remote sensing

Article

The First Comprehensive Accuracy Assessment of GlobeLand30 at a National Level: Methodology and Results

Maria Antonia Brovelli [1],*, Monia Elisa Molinari [1], Eman Hussein [1], Jun Chen [2] and Ran Li [2]

[1] Hydroinformatics Lab, DICA, Politecnico di Milano, Como Campus, via Valleggio 11, Como 22100, Italy; E-Mails: moniaelisa.molinari@polimi.it (M.E.M.); eng.emanhussein@gmail.com (E.H.)

[2] National Geomatics Center of China, 28 Lianhuachi West Road, Haidian District, Beijing 100830, China; E-Mails: chenjun@nsdi.gov.cn (J.C.); liran@nsdi.gov.cn (R.L.)

* Author to whom correspondence should be addressed; E-Mail: maria.brovelli@polimi.it; Tel.: +39-31-332-7336; Fax: +39-31-332-7321.

Academic Editors: Ioannis Gitas and Prasad S. Thenkabail

Abstract: As result of the "Global Land Cover Mapping at Finer Resolution" project led by National Geomatics Center of China (NGCC), one of the first global land cover datasets at 30-meters resolution (GlobeLand30) has been produced for the years 2000 and 2010. The first comprehensive accuracy assessment at a national level of these data (excluding some comparisons in China) has been performed on the Italian area by means of a benchmarking with the more detailed land cover datasets available for some Italian regions. The accuracy evaluation was based on the cell-by-cell comparison between Italian maps and the GlobeLand30 in order to obtain the confusion matrix and its derived statistics (overall accuracy, allocation and quantity disagreements, user and producer accuracy), which help to understand the classification quality. This paper illustrates the adopted methodology and procedures for assessing GlobeLand30 and reports the obtained statistics. The analysis has been performed in eight regions across Italy and shows very good results: the comparison of the datasets according to the first level of Corine Land Cover nomenclature highlights overall accuracy values generally higher than 80%.

Keywords: land cover; GlobeLand30; accuracy assessment; Italy

1. Introduction

Global land cover data represent highly valuable information to monitor the extension and status of land resources and give an important contribution to understanding the balance between global land cover pattern, climate and biochemistry of the earth system [1,2]. Thanks to the continuous advances in remote sensing sensors and mapping technologies, the availability of these datasets is rapidly increasing. It is well established that the knowledge of the classification accuracy is a key factor to improve the production of these datasets and understand their suitability in many different applications. Classification accuracy assessment of land cover maps represents a very important research topic and many studies, recommendations and guidelines about it have been published over the years [3–6].

Several global land cover maps derived from different remotely sensed sources are now freely available for research purposes: the 1 km International Geosphere-Biosphere Programme Data and Information System Cover (IGBP-DISCover) map [7]; the 1 km University of Maryland (UMD) land cover map [8]; the 1 km Global Land Cover 2000 (GLC2000) map [9]; the 500 m MODIS land cover maps [10]; and the 300 m GlobCover land cover maps [11]. Nowadays, a new global land cover dataset at 30 m resolution has been added. This dataset, named GlobeLand30 [12], refers to the years 2000 and 2010. It was produced by the Chinese government and then donated to the United Nations with the aim of contributing to research on sustainable development and climate change. Since its release for open access at the end of September 2014, GlobeLand30 has been downloaded by more than 1000 scientists from 70 countries. Since then, it has been also used in some analysis on spatial distribution pattern and temporal fluctuation of global land surface water [13,14].

According to preliminary tests, the GlobeLand30 achieves values of overall accuracy of over 80% [12]. Obviously, an independent assessment of its quality could greatly increase the interest and usage of the product. This study wants to further contribute in the classification quality evaluation presenting the first comprehensive thematic accuracy assessment carried out on the available Italian datasets. In particular, Section 2 describes the benchmarking analysis performed between GlobeLand30 and the different Italian regional reference data by illustrating the adopted methodology, based on the confusion matrix approach, and the main characteristics of the datasets; Section 3 reports the statistics obtained for the different Italian regions in which global land coverage data are available and the multiple land cover classes considered; finally, the conclusion in Section 4 presents the main research results and needs.

2. Methods and Data

2.1. Confusion Matrix and Derived Indexes

Many different methods and indexes have been proposed in the literature to describe the classification quality of thematic maps. Within this study, the accuracy evaluation of GlobeLand30 has been performed through a comparison with reference datasets by following the "good practices" suggested by Foody [5] and Olofsson *et al.* [6,15]. According to these authors, a scientifically rigorous method for the accuracy assessment should involve three important steps: the response design, the sampling design and the accuracy analysis.

The response design is defined as "the protocol for the acquisition of the reference data label" [5]. The spatial assessment unit [16], the source of reference data and the reference labeling process are some of the most important factors to take into account within this step. Another key point is to ensure the higher quality of reference classification with respect to the land map being evaluated. Furthermore, if the collection of reference data for the whole region of interest cannot be feasible, a probability sampling design is needed for selecting a sample of data on which the accuracy analysis will be performed. This step requires an accurate choice of both the sample size and sampling methods [17].

For the current study, the reference data have been selected among the already existing Italian land cover datasets. At present, the only land coverage maps covering the whole national area have been produced in the frame of Corine Land Cover (CLC), a European project specifically intended for detection and monitoring of the characteristics of land cover and use; although homogeneous and available for different time periods, these datasets are not suitable as reference for the comparison with GlobeLand30 because of their low resolution (100 m). Fortunately, land cover maps characterized by higher quality than GlobeLand30 have been produced and made available by some Italian regions. This allowed us to perform an accuracy analysis for each of them. Furthermore, all the available reference data were taken into consideration in the accuracy analysis process, bringing to a regular distributed sample for each region.

The accuracy analysis has been carried out by means of the confusion or error matrix [18], which is derived from the spatial comparison between the classified dataset and the reference one by selecting the pixel as the spatial assessment unit. The result of this comparison is a square matrix characterized by a number of rows and columns equal to the total number of the considered land cover classes; in particular, the elements belonging to the main diagonal represent the correctly classified data, while the off-diagonal elements identify the classification errors.

Many indexes describing the quality of the classification can be derived from the error matrix [19]. In accordance with the "good practices", the most commonly used overall accuracy (OA), user (UA) and producer (PA) accuracy have been calculated. Furthermore, the more recently proposed allocation (AD) and quantity (QD) disagreement have been considered.

Overall accuracy is one of the most popular agreement measures and indicates the percentage of correctly classified pixels. It can be simply computed as shown in Equation (1), where q is the number of the classes, n_{ii} are the diagonal elements and n represents the total number of considered pixels. Unfortunately in the literature, it is impossible to find a unique threshold that defines the acceptable values of overall accuracy for the image classification; Anderson *et al.* [20] propose a value of at least 85%, Pringle *et al.* [21] recommend a value over 70%, whereas Thomlinson *et al.* [22] consider accurate a classification when the OA is at least equal to 85% and no class is less than 70%.

$$OA = \frac{\sum_{i=1}^{q} n_{ii}}{n} \times 100 \tag{1}$$

Unlike the overall accuracy, user and producer accuracies allow understanding the classification quality of each single class i. In particular, the former expresses the accuracy of the classification from the user prospective and can be defined as the percentage of classified pixels that correctly match the ground truth. The latter refers to the accuracy from the producer prospective and can be defined as the percentage of the pixels of ground truth correctly detected in the classified map. User and producer

accuracies, respectively, can be computed as proposed in Equations (2) and (3); n_{+i} and n_{i+} represent the marginal sum of the columns and the marginal sum of the rows, respectively.

$$UA_i = \frac{n_{ii}}{n_{i+}} \times 100 \tag{2}$$

$$PA_i = \frac{n_{ii}}{n_{+i}} \times 100 \tag{3}$$

For each of the above-mentioned parameters, the error/uncertainty analysis has also been performed by calculating variance, standard error, and confidence intervals at 95% probability. For more information about the adopted equations refer to Olofsson *et al.* [6].

Moreover, in recent years Pontius and Millones [23] proposed two new parameters specifically intended for estimating the disagreement component between classified and reference datasets. The allocation disagreement (Equation (4)) is defined as the disagreement value that "is due to the less than optimal match in the spatial allocation of the categories" while the quantity disagreement (Equation (5)) is the part of disagreement "due to the less than perfect match in the proportion of the categories".

$$AD = \frac{\sum(2 * min\left(\frac{n_{+i}}{n} - \frac{n_{ii}}{n}, \frac{n_{i+}}{n} - \frac{n_{ii}}{n}\right))}{2} \times 100 \tag{4}$$

$$QD = \frac{\sum\left|\frac{n_{+i}}{n} - \frac{n_{i+}}{n}\right|}{2} \times 100 \tag{5}$$

2.2. Available Data

2.2.1. GlobeLand30

GlobeLand30 represents one of the first efforts in mapping a global land cover at 30 m resolution and it has been realized under the "Global Land Cover Mapping at Finer Resolution" project led by National Geomatics Center of China (NGCC). The dataset has been generated for the two baseline years of 2000 and 2010 and required the collection and classification of more than 10,000 scenes, primarily obtained from Landsat Thematic Mapper (TM) and Enhanced TM plus (ETM+) satellites; for the years 2010, also the images of the Chinese Environmental and Disaster (HJ-1) satellite were considered. For more details about the pixel-object-knowledge-based (POK-based) classification approach used to produce GlobeLand30 data, refer to Chen *et al.* [12].

