

REVIEW OF HIGH-RESOLUTION GLOBAL LAND COVER

G. Bratic¹, A. Vavassori¹, M.A. Brovelli^{2*}

¹ Politecnico di Milano, Dept. of Civil and Environmental Engineering, Via Gaetano Previati 1/c, 23900 Lecco, Italy -
(gorica.bratic, alberto.vavassori)@polimi.it

² Politecnico di Milano, Dept. of Civil and Environmental Engineering, Piazza Leonardo da Vinci 32, 20133 Milan, Italy -
maria.brovelli@polimi.it

Commission IV, WG IV/4

KEY WORDS: Global Land Cover, High-Resolution Land Cover, Land Cover Legend, Review, Land Cover Accuracy

ABSTRACT:

The land cover detection on our planet at high spatial resolution has a key role in many scientific and operational applications, such as climate modeling, natural resources management, biodiversity studies, urbanization analyses and spatial demography. Thanks to the progresses in Remote Sensing, accurate and high-resolution land cover maps have been developed over the last years, aiming at detecting the spatial resolution of different types of surfaces. In this paper we propose a review of the high-resolution global land cover products developed through Earth Observation technologies. A series of general information regarding imagery and data used to produce the map, the procedures employed for the map development and for the map accuracy assessment have been provided for every dataset. The land cover maps described in this paper concern the global distribution of settlements (Global Urban Footprint, Global Human Settlement Built-Up, World Settlement Footprint), water (Global Surface Water), forests (Forest/Non-forest, Tree canopy cover), and a two land cover maps describing world in 10 generic classes (GlobeLand30 and Finer Resolution Observation and Monitoring of Global Land Cover). The advantages and shortcomings of these maps and of the methods employed to produce them are summarized and compared in the conclusions.

1. INTRODUCTION

The number of high-resolution (HR) global land cover (LC) maps has increased. This is not a surprise given the advances in Remote Sensing. Moreover, HRLC is useful for numerous applications such as climate modeling, biodiversity studies, natural resource management, inter-comparison, etc. LC production, including HRLC production, is not standardized. As a result, available HRLC is produced in different ways and have different characteristics. We made a review of the relevant literature of the existing HRLC maps to make an overview of their main characteristics, and to make a comparison among them.

The scope of the review is to make an outline of the available HRLC and their characteristics. This will allow the potential users to find all the details necessary for the proper HRLC exploitation in one place. Furthermore, thanks to the overview we will be able to observe if the information about different characteristics is complete and suitable, and we will suggest improvements of HRLC literature/documentation. The review will be based on the scientific literature on binary and multi-class global HRLCs. Binary datasets include Global Urban Footprint (GUF) (Esch et al., 2018), Global Human Settlement Built-Up Grid – Sentinel-1 (GHS BU S1NODSM) (Corbane et al., 2017), Global Human Settlement Built-Up Grid – Landsat (GHS BU LDSMT) (Corbane et al., 2017), Global Surface Water (GSW) (Pekel et al., 2016), Forest / Non-Forest (FNF) (Shimada et al., 2014), Tree canopy cover (Hansen et al., 2013), World Settlement Footprint (WSF) (Marconcini et al., 2020), GlobeLand30 (GL30) (Chen et al., 2015), Finer Resolution Observation and Monitoring of Global Land Cover (FROM-GLC)

(Gong et al., 2013). Most of the references were cited hundreds of times which is indicating that the datasets were found useful for research.

For the review, we will take into consideration different aspects of HRLC production and validation. We will enter into details regarding input satellite imagery (e.g. type, acquisition date, resolution, etc) and auxiliary data used for derivation of HRLC. Furthermore, we will include a description of the sampling schemes and sources of training and validation data. Lastly, we will outline legend definition and accuracy assessment and results. Table 1 provides an overview of the existing HRLC, reporting the datasets producer, spatial resolution and reference years.

2. HIGH-RESOLUTION GLOBAL LAND COVER

2.1 GlobeLand30

GL30 is a set of global land cover maps at 30 m resolution (Chen et al., 2015). Three version of GL30 are available for 3 different years: 2000, 2010, and since recently for 2020. National Geomatics Center of China, producer of this map, provides the two datasets as open-access in the official web-site (<http://www.globallandcover.com/>). It is provided in the Universal Transverse Mercator (UTM) projection. Legend of GL30 consists of 10 main classes: Cultivated land, Forest, Grassland, Shrubland, Wetland, Water bodies, Tundra, Artificial surfaces, Bare land and Permanent snow and ice. Description of every class can be found at the official website.

* Corresponding author

Name	Producer	Resolution	Year
GL30	NGCC	30 m	2000, 2010, 2020
FROM-GLC	Tsinghua University	30 m	2010, 2015, 2017
GHS BU S1NODSM	Joint Research Center	20 m	2016
GHS BU LDSMT	Joint Research Center	30m	1990, 2000, 2014
GUF	German Aerospace Center	12 m	2011
GSW	Joint Research Center	30 m	1984 – 2019 (map for every year)
FNF	The Japan Aerospace Exploration Agency – Earth Observation Research Center	25 m	2007 – 2010 (map for every year); 2015 – 2017 (map for every year)
Tree canopy cover	Hansen - University of Maryland (UMD) - Google - United States Geological Survey (USGS) - National Aeronautics and Space Administration (NASA)	30 m	2000
WSF	German Aerospace Center	10 m	2015

Table 1. Overview of the existing HRLC

2.1.1 Imagery Main data source of satellite imagery for deriving GL30 were Landsat 5, 7, and 8 collections. Moreover, Chinese Environmental and Disaster satellite (HJ-1) and China High Resolution Satellite (GF-1) were supplementary image sources for v.2010 and v.2020, respectively. The imagery selected for production was imagery sensed during the vegetation growing season. Data gaps were filled with imagery sensed in the time window of 2 years prior or after the product baseline year. GL30 v.2020 is fairly new, and information about this version are limited, so the following subheadings are referring to the first two GL30 versions mostly.

