

Article

Use of LUCAS LC Point Database for Validating Country-Scale Land Cover Maps

Christos G. Karydas ^{1,*}, Ioannis Z. Gitas ¹, Steffen Kuntz ² and Chara Minakou ¹

¹ Lab of Forest Management and Remote Sensing, School of Agriculture, Forestry and Natural Environment, Aristotle University of Thessaloniki, 54124 Thessaloniki, Greece; E-Mails: igitas@for.auth.gr (I.Z.G.); minakou@for.auth.gr (C.M.)

² Airbus Defence & Space, DE-88090 Immenstaad, Germany; E-Mail: Steffen.Kuntz@astrium.eads.net

* Author to whom correspondence should be addressed; E-Mail: xkarydas@auth.gr; Tel.: +30-2310-992-689; Fax: +30-2310-992-699.

Academic Editors: Parth Sarathi Roy and Prasad S. Thenkabail

Received: 16 January 2015 / Accepted: 13 April 2015 / Published: 23 April 2015

Abstract: In this study, the Land Use/Cover Area frame statistical Survey (LUCAS) of 2009 was used as a reference dataset for validating a Land Cover Map of Greece for 2007, produced with remote sensing by the Greek Office of the World Wildlife Fund (WWF Hellas). First, all class definitions were decomposed in terms of four vegetation parameters (type, height, density, and composition), considered as critical in indicating unconformities between LUCAS and the WWF Hellas map; their inter-class relations were described in a table of correspondence. Then, a two-tier methodology was applied: an “automated” process, where thematic agreement was based exclusively on the main land cover attribute of LUCAS (LC1); and a “supervised” process, where thematic agreement was based on the reinterpretation of LUCAS ground photos and use of ancillary earth observation imagery; non-square error matrix was deployed in both processes. For the supervised process specifically, a decision-tree was designed, using the critical vegetation parameters (mentioned above) as quantified criteria, thus allowing objective labelling of testing points in both systems. The results show that only a small proportion of the reassessed points verified the WWF Hellas map predictions and that the overall accuracy of the supervised process was reduced compared to that of the automated process. In conclusion, the LUCAS point database was found to be supportive, but not fully efficient, for identifying the various sources of error in country-scale land cover maps derived with remote sensing. Synergy with very high resolution satellite images and air photos, or a

dedicated ground truth campaign, seems to be inevitable in order to validate their thematic accuracy, especially in highly heterogeneous environments. In this direction, LUCAS could be used as a verification, rather than a validation, dataset.

Keywords: WWF Hellas; ground truth; non-square error matrix; supervised validation; reinterpretation

1. Introduction

Validation of land use/cover (LU/LC) maps derived from remote sensing is a challenging and expensive task. Especially when large areas have to be mapped, the availability of recent and accurate reference data for an independent and sound validation becomes critical.

In early mapping studies, accuracy assessment was frequently an afterthought. Through time, statistical observation and use of confusion (or error) matrices, have ultimately become the key tool for accuracy assessment of classified images or land cover maps in general [1,2].

In the framework of GMES/Copernicus initiative, strong effort was made to assure high quality of European-wide mapping and monitoring approaches such as the Copernicus Land Monitoring Core Service. Pre-cursor services such as the Fast Track Services (FTS) already set the frame for validation campaigns on European level (see for instance, [3]).

In parallel to the Copernicus program, which focuses on remote sensing methodologies, every three years since 2000, the Land Use/Cover Area frame statistical Survey (LUCAS) point database provides detailed attributes for land cover and land use in large parts of Europe, accompanied by a set of ground photographs for every record. The dataset is based on a sampling approach assessing land cover/use conditions *in situ* on a relatively high-density grid.

Using separate classification systems for land cover and land use, the primary objective of LUCAS is the extraction of comparable statistical information in all participating countries over time. LUCAS is compatible with other existing land cover/use systems (e.g., FAO, NACE and Farm Structure Survey) and fulfills the specifications of the European INSPIRE standardization initiative on land cover/use. Data quality is assured regularly by a two-level quality control, in which all points are evaluated by external controllers [4].

Land cover in LUCAS is broken down in eight main categories, which are indicated by capital letters: A: Artificial land; B: Cropland; C: Woodland; D: Shrubland; E: Grassland; F: Bare land; G: Water areas; H: Wetlands. Every main category contains subclasses, which are indicated by the combination of the letter of the category and up to three digits. Altogether there exist 76 classes [5]. For validation purposes of remote sensing based thematic mapping, the 2001–2002 database of LUCAS was already used by the European Environmental Agency (EEA) for checking the quality of CORINE Land Cover (CLC) 2000 per country [6]. Furthermore, LUCAS 2003 database was combined with CLC 2000 for areal estimates in coastline zones throughout Europe [7,8]. (On a proposal from the Commission in 27 June 1985, the European Council adopted a decision on the CORINE program (Coordination of information on the environment) to compile information on the state of the environment; to coordinate the compilation of data and the organization of information within the Member States or at international level; and to ensure

that information is consistent and that data are compatible. CORINE Land Cover comprises 44 thematic classes with a minimum mapping unit (MMU) of 25 ha for stock, and 5 ha for changes, respectively. (for more details see <http://www.eea.europa.eu/publications/COR0-landcover>)

In the framework of geoland2, the main aim of this research was to investigate the extent to which point databases of land use/cover from statistical sampling can be deployed as reference datasets for validating country-scale land cover maps generated from remote sensing. In this view, the LUCAS point database of 2009 was used as a reference dataset for validating a land cover map of Greece from 2007, produced with remote sensing by the Greek Office of the World Wildlife Fund (WWF Hellas) in the framework of “The Future of the Greek Forest” project, with a view to monitor land cover changes in Greece between years 1987 and 2007. (Geoland2 was an EU funded 7th Framework Programme (FP7) Research Project which was responsible for the development and pre-operational validation of the GMES Land Monitoring Core Service. The project started in September 2008 and finished in 2012. It comprised 51 project partners involving more than 80 major national and international user organizations.)

The basic hypothesis of the study was that the LUCAS point database could be used for revealing sources of error and identifying weaknesses of land cover maps produced with satellite image classification on a country scale.

2. Study Area and Dataset

2.1. The Study Area

Greece is located in SE Europe, at the southern edge of the Balkan Peninsula. It covers a total area of 131,957 km², of which land area accounts for 130,846 km² and internal waters (lakes and rivers) account for 1144 km² (0.87%). Land boundaries measure 1160 km and coastline 13,676 km. The country is one of the most mountainous countries in Europe, with 80% of the territory covered by mountains (Mt. Olympus is the highest peak, with an altitude of 2918 m).

The climate of Greece is predominantly Mediterranean, with some parts of the country with alpine or temperate climate. Characterized by dry summers and rainy winters, Greece is mainly covered by Mediterranean ecosystems, such as forests, woodlands, and scrub biomes. Mediterranean forests are generally composed of broadleaved trees, such as the oak and mixed sclerophyllous forests.

According to the CORINE LC 2000 database, Greece is composed of the following land covers at the first level of the classification hierarchy: Artificial surfaces: 2.1%; Agricultural areas: 39.9%; Forests and semi-natural areas: 56.5%; Wetlands: 1.0%; and Water bodies: 0.5% (overall accuracy estimated at 80%; source: EEA).

2.2. The Land Cover Map

For the needs of WWF Hellas, LC maps of Greece for the years 1987 and 2007 were produced from a set of 54 Landsat TM images (2 years × 27 scenes, covering the entire country). Landsat TM imagery is regarded worldwide as appropriate for land cover mapping on a country scale. Therefore, many organizations have used Landsat databases for their purposes, indicatively, the Natural Environment Research Council (NERC) of the United Kingdom [9], the Joint Research Centre (JRC) of the European

Union [10] and the National Land Cover Database (NLCD) of the United States of America [11]. There also exist examples of land cover mapping using imagery with coarser spatial resolution, such as MERIS (300 m) and MODIS (500 m) in Siberia [12] and Portugal [13].

Object-based image analysis (OBIA) was performed with the Landsat scenes using eCognition Developer 8.0 ©. The thematic outputs were mosaicked to produce two seamless LC maps. These maps are available online: <http://oikoskopio.gr/map/>. Focusing on monitoring natural vegetation categories indigenous in Greece, WWF Hellas employed the following classification scheme: “Bare land and artificial surfaces” (1), “Agricultural land” (2), “Broadleaved forests” (3) “Coniferous forests” (4), “Sclerophyllous vegetation” (5), “Shrublands” (6), “Sparsely vegetated areas” (7), “Burnt areas” (8), and “Water bodies” (9). The map of 2007 was validated by independent surveyors hired by the producer; a set of 1376 check points was selected throughout the country during the same year, resulting in an overall accuracy of 78.4% and a Cohen’s kappa coefficient value of 0.742 [14].

2.3. The Validation Dataset

In the LUCAS 2009 survey for Greece (the first to be conducted for the country), the sampling points were selected from the mainland and the islands of Crete and Evia, the two major islands of Greece. The remaining islands (covering an area of 11,429 km²) together with the high peaks of the country were excluded from the sampling for cost reasons (Figure 1). According to the LUCAS standards, the sampling points are recorded on a regular grid of about 2.8 km × 2.8 km. The LUCAS protocol describes sampled points as circles of either 3 or 20 m radius, *i.e.*, covering a surface of about 7 m² or 1257 m², for homogeneous or heterogeneous categories, respectively [7,8].

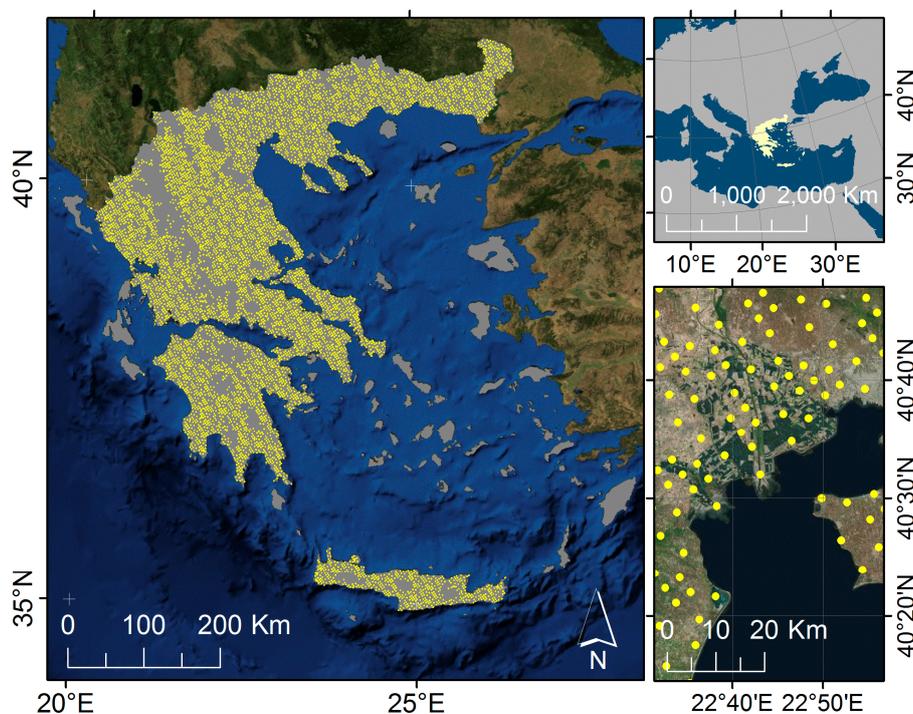


Figure 1. Located at the southern edge of the Balkan Peninsula, SE Europe (top right inset), Greece was covered by 73% by the LUCAS 2009 Survey (yellow dots). In the (bottom right inset), a closer view of LUCAS points is shown (approx. regular grid of 2.8 km).