For the current study, GlobeLand30 is available in raster format with 6 different tiles covering the Italian territory; the data are provided in WGS84 (World Geodetic System 1984) reference system and UTM (Universal Transverse Mercator) projection: 3 tiles refer to UTM zone 32 N and 3 to UTM zone 33 N. The legend is based on eleven land cover categories as shown in Figure 1.
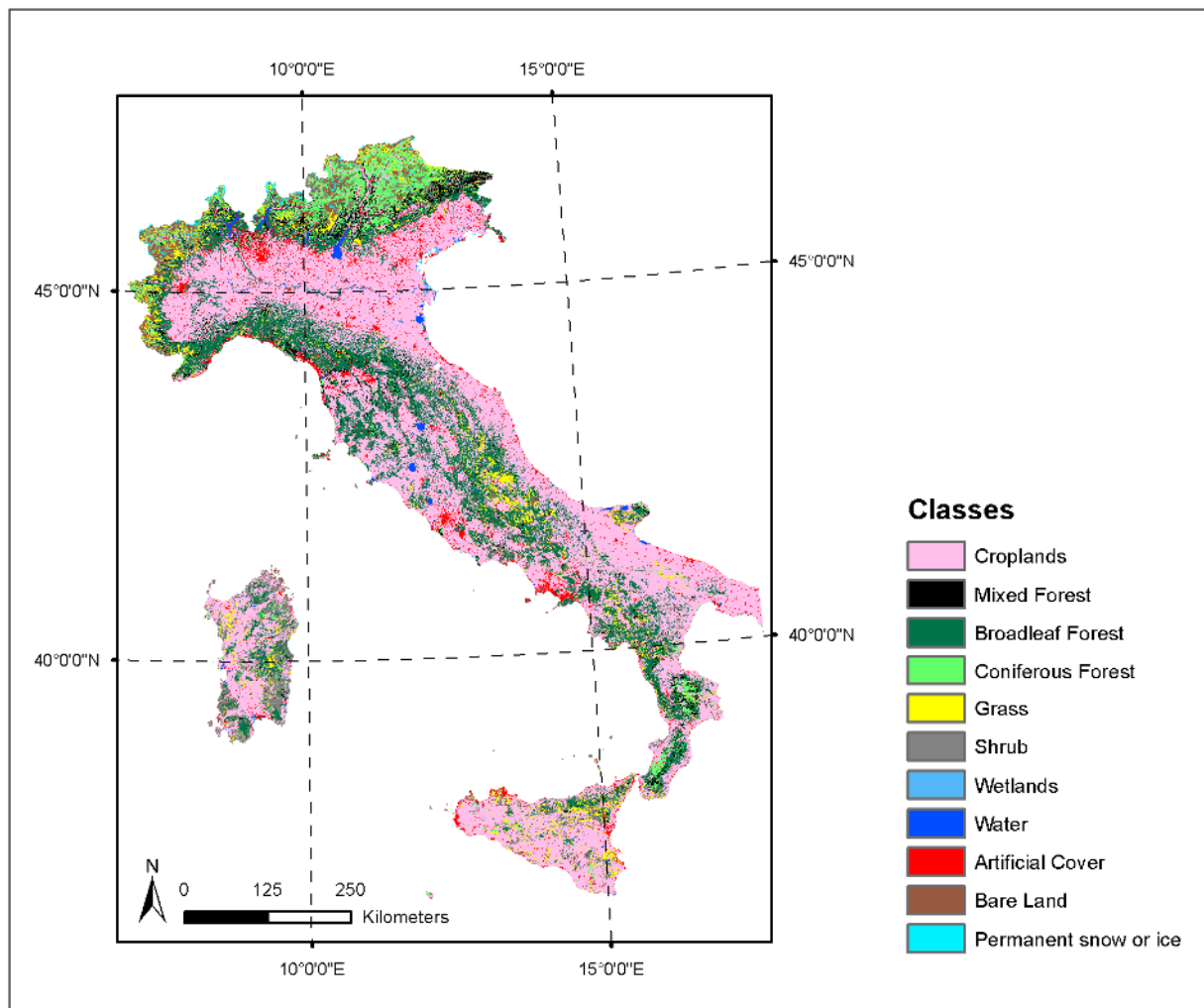
Figure 1. GlobeLand30 data for the Italian territory.

2.2.2. Italian Land Coverage Dataset

As mentioned in Section 2.1, very detailed land cover maps have been produced for some Italian regions and, in most cases, they are available as open data for users. Table 1 lists the available regional datasets and specifies some characteristics to take into account in the benchmarking process; currently we have available data for eight out of twenty regions, and four of them provide data for comparison with both years GlobeLand30 maps. Figure 2 shows the eight regions of interest for comparison. It is evident that the geographic distribution cannot be considered representative for the whole of Italy. In fact the paper proposes statistics and indexes for single Italian regions.

All the regional datasets are available in ESRI (Environmental Systems Research Institute) Shapefile format, but they refer to different years, different reference systems, and sources, like QuickBird or aerial photos interpretation. Furthermore the datasets scales range between 1:10,000 and 1:25,000 with accuracies equal or better than 5 m. Finally, most of them are completely compliant with the CLC nomenclature, a hierarchical classification system based on three levels whose first level includes five classes: *artificial surfaces, agricultural areas, forests and semi natural areas, wetlands* and *water bodies*.

Table 1. Characteristics of the Italian datasets available for the study; it is worth noting that Bolzano and Trento are the two Autonomous Provinces of Trentino-Alto Adige region. (ETRS: European Terrestrial Reference System; TM: Transverse Mercator).

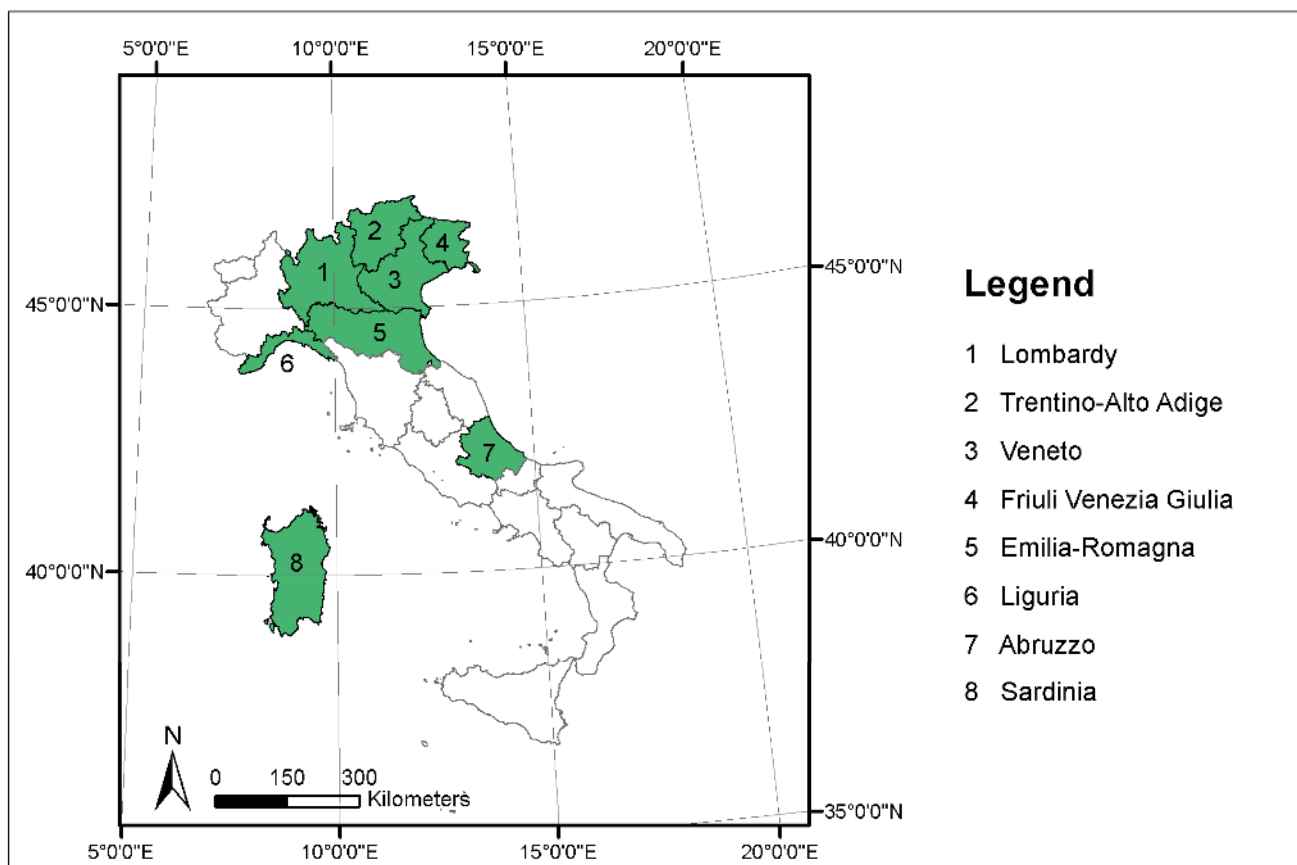| Dataset	Year	Reference System
LOMBARDY	1999–2000 2012	WGS84/UTM zone 32N
LIGURIA	2000 2009–2012	ETRS89/UTM zone 32N
SARDINIA	1997–2000 2003–2006	Roma40/Gauss Boaga Ovest
EMILIA-ROMAGNA	2003 2008	ETRS89/UTM zone 32N
VENETO	2007–2009	Roma40/Gauss Boaga Ovest
FRIULI VENEZIA GIULIA	2000	ETRS89/TM33
BOLZANO	1992–1997	ETRS89/UTM zone 32N
TRENTO	2000	ETRS89/UTM zone 32N
ABRUZZO	1997	WGS84/UTM zone 33N

Figure 2. Italian land coverage datasets currently available for the study.