2.1.2 Training data In case of GL30 v.2000 and v.2010 there were various sources of reference data for training. These include existing global and regional LC, global Digital Elevation Models (DEM) - SRTM (Shuttle Radar Topographic Mission) and ASTER (Advanced Spaceborne Thermal Emission and Reflection Radiometer), global topographic data with 1 : 1 000 000 scale (Hayakawa et al., 2008), and ecological zones (Olson et al., 2001). Furthermore, ancillary sources such as Google map, Map World, Open Street Map (OSM) and Geo-Wiki were exploited as well. To be able to integrate reference data from different sources and with diverse format, accuracy and spatial resolution specific tool with unique user interface was developed.

2.1.3 Algorithms For production of GL30 Pixel-Object-Knowledge-based (POK-based) classification approach was adopted. This approach is initialized with pixel-based classification, then different thresholds are applied to define objects, and finally nature-based, culture-based or temporal constraints knowledge is introduced for verification purpose. Pixel-based classification was hierarchical with aim to reduce spectral confusion among pixels. This means that one class at the time was classified in the following sequence - water bodies, wetland, permanent ice/snow, artificial surfaces, cultivated land, forest, shrubland, grassland, bareland and tundra. Multiple classification algorithms were used for pixel-based classification, such

as Maximum Likelihood Classifier (MLC), Support Vector Machine (SVM), Decision Tree (DT) or Automated thresholding. Some classes were extracted based on only one algorithm, while in some other classes were extracted by combining outcomes of several algorithms.

2.1.4 Validation data and results The validation exercise of NGCC was a preliminary validation of land cover map for 2010. Validation dataset comprised of 154 586 valid pixel samples out of the initial 159 874 pixel samples. The samples were extracted by a two-rank sampling strategy. In the first rank sampling, 80 out of 847 map sheet were selected. The map sheets were distributed among 5 continents taking into consideration land-area ratio. In the second rank, sampling stratification by land cover type was performed. Number of samples per each land cover type was estimated with landscape index and layer are ratio in the total sample size. Finally, location of samples was determined based on spatial correlation analyses.

The estimated Overall Accuracy (OA) in the preliminary accuracy of GL30-2010 (Table 2) was 80%. User's Accuracy (UA) was ranging 72.16% - 86.7% (Chen et al., 2017). Preliminary accuracy assessment does not report Producer's Accuracy (PA). Also, accuracy for GL30-2000 on a global level was not reported. Independent accuracy assessment of GL30 conducted at regional (Manakos et al., 2014), national (Brovelli et al., 2015; Yang et al., 2017), and subcontinental level (Jacobson et al., 2015) show similar results to the preliminary accuracy assessment, with an exception in Central Asia where accuracy is 46% (Sun et al., 2016).

Class of GL30 v.2010	UA [%]
Cultivated land	82.76
Forest	83.58
Grassland	72.16
Shrubland	72.64
Wetland	74.87
Water bodies	84.7
Artificial surfaces	86.7
Bareland	81.76
Permanent snow and ice	75.79
OA [%]	80.33% ± 0.2%
Kappa	0.75

Table 2. Accuracy of GL30 2010

2.2 Finer Resolution Observation and Monitoring of Global Land Cover

FROM-GLC is a product of University of Tsinghua. Currently, this map has 3 editions for years 2010, 2015, and 2017 (Gong et al., 2013). Each edition has slightly different legend that can be summarized into 10 main classes: Cropland, Forest, Grass, Shrub, Wetland, Water, Tundra, Impervious, Bare land, and Snow/Ice. Further details about FROM-GLC legend, as well as free data download are provided on the official website (<http://data.ess.tsinghua.edu.cn/>). Data is provided in WGS84 coordinate reference system (CRS).

2.2.1 Imagery Landsat 5 imagery acquired in 2009 and 2010 constitutes around 63% of imagery for production of 2010 edition. Landsat 5 and Landsat 7 collections from earlier years were used to fill the data gaps. For 2015 edition, Landsat 8 imagery from 2014-2015 was predominant data source, but imagery from 2013 and 2016 was used when necessary to replace

missing imagery. Sentinel-2 imagery from 2017 was the only source of imagery for 2017 edition of FROM-GLC.

2.2.2 Training data When edition 2010 is concerned, there were two phases of collecting training data. In the first phase training samples were extracted and interpreted based on what can be identified in Landsat imagery, and in some occasions Google Earth imagery was used as a reference. As a results, 91 433 training samples were collected. From each Landsat scene 10-20 samples were extracted, with maximum 3 samples per class. They were limited to homogeneous landscape in 8x8 pixels area. The 2010 training samples were distributed in such a way to be representative of major classes in the scene. The second phase was refinement of the first-phase samples based on additional information from time series vegetation index from MODIS EVI for 2010. There were 10 photo-interpreters involved in the second-phase.

For edition 2015, in the scope of sample collection, 13 image interpreters were working with Global Mapper, the tool that integrates Landsat images, MODIS (Moderate Resolution Imaging Spectroradiometer) EVI (Enhanced Vegetation Index) time series, monthly temperature and precipitation, and Google Earth images. They interpreted samples in 4 different seasons, in order to understand if samples from different season can improve accuracy. The samples were cross-checked by senior specialist for photo-interpretation. Final number of training sample points was 91 433.

For edition 2017, training samples from 2015 were exploited for classifying Sentinel-2 images.

