ-1); all of them are multispectral images with 30m spatial resolution. Additionally, auxiliary information from existing land cover data and online resources significantly contributed to the map making process.

The spatial resolution of GL30 (30m) means that this map is appropriate enough to capture even small LC changes, including those due to human impact. GL30 is provided in a form of Geo-Tiff raster tiles in WGS84 (World Geodetic System 1984) reference system and UTM (Universal Transverse Mercator) projection. EPSG code for WGS 84 / UTM projection is 32yxx, where y is 7 where the tiles are located in the southern hemisphere or 6 when tiles are in the northern hemisphere, while xx depends on which out of 60 zones (each 6° of longitude in width) the tile belongs to.

In this paper, validation was conducted for 23 European countries (those containing LUCAS points) covered by 30 tiles (size: 6° (longitude) x 5° (latitude)) of the more recent NGCC product (for year 2010) with 10 land cover types: Cropland, Forest, Grassland, Shrubland, Wetland, Water Bodies, Tundra, Artificial Surfaces, Bareland and Permanent Ice and Snow.

2.2 LUCAS

LUCAS (http://ec.europa.eu/eurostat/web/lucas) is a project led by the Statistical Office of the European Commission (Eurostat) (ESTAT, 2015). The LUCAS survey is conducted every 3 years and new versions of the dataset are published accordingly. The selected dataset is composed of points referenced in the WGS84 coordinate system (EPSG:4326) and obtained in 2009 in 23 European countries (hereafter Europe)(Belgium, Czech Republic, Denmark, Germany, Estonia, Ireland, Greece, Spain, France, Italy, Latvia, Lithuania, Luxembourg, Hungary, Netherlands, Austria, Poland, Portugal, Slovenia, Slovakia, Finland, Sweden, and the United Kingdom), where field surveyors created a collection of around 235,000 points including land use, LC, and many other types of information.

The LUCAS sampling strategy (Martino et al., 2009) consists of two phases. First, in the stratification phase, points found on the intersection of 2km^2 x 2km^2 systematic grid are photo-interpreted and assigned to one of the predetermined classes. Based on the first-phase (master) sample, the second-phase (field) sample is chosen using a simple random sampling method. The overall sampling rate (rate between number of points in field samples and in the master sample) is about 25%. Samples are extracted independently in each NUTS2 (a basic region for the application of regional policies) and in every stratum, and the sampling rate per stratum is modified separately for each NUTS2 region.

Inaccessible points (above 1000 meters and far from the road network) are discarded from the field samples, and often photo-interpreted, to reduce the cost of the survey (Martino et al., 2009).

LUCAS land cover nomenclature has 8 main classes: Artificial land, Cropland, Woodland, Shrubland, Grassland, Bareland, Water and Wetlands. Every main class has several subclasses.

2.3 DUSAF

DUSAF is a regional land cover database for the Lombardy region, in Northern Italy, produced since 2000 (Bonomi et al., 2011). In the Lombardy Region geoportal (http://www.geoportale.regione.lombardia.it/en) DUSAF land cover vector maps for different time periods are available at 1:10,000 scale, in Shapefile format and in WGS84 reference system with UTM 32 N projection (EPSG:32632). The adopted legend (i.e. the existing classes) is designed in five hierarchical

levels. The first level consists of five classes: Artificial surfaces, Agricultural areas, Forest and Semi Natural Areas, Wetlands, and Water Bodies; every next level contains subclasses of the previous level. The map selected for this study is DUSAF 4.0 from 2012. DUSAF has a higher spatial resolution than GL30 (its accuracy is around 2m) and better detail in the definition of the classes. Therefore, it can be considered as a ground truth reference for assessing the quality of GL30.

3. PROCESSING

The study performs preprocessing (i.e. preparation of the two datasets), validation, and analysis of the accuracy indexes derived from validation. Two open-source software applications, QGIS (QGIS Development Team, 2018) and GRASS GIS (GRASS Development Team, 2017) were used for the preprocessing, while for validation two tools were developed through an ad hoc Python script for the QGIS desktop GIS software. The tools samples the pixel values of the map under assessment (in this case GL30) corresponding to points or pixels where ground truth observations (in this case LUCAS or the rasterized DUSAF) are present. Based on these sampled pixels and the ground truth observations, an error matrix is created. This is the basic element for the computation of several accuracy indexes proposed in literature. Some indexes indicate the global accuracy of the map (e.g. the Overall Accuracy, Kappa index, etc.), while others are used for accuracy assessment of the individual LC classes available in the classified map (e.g. the Producer's accuracy and the User's accuracy).