2.3. Data Processing

Due to the different characteristics in terms of format, legend, scale and reference system between GlobeLand30 and the Italian regional maps, a data pre-processing phase was required before calculating the confusion matrix and its derived statistics; the workflow carried out on each dataset is shown in Figure 3, which highlights the main steps needed for all the datasets (continuous boxes) together with the additional ones applied only to some maps (dashed boxes).

As reported in the workflow, the first main step of the data processing involves the rasterization of the Italian vector datasets in order to allow the benchmarking with the GlobeLand30, which is provided in raster format. Taking into account the different levels of detail of the two datasets, we decided to calculate two raster maps characterized by cell sizes, respectively, equal to 30 m and 5 m: the former corresponds to the GlobeLand30 resolution; the latter is representative of the regional data. Two methods of rasterization, center and area prevalence [24], were compared. This step led to the generation of three raster maps: 30 m and 5 m with the rasterization center method and 30 m with the area prevalence one. All these maps were used in the subsequent steps to evaluate if the rasterization resolution and method can significantly affect the comparison results.
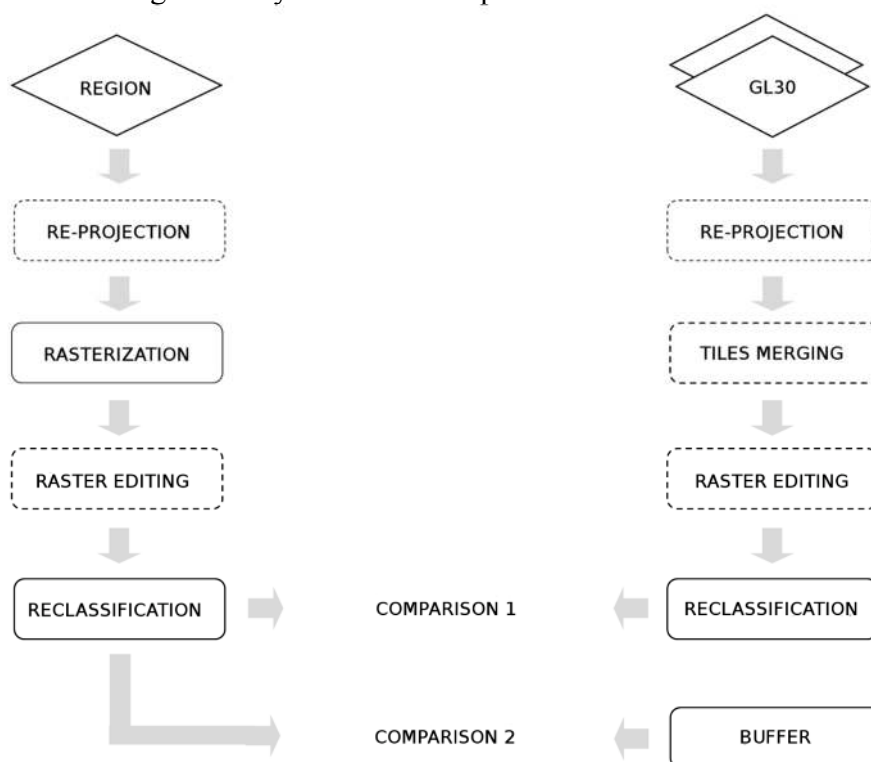

Figure 3. Data processing workflow performed on Italian datasets and GlobeLand30 (GL30). Dashed boxes represent additional steps required for only some datasets.

The reclassification is the other important step highlighted in the workflow as no comparison process can be carried out since GlobeLand30 and the Italian datasets are characterized by different thematic classifications. Both of the maps were reclassified according to the same legend; in particular, two classification methods were considered. The first legend (hereinafter referred as "first reclassification method") was based on the first five levels of CLC nomenclature previously listed (*artificial surfaces, agricultural areas, forests and semi natural areas, wetlands and water bodies).*

Considering the wide availability of GlobeLand30 classes for *forests and semi natural areas* Corine category, a second thematic legend (hereinafter referred as "second reclassification method") was defined by replacing this class with four subclasses (*forests, grass and/or herbaceous vegetation associations, open spaces with little or no vegetation, glaciers and perpetual snow*). As shown in Table 2, GlobeLand30 classes have a good correspondence with the new ones and were easily reclassified as well as the Italian Corine-based datasets, which automatically led back to the first level of classification; not completely CLC compliant datasets were instead reclassified by manually assigning to each category the more suitable Corine class. It is worth noting that, in some cases, errors due to interpretation of ambiguous classes may have been introduced.

Table 2. First and second reclassification methods: correspondence between GlobeLand30 and Corine classes.

Corine Legend	GlobeLand30 Legend
1 Artificial surfaces	1 Artificial cover
2 Agricultural areas	2 Croplands
3 Forests and semi natural areas	3 Mixed forest, Broadleaf forest, Coniferous forest, Grass, Shrub, Bare land, Permanent ice or snow
OR	OR
3.1 Forests	3.1 Mixed forest, Broadleaf forest, Coniferous forest
3.2 Scrub and/or herbaceous vegetation associations	3.2 Grass, Shrub
3.3.0 Open spaces with little or no vegetation	3.3.0 Bare land
3.3.5 Glaciers and perpetual snow	3.3.5 Permanent ice or snow
4 Wetlands	4 Wetlands
5 Water bodies	5 Water

Besides the main above-mentioned steps, some case studies required further additional procedures to enable the land cover maps comparison. Some Italian datasets were available in a different reference system compared with GlobeLand30; therefore a re-projection procedure was performed; on the other hand, Italian regions often cover different GlobeLand30 tiles, corresponding to different UTM zones: in this case both a re-projection and a merging of the tiles were needed to obtain a unique GlobeLand30 map. Finally, GlobeLand30 and the Italian datasets present differences in the identification of the coastline, mainly due to their different level of detail; although the pixels involved are few compared to the total (less than 0.50%), they could alter the estimate of user and producer accuracy for the smallest classes. Therefore a raster editing procedure was needed to take into account the error around this line. As the sea is not considered in the classification (null value), in some cases a classified pixel on one map corresponds to a null cell on the other; in order to solve this problem a new class was added; this class contains the areas equal to sea at least in one of the two maps.

As shown in the workflow, once the two datasets were reclassified, a first comparison was carried out. A second comparison, similar to what was done in Gallego [25], was then performed to evaluate the influence of the GlobeLand30 co-location tolerance on the classification quality, as it was evaluated by the data producers (70 m): this means that a pixel could be located in a buffer of 70 m with respect to its true position. Classification errors due to this spatial mismatching should be particularly evident at the border between two different classes. To verify this behavior, all cells

belonging to a buffer of 70 m around GlobeLand30 classes border were eliminated and the confusion matrix and statistics were calculated on the other pixels.

3. Results and Discussion

This section shows the outcomes of the benchmarking between the GlobeLand30 data and the Italian regional datasets. In order to avoid a long and repetitive report about all the different statistics calculated for each Italian region, we decided to propose with larger details only the accuracies obtained for a single case study, the Lombardy Region, and provide a general summary of the other regional results.

3.1. Lombardy Case Study

In line with the INSPIRE Directive recommendations, the Lombardy Region has launched the development of Infrastructure for Spatial Information with the aim of collecting and sharing the regional geographical data that are thus made easily accessible to users. Among the several datasets and services provided, the Infrastructure includes the DUSAF ("Destinazione d'Uso dei Suoli Agricoli e Forestali") Database [26], a valuable instrument to monitor land cover regional changes.

DUSAF, Italian acronyms for "Use Categories of Agricultural and Forest Soil", is a land cover database created in 2000–2001 within a project funded by the Lombardy Region and carried out by the Regional Authority for Services to Agriculture and Forests (ERSAF) with the cooperation of the Regional Agency for the Protection of the Environment (ARPA). The database is periodically updated through the photointerpretation of aerial photos integrated with regional databases information and it is currently available for five temporal periods (Table 3). Each release is freely downloadable from Lombardy Region Geoportal [27] and includes a polygonal layer at 1:10,000 scale information that depicts the regional land use and cover. The adopted legend is structured in a hierarchical way and it is based on five levels of investigation: the first three levels comply with the CLC nomenclature while the fourth and fifth ones represent additional levels specifically intended for identifying characteristics of the Lombardy territory.

Table 3. Characteristics of DUSAF database releases.

Characteristics	DUSAF 1.1	DUSAF 2.0	DUSAF 2.1	DUSAF 3.0	DUSAF 4.0
Year	1999–2000	2005–2007	2007	2009	2012
Extension	Whole region	Whole region	Whole region	Limited to some provinces	Whole region

Due to the wide availability of the Lombardy land cover maps, the benchmarking has been performed on both the GlobeLand30 datasets provided for years 2000 and 2010; in particular, a first comparison has been carried out between GlobeLand30 2000 and DUSAF1.1, since the two maps refer to the same year. Furthermore, a comparison between GlobeLand30 2010 and DUSAF4.0, which has been selected for its areal completeness instead of DUSAF3.0, has been performed. This analysis was preceded by an evaluation of differences between DUSAF3.0 and DUSAF4.0 in their overlapping areas, an evaluation that highlights an agreement of 99.8% between the two maps.

