



LUCAS cover photos 2006–2018 over the EU: 874 646 spatially distributed geo-tagged close-up photos with land cover and plant species label

Raphaël d’Andrimont¹, Momchil Yordanov¹, Laura Martinez-Sanchez¹, Peter Haub², Oliver Buck³, Carsten Haub³, Beatrice Eiselt⁴, and Marijn van der Velde¹

¹European Commission, Joint Research Centre (JRC), Ispra, Italy

²Imaging Consulting, Altlussheim, Germany

³EFTAS, EFTAS Fernerkundung Technologietransfer GmbH, Münster, Germany

⁴European Commission, Eurostat (ESTAT), Luxembourg, Luxembourg

Correspondence: Raphaël d’Andrimont (raphael.dandrimont@ec.europa.eu) and Marijn van der Velde (marijn.van-der-velde@ec.europa.eu)

Received: 4 April 2022 – Discussion started: 15 June 2022

Revised: 24 August 2022 – Accepted: 6 September 2022 – Published: 30 September 2022

Abstract. In the European Union, since 2006, a tri-annual survey has sampled land cover and land use information under the Land Use/Cover Area frame Survey (LUCAS). A total of 1 351 293 observations from 651 780 unique locations for 106 variables were collected during the five LUCAS surveys, including a total of 5.4 million landscape photos representing the observer view in the four cardinal directions as well as point photos showing the actual surveyed point. In addition to these previously published photos, additional LUCAS cover photos were recorded, showing a close-up view and thus more detail of the sampled tree, crop and plant species. Between 2006 and 2018, 875 661 LUCAS cover close-up photos were collected, showing the relevant land cover and plant species on the entire photo with the absence of any other elements of the landscape in the frame. Photos containing potential privacy content were identified following a two-stage deep learning anonymisation process, resulting in the removal of 346 photos before publication. This paper summarises the collection of LUCAS cover photos and the filtering for mandatory privacy issues and also provides links to download the data along with the photo metadata and cross-links to the corresponding LUCAS-harmonised survey data. Moreover, after presenting the final public and open dataset consisting of 874 646 photos, potential applications relying on recent advances in geo-spatial analysis and statistical learning, such as large-scale biodiversity monitoring, are discussed. The data are available on the JRC Open Data Catalogue: <http://data.europa.eu/89h/c83906d7-1817-495f-b4ba-f2e4b8859d48> (European Commission, Joint Research Centre, 2022).

1 Introduction

In the European Union (EU), a tri-annually surveyed sample of land cover and land use has been collected since 2006 under the Land Use/Cover Area frame Survey (LUCAS) (Gallego and Delincé, 2010). LUCAS was carried out in 2006, 2009, 2012, 2015, and 2018 and is planned for 2022. During the five campaigns already carried out, a total of 1 351 293 point surveys at 651 780 unique locations were performed

along with 5.4 million landscape photos. On each of these surveyed points, depending on the year, observations of up to 109 variables were recorded. The combination of the information collected in the five LUCAS surveys has resulted in the most comprehensive in situ database on land cover and land use in the EU (d’Andrimont et al., 2020).

In addition to the landscape and point photos already published (d’Andrimont et al., 2020), other specific photos were taken, including the LUCAS cover photos, which offer a

close-up view of the land cover on which plant species should be identifiable. These photos were not taken to be published but to support simultaneous visual quality control alongside the field survey. Between 2006 and 2018, 875 661 of such LUCAS cover photos were collected. However, as this specific LUCAS cover photo dataset was not designed as an output of the survey, it has not been published yet.

The objective of this paper is to make this rich dataset available in analysis-ready form to the research community for various use cases. The prerequisite for using the LUCAS cover data and photos in other applications (e.g. biodiversity monitoring or machine readable calibration sources for EO) requires organising, curating, documenting and publishing the photos following FAIR (findability, accessibility, interoperability, and reuse) principles (Wilkinson et al., 2016). This paper summarises the collection of LUCAS cover photos and the filtering for mandatory privacy issues and also provides links to download the data along with the photo metadata and cross-links to the corresponding LUCAS-harmonised survey data.