The validation can be performed by using two types of reference datasets: vector points or raster data. Moreover, the methodology requires reference and classified datasets with the same classification system, the same coordinate/reference system, and in case of raster data, the same spatial resolution as well. Therefore, some preprocessing steps of datasets were required to successfully perform the validation. LUCAS preprocessing includes: i) removal of the classified points with not valid (null) coordinates due to interrupted GPS signal during survey (denoted as 'X' in GPS column of LUCAS attribute table, ii) reprojection to ETRS89/ETRS-LAEA (EPSG:3035) projection to have a unique coordinate/reference system for all of Europe, and iii) addition of a column to the attribute table with current LUCAS notation translated to the notation of GL30 (by rules shown in Table 2) so that they can be directly compared during validation.

To prepare the raster data, GL30 tiles were firstly reprojected from UTM projection to ETRS89/ETRS-LAEA (EPSG:3035) and then merged to obtain a unique dataset for the 23 countries in Europe. The tiles contain value 0 due to the reprojection in UTM, since a regular tile in geographic coordinates does not correspond to a rectangle in projected coordinates. The 0 values were removed since they do not contain information about classification. Finally, GL30 was reclassified according to the rules in the Table 2.

Validation was made by an ad-hoc PyQGIS script named *pts_lcval* (https://github.com/GoricaB/Land-cover-validation). *pts_lcval* tool samples the classified map (GL30) in the position of the points of the reference map (LUCAS) and adds the sampled value in the new column of the reference map. In this way, values of the classified map are as vector points, not raster. Next, the confusion matrix, and all of the indexes of accuracy (global and per-class) were computed. For validation of GL30 in the specific country, the points of LUCAS covering the country are selected prior to running the tool.

	LUCAS		GlobeLand 30
A	Artificial land	80	Artificial surfaces
B	Cropland	10	Cultivated land
C	Woodland	20	Foresst
D	Shrubland + Tundra	40	Shrubland
		70	Tundra
E	Grassland	30	Grassland
F	Bareland	90	Bareland
G	Water	60	Water Bodies
G50	Permanent Snow and Glaciers	100	Permanent Snow and Ice
H	Wetland	50	Wetland

Table 2. Reclassification rules for GL30

Validation based on the vector dataset as a reference required different preprocessing. Being a vector layer, DUSAF 4.0 needed to be rasterized. Although, the scale of DUSAF is 1:10,000 (about 4m resolution), it was rasterized to 30m resolution to correspond to GL30. The change of resolution does not significantly affect the outcome of the analysis (Brovelli et al., 2015). It was also reclassified to match the classes of GL30 using rules from Table 3. Tundra class does not exist in Lombardy, so it was left out. Two tiles of GL30 are covering the whole area of Lombardy, and therefore GL30 preprocessing for this type of validation merged the two tiles. Moreover, since the two maps are originally in the same coordinate/reference system, reprojection was not needed. At this point, both datasets were prepared to be input for the validation tool *raster_lcval*, which compares datasets pixel by pixel, computing the confusion matrix and many accuracy indexes.

DUSAF level 3		GL30
211, 213, 221, 222, 223, 224, 231	10	Cropland
311, 312, 313, 314	20	Forest
321	30	Grasslands
322, 324	40	Shrubland
411	50	Wetlands
511, 512	60	Water Bodies
111, 112, 121, 122, 123, 124, 131, 132, 133, 134, 141, 142	80	Artificial surfaces
331, 332, 333	90	Barelends
335	100	Permanent Ice and Snow

Table 3. Reclassification rules for DUSAF

4. VALIDATION OF GL30 WITH LUCAS IN EUROPE

GL30 was validated using the LUCAS database in two cases; in the former, the validation was performed for all of Europe, while in the latter the indexes were computed for each country separately.

4.1 Europe

Several indexes are drawn from the error matrix, which, according to Labatut and Cherifi (2011), Congalton and Green (2009), Koukoulas and Blackburn (2001), Pontius and Millones (2011), can be useful for map accuracy assessment. Global indexes are shown in Table 4, while per-class indexes are shown in Table 5. The values of the indexes range from 0 to 1. For indexes of agreement (Overall Accuracy - P0, Margfit Overall Accuracy - $P0_{marg}$, Classification Success Index - CSI, Kappa - K, User's Accuracy

- UA and Producer's Accuracy - PA), the value 1 denotes perfect agreement and value 0 denotes absolute disagreement; for indexes of disagreement, Allocation Disagreement - AD and Quantity Disagreement - QD, the opposite is true (i.e. 1 is perfect disagreement).

The most common index is P0, whose value should be at least 0.85 according to Anderson et al. (1976) and Thomlinson et al. (1999), or 0.7 according to Pringle et al (2009) in order to say that the map is accurate enough. The value of P0 is 0.55. $P0_{marg}$, which is similar to P0 but takes into account the differences in class sample size, shows the same, low accuracy (0.57). The values obtained for CSI (0.37) and K (0.38) confirm the low accuracy since they are lower than the expected minimum value (0.8) proposed by Koukolas and Blackburn (2001) and Landis and Koch (1977).

QD value (0.3) is double the value of AD (0.15). This implies that the total error is predominantly due to the difference in the number of sample units in a class of the reference and classified maps. However, allocation error is not negligible. Considering the six global indexes, it can be claimed that either classification went wrong or the reference data is not accurate.