The designated number of LUCAS 2009 points for Greece was 7789, from which a subset of 6899 points were visited by the surveyors, thus 88.5% of the total points; the rest of the points were not visited due to accessibility reasons. Finally, about 73% of the country was visited and recorded. The mean elevation in the visited areas was 402 m with a mean slope of 19.3%, whereas in the excluded areas, the mean elevation was 684 m and the mean slope 28.7%; the mean elevation of Greece is 478 m with 21.9% mean slope. In the LUCAS 2012 survey, the number of visited points was raised to 7626, as more islands were included, thus resulting in an overall coverage of 80%.

Every point in the LUCAS database contains by default two pairs of coordinates: one of the designated point (in the ETRS89-LAEA projection system); and one of the location from where the surveyor observed that point (in geographic decimal degrees) [15]. In Greece, the distance between locations of observation (as recorded by the surveyors' GPS) from the designated points, was found to vary between 0 and 7310 m. The "Artificial areas", "Cropland", and "Grasslands" were found to be observed from a mean distance less than 100 m, which is a critical threshold according to the LUCAS instructions [15]. "Woodland", "Shrubland", "Bare land", "Water areas", and "Wetlands", on the other hand, were observed from mean distances longer than 100 m; especially the two latter categories were observed from mean distances longer than 500 m. In general, 28.8% of the visited LUCAS 2009 points in Greece were observed from a distance longer than 100 m (Figure 2).

As it is known from the LUCAS standards, for points corresponding to artificial surfaces or arable land, the window of observation is that of 3-m radius, whereas an extended window of observation (20-m radius) is used for the rest of the classes, *i.e.*, natural and heterogeneous agricultural land covers. In Greece, 2615 points were observed using the narrow window of observation, while the rest of the points (5174) were observed using the extended window of observation.

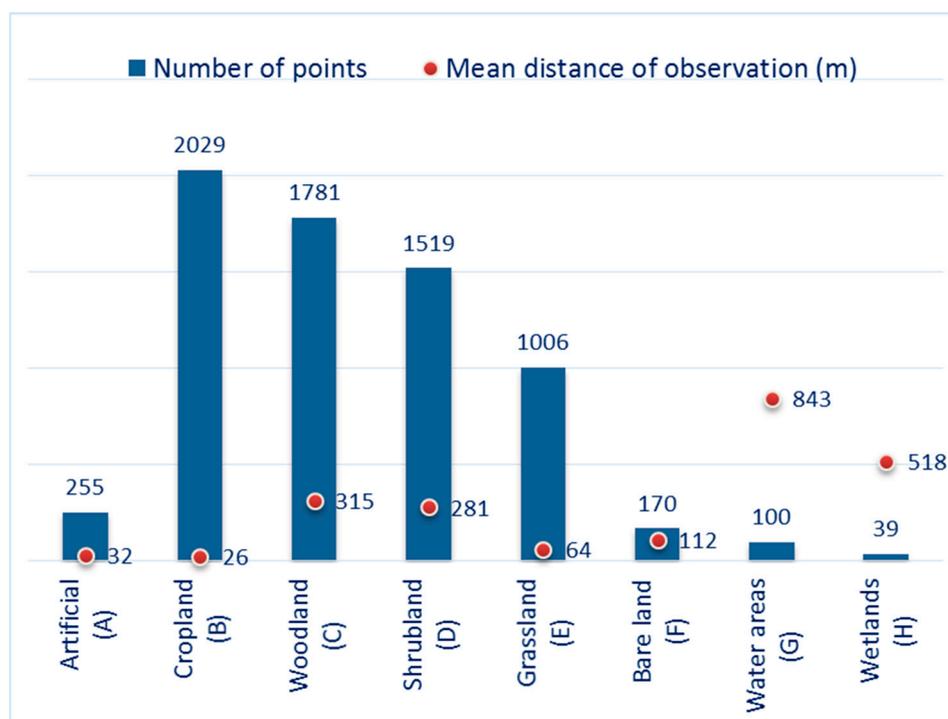


Figure 2. Distribution of the number of visited LUCAS 2009 points in Greece and their mean observation distance per major category.

3. Methodology

3.1. Overview

Several methodologies can be implemented for validating LU/LC maps derived with remote sensing [16]. In the current study, the error matrix method was applied, in order to indicate the thematic quality of the land cover map from point-based evaluations [17]. According to [18], the error matrix method puts the basis for a series of descriptive and analytical statistical techniques.

The two datasets (classified and reference) were co-registered, by adjusting the WWF Hellas polygon layer, which carries a higher locational error (as is derived from imagery), to the LUCAS point layer geometry and then they were intersected.

A non-square error matrix was employed, *i.e.*, one with different classification schemes for classified and reference data [19,20]. The thematic agreement between the two classification schemes was depicted in a table of correspondence through an adjustment of the WWF Hellas scheme to the LUCAS nomenclature.

Two validation processes were applied, independently from one another (Figure 3):

- In the first process, reference information was taken exclusively from the LUCAS main land cover attribute, namely LC1. This process was called “automated” and those WWF Hellas classes that were not identical to a specific LUCAS category were excluded (5/9 classes were assessed).
- In the second process, reference information was extracted not only from LC1, but also from the ground photos taken by the LUCAS surveyors and by interpreting available earth observation imagery, either satellite or air photographs. This process was called “supervised” and demanded a different (more extensive) error matrix (9/9 classes were assessed).

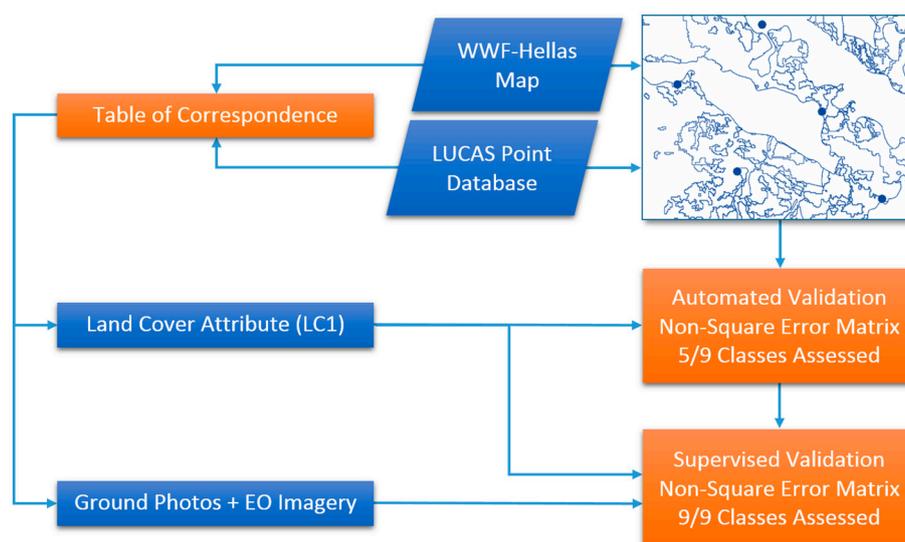


Figure 3. Depiction of the designed validation processes (automated and supervised) for the accuracy assessment of the WWF Hellas map of 2007 using the LUCAS 2009 points as main reference dataset.

The use of additional data in the supervised validation allowed for a better understanding of the critical differences between the classification scheme employed by the WWF Hellas map and the LUCAS

nomenclature. The overall approach follows the recommendations of [20], who states that “the matrix may reveal interclass confusion that could be resolved with the use of additional discriminatory information”. A similar two-tier approach (*i.e.*, automated comparison and reinterpretation) was followed by the European Environmental Agency (EEA) within the framework of CLC2000 project [6].

3.2. Data Co-Registration

Precise co-registration of the two layers was a prerequisite before the analysis, because possible mismatching could affect assessment of heterogeneous land cover classes substantially [20].

In order to match geometrically the polygon-layer of the WWF Hellas map with the point layer of LUCAS, the basemap embedded in ArcGIS© software was used as spatial reference, provided that the LUCAS point-layer showed to match perfectly with it. For Greece, this basemap comprises 1-m pixel size GeoEye and IKONOS color imagery for the major part of the country and 15-m Landsat compositions for few areas, where very high resolution imagery is not available.

Checking with the basemap at 100 random points throughout the country, the degree of the WWF Hellas map’s shifting was found to be about 310 m to the west and 225 m to the south on average, close to the typical reported values of positional accuracy of Landsat TM imagery [21].

Before co-registration, both layers were re-projected to the WGS84 system (code: EPSG 4326), which is the original projection of Landsat imagery. Then, the WWF Hellas map was manually adjusted to the basemap, using control points from many characteristic sites (shorelines, highways, rivers, buildings, *etc.*), selected throughout the country. Finally, visual inspection of the adjusted polygons verified a very reliable output.

3.3. Nomenclature Differences

3.3.1. Generic

Validation processes (either automated or supervised) require clear rules for the comparison of classified and reference data. Any uncertainties, which may arise either at the conception, the measurement, or the analysis stage of the process, have to be recognized, evaluated, and managed in advance. A list of the class codes and names for the first level of LUCAS and the single level of the WWF Hellas map provided in Table 1 will help the reader to follow easily the analysis of the classes.

Table 1. A summary of codes and names for the first level of LUCAS and for the single level of the WWF Hellas map.

LUCAS Categories (1st Level)	WWF Hellas Categories (Single Level)
A: Artificial land	1: Bare land and artificial surfaces
B: Cropland	2: Agricultural land
C: Woodland	3: Broadleaved forests
D: Shrubland	4: Coniferous forests
E: Grassland	5: Sclerophyllous vegetation
F: Bare land	6: Shrublands
G: Water areas	7: Sparsely vegetated areas
H: Wetlands	8: Burnt areas
	9: Water bodies

3.3.2. Artificial Surfaces, Agricultural Land, and Bare Land

“Bare land and artificial surfaces” of the WWF Hellas classification scheme contains features corresponding to more than one distinctive categories of LUCAS. More specifically, the features “settlements”, “industrial areas”, or “quarries” are included in “Artificial land” of LUCAS, whereas the feature “rock outcrops” is included in “Bare land”. Therefore, “Bare land and artificial surfaces” of WWF Hellas was considered as a mixture of classes compared to the reference data.

“Agricultural land” of WWF Hellas contains “greenhouses”, which however, are contained in “Artificial land” of LUCAS. Therefore, “Greenhouses” (code A13) had to be excluded from “Artificial land” and to be grouped with the reference set of “Cropland”.

Possible lack of evidence by the LUCAS field surveyors in identifying undoubtedly agricultural land could also raise uncertainty. According to the LUCAS survey protocol, if a crop is not possible to be recognized (e.g., after harvesting), the point should be recorded either as “Bare land” in case of spontaneous weed covering less than 50% of the surface, or as “Spontaneous re-vegetated surfaces” (E30) in case of spontaneous weed cover more than 50% of the surface.

3.3.3. Forest Land

LUCAS follows the “Reg (EC) No 2152/2003 of 17/11/2003” of the European Union, where “forest” is defined as “land with tree crown cover of more than 10% and area of more than 0.5 ha; the trees should be able to reach a minimum height of 5 m at maturity *in situ*”. The LUCAS definition of “Woodland” relies mostly on tree density in a 20-m radius surface (the “extended window of observation”, corresponding to a circular extent of 1257 m²). The threshold of tree density is set to 10%, whereas other parameters are not deliberately included in the definition. For the WWF Hellas map, the determining parameter is tree height, with a minimum threshold set at 2 m. Implicitly from Landsat TM multispectral image pixel size, the minimum mapping unit (MMU) for WWF Hellas map is a square of 900 m², whereas for LUCAS, MMU is by default a circle of 20 m radius (Figure 4).