2 In-situ LUCAS survey protocol

LUCAS is a two-phase sample survey. The first sample is a systematic selection of points on a grid with a 2 km spacing in eastings and northings, covering the whole EU territory (Gallego and Bamps, 2008). Currently, it includes around 1.1 million points (Fig. 1) and is referred to as the master sample. Each point of the first-phase sample is classified into one of 10 land-cover classes via visual interpretation of ortho-photos or satellite images (ESTAT, 2018). Then a stratified sample is selected to obtain the desired statistically representative spatial distribution of sampled land cover classes according to the first-phase visual interpretation (European Commission et al., 2018).

3 LUCAS cover photo collection protocol

As described in Eurostat (2018), LUCAS surveyors document their observations in several sets of photos. The type of photo to be taken depends on the type of observation, the land cover, the presence or absence of water management, the need to collect a soil sample, and the need to document conflicting cases. Photos are taken for each observed in situ point covering the actual point (P) and the four cardinal direction views of north, east, south, and west (N, E, S, W) (examples of these are the five first photos from left to right in Fig. 2). This P, N, E, S, W photo dataset corresponds to 5 440 459 photos for the five surveys. These photos are publicly available for download along with an EXIF database in d'Andrimont et al. (2020), containing image metadata and an explanation as to the difference between landscape and point photos. The background image is from Map data © Google 2020.

However, as described in Eurostat (2018), other non-publicly available photos were taken. Among these photos, the LUCAS cover (C) photos were collected mainly from croplands (class B), woodlands (class C), shrublands (class D), and grasslands (class E). The aim of these cover photos is to enable the identification of the recorded crops and plants during simultaneous quality controls in the office by means of the photo on screen (Eurostat, 2018). The cover photo should be taken at a close distance, so that the structure of, for example, leaves, barks, flowers or fruits can be clearly seen. See an example in Fig. 2 on the right.

4 Photo metadata extraction

The LUCAS cover photos were obtained from the Eurostat archive via portable hard drives. The photos' metadata were then extracted with the ExifTool (v 12.10) (Harvey and Körtner, 2016), resulting in a database of photos that was compared for completeness with the survey data records. The EXIF metadata were extracted for 82 fields (Table 1). Finally, the LUCAS cover EXIF table was joined to the LUCAS-harmonised database to provide all survey information into one unique table.

5 Automated identification of photos with potential privacy content

5.1 Check for the presence of privacy content and manual anonymisation

According to the guidelines of the LUCAS project, it must be ensured that no private content is included in the published images. This applies in particular to vehicle registration plates and recognisable persons and faces, which have to be blurred or removed in the photos. Since this anonymisation requirement applies equally to the previously unpublished cover photos, all cover photos must be checked for the presence of private content.

In order to fulfil this essential quality requirement for the image data, the checking of the images was carried out in a purely manual capacity in previous LUCAS campaigns via a visual inspection of the photos. To reduce the manual effort and to improve the anonymisation quality, an automated procedure was used for the first time to support the image anonymisation process in the 2018 LUCAS campaign.

The method developed during the 2018 survey is based on the highly efficient convolution neural network (CNN) YOLO (Redmon and Farhadi, 2018). This neural network enables the recognition of a large number of different object classes as well as multiple objects per image at a very high speed.

In parallel to the established manual control procedure, the CNN approach was independently tested. It confirmed that YOLO ideally fulfils the basic requirements for the task of pre-classification and can be used as a binary classifier

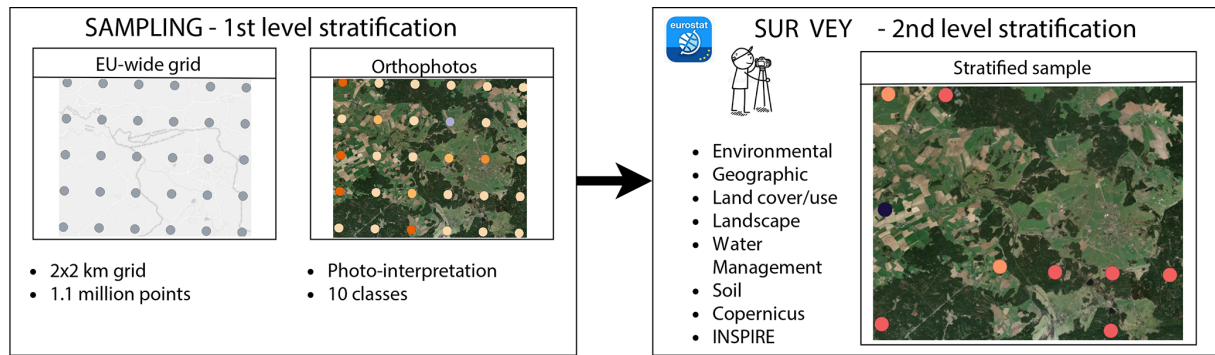


Figure 1. Schematic overview of the LUCAS and harmonisation methodologies. This illustrates the sampling at the basis of the production of the LUCAS primary data. LUCAS cover photos are close-up photos originally collected to support the surveyors' interpretation and control.



