Report

for

EUROSTAT
European Commission

Sub-Task 2.1

Overview of existing information on grassland
&
Area and trend estimation for grassland classes: methods, validation and results

Methodological support for the LUCAS project

Supply of statistical services in statistical methodology,
Lot 1: Methodological support

Specific Contract N° 000069

under the Framework Contract 2018.0086

July 2021

in joint-venture with

GOPA Consultants
Hindenburgring 18, 61348 Bad Homburg, Germany
Phone +49 6172 930-521 Fax: +49 6172 930-130 Email: statistics@gopa.de

Table of content

Report

Tables

Figures

1 Overview of existing information on grassland

1.1 Introduction

The aim of the subtask 2.1 is to provide grassland area estimations and grassland area trends across time following the definition and scope of the LUCAS Survey to support the information needs of DG AGRI, in primis. Grassland estimates can increase the already available information on the implementation of the policies (CAP), fill some gaps and enable the monitoring of the interactions between agriculture and environment. Currently, data are available from MSs (e.g., temporary grassland and permanent pastures) are collected in the Eurostat statistics. However, the information retrievable from the LUCAS surveys can depict the grassland areas according to the specific nomenclature system adopted that can be clearly different for some aspect from the CAP grassland definitions implemented specifically by MSs according to their peculiarities (e.g., the presence of ineligible elements such as trees and shrubs).

The activity was piloted on the last LUCAS survey (2018) that benefits from the availability of additional variable missing in the previous surveys. After the consolidation of the procedures, the work was extended to the previous LUCAS campaigns (2009, 2012 and 2015). Estimations concerns status, namely grassland area estimates, and trends (2009-2018) at different NUTS levels (0, 1, 2 and 3).

The results were computed for the version 2016 of the Nomenclature of territorial units for statistics (NUTS). As of 1 January 2021, the new NUTS 2021 classification will come into force. The task results produced by the modelling exercises cannot be extrapolated for the statistical regions changed between 2016 and 2021 while the rules on the LUCAS survey defined for the grassland classes' results can be easily adapted for the new statistical regions with aggregation operations.

Objectives from the Task description

- To define specific rules for extracting all the relevant LUCAS survey points associated with grassland areas.
- To define a procedure for producing area estimates from the LUCAS grassland points available for the years 2009-2018.
- To assess the temporal trend and spatial distribution of grassland through the multi-annual LUCAS datasets.
- To analyse the quality of the grassland area estimates.

Report

Further specifications on the expected outcomes

- To provide a wider set of information than the classical base for grassland reported in EUROSTAT´s traditional statistics for crops (Farm Structure Survey – Integrated Farm Statistics and annual crop statistics) where enough data are available for temporary grassland while permanent grassland is much less detailed and reliable as includes many different kinds that are not differentiated.
- To provide estimates for permanent grassland as aggregated class and subclasses according to the LUCAS classification: E00 (Grassland), E10 (Grassland with sparse tree/shrub cover), E20 (Grassland without tree/shrub cover) and E30 (Spontaneously re-vegetated surfaces), by using LC1/LC2, LU1/LU2 and additional LUCAS parameters. Subclasses are of interest for specific analysis at EU and NUTS0 level (i.e. the EU protein balance could be possibly used to better estimate the quantity of proteins coming from permanent grassland used for feeding animals).

The whole Task 2 will be implemented in the timeframe October 1, 2020 - August 31, 2021, while subtask 2.1 was carried out in the period October 1, 2020 – April 30, 2020.

2 Datasets

The definition and application of rules for grassland classes and the modelling phase (model-based area estimation and grassland trend analysis) were carried out by exploiting multiple datasets used alone or in an integrated way. The datasets derived both from the LUCAS projects and from external datasets with European coverage:

1. Harmonized multi-temporal LUCAS Surveys microdata: field points (2009, 2012, 2015 and 2018) and photo-interpreted points (2015 and 2018).
2. Datasets of photo-interpreted LUCAS points for the 2009 and 2012 LUCAS Surveys.
3. LUCAS Master Grid.
4. CORINE Land Cover 2018.
5. Copernicus High Resolution Layer Grassland (2015 and 2018).

Specific variables form the datasets 3, 4 and 5 were used as covariates during the modelling exercises.

2.1 The LUCAS Surveys

The Land Use/Cover Area frame Survey (LUCAS) is the project aimed at collecting in situ land cover (LC) and land use (LU) data collection over the whole of the European Union (EU). LUCAS carried out five surveys in 2006, 2009, 2012, 2015, and 2018. These campaigns involved a total of 1 351 293 points at 651 780 unique locations by collecting 5.4 million photos and up to 109 variables. Such great deal of data can be considered the most comprehensive in situ database on land cover and land use in the EU (d'Andrimont et al., 2020).

LUCAS surveys are made up of different components: Core and Modules. In 2018, LUCAS comprises the Core, the Grassland module and the Soil module. LUCAS Core includes the identification of the point, different aspects of land cover and land use information and land and water management and a new part on EUNIS and a test for Copernicus. The Grassland module is a test module on 3734 points, to assess the practical and scientific feasibility to collect the relevant information. The Soil module collects soils sample in 26,014 LUCAS points to measure different soil parameters (e.g., bulk density, organic matter, biodiversity).

The typology of information collected, and the specific modules can change across campaigns. In general, the new campaigns collect additional information with new modules that cannot traced back in time while the Core data ensure enough temporal consistency across the surveys (2009-2018).

The LUCAS Survey classification adopts a separate classification system for LC and LU. LC is the physical cover of the Earth’s surface and LU is the socioeconomic function of the land. The same classification is applied in all EU countries in the LUCAS survey. It also allows comparisons in time. It is as much as possible compatible with the existing LC/LU systems (e.g., FAO, NACE and Farm Structure Survey). Slights variation on the LUCAS LC/LU classes occurred during the surveys 2009-2018.

LUCAS survey registers LC, LU and other variables for each point that corresponds to a circle of 3 m diameter (7m2). The observation window must be enlarged by surveyors when the point falls in an area with non-homogeneous LC (e.g. Trees or shrubs interspersed with grass and bare land). In these cases, an extended window of observation with a circle of 40 m diameter (0.13 ha) must be used.

The harmonized multi-temporal LUCAS Surveys dataset (years 2009, 2012, 2015 and 2018) were produced by Joint Research Centre of the European Commission (JRC) by processing the LUCAS Survey raw data with a harmonisation procedure. The harmonised datasets allow a seamless application of the rules across the LUCAS campaigns, facilitate the final users of the results as well as the replication of the rules, and query on the datasets. In general, the LUCAS harmonised datasets should ensure the usability of the LUCAS microdata, especially for the scientific community, thanks to the extensive process of cleaning, semantic and topological harmonisation, creation of one consolidated database with hard-coded links to the full-resolution photos, openly accessible (d'Andrimont et al., 2020).

The harmonized datasets were integrated with a dataset of photo-interpreted (PI) LUCAS points for the year 2009 and 2012 resulting in a much higher number of survey points for the mentioned years. The additional PI points are also panel points in the LUCAS Survey 2015. The resulting dataset was used for the grassland trend analysis ensuring a good number of points across the time range 2009-2012 and allowed at the same time the full coherence with the calibration procedure carried out by Eurostat for each survey year for the LC/LU area estimation carried out with the design-based approach.

Table 1 - Number of LUCAS Survey points by year (2009-2018) at EU-level.

NUTS	Year	# LUCAS points
EU-28	2018	337,854
EU-28	2015	338,725
EU-27	2012	333,916
EU-23	2009	261,610

Table 2 - Number of LUCAS Survey points by country and year (2009-2018) at EU-level (2018: EU28, 2015: EU28, 2012: EU27, 2009: EU23).

NUTS0	# LUCAS points			
	2018	2015	2012	2009
AT	8840	8839	8519	7057
BE	3659	2899	2596	1804
BG	7678	7677	7692	
CY	2313	1726	1442	
CZ	5713	5712	5730	4706
DE	26777	26598	26634	21529
DK	3703	3665	3646	2567
EE	2665	2637	2585	2670
EL	12622	12521	12435	10389
ES	45314	50281	50267	38467
FI	16182	16116	16168	19955
FR	48215	48188	48105	37602
HR	4239	3532		

NUTS0	# LUCAS points			
	2018	2015	2012	2009
HU	5514	5169	5133	5512
IE	4975	4907	4924	4183
IT	28294	28693	28354	23601
LT	4584	4505	4493	3860
LU	340	251	259	152
LV	5376	5374	5186	3827
MT	79	79	79	
NL	5011	2521	2519	2460
PL	23086	22980	23064	18551
PT	7168	9006	9025	5547
RO	16725	16720	16731	
SE	26709	26648	26721	27511
SI	1922	1923	1884	1401
SK	2898	2755	2761	3052
UK	17253	16803	16964	15207

NUTS0	# LUCAS points			
	2018	2015	2012	2009
EU	337854	338725	333916	261610

2.1.1 List of LUCAS Survey variables

In the following table, the LUCAS Survey variables used for the grassland rules are listed and described.

Table 3 - LUCAS Survey variables (Source LUCAS Survey 2018). The variables PARCEL_AREA_HA and GRAZING were used only in the data exploration and rules testing and not in the final grassland rules.

Survey variable	Definition	Levels
LC1	Land cover is the biophysical coverage of land (e.g., crops, grass). LC1 refers to the primary LC to be registered for each LUCAS survey point according to the LUCAS nomenclature. In case of layered, competing (no clear dominance) land covers the classification starts from the top.	75 LC classes aggregated in the categories: A00, B00, C00, D00, E00, F00 and H00 Cf. 1.4
LC2	The secondary LC to be registered when more than one LC co-exists in the same area (e.g., agro-forestry areas, where wooded land is combined with crops; arable land, when different crops are intrinsically mixed in the same field).	
LU1	Land use indicates the socio-economic use of land (e.g., agriculture, forestry, recreation or residential use). LU1 refers to the primary LU to be register for each point according to the LUCAS nomenclature. It is possible to record 2 LCs and 1 LU or 1 LC and 2 LUs. LU codes are to be associated with the respective LC code, thus LC1 -> LU1 and LC2 -> LU2 if the LC and LU are related.	53 LU classes aggregated in the categories: U100, U200, U300 and U400. Cf. 1.4
LU2	The secondary LU to be registered in case of the occurrence of different LUs for the same point (e.g., a building with several floors holding residential and commercial use).	
INSPIRE_PLCC4	The percentage of herbaceous plants from the INSPIRE PLCC module to be collected only for the points where LC (LC1 or LC2) is	0-100

	either woodland (C00), shrub land (D00), grassland (E00) or bare land (F00) and is to be assessed within the homogeneous plot inside the extended window of observation (20m radius). Assessment of these percentages is made using the “birds-eye” view	
PARCEL_AREA_HA	Rough estimate of the area of the parcel (in ha) to which the point belongs. The area assessment takes into account the whole parcel having the same LC and LU as the observation point, except for wooded land.	<0.1; 0.1-0.5; 0.5-1; 1-10; >=10;
GRAZING	Information on grazing collected for points with the following land cover: B00, C00, D00, E00, F00 and H00. The reference area is the field, or if there are no field borders the area up to a distance of some 500 m, only the area visible from the point reached in the field is assessed.	Signs of grazing/ No signs of grazing/ Not relevant

2.2 The LUCAS Master Grid

The LUCAS Master Grid is the starting list for the LUCAS survey design, the so-called First Phase; it consists of around 1,100,000 geo-referenced points (corresponding to a 2*2 Km) grid covering the EU-28 territory, also systematically selected). Each of these points is classified into k LC categories (the strata) based on photointerpretation of aerial photos or satellite images. In 2005 these points were stratified into 7 aggregated strata and 10 in the 2016 version (1=Arable land, 2=Permanent crops, 3=Grass, 4=Wooded areas, 5=Shrubs, 6=Bare surface, low or rare vegetation, 7=Artificial constructions and sealed areas, 8=Inland water, 9=Transitional and coastal waters, 10=Impossible to PI). In the Second Phase, the final field sample is a selected from the master by strata and by NUTS2 and visited to register LC, LU and other variables in situ. In this way, it is possible to combine the information resulting from the photointerpretation with the information collected during the ground inspection of a portion of the N points selected in the first phase. The final statistical estimates are based on the weights derived from both the master and the field observations.

Table 4 - Number of Master points by country. Source: Master_190517.

NUTS 0	# Master points
AT	20982
BE	7680
BG	27748
CY	2317
CZ	19716
DE	89442
DK	10806
EE	11328
EL	32937
ES	126471
FI	84361
FR	137244
HR	14148
HU	23269
IE	17458

NUTS 0	# Master points
IT	75140
LT	16234
LU	644
LV	16139
MT	80
NL	9326
PL	77985
PT	22963
RO	59589
SE	112424
SI	5063
SK	12265
UK	61133
EU-28	**1094892**

Table 5 - Number of Master points by strata. Source: Master_190517

Stratum	# Master points
na	32
1 = Arable land	321035
2 = Permanent crops	31323
3 = Grass	130745
4 = Wooded areas	455754
5 = Shrubs	67705
6 = Bare surface, low or rare vegetation	12170
7 = Artificial constructions and sealed areas	41815
8 = Inland water	30695
9 = Transitional and coastal waters	1162
10 = Impossible to PI	2456
EU-28	**1094892**

2.3 Covariates

To improve the reliability and power of prediction of the models used for grassland area estimation and grassland trend analysis a set of covariates with relationships with the spatio-temporal characteristics of the grassland classes were selected and used. The following table lists the covariates, their sources and provides a brief description of the covariates.

Table 6 – List, source and description of the covariates used in the modelling phase

Covariate	Source	Definition	Levels
X_LAEA	LUCAS Master Grid (Master_190517)	Longitude in meters (Lambert Azimuthal Equal Area projection -LAEA[1])	-
Y_LAEA	LUCAS Master Grid (Master_190517)	Latitude in meters (Lambert Azimuthal Equal Area projection -LAEA)	-
ELEVATION	LUCAS Master Grid (Master_190517)	Elevation in meters of the point (source EUDEM[2])	-
NUTS1_16	LUCAS Master Grid (Master_190517)	NUTS 1 from GISCO DB 2016	
NUTS2_16	LUCAS Master Grid (Master_190517)	NUTS 2 from GISCO DB 2016	
STRATUM_LABEL (STR_18)	LUCAS Master Grid (Master_190517)	Strata variable for 2018	Eight strata: 1 = Arable land 2 = Permanent crops 3 = Grass

[1] The coordinate reference system used for pan-European statistical mapping at all scales or other purposes where true area representation is required.
[2] The Digital Elevation Model over Europe from the GMES RDA project (EU-DEM) is a Digital Surface Model (DSM) representing the first surface as illuminated by the sensors. The EU-DEM dataset is a realisation of the Copernicus programme, managed by the European Commission, DG Enterprise and Industry. Source: https://land.copernicus.eu/imagery-in-situ/eu-dem/eu-dem-v1.1

Covariate	Source	Definition	Levels
			4 = Wooded areas 5 = Shrubs 6 = Bare surface, low or rare vegetation 7 = Artificial constructions and sealed areas 8 = Inland water 9 = Transitional and coastal waters 10 = Impossible to PI
CLC18 classes 231 and 321	CORINE Land Cover 2018 (EEA)		
GRAVPI_2015 and GRAVPI_2018	Copernicus High Resolution Layer Grassland 2015 and 2018 (Copernicus Land Monitoring Service)		

2.3.1 CORINE Land Cover

Each point of the Master of the LUCAS survey, and therefore each point of the surveys, are associated to the corresponding CORINE Land Cover (CLC) classes of the land cover map. Such variable has been updated in the LUCAS Master Grid throughout the years by attaching the corresponding CLC class for the available CLC updates (2006, 2012 and 2018). The reference CLC products are available as raster maps with 100 m spatial resolution.

The CLC class for the year 2018 associated to each Master point was selected as one of the covariates to be used for the model-based area estimates of the grassland classes. Due to the wide interval of update (6 years) the variable CLC18 was used both for the modelling exercises performed with the LUCAS surveys 2015 and 2018.

Semantic analysis of the nomenclature has been carried out to identify the CLC classes having the strongest relationship with the grassland classes defined (TG, PGs, PGsn and OG). In general, the latter can be linked to one or more CLC classes due to the ample definitions adopted by the CLC nomenclature. To avoid the one-to-many relationships, the following strict associations have been considered. Following the semantic analysis, the grassland classes were associated to specific CLC classes as reported in the table below.

Table 7 – CLC classes selected as representative for the defined grassland classes.

Grassland class	CLC class code	CLC class label	Notes
Temporary grassland (TG)	none	none	Temporary grasslands are not included as separate CLC category. Fodder crops are included in class 211.
Permanent grasslanf (PG)	231	Pastures, meadows and other permanent grasslands under agricultural use	Pastures refer to permanent grassland (> 5 years not in rotation). The class includes temporary and artificial pastures not under a rotation system, which become permanent grasslands five years after ploughing.
Other grassland (OG)	321	Natural grassland	Natural grasslands are better distinguished compared to many agricultural related classifications.

2.3.2 Copernicus High Resolution Layers (HRLs) - Grassland

The HRL Grassland products produced in the framework of the Copernicus Land Monitoring Service were analysed to select suitable variables to be used as covariates for the whole Europe. These grassland maps were deemed relevant to the modelling exercise both for the thematic and spatial characteristics. In terms of definition, the grassland class is defined as *natural, semi-natural and managed grasslands (according to their origin and utilization) as well as all types of grassland (permanent or seasonal) under highly heterogeneous biogeographic conditions (wet or dry climate, fertile or poor soil). Additional non-woody plants such as lichens, mosses and ferns can be included*.

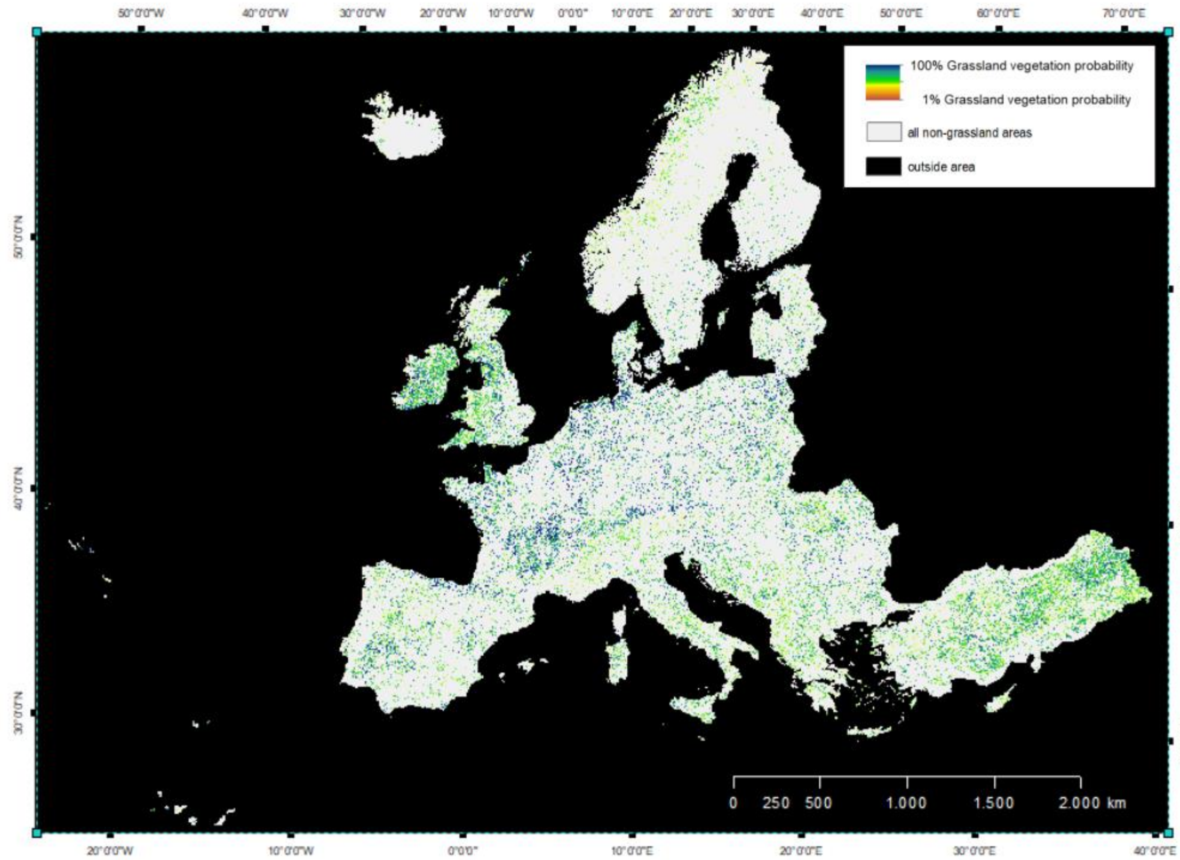
The HRL Grassland products were consolidated since 2015 with two series 2015 and 2018. The latter ensures a higher level of quality and both series were generated by long time series of imagery from several sensors allowing good performances in the grassland detection and separation from crops. The available products span from status maps to change maps for the reference years. Status maps are provided as binary products grassland/non-grassland in raster format (10 m for 2018 and 20 m for 2015) that includes the full spectrum of grassland use intensity (from natural to managed grasslands). In addition, the Grassland Vegetation Probability Index (GRAVPI) raster maps are available (10 m and 20 m for 2018 and 2015, respectively) expressing the probability of occurrence of grassland in a specific pixel (range 1-100). This type of product, also called "expert product", provides additional information for advanced users on the spatial variation in the reliability of the mapping (e.g., low reliability can occur when the quality of images in the time series is low or the number of images is limited).

This expert product was selected as the most suitable to be used as covariate in the modelling exercise for the area estimation carried out in 2015 and 2018. In general, the quality and the production process of GRAVPI is higher in the 2018 product in terms of quality and spatial resolution. In terms of consistency and comparability, the products 2015 and 2018 were developed maintaining the full thematic correspondence. GRAVPI 2015 and GRAVPI 2018 pixel values were associated to the corresponding LUCAS Master point with spatial overlay operations[3]. Finally, each LUCAS Master point was endowed of the additional variables GRAVPI_2015 and GRAVPI_2018 with the following modalities:

- 0 (all non-grassland areas);
- 1-100 (1-100% Grassland vegetation probability index);
- 254 (unclassifiable, no satellite image available, or clouds, shadows, or snow);
- 255 (outside area).

[3] Two methods were tested for linking GAVPI pixel values to each LUCAS Master point: systematic association and bilinear interpolation. The first one associate the GRAVPI upper right pixel to the corresponding LUCAS Master point, while the second one performs the mean of the GRAVPI values of the 4 pixels surrounding the point. The bilinear interpolation was selected as final method for computing the covariates for 2015 and 2018.

Figure 1 – The product GRAVPI 2018 with the full geographical coverage of the EEA 39 countries. Spatial resolution is 10m and the coordinate reference system is Lambert Azimuthal Equal Area Projection (LAEA).



3 Definition of grassland classes

The following classes were defined for the analysis of the LUCAS Surveys points and their subsequent use for the model-based area estimation:

- Temporary grassland (TG);
- Permanent agricultural grassland (PGa);
- Permanent semi-natural grassland (PGsn);
- Other grassland (OG).

The aggregated class PG = PGa+PGsn was also defined to allow the computation of estimations whenever the LUCAS survey points were not sufficient to populate the subclasses (PGa and PGsn), hence, to provide reliable results. In addition, the Non-grassland (NG) class was defined to collect all the remaining LUCAS survey points. The sum of the LUCAS survey points associated to all classes must correspond to the number of LUCAS survey points for each campaign (2009, 2012, 2015 and 2018).

From a statistical/visual point of view, the advantage of using these categories will be to highlight the presence of clearly agricultural areas (TG and PGa) vs the others (PGsn and OG). This can help to visualise the areas with a clear agricultural use (e.g., in central EU) and the areas with semi-natural features (e.g., Mediterranean).

The initial focus was on the definition of rules for extracting LUCAS Survey points belonging to permanent and temporary grassland according to the available LUCAS definition for LC and LU classes and additional variables collected through the campaigns. Rules definition was addressed to avoid an overlap between the temporary and permanent grassland in order to represent two distinct concepts.

The use of additional variables allows defining specific rules to identify other types of grassland that might be hidden by the primary LC/LU associated to each point. These “hidden grassland” points should be placed outside the boundaries of the TG and PG classes. Therefore, the additional class Other Grassland (OG) was defined, to collect these additional points. OG contains grassland points associated with specific LCs such as woodland and shrub land, with a current or potential agricultural use or marginal agricultural lands with permanent grassland. The following table reports the description and details for each grassland class and is the results of the LUCAS data exploration and rules testing carried out during the subtask.

Table 8 – Grassland classes terms and definitions with the main LUCAS Survey elements considered in the classification.

Class code	Class name	Description	LUCAS survey points included	LUCAS survey points excluded
TG	Temporary agricultural grassland and pastures	Agriculturally improved grassland and temporary pastures typically used for fodder or for grazing included in the agricultural rotation.	Points with the primary LC (LC1) classified as Fodder crops (B50) - Clovers (B51), Lucerne (B52), Other leguminous and mixtures for fodder (B53), Mixed cereals for fodder (B54) and Temporary grasslands(B55) - with the primary (LU1) or secondary (LU2) agricultural LU (Agriculture, excluding fallow land and kitchen gardens, or Fallow land). Points with the secondary LC (LC2) classified as B50 and LC1 as Cereals (B10), Root crops (B20), Non-permanent industrial crops (B30), Dry pulses, vegetables and flowers (B40), Woodland (C00) with primary or secondary agricultural LU. Points with LC2 classified as B50 and LC1 as Other artificial areas (A30) with primary or secondary agricultural LU.	Permanent grassland areas with or without tree/shrub cover (E10 or E20), not included in the rotation.
PGa	Permanent agricultural grassland	Permanent grassland and permanent pasture that is not part of a crop rotation	Points with primary LC (LC1) classified as Grassland (E00) with the primary (LU1) or secondary (LU2) agricultural LU.	Points with the primary LU (LU1) classified as Road transport (U312), Air transport (U314), Construction (U330),

		(i.e., >=5 years), which can be used to grow grasses and other herbaceous forage naturally (self-seeded) or through cultivation (sown).	Points with the secondary LC (LC2) classified as E00 and LC1 as Other artificial areas (A30) with primary agricultural LU and with an herbaceous cover >=30%[4].	Commerce, financial, professional and information services (U340), Community services (U350), Recreation, leisure, sport (U360) and Residential (U370).
PGsn	Permanent semi-natural grassland	Permanent grassland areas occurring in shrubland and woodland areas with an agricultural use	Points with the primary LC (LC1) classified as Woodland (C00) or Shrubland (D00) with the primary (LU1) or secondary (LU2) agricultural LU and with an herbaceous cover >=30%.	Points with primary LC (LC1) classified as Woodland and secondary LC (LC2) classified as Fodder crops (B50) are moved to the TG class.
PG	Permanent grassland	The aggregation of PGa and PGsn.	See PGa and PGsn	-
OG	Other grassland	Grassland, woodland and shrubland with a minimum grassland coverage abandoned or with semi-natural and natural areas not in use. These	Points with the primary LC (LC1) classified as Grassland (E00) or Shrubland (D00) or Woodland (C00) with primary or secondary LU classified as Unused and abandoned areas (U400) and a minimum threshold of herbaceous coverage (>=30%).	Points with the secondary LC (LC2) classified as Permanent crops: fruit trees (B70) or Other permanent crops (B80).

[4] Assessed within the homogeneous plot inside the extended window of observation (20m radius) for each LUCAS survey point (INSPIRE PLCC module).

		can be areas with a current or potential agricultural use or marginal agricultural lands with permanent grassland not included in the permanent and temporary grassland classes.		
NG	Non grassland	All other areas	-	-

Since not all the variables used to assign LUCAS survey points to grassland classes are available for all the LUCAS campaigns (2009-2018) different classes were used for the analysis carried out within the task, see table below.

Table 9 – Grassland classes considered for the model-based area estimation and grassland trend analysis.

Analysis	Grassland classes	Years	NUTS level
Model-based area estimation	TG, PGa, PGsn, PG and OG	2015 and 2018	NUTS 0/1/2/3
Grassland trend	TG and PGa	2009, 2012, 2015 and 2018	NUTS 0/1/2

3.1 General criteria for rules definition

Hereafter the main criteria used for the definition of rules for assigning LUCAS survey points to grassland classes are listed. The criteria were defined after the analysis of the LUCAS classification rules, nomenclature, variables definition and data exploration based on a different combination of the most grassland-relevant variables by testing several rules for each grassland class (Cf. 4.1).

1. LC is the most relevant variables for defining the membership of a point to a grassland class.
2. The primary LC (LC1) is the most relevant LC for the rules concerning OG and PGsn classes.
3. The agricultural LUs (U111 and U112) are the only LUs relevant for the eligibility of points for the PG and TG classes.
4. The agricultural LU U113 (Kitchen garden) is not relevant for the definition of the grassland classes.
5. The presence of herbaceous cover for points that are classified as B7* or B8* is deemed not relevant for their eligibility for the OG class.
6. The presence of an herbaceous cover with a minimum threshold is a prerequisite for the definition of rules for PGsn and OG classes.
7. The 30% threshold can be acceptable for defining the herbaceous coverage (variable INSPIRE_PLCC4[5]). The threshold is also in line with the definition of grassland coverage in the Copernicus HRLs products.

3.2 LUCAS variables and modalities considered in the definition of selection criteria

The following table reports the selected LUCAS variables and modalities for the final selection of eligible LUCAS Survey points classified as TG, PGa, PGsn and OG. Availability of the variables across the LUCAS Survey 2009-2019 is also reported.

Table 10 – List of LUCAS survey variables used in the rules for grassland classes and availability for each campaign.