In LUCAS nomenclature, mixed forests form a separate category, namely “Mixed Woodland” (C30), whereas WWF Hellas map does not contain a mixed forest class. The threshold in LUCAS for classifying a forest in a pure type is set to 75%; otherwise, it belongs to mixed cover. In WWF Hellas, clear majority (*i.e.*, >50% of the surface) is the default rule in order to classify a forest either in “Broadleaved” or “Coniferous” forests. As a result, those surfaces with a mixture between 25% and 75% will be classified in one of the two WWF Hellas pure forest groups, but be recorded as “mixed” in the reference dataset (LUCAS) (Table 2).

3.3.4. Shrublands and Grasslands

“Sclerophyllous vegetation” of WWF Hellas is not a distinctive class in LUCAS, as sclerophyllous associations may be contained in “Broadleaved and Evergreen Woodland”, in “Mixed Woodland”, or in “Shrubland” categories. The classification of WWF Hellas in this case is closer to that of CORINE nomenclature, where sclerophyllous vegetation forms a single category at the more detailed level of the system (CORINE category 3.2.3), under the class of “Scrub and/or herbaceous vegetation associations” (CORINE category 3.2). According to CORINE LC 2000, Sclerophyllous vegetation surfaces in Greece cover about 17% of the country.

Table 2. Comparative list of terms used for the definition of forestland cover according to the WWF Hellas map and LUCAS nomenclature.

Forest Parameter	WWF Hellas	LUCAS
MMU	900 m ²	1257 m ²
Shape	Square	Circular
Density	-	>10%
Tree height	>2 m	-
Purity (Mixture)	>50% (mixture N/A)	>75% (25%–75%)

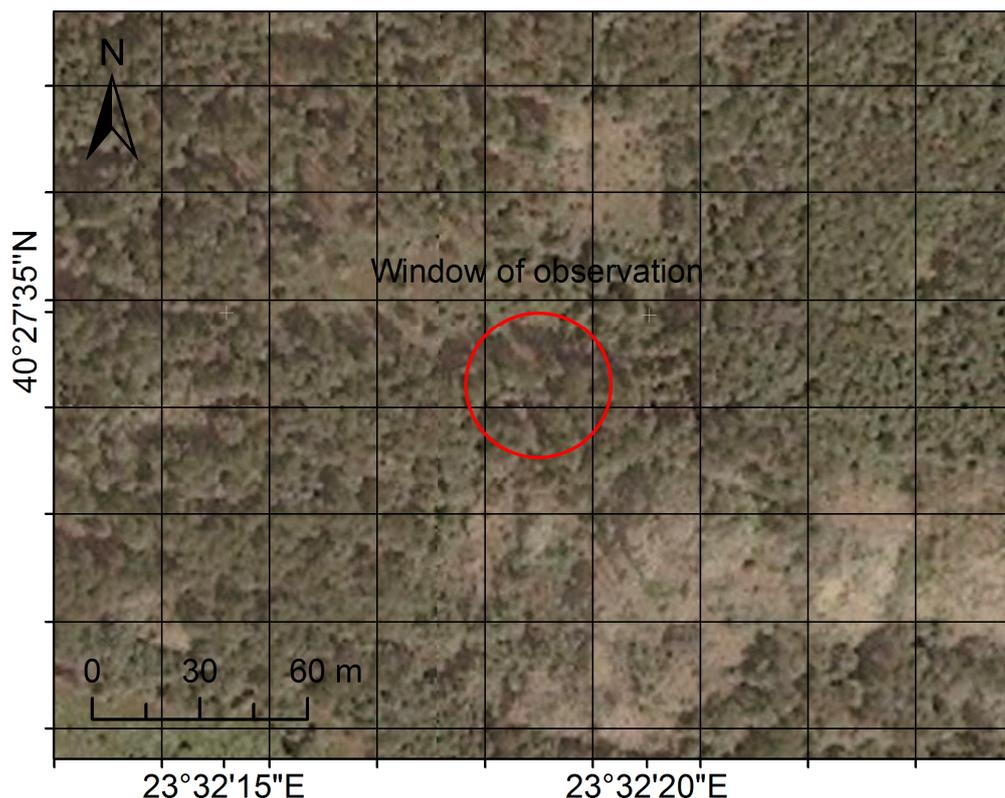


Figure 4. Comparison of the minimum mapping unit of the WWF Hellas map (identical to the Landsat pixel size, 30 m; here depicted as a 30 × 30 m grid) and the LUCAS extended window of observation (20-m radius; here depicted as a red circle) in a forestland in Greece (image background: ArcGIS© basemap).

“Shrublands” of WWF Hellas differ from the same category of LUCAS. The definition WWF Hellas focuses on high evergreen shrubs and ignores low shrubs, which instead are summarized in the heterogeneous class of “Sparsely vegetated areas”. On the other hand, the “Shrubland” class of LUCAS is characterized by the feature of “small woody plants” and is divided into two subclasses in terms of density of trees found inside the shrub associations: D10, corresponding to tree density between 5% and 10%; and D20, to less than 5%. Although in practice it is difficult to assess tree density in that detail, the threshold of 10% coverage by trees remains critical for discriminating shrubland from forest in LUCAS. This threshold can be considered as a simplification of FAO’s discrimination of “forests” from “other wooded land” (“Other wooded land” means land either with a tree crown cover of 5% to 10% of trees able to reach a height of 5 m at maturity *in situ*, or a crown cover of more than 10% of trees not

able to reach a height of 5 m at maturity *in situ* and shrub or bush cover [5]. Presence of trees in the “Shrublands” of the WWF Hellas map is not quantified, but only implied by the term “transitional woodland; usually, surfaces adjacent to forests”. In ecological terms, shrubs in Mediterranean areas correspond to permanent vegetation in regions where soil and climate conditions are poor.

“Sparsely vegetated areas” of WWF Hellas is not a distinctive class in LUCAS, too. Moreover, the terms “low vegetation” and “sparse vegetation” contained in the definition of this class, are not linked to any specific LUCAS category. Instead of “low” and “sparse”, LUCAS nomenclature adopts the term “dominant” in order to characterize a vegetated surface as dense enough to belong to a category. Based on the fact that bare land is defined (by LUCAS) as land with less than 50% vegetation cover, the threshold of grass associations cover in “Grassland” or “Shrubland” can be taken (inversely) as larger than 50%. Furthermore, “low vegetation” is not quantified by any of the two systems. By implication, however, “low vegetation” could be taken as lower than 2 m, based on the fact that forests are defined (by WWF Hellas, this time) as containing trees higher than 2 m.

Bare land for WWF Hellas is included in the mixed category of “Bare land and artificial surfaces” and is defined relatively strictly, as it implies the absolute absence of vegetation; for example, in rock outcrops, quarries, land permanently covered by snow, *etc.* LUCAS is more flexible in the identification of bare land, as it requires a percentage larger than 50%, *i.e.*, bare land has only to be predominant in the examined surface.

A summary of the categories corresponding to dominant shrub or grass associations and bare land for both classification systems and their respective criteria (group, height, density, presence of trees, species compositions, *etc.*) is provided in Table 3.

3.3.5. Other Categories

The “Burnt areas” of the WWF Hellas map are defined as areas with any natural land cover, burned within a period of less than a year before mapping. For LUCAS, burnt areas do not form a distinctive category, but are classified according to their present cover (e.g., shrubland, bare land, *etc.*); they may also correspond to “fire breaks” found within forests. The class “Water bodies” of WWF Hellas contains both water bodies and wetlands, which for LUCAS correspond to different categories, *i.e.*, “Water areas” and “Wetlands”, respectively. Therefore, “Water bodies” could be considered as a mixture of classes compared to the reference data (LUCAS).

3.3.6. Inter-Class Relations and Decision-Tree

From the analysis of all class definitions, four vegetation parameters were found to be critical for the establishment of a set of inter-class relations between the WWF Hellas map classification scheme and the LUCAS nomenclature, namely:

- Vegetation type, which refers to the ecologically dominant groups in terms of trees, shrubs, or grasses.
- Vegetation height, concerning shrub associations only.
- Vegetation density, in terms of thresholds associated with specific categories.
- Composition, in terms of species thresholds associated with specific groups.

Table 3. A summary of shrubland and grassland categories for LUCAS and WWF Hellas classification systems and respective criteria (criteria in bullets; LUCAS criteria before slash and WWF Hellas criteria after slash, respectively).

LUCAS	WWF Hellas	“Shrublands” (6) or “Sclerophyllous Vegetation” (5) if Sclerophyllous Species are Predominant	“Sparsely Vegetated Areas” (7)	“Bare land and Artificial Surfaces” (1)
D10: Shrubland with trees		<ul style="list-style-type: none"> shrubs > 20%/high shrubs 10% < tree cover < 20%/transitional woodland 	<ul style="list-style-type: none"> small woody plants > 20%/low sparse vegetation 10% < tree cover < 20%/transitional woodland 	N/A
D20: Shrubland without trees		<ul style="list-style-type: none"> shrubs > 20%/high shrubs tree cover < 10%/transitional woodland 	<ul style="list-style-type: none"> small woody plants > 20%/low sparse vegetation tree cover < 10%/transitional woodland 	N/A
E10: Grassland with trees		N/A	<ul style="list-style-type: none"> grass > 50%/low sparse vegetation 5% < tree cover < 10%/N/A 	N/A
E20: Grassland without trees		N/A	<ul style="list-style-type: none"> grass > 50%/low sparse vegetation plants tree cover < 5%/N/A 	N/A
F: Bare land		N/A	<ul style="list-style-type: none"> bare land > 50% (but not 100%)/low sparse vegetation 	<ul style="list-style-type: none"> bare land = 100%/rock outcrops

Table 4. A summary of the thematic agreement between the classification scheme (WWF Hellas map 2007) and the reference nomenclature (LUCAS 2009) in terms of inter-class relations in a table of correspondence.

WWF Hellas Class Name	WWF Hellas Class Definition	Automated Validation (Use of LC1 Attribute)	Supervised Validation (Use of Ground Photos & Imagery)	Justification
1. Bare land and artificial surfaces	Settlements, industrial areas, rocky outcrops, quarries, and land covered by snow	A: Artificial land (excl. A13) F: Bare land	F: Bare land	A13: Greenhouses correspond to “2” “Rocky outcrops” of “1” correspond to F
2. Agricultural land	Land used for crop production including greenhouses and bare arable land (rangelands excluded)	B: Cropland A13: Greenhouses	F: Fallow/abandoned land E30: Spontaneous re-vegetated surfaces D: Shrubland	F, E30, or D can be used for crops difficult to identify after harvesting (not rangelands)
3. Broadleaved forests	Areas covered by broadleaf trees >2 m high with sparse or dense cover	C10: Broadleaved and Evergreen Woodland	C30: Mixed Woodland	Due to lack of a mixed forest class in WWF Hellas, C30 may be used for reinterpretation of class “3”
4. Coniferous forests	Areas covered by coniferous trees >2 m high with sparse or dense cover	C20: Coniferous Woodland	C30: Mixed Woodland	Due to lack of a mixed forest class in WWF Hellas, C30 can be used as reference to “Coniferous forests”
5. Sclerophyllous vegetation	Surfaces covered by sclerophyllous shrubs or tree types	N/A	C10: Broadleaved and Evergreen Woodland C30: Mixed Woodland D: Shrubland	Sclerophyllous surfaces do not form a single class in LUCAS; instead, sclerophyllous species may be contained in C10, C30, or D classes
6. Shrublands	Areas covered by high evergreen broadleaved shrubs and transitional woodland; usually, surfaces adjacent to forests	N/A	D: Shrubland	“Small woody plants” of LUCAS are not included in class “6” of WWF Hellas; instead, they are included in class “7”
7. Sparsely vegetated areas	Natural surfaces covered with sparse low vegetation	N/A	D: Shrubland E: Grassland	D of LUCAS can be low and sparse; E is low and can be sparse
8. Burned areas	Areas burned within a period less than a year before image acquisition	N/A	C: Woodland D: Shrubland F: Bare land	Burnt areas by WWF Hellas may be any natural land cover; in LUCAS terminology, they are classified as the present natural category, or they can be “fire breaks” within forests
9. Water bodies	Lakes, rivers, wetlands, <i>etc.</i>	G: Water areas H: Wetlands	N/A	Wetland is not independent class in WWF Hellas map; instead, it is contained in “Water bodies” class

These many-to-many relations were a prerequisite for the judgment of thematic agreement between classified and reference data and were used by both the automated and supervised validations processes. A summary of the thematic agreement (with short justifications) is provided in a table of correspondence (Table 4).