Figure 2. Example of LUCAS photos collected on a LUCAS point located in France in a barley field (lat: 48.1645, long: -2.4970). For each LUCAS point, photos are collected for north, east, south, west, point, and cover.

with the classes (1) anonymisation potentially necessary and (2) anonymisation not necessary.

Essential for such an approach is the guarantee of a low false negative rate for the automated binary pre-classification by the CNN, i.e. the acceptance of a low specificity during the first step of this two-step procedure. This was achieved by a suitable choice of object classes from the pool of all available classes and the use of a suitable detection threshold (0.1). The subsequent manual control in the second step ensures an almost vanishing false positive rate and thus a high sensitivity (true positive rate, recall or hit rate).

Thus, this two-step approach maximises specificity and sensitivity and achieves a very high overall accuracy of image anonymisation. The cover photos analysed in this study do not differ technically from the previously anonymised LUCAS images, and thus the described procedure could be applied directly to the cover photos without modification.

5.2 Results of the anonymisation check

This two-step anonymisation approach from the 2018 campaign was also applied to the LUCAS cover photos from the 2006–2018 campaigns. Table 2 provides an overview of the photo anonymisation checks for all LUCAS campaigns. In total, 875 661 were processed, out of which 66 were corrupted, meaning they failed to ingest. In turn, 16 880 photos or 1.92 % were flagged by the CNN as containing elements with potential anonymisation need. Finally, a total of 346 from the previously flagged set were manually short-

listed as actually containing elements worth anonymising for. Added to these are the 604 that did not have a match within the LUCAS-harmonised product (d'Andrimont et al., 2020). The difference between columns 5 (i.e. first step) and 6 (i.e. second step) of Table 2 shows that the number of photos containing potential non-anonymised elements was reduced by ca. 97.5–99 % for each campaign, i.e. roughly only 1–2.5 % of the photos have to be checked visually. Still, within this reduced amount, most photos will not contain elements that need to be anonymised. The main known reasons why photos are marked as containing potential non-anonymised elements while not containing any are listed below:

1. *Photo is already anonymised.* During the LUCAS campaigns, strict anonymisation procedures were in place to cover persons and car plates with white bars. Despite this, the neural network will detect a car as a car, even if its car plate is already correctly anonymised, i.e. covered with a white bar.
2. *Photo shows only very small elements.* Very small elements – i.e. cars or persons in the background of an image – that are not recognisable do not have to be anonymised. Nevertheless, the neural network will detect these small cars, trucks, persons, etc. and thus mark the photos as containing potential non-anonymised elements.
3. *Photo shows only a non-recognisable part of a car or person.* The neural network detects objects even if

Table 1. The LUCAS cover data set is provided with two tables: the EXIF table, with 82 metadata attributes extracted from the photos, and the LUCAS-harmonised table containing 121 attributes.