Accuracy index	Value
P0	0.55
$P0_{marg}$	0.57
CSI	0.37
K	0.38
AD	0.15
QD	0.3

Table 4. Global accuracy indexes for Europe

Per-class accuracy indexes, like UA and PA, can reveal the classes with lowest accuracy and thus the highest contribution to low overall accuracy.

From Table 5 it can be seen that from the user's perspective, Forest (0.8) and Water Bodies (0.77) are the only classes with good agreement, since their accuracy values are higher than the threshold of 0.7 suggested by Thomlinson et al. (1999) for UA and PA. From the producer's point of view, Cropland is the only accurate class, but it is highly accurate (0.95). Five out of nine classes show very low accuracy according to both of the indexes, so it can be said that Grassland, Shrubland, Wetland, Bareland and Permanent Ice and Snow are the main source of error.

Class	PA	UA
Cropland	0.95	0.48
Forest	0.64	0.8
Grassland	0.03	0.4
Shrubland	0.3	0.25
Wetland	0.21	0.35
Water Bodies	0.24	0.77
Artificial surface	0.49	0.42
Bareland	0.02	0.11
Permanent Ice and Snow	0.11	0

Table 5. Per-class accuracy indexes for Europe

4.2 Country by country

Since the LUCAS data were surveyed by independent teams for each of 23 countries, the validation was also performed country

by country to understand variations at this level. In the Table 6 it can be seen that P0 is the lowest in Ireland and highest in the Czech Republic and Hungary, but even in these cases it is below 0.7. In most of the cases, $P0_{marg}$ is lower than P0, which means that difference in sample size is contributing to the Overall Accuracy. Being always below 0.5, K is indicating poor agreement in all of the countries. As in the case of Europe, the trend of AD and QD behavior remains the same. For all of the countries except France, QD is larger than AD, and may be up to 5 times larger.

Per-class indexes are shown for each class and each country in the Figure 1 for PA and Figure 2 for UA. The darker the color on the map, the higher the value of the accuracy.

According to PA results, Cropland is the most accurate class in all the considered countries, while Forest and Artificial Surface are moderately or poorly accurate depending on the country. The rest of the classes like Grassland, Shrubland, Wetland, Water Bodies, Artificial Surface and Bareland have low accuracy in all of the countries.

According to UA, Forest is highly accurate for many countries and Water Bodies is accurate only for several countries. Remaining classes (Cropland, Grassland, Shrubland, Wetland, Artificial Surface and Bareland) have poor agreement.

Country	P0	$P0_{marg}$	CSI	K	AD	QD
AT	0.56	0.5	0.46	0.39	0.13	0.31
BE	0.48	0.41	0.35	0.33	0.09	0.43
CZ	0.66	0.45	0.36	0.49	0.07	0.27
DE	0.61	0.46	0.39	0.43	0.08	0.31
DK	0.62	0.34	0.3	0.3	0.08	0.31
EE	0.60	0.46	0.34	0.39	0.1	0.34
EL	0.51	0.47	0.37	0.35	0.22	0.26
ES	0.53	0.45	0.35	0.32	0.18	0.29
FI	0.63	0.4	0.36	0.39	0.22	0.14
FR	0.52	0.55	0.36	0.32	0.11	0.37
HU	0.66	0.55	0.42	0.45	0.09	0.25
IE	0.12	0.41	0.27	0.04	0.13	0.75
IT	0.56	0.55	0.39	0.34	0.15	0.29
LT	0.48	0.5	0.35	0.3	0.08	0.44
LU	0.55	0.34	0.34	0.4	0.09	0.36
LV	0.46	0.4	0.28	0.27	0.13	0.4
PL	0.59	0.45	0.33	0.38	0.08	0.33
PT	0.44	0.42	0.3	0.27	0.21	0.35
SE	0.65	0.52	0.38	0.32	0.19	0.17
SI	0.6	0.44	0.34	0.38	0.1	0.3
SK	0.6	0.42	0.36	0.42	0.09	0.31
UK	0.37	0.54	0.38	0.21	0.13	0.49