With a view to support an objective class identification and labeling of testing points in the supervised process specifically, a classification decision-tree was designed using respective quantified criteria based on the critical vegetation parameters mentioned above (examples of criteria are: shrubs height >1 m, grass >50%, conifers >75%, *etc.*).

A 2-year temporal inconsistency between LUCAS 2009 and the WWF Hellas map of 2007 was not considered prohibitive for their comparison. The degree of land cover changes between 1987 and 2007 (*i.e.*, a 20-year period) in Greece was found to be less than 1.5% on average [14]. Some areas in Peloponnese affected by extensive wild fires in August of 2007 and thus not mapped by the 2007 product [22], cover about 1700 km², which corresponds only to 1.76% of the total examined surface in the current study. Moreover, burnt areas are treated by LUCAS according to their present land cover and therefore they could be compared with the same surface before fire.

4. Results and Discussion

4.1. Automated Validation

4.1.1. Generic

In the automated process, the only reference information used was the LUCAS primary land cover attribute (LC1). Four WWF Hellas classes were excluded from the error matrix, namely “Sclerophyllous vegetation”, “Shrublands”, “Sparsely vegetated areas”, and “Burnt areas”, as they did not correspond clearly to any of the LUCAS categories. Similarly, the subsets of those reference (LUCAS) classes that did not correspond to any WWF Hellas class were also excluded from the assessment. More specifically, “Mixed Woodland” was excluded from the assessment of “Broadleaved forests” and “Coniferous forests”, as there is no mixed forest class in the WWF Hellas map; while, “Spontaneous re-vegetated surfaces” (E30) and “Bare land” (F) were excluded from the assessment of “Agricultural land”, as a field would possibly be misclassified if it were covered by unknown crop residuals (see Section 3.3.2). As a result, a total number of 4327 points were assessed in the automated validation (62.7% of the full visited set). The excluded subsets were included again in the supervised validation process. The potential use of LC2 (*i.e.*, the secondary land cover of LUCAS), which theoretically could assist automated validation in case of some complicated classes, was completely missing from the Greek LUCAS 2009 database.

4.1.2. Error Matrix

For reasons of comparison of the automated process results with those derived from the supervised process (where all classes take part), the non-assessed classes by the automated process were kept in the error matrix, though marked with “N/A” in the totals and accuracy cells. Therefore, after necessary splits and regrouping of the LUCAS categories, a 9 × 12 error matrix was used in both cases.

Table 5. The error matrix and accuracy totals of the automated validation process (underlined numbers indicate points resulted from clear relations of agreement; numbers in brackets indicate excluded class relations).

WWF Hellas classes	LUCAS categories	A: Artificial Land Except Greenhouses (A13)	A13: Greenhouses	B: Cropland	C10: Broadleaved and evergreen woodland	C20: Coniferous woodland	C30: Mixed woodland	D: Shrubland	E10,E20: Grassland with, without sparse tree/shrub cover	E30: Spontaneous re-vegetated surfaces	F: Bare land	G: Water areas	H: Wetlands	Total points	User's accuracy
1: Bare land and artificial surfaces		<u>90</u>	0	18	4	3	1	23	39	2	<u>23</u>	6	2	211	53.6%
2: Agricultural land		136	<u>5</u>	<u>1972</u>	290	43	57	351	430	(257)	(99)	17	15	3316	59.6%
3: Broadleaved forests		6	0	6	<u>389</u>	20	(61)	42	12	0	3	2	1	481	80.9%
4: Coniferous forests		0	0	2	52	<u>109</u>	(32)	32	7	1	0	0	0	203	53.7%
5: Schlerophyllous vegetation		1	0	1	(130)	47	(36)	(138)	12	0	2	0	0	N/A	N/A
6: Shrublands		3	0	6	175	43	38	(284)	42	2	6	0	0	N/A	N/A
7: Sparsely vegetated areas		13	0	21	142	63	30	(636)	(188)	(6)	(33)	2	3	N/A	N/A
8: Burnt areas		0	0	0	(0)	(7)	(1)	(3)	(2)	(0)	(0)	0	0	N/A	N/A
9: Water bodies		0	0	3	6	0	0	9	5	1	3	<u>72</u>	<u>17</u>	116	76.7%
Total points		232	5	2001	741	175	N/A	N/A	N/A	N/A	29	97	35	4327	
Producer's accuracy		38.8%	100.0%	98.6%	52.5%	62.3%	N/A	N/A	N/A	N/A	79.3%	74.2%	48.6%		61.9%

The automated validation resulted in an overall accuracy of 61.9% and very diverse user's and producer's accuracies (Table 5). As it is conceived from the results, a large number of points classified as "Agricultural land" (356) was found to be "Spontaneous re-vegetated surfaces" or "Bare land" according to LUCAS. This fact justifies the exclusion of these points prior to the automated process. These points (almost 11% of "Agricultural land") were examined in the supervised processes by reinterpreting the respective ground photos. Another large number of points classified as "Agricultural land" were found to be "Grassland with sparse tree/shrub cover" (E10) or "Grassland without tree/shrub cover" (E20), but this fact could be attributed only to possible confusion of the classification algorithm developed by the WWF Hellas team for the mapping, and not to any other conceptual or measuring uncertainty of the validation process.

Furthermore, quite a big number of points classified as "Agricultural land" (almost 14% of the class) were found to be "Artificial land" or "Shrubland" according to LUCAS (136 and 351, respectively; 487 in total). An explanatory hypothesis was that in the case of "Artificial land", these points correspond mostly to roads and small buildings found inside agricultural areas; while in the case of "Shrubland", they correspond to natural vegetation patches found within agricultural areas, either in terms of riparian vegetation, or hedges between different farms (tree plantations, vineyards, or arable land). "Agricultural land" was mapped by WWF Hellas based on ancillary information from the Land Parcel Identification System (LPIS) of Greece at the block level (*i.e.*, the land use level). This led to a definition of agricultural land by WWF as a land use rather than land cover and on a scale broader than that of the field scale. While for LPIS at this scale, small or linear features are ignored, for LUCAS they are recorded, as the principle of observation is a minimum of 3 m diameter or width. These subsets were re-examined in the supervised process by reinterpreting the respective ground photos, occasionally assisted by available imagery.

In order to calculate the Cohen's kappa coefficient, the error matrix was squared by aggregating complementary reference classes (*i.e.*, A + F, B + A13 and G + H; see Table 4) and removing those ones that did not correspond to all WWF Hellas classes (*i.e.*, C30, D, E10-20, E30, and F). In this (simplified) square error matrix, the Cohen's kappa coefficient was found to be 0.625; therefore, the two datasets demonstrated a moderate agreement [20]. It is worth noting that the overall accuracy in the simplified square error matrix was found to be 81%, *i.e.*, significantly higher than the accuracy of the original matrix (61.9%). Therefore, class regroupings in the error matrix seem to work on the expense of the classification performance.

4.1.3. Error Distribution

Spatial distribution of classification errors can assist in the recognition of some uncertainty sources, because it allows to link error with other spatial properties of the mapped area. Among several ways of presenting spatial error distribution, a Boolean method was selected here [23]. Indicatively, the spatial error distribution maps for "Agricultural land" and "Broadleaved forests", are demonstrated in Figures 5 and 6, respectively. For the "Agricultural land", classification errors show a preference to mountainous or semi-mountainous areas, whereas for the "Broadleaved forests", error seems to prefer the western and southern parts of the country.

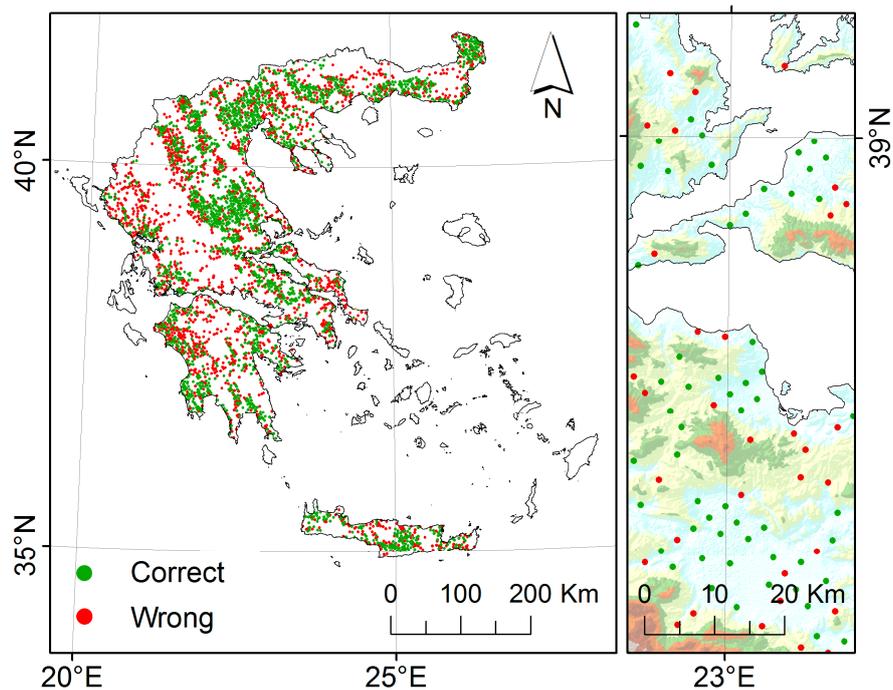


Figure 5. Spatial distribution of classification errors for the class “Agricultural land” of the WWF Hellas map; in the inset, a closer view from central Greece, on an elevation background.