Origin of attributes	#	Attribute names
EXIF fields	82	ApertureValue, BrightnessValue, ColorSpace, ComponentsConfiguration, CompressedBitsPerPixel, Contrast, Copyright, CustomRendered, DateTime, DateTimeDigitized, DateTimeOriginal, DeviceSettingDescription, DigitalZoomRatio, ExifImageLength, ExifImageWidth, ExifOffset, ExifVersion, ExposureBiasValue, ExposureIndex, ExposureMode, ExposureProgram, ExposureTime, FileSource, Flash, FlashPixVersion, FNumber, FocalLength, FocalLengthIn35mmFilm, FocalPlaneResolutionUnit, FocalPlaneXResolution, FocalPlaneYResolution, GainControl, Gamma, GPSAltitude, GPSAltitudeRef, GPSDate, GPSInfo, GPSPLatitude, GPSPLatitudeRef, GPSPLongitude, GPSPLongitudeRef, GPSPMapDatum, GPSPSatellites, GPSPTimeStamp, GPSPVersionID, ImageDescription, InteroperabilityIndex, InteroperabilityOffset, InteroperabilityVersion, ISOSpeedRatings, LightSource, Make, MaxApertureValue, MeteringMode, Model, OECF, OffsetSchema, Orientation, Padding, PrimaryChromaticities, Rating, RelatedImageLength, RelatedImageWidth, ResolutionUnit, Saturation, SceneCaptureType, SceneType, SensingMethod, Sharpness, ShutterSpeedValue, Software, SubjectDistanceRange, WhiteBalance, WhitePoint, XResolution, YCbCrCoefficients, YCbCrPositioning, YResolution, year, pointid, file_path_ftp_cover, id
LUCAS-harmonised fields	121	id, point_id, year, nuts0, nuts1, nuts2, nuts3, th_lat, th_long, office_pi, ex_ante, survey_date, car_latitude, car_ew, car_longitude, gps_proj, gps_prec, gps_altitude, gps_lat, gps_ew, gps_long, obs_dist, obs_direct, obs_type, obs_radius, letter_group, lc1, lc1_label, lc1_spec, lc1_spec_label, lc1_perc, lc2, lc2_label, lc2_spec, lc2_spec_label, lc2_perc, lu1, lu1_label, lu1_type, lu1_type_label, lu1_perc, lu2, lu2_label, lu2_type, lu2_type_label, lu2_perc, parcel_area_ha, tree_height_maturity, tree_height_survey, feature_width, lm_stone_walls, crop_residues, lm_grass_margins, grazing, special_status, lc_lu_special_remark, cprn_cando, cprn_lc, cprn_lc_label, cprn_lc1n, cprnc_lc1e, cprnc_lc1s, cprnc_lc1w, cprn_lc1n_brdth, cprn_lc1e_brdth, cprn_lc1s_brdth, cprn_lc1w_brdth, cprn_lc1n_next, cprn_lc1s_next, cprn_lc1e_next, cprn_lc1w_next, cprn_urban, cprn_impervious_perc, inspire_plcc1, inspire_plcc2, inspire_plcc3, inspire_plcc4, inspire_plcc5, inspire_plcc6, inspire_plcc7, inspire_plcc8, eunis_complex, grassland_sample, grass_cando, wm, wm_source, wm_type, wm_delivery, erosion_cando, soil_stones_perc, bio_sample, soil_bio_taken, bulk0_10_sample, soil_blk_0_10_taken, bulk10_20_sample, soil_blk_10_20_taken, bulk20_30_sample, soil_blk_20_30_taken, standard_sample, soil_std_taken, organic_sample, soil_org_depth_cando, soil_taken, soil_crop, photo_point, photo_north, photo_south, photo_east, photo_west, transect, revisit, th_gps_dist, file_path_gisco_north, file_path_gisco_south, file_path_gisco_east, file_path_gisco_west, file_path_gisco_point, gps_geom, th_geom, trans_geom, file_path_ftp_cover