2.2.3 Algorithms Four different classification algorithms were used for classification in 2010 : MLC, J4.8 DT, Random Forest (RF) and SVM. For 2015 and 2017 editions, RF was chosen after global and continental experiments.

2.2.4 Validation data and results Validation samples for 2010 were extracted by partitioning globe into 7 000 equal area hexagons, and by selecting 5 samples for each of them. The sample interpretation had 2 phases. In the first phase samples were photo-interpreted from Landsat imagery, while in the second phase they were refined with reference to the MODIS EVI time series. 36 630 validation samples were extracted. Validation data collection strategy for 2015 was the same, and it resulted in 36 000 samples. The validation samples from 2015 were reused in the 2017 edition validation exercise. Outcome of all 4 algorithms used for 2010 production was assessed. OA shows that the SVM was the best algorithm with 64.89%, followed by RF, J4.8 and MLC with 59.83%, 57.88%, and 53.88%, respectively. There is a slight increase in the OA in 2017 with respect to OA in 2010 and 2015 (Table 3). Moreover, there is evident increase of class accuracy in case of Cropland, Grass and Shrub in 2015 compared to 2010. OA by seasons for the FROM-GLC 2015 and FROM-GLC 2017 product is reported in Table 4. It is evident that the highest OA is for a season from December to February (70.95%), however it is not significantly different with respect to all year OA (67.16%). In case of FROM-GLC 2015 per-class accuracy metrics, UA and PA, were not computed in the validation.

2.3 Global Human Settlements

GHS BU is a set of data that represents the built up surfaces evolution and urban-rural delimitation. There are two different

Classes	FROM-GLC v.2010		FROM-GLC v.2017	
	PA [%]	UA [%]	PA [%]	UA [%]
Cropland	39.23	45.27	61.07	58.14
Forest	76.45	80.49	84.2	83.47
Grass	34.88	44.13	64.49	58.51
Shrub	33.94	49.65	50.77	62.75
Wetland	-	-	10.14	34.48
Water	88.76	80.56	92.32	86.39
Tundra	-	-	77.73	75.89
Impervious	10.53	30.77	72.19	61.76
Bare land	89.91	62.72	84.58	81.4
Snow/Ice	85.66	95.61	71.17	91.96
Clouds	77.6	69.65	-	-
OA [%]	64.89		72.76	

Table 3. Accuracy of FROM-GLC 2010 (SVM algorithm) and FROM-GLC 2017

FROM-GLC 2015 season	OA [%]
Dec-Feb	70.95
Mar-May	67.81
Jun-Aug	63.97
Sep-Nov	65.91
All seasons	67.16

Table 4. Accuracy of FROM-GLC 2015 by seasons and overall

streams of GHS BU production, one multi-temporal 30 m resolution product based on Landsat optical imagery - GHS BU LDSMT and one 20 m resolution product based on Sentinel-1 SAR imagery - GHS BU S1NODSM (Corbane et al., 2017). GHS BU LDSMT data refer to four different epochs - 1975, 1990, 2000 and 2014, while GHS BU S1NODSM is only for 2016. First release of GHS BU LDSMT (v.2015) had 38 m resolution and was reprocessed to second/current (v.2017) release with 30 m resolution. Original GHS BU products are provided in the Web Mercator projection (EPSG:3857). Resampled versions of original products at resolutions of 250 m and 1 km are available in World Mollweide projection (EPSG:54009). Free download is available at website: <https://ghsl.jrc.ec.europa.eu/download.php>. Table 5 contains the legend of the two products GHS BU S1NODSM and GHS BU LDSMT.

	Class code	Class
S1	0	Non-built-up areas
	1	Built-up areas
LDSMT	0	No data
	1	Water surface
	2	Land not built-up in any epoch
	3	Built-up from 2000 to 2014
	4	Built-up from 1990 to 2000
	5	Built-up from 1975 to 1990
	6	Built-up up to 1975

Table 5. Legend of GHS BU S1NODSM and GHS BU LDSMT

2.3.1 Imagery A total number of 32 808 Landsat images were used for the production of GHS BU LDSMT (v.2015 and v.2017). In particular, 7 588 scenes acquired by the Multispectral Scanner (1975 collection), 7 375 by Landsat 4-5 (1990 collection), 8 756 by Landsat 7 (2000 collection) and 9 089 by Landsat 8 (2014 collection). For the GHS BU S1NODSM 2016 product, 5 026 Sentinel-1A images acquired between December

2015 and October 2016, as well as 1 695 Sentinel-1A and 329 Sentinel-1B images from November 2016 to December 2017 were employed.

2.3.2 Training data A quite heterogeneous training dataset sources were used for the classification of the Landsat images for the first release. It includes MERIS Globe Cover artificial surfaces, LandScan population grids, OSM, Geonames and MODIS 500 m (MOD500) global urban extents. For the classification of Sentinel-1 images, the GL30 product together with the built-up areas derived from the first release of GHS BU LDSMT product were used. Finally, for the reprocessing of Landsat data for the second (current) release, the artificial surfaces from GL30 and the built-up areas from GHS BU S1NODSM 2016 were employed. Basically, an incremental learning was tested, leading to a progressive refinement of the product. Moreover, ancillary data were used for the processing of Sentinel-1 data, namely SRTM DEM (with 1 arcsec resolution) and ASTER GDEM V2 (Global DEM Version 2).

