Regional statistics and
Geographic Information
Author: Savvas Zachariadis
E4.LUCAS

eurostat
Technical Documents
2018

LUCAS 2018

(Land Use / Cover Area Frame Survey)

Quality Report

Table of Contents

List of Figures

List of tables

Abbreviations

CAP	Common Agricultural Policy
CAPRI	Common Agricultural Policy Regionalised Impact
CLC	CORINE Land Cover
CRAN	Comprehensive R Archive Network
CV	Coefficient of Variation
DG	Directorate General
DLV	Deliverable
DMT	Data Management Tool
EEA	European Environment Agency
ELEV	Elevation
ESDI	European Spatial Data Infrastructure
ESS	European Statistical System
ESTAT	Statistical Office of the European Union
EU	European Union
FAO	Food and Agriculture Organization
GPS	Global Positioning System
INE	Instituto Nacional de Estatística (Portugal)
INSPIRE	Infrastructure for Spatial Information
LAEA	Lambert Azimuthal Equal Area
LC	Land Cover
LU	Land Use
LUCAS	Land Use/Cover Area Frame Survey
MS	Member States

NACE	Nomenclature statistique des activités économiques dans la Communauté européenne
NUTS	Nomenclature des Unités Territoriales Statistiques
PI	Photointerpretation
QR	Quality Report
SAS	Statistical Analysis System
SRS	Simple Random Sample
STR	Strata
SQL	Structured Query Language
TW	Transitional Water

1 Introduction

1. Introduction

The scope of this document is to report on the quality of the "Land cover and Land use Area Frame Survey 2018" (LUCAS 2018), including the process, the micro data produced and the derived statistical tables. The structure of the report is defined by the ESS handbook for quality reports[1].

Eurostat has carried out the LUCAS survey every 3 years, since 2009, based on standardized definitions and a standardized methodology. A pilot survey was run in 2006, using a slightly different sample design.

The data collected includes land cover and land use information in the strict sense, as well as territorial information (e.g. irrigation and land management). The reference area in 2018 is the total area of all EU countries including UK.

The LUCAS survey is divided: an in-situ part or field survey (data is collected in the field) and a part where data are produced by photo-interpretation in the office. Photo-interpretation is used for areas that are difficult to access. The statistical tables derived are based on the data of both parts.

The sample for both parts is stratified by main land cover classes and includes more than 238 000 points for the field sample and some 100 000 for the sample that is photo-interpreted. Around 2/3 of the points are visited in subsequent surveys.

The legal base of the LUCAS survey has evolved over the years. A pilot a "Land Use and Cover Area frame Survey (LUCAS)" was launched by DG Agriculture and Eurostat in 2000, based on Decision 1445/2000/EC of 22/5/2000 of the Council and the European Parliament[2], dealing with the application of area frame techniques. In 2001 (postponed to 2002), the first LUCAS pilot survey was carried out in 13 of the 15 Member States of the European Union. The survey was carried out again in 2003 in all EU-15 Member States plus Hungary, allowing improvement of the data collection system and analyses of land use and land cover changes (2001-2003). The project was extended in duration from 2004 to 2007 by Decision 2066/2003/EC of 10/11/2003[3]. The coverage of the EU Member States and the related financing are laid down by Decision 786/2004/EC of 21/4/2004[4]. In 2006, the survey was carried out on 11 Member States (Luxembourg, Belgium, Czech Republic, Germany, Spain, Poland, Italy, France, the Netherlands, Hungary and Slovakia)

[1] *Eurostat (2020). The European Statistical System (ESS) handbook for Quality and Metadata Reports – 2020 Edition. ISBN: 978-92-76-09154-7 Available at: https://ec.europa.eu/eurostat/documents/3859598/10501168/KS-GQ-19-006-EN-N.pdf/bf98fd32-f17c-31e2-8c7f-ad41eca91783?t=1583397712000*

[2] *Decision No 1445/2000/EC of the European Parliament and of the Council of 22 May 2000 on the application of aerial-survey and remote-sensing techniques to the agricultural statistics for 1999 to 2003. Available at: http://eur-lex.europa.eu/legalcontent/EN/TXT/?uri=CELEX:32000D1445*

[3] *Decision No 2066/2003/EC of the European Parliament and of the Council of 10 November 2003 on the continued application of area survey and remote-sensing techniques to the agricultural statistics for 2004 to 2007 and amending Decision 1445/2000/EC. Available at: http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32003D2066*

[4] *Decision No 786/2004/EC of the European Parliament and of the Council of 21 April 2004 amending Decisions No 1720/1999/EC, No 253/2000/EC, No 508/2000/EC, No 1031/2000/EC, No 1445/2000/EC, No 163/2001/EC, No 1411/2001/EC, No 50/2002/EC, No 466/2002/EC, No 1145/2002/EC, No 1513/2002/EC, No 1786/2002/EC, No 291/2003/EC and No 20/2004/EC with a view to adapting the reference amounts to take account of the enlargement of the European Union. Available at: http://eur-lex.europa.eu/legalcontent/EN/TXT/?uri=CELEX:32004D0786*

to test the methodology at EU level with a restricted budget, by starting the current data collection frequency: every three years. From January 2008 onwards, LUCAS has been part of Eurostat's activities and budget. As from 2012, it is financially supported by other DGs of the Commission.

According to the handbook on quality reports, this document includes the following chapters:

1.	*Statistical Presentation – Concepts and Definitions of LUCAS survey*
2.	*Statistical Processing – Methodology and Survey Design*
3.	*Quality Management - Quality Assurance*
4.	*Post Processing – Post Data Collection Procedures*
5.	*Relevance – User Needs*
6.	*Accuracy and Reliability*
7.	*Timeliness and Punctuality*
8.	*Coherence and Comparability*
9.	*Accessibility and Clarity*

This report covers the whole route of the LUCAS 2018 survey i.e. from the sample design process to the implementation of the LUCAS survey (in-situ, photo-interpretation), the quality management and the post-collection process to the presentation of final results. During the 2018 round, photo-interpretation has been widely introduced (29% of totals), notwithstanding, the field collection modality had remained the main norm involving more than 2/3 of total points. Besides, the in-situ mode is the unique characteristic of LUCAS survey.

2 Statistical Presentation

2. Statistical Presentation

2.1. Statistical concepts and definitions

LUCAS is the acronym of Land Use and Cover Area frame Survey. The aim of the LUCAS survey is to gather harmonized information on land use, land cover and environmental parameters. The survey also provides territorial information to analyze the interactions between agriculture, environment and countryside, such as irrigation and land management. Since 2006, EUROSTAT has carried out LUCAS surveys every three years. The most recent surveys happened in the spring-summer of 2009, 2012, 2015 and 2018. Since the LUCAS surveys are mainly carried out in-situ, this means that observations are made and registered on the ground by field surveyors.

The main statistical variables collected in the LUCAS survey are the Land Cover and Land Use. On the sampling units (points), two different modalities for land cover (LC1 - the primary information and LC2 - the secondary one) and land use (LU1 and LU2) can be collected.

The list of all variables collected during the survey can be found in LUCAS primary data https://ec.europa.eu/eurostat/web/lucas/data/primary-data/2018

Definitions relate to total official area of the country, which includes land area and land under inland water. Even if in LUCAS, the concept of land is extended to inland water areas (lakes, rivers, coastal areas such as estuaries, lagoons); it does not embrace uses below the earth's surface (mine deposits, subways, mushroom beds, and ground levels of buildings).

The Land Cover is the physical cover of the earth's surface and the Land Use is the socio-economic function of the land. As far as the land use is concerned, it is worthwhile to specify that the figures refer specifically to the use of the land for which any sign is visible in the ground. Therefore, data reported in any table referring to the use has to be interpreted as the 'visible use'. As an example if a piece of land is regularly used for leisure purposes but no signs are visible on the spot, such a use will not be recorded by the surveyor and will not appear in the figures unless auxiliary data have been used for supporting data collection.

In the field, the surveyor classifies the land cover and the visible land use according to the harmonized LUCAS Survey land cover and land use classification. Landscape pictures are taken in the four cardinal directions. A specific topsoil module was implemented in 2009, in 2012 (partly), 2015 and in 2018.

From the LUCAS survey in situ data collection, different types of information are obtained:

1. Micro data
2. Images
3. Statistical tables

1. Micro data

Land cover, land use and environmental parameters associated to the single surveyed points are available freely for download in the LUCAS dedicated section. Information on landscape features as well as on specific ad hoc modules (soil, grassland) are available. Topsoil samples are taken on 10% about of total LUCAS points. The soil samples of the 2018 collection are currently being analyzed in laboratories.

2. Images

Point and landscape photos taken in the four cardinal directions at each point are available freely by request either via e-mail contact to estat-user-support@ec.europa.eu or by using the online order form.

3. Statistical tables

Statistical tables with aggregated results by land cover, land use at geographical level are available in Eurobase (https://ec.europa.eu/eurostat/web/lucas/data/database) under the domain land cover, land use and landscape (LUCAS). The statistics are presented at NUTS0, NUTS1 and NUTS2 levels using the classification for NUTS 2016.

2.1.1. Data description

LUCAS surveys are carried out in-situ by collecting information on the ground by field surveyors on a set of points that might be also visited in subsequent years. The surveyor classifies the land cover and the visible land use according to the harmonized LUCAS Survey land cover and land use classifications. The classification system has been defined to obtain a clear separation of land cover and land use, a full hierarchy and a comparability with other existing land cover/use systems.

A specific soil module was implemented in 2009, in 2012 (partly), 2015 and 2018. In addition, a new module for verifying Copernicus data has been introduced in 2018 as well as a pilot grassland survey. In coherence with the previous rounds of the LUCAS survey, the 2018 edition includes improvements on some aspects of the survey characteristics.

In 2018 round, the survey is used as a sort of “multi-purpose” survey because it integrates different samples with different objectives:

(i) the estimates on land cover and land use,

(ii) an extended soil module where a topsoil sample is collected, for bulk density, soil biodiversity and organic horizon,

(iii) a test module for grassland,

(iv) additional points for Copernicus programme.

The LUCAS surveys are used to monitor social and economic use of land as well as to monitor ecosystems and biodiversity. Sustainable Development Indicators and Agro Environmental indicators on soil are examples of LUCAS data use, while the collected micro-data collected also serve to produce, verify and validate CORINE Land Cover (CLC) and Copernicus.

2.1.2. Reference Area

The reference area in 2018 was the total area of the EU countries including the UK. The territories/islands of France, Spain and Portugal listed below were not included in the field survey; they are excluded from the reference population and hence the area is not considered in the estimation process. The area of these territories sums up to less than 2.5 % of the total area of EU:

- ES63 (Ciudad Autonoma de Ceuta)
- ES64 (Ciudad Autonoma de Melilla) (ES63 + ES 64 = 0.03% of ES6 (SUR))
- ES70 (Canarias)
- FR9 (Departements D’outre-Mer)
- PT20 (Região Autónoma dos Açores)
- PT30 (Região Autónoma da Madeira)

All the survey has been conceived and designed by Eurostat with an ad hoc technical support from Joint Research Center (JRC). The Contractors were responsible for the data collection in the 28 countries arranged in five lots, also for the recruitment and management of the surveyors and the data delivery. The data collection started in field in March 2018 and was completed in office in March 2019, with the last quality checks; in the 2018 round more than 700 surveyors were recruited for a total of more than 283 000 points to be visited in the ground.

2.1.3. Classification system

The LUCAS classification is characterized by:

- clear separation of land cover and land use
- full hierarchy
- comparability with other existing land cover/use systems

While reading the results and comparing them with other sources, it is important to have in mind that the LUCAS survey clearly distinguishes between land cover and land use. Most of the existing information on land cover and land use is based on mixed classification of land cover and land use (as the CORINE Land Cover classification).

When data from the two different dimensions needs to be matched, compared and/or combined this distinction is particularly worthwhile. For example, land cover 'grassland' relates to the actual coverage of the soil (basically spontaneous vegetation) while its use can vary from private gardens to public parks to agriculture and others. Grassland with agricultural use is an important component of the Utilized Agricultural Area and can be derived combining land cover and use.

LUCAS classifications is hierarchical, having the ability to accommodate different levels of information, starting with structured broad-level classes, which allow further systematic subdivision into more detailed sub-classes. At each level, the defined classes are mutually exclusive.

The LUCAS 2018 Survey classification does not differ from the 2015 survey classification. Main changes for land use compared to 2015 classification involve LU4 and precisely:

U410 Abandoned areas has been further subdivided:
U411 Abandoned industrial areas
U412 Abandoned commercial areas
U413 Abandoned transport areas
U414 Abandoned residential areas
U415 Other abandoned areas

For detailed information, see the LUCAS 2018 classification document.

2.1.4. Statistical unit

The statistical unit in a LUCAS survey is a portion of land of circular shape and a conventional dimension of 1.5-meter radius (extended to 20 meters radius in specific cases). More information that is detailed could be found in the following document: LUCAS 2018 - Instructions

2.1.5. Statistical population

The statistical population in LUCAS survey consists of the "reference population", that is the area of EU territory included in the survey.

2.1.6. Reference period

The current quality report refers to LUCAS survey that had taken place in 2018, starting from March 2018 until October 2018. It should be noted that the above period refers to data collection in the field. Information from a considerable amount of points is collected by means of Photo-interpretation, which had been finalized few months later on.

2.1.7. Frequency of Distribution

As soon as the survey ends, LUCAS aggregated tables are available every three years at time t+18 months. Microdata are downloadable at time t+7months.

2.1.8. Unit of Measure

The unit of measure for Land cover and Land use are expressed in square kilometers (Km2), percentage (%).
For topsoil, the 2009, 2015 and 2018 data samples have been analyzed for:

- the percentage of coarse fragments
- particle size distribution (% clay, silt and sand content)
- pH (in CaCl2 and H2O)
- organic carbon (g/kg)
- carbonate content (g/kg)
- phosphorous content (mg/kg)

3 Statistical Processing

3. Statistical Processing

3.1. Methodology and Survey design

The base list of the survey is obtained using the one km2 grid resulting from the INSPIRE (Infrastructure for Spatial Information in Europe[5]) recommendations. It includes around 4,000,000 points in the entire European Union territory. The projection used is the Lambert Azimuthal Equal-Area coordinate reference system (ETRS 1989 LAEA). From the above grid, the LUCAS points are selected from a standard 2 km grid, which comprises around 1 million points all over the EU. Only a sample of the LUCAS points is visited in each campaign. The survey consists of a two phases area sample; in the first phase a frame of more than 1 million geo-referenced points (the so-called Master sample or first phase sample) is systematically selected from a 2 square km grid built all over the EU territory. The frame is stratified according to land cover classes. From the Master sample, a second phase sample is selected; on these points, statistical information is collected by surveyors in the field or by photo interpretation in the office.

LUCAS 2018 survey focused on:

(i) a different specification and use of the non-eligibility concept,

(ii) a review of the rules for assigning photo-interpreted and field points in the sample,

(iii) a finer stratification.

In this chapter, we present the characteristics of the new sampling design for 2018, and the innovative methodology that had been applied.

3.1.1. Sample design – First Phase (Master)

As mentioned, the LUCAS Master data set is obtained by using a 4 km² grid (2x2 km) which includes around 1 100 000 points covering the EU territory. Each of these points was classified into 10 land cover categories (the strata), based on photointerpretation (PI) of aerial photos or satellite images. Beyond the geographical characteristics of the point (i.e. its GPS coordinates, the values of the corresponding NUTS3, NUTS2, NUTS1 and NUTS0), some specific information was added to each point such as the elevation, the distance to the nearest road, the population density in the most internal 1 km², etc.

---

[5] *INSPIRE. Available at: http://inspire.ec.europa.eu/about-inspire/563*

3.1.1.1. The new stratification variable in Master 2018

A first important improvement had been the update of the information related to the Master sample. Each point of the 2 by 2 Km grid was assigned with an updated stratification and all the related geographical and administrative information available. In comparison to 2005, (date of the previous grid stratification), the variable "STR05" used until 2015 round had been replaced and updated by a new variable "STR18" including an enlarged classification. STR18 is classified in 10 modalities. The modalities of STR05 “wooded area and shrubland” had been split into two (“wooded area” and “shrubland”), while two new modalities “transitional water” (estuaries, intertidal areas, coastal lagoons, etc.) and “impossible to photo-interpret” have been inserted. As a result, a fair proportion of the points (about 26.5%) changed classification (table)

Table 1: Classification of "stratum" variable in 2018 and in 2015 surveys

2018 (STR18)	2015 (STR05)
1-Arable land	1-Arable land
2-Permanent crops	2-Permanent crops
3-Grassland	3-Grassland
4-Wooded	4-Wooded areas and shrub land
5-Shrub land	5-Bare land, low or rare vegetation
6-Bare land	6-Artificial land
7- Artificial	7-Water
8 – Inland water	
9 – Transitional water	
10 – Impossible to PI	

The stratum variable adopted in 2015 was first collected in the year 2005 (STR05). In 2018 survey, the updated (STR18) had led part of the points to be classified differently because they actually changed their characteristics.

3.1.1.2. Land Cover Assignment in Master

In order to develop the sampling strategy, it was necessary to estimate the most probable land cover (LC) that could be observed in each point. Such information is important because it permits to estimate the distribution of the target variables in the different strata.

The most probable Land Cover is assigned to each point of the Master, forecasted by a linear logistic regression model, estimated on the basis of the real data from the 2015 LUCAS survey, also considering about 16 covariates. This information is used to calculate the coefficients of variation (CVs) for the 16 target variables (table 2).

In particular, it was assumed that it is possible to estimate the land cover by referring to a proper classification model in which the value in 2015 could be derived considering some covariates, like the strata to which the point was classified in 2005 (STR05), the land cover as from CORINE 2012, etc. Once the parameters of the model were estimated, there were applied to all the information in the Master data set, to obtain the predicted probability to observe a given land cover for all its records. It has to be noted that the land cover can assume different values, considering all the possible biophysical coverage of land (e.g. natural areas, forests, buildings, roads or lakes, etc.). In our case, we considered the classification referred to the upper bound of the expected errors for the next LUCAS survey. This leads to have 16 classes, as in table 2.

Table 2: Rules used to classify LUCAS land covers in typologies referred to the upper bounds expected errors

Name of the Recoded LC	Land cover	Original classification of land cover accordingly to the LUCAS standards (two digits)
A	Roofed built-up areas	A1
B	Artificial non-built up areas	A2
C	Cereals	B1
D	Root, non permanent industrial crops, dry pulses, etc.	B2, B3, B4 and B5
E	Permanent crop	B7, B8
F	Broadleaved woodland	C1
G	Coniferous woodland	C2
H	Mixed woodland	C3
I	Shrubland with sparse tree cover	D1
L	Shrubland without tree cover	D2
M	Grassland with sparse tree/shrub cover	E1
N	Grassland without sparse tree/shrub cover	E2
O	Spontaneously re-vegetated surfaces	E3
P	Bare land and lichens/moss	F
Q	Water areas	G
R	Wetlands	H

Concerning the active variables used in the model, new information had been derived by an automatic synthesis of the satellite image centered in each point of the Master data set.

Figure 1: Image derived from Google Satellite

From the 49 (7x7) pixels it is possible to obtain the mean and standard deviations of the values referred to each color channel. Such statistics were evaluated for all the points of the LUCAS master data set by a proper procedure able to download automatically the image (in JPEG format) centred at their GPS coordinates.

3.1.1.3. Indexes to evaluate the results of the model

The classification capacity of each model was tested by considering in a first step all the records belonging to the LUCAS 2015 survey (for each selected country). For these records, as the land cover is known it permitted to split the data set in two parts (of almost equal size). The first, called train, was used to estimate the parameters of the model, the second, test, to verify its classification performance.

In a second step, after having evaluated the capacity of the model, all the records of the selected Country were considered, thus permitting to estimate the parameters of the final linear logistic regression model (for all the records that belong to the LUCAS 2015 survey). Then, the model was applied to all the remaining records, having the score of presenting a given land cover.

It is important to observe that the score of the linear logistic regression had been transformed in a specific value of land cover by means of a threshold, able to reproduce the original ratio of points having the considered land cover in the train set.

Moreover, specific indexes were considered to evaluate the capacity of the models; these are based on the confusion matrixes obtained at the end of the estimation of each of the above-described steps (train and test and all the records belonging to the 2015 survey). It has to be noted that each column of the confusion matrix represents the instances in the predicted class while each row represents those of the observed one. In particular:

Table 3: General Confusion matrix

		Predicted class		
		0	1	Total
Observed class	0	True Negative (TN)	False Positive (FP)	TN+FP
	1	False Negative (FN)	True Positive (TP)	FN+TP
	Total	TN+FN	FP+TP	N

The indexes that were considered are those usually adopted to evaluate the results of a classification model:

- accuracy: (TN+TP)/N,
- error rate: (FP+FN)/N,
- sensitivity: TP/(TP+FN),
- specificity: TN/(TN+FP),
- ratio of original positive: (FN+TP)/N,
- ratio of predicted positive: (FP+TP)/N.

Except for the *Error rate*, higher values of the indexes show a good discriminant classification.

In following tables, the confusion matrixes and the classification performance are depicted for each land cover. It has to be noted that these results refer to all the European countries and to all the points in the Master dataset having an observed value of LC in the 2015 survey.

Table 4 Confusion Matrix for each land cover

Land cover	Observed:0, predicted: 0	Observed:0, predicted: 1	Observed:1, predicted: 0	Observed:1, predicted: 1
Roofed built-up areas	332 180	1 879	1 896	2 420
Artificial non-built up areas	326 687	1 964	1 968	7 756
Cereals	274 153	18 003	17 829	28 390
Root, non permanent industrial crops, dry pulses, etc.	304 267	12 692	15 009	6 407
Permanent crop	325 439	466	468	12 002
Broadleaved woodland	272 216	13 241	15 529	37 389
Coniferous woodland	278 177	17 934	17 946	24 318
Mixed woodland	292 469	16 827	16 825	12 254
Shrubland with sparse tree cover	324 939	3 193	5 692	4 551
Shrubland without tree cover	317 183	4 148	4 149	12 895
Grassland with sparse tree/shrub cover	321 608	6 182	6 193	4 392
Grassland without sparse tree/shrub cover	277 751	13 220	14 259	33 145
Spontaneously re-vegetated surfaces	324 319	5 541	5 789	2 726
Bare land and lichens/moss	324 361	3 708	3 813	6 493
Water areas	330 473	4	2	7 896
Wetlands	333 684	368	34	4 289

Table 5: Classification performance

Land cover	Accuracy	Error rate	Sensitivity	Specificity	Original percentage (%)	Percentage from model
Roofed built-up areas	0.989	0.011	0.561	0.994	0.013	0.013
Artificial non-built up areas	0.988	0.012	0.798	0.994	0.029	0.029
Cereals	0.894	0.106	0.614	0.938	0.137	0.137
Root, non permanent industrial crops, dry pulses, etc.	0.918	0.082	0.299	0.960	0.063	0.056
Permanent crop	0.997	0.003	0.62	0.999	0.037	0.037
Broadleaved woodland	0.915	0.085	0.707	0.954	0.156	0.150
Coniferous woodland	0.894	0.106	0.575	0.939	0.125	0.125
Mixed woodland	0.901	0.099	0.421	0.946	0.086	0.086
Shrubland with sparse tree cover	0.974	0.026	0.444	0.990	0.030	0.023
Shrubland without tree cover	0.975	0.025	0.757	0.987	0.050	0.050
Grassland with sparse tree/shrub cover	0.963	0.037	0.415	0.981	0.031	0.031

Grassland without sparse tree/shrub cover	0.919	0.081	0.699	0.955	0.140	0.137
Spontaneously re-vegetated surfaces	0.967	0.033	0.320	0.983	0.025	0.024
Bare land and lichens/moss	0.978	0.022	0.630	0.989	0.030	0.030
Water areas	1.000	0.000	1.000	1.000	0.023	0.023
Wetlands	0.999	0.001	0.992	0.999	0.013	0.014

3.1.2. Sample design - Second Phase

The second phase sample design is a stratified one but the stratification was not fixed like in previous surveys (given by the combinations of NUTS 2 level regions by STR05) but rather obtained in a dynamic way.

Starting from the "atomic strata" (given by the Cartesian product of STR18, CLC and ELEV (elevation) classifications) the final strata and sample size had been identified by aggregating the atomic strata with an iterative algorithm that optimizes the coefficients of variations of the target variables at NUTS2 level and taking into account the related, desired sampling errors fixed ex ante. Therefore, the final stratification depends on the most correlated combinations of modalities of the stratification characteristics with the target variables. The stratification "criteria" vary according to the specificity of the country and NUTS2 territories. Finally, as the sample size corresponding to the optimized solution does not equalize the predetermined contractual amount of units to be selected for each country, in a second step this sample size is adjusted accordingly by decreasing or increasing proportionally the allocation in each stratum based on the difference between optimized and acceptable sample sizes.

Some of the master sample points had been excluded for the second phase sample taking into account the following accessibility criteria:

- Altitude;
- Distance to roads;
- Accessibility indicator from CORINE Land Cover (CLC);
- Rule for eligibility.

Concerning altitude, the points above 1,500 m were deemed difficult to reach. The second criterion is the distance to the closest road. The distance had been computed based on Tele-atlas road network. The road network generally excludes rural dirt roads used for the access to agricultural fields, usually good enough to allow the access of enumerators by car. All points in agricultural landscapes are regarded reachable thanks to dirt roads, although other obstacles may appear, such as private property delimited by fences.

For the criterion relative to accessibility CORINE Land Cover (CLC), agricultural areas are assumed rich in drivable dirt roads, in particular where there is a low density of paved roads. There is also an implicit assumption that the density of drivable dirt roads is much lower in other landscape types: forest, shrub, wetland, etc. To this end, CLC was split into two categories: potentially easy and difficult accessibility. Difficult accessibility includes forest, scrub, non-agricultural bare land, wetland and water.

Concerning the criterion based on the eligibility rule, the CLC-based accessibility was combined with distance to roads and altitude. The following thresholds are defined:

a) Points above 1,500 m and distant > 600 m from the closest roads or with an elevation change >100 m from the closest road.

b) Points below 1,500 m with a land cover type neighborhood (600 m circle) classified as potentially problematic accessibility (forest, shrub, water and wetland) and distant > 600 m from the closest roads or with an elevation change > 100 m from the closest road.

c) Points that would have been eligible with the general rules, but could not be reached in 2015 (OBS_TYPE = 3 or 4) and were considered non-eligible in 2015.

Categories a) b) and c) were merged in a set of strata to be treated with photo-interpretation. In addition, there were 6975 points excluded from the second phase sample: points with a stratification code equal to "transitional water[6]" and points outside the reference NUTS area.

The subsampling method used to determine the sample for the field survey is a systematic procedure with multiple ranked replicates that ensure a certain spatial homogeneity in the distribution. The rule of having a minimum of two sample points per stratum in each NUTS 2 had been applied, unless there were not enough points in the master sample.

3.1.2.1. Optimization of the sample

In general, a sample could be defined as optimal in terms of both its costs (i.e. the number of units to be interviewed) and its accuracy (related to the sampling variance of target estimates).

In order to optimize a stratified sampling design of a given population of interest, its members must be assigned to groups (strata), which should be homogeneous with respect to the target variables, whose estimation is the aim of the survey. Simple random sampling could then be applied within each stratum, having defined the overall allocation, i.e. the number of units to be selected in each stratum (Cochran, 1977). The allocation is in general proportional to the variability of target variables in strata (Neyman, 1934).

Many studies dealing with the problem of stratified sample design optimization have been conducted; a general review of the proposed methods is contained in Gonzales (2010). From a global view, optimization of stratified sampling has been considered as a two-step process: first, a stratification is chosen by exploiting all the auxiliary information available on sampling units, or only a subset, selected on the basis of known correlations between target and stratification variables. Then, given the chosen stratification, the problem of allocation is solved (Dalenius and Hodges, 1959).

Well-known solutions in the multivariate case (more than one target variable) are the ones given by Bethel (1985, 1989) and Chromy (1987). Together with many others, these solutions assume that stratification of population is given.

The approach followed in the optimization process of LUCAS sampling design is based on the joint determination of the optimal stratification of a sampling frame, together with the optimal sample size determination and allocation. This approach is the most general one, as it can operate in the full multivariate case (i.e. concerning both stratification and target variables), without being obliged to choose the number of strata. Its implementation is based on the use of the genetic algorithm. The general procedure had been

[6] These areas correspond to what is defined in the water framework directive (Directive 2000/60/EC) and refer to bodies of surface water near river mouths which are partly saline in character because of their proximity to coastal waters but which are substantially influenced by freshwater flows. They also include water surfaces in estuaries (the wide portion of rivers at their mouths subject to the influence of the sea into which the water course flows) and lagoons (water areas cut off from the sea by coastal banks or other forms of relief with, however, certain possible openings). These areas are not part of the NUTS definition and therefore excluded from the LUCAS reference area.

implemented in an R statistical package named SamplingStrata, which is available on the CRAN (Barcaroli et al, 2018).

The optimization of the sampling design starts by making the sampling frame available, defining the target estimates of the survey and establishing the precision constraints on them. It is then possible to determine the best stratification and the optimal allocation. Finally, the selection of the sample can be carried out. When formalizing the description above, these are the required steps:

1. *analysis of the frame data*: identification of available auxiliary information,
2. *construction of atomic strata*: on the basis of the categorical auxiliary variables available in the sampling frame, a set of strata can be constructed by calculating the Cartesian product of the values of all the auxiliary variables and assigning the information on the distributions of the target variables (means and standard deviations) to each stratum;
3. *choice of the precision constraints* for each target estimate, possibly differentiated by domain,
4. *optimization of stratification* and *determination of required sample size and allocation* in order to satisfy precision constraints on target estimates,
5. *adjustment of the final sampling size*,
6. *selection of units* from the sampling frame with a stratified random sample selection scheme,
7. *evaluation of the found optimal solution* in terms of expected precision.

3.1.2.2. The stratification of the sample

In 2015 survey, the sample was stratified considering the variables NUTS2 and the stratum variable STR05 by country. Therefore, the number of strata was ex ante fixed and it was given by the Cartesian product (combinations) of the number of NUTS2 by all the available modalities of STR05 in each region. In this schema, furthermore, the regions (NUTS2) are the minimum territorial study domain.

In 2018 survey, the strata had been identified by an iterative optimization algorithm that, starting from the “atomic strata”, aggregates them considering the coefficient of variations of the target variables and the related desired sampling errors. The optimization is carried out distinctly for each value of NUTS2 domain, and then aggregating the results at country level. For each NUTS2 value, the Cartesian product of STR18, CLC and ELEV classifications gives the atomic strata. As ELEV is a continuous variable, a preliminary step of categorization has been performed, utilizing the K-means algorithm to produce four distinct classes for this variable. The coefficients of variations are related to the estimates of the 16 target variables, whose values have been previously predicted for each point of the master by a logistic model. The iterative algorithm optimizes the stratification, aggregating the atomic strata with the aim of minimizing the overall sample size required to fulfil the precision constraints (the CVs of the target variables). Therefore, the stratification is not produced by a fixed combination of variables but it depends on the most correlated combinations of modalities of the stratification characteristics with the target variables; the stratification “criteria” vary according to the specificity of the country and of the NUTS2 territories, which are assumed to be, as in 2015 survey, the minimum territorial study domain.

In the following table, a comparison between the actual number of strata and the one obtained only considering the combinations of NUTS2 and STR18 (instead of STR05 as in 2015, that is using the same 2015 criteria) is given at country level.

Table 6: Number of strata according to the 2018 actual stratification and the hypothetical one obtained using the 2015 criteria

Country	Strata number in Actual stratification (a)	Strata number in Hypothetical stratification (b)	ratio : (a)/(b)
Belgium	400	82	4.9
Bulgaria	586	48	12.2
Czech Republic	526	63	8.3
Denmark	406	40	10.2
Germany	1 930	294	6.6
Estonia	147	8	18.4
Ireland	251	16	15.7
Greece	1 374	104	13.2
Spain	2 172	128	17.0
France	1 920	176	10.9
Croatia	236	16	14.8
Italy	1 967	167	11.8
Latvia	171	8	21.4
Lithuania	189	8	23.6
Luxembourg	55	8	6.9
Hungary	568	55	10.3
Netherlands	418	89	4.7
Austria	558	70	8.0
Poland	1 279	128	10.0
Portugal	725	40	18.1
Romania	841	64	13.1
Slovenia	152	15	10.1
Slovakia	284	32	8.9
Finland	372	39	9.5
Sweden	657	62	10.6
United Kingdom	1 776	284	6.3
Total	19 962	2 060	9.7

Note: Cyprus and Malta are not reported because they are entirely collected

As the table shows, the number of strata in the actual stratification is higher (about 10 times) than the number of the hypothetical one (implemented by the same criteria of 2015 survey).

3.1.2.3. Adjustment of the final sampling size

After the optimization step, the final sample size is the result of the allocation of units in optimized strata. This allocation is such that the precision constraints are expected to be satisfied.

Actually, three possible situations may occur:

1. The resulting sample size is acceptable. In this case, no action is required.
2. The resulting sample size is too high, it is not compatible with the available budget,

3. The resulting sample size is too low; the available budget permits an increase in the number of units.

Whenever the sample size corresponding to the optimized solution does not equalize the predetermined contractual amount of units to be selected in each country, this sample size is adjusted in a second step by proportionally varying the allocation in each stratum (by decreasing or increasing it accordingly to the sign of the difference between adjusted and optimal sample sizes). The function “adjustSize” permits to obtain the desired final sample size. These differences are reported in the following table.

Table 7: Contractual, optimal and adjusted sample size by country

Country	Points in Master	Contractual Sample size	Optimal sample size	Adjusted sample size	(Adjusted/Optimal) / Optimal (%)
Belgium	7 673	3 659	5 522	3 659	-33.7
Bulgaria	27 731	7 680	13 512	7 680	-43.2
Czechia	19 716	5 713	10 069	5 713	-43.3
Denmark	10 771	3 703	6 422	3 703	-42.3
Germany	89 399	26 777	50 196	26 777	-46.7
Estonia	11 322	2 665	2 874	2 665	-7.3
Ireland	17 399	4 975	7 206	4 975	-31.0
Greece	32 817	12 622	13 388	12 622	-5.7
Spain	124 543	45 314	40 016	45 314	13.2
France	137 047	48 215	61 786	48 215	-22.0
Croatia	14 141	4 239	6 835	4 239	-38.0
Italy	75 034	28 294	36 338	28 294	-22.1
Latvia	16 135	5 376	2 695	5 376	99.5
Lithuania	16 234	4 584	4 685.7	4 584	-2.2
Luxembourg	644	340	463	340	-26.6
Hungary	23 267	5 513	11 824	5 513	-53.4
Netherlands	8 882	5 011	5 837	5 011	-14.2
Austria	20 982	8 840	8 509	8 840	3.9
Poland	77 964	23 086	32 265	23 086	-28.4
Portugal	22 144	7 168	9 377	7 168	-23.6
Romania	59 558	16 723	16 828	16 723	-0.6
Slovenia	5 064	1 923	2 252	1 923	-14.6
Slovakia	12 265	2 898	5 711	2 898	-49.3
Finland	84 316	16 182	9 279	16 182	74.4
Sweden	112 385	26 709	20 197	26 709	32.2
UK	61 038	17 253	36 260	17 253	-52.4

Note: Contractual size: the size fixed before the running of the procedure to assign the batches;
Optimal size: the size calculated by the procedure based on CVs of estimated target variables;
Adjusted size: the calculated size normalized with the contractual ones;

Given the same sample size (and related percentages of PI and direct data collection) and the way the PI are chosen, the strata are in average smaller and the sample units are much more spread and mixed (PI and direct data collection) over the countries. In the following figures, Italy is reported as an example of the distribution of samples.

Figure 2: Distribution of sample units in 2015 survey – Italy

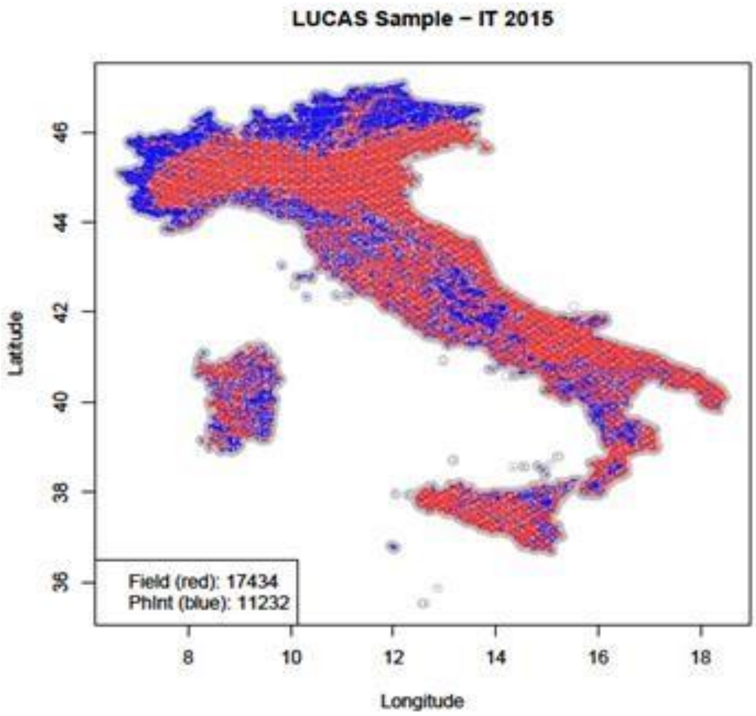
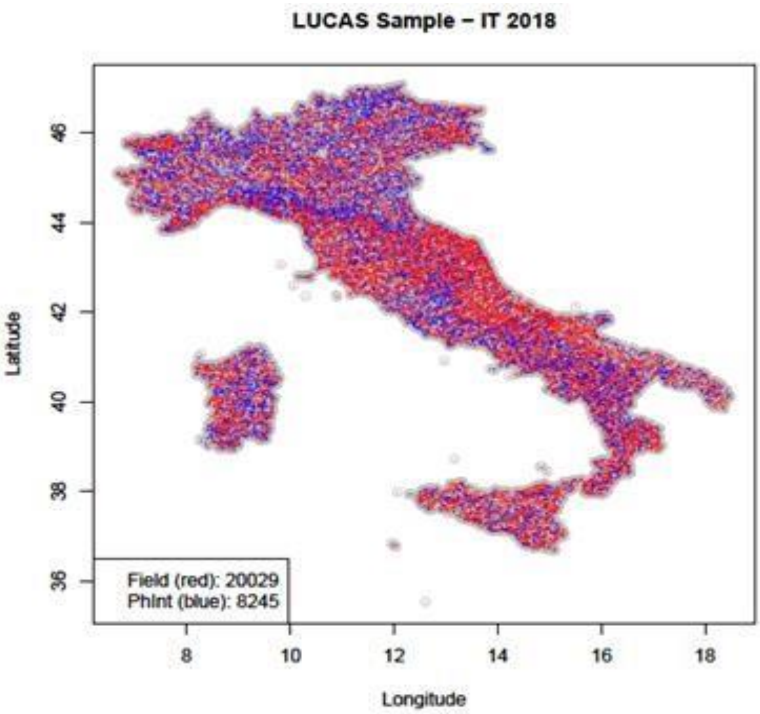


Figure 3: Distribution of sample units in 2018 survey – Italy



3.2. Differences in 2018 methodology design from 2015

The basic scheme in 2018 is essentially the same as the previous survey but some changes have been implemented in the Master and in the second phase sample. In this chapter, the main differences between 2015 and 2018 sample designs of LUCAS survey are described with regards:

1. the new stratification variable in Master 2018,
2. eligibility and photo interpretation,
3. the use of photo interpretation,
4. the stratification of the second phase sample,
5. The calculation of the sample size and the allocation of the sampling units.

In 2015 survey the sample size was calculated according to the requested precision at level NUTS1 for the more important modalities of land cover (see table below) on the basis of the previous survey results. Once the sample size was fixed, in every country the allocation of the sample units in the strata (identified as the combinations of regions by the STR05 variable available in the Master) was more or less proportional to the strata population because the points were taken by a systematic selection (excluding strata related to smaller subpopulations).