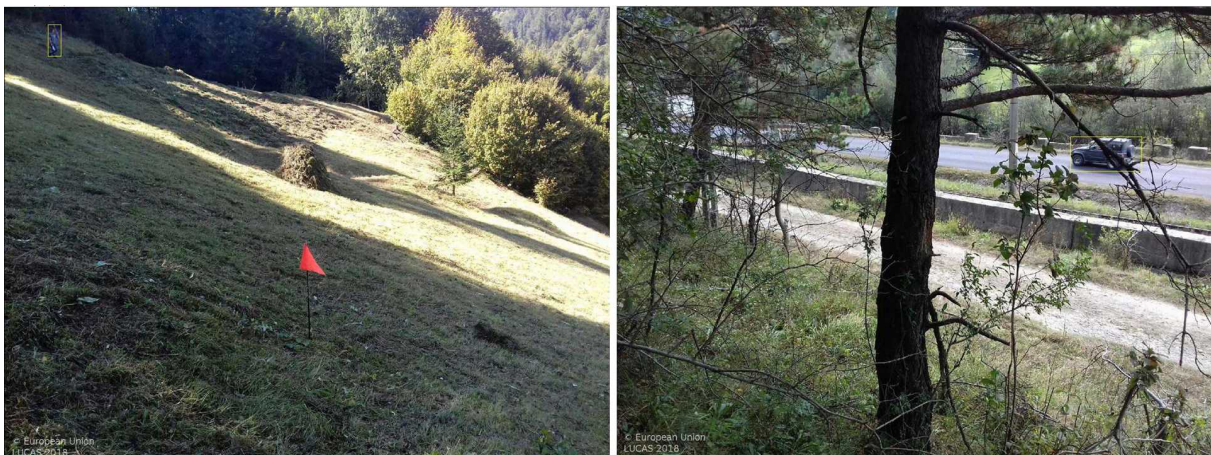


Figure 3. Examples showing correct identification of potential anonymisation elements (human person left and car right) that needed no anonymisation due to lacking face or number plate recognition.

they are only partially visible. It therefore marks photos as containing potential non-anonymised objects even if only a hand, an arm or a car door, which does not have to be anonymised, is visible.

4. *Wrongly classified objects.* The neural network wrongly classified images as containing a certain element that they did not, e.g. an animal or plant, classified as a person. The priority of the classification process was to miss no (or as few as possible) non-anonymised objects. Therefore, the object threshold score was set to a very low value of 0.1 to avoid missing any non-anonymised elements. Albeit ensuring this, the very low threshold score also delivers wrongly classified objects.

In total, 1016 images were removed from the original source set, including the ones from columns “no harmo”, “second step”, and “corrupted”.

6 Harmonisation of the filename and watermark

A number of changes had to be done to the filename and directory tree in order to harmonise both between the survey years. Namely, certain years had listed countries with a different country code (e.g. “GR” instead of “EL” or “GB” instead of “UK”), filenames had used an upper or a lowercase for either the “c” to indicate the “cover” status of the LUCAS image, or there were differences in the “.jpg” extension. To coherently catalogue this, all images were renamed to fit the convention “LUCASYYYY_PointID_Cover.jpg”, where YYYYY is a placeholder for the year of the survey. Additionally, watermarks were added to the 2009 and 2012 images in cases where these were lacking.

7 Final data overview

The distribution of LUCAS cover photos per land cover and per year is shown in Fig. 5. The distribution per country is presented in Table 3. For each survey year, a number ranging from 107 022 (in 2006) to 217 609 (in 2015) were collected, totalling 874 666 for the five surveys. The sampling of LUCAS aims to revisit some of the points in successive surveys, resulting thus in point revisits for each point surveyed ranging from 1 to 5, as shown in Fig. 4. In Fig. 6, one random photo example is shown for each land cover class.

8 Limitations

Several limitations inherent to the dataset are briefly discussed. The first limitation is linked to the survey protocol, which does not detail the field of view required to take the photos. Indeed, the instruction to the surveyor is to make sure the plant can be recognised on the picture, which is subjective to the surveyor. As illustrated in Fig. 6, the photos could

be taken with a diversity of views: whole plant views, landscape views (sometimes with sky), plants with artificial background (red support), or plants with bare soil. Also, sometimes, the surveyor takes the plant from the soil, and the pictures thus contain the root system and vegetative organs. Future surveys could consider specifically collecting information on the view type. Applications that are dedicated to recognising plant species (e.g. Pl@ntNet) ask the surveyor to select the type of view as leaf, flower, fruit, bark, whole plant, or other (Goëau et al., 2013).

Another limitation is the lack of EXIF information for some photos, as highlighted in Table 2. Also, the quality of the EXIF data collected, when available, depends on the quality of the sensors and its calibration. Standardising the type of sensors used to collect the images would greatly facilitate this routine and the uptake of the data. Another option would be to hardcode some variables, such as looking direction from time of day and year and the angle of shadows when sunny, when possible.