Table 6. Global accuracy indexes for each country

5. VALIDATION OF GL30 WITH LUCAS AND DUSAF IN LOMBARDY

From previous experience (Brovelli et al., 2015) it was expected that the results of validation would definitely be more satisfactory, at least in Italy. Considering, the Lombardy region (Northern Italy) Bratic et al. (2018) obtained significantly higher results in GL30 accuracy. The reference map used for the validation was, in this case, DUSAF 4.0, a regional authoritative map with 2m accuracy. This section compares the results of the two GL30 validations using the LUCAS and DUSAF datasets, respectively.
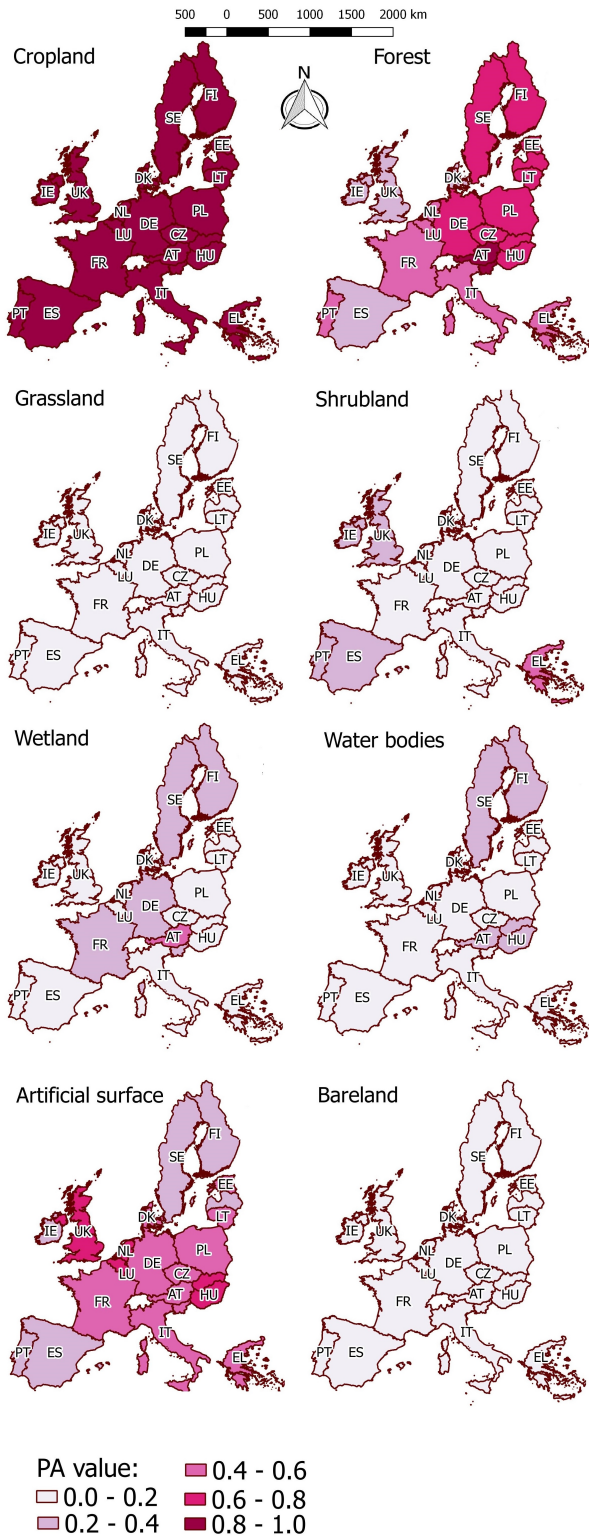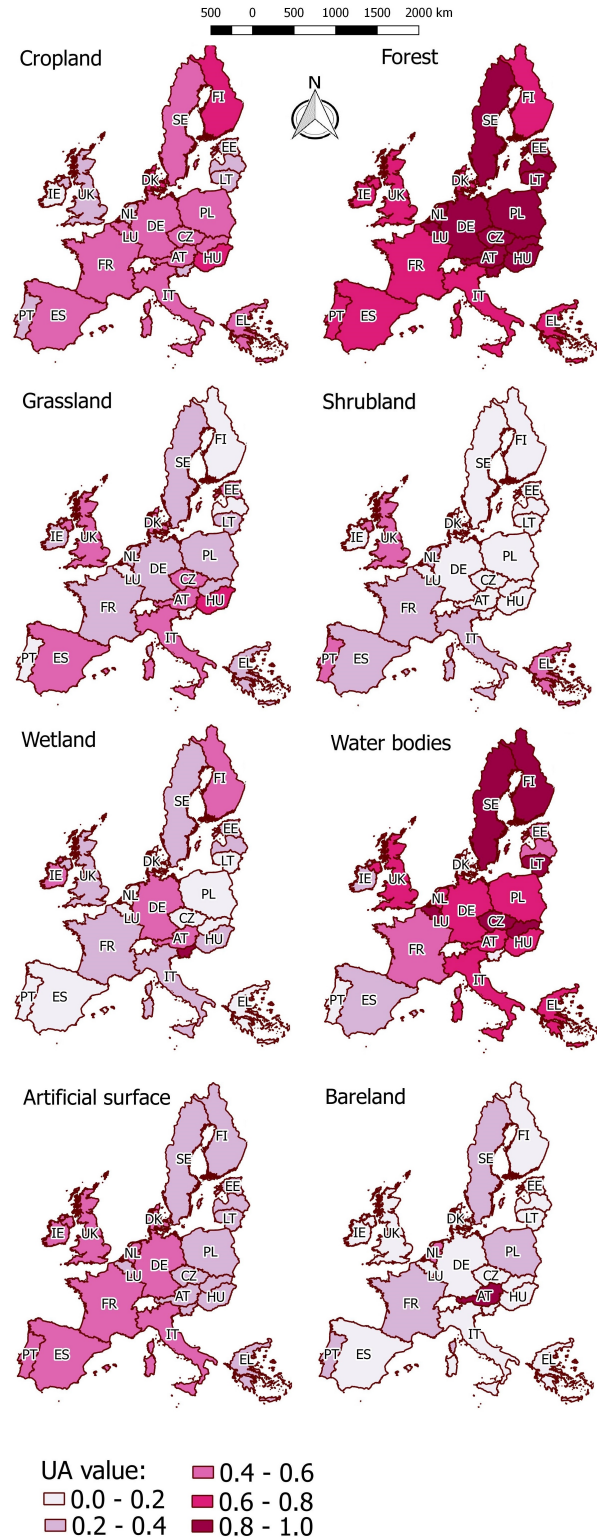
Figure 1. Producer's accuracy for each country