In 2018 survey, given the budget and the timing of the contractual steps to assign the batches, a calculation was made in order to confirm the 2015 sizes in terms of direct and PI data collection; these data states the constraints to be respected in the final allocation of sampling units by country.

Table 8: 2015 Survey-requested expected accuracy (relative error) by different land cover modalities at NUTS 0 level

Land cover class	Relative error	Land cover class	Relative error
A	0.15	**C**	0.15
B	0.15	**C1**	0.2
B1	0.15	**C2**	0.2
B2	0.25	**C3**	0.2
B3	0.25	**D**	0.02
B5	0.25	**E**	0.075
B7	0.25	**F**	0.2
		G	0.2

As soon as sampling precisions have been set *(table 9)* as ”desired” target precision at region level, they were used, by the same procedure for optimizing the stratification, to calculate the desired sample size, according to the desired accuracy, and to allocate the number of units in the strata of each country. According to the sample size in every strata, the sampling units had been selected from the corresponding population in the strata by a simple random selection procedure.

Table 9: 2018 survey-desired accuracy (relative error) by different land cover modalities at territorial level NUTS1/NUTS2

Land cover code	Land cover	Relative error (%)
A10	ROOFED BUILT-UP AREAS	15
A20	ARTIFICIAL NON-BUILT UP AREAS	15
B10	CEREALS	15
B2-B5	ROOT, NON-PERMANENT INDUSTRIAL CROPS, DRY PULSES, VEGETABLES AND FLOWERS, FODDER CROPS	20
B7-B8	PERMANENT CROP	20
C10	BROADLEAVED WOODLAND	20
C20	CONIFEROUS WOODLAND	20
C30	MIXED WOODLAND	20
D10	SHRUB LAND WITH SPARSE TREE COVER	20
D20	SHRUB LAND WITHOUT TREE COVER	20
E10	GRASSLAND WITH SPARSE TREE/SHRUB COVER	15
E20	GRASSLAND WITHOUT TREE/SHRUB COVER	15
E30	SPONTANEOUSLY RE-VEGETATED SURFACES	20
F00	BARE LAND AND LICHENS/MOSS	20
G00	WATER AREAS	20
H00	WETLANDS	20

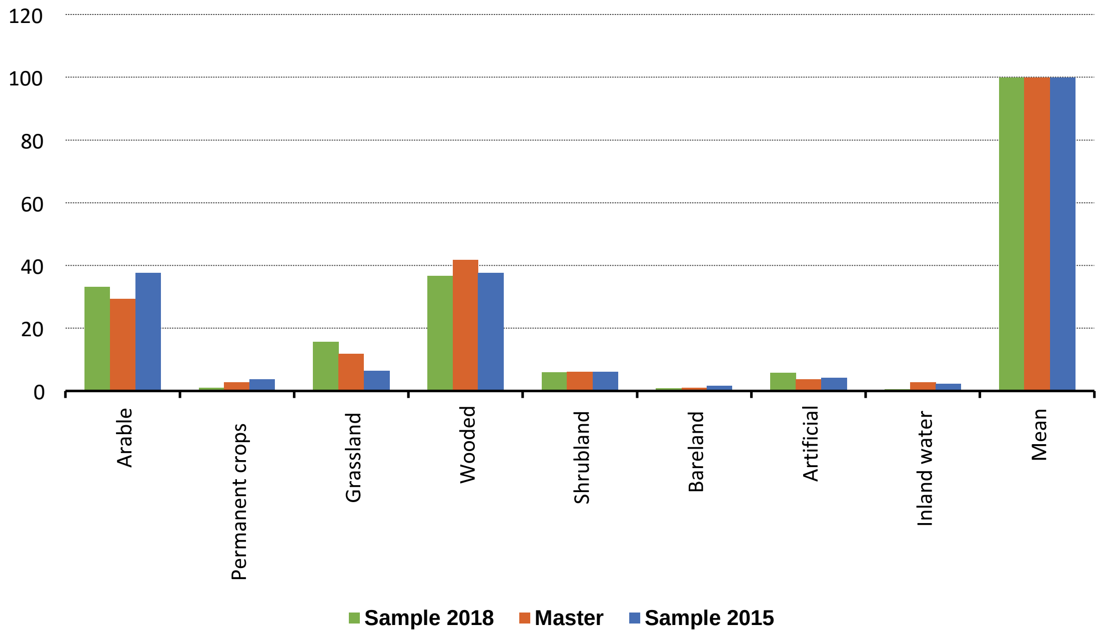
The land cover modalities had been estimated for each point in the Master from the previous LUCAS surveys by a statistical model. Therefore, their values were not the observed ones but the predicted values, the sampling errors and the generated sample size are hypothetical ones given under the condition that the statistical model is adequate.

Summarizing, while the sample size at level of country has been fixed with the same procedure in 2015 and 2018 surveys, the allocation of the units in strata is quite different in the two LUCAS occasions. In 2015 survey, the units have been allocated more or less proportionally while in 2018 the optimization algorithm provided the allocation. Moreover, the selection procedure in 2015 is systematic while in 2018 survey the units were selected by a simple random sample (SRS) procedure.

The effects of the different techniques of sample allocation are described by the following graphs that compare the distributions of the 2015 and 2018 sample units with the point of Master with regard to the main characteristics. Both the collected points by field operations and by photo-interpretation in the office ex ante are considered. The differences depend on various factors: stratification criteria, the allocation procedure used, the estimated variability of the target variables, the use of PI points etc.

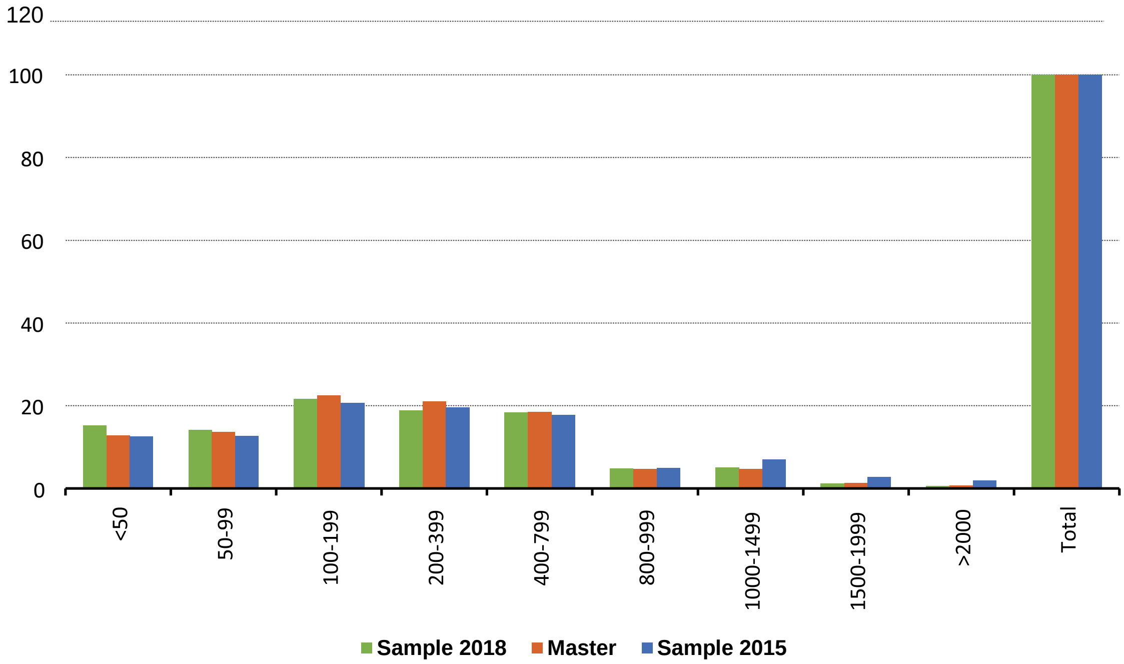
In first figure are reported the percentage distributions of points in the 2015 and 2018 samples and in the Master by the stratum variable STR18. For “arable land” and “artificial”, the two samples are over-represented with respect to the Master while for “wooded area” and “inland water” the percentage in the Master is higher than in the two samples. For “permanent crops” and “bareland” the points in 2015 sample are more than in Master and in Sample 2018 while the vice versa holds for “grassland”.

Figure 4: Percentage distributions of points in Master and in 2015/2018 samples by STR18 (Percentage)



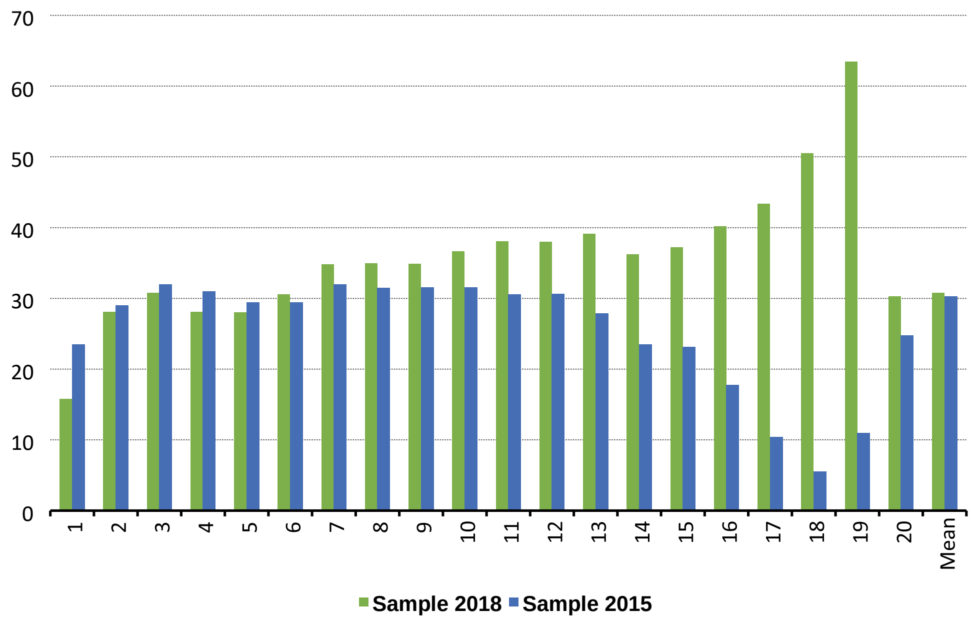
The distribution of points by class elevation is depicted in the following figure. Because the different procedures in using the PI points and the introduction of the probability of change, in 2015 sample the percentages of points with an elevation more than 1 000 m are higher than in Master and 2018 sample while the points less than 100 m are more present in sample 2018. For elevation between 400 and 1 000 m the percentage allocation of two samples is the same as in Master while small differences are found in the intermediate elevation (100-400 m).

Figure 5: Percentage distributions of points in Master and in 2015/2018 samples by elevation (Percentage)



In the next figure the ratios between the number of points in the two samples and the points in the Master belonging to the same class of probability of change (that has been estimated for all the point of the Master) is reported. The ratios have the meaning of the “coverage” of a sample with respect to its frame. The coverage of 2015 sample is slightly greater for the lower classes of probability to change while the points with higher probability to change are much more present in 2018 sample.

Figure 6: Ratios between the number of points in 2015/2018 samples and in the Master (The probability to change classes are obtained dividing the interval 0-1 by 20) (Percentage)



3.3. Data Collection – LUCAS in the field

LUCAS survey could be regarded as a multimode survey since 2015, comprising information collected both in the field and by Photo Interpretation (PI) in the office. In the 2018 LUCAS, a total of almost 338 000 points were selected for the second phase sample including about 240 000 for the in situ data collection and 98 000 to be photo-interpreted in the office. As mentioned, the in situ survey not only collects data on land cover and land use but also includes:

1. An extended soil module where a topsoil sample is collected on a maximum of 26 014 points. Out of these points, some 9 000 points will be evaluated for bulk density (this evaluation is done by the surveyor). On 1 000 locations out of these 9 000 points, a sample for assessing soil biodiversity is also to be taken. Additionally, on 1 470 points, the depth of the organic horizon is to be measured by the surveyor (up to 40 cm),
2. a test module for grassland on a maximum of 3 734 points,
3. An additional observation on 94 013 points for the Copernicus programme.

LUCAS as an in situ survey has been chiefly transacted in the field. Each point belonging to the field sample is investigated by collecting a set of detailed information using a specific field form (LUCAS 2018 – Technical reference document C2 - Field Form and Ground Document (template)) with the guidance of comprehensive instructions for surveyors (LUCAS 2018 – Technical reference document C1 - Instructions for Surveyors LUCAS 2018 - Instructions). Surveyors receive training before going into the field: they have a set of supporting documents, instructions on how to carry out the survey, and a set of quality control procedures. In the field, the surveyor classifies the land cover and the visible land use according to the harmonized LUCAS Survey land cover and land use classifications. Landscape pictures are taken in the four cardinal directions. The surveyor also collects information relating to the percentage of land cover within a specific window of observation, the area size, the width of any specific features, the height of any trees, as well as information on land and water management (for example, grazing or irrigation). In addition, surveyors carry out the implementation of LUCAS modules for landscape and linear features, grassland and soil. In addition to the obligatory fields, the surveyor can - and in specific situations has the obligation - to add comments and remarks. The information collected in the field can be grouped into the following categories:

1. Identification of the point
2. Access to point
3. Comments on the way to the point
4. Point observation
5. Land cover and land use
6. Land management, special status and special remarks on land cover/use
7. INSPIRE Pure Land Cover Classes
8. Water management on the field
9. Soil
10. Grassland module
11. Photo (minimum 6 pictures N, E, S, W (4 photos) close-up of crop (not on artificial or vegetation-free areas), point in context (to be able to relocate)

The ground document indicates the location of the LUCAS point. The point as drawn on this orthophoto is the reference for locating the LUCAS point in the field. This is the point on which information has to be collected. The LUCAS point location and the real position of the surveyor might not be identical.

While the information of GPS coordinates and precision are referring to the position of the surveyor doing the observation, the information on LC/LU, environmental information and the photos of the point and of the crop/cover have to refer to the LUCAS point itself as determined by the orthophoto, even if it is further away from the real position of the surveyor.

In LUCAS 2018 a collection on INSPIRE pure land cover classes was included. Data are collected for the points where LC1 is either woodland (CXX), shrub land (DXX), grassland (EXX) or bare land (FXX) and is assessed within the homogeneous plot inside the extended window of observation (20m radius). Unlike what happens in LUCAS classes, where the sum of percentage of combined land cover can be more than 100%, in this case the sum of INSPIRE classes must be 100%. Assessment of the percentages had been made using the “birds-eye” view.

Concerning water management, this is only relevant for points where LU = U111 or U112. In case of more than one source of irrigation or delivery system, the surveyor is requested to report the most important source.

It is mandatory that the surveyor does the anonymization directly before sending the photos to the upper level (i.e. the Regional or the Central Office). According to the LUCAS 2018 tender, noncompliance to this rule is considered a breach of contract and will lead to legal consequences.

3.4. Photo-interpretation

In the 2018 survey, data collection had been also carried out by photo-interpretation (PI) in the office. It is a cost effective method and this enables to enlarge the sample size under the budget constraints. Photo interpretation through satellite images or orthophotos consists of the most common method for other EU and World land cover observation. Photo-interpretation played an important role during the 2018 data collection. Access to points can be difficult in absence of adequate road network, for the landscape characteristics. The territory was classified in eligible and not eligible for the field survey, using all geographical information available. However, the exclusion of points from the sample is a likely source of bias, which had to be treated separately from the field survey. Therefore, the non-eligible excluded area needed to be covered with a complementary photo-interpretation. Photo interpretation was not included in LUCAS 2009 and 2012, but in the subsequent surveys (2015 and 2018) was used as a reasonable method in order to:

(i) Deal with the missing units found in field work and

(ii) Take into account unattainable units or units that are too costly to be reached.

In LUCAS, photo-interpretation is used substantially in two ways:

A. Points of the planned sample that are photo-interpreted in the office; up to LUCAS 2015 to cover all non-eligible points excluded from field survey, while in 2018, selected as sampling units in the whole EU territory, according to a change probability assigned a priori to all the points in the Master first phase.

B. Points that had to be assessed in the field and therefore were approached by a surveyor but revealed themselves not to be visible in the field (e.g. hidden by a high wall delimiting the property) and therefore had to be photo-interpreted in the field or points that had to be assessed in the field but were identified as impossible to be reached (e.g. military area) during the planning of the survey. Consequently, they were not approached by a surveyor but were directly photo-interpreted in the office ex-ante.

The points defined under (B) can be considered as “missing units” and the photo interpretation as a method to deal with them. Nevertheless, in this case their contribution to the final estimates is much reduced (e.g. for 2015 and 2018 surveys, the percentage of (B) points over the sample size was about 7%).

On the contrary, the points under (A) are part of the planned sample and their amount have been relevant in the previous surveys (e.g. about 21% in 2015 and 29% in 2018) and will be much more enlarged in following surveys.

Contrary to 2015 approach, in 2018 survey all the points were considered suitable to be selected and surveyed by either of twofold modes. The choice to assign one modality to a selected point had been done after the sample selection and not, as in the previous surveys, by dividing the Master into eligible and non-eligible points and hence proceeding to the selection step from these two subpopulations.

Besides, it is convenient as the probability of the point to change its land cover characteristics is low. Photointerpretation could, nevertheless, produce underestimates when assessing the changes of land cover characteristics, because the available photos had taken in a previous year. To avoid or to reduce the risk of biases, the use of photo-interpretation was limited to unchangeable points or those with a very low probability to change during the time between surveys.

Therefore, a point was considered in situ or photo-interpreted based on two indices calculated for all the points of the Master:

1. The reachability index
2. The propensity to change index.

The choice depends also on the constraints of the PI quotas in each country, fixed by the technical specifications of each contract.

3.4.1. The index of reachability

The index of reachability was introduced to represent the difficultness that an enumerator could encounter in reaching a given point. More precisely, it synthesizes the possibility that the point is far from a road, or on a cliff, etc. According to the variables in the Master data set, the following ones were considered useful in determining such index:

- The absolute difference in elevation between the altitude of the point and the one referred to the nearest road (ABS_RATIO),
- The distance to the nearest point on a road (NEARDIST),
- The angle to the nearest point in a road (NEARANGLE).

The index is obtained by combining these variables with proper coefficients, which were estimated by means of a Principal Component Analysis. Such statistical technique permits to obtain combinations of the active variables that took into account their correlation structure. These combinations, considered as net variables, are orthogonal (not correlated) between them.

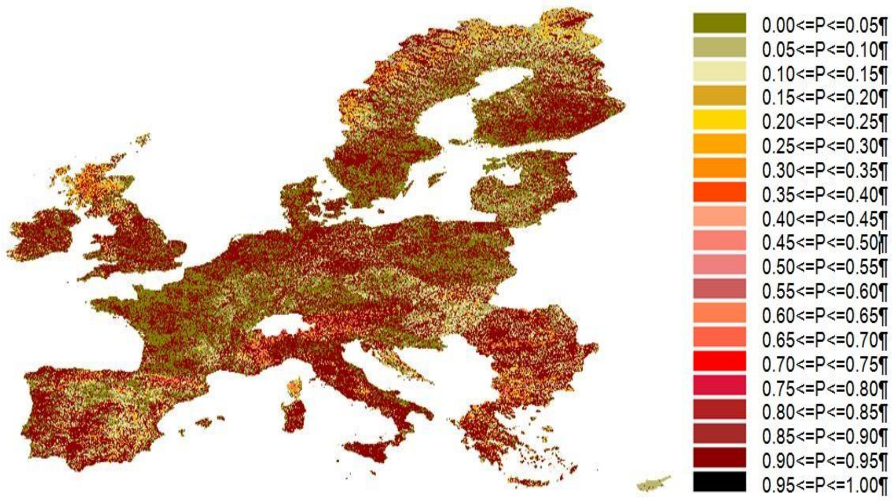
With this procedure, the index of reachability was built assuming higher values (0,1) for those points resulting more difficult for an in-situ visit. Moreover, special values had been added to this new variable in order to take into account those points that were previously observed as of difficult access.

In particular, the following conditions were considered (and for these the value of the index was imposed as 1):

- Previously considered as a point to be photo-interpreted;
- Value of the stratification variable specifying that the points should be photo-interpreted;
- Points with difficult access comment, or points that landowner refused access or points that landowner refused to collect SOIL data.

The next image represents the distribution of such index for all the points in the master data set.

Figure 7 Distribution of index reachability



3.4.2. The probability of change

Another additional information added to each record of the LUCAS master data set refers to the propensity of change in the estimated land cover. Such propensity also depends on the type of land cover that was associated to the point. For instance, it could be considered that the propensity to change for a point associated to an “Artificial land” should be less than the one associated to a “Crop” or “Grassland”.

To estimate such variable, a linear logistic regression model was introduced. The dependent variable was obtained by considering the results observed in the LUCAS surveys related to the years 2009, 2012 and 2015. In particular, it was supposed to have the same land cover if:

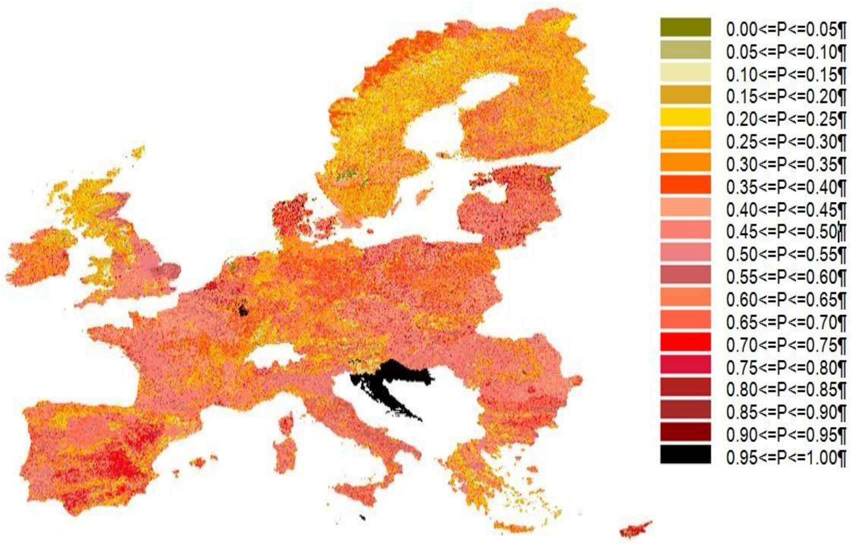
- The land cover in LUCAS 2015 was the same as the one observed in LUCAS 2012,
- The land cover in LUCAS 2015 was the same as observed in LUCAS 2009 (and the point was not observed in 2012),
- The land cover in LUCAS 2012 was the same as observed in LUCAS 2009 (and the point was not observed in 2015).

Instead, all the records observed in at least two LUCAS surveys were associated to a change in the land cover if any of the above conditions was not met.

The covariates of the linear logistic regression model were the same of those used when estimating the land cover, except for the characteristics of the satellite images which had not been considered, while the estimated land covers entered in the model as independent variables.

It has to be noted that the estimated score is related to the “not change” in land cover. The results of this model are analyzed by considering the graphical representation of this probability as distributed in the European countries (Figure 8).

Figure 8: Probability not to change in the European countries (Probability)



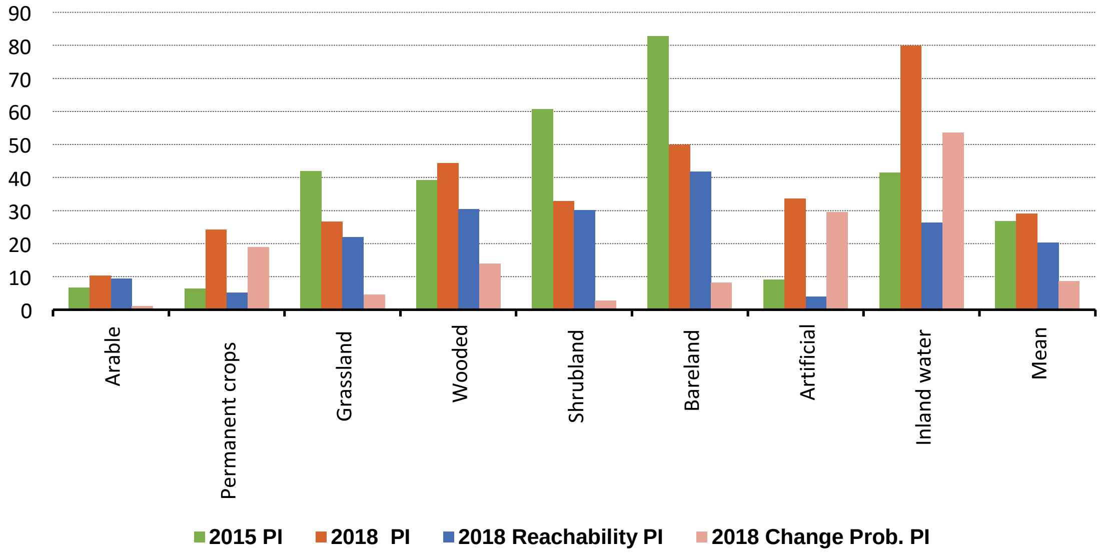
3.4.3. The implications of new approach of Photointerpretation in 2018 survey

The two different approaches in using photo interpretation followed in 2015 and 2018 surveys, had resulted in different distributions of PI points in the two samples. In the following graphs, the percentages’ ratios of PI points over the totals of two structural characteristics (STR18 variable and elevation class) are depicted. The features of probability of change and the reachability are reported as well.

The first figure reports the ratios by STR18. By average the percentage of PI in the two samples are similar (about 29% and 27% respectively) but the modalities of STR18 show different figures.

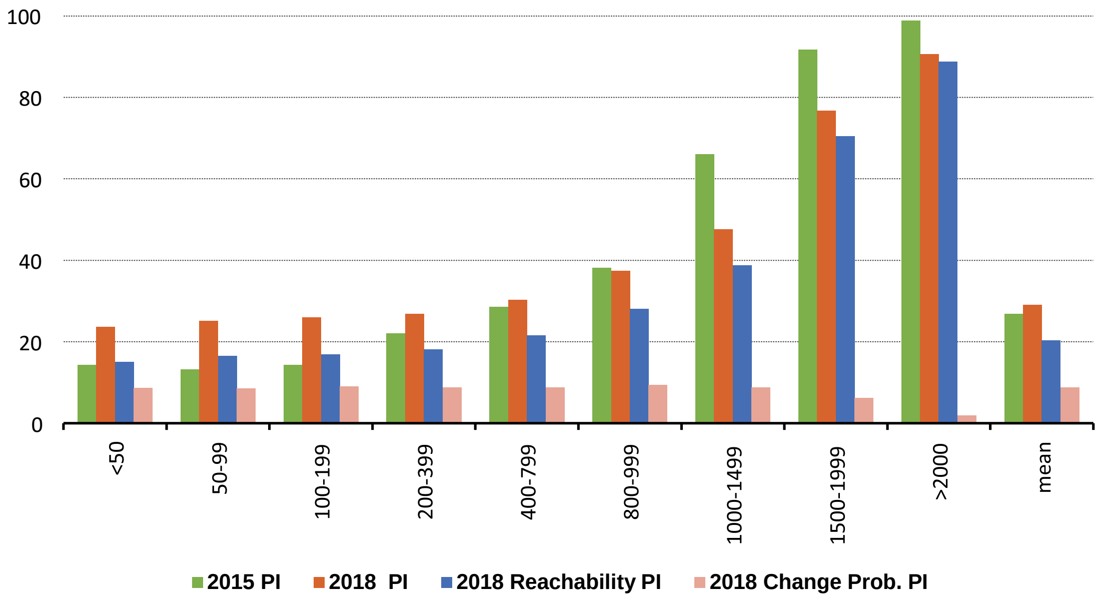
The STR18 modalities can be divided in two groups, according to the greater differences showed by the graph; in the first group - permanent crops, artificial and inland water - the percentage of PI in sample 2018 are higher than in the 2015 one, while in the second group - grassland, shrubland and bareland - the vice versa holds. The figures in the two groups are correlated to the different composition of PI_prob and PI_reach; in the first group the PI_Prob are more than the PI_Reach while the contrary occurs in the second group.

Figure 9: Percentage ratios of PI points over totals by STR18 - 2015 and 2018 samples (Percentage)



In next figure, the same analysis is carried out for the variable class of elevation. The ratios in 2018 sample are higher than the ones obtained in 2015 survey up to 800 meters of elevation and then the trend inverts. Below that threshold, the ratios are substantially steady in both samples while above they strongly increase. This pattern for 2018 depends on the PI assigned because of the reachability index while the ratios for the PI assigned from the probability to change were slightly decreasing.

Figure 10: Percentage ratios of PI points over totals by class of elevation - 2015 and 2018 samples (Percentage)



4 Quality Management

4. Quality Management

4.1. Quality assurance

Quality assurance is a central component throughout all the phases of the LUCAS survey to assure the quality and the comparability of results. Quality assurance includes a common framework or harmonized approach, automated quality controls implemented in with common IT tools.

Quality assurance covers different aspects, starting from the provision of a common framework for all participants. This is especially important as the survey has been split up in several Lots, which have been contracted to different entities and a common understanding across the lots needs to be assured.

To this end, the following actions have been foreseen:

- Common documentation and instructions for all surveyors;
- Common “Frequently Asked Questions and Answers” document updated regularly based on issues raised by the contractors during the running of the survey;
- Standardized and automated Data Management Tool (DMT);
- Common training for all the Survey Managers;
- Common set-up and follow-up visit to each country by a team of experts.

4.2. The Data Management Tool 2018 (DMT) for the Standardization and computerization of the main phases of the data management

The LUCAS data collection process aims at collecting raw micro-data (e.g. tabular data, pictures and GPS tracks) at geo-referenced points belonging to a representative sample. The volume of these datasets is quite considerable and requires specific tools to manage transmission, editing, storage, etc. Due to these specificities, the standardization and computerization of the phases of the LUCAS data production process was reinforced with the development of an ad-hoc IT tool, named Data Management Tool (DMT). Data collection and validation of internal consistency is assured through DMT, which is also linked to the visual quality control, including acceptance and rejection of points. Working on the local client requires download and upload of data and files from and to the central database.

The DMT records the data and analyses the quality of the recorded values through an automatic quality control. All the information collected for one point (i.e. photos, ground documents, GPS-tracks on map, and data) should present compatibles values, both in the combinations of the land cover/use, and for their values when the same point was observed in different rounds. For example, it is unlikely to observe an urban area changing into an agricultural field. Moreover, there are combinations of land covers/uses that are not allowed or cannot be consistent. The Automatic Quality Control (AutoQC) business rules included in the DMT, render impossible the entry of incoherent values, requesting explanation for the registration of unlikely values. By this way, the rules perform automatic consistency and logical checks at the first stage of data entry in order to guarantee coherence and good quality of the data.

The DMT provides support in all the phases of the survey with the following main modules:

1. Point management: Data Entry Tool (supervised data entry, consistency and ranges check);
2. Data Import (sent forward by one-step lower level or sent backward by one-step higher level);
3. Data export;
4. Point assignment;
5. Report builder;
6. Language choice

4.3. Quality control during the field work

Data quality checks run parallel with the data collection. The goal is to identify and correct systematic errors during the data collection as early as possible. The results collected by the surveyors are subject to a detailed, quality check. A hierarchical control structure has been set up to serve this purpose. Depending on the country size, it could include up to five levels of control:

1. Surveyor (SU) or Photo interpreter (PI)
2. Supervisor
3. Central or Regional Offices (CO/RO)
4. External Quality Control (XQC)
5. Eurostat (ESTAT)

At each step, data are checked before being forwarded to the next level. The first level is (verifying completeness and consistency) is carried out either during the compilation phase or when the data collected in the field are uploaded to the central data repository. The automated control takes place through the DMT application.

A second level of quality controls is carried out at the regional or central offices, where all of the surveyed points were visually checked. Central Offices and project managers receive training directly from Eurostat in Luxembourg covering the overall approach, the survey instructions and the Data Management Tool - as well as a field trip to allow for hands-on experience.

An external company performs data quality check on about 40% of the points. Both automatic and manual controls were applied. The quality control includes:

- interactive control of accuracy and compliance to the quality requirements as defined in the LUCAS framework;
- The first 20 % of points assigned to a surveyor are controlled in their entirety to detect early on any systematic errors being made.

All available information (ancillary information, ground documents, metadata on the survey, land cover and land use classification, transect data, GPS tracks, photos, justification for photo-interpretation) is analyzed to evaluate the reliability of the results. Point data that clearly requires correction or clarification is rejected and send back to the fieldwork contractors, while the other points are transmitted to Eurostat. After a revision by the fieldwork contractors of the points rejected, these points go once more to external quality control.

The second control of the data can lead to acceptance or rejection. In both cases, the data was forwarded to Eurostat, where points rejected twice are checked to guarantee the compliance with the tender specifications.

Detailed reports were delivered to Eurostat on a weekly basis and quick feedback provided when needed. A continuous help-desk was assured by the LUCAS team and by the JRC-soil team to the contractors. A FAQ list has been continuously updated and circulated to support the various actors and provide additional training on specific issues.

5 Post Processing

5. Post Processing

5.1. Post-data collection process

Post data collection involve additional quality check of individual data points as part of the correction of 2018 LUCAS data. Eurostat performs a number of macro and micro editing techniques in order to fine-tune the final estimates. The identification of possible influent errors might be fed into the validation process and imply further corrections to the micro data.

Eurostat control first includes the consolidation of the “raw” data set. Further steps of the validation process are associated for example the consistency checks with other datasets of the same domain (previous years LUCAS data) and consistency with data of other providers.

The activity was implemented by visual checks on sample points, cross-sectional checks (derived from the DMT validation rules) taken over a procedure implemented in SAS language, and validation of the corrections performed.

5.1.1. Cleaning of Microdata

At an early stage of the post-processing phase a first process of cleaning the microdata, it was taken care to apply a first set of corrections to the database starting with the correction of all “No values”. The microdata sometimes contained no value for a certain feature like Special remarks and in other cases “8” which stands for “not applicable”. All the empty values were then changed to “8” in this case. This procedure was repeated accordingly for every feature where it was possible.

5.1.2. Visual checks

The aim of the visual checks of individual data points is to provide results in the form of concrete corrections. These corrections can be manual corrections for single individual points, or automatic corrections for a group of points.

The first checks for coherence were the main source for defining the scope of the visual checks together with the matrices of LC/LU combinations, where unlikely changes and combinations can be identified.

Finally, the exact samples of points for different categories were calculated after having defined the scope of the visual checks. For this purpose, the 2018 database and/or the joint databases of 2015 and 2018 (i.e. panel points) were queried resulting in a list of point identifiers (IDs) which represent the amount of points connected to a certain category.

The visual checks also focused on points that showed a change in direction of observation and at the same time an unlikely LC change, which was chosen to be a LC that changed from artificial in one campaign to non-artificial in another and vice versa. All these points are located on or near a linear feature or border of two LC and it was observed that a general issue is the replication of decisions taken in former campaigns. This replication is important, because if it is not clear why a certain land cover was classified in previous survey it leads automatically to the observed unlikely changes.

The main parameters recorded for the survey, the potentially two LC and two LU as well as the corresponding percentages. Four different combinations of unlikely changes and the use of two different Land Use classes (U140 Mining / U361 Leisure) were examined. One combination was fishing and leisure, which is present in waterbodies and shows the problem to determine clear uses for lakes and rivers in general. Often this combination or only one of these both uses is coded for the same lake and in a panel these uses are likely to oscillate. Another combination checked the combinations for the use of energy production and it showed that it was often misused for electric lines that serve for energy transport instead. Mostly the use was correctly applied for water reservoirs / dams. Also the combinations for energy transport and for protection infrastructures with other land uses were examined and showed that these combinations are correctly applied. The checks for points with leisure and for mining showed as well that the classifications are fine and no systematic errors could be identified.

Another check for spontaneous vegetation changing to bare soil showed that almost all of these changes were correct and represented crop fields affected by crop rotation. Another check analyzed the changes from grassland to rocks and stones and most of these changes occurred in mountainous areas. The majority of these changes were mistakes that occurred by different photo interpretations between the years, even if the orthophotos showed the same situation.

The visual checks of unlikely LU changes dealt altogether with eleven different changes of LU of which alone six dealt with different aspects of abandoned LU. The so far only Abandoned LU class (U415) was extended to five different classes and more specific LU classes for LUCAS 2018 (U411-U415) and it was examined if they have been correctly classified. Checks were dealing with points that changed from unused/semi-natural (U420) to abandoned (U411-415) and from specific previous uses (Industrial, residential...) to the unspecific class of other abandoned (U415). It was observed that the abandoned and semi-natural classes were often used for very similar situations that produce incorrect/virtual changes and especially the use of abandoned classes is mostly unreasonable when signs of any previous use can be found.

Finally, individual points where Forestry Use changed to Agricultural Use were checked and most of the time a shift of the point location was responsible for them.

5.1.3. Correction and Validation

Data validation is important to ensure the data is clean, correct and coherent. The object of the validation process was to remove any inconsistencies, no valid values and blanks inside the database. The validation process consisted of the identification of not valid or inconsistent values and the correction of these values. The validation of LUCAS micro-data produced a list of errors that had to be corrected. These corrections were usually possible by automatically correcting unreasonable or impossible values, but for some cases, additional visual checks were needed to determine the correct values. This was for example the case for INSPIRE and FAO values that were not correct or missing. These values could not be corrected automatically by using a standard value, instead the correct values were identified by analyzing photos and ground documents.

Thus, visual checks and cross-sectional checks from statistical software had provided input for validation. It was selected over 100 different samples, which represented groups of LUCAS points that all share certain inconsistent or at least questionable criteria. Then data analysis followed of the individual LUCAS points, supported by a software tool, which provided the possibilities to view photos, ground documents and for panel points the recorded data of previous LUCAS campaigns.

The activity for validating the microdata included the application of all the rules from the automatic quality control of the Data Management Tool to identify the invalid values. The reason why this set of rules needed to be applied to the complete microdata again was that these rules changed during the survey and therefore not every point was recorded with the same and final set of rules. Besides, there is also a good reason why the aim, of preventing any incorrect or invalid value, cannot be completely ensured. Because it lies in the nature of the entered data that not all interdependencies of the entered values can be automatically checked. This has to be done by personally checking the values of the LUCAS points as it is done by the Regional and Central Offices, the external Quality Control during the production phase of the survey.

Therefore, the rules from the automatic quality were applied by using a validation tool that was developed in SAS software. It was based on the 722 quality controls that were included in the Data Management Tool of the LUCAS 2018 campaign. The validation process and the corresponding corrections were iterated until all issues were eliminated. The interdependencies of all the different values was the main reason for these newly appearing errors. This series of processes were carried out several times until only valid values were left. Finally, 1902 values of 1745 points that were identified during the validation process were corrected.

5.2. Calculation of variables

As soon as, the microdata were cleaned and validated then the calculation process of the estimates took place. That includes the calculation of settlement area, the FAO variable, the calibrated weights and the estimates production.

5.2.1. Settlement Area calculation

The relevance of settlement area is meaningful since urban population tends to wax over years. Nevertheless, certain demographic and lifestyle trends impeded efficient land use in urban areas. On this account, settlement areas are expanding more quickly than populations are growing and consequently there is a loss of land and ecosystem services, providing environmental challenges that would need to be encountered. LUCAS data for settlement serves as an important indicator of Sustainable Development Growth (SDG).