9 Potential use of the data and perspectives

The LUCAS cover dataset with systematically sampled geolocated observations and photos of crops, trees, shrubs, grasses, and other plants can be the source for different uses and drive the development of various applications. The specific advantages are (1) the sample design, where the regular systematic 2 km LUCAS grid ensures an exhaustive EU-wide coverage; (2) the observations were done over a period of 15 years, with several of the points having been revisited up to 5 times, providing a unique historical perspective; (3) the photos are annotated with a label following the LUCAS legend; and (4) while this label may not be precise enough for various applications, computer-vision-based methods could extract information from the image and enrich the label. Indeed, LUCAS has been designed to collect statistics about land use and land cover, and specific applications will have different needs. The precision of the legend is, for example, not sufficient for botanical applications that need species-level information on observed plants. However, the dataset could provide training data to build deep learning convolutional neural networks to recognise and classify trees, plants, and crop types along with their phenological stages on photos, such as in d'Andrimont et al. (2022b).

Indeed, recent advances in combining photo sources from citizens and with those from experts in combination with computer vision are simultaneously enabling species identification and the gathering of occurrence data (e.g. Pl@ntNet (Goëau et al., 2018), iNaturalist (Nugent, 2018), or Flora Incognita (Mäder et al., 2021)). The collection of such geolocated plant species occurrences contributes to collaborative data platforms such as the Global Biodiversity Information Facility (GBIF). This can be complementary to long-term and high-quality but resource-intensive biodiversity assess-

Table 2. Results of the photo screening with potential anonymisation issues. The total number of LUCAS cover images on disk is specified by the term *source*; *no harmo* indicates LUCAS cover images on disk that do not have a corresponding row in the LUCAS-harmonised product (d'Andrimont et al., 2020); *no EXIF* is for images on disk that do not have any EXIF information encoded; *first step* are images flagged by the YOLO network; *second step* are images flagged manually from the YOLO set as having some element that is subject to anonymisation; *corrupted* are the images that have a visual distortion that makes the image unusable; the final column shows the images per year that are part of the published set.

Year	Total number of photos surveyed			Flagged photos			Total number of photos published
	Source	No harmo	No EXIF	First step	Second step	Corrupted	
2006	107 140	54	3	1007	63	1	107 022
2009	150 125	39	9559	2239	80	13	149 993
2012	204 944	0	9652	3930	88	50	204 806
2015	217 638	0	1654	4654	28	1	217 609
2018	195 814	511	2085	5050	87	1	195 216
Total	875 661	604	22 953	16 880	346	66	874 646

Table 3. Number of LUCAS cover photos per country and per year.

	2006	2009	2012	2015	2018	Total
AT	0	3571	5249	4965	3659	17 444
BE	1898	1066	1856	1907	2187	8914
BG	0	0	5794	5226	4977	15 997
CY	0	0	1045	1127	1136	3308
CZ	4568	4230	5097	5068	4779	23 742
DE	5311	16 333	21 497	21 364	19 765	84 270
DK	0	1881	2857	3026	2610	10 374
EE	0	1507	1773	1806	1418	6504
EL	0	4461	5938	6202	6881	23 482
ES	23 451	18 762	28 447	28 623	25 525	124 808
FI	0	10 230	9281	10 422	6545	36 478
FR	29 187	22 614	31 005	32 674	31 570	147 050
HR	0	0	0	2674	2053	4727
HU	6657	4377	4124	4047	3276	22 481
IE	0	1414	2336	2495	2505	8750
IT	12 315	10 345	15 086	15 230	15 131	68 107
LT	0	2780	3491	3420	2608	12 299
LU	156	133	178	181	210	858
LV	0	2543	3341	3669	2826	12 379
MT	0	0	43	50	48	141
NL	2479	1652	1609	1693	3122	10 555
PL	18 433	14 709	19 059	18 963	18 064	89 228
PT	0	3848	6018	5714	4624	20 204
RO	0	0	641	10 385	8473	19 499
SE	0	13 624	16 876	14 646	9556	54 702
SI	0	1036	1381	1440	1382	5239
SK	2567	2230	2066	2101	1760	10 724
UK	0	6642	8709	8475	8526	32 352
NOT EU	0	5	9	16	0	30
Total	107 022	149 993	204 806	217 609	195 216	874 646