Grassland class	LUCAS variable	LUCAS variable modalities/threshold value	Availability of variables by LUCAS surveys			
			2009	2012	2015	2018
TG	LC1	B10, B20, B30, B40, B50, C00	x	x	x	x
	LC1	A30			x	x
	LC2	B50	x	x	x	x
	LU1	U111, U112	x	x	x	x

[5] The variable is collected within the LUCAS module INSPIRE Pure Land Cover Component (PLCC) to improve the mapping with other products by collecting additional information for each surveyed point relative to the composition (0-100%) of the LC. Data are collected for the points where LC1 is either woodland (C00), shrub land (D00), grassland (E00) or bare land (F00) and is to be 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%, the sum of PLCC must be 100%. Assessment of these percentages is made using the "birds-eye" view.

	LU2	U111, U112	x	x	x	x
PGa	LC1	E00	x	x	x	x
	LC2	A30			x	x
	LU1	U111, U112, U312, U314, U330, U340, U350, U360, U370	x	x	x	x
	LU2	U111, U112	x	x	x	x
	INSPIRE_PLCC4	>=30%			x	x
PGsn	LC1	C00, D00	x	x	x	x
	LC2	B50	x	x	x	x
	LU1	U111, U112	x	x	x	x
	LU2	U111, U112	x	x	x	x
	INSPIRE_PLCC4	>=30%			x	x
OG	LC1	E00, C00, D00	x	x	x	x
	LC2	B70 and B80	x	x	x	x
	LU1	U400	x	x	x	x
	LU2	U400	x	x	x	x
	INSPIRE_PLCC4	>=30%			x	x

3.3 Rules for selecting LUCAS surveys points for the grassland classes

The following table summarises the rules defined for the different grassland classes: TG, PGa, PGsn and OG. Rules are reported in pseudo-code and can be converted in SQL language to extract the corresponding LUCAS Survey points from the microdata. Each grassland class is defined by more than one rule; hence, the extraction of LUCAS points needs to be carried out by combining the different rules: the rules are applied with a logical operator OR.

Report

Table 11 – Final rules defined for the different grassland classes.

ID	Grassland class	LUCAS variables	LC classes	LU classes	Rule	notes
1_11	TG	LC1, LU1, LU2	B50	U111, U112	LC1 = B5* AND (LU1 = (U111 OR U112) OR LU2 = (U111 OR U112))	
1_12	TG	LC1, LC2, LU1, LU2	A30, B50	U111, U112	(LC1= A30 AND LC2=B5*) AND (LU1 = (U111 OR U112) OR LU2 = (U111 OR U112))	A30 is missing in 2009 and 2012
1_13	TG	LC1, LC2, LU1, LU2	B10, B20, B30, B40, C00	U111, U112	(LC1= B1* OR B2* OR B3* OR B4* OR C*) AND LC2=B5*) AND (LU1 = (U111 OR U112) OR LU2 = (U111 OR U112))	
2_11	PGa	LC1, LU1, LU2	E00	U111, U112	(LC1 = E* AND LU1 = (U111 OR U112)) OR (LC1 = E* AND LU1 <> (312 AND 314 AND 330 AND 340 AND 350 AND 36* AND 370) AND LU2 = (U111 OR U112))	
2_12	PGa	LC1, LC2, LU1, LU2, INSPIRE_PLCC4	A30, E00	U111, U112	((LC1=A30 AND LC2=E*) AND (LU1 = (U111 OR U112) OR (LC1=A30 AND LC2=E*) AND LU1 <> (312 AND 314 AND 330 AND 340 AND 350 AND 36* AND 370) AND LU2 = (U111 OR U112))) AND INSPIRE_PLCC4>=30	The rule produces a very limited number of points (126 in 2018 and 0 in 2015). The rules cannot be applied in 2009 and 2012 due to the missing variable INSPIRE_PLCC4.
170	PGsn	LC1, LU1, LU2, INSPIRE_PLCC4	C00	U111, U112	LC1=C* OR (LC1=C* AND LC2 <> B5*)) AND (LU1= (U111 OR U112) OR LU2= (U111 OR U112)) AND INSPIRE_PLCC4>=30	The points with LC1=C* AND LC2= B5* belonging to TG are removed
180	PGsn	LC1, LU1, LU2, INSPIRE_PLCC4	D00	U111, U112	LC1=D* OR (LC1=C* AND LC2 <> B5*)) AND (LU1= (U111 OR U112) OR LU2= (U111 OR U112)) AND INSPIRE_PLCC4>=30	
2_2	OG	LC1, LU1, LU2	E00	U400	LC1 = E* AND (LU1 = U4* OR LU2 = U4*) AND INSPIRE_PLCC4>=30	INSPIRE_PLCC4 is missing in 2009 and 2012
17_2	OG	LC1, LU1, LU2, INSPIRE_PLCC4	C00	U400	(LC1=C* AND (LC2 <> B7* AND B8*)) AND (LU1=U4* OR LU2=U4*) AND INSPIRE_PLCC4>=30	
18_2	OG	LC1, LU1, LU2, INSPIRE_PLCC4	D00	U400	LC1=D* AND (LU1=U4* OR LU2=U4*) AND INSPIRE_PLCC4>=30	

3.4 Application of rules for the grassland classes to LUCAS Surveys 2009-2018

Hereafter a series of graphs is reported for each LUCAS Survey year at EU-level and NUTS 0 by combining all the defined rules for each grassland class.

Report

Figure 2 – Number of LUCAS survey points for each grassland class by year at EU-level (2018: EU28, 2015: EU28, 2012: EU27, 2009: EU23). According to the rules, the four grassland classes can be defined only for 2015 and 2018.

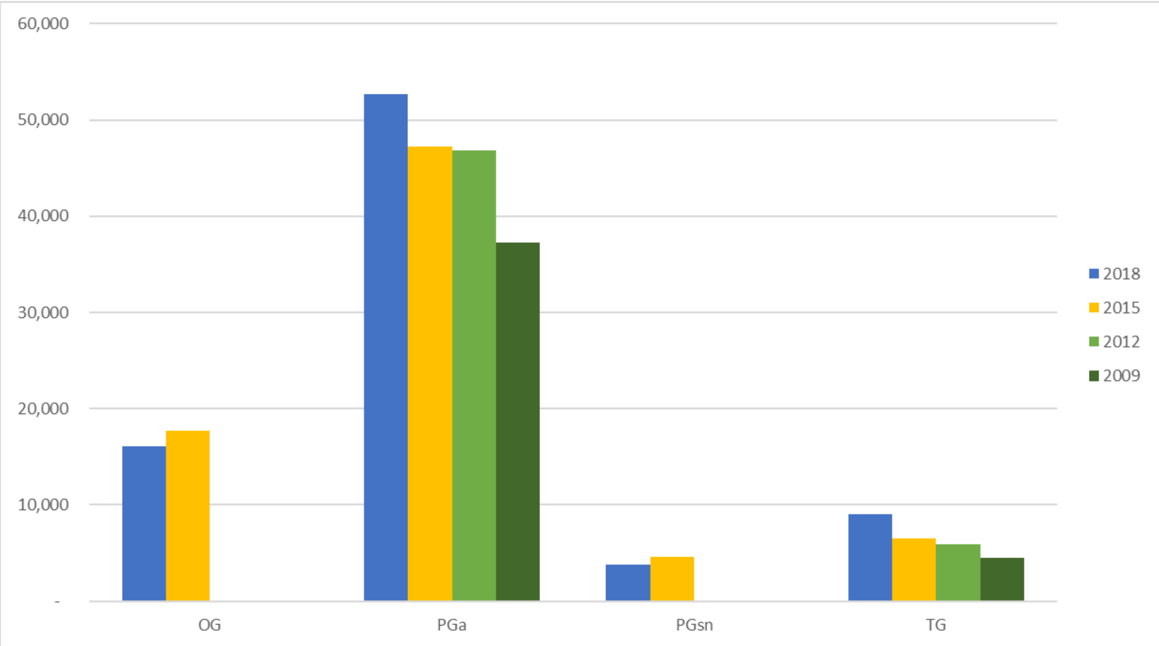
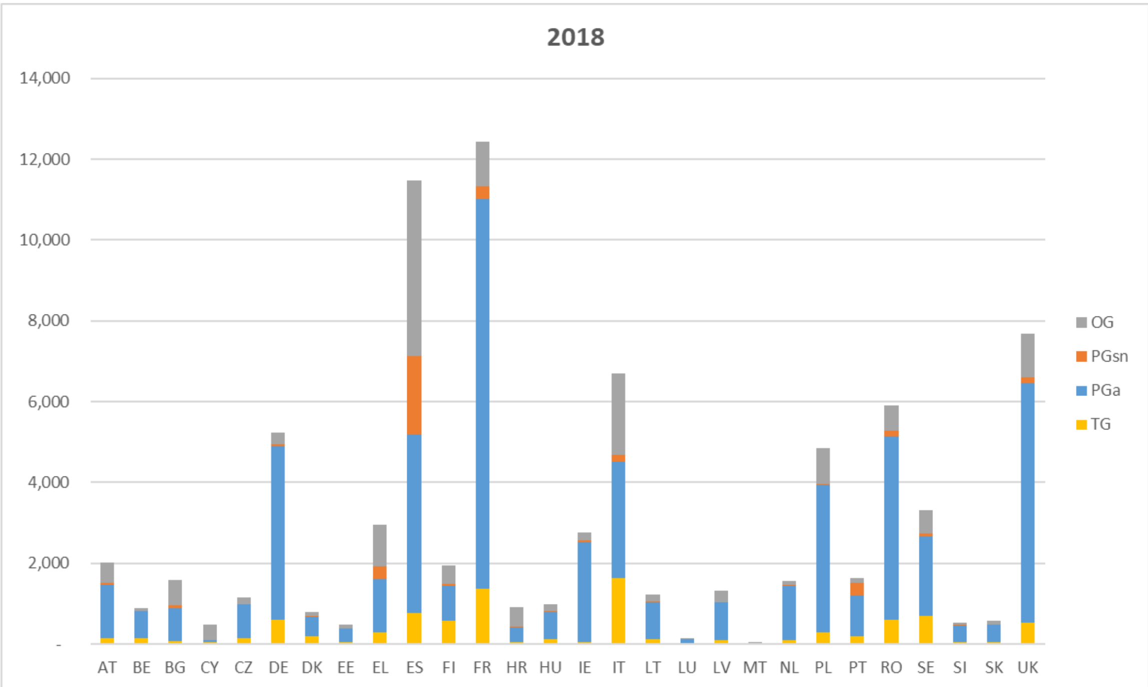


Figure 3 – Number of LUCAS survey points for each grassland class by country (Survey year: 2018).



Report

Figure 4 – Number of LUCAS survey points for each grassland class by country (Survey year: 2015).

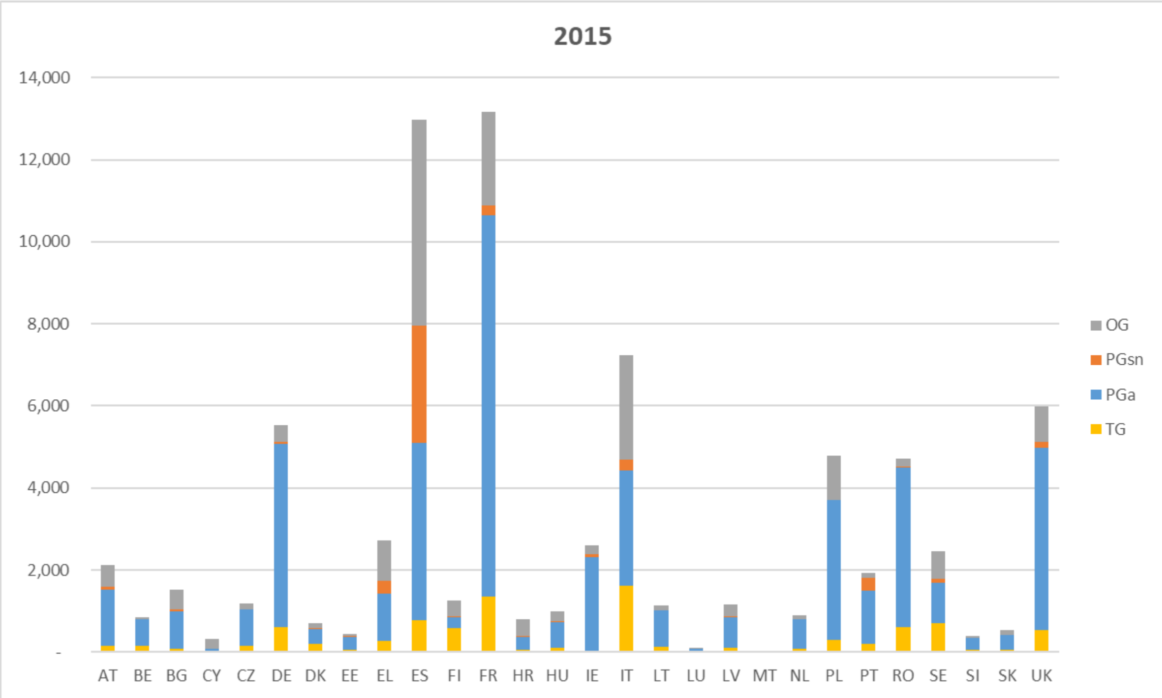


Figure 5 – Number of LUCAS survey points for each grassland class by country (Survey year: 2012).

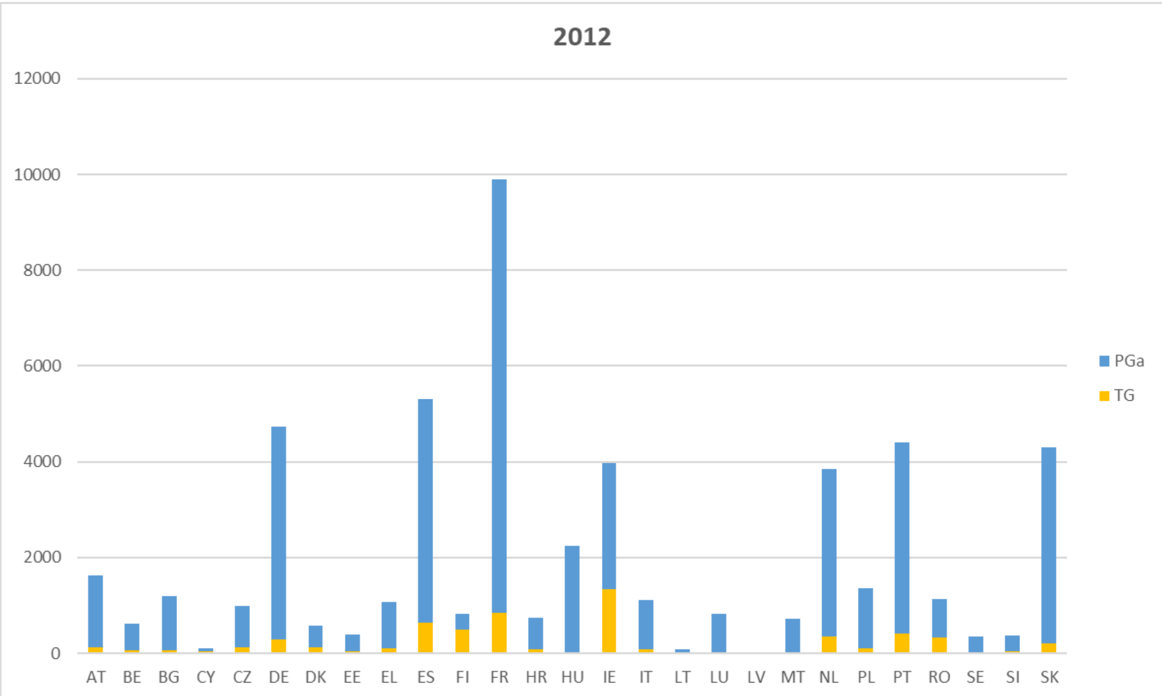
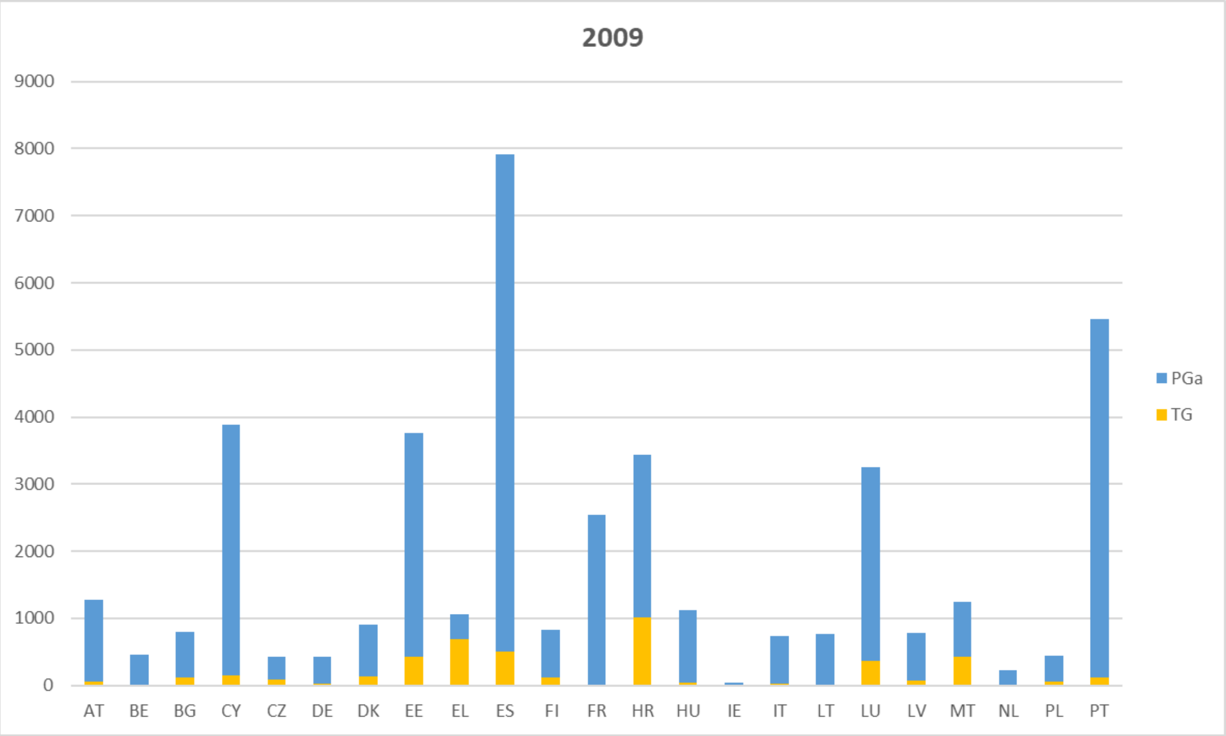


Figure 6 – Number of LUCAS survey points for each grassland class by country (Survey year: 2009).



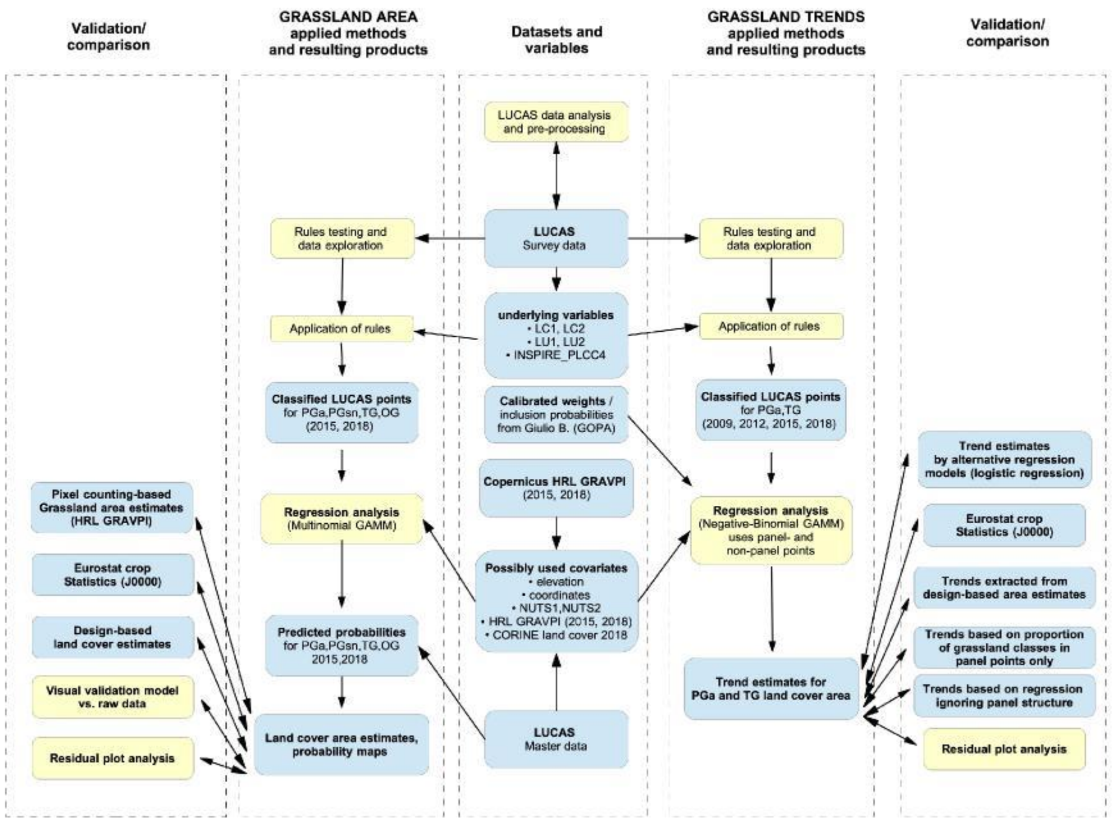
Report

4 Area and trend estimation for grassland classes: methods, validation, and results

4.1 Introduction

In the following, we present methods, underlying data, validation, and comparison steps with respect to land cover area and trend estimates for different grassland classes in the EU. An overview scheme is given in Fig. 7. With respect to the definition and application of rules in order to define the different grassland classes, we kindly refer to the previous chapters of this document.

Figure 7 – Overview scheme of the underlying data and applied approaches to obtain land cover area and trend estimates for different grassland classes in the EU.



5 LAND COVER AREA ESTIMATIONS

5.1 Statistical methods

In order to statistically estimate areas from in situ and/or remote sensing samples, there exist various methods which can be divided into three main classes: design-based methods (e.g., [12]), model-based methods (e.g., [26]), as well as model-assisted methods (e.g., [36]). Here, design-based methods strongly rely on a correct sampling design. Simply speaking, they estimate the area by multiplying each sample with a factor being related to the total area, which this sample point represents, but in fact, this factor does not necessarily represent the surrounding of the point, but instead it´s only a measure of the statistical representativeness of the points of the frame. The issue is that this area could be fragmented and spread everywhere. This factor is proportional to the inverse of a sample point's inclusion probability: the higher the inclusion probability (e.g., defined per stratum) is, the lower is of course the total area associated with a point from the corresponding stratum. Such design-based approaches depict the traditional way to perform land cover area estimation.

During the last decades, an alternative approach to design-based methods has been established, being increasingly used in the context of geo- and biostatistics, namely model-based methods [19, 2, 27]. Here, in a first step, statistically valid correlations between several possible auxiliary variables ("co-variables"/"predictors") and the variable of interest (e.g., grassland type) are estimated with appropriate regression methods applied to the sample. Then in a second step, the variable of interest can be predicted for each point in the area of interest – including those where no sample (but of course co-variable values) exist (c.f. e.g., [8, 9, 16, 19, 22]).

The advantage of such approaches compared to sampling-based methods are versatile:

1 they do not any longer sensitively depend on a proper sampling design, since model-based predictions for all existing points are used instead of inclusion probabilities;
2 the consideration of additional co-variables may strongly increase the area predictions, since the effect of continuously and locally changing variables is considered instead of the use of artificial/coarse strata;
3 model-based approaches can be used for predictions (e.g., with respect to future changes in co-variables) [28, 6];
4 various difficulties frequently connected to ecological/geological data can be considered within modern regression approaches, such as temporal and/or spatial autocorrelation, various types of predictor or outcome variables, and nonlinear dependencies between predictors and outcome variables [27];
5 there is no lower limit of model-based area estimates: they can in principle be calculated for arbitrary small sub-regions; even if not a single sampled point lies within such a sub-region, land cover area can be calculated (and are based on interpolation of surrounding sample points in conjunction with covariate values in the sub-region).

Thus, it is not surprising that under some circumstances, model-based estimated can produce more reliable results compared to design-based methods [7, 4, 25] – importantly without the need to change the underlying stratified study design. However, model-based approaches are only finding its way into geo- and bio statistical practice relatively slowly, as it is relatively young – compared to sampling-based approaches – and also requires advanced statistical regression software, which has only been used primarily in the last decade (e.g., through the establishment of the open source software R [32]) has experienced wide application and flexibility. For the sake of completeness, we want to point out that model-assisted approaches already exist, which are a compromise between design-based and model-based approaches for land cover area estimation [36].

5.2 Model-based approach

Area estimations have been performed separately for each country and year (2015 and 2018). In particular, we used multinomial regression model approaches [35] in order to model the outcome variable consisting of the (non-ordered) levels ”permanent grassland with agricultural use” (PGa), ”permanent grassland semi-natural” (PGsn), ”temporary grassland” (TG), ”other grassland” (OG) and ”non-grassland” (NG). Models have been fitted to LUCAS *in situ* data from which the above-mentioned grassland classes have been defined based on the rules presented in section 3.

In particular, these models (belonging to the class of generalized linear models/GLMs [5, 11]) estimate the probability for each of the different above-mentioned classes, where the sum of the 5 probabilities in each point always sums up to 1 [35]. As possible fixed-effect co-variables, we used the variables ”X_LAEA” and ”Y_LAEA” to account for spatial inhomogeneity in the probabilities, as well as ”ELEVATION”. Furthermore, we used two variables being strongly related to the probability of grassland vs. non-grassland: ”CLC18” (based on CORINE land Cover) as well as ”GRASVPI” (a probability of grassland occurrence based on photointerpretation). Finally, we additionally included the variables ”NUTS1_16 as well as ”STRATUM_LABEL”, the latter being a variable defining the different strata used during stratified sampling. Since ”NUTS1_16” and ”STRATUM_LABEL” frequently comprise several levels, we included them as random instead of fixed effects (thus, using generalized mixed modelling approaches [31, 42]). In case ”NUTS1_16” comprised less than 4 levels, ”NUTS2_16” has been used instead as a random intercept. To select the most appropriate model (separately for each country and year), we compared 9 different models with different combination and types of the above mentioned co-variables as predictors. Of course by also including a 2D smooth term of the variables ”X_LAEA” and ”Y_LAEA” as well as an 1D smooth of ”ELEVATION” (both realized in the framework of generalized additive modelling /GAMs [17, 18, 39]). We further selected the model with the lowest AIC (”Akaike information criterion”) value [1] which is a frequently used approach during model selection [13, 42, 20]. Thus, since we combined generalized, mixed, and additive modelling, we finally worked with generalized additive mixed models (GAMMs) for the model-based estimation of land cover area.

It is important to point out that by providing only partially matching variables as co-variables we do not introduce any bias with respect to the model-based estimates: if a provided co-variable does not correlate with the outcome variable (i.e., the variable of interest), it will be excluded in the AIC-based model selection procedure. If a co-variable is kept within this procedure, the model will use only existing/detected correlation for final predictions/area calculations. Thus, there is no *a priori* input, which can produce bias, in contrast, model-based estimations are purely data driven.

After models have been fitted to the data, for some randomly selected countries, model residuals have been checked to determine if all assumption underlying a regression analysis were met (such as independence, homogeneity, and linearity [13, 40, 42, 20]). In particular, we checked possible spatial autocorrelation of model residuals [23, 10, 41] with bubble-plots and semi-variogram analyses [20, 42] in order to prevent underestimated standard errors and/or over-smoothing. In the spot-checked countries, there was no distinct spatial autocorrelation detected, most probably since various spatial covariates (including the spatial 2D-Smooth) have been used which may strongly reduce the autocorrelation within the model residuals.

Final land cover area calculations were based on model predictions. In particular, all co-variables used during GAMM-fits applied to the LUCAS data were provided similarly defined for the entire master grid, the latter consisting of point data in a 2x2 *km* grid. Subsequently, probabilities for the different grassland classes have been calculated based on model predictions and subsequently multiplied by 4 (since each master grid point is in the middle of a 2x2 *km* grid cell). Further thoughts/validations with respect to this approach (particularly the question how well the entire area of the 4 km^2 patch is represented by the master point in its center) are presented/discussed in Section 5.4.

The uncertainties connected to each area estimation (for each NUTS0, NUTS1, NUTS2, and NUTS3 level) have been calculated based on model prediction matrices in conjunction with simulations from the posterior distribution of the different model parameters, of course taking the variance-covariance structures appropriately into account [39]. Final variances, the coefficient of variation (COV) and 95 %-confidence intervals have been calculated directly from 1000 resamples of model-based area estimations with respect to the considered NUTS-level and -ID.

5.3 Software

For all statistical calculations and visualizations, the statistical open-source software R [32] has been used. For regression analysis, we used the R-package mgcv [39], and for visualizations the package ggplot2 [38].

5.4 Possible problems, validation, and comparison

5.4.1 The problem of in situ small-scale data

In the approach described above, we associate with each point with in situ LUCAS information the corresponding master pixel, along with some additional covariates (e.g. based on HRL information and Corine land cover). Subsequently, we use these pixels (corresponding to the number of surveyed points) to fit an appropriate regression model (GAMM), correlating the LUCAS-based variable of interest (describing the different classes of grassland - c.f., above) with the covariates. Finally, we use the fitted GAMM to predict the different grassland classes for all pixels of the master including those where no in situ information is available. Final land cover areas are calculated by summing up the predicted probabilities to belong to a certain class, and multiplying this number by 4 (in order to relate it to the entire 4 km^2 grid cell).