Considering LUCAS classification in greater detail, settlement area had been calculated as a result of the following land cover and land use values:

Land cover:

- A10 Roofed built-up areas including buildings and greenhouses;
- A20 Artificial non built-up areas including sealed area features, such as yards, farmyards, cemeteries, car parking areas etc. and linear features, such as streets, roads, railways, runways;
- A30 Other artificial areas including bridges and viaducts, mobile homes, solar panels, power plants, electrical substations, pipelines, water sewage plants, open dump sites;

Land use:

- U210 Energy production including areas used for production of electricity (including renewable energy), manufacturing of gas by purification, production of steam;
- U220 Industry and manufacturing including areas used for manufacturing of food, manufacturing of beverages and tobacco products, manufacturing of textile products, processing of coal, processing of oil and metal, production of non-metal mineral goods, industrial and manufacturing of chemical and related products, production of machinery and equipment, production of wood-based products and articles of cork and straw, printing of products such as newspapers and books, reproduction of recorded media such as compact discs, videos, software on discs or tapes, records etc.;

- U310 Transport, communication networks, storage, protection works (except U313 Water transport) including areas used for all types of railways, TGV traces, railway stations, streets, roads, highways, car parking, bus stations, tramways and tram stations, funiculars, airports, transport via pipelines, postal services and telecommunication infrastructures, logistics and storage of goods and warehousing, protection infrastructures against landslides or avalanches, dikes, electricity distribution, gas and thermal power distribution;
- U320 Water and waste management including areas used for water collection, water treatment and supply, sewerage, waste treatment;
- U330 Construction including areas used for construction of buildings and civil engineering works, specialized construction activities (e.g. demolition);
- U340 Commerce, financial, professional and information services including areas used for repair and installation of machinery and equipment, wholesale and retail trade, real estate activities, hotels and similar accommodation, food and beverage service activities, holiday and other short-stay accommodation (holiday apartment lots), camping grounds, recreational vehicle parks and trailer parks (also closed holiday camps), financial and insurance activities, professional activities, scientific and technical activities, administrative and support service activities;
- U350 Community services including areas used for public administration, local authorities, defense, education, health and social work, religion, as well as other services if provided through community services;
- U362 Sport including areas used for sport activities;
- U370 Residential including areas used for housing purpose.

5.2.2. FAO variable calculation

LUCAS data are also useful in order to estimate the forest area according to FAO classification. However, several aspects need to be tackled in the alignment of LUCAS and FAO classification for forest classes. First of all the differences in the semantic definition of LUCAS wooded areas and FAO forest definitions: if an area has > 10% of trees (excluding fruit trees in permanent crops) in LUCAS is labelled as "wooded area", but FAO takes this into account only if it is greater than 0.5 Ha.

In fact, variations in the definitions may cause inconsistencies when datasets are compared over time.

Data collection process during field campaigns could be also affected by errors that have an impact on forest areas (Woodland (C00)). The key elements and definitions for the forest classes used in LUCAS 2018 and in FAO (FRA 2015) are reported in the following tables respectively.

Table 10: Terms, definitions and remarks for the "woodland" class in LUCAS 2018 (Source Eurostat, 2018)

Term	Definition	Remark
Woodland (C00)	Areas covered by trees with a canopy of at least 10%. Also woody hedges and palm trees are included in this class	Height of trees at maturity and width of woody features have to be assessed.
		The 10% of canopy cover has to be assessed in the extended window of observation (Area 0.13 ha).
		If the wooded area is larger than 0.5 ha, the height of trees is above 5 m at maturity and the width of the wooded feature is more than 20 m, the surveyor has to indicate the forest cover code in the respective "LC plant species" field, according to the forest type classification of the European Environment Agency.
		Trees that are known as forest trees can also be grown as an orchard

Table 11: Terms and definitions of the FAO forestry-related classed

Term	Definition
Forest	Land spanning more than 0.5 hectares with trees higher than 5 meters and a canopy Cover of more than 10 percent, or trees able to reach these thresholds in situ. It does not include land that is predominantly under agricultural or urban land use.
Other wooded land	Land not defined as "Forest", spanning more than 0.5 hectares; with trees higher than 5 meters and a canopy cover of 5-10 percent, or trees able to reach these thresholds; or with a combined cover of shrubs, bushes and trees above 10 percent. It does not include land that is predominantly under agricultural or urban land use.
Other land	All land that is not classified as forest or other wooded land.
Other land with tree cover: sub-category of other land	Land considered as "other land", that is predominantly agricultural or urban lands use and with has patches of tree cover that span more than 0.5 hectares with a canopy cover of more than 10 percent of trees able to reach a height of 5 meters at maturity.

It was possible, therefore, to add a new variable to each LUCAS survey representing the FAO forestry classification, i.e. a new variable characterized by three values:

- 1: FAO forest
- 2: other wooded land
- 3: other land with tree cover.

This was done by considering different combinations of the values of the following variables:

- primary and secondary land cover (LC1 and LC2);
- primary and secondary land use (LU1 and LU2);
- species associated to LC1 (value of the variable LC1_Species);
- size of the area referred to LC1 (AREA_SIZE)

- variable referring to the TREE_HEIGHT;
- variable referring to the FEATURE_WIDTH;
- variable referring to the SURVEY_LC_LU_SPECIAL_REMARK.

All the procedure to derive the variable from combinations is given in a detailed syntax script in annex. The table contains the conditions (combination of the active variables) that were used to obtain the FAO forestry classification.

5.2.3. Weights calculation and Calibration

The weight of the single point is obtained, starting from the inverse of probability of selection, by an iterative proportional fitting (IPF) procedure that associates, in each iteration, new weights to each point up to equalize the sum of weights and the known totals of the domains to which the units belong. As soon as the 2018 sample units were selected, they have been automatically weighted as the inverse of inclusion probabilities. The sum of these weights provide the total population area.

The estimating procedure is based on a calibrated estimator. It assures that the estimates of some structural variables are forced to equalize "known totals" in some domains: other than in "administrative entities" (NUTS0, NUTS1 and NUTS2), also aggregated classifications of five elevation classes are taken into account. Certain number of points is equivalent to consider the "area", which it had been derived by multiplying the number of points by a constant, the averaged area in the NUTS2. Because it had been obtained by an external reliable source, the "known total areas" of NUTS2, NUTS1 and NUTS0 are "true".

The calibrated estimator also takes over the correction for missing units, where the "average collected point" is conceptually averaged taking into consideration the strata and the class of elevation at different level of NUTS area. The final-weighting procedure was based on the use of the statistical package, which allows calculating the sampling estimates by using calibration estimators. The weight assigned to each unit is obtained according to a procedure divided in several steps:

1. The "*starting weight*" of each sample unit, named "*direct weight"*, is calculated according to the sampling design, as the reciprocal of the inclusion probability;

2. The starting weight is adjusted in order to account for non-response, obtaining the "*base weight"*;

3. Correction factors of the base weight basis are computed to take into account equality constraints between some known parameters of the population and the corresponding sample estimates;

4. The *"final weight"* is obtained as the product between the base weight and the correction factors.

Steps 2 and 3 had not necessarily been carried out distinctly: if the non-response model is the same than that for the overall calibration, they can be executed jointly. The gist was to minimize the distance between the weight before and after the calibration phase ("Calibration Estimators in Survey Sampling", Deville and Sarndal, JASA, 1992).

A calibration estimator requires the definition of a calibration model, where the indicated variables with respect to which the known totals in the sampling frame are calculated. Apart from the known totals related to NUTS2 and elevation (5 classes), further integration was implemented with other known totals derived from Copernicus estimates (CORINE Land Cover and High Resolution Layers). Thus, the final model further included the following parameters from CLC and HRL at NUTS2 level:

- imperviousness
- artificial
- agricultural
- woodland
- wetland
- water

All these variables were added to each point in the Master, as binary variables: for instance, for variable "artificial" the value 1 indicates if the point is artificial, 0 is not. When the calibration procedure is invocated in the software application, it calculates the totals in the master by summing the area of all points having values equal to one.

5.2.4. Estimates Production

Having defined the calibration model and by using final (calibrated) weights, the statistical estimates had been obtained by considering the LUCAS microdata (values of the variables observed in sampled points) that had been previously checked and corrected where necessary. The estimates production process was implemented in R statistical package. The estimates refer:

- Country level: Land Cover (1, 2 and 3 digits), Land Use (1, 2 and 3 digits), at NUTS0, NUTS1 and NUTS2,
- Specific Land use units: Land Use in Heavy Environment Areas (LUD), Land Use for Services and Residential Area (LUE), at NUTS0, NUTS1 and NUTS2,
- FAO and Settlement Area estimates,
- EU level total: Land Cover (1, 2 and 3 digits), Land Use (1, 2 and 3 digits), FAO class and Settlement, for 23, 28 and 27 countries minus UK.

This was the standard calibration procedure adopted for the generality of countries. Hence, the focus was on the detection of non-plausible variations for some aggregates of land cover (i.e., "artificial" and "water") and settlement. Non-plausible variations have been accepted if they could be explained by sampling errors, that is, as a rule of thumb, when related confidence intervals were intersecting. Otherwise, a different ad hoc procedure was adopted, consisting in interpolating values of estimates in the four years (2009, 2012, 2015 and 2018) and constraining the anomalous values to assume the ones derived from the interpolation of the imperviousness values. This had been done in "Artificial" class for Czech Republic and Croatia.

An additional treatment was applied for Sweden and Slovenia in 2018: the value for "settlement" estimate was calculated by multiplying the value of settlement 2015 and the ratio between CLC2018 and CLC2015. An even more specific treatment was dedicated to Finland and Netherlands. In these countries, non-negligible deviations of the estimated total areas from known total areas for some NUTS 2 regions had been detected. That was due to the non-convergence of calibration procedure for the known totals of Copernicus variables, when there were no observed units in domains related to some of the known totals. As in the calibration model the NUTS2 totals were associated to Copernicus totals, this was the cause of the problem. For this reason, the calibration procedure was differentiated for these two countries, where the known totals of NUTS2 areas were considered also independently from Copernicus known totals. This ad hoc treatment allowed obtaining a full compliance to known totals of areas at NUTS2 level in these two countries. In other countries, due to conflicting known totals, some minor deviations remain.

6 Relevance

6. Relevance

6.1. Relevance, User Needs and Completeness

The LUCAS survey was initially conceived with the aim of providing early estimates for crop production but after some pilot tests, it became evident that the tool was not adequate for that purpose due to the time span until results were available. However, the results of LUCAS proved valuable for other uses and the scope and purpose were modified and broadened.

The needs of the Commission services related to LUCAS were assessed, confirmed and integrated at the strategic level by the LUCAS Advisory Group. They span from reporting obligations linked to the Common Agricultural Policy and to Rural development Policy, to implementation of the Kyoto Protocol, from the monitoring EU biodiversity policy to Soil Thematic Strategy and to the context of the implementation of Copernicus (formerly GMES) covering earth observation by means of satellites, ground based, sea-borne and airborne facilities in order to provide environmental information.

LUCAS data are used for Agro Environmental Indicators (AEI), LULUCF (land use, land use change and forestry) indicators, Europe Resource Efficiency indicators and are planned to be used in assessing the Good Agricultural and Environmental Condition (GAEC). Moreover, in the context of CORINE Land Cover (CLC) and all other pan-EU mapping initiatives, such as the Copernicus HRL (High Resolution Layers) LUCAS is used for production, verification and validation processes.

Four main types of users mainly use LUCAS data:

1. Eurostat internal use
2. Other DGs and European Institutions: mainly the JRC, EEA, RTD, AGRI, ENV, CLIMA, and GROW either directly by the DGs or through external contracts
3. For national purposes by national authorities
4. Research purposes by universities and research institutions

LUCAS also provides information for monitoring for a range of socio-environmental challenges, such as land take, soil degradation, environmental impact of agriculture or the degree of landscape fragmentation. More specifically data from LUCAS can be used to help analyze and contribute to the development of various EU policy areas such as:

<u>-Common Agricultural Policy:</u> Integrating environmental concerns into the Common Agricultural Policy,

<u>-Soil thematic strategy:</u> Protecting the soil, as detailed in the soil thematic strategy,

<u>-EU biodiversity strategy:</u> Promoting biodiversity and conservation, through the EU's biodiversity strategy

-Green Deal: Encouraging the efficient use of resources for sustainable growth, as in the resource-efficient Europe initiative,

-Copernicus: Land monitoring, spatial planning and resource management, as carried out by the Copernicus Earth Observation Programme,

-Climate change: Tackling climate change, through monitoring conducted by the European Environment Agency, as well as actions under the European climate change programme).

In addition, LUCAS data also provides a rich source of information for the research community, general public, business community, media, and international organizations. Requests for access to the LUCAS photos are regularly received. Micro-data is freely accessible and the access to it is not monitored.

An important issue in the user requirements is the timing of surveys. Commission services declared their need for the core part of the LUCAS survey to be carried out at three-year intervals, in synchronization with CORINE Land Cover and the update of the HRL (High Resolution Layers).

LUCAS use includes the microdata, the photos, the soil and the statistical tables produced by Eurostat from the microdata. In the Commission departments, the LUCAS micro data is particularly relevant for modeling as can be seen in the collection of use cases presented on the Eurostat website: http://ec.europa.eu/eurostat/web/lucas/publications/use-cases.

Concerning the satisfaction request there is no feedback analysis for this issue. On photo requests, feedback is always asked; nevertheless, there are no measures to determine user satisfaction.

The completeness rate is not applicable in LUCAS survey. According to the methodological design of survey (previous chapter) the sample is derived with the points need to be assessed. The missing data phenomena is almost negligible in the survey. In case surveyors could not reach the points, they were obliged to fill in the field form based on the information that he/she could collect from a relative distance. In different case, additional methods are used, such as photointerpretation or imputation.

7 Accuracy and Reliability

7. Accuracy and reliability

7.1. Overall accuracy

Accuracy of collected data has been evaluated by means of different indicators calculated in data collection and in data elaboration steps, generally analyzing the number of errors found (and corrected) and their impact on the target variables. In addition, the impact of data corrections on final estimates has been evaluated by considering the weights associated to the changes.

Accuracy of estimates has been evaluated by calculating their precision, expressed in terms of coefficients of variation, and comparing the obtained precisions with the ones set as constraints when designing the 2018 sample.

The accuracy is tackled at Eurostat level, by eliminating as much as possible non-sampling errors and by calculating sampling errors. The missing data phenomena is almost negligible in the survey. In case surveyors could not reach the points, they were obliged to fill in the field form based on the information that he/she could collect from a relative distance. The majority of points were surveyed at a distance lower than 100 m (85%), while a small percentage (9%) was photo-interpreted (PI) in the field due to accessibility problems (etc.) with the auxiliary use of the point's orthophoto.

Another important issue is the impact of the calibration procedure (see Post-processing chapter), adopted to re-weight data, on the final estimates, evaluated by comparing the latter with the ones obtained by applying a Horvitz-Thompson estimator. The comparison has been carried out at different geographical and LC/LU classifications.

7.2. Sampling error

In LUCAS survey, sampling error consists of the coefficients of variations of the estimates (CVs %). Sampling error had thus been calculated for each estimate and disseminated in the Statistical Tables for each domain of estimation.

As already mentioned, LUCAS is a two phase stratified survey. Therefore, the first phase (Master) is to be considered as the population area since its total points comprise all EU territory in a frame of 2*2 Km2 distance from one point to another. The second phase, sample to be surveyed, is derived from the Master and that signifies the representation rate of total population. The following table provides relevant information on the representation rate for each country in 2018 survey.

Table 12: Adjusted sample size by country

Country	Points in Master	Adjusted sample size
Belgium	7 673	3 659
Bulgaria	27 731	7 680
Czech Republic	19 716	5 713
Denmark	10 771	3 703
Germany	89 399	26 777
Estonia	11 322	2 665
Ireland	17 399	4 975
Greece	32 817	12 622
Spain	124 543	45 314
France	137 047	48 215
Croatia	14 141	4 239
Italy	75 034	28 294
Latvia	16 135	5 376
Lithuania	16 234	4 584
Luxembourg	644	340
Hungary	23 267	5 513
Netherlands	8 882	5 011
Austria	20 982	8 840
Poland	77 964	23 086
Portugal	22 144	7 168
Romania	59 558	16 723
Slovenia	5 064	1 923
Slovakia	12 265	2 898
Finland	84 316	16 182
Sweden	112 385	26 709
United Kingdom	61 038	17 253

7.2.1. Accuracy of Coefficients of variations

The coefficient of variations for each estimate as an additional unit of measure in LUCAS 2018 survey are provided online at https://ec.europa.eu/eurostat/web/lucas/data/database. Furthermore, an analysis of the precision levels obtained in 2018 LUCAS survey compared with those in 2015 has been carried out. Graphical results are reported depicting the range of coefficient of variation. The first set of graphs presents the CVs calculated for the Land Cover at first digit level, distinct for each country. The second set reports the distribution of the CVs of all the Land Cover at third digit level estimates produced for each country, in 2018 and 2015. The third set reports the same CVs calculated for the Land Cover at first digit level, but considering NUTS2 geographical level. The same sets have been produced also for Land Use, with the same characteristics (*see annex*).

Based on these graphs, it is possible to say that the precision levels obtained in 2018 are not sensibly different from those obtained in 2015, even if the correspondent sample designs are quite different. This is possibly because the 2018 sample design, though strongly optimized, as a side effect produced a high variability of inclusion probabilities, with a negative effect on sampling variance, thus compensating the gains obtained with the optimization. This factor has already been taken into account when designing the next LUCAS sample.

7.3. Non-sampling error

Conversely, non-sampling error could arise from human error, such as error in problem identification, method or procedure used. In LUCAS survey, thus, the non-sampling errors are associated with classification errors made by surveyors or photo interpreters during the collection process. Even though most of them were identified by the DMT Automatic rules, still misclassifications were possible. Among numerous reasons, we could refer the distance of observation, the quality of the orthophotos and certainly the possibility of a natural error by the enumerator. Data acquisition is a phase of survey process particularly important and sensitive. Its main characteristic is that once the data are collected it is not possible to come back and to collect again, contrary to other phases as, for example, data treatment where it is possible to revise and to adjust rules and algorithms.

In all the survey data (from DMT and post-phase), the number of points with at least one error (erroneous points) is 18571 and therefore the percentage of the erroneous point over the sample size is 5.5%. The total number of errors (failed rules) amounts to 22815 and so the average number of errors by the amount of erroneous points is equal to 1.23 while the overall average number of errors is 0.068. Most of erroneous points (81.3%) presents only one error while about the totality (99.7%) presents less than four errors per point.

In the following table the rule code, the meaning and the related percentage of errors are given; only the rules with a percentage more than 1% are reported. Most of the errors (93.1%) are due to only 10 rules and one third of them (31.6%) to only one; eight rules are related to the content of the survey (67.2%) while two rules (10011 and 10012) regard the modes of data collection (25.9%).

Table 13: Percentages of DMT failed rules and meaning

Rule	%	Meaning
10367	31.6	IF LC1 = D10 THEN MORE THAN 10% HAS TO BE DECLARED IN SHRUBS AND CONIFEROUS + BROADLEAVED EQUAL 5%
10012	13.2	End time in the limits fixed
10011	12.7	Start time in the limits fixed
200003	11.6	If there is no LC2 and LC1= Axx, Cxx, Dxx, Exx, Fxx, Hxx then LC% must be 100%
10370	7.6	IF LC1 = E20 OR E30 THEN MORE THAN 10% HAS TO BE DECLARED IN HERBACEOUS PLANTS AND CONIFEROUS + BROADLEAVED + SHRUBS UNDER OR EQUAL TO 5%
10368	7.0	IF LC1 = D20 THEN MORE THAN 10% HAS TO BE DECLARED IN SHRUBS AND CONIFEROUS + BROADLEAVED EQUAL 0%
10381	5.7	Water Management - Presence of water management not needed
10500	1.4	EUNIS COMPLEX needed when LC1/2 has trees
200008	1.2	Special remark 'Harvested field' is not needed if LC1 = Bxx
10371	1.1	If LC1 = F30 then more than 90% has to be declared in Lichens and mosses
Total	93.1	

The following table reports the observed points, the absolute number of errors and the number of erroneous points (that is the points with at least one error) by country.

Table 14: Number of sampled points, errors and erroneous points by country

Country	Number of points in survey	Number of errors	Number of erroneous points
AT	8840	493	362
BE	3659	146	58
BG	7680	637	559
CY	2313	452	379
CZ	5713	155	128
DE	26777	666	552

Country	Number of points in survey	Number of errors	Number of erroneous points
DK	3703	365	303
EE	2665	153	122
EL	12622	2030	1648
ES	45314	4270	3722
FI	16182	1703	1483
FR	48215	1308	1047
HR	4239	547	471
HU	5513	346	327
IE	4975	144	110
IT	28294	2837	1904
LT	4584	480	300
LU	340	4	4
LV	5376	224	154
MT	79	9	9
NL	5011	438	382
PL	23086	1159	815
PT	7168	766	676
RO	16723	704	564
SE	26709	1633	1506
SI	1923	49	48
SK	2898	149	131
UK	17253	948	807
EU	337854	22815	18571

7.3.1. Indicators on non-sampling errors

Based on the information of the above table, five indicators had been calculated and reported in the following table depicting the impact of errors per country.

Table 15: Indicators of impact of errors in the country

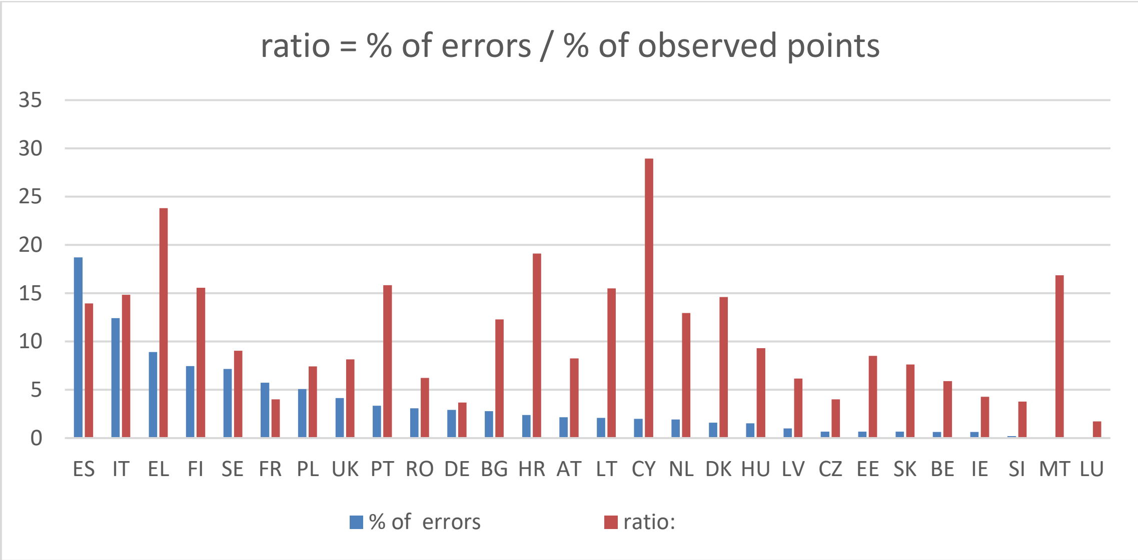
Country	Percentage of erroneous points over survey points	Percentage of points with more than 1 error over erroneous points	Percentage distribution of errors over total errors	Percentage distribution of sampled points	Ratios: (d) / (c)
	(a)	(b)	(c)	(d)	(e)
AT	4,1	26,8	2,2	2,6	0,8
BE	1,6	12,1	0,6	1,1	0,6
BG	7,3	9,3	2,8	2,3	1,2
CY	16,4	12,4	2,0	0,7	2,9
CZ	2,2	18,0	0,7	1,7	0,4
DE	2,1	13,9	2,9	7,9	0,4
DK	8,2	18,5	1,6	1,1	1,5
EE	4,6	24,6	0,7	0,8	0,9
EL	13,1	21,5	8,9	3,7	2,4
ES	8,2	13,4	18,7	13,4	1,4
FI	9,2	12,6	7,5	4,8	1,6
FR	2,2	4,7	5,7	14,3	0,4
HR	11,1	8,9	2,4	1,3	1,9
HU	5,9	4,3	1,5	1,6	0,9
IE	2,2	13,6	0,6	1,5	0,4
IT	6,7	45,7	12,4	8,4	1,5
LT	6,5	55,3	2,1	1,4	1,6

Country	Percentage of erroneous points over survey points	Percentage of points with more than 1 error over erroneous points	Percentage distribution of errors over total errors	Percentage distribution of sampled points	Ratios: (d) / (c)
LU	1,2	0,0	0,0	0,1	0,2
LV	2,9	42,2	1,0	1,6	0,6
MT	11,4	0,0	0,0	0,0	1,7
NL	7,6	9,9	1,9	1,5	1,3
PL	3,5	41,0	5,1	6,8	0,7
PT	9,4	12,3	3,4	2,1	1,6
RO	3,4	23,6	3,1	4,9	0,6
SE	5,6	8,0	7,2	7,9	0,9
SI	2,5	2,1	0,2	0,6	0,4
SK	4,5	8,4	0,7	0,9	0,8
UK	4,7	13,3	4,2	5,1	0,8
EU	5,5	18,7	100,0	100,0	1,0

The percentage distribution of errors (c) gives the contribution of every country to the total error in the survey. It partially depends on the size of the sample in the country whose distribution is reported in the column (d). The largest impact is due to Spain and Italy followed by Greece, Finland and Sweden while the smallest are from Luxemburg, Malta and Slovenia. However, a different picture is obtained considering also the ratios (e).

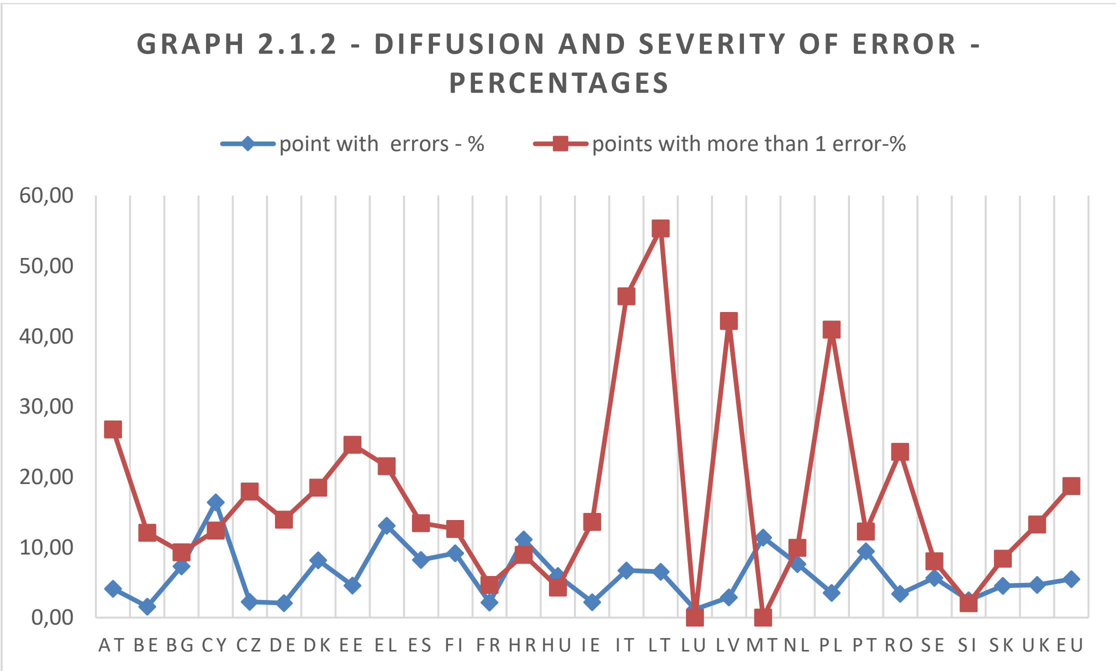
In the following figure the ratio is plotted together the percentages of errors, ranked by countries. The two indicators are reported with a different scale; in order to highlight the values of the ratio, it is multiplied by 10. Ratios depict the figures of the importance of errors in each country regardless their sizes. Under the hypothesis of a uniform distribution of errors (equal to the average percentage), the ratio should be equal to one. The two rankings are quite different because they express different points of view.

Figure 11: Ratio between the percentages of errors and observed points by country (Column C+E)



The indicator percentage of erroneous point (a) can be considered as “diffusion” indicators while the percentage of erroneous point with more than one error (b) could be taken as “severity” one; their distributions among the countries are quite different. The percentage of erroneous points varies from 1.2% for Luxemburg to the 16.4% for Cyprus; the higher percentages, over 10%, were found in Greece, Croatia and Malta. The higher percentages of points with a number of errors greater than 1 is reported for Italy, Lithuania, Latvia and Poland (more than 40%) while a second group (Estonia, Greece and Romania) presents percentages between 21% and 25%. The two indicators are compared in the following Figure.

Figure 12: Diffusion and Severity of Error - Percentages



7.4. Accuracy and Reliability among observation type

The variable “observation type” in LUCAS survey is a matter of great significance in terms of accuracy and reliability. The variable consists of seven modalities namely:

1-Field survey, point visible, <= 100 m

2-Field survey, point visible, >100 m to point

3-Photo-interpretation in the field

4-Point not visible. PI not possible

5-Out of national territory

6-Out of EU 28

7-Photo-interpretation in the office

The PI in office is ex ante established in the sampling design while the modes coded from 4 to 6 amount to about 40 points and that is a negligible percentage of the total points. Hence, concerning the analysis of data acquisition phase, the more interesting modalities are those related to field activities that are coded 1-3. Their percentage distribution by countries is reported in the following table. Croatia and Italy can be considered as outliers because they have the lowest percentages (65% and 57%) in “field <100 mt” (the modality more precise in comparison with the other two) and in the same time the highest percentages of “PI in field” (38% and 24%) that is the less precise mode of data collection. In a second group (Bulgaria, Ireland and Romania) low percentages of “in field <100mt” (77% - 78%) are paired with high percentages in PI in field (17% - 18%) or in Greece where the low percentage (77%) “in field<100mt is matched with high percentage in the other two modes of data collection (11%). A third group composed by Austria, Spain, Slovenia, Slovakia and United Kingdom presents high percentages (10%-12%) in “PI in field” even if the levels of “in field <100mt” and “in field>100mt” are below or in the average.

Table 16: Distribution (percentage) of points by in field observation mode and country

Country	Field observation type			
	1	2	3	Total
AT	88%	2%	10%	100%
BE	94%	1%	4%	100%
BG	77%	6%	17%	100%
CY	88%	6%	6%	100%
CZ	98%	1%	1%	100%
DE	90%	6%	3%	100%
DK	89%	8%	3%	100%
EE	85%	6%	9%	100%

Country	Field observation type			
	1	2	3	Total
EL	77%	11%	11%	100%
ES	87%	4%	10%	100%
FI	94%	2%	4%	100%
FR	92%	2%	6%	100%
HR	57%	4%	38%	100%
HU	84%	6%	9%	100%
IE	73%	9%	18%	100%
IT	65%	10%	24%	100%
LT	97%	1%	2%	100%
LU	96%	1%	3%	100%
LV	94%	2%	4%	100%
MT	90%	4%	6%	100%
NL	85%	11%	4%	100%
PL	91%	3%	6%	100%
PT	91%	3%	7%	100%
RO	78%	3%	18%	100%
SE	96%	1%	3%	100%
SI	86%	2%	12%	100%
SK	85%	6%	10%	100%
UK	82%	8%	10%	100%
EU	86%	5%	10%	100%

In following table, it is analyzed how the main variables, land cover and land use, are collected. In general, land cover has a better performance, in terms of the percentages of “in field <100mt” mode than the variable land use. Within the land cover modalities, as expected, water areas and wetlands show the lowest percentages of points observed by “in field<100mt” mode and high percentages for the remaining two modalities of observation. Shrubland and woodland have a different figure with only high percentages of points observed by “PI in field” mode.

For what concerns the variable land use, the modality “energy, industry and manufacturing” present the lowest percentages of points observed by “in field<100mt” mode and the highest percentages for “in field >100mt” and “PI in field”. Except “agriculture”, the other modalities present high percentage of points observed by “PI in field” type. While the figures of observation modes for land cover are substantially adequate, the land use ones are to some extent surprising, because it is not easily to understand the differences in the figures of observation types for LU2, LU3 and LU4 where the content of these classification modalities are not so different.

Table 17: Distribution (percentage) of points by observation type and land cover /land use

	field observation type			
land cover	**1**	**2**	**3**	**Total**
A - ARTIFICIAL	90%	2%	8%	100%
B – CROPLAND	90%	7%	4%	100%
C – WOODLAND	79%	4%	18%	100%
D – SHRUBLAND	79%	6%	15%	100%
E – GRASSLAND	88%	4%	8%	100%
F - BARE LAND	90%	5%	5%	100%
G - WATER AREAS	74%	12%	14%	100%
H – WETLANDS	69%	10%	21%	100%
land use	**1**	**2**	**3**	**Total**
1 - AGRICULTURE	87%	5%	8%	100%
2 - ENERGY INDUSTRY AND MANUFACTURING	67%	12%	21%	100%
3 - TRANSPORT, UTILITIES AND RESIDENTIAL	87%	2%	11%	100%
4 - UNUSED AND ABANDONED AREAS	76%	6%	17%	100%
Total	86%	5%	10%	100%

7.4.1. Relation of different modalities using DMT errors

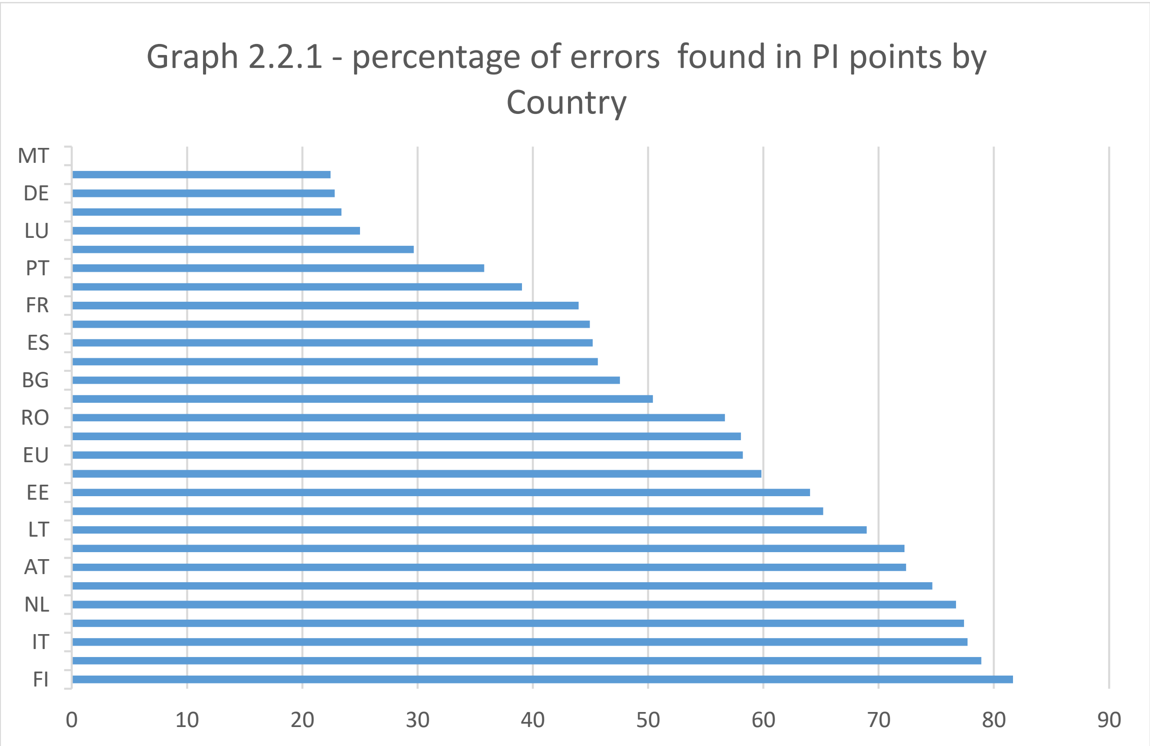
The relationship between the main modalities of data collection and DMT errors are analyzed by the following table that reports the indicators previously calculated. The highest percentage of erroneous points 8.6% is found for PI in the office that shows also an anomalous percentage of points with a number of errors greater than 1 (33.3%). Apart from this outlier, the variability of this indicator between the data collection modes is higher than the ones of the other indicators.

Table 18: indicators based on DMT errors

survey observation type	erroneous points over survey points %	average number of errors of erroneous points	number of errors % (a)	survey points % (b)	ratio: (a) / (b) *100
field <100 mt	4,1	1,07	40,1%	60,4%	0,66
field > 100 mt	4,3	1,06	2,3%	3,3%	0,69
PI in field	4,9	1,04	5,2%	6,8%	0,77
Pi in office	8,6	1,35	52,3%	29,5%	1,77
Total	5,5	1,20	100,0%	100,0%	1,00

In order to show the variability over the Countries of the errors by the data collection modalities, in the following Figure are given the percentages of the number of errors found in points collected by PI modalities over the total errors (in field plus PI) by Country.

Figure 13: Percentage of errors found in PI points by country



7.4.2. Changes in data collection modes between last surveys

In LUCAS, missing units are not allowed. The refusal of an owner to admit the surveyor or the impossibility to reach a point are treated resorting to photo interpretation. Nevertheless, considering that PI may result less precise than in field mode for many of the requested information and if the substitution of data collection type is "selective" (i.e. it depends on the characteristics of the points), the estimates may be biased and these effects cumulate in each wave. The substitution of a direct observation of the point with an "indirect" one (such as PI or observation from a large distance) can be considered similar to "panel attrition".

Information on the changes in data collection modes are useful not only to have some indication on data quality but also to plan a future panel, can be obtained by analyzing the set of points common to 2015 and 2018 samples considering the points surveyed by the modalities "in field" in 2018 and in 2015. In the following table, the absolute values of the transitions are reported. The yellow cell is related to the points that did not change the "in field with a distance<=100 mt" modality (88.9% over the total points). The red cells represent the "attrition" generated in 2018 (6.7%), while the blue ones are the attrition acquired from the previous survey (2.3%); finally the green cells represent the attrition in 2015 recovered in 2018. These results show that in the subset of "panel" points the methods of data collection are less precise in 2018 with respect to 2015, due to the moving from the modality "field survey <100 mt" in 2015 to " PI in field" in 2018.

Table 19: Observation method "in field" of points common to 2018 and 2015 samples

Absolute frequencies

Observation method in 2018	Observation method in 2015			
	Field survey, distance <= 100m	Point observed in the distance (>100 mt)	Photo-interpretation in the field	Total
Field survey, distance <= 100m	72154 (88.9%)	1454 (1.8%)	239 (0.3%)	73847 (91.0%)
Point observed distance >100m	1861 (2.3%)	1083 (1.3%)	45 (0.06%)	2989 (3.7%)
PI in the field	3572 (4.4%)	345 (0.4%)	434 (0.5%)	4351 (5.3%)
total	77587 (95.6%)	2882 (3.5%)	718 (0.9%)	81187 (100.0%)

7.4.3. Indicators on Differences between PI and in field modalities

Photo interpretation is one of the modes of data collection in LUCAS survey that has been used to greater extent in 2018 campaign. Therefore, the two modalities of data collection may have different precision in collecting information and hence a relative analysis is subsequently presented. The amount of PI points, regardless in field or in office, was 36.3% of the total observed points (Table 20).

Table 20: PI and in field points in LUCAS 2018

observation type	survey points	%
in field	215145	63.7%
PI in field	22865	6.8%
PI in office	99783	29.5%
Total PI	122648	36.3%
Total Points	337835	100%

The PI data collection mode, nevertheless, could have the following shortcomings:

- The photos of a PI point are kept some years (three or even more) before the Lucas survey; therefore they do not represent the actual status of the points because its characteristics could have changed.
- For some variables, the photo interpretation cannot be as accurate as the in-field observation.

To estimate the bias due to the different modes (PI and in field) of collecting a specific variable we would need three measures of a variable on the same point: the PI, the surveyor observation and a “conciliation” done by more qualified personnel to assess the “true” value”. Therefore, the bias estimation needs a planned random sample of points to avoid further bias due to the point selection. In absence of this tool, we can however obtain from survey data some indicators on the performances of the two modalities of data collections.

For this purpose, indicators could be calculated as “distance” between the two modes. It is not an estimate of bias but could help to give information on the shortcomings of PI. To this end, we use the longitudinal structure between consecutive rounds (that is the points present in both the surveys) analyzing the points that changed from in field modality to PI modality and vice versa. In doing so, the reference periods are different; assuming the photo are referred 3 years before we have that the difference between in field point at T and PI point at T-3 are related to the period (T-6,T) while the difference between in field point at T-3 and PI point at T are referred to T itself. For example by analyzing the points, common to surveys 2015 and 2018, if we consider the in-field points at 2018 and PI points at 2015 the period is 2012-2018 while for the in-field points at 2015 and PI points at 2018 the period is 2015 itself. In the table below the eight combinations of data collection modes to be analyzed are shown; the number are related to the longitudinal 2015/2018 structure. The survey data collection modes and related reference years for survey held in 2015 and 2018 are reported.