Figure 4 shows the processing steps performed on the datasets for the benchmarking analysis. Since both the datasets are provided in the WGS84/UTM32N reference system and the Lombardy case study does not present problems with coastlines, no re-projection and raster editing steps were necessary. Thus, only the rasterization and reclassification procedures were applied to DUSAF maps, whereas GlobeLand30 data required, before the reclassification and buffering, an additional step in order to merge the two tiles covering the case study area.
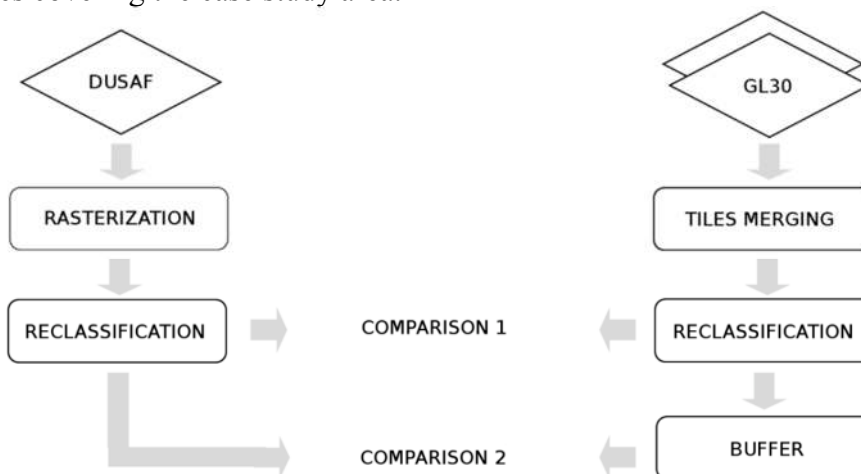
Figure 4. Lombardy Region case study: data processing workflow performed on DUSAF and GlobeLand30 (GL30) datasets. (DUSAF, Italian acronyms for "Use Categories of Agricultural and Forest Soil").

3.1.1. Comparison between GlobeLand30 2000 and DUSAF1.1

Table 4 reports the overall results obtained from the comparison that has been performed between DUSAF1.1 and GlobeLand30 2000; both maps have been reclassified according to the first reclassification method (CLC first level) mentioned in Section 2.3. The most recommended statistics, such as overall accuracy (confidence intervals at 95% probability), allocation disagreement and quantity disagreement, have been calculated for two different cases (with and without buffer application) and, for each of them, three resolution datasets (30 m, 30 m "prevalence", and 5 m) have been considered.

Regarding the *no buffer* case, the statistics show the high values for the OA coefficient, 86%, and the disagreement percentage mainly due to the allocation component, of 11%, rather than the quantitative one, of only 2%. Moreover, the removal of the cells influenced by co-location tolerance (applied in the *buffer* case) leads to a reduction of the AD value by 4%, which explains the consequent increase of the OA value, from 86% to 90%. In both the considered cases, the analysis of the data highlights that there are no significant changes in the results with different input dataset resolutions.

In addition to the above presented general results, specific statistics have been derived from the confusion matrix in order to better understand the accuracy distribution among the CLC classes (*artificial surfaces*, *agricultural areas*, *forests and semi natural areas*, *wetlands*, *water bodies*); Figure 5 presents the per-class values of user accuracy and producer accuracy, divided as usual into the different cases of *no buffer* and *buffer*. Table 5 reports the confidence intervals at 95% probability calculated for PA and UA indexes. Since input resolution does not imply changes in the results, the proposed statistics refer only to the 30 m resolution dataset.

Table 4. Comparison between GlobeLand30 2000 and DUSAF1.1 (I reclassification): overall statistics and number of samples for the two cases of *no buffer* and *buffer* and for different resolutions. OA: overall accuracy (confidence intervals at 95% probability); AD: allocation disagreement; QD: quantity disagreement.

Case Study	Resolution	OA (%)	AD (%)	QD (%)	Samples (-)
	30 m	86.4 ± 0.013	11.6	2.0	26,520,624
NO BUFFER	30 m (prevalence)	86.5 ± 0.013	11.5	2.0	26,547,419
	5 m	86.3 ± 0.002	11.7	2.0	954,739,157
	30 m	90.5 ± 0.012	7.4	2.0	22,691,322
BUFFER	30 m (prevalence)	90.6 ± 0.012	7.3	2.0	22,714,156
	5 m	90.8 ± 0.002	7.1	2.0	800,454,633
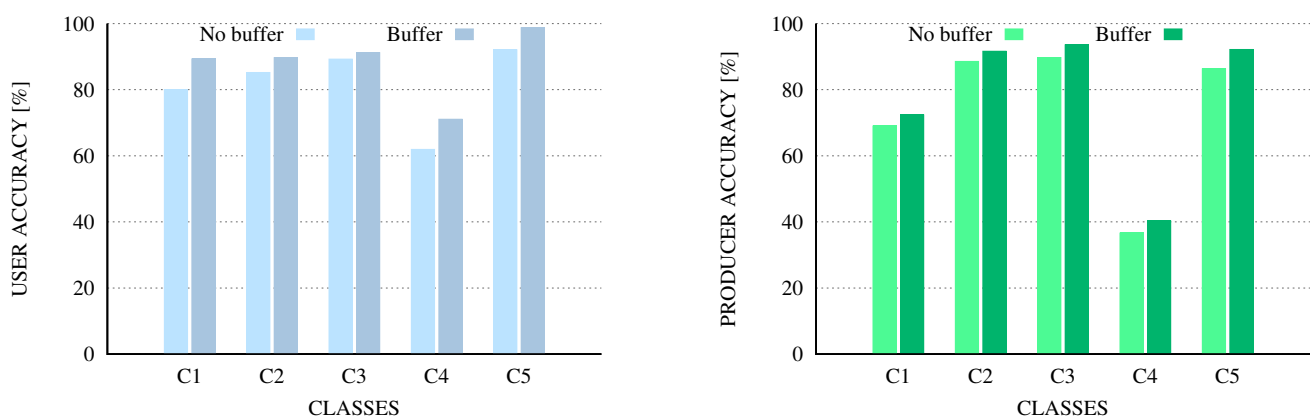
Figure 5. Comparison between GlobeLand30 2000 and DUSAF1.1 (I reclassification): per-class statistics referred to 30 m resolution input datasets. C1: artificial surfaces, C2: agricultural areas, C3: forest and semi natural areas, C4: wetlands, C5: water bodies.

Table 5. Comparison between GlobeLand30 2000 and DUSAF1.1 (I reclassification): per-classes confidence intervals at 95% probability calculated for user (UA) and producer accuracies (PA); data refer to 30 m resolution input datasets.

Index	C1		C2		C3		C4		C5	
	No Buffer	Buffer	No Buffer	Buffer	No Buffer	Buffer	No Buffer	Buffer	No Buffer	Buffer
UA	±0.046	±0.045	±0.020	±0.018	±0.019	±0.018	±0.663	±0.821	±0.059	±0.027
PA	±0.042	±0.050	±0.016	±0.015	±0.017	±0.015	±0.394	±0.507	±0.067	±0.061

The data analysis identifies *water bodies*, *forests and semi natural areas* and *agricultural areas* as the best detected CLC classes, with accuracies in the *no buffer* case always greater than 85%; in particular, *water bodies* class registers the maximum value of UA, equal to 92%, while *forest and semi natural areas* is the class with the greatest PA percentage, with 90%; *artificial surfaces* class presents a high UA value, around 80%, while the PA, compared to the other mentioned classes, decreases to 69%. Finally, results highlight major difficulties in the detection of the *wetlands* class, which is characterized by lower values of accuracies (62% for UA and 37% for PA). However, the smaller accuracy does not affect in a significant way the overall accuracy values for Lombardy due to the fact that this class covers only 0.1% of the whole region. With regard to the *buffer case*, as already

noted above, the buffer cells removal allows us to obtain further improvement in all the considered statistic values.

A second comparison between GlobeLand30 2000 and DUSAF1.1 has been conducted with regard to the second reclassification method, which divides the *forests and semi natural areas* class into the four subclasses: *forests, herbaceous and/or grass vegetation associations, open spaces with little or no vegetation, glaciers and perpetual snow*. The comparison between the first method results (Table 4) and the new calculated statistics (Table 6) shows that the introduction of a greater level of classification detail entails an increase of both the allocation and quantity disagreement values; consequently, the OA value decreases: from 86% to 77% in the *no buffer* case, and from 90% to 82% in the *buffer* case.

Table 6. Comparison between GlobeLand30 2000 and DUSAF1.1 (II reclassification): overall statistics and number of samples for the two cases of *no buffer* and *buffer* and for different resolutions. OA: overall accuracy (confidence intervals at 95% probability); AD: allocation disagreement; QD: quantity disagreement.