A potential problem with this approach is that LUCAS data is assessed at small patches of approx. 20x20 meters. Covariate information is obtained on a much coarser spatial scale, e.g., 100x100 or 250x250 Meters. In our current approach, however, we implicitly assume that both - LUCAS data and covariate values - represent the values of the entire 4 km^2 belonging to a master pixel, which is obviously not true. Consequently, the calculated model-based uncertainties (confidence intervals/variance/COV) reflect only the uncertainty related to those pixels where no LUCAS data have been obtained (and thus the model has to predict the values), but not the uncertainties connected to the fact that LUCAS data and covariates may not always be representative for the entire patch of 4 km^2. Thus, simply speaking we pretend that the local covariate values (e.g., evaluated at the LUCAS point or in its surroundings) apply to the entire 4 km^2 area surrounding each master point – which is of course never the case, they can be more or less representative. However, one can assume that the larger the evaluated area, the smaller the problem, since it will average out.

A statistically more proper way would be calculating the covariates (if possible) for the exact 20x20 meter patch where LUCAS information has been obtained and subsequently fitting the regression models to these data. Ideally, we then use a grid of 20x20 meters for the entire study area for our predictions. However, this is not possible, since this would be far beyond feasible computation times (since for uncertainty estimations resampling with 1000 resamples per region is involved – c.f., above). Alternatively, in order to reduce the computation time, we could pre-process all covariates by (1) evaluating them indeed for 20x20 meter patches for the entire study area, and (2) pooling/averaging them on a coarser grid (e.g., 1 km^2) more feasible for the time-consuming regression-based predictions.

Since such a pre-processing of the covariates is connected to some effort, we wanted to know if this approach is worth the effort and performed a simulation study with a scenario strongly related to our real situation. In particular, we simulated a grid of N cells (their centre representing master points), and each cell is again divided into 100 sub-cells (representing 20x20 m covariate and possibly LUCAS information). Subsequently, we randomly distributed a covariate in these N × 100 sub-cells and based on this we created a binary random grassland variable correlating with the covariate. So finally, we obtain a virtual study area consisting of N master points, but grassland and covariate information is given on a 100 times finer spatial scale. In a second step, we created a random subset of N/2 master points with grassland and covariate information (which is our virtual LUCAS sample) and fitted a regression model. Finally, we estimated the area by (A) the ”old” approach, where we predict the grassland type only based on the information given in the sub-cell in the centre of each large cell and multiply this by the factor 100, (B) the ”new” approach, where we predict the grassland type for each sub cell of the entire study area. In both cases, we also calculate 95 % confidence intervals for the area estimates. Eventually we tested if the true area value falls into the calculated confidence intervals. The entire process was repeated 2000 times for N=81 and N=25.

Since this is a simulation study, the difference to the real data is, however, that the true area occupied by a certain grassland type is known here. The results reveal that both approaches – our "old" approach as well as the proposed "new" approach – give very similar results with unbiased area estimates as well as confidence intervals showing type I error rates at or below the nominal level of alpha = 0.05 – even if area calculation is performed for small regions, namely 25 master points only:

- 25 Master points: old approach: alpha = 0.0285; new approach: alpha = 0.027;
- 81 Master points: old approach: alpha = 0.048; new approach: alpha = 0.0465;

where alpha is the type I error rate (with a nominal level of alpha = 0.05). From this we can conclude that our "old" approach is still valid and produces (if at all) negligible bias – even on small spatial scales. Several discussions with other LUCAS statisticians confirmed this view. In addition, they pointed out: ...given that your simulation has demonstrated that the results obtained with the two approaches are not significantly different, and the old one is also preferable from the implementation point of view, there should be no doubts about the final choice.

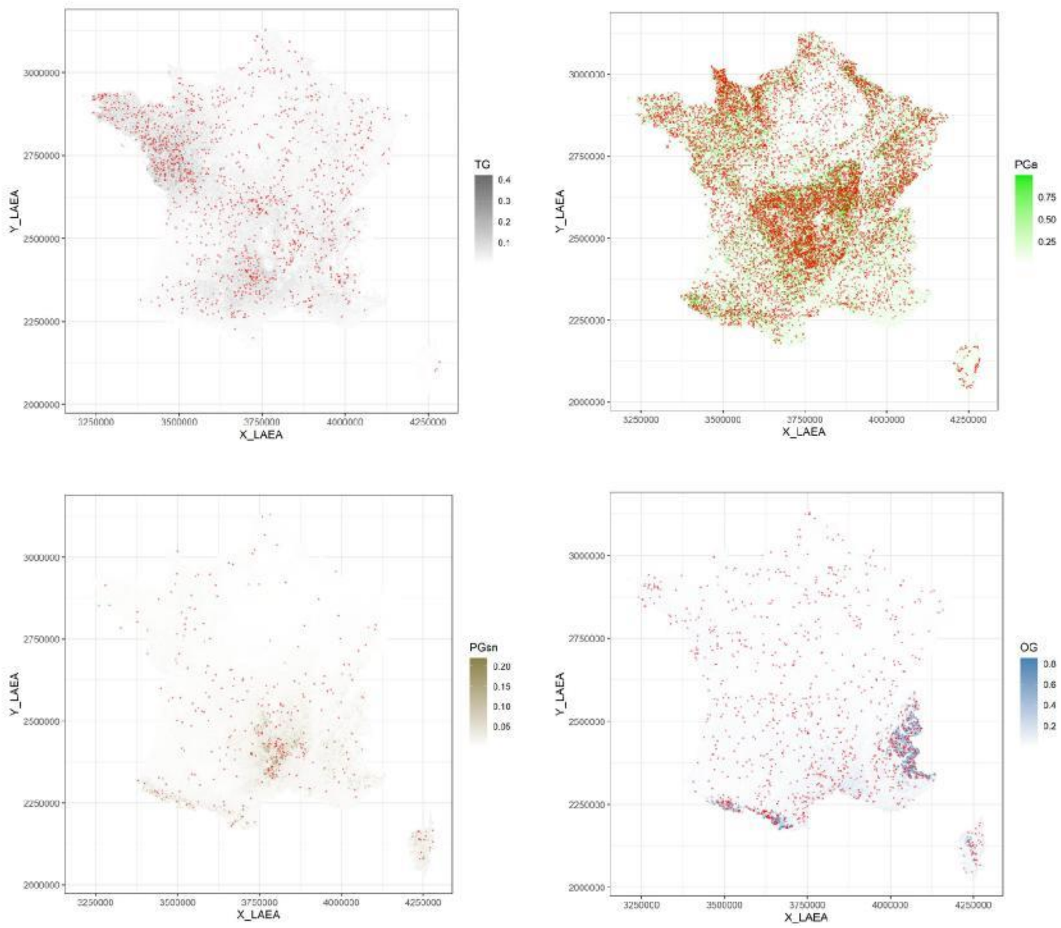
As an additional validation, Flavio Lupia provided the "averaged covariates" as described above ("new" approach) exemplarily generated for Italy for all NUTS levels 1-3. The land cover area estimations using the "old" and the "new" approach have been performed and compared. It appeared that even on NUTS3-level, estimated areas differ between both approaches by 9 % only – the same order of magnitude holds for the outer values of the corresponding confidence bands. Keeping in mind that this difference will even decrease if we consider higher NUTS-levels, since we will have more total points available – the problem will increasingly vanish for higher levels by "averaging out".

Thus, both – simulated data as well as real data – suggest that there is only a minor difference between both approaches, and we decided to stick to the old approach (mainly for the reasons of feasibility with respect to implementation and computation times).

5.4.2 Visual Validation

For additional validation, we spatially plotted the predicted probabilities along with the underlying LUCAS in situ point data (separately for each country and grassland level combination) to validate separately for each country if model-based patterns reasonably reproduce the patterns observed in the underlying point data. These plots are provided (deliverable D2.1.1) and no obvious deviations were detected. Furthermore, we plotted (again separately for each country and grassland level combination) our model-based land coverage area estimations (along with confidence intervals) and compared them with the corresponding design-based (model-assisted) estimates. Here as well, no problems were detected (for an example, c.f., Fig. 8).

Figure 8 – Visual validation of model based estimated probabilities (continuous colours) vs. in situ LUCAS points belonging to the grassland class of interest (red dots) – taking the example of France

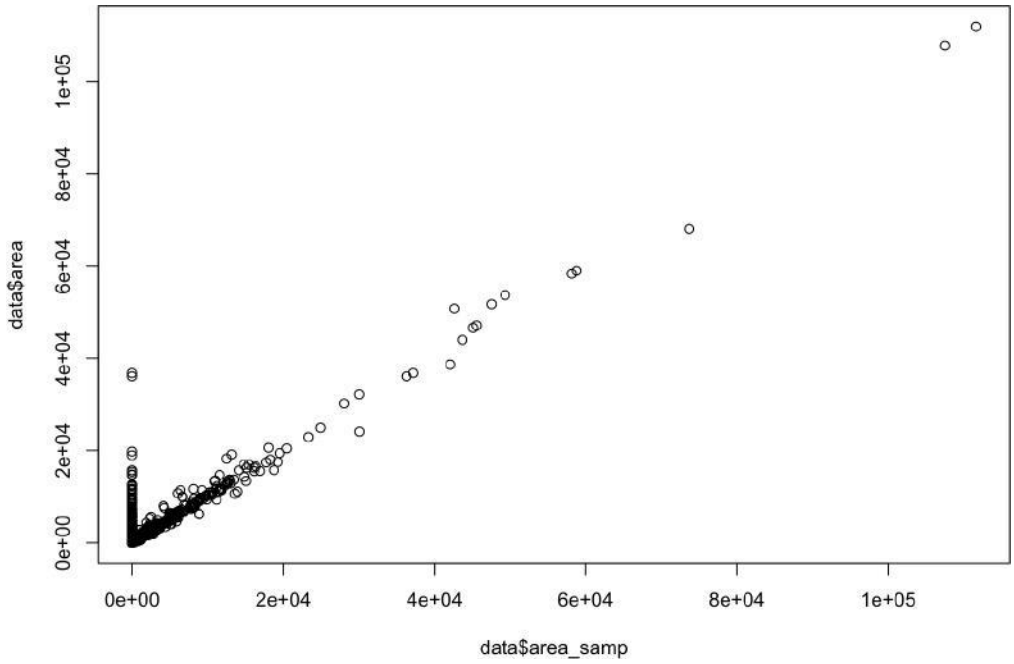


5.4.3 Validation/comparison with design-based (model-assisted) estimates

An important step is the comparison between model-based and design-based estimates. In particular, LUCAS statisticians have previously calculated the weights for the design-based estimates. Actually, their approach was a model-assisted approach, since they calculated calibrated weights from the initial weights (inverse of the probabilities of inclusion assigned to each point in the master) taking into account some important parameters in the population (as the total areas, the elevation classes and HRL and Corine Land Cover classes) using machine learning techniques.

A plot of the model-based vs. design-based (model-assisted) land cover area estimates is given in Fig. 9 where the x-axis represents the design-based Estimates and the y-axis the model-based estimates. The plot is based on all existing NUTS0-NUTS3-levels in the EU. It appears that both measures are strikingly correlated; there is not one single strong outlier. Only where design-based estimates are zero, there are several non-zero model-based estimates. These values represent NUTS-level where not one single LUCAS point is inside. Here, design-based estimates fail where model-based estimates are still possible. If we remove these values (i.e., restricting the data to those where design-based estimates are larger than zero) the correlation between both approaches is with $r = 0.9956$ and $p < 2.2e-16$ very high. Thus, we can conclude that for those subareas where LUCAS points are available, point estimates from both methods perform equally well. However, there could be differences in variance estimates between both methods (e.g., since model-based estimates may cope better with spatial autocorrelation) which we did not systematically evaluate.

Figure 9 – Comparison of model-based vs. design-based (model-assisted) land cover area estimates. X-axis: design-based estimates, Y-axis: model-based estimates. The plot is based on all existing NUTS0-NUTS3-levels in the EU

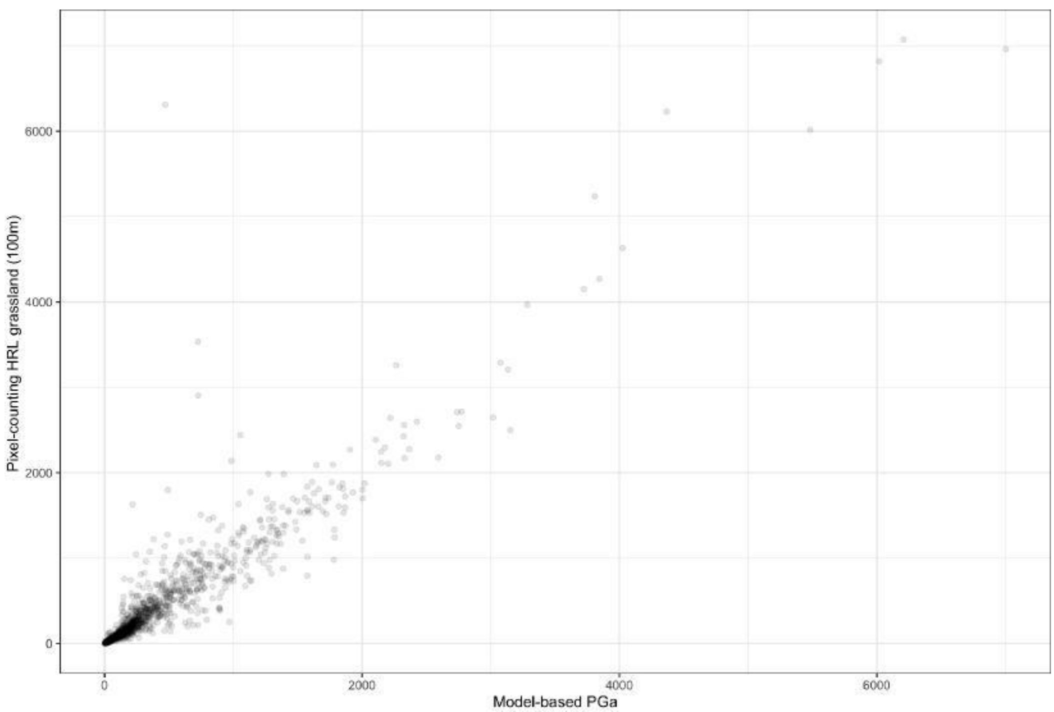


5.4.4 Validation/comparison with external data sets.

Another comparison/validation has been done by comparing the model-based land cover area estimates with respect to "permanent grassland" (PG) respectively "permanent grassland with agricultural use" (PGa) to pixel count results based on HRL Grassland at 100m respectively 10m resolution at NUTS2 and NUTS3 level. In addition, here, for all resolutions (10 m vs. 100 m) and both types of grassland (PG vs. PGa) there is a good correlation with the pixel count-based estimates of $r \approx 0.9$ (for an example, c.f., Fig. 10). Since here (and e.g., in contrast to our model- vs. design-based comparison), the correlated measures rely on independent data sources, the value of $r \approx 0.9$ can be assessed as a good result. Furthermore, the correlation line would have a slope of approx. 1, which means that even in absolute values of land cover area, both measures are strongly comparable. These results are surprisingly good considering the facts that (1) we consider NUT3-levels, i.e., small areas with the highest model uncertainties; (2) pixel count methods are prone to systematic bias (several studies exist discussing this topic, e.g., [15]); and (3) the definitions of grassland in the HRL context and our definitions of PG respectively PGa may differ in some details.

Finally, we also compared our model-base land cover area estimates for PGa with the area estimates as provided by official crop statistics from Eurostat respectively Faostat on NUTS0 level. The correlation plots (based on Kendal's tau) are given in Fig. 11. It appears that there is a good positive and highly significant correlation between our estimates and the official crops statistics. In particular, the slope of regression lines is approximately 1, indicating that not only relative spatial differences but also absolute values are very similar.

Figure 10 – Comparison of model-based (PGa) vs. pixel-count HRL-based (grassland) land cover area estimates. X-axis: model-based PGA estimate, Y-axis: pixel-count HRL-based estimates. The plot is based on all existing NUTS3-levels in the EU



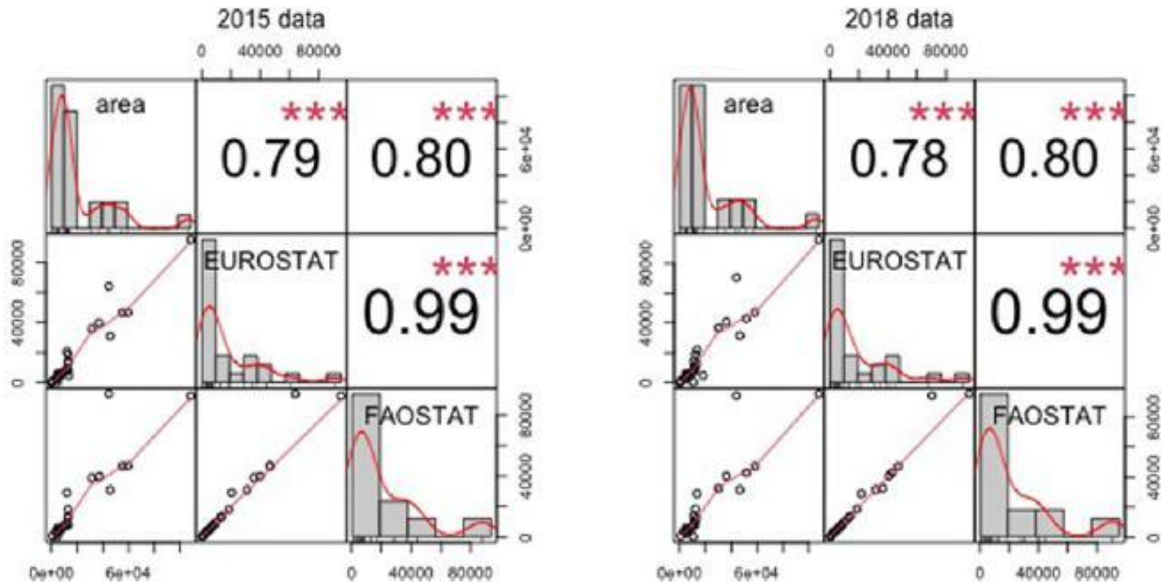
5.5 Results

The final results (separately evaluated for 2015 and 2018 as well as for the levels "permanent grassland with agricultural use" (PGa), "permanent grassand semi-natural" (PGsn), the merged level "permanent grassland" (PG) consisting of PGa + PGsn, "temporary grassland" (TG), and "other grassland" (OG)) are given in two ways:

1 by a summary data frame comprising model-based land cover area estimates along with several measures of uncertainties (such as 95 % confidence limits as well as COV) and the design-based (model-assisted) area estimates for comparison for each NUTS0, NUTS1, NUTS2, and NUTS3 level (deliverable D2.1.2);

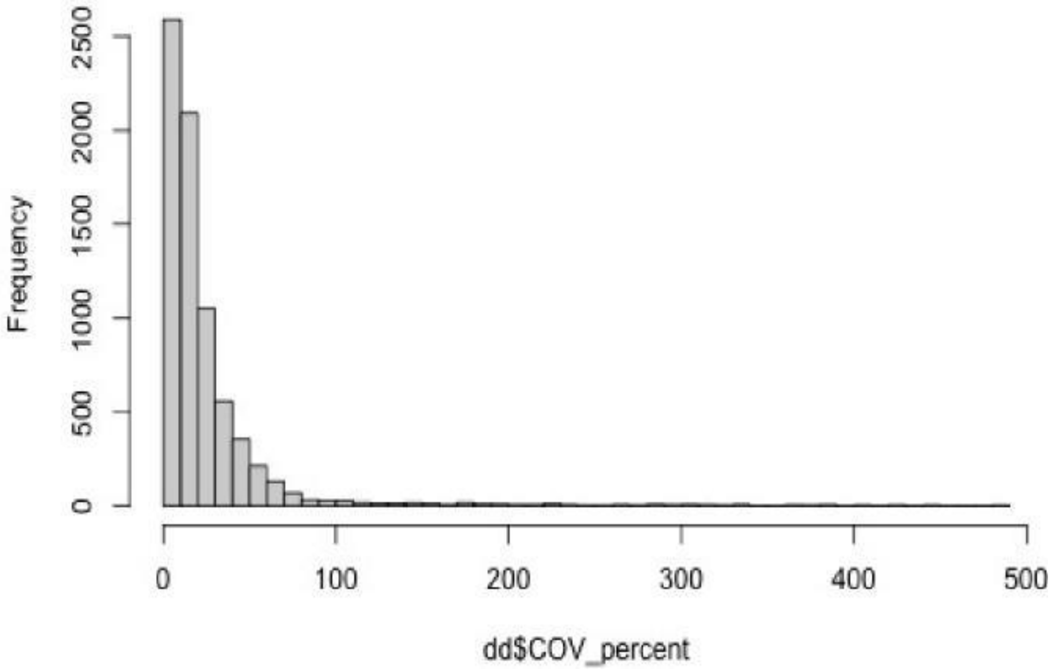
2 by the master data frame where the model-based probabilities for the different grassland classes are attached to each master pixel (deliverable D2.1.3).

Figure 11 – Pairwise correlations (based on Kendal's tau) of model-based PGa land cover area estimates with official grassland area estimates from Eurostat respectively Faostat for 2015 and 2018 (NUTS0 level).



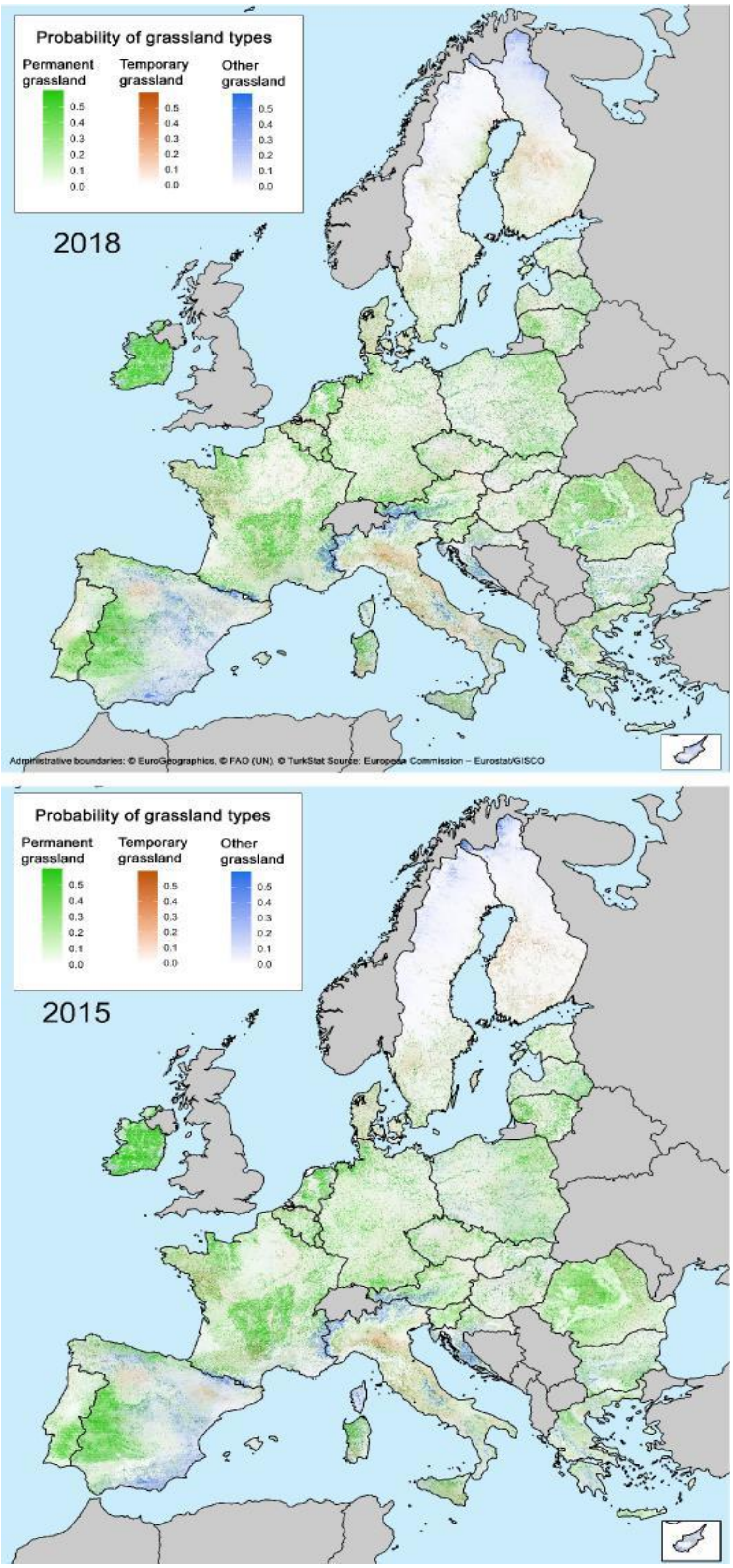
The model-based estimated probabilities of different grassland types for 2015 and 2018 are shown in Fig. 13 (related to point 2 above). Comparisons between the total and relative land cover area estimates for 2015 vs. 2018 are also shown in Fig. 14-16. Although the estimates of both years rely on completely independent data sets, similarities are striking, demonstrating both, a dense data basis as well as robust model results. Furthermore, although probabilities are independently estimated for each country, some probability patterns obviously cross boundaries (c.f., for example "permanent grassland" (PG) in the Extramadura in Spain and adjacent areas in Spain in Fig. 14) also underlining the robustness of the applied approach. Other patterns in contrast show strong relations to country boundaries (c.f., for example "temporary grassland" (TG) at the border between Italy and France in Fig. 14) – however, these borders are frequently connected to geographic barriers such as mountains. Numerical data on land cover area estimates for 2015 and 2018 on NUTS0 level are given in Tab. 12-14.