Table 21: Actual reference year and panel size by data collection mode in survey at year 2015 and 2018

Data collection mode		Actual reference year		Panel size	
survey 2015	survey 2018	survey 2015	survey 2018	Abs.values	%
Field	PI	2015	2015	18231	14,5
PI	Field	2012	2018	1594	1,3
PI	PI	2012	2015	23011	18,3
Field	Field	2015	2018	82675	65,9
				125511	100

The most interesting data set analyzing the distance between field and PI data collection is the group characterized by Field in 2015 and PI in 2018 where the collected information can be considered as a replicated observation of a point in the same year (2015) carried out by two different modalities. Of course, the differences in the variable collected in the two years could be due also to material errors of the surveyors. We could assume these errors are equally distributed by survey/years.

The differences in variables in the group composed by the points observed in field in 2015-2018 can be considered as gross variations affected only by random errors as well as the group PI-PI, referred to a previous period. Finally, the group PI-field produces variations over six years affected however by biases due to different ways to take information.

In the next table, the percentages of differences in LC1, LC2, LU1 and LU2 between 2015 and 2018, over the total in the related sub group, are reported; land cover and land use are used in the higher level of classification (1 digit for LC and 2 digits for LU). In the group field-PI, the percentage of differences between land cover is about 6% and for land use is 4.5% that can be considered due to different modalities in data collection.

Table 22: Differences in land cover and land use between 2015 and 2018 Considering LC1 coded by 1 digit, LU1 coded by two digits

		LC1 2018 ≠ LC1 2015	LC2 2018 ≠ LC2 2015	LU1 2018 ≠ LU1 2015	LU2 2018 ≠ LU2 2015
2015	**2018**	**%**	**%**	**%**	**%**
Total		12,1	59,3	5,8	60,1
Field	PI	6,1	63,3	4,5	66,5
PI	Field	17,1	64,5	14,0	52,1
PI	PI	8,7	49,4	7,6	47,5

		LC1 2018 ≠	LC2 2018 ≠	LU1 2018 ≠	LU2 2018 ≠
		LC1 2015	LC2 2015	LU1 2015	LU2 2015
Field	Field	14,2	59,8	5,4	62,1

For LC2 and LU2 the differences are much higher but the very small amount of points that have this information, makes the results not enough reliable.

The absolute number of differences in land cover and the percentage ratio of these differences over the no changes (that is the points that did not change modality from 2015 to 2018) are reported by LC1 in 2015 in the next table. In artificial modality, for instance, the indicator shows that we have 6.5 differences every 100 no changes. The most problematic modality is bareland (with a ratio = 122), for which the differences exceed the no changes; shrubland and grassland also show significantly high ratios.

Table 23: Differences in land cover and land use modalities between 2015 and 2018 Considering LC1 coded by 1st digit, LU1 coded by two digits

LC1 in 2015	Differences field - PI	% Ratio: difference/no changes
land cover – LC1		
A - artificial	139	6,5
B – cropland	226	7,8
C – woodland	140	1,6
D - shrubland	146	22,1
E – grassland	311	15,4
F – bareland	121	122,2
G - water areas	18	5,2
H – wetland	19	4,0
Total	1120	6,5
land use – LU1		
U1 agr, fishing, mining	231	1,8
U2 energy & manufacturing	14	15,1

LC1 in 2015	Differences field - PI	% Ratio: difference/no changes
U3 transport, utilities & residential	121	4,7
U4 unused and abandoned areas	454	23,2
Total	820	4,7

In order to calculate the indicators we also consider the points coded BX1 and BX2 that are the 7.8% of the B codes in the subsample we are analyzing. In field, they were classified and hence an actual difference exists between the two modalities of survey taking.

The above analysis has been carried out using the higher level of classification for LC (one digit i.e. A) and LU (two digits i.e. U11) where the differences inside the classes are cancelled in the calculation. The results obtained using the complete codes of classification are reported in the following table, where the percentages of difference for LC1 and LU1 greatly increase with respect to the ones reported in previous Table.

Table 24: Differences in land cover and land use between 2015 and 2018.

		LC1 2018 ≠	LC2 2018 ≠	LU1 2018 ≠	LU2 2018 ≠
		LC1 2015	LC2 2015	LU1 2015	LU2 2015
2015	2018	%	%	%	%
Total		31.3	65.2	10.8	60.9
Field	PI	21.5	67.7	8.0	67.1
PI	Field	37.6	64.5	18.0	54.8
PI	PI	15.4	53.2	9.3	48.8
Field	Field	37.7	66.3	11.6	62.8

For the target combination “Field in 2015 and PI in 2018” no difference theoretically should be found because in this case there are two repeated observations on the same points in the same reference year 2015 (see previous table). Nevertheless, differences are found and they can be assumed as an indication of bias, largely due to the different collection mode. The percentages of these differences can be decomposed according to the contribution of LC1 distinguishing between (i) the transition within the higher code classes and (ii) the ones among these classes. The data related to the first case are showed in the following Table. The percentage of differences from one code to another over the class total is particularly relevant for cropland due to changes in classification. Apart from this outlier, the classes more affected from the differences are artificial, shrubland and grassland.

Table 25: Differences in land cover between 2015 (in field mode) and 2018 (PI mode) within 1 digit code classes

LC1 in 2015	LC1 in 2018	LC1 2018 ≠ LC1 2015		class totals
in field	PI	%	abs value	abs value
A	A	5.1	108	2130
B	B	81.2	2351	2897
C	C	1.9	160	8493
D	D	5.4	36	661
E	E	6.7	136	2017
F	F	1.0	1	99
G	G	1.5	5	343
H	H	0.6	3	471
Total		16.4	2800	17111

In following table, the differences in percentages over the subclass total and in absolute values between the two data collection modalities (field in 2015 and PI 2018) are reported. The data are showed by the LC1 collected in field in 2015. Neglecting these cases and analyzing both the percentages and the absolute values, we can say that the subclasses more affected by differences are A22, C10, D10, E10, E20 and E30.

Table 26: Differences in land cover between 2015 (in field mode) and 2018 (PI mode) within 3 digit code subclasses

LC1 – 2015 in field	diff %	abs.value	LC1 – 2015 in field	diff %	abs.value	LC1 – 2015 in field	diff %	abs.value
A11	3.7	28	B11	3.4	22	C10	2.3	93
A12	1.4	1	B12	3.5	3	C21	0.9	9
A13	7.1	4	B13	19.0	58	C22	1.3	19
A21	9.2	50	B14	5.6	3	C23	1.9	5
A22	6.2	51	B15	10.0	9	C31	1.2	9
A30	16.1	5	B16	5.6	17	C32	0.1	1
			B17	29.9	20	C33	1.0	4
			B18	2.9	1			
			B19	9.1	1	D10	24.5	115

LC1 – 2015 in field	diff %	abs.value	LC1 – 2015 in field	diff %	abs.value	LC1 – 2015 in field	diff %	abs.value
			B21	3.6	1	D20	9.2	31
			B31	5.2	8			
			B32	2.8	5	E10	16.4	80
			B34	33.3	1	E20	7.4	108
			B37	10.0	1	E30	31.9	123
			B41	2.2	1			
			B42	14.3	1	F10	18.5	5
			B43	12,5	2	F30	25.0	1
			B44	18,2	2	F40	64.6	115
			B51	8,3	1			
			B52	6,4	3	G11	5.3	11
			B53	9,5	4	G12	14.3	1
			B54	11,8	2	G21	4.1	6
			B55	12,4	15			
			B72	11,1	2	H11	6.7	12
			B74	9,3	9	H12	2.1	6
			B76	9,5	2	H21	7.7	1
			B81	6,6	14			
			B82	7,9	12			
			B83	35,3	6			

Under the assumption that the PI points in great part cancel the status variations between two surveys, it is reasonable to expect an under estimate of variations in land cover. An indication (*table 27*) of this effect is done by the percentage variations between 2018 and 2015 over 2015 found in the panel 2015-2018 considering the mode “field in 2015 and field in 2018” versus the variations obtained considering all the points in the panel. The variations are higher in the first case and generally of opposite sign. These differences are not only due not to the observation mode but also to the relationship between this variable and land cover, that is the distribution of PI / field modes over land cover modalities is not the same.

Table 27: Percentage of changes in the points of 2015-2018 panel – field mode in both the years

Observation mode	A	B	C	D	E	F	G	H
field 2015 / field 2018	-1.3	1.6	-4.5	6.1	2.8	-7.7	-17.7	1.1
all the panel points	-0.8	-1.3	3.3	-3.9	-3.1	3.9	5.1	-1.3

7.4.4. Accuracy Indicators on distance and length of survey

Another interesting issue in terms of accuracy deals with the distance and the length of the survey. In the following table, the survey start-end dates are reported, broken down by the main observation types and country. It could be seen in the last two columns the length of the data collection phase and the ratios between each length and the total length of data acquisition period. The lengths as well as the month of the start and the end of data collection influence the content of data in particular for the crops and inland water.

The lengths of the periods related to field observation modalities (coded 1,2,3) are quite different while they should be more or less the same, that is the different field modalities should be randomly distributed over a unique period. The indicators related to the countries show a large variability. They range from 10 days in Malta to 369 days in Poland. The greatest lengths are observed in Belgium, France and Poland.

Table 28: Observation period and length of survey data collection by observation type and country

	Starting date				Ending date				
Observation type	Y	M	D		Y	M	D	Length in days	% on total period
TOTAL	18	01	12		19	03	14	422	100%
1	18	01	12		19	02	21	399	95%
2	18	03	28		18	12	23	265	63%
3	18	04	02		19	02	17	315	75%
7	18	03	21		19	03	14	353	84%
Country	Y	M	D		Y	M	D	Length	% on tot
TOTAL	18	01	12		19	03	14	422	100%
AT	18	04	17		18	10	21	184	44%
BE	18	05	07		19	03	14	307	73%
BG	18	04	17		18	12	03	226	54%

	Starting date				Ending date				
CY	18	03	25		18	10	04	189	45%
CZ	18	05	09		18	11	30	201	48%
DE	18	04	27		18	12	20	233	55%
DK	18	03	21		18	09	29	188	45%
EE	18	04	28		18	10	16	168	40%
EL	18	04	02		18	10	31	209	50%
ES	18	03	29		18	12	19	260	62%
FI	18	03	29		18	12	04	245	58%
FR	18	04	27		19	03	07	310	73%
HR	18	05	02		18	12	22	230	55%
HU	18	04	17		18	09	24	157	37%
IE	18	05	08		18	10	22	164	39%
IT	18	04	13		18	12	19	246	58%
LT	18	05	09		18	09	28	139	33%
LU	18	05	07		18	09	16	129	31%
LV	18	04	27		19	01	03	246	58%
MT	18	05	02		18	05	12	10	2%
NL	18	04	12		18	09	30	168	40%
PL	18	01	12		19	01	21	369	87%
PT	18	04	18		18	10	21	183	43%
RO	18	05	13		19	02	21	278	66%
SE	18	04	16		18	12	11	235	56%
SI	18	05	13		18	11	19	186	44%
SK	18	04	30		18	12	10	220	52%
UK	18	03	27		18	11	21	234	55%

The distance from the observer and the sampled points could be read from two different points of view. From one side as a factor influencing the data quality and on the other as an explanation of the used observation mode. In the following Table, the mean and the maximum value of the distance from enumerator and the point to be visit are reported by observation type and by country.

Table 29: Distance in meters (mt) from observer and sampled point by Country

Country	observation Type					
	in field<100mt		in field>100mt		PI in field	
	mean	max	mean	max	mean	max
AT	11,7	100	162	600	1511	15269
BE	7.8	96	152	290	95	889
BG	9,0	99	303	1980	1108	9463
CY	8,9	100	215	831	813	3980
CZ	5,6	100	196	495	518	7800
DE	13,5	100	551	92117	601	88579
DK	14,3	99	198	955	502	5290
EE	10,1	99	225	710	1480	66602
EL	13,7	100	341	3453	876	82124
ES	6,6	100	243	1045	1162	61514
FI	7,9	100	162	698	1587	21475
FR	6,9	100	193	1965	574	374523
HR	9,9	99	287	1724	1415	71749
HU	10,3	100	245	813	625	6073
IE	14,0	100	209	627	327	6397
IT	16,6	100	257	2381	783	287333
LT	4,6	94	223	639	605	5990
LU	7,7	97	168	234	96	290
LV	4,6	99	182	429	529	6178

Country	observation Type					
	in field<100mt		in field>100mt		PI in field	
	mean	max	mean	max	mean	max
MT	10,9	77	163	252	633	2332
NL	16,0	100	210	2111	279	4158
PL	8,5	100	216	1240	598	6073
PT	7,6	100	213	752	792	12118
RO	7,8	100	297	1364	1530	49764
SE	5,4	100	159	464	1941	22179
SI	8,4	96	381	6809	772	8961
SK	8,4	99	269	846	1027	7320
UK	14,4	100	199	2056	890	19224
EU	9,7	100	236	92117	845	374523

The means of in field<100mt range from 4,6mt (Latvia and Lithuania) to 16,6 mt (Italy). For in field >100mt from 162 mt. (Austria and Finland) to 551 mt (Germany). For PI in field (obs type =3) from about 95 mt (Luxemburg and Belgium) to 1941 mt. (Sweden). The highest values of distance mean of PI in field could be considered as a signal of not reachability and justify to some extent the change of direct observation with PI. The means of distance for in field>100mt are also high in terms of accuracy classification from such distance. The following table reports the two modes of observation the percentages of points by the modalities of LC and LU that in general are not negligible. It should be noted that the observation done at a large distance is considered reliable in case of artificial, grassland, water areas and wetland, nevertheless, but not on the same scale for cropland or woodland or even more for the classification of land use.

Table 30: Percentages of points by LC and LU modalities and observation types

Land cover	In field >100mt	PI in field	Land use	In field >100mt	PI in field
A - ARTIFICIAL	2%	8%	**1 – AGRICULTURE**	5%	8%
B – CROPLAND	7%	4%	**2 - ENERGY INDUSTRY AND MANUFACTURING**	12%	21%
C – WOODLAND	4%	18%	**3 - TRANSPORT, UTILITIES AND RESIDENTIAL**	2%	11%

Land cover	In field >100mt	PI in field	Land use	In field >100mt	PI in field
D – SHRUBLAND	6%	15%	4 - UNUSED AND ABANDONED AREAS	6%	17%
E – GRASSLAND	4%	8%			
F - BARE LAND	5%	5%			
G - WATER AREAS	12%	14%			
H – WETLANDS	10%	21%			

In next table, the above analysis is further developed by considering additional indicators for which it is possible to quantify the corresponding percentage of points. For the points directly observed from a distance >100mt, in the first column are reported the 3rd quartiles of the distance distributions by country, i.e. the minimum values assumed by the last fourth of the points ordered by distance. For some cases, the distance is an important determinant of the precision of the observation that are quantified in the second column. For the PI in field mode, the 1st quartiles of the distance distributions by country are presented, i.e. the maximum value for the first fourth of the points ordered by distance. The distances of these points are generally not so large and hence part of them could be even directly observed. So in the first case there is a risk in data quality where the obs type = 2 may not have been the best option by the enumerators, whilst in the second case obs type = 2 could have been applied for many countries.

Table 31: Distance from observer to sampled point by Country 3rd quartile (Q3) for in field>100mt and 1st quartile (Q1) of PI in field

	in field >100mt		PI in field	
Country	Q3 - distance in mt	% of points with a distance >Q3	Q1 - distance in mt	% of points with a distance < Q1
AT	177	1,7	169	2,8
BE	168	1,3	29	1,2
BG	381	6,2	280	5,4
CY	243	4,7	230	1,5
CZ	234	1,3	23	0,6
DE	259	5,5	56	1,0
DK	242	6,8	66	0,8
EE	256	6,2	214	3,1
EL	404	9,8	200	3,3
ES	303	3,2	348	2,9

	in field >100mt		PI in field	
FI	182	1,9	144	1,2
FR	224	1,7	58	1,8
HR	315	3,3	230	10,4
HU	291	5,8	90	2,8
IE	259	7,0	135	4,5
IT	301	8,3	203	6,5
LT	271	1,5	117	0,7
LU	234	1,1	45	0,7
LV	222	1,7	115	1,4
MT	252	2,9	126	1,6
NL	250	9,2	48	1,1
PL	258	2,1	94	1,5
PT	256	2,2	154	1,8
RO	374	2,9	234	5,7
SE	182	0,9	144	0,7
SI	247	1,7	125	3,2
SK	320	5,4	151	3,2
UK	230	6,3	81	2,6
EU	262	4,1	140	2,8

The above analysis is completed by analyzing the comments recorded by the enumerators to describe the operations carried out while observing each sample point; they are useful to understand the fields' conditions and the difficulties met by them. Such comments have been classified in more than 150 categories concerning several aspects of the field operations. The categories concerning the difficulties in reaching the points are 21 and they are detected by the first two digits code equal to AB. This information are useful to comprehend the differences in using the "PI in field" or "in field>100mt" that are the two modalities more critical in collecting reliable data. The points distributions among comments are similar for observation method 2 (field >100 mt) and 3 (PI in field), as reported in next table. As expected, for each method the most common comment was AB01 = "Fence, wall, locked gate/door, natural obstacle" (20.7% for method 2 and 24.6% in case of method 3).

Most of the other categories have similar frequencies for the two methods with the following exceptions:

AB09="Shortest possible walking time more than 1 hour" (2.7% in case of method 2 and 11.8% in case of method 3);

AB12="High crop on the ground" (27.6% in case of method 2 and 1.19% in case of method 3);

AB13="Non-drivable road/track" (2.6% in case of method 2 and 12.0% in case of method 3).

Table 32: Distribution of points by observation method and comment code (LUCAS 2018)

	in field>100mt		PI in field		total	
	abs. values	%	abs. values	%	abs. values	%
AB01	2934	20.7	8509	24.6	11443	23.5
AB02	351	2.5	2954	8.5	3305	6.8
AB03	691	4.9	2344	6.8	3035	6.2
AB04	69	0.5	384	1.1	453	0.9
AB05	723	5.1	1021	3.0	1744	3.6
AB06	1962	13.9	4329	12.5	6291	12.9
AB07	93	0.7	54	0.2	147	0.3
AB08	153	1.1	219	0.6	372	0.8
AB09	388	2.7	4072	11.8	4460	9.1
AB10	42	0.3	558	1.6	600	1.2
AB11	1529	10.8	2334	6.7	3863	7.9
AB12	3903	27.6	662	1.9	4565	9.4
AB13	372	2.6	4169	12.0	4541	9.3
AB14	46	0.3	248	0.7	294	0.6
AB15	4	0.0	71	0.2	75	0.2
AB16	11	0.1	125	0.4	136	0.3
AB17	11	0.1	10	0.0	21	0.0

	in field>100mt		PI in field		total	
AB18	10	0.1	76	0.2	86	0.2
AB19	25	0.2	114	0.3	139	0.3
AB20	6	0.0	81	0.2	87	0.2
AB99	834	5.9	2275	6.6	3109	6.4
Total	14157	100.0	34609	100.0	48766	100.0

In the following tables, the distribution parameters of length of data collection are reported (length is expressed in minutes). Apart from the average, other parameters are those values that divide the length distribution according to 50% / 50% for the median, 75% / 25% for the third quartile (Q3), 95% /5% for P95, 99% / 1% for P99 and finally max as the maximum value.

Considering all the points, the average of observation length is about 13 minutes. The 50% of the length are less than 10 minutes (the median), the 25% of the points are observed with a length from 10 to 17 minutes (Q3), the further 20% with a length from 17 to 37 minutes; only 4% of the points range from 37 to 64 minutes while 1% has a length more than 64 minutes.

The length averages range from 3,6 minutes for “PI in office” to 17,4 minutes for “in field < 100mt” while the remaining ones are both about 13 minutes. The length distribution parameters of “in field <100mt“ mode is quite different from the “in field >100mt” and the two distributions of the field >100mt and the “PI in field” are very similar. This is related to the information completeness with respect to the ones collected in field from a distance <100mt.

Table 33: Parameters of distribution of observation length by observation type (minutes)

observation type	number of points	average	Median	Q3	P95	p99	max
Total	337854	12.9	10	17	37	64	830
In field <100mt	203961	17.4	14	21	43	70	740
In field >100mt	11159	13.3	10	16	33	55	367
PI in field	22894	13.0	10	16	35	64	830
PI in office	99803	3.6	3	5	8	15	782

In the following table, the parameters of the distribution of observation length are provided, according to the land cover and land use modalities. The averages and the medians of cropland and grassland are the largest (more than 14 minute and 11 respectively) while, as expected, the lowest are those related to water area, wetland and artificial. With regard to land use, LU1 (agriculture, fishing and mining) show a length distribution skewed to right, that is with all the parameters greater than the other modalities.

Table 34: parameters of the distribution of the observation length by land cover and land use (minutes)

land cover	number of points	Average	Median	Q3	P95	P99	Max
Total	337854	12,9	10	17	37	64	830
A - artificial	21545	9,0	7	12	23	40	614
B – cropland	86693	14,3	11	17	36	61	740
C – woodland	120087	12,2	7	16	39	66	740
D - shrubland	19030	12,6	8	17	38	64	677
E – grassland	73108	14,6	11	18	40	69	830
F – bareland	7987	11,5	8	15	34	55	357
G - water areas	2641	7,5	4	10	25	49	356
H – wetland	6726	6,1	2	6	25	50	321
Land use							
U1 agr, fishing, mining	253149	13,5	10	17	38	65	782
U2 energy & manufacturing	1201	10,1	7	13	27	50	245
U3 transport, utilities & residential	34675	10,3	8	13	26	48	614
U4 unused and abandoned areas	48792	11,8	7	16	36	63	830

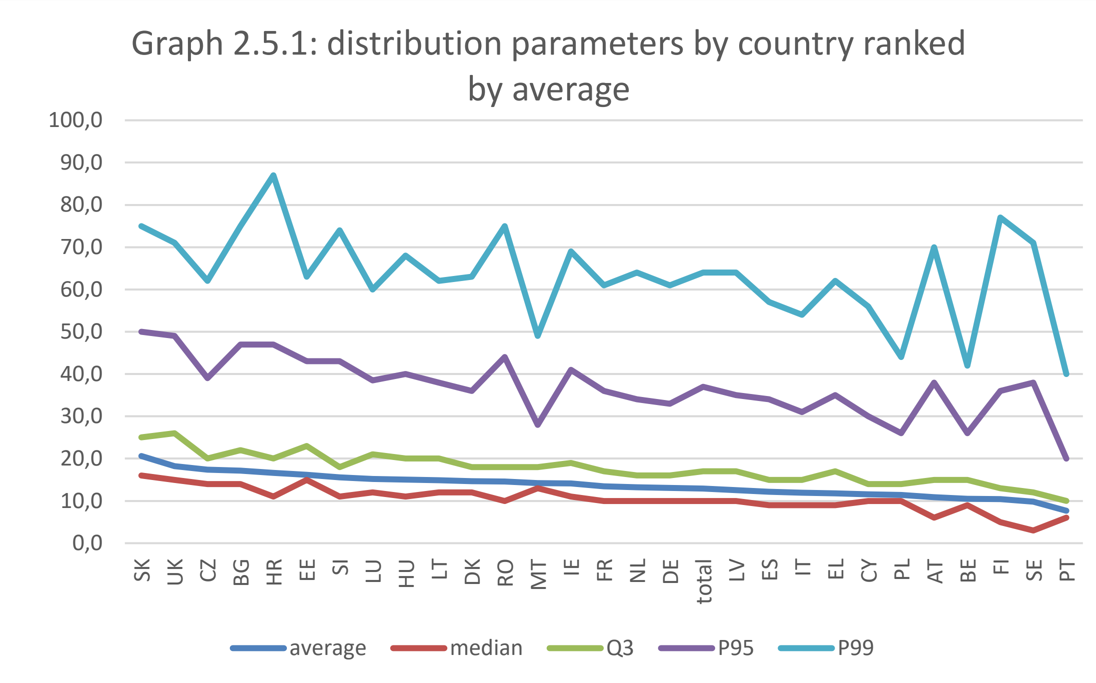
In the next table and figure, the countries are ranked by the length averages.

Table 35: Parameters of the distribution of the observation length by countries ranked by length average (minutes)

Country	number of points	average	Median	Q3	P95	P99	Max
SK	2898	20,6	16	25	50	75	514
UK	17253	18,2	15	26	49	71	166
CZ	5713	17,4	14	20	39	62	677
BG	7678	17,2	14	22	47	75	340
HR	4239	16,6	11	20	47	87	614
EE	2665	16,2	15	23	43	63	119

Country	number of points	average	Median	Q3	P95	P99	Max
SI	1922	15,6	11	18	43	74	562
LU	340	15,2	12	21	38,5	60	110
HU	5514	15,0	11	20	40	68	575
LT	4584	14,9	12	20	38	62	384
DK	3703	14,7	12	18	36	63	372
RO	16725	14,6	10	18	44	75	668
MT	79	14,2	13	18	28	49	49
IE	4975	14,2	11	19	41	69	167
FR	48215	13,4	10	17	36	61	782
NL	5011	13,2	10	16	34	64	353
DE	26777	13,1	10	16	33	61	325
LV	5376	12,6	10	17	35	64	362
ES	45314	12,1	9	15	34	57	740
IT	28294	11,9	9	15	31	54	605
EL	12622	11,8	9	17	35	62	370
CY	2313	11,6	10	14	30	56	142
PL	23086	11,4	10	14	26	44	830
AT	8840	10,8	6	15	38	70	362
BE	3659	10,5	9	15	26	42	301
FI	16182	10,5	5	13	36	77	479
SE	26709	9,8	3	12	38	71	366
PT	7168	7,7	6	10	20	40	178

Figure 14: Distribution parameters by country ranked by average



7.5. Effects of model estimators on the final results

The introduction of calibration is significant in order to achieve greater accuracy on the final estimates. Under this scope, further analysis has been carried out in order to assess the impact of the calibration procedure on the final estimates. Usually, calibration is performed with the aim to reduce bias and variance of the estimates, and its impact is very relevant when non-response rates are high. In LUCAS case, non-response rate is negligible; nonetheless, bias could be introduced by other above-mentioned factors such as the observation type and difficult access to the point.

Bias cannot be directly estimated, but the comparison between calibrated and non-calibrated estimates could provide information of its possible magnitude. In particular, the calibration model was introduced to make equal to available known totals in each country (NUTS0), all the estimates related to total area by:

- elevation class (ELEV);
- imperviousness;
- cropland;
- woodland;
- wetland;
- water.

Known totals had been calculated for each country by summarizing the values of the correspondent variables in the Master dataset. In particular, values from 2 to 6 have been assigned to each point in the Master by making use of the datasets of Copernicus CORINE Land Cover and High Resolution Layers.

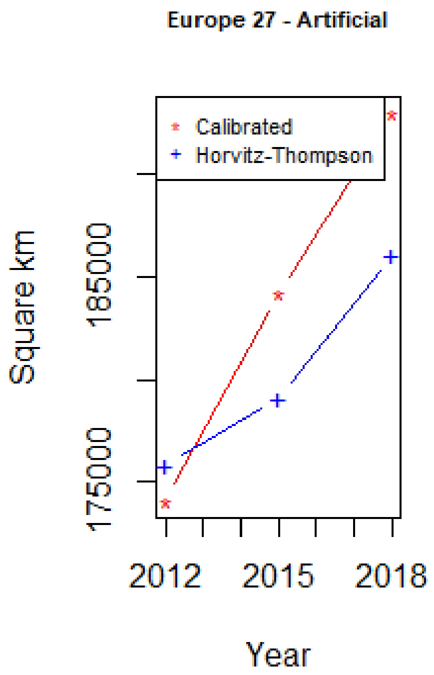
7.5.1. LUCAS Horvitz-Thompson estimates

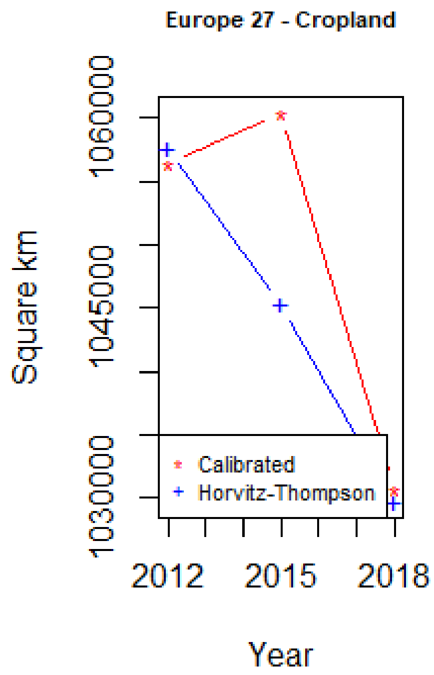
LUCAS estimates are being produced using the methods of calibration in which the estimates are forced to equalize known totals deriving from reliable sources. By this way, the variances of estimates and biases are reduced. Assuming that the known totals are "true" values, relevant distance between calibrated and Horvitz-Thompson estimates could be a signal of possible bias in data or in any case of large variability.

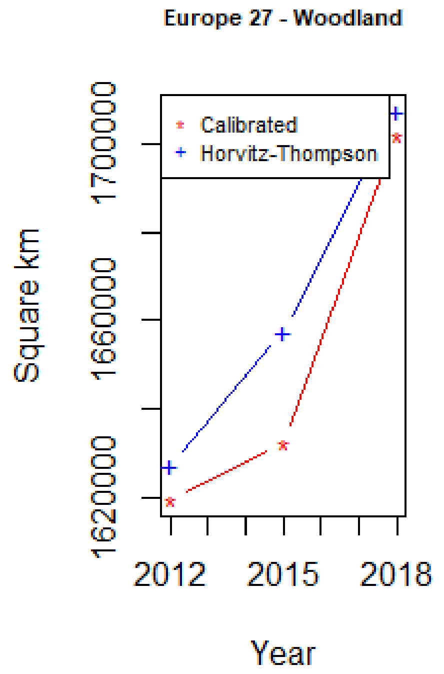
In order to evaluate the impact of calibration in the production process of estimates, also Horvitz-Thompson (HT) estimates have been produced. HT estimates are the ones obtained by making use of initial weights, i.e. those obtained by the inverse of the inclusion probabilities of the sampled points. A subset of the HT estimates have been compared to calibrated ones, that is:

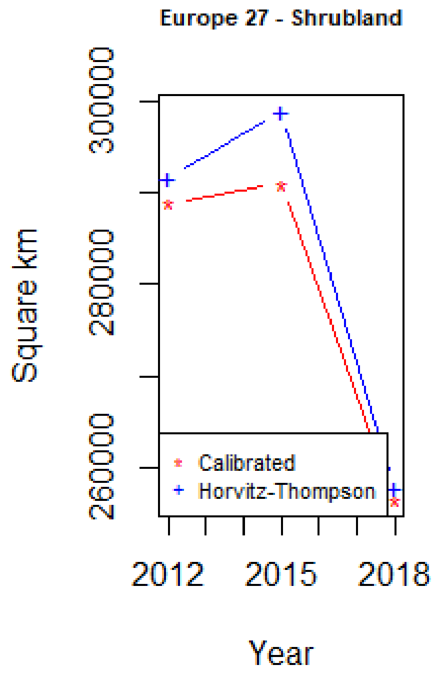
- Land cover (1 digit);
- Land use (1 digit);
- settlement;
- FAO classes (1,2,3);
- LUE (Land use with heavy environmental impact);
- LUD (Services and residential area).

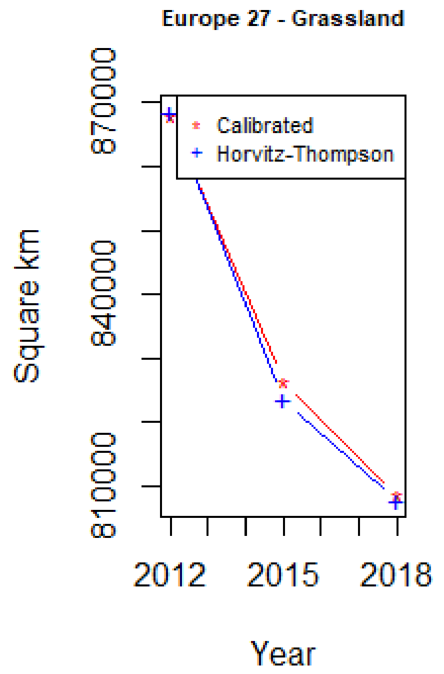
All of them are produced at EU level. We thus report the comparison results for Europe with 27 countries.

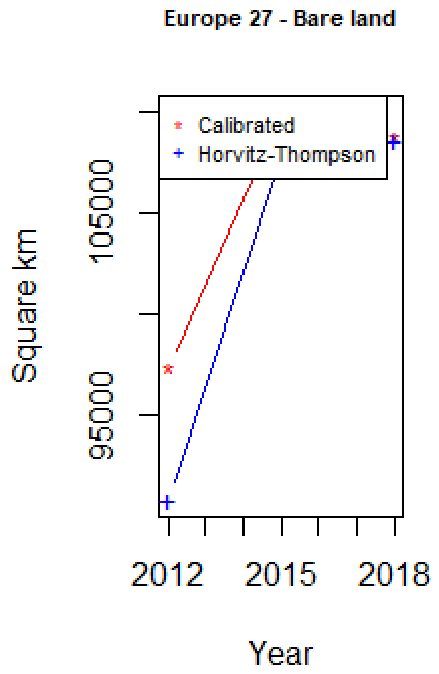


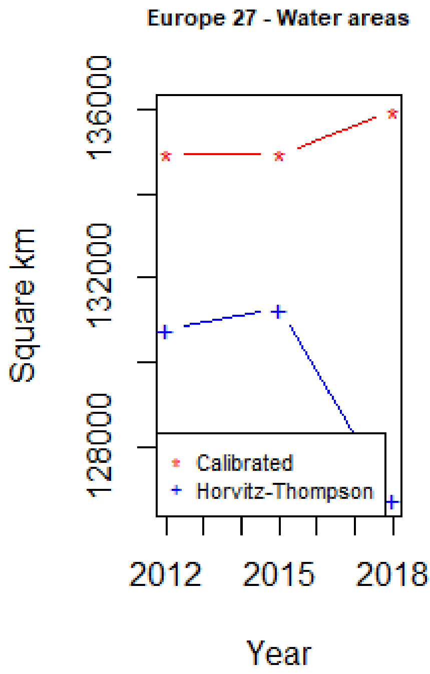


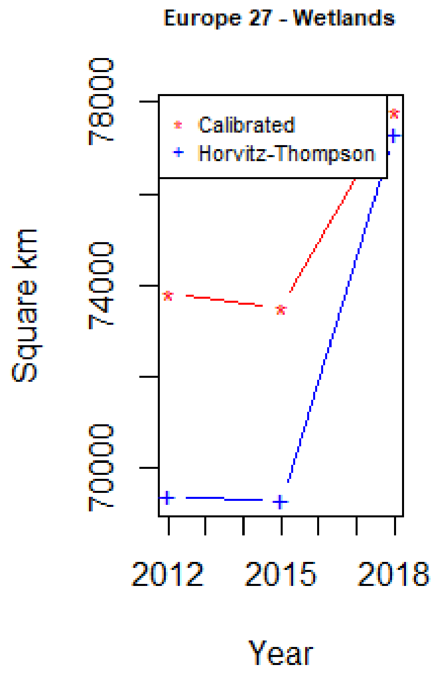

Europe 27 - Woodland
Calibrated
Horvitz-Thompson
Square km
1620000
1660000
1700000
2012
2015
2018
Year

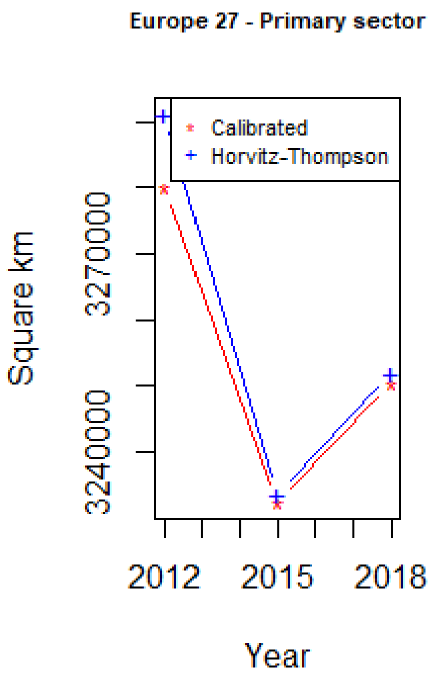

Europe 27 - Shrubland
Calibrated
Horvitz-Thompson
Square km
260000
280000
300000
2012
2015
2018
Year

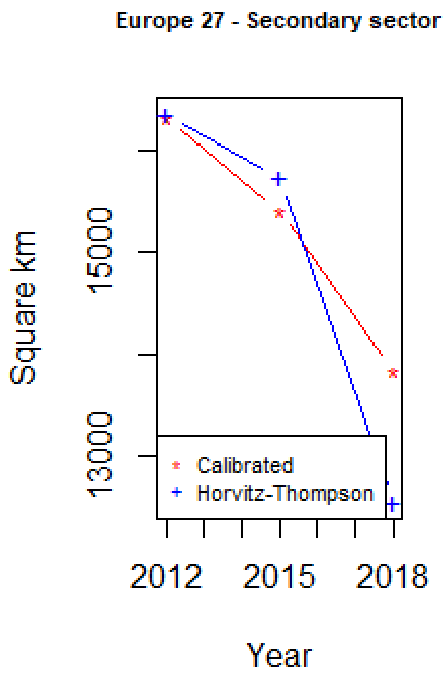

Europe 27 - Grassland
Calibrated
Horvitz-Thompson
Square km
810000
840000
870000
2012
2015
2018
Year

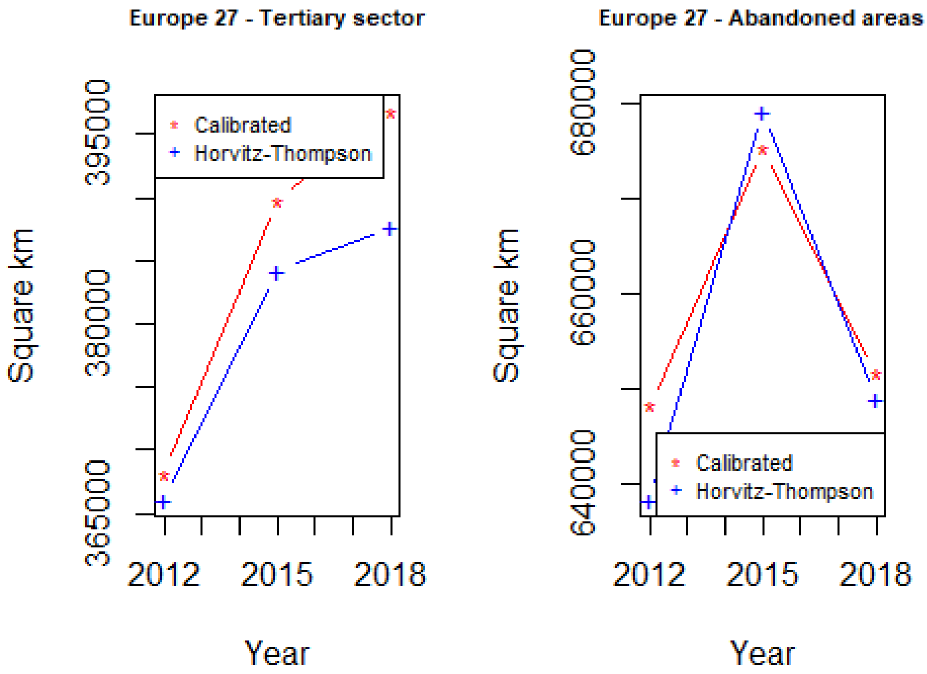

Europe 27 - Bare land
Calibrated
Horvitz-Thompson
Square km
95000
105000
2012
2015
2018
Year


Europe 27 - Water areas
Square km
136000
132000
128000
Calibrated
Horvitz-Thompson
2012
2015
2018
Year