Figure 2. User's accuracy for each country

Since the accuracy of DUSAF is high enough to consider it as a confident ground truth reference, the comparison is important in order to understand if the poor validation results obtained in section 4 are due to errors in the LUCAS dataset.

Figure 3 presents a comparison between the global indexes obtained against LUCAS and DUSAF, respectively. It can be seen that all of the indexes computed based on DUSAF, except indexes of disagreement, exceed the ones computed based on LUCAS. On the opposite side, the disagreement indexes have switched places in the terms of higher one, because in the case of DUSAF, AD surpasses QD. As the DUSAF dataset is known to be an accurate one, values of the indexes raise doubts that LUCAS might be suitable for the assessment of GL30.
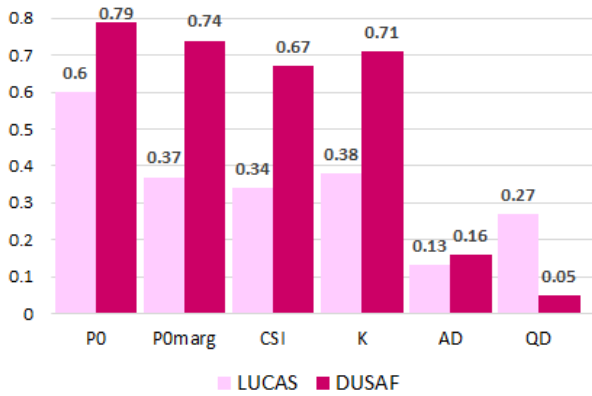
Figure 3. Global accuracy indexes for Lombardy

Furthermore, the Figure 4 and 5 depict differences in UA and PA, for each per-class index, when the different reference datasets are used. Numbers on the horizontal axis denote the classes of GL30 as shown by tables 2 and 3. Once more, it is evident that DUSAF outperformed LUCAS, especially in the case of the classes of Bareland and Permanent Ice and Snow, which for LUCAS results are equal to 0, while for DUSAF they are 0.8 for both UA and PA.
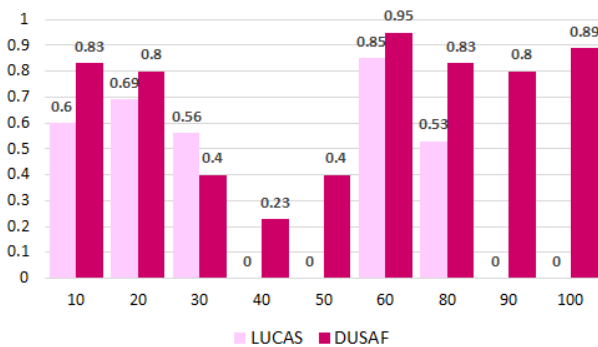

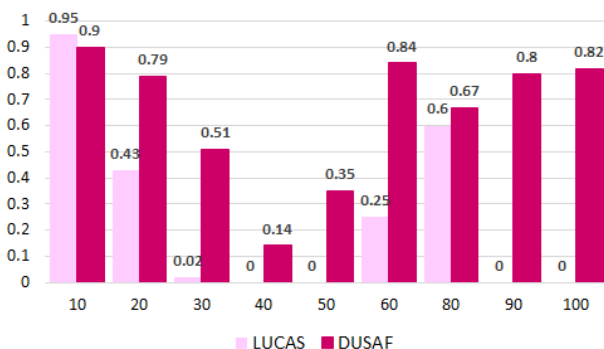Figure 4. User's accuracy for each LC class in Lombardy

Figure 5. Producer's accuracy for each LC class in Lombardy

6. VISUAL CHECK OF LUCAS

The comparison of results, presented in the previous section, points out irregularities in the LUCAS data. Hence, a visual check of a certain number of LUCAS points was conducted. It

was done with an attempt to understand if LUCAS has systematic errors, since this type of error can sometimes be removed. If this is the case for LUCAS, LUCAS might become more valuable for GL30 validation. The points were photo interpreted on the basis of three imagery sets:

1. Orthophoto of Lombardy region for 2007 at scale 1:10,000 (hereafter Ortho2007)
2. Orthophoto of Lombardy region for 2012 at scale 1:10,000 (hereafter Ortho2012)
3. Bing Aerial imagery (hereafter Bing). Since it is composed of tiles, the dates of imagery acquisition might vary from April 2015 to September 2017. Since these images are temporally more distant from the time of production of LUCAS data then the previous ones, it is expected that greater differences will be found. Nevertheless these images are considered due to their high resolution (around 2m) and therefore greater details.