Figure 12 – Distribution of COVs for land cover area estimation pooled over all grassland classes and considered NUTS levels



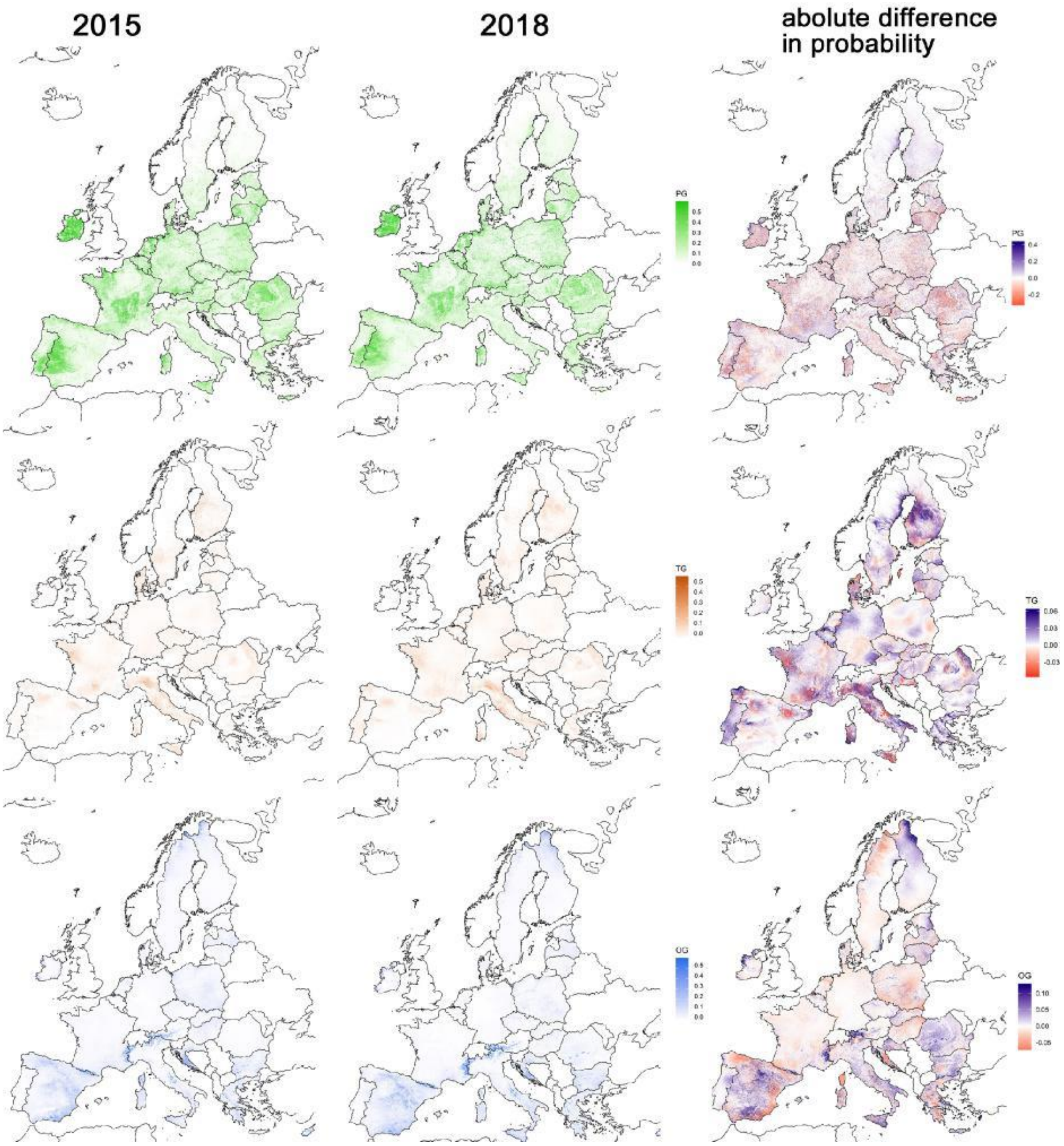
Report

Figure 13 – Model-based estimated probabilities of different grassland types for 2018 and 2015



Report

Figure 14 – Model-based estimated probabilities of different grassland types for 2015, 2018 and the absolute difference in probability between both years



Report

Figure 15 – Model-based estimates of total land cover area of the different grassland classes for each country and year

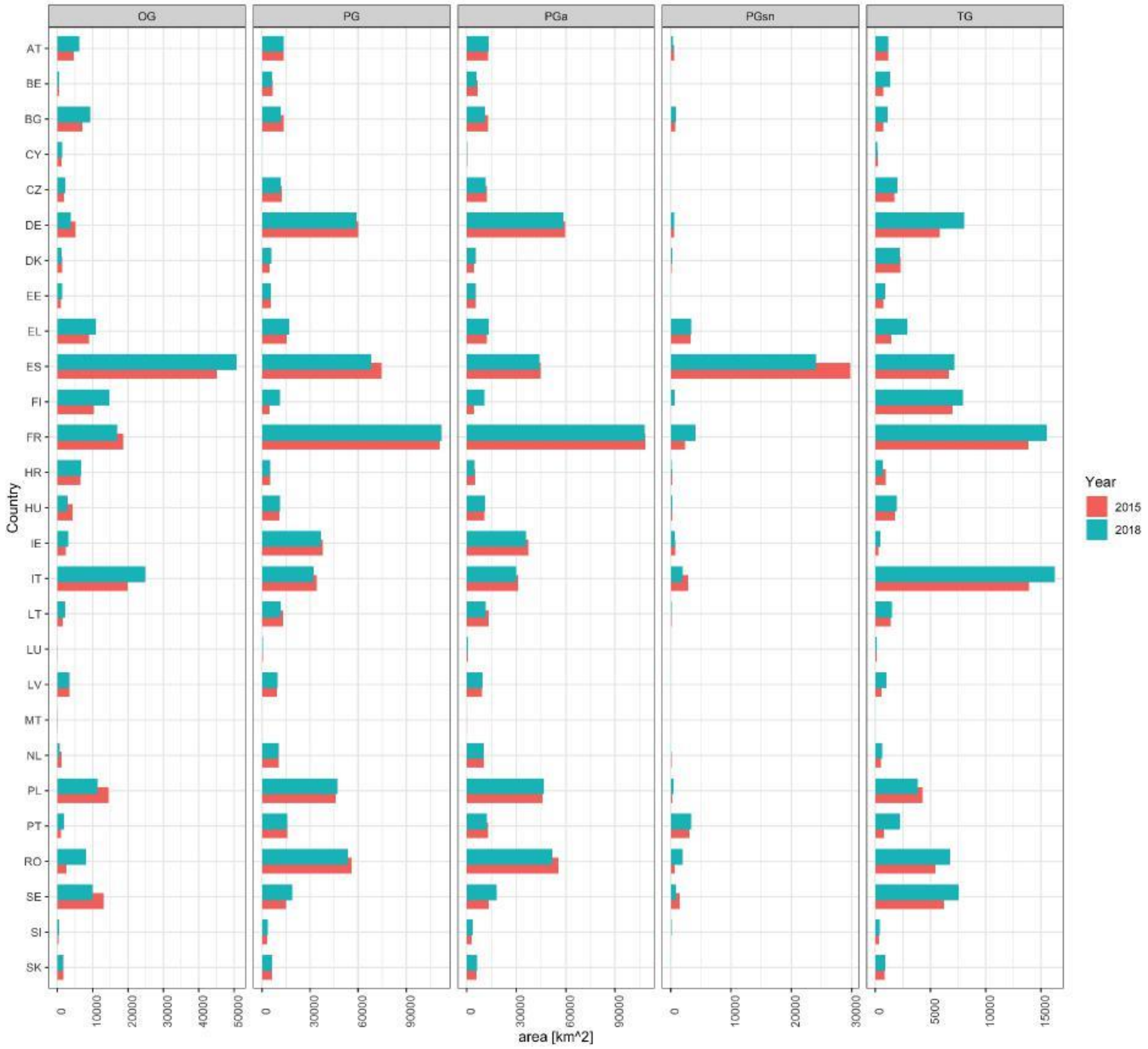


Figure 16 – Model-based estimates of land cover area percentage (related to the total area per country) of the different grassland classes for each country and year.

Table 12 – Model-based area estimates for PGa 2015 and 2018

country	year1	area1	year2	area2
AT	2015	13010.50	2018	13272.20
BE	2015	6648.50	2018	6047.00
BG	2015	12940.00	2018	10905.00
CY	2015	223.90	2018	216.00
CZ	2015	12286.80	2018	11567.60
DE	2015	59422.70	2018	58331.40
DK	2015	4548.40	2018	5516.50
EE	2015	5498.30	2018	5500.20
EL	2015	12031.80	2018	13484.00
ES	2015	44699.20	2018	43978.20
FI	2015	4517.50	2018	10716.60
FR	2015	108332.90	2018	107761.40
HR	2015	4993.10	2018	4755.70
HU	2015	10723.20	2018	11061.30
IE	2015	37082.80	2018	36089.90
IT	2015	31331.30	2018	30170.20
LT	2015	13219.50	2018	11599.00
LU	2015	687.70	2018	754.70
LV	2015	9383.10	2018	9672.40
MT	2015	40.40	2018	24.00
NL	2015	10422.30	2018	10255.00
PL	2015	45731.30	2018	46607.80
PT	2015	12921.00	2018	12273.00
RO	2015	55257.90	2018	51728.80
SE	2015	13518.00	2018	18215.60
SI	2015	3220.10	2018	3682.30
SK	2015	6073.80	2018	6321.30

Table 13 – Model-based area estimates for PGsn 2015 and 2018

country	year1	area1	year2	area2
AT	2015	534.80	2018	400.60
BE	2015	23.50	2018	60.30
BG	2015	751.70	2018	843.10
CY	2015	12.00	2018	24.00
CZ	2015	83.00	2018	91.10
DE	2015	561.00	2018	612.30
DK	2015	193.90	2018	211.50
EE	2015	83.40	2018	66.10
EL	2015	3294.30	2018	3455.00
ES	2015	29736.10	2018	24073.20
FI	2015	87.10	2018	658.90
FR	2015	2399.80	2018	4115.20
HR	2015	227.30	2018	173.10
HU	2015	236.00	2018	286.90
IE	2015	767.30	2018	708.10
IT	2015	2922.90	2018	1977.10
LT	2015	102.30	2018	149.50
LU	2015	11.80	2018	9.30
LV	2015	44.00	2018	67.10
MT	2015	4.00	2018	4.00
NL	2015	113.70	2018	95.70
PL	2015	258.80	2018	487.70
PT	2015	3057.60	2018	3443.80
RO	2015	656.10	2018	1962.50
SE	2015	1491.20	2018	897.60
SI	2015	99.60	2018	135.00
SK	2015	94.40	2018	65.10

Table 14 – Model-based area estimates for PG 2015 and 2018

country	year1	area1	year2	area2
AT	2015	13545.30	2018	13672.80
BE	2015	6672.00	2018	6107.30
BG	2015	13691.80	2018	11748.10
CY	2015	235.90	2018	240.00
CZ	2015	12369.90	2018	11658.80
DE	2015	59983.80	2018	58943.70
DK	2015	4742.30	2018	5728.00
EE	2015	5581.70	2018	5566.30
EL	2015	15326.10	2018	16939.00
ES	2015	74435.40	2018	68051.40
FI	2015	4604.50	2018	11375.40
FR	2015	110732.70	2018	111876.60
HR	2015	5220.40	2018	4928.80
HU	2015	10959.20	2018	11348.30
IE	2015	37850.00	2018	36798.00
IT	2015	34254.20	2018	32147.20
LT	2015	13321.90	2018	11748.50
LU	2015	699.50	2018	764.00
LV	2015	9427.10	2018	9739.50
MT	2015	44.40	2018	28.00
NL	2015	10536.00	2018	10350.70
PL	2015	45990.20	2018	47095.60
PT	2015	15978.60	2018	15716.80
RO	2015	55914.00	2018	53691.30
SE	2015	15009.20	2018	19113.20
SI	2015	3319.60	2018	3817.30
SK	2015	6168.20	2018	6386.40

Table 15 – Model-based area estimates for TG 2015 and 2018

country	year1	area1	year2	area2
AT	2015	1192.90	2018	1181.90
BE	2015	717.00	2018	1336.20
BG	2015	768.00	2018	1101.80
CY	2015	220.60	2018	172.00
CZ	2015	1723.90	2018	2003.10
DE	2015	5846.80	2018	8039.80
DK	2015	2313.10	2018	2230.80
EE	2015	760.20	2018	897.80
EL	2015	1453.10	2018	2904.50
ES	2015	6638.80	2018	7196.80
FI	2015	6993.30	2018	7969.10
FR	2015	13866.20	2018	15545.50
HR	2015	934.40	2018	684.70
HU	2015	1773.30	2018	1920.20
IE	2015	292.90	2018	440.80
IT	2015	13908.00	2018	16253.50
LT	2015	1379.80	2018	1546.10
LU	2015	113.20	2018	123.60
LV	2015	545.00	2018	1026.90
MT	2015	28.40	2018	28.00
NL	2015	542.00	2018	648.30
PL	2015	4296.30	2018	3851.60
PT	2015	802.20	2018	2260.00
RO	2015	5438.10	2018	6747.50
SE	2015	6250.90	2018	7544.40
SI	2015	330.60	2018	407.40
SK	2015	868.80	2018	900.70

Report

Table 16 – Model-based area estimates for OG 2015 and 2018

country	year1	area1	year2	area2
AT	2015	4581.80	2018	6274.20
BE	2015	528.50	2018	546.30
BG	2015	7143.20	2018	9229.30
CY	2015	1283.40	2018	1468.00
CZ	2015	1859.10	2018	2247.30
DE	2015	5109.90	2018	3918.70
DK	2015	1481.70	2018	1267.60
EE	2015	1038.50	2018	1495.30
EL	2015	9048.40	2018	10930.30
ES	2015	45099.40	2018	50749.60
FI	2015	10350.80	2018	14686.80
FR	2015	18673.90	2018	16920.20
HR	2015	6564.90	2018	6739.60
HU	2015	4330.10	2018	2923.40
IE	2015	2418.50	2018	3160.70
IT	2015	19893.90	2018	24890.00
LT	2015	1602.60	2018	2189.10
LU	2015	41.90	2018	8.00
LV	2015	3503.20	2018	3411.30
MT	2015	44.30	2018	40.00
NL	2015	1312.60	2018	705.50
PL	2015	14495.50	2018	11403.90
PT	2015	1155.30	2018	1875.70
RO	2015	2613.70	2018	8210.30
SE	2015	13117.60	2018	9981.10
SI	2015	388.40	2018	452.60
SK	2015	1767.70	2018	1668.30

6 LAND COVER TREND ESTIMATIONS

6.1 Statistical Methods

In the following we want to develop and apply methods in order to estimate for a sub-region (i.e., a specific NUTS-ID) the average percentage change per year for the land cover area of a grassland class – along with certainty estimates (such as variance, confidence intervals or COVs).

Modern and flexible approaches to estimate linear or nonlinear ecological trends are frequently based on regression methods (e.g., [29, 33, 30, 21]) and are capable to integrate various difficulties frequently associated with ecological data, such as spatio-temporal autocorrelation, nonlinear dependencies, or partially panelled data (e.g., [37, 34, 27, 41]). In the following analysis, we have to restrict the trend analysis to the two grassland classes "temporary grassland" (TG) and "permanent grassland with agricultural use" (PGa), since only for these two levels all LUCAS-based variables required for a proper definition are consistently available in the considered data from 2009, 2012, 2015 and 2018. Trend estimates are based on in situ data only, i.e., they do not use predictive modelling (as in area estimates – c.f. Section 2) and thus can only be calculated for subareas (respectively NUTS levels) where in situ data from different surveys exist.

Conceptually, trend as well as area estimation models are strongly related to each other, because the aim of both approaches is an appropriate description of the spatio-temporal distribution pattern of the grassland type variables under consideration. The key difference lies in the focus: in the trend model, the focus is on the relative temporal area development (i.e., the relative inter-annual change) including data from several years, and all other covariates (such as the spatially varying distribution or the dependency on other covariates) are only included in order to account for corresponding bias. Furthermore, panel data points play an outstanding role here, since they provide an important source for trend estimation without additional variance introduced by annually changing sample locations. Nevertheless, non-panel points can be augmented to the data and further increase the statistical power. The trend models are optimized to detect trends with a high statistical power but are not appropriate for the estimation of total area values. Mathematically speaking, here, only the slope of the regression line is estimated (representing an average change per year) and presented uncertainties concern only this slope and not the intercept of the regression line.

The land cover are estimation models (c.f., previous Section) are in contrast specialized to estimate land cover areas, including several corresponding features. Like for example the constraint that the local probability for the different grassland classes sum up to 1, and the fact that they are used for predictive modelling in order to predict land cover also for master grid points where no in situ data are available. Since they are separately applied to the different survey data (2015 vs. 2018), they do not make use of the partial panel structure (which is indeed not required here).

6.2 Trend model approach

For the sake of additional internal validation, we applied two distinctly different trend model approaches to the data:

First, we used a binomial regression (also called logistic regression) model realized in the framework of generalized linear models (GLMs) [14, 17]. In particular, separately for each grassland type, we defined the binomial variable "PGa" (respectively "TG") vs. "non-PGa" (respectively "non-TG"), and used the master-point-ID as a random intercept in the context of generalized linear mixed modelling (GLMMs) [5, 11, 42] in order to appropriately account for the partial panel structure of the data. Additionally, the calibrated regression weights (c.f., Section 2.2.3) have been used as a priori regression weights in order appropriately account for the sampling design, respectively for the fact that points represent different sizes of area, and finale trend estimates should represent changes in land cover area. This approach is however connected with two serious problems:

1. Additive modelling software (and thus an appropriate nonlinear description of the spatial distribution patterns) cannot be applied. The reason is that the required mixed modelling in conjunction with additional iterations related to generalized cross-validation procedures (c.f., [39]) and additional model selection procedures (c.f., below) lead to unfeasible computation times – even on high performance computers. Thus, non-additive models (provided by the glmer() function in the R-package lme4) have to be applied, leading to the problem of spatial autocorrelation. Thus, results can be biased and variance estimates are most probably underestimated [23, 10, 20];

2. The trend estimates of logistic regression models are given in terms of odd ratios [13, 20] instead of percentage change per year. The interpretation of odd ratios is much less intuitive, an approximation of the percentage change can however be obtained using rescaled covariates and calculations based on the estimated baseline probability of the considered grassland class as well as the trend estimation. Due to these two disadvantages, we finally used this logistic modelling approach only to safeguard/compare them to the results of the second approach (presented below). As possible fixed-effect co-variables in the logistic regression model, we used again the variables "X_LAEA" and "Y_LAEA" (as main effects as well as interaction terms) to account for spatial inhomogeneity in the probabilities, as well as "ELEVATION" and, of course, "YEAR".

Second, we used a negative binomial model (also realized in the framework of GLMs [24]) applied to spatially pooled data. In particular, for each sub-region (represented by a NUTS code on the level NUTS0-NUTS2) we first created an artificial spatial grid of N=100 grid cells with outer grid boundaries touching the boundaries of the sub-regions. Secondly, separately for each available year and grassland class ("PGa" and "TG"), we summed up the calibrated weights of the in situ LUCAS points belonging to the considered class and lying within the spatial grid cell (for using it as the outcome variable during regression), and averaged all other co-variables over these points (for using them as co-variables in regression). Further, we summed up the calibrated weights of all in situ points (independent of the grassland class) lying within the cell (for using them as a priori regression weights for the subsequent regression analysis). To account for the partial panel structure of the data, we actually applied the procedure as described above separately for different "temporal abundances" of LUCAS points belonging to the same grid cell: each "temporal abundance class" was defined by a unique combination of years for which points within a grid cell were available. Finally, each unique combination of grid cell ID and temporal abundance class have been labelled with a unique ID in order to use it as a random factor in the context of generalized linear mixed modelling (GLMMs) [5, 11, 42]. As co-variables, also here we used the variables "X_LAEA" and "Y_LAEA" to account for spatial inhomogeneity in the probabilities, as well as "ELEVATION" and, of course, "YEAR". Since pooled data are of smaller total size, we also tested a 2D smooth term of the variables "X_LAEA" and "Y_LAEA" realized in the framework of generalized additive modelling /GAMs [17, 18, 39]. Thus, in summary, a generalized additive mixed model (GAMM) with a negative binomial probability distribution has been applied to the spatially pooled data. This was achieved by using the sum of the calibrated weights per grid cell belonging to a certain grassland class (i.e., representing the land cover area of this class in the grid cell) as the outcome variable, and the total area of the grid cell as a priori weights. In particular, the partial panel structure has been appropriately integrated via mixed modelling.

For both types of models, we again selected (separately for each grassland class and NUTS-ID) the model with the lowest AIC ("Akaike information criterion") value [1] which is a frequently used approach during model selection [13, 42, 20].

6.3 Software

For all statistical calculations and visualizations, the statistical open-source software R [32] has been used. For regression analysis, we used the R-package mgcv [39] as well as lme4 [3], and for visualizations the package ggplot2 [38].

6.4 Possible Problems, validation and comparison

6.4.1 Correlations between different trend measures

In order to validate the trend estimates, we compared the estimates of the statistically most proper method (the above-described negative binomial GAMM) with several other measures, on the one hand given by alternative (less appropriate) methods applied to the same underlying data, on the other hand from external sources. In particular, we compared the preferred negative binomial GAMM with seven other trend estimates (c.f., Fig. 17), where 5 out of these 7 methods rely (at least partially) on the same data, and 2 (out of 7) are based on external sources. The main motivation of this intense comparison was to validate the robustness of the estimated trends. If our GAMM-based estimations are reasonable, they should positively correlate with most of the alternative (possibly less appropriate) methods. The alternative estimates are briefly described in the following.

As explained above, we used two different model types to estimate trends: a logistic regression model as well as a negative-binomial model. From the theoretical point of view (also explained in more detail above), the negative binomial model is more reliable, since it considers possible spatial autocorrelation (via the spatial smooth predictor) and leads additionally to straight-forward percentage trend estimates, whereas the logistic model allows only for approximations of the percentage change per year.

As a third way to estimate trends, one can extract the sampling-based area estimates (c.f., chapter 5) from each survey year by summing up the calibrated weights for the sub-region of interest. In this case trends are calculated directly as the average percentage change of these point estimates between the surveys (i.e., first the percentage change between each two subsequent surveys has been calculated, and secondly the mean from all available values). However, this approach totally neglects any panel structure in the data and is therefore most probably much less reliable.

A fourth way to estimate trends is to apply a negative binomial model (as described above), but ignoring the partial panel point structure. These models are much less complex (since GAMs instead of GAMMs are applied); however, due to the neglect of the partial panel structure, results are assumed to be biased.

A fifth way is given by a naive approach, where only panel points between each two subsequent surveys are extracted, and the percentage change in the number of points belonging to the grassland class of interest is calculated. Eventually, the mean over the resulting 3 values is calculated. We want to point out that in this approach; the total area associated with each point (strongly differing due to the stratified sampling design) is ignored, as well as the information from non-panel points.

A sixth way is given by calculating the percentage change from 2015 to 2018 based on the model-based area estimates (c.f., Section 2). However, this approach ignores on the one hand the other hand only the most recent change is reflected. As external sources, we used crop statistics from Eurostat respectively Faostat.

The correlation plot (using Kendall's tau, since data are not-normally distributed) are presented in Fig. 17. It appears that there is an obvious positive (and often significant) correlation between all six different trend estimation methods applied to the LUCAS data (green boxes). In particular, the first two approaches (the negative binomial model as well as the logistic regression model) show significant correlations to all other methods, being in line with the assumption that these two methods provide the most reliable estimates. Only the correlation with the "Two-years-methods" is weak, which is not surprising, since these estimates reflect only changes from 2015 to 2018.

Surprisingly, there is no positive or significant correlation of our trend estimates with the data from official crops statistics, although the spatial correlation of our model-based land cover area estimates are highly correlated with these data (c.f., Fig. 11). Reasons for this mismatch can be versatile:

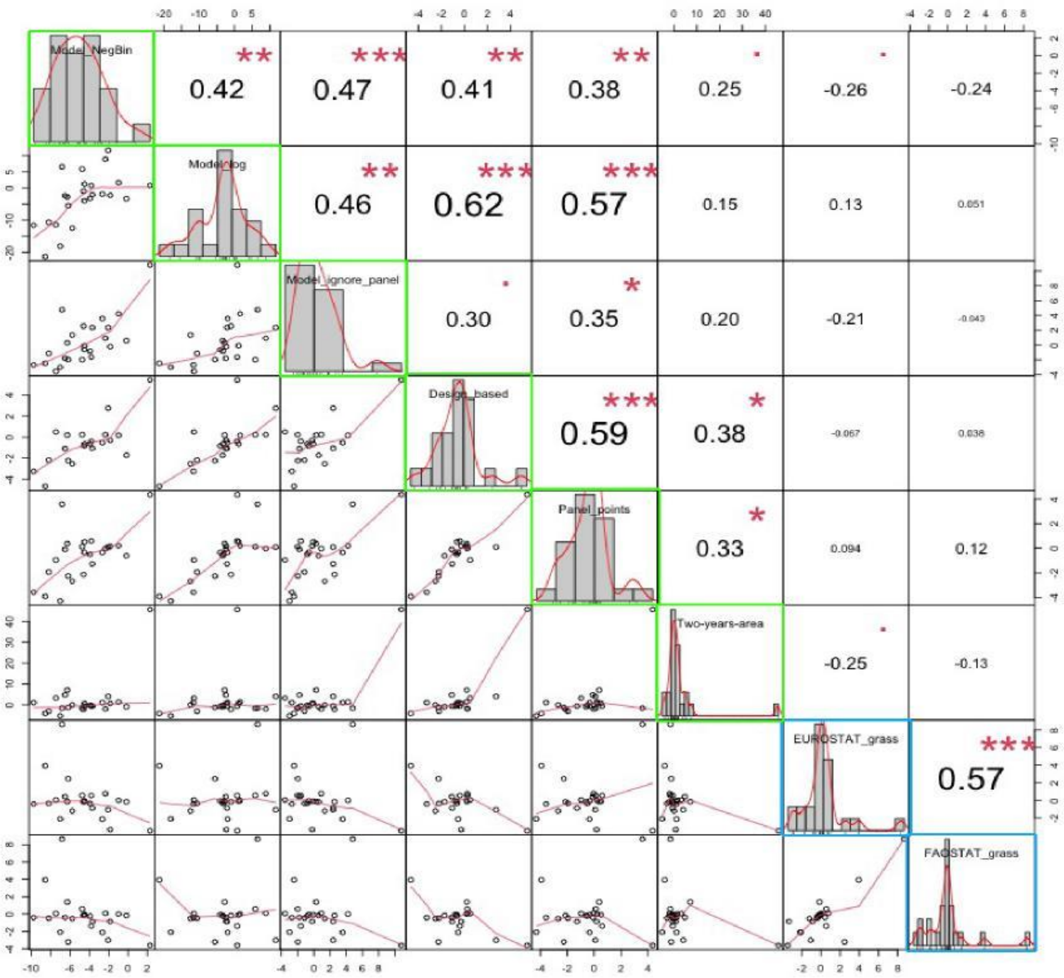
- First and foremost, the different production methods between LUCAS surveys and Eurostat/Faostat crop statistics;

- Already from the Eurostat/Faostat raw data, it was obvious that these data comprise some problems, since in several cases values were identical in subsequent years (which is possibly related to the imputation of missing data but, however, may bias trend analyses);
- The latter is also true for some of the LUCAS data (c.f., discussion with respect to the inclusion of photo-interpreted points from 2009 and 2012);
- We consider PGa which most probably differs from the definition of "grassland" in crops statistics (e.g., the matter may comprise PGsn, TG or OG as well).

We want to point out that we cannot totally exclude that trends extracted from our methods are biased – particularly in the view that the LUCAS sampling design strongly changed between 2009-2015 vs. 2018. However, we did not find any indication of such bias, and we think that the sampling design problem has been appropriately considered. Nevertheless, from our various correlation analyses with respect to different methods for trend estimation and different underlying data, we observe that trend estimates are in general not that robust as land cover area estimates are. For example, Kendall's correlation coefficient is $\tau = 0.99$ for Eurostat Crop Statistics vs. Faostat area estimation, but for trends (extracted from the same data) it holds that $\tau = 0.57$ only. In principle, the same is observed for the correlations between the area vs. trend estimates in our analysed LUCAS-based data. Thus, trend estimates appear to be much less robust than land cover area estimates – independent of the data source.

Furthermore, we want to emphasize that PGa trend estimates based on our sophisticated models (the Negative-binomial model and the logistic regression model – both integrating panel- as well as non-panel points) lead in average to stronger negative values for trends – compared to the various alternative (but statistically less appropriate) approaches. We do not have a straightforward explanation for that but do not assume a methodological error, since (1) both sophisticated models show approximate the same average trend value, and (2) this difference between "sophisticated" and "non-sophisticated" models is not observed for TG estimates, where correlations between all trend measures are much higher. Thus, particularly since for PGa and TG the same approach has been applied, it seems not to be a general bias of our approach. This difference between PGa and TG may indicate that only the sophisticated models appropriately capture the above-mentioned sudden change in the sampling design, and that this mainly concerns areas with PGa rather than areas with a high probability of TG.

Figure 17 – Pairwise correlations (based on Kendal's tau) of PGa land cover area trend estimates based on different methods applied to in situ LUCAS data (green boxes) and grassland trend estimates from external data based on Eurostat crop statistics (blue boxes). On the diagonal, there are variable names and histrogram plots of the raw data, below the diagonal, paired scatter plots, and above the diagonal, the P-values for the paired correlations. In particular, the first variable (Model_Negbin) is the statistically most valid GAMM integrating panel and non-panel points, the second variable (Model_Log) is the less appropriate logistic regression GLMM, the third variable (Model_ignore_panel) is a negative binomial GAM ignoring the partial panel structure, the fourth variable (Design_based) are trend estimates extracted from design-based area estimates and also ignoring the panel structure, the fifth variable (Panel_points) is based on the percentage change of panel points (only) belonging to PGa vs. non-PGa (and thus ignores the area associated with each point), and the sixth variable (Two_years_area) is the trend between 2015 and 2018 based on model-based land cover area estimates, which also ignores the partial panel structure. The last two variables (EUROSTST_grass and FAOSTAT_grass) are trends extracted from official Eurostat sources for grassland area estimates. All evaluations are performed on the NUTS0 level.



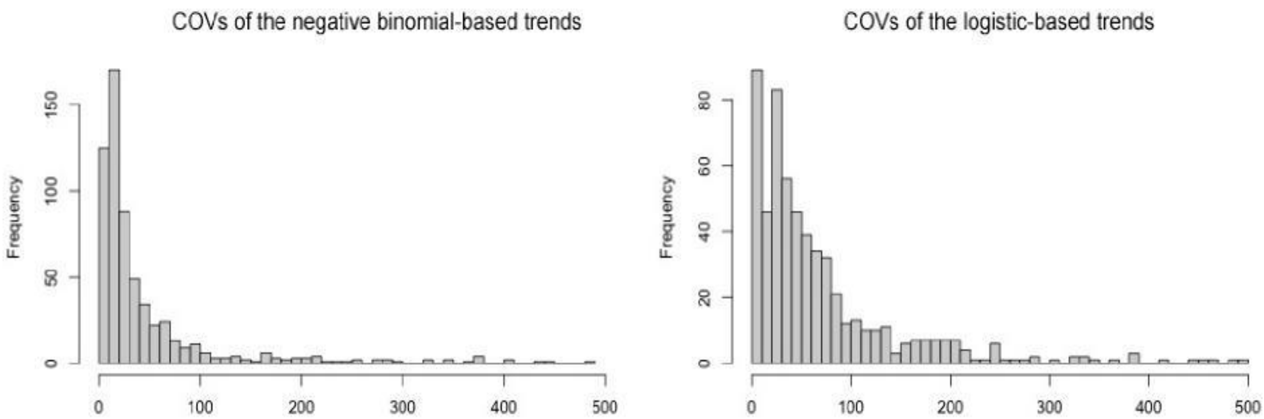
Report

For further validation of the estimates, we correlated the trend estimates for all considered NUTS0-3 levels (separately for TG- vs. PGa-trends) pairwise between the negative-binomial, the logistic regression and the design-based estimates – the three approaches assumed to lead to the most reliable estimates. It appears that indeed there is a distinct correlation in all pairwise correlations with an r ranging between 0.5 and 0.7 and *P* being always highly significant. It is however not surprising that the correlation is not even higher, since the trend estimates rely on three distinct different methods, where the logistic approach and even more the naive approach may violate underlying statistical assumptions and thus can be biased. Furthermore, NUTS3 levels with only relatively sparse data dominate the analysis, i.e., trends are estimated including a high amount of uncertainty/stochasticity.

Here, at least with respect to PGa, the negative binomial model (which is from theoretical side the model that performs best) in comparison to the design-based estimate performed much better (r = 0.65) than the logistic regression model compared to the design-based trend estimate (r = 0.49). With respect to TG, the performance of both models was comparable. Furthermore, the COVs related to the trend estimates based on the negative binomial model were in average much smaller compared to those of the logistic approach (Fig. 18). Together with the fact that the logistic approach is possibly biased (due to the neglect of autocorrelation as well as only an approximate estimate of percentage trend values) in contrast to the negative binomial approach, it is obvious that the negative binomial approach is much more reliable and will be used in the following.

A validation with respect to the estimated COVs is given when re-calculating the above-described correlation between the negative binomial-based estimates and the design-based estimates – but this time weighting the correlation using the inverse COV values as a priori regression weights. Here, the correlation strength distinctly increases from r = 0.65 to r = 0.8.