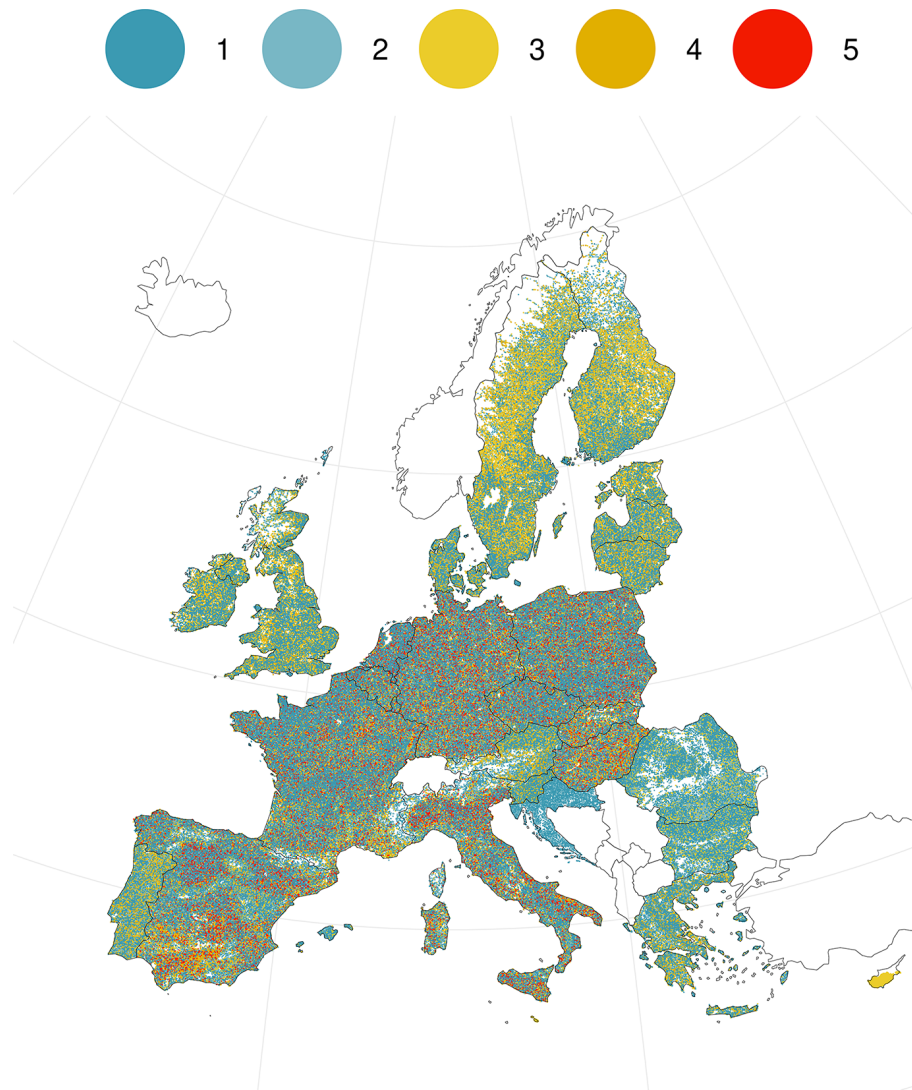


Figure 4. LUCAS cover photos' geographical distribution and number of (re)visits to each survey over the five surveys between 2006 and 2018. Visits range from 1 to 5. Map projection: EPSG 3035.

ments by professional botanists (Miller-Rushing et al., 2012). The photos in this new LUCAS cover photo dataset could be ingested in such applications. In fact, the LUCAS cover photos of crops (letter group “B”) are currently used to generate a specific application within PI@ntNet to recognise crops.

10 Data availability

This section describes the dataset provided along with this article, including the tables and photos. The data are available at https://jeodpp.jrc.ec.europa.eu/ftp/jrc-opendata/LUCAS/LUCAS_COVER/, (d'Andrimont et al., 2022a).

1. *Photos.* The 874 646 LUCAS cover photos are available on the FTP, downloadable here: https://jeodpp.jrc.ec.europa.eu/ftp/jrc-opendata/LUCAS/LUCAS_COVER/

(d'Andrimont et al., 2022a). The directory tree for the photo database follows the standard of the original EUROSTAT data. It is thereby organised by starting from the folder shown in the provided link, descending into a folder specifying the year of the survey (“LUCASYYYYY”), followed by folders of each respective NUTS0 code that the point is located in, in turn preceded by folders of the first three digits and the second three digits of the point ID. The JPG files that constitute the LUCAS cover photos are to be found in the folder named after the second three digits of the point ID according to the convention described in Sect. 6.