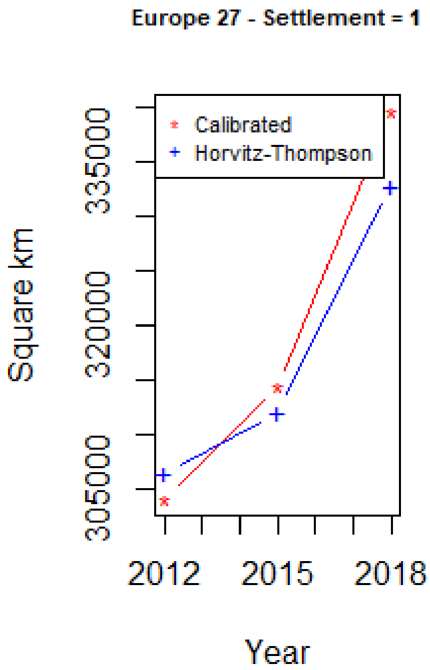

Europe 27 - Wetlands
Square km
78000
74000
70000
Calibrated
Horvitz-Thompson
2012
2015
2018
Year

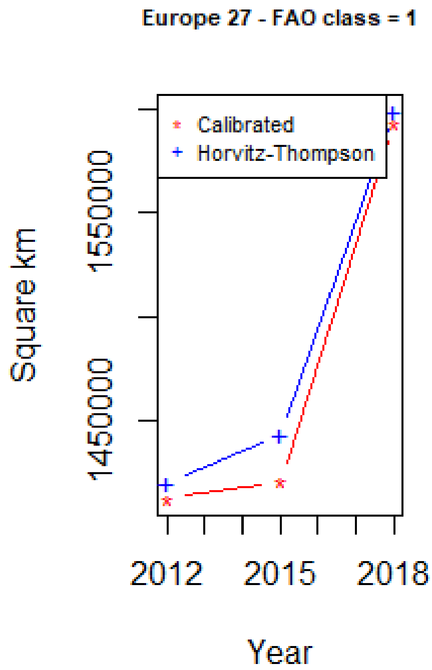

Europe 27 - Primary sector
Square km
3270000
3240000
Calibrated
Horvitz-Thompson
2012
2015
2018
Year

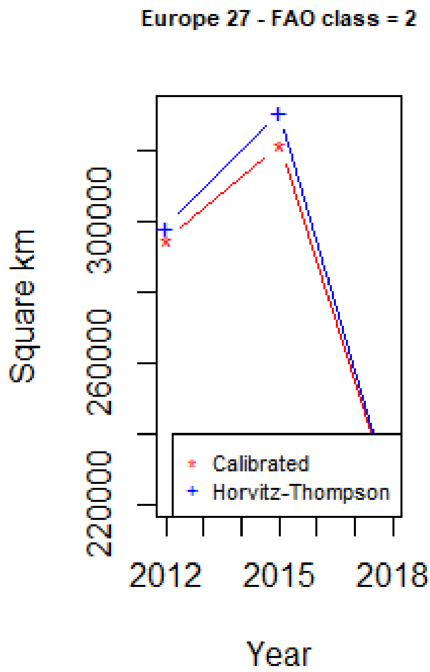

Europe 27 - Secondary sector
Square km
15000
13000
Calibrated
Horvitz-Thompson
2012
2015
2018
Year

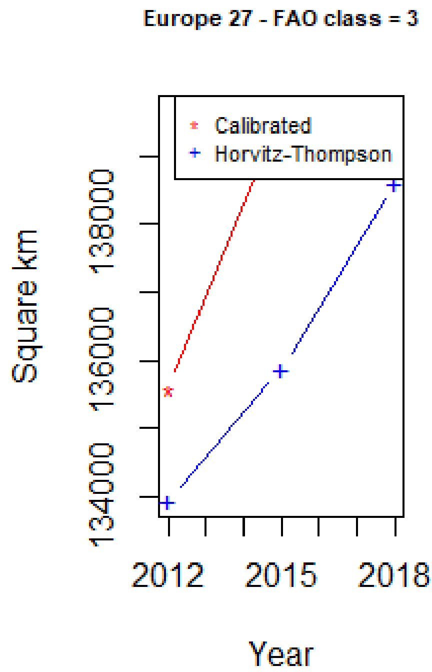


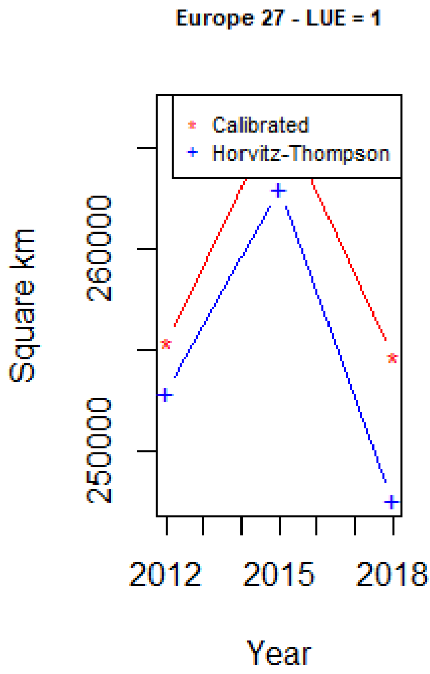
Europe 27 - Tertiary sector
Calibrated
Horvitz-Thompson
Square km
365000
380000
395000
2012
2015
2018
Year
Europe 27 - Abandoned areas
Calibrated
Horvitz-Thompson
Square km
640000
660000
680000
2012
2015
2018
Year

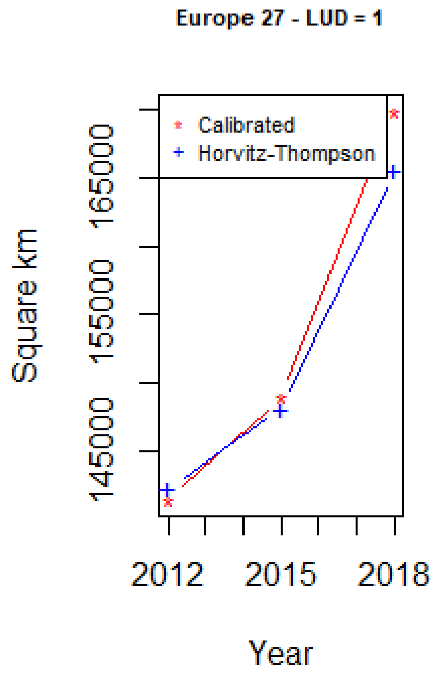

Europe 27 - Settlement = 1
Calibrated
Horvitz-Thompson
Square km
305000
320000
335000
2012
2015
2018
Year


Europe 27 - FAO class = 1
Calibrated
Horvitz-Thompson
Square km
1450000
1550000
2012
2015
2018
Year


Europe 27 - FAO class = 2
Calibrated
Horvitz-Thompson
Square km
220000
260000
300000
2012
2015
2018
Year


Europe 27 - FAO class = 3
Calibrated
Horvitz-Thompson
Square km
134000
136000
138000
2012
2015
2018
Year





In terms of impact of the calibration procedure on the LUCAS estimates, thus, there is no clear evidence of differences between trends observed in EU estimates. Therefore, the following considerations are valid in general for the three groups of estimates, the one involving all rounds of the LUCAS survey.

The estimates mostly affected by the calibration procedure are those related to Land Cover, rather than Land Use.

As for Land Cover, the value “Grassland” is the one less affected by the calibration, as its HT and calibration estimates are more or less the same in all the rounds of the survey. The opposite case is for “Water areas”, whose estimates diverge in all rounds. These estimates are based generally on small amount of sampling points and hence are subject to a large variability. In an intermediate position are the other cases. Except for Woodland and Shrubland the calibration estimates are higher than the corresponding HT ones; it means that the direct weights have redistributed the probability mass in favor of Woodland and Shrubland points.

Land Use estimates only present a noticeable divergence for “Secondary sector” in 2018. The other groups of estimates, related to “Settlement”, “FAO classes”, “LUE” and “LUD”, do not show significant divergences between HT and calibration estimates.

In the following table, there is a focus on the impact of calibration on estimates of “artificial“- and “water” Land Cover by country.

Table 36: Horvitz-Thompson and calibrated estimates of “artificial – water”, Land Cover by country (km2)

Country	Estimate	Area_2018	Variable	Area_2018
AT	Artificial HT	3494	Water HT	1361
AT	Artificial calibrated	3495	Water calibrated	1505
BE	Artificial HT	3433	Water HT	373

Country	Estimate	Area_2018	Variable	Area_2018
BE	Artificial calibrated	3601	Water calibrated	340
BG	Artificial HT	2585	Water HT	1168
BG	Artificial calibrated	2555	Water calibrated	1271
CY	Artificial HT	579	Water HT	32
CY	Artificial calibrated	573	Water calibrated	33
CZ	Artificial HT	3254	Water HT	991
CZ	Artificial calibrated	3450	Water calibrated	978
DE	Artificial HT	26919	Water HT	6098
DE	Artificial calibrated	27016	Water calibrated	6140
DK	Artificial HT	2874	Water HT	593
DK	Artificial calibrated	2947	Water calibrated	596
EE	Artificial HT	765	Water HT	2557
EE	Artificial calibrated	785	Water calibrated	2185
EL	Artificial HT	5259	Water HT	1536
EL	Artificial calibrated	5332	Water calibrated	1731
ES	Artificial HT	18648	Water HT	4969
ES	Artificial calibrated	18440	Water calibrated	5209
FI	Artificial HT	5661	Water HT	31140
FI	Artificial calibrated	5636	Water calibrated	36746
FR	Artificial HT	30368	Water HT	6292
FR	Artificial calibrated	30893	Water calibrated	6835
HR	Artificial HT	1728	Water HT	468
HR	Artificial calibrated	1785	Water calibrated	470
HU	Artificial HT	3624	Water HT	2013
HU	Artificial calibrated	3742	Water calibrated	1767
IE	Artificial HT	2976	Water HT	1402
IE	Artificial calibrated	2958	Water calibrated	1480
IT	Artificial HT	19562	Water HT	5310

Country	Estimate	Area_2018	Variable	Area_2018
IT	Artificial calibrated	19809	Water calibrated	5382
LT	Artificial HT	1402	Water HT	1572
LT	Artificial calibrated	1392	Water calibrated	1657
LU	Artificial HT	186	Water HT	17
LU	Artificial calibrated	189	Water calibrated	15
LV	Artificial HT	1085	Water HT	1925
LV	Artificial calibrated	1123	Water calibrated	1960
MT	Artificial HT	79	Water HT	4
MT	Artificial calibrated	87	Water calibrated	4
NL	Artificial HT	4555	Water HT	2485
NL	Artificial calibrated	4717	Water calibrated	3834
PL	Artificial HT	11275	Water HT	5556
PL	Artificial calibrated	11233	Water calibrated	5561
PT	Artificial HT	5660	Water HT	1557
PT	Artificial calibrated	5707	Water calibrated	1626
RO	Artificial HT	6672	Water HT	3826
RO	Artificial calibrated	6790	Water calibrated	3735
SE	Artificial HT	6911	Water HT	39956
SE	Artificial calibrated	7988	Water calibrated	39798
SI	Artificial HT	774	Water HT	109
SI	Artificial calibrated	880	Water calibrated	108
SK	Artificial HT	1697	Water HT	593
SK	Artificial calibrated	1679	Water calibrated	481
UK	Artificial HT	15717	Water HT	3332
UK	Artificial calibrated	15726	Water calibrated	5004

8 Timeliness and Punctuality

8. Timeliness and Punctuality

8.1. Timeliness

Data collection on the field takes place between spring and autumn on the year of the survey (t). LUCAS 2018 data field collection was completed in October 2018 and the Photointerpretation campaign in March 2019. The statistics are published according to the schedule of t+18months.

The first version of the LUCAS microdata is published in the summer after the survey at the latest. Successive versions of the microdata and/or the statistics become available after additional quality controls.

As soon as LUCAS 2018 had officially ended in March 2019 (end of PI campaign) almost at the same time post processing procedures on first data had been started. The dissemination phase of the LUCAS 2018 survey started in April 2019 with the release of the LUCAS primary (micro) data by EUROSTAT in LUCAS dedicated web page: https://ec.europa.eu/eurostat/web/lucas/data/primary-data/2018

The statistical tables encompass the estimates of Land cover and Land use by NUTS 2 regions as well as the Land cover for Forest FAO category and the Settlement Area. The corresponding units of measure contain the land cover/use by square kilometers, percentages, and their coefficient of variations. The first produced estimates were related to Settlement Area and Land Cover for Forest FAO, which had been disseminated in May 2020.

The dissemination of first LUCAS statistical tables took place in May 2020. Therefore, the interval between LUCAS official end survey date (T = March 2019) and preliminary results release (April 2019) is a month difference (T+1). As far as the first statistical tables is concern (first dissemination took place in May 2020), the corresponding interval is fourteen months (T+14). The final statistical tables had been disseminated in May 2021.

8.2. Punctuality

The punctuality of actual delivery of the data and the target date had been respected for LUCAS 2018 survey at 100% concerning primary data and the first statistical tables. The delivery of the final statistical tables was lagged by eight months. LUCAS webpage now contains six statistical tables, enhanced compared to previous campaigns, where each one of them involves a significant amount of measure units for every NUTS 2 region. The process of final dissemination, thus, contained a significant amount of data that needed to be checked as well as the performance of the new calibrated parameters. In addition, the 2018 final estimates had been the outcome of a new advanced methodological design that produced less common points than previous surveys. This has resulted in a break of series of final estimates in trend level analysis with previous rounds, which had been more evident for small land cover classes, such as artificial, water and wetland. Therefore, relevant crosscheck analysis had to be repeatedly performed.

9 Coherence and Comparability

9. Coherence and Comparability

9.1. Coherence

9.1.1. Coherence - cross domain

Coherence of statistics is their adequacy to be reliably combined in different ways and for various uses. Various sources of data currently provide information on land uses and agro-environmental topics. They include, among others, area sample surveys conducted by member States, NATURA 2000 maps and CORINE Land Cover (CLC). These sources are not often completely coherent with LUCAS data. While reading the results and comparing them with other sources it is important to have in mind that the LUCAS survey clearly distinguishes between land cover and land use. Despite the effort of harmonization of the definitions, some differences (sometimes not negligible) can be observed when comparing different sources. These differences can be due to the following reasons:

- Different methodologies;
- Certain margin of subjectivity in the application of the definitions;
- The (im)-possibility to clearly distinguish between coverage and use in the figures available from other domains;
- Variability of the estimates due to the sampling methodology.

Consequently, mapping LUCAS with other sources is really challenging. For instance, CLC and HRLs comprise five total land cover classes that are analyzed to a third digit level whilst in LUCAS the classes are eight. Additionally, the reference unit for LUCAS data is the point whilst for CORINE land cover and Copernicus HRLs, for instance, is the map extraction.

The coherence among other international and national land cover data sources is not an easy task due to above reasons. The mapping is even more difficult because the data collected by an international or organization could not be homogeneous. In addition, land cover classification also includes the agroforestry in some national sources. Despite the differences and under the scope of coherence and validation, it was possible to provide the forest areas from LUCAS datasets by adopting the FAO forest classification. In the following subsections, the main characteristics of coherence are outlined concerning the FAO, CLC.

9.1.1.1. FAO forest definitions

In general, the LC/LU classification is comparable with others LC/LU systems (e. FAO, CLC), hence compatibility of the adopted definitions with the main international concepts and definitions is guaranteed. Additional parameters though have been introduced where needed to allow the match, while keeping an independency and flexibility in the main item classification.

In LUCAS, Woodland has been defined in a way that allows providing estimates compatible with the FAO results. In particular, the comparability with FAO forest classification has been strengthened with the inclusion of variables area size, height of trees, width of features and percentage of land cover. However, differences between the semantic definition of LUCAS wooded areas and FAO forest definitions have to be taken into account: if an area has > 10% of trees (excluding fruit trees in permanent crops) in LUCAS is labeled as "wooded area", FAO take this into account only if it is > than 0.5 Ha. According to the above, it is expected higher forest values for FAO results than in LUCAS. However, the comparison between official FAO (FAOSTAT) and LUCAS FAO forest results (based on FAO classification) display great coherence and provide a quite satisfactory match.

9.1.1.2. CORINE Land Cover definitions

The CORINE Land classification comprises of five land cover classes further analysed to three sub levels totally account for 44 sub classes. Thus, different combination of CLC codes need to be included in order to be as much as possible coherent with LUCAS. In case of forest areas, the first digit classification “3 = Forest and semi-natural areas” includes areas that cannot be compared with LUCAS. However, the second digit of the CLC class “31=Forest” is considered as the closest to the definition adopted by LUCAS Woodland (C) class. Specifically, forest=31 class of CLC nomenclature consists of areas occupied by forests and woodlands with a vegetation pattern composed of native or exotic coniferous and/or deciduous trees and which could be used for the production of timber or other forest products. The forest trees are under normal climatic conditions higher than 5 m with a canopy closure of 30% at least. Similarly, in case of LUCAS (A) artificial the closest is CLC=1 artificial surfaces with the exclusion of 2nd digit = 14 (artificial, agricultural vegetated areas).

The second CLC class summarizes better part of LUCAS cropland, woodland, shrubland, grassland but unfortunately, that is still not sufficient to map classes. The map between CLC and LUCAS in third digit is required for some classes so to capture efficiently the similarities and improve the coherence. For instance in LUCAS (E) grassland, a combination of 3rd digit CLC codes from different 1st CLC codes needs to be considered as shown below:

- 142 - Sport and leisure facilities
- 243 - Agriculture mosaics with significant natural vegetation
- 321 - Natural grasslands
- 333 - Sparsely vegetated areas
- 141 - Green urban areas
- 231 - Pastures

Still discrepancies cannot be fully eliminated because of the different approach of land cover estimation. For instance if we consider LUCAS A (artificial) and CLC = 1 (Artificial surfaces) and exclude the 2nd CLC classes i.e. CLC13 (mine dump and construction sites) and 14 (artificial non-agricultural vegetated areas) that do not fully correspond to LUCAS artificial in terms of size area, the total CLC values are way higher. Similarly, LUCAS Artificial A1 (built-up) or even A11 (more than three floors) do not entirely match with CLC=11 (urban fabric) and third digit CLC=111 (continuous urban fabric) respectively. In addition, the inclusion of the different codes from CLC does not work in the same way for all countries and regions.

So, methodology, classification and definition in terms of what and how should a specific parcel be attributed remain the main questionable aspects from a coherence point of view. It should not also be forgotten that even if LUCAS microdata had been subjected to numerous quality checks both automatically and manually, therefore, the vast majority of LUCAS microdata are accurate and their classifications are robust, notwithstanding, CLC and HRLs datasets are both derived from minimum mapping units, and consequently the CLC polygons would always vary from LUCAS points.

9.1.2. Coherence - internal

The coherence between the total area of the countries and their split according to land cover and land use is guaranteed by definition. A standardized methodology and classification has been applied in all the countries and from one round to another since the 2006 pilot survey. Therefore, the internal coherence is perfectly assured.

9.2. Comparability

9.2.1. Comparability – geographical

The survey is fully harmonized and comparable, since the surveyors use the same methodology in all countries.

9.2.2. Comparability - over time

Different aspects of comparability have been assessed through:

- Comparison of the main features of 2009, 2012, 2015, and 2018 surveys by focusing on the following elements: sample design, sample size, countries involved, sampling unit and data collection method;

- Comparison of the information collected with the previous surveys (comparison of the variables reported in the field forms);

- Comparison of the definition of the variables collected with the previous surveys (information reported in the metadata and/or in the Technical Reference Documents).

The following table consolidates the relevant information collected of LUCAS surveys 2009 – 2018. It should also be noted that the information on linear features collected through transect had not been applied in 2018 round.

Table 37: Main features of LUCAS surveys 2009 - 2018

Item	2009	2012	2015	2018
Reference population	All EU Member States except for Bulgaria, Cyprus, Malta and Romania	All EU Member States	All EU Member States	All EU Member States
Sampling Unit	Point	Point	Point	Point
Sampling Scheme	Two-phase design with stratification	Two-phase design with stratification	Two-phase design with stratification	Two-phase design with stratification
First Phase Sample - Master Grid (size)	Less than 1 million (approx.)	Over 1 million (approx.)	Over 1 million (approx.)	Over 1 million (approx.)
Second Phase - Field Sample (size) (N. points surveyed)	234 500	270 000	273,000 field points + 67.000 (PI)	238 000 field points + 100 000 (PI)
Member States	23	27	28	28
Main information collected	Land Cover/Land Use details (i.e. height of trees, width of feature, plant species and degree of coverage (percentage); soil data; water management information and transect data. Soil	Land use data; land cover details (i.e. height of trees, width of feature, plant species and degree of coverage (percentage); soil data; water management information and transect data.	Land use data; land cover details (i.e. height of trees, width of feature, plant species and degree of coverage (percentage); soil data; water management information and transect data. soil	Land use data; land cover details (i.e. height of trees, width of feature, plant species and degree of coverage (percentage); soil data; water management information, extended soil module, grassland module, Copernicus programme
Stratification	Yes	Yes	Yes	Yes
Information collected in LF walking a transect	Yes	Yes	Yes	No

9.2.2.1. The break of series in 2018 data

The methodological revision of the 2018 sample design has resulted in relevant adjustments to the final estimates of statistical tables, and consequently, produced implications on the comparability of 2018 values with previous rounds. Any possible lack of consistency could be attributed in the 2018 sample design in comparison with the previous campaigns.

The 2018 sample design had commenced with the update of the Master by adding new elements, (CLC, HRLs, NUTS 2 totals, elevation classes) to count in the estimation of the target variables. A more detailed variable of STR18 has replaced the limited STR05 variable resulting to a higher number of strata. The build of 2018 stratification had been the outcome of possible combinations of NUTS 2, STR18, CLC modalities and elevation classes via an iterative algorithm that optimized the stratification given the coefficient of variations of the target variables. The sample design process had also been restructured in terms of the eligibility criteria. Until 2015 survey, eligibility criterion had been introduced to divide points as eligible or non-eligible, depending on whether a point was considered unattainable or too costly to be reached. It had thus divided the Master in two parts before the sample selection. However, the eligibility criterion has been removed in 2018 and all points were considered as eligible to be selected in the survey sample.

The consequences of the above modifications were the upshot of less common points in the 2018 sample than those selected in 2009-2015. In addition, a larger number of Photo-interpreted points had been assigned in 2018 survey comparing with previous rounds. Considering that Photointerpretation is different from in-situ observation, PI points are more likely to be less consistent with previous surveys. Therefore, possible variations are contingent particular for small countries/regions and land cover classes, such as artificial, shrubland, grassland, bareland, water, and wetland.

A final aspect is related to the comparability over time among NUTS 2 regions. The final estimates of last LUCAS survey have been estimated according to the NUTS 2 Classification of 2016. The latter involves the restructure of some specific NUTS 2 regions in France, Poland, Ireland, Lithuania, Croatia and Hungary, rending thus impossible the comparison over time for these regions.

10 Accessibility and Clarity

10. Accessibility and Clarity

10.1. News release

News releases are published periodically on-line.

10.2. Publications

Land Cover and Land Use 2018 articles are provided in Eurostat Statistics Explained section available at

https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Land_cover_statistics

https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Land_use_statistics

Moreover, LUCAS chapters have been included in the Eurostat yearbook and the Eurostat regional yearbook. In addition, publications related to technical documents, landscape indicators and metadata are available at LUCAS dedicated section

10.3. Online database

LUCAS statistical tables are available on Eurobase, under the land cover, land use and landscape LUCAS (lan) heading.

10.4. Microdata access

The LUCAS 2018 primary data are online available at:

http://ec.europa.eu/eurostat/web/lucas/data/primary-data/2018.

Similarly, previous LUCAS surveys (2015-2012-2009-2006) microdata are provided in:

https://ec.europa.eu/eurostat/web/lucas/data/primary-data/2015

https://ec.europa.eu/eurostat/web/lucas/data/primary-data/2012

https://ec.europa.eu/eurostat/web/lucas/data/primary-data/2009

https://ec.europa.eu/eurostat/web/lucas/data/primary-data/2006

Additionally, alphanumerical variables and photographs linked to the geo-referenced points are included. The LUCAS data are free of charge to all users.

The LUCAS photos can be obtained by contacting Eurostat (estat-user-support@ec.europa.eu) or by using the online order form. A LUCAS photo viewer allows visualizing maps of the data: http://ec.europa.eu/eurostat/web/lucas/lucas-photo-viewer.

10.5. Confidentiality - policy

Regulation (EC) No 223/2009 on European statistics (recital 24 and Article 20(4)) of 11 March 2009 (OJ L 87, p. 164), stipulates the need to establish common principles and guidelines ensuring the confidentiality of data used for the production of European statistics and the access to those confidential data with due account for technical developments and the requirements of users in a democratic society. LUCAS data are not confidential.

10.6. Other - Soil information

Concerning LUCAS topsoil data, the datasets for the LUCAS 2009 and 2012 Topsoil Module include data from 19,969 and 2034 samples respectively, from 25 Member States and can be downloaded from https://esdac.jrc.ec.europa.eu/content/lucas-2009-topsoil-data

The datasets for the LUCAS 2015 Topsoil Module include data from 21,859 samples from 28 Member States, together with reference data describing a range of environmental conditions for the LUCAS Soil locations can be downloaded from: https://esdac.jrc.ec.europa.eu/content/lucas2015-topsoil-data
Data were also collected during 2015 in Switzerland, Albania, Bosnia and Herzegovina, North Macedonia, Montenegro, Serbia collected during 2015 using the same methodology.

The datasets for the LUCAS 2018 Topsoil Module include data from 18,279 samples from 28 Member States will be available from https://esdac.jrc.ec.europa.eu during 2021. Data can be downloaded after prior registration.

10.7. Documentation on methodology

The required applicable documentation for LUCAS survey, aggregated data and landscape indicators is provided in the LUCAS dedicated page under methodology available at: https://ec.europa.eu/eurostat/web/lucas/methodology.
For the topsoil data, the paper "LUCAS Soil, the largest expandable soil dataset for Europe": A review by Orgiazzi et al provides a detailed insight into the design and methodology of the data collection and laboratory analysis at: https://onlinelibrary.wiley.com/doi/full/10.1111/ejss.12499
The LUCAS pages on ESDAC contain a wealth of supporting material and reports.

10.8. Quality documentation

All relevant quality documentation are available on LUCAS dedicated online session. More precisely, there are available:

LUCAS 2009 - Quality Checks

LUCAS 2012 - Quality Checks

LUCAS 2015 - Quality Checks

LUCAS 2018 - Quality Checks

11 References

11. References

-Baillargeon S. and Rivest L.-P. (2012). *The construction of stratified designs in R with the package stratification*. Survey Methodology, Vol. 37, No. 1, pp. 53-65

-Baillargeon S. and Rivest L.-P. (2014). *Stratification: Univariate Stratification of Survey Populations*. R package version 2.2-5. https://CRAN.R-project.org/package=stratification

-Ballin M., Barcaroli G. (2016). *Optimization of stratified sampling with the R package SamplingStrata: Applications to network data*. Computational Network Analysis with R: Applications in Biology, Medicine and Chemistry, Wiley

-Ballin M., Barcaroli G., Catanese E, D'Orazio M. (2016). *Stratification in Business and Agriculture Surveys with R*. Romanian Statistical Review 2/2016, pp. 43-58

Barcaroli G. (2018). *Optimization of sampling strata with the SamplingStrata package*. https://barcaroli.github.io/SamplingStrata/articles/SamplingStrata.html

-Barcaroli, G. (2014). *SamplingStrata: An R package for the optimization of stratified sampling*. Journal of Statistical Software 61 (4), 1-24.

-Barcaroli G., Pagliuca D., Willighagen E., Zardetto D. (2018). *SamplingStrata: Optimal stratification of sampling frames for multipurpose sampling surveys.* R package version 1.2. http://cran.rproject.org/web/packages/SamplingStrata/index.html

-Bethel J. (1985). *An Optimum Allocation Algorithm for Multivariate Surveys*. American Statistical Proceedings of the Survey Research Methods Section, pp. 209-212

-Bethel J. (1989). *Sample Allocation in Multivariate Surveys*. Survey Methodology, Vol. 15, pp. 47-57

-Chromy J.B. (1987). *Design Optimization with Multiple Objectives*. Proceedings of the American Statistical Association Section on Survey Research Methods 1987, pp. 194-199

-Cochran, W.G. (1977). *Sampling Techniques*. John Wiley & Sons, New York

-Copernicus High Resolution Layers – Methodology and HRLs available at: https://land.copernicus.eu/pan-european/high-resolution-layers

-CORINE Land Cover dedicated page: https://land.copernicus.eu/pan-european/corine-land-cover

-Dalenius, T., Hodges, J.L. (1959). *Minimum Variance Stratification*. Journal of American Statistical Association, Vol. 54, pp. 88-101

-Day C. D. (2006). *Application of an Evolutionary Algorithm to Multivariate Optimal Allocation in Stratified Sampling Designs*. Proceedings of the American Statistical Association Section on Survey Research Methods 2006 [CD-ROM]

-Day C. D. (2010). *A Multi-Objective Evolutionary Algorithm for Multivariate Optimal Allocation*. Section on Survey Research Methods JSM 2010 pp.3351-3358

-Decision No 1445/2000/EC of the European Parliament and of the Council of 22 May 2000 on the application of aerial-survey and remote sensing techniques to the agricultural statistics for 1999 to 2003. Available at: http://eur-lex.europa.eu/legalcontent/EN/TXT/?uri=CELEX:32000D1445

-Decision No 2066/2003/EC of the European Parliament and of the Council of 10 November 2003 on the continued application of aerial survey and remote-sensing techniques to the agricultural statistics for 2004 to 2007 and amending Decision 1445/2000/EC. Available at: http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32003D2066

-Decision No 786/2004/EC of the European Parliament and of the Council of 21 April 2004 amending Decisions No 1720/1999/EC, No 253/2000/EC, No 508/2000/EC, No 1031/2000/EC, No 1445/2000/EC, No 163/2001/EC, No 1411/2001/EC, No 50/2002/EC, No 466/2002/EC, No 1145/2002/EC, No 1513/2002/EC, No 1786/2002/EC, No 291/2003/EC and No 20/2004/EC with a view to adapting the reference amounts to take account of the enlargement of the European Union. Available at: http://eur-lex.europa.eu/legalcontent/EN/TXT/?uri=CELEX:32004D0786

-Eurostat (2020). *The European Statistical System (ESS) handbook for Quality and Metadata Reports –2020 Edition.* ISBN: 978-92-76-09154-7 Available at: https://ec.europa.eu/eurostat/documents/3859598/10501168/KS-GQ-19-006-EN-N.pdf/bf98fd32-f17c-31e2-8c7f-ad41eca91783?t=1583397712000

-Gonzales J.M., Eltinge J.L. (2010). *Optimal Survey Design: a Review*. Section on Survey Research Methods - JSM, October 2010 https://www.bls.gov/osmr/abstract/st/st100270.htm

-Gunning P., Horgan J.M. (2004). *A New Algorithm for the Construction of Stratum Boundaries in Skewed Populations*. Survey Methodology, Vol. 30, No. 2, pp. 159-166

-Hartigan, J.A., Wong M.A. (1979). *A k-means clustering algorithm*. Applied Statistics, 28, pp. 100108

-Keskinturk T., Er S. (2007). *A Genetic Algorithm Approach to determine Stratum Boundaries and Sample Sizes of Each Stratum in Stratified Sampling*. Computational Statistics and Data Analysis, Vol. 52, Issue 1, 15 September 2007, s.53-67.

-Kozak M., Wang H.Y. (2010). *On stochastic optimization in sample allocation among strata*. Metron – International Journal of Statistics 2010, vol. LXVIII, n.1, pp. 95-103

-Lavallée P., Hidiroglou M.A. (1988). *On the stratification of skewed populations*. Survey Methodology, Vol.14, pp.33-43

-LUCAS dedicated web page - http://ec.europa.eu/eurostat/web/lucas/overview

-LUCAS 2018 – *Technical reference document C1 - Instructions for Surveyors*, available at: https://ec.europa.eu/eurostat/documents/205002/8072634/LUCAS2018-C1-Instructions.pdf

-LUCAS 2018 – *Technical reference document C2 - Field Form and Ground Document*, available at: https://ec.europa.eu/eurostat/documents/205002/8072634/LUCAS2018-C2-FieldForm-GD-Template.pdf

-LUCAS 2018 – *Technical reference document C3 - Classification*, available at: https://ec.europa.eu/eurostat/documents/205002/8072634/LUCAS2018-C3-Classification.pdf

-LUCAS 2018 – *Technical reference document C4 – Quality Control Procedures*, available at: https://ec.europa.eu/eurostat/documents/205002/8072634/LUCAS2018-C4-QCProcedures.pdf

Orgiazzi, A., Ballabio, C., Panagos, P., Jones, A., Fernández-Ugalde, O. - *LUCAS Soil, the largest expandable soil dataset for Europe*, November 2017, https://onlinelibrary.wiley.com/doi/full/10.1111/ejss.12499

-Neyman, J. (1934). *On the two different aspects of the representative methods*. The method stratified sampling and the method of purposive selection. Journal of Royal Statistical Society, 97, 558-606.

-Vose M. D. (1999). *The Simple Genetic Algorithm: Foundations and Theory*, MIT Press, Cambridge, MA

12 ANNEXES

12. Annexes

Annex 1: Calculation of the FAO variable

FAO Forestry value	Condition FAO_CLASS	Original SQL syntax	SAS syntax
0	1	WHERE (((Export20160121.SURVEY_LC1) Like 'G*' Or (Export20160121.SURVEY_LC1) Like 'H*'));	if (upcase(substr(&lc1_name,1,1))='G' or upcase(substr(&lc1_name,1,1))='H')
1	2	WHERE (((Export20160121.SURVEY_LC1)='A22') AND ((Export20160121.SURVEY_LU1)='U312') AND ((Export20160121.SURVEY_LU2)='U120') AND ((Export20160121.FAO_CLASS) Is Null));	if (upcase(&lc1_name)='A22' and upcase(&lu1_name)='U312' and upcase(&lu2_name)='U120' and &fao_class_name='')
1	3	WHERE (((Export20160121.SURVEY_LC1)='A30') AND ((Export20160121.SURVEY_LC2) Like 'c*' Or (Export20160121.SURVEY_LC2) Like 'D*' Or (Export20160121.SURVEY_LC2) Like 'E*' Or (Export20160121.SURVEY_LC2) Like 'F*') AND ((Export20160121.SURVEY_LU1)='U319') AND ((Export20160121.SURVEY_LU2)='U120') AND ((Export20160121.FAO_CLASS) Is Null));	if (upcase(&lc1_name)='A30' and (upcase(substr(&lc2_name,1,1))='C' or upcase(substr(&lc2_name,1,1))='D' or upcase(substr(&lc2_name,1,1))='E' or upcase(substr(&lc2_name,1,1))='F') and upcase(&lu1_name)='U319' and upcase(&lu2_name)='U120' and &fao_class_name='')
3	4	WHERE (((Export20160121.SURVEY_LC1) Like 'B7*') AND ((Export20160121.SURVEY_LU1)='U111' Or (Export20160121.SURVEY_LU1)='U112' Or (Export20160121.SURVEY_LU1)='U113' Or (Export20160121.SURVEY_LU1) Like 'U4*') AND (Not (Export20160121.SURVEY_LC1_SPECIES)='B75E' And Not (Export20160121.SURVEY_LC1_SPECIES)='B75P') AND ((Export20160121.SURVEY_AREA_SIZE)>'1') AND ((Export20160121.FAO_CLASS) Is Null));	if (upcase(substr(&lc1_name,1,2))='B7' and (upcase(&lu1_name)='U111' or upcase(&lu1_name)='U112' or upcase(&lu1_name)='U113' or upcase(substr(&lu1_name,1,2))='U4') and (upcase(&lc1_species_name) ne 'B75E' and upcase(&lc1_species_name) ne 'B75P') and &survey_area_size_name>1 and &fao_class_name='')
3	5	WHERE (((Export20160121.FAO_CLASS) Is Null) AND ((Export20160121.SURVEY_LC1)='B81') AND ((Export20160121.SURVEY_LU1)='U111' Or (Export20160121.SURVEY_LU1)='U112' Or (Export20160121.SURVEY_LU1)='U113' Or (Export20160121.SURVEY_LU1) Like 'U4*') AND ((Export20160121.SURVEY_AREA_SIZE)>'1'));	if (upcase(&lc1_name)='B81' and (upcase(&lu1_name)='U111' or upcase(&lu1_name)='U112' or upcase(&lu1_name)='U113' or upcase(substr(&lu1_name,1,2))='U4') and &survey_area_size_name>1 and &fao_class_name='')

1	6	WHERE (((Export20160121.FAO_CLASS) Is Null) AND ((Export20160121.SURVEY_LC1_SPECIES)='B83F') AND ((Export20160121.SURVEY_AREA_SIZE)>'1'));	if (upcase(&lc1_species_name)='B83F' and &survey_area_size_name>1 and &fao_class_name='')
1	7	WHERE (((Export20160121.FAO_CLASS) Is Null) AND ((Export20160121.SURVEY_LC1) Like 'C*') AND ((Export20160121.SURVEY_LC2)='8') AND ((Export20160121.SURVEY_LU1)='U111') AND ((Export20160121.SURVEY_LU2)='8') AND ((Export20160121.SURVEY_AREA_SIZE)>'1') AND ((Export20160121.SURVEY_TREE_HEIGHT_MATURITY)>'1') AND ((Export20160121.SURVEY_FEATURE_WIDTH)>'1'));	if (upcase(substr(&lc1_name,1,1))='C' and &lc2_name='8' and upcase(&lu1_name)='U111' and &lu2_name='8' and &survey_area_size_name>1 and &survey_tree_height_maturity_name >1 and &survey_feature_width_name>1 and &fao_class_name='')
3	8_1	WHERE (((Export20160121.FAO_CLASS) Is Null) AND ((Export20160121.SURVEY_LC1) Like 'C*') AND ((Export20160121.SURVEY_LU1)='U111' Or (Export20160121.SURVEY_LU1)='U112' Or (Export20160121.SURVEY_LU1)='U113') AND ((Export20160121.SURVEY_AREA_SIZE)>'1') AND ((Export20160121.SURVEY_TREE_HEIGHT_MATURITY)>'1') AND ((Export20160121.SURVEY_FEATURE_WIDTH)>'1')) OR (((Export20160121.FAO_CLASS) Is Null) AND ((Export20160121.SURVEY_LC1) Like 'C*') AND ((Export20160121.SURVEY_LU2)='U111' Or (Export20160121.SURVEY_LU2)='U112' Or (Export20160121.SURVEY_LU2)='U113') AND ((Export20160121.SURVEY_AREA_SIZE)>'1') AND ((Export20160121.SURVEY_TREE_HEIGHT_MATURITY)>'1') AND ((Export20160121.SURVEY_FEATURE_WIDTH)>'1')) OR (((Export20160121.FAO_CLASS) Is Null) AND ((Export20160121.SURVEY_LC1) Like 'C*') AND ((Export20160121.SURVEY_LC2) Like 'B*') AND ((Export20160121.SURVEY_LU1) Like 'U4*' Or (Export20160121.SURVEY_LU1)='U120') AND ((Export20160121.SURVEY_AREA_SIZE)>'1') AND ((Export20160121.SURVEY_TREE_HEIGHT_MATURITY)>'1') AND ((Export20160121.SURVEY_FEATURE_WIDTH)>'1'));	if (upcase(substr(&lc1_name,1,1))='C' and (upcase(&lu1_name)='U111' or upcase(&lu1_name)='U112' or upcase(&lu1_name)='U113') and &survey_area_size_name>1 and &survey_tree_height_maturity_name >1 and &survey_feature_width_name>1 and &fao_class_name='')
	8_2		if (upcase(substr(&lc1_name,1,1))='C' and (upcase(&lu2_name)='U111' or upcase(&lu2_name)='U112' or upcase(&lu2_name)='U113') and &survey_area_size_name>1 and &survey_tree_height_maturity_name >1 and &survey_feature_width_name>1 and &fao_class_name='')
	8_3		if (upcase(substr(&lc1_name,1,1))='C' and upcase(substr(&lc2_name,1,1))='B' and (upcase(substr(&lu1_name,1,2))='U4' or upcase(&lu1_name)='U120') and &survey_area_size_name>1 and &survey_tree_height_maturity_name >1 and &survey_feature_width_name>1 and &fao_class_name='')
1	9_1	WHERE (((Export20160121.FAO_CLASS) Is Null) AND ((Export20160121.SURVEY_LC1) Like 'C*') AND ((Export20160121.SURVEY_LU1)='U120') AND ((Export20160121.SURVEY_LU2)='8' Or (Export20160121.SURVEY_LU2)='U140' Or (Export20160121.SURVEY_LU2)='U150' Or (Export20160121.SURVEY_LU2)='U318' Or (Export20160121.SURVEY_LU2)='U321' Or (Export20160121.SURVEY_LU2)='U322' Or (Export20160121.SURVEY_LU2)='U350' Or (Export20160121.SURVEY_LU2)='U361' Or (Export20160121.SURVEY_LU2)='U362' Or (Export20160121.SURVEY_LU2)='U370') AND ((Export20160121.SURVEY_AREA_SIZE)>'1') AND ((Export20160121.SURVEY_TREE_HEIGHT_MATURITY)>'1') AND ((Export20160121.SURVEY_FEATURE_WIDTH)>'1'))	if (upcase(substr(&lc1_name,1,1))='C' and upcase(&lu1_name)='U120' and (upcase(&lu2_name)='8' or upcase(&lu2_name)='U140' or upcase(&lu2_name)='U150' or upcase(&lu2_name)='U318' or upcase(&lu2_name)='U321' or upcase(&lu2_name)='U322' or upcase(&lu2_name)='U350' or upcase(&lu2_name)='U361' or upcase(&lu2_name)='U362' or upcase(&lu2_name)='U370') and &survey_area_size_name>1 and &survey_tree_height_maturity_name >1 and &survey_feature_width_name>1 and &fao_class_name='')