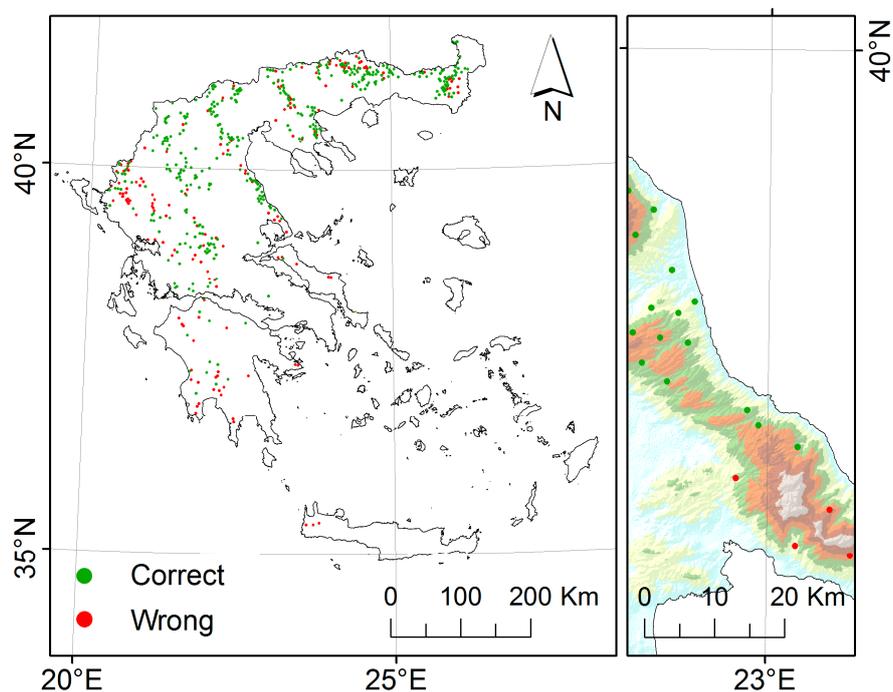


Figure 6. Spatial distribution of classification errors for the class “Broadleaved forests” of WWF Hellas map; in the inset, a closer view from central Greece, on an elevation background.

4.2. Supervised Validation

4.2.1. Generic

By including all WWF Hellas classes in the assessment, the supervised validation process broadened the potential of agreement between classified and reference datasets. In many cases, especially for the

“Agricultural land”, the evaluation was assisted by visual interpretation of the very-high resolution imagery provided by the ArcGIS © basemap, in parallel with the use of ground photos; and where this basemap was not detailed enough to support the interpretation, a country-scale mosaic of colored air photographs, captured in the period 2007–2009 and provided by the Greek cadastre institution, was employed [24].

4.2.2. Artificial Surfaces, Agricultural Land, and Bare Land

From a total of 23 “Bare land” (F) points found inside “Bare land and artificial surfaces” (1), 20 were verified as “Bare land and artificial surfaces” (87%), two were rejected, and one could not be identified. This result justifies *a priori* merging of “Artificial land” (excluding “Greenhouses”, A13) and “Bare land” (F) into a single group, which was conducted during the automated process.

For a total of 356 points and more specifically, 257 of “Spontaneous re-vegetated surfaces” (E30) and 99 of “Bare land” (F) found inside “Agricultural land” (2), the available ground photos were interpreted with a view to discover recent signs of agricultural management, e.g., wheel tracks, tilled soil, crop residues, irrigation installations, a regular field shape, *etc.* These signs could be supportive for land under cultivation or rotation; existence of neighboring agricultural parcels or a nearby village was also in favor of a positive assessment (Figure 7). Most of the positive assessments were associated with short observation distances; only in seven cases the observation distance was larger than 100 m.



Figure 7. A case where agricultural land was classified as “Spontaneous re-vegetated surfaces” (E30) by the LUCAS surveyors, due to the fact that crop was not possible to be identified from crop residuals; in the inset: a close view of the residuals.

From a total of 257 points of “Spontaneous re-vegetated surfaces” (E30) found inside “Agricultural land” (2), 9% were verified as “Agricultural land”, 90% were rejected, and 1% could not be identified (no photos available). In some cases, the crop or residuals could be recognized in the photos (e.g., corn or other cereals), while in some other cases, there were signs of tillage, visible cultivation rows, wheel tracks, or contingency with other fields. From a total of 99 points of “Bare land” (F) found inside “Agricultural

land” (2), 26% were verified, 71% were rejected, and 3% could not be identified (in one case, there were no photos available). Many points were found with signs of tillage, irrigation installments, or presence of weeds. As a result, 14% of the reassessed points of “Spontaneous re-vegetated surfaces” (E30) and “Bare land” (F) were reclassified as “Agricultural land”.

Reinterpretation of the 136 points of “Artificial land” (A) found inside “Agricultural land” (2), reaffirmed an assumption of small or linear artificial features found inside—otherwise—agricultural areas, e.g., small agricultural installations, rural roads, *etc.* From a total of 351 points of “Shrubland” (D) found inside “Agricultural land”, 5% were verified, 92% were rejected, and 3% could not be identified. Most of the examined cases (76%) were found to have an observation distance smaller than 100 m. In addition to that, there was another 13% of the points found either very close to agricultural land (*i.e.*, in a distance smaller than 60 m), or partially covering it, or at the edge of an agricultural field. In a characteristic case in Chalkidiki, the 20-m radius surface (*i.e.*, the extended window of observation) covers partially a natural vegetation patch, an olive plantation, and arable land. The polygon covering this surface is recorded as “Mixed Arable Land” (ID = 41) by the LPIS, whereas for LUCAS the point is evaluated as D20 (Shrubland without tree cover) (Figure 8). These findings are in accordance with the hypothesis that the use of LPIS as ancillary information in image classification conducted by WWF Hellas, resulted in the lowering of classification performance, considering that the extended mixture of agricultural and natural land was overlooked.

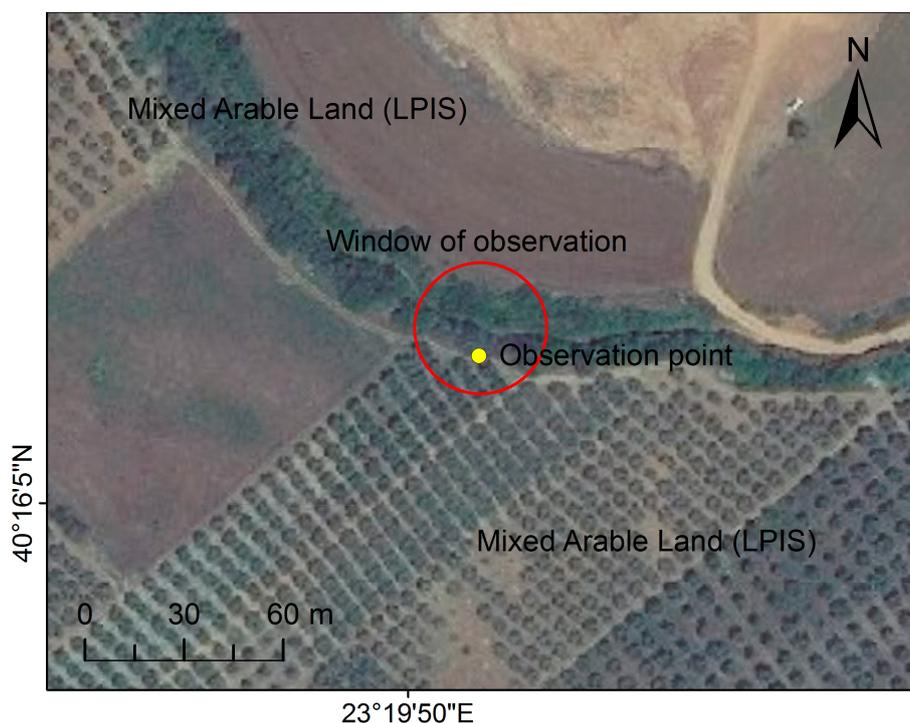


Figure 8. A particular case in Chalkidiki, where the extended window of observation (red circle) covers partially natural vegetation and agricultural land (olive trees and arable land). The surface is classified as “Shrubland without tree cover” (D20) by the LUCAS surveyors, whereas for LPIS is considered as “Mixed Arable Land”. The point of observation was found inside the extended window of observation.



Figure 9. A particular case in Thessaloniki, where the point was classified as “Agricultural land” by WWF Hellas and as “Shrubland without tree cover” (D20) by the LUCAS surveyors, possibly due to the existence of a shrub patch located at the edge of the olive plantation.

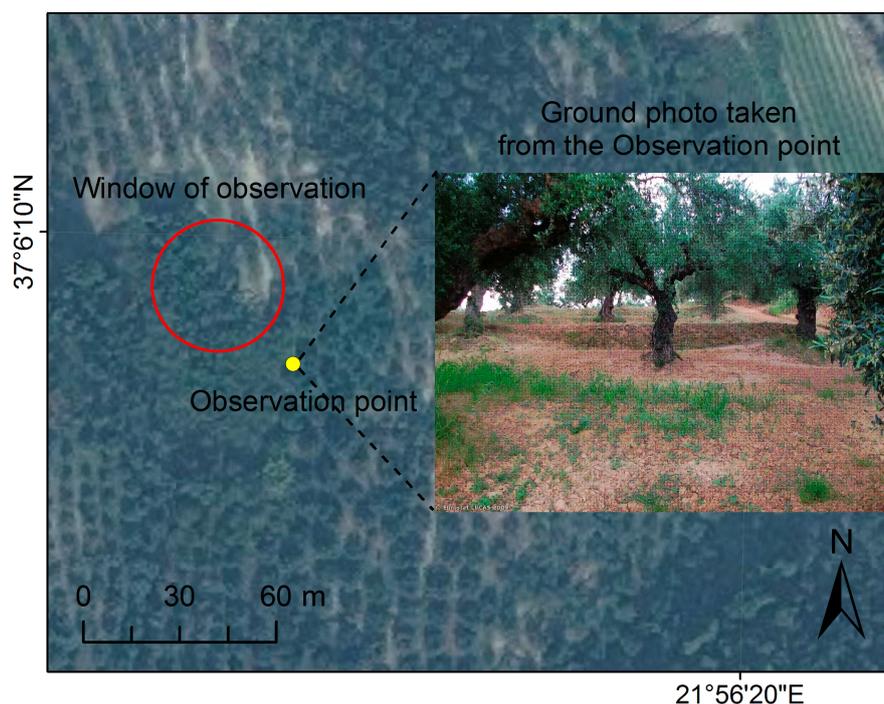


Figure 10. A particular case in Messinia, where the point to be assessed is found in a transitional zone between olive trees and high shrub patches. The point was classified as agricultural land by WWF Hellas and as “Shrubland with sparse tree cover” (D10) by the LUCAS surveyors; in the inset: an eastward photograph taken by the surveyors.

During reinterpretation of ground photos taken in points classified as “Agricultural land” (2), several cases were discovered, where the assessment conducted by LUCAS surveyors could be considered as

questionable. In a characteristic case in Thessaloniki, a small patch of shrubs, located at the edge of an olive plantation, overlays only partially the 20-m radius extended window of observation (as it can be perceived from the basemap), whereas the window covers also a part of the olive plantation beside it (Figure 9).

In another case in Messinia, the point to be assessed was found exactly at the interface between an olive plantation and a natural vegetation patch (shrubs, trees, or mixture). According to the LUCAS instructions, the rule “look to East” could be applied in this vague case, which would be in favor of the “Olive plantation” (B81), rather than of “Shrubland” (D), *i.e.*, the assessment recorded by the LUCAS surveyors (Figure 10).

4.2.3. Forest Land

From a total of 61 “Mixed Woodland” (C30) points found inside “Broadleaved forests” (3), 16% were found to be mixed, 5% were found not to be mixed, and 79% could not be identified. From a total of 32 points of “Mixed Woodland” (C30) found inside “Coniferous forests” (4), 19% were found to be mixed, 16% were found not to be mixed, and 66% could not be identified. The most prominent reason for the remaining high uncertainty in both cases was the long distance and invisibility of the targeted points from the location of observation. Only in few cases, identification of the forest type was possible even from a long distance (*i.e.*, >100 m), when a ground photo was available to the same direction. In these cases, judgment was implicit, based on the fact that the area was covered by a homogeneous vegetation surface between the point of observation and the targeted site. A broadleaved forest in Arcadia is a good example of this kind of remote assessment (Figure 11).