2.3.3 Algorithms The method used for the built-up recognition is based on Symbolic Machine Learning (SML) (Pesaresi et al., 2016). For the processing of Sentinel-1 images, input features (Sentinel-1 GRD image features and topographic features) are reduced to unique sequences and then associated with the built-up learning set to obtain a built-up confidence ENDI (Evidence-based Normalized Differential Index); through the Otsu binarization (Otsu, 1979), a binary built-up map is obtained. For the processing of Landsat data, input features (radiometric and textural features) are associated with the training set to generate built-up maps per date (1975, 1990, 2000 and 2014); from the multi-temporal fusion of these maps, the new product is obtained.

2.3.4 Validation data and results The GUF dataset was used as reference for the quantitative assessment of the outputs and in particular for a cross-comparison of the different SML outputs. The Kappa coefficient, commission and omission errors were evaluated. The median and standard deviation of these metrics were calculated for 23 134 tiles of 150x150 km size. An increasing value of Kappa coefficient and a decreasing commission and omission errors can be observed in GHS BU S1NODSM 2016 and GHS BU LDSMT v.2017 with respect to the GHS BU LDSMT v.2015. Moreover, an increasing agreement was observed moving from GHS BU S1NODSM 2016 to GHS BU LDSMT v.2015 product, confirming the utility of the incremental learning that characterizes the SML classifier. Significant improvements were observed in Africa (reduction of commission errors with the GHS BU LDSMT v.2017), while the best result in terms of omission error was obtained in Asia. The lowest median commission error (0.27) is observed in South America with the GHS BU LDSMT v.2017 product, the lowest median omission error (0.35) is observed in Europe with the GHS BU S1NODSM 2016 product. Finally, South America is the area showing the highest overall agreement with the reference data (Kappa coefficient equal to 0.50 obtained with the GHS BU LDSMT v.2017 product); Kappa resulted equal to 0.42 and 0.40 respectively for Europe and North America for GHS BU LDSMT v.2017.

2.4 Global Urban Footprint

GUF is a raster dataset based on satellite radar imagery that represents the human settlements pattern in urban and rural environment with a spatial resolution of 0.4 arcsec (about 12 m near the equator) (Esch et al., 2018). The dataset is

given in WGS84 CRS and is provided by the DLR (German Aerospace Center) German Remote Sensing Data Center (<https://geoservice.dlr.de/web/maps/eoc:guf:4326>). The dataset at full spatial resolution is freely available for scientific use, whereas the generalized version with lower resolution (2.8 arcsec, corresponding to about 84 m near the equator) is available for non-profit and non-scientific applications. The product is a binary raster data in 8-bit (LZW-compressed GeoTIFF format) that shows three coverage categories in a black-and-white representation: built-up areas in black (value = 255), non-built up surfaces in white (value = 0) and no-data in grey (value = NoData). Built-up areas are defined as regions featuring man-made building structures with a vertical component.

2.4.1 Imagery A total of 182 249 TanDEM-X and TerraSAR-C radar images with a 3 m ground resolution were used for the GUF map production. These satellite images were mostly collected between 2011 and 2012 (93%) in the context of the TanDEM-X German mission. Single scenes sensed between 2013 and 2014 were employed to fill data gaps.

2.4.2 Training data Training samples are automatically identified on the basis of certain thresholds derived from image statistics in terms of amplitude and texture data. Auxiliary data were used to improve the classification performance. The following layers were used as ancillary data in the post-processing phase: OSM Settlements and Roads, DLR Relief mask and Road Clusters, GL30 Settlements, Water and Wetlands, Copernicus Imperviousness Layer 2012, US National Land Cover Dataset 2011, TimeScan-ASAR and TimeScan-Landsat. These layers were used to confirm or exclude pixels that were classified as built-up areas.

2.4.3 Algorithms In order to produce the GUF layer a fully automated processing environment was used, namely the Urban Footprint Processor (UFP). The UFP is able to manage every part of the processing chain: feature extraction, unsupervised classification, mosaicking and post-editing. An unsupervised classification method called Support Vector Data Description (SVDD), combining data of backscatter amplitude and of local texture, was used in this context. The classifier determines the hypersphere with minimum radius that encloses all the training samples for the built-up class and then it associates the unknown samples falling within the hypersphere boundary to the built-up class. The procedure allows to detect essentially vertical structures of urban habitations and not impervious surfaces such as roads or paved elements.

2.4.4 Validation data and results Accuracy of GUF was evaluated on the basis of ground truth data and considering established settlements maps (GHS BU, GL30 and MOD500) for a relative comparison. Ground truth data were manually collected from very high resolution optical images for 12 urban sites (see Table 6) across the world, each one covering an area of 1° latitude by 1° longitude. A total of 1 000 points for each class (built-up and non-built-up) and for each test area were randomly extracted and used as reference for the accuracy estimate. The validation showed that the GUF has the highest accuracy (OA = 85.04%, Kappa = 0.686), with the lowest standard deviation (3.96% of the overall accuracy), followed by GHS BU, GL30 and MOD500, that has the lowest accuracy measures. Beside this regional validation, the four data sets were compared at a global level: basically, a great difference between existing inventories exists. Table 6 summarizes the accuracy assessment results for 12 urban study sites.

	OA [%]	Kappa
Addis Ababa	84.87	0.695
Athens	85.42	0.672
Beijing	87.35	0.682
Dar Er Salaam	85.01	0.689
Kampala	84.95	0.690
Kigali	82.60	0.623
Lagos	84.49	0.688
Mexico City	89.40	0.786
Milan	89.33	0.784
Nairobi	85.01	0.688
New York	74.28	0.480
Perth	87.75	0.753
Mean	85.04	0.686
Standard deviation	3.96	0.080

Table 6. Accuracy of GUF

2.5 World Settlement Footprint

WSF is a raster dataset representing the spatial distribution of human settlements for the year 2015 at a global level (Marconcini et al., 2020). The dataset is available with a resolution of about 10 m (0.32 arcsec) and it is produced by the DLR. The dataset is publicly and freely available through figshare (<https://doi.org/10.6084/m9.figshare.c.4712852>) and can be downloaded in 306 GeoTIFF files in WGS84 CRS (EPSG:4326). The legend of WSF has value 255 that correspond to settlements, while the other surfaces are labelled with the value 0. Settlements are defined as a region featuring man-made structures.