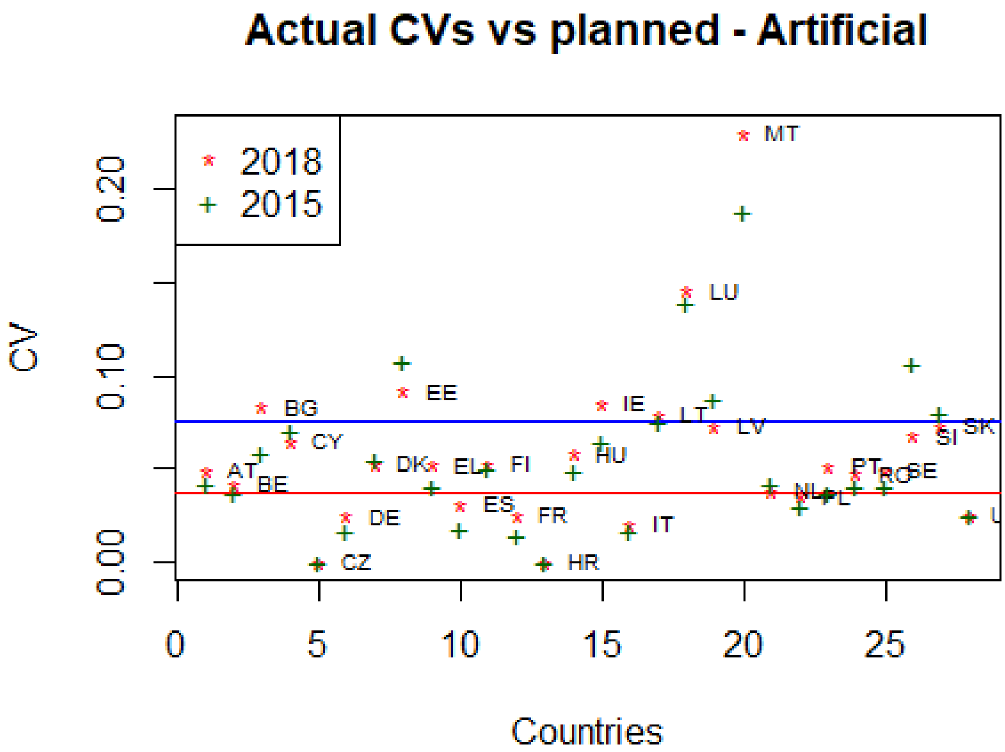
	9_2	OR (((Export20160121.FAO_CLASS) Is Null) AND ((Export20160121.SURVEY_LC1) Like 'C*') AND ((Export20160121.SURVEY_LU1)='U140' Or (Export20160121.SURVEY_LU1)='U150' Or (Export20160121.SURVEY_LU1) Like 'U4*') AND ((Export20160121.SURVEY_LU2)='8') AND ((Export20160121.SURVEY_AREA_SIZE)>'1') AND ((Export20160121.SURVEY_TREE_HEIGHT_MATURITY)>'1') AND ((Export20160121.SURVEY_FEATURE_WIDTH)>'1')) OR (((Export20160121.FAO_CLASS) Is Null) AND ((Export20160121.SURVEY_LC1) Like 'C*') AND ((Export20160121.SURVEY_LU1)='U350') AND ((Export20160121.SURVEY_LU2)='8' Or (Export20160121.SURVEY_LU2)='U120') AND ((Export20160121.SURVEY_AREA_SIZE)>'1') AND ((Export20160121.SURVEY_TREE_HEIGHT_MATURITY)>'1') AND ((Export20160121.SURVEY_FEATURE_WIDTH)>'1')) OR (((Export20160121.FAO_CLASS) Is Null) AND ((Export20160121.SURVEY_LC1) Like 'C*') AND ((Export20160121.SURVEY_LU1)='U318') AND ((Export20160121.SURVEY_LU2)='8') AND ((Export20160121.SURVEY_AREA_SIZE)>'1') AND ((Export20160121.SURVEY_TREE_HEIGHT_MATURITY)>'1') AND ((Export20160121.SURVEY_FEATURE_WIDTH)>'1'));	if (upcase(substr(&lc1_name,1,1))='C' and (upcase(&lu1_name)='U140' or upcase(&lu1_name)='U150' or upcase(substr(&lu1_name,1,2))='U4') and upcase(&lu2_name)='8' and &survey_area_size_name>1 and &survey_tree_height_maturity_name >1 and &survey_feature_width_name>1 and &fao_class_name='')
	9_3		if (upcase(substr(&lc1_name,1,1))='C' and upcase(&lu1_name)='U350' and (upcase(&lu2_name)='8' or upcase(&lu2_name)='U120') and &survey_area_size_name>1 and &survey_tree_height_maturity_name >1 and &survey_feature_width_name>1 and &fao_class_name='')
	9_4		if (upcase(substr(&lc1_name,1,1))='C' and upcase(&lu1_name)='U318' and upcase(&lu2_name)='8' and &survey_area_size_name>1 and &survey_tree_height_maturity_name >1 and &survey_feature_width_name>1 and &fao_class_name='')
2	10_1	WHERE (((Export20160121.FAO_CLASS) Is Null) AND ((Export20160121.SURVEY_LC1) Like 'C*') AND ((Export20160121.SURVEY_LU1)='U120') AND ((Export20160121.SURVEY_LU2)='8' Or (Export20160121.SURVEY_LU2)='U140' Or (Export20160121.SURVEY_LU2)='U150' Or (Export20160121.SURVEY_LU2)='U318' Or (Export20160121.SURVEY_LU2)='U321' Or (Export20160121.SURVEY_LU2)='U322' Or (Export20160121.SURVEY_LU2)='U350' Or (Export20160121.SURVEY_LU2)='U361' Or (Export20160121.SURVEY_LU2)='U362' Or (Export20160121.SURVEY_LU2)='U370') AND ((Export20160121.SURVEY_AREA_SIZE)>'1') AND ((Export20160121.SURVEY_TREE_HEIGHT_MATURITY)='1')) OR (((Export20160121.FAO_CLASS) Is Null) AND ((Export20160121.SURVEY_LC1) Like 'C*') AND ((Export20160121.SURVEY_LU1)='U140' Or (Export20160121.SURVEY_LU1)='U150' Or (Export20160121.SURVEY_LU1) Like 'U4*') AND ((Export20160121.SURVEY_LU2)='8') AND ((Export20160121.SURVEY_AREA_SIZE)>'1') AND ((Export20160121.SURVEY_TREE_HEIGHT_MATURITY)='1')) OR (((Export20160121.FAO_CLASS) Is Null) AND ((Export20160121.SURVEY_LC1) Like 'C*') AND ((Export20160121.SURVEY_LU1)='U350') AND ((Export20160121.SURVEY_LU2)='8' Or (Export20160121.SURVEY_LU2)='U120') AND ((Export20160121.SURVEY_AREA_SIZE)>'1') AND ((Export20160121.SURVEY_TREE_HEIGHT_MATURITY)='1')) OR (((Export20160121.FAO_CLASS) Is Null) AND ((Export20160121.SURVEY_LC1) Like 'D*' Or	if (upcase(substr(&lc1_name,1,1))='C' and upcase(&lu1_name)='U120' and (upcase(&lu2_name)='8' or upcase(&lu2_name)='U140' or upcase(&lu2_name)='U150' or upcase(&lu2_name)='U318' or upcase(&lu2_name)='U321' or upcase(&lu2_name)='U322' or upcase(&lu2_name)='U350' or upcase(&lu2_name)='U361' or upcase(&lu2_name)='U362' or upcase(&lu2_name)='U370') and &survey_area_size_name>1 and &survey_tree_height_maturity_name =1 and &fao_class_name='')
	10_2		if (upcase(substr(&lc1_name,1,1))='C' and (upcase(&lu1_name)='U140' or upcase(&lu1_name)='U150' or upcase(substr(&lu1_name,1,2))='U4') and upcase(&lu2_name)='8' and &survey_area_size_name>1 and &survey_tree_height_maturity_name =1 and &fao_class_name='')
	10_3		if (upcase(substr(&lc1_name,1,1))='C' and upcase(&lu1_name)='U350' and (upcase(&lu2_name)='8' or upcase(&lu2_name)='U120') and &survey_area_size_name>1 and &survey_tree_height_maturity_name =1 and &fao_class_name='')

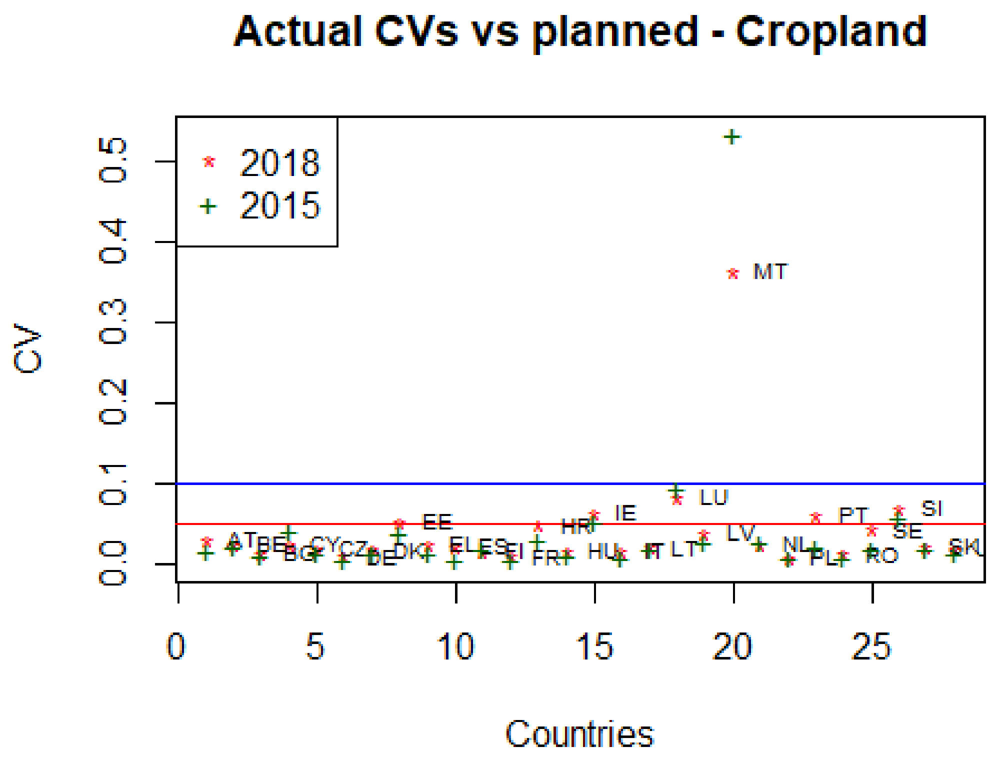
<table>
<tr><td rowspan="3"></td><td>10_4</td><td rowspan="3">(Export20160121.SURVEY_LC1)='E10') AND ((Export20160121.SURVEY_LU1)='U120') AND ((Export20160121.SURVEY_LU2)='8' Or (Export20160121.SURVEY_LU2)='U140' Or (Export20160121.SURVEY_LU2)='U150' Or (Export20160121.SURVEY_LU2)='U318' Or (Export20160121.SURVEY_LU2)='U321' Or (Export20160121.SURVEY_LU2)='U322' Or (Export20160121.SURVEY_LU2)='U350' Or (Export20160121.SURVEY_LU2)='U361' Or (Export20160121.SURVEY_LU2)='U362' Or (Export20160121.SURVEY_LU2)='U370') AND ((Export20160121.SURVEY_AREA_SIZE)>'1') AND ((Export20160121.SURVEY_TREE_HEIGHT_MATURITY)>'1') AND ((Export20160121.SURVEY_FEATURE_WIDTH)>'1')) OR (((Export20160121.FAO_CLASS) Is Null) AND ((Export20160121.SURVEY_LC1) Like 'D*' Or (Export20160121.SURVEY_LC1)='E10') AND ((Export20160121.SURVEY_LU1)='U140' Or (Export20160121.SURVEY_LU1)='U150' Or (Export20160121.SURVEY_LU1) Like 'U4*') AND ((Export20160121.SURVEY_LU2)='8') AND ((Export20160121.SURVEY_AREA_SIZE)>'1') AND ((Export20160121.SURVEY_TREE_HEIGHT_MATURITY)>'1') AND ((Export20160121.SURVEY_FEATURE_WIDTH)>'1')) OR (((Export20160121.FAO_CLASS) Is Null) AND ((Export20160121.SURVEY_LC1) Like 'D*' Or (Export20160121.SURVEY_LC1)='E10') AND ((Export20160121.SURVEY_LU1)='U350') AND ((Export20160121.SURVEY_LU2)='8' Or (Export20160121.SURVEY_LU2)='U120') AND ((Export20160121.SURVEY_AREA_SIZE)>'1') AND ((Export20160121.SURVEY_TREE_HEIGHT_MATURITY)>'1') AND ((Export20160121.SURVEY_FEATURE_WIDTH)>'1'));</td><td>if ((upcase(substr(&lc1_name,1,1))='D' or upcase(&lc1_name)='E10') and upcase(&lu1_name)='U120' and (upcase(&lu2_name)='8' or upcase(&lu2_name)='U140' or upcase(&lu2_name)='U150' or upcase(&lu2_name)='U318' or upcase(&lu2_name)='U321' or upcase(&lu2_name)='U322' or upcase(&lu2_name)='U350' or upcase(&lu2_name)='U361' or upcase(&lu2_name)='U362' or upcase(&lu2_name)='U370') and &survey_area_size_name>1 and &survey_tree_height_maturity_name >1 and &survey_feature_width_name>1 and &fao_class_name='')</td></tr>
<tr><td>10_5</td><td>if ((upcase(substr(&lc1_name,1,1))='D' or upcase(&lc1_name)='E10') and (upcase(&lu1_name)='U140' or upcase(&lu1_name)='U150' or upcase(substr(&lu1_name,1,2))='U4') and upcase(&lu2_name)='8' and &survey_area_size_name>1 and &survey_tree_height_maturity_name >1 and &survey_feature_width_name>1 and &fao_class_name='')</td></tr>
<tr><td>10_6</td><td>if ((upcase(substr(&lc1_name,1,1))='D' or upcase(&lc1_name)='E10') and upcase(&lu1_name)='U350' and (upcase(&lu2_name)='8' or upcase(&lu2_name)='U120') and &survey_area_size_name>1 and &survey_tree_height_maturity_name >1 and &survey_feature_width_name>1 and &fao_class_name='')</td></tr>
<tr><td rowspan="2">1</td><td>11_1</td><td rowspan="2">WHERE (((Export20160121.FAO_CLASS) Is Null) AND ((Export20160121.SURVEY_LC1)='D10') AND ((Export20160121.SURVEY_LU1)='U120') AND ((Export20160121.SURVEY_LU2)='8' Or (Export20160121.SURVEY_LU2)='U140' Or (Export20160121.SURVEY_LU2)='U150' Or (Export20160121.SURVEY_LU2)='U318' Or (Export20160121.SURVEY_LU2)='U321' Or (Export20160121.SURVEY_LU2)='U322' Or (Export20160121.SURVEY_LU2)='U350' Or (Export20160121.SURVEY_LU2)='U361' Or (Export20160121.SURVEY_LU2)='U362' Or (Export20160121.SURVEY_LU2)='U370') AND ((Export20160121.SURVEY_AREA_SIZE)='1')) OR (((Export20160121.FAO_CLASS) Is Null) AND ((Export20160121.SURVEY_LC1)='D10') AND ((Export20160121.SURVEY_LU1) Like 'U4*') AND</td><td>if (upcase(&lc1_name)='D10' and upcase(&lu1_name)='U120' and (upcase(&lu2_name)='8' or upcase(&lu2_name)='U140' or upcase(&lu2_name)='U150' or upcase(&lu2_name)='U318' or upcase(&lu2_name)='U321' or upcase(&lu2_name)='U322' or upcase(&lu2_name)='U350' or upcase(&lu2_name)='U361' or upcase(&lu2_name)='U362' or upcase(&lu2_name)='U370') and &survey_area_size_name=1 and &fao_class_name='')</td></tr>
<tr><td>11_2</td><td>if (upcase(&lc1_name)='D10' and upcase(substr(&lu1_name,1,2))='U4' and upcase(&lu2_name)='8' and</td></tr>
</table>

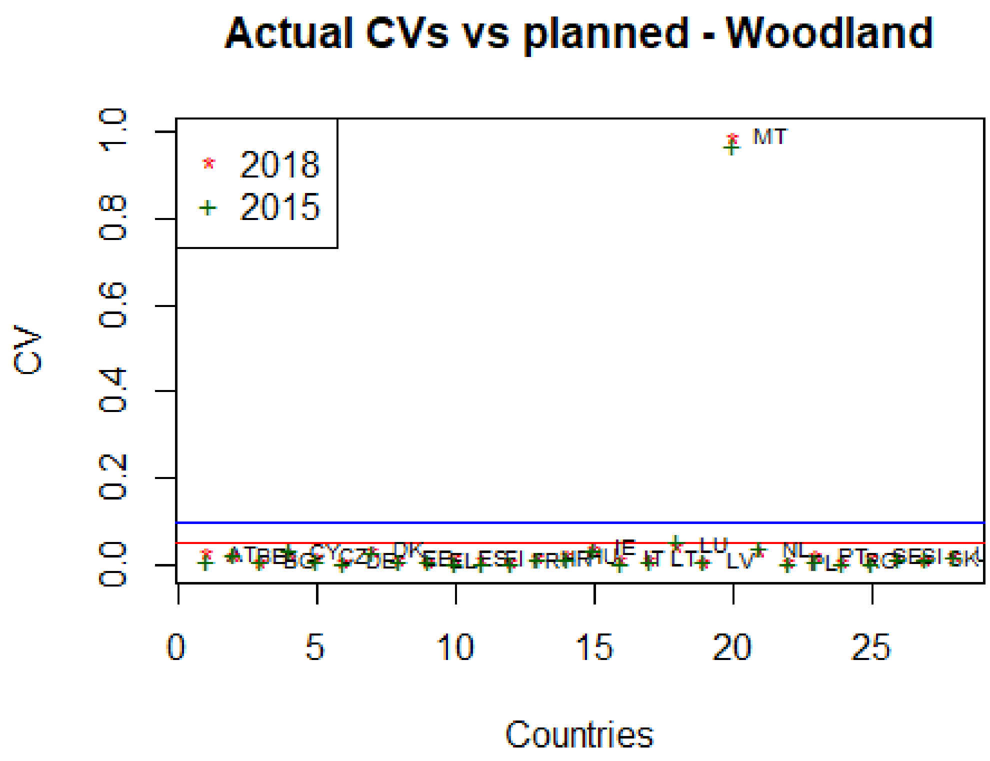
<table>
<tr><td rowspan="4"></td><td></td><td rowspan="4">((Export20160121.SURVEY_LU2)='8') AND ((Export20160121.SURVEY_AREA_SIZE)='1')) OR (((Export20160121.FAO_CLASS) Is Null) AND ((Export20160121.SURVEY_LC1)='E10') AND ((Export20160121.SURVEY_LU1)='U120') AND ((Export20160121.SURVEY_LU2)='8' Or (Export20160121.SURVEY_LU2)='U140' Or (Export20160121.SURVEY_LU2)='U150' Or (Export20160121.SURVEY_LU2)='U318' Or (Export20160121.SURVEY_LU2)='U321' Or (Export20160121.SURVEY_LU2)='U322' Or (Export20160121.SURVEY_LU2)='U350' Or (Export20160121.SURVEY_LU2)='U361' Or (Export20160121.SURVEY_LU2)='U362' Or (Export20160121.SURVEY_LU2)='U370') AND ((Export20160121.SURVEY_AREA_SIZE)='1')) OR (((Export20160121.FAO_CLASS) Is Null) AND ((Export20160121.SURVEY_LC1)='E10') AND ((Export20160121.SURVEY_LU1) Like 'U4*') AND ((Export20160121.SURVEY_LU2)='8') AND ((Export20160121.SURVEY_AREA_SIZE)='1')) OR (((Export20160121.FAO_CLASS) Is Null) AND ((Export20160121.SURVEY_LC1) Like "C*") AND ((Export20160121.SURVEY_LU1)='U120') AND ((Export20160121.SURVEY_LU2)='8' Or (Export20160121.SURVEY_LU2)='U140' Or (Export20160121.SURVEY_LU2)='U150' Or (Export20160121.SURVEY_LU2)='U318' Or (Export20160121.SURVEY_LU2)='U321' Or (Export20160121.SURVEY_LU2)='U322' Or (Export20160121.SURVEY_LU2)='U350' Or (Export20160121.SURVEY_LU2)='U361' Or (Export20160121.SURVEY_LU2)='U362' Or (Export20160121.SURVEY_LU2)='U370') AND ((Export20160121.SURVEY_AREA_SIZE)='1'));</td><td>&survey_area_size_name=1 and &fao_class_name=''</td></tr>
<tr><td>11_3</td><td>if (upcase(&lc1_name)='E10' and upcase(&lu1_name)='U120' and (upcase(&lu2_name)='8' or upcase(&lu2_name)='U140' or upcase(&lu2_name)='U150' or upcase(&lu2_name)='U318' or upcase(&lu2_name)='U321' or upcase(&lu2_name)='U322' or upcase(&lu2_name)='U350' or upcase(&lu2_name)='U361' or upcase(&lu2_name)='U362' or upcase(&lu2_name)='U370') and &survey_area_size_name=1 and &fao_class_name='')</td></tr>
<tr><td>11_4</td><td>if (upcase(&lc1_name)='E10' and upcase(substr(&lu1_name,1,2))='U4' and upcase(&lu2_name)='8' and &survey_area_size_name=1 and &fao_class_name='')</td></tr>
<tr><td>11_5</td><td>if (upcase(substr(&lc1_name,1,1))='C' and upcase(&lu1_name)='U120' and (upcase(&lu2_name)='8' or upcase(&lu2_name)='U140' or upcase(&lu2_name)='U150' or upcase(&lu2_name)='U318' or upcase(&lu2_name)='U321' or upcase(&lu2_name)='U322' or upcase(&lu2_name)='U350' or upcase(&lu2_name)='U361' or upcase(&lu2_name)='U362' or upcase(&lu2_name)='U370') and &survey_area_size_name=1 and &fao_class_name='')</td></tr>
<tr><td rowspan="2">3</td><td>12_1</td><td rowspan="2">WHERE (((Export20160121.FAO_CLASS) Is Null) AND ((Export20160121.SURVEY_LC1) Like 'C*') AND ((Export20160121.SURVEY_LU1) Like 'U2*' Or (Export20160121.SURVEY_LU1) Like 'U31*' Or (Export20160121.SURVEY_LU1) Like 'U32*' Or (Export20160121.SURVEY_LU1) Like 'U34*' Or (Export20160121.SURVEY_LU1) Like 'U36*' Or (Export20160121.SURVEY_LU1)='U370') AND ((Export20160121.SURVEY_AREA_SIZE)>'1') AND ((Export20160121.SURVEY_TREE_HEIGHT_MATURITY)>'1') AND ((Export20160121.SURVEY_FEATURE_WIDTH)>'1')) OR (((Export20160121.FAO_CLASS) Is Null) AND ((Export20160121.SURVEY_LC1) Like 'C*') AND ((Export20160121.SURVEY_LU1)='U350') AND ((Export20160121.SURVEY_AREA_SIZE)>'1') AND ((Export20160121.SURVEY_TREE_HEIGHT_MATURITY)>'1') AND ((Export20160121.SURVEY_FEATURE_WIDTH)>'1') AND ((Export20160121.SURVEY_LU2)='U361' Or (Export20160121.SURVEY_LU2)='U362'));</td><td>if (upcase(substr(&lc1_name,1,1))='C' and (upcase(substr(&lu1_name,1,2))='U2' or upcase(substr(&lu1_name,1,3))='U31' or upcase(substr(&lu1_name,1,3))='U32' or upcase(substr(&lu1_name,1,3))='U34' or upcase(substr(&lu1_name,1,3))='U36' or upcase(&lu1_name)='U370') and &survey_area_size_name>1 and &survey_tree_height_maturity_name >1 and &survey_feature_width_name>1 and &fao_class_name='')</td></tr>
<tr><td>12_2</td><td>if (upcase(substr(&lc1_name,1,1))='C' and upcase(&lu1_name)='U350' and &survey_area_size_name>1 and &survey_tree_height_maturity_name >1 and &survey_feature_width_name>1 and (upcase(&lu2_name)='U361' or</td></tr>
</table>

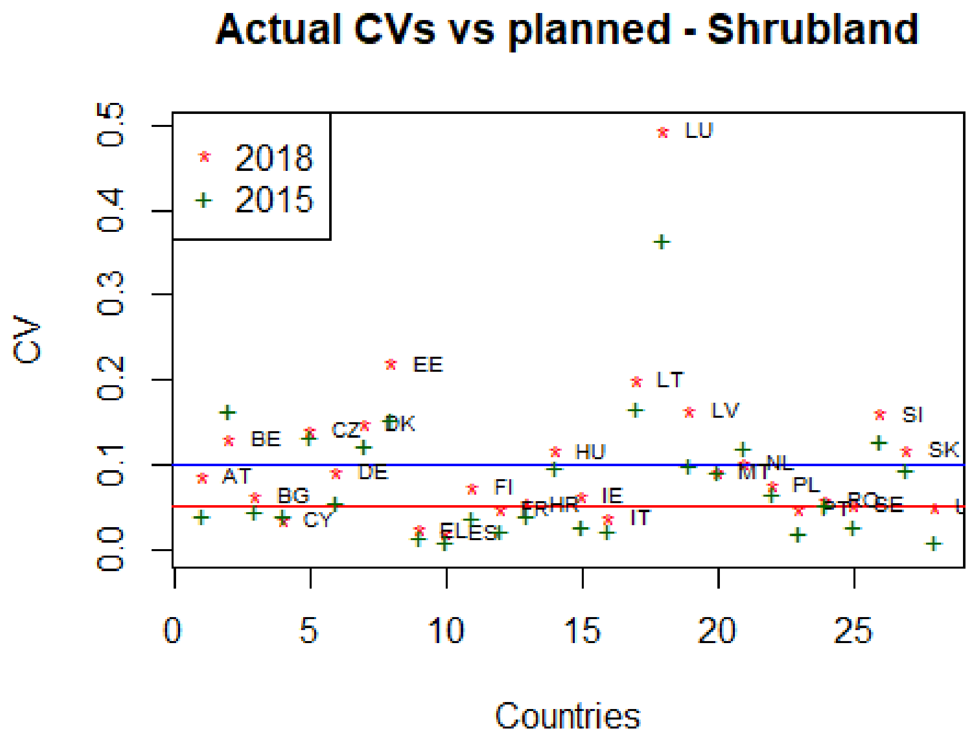
			upcase(&lu2_name)='U362') and &fao_class_name='')
1	13	WHERE (((Export20160121.FAO_CLASS) Is Null) AND ((Export20160121.SURVEY_LU1)='U120') AND ((Export20160121.SURVEY_LC_LU_SPECIAL_REMARK)='3' Or (Export20160121.SURVEY_LC_LU_SPECIAL_REMARK)='4' Or (Export20160121.SURVEY_LC_LU_SPECIAL_REMARK)='5'));	if (upcase(&lu1_name)='U120' and (&survey_lc_lu_special_remark_name =3 or &survey_lc_lu_special_remark_name= 4 or &survey_lc_lu_special_remark_name= 5) and &fao_class_name='')
0	In all the other cases		

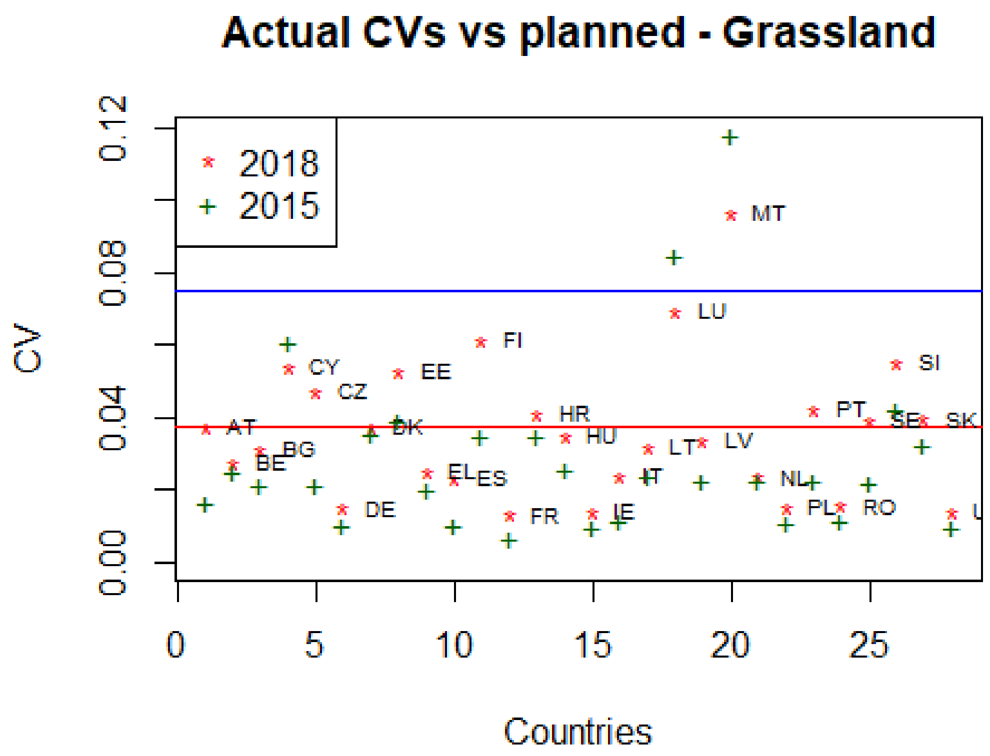
Annex 2: Coefficients of Variations for the Land Cover

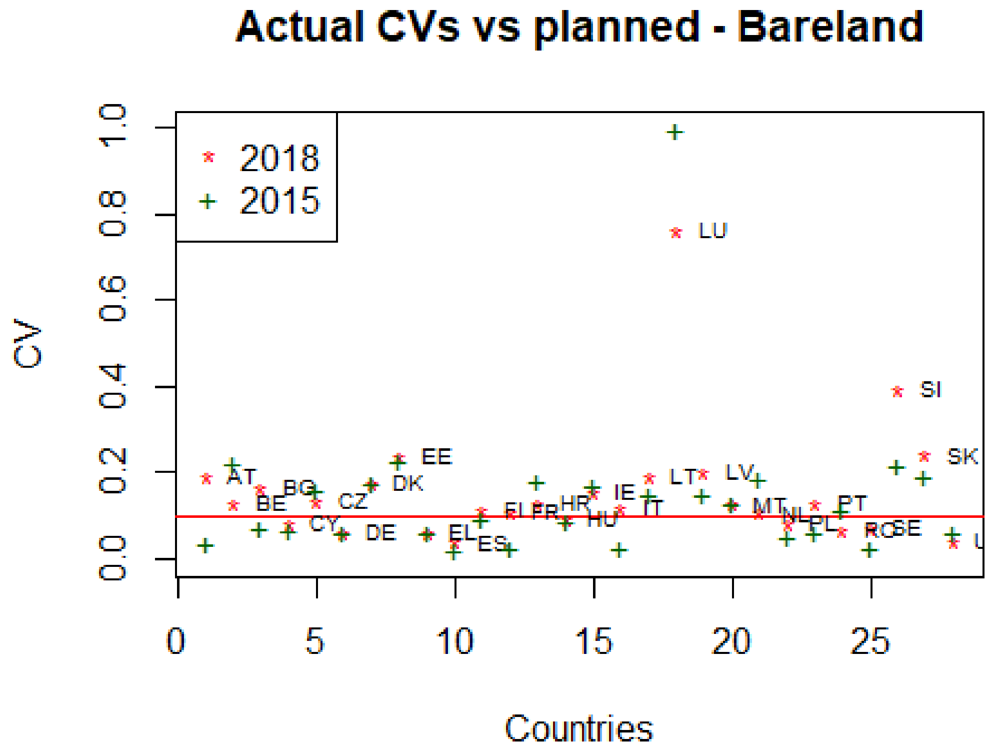


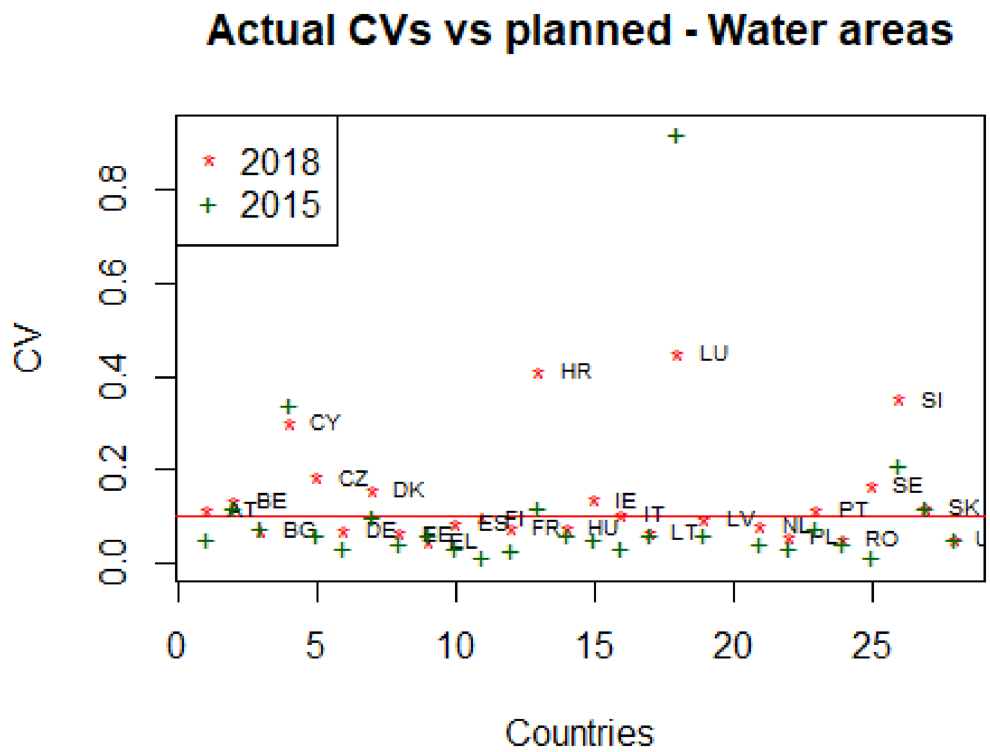

Actual CVs vs planned - Cropland
2018
2015
CV
0.0
0.1
0.2
0.3
0.4
0.5
0
5
10
15
20
25
Countries
MT
LU
IE
EE
HR
PT
SI
SE
LV
NL
RO
SK
PL
HU
IT
LT
FR
FI
EL
ES
DK
DE
CY
CZ
BG
AT
BE

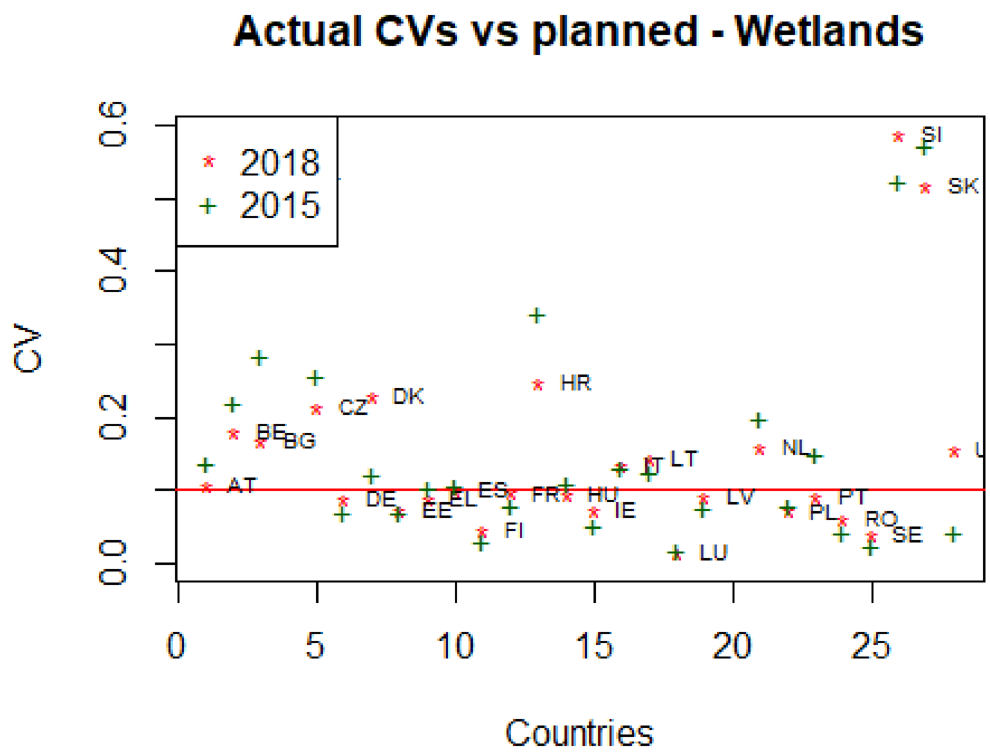

Actual CVs vs planned - Woodland
2018
2015
CV
0.0
0.2
0.4
0.6
0.8
1.0
0
5
10
15
20
25
Countries
MT
LU
IE
DK
NL
LV
LT
IT
PT
RO
SE
SI
SK
PL
HU
HR
FR
FI
EL
ES
EE
DE
CY
CZ
BG
AT
BE


Actual CVs vs planned - Shrubland
2018
2015
CV
Countries
0.0
0.1
0.2
0.3
0.4
0.5
0
5
10
15
20
25
LU
EE
LT
LV
SI
SK
BE
CZ
DK
HU
NL
MT
AT
DE
FI
BG
IE
PL
RO
SE
CY
IT


Actual CVs vs planned - Grassland
2018
2015
CV
Countries
0.00
0.04
0.08
0.12
0
5
10
15
20
25
MT
LU
FI
CY
EE
SI
CZ
HR
PT
SK
AT
DK
HU
LV
LT
BG
BE
ES
NL
DE
FR
IE
PL
RO