Only the points of the classes that had very low PA and UA were verified. The classes with a large number of points, Grassland and Water Bodies, were randomly sampled with a rate of about 20% to reduce number of points to be verified. Hence, the following classes were checked independently by three operators:

1. Grassland - 40 points
2. Shrubland - 15 points
3. Wetland - 1 point
4. Water Bodies - 14 points
5. Bareland - 14 points

After the photo-interpretation of the points on the different imagery sets, the percentage of correctly classified LUCAS points according to each imagery set and class was calculated. Table 7 displays the percentage of correct LUCAS points found in every analyzed class according to photo-interpretation of Bing, Ortho2007 and Ortho2012. Unambiguously, all of the classes contain certain errors, but it can be said that Grassland and Shrubland are classified much better then the rest of the classes. The Wetland class is completely incorrect, and this is an effect of the small sample size in this category. Namely, there is only one point that is classified as Wetland in Lombardy, but it is wrongly classified according to all of the photo-interpreted maps.

Class	Bing	Ortho2012	Ortho2007
Grassland	50%	60%	65%
Shrubland	67%	73%	67%
Wetland	0%	0%	0%
Water Bodies	21%	29%	29%
Bareland	29%	36%	57%

Table 7. Percentage of correctly classified LUCAS points

Several attributes collected during the LUCAS survey are recognized as potentially useful for detecting the cause of errors in LUCAS data:

1. Observation distance
2. Observation type and
3. GPS precision

Observation distance is the distance of the surveyor from the observation point at the moment of point acquisition. It is expected that the observations from the smaller distance are more accurate. The Table 8 is showing which percentage of the correctly classified LUCAS points in a class are surveyed from <5m, <10m, <100m and >100m observation distances. The columns "<10m" and "<100m" are cumulative of the previous columns. The results shown in the table are not consistent for all classes. In fact, the Grassland and Shrubland points have more probability to be correct if they were surveyed from a distance of less than 100m. The increase of proximity to the observation (<5m) seems not to be important in this case. Results for Water are different depending on the reference imagery, and results for Bareland are not showing any dependence on the observation distance. Taking into consideration all of the above mentioned results, it may be reasonable to think that observation distance does not have a significant impact on the classification of the points.

Map	Class	Observation distance			
		<5m	<10m	<100m	>100m
Bing	Grassland	35%	65%	90%	10%
	Shrubland	20%	50%	80%	20%
	Water Bodies	0%	0%	100%	0%
	Bareland	25%	25%	50%	50%
Ortho2012	Grassland	36%	64%	86%	14%
	Shrubland	18%	45%	82%	18%
	Water Bodies	0%	0%	67%	33%
	Bareland	40%	40%	60%	40%
Ortho2007	Grassland	46%	69%	92%	8%
	Shrubland	20%	50%	80%	20%
	Water Bodies	0%	0%	67%	33%
	Bareland	22%	33%	56%	44%

Table 8. Percentage of correctly classified points with respect to distance of observation point

Observation type attribute is, to some extent, similar to observation distance. There are 5 types of observations in the LUCAS survey, but in Lombardy, only 3 are present:

- Type I: field survey of points visible on distance smaller than 100m

- Type II: field survey of points visible on distance larger than 100m

- Type III: field photo-interpretation in circumstances of unexpectedly inaccessible points

Percentage of the correctly classified LUCAS points in a class with respect to the three types of observation are displayed in the Table 9. It was expected that type I observations will have high percentages for each class. However, the results of Bareland showed high deviation with respect to other classes. The highest number of correctly classifed Bareland points were obtained by type III observations. Furthermore, none of the observation types had homogeneous results for all of the classes. Hence the observation type seems not to to influence correctness of the LUCAS points.

GPS precision is an average location error as given by GPS receivers (in m). It is expected that the points for which the value of GPS precision was smaller would be more accurate, but Table 10 does not support this expectation. In the table, the column "<10m" is cumulative of the column "<5m".