2.5.1 Imagery The satellite images used for the WSF2015 production include both radar (Sentinel-1) and optical (Landsat 8) data. The former are about 107 000 images with a resolution of 10 m, the latter are about 217 000 images with a resolution of 30 m. Optical data are used to overcome the limitations of radar data, and vice versa.

2.5.2 Training data In order to train the classifier, different training samples were selected. For optical data, training samples for the two classes (settlement and non-settlement) were selected based on specific thresholds applied to the indexes NDBI (Normalized Difference Built-up Index), NDVI (Normalized Difference Vegetation Index) and MNDWI (Modified Normalized Difference Water Index). For radar data, thresholds in terms of mean backscattering were determined instead. For the classification procedure, ancillary data were employed, namely the SRTM and the ASTER DEM: these data were used in the processing of radar data, in order to remove pixels characterized by a slope higher than 10°.

2.5.3 Algorithms An advanced classification system that jointly exploits optical and radar satellite imagery was employed. First, images were pre-processed and temporal statistics were extracted for the automatic selection of the training set. Afterwards, a SVM algorithm with Radial Basis Function (RBF) Gaussian Kernel classifier was applied for the two types of images. A final post-classification procedure was performed to combine the maps derived from optical and radar images and to remove false alarms.

2.5.4 Validation data and results The final product of WSF represents the global distributions of settlements referred to the year 2015. In order to evaluate the map accuracy, 900

000 samples labelled with a crowdsourcing activity, through the visual interpretation of very high resolution Google Earth satellite imagery (relative to the period 2014-2015 and with a spatial resolution of 1.5, 0.5 and 0.15 m) were employed. As for sampling design, a stratified random sampling was applied, with 50 tiles of 1x1 degree size; a 3x3 block spatial assessment unit, composed of 9 cells with dimensions 10x10 m was chosen: for every block and for every tile, 1 000 samples for settlement and 1 000 samples for non-settlement class were extracted. The evaluation was performed on 3 different settlement definitions and 4 assessment criteria. The result shows a mean Average Accuracy (AA%) equal to 86.37% and an average Kappa equal to 0.6885. AA% is a balanced measure of correct settlement and non-settlement detection obtained as the mean between PA for settlements - PA_S and PA for Non-settlements-PA_{NS} (%). The best accuracy results were obtained when considering the combination buildings-building lots as settlement definition. Furthermore, the average PA are equal to 88.71% and 84.04% respectively for the class settlement and non-settlement. Compared to similar product, such as GUF, GHSL and GLC30, the WSF2015 accuracy is higher. Also, WSF2015 product shows a significant improvement in the detection of small settlements in rural regions as well as scattered suburban areas. Table 7 reports the worst and the best results in terms of accuracy. The former was obtained considering buildings as settlement and employing the assessment criterion 1 (per-cell matching); the latter considering buildings and building lots as settlement and employing the assessment criterion 4 (per-block matching).

	Settlement = buildings	Settlement = buildings + building lots
	Assessment criterion 1	Assessment criterion 4
AA [%]	83.27	89.33
Kappa	0.5486	0.7822
PA_S [%]	91.99	89.61
PA_{NS} [%]	74.56	89.06
UA_S [%]	59.41	85.85
UA_{NS} [%]	96.40	92.04

Table 7. Accuracy of WSF. S indicates settlements and NS indicates non-settlements

2.6 Global Surface Water

GSW is a dataset representing the spatio-temporal variability of the global water surface and the changes occurred over the time span 1984-2016 (32 years), with a resolution of 30 m (Pekel et al., 2016). The dataset is now available for the time interval 1984-2019. The layer has been developed in the framework of the Copernicus Programme by the EC JRC, with the aim of giving support to many applications including water resources management, climate modelling, biodiversity conservation and food security. The products are freely available from <https://global-surface-water.appspot.com> and they include maps of water occurrence, occurrence change intensity, seasonality, recurrence, transitions, maximum water extent, monthly recurrence, yearly history and monthly water history. Legends and more detailed description of all the maps are available at the above mentioned web site.

2.6.1 Imagery The data used in this context include the entire multi-temporal and orthorectified archive of Landsat 5 Thematic Mapper (TM), Landsat 7 Enhanced Thematic

Mapper-plus (ETM+) and Landsat 8 Operational Land Imager (OLI), Top-Of-Atmosphere (TOA) reflectance and brightness temperature images. In particular, the period of acquisition between 16 March 1984 and 10 October 2015 was considered.

2.6.2 Training data In order to classify the Landsat images, a spectral library, describing the behaviour of the three target classes (namely, water, land and non-valid observations) was built. A total number of 64 254 samples, obtained through visual interpretation of 9 149 Landsat scenes, were used, so that the intrinsic variability of the classes could be detected. These records were integrated with the NDVI and the Hue-Saturation-Value (HSV). Ancillary data were used as well: they include DTMs, glacier data (Randolph Glacier Inventory 5.0), urban areas data (GHSL) and a global-scale lava mask.

2.6.3 Algorithms Expert systems were employed for image classification, in order to take into account data uncertainty, image interpretation expertise and multiple data sources. The classification procedure consists in a sequential decision tree using either multispectral and multitemporal attributes of the Landsat images and ancillary data. For improving the classifier performance, visual analytics and evidential reasoning were used as well. The expert system was run in Google Earth Engine and the code can be provided on specific request.