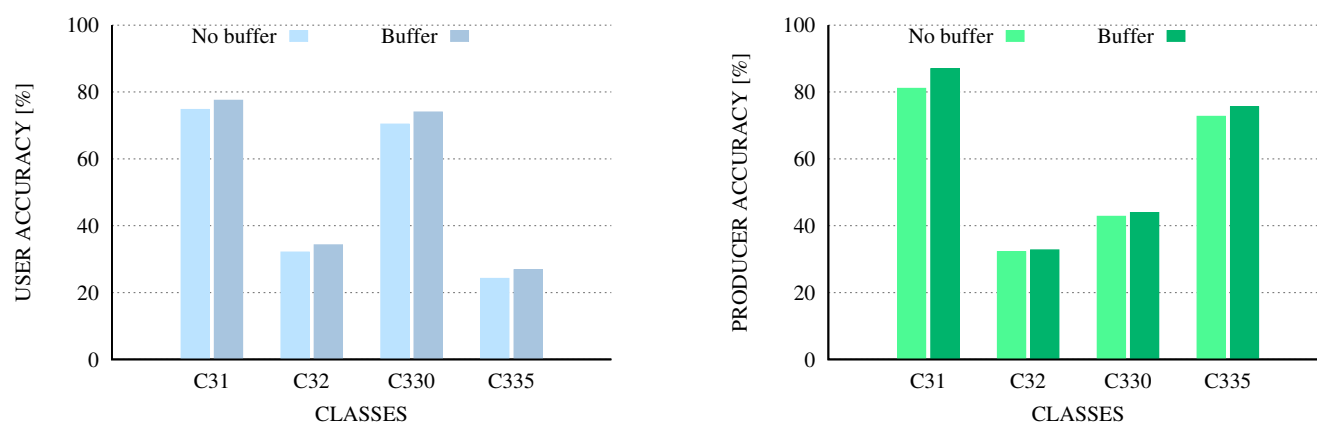| Case Study	Resolution	OA (%)	AD (%)	QD (%)	Samples (-)
NO BUFFER	30 m	77.2 ± 0.015	18.0	4.9	26,520,624
	30 m (prevalence)	77.3 ± 0.015	17.8	4.9	26,547,419
	5 m	77.1 ± 0.003	18.0	4.9	954,739,157
BUFFER	30 m	82.0 ± 0.015	12.8	5.2	21,531,919
	30 m (prevalence)	82.2 ± 0.015	12.7	5.1	21,551,669
	5 m	82.4 ± 0.003	12.4	5.2	753,490,062

Figure 6. Comparison between GlobeLand30 2000 and DUSAF1.1 (II reclassification): per-class statistics referred to 30 m resolution input datasets. C31: forests; C32: herbaceous or/and grass vegetation associations; C330: open spaces with little or no vegetation; C335: glaciers and perpetual snow.

Among the newly derived classes (Figure 6), *forests* are classified with the best accuracies (75% for UA and 81% for PA) while the *herbaceous and/or grass vegetation associations* is the class with the lowest accuracies, below 35%. The *open spaces with little or no vegetation* and the *glaciers and perpetual snow* classes present an opposite behavior: the former registers a high user accuracy (70%) with respect to the producer one (43%), the latter is characterized by a UA equal to 24% and a PA of

73%. The same trends observed with regard to the *no buffer* case statistics are valid also for the *buffer* case ones. Table 7 shows the calculated confidence intervals at 95% probability for UA and PA indexes.

Table 7. Comparison between GlobeLand30 2000 and DUSAF1.1 (II reclassification): per-classes confidence intervals at 95% probability calculated for user (UA) and producer accuracies (PA). Data refer to 30 m resolution input datasets.

Index	C31		C32		C330		C335	
	NO BUFFER	BUFFER	NO BUFFER	BUFFER	NO BUFFER	BUFFER	NO BUFFER	BUFFER
UA	±0.032	±0.034	±0.071	±0.084	±0.083	±0.092	±0.136	±0.170
PA	±0.027	±0.026	±0.061	±0.069	±0.046	±0.052	±0.235	±0.265

3.1.2. Comparison between GlobeLand30 2010 and DUSAF4.0

The same comparison analyses performed between GlobeLand30 2000 and DUSAF 1.1 have also been carried out on the datasets referred to year 2010 (GlobeLand30 2010 and DUSAF4.0). Table 8 reports the overall statistics related to the first classification method and calculated for different resolutions and for the two cases of *no buffer* and *buffer*. Data show a behavior similar to that observed in the previous datasets comparison: considering the *no buffer* case, OA value amount to 86% and input resolution variations can be considered negligible (no significant changes in the results were highlighted). Furthermore, the removal of co-location tolerance buffer cells entails an increase in OA of about 4%; this is mostly due to a reduction in the AD value, from 9.8% to 6.3%–6.6%.

Table 8. Comparison between GlobeLand30 2010 and DUSAF4.0 (I classification): overall statistics and number of samples for the two cases of *no buffer* and *buffer* and for different resolutions. OA: overall accuracy (confidence intervals at 95% probability); AD: allocation disagreement; QD: quantity disagreement.

Case Study	Resolution	OA (%)	AD (%)	QD (%)	Samples (-)
NO BUFFER	30 m	86.2 ± 0.013	9.8	4.0	26,520,835
	30 m (prevalence)	86.3 ± 0.013	9.8	3.9	26,547,614
	5 m	86.1 ± 0.002	9.9	4.0	954,746,538
BUFFER	30 m	90.3 ± 0.012	6.6	3.1	22,678,952
	30 m (prevalence)	90.4 ± 0.012	6.5	3.0	22,701,839
	5 m	90.7 ± 0.002	6.3	3.0	798,071,817

Among the CLC classes (Figure 7), *water bodies*, *agricultural areas* and *forests and semi natural areas* are detected with the best accuracies, always greater than 80%. Considering the *no buffer* case, the *water bodies* class reaches the maximum value for UA (95%) while the *agricultural areas* class is characterized by the best PA (about 90%); *wetlands* is the class with the lowest values of accuracies (below 40%). Table 9 presents the calculated confidence intervals at 95% probability for PA and UA indexes.

Focusing on the second reclassification method, statistics (Table 10) highlight an increase of the allocation and quantity disagreement percentages and the consequent decreasing in overall accuracies, which assume values around 80% for the *no buffer* case and 85% for the *buffer* one. Compared to the results obtained for the year 2000 (Table 6), the overall accuracies are up to 3% higher: this is

explained by the significant improvements observed in sub-classes detection, particularly noticeable in the *open spaces with little or no vegetation* and the *glaciers and perpetual snow* classes. As can be seen in Figure 8, the latter represents the best classified land coverage and it is characterized by values of UA and PA, respectively, equal to 89% and 82% (in *no buffer* case); also *forests* and *open spaces with little or no vegetation* have high accuracies, around 80%, while *herbaceous and/or grass vegetation associations* is instead the worst identified class. Table 11 reports the confidence intervals at 95% probability calculated for UA and PA indexes.
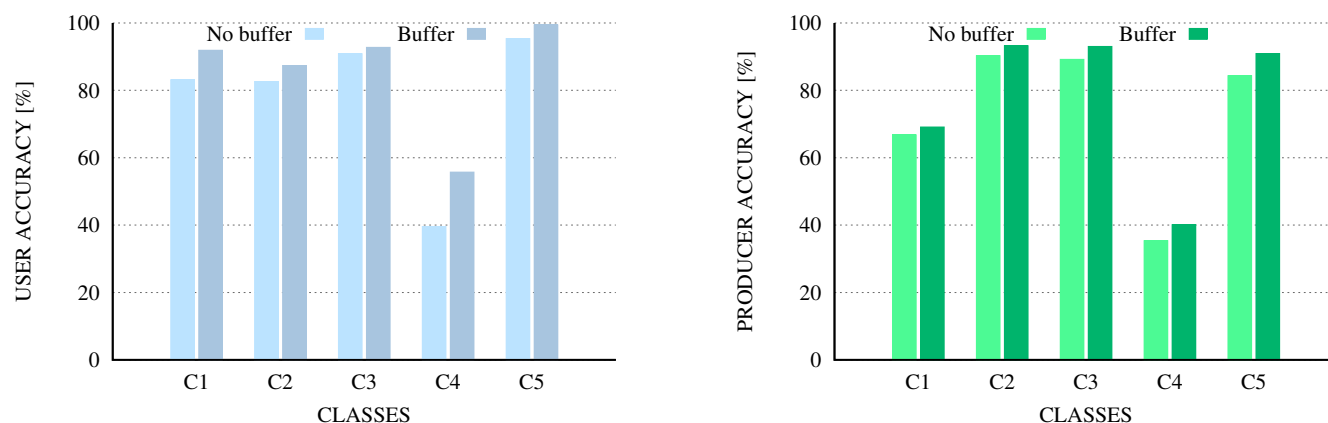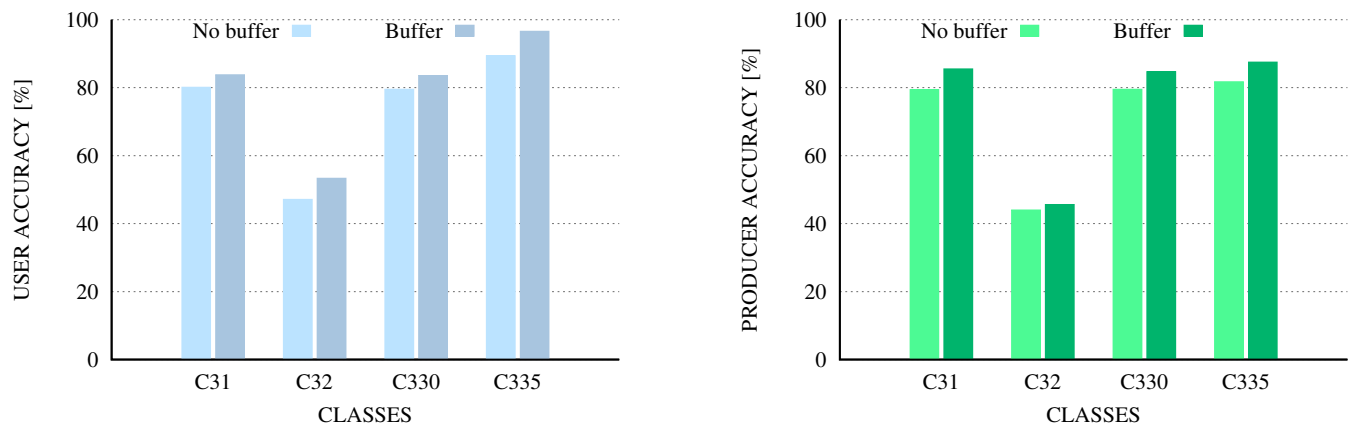

Figure 7. Comparison between GlobeLand30 2010 and DUSAF4.0 (I reclassification): per-class statistics referred to 30 m resolution input datasets. C1: artificial surfaces, C2: agricultural areas, C3: forest and semi natural areas, C4: wetlands, C5: water bodies.