Map	Class	Type of observation		
		I	II	III
Bing	Grassland	80%	5%	15%
	Shrubland	80%	0%	20%
	Water Bodies	100%	0%	0%
	Bareland	25%	0%	75%
Ortho2012	Grassland	73%	5%	23%
	Shrubland	82%	0%	18%
	Water Bodies	67%	33%	0%
	Bareland	40%	0%	60%
Ortho2007	Grassland	81%	4%	15%
	Shrubland	80%	0%	20%
	Water Bodies	67%	33%	0%
	Bareland	44%	0%	56%

Table 9. Percentage of correctly classified points with respect to observation type

Map	Class	GPS precision		
		<5m	<10m	>10m
Bing	Grassland	15%	95%	5%
	Shrubland	20%	50%	50%
	Water Bodies	0%	100%	0%
	Bareland	25%	100%	0%
Ortho2012	Grassland	23%	95%	5%
	Shrubland	18%	45%	65%
	Water Bodies	0%	100%	0%
	Bareland	20%	100%	0%
Ortho2007	Grassland	12%	96%	4%
	Shrubland	20%	50%	50%
	Water Bodies	0%	100%	0%
	Bareland	22%	100%	0%

Table 10. Percentage of correctly classified points with respect to GPS precision

The percentage of correctly classified LUCAS points corresponding to different GPS precision thresholds, as presented in the table, shows that for all classes except Shrubland, error are fewer for points taken with a GPS precision of less than 10m. Since there is an exception in the case of Shrubland, GPS precision cannot be considered as an attribute that affected the classification accuracy of LUCAS points.
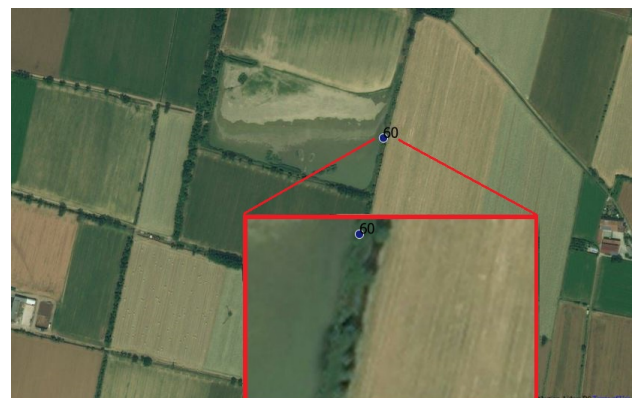

Figure 6. Example of an error in the LUCAS data

In addition to the analysis, on the Figure 6, which is not an isolated case, it is visible that one point belonging to the Water Bodies class (60) in LUCAS is wrongly classified. That point is observed as a type I observation, from an observation distance of

96m and with a GPS precision of 5m. Despite that water is not expected to be misclassified, it is wrongly classified looking at the Bing tile from May 2015. As can be seen, water is nowhere even close to the point, as the point belongs to the Shrubland class, and is surrounded by Cropland.

The same is evident from the Ortho2007 and Ortho2012, therefore this error is not due to landscape change over the years.

7. CONCLUSION

Validation of global land cover products is a challenging task considering the area for which ground truth should be provided at a reasonable cost. Initially, LUCAS 2009 data seemed to be a good dataset to be exploited as a ground truth reference since it was mostly collected during in-situ field surveys. Other advantages of LUCAS are that it covers an enormous area (almost a continent) and that it is free of charge. Some former studies have shown that LUCAS might not be a good choice of reference, but since characteristics of GL30 are more similar to LUCAS characteristics with respect to datasets validated in these studies (smaller differences in MMU and the same class definition as for reference data), LUCAS was used as a ground truth reference for the validation of GlobeLand30. According to case studies found in the literature, GL30's accuracy is expected to be higher than the validation outcome that our study is showing. This brings into question whehter LUCAS is an adequate ground truth reference, even in case of similar features with comparison maps (in this case GL30). Further examination was done to determine if it is justified to use LUCAS as a reference by comparing validation results where LUCAS served as a reference to other results obtained with authoritative, high-accuracy data. The analysis for comparison were restricted to the Lombardy region, Italy because here highly accurate data were easily accessible (DUSAF). In general, accuracy indexes for validation with LUCAS are lower than the one with DUSAF, which points towards errors in LUCAS since DUSAF is known to be very accurate. Furthermore, a visual check of LUCAS points was done to understand if some specific characteristic (observation distance, observation type, and GPS precision) may be the cause for errors in the dataset in such a way that these specific points could be removed from the dataset. The analysis was not able to reveal any obvious systematic error. Since the analysis of LUCAS errors were limited to one small region, it cannot be said that the results can be extended to the whole area covered by LUCAS. Certain caution is recommended when utilizing LUCAS as ground truth.

ACKNOWLEDGEMENT

REFERENCES

Anderson, J. R., Hardy, E. E., Roach, J. T. and Witmer, R. E., 1976. A land use and land cover classification system for use with remote sensor data. Technical report. Report.

Bonomi, A., Scaramellini, G., Treu, M. C., Lanzani, A., Bolognesi, C., Pretolani, R., Agnoletti, M., Santoro, A., Pettenella, D., Sartori, F., Bracco, F. and Pileri, P., 2011. Tools for territorial knowlege and governments. In: D. Fasolini, S. Pezzoli, S. Vanna Maria, M. Credali, L. Minnella and M. Peggion (eds), *Land cover changes in Lombardy over the last 50 years*, ERSAF - Regional Agency for Agriculture and Forestry Service, Lombardy Region - Directorate of Territory and Urban Planning, Milan, Italy, chapter 2, pp. 17–28.