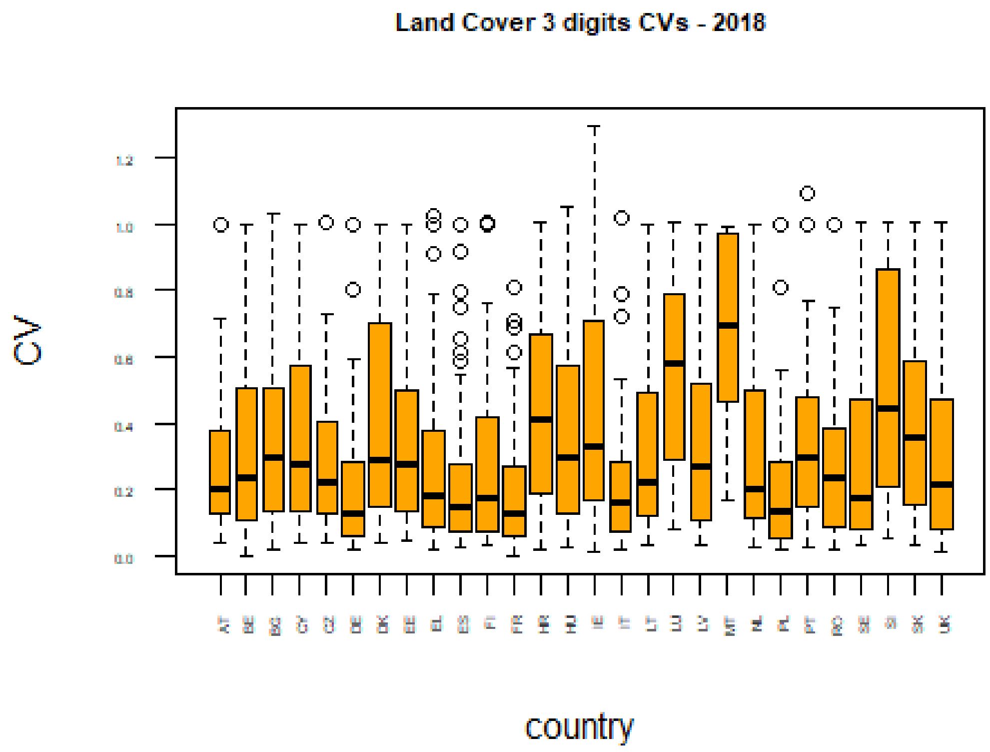

Actual CVs vs planned - Bareland
2018
2015
CV
0.0
0.2
0.4
0.6
0.8
1.0
0
5
10
15
20
25
Countries
LU
SI
SK
EE
AT
BG
DK
BE
CZ
CY
DE
EL
ES
FI
FR
HR
HU
IE
IT
LT
LV
MT
NL
PL
PT
RO
SE

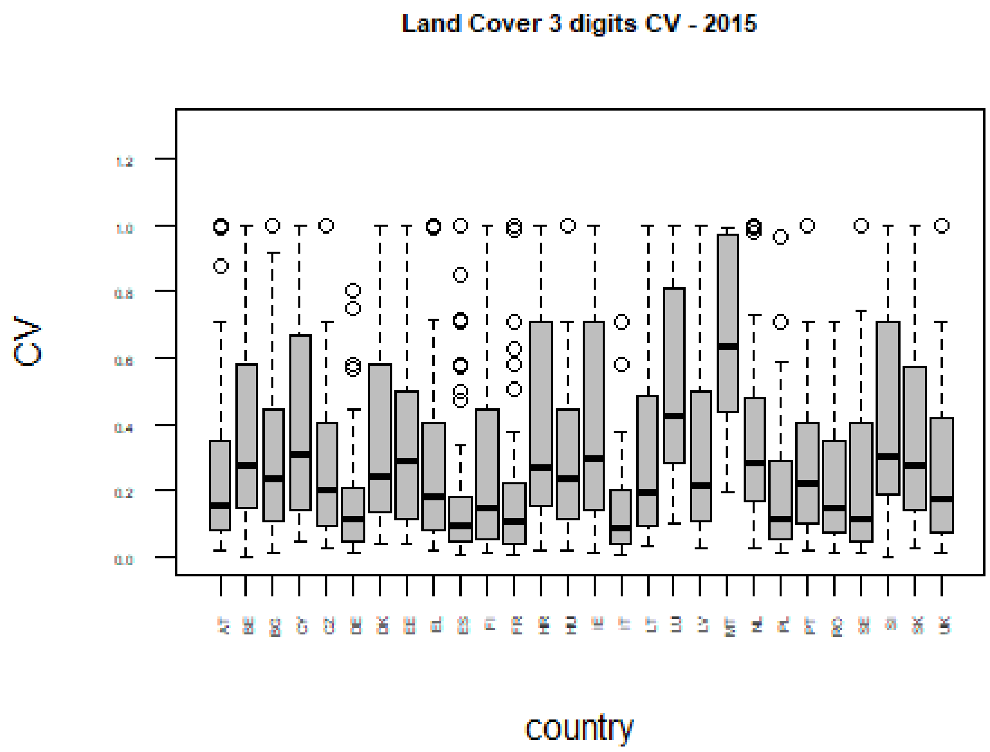

Actual CVs vs planned - Water areas
2018
2015
CV
0.0
0.2
0.4
0.6
0.8
0
5
10
15
20
25
Countries
HR
LU
SI
CY
CZ
DK
AT
BE
BG
DE
EE
EL
ES
FI
FR
HU
IE
IT
LT
LV
NL
PL
PT
RO
SE
SK



Note: Red and blue lines represent the minimum and maximum values of planned CVs for the 2018 LUCAS survey)

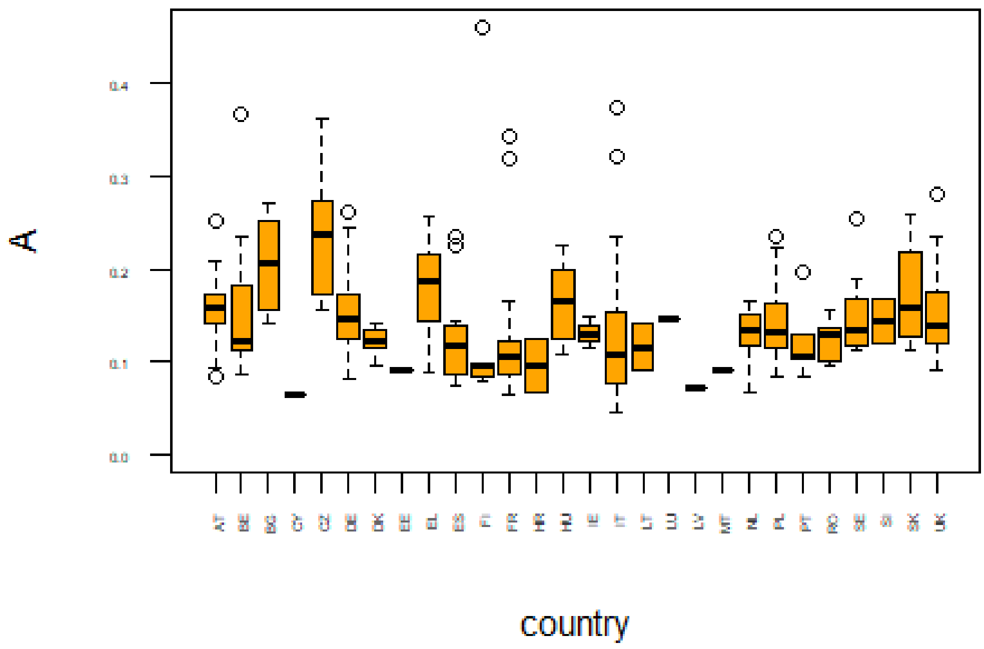
CVs calculated for the Land Cover at 3rd digit level



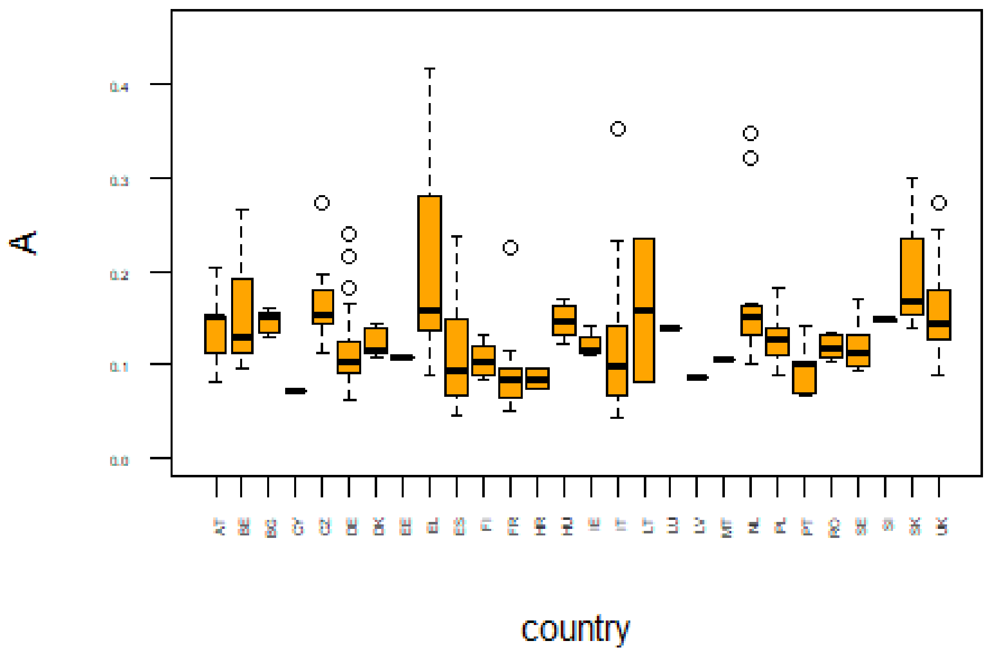


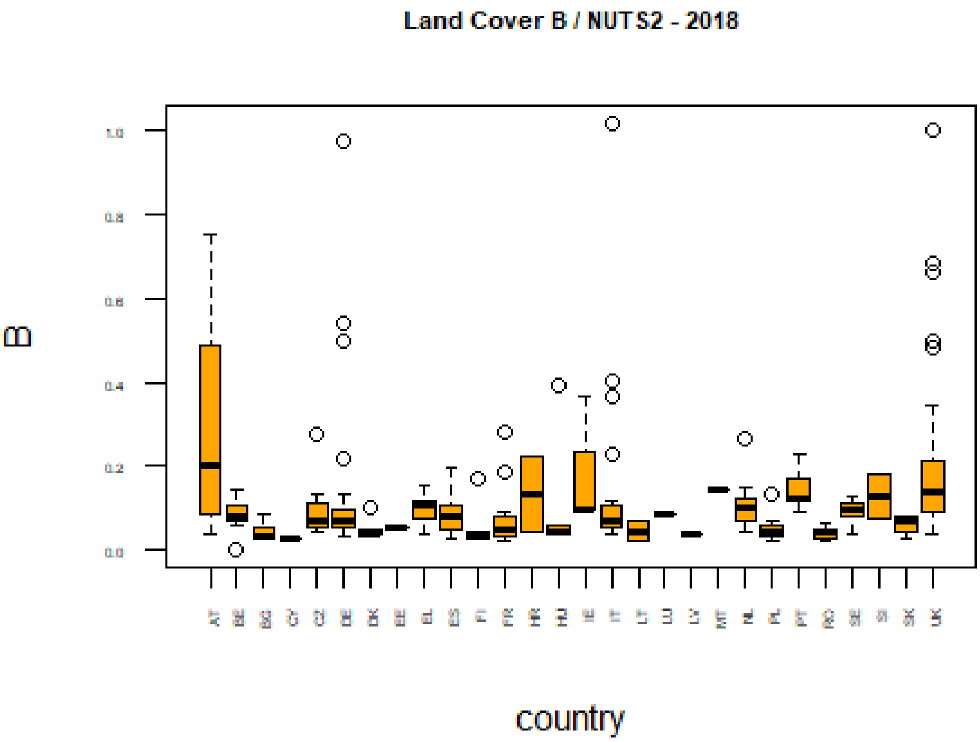
CVs calculated for the Land Cover at NUTS2 geographical level

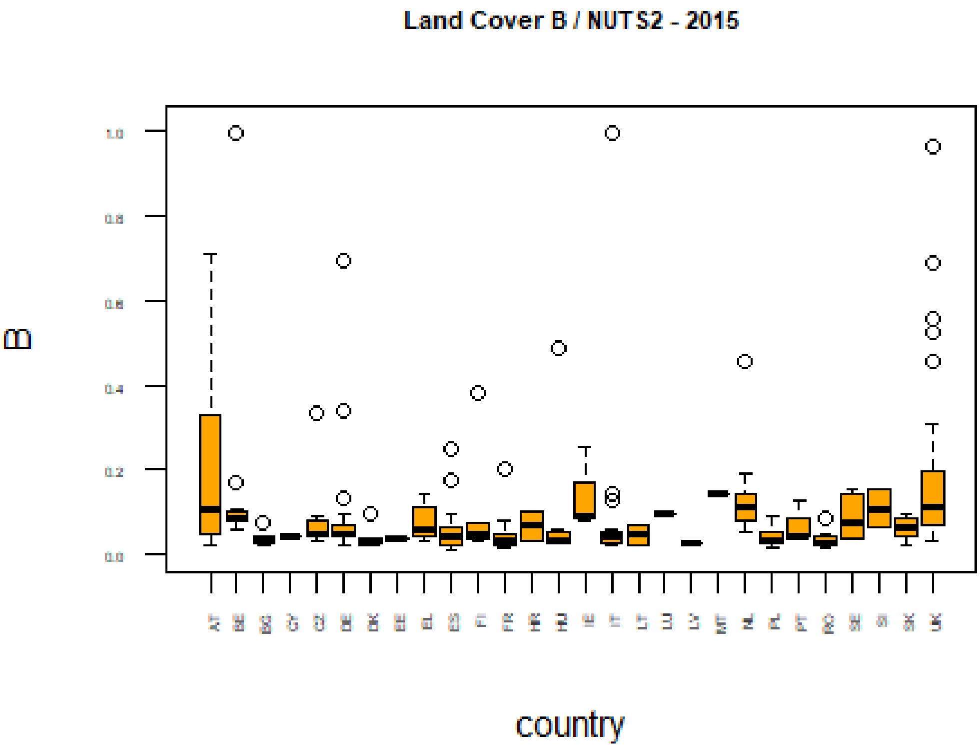
Land Cover A / NUTS2 - 2018



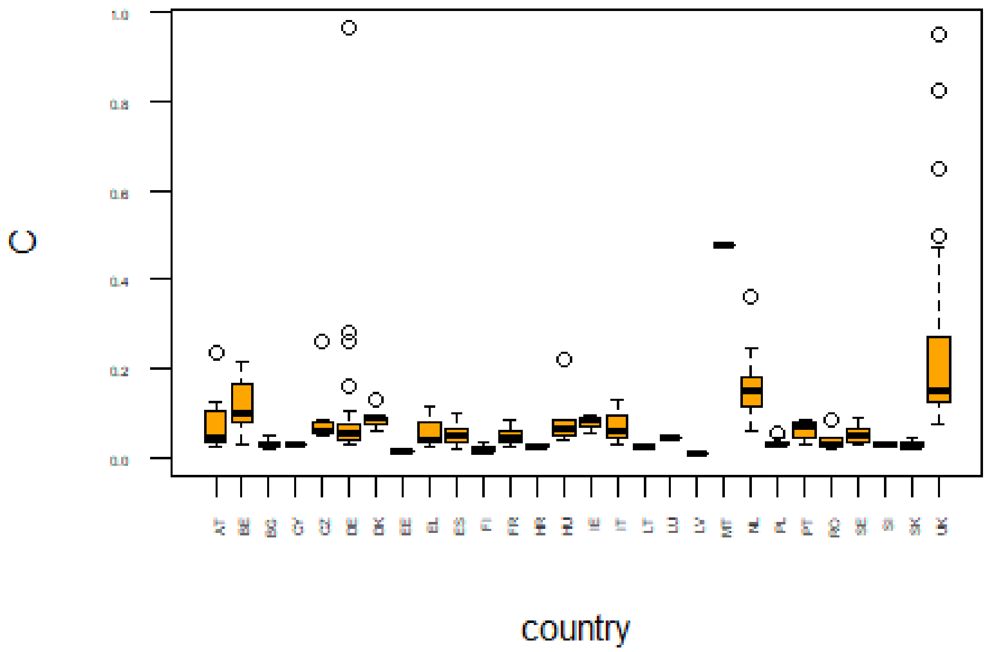
Land Cover A / NUTS2 - 2015



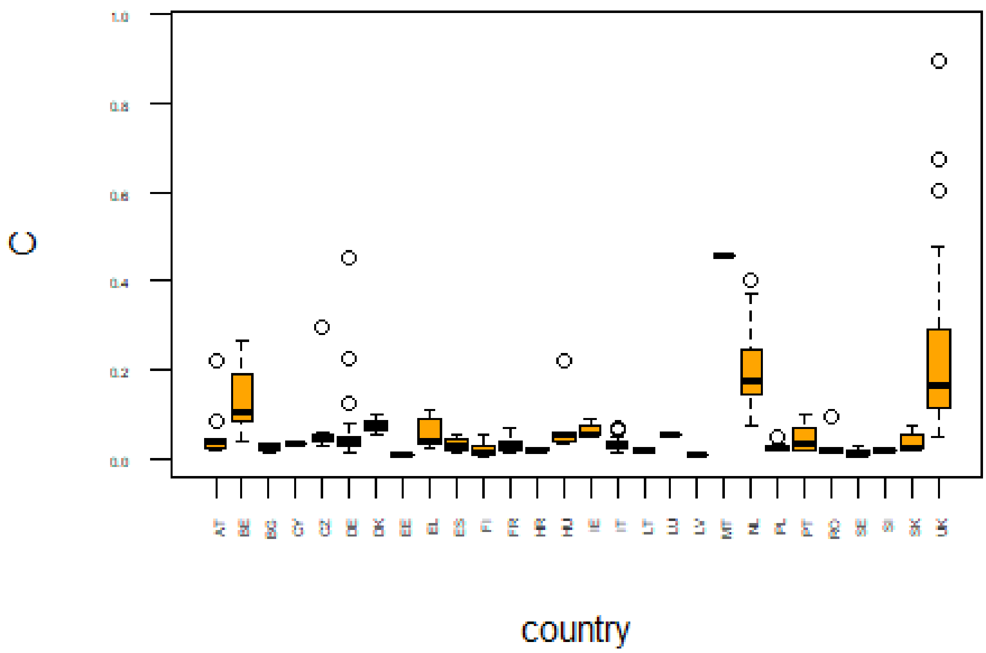

Land Cover B / NUTS2 - 2018
B
1.0
0.8
0.6
0.4
0.2
0.0
AT BE BG CY CZ DE DK EE EL ES FI FR HR HU IE IT LT LU LV MT NL PL PT RO SE SI SK UK
country

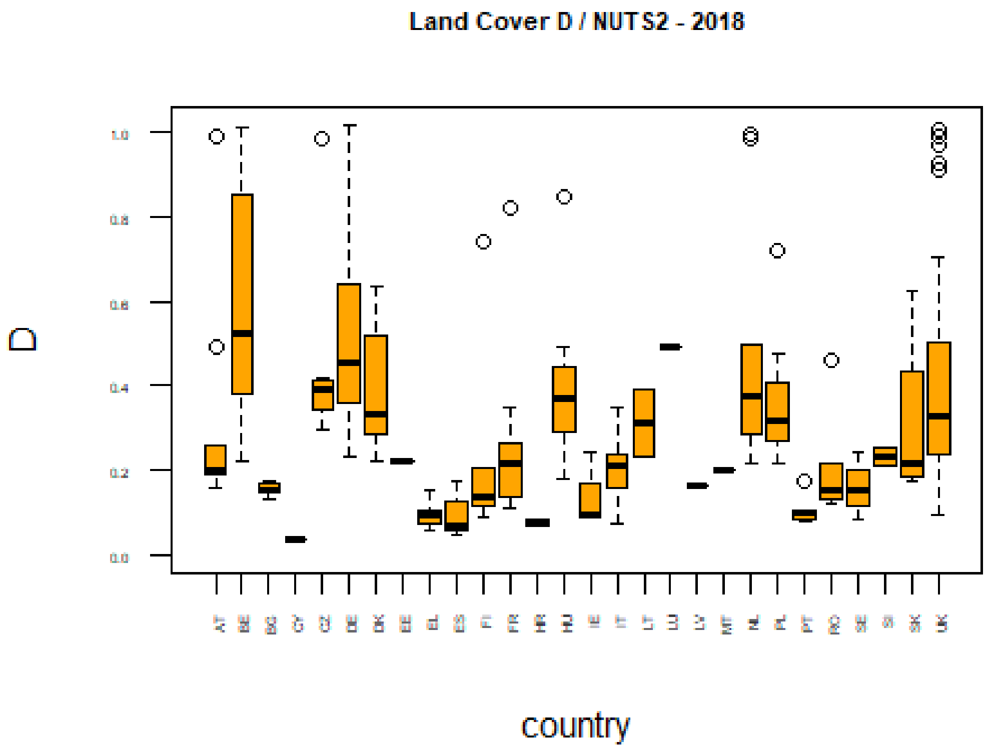

Land Cover B / NUTS2 - 2015
B
1.0
0.8
0.6
0.4
0.2
0.0
AT BE BG CY CZ DE DK EE EL ES FI FR HR HU IE IT LT LU LV MT NL PL PT RO SE SI SK UK
country

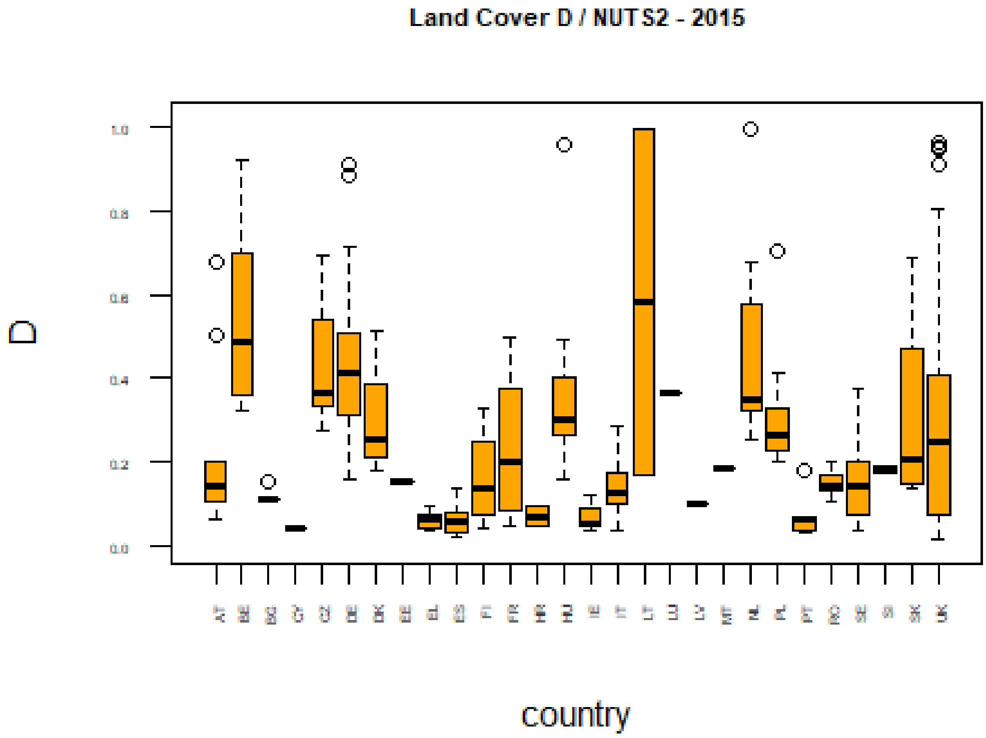
Land Cover C / NUTS2 - 2018



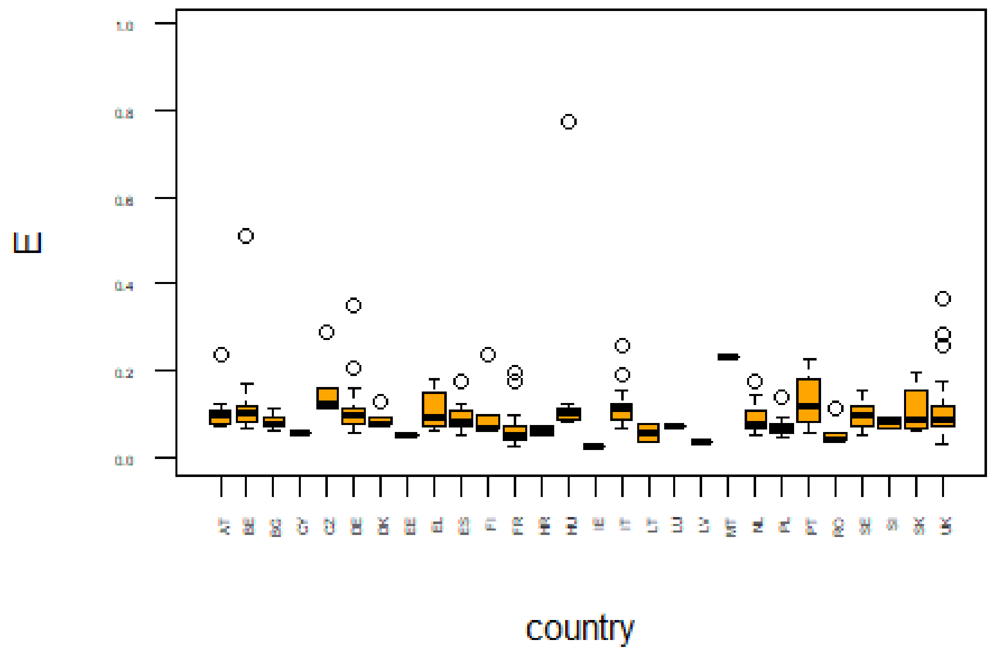
Land Cover C / NUTS2 - 2015



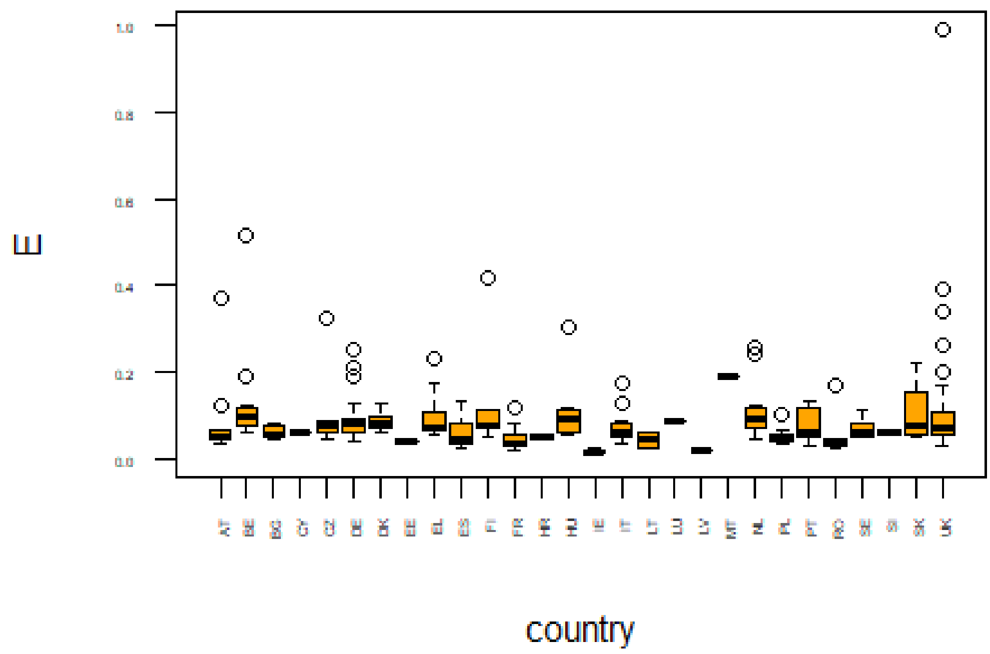

Land Cover D / NUTS2 - 2018
1.0
0.8
0.6
0.4
0.2
0.0
D
country

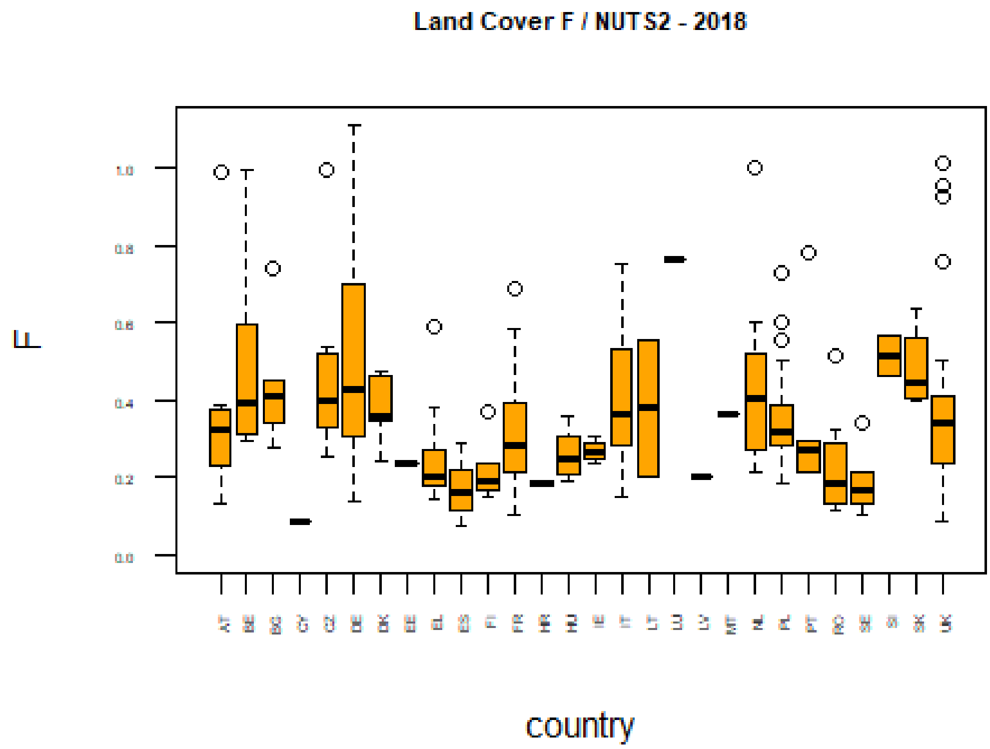

Land Cover D / NUTS2 - 2015
1.0
0.8
0.6
0.4
0.2
0.0
D
country

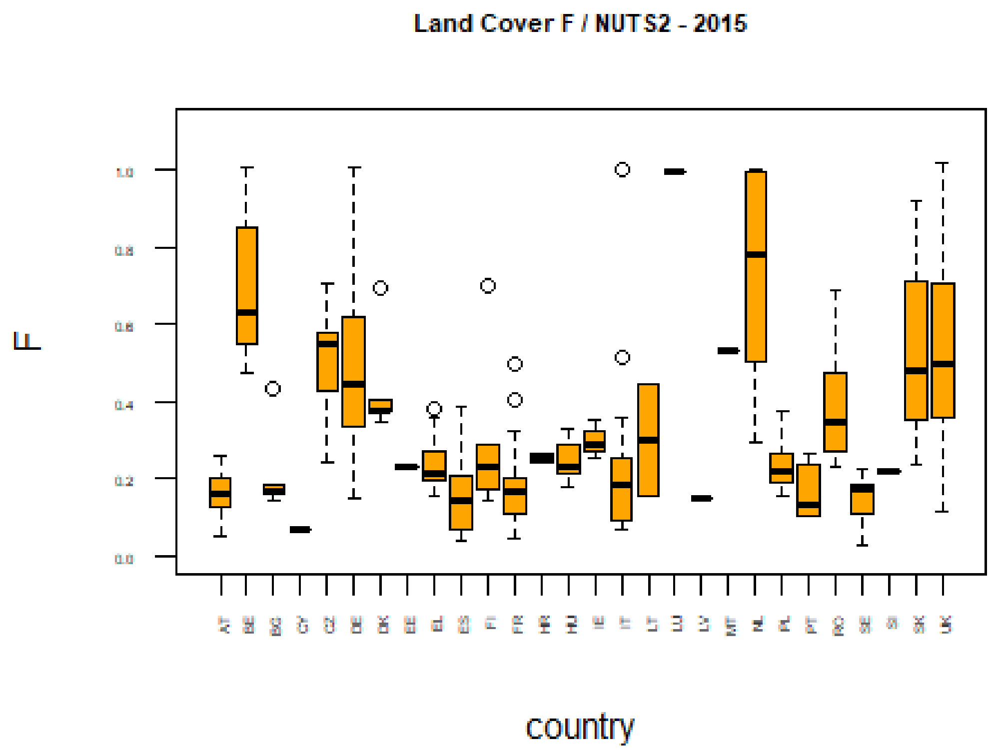
Land Cover E / NUTS2 - 2018

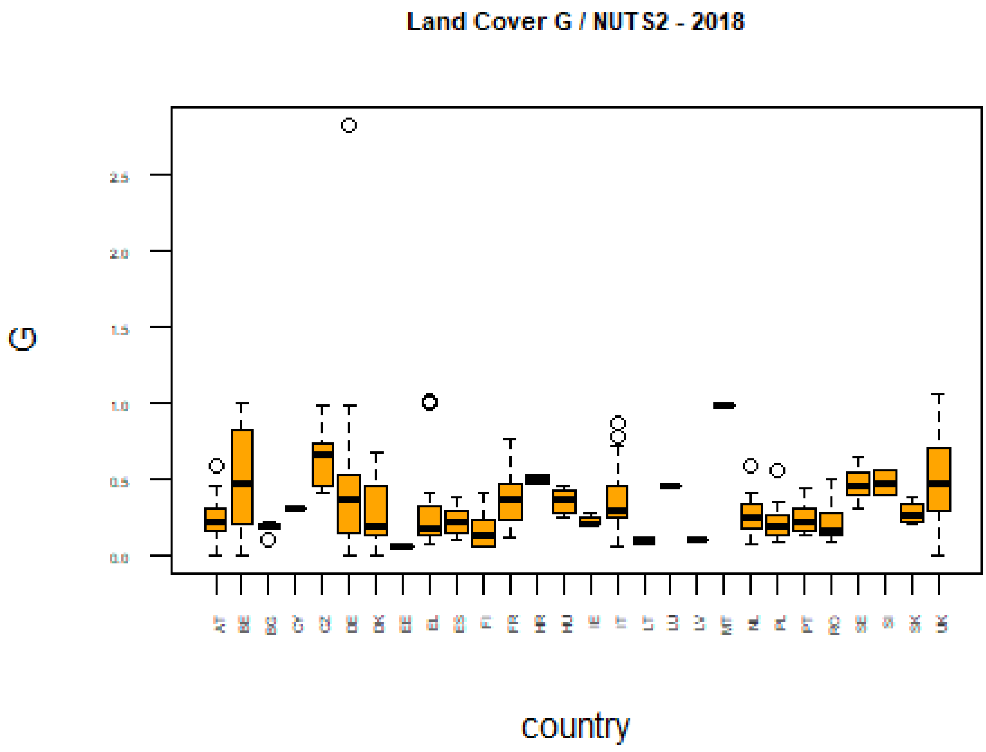


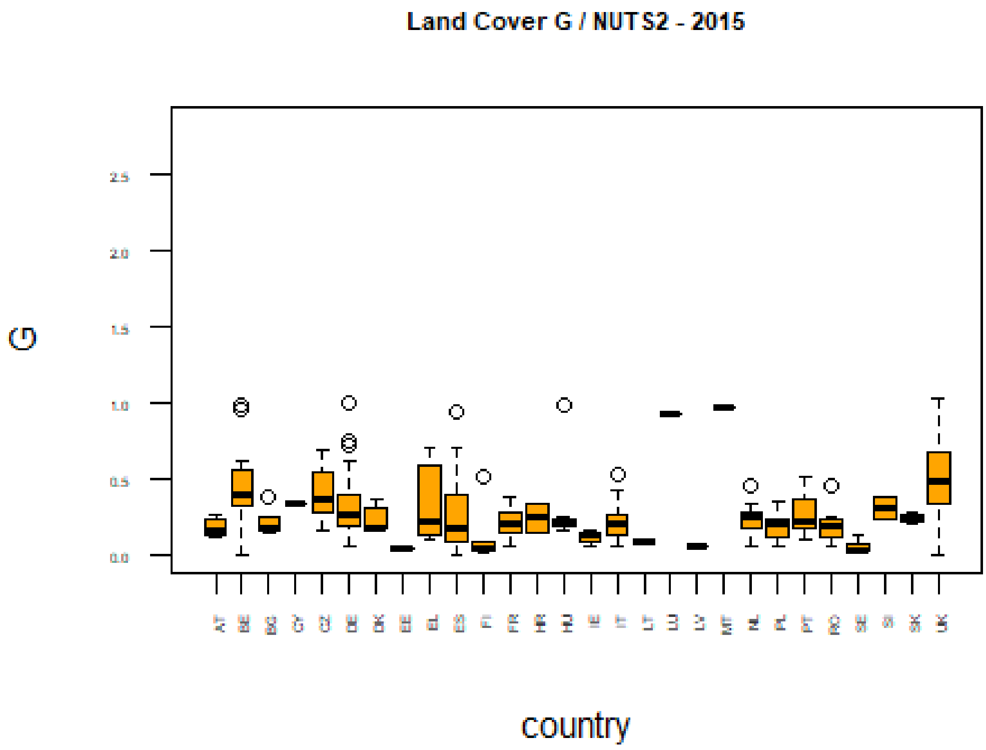
Land Cover E / NUTS2 - 2015

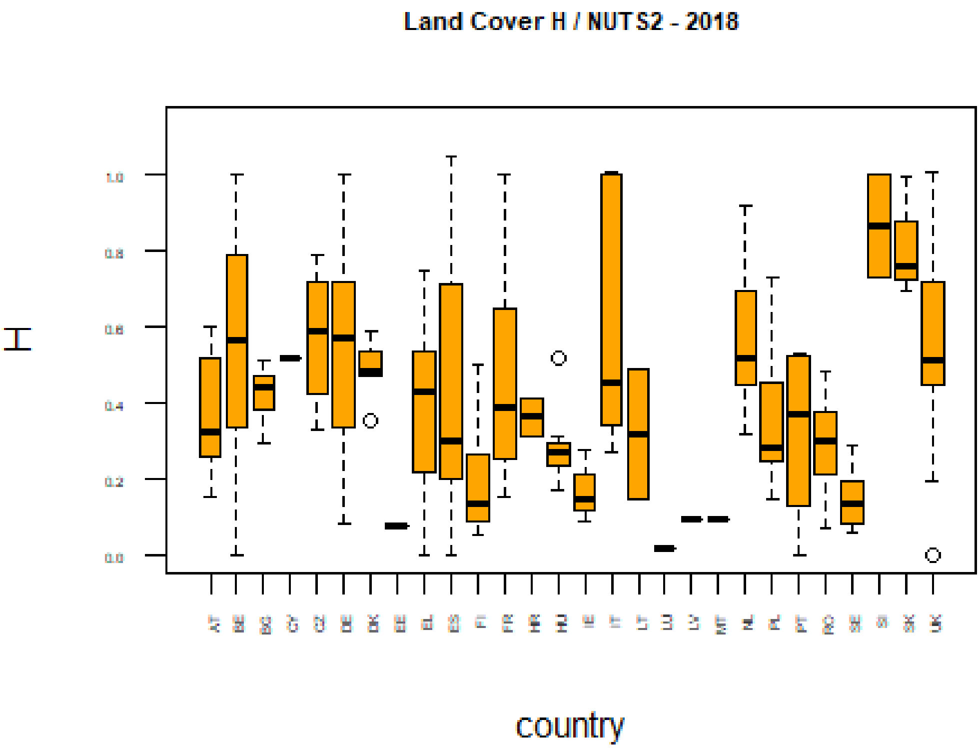



Land Cover F / NUTS2 - 2018
F
0.0
0.2
0.4
0.6
0.8
1.0
AT BE BG CY CZ DE DK EE EL ES FI FR HR HU IE IT LT LU LV MT NL PL PT RO SE SI SK UK
country