Figure 18 – Distribution of trend-related COVs for the two different trend regression model approaches

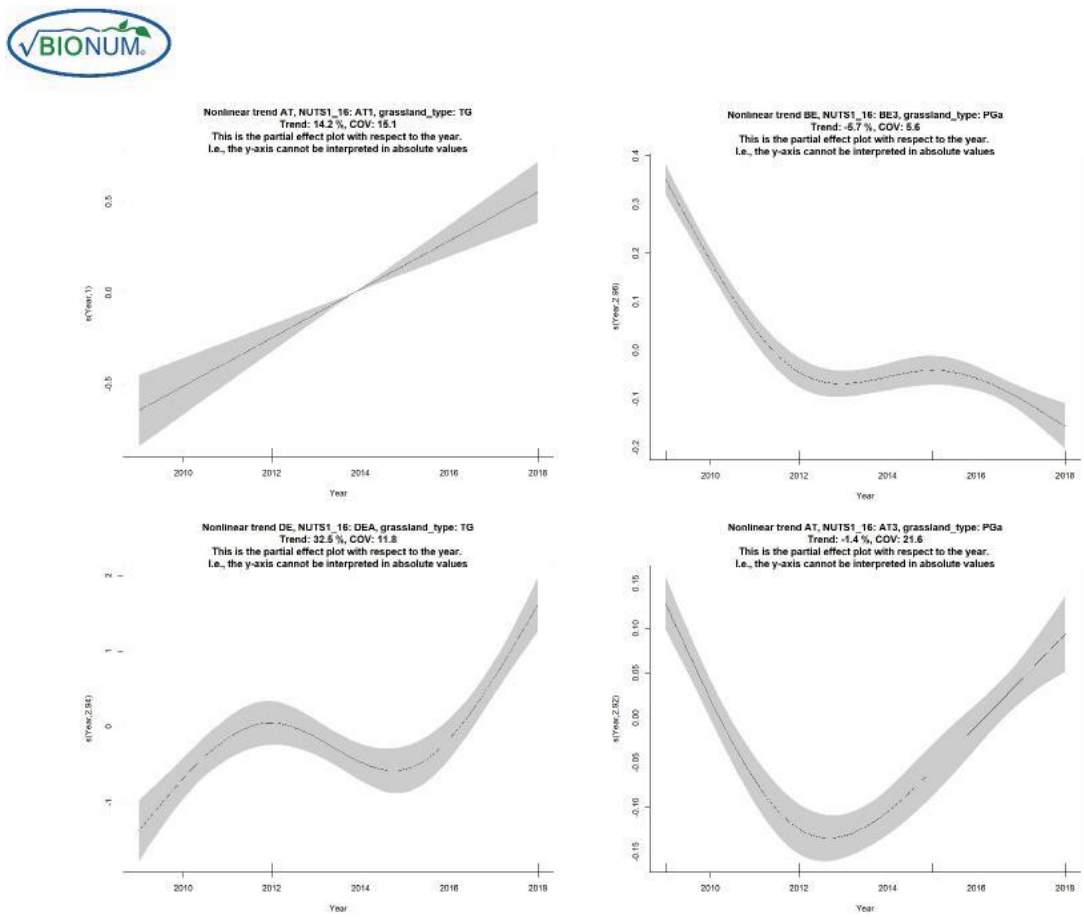


Report

6.4.2 Visual Validation

Based on the final selected GAMM (obtain during the AIC-based model selection procedure as described above – separately performed for each NUTS-ID) we also fitted a model where the variable ”YEAR” has not been used as a linear predictor but a smooth (i.e., possibly nonlinear) predictor instead (using regression splines in the context of generalized additive modelling [18, 39]). Using this method in combination with generalized cross-validation, we can investigate if the trend is linear or rather nonlinear. For each grassland class and NUTS-ID separately, we generated a corresponding model plot including an estimate of the confidence bands (deliverable D.2.1.4). These figures can be used on the one hand in order to validate the COV estimate via the width of estimated confidence bands, on the other hand to detect strongly nonlinear trend behaviour (some examples are given in Fig. 19).

Figure 19 –Some examples of nonlinear trends used to obtain more detailed information on the trend behaviour in time than linear models can provide. Shaded regions depict 95%-confidence bands.

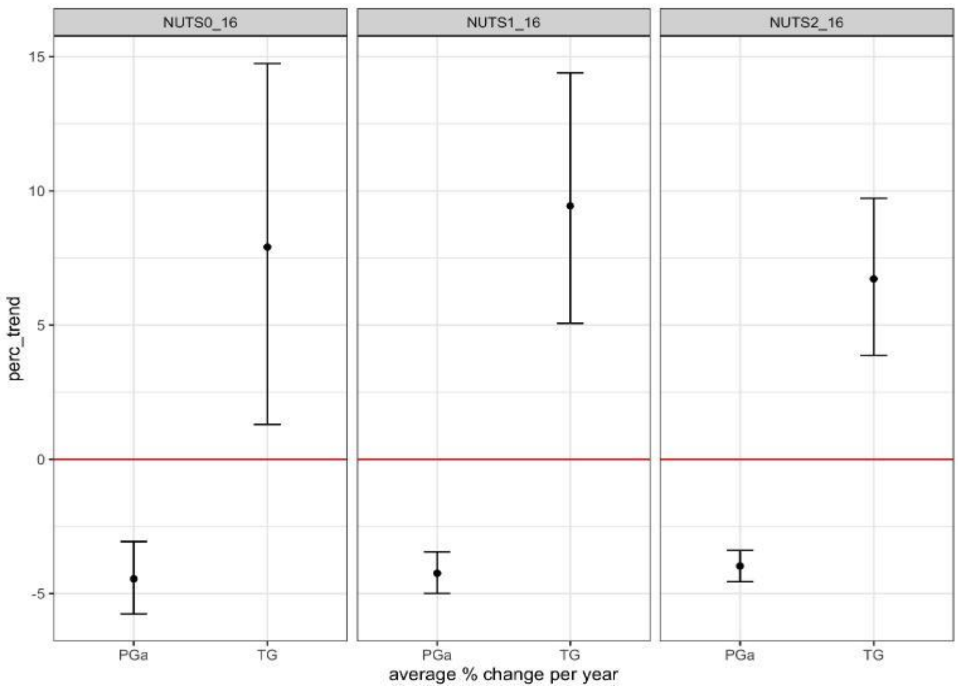


6.4.3 The problem of data augmentation by photo-interpreted data

Relatively late in the context of Task II it appeared that the surveys 2009 and 2012 have been artificially enriched by photo-interpreted data from 2015. In particular, in 2009, survey data were added about 27.000 and in 2012 about 64.000 points. Thus, from theoretical side, there was the strong initial concern that including these points into the trend analysis will introduce a bias by weaken the trend estimates, since artificial temporally constant points have been added.

Fruitful discussions with other LUCAS statisticians/experts revealed the following: photos of the 2015 PI points were effectively taken 2/3 years before and so they are contemporary to the 2012 ”in field” points while, from a logical point of view, they represent the ”future status” with respect to 2009 and the ”old status” with respect to 2015. As these points were selected in areas with a low probability of change (e.g., remote areas), one can thus assume that the introduced trend-bias is only minor. Moreover, because the eligibility criteria changed in 2015, the aim of adding the PI points to 2009 and 2012 was to increase the comparability between 2009, 2012 and subsequent surveys. Furthermore, inclusion- and calibrated weights (indispensable for an unbiased trend regression analysis) were only available for the data sets including these points.

Figure 20 – Model-based average trend estimates (average yearly percentage change – based on a negative binomial GAMM and across all NUTS-IDs of the considered NUTS level) separately shown for different grassland classes and NUTS levels. Black bars indicate 95%-confidence intervals calculated based on bootstrapping.



Taken together, and as advised by the other LUCAS statisticians, we proceeded the trend estimates including the additional PI points.

6.5 Results

The results of the trend estimates (based on the negative binomial GAMM) are shown in Fig. 20-21. Particularly Fig. 20 demonstrates that trend estimates are in average quite robust: independent of the NUTS level (NUTS0-2), average mean values of trends are very similar. In particular, there is in average a slight yearly negative trend for the PGa grassland class (approx. -4 % decay per year) and a distinct positive trend for the TG grassland class (approx. +7.5 % increase per year).

A map of the estimated trends (based on NUTS0-NUTS1 level evaluation) is given in Fig. 21. Due to sparse data, however, some of the trends (particularly for TG in small NUTS1 areas) have not been estimated.

Numerical data on TG and PGa trend estimates on NUTS0 level are given in Tab. 17-18, and the trends for all country and NUTS level combinations (along with different certainty estimates) are given in deliverable D.2.1.5.

Table 17 – Estimated PGa trends on NUTS0-level including the coefficient of variation (COV), as well as the limits of the 95%-confidence intervals

country	trend [% change/year]	COV [%]	95%-CI_down	95%-CI_up
AT	-1.80	13.00	-2.30	-1.40
BE	-8.20	3.40	-8.70	-7.70
BG	-7.00	6.60	-7.90	-6.10
CY	-11.30	27.40	-16.80	-5.40
CZ	-2.30	11.60	-2.90	-1.80
DE	-4.50	1.70	-4.60	-4.30
DK	-6.30	6.50	-7.10	-5.50
EE	-5.80	7.90	-6.60	-4.90
EL	-2.00	15.30	-2.60	-1.40
ES	-3.80	5.70	-4.20	-3.40
FI	2.30	28.10	1.00	3.60
FR	-3.90	3.30	-4.20	-3.70
HR	-6.80	19.90	-9.30	-4.20
HU	-2.60	14.40	-3.40	-1.90
IE	-0.20	38.10	-0.30	-0.00
IT	-6.20	3.10	-6.50	-5.80
LT	-8.50	2.70	-9.00	-8.10
LU	-7.40	12.10	-9.10	-5.70
LV	-9.70	2.90	-10.30	-9.20
MT	7.10	76.00	-3.30	18.60
NL	-3.70	6.70	-4.20	-3.20
PL	-4.50	2.70	-4.70	-4.20
PT	-4.70	5.90	-5.30	-4.20
RO	-4.60	6.70	-5.20	-4.00
SE	-6.20	5.70	-6.90	-5.50
SI	-6.50	4.10	-7.00	-6.00
SK	-1.00	40.60	-1.70	-0.20

Report

Figure 21 – Model-based trend estimates (average yearly percentage change – based on a negative binomial GAMM) on NUTS0 (wbove) and NUTS1 (below) level for PGa (left-hand side) and TG (right-hand side) grassland class. Due to sparse data, some of the trends (particularly for TG) have not been estimated.

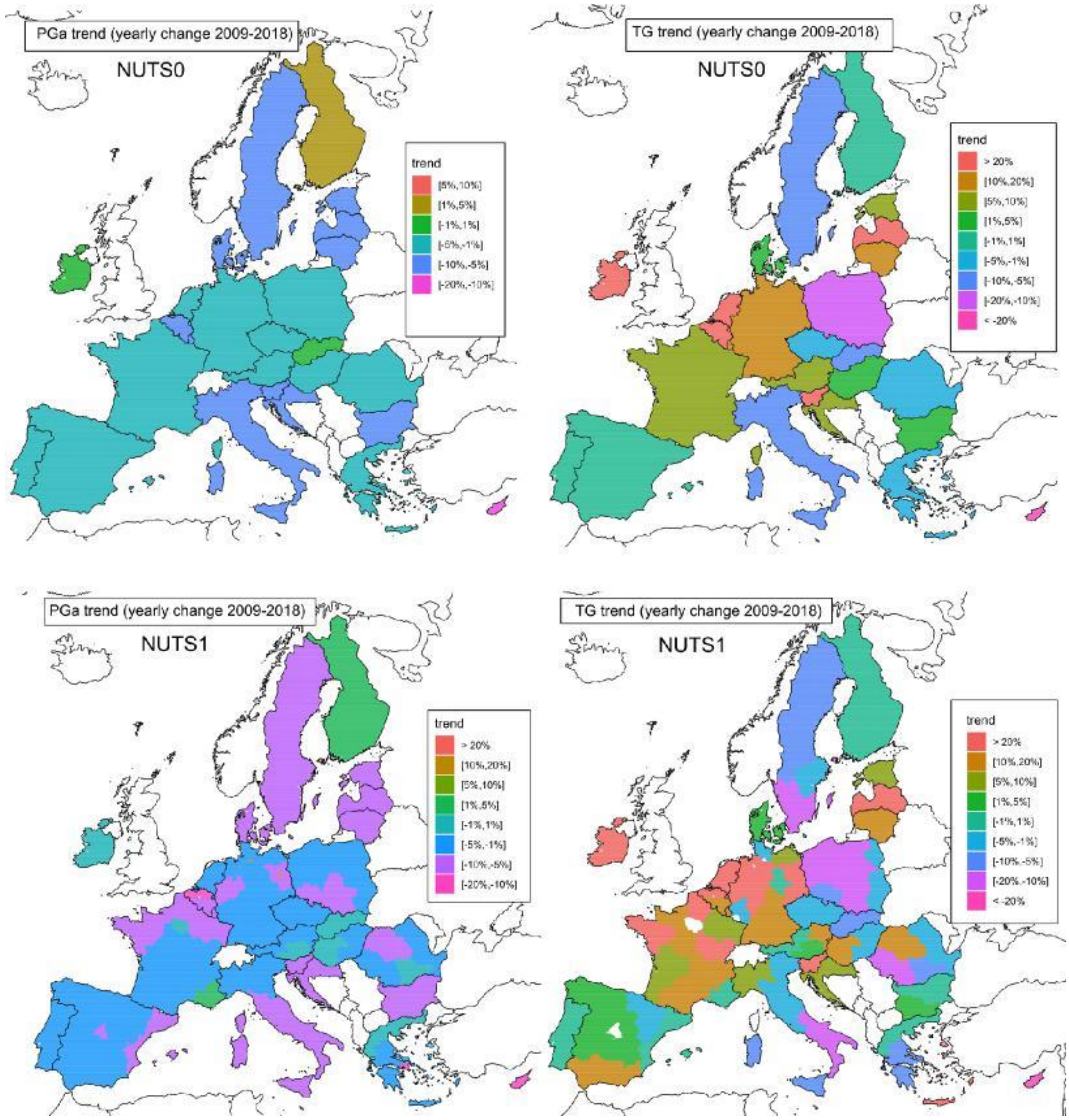


Table 18 – Estimated TG trends (% change per year) on NUTS0-level including the coefficient of variation (COV), as well as the limits of the 95%-confidence intervals.

country	trend [% change/year]	COV [%]	95%-CI_down	95%-CI_up
AT	6.50	22.90	3.60	9.60
BE	30.30	8.20	24.90	35.90
BG	4.10	73.90	-1.80	10.30
CY	-21.50	17.70	-27.80	-14.60
CZ	-4.60	28.90	-7.10	-2.00
DE	15.50	6.30	13.50	17.60
DK	1.40	96.30	-1.20	4.10
EE	8.80	29.90	3.50	14.30
EL	-4.30	13.80	-5.50	-3.20
ES	0.80	91.70	-0.60	2.20
FI	0.70	87.60	-0.50	1.80
FR	5.40	10.00	4.30	6.50
HR	9.60	46.80	0.80	19.20
HU	3.10	40.20	0.70	5.60
IE	41.50	9.90	32.30	51.40
IT	-5.90	6.30	-6.60	-5.20
LT	12.40	15.30	8.60	16.40
LU	18.70	64.50	-4.40	47.40
LV	21.40	12.50	15.80	27.30
MT	0.90	512.80	-8.10	10.90
NL	65.70	5.60	56.70	75.20
PL	-10.70	7.60	-12.20	-9.20
PT	-0.90	167.90	-4.00	2.20
RO	-4.40	25.30	-6.50	-2.20
SE	-9.90	6.10	-11.00	-8.80
SI	35.00	11.40	26.20	44.30
SK	-6.00	33.60	-9.80	-2.10

7 References

- Beaufoy, G., Blom, S., Hartel, T., Jones, G., Popa, R., Poux, X., & Ruiz, J. (2015, November). Europe's wood pastures: condemned to a slow death by the CAP. In *Booklet produced for the woodpasture policy seminar in the European Parliament, Brussels, Belgium* (Vol. 17).
- d'Andrimont, R., Yordanov, M., Martinez-Sanchez, L., Eiselt, B., Palmieri, A., Dominici, P., ... & van der Velde, M. (2020). Harmonised LUCAS in-situ data and photos on land cover and use from 5 tri-annual surveys in the European Union. arXiv preprint arXiv:2005.05272.
- EEA 2017 Copernicus Land Monitoring Service – High Resolution Layer Grassland: Product Specifications Document. 2017-03-15
- Copernicus Land Monitoring Service High Resolution land cover characteristics Grassland 2018 and Grassland change 2015-2018 https://land.copernicus.eu/pan-european/high-resolution-layers/grassland. (last access: 1 January 2021), 2021
- Beaufoy, G., Blom, S., Hartel, T., Jones, G., Popa, R., Poux, X., & Ruiz, J. (2015, November). Europe's wood pastures: condemned to a slow death by the CAP. In *Booklet produced for the woodpasture policy seminar in the European Parliament, Brussels, Belgium* (Vol. 17).
- Eurostat: LUCAS Grid Record Descriptor, https://ec.europa.eu/eurostat/documents/205002/7329820/LUCAS-Grid-Record-Descriptor.pdf (last access: 22 May 2019), 2018.
- Eurostat: Technical reference document S1: Stratification Guidelines, https://ec.europa.eu/eurostat/documents/205002/7329820/LUCAS2018_S1-StratificationGuidelines_20160523.pdf (last access: 22 May 2019),
- Eurostat: LUCAS web site, https://ec.europa.eu/eurostat/web/lucas (last access: 31 March 2021), 2021.
- Eurostat: Technical reference document C-1: Instructions for surveyors, https://ec.europa.eu/eurostat/documents/205002/8072634/LUCAS2018-C1-Instructions.pdf (last access: 31 March 2021), 2018.
- Eurostat: Technical reference document C-3: Classification, https://ec.europa.eu/eurostat/documents/205002/8072634/LUCAS2018-C3-Classification.pdf (last access: 31 March 2021), 2018.
- Eurostat: LUCAS Grid – Eurostat, https://ec.europa.eu/eurostat/web/lucas/data/lucas-grid (last access: 31 March 2021), 2021.
- Eurostat: LUCAS SURVEY2018 WEB CSV Record Descriptor, https://ec.europa.eu/eurostat/documents/205002/8072634/LUCAS2018-RecordDescriptor-190611.pdf (last access: 31 March 2021), 2021.
- H. Akaike. Information theory and an extension of the maximum likelihood principle. *International Sympossium on Information Theroy*, Second Edition:267–281, 1973.
- David R Anderson. *Model based inference in the life sciences: a primer on evidence*. Springer Science & Business Media, 2007.
- Douglas Bates, Martin Mächler, Ben Bolker, and Steve Walker. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1):1–48, 2015.

- Edwige Bellier, Pascal Monestiez, Grégoire Certain, Joel Chadœuf, and Vincent Bretagnolle. Reducing the uncertainty of wildlife population abundance: model-based versus design-based estimates. *Environmetrics*, 24(7):476–488, 2013.
- Benjamin M. Bolker, Mollie E. Brooks, Connie J. Clark, Shane W. Geange, John R. Poulsen, M Henry H. Stevens, and Jada-Simone S. White. Generalized linear mixed models: a practical guide for ecology and evolution. *Trends Ecol Evol*, 24(3):127–135, Mar 2009.
- George EP Box, Gwilym M Jenkins, Gregory C Reinsel, and Greta M Ljung. *Time series analysis: forecasting and control*. John Wiley & Sons, 2015.
- KRW Brewer. Design-based or prediction-based inference? stratified random vs stratified balanced sampling. *International Statistical Review*, 67(1):35–47, 1999.
- Dick J Brus and Jaap J DeGruijter. Design-based versus model-based estimates of spatial means: Theory and application in environmental soil science. *Environmetrics*, 4(2):123–152, 1993.
- DJ Brus and JJ De Gruijter. Random sampling or geostatistical modelling? choosing between design-based and model-based sampling strategies for soil (with discussion). *Geoderma*, 80(1-2):1–44, 1997.
- Carsten F. Dormann, Jana M. McPherson, Miguel B. Araújo, Roger Bivand, Janine Bolliger, Gudrun Carl, Richard G. Davies, Alexandre Hirzel, Walter Jetz, W Daniel Kissling, et al. Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. *Ecography*, 30(5):609–628, 2007.
- J.J. Faraway. *Extending the linear model with R, generalized linear, mixed effects and non-parametric regression models*. Chapman & Hall/CRC, Boca Raton, 2006.
- Lorenzo Fattorini, Piermaria Corona, Gherardo Chirici, and Maria Chiara Pagliarella. Design-based strategies for sampling spatial units from regular grids with applications to forest surveys, land use, and land cover estimation. *Environmetrics*, 26(3):216–228, 2015.
- A. Field, J. Miles, and Z. Field. *Discovering statistics using R*. SAGE Publications Ltd, 2012.
- J. Fox. *Applied Regression Analysis and Generalized Linear Models*. Sage Publications, CA., 2008.
- F.J. Gallego. Remote sensing and land cover area estimation. *International Journal of Remote Sensing*, 25(15):3019–3047, 2004.
- Timothy G Gregoire. Design-based and model-based inference in survey sampling: appreciating the difference. *Canadian Journal of Forest Research*, 28(10):1429–1447, 1998.
- A. Guisan, T.C. Edwards, and T Hastie. Generalized linear and generalized additive models in studies of species distributions: setting the scene. *Ecological Modelling*, 157:89–100, 2002.
- T. Hastie and R.J. Tibshirani. *Generalized Additive Models*. London, UK: Chapman and Hall, 1990.
- Hannes Kazianka and Jürgen Pilz. Model-based geostatistics. In *International Encyclopedia of Statistical Science*, pages 833–836. Springer, 2011.
- F. Korner-Nievergelt, T. Roth, S. von Felten, J. Guelat, B. Almasi, and P. Korner-Nievergelt. *Bayesian Data Analysis in Ecology Using Linear Models with R, BUGS, and Stan*. Elsevier, London, 2015.
- I. K.Petersen, R. D. Nielsen, and M. L. Mackenzie. Post-construction evaluation of bird abundances and distributions in the horns rev 2 offshore wind farm area, 2011 and 2012. *Aarhus University, Department of Bioscience: Report commissioned by DONG Energy*, 2014.

- G Laslett, G Heuvelink, N Cressie, N Urquhart, R Webster, and A McBratney. Random sampling or geostatistical modelling? choosing between design-based and model-based sampling strategies for soil: Discussion. 1997.

- P Legendre. Spatial autocorrelation: trouble or new paradigm? *Ecology*, 74:1659–1673, 1993.

- A. Linden and S. Maentyniemi. Using the negative binomial distribution to model overdispersion in ecological count data. *Ecology*, 92(7):1414–1421, Jul 2011.

- Roderick J Little. To model or not to model? competing modes of inference for finite population sampling. *Journal of the American Statistical Association*, 99(466):546–556, 2004.

- Ronald E. Mcroberts. A model-based approach to estimating forest area. *Remote Sensing of Environment*, 103(1):56–66, 2006.

- M. Mercker, N. Markones, J. Schwemmer, K. Borkenhagen, and S. Garthe. An integrated framework to estimate seabird population numbers and trends. *Journal of Wildlife Management*, page https://doi.org/10.1002/jwmg.22026, 2020.

- Douglas C Montgomery, Lynwood A Johnson, and John S Gardiner. *Forecasting and time series analysis*. McGraw-Hill Companies, 1990.

- Michelle Paleczny, Edd Hammill, Vasiliki Karpouzi, and Daniel Pauly. Population trend of the world's monitored seabirds, 1950-2010. *PLoS One*, 10(6):e0129342, 2015.

- S.F. Pearson, M.M. Lance, and M.G. Raphael. Washington 2013 at-sea marbled murrelet population monitoring: Research progress report. *Washington Department of Fish and Wildlife, Wildlife Science Division and USDA Forest Service Pacific Northwest Research Station, Olympia, WA*, 2014.

- J.C Pinheiro and D.M. Bates. *Mixed-effect models in S and S-Plus*. New York: Springer Verlag, 2000.

- R Core Team. R: A language and environment for statistical computing. *R Foundation for Statistical Computing, Vienna, Austria.*, ISBN 3-900051-07-0, URL http://www.R-project.org/, 2016.
- R.A. Robinson, J.H. Marchant, D.I. Leech, D. Massimino, M.J.P. Sullivan, S.M. Eglington, C. Barimore, D. Dadam, I.S. Downie, M.J. Hammond, S.J. Harris, D.G. Noble, R.H. Walker, and S.R. Baillie. Birdtrends 2015: trends in numbers, breeding success and survival for uk breeding birds. *Research Report 678. BTO, Thetford. www.bto.org/birdtrends*, 2015.

- C.J. Schwarz. Detecting trends over time. *In "Course Notes for Beginning and Intermediate Statistics"*, 2015.

- J. Starkweather and A. K. Moske. Multinomial logistic regression. *Retrieved from: http : / /www.unt.edu/rss / class/Jon/ Benchmarks/MLRJDSAug2011.pdf*, 2011.

- Stephen V. Stehman. Model-assisted estimation as a unifying framework for estimating the area of land cover and land-cover change from remote sensing. *Remote Sensing of Environment*, 113(11):2455–2462, 2009.

- N.S. Urquhart and T.M. Kincaid. Designs for detecting trend from repeated surveys of ecolgical resources. *Journal of Agricultural, Biological, and Environmental Statistics*, 4(4):404–414, 1999.

- H Wickham. *ggplot2: Elegant Graphics for Data Analysis*. Springer Dordrecht Hiedelberg London New York, 2009.

- S Wood. *Generalized Additive Models: An Introduction with R*. Chapman & Hall/CRC., 2017.
- A. Zuur, E. Ieno, and G.M. Smith. *Analysing Ecological Data*. Springer Science+Business Media, LLC, 2007.
- A. F. Zuur, E. N. Ieno, and A. A. Saveliev. *Spatial, Temporal and Spatial-Temporal Ecological Data Analysis with R-INLA*. Highland Statistics Ltd, 2017.
- A.F. Zuur, E.N. Ieno, N.J. Walker, A.A. Saveliev, and G.M. Smith. *Mixed Effect Models and Extensions in Ecology with R*. Springer Science+Business Media, LLC, New York, 2009.

8 Annexes

8.1 Annex 1 – Testing of a set rules relevant for grassland classes.

Before the definition of the final rules for assigning LUCAS survey points to the grassland classes, several tests were made by considering different combinations of LUCAS variables. Such variables belong to the LUCAS Core where the main LC/LU and associated variables are collected, or to LUCAS Modules[6], devoted to additional variables (e.g., INSPIRE Pure Land Cover Component - PLCC). Different combinations of LUCAS variables (mainly LC1, LC2, LU1 and LU2) were tested, to explore the number of points eligible for the classes as well as the LC and LU characteristics. Combinations considered "unlikely" or "impossible" according to the LUCAS Survey were specifically checked in terms of compliance with the LUCAS classification system (i.e., allowed combination of LCxLU, LC1xLC2 and LU1xLU2).

Hereafter some of the most relevant tests are reported. The tests are based on the application of several rules on the LUCAS survey 2018 data followed by data exploration and analysis that allowed keeping or removing specific variables not relevant for the grassland classes from the final set of rules. For each data exploration, a visual analysis on a sample of LUCAS survey photos was also carried out to clarify the relevance of specific variable combination for the grassland classes.

8.1.1 Relevance of F40 (Other bare soil) for the grassland classes

By definition, the class "Other bare soil" (F40) has no dominant land cover on at least 90% of the area and can have a link with agricultural use (U111, U112). An analysis was performed to check the correctness of the herbaceous coverage for the LUCAS Survey points 2018 with the variable INSPIRE_PLCC4 that assess for each LUCAS point the percentage of herbaceous plants with a percentage value between 0 and 100. The value of the variable resulted always <=10% assuring the full coherence with the F40 class definition.

To exclude any eligibility for the grassland classes and explore the agricultural LU of the points classified as F40, the 2018 points were extracted with following rule by retrieving the variables INSPIRE_PLCC4, PARCEL_AREA_HA, CROP_RESIDUES[7], GRAZING[8]:

LC1 = F40 AND (LU1 = U111 OR LU2 = U112).

[6] Specific LUCAS protocols were carried out on demand such as the transect of 250 m to assess transitions of land cover and existing linear features (2009, 2012, 2015), the topsoil module (2009, 2012 (partly), 2015, and 2018), the grassland module (2018), and the Copernicus module collecting the homogeneous and continuous extent of land cover in a 50 m radius (2018).
[7] The presence of crop residues is assessed by surveyors for the LUCAS points (LC = B00) to register the land management activities. The reference area is the field, as far as visible from the point reached by the surveyor.
[8] The presence of signs of grazing in the plot relative to the survey points are registered as special remarks on LC/LU for each point (LC = B00, C00, D00, E00, F00 and H00). Signs of permanent or occasional grazing (e.g. animal tracks) can occur in the field/parcel where the point fall in, if there are no field borders the area up to a distance of some 500 m.

Results show 5014 at EU level with almost two-thirds concentrated in few countries: ES, UK, FR, DE and RO, with the largest share in ES (39%).

Figure 22 – Distribution of LUCAS survey points (2018) among countries extracted with the rule on F40 (total number of points at EU level is 5014).