2.6.4 Validation data and results The products were validated by calculating commission and omission errors on a sample of 40 124 control points, adequately selected to be geographically and temporally well-distributed. In particular, a 27 268 pixels sample was used for the omission error (i.e. 1-PA) estimate and a 12 856 pixels sample for the commission error (i.e. 1-UA) estimate. The evaluation was performed using a systematic sample frame (1° latitude by 1° longitude grid) and the stratification to areas of high and low water probability. A point was randomly selected within each cell and each strata in order to evaluate commission and omission errors for different images, randomly selected for each sensor across the archive time span. Table 8 shows in detail the accuracy assessment results. Errors of omission were overall less than 5% and errors of commission less than 1%. Furthermore, all three sensors provide similar and good results, with differences less than 0.2% for commission errors and 1.2% for omission errors. Also, all sensors provided a better performance in detecting permanent water with respect to seasonal water (the omission error for seasonal water is always higher than the omission error for permanent water).

		UA [%]	PA [%]
Landsat 5	Overall	99.45	97.01
	Seasonal	98.80	74.91
	Permanent	99.56	98.79
Landsat 7	Overall	99.35	95.79
	Seasonal	98.38	73.82
	Permanent	99.50	97.72
Landsat 8	Overall	99.54	96.25
	Seasonal	98.53	77.40
	Permanent	99.66	99.10

Table 8. Accuracy of GWS

2.7 Forest/Non-forest

The FNF dataset developed by the Japan Aerospace Exploration Agency (JAXA) represents the global distribution

of forests for the years 2007-2010 with a spatial resolution of about 25 m (0.8 arcsec) (Shimada et al., 2014). Tree-covered areas larger than 0.5 ha and with a canopy cover over 10% are here defined as forests. Data are represented in the WGS84 CRS and can be accessed at https://www.eorc.jaxa.jp/ALOS/en/palsar_fnf/fnf_index.htm#. The legend of FNF consists of Forest, Non forest, and Water classes represented by values 1, 2, and 3, respectively. Forest is defined as the tree covered land with area larger than 0.5 ha and canopy cover over 10%.

2.7.1 Imagery For the 2007-2010 FNF land cover production, four global mosaics of ALOS PALSAR (Advanced Land Observing Satellite Phased Arrayed L-band Synthetic Aperture Radar) HH and HV data were created and classified. These mosaics were obtained from data acquired annually between June and October for the years 2007, 2008, 2009 and 2010. Similarly, global mosaics of ALOS-2 PALSAR-2 are created for years 2015-2020, but currently land cover maps are available only for 2015, 2016, and 2017. Mosaics are expressed in terms of gamma-naught (γ^0 , normalized radar cross section), a measure of the radar backscatter normalized with respect to the realistic illumination area.

2.7.2 Training data In order to discriminate forest from non forest areas, specific thresholds in terms of HV γ^0 was derived from histograms and cumulative distribution functions. The threshold estimations were based on training data from 15 regions of the world: Sumatra Indonesia, Papua New Guinea, Borneo, Malaysia, Philippines, East Asia, Japan, India, Europe-Russia, Australia, Amazon, Chile, Africa, North America and Central America. The training data represent 3-10 subareas in each of the 15 region for each class. They were selected with reference to Google Earth Imagery to be representative of diverse forest types as well as non-forest categories. There was 90 subareas representing forest, and 72 subareas representing non-forest. Each subarea comprise around 65 000 to 80 000 pixels (40 - 50 km²).

2.7.3 Algorithms The procedure followed for the FNF maps production consists in a multi-resolution segmentation of each mosaic, with the application of a 5x5 median filter. The software eCognition was employed for this purpose. Afterwards, the previously selected thresholds were applied for the detection of settlement areas (thresholds in terms of HH γ^0), forest areas (thresholds in terms of HV γ^0) and water bodies (thresholds in terms of HH γ^0 , standard deviation of HH and Geometric Density functions); non-forest areas are then computed accordingly.

2.7.4 Validation data and results The products were validated considering as references the Degree Confluence Project (DCP), the Forest Resource Assessment (FRA) and Google Earth Images (GEI). In particular, GEI data available from 2000 to 2012 were used, with a total of 4 114 reference points (1 456 points for forest and 2 548 for non-forest). As for the DCP is concerned, data were collected by volunteers, for a total number of 2 652 validation points. Finally, the FNF maps were compared with the FRA2005 and FRA2010 products of Food and Agriculture Organization (FAO). Accuracies resulted equal to 85%, 91% and 95% with respect to DCP, GEI and FRA 2005/2010. Table 9 reports more detailed information about the accuracy assessment based on DCP and GEI.

Product	Class	GEI			DCP		
		PA [%]	UA [%]	OA [%]	PA [%]	UA [%]	OA [%]
FNF 2007	F	82.59	95.31	91.49	68.8	86.43	85.19
	NF	97.33	89.49		91.26	91.26	
FNF 2008	F	81.13	94.31	90.58	68.65	83.01	84.78
	NF	96.79	88.65		90.2	92.72	
FNF 2009	F	81.51	95.04	90.98	61.38	74.79	82.36
	NF	97.2	88.89		88.57	92.74	
FNF 2010	F	81.64	95.33	91.13	66.67	89.8	87.14
	NF	97.37	88.98		93.46	93.46	

Table 9. Overall Accuracy of PALSAR FNF (2007-2010) with respect to DCP and GEI. F stands for Forest and NF for Non-forest

2.8 Tree canopy cover

The Tree canopy cover is one of the many datasets of the Global Forest Change project carried out by the University of Maryland, Google, USGS and NASA (Hansen et al., 2013). This dataset represent Tree canopy closure of the trees with at least 5m height. It has values from 0 to 100 which correspond to percentage of canopy closure. It represent a state of forest in 2000 and it was used as a starting point to compute products of forest gain and forest loss in the subsequent years. The Tree canopy cover has a resolution of 30 m at the equator (1 arcsec). It is available in the geographic reference system WGS84 and can be accessed at https://earthenginepartners.appspot.com/science-2013-global-forest/download_v1.7.html.