Table 9. Comparison between GlobeLand30 2010 and DUSAF4.0 (I reclassification): per-classes values of confidence interval at 95% probability calculated for user (UA) and producer accuracies (PA); data refer to 30 m resolution input datasets.

Index	C1		C2		C3		C4		C5	
	No Buffer	Buffer	No Buffer	Buffer	No Buffer	Buffer	No Buffer	Buffer	No Buffer	Buffer
UA	±0.045	±0.038	±0.021	±0.020	±0.018	±0.017	±0.524	±0.751	±0.446	±0.019
PA	±0.038	±0.046	±0.016	±0.015	±0.017	±0.015	±0.418	±0.513	±0.069	±0.065

Table 10. Comparison between GlobeLand30 2010 and DUSAF4.0 (II classification): overall statistics and number of samples for the two cases of *no buffer* and *buffer* and for different resolutions. OA: overall accuracy (confidence intervals at 95% probability); AD: allocation disagreement; QD: quantity disagreement.

Case Study	Resolution	OA (%)	AD (%)	QD (%)	Samples (-)
NO BUFFER	30 m	79.8 ± 0.015	16.1	4.0	26,520,835
	30 m (prevalence)	80.0 ± 0.015	16.0	3.9	26,547,614
	5 m	79.8 ± 0.002	16.2	4.0	954,746,538
BUFFER	30 m	85.0 ± 0.015	10.7	4.2	21,492,287
	30 m (prevalence)	85.2 ± 0.015	10.7	4.2	21,513,333
	5 m	85.6 ± 0.002	10.2	4.2	749,419,192

Figure 8. Comparison between GlobeLand30 2010 and DUSAF4.0 (II reclassification): per-classes statistics referred to 30 m resolution input datasets. C31: forests; C32: herbaceous or/and grass vegetation associations; C330: open spaces with little or no vegetation; C335: glaciers and perpetual snow.

Table 11. Comparison between GlobeLand30 2010 and DUSAF4.0 (II reclassification): per-classes confidence intervals at 95% probability calculated for user (UA) and producer (PA) accuracies. Data refer to 30 m resolution input datasets.

Index	C31		C32		C330		C335	
	No Buffer	**Buffer**	**No Buffer**	**Buffer**	**No Buffer**	**Buffer**	**No Buffer**	**Buffer**
UA	±0.031	±0.032	±0.074	±0.089	±0.057	±0.057	±0.210	±0.148
PA	±0.027	±0.0003	±0.059	±0.001	±0.050	±0.0005	±0.229	±0.02

3.2. Italian Area

The benchmarking methodology proposed for the Lombardy Region has been adopted for all the available Italian datasets; in the following a summary of the results is proposed for both years 2000 and 2010. All the statistics refer to the 30 m resolution case study. In all the cases considered, the confidence intervals at 95% probability are, at most, equal to 1%, so we decided not to report them.

3.2.1. Comparison between GlobeLand30 2000 and Italian Datasets

Figure 9 shows, for each regional dataset, the values of overall accuracy, allocation and quantity disagreement obtained through the first reclassification method in *no buffer* case. The results highlight that GlobeLand30 presents an OA that ranges between 81% (Liguria) and 92% (Autonomous Province of Bolzano). In most cases, the percentage of disagreement is almost equally divided between AD and QD; the only exceptions are represented by the Liguria and Abruzzo case studies that, as observed also for the Lombardy one, present values of AD significantly higher than QD. Generally, 3%–4% overall accuracy can be gained if the co-location tolerance influence is taken into consideration (*buffer* case).

More in detail, by analyzing user and producer accuracy values (calculated for each dataset and for each Corine class, as shown in Figures 10 and 11) it is possible to identify *forests and semi natural areas* as the best detected class: with regard to this class, the trend is homogeneous for all datasets, with UA and PA values generally higher than 80%. Other classes are characterized by a

variable behavior among the different regions, which is particularly marked with regard to *water bodies* and *wetlands*.
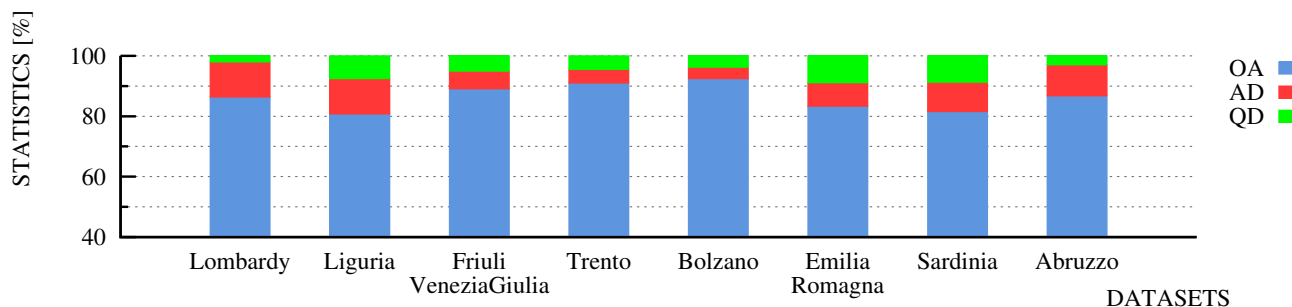

Figure 9. Comparison between GlobeLand30 2000 and Italian regional datasets (I reclassification, no buffer case): overall accuracy (OA), allocation disagreement (AD), quantity disagreement (QD) for each dataset.
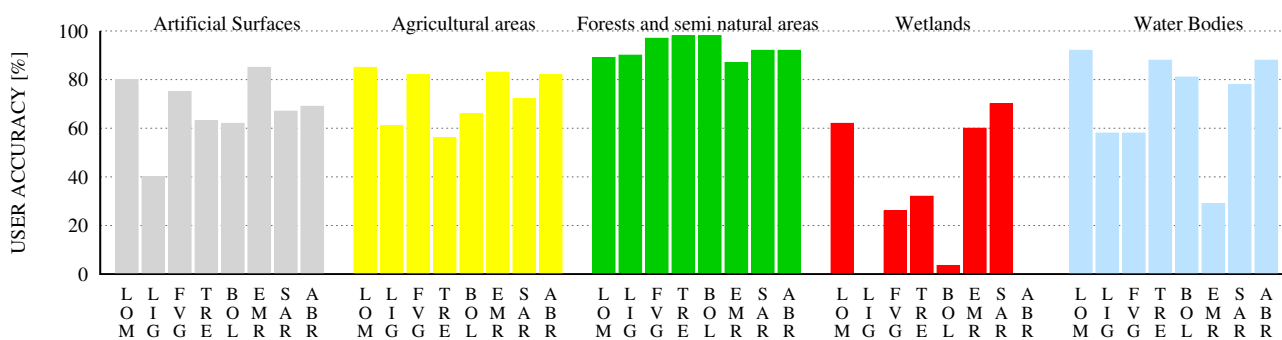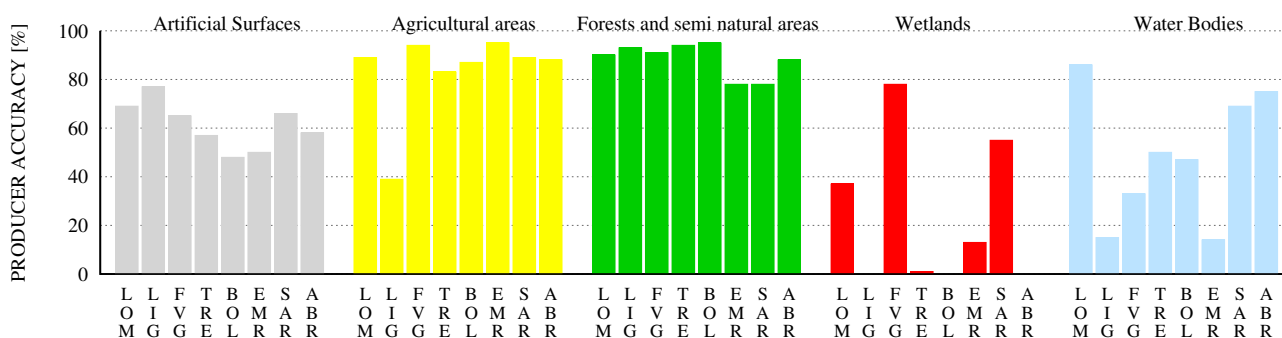

Figure 10. Comparison between GlobeLand30 2000 and Italian regional datasets (I reclassification, *no buffer* case): user accuracies for each class and dataset.