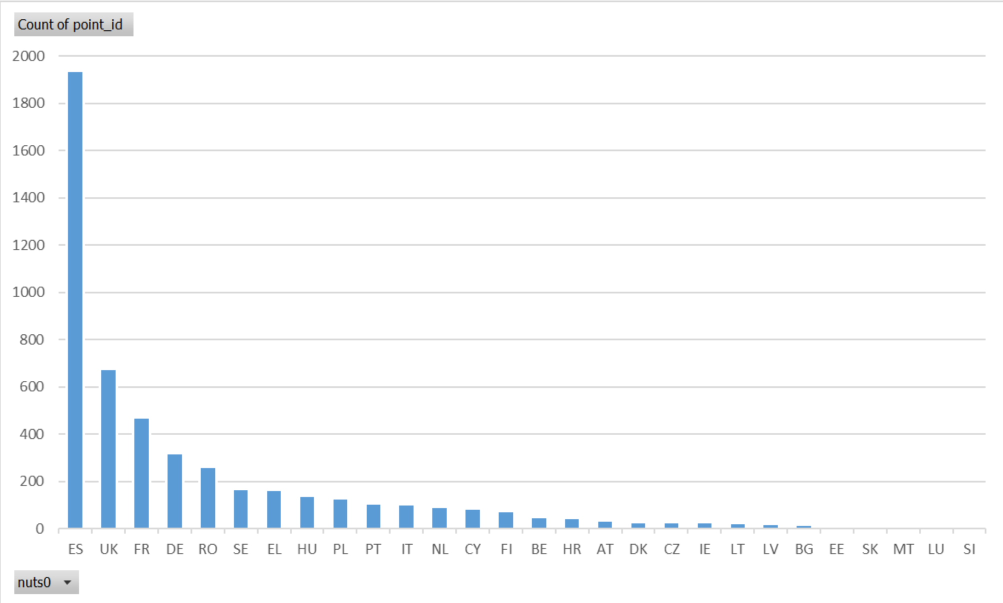
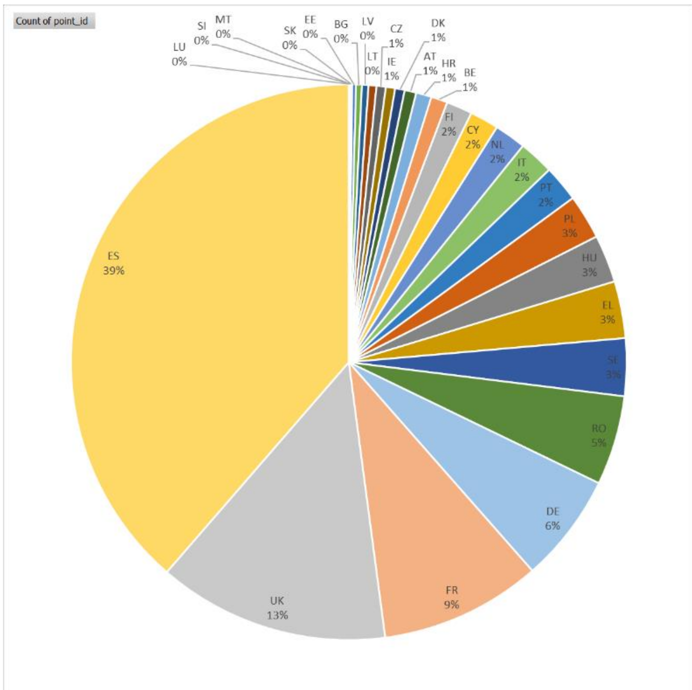
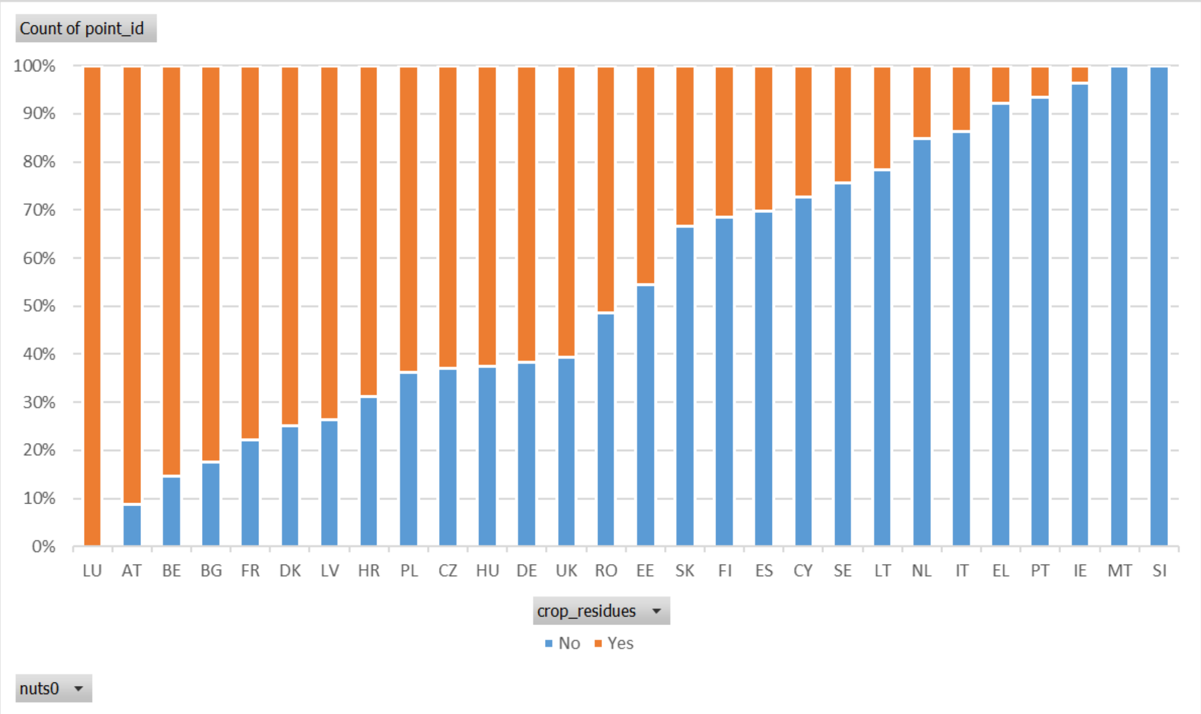


Figure 23 – Percentage of LUCAS survey points (2018) by country extracted with the rule on F40 (total number of points at EU level is 5014).



The presence of crop residues (variable CROP_RESIDUES) is registered for 43.5% of the points with the following distribution at country level.

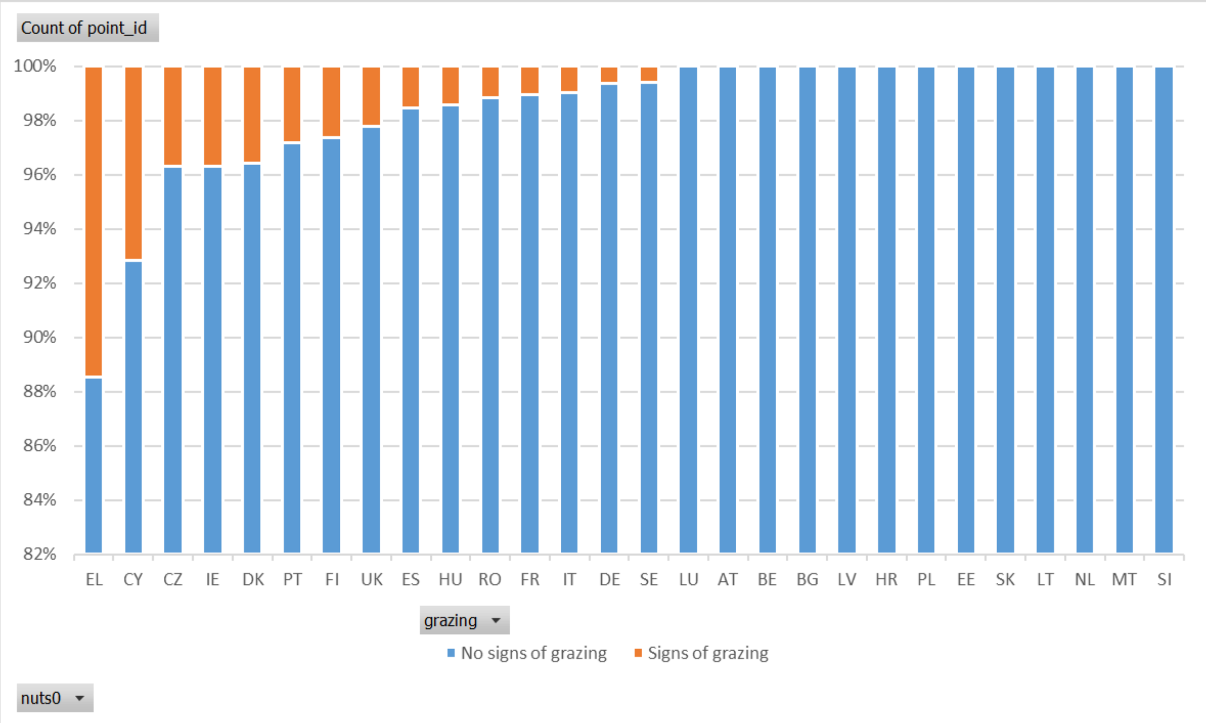
Figure 24 - Percentage of points extracted with the rule on F40 with or without crop residues by country.



Signs of grazing are negligible (1.8%) at EU level. The share of points extracted with signs of grazing is variable among countries with the highest value occurring for EL (11.5%). The occurrence of grazing is assessed by surveyors also for F00 points. An analysis of a set of photos of the points was done but was not sufficient to understand the presence of grazing and of a relevant herbaceous cover. Probably some signs were visible at the time of the survey and/or the herbaceous coverage was not correctly assessed or it was ignored and more relevance was given to other signs (e.g. fences, trampling, etc.).

Report

Figure 25 - Percentage of points extracted with the rule on F40 with or without signs of grazing by country.



Given the intrinsic nature of the points classified as F40 and the data exploration results it can be deemed not relevant the use of the F40 class for extracting points for the grassland classes.

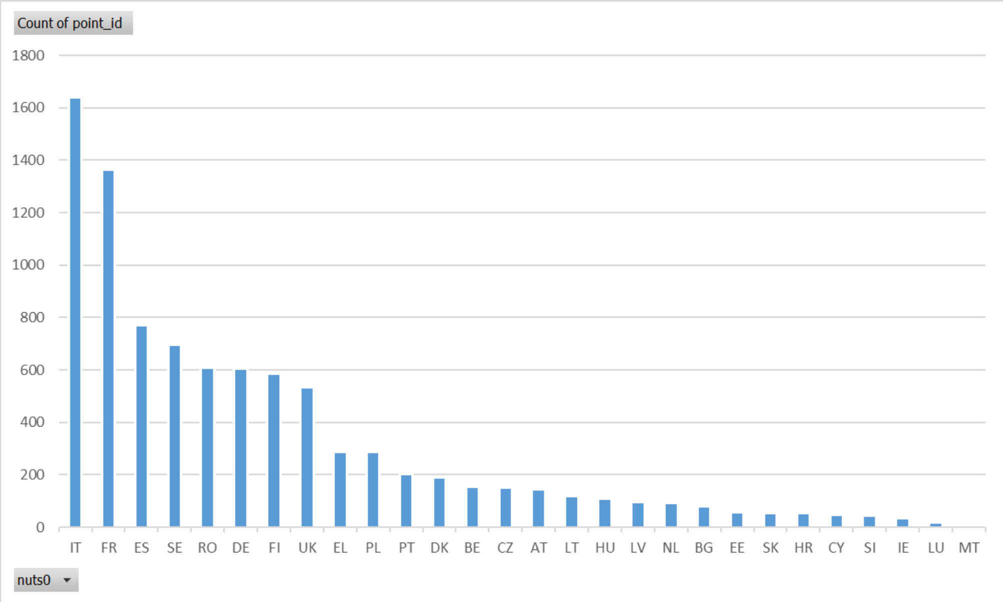
8.1.2 Temporary Grassland (TG)

The rules definition was driven by the selection of LUCAS points with LC representing areas used for fodder crops and grazing included in the agricultural rotation, namely the B50 (Fodder Crops) class and subclasses: B51 (Clovers), B52 (Lucerne), B53 (Other leguminous and mixtures for fodder), B54 (Mixed cereals for fodder) and B55 (Temporary grasslands). Among these points, only those having an agricultural LU were considered relevant.

In the following, the main rules tested before the definition of the final rules (Table 11) are analysed with focus on the main variables and modalities associated to the extracted LUCAS Survey points.**Rule 1_101: LC1=B5* OR LC2=B5***

The rule is based only on LC and extracts all the points with primary or secondary LC classified with the classes B51, B52, B53, B54 and B55. A total number of 9034 survey points are available at EU lever for the 2018 campaign.

Figure 26 – Distribution of LUCAS survey points (2018) among countries extracted with the Rule 1_101 (total number of points at EU level is 9034).



Most points at EU level (97.8%) have LC1=B5*, the remaining belongs to belongs to classes B00 (Cropland), C00 (Woodland) or A30 (Other artificial areas).

The majority of points miss a second LC variable (94% ca.). The remaining have a LC2 equal to B00 (3.4% ca.) or E30 (2.6% ca.). The B00 classes concerned are B10 (Cereals), B30 (Non-permanent industrial crops) and B40 (Dry pulses, vegetables and flowers) and B50 (Fodder crops). The B50 class cover 3% ca. corresponding to 275 points. All the combinations of LC1 and LC2 reported in the following table are compliant with the LUCAS classification specifications.

Table 19 – Number of LUCAS Survey points (2018) for each combination of LC1 - LC2 (Rule 1_101).

LC1/LC2	8	B11	B13	B14	B15	B16	B18	B31	B32	B35	B41	B44	B51	B52	B53	B54	B55	E30	EU-28
A30													1	1	4		6		12
B11													6	5	7		11		29
B12															1				1
B13													7	3	6	1	23		40
B14													1				4		5
B15													9		3		7		19
B16													1		1		4		6
B18																	2		2

B19															2				2
B22																	1		1
B23														1	2		2		5
B31														1	1		1		3
B32													2		1		2		5
B35															2				2
B36																	1		1
B37																	2		2
B41													1			1	2		4
B43													1				1		2
B44																	1		1
B51	547		6		2				2					5	1		23	34	620
B52	1843	1	1		3				1				3		1		11	136	2000
B53	1178				2		1	3				1				2	2	17	1206
B54	835									1	1				1			22	860
B55	4100	1	3	1	1	2			1	1			22	3	5			12	4152
B71													1			1	1		3
B74																2			2
B75													1		1	1	1		4
B76																	1		1
B81														5	2	2	5		14
B82													1				1		2
B83																	2		2
C10															11	11	3		25
C21																	1		1
EU-28	**8503**	**2**	**10**	**1**	**8**	**2**	**1**	**3**	**4**	**2**	**1**	**1**	**57**	**24**	**52**	**21**	**121**	**221**	**9034**

The LC1/LC2 combinations obtained and the further visual analysis of a set of points with the LUCAS photos suggest considering eligible for the TG class all the points with LC1= B1*, B2*, B3*, B4* or C* when LC2=B5*. The points having the primary LC classified as B7* (Permanent crops: fruit trees) or B8* (Other permanent crops) and LC2=B5* can be considered not relevant for the TG class since the permanent cultivation is deemed dominant.

The points with LC1=A30 (Other artificial areas) and LC2=B50 are 12 at EU level and have the first LU1=U319 (Electricity, gas and thermal power distribution) and LU2=U111. The visual analysis of the LUCAS photos of the points reveals that 10 points are located below electric power distribution lines, one is located inside a wind power plant, and one misses clear features useful to identify the A30 class (point_ID = 48244768). These points can be considered eligible for the TG class and will be included in one of the final TG rules.

Most points extracted with the Rule 1_101 have LU1=U111 (99.3%ca.), the remaining LU1 are U112, U113, U120, U319, U350, U361, U415.

Table 20 - Number of LUCAS Survey points (2018) by LU1 (Rule 1_101).

LU1	# LUCAS points
U111	8970
U112	6
U113	16
U120	24
U319	12
U350	2
U361	1
U415	3
EU-28	**9034**

The majority of points miss the secondary LU (99.5% ca.), the remaining points have the following LU2: U111, U350, U361 and U370.

Table 21 - Number of LUCAS Survey points (2018) by LU2 (Rule 1_101). The code 8 indicates that the points misses a secondary LU.

LU2	# LUCAS points
8	8991

LU2	# LUCAS points
U111	37
U350	3
U361	2
U370	1
EU-28	**9034**

According to the LUCAS classification, generally B50 subclasses can be linked to a LU with agricultural production (U111), kitchen gardens (U113) and fallow land (U112) except for B55 (Temporary grasslands), which can be linked only to U111 and U113.

Table 22 - Number of LUCAS Survey points (2018) for each combination of LC1 - LU1 (Rule 1_101).

LC1/LU1	U111	U112	U113	U120	U319	U350	U361	U415	EU-28
A30					12				12
B11	29								29
B12	1								1
B13	40								40
B14	5								5
B15	19								19
B16	6								6
B18	2								2
B19	2								2
B22	1								1
B23	5								5

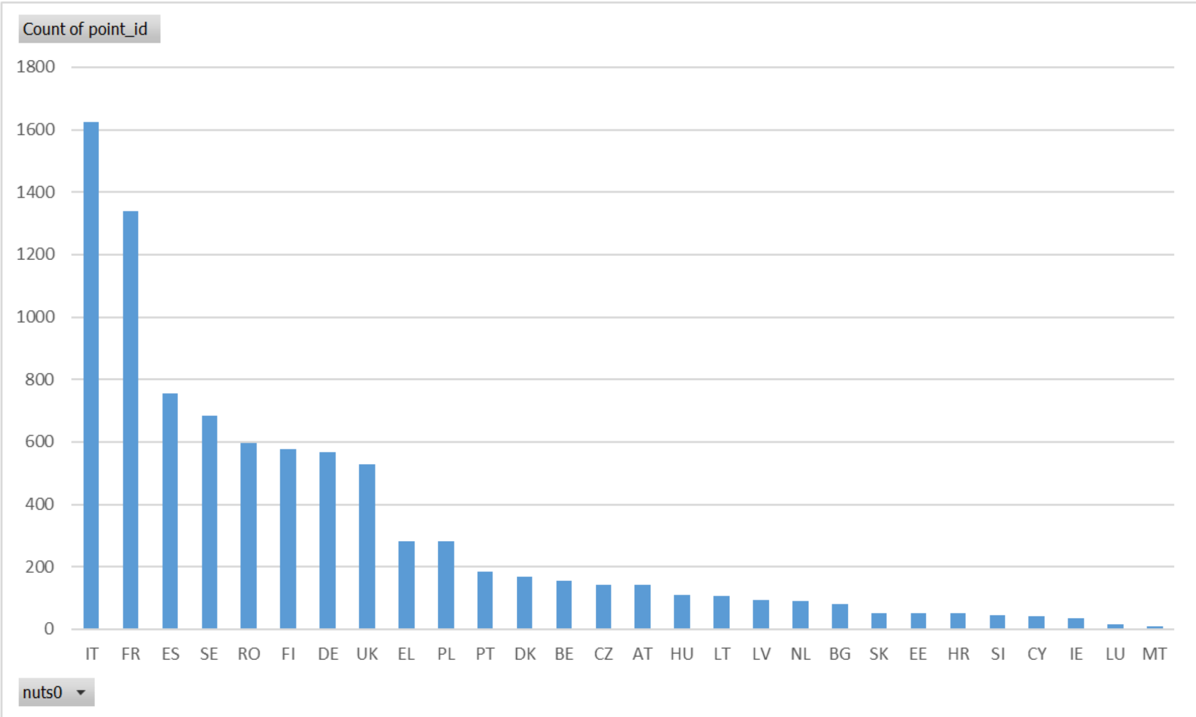
B31	3								3
B32	5								5
B35	2								2
B36	1								1
B37	2								2
B41	4								4
B43	2								2
B44			1						1
B51	619	1							620
B52	1995	3	2						2000
B53	1205		1						1206
B54	857	2						1	860
B55	4138		9			2	1	2	4152
B71	3								3
B74	2								2
B75	3		1						4
B76			1						1
B81	14								14
B82	1		1						2
B83	2								2
C10	2			23					25
C21				1					1
EU-28	**8970**	**6**	**16**	**24**	**12**	**2**	**1**	**3**	**9034**

Report

Rule 1_11: LC1 = B5* AND (LU1 = (U111 OR U112) OR LU2 = (U111 OR U112))

The rule is based only on LC1 by collecting the points with fodder crops classes and subclasses (B50) and includes a condition on the primary and secondary LUs that must be U111 (Agriculture (excluding fallow land and kitchen gardens) or U112 (Fallow land). A total number of 8821 survey points are available at EU lever for the 2018 campaign.

Figure 27 – Distribution of LUCAS survey points (2018) among countries extracted with the Rule 1_11 (total number of points at EU level is 8821).



Most of the points (96% ca.) misses a secondary LC (LC2=8). The remaining points have a LC2 equal to E30 (2.5 % ca.), B50 subclasses (0.9% ca.) and other B00 subclasses: B10 (Cereals), B30 (Non-permanent industrial crops) and B40 (Dry pulses, vegetables and flowers). The combinations of LC1/LC2 reported in the following table are all compliant with the LUCAS classification specifications. Combination between Fodder crops (B50) and Cereals (B10) are always “likely combinations”; only B50-B17 (Rice) is an “impossible combination”. For the future LUCAS implementation, it could be verified the extent to which the combination LC1=B5* and LC2=B1*, B2*, B3* or B4* are plausible.

Table 23 - Number of LUCAS Survey points (2018) for each combination of LC1 and LC2 generated with the Rule 1_11.

LC1/LC2	8	B11	B13	B14	B15	B16	B18	B31	B32	B35	B41	B44	B51	B52	B53	B54	B55	E30	EU-28

B55	4087	1	3	1	1	2			1	1			22	3	5			12	4139
B52	1842	1	1		3				1				3		1		11	135	1998
B53	1177				2		1	3				1				2	2	17	1205
B54	834									1	1				1			22	859
B51	547		6		2				2					5	1		23	34	620
EU-28	8487	2	10	1	8	2	1	3	4	2	1	1	25	8	8	2	36	220	8821

The dominant LU1 is U111 (99.9%ca.). The remaining LU1 are U112 and U350 (Community services), that is considered an unlikely combination for the LUCAS specifications.

Table 24 - Number of LUCAS Survey points (2018) by LU1 generated with the Rule 1_11.

LU1	# LUCAS points
U111	8814
U112	6
U350	1
Grand Total	8821

Most points miss a second LU (99.9% ca.), while the remaining (7 points) have the following LU2: U111, U350 (Community services), U361 (Amenities, museums, leisure) and U370 (residential). The points having LU2=350 or 361 or 370 have LU1=U111 are considered "unlikely" combinations according to the LUCAS specifications.

Table 25 - Number of LUCAS Survey points (2018) by LU2 generated with the Rule 1_11. The code 8 indicates that the points miss a secondary LU.

LU2	# LUCAS points
8	8814
U111	1
U350	3
U361	2
U370	1
EU-28	8821

Table 26 - Number of LUCAS Survey points (2018) for each combination of LC1 - LU1 generated with the Rule 1_11.

LC1/LU1	U111	U112	U350	EU-28
B55	4138		1	4139
B52	1995	3		1998
B53	1205			1205
B54	857	2		859
B51	619	1		620
EU-28	8814	6	1	8821

1.1.1 Permanent grassland agricultural (PGa)

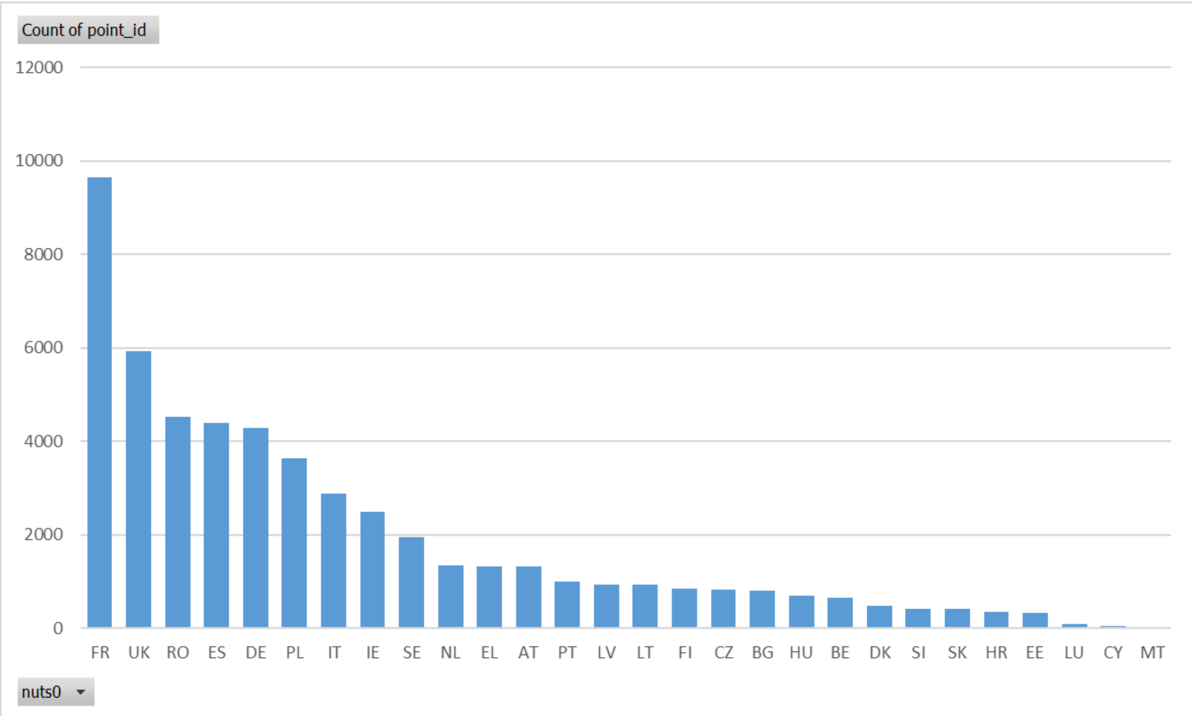
The rules tested were based on the selection of points with a LC classified as Grassland classes and subclasses (E00) with an agricultural LU.

In the following, the main rules tested before the definition of the final rules (Table 11) are analysed with focus on the main variables and modalities associated to the extracted LUCAS Survey points.

Rule 2_10: LC1 = E* AND (LU1 = (U111 OR U112) OR LU2 = (U111 OR U112))

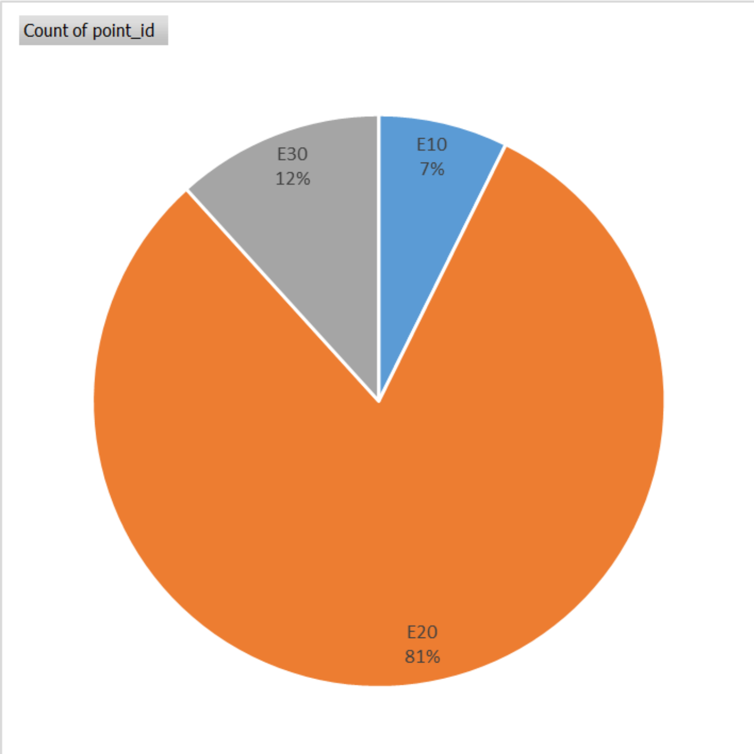
The rule extracts all the LUCAS survey points with the primary LC classified as Grassland (E00) with primary or secondary agricultural LU (U111 or U112).

Figure 28 – Distribution of LUCAS survey points (2018) among countries extracted with the Rule 2_10 (total number of points at EU level is 52639).



Most of the points extracted by the rule (81% ca.) have E20 (Grassland without tree/shrub cover). Almost all points miss the secondary LC (LC2) except one having LC2=E20 and LC1=E20 that can be considered an error.

Figure 29 – Percentage of LUCAS survey points (2018) by LC1 extracted with the Rule 2_10.



Considering the LU associated to the points, LU1 is represent mainly agriculture: U111 (93% ca.) and U112 (7% ca.). The secondary LU (LU2) is missing for almost all the points (99% ca.).

Table 27 - Number of LUCAS Survey points (2018) by LU1 generated with the Rule 2_10.

LU1	# Lucas points
U111	48711
U112	3732
U120	51
U210	3
U312	48

LU1	# Lucas points
U314	8
U316	3
U317	5
U318	22
U319	7
U321	8
U330	1
U350	17
U370	23
EU-28	**52639**

Table 28 – Number of LUCAS Survey points (2018) by LU2 generated with the Rule 2_10.

LU2	# Lucas points
8	52205
U111	194
U112	2

LU2	# Lucas points
U120	9
U150	10
U341	5
U350	28
U361	80
U362	68
U370	38
EU-28	**52639**

The following table reports the combination LU1-LU2 for all the points extracted with rule to identify specific LUs that may be not relevant for the grassland class. Specifically, the following can be considered not suitable for defining PGa and can be removed from the final PGa rules: Road transport (U312), Air transport (U314), Construction (U330), Commerce, financial, professional and information services (U340), Community services (U350), Recreation, leisure, sport (U360) and Residential (U370). Concerning U312 and analysis of the LUCAS survey photos was carried out resulting in several roads in between or inside agricultural fields with a relevant herbaceous coverage.