2.8.1 Imagery For the production of the Tree canopy cover Landsat 7 ETM+ images, collected during the growing season, were employed.

2.8.2 Training data The set of training data was extracted using ancillary information, in particular Quickbird imagery and existing layers of percent tree cover derived from Landsat in Web-Enabled Landsat Data (WELD) project (Hansen et al., 2011) and from MODIS (Hansen et al., 2003) data. Training data concern percent tree cover (i.e. Tree canopy cover), forest loss and forest gain.

2.8.3 Algorithms The processing of Landsat data followed a standard pre-processing schema that includes image re-sampling, Digital Numbers conversion to TOA reflectance, cloud detection and image normalization. The classification procedure consists in a bagged decision tree that relates the above-mentioned training data to Landsat data metrics. Landsat metrics were extracted for each band, namely statistical values of reflectance (mean, maximum, minimum and selected percentile values) and slope of the linear regression between band reflectance and image date.

2.8.4 Validation data and results Since estimation of forest loss and forest gain was primary goal of the Global Forest Change project, the validation on global level was performed only for forest loss and forest gain products, and not for the Tree canopy cover.

3. CONCLUSIONS

In this paper nine global high-resolution land cover are described. All of the datasets have same purpose - detect LC on global level. However, approach for doing so is different depending on the data producer.

In most of the cases, primary source of satellite imagery comes from the optical satellite sensors such as Landsat ((GL30, GSW, GHS LDSMT, Tree Canopy Cover, FROM GLC v.2010 and v.2015) and Sentinel-2 (FROM GLC v. 2015). Few HRLC are derived starting from radar images such as Sentinel-1 (GHS BU S1NODSM) and ALOS PALSAR/ALOS2 PALSAR2 (FNF). Lastly, one HRLC took advantage of both optical and radar sensors - Sentinel-1 and Landsat 8 (WSF). In case of data gaps, common solution was to use imagery from the same sensor in the period prior or after product baseline year, but in the rare cases data from other sensors were exploited.

Most of the HRLC are focused on detection of single class such as built-up area (GHS BU S1NODSM, GHS LDSMT, GUF and WSF), forest (FNF, Tree Canopy Cover) and water (GSW). Nevertheless, GL30 and FROM-GLC have 10 classes which in a generic way capture different types of Earth cover.

Selection of training data prevalently relied on photo-interpretation of satellite imagery in combination with multiple ancillary information (e.g. previous versions of the same dataset, existing land cover data, MODIS EVI, OSM, etc). In particular, this is valid for GL30, FROM-GLC, first release of GHS BU LDSMT, and Tree canopy cover. FROM-GLC v.2017 was derived from the same samples as FROM-GLC v.2015. GSW and FNF relied mostly on photo-interpretation and band statistics, while WSF and GUF used automatic training extraction based on certain thresholds. It is interesting to notice that for GHS BU products incremental learning was used. I.e. each previous version of product was used in combination with other ancillary data for creating training for the following product.

FROM-GLC, WSF and Tree canopy cover were classified with with single classification algorithm. In case of former two the SVM was used, while for the latter one used bagged decision tree. Other datasets were derived by custom made processing chains created to address the issues of global variability of land cover and/or to accommodate specific characteristics of input imagery and large volumes of data. GL30 had POK-based classification; GUF was computed in UFP processing environment with unsupervised classification method SVDD; GSW was produced by expert systems by a sequential decision tree taking into account multi-temporal and multi-spectral attributes of Landsat images; FNF was developed in the rule-based approach depending on the thresholds of HV or HH γ^0 .

Different sampling schemes were applied for validation of different products. Often splitting world into a regular grid was adopted, and then either some cells of the grid were selected for extracting samples (e.g. GL30) or there was certain number of samples determined for each cell of the grid (FROM-GLC, GSW, WSF). Stratified random sampling was often exploited for the sample allocation (GUF, GSW, WSF). GUF validation samples were extracted for 12 urban cities. FNF does not report sampling strategy for GEI, while DCP are distributed on integer degree coordinates worldwide. GHS BU products were inter-compared with GUF. Tree Canopy Cover does not report validation. GHS BU products were inter-compared with GUF.

Even though the approach for producing land cover is different, the aspect that matter the most to users is the accuracy. OA gives an overall impression about product accuracy, however it is important also to look at the class accuracy as well. For example, OA GL30 is better than FROM-GLC, but looking at the UA FROM-GLC has significantly better accuracy for Ice and snow class. OA in case of binary maps is less relevant for the

comparison. This is because binary map consists of class of interest and class that includes everything else. The latter one is usually more accurate and covers larger area, thus it unjustifiably increases OA value. Therefore, in case of binary maps it is more appropriate to look at the accuracy of class of interest. Binary map of forest - FNF - outperforms both GL30 and FROM-GLC regarding the UA of Forest class. The lowest UA of FNF with respect to GEI is 94%, while UA of GL30 and FROM-GLC for the same class is 80.49% and 83.47%, respectively. GSW has impressive UA for water 99% that is significantly better than UA of GL30 or FROM-GLC for water. Among binary products for built up class, analyses made within validation of WSF show that on average WSF has the best accuracy with AA% of 86.37, followed by GUF (80.13), GHSL(71.09) and GLC30 (67.79).