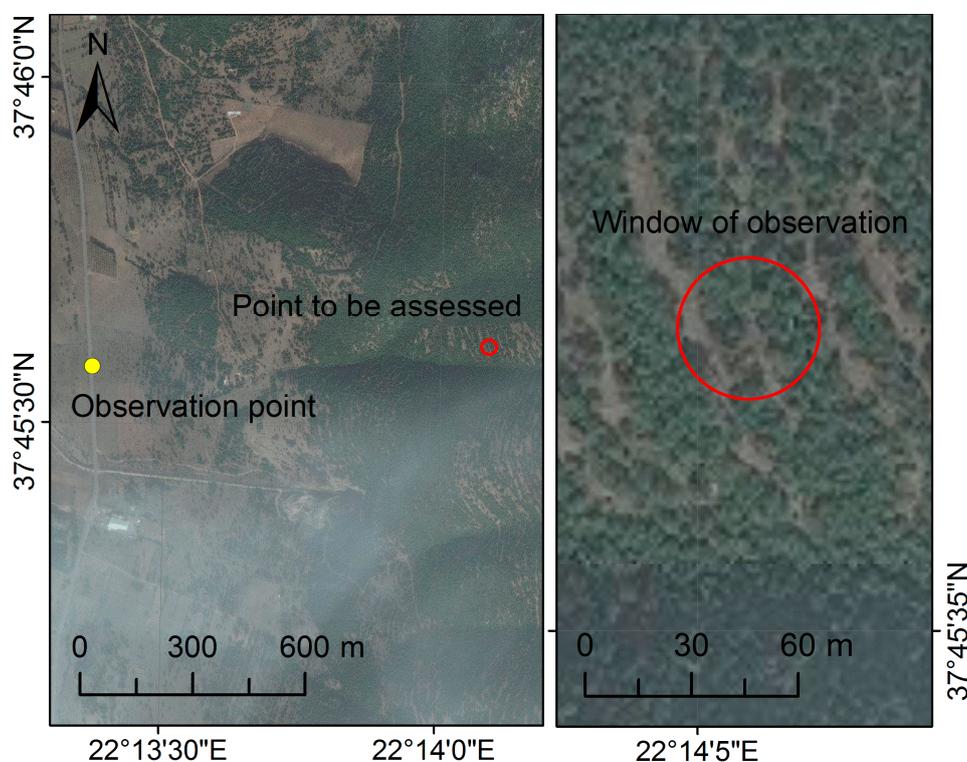


Figure 11. Identification of forest type from a very long distance (about 1 km) by the LUCAS surveyors in Arcadia, possibly using far ground pictures and available air photos; the assessment (“Broadleaved and Evergreen Woodland”, C10) verifies the WWF Hellas map.

A particular difficulty of the LUCAS survey, that may have possibly affected assessment of several points, was discovered with regard to “Mixed Woodland” (C30) category. The mean distance of these points found inside “Coniferous forests” (4) was estimated at 431 m (with a maximum at 3230 m). Considering that tree stands have to be evaluated for their composition, height, biomass, and density, judgment of forest mixture at so long distances render it to be almost impossible, therefore it should not be attempted. In a case of “Mixed Woodland” in Ioannina, the observation distance was 2380 m and the point would not be very possible to be assessed correctly. This questionable assessment could certainly be safer if conducted from an available forest road quite closer to the point (Figure 12).

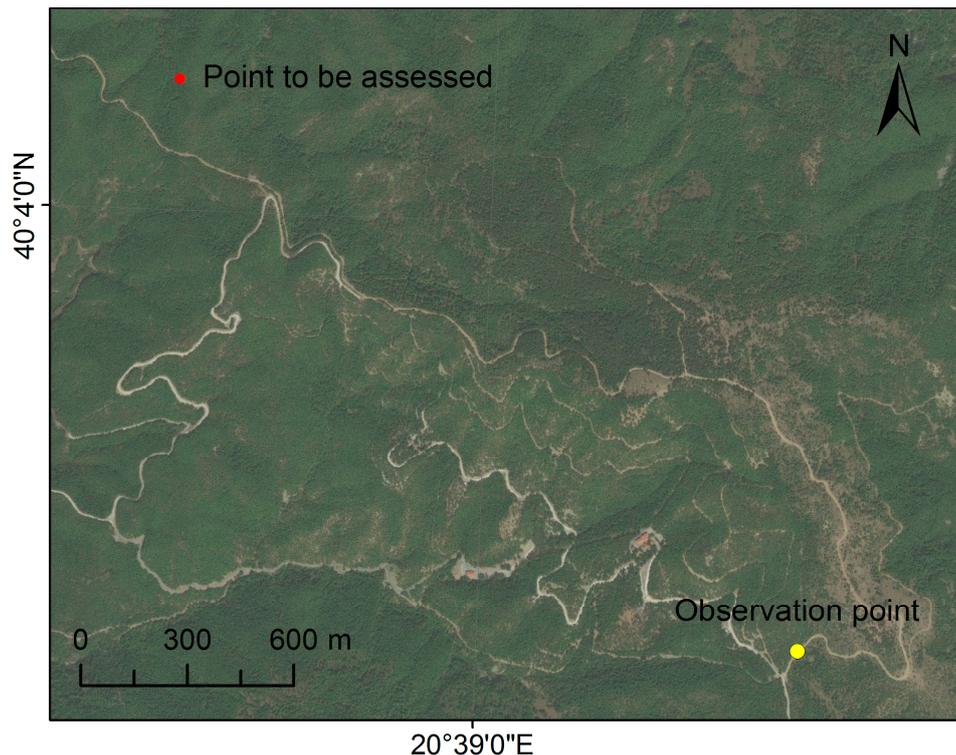


Figure 12. A particular case in Ioannina, where identification of “Mixed Woodland” (C30) was attempted by the LUCAS surveyors from a distance longer than 2 km, although available forest roads closer to the point could provide a safer assessment.

4.2.4. Shrublands and Grasslands

From a total of 284 points of “Shrubland” (D) found inside “Shrublands” (6), 72% were confirmed as “Shrublands”, 17% were rejected, and 11% were not possible to identify. Most of the rejected cases were found to form low shrub associations, therefore according to WWF Hellas nomenclature, they should belong to class “Sparsely vegetated areas” (7). The non-identified cases were associated with surfaces covered by both high and low shrubs, but their composition was not possible to be quantified.

Reference data corresponding to the class of “Sclerophyllous vegetation” (5) are not provided by LUCAS 2009. However, LUCAS groups that could potentially correspond to sclerophyllous species compositions are found inside “Broadleaved and Evergreen Woodland” (C10), “Mixed Woodland” (C30), or “Shrubland” (D). As a result of reinterpretation of the points classified as “Sclerophyllous vegetation”, from a total of 130 points of “Broadleaved and Evergreen Woodland”, 18% were verified as “Sclerophyllous vegetation”, 9% were rejected, and 72% could not be identified. From a total of 36

points of “Mixed Woodland”, 8% were verified as “Sclerophyllous vegetation”, none were rejected, and 92% could not be identified. From a total of 138 points of “Shrubland”, 39% were verified as “Sclerophyllous vegetation”, 1% were rejected, and 60% could not be identified. Again, the most uncertain cases were those with long distances (>100 m) from the point of observation, due to the fact that these points were either non-visible or there were no ground photos taken to the direction of the point to be assessed. The quite large number of verified points of “Shrubland” as sclerophyllous vegetation surfaces (almost all for a distance <150 m) is in favor of the assumption that a large part of shrublands in Greece consists of sclerophyllous species compositions.

Similarly, for surfaces classified as “Sparsely vegetated areas” (7), there are no reference data available in LUCAS 2009. The potential LUCAS groups that could correspond to this WWF Hellas class, can be found inside “Shrubland” (D), “Grassland” (E), or “Bare Land” (F) categories. According to the thematic agreement between the two systems, “Sparsely vegetated surfaces” correspond mainly to “low shrubs” defined as a component of “Shrubland” category, to “grasslands” defined as a component of “Grassland” category, or to “Bare land” category, *i.e.*, when there is no dominant vegetation cover in at least 50% of the surface. In a case in Viotia, the ground photos from inside the window of observation and the basemap show clearly a surface covered by “low woody plants” by more than 20% and therefore, the site was correctly classified as “Sparsely vegetated areas” in the WWF Hellas map. However, the LUCAS surveyors classified the point wrongly as “Grassland without tree/shrub cover” (E20), although it contains shrubs by more than 20% and therefore should be classified more appropriately as “Shrubland without tree cover” (E20) (Figure 13).

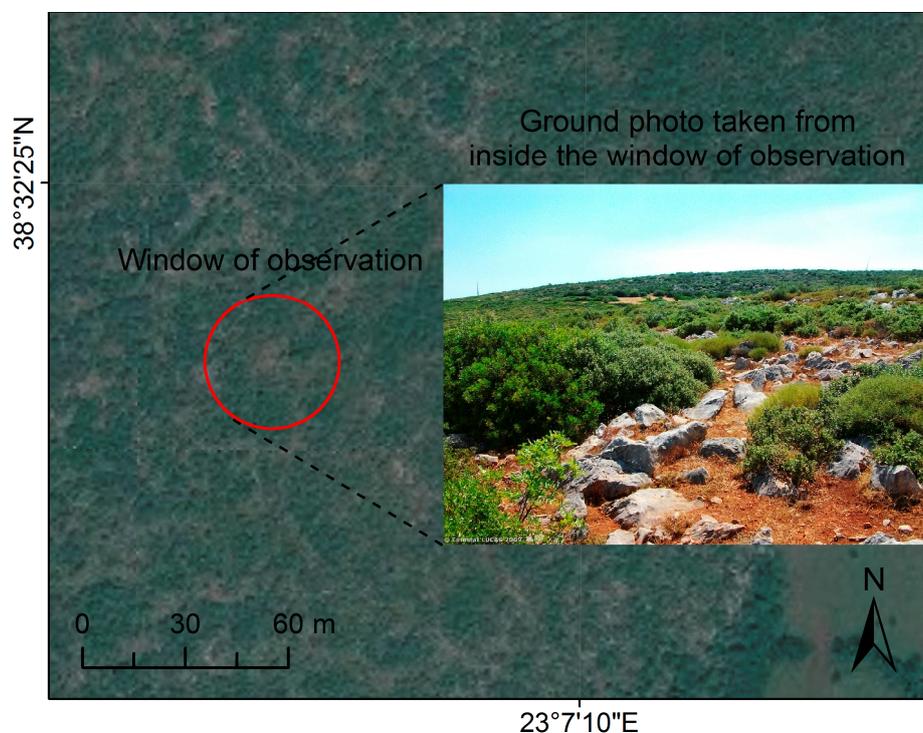


Figure 13. A particular case in Viotia, where the point was classified correctly as “Sparsely vegetated areas” by WWF Hellas, but wrongly as “Grassland without tree/shrub cover” (E20) by the LUCAS surveyors. Both the ground photos and the basemap reveal shrubs lower than 1 m and denser than 20%, which according to LUCAS should be defined as “Shrubland without tree cover” (D20).

The reinterpretation of ground photos of points of “Shrubland” (D) found inside “Sparsely vegetated areas” (7), focused only on those points found at the exact location of assessment (in practice at observation distances less than 2 m), so as to gain a clear idea of the surface details. From a number of 134 “Shrubland” checked points (out of a total of 363 ones), 66% were verified as “Sparsely vegetated areas”, 13% were rejected, and 21% could not be identified. From a number of 30 “Grassland” (E10–20) checked points (out of a total of 188), 90% were verified as “Sparsely vegetated areas” and 10% were rejected. From a total of six “Spontaneous re-vegetated surfaces” (E30) checked points found inside “Sparsely vegetated areas”, 83% were verified as “Sparsely vegetated areas” and 17% were rejected. From a number of 22 “Bare land” (F) points found inside “Sparsely vegetated areas”, 41% were verified as “Sparsely vegetated areas”, 51% were rejected, and 5% could not be identified. Finally, from a total of 13 points belonging to different LUCAS natural categories (C20, C30, D20, and E10) found inside “Burnt areas” (8), 38% were verified as “Burnt areas”, none were rejected, and 62% could not be identified.