Table 29 - Number of LUCAS Survey points (2018) for each combination of LU1 – LU2 generated with the Rule 2_10.

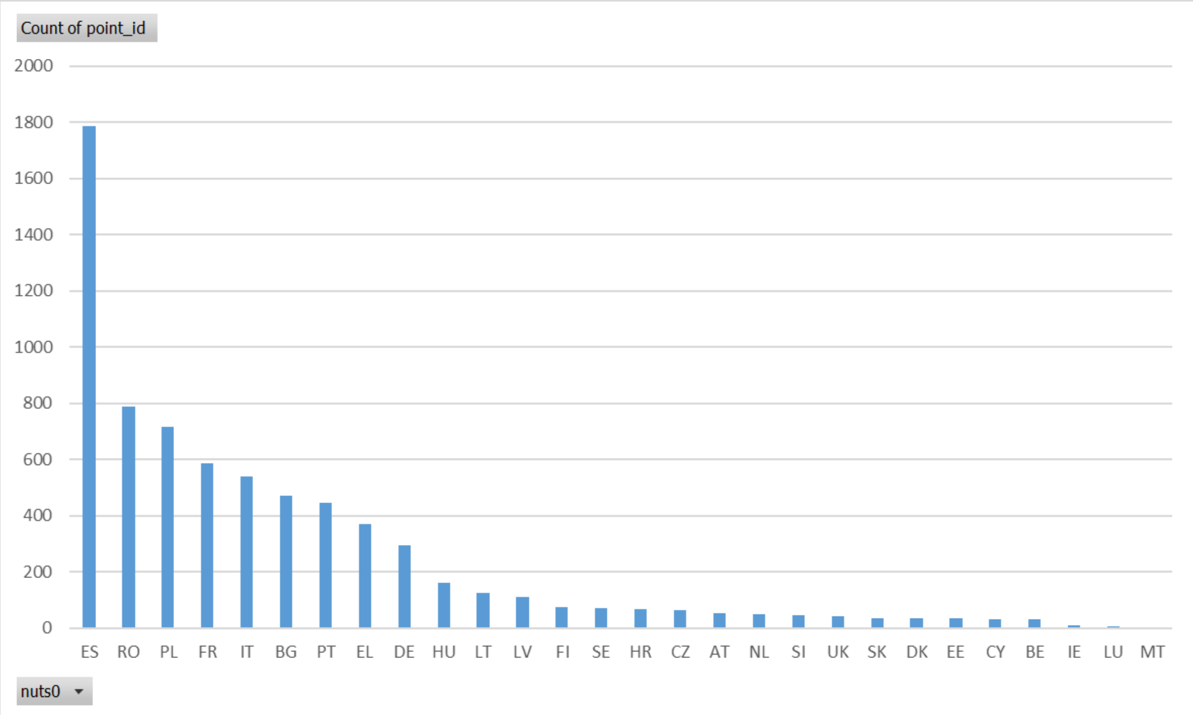
LU1/LU2	8	U111	U112	U120	U150	U341	U350	U361	U362	U370	EU-28
U111	48474			9	10	5	28	80	67	38	48711
U112	3731								1		3732
U120		50	1								51

U210		3									3
U312		48									48
U314		8									8
U316		3									3
U317		5									5
U318		22									22
U319		6	1								7
U321		8									8
U330		1									1
U350		17									17
U370		23									23
EU-28	**52205**	**194**	**2**	**9**	**10**	**5**	**28**	**80**	**68**	**38**	**52639**

Rule 6_01: LC2 = E* AND (LU1= (U111 OR U112) OR LU2= (U111 OR U112))) AND INSPIRE_PLCC4>=30

The rule allows to extract the points with the secondary LC belonging to Grassland (E00) with primary and secondary agricultural LU (U111 or U112) and a minimum threshold of herbaceous cover (>=30%) defined by the variable INSPIRE_PLCC4.

Figure 30 – Distribution of LUCAS survey points (2018) among countries extracted with the Rule 6_01 (total number of points at EU level is 7025).



The following table reports all the points for each combination LC1-LC2. An analysis of a sample of LUCAS survey photos was also carried out to better analyse the points having specific LC1. Results of the analysis suggest excluding for the final PGa rules the points with LC1= A21, A22 and B*. The points having LC1=C* can be relevant for the PGsn class.

Table 30 - Number of LUCAS Survey points (2018) for each combination of LC1 – LC2.

LC1/LC2	E10	E20	E30	EU-28
A21		15	13	28
A22		210	68	278
A30	9	111	7	127
B11			730	730
B12			31	31
B13			250	250
B14			126	126

B15			131	131
B16			259	259
B18			67	67
B19			18	18
B21			38	38
B22			22	22
B23			17	17
B31			153	153
B32			297	297
B33			12	12
B34			1	1
B35			8	8
B36			1	1
B37			9	9
B41			94	94
B43			38	38
B44			1	1
B45			10	10
B51			14	14
B52			68	68
B53			11	11
B54			9	9
B55			6	6
B71		391	36	427

B72		47	18	65
B73		72	37	109
B74		104	155	259
B75		225	93	318
B76		4	38	42
B77		4	10	14
B81		154	488	642
B82		243	221	464
B83		2	13	15
B84			20	20
C10		1533	153	1686
C21		7	2	9
C22		40	9	49
C23		15		15
C31		4		4
C32		14	2	16
C33		14	2	16
E20		1		1
EU-28	**9**	**3210**	**3806**	**7025**

1.1.2 Permanent grassland semi-natural (PGsn)

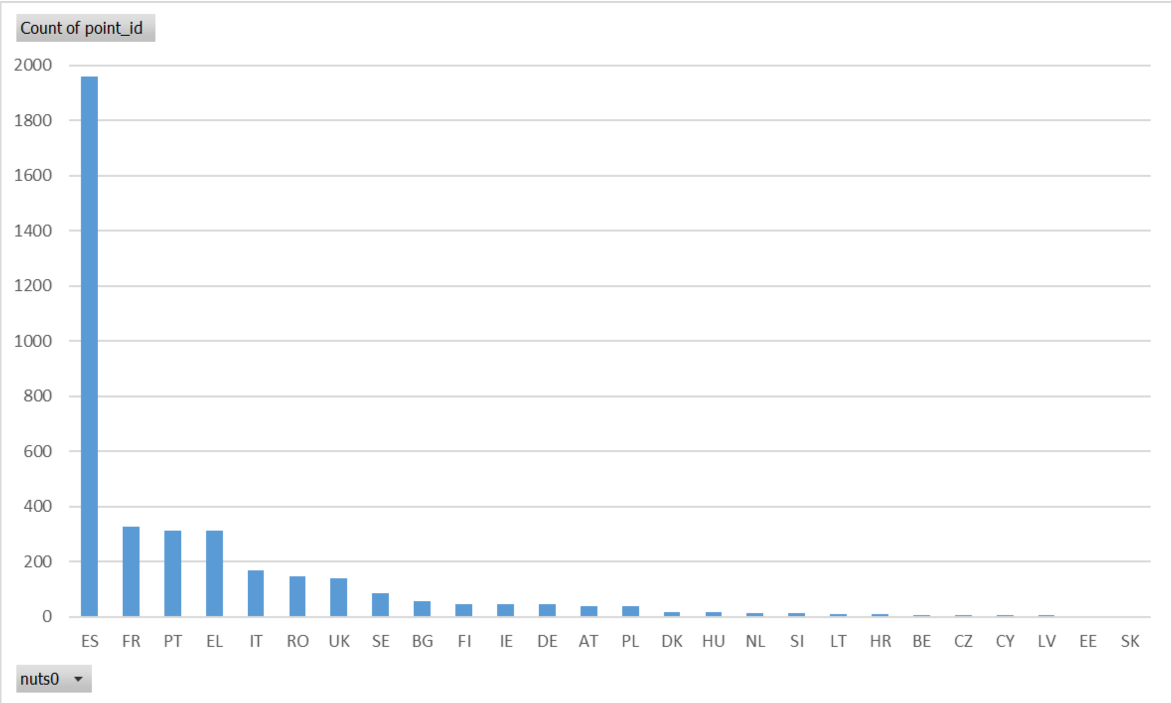
The rules explored were based on the selection of points with a LC classified as Woodland (C00) or Shrubland (D00) classes and sub-classes and with an agricultural LU. In addition, a minimum herbaceous cover was set as pre-condition trough the variable INSPIRE_PLCC4.

Rule 1718: (LC1=C*OR LC1=D*) AND (LU1= (U111 OR U112) OR LU2= (U111 OR U112)) AND INSPIRE_PLCC4>=30

The rule extracts all the LUCAS survey points having as primary LC Woodland (C00) or Shrubland (D00) with agricultural LU (U111 or U112) and with a minimum herbaceous cover (>=30%).

In the following, the main rules tested before the definition of the final rules (Table 11) are analysed with focus on the main variables and modalities associated to the extracted LUCAS Survey points.

Figure 31 – Distribution of LUCAS survey points (2018) among countries extracted with the Rule 1718 (total number of points at EU level is 3840).



As concerns the points extracted at EU level, almost 95% of LC1 of is made up of C10, D10 and D20.

Table 31 – Number and percentage of LUCAS survey points (2018) by LC1 (Rule 1718).

LC1	# LUCAS points	% LUCAS points
C10	2247	58.52%
C21	30	0.78%
C22	86	2.24%

LC1	# LUCAS points	% LUCAS points
C23	24	0.63%
C31	10	0.26%
C32	25	0.65%
C33	29	0.76%
D10	454	11.82%
D20	935	24.35%
EU-28	**3840**	**100.00%**

The following table reports the combination LC1-LC2 for each LUCAS point extracted with the rule. A set of points are classified as Fodder crops (B50) as LC2 resulting in an overlap with the points extracted with the rules for TG. The final rule needs to avoid the overlap and remove some points that will belong to the TG class.

Table 32 - Number of LUCAS Survey points (2018) for each combination of LC1 – LC2 (Rule 1718).

LC1/LC	8	E20	E30	D20	B15	B53	B54	B55	B14	B11	B13	B81	B84	B43	B19	B12	B23	B18	B32	B31	EU-28
C10	426	1533	153	72	18	11	9	3	4	4	4	3	1	1	1	1	1		1	1	2247
D20	935																				935
D10	454																				454
C22	30	40	9	5	2																86
C21	20	7	2					1													30
C33	12	14	2															1			29
C32	9	14	2																		25

C23	9	15																			24
C31	6	4																			10
EU-28	1901	1627	168	77	20	11	9	4	4	4	4	3	1	1	1	1	1	1	1	1	3840

The analysis of the number of LUCAS points at EU level by LC2 shows the prevalence of points with missing LC2 (49.5%), followed by E20 (42% ca.), E30 (4% ca.) and other B00 subclasses (1.7% ca.).

Table 33 - Number of LUCAS Survey points (2018) for each combination of LC1 – LU1 (Rule 1718).

LC1/LU1	U111	U112	U120	U210	U311	U312	U370	EU-28
C10	450		1786			6	5	2247
D20	877	51	2			4	1	935
D10	413	30	7	1	1	1	1	454
C22	25		60			1		86
C21	11		19					30
C33	12		17					29
C32	8		17					25
C23	6		18					24
C31	7		3					10
EU-28	**1809**	**81**	**1929**	**1**	**1**	**12**	**7**	**3840**

Considering the LU associated to the points, LU1 is represent mainly agricultural and forestry areas: U111, U112 (49% ca.) and U120 (50 % ca.). The remaining LUs are U210, U311, U312 and U370.

Table 34 – Number and percentage of LUCAS survey points (2018) by LU1 (Rule 1718).

LU1	# LUCAS points	% LUCAS points
U111	1809	47.11%
U112	81	2.11%
U120	1929	50.23%
U210	1	0.03%
U311	1	0.03%
U312	12	0.31%
U370	7	0.18%
EU-28	**3840**	**100.00%**

The second LU (LU2) is missing for almost the half of the points (49% ca.), the other half is U111 (51% ca.). A negligible share belongs to U120, U150, U350, U361 and U362.

Table 35 – Number and percentage of LUCAS survey points (2018) by LU2 (Rule 1718).

LU1	# LUCAS points	% LUCAS points
8	1869	48.67%
U111	1948	50.73%
U112	2	0.05%

LU1	# LUCAS points	% LUCAS points
U120	11	0.29%
U150	2	0.05%
U350	1	0.03%
U361	6	0.16%
U362	1	0.03%
EU-28	**3840**	**100.00%**

The following tables report the combination LU1/LU2 split by the two distinct LC1: C00 and D00.

Table 36 - Number of LUCAS Survey points (2018) for each combination of LU1 – LU2 when LC1=D00 (Rule 1718).

LU1/LU2	8	U111	U112	U361	U362	EU-28
U111	1286			3	1	1290
U112	81					81
U120		9				9
U210		1				1
U311		1				1
U312		5				5
U370		1	1			2
EU-28	**1367**	**17**	**1**	**3**	**1**	**1389**

Table 37 - Number of LUCAS Survey points (2018) for each combination of LU1 – LU2 when LC1=C00 (Rule 1718).

LU1/LU2	8	U111	U112	U120	U150	U350	U361	EU-28
U111	502			11	2	1	3	519
U120		1919	1					1920
U312		7						7
U370		5						5
EU-28	**502**	**1931**	**1**	**11**	**2**	**1**	**3**	**2451**

In terms of size of the parcel of the points, large parcels (> 10ha) are dominant followed by the range 1-10 ha (25% ca.).

Table 38 - Number and percentage of LUCAS survey points (2018) by parcel size (Rule 1718).

Parcel size	# LUCAS points	% LUCAS points
> 10 ha	2294	59.74%
1 - 10 ha	961	25.03%
< 0.5 ha	366	9.53%
0.5 - 1 ha	219	5.70%
EU-28	**3840**	**100.00%**

1.1.3 Other Grassland (OG)

The definition of rules was based on the identification of LUCAS survey points with specific LC such as: Woodland (C00) and Shrubland (D00) classes and sub-classes with potential agricultural use[9]. A minimum threshold for the herbaceous cover is always a pre-condition to be set with the variable INSPIRE_PLCC4. These areas can be often either marginal agricultural lands belonging to farms or outside area managed by farms that might be potentially eligible in the CAP framework. Purely residential, industrial, commercial grassland area are excluded from this class. It is also expected that a subset of points classified as OG will be also eligible for the agroforestry classes.

In terms of LU, LUCAS survey points classified as U400 (Unused and abandoned areas) were considered relevant for OG after an analysis of LUCAS classification rules and visual inspection of a subset of points with LUCAS 2018 survey photos. U400 is made up of the following subclasses:

- U410 Abandoned areas. This class consists of abandoned areas with signs or structures of previous use of any kind. Areas belonging to the abandoned class are not in use and cannot be used anymore for the original purpose without major reparation/renovation work. The following classes are included: U411, U412, U413, U414 and U415. By definition, agricultural areas and fallow land are excluded.
- U420 Semi-natural and natural areas not in use. This class includes areas, which are in natural/semi-natural state. By definition, agricultural areas and fallow land are excluded.

Preliminary analysis of a sample of photos

A visual analysis of a small set of photos (n=20) in areas with the rule LC1=E00 AND LU1=U400 shows points representing mainly:

- marginal areas, probably non managed for agricultural activities;
- Areas located in complex agricultural patterns cultivated and non-cultivated.

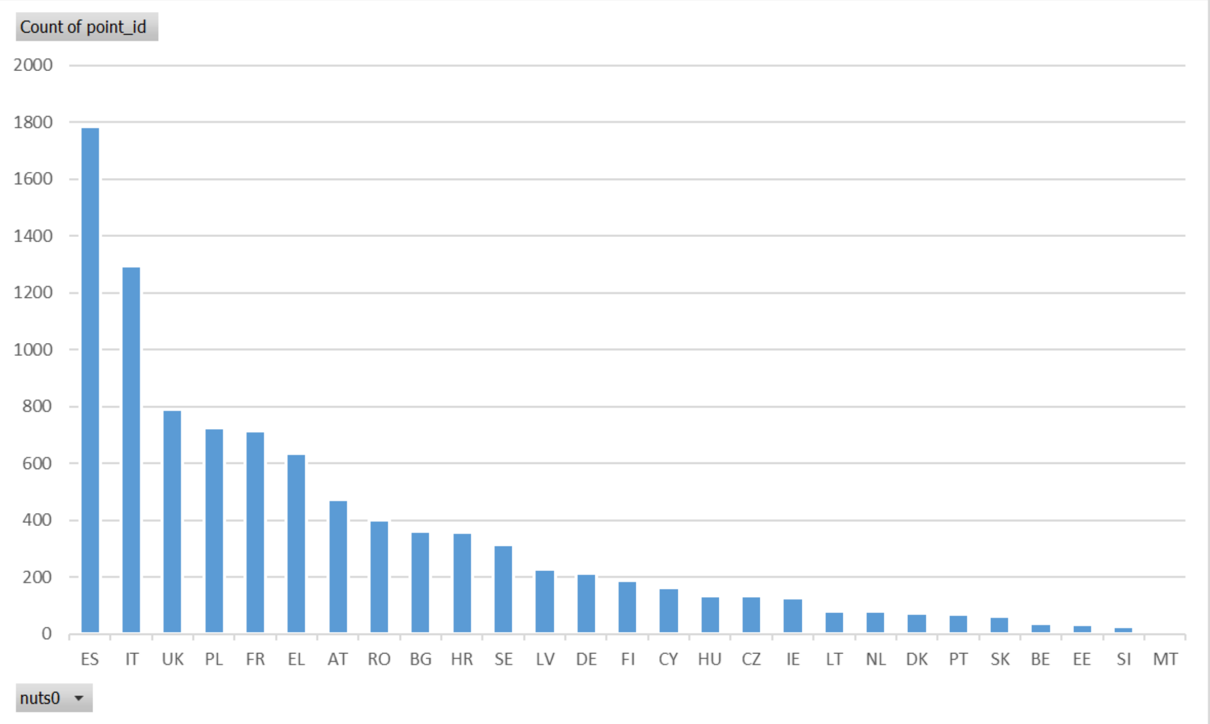
In the following, the main rules tested before the definition of the final rules (Table 11) are analysed with focus on the main variables and modalities associated to the extracted LUCAS Survey points.

Rule 2_2: LC1 = E* AND (LU1 = U4* OR LU2=U4*) AND INSPIRE_PLCC4>=30

[9] An extraction of points was also performed with the rule LC1=B50 AND LU1=U400 resulting in only three points with LU1=U415 (Other abandoned areas). The specific combination B50-U415 is classified as “unlikely” according to the LUCAS Survey 2018 instructions. There is a contradiction between an agriculture use (B50) and abandoned land. As it concerns only three points this will not be further investigated. These points are excluded from the grassland classes analysed.

The rule extracts all the LUCAS survey points having as primary LC Grassland (E00), with a primary or secondary LU classified as unused and abandoned areas (U400) and with a minimum herbaceous cover (>=30%).

Figure 32 – Distribution of LUCAS survey points (2018) among countries extracted with the Rule 2_2 (total number of points at EU level is 9486).



Total number of points at EU level 9486. As reported by the following table most of the points (57% ca.) have primary LC (LC1) classified as Grassland without tree/shrub cover (E20). None of the points has a secondary LC (LC2).

Table 39 - Number and percentage of LUCAS survey points (2018) by LC1 (Rule 2_2).

LC1	# LUCAS points	% LUCAS points
E20	5361	56.51%
E10	2371	24.99%
E30	1754	18.49%

LC1	# LUCAS points	% LUCAS points
EU-28	9486	100.00%

Table 40 - Number of LUCAS Survey points (2018) for each combination of LC1 – LU1 (Rule 2_2).

LC1/LU1	U411	U412	U413	U414	U415	U420	EU-28
E20	19	7	10	39	356	4930	5361
E10	10	2	5	37	234	2083	2371
E30	38	5	10	53	330	1318	1754
EU-28	**67**	**14**	**25**	**129**	**920**	**8331**	**9486**

The majority (88% ca.) of eligible points belong to the classes U420 (Semi-natural and natural areas not in use) and U415 (Other abandoned areas). The secondary LU (LU2) is missing.

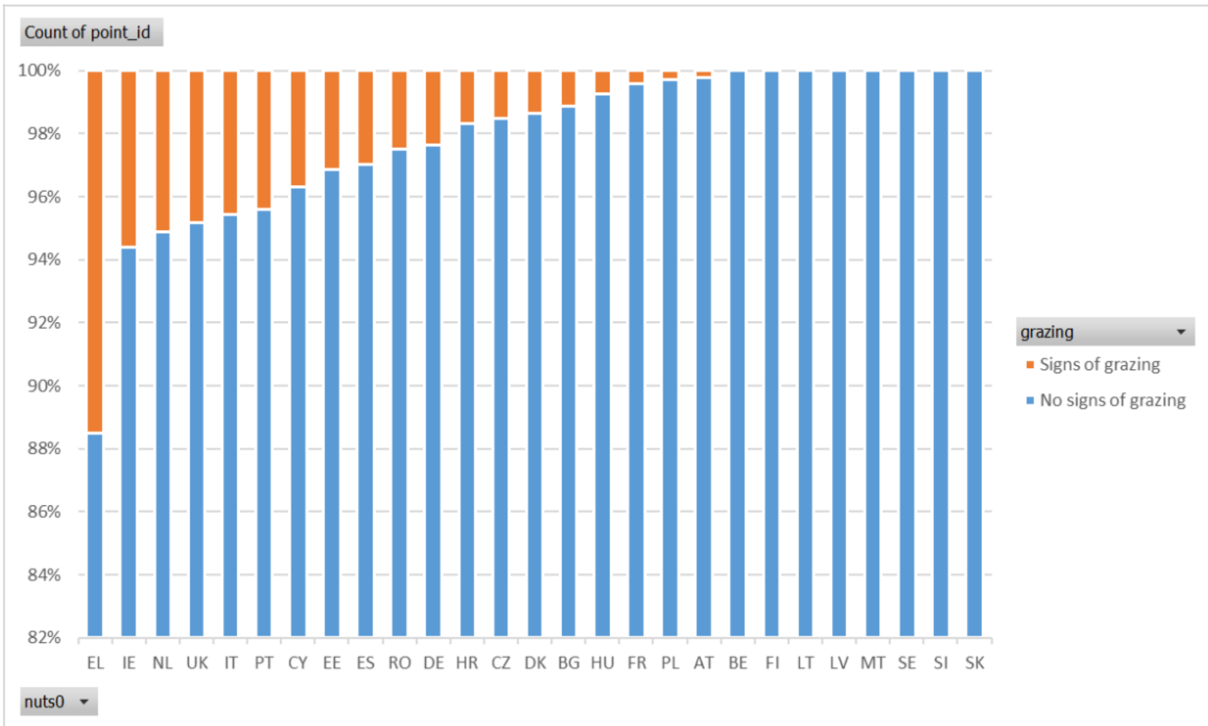
Table 41 - Number and percentage of LUCAS survey points (2018) by LU1 (Rule 2_2).

LU1	# LUCAS points	% LUCAS points
U411	67	0.71%
U412	14	0.15%
U413	25	0.26%
U414	129	1.36%

LU1	# LUCAS points	% LUCAS points
U415	920	9.70%
U420	8331	87.82%
EU-28	**9486**	**100.00%**

At EU level a small share of points exhibits signs of grazing (3% ca.). The share is widely variable at NUTS0, values greater than 5% are registered for EL, IE, and NL.

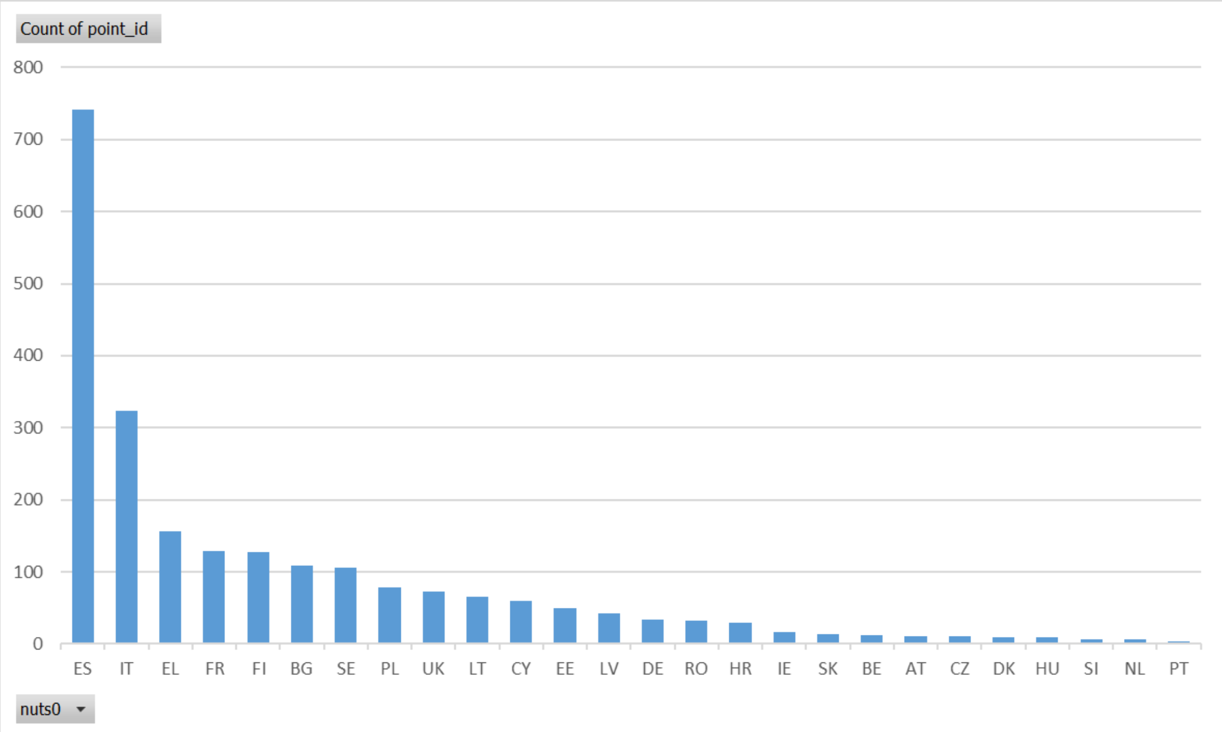
Figure 33 – Percentage of LUCAS survey points (2018) among countries extracted with the Rule 2_2 with or without sign of grazing.



Rule 17_2: LC1=C* AND (LU1=U4* OR LU2=U4*) AND INSPIRE_PLCC4>=30

The rule extracts all the LUCAS survey points having as primary LC Woodland (C00) with a primary or secondary LU classified as abandoned (U400 - Unused and abandoned areas) and with a minimum herbaceous cover (>=30%).

Figure 34 – Distribution of LUCAS survey points (2018) among countries extracted with the Rule 17_2 (total number of points at EU level is 2254).



Total number of points at EU level is 2254. As reported by the following table most of the points are classified as C10 (Broadleaved woodland), 66% ca. and C22 (Pine dominated coniferous woodland), 19.5% ca.. More than 99.5% of the points miss the secondary LC (LC2), the remaining share belongs to the classes E30 (0.31%), E20 (0.13%) and B75 (0.04%).

Table 42 - Number and percentage of LUCAS survey points (2018) by LC1 (Rule 17_2).

LC1	# LUCAS points	% LUCAS points
C10	1489	66.06%
C21	64	2.84%
C22	439	19.48%
C23	94	4.17%

LC1	# LUCAS points	% LUCAS points
C31	12	0.53%
C32	74	3.28%
C33	82	3.64%
EU-28	**2254**	**100.00%**

By analysing the combinations LC1-LC2, one point has B75 (Other fruit trees and berries as secondary LC, a combination "unlikely" according to the LUCAS instructions. The analysis of the photos cannot reveal useful elements to verify the combination. The points having LC2 classified as B70 (Permanent crops: fruit trees) or B80 (Other permanent crops) are removed from the final rule.

Table 43 - Number of LUCAS Survey points (2018) for each combination of LC1 - LC2 (Rule 17_2).

LC1/LC2	8	B75	E20	E30	EU-28
C10	1483	1	2	3	1489
C21	64				64
C22	434		1	4	439
C23	94				94
C31	12				12
C32	74				74
C33	82				82
EU-28	**2243**	**1**	**3**	**7**	**2254**

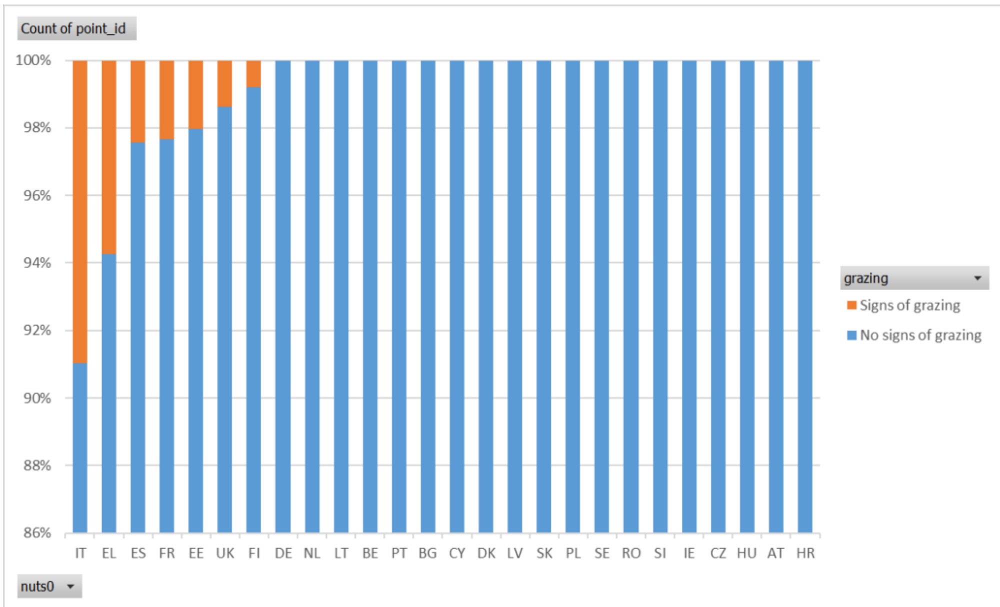
The majority (96% ca.) of eligible points belong to the LU1 class U420 (Semi-natural and natural areas not in use) and U415 (Other abandoned areas). The second LU is missing.