Figure 11. Comparison between GlobeLand30 2000 and Italian regional datasets (I reclassification, *no buffer* case): producer accuracies for each class and dataset.

Considering the second classification method, results presented in Figure 12 show for all the datasets a general lowering in accuracies values, variables between 67% (Sardinia) and 81% (Emilia-Romagna); this is mostly due to an increase of the allocation disagreement, which is particularly high (about 13%–14%) for the Sardinia Region and the Autonomous Provinces of Trento and Bolzano. The analysis of the results shows that among the sub-classes *forests* is classified with the best accuracies, generally over 70% in most of the Italian regions.

Finally, the comparisons performed in the *buffer case* showed an increase in overall accuracy of about 4%–5%.
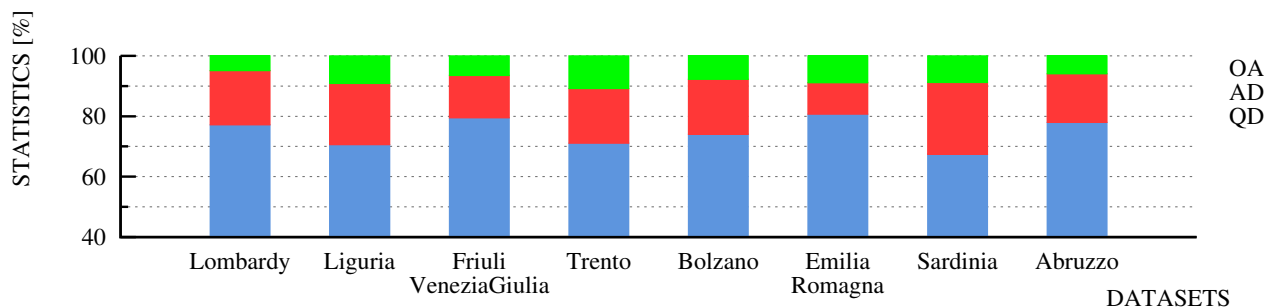

Figure 12. Comparison between GlobeLand30 2000 and Italian regional datasets (II reclassification, *no buffer* case): overall accuracy (OA), allocation disagreement (AD), quantity disagreement (QD) for each regional dataset.

3.2.2. Comparison between GlobeLand30 2010 and Italian Datasets

The overall statistics obtained from the comparison according to the first classification method and the *no buffer* case between GlobeLand30 and the five available Italian datasets for the year 2010 are reported in Figure 13 and show that the overall accuracy values vary between 81% (Sardinia) and 86% (Lombardy); an increase of about 3%–4% could be achieved through the application of the buffer method, which allows us to considerably reduce the allocation disagreement. As observed for the year 2000, *forests and semi natural areas* is the class identified with the best accuracies for all the considered datasets, while the other classes are characterized by variable behavior (Figures 14 and 15).

Finally, the bar chart proposed in Figure 16 shows the overall statistics resulting by applying the second classification method and it highlights, once again, a significant increase of disagreement values that entail a decrease in OA percentages: new values vary between 62% (Sardinia) and 80% (Lombardy and Emilia-Romagna). The application of the *buffer* case leads to a general increase of OA of about 3%–5%.
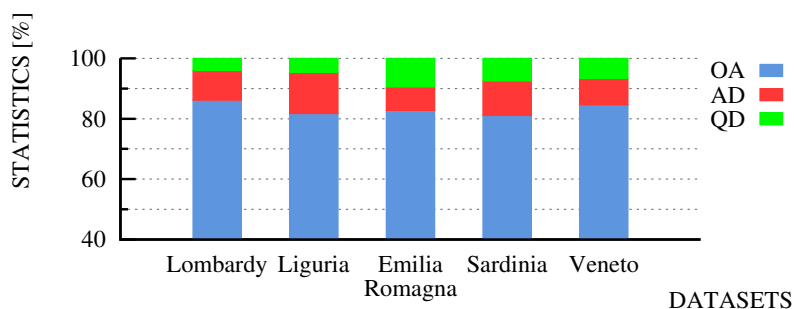

Figure 13. Comparison between GlobeLand30 2010 and Italian regional datasets (I reclassification, *no buffer* case): overall accuracy (OA), allocation disagreement (AD), quantity disagreement (QD) for each dataset.
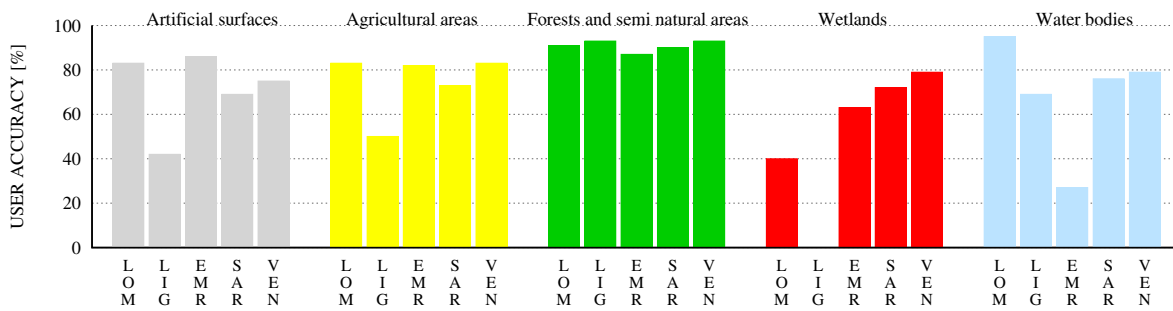
Figure 14. Comparison between GlobeLand30 2010 and Italian regional datasets (I reclassification, *no buffer* case): user accuracies for each class and dataset.
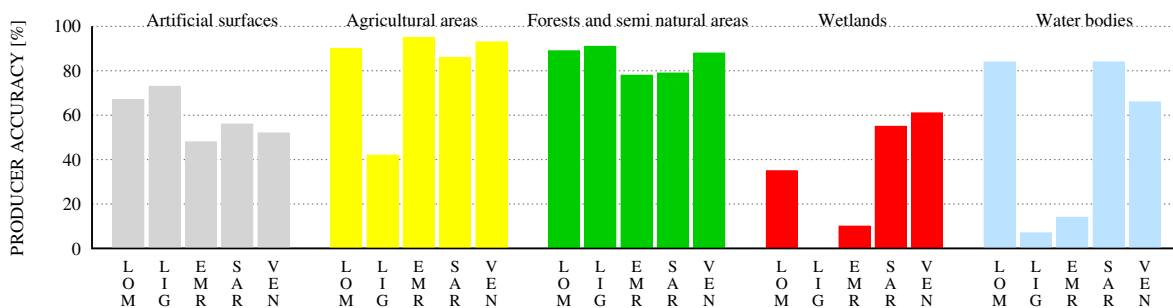

Figure 15. Comparison between GlobeLand30 2010 and Italian regional datasets (I reclassification, *no buffer* case): producer accuracies for each class and dataset.

With respect to the sub-classes, *forests* is the best detected class with UA and PA values higher than 75% in most cases.
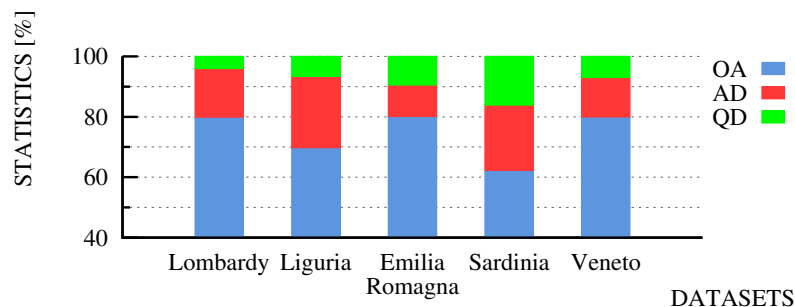

Figure 16. Comparison between GlobeLand30 2010 and Italian regional datasets (II reclassification, *no buffer* case): overall accuracy (OA), allocation disagreement (AD), quantity disagreement (QD) for each regional dataset.

4. Conclusions

Depending on the different reclassification scheme considered, the assessment of the thematic accuracy performed between GlobeLand30 and eight different Italian regional land coverage maps, does not show negligible differences in the overall accuracy values. In fact, the first reclassification method, based on the five Corine classes, allows overall accuracy to be obtained between 81% and 92% (both for 2000 and 2010). While the second classification approach, based on Corine subclasses,

leads to overall accuracies that vary between 62% and 81%. As expected, the subdivision of a broad class in more detailed subclasses determines greater difficulties in the detection process.

Whatever the level of classification is, one part of the disagreement observed between GlobeLand30 and regional datasets can likely be attributed to the fact that, in some cases, the images are taken in different time periods (see Table 1); another part can be due to the different thematic classification and resolution used by the original developers. Moreover, the co-location tolerance can also play an important role: in fact, if we eliminate the amount of the disagreement due to this factor by introducing a buffer around the class polygon borders, the overall accuracy for both the reclassification methods increases (84% to 96 % in the former case, 65% to 86% in the latter one).

Another important point of discussion is related to the reference data. Although used in the benchmarking as an accurate representation of reality, they are in fact another type of classification that may also contain errors. These errors are by default included in the comparison analysis and they can cause bias to the accuracy assessment [28,29]. Even though the reference data accuracy is adequate for assessing the Globeland30 quality, a further analysis on the quality of the reference data itself could be taken into account in future research. Traditional monitoring done *in situ* by experts and new approaches based on the citizen science paradigm [30] will be applied. Moreover, some other theoretical and practical items have to be considered. First of all we want to make some analysis about co-variance and correlation of errors associated with the different classes. Secondly, we want to complete the analysis considering the detailed Italian land cover maps for the remaining regions.