4.2.5. Revised Error Matrix

The revised error matrix contains all thematic classes of WWF Hellas (1–9) and all reference categories from LUCAS, *i.e.*, A–H. Once more, the reference categories were regrouped in terms of thematic agreement; *e.g.*, “Artificial land” except “Greenhouses” (A–A13), “Grassland with sparse tree/shrub cover” together with “Grassland without tree/shrub cover” (E10–E20), “Spontaneously re-vegetated surfaces” (E30), *etc.* One thousand three hundred sixty one extra points were included in the reassessment of all classes, thus raising the number of the total assessed points to 5688 (82.4% of the full visited set; 19.7% more points than the automated process). It should be noted that the points “not possible to be assessed” were excluded from the “Total points” of the matrix. Finally, the revalidation resulted in an overall accuracy of 51.8% (*i.e.*, 10.1% less than the automated process) and very diverse user’s and producer’s accuracies (Table 6).

A comparison of the two error matrices (*i.e.*, the one of the automated *vs.* the one of the supervised validation process), reveals lower performances for the latter, which seems to be associated with two specific classes: “Shrublands” (6) and “Sparsely vegetated areas” (7). In fact, these two classes gained extremely low user’s accuracy values (14.4% and 10.4%, respectively), which can be attributed to the high number of points found primarily in the “Broadleaved and Evergreen Woodland” (C10) reference subset and secondarily to “Coniferous Woodland” (C20) and “Mixed Woodland” (C30) subsets. This shows that there was a significant confusion of “Shrublands” and “Sparsely vegetated areas” classes of the WWF Hellas map with all the forest classes of LUCAS (C). If these two WWF Hellas classes remained out of the reassessment, the overall accuracy would be 58.6%, *i.e.*, only 3.3% lower than the one achieved by the automated process.

Table 6. The error matrix and accuracy totals of the supervised validation process (underlined numbers indicate points resulted from clear relations of agreement; number in brackets indicate interpreted points).

WWF Hellas classes	LUCAS categories	A: Artificial land except Greenhouses (A13)	A13: Greenhouses	B: Cropland	C10: Broadleaved and evergreen woodland	C20: Coniferous woodland	C30: Mixed woodland	D: Shrubland	E10,E20: Grassland with, without sparse tree/shrub cover	E30: Spontaneously re-vegetated surfaces	F: Bare land	G: Water areas	H: Wetlands	Total points	User's accuracy (%)
1: Bare land and artificial surfaces		<u>90</u>	0	18	4	3	1	23	39	2	(20)	6	2	210	52.4
2: Agricultural land		(0)	<u>5</u>	<u>1972</u>	290	43	57	(18)	430	(24)	(26)	17	15	3658	55.9
3: Broadleaved forests		6	0	6	<u>389</u>	20	(10)	42	12	0	3	2	1	494	80.8
4: Coniferous forests		0	0	2	52	<u>109</u>	(6)	32	7	1	0	0	0	214	53.7
5: Schlerophyllous vegetation		1	0	1	(24)	47	(3)	(54)	12	0	2	0	0	157	51.6
6: Shrublands		3	0	6	175	43	38	(55)	42	2	6	0	0	383	14.4
7: Sparsely vegetated areas		13	0	21	142	63	30	(89)	(27)	(5)	(9)	2	3	451	10.4
8: Burnt areas		0	0	0	(0)	(2)	(0)	(1)	(2)	(0)	(0)	0	0	5	100.0
9: Water bodies		0	0	3	6	0	0	9	5	1	3	<u>72</u>	<u>17</u>	116	76.7
Total points		113	5	2029	1082	330	145	323	576	35	69	97	38	5688	
Producer's accuracy (%)		79.6	100.0	97.2	38.2	33.6	13.1	50.2	5%	82.9	49.5	72.7	44.7		51.8

4.2.6. Effectiveness of the Reassessment

From a total number of 1361 reinterpreted points, 292 (20.6%) were in favor of confirmation of the WWF Hellas map predictions, 788 (55.6%) were against, and another 336 (23.7%) could not assist in the reassessment. The large number of points left out as “not possible to be assessed” may have biased somehow the sampling design of the supervised process and consequently the overall and partial accuracies. In general, the various categories that underwent reassessment were affected in different degrees in regard to the number of correct and wrong points, or the number of points not able to provide a clear assessment (Figure 14).

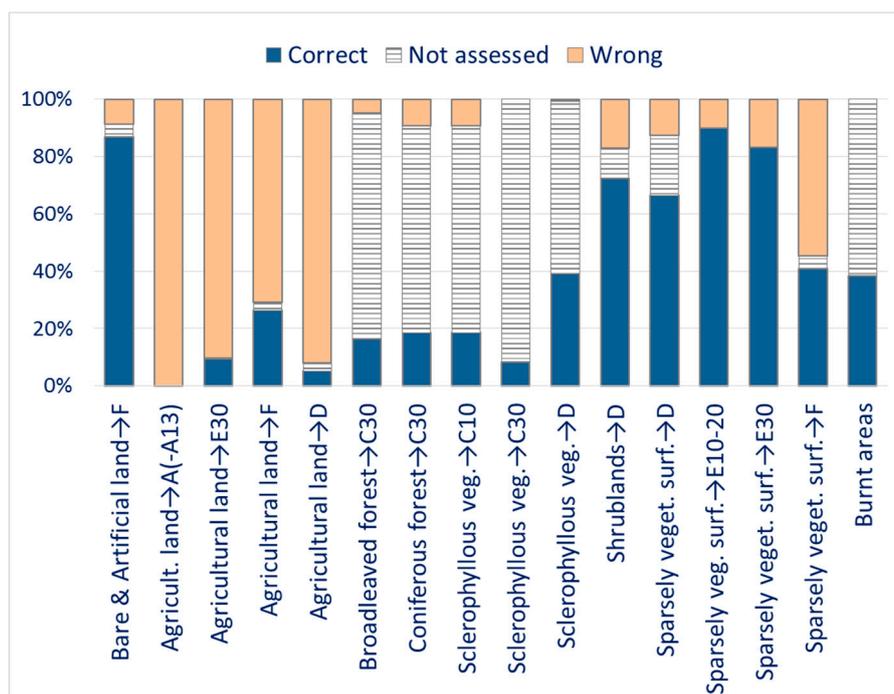


Figure 14. Distribution of correct, wrong, and not able to provide a clear assessment points per WWF Hellas category, as resulted from the supervised validation process.

The close examination of several cases from so many and different sources, has revealed some paradoxes in the implementation of the LUCAS survey protocol. For example, in a case in Fthiotis, the 3-m radius window (which is used for arable land) captures partially the crop of the field and partially a patch of shrubs; whereas, the extended window of observation (20-m radius), which is used for natural land covers, indicates a surface where agriculture is dominant. The surveyor assessed this point as “Shrubland without tree cover” (D20). The fact that the distance of observation is 63 meters, reveals that the judgment was based on available air photographs (Figure 15).

In a similar case in Larissa, central Greece, the point to be assessed is clearly inside a natural patch. However, the extended window of 20 m used to judge the point, covers partially (almost by half) the natural patch and neighboring arable land. Furthermore, part of the targeted natural patch is maintained as a hedge between two agricultural fields. As a result, the agricultural land cover is predominant in three directions. On the other hand, the 3-m radius surface used for arable crops falls exactly inside the natural patch. Judging from an observation distance of 130 m and without visibility, the surveyor obviously used air photographs to assess this point as “Shrubland with sparse trees” (D10) (Figure 16).

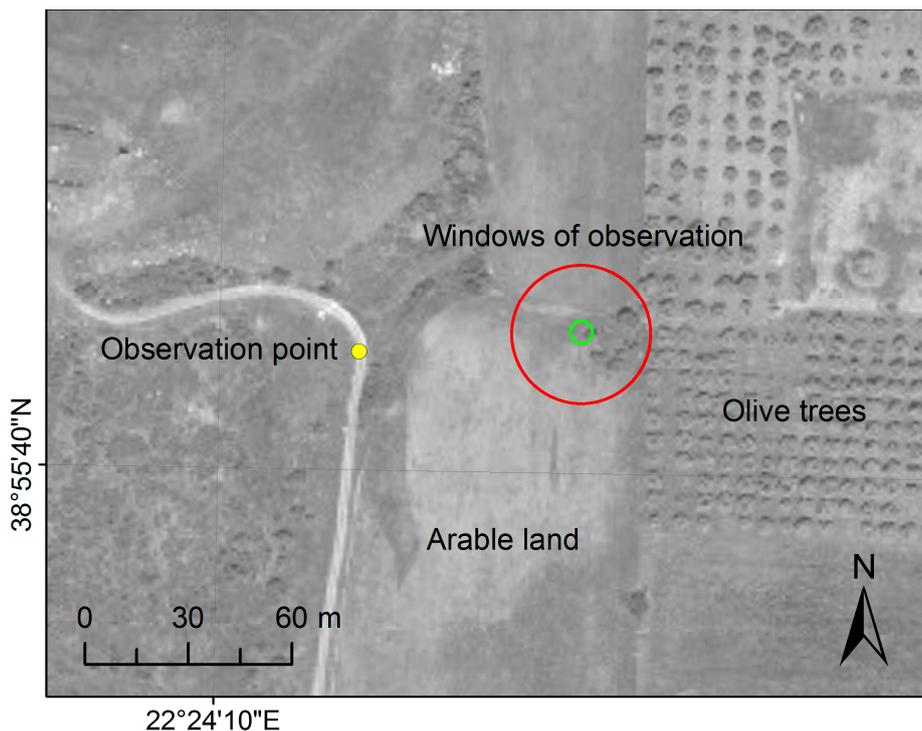


Figure 15. A paradox case in Fthiotis, where a small natural vegetation patch found in agricultural land renders identification impossible, either by using the 3-m radius (small green circle) or the 20-m radius (big red circle) observation windows.

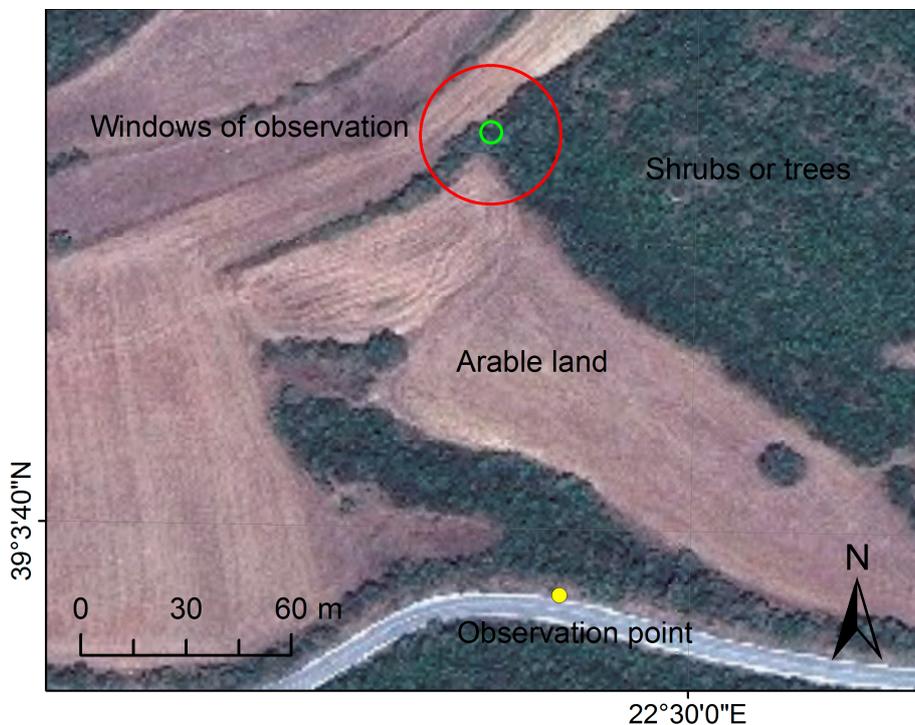


Figure 16. A paradox case in Larisa, where a mixture of agricultural land with an extended shrubland surface is not possible to be identified, neither by using the 3-m radius (small green circle) nor the 20-m radius (big red circle) observation windows.