Land Cover F / NUTS2 - 2015
F
0.0
0.2
0.4
0.6
0.8
1.0
AT BE BG CY CZ DE DK EE EL ES FI FR HR HU IE IT LT LU LV MT NL PL PT RO SE SI SK UK
country


Land Cover G / NUTS2 - 2018
G
2.5
2.0
1.5
1.0
0.5
0.0
country


Land Cover G / NUTS2 - 2015
G
2.5
2.0
1.5
1.0
0.5
0.0
country


Land Cover H / NUTS2 - 2018
H
country

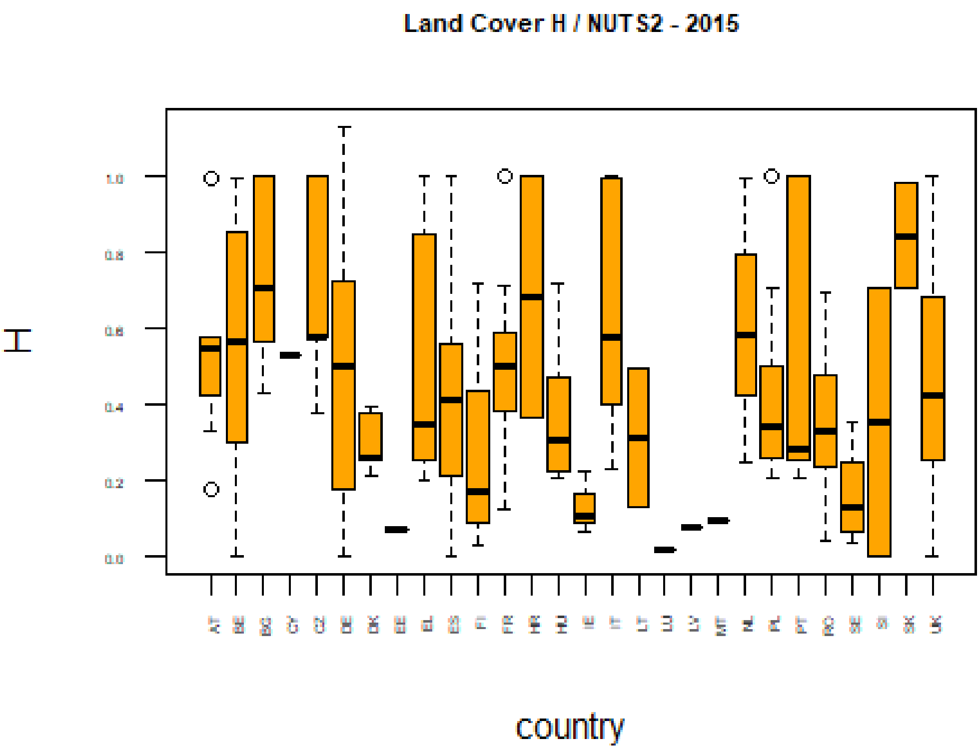
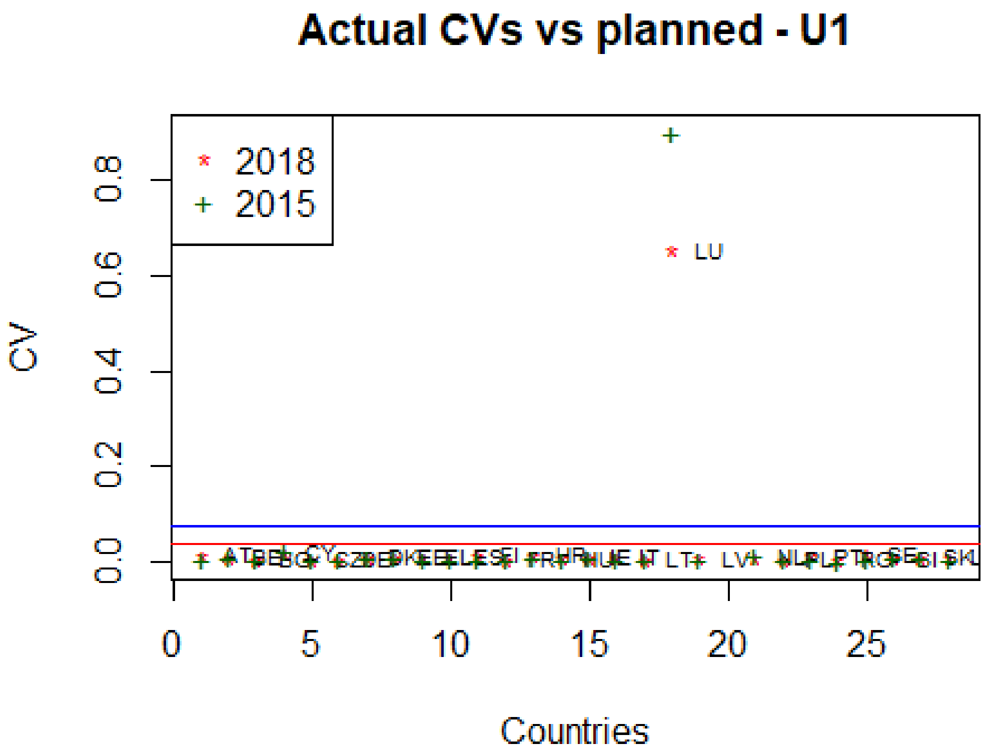

Land Cover H / NUTS2 - 2015
H
country

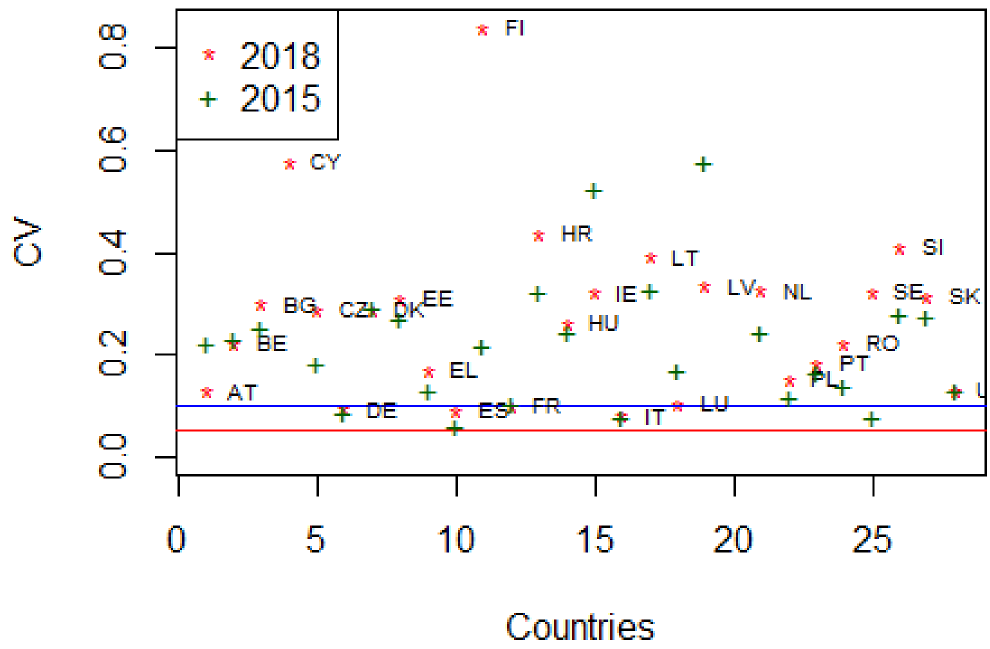
Table 38: Coefficients of Variations of LUCAS Land Cover 2018 (1st digit)

MS	A=Artificial	B=Cropland	C=Woodland	D=Shrubland	E = Grassland	F = Bareland	G = Water	H =Wetland
AT	0.049	0.030	0.028	0.088	0.037	0.192	0.114	0.107
BE	0.042	0.025	0.023	0.131	0.028	0.129	0.135	0.180
BG	0.084	0.013	0.011	0.065	0.032	0.166	0.071	0.170
CY	0.065	0.025	0.030	0.036	0.054	0.084	0.304	0.518
CZ	0.051	0.019	0.019	0.141	0.047	0.135	0.184	0.215
DE	0.025	0.010	0.009	0.093	0.015	0.061	0.071	0.090
DK	0.053	0.018	0.036	0.148	0.037	0.174	0.159	0.229
EE	0.092	0.054	0.015	0.220	0.053	0.237	0.065	0.073
EL	0.052	0.024	0.012	0.024	0.025	0.061	0.048	0.090
ES	0.031	0.023	0.018	0.020	0.023	0.039	0.086	0.102
FI	0.053	0.016	0.016	0.073	0.061	0.112	0.100	0.046
FR	0.025	0.010	0.010	0.049	0.013	0.108	0.075	0.096
HR	0.013	0.049	0.018	0.056	0.041	0.132	0.411	0.247
HU	0.058	0.016	0.023	0.117	0.035	0.095	0.075	0.096
IE	0.085	0.064	0.043	0.063	0.014	0.156	0.137	0.075
IT	0.021	0.017	0.014	0.040	0.024	0.118	0.107	0.135
LT	0.079	0.021	0.015	0.201	0.032	0.192	0.067	0.144
LU	0.146	0.084	0.045	0.494	0.070	0.762	0.452	NA
LV	0.073	0.039	0.012	0.165	0.034	0.203	0.096	0.092
MT	0.092	0.142	0.478	0.199	0.230	0.365	0.990	NA
NL	0.037	0.026	0.034	0.109	0.022	0.108	0.055	0.159
PL	0.035	0.009	0.008	0.078	0.015	0.081	0.059	0.072
PT	0.051	0.060	0.024	0.049	0.042	0.129	0.117	0.092
RO	0.047	0.012	0.011	0.060	0.016	0.068	0.053	0.062
SE	0.049	0.043	0.020	0.053	0.039	0.072	0.169	0.039
SI	0.068	0.069	0.020	0.161	0.056	0.394	0.353	0.589
SK	0.074	0.022	0.014	0.119	0.040	0.241	0.121	0.518
UK	0.025	0.019	0.025	0.051	0.014	0.043	0.052	0.157

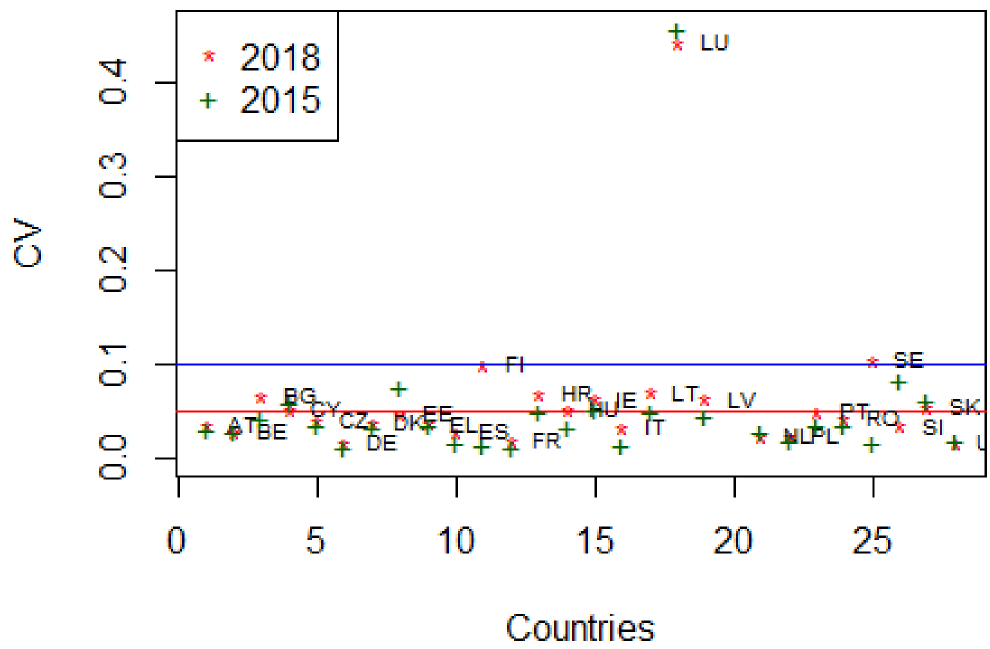
Annex 3: Coefficients of Variations for the Land Use

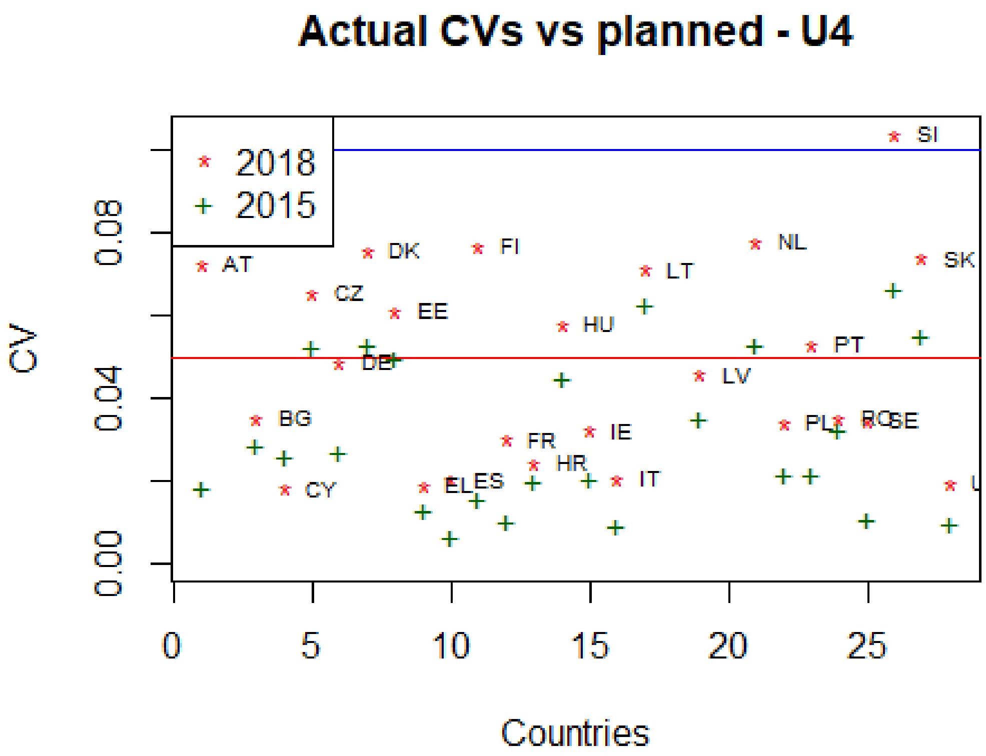


Actual CVs vs planned - U2



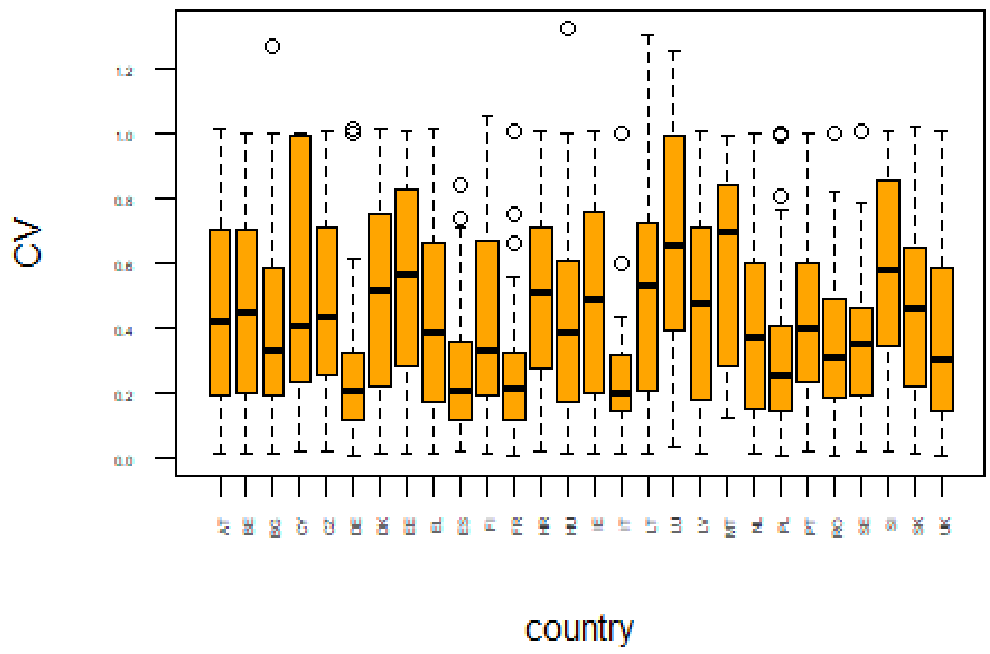
Actual CVs vs planned - U3



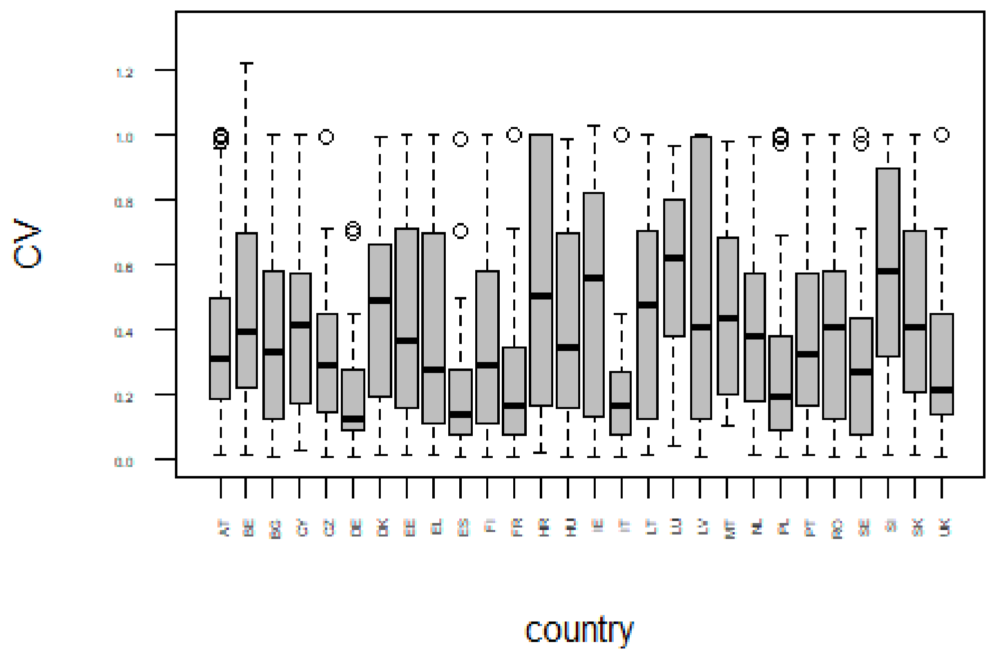

Actual CVs vs planned - U4
2018
2015
CV
0.00
0.04
0.08
Countries
0
5
10
15
20
25
AT
BG
CY
CZ
DE
DK
EE
EL
ES
FI
FR
HR
HU
IE
IT
LT
LV
NL
PL
PT
RO
SE
SI
SK

CVs calculated for the Land Use at 3rd digit level

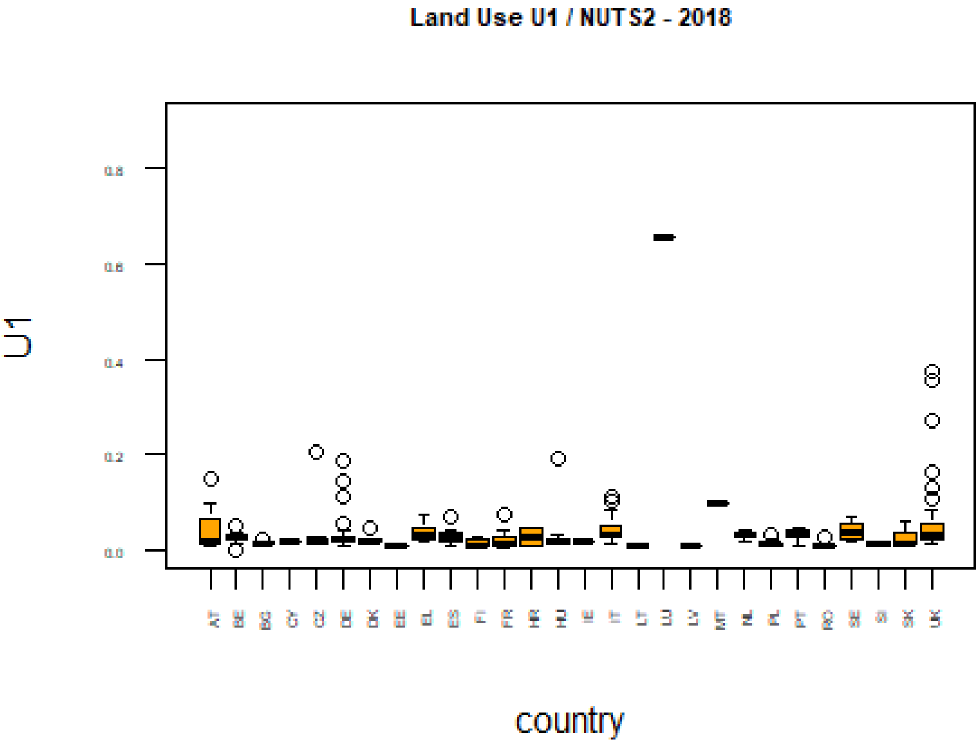
Land Use 3 digits CVs - 2018

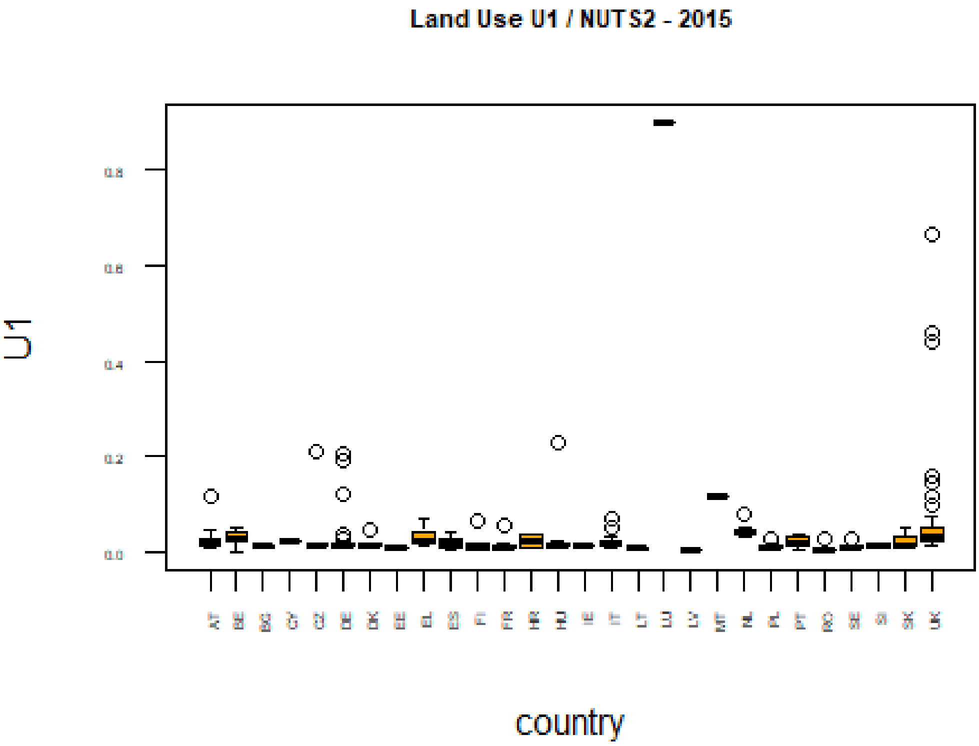


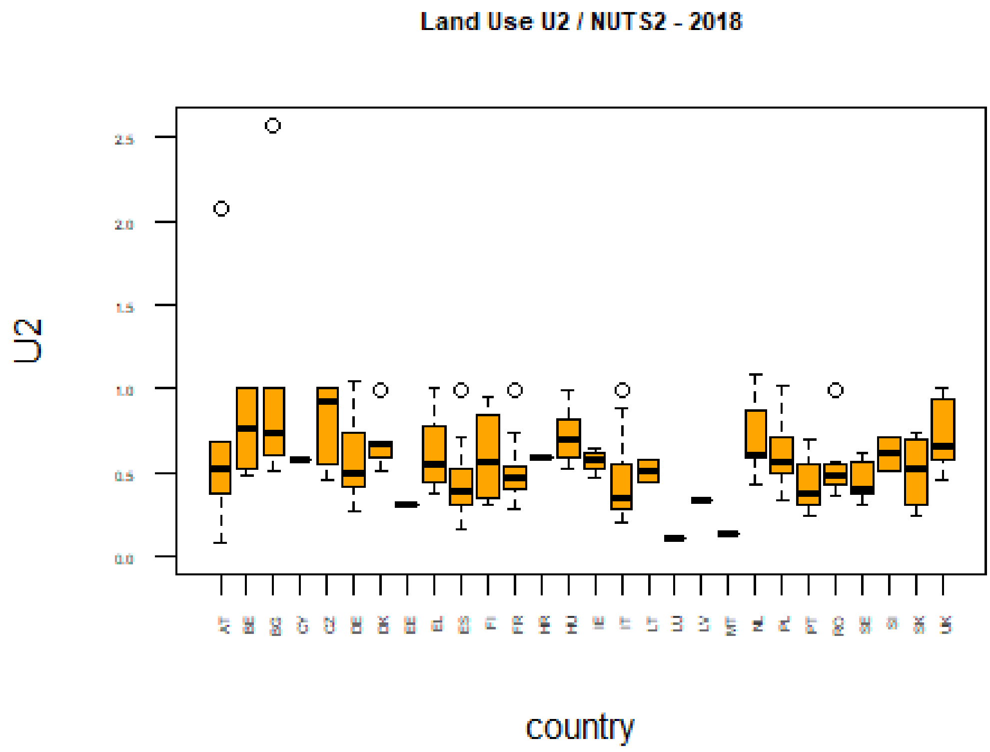
Land Use 3 digits CV - 2015

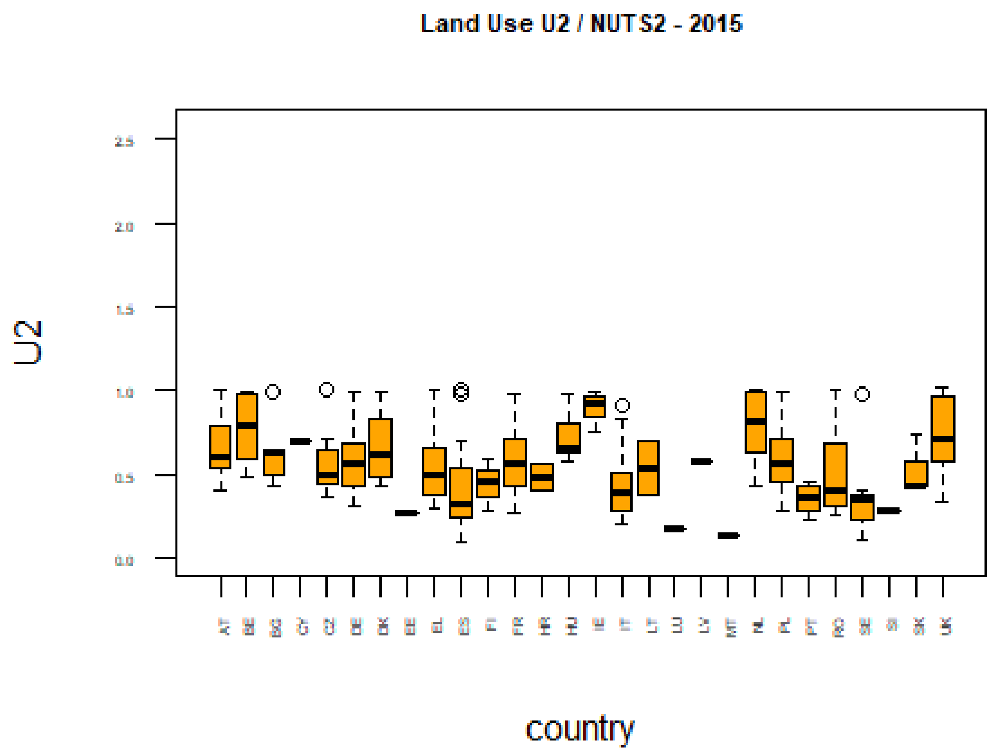


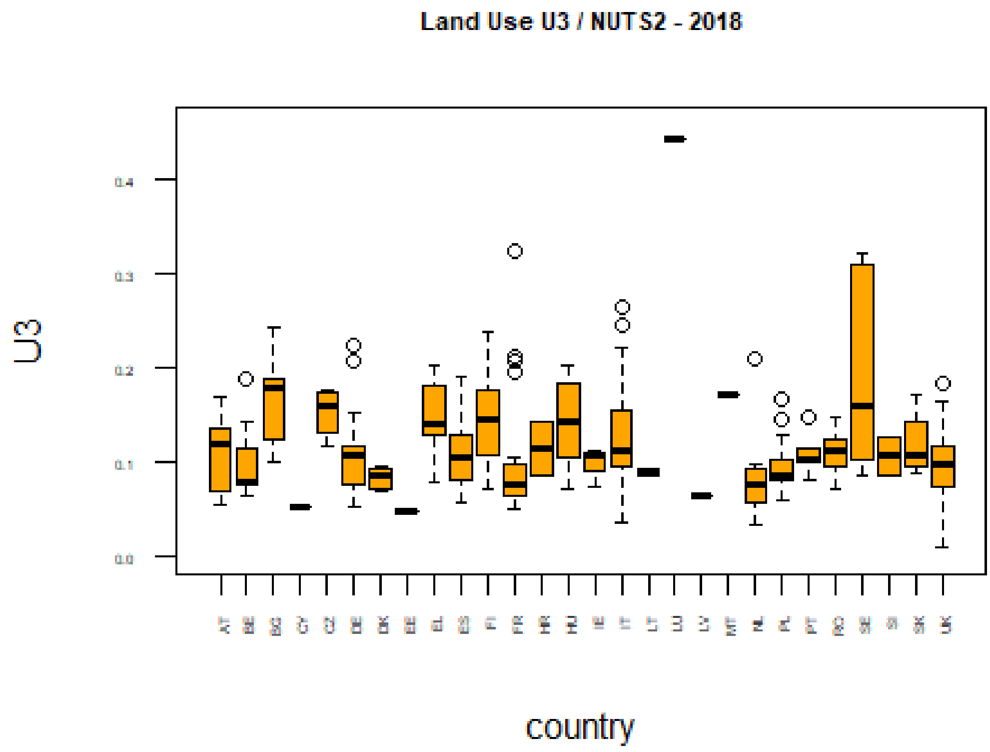
CVs calculated for the Land Use at NUTS2 geographical level

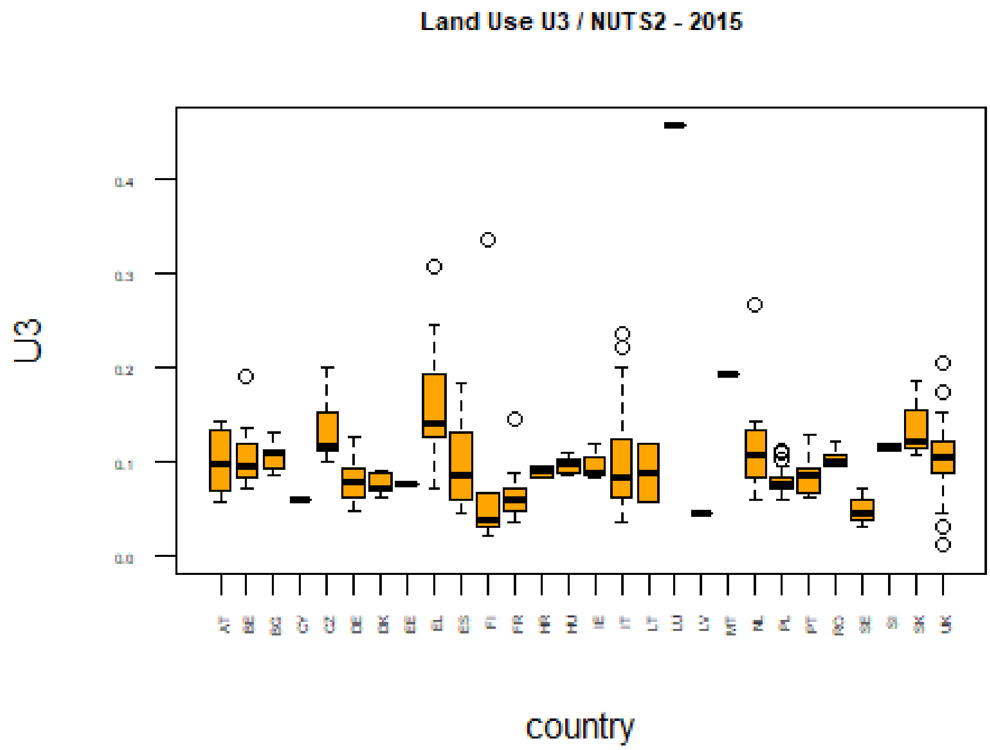


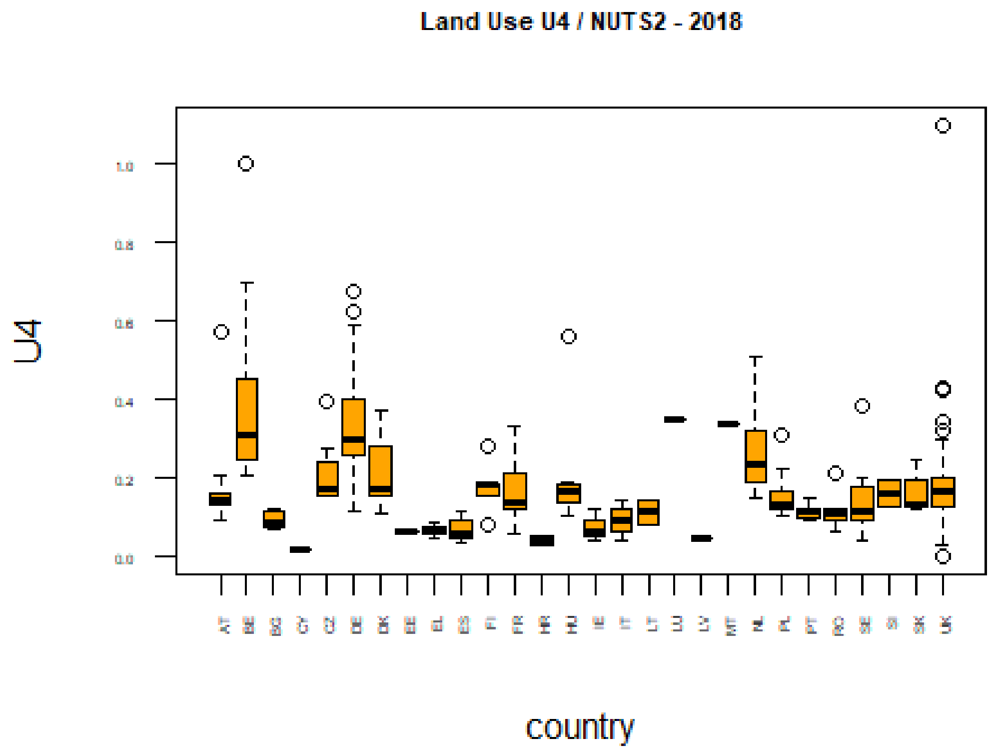



Land Use U2 / NUTS2 - 2018
U2
0.0
0.5
1.0
1.5
2.0
2.5
AT BE BG CY CZ DE DK EE EL ES FI FR HR HU IE IT LT LU LV MT NL PL PT RO SE SI SK UK
country


Land Use U2 / NUTS2 - 2015
U2
0.0
0.5
1.0
1.5
2.0
2.5
AT BE BG CY CZ DE DK EE EL ES FI FR HR HU IE IT LT LU LV MT NL PL PT RO SE SI SK UK
country


Land Use U3 / NUTS2 - 2018
U3
0.0
0.1
0.2
0.3
0.4
country


Land Use U3 / NUTS2 - 2015
U3
0.0
0.1
0.2
0.3
0.4
country


Land Use U4 / NUTS2 - 2018
U4
country

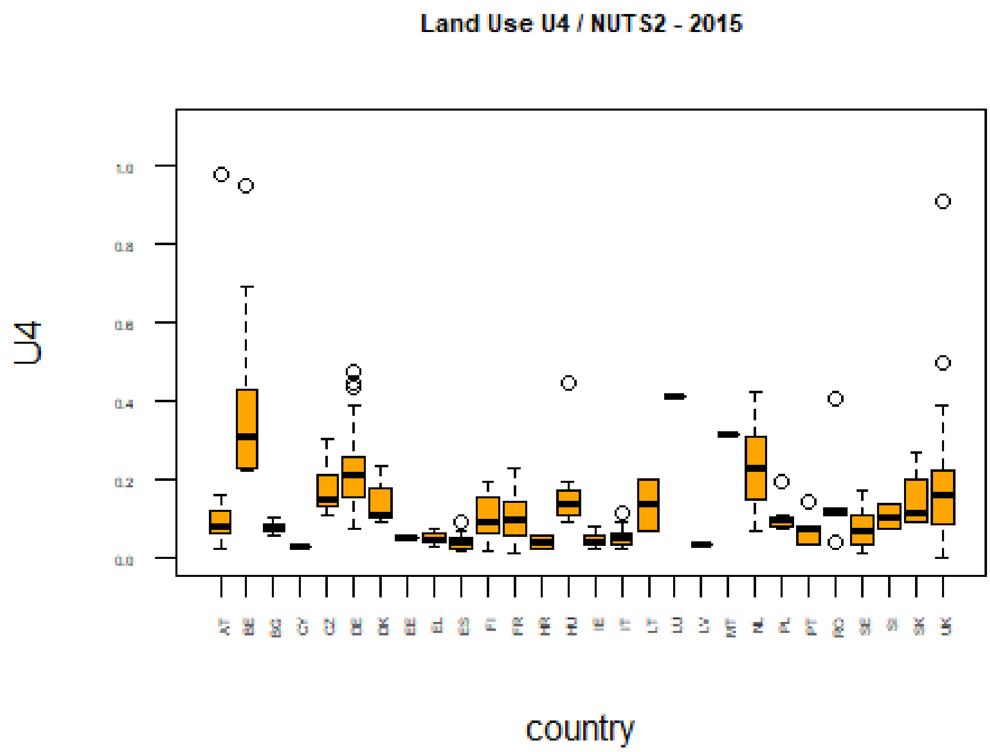

Land Use U4 / NUTS2 - 2015
U4
country

Table 39: Coefficients of Variations of LUCAS Land Use 2018 (1st digit)

	U1 = Primary Sector	U2= Secondary Sector	U3=Tertiary Sector	U4=Unused and abandoned
AT	0.015	0.129	0.037	0.073
BE	0.008	0.222	0.028	0.098
BG	0.006	0.300	0.067	0.035
CY	0.020	0.577	0.053	0.018
CZ	0.005	0.290	0.041	0.065
DE	0.003	0.095	0.017	0.049
DK	0.009	0.289	0.038	0.076
EE	0.008	0.312	0.048	0.061
EL	0.008	0.172	0.037	0.019
ES	0.008	0.094	0.028	0.020
FI	0.015	0.860	0.095	0.072
FR	0.004	0.102	0.019	0.030
HR	0.014	0.438	0.069	0.024
HU	0.007	0.262	0.053	0.058
IE	0.010	0.322	0.063	0.032
IT	0.009	0.081	0.033	0.021
LT	0.006	0.392	0.070	0.071
LU	0.016	0.655	0.107	0.443
LV	0.007	0.336	0.063	0.046
MT	0.093	NA	0.096	0.127
NL	0.010	0.322	0.020	0.073
PL	0.003	0.153	0.024	0.034
PT	0.012	0.183	0.050	0.053
RO	0.003	0.222	0.043	0.035
SE	0.017	0.326	0.104	0.035
SI	0.007	0.411	0.035	0.104
SK	0.008	0.315	0.055	0.074
UK	0.008	0.133	0.017	0.019