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, pp. 87–92.

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), pp. 4191–4212.

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.

Chen, L., 2010. Globeland30. NGCC - National Geomatic Center of China http://www.globallandcover.com/home/Enbackground.aspx (2018-06-12).

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

Esch, T., Marconcini, M., Felbier, A., Roth, A., Heldens, W., Huber, M., Schwinger, M., Taubenböck, H., Müller, A. and Dech, S. W., 2013. Urban footprint processorâĂŤfully automated processing chain generating settlement masks from global data of the tandem-x mission. *IEEE Geoscience and Remote Sensing Letters* 10, pp. 1617–1621.

ESTAT, 2015. LUCAS 2009 (land use/cover area frame survey). Technical report, Directorate E: Agriculture and environment statistics; Statistical Cooperation, Unit E-1: Farms, agro-environment and rural development.

Gallego, F. J., 2011. Validation of GIS layers in the EU: getting adapted to available reference data. *International Journal of Digital Earth* 4(sup1) pp. 42–57.

Gong, P., Wang, J., Yu, L., Zhao, Y., Zhao, Y., Liang, L., Niu, Z., Huang, X., Fu, H., Liu, S., Li, C., Li, X., Fu, W., Liu, C., Xu, Y., Wang, X., Cheng, Q., Hu, L., Yao, W., Zhang, H., Zhu, P., Zhao, Z., Zhang, H., Zheng, Y., Ji, L., Zhang, Y., Chen, H., Yan, A., Guo, J., Yu, L., Wang, L., Liu, X., Shi, T., Zhu, M., Chen, Y., Yang, G., Tang, P., Xu, B., Giri, C., Clinton, N., Zhu, Z., 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.

GRASS Development Team, 2017. Geographic Resources Analysis Support System (GRASS) Software. Open Source Geospatial Foundation http://grass.osgeo.org (20 September 2017).

Hansen, M. C., Potapov, P. V., Moore, R., Hancher, M., Turubanova, S. A., Tyukavina, A., Thau, D., Stehman, S. V., Goetz, S. J., Loveland, T. R., Kommareddy, A., Egorov, A., Chini, L., Justice, C. O. and Townshend, J. R. G., 2013. High-resolution global maps of 21st-century forest cover change. *Science* 342(6160), pp. 850–853.

Karydas, C. G., Gitas, I. Z., Kuntz, S. and Minakou, C., 2015. Use of LUCAS LC point database for validating country-scale land cover maps. *Remote Sensing* 7(5), pp. 5012–5041.

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), pp. 499–510.

Labatut, V. and Cherifi, H., 2011. Accuracy Measures for the Comparison of Classifiers. In: A.-D. Ali (ed.), *The 5th International Conference on Information Technology*, Al-Zaytoonah University of Jordan, amman, Jordan, pp. 1,5.

Landis, J. R. and Koch, G. G., 1977. The measurement of observer agreement for categorical data. *Biometrics* 33(1), pp. 159–174.

Martino, L., Palmieri, A. and Gall, J., 2009. Use of auxiliary information in the sampling strategy of a european area frame agro-environmental survey. Technical report, European Commission: DG-ESTAT, European Commission: DG -JRC.

Pekel, J.-F., Cottam, A., Gorelick, N. and Belward, A. S., 2016. High-resolution mapping of global surface water and its long-term changes. *Nature* 540, pp. 418 EP –.

Pesaresi, M., Ehrlich, D., Ferri, S., Florczyk, A. J., Freire, S., Halkia, M., Julea, A. M., Kemper, T., Soille, P. and Syrris, V., 2016. *Operating procedure for the production of the global human settlement layer from Landsat data of the epochs 1975, 1990, 2000, and 2014*. Vol. 27741 EN, Publications Office of the European Union, Luxembourg.

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), pp. 4407–4429.

Pringle, M., Schmidt, M. and Muir, J., 2009. Geostatistical interpolation of SLC-off landsat ETM+ images. *ISPRS Journal of Photogrammetry and Remote Sensing* 64(6), pp. 654 – 664.

QGIS Development Team, 2018. QGIS Geographic Information System. Open Source Geospatial Foundation.

Shimada, M., Itoh, T., Motooka, T., Watanabe, M., Shiraishi, T., Thapa, R. and Lucas, R., 2014. New global forest/non-forest maps from alos palsar data (2007âĂŞ2010). *Remote Sensing of Environment* 155, pp. 13 – 31.

Thomlinson, J. R., Bolstad, P. V. and Cohen, W. B., 1999. Coordinating methodologies for scaling landcover classifications from site-specific to global: Steps toward validating global map products. *Remote Sensing of Environment* 70(1), pp. 16 – 28.