Familiarization of surveyors with spatial patterns captured by available imagery, together with their experience about common vegetation types in the surveyed area, could assist identification from a distance. In characteristic a case in Herakleon, it is perceived that judgment was based exclusively on the interpretation of ancillary EO data in combination with good knowledge of the local ecosystems (Figure 17).



Figure 17. In a characteristic case in Herakleon, the targeted point was classified as “Shrubland with sparse tree cover” (D10) by the LUCAS surveyors from remotely (>100 m), possibly based on use of EO data and knowledge of the local ecosystems.

5. Conclusions

In this study, LUCAS 2009, a pan-European statistical point database, was used as reference for validating a land cover map of Greece for 2007, produced with remote sensing by the Greek Office of the World Wildlife Fund (WWF Hellas). Validation went through a two-tier methodology: an “automated” process using the main land cover attribute of LUCAS (LC1); and a “supervised” process by reinterpreting all the available ground photos taken by the LUCAS surveyors and ancillary earth observation imagery.

Although, this approach has been used in the past [6–8], the current paradigm was the first to employ LUCAS points for testing a country-scale land cover map with a clearly different nomenclature. This was achieved by decomposing class definitions into four vegetation parameters, namely, type, height, density, and composition, thus assisting description of inter-class relations and thematic agreement in a table of correspondence. The same parameters were used to design a decision-tree of quantified criteria, which allowed for an objective labeling of testing points in LUCAS nomenclature and the WWF Hellas map simultaneously.

Interesting findings were derived from the application of the non-square error matrix method in the automated and supervised validation processes. With the automated process, only five out of nine WWF Hellas categories was possible to assess, due to significant nomenclature differences between LUCAS and the WWF Hellas map. With the supervised process, on the other hand, all the categories were assessed, but only after a systematic use of the LUCAS ground photos together with very high resolution imagery. Moreover, only a small part of the reassessed points (20.6%) verified the WWF Hellas map predictions, whereas the majority of the points was either against (55.6%), or neutral (23.7%) to the map predictions.

As a consequence, the supervised process resulted in a decrease of the overall accuracy by 10.1% compared to the automated process (*i.e.*, from 61.9% to 51.8%) and of all user's accuracies. The categories that were excluded from the automated process were found now to have either moderate (51.6% for "Sclerophyllous vegetation") or low user's accuracies (10.4% for "Sparsely vegetated areas" and 14.4% for "Shrublands"), with only exception that of "Water bodies" (100%). These results indicate that mere automated validations (*i.e.*, without examining class definitions in depth) might result in fake (and most possibly upwards) classification performances.

In summary, the LUCAS 2009 point database was found to be supportive, but not fully efficient, in identifying sources of error in country-scale land cover maps derived from remote sensing. A dedicated ground truth campaign, as the one already carried out with good results for WWF Hellas, seems to be more appropriate, especially if landscapes are as complex as in Greece (achieved accuracy of 87.4%).

From studies carried out in geoland2 project, using very high resolution imagery in a multi-stage area-frame sampling is recommended [25]. With such an approach, statistical representativeness can be assured and heterogeneities in landscapes can be analyzed more efficiently. In addition, the status quo in the sample areas will be documented reliably, allowing reinterpretation in later stages if necessary. In this direction, LUCAS could be used as a verification, rather than a validation, dataset.

Acknowledgments

This research was conducted in the framework of "geoland2" project (EC 218795) and "The Future of the Greek Forest" project, which was funded by WWF Hellas. The authors would like to thank Eurostat for the provision of LUCAS databases and the anonymous reviewers for their comments and suggestions which resulted in improving the manuscript.

Author Contributions

Christos Karydas carried out the construction of the geodatabase, the main part of the analytical work and the major part of manuscript writing. Ioannis Gitas set the idea of the research, discussed and contributed to all steps of the analysis, and commented on the manuscript. Steffen Kuntz elaborated the policy dimension of the research and wrote the respective parts. Chara Minakou was involved in the early steps of the research and conducted part of the reinterpretation. All the contributors were involved in the geoland2 project.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Congalton, R.G. A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sens. Environ.* **1991**, *37*, 35–46.
2. Lu, D.; Weng, Q. A survey of image classification methods and techniques for improving classification performance. *Int. J. Remote Sens.* **2007**, *28*, 823–870.
3. Kuntz, S.; (Ed.). Quality Assurance Approach for the Land FTS, 2007, BOSS4GMES Doc. D324–3, Public Document from the BOSS4GMES Consortium. Available online: <http://www.boss4gmes.eu> (accessed on 14 January 2015).
4. EC EUROSTAT. LUCAS 2009 (Land Use/Cover Area Frame Survey), M3—Non Sampling Error, 2013. Available online: <http://ec.europa.eu/eurostat/documents/205002/769457/LUCAS2009-M3-NonSamplingError-20130414.pdf> (accessed on 14 January 2015).
5. EC EUROSTAT. LUCAS 2009 (Land Use/Cover Area Frame Survey) Technical Reference Document C-3: Land Use and Land Cover: Nomenclature (Version of 16 March 2009), 2009. Available online: http://ec.europa.eu/eurostat/documents/205002/208012/LUCAS2012_C3-Classification_20131004_0.pdf (accessed on 14 January 2015).
6. European Environmental Agency (EEA). The Thematic Accuracy of Corine Land Cover 2000 Assessment Using LUCAS (Land Use/Cover Area Frame Statistical Survey, 2006, Technical Report 7/2006, Copenhagen, ISBN 92-9167-844-9. Available online: <http://www.eea.europa.eu/data-and-maps/figures/corine-land-cover-types-2006> (accessed on 15 January 2015).
7. Gallego, F.J.; Bamps, C. Using CORINE land cover and the point survey LUCAS for area estimation. *Int. J. Appl. Earth Obs. Geoinf.* **2008**, *10*, 467–475.
8. Gallego, F.J. Validation of GIS layers in the EU: Getting adapted to available reference data. *Int. J. Digit. Earth* **2011**, *4*, 42–57.
9. Fuller, R.M.; Groom, G.B.; Jones, A.R. The Land Cover Map of Great Britain: An automated classification of Landsat Thematic Mapper data. *Photogramm. Eng. Remote Sens.* **1994**, *60*, 553–562.
10. Pekkarinen, A.; Reithmaier, L.; Strobl, P. Pan-European forest/non-forest mapping with Landsat ETM+ and CORINE Land Cover 2000 data. *ISPRS J. Photogramm. Remote Sens.* **2009**, *64*, 171–183.
11. Fry, J.A.; Xian, G.; Jin, S.; Dewitz, J.A.; Homer, C.G.; Yang, L.; Barnes, C.A.; Herold, N.D.; Wickham, J.D. Completion of the 2006 national land cover database for the conterminous United States. *Photogramm. Eng. Remote Sens.* **2011**, *77*, 858–864.
12. Skinner, L.; Luckman, A. Introducing a landcover map of Siberia derived from MERIS and MODIS data. In Proceedings of the 2004 IEEE International Conference on Geoscience and Remote Sensing Symposium, Anchorage, AK, USA, 20–24 September 2004.
13. Carrao, H.; Sarmiento, P.; Araujo, A.; Caetano, M. Retrieving land cover information from MERIS and MODIS Data: A comparative study for landscape characterization in Portugal. In Proceedings of the IEEE International Conference on Geoscience and Remote Sensing Symposium, Barcelona, Spain, 23–28 July 2007; pp. 1271–1274.
14. Gitas, I. Greece Then and Now—Diachronic Mapping of Land Cover, 1987–2007. WWF Hellas, 2012. Available online: <http://issuu.com/wwf-greece/docs/diahroniki-hartografisi> in Greek; (accessed on 26 February 2015).

15. EC EUROSTAT. LUCAS 2009 (Land Use/Cover Area Frame Survey), Technical Reference Document C-1: General Implementation Land Cover and Use Water Management Soil Transect Photos, Instructions for Surveyors (Version of 20 March 2009). Available online: http://ec.europa.eu/eurostat/documents/205002/208012/LUCAS2012_C1-InstructionsRevised_20130110a.pdf (accessed on 14 January 2015).
16. Strahler, A.H.; Boschetti, L.; Foody, G.M.; Friedl, M.A.; Hansen, M.C.; Herold, M.; Mayaux, P.; Morissette, J.T.; Stehman, S.V.; Woodcock, C.E. Global Land Cover Validation: Recommendations For Evaluation And Accuracy Assessment Of Global Land Cover Maps. GOFc-GOLD, Report No. 25, Luxembourg: Office for Official Publications of the European Communities, 2006. Available online: http://nofc.cfs.nrcan.gc.ca/gofc-gold/Report%20Series/GOLD_25.pdf (accessed on 14 January 2015).
17. Zhan, Q.; Molenaar, M.; Tempfli, K.; Shi, W. Quality assessment for geo-spatial objects derived from remotely sensed data. *Int. J. Remote Sens.* **2005**, *26*, 2953–2974.
18. Liu, W.; Gopal, S.; Woodcock, C.E. Uncertainty and Confidence in Land Cover Classification Using a Hybrid Classifier Approach. *Photogramm. Eng. Remote Sens.* **2004**, *70*, 963–971.
19. Finn, J.T. Use of the average mutual information index in evaluating classification error and consistency. *Int. J. Geogr. Inf. Syst.* **1993**, *7*, 349–366.
20. Foody, G.M. Status of land cover classification accuracy assessment. *Remote Sens. Environ.* **2002**, *80*, 185–201.
21. Storey, J. *Landsat 7 (L7) Image Assessment System (IAS) Geometric Algorithm Theoretical Basis Document (ATBD)*; U.S. Geological Survey Report LS-IAS-01 Version 1.0; EROS, South Dakota, SD, USA, 2006.
22. Gitas I.Z.; Polychronaki, A.; Katagis, T.; Mallinis, G. Contribution of remote sensing to disaster management activities: A case study of the large fires in the Peloponnese, Greece. *Int. J. Remote Sens.* **2008**, *29*, 1847–1853.
23. Comber, A.; Fisher, P.; Brunson, C.; Khmag, A. Spatial analysis of remote sensing image classification accuracy. *Remote Sens. Environ.* **2012**, *127*, 237–246.
24. National Cadastral and Mapping, S.A. Available online: <http://gis.ktimanet.gr/wms/ktbasemap/default.aspx> (accessed on 14 January 2015).
25. GEOLAND2 Report Summary—(Geoland2—Towards an Operational GMES Land Monitoring Core Service), Project Reference: 218795, Last Updated on 1-8-2014. Available online: http://cordis.europa.eu/result/rcn/53368_en.html (accessed on 15 January 2015).