Table 44 - Number and percentage of LUCAS survey points (2018) by LU1 (Rule 17_2).

LU1	# LUCAS points	% LUCAS points
U420	2156	95.65%
U415	66	2.93%
U414	20	0.89%
U411	7	0.31%
U412	3	0.13%
U413	2	0.09%
EU-28	**2254**	**100.00%**

At EU level a small share of points exhibits signs of grazing (3% ca.). The share is widely variable at NUTS0 with the highest values (>=5%) for IT and ES.

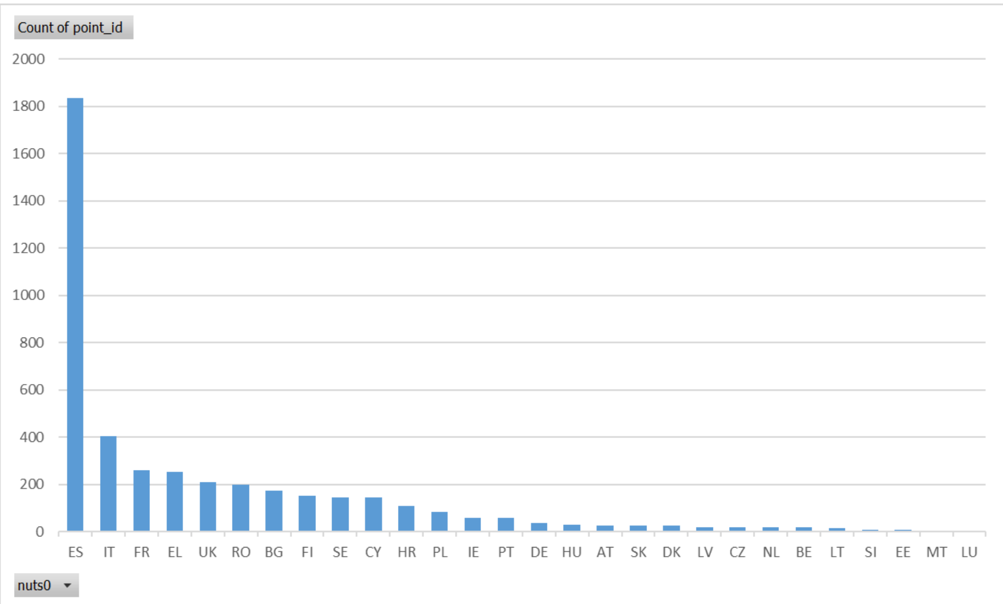
Figure 35 – Percentage of LUCAS survey points (2018) among countries extracted with the Rule 17_2 with or without sign of grazing.



Rule 18_2: LC1=D* AND (LU1=U4* OR LU2=U4*) AND INSPIRE_PLCC4>=30

The rule extracts all the LUCAS survey points having as primary LC Shrubland (D00) with a primary or secondary LU classified as abandoned (U400 - Unused and abandoned areas) and with a minimum herbaceous cover (>=30%).

Figure 36 – Distribution of LUCAS survey points (2018) among countries extracted with the Rule 18_2 (total number of points at EU level is 4342).



Most of the points (2849, 66% ca.) have as primary LC (LC1) Shrubland without tree cover (D20). The second LC (LC2) is missing.

Table 45 - Number and percentage of LUCAS survey points (2018) by LC1 (Rule 18_2).

LC1	# LUCAS points	% LUCAS points
D20	2849	65.61%
D10	1493	34.39%
EU-28	**4342**	**100.00%**

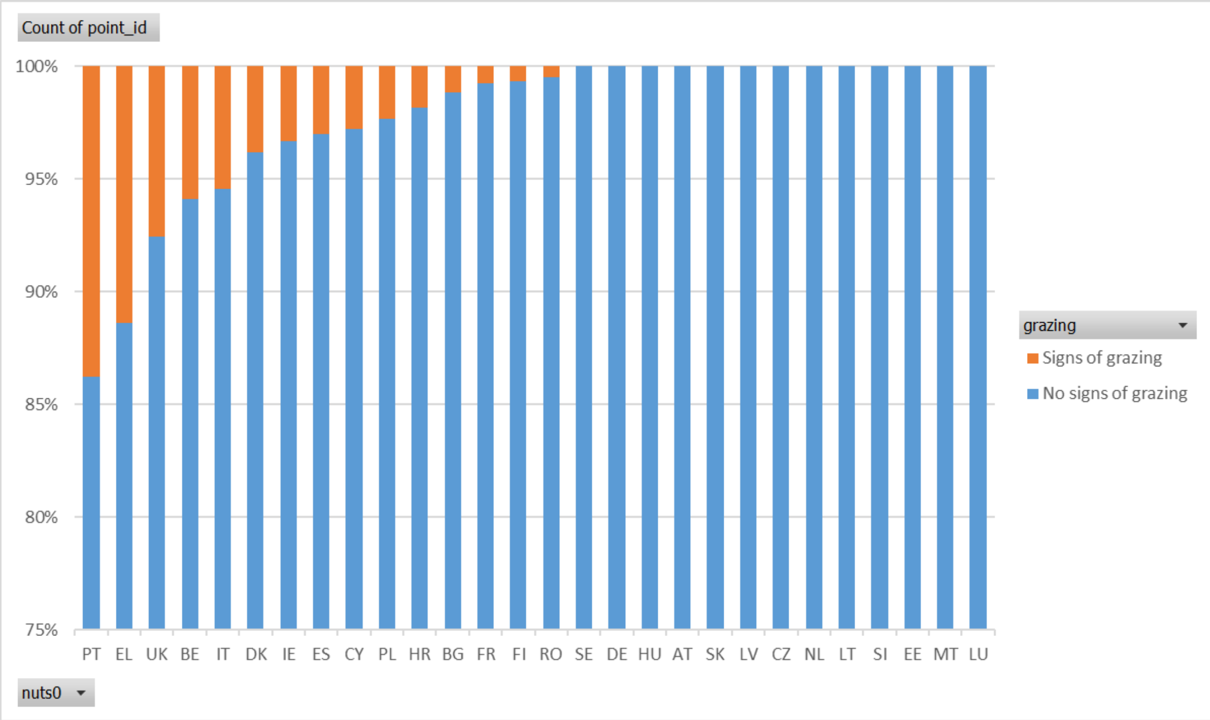
The majority (95% ca.) of eligible points belong to the classes U420 (Semi-natural and natural areas not in use) and U415 (Other abandoned areas). The second LU is missing.

Table 46 - Number and percentage of LUCAS survey points (2018) by LU1 (Rule 18_2).

LU1	# LUCAS points	% LUCAS points
U420	4118	94.84%
U415	199	4.58%
U414	14	0.32%
U411	8	0.18%
U413	2	0.05%
U412	1	0.02%
EU-28	**4342**	**100.00%**

At EU-level a small share of points exhibits signs of grazing (3% ca.). The share is widely variable at NUTS0 with the highest values (>=5%) registered for PT, EL, UK, BE and IT.

Figure 37 – Percentage of LUCAS survey points (2018) among countries extracted with the Rule 18_2 with or without sign of grazing.



1.2 Annex 2 - First analysis of LUCAS survey points with signs of grazing.

A first analysis of LUCAS survey points with sign of grazing was carried out to gain insight on the geographical distribution as well as the associate LCs and LUs as reported in the following tables.

Table 47 - Number of LUCAS survey points (2018) with signs of grazing by country.

NUTS 0	# LUCAS points	% of LUCAS points
AT	314	1.08%
BE	447	1.54%
BG	445	1.53%
CY	73	0.25%
CZ	265	0.91%
DE	1349	4.65%
DK	206	0.71%
EE	56	0.19%
EL	2092	7.21%
ES	5148	17.75%
FI	169	0.58%
FR	6210	21.42%
HR	98	0.34%

NUTS 0	# LUCAS points	% of LUCAS points
HU	173	0.60%
IE	1459	5.03%
IT	1410	4.86%
LT	234	0.81%
LU	77	0.27%
LV	142	0.49%
NL	641	2.21%
PL	393	1.36%
PT	938	3.23%
RO	1653	5.70%
SE	922	3.18%
SI	122	0.42%
SK	103	0.36%
UK	3858	13.30%
EU-28	**28997**	**100.00%**

Table 48 - Number of LUCAS survey points (2018) with signs of grazing by LC1 at EU-level.

LC1	# LUCAS points	% of LUCAS points
E20	18898	65.172%
C10	3004	10.360%
E10	2039	7.032%
D20	1300	4.483%
D10	816	2.814%
E30	679	2.342%
B55	542	1.869%
C22	283	0.976%
B81	117	0.403%
F40	101	0.348%
C33	95	0.328%
B54	92	0.317%
C21	92	0.317%
C32	86	0.297%
B71	68	0.235%

LC1	# LUCAS points	% of LUCAS points
B11	67	0.231%
C23	64	0.221%
B13	63	0.217%
B53	57	0.197%
B15	56	0.193%
C31	47	0.162%
B52	47	0.162%
H12	46	0.159%
B74	40	0.138%
BX1	35	0.121%
B75	34	0.117%
B51	28	0.097%
B16	25	0.086%
B12	21	0.072%
H11	17	0.059%

LC1	# LUCAS points	% of LUCAS points
B73	16	0.055%
B14	15	0.052%
F10	15	0.052%
B31	11	0.038%
B72	10	0.034%
B32	9	0.031%
B19	8	0.028%
BX2	7	0.024%
H21	7	0.024%
B84	6	0.021%
B43	5	0.017%
B82	5	0.017%
B41	4	0.014%
F20	4	0.014%
B18	4	0.014%

LC1	# LUCAS points	% of LUCAS points
B21	3	0.010%
B76	2	0.007%
B22	2	0.007%
F30	1	0.003%
B45	1	0.003%
B17	1	0.003%
B35	1	0.003%
B37	1	0.003%
EU-28	**28997**	100.000%

Table 49 - Number of LUCAS survey points (2018) with signs of grazing by LU1 at EU-level.

LU1	# LUCAS points	% of LUCAS points
U111	24733	85.295%
U112	205	0.707%
U113	20	0.069%
U120	2750	9.484%

LU1	# LUCAS points	% of LUCAS points
U140	3	0.010%
U150	9	0.031%
U210	2	0.007%
U311	1	0.003%
U312	5	0.017%
U313	2	0.007%
U314	1	0.003%
U316	1	0.003%
U318	10	0.034%
U319	2	0.007%
U321	3	0.010%
U322	1	0.003%
U330	1	0.003%
U350	12	0.041%
U361	35	0.121%

LU1	# LUCAS points	% of LUCAS points
U362	19	0.066%
U370	72	0.248%
U411	1	0.003%
U412	1	0.003%
U414	1	0.003%
U415	95	0.328%
U420	1012	3.490%
EU-28	**28997**	100.000%

Table 50 - Number of LUCAS survey points (2018) with signs of grazing for each combination of LU1 - LU2 at EU-level.

LU1/LU2	8	U111	U120	U150	U318	U341	U350	U361	U362	U370	EU-28
U111	24592		30	9		2	15	57	22	6	24733
U112	205										205
U113	19							1			20
U120	545	2191		2				11	1		2750
U140	3										3
U150	9										9
U210	1	1									2

Report

U311	1										1
U312	2	1	1		1						5
U313	2										2
U314		1									1
U316		1									1
U318	3	7									10
U319		2									2
U321	2	1									3
U322	1										1
U330		1									1
U350	5	7									12
U361	32								2	1	35
U362	18							1			19
U370	59	12						1			72
U411	1										1
U412	1										1
U414	1										1
U415	95										95
U420	1012										1012
EU-28	**26609**	**2225**	**31**	**11**	**1**	**2**	**15**	**71**	**25**	**7**	**28997**

1.3 Annex 3 – LUCAS Survey points by grassland class and survey year

Table 51 - Number of LUCAS survey points extracted for each grassland class by survey year at EU level. The classes PGsn and OG cannot be defined for the survey 2009 and 2012 due to the lack of the INSPIRE Module and the variables accounting for the herbaceous cover (INSPIRE_PLCC4) required by the corresponding rules.

LUCAS survey year	NUTS	NG	OG	PGa	PGsn	PG=PGa+PGsn	TG	# LUCAS survey points
2018	EU28	256,301	16,081	52,668	3,816	56,484	8,988	337,854
2015	EU28	262,668	17,729	47,252	4,559	51,811	6,517	338,725
2012	EU27	281,136	-	46,824	-	46,824	5,956	333,916
2009	EU23	19,796	-	37,311	-	37,311	4,503	261,610

Table 52 - Number of LUCAS survey points (2018) extracted for each grassland class by NUTS 0.

NUTS 0	NG	OG	PGa	PGsn	PG=PGa+PGsn	TG
AT	6,828	512	1,316	40	1,356	144
BE	2,780	66	651	7	658	155
BG	6,088	642	810	58	868	80
CY	1,843	367	54	6	60	43
CZ	4,556	162	836	7	843	152
DE	21,541	284	4,305	45	4,350	602
DK	2,911	108	476	18	494	190
EE	2,183	88	334	4	338	56
EL	9,659	1,045	1,323	311	1,634	284
ES	33,828	4,359	4,404	1,953	6,357	770
FI	14,233	464	855	45	900	585
FR	35,767	1,104	9,659	325	9,984	1,360
HR	3,323	494	360	11	371	51
HU	4,537	170	680	17	697	110
IE	2,208	201	2,485	46	2,531	35
IT	21,589	2,020	2,889	169	3,058	1,627
LT	3,370	160	925	11	936	118
LU	224	1	98	-	98	17
LV	4,058	289	927	5	932	97
MT	55	10	6	-	6	8
NL	3,447	102	1,357	13	1,370	92
PL	18,227	888	3,648	37	3,685	286
PT	5,530	129	1,011	300	1,311	198
RO	10,819	633	4,523	147	4,670	603
SE	23,400	567	1,962	85	2,047	695
SI	1,405	40	421	12	433	44
SK	2,321	102	417	4	421	54
UK	9,571	1,074	5,936	140	6,076	532
EU28	**256,301**	**16,081**	**52,668**	**3,816**	**56,484**	**8,988**

Report

Table 53 - Number of LUCAS survey points (2015) extracted for each grassland class by NUTS 0.

NUTS 0	NG	OG	PGa	PGsn	PG=PGa+PGsn	TG
AT	6727	550	1376	61	1437	125
BE	2127	51	647	1	648	73
BG	6178	491	901	52	953	55
CY	1393	242	40	3	43	48
CZ	4554	137	891	6	897	124
DE	21241	406	4481	42	4523	428
DK	2944	121	382	16	398	202
EE	2204	61	322	5	327	45
EL	9967	974	1133	325	1458	122
ES	37470	5025	4332	2847	7179	607
FI	15007	395	273	4	277	437
FR	35239	2293	9292	222	9514	1142
HR	2728	409	322	14	336	59
HU	4194	244	620	13	633	98
IE	2329	219	2287	54	2341	18
IT	21897	2538	2803	259	3062	1196
LT	3387	112	903	7	910	96
LU	171	2	67		67	11
LV	4277	288	762	2	764	45
MT	50	12	10		10	7
NL	1684	93	703	4	707	37
PL	18161	1076	3408	19	3427	316
PT	7192	116	1305	313	1618	80
RO	12229	183	3884	46	3930	378
SE	24433	663	994	107	1101	451
SI	1546	50	289	9	298	29
SK	2218	113	367	5	372	52
UK	11121	865	4458	123	4581	236
EU28	**262,668**	**17,729**	**47,252**	**4,559**	**51,811**	**6,517**

Table 54 - Number of LUCAS survey points (2012) extracted for each grassland class by NUTS 0.

NUTS 0	NG	PGa	TG
AT	6902	1502	115
BE	1981	551	64
BG	6494	1128	70
CY	1345	57	40

NUTS 0	NG	PGa	TG
CZ	4742	868	120
DE	21898	4459	277
DK	3076	443	127
EE	2191	356	38
EL	11371	957	107
ES	44957	4682	628
FI	15347	331	490
FR	38216	9044	845
HU	4391	653	89
IE	2691	2215	18
IT	24378	2647	1329
LT	3379	1039	75
LU	172	79	8
LV	4360	810	16
MT	66	5	8

NUTS 0	NG	PGa	TG
NL	1790	708	21
PL	19224	3498	342
PT	7659	1264	102
RO	12326	3984	421
SE	25585	807	329
SI	1542	322	20
SK	2388	329	44
UK	12665	4086	213
EU27	**281,136**	**46,824**	**5,956**

Table 55 - Number of LUCAS survey points (2009) extracted for each grassland class by NUTS 0.

NUTS 0	NG	PGa	TG
AT	5784	1210	63
BE	1351	439	14
CZ	3902	682	122
DE	17644	3737	148

NUTS 0	NG	PGa	TG
DK	2132	346	89
EE	2249	388	33
EL	9490	765	134
ES	34699	3338	430
FI	18901	368	686
FR	29681	7422	499
HU	4677	716	119
IE	1647	2533	3
IT	20162	2420	1019
LT	2735	1079	46
LU	107	43	2
LV	3092	711	24
NL	1688	768	4
PL	15294	2889	368
PT	4769	698	80

NUTS 0	NG	PGa	TG
SE	26266	815	430
SI	1171	226	4
SK	2606	382	64
UK	9749	5336	122
EU23	219,796	37,311	4,503

Table 56 - Number of LUCAS survey points for the class PGa by LC1 by survey year at EU-level (2018: EU28, 2015: EU28, 2012: EU27, 2009: EU23). The primary LC classes involved in the PGa rule are: A30 (Other artificial areas), E10 (Grassland with sparse tree/shrub cover), E20 (Grassland without tree/shrub cover) and E30 (Spontaneously re-vegetated surfaces).

NUTS 0	2018				2015			2012			2009		
	A30	E10	E20	E30	E10	E20	E30	E10	E20	E30	E10	E20	E30
AT	6	72	1221	17	62	1280	34	62	1408	32	58	1143	9
BE		30	600	21	42	592	13	42	484	25	22	411	6
BG	2	202	479	127	306	407	188	523	508	97			
CY		2	8	44	4	2	34	6	5	46			
CZ	3	20	793	20	24	826	41	32	775	61	24	627	31
DE	28	102	3944	231	121	4202	158	173	4045	241	163	3527	47
DK	1	23	425	27	7	333	42	12	358	73	14	321	11
EE		28	266	40	26	262	34	43	233	80	29	339	20
EL		252	268	803	284	314	535	310	308	339	249	289	227

ES	17	607	2626	1154	745	2524	1063	1060	2335	1287	756	1941	641
FI	6	47	719	83	16	177	80	18	219	94	18	241	109
FR	25	481	8563	590	748	8208	336	954	7506	584	832	6427	163
HR		68	253	39	46	195	81						
HU	2	38	512	128	48	462	110	81	443	129	105	517	94
IE		54	2413	18	70	2192	25	82	2097	36	84	2434	15
IT		225	1640	1024	397	1543	863	562	1400	685	556	1293	571
LT		37	798	90	54	715	134	126	685	228	71	967	41
LU		9	89		8	59		6	70	3	3	40	
LV	1	34	799	93	22	690	50	58	637	115	48	637	26
MT				6		1	9		1	4			
NL	4	17	1311	25	3	692	8	8	651	49	4	754	10
PL	11	179	3138	320	170	2827	411	231	2629	638	169	2549	171
PT		145	641	225	237	832	236	200	774	290	116	194	388
RO	4	776	3165	578	889	2189	806	898	2516	570			
SE	8	194	1658	102	161	785	48	146	558	103	152	579	84
SI	2	56	356	7	61	226	2	68	245	9	58	161	7
SK		29	368	20	21	322	24	33	261	35	44	306	32
UK	6	112	5490	328	154	4217	87	182	3679	225	235	5049	52
EU				**52668**			**47252**			**46824**			**37311**

1.4 Annex 3 – LUCAS nomenclature

Table 57 – LUCAS Survey nomenclature (2018) adopted for LC.

LC category code	LC category label	LC class code	LC class label
8	8	8	Not relevant
A00	ARTIFICIAL LAND	A11	Buildings with 1 to 3 floors
A00	ARTIFICIAL LAND	A12	Buildings with more than 3 floors
A00	ARTIFICIAL LAND	A13	Greenhouses
A00	ARTIFICIAL LAND	A21	Non built-up area features
A00	ARTIFICIAL LAND	A22	Non built-up linear features
A00	ARTIFICIAL LAND	A30	Other artificial areas
B00	CROPLAND	B11	Common wheat
B00	CROPLAND	B12	Durum wheat
B00	CROPLAND	B13	Barley
B00	CROPLAND	B14	Rye
B00	CROPLAND	B15	Oats
B00	CROPLAND	B16	Maize
B00	CROPLAND	B17	Rice
B00	CROPLAND	B18	Triticale
B00	CROPLAND	B19	Other cereals
B00	CROPLAND	B21	Potatoes
B00	CROPLAND	B22	Sugar beet
B00	CROPLAND	B23	Other root crops
B00	CROPLAND	B31	Sunflower

B00	CROPLAND	B32	Rape and turnip rape
B00	CROPLAND	B33	Soya
B00	CROPLAND	B34	Cotton
B00	CROPLAND	B35	Other fibre and oleaginous crops
B00	CROPLAND	B36	Tobacco
B00	CROPLAND	B37	Other non-permanent industrial crops
B00	CROPLAND	B41	Dry pulses
B00	CROPLAND	B42	Tomatoes
B00	CROPLAND	B43	Other fresh vegetables
B00	CROPLAND	B44	Floriculture and ornamental plants
B00	CROPLAND	B45	Strawberries
B00	CROPLAND	B51	Clovers
B00	CROPLAND	B52	Lucerne
B00	CROPLAND	B53	Other leguminous and mixtures for fodder
B00	CROPLAND	B54	Mixed cereals for fodder
B00	CROPLAND	B55	Temporary grasslands
B00	CROPLAND	B71	Apple fruit
B00	CROPLAND	B72	Pear fruit
B00	CROPLAND	B73	Cherry fruit
B00	CROPLAND	B74	Nuts trees
B00	CROPLAND	B75	Other fruit trees and berries
B00	CROPLAND	B76	Oranges
B00	CROPLAND	B77	Other citrus fruit

Report

B00	CROPLAND	B81	Olive groves
B00	CROPLAND	B82	Vineyards
B00	CROPLAND	B83	Nurseries
B00	CROPLAND	B84	Permanent industrial crops
B00	CROPLAND	BX1	Arable land (only in case pi)
B00	CROPLAND	BX2	Permanent crops (only in case pi)
C00	WOODLAND	C10	Broadleaved woodland
C00	WOODLAND	C21	Spruce dominated coniferous woodland
C00	WOODLAND	C22	Pine dominated coniferous woodland
C00	WOODLAND	C23	Other coniferous woodland
C00	WOODLAND	C31	Spruce dominated mixed woodland
C00	WOODLAND	C32	Pine dominated mixed woodland
C00	WOODLAND	C33	Other mixed woodland
D00	SHRUBLAND	D10	Shrubland with sparse tree cover
D00	SHRUBLAND	D20	Shrubland without tree cover
E00	GRASSLAND	E10	Grassland with sparse tree/shrub cover
E00	GRASSLAND	E20	Grassland without tree/shrub cover
E00	GRASSLAND	E30	Spontaneously vegetated surfaces
F00	BARE LAND AND LICHENS/MOSS	F10	Rocks and stones
F00	BARE LAND AND LICHENS/MOSS	F20	Sand
F00	BARE LAND AND LICHENS/MOSS	F30	Lichens and moss
F00	BARE LAND AND LICHENS/MOSS	F40	Other bare soil
G00	WATER AREAS	G11	Inland fresh water bodies

G00	WATER AREAS	G12	Inland salty water bodies
G00	WATER AREAS	G21	Inland fresh running water
G00	WATER AREAS	G22	Inland salty running water
G00	WATER AREAS	G30	Transitional water bodies
G00	WATER AREAS	G50	Glaciers, permanent snow
H00	WETLANDS	H11	Inland marshes
H00	WETLANDS	H12	Peatbogs
H00	WETLANDS	H21	Salt marshes
H00	WETLANDS	H22	Salines and other chemical deposits
H00	WETLANDS	H23	Intertidal flats

Table 58 – LUCAS Survey nomenclature (2018) adopted for LU.

LU category code	LC category label	LU class code	LU class label
8	8	8	Not relevant
U100	PRIMARY SECTOR	U111	Agriculture (excluding fallow land and kitchen gardens)
U100	PRIMARY SECTOR	U112	Fallow land
U100	PRIMARY SECTOR	U113	Kitchen garden
U100	PRIMARY SECTOR	U120	Forestry
U100	PRIMARY SECTOR	U130	Aquaculture and fishing
U100	PRIMARY SECTOR	U140	Mining and quarrying
U100	PRIMARY SECTOR	U150	Other primary production
U200	SECONDARY SECTOR	U210	Energy production
U200	SECONDARY SECTOR	U221	Manufacturing of food, beverages and tobacco products

U200	SECONDARY SECTOR	U221	Manufacturing of food, beverages and tobacco products
U200	SECONDARY SECTOR	U222	Manufacturing of textile products
U200	SECONDARY SECTOR	U222	Manufacturing of textile products
U200	SECONDARY SECTOR	U222	Manufacturing of textile products
U200	SECONDARY SECTOR	U223	Coal, oil and metal processing
U200	SECONDARY SECTOR	U223	Coal, oil and metal processing
U200	SECONDARY SECTOR	U223	Coal, oil and metal processing
U200	SECONDARY SECTOR	U224	Production of non-metal mineral goods
U200	SECONDARY SECTOR	U224	Production of non-metal mineral goods
U200	SECONDARY SECTOR	U225	Chemical and allied industries and manufacturing
U200	SECONDARY SECTOR	U225	Chemical and allied industries and manufacturing
U200	SECONDARY SECTOR	U225	Chemical and allied industries and manufacturing
U200	SECONDARY SECTOR	U226	Machinery and equipment
U200	SECONDARY SECTOR	U226	Machinery and equipment
U200	SECONDARY SECTOR	U226	Machinery and equipment
U200	SECONDARY SECTOR	U227	Wood based products
U200	SECONDARY SECTOR	U227	Wood based products
U200	SECONDARY SECTOR	U228	Printing and reproduction
U200	SECONDARY SECTOR	U228	Printing and reproduction
U300	TERTIARY SECTOR, TRANSPORT, UTILITIES & RESIDENTIAL	U311	Railway transport
U300	TERTIARY SECTOR, TRANSPORT, UTILITIES & RESIDENTIAL	U312	Road transport
U300	TERTIARY SECTOR, TRANSPORT, UTILITIES & RESIDENTIAL	U313	Water transport

U300	TERTIARY SECTOR, TRANSPORT, UTILITIES & RESIDENTIAL	U314	Air transport
U300	TERTIARY SECTOR, TRANSPORT, UTILITIES & RESIDENTIAL	U315	Transport via pipelines
U300	TERTIARY SECTOR, TRANSPORT, UTILITIES & RESIDENTIAL	U315	Transport via pipelines
U300	TERTIARY SECTOR, TRANSPORT, UTILITIES & RESIDENTIAL	U316	Telecommunication
U300	TERTIARY SECTOR, TRANSPORT, UTILITIES & RESIDENTIAL	U317	Logistics and storage
U300	TERTIARY SECTOR, TRANSPORT, UTILITIES & RESIDENTIAL	U318	Protection infrastructures
U300	TERTIARY SECTOR, TRANSPORT, UTILITIES & RESIDENTIAL	U319	Electricity, gas and thermal power distribution
U300	TERTIARY SECTOR, TRANSPORT, UTILITIES & RESIDENTIAL	U321	Water supply and treatment
U300	TERTIARY SECTOR, TRANSPORT, UTILITIES & RESIDENTIAL	U322	Waste treatment
U300	TERTIARY SECTOR, TRANSPORT, UTILITIES & RESIDENTIAL	U330	Construction
U300	TERTIARY SECTOR, TRANSPORT, UTILITIES & RESIDENTIAL	U341	Commerce
U300	TERTIARY SECTOR, TRANSPORT, UTILITIES & RESIDENTIAL	U342	Financial, professional and information services
U300	TERTIARY SECTOR, TRANSPORT, UTILITIES & RESIDENTIAL	U350	Community services
U300	TERTIARY SECTOR, TRANSPORT, UTILITIES & RESIDENTIAL	U361	Amenities, museums, leisure
U300	TERTIARY SECTOR, TRANSPORT, UTILITIES & RESIDENTIAL	U362	Sport

U300	TERTIARY SECTOR, TRANSPORT, UTILITIES & RESIDENTIAL	U370	Residential
U400	UNUSED AND ABANDONED AREAS	U411	Abandoned industrial areas
U400	UNUSED AND ABANDONED AREAS	U412	Abandoned commercial areas
U400	UNUSED AND ABANDONED AREAS	U413	Abandoned transport areas
U400	UNUSED AND ABANDONED AREAS	U414	Abandoned residential areas
U400	UNUSED AND ABANDONED AREAS	U415	Other abandoned areas
U400	UNUSED AND ABANDONED AREAS	U420	Semi-natural and natural areas not in use