More in general, considering that different organizations have more or less independently been producing land cover maps, we can summarize the following problems:

1. The classification methods used are heterogeneous;
2. The number and type of classes is not the same; and
3. Some land coverage areas have been misclassified due to various problems related to image acquisition (period of the year, geometric and radiometric correction, *etc.*) or processing.

With respect to these problems, the implementation of an intelligent and open global land coverage geo-platform will allow scientists to share data and citizen to participate for the improvement of the land cover classification quality. This can be obtained by providing web-based services to share and compare land cover data sets, to evaluate the coherency, to highlight the differences between the uploaded land cover maps and to solve these differences, exploiting geo-visualization and geo-crowdsourcing by means of mobile platforms. VIEW-IT (Virtual Interpretation of Earth Web-Interface Tool) [31] and Geo-Wiki Project [32] are examples of such kind of platforms.

Acknowledgments

Author Contributions

Conflicts of Interest

References

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

2 Sellers, P.J.; Tucker, C.J.; Collatz, G.J.; Los, S.O; Justice, C.O; Dazlich, D.A.; Randall, D.A. A global 1-degree-by-1-degree NDVI data set for climate studies. Part 2: The generation of global fields of terrestrial biophysical parameters from the NDVI. *Int. J. Remote Sens.* **1994**, *15*, 3519–3545.

3 Stehman, S.V.; Czaplewski, R.L. Design and analysis for thematic map accuracy assessment: Fundamental principles. *Remote Sens. Environ.* **1998**, *64*, 331–344.

4 Strahler, A.H.; Boschetti, L.; Foody, G.M.; Friedl, M.A.; Hansen, M.C.; Herold, M.; Mayaux, P.; Morisette, J.T; Stehman, S.V.; Woodcock, C.E. Global land cover validation: Recommendations for evaluation and accuracy assessment of global land cover maps. In *Technical Report EUR 22156 EN-DG 2006*; Office for Official Publications of the European Community: Luxembourg, 2006.

5 Foody, G.M. Classification accuracy assessment. *IEEE Geosci. Remote Sens. Soc. Newsl.* **2011**, *159*, 8–14.

6 Olofsson, P.; Foody, G.M.; Herold, M.; Stehman, S.V.; Woodcock, C.E.; Wulder, M.A. Good practices for estimating area and assessing accuracy of land change. *Remote Sens. Environ.* **2014**, *148*, 42–57.

7 Loveland, T.R.; Reed, B.C.; Brown, J.F.; Ohlen, D.O.; Zhu, Z.; Yang, L.; Merchant, J.W. Development of a global land cover characteristics database and IGBP DISCover from 1 km AVHRR data. *Int. J. Remote Sens.* **2000**, *21*, 1303–1330.

8 Hansen, M.C.; Defries, R.S.; Townshend, J.R.G.; Sohlberg, R. Global land cover classification at 1 km spatial resolution using a classification tree approach. *Int. J. Remote Sens.* **2000**, *21*, 1331–1364.

9 Fritz, S.; Bartholome, E.; Belward, A.; Hartley, A.; Stibig, H.J.; Eva, H.; Mayaux, P; Bartalev, S.; Latifovic, R.; Kolmert, S.; *et al*. *Harmonization, Mosaicking, and Production of the Global Land Cover 2000 Database*; JRC Scientific and Technical Research Reports 2003. Available online: http://publications.jrc.ec.europa.eu/repository/handle/JRC26168 (accessed on 5 February 2015).

10 Friedl, M.A.; McIver, D.K.; Hodges, J.C.F.; Zhang, X.Y.; Muchoney, D.; Strahler, A.H.; Woodcock, C.E.; Gopal, S.; Schneider, A.; Cooper, A.; *et al*. Global land cover mapping from MODIS: Algorithms and early results. *Remote Sens. Environ.* **2002**, *83*, 287–302.

11 Arino, O.; Bicheron P.; Achard, F.; Latham, J.; Witt, R.; Weber, J.L. GLOBCOVER the Most Detailed Portrait of Earth. *ESA Bull. Eur. Space Agency* **2008**, *136*, 24–31.

12 Chen, J.; Chen, J.; Liao, A.; Cao, X.; Chen, L.; Chen, X.; He, C.; Han, G.; Peng, S.; Lu, M; *et al*. Global land cover mapping at 30 m resolution: A POK-based operational approach. *ISPRS J. Photogram. Remote Sens.* **2014**, doi:10.1016/j.isprsjprs.2014.09.002.

13 Cao, X.; Chen, J.; Chen, L.J.; Liao, A.N.; Sun, F.D.; Li, Y.; Li, L.; Lin, Z.H.; Pang, Z.G.; Chen, J.; *et al*. Preliminary analysis of spatiotemporal pattern of global land surface water. *Sci. China Earth Sci.* **2014**, *57*, 2330–2339.

14 Manakos, I.; Chatzopoulos-Vouzoglanis, K.; Petrou, Z.I.; Filchev, L.; Apostolakis, A. Globalland30 mapping capacity of land surface water in Thessaly, Greece. *Land* **2015**, *4*, 1–18.

15 Olofsson, P.; Foody, G.M.; Stehman, S.V.; Woodcock, C.E. Making better use of accuracy data in land change studies: Estimating accuracy and area and quantifying uncertainty using stratified estimation. *Remote Sens. Environ.* **2013**, *129*, 122–131.

16 Stehman, S.V.; Wickham, J.D. Pixels, blocks of pixels, and polygons: Choosing a spatial unit for thematic accuracy assessment. *Remote Sens. Environ.* **2011**, *115*, 3044–3055.

17 Stehman, S.V. Sampling design for accuracy assessment of land cover. *Int. J. Remote Sens.* **2009**, *30*, 5243–5272.

18 Congalton, R.G.; Green, K. *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices*; Lewis Publishers: Boca Raton, FL, USA, 1999; p. 137.

19 Liu, C.; Frazier, P.; Kumar, L. Comparative assessment of the measures of thematic classification accuracy. *Remote Sens. Environ.* **2007**, *107*, 606–616.

20 Anderson, J.R.; Hardy, E.E; Roach, J.T.; Witmer, R.E. *A land Use and Land Cover Classification System for Use with Remote Sensor Data*; Geological Survey Professional Paper 964; U.S. Geological Survey: Reston, VA, USA, 1976.

21 Pringle, M.J.; Schmidt, M.; Muir, J.S. Geostatistical interpolation of SLC-off Landsat ETM+ images. *ISPRS J. Photogramm. Remote Sens.* **2009**, *64*, 654–664.

22 Thomlinson, J.R.; Bolstad, P.V.; Cohen, W.B. Coordinating methodologies for scaling land cover classifications from site-specific to global: Steps toward validating global map products. *Remote Sens. Environ.* **1999**, *70*, 16–28.

23 Pontius, R.G.; Millones, M. Death to Kappa: Birth of quantity disagreement and allocation disagreement for accuracy assessment. *Int. J. Remote Sens.* **2011**, *32*, 4407–4429.

24 Biagi, L.; Negretti, M. A New Approach to the Polygons Rasterization in GRASS. Available online: http://geomatica.como.polimi.it/workbooks/n2/articoli/lbmn.pdf (accessed on 4 February 2015).

25 Gallego, J. *Comparing CORINE Land Cover with a more Detailed Database in Arezzo (Italy). Towards Agri-Environmental Indicators*; Topic report 6/2001 European Environment Agency 2001; European Environment Agency: Copenhagen, Danmark, 2001; pp. 118–125.

26 Credali, M.; Fasolini, D.; Minnella, L.; Pedrazzini, L.; Peggion, M.; Pezzoli, S. Tools for territorial knowledge and government. In *Land Cover Changes in Lombardy over the Last 50 Years*; Fasolini, D., Pezzoli, S., Sale, V.M., Cesca, M., Coffani, S., Brenna S., Eds.; ERSAF-Lombardy Region: Milan, Italy, 2011; pp. 17–19.

27 Geoportale della Lombardia. Available online: http://www.cartografia.regione.lombardia.it/geoportale (accessed on 3 October 2014).

28 Foody, G.M. Assessing the accuracy of land cover change with imperfect ground reference data. *Remote Sens. Environ.* **2010**, *114*, 2271–2285.

29 Foody, G.M. The impact of imperfect ground reference data on the accuracy of land cover change estimation. *Int. J. Remote Sens.* **2009**, *30*, 3275–3281.

30 Comber, A.; See, L.; Fritz, S.; van der Velde, M.; Perger, C.; Foody, G.M. Using control data to determine the reliability of volunteered geographic information about land cover. *Int. J. Appl. Earth Obs. Geoinf.* **2013**, *23*, 37–48.

31 Clark, M.L.; Aide, M. Virtual Interpretation of Earth Web-Interface Tool (VIEW-IT) for collecting land-use/land-cover reference data. *Remote Sens.* **2011**, *3*, 601–620.

32 Fritz, S.; McCallum, I.; Schill, C.; Perger, C.; Grillmayer, R.; Achard, F.; Kraxner, F.; Obersteiner, M. Geo-Wiki.Org: The use of crowdsourcing to improve global land cover. *Remote Sens.* **2009**, *1*, 345–354.