After the review of literature it is evident that different methodologies are needed to exploit information from different imagery sources and for detecting different LC classes. However, couple of improvements could be made in reporting, as well as in a production to facilitate exploitation of the HRLC. One flaw of the HRLC documentation is that sometimes procedure of validation is not clear enough, so it is not possible to understand if accuracy was determined with a statistically valid number of samples and sampling strategy (e.g. FNF). Even bigger issue is when accuracy is not determined or reported (e.g. Tree Canopy Cover). Another difficulty is use of different metrics or style for reporting accuracy (e.g. GHSL BU). Lastly, classes with the same name have different definition (i.e. do not represent exactly the same features).

The mentioned issues could be addressed by a standard/de facto standard which regulates definition of the classes (e.g. LCCS of FAO), accuracy assessment procedure and reporting. This would increase inter-operability and make it easier for user to decide which HRLC to use or give them possibility of using multiple HRLC simultaneously. The need of standard is critical now given that production of land cover is an ongoing process.

REFERENCES

- Brovelli, M. A., Molinari, M. E., Hussein, E., Chen, J., Li, R., 2015. The first comprehensive accuracy assessment of GlobeLand30 at a national level: methodology and results. *Remote Sensing*, 7(4), 4191–4212.
- Chen, J., Cao, X., Peng, S., Ren, H., 2017. Analysis and Applications of GlobeLand30: A Review. *ISPRS International Journal of Geo-Information*, 6(8), 230.
- Chen, J., Chen, J., Liao, A., Cao, X., Chen, L., Chen, X., ..., Mills, J., 2015. Global land cover mapping at 30m resolution: A POK-based operational approach. *ISPRS Journal of Photogrammetry and Remote Sensing*, 103, 7–27.
- Corbane, C., Pesaresi, M., Politis, P., Syrri, V., Florczyk, A. J., Soille, P., ..., Kemper, T., 2017. Big earth data analytics on Sentinel-1 and Landsat imagery in support to global human settlements mapping. *Big Earth Data*, 1(1-2), 118–144.
- Esch, T., Bachofer, F., Heldens, W., Hirner, A., Marconcini, M., Palacios-Lopez, D., ..., Gorelick, N., 2018. Where we live—A summary of the achievements and planned evolution of the Global Urban Footprint. *Remote Sensing*, 10(6), 895. Number: 6.
- Gong, P., Wang, J., Yu, L., Zhao, Y., Zhao, Y., Liang, L., ..., 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), 2607–2654.
- Hansen, M. C., DeFries, R. S., Townshend, J. R. G., Carroll, M., Dimiceli, C., Sohlberg, R. A., 2003. Global Percent Tree Cover at a Spatial Resolution of 500 Meters: First Results of the MODIS Vegetation Continuous Fields Algorithm. *Earth Interactions*, 7(10), 1 - 15.
- Hansen, M. C., Egorov, A., Roy, D. P., Potapov, P., Ju, J., Turubanova, S., Kommareddy, I., Loveland, T. R., 2011. Continuous fields of land cover for the conterminous United States using Landsat data: first results from the Web-Enabled Landsat Data (WELD) project. *Remote Sensing Letters*, 2(4), 279-288.
- Hansen, M. C., Potapov, P. V., Moore, R., Hancher, M., Turubanova, S. A., Tyukavina, A., ..., Townshend, J. R. G., 2013. High-Resolution Global Maps of 21st-Century Forest Cover Change. *Science*, 342(6160), 850–853. Number: 6160 Publisher: American Association for the Advancement of Science Section: Report.
- Jacobson, A., Dhanota, J., Godfrey, J., Jacobson, H., Rossman, Z., Stanish, A., Walker, H., Riggio, J., 2015. A novel approach to mapping land conversion using Google Earth with an application to East Africa. *Environmental Modelling Software*, 72, 1 - 9.
- Manakos, I., Chatzopoulos-Vouzoglani, K., Petrou, Z., Filchev, L., Apostolakis, A., 2014. Globalland30 Mapping Capacity of Land Surface Water in Thessaly, Greece. *Land*, 4(1), 1–18.
- Marconcini, M., Metz-Marconcini, A., Üreyen, S., Palacios-Lopez, D., Hanke, W., Bachofer, F., ..., Strano, E., 2020. Outlining where humans live, the World Settlement Footprint 2015. *Scientific Data*, 7(1), 242. Number: 1 Publisher: Nature Publishing Group.
- Otsu, N., 1979. A Threshold Selection Method from Gray-Level Histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, 9(1), 62-66.
- Pekel, J.-F., Cottam, A., Gorelick, N., Belward, A. S., 2016. High-resolution mapping of global surface water and its long-term changes. *Nature*, 540(7633), 418–422. Number: 7633 Publisher: Nature Publishing Group.
- Pesaresi, M., Syrri, V., Julea, A., 2016. A New Method for Earth Observation Data Analytics Based on Symbolic Machine Learning. *Remote Sensing*, 8(5).
- Shimada, M., Itoh, T., Motooka, T., Watanabe, M., Shiraishi, T., Thapa, R., Lucas, R., 2014. New global forest/non-forest maps from ALOS PALSAR data (2007–2010). *Remote Sensing of Environment*, 155, 13–31.
- Sun, B., Chen, X., Zhou, Q., 2016. Uncertainty assessment of GlobeLand30 land cover data set over Central Asia. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLI-B8, 1313-1317.
- Yang, Y., Xiao, P., Feng, X., Li, H., 2017. Accuracy assessment of seven global land cover datasets over China. *ISPRS Journal of Photogrammetry and Remote Sensing*, 125, 156 - 173.