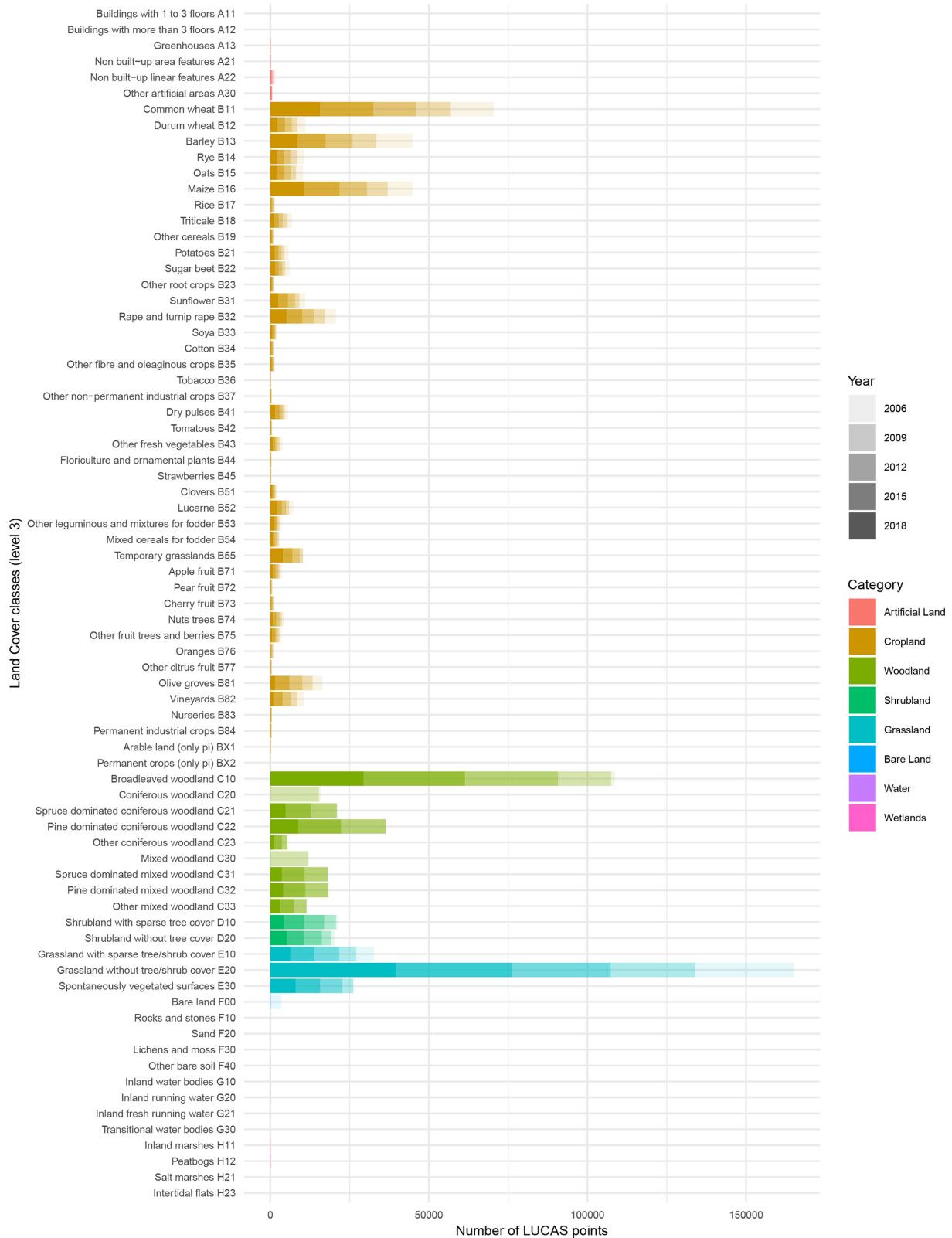


Figure 5. Distribution of LUCAS cover photos in land cover classes in the multi-year harmonised LUCAS database. In cases where survey years are not present, please orientate oneself with reference to adjacent classes of the same colour. Counting for the distribution of each class begins at 2018 and ends with 2006 due to the relative abundance of 2018 in terms of classes compared to other years.



Figure 6. Examples of LUCAS cover photos for all classes. The land cover class is shown in the top left corner of the image in white. The letter at the top of the row represents the LUCAS level 1 label (variable *letter_group*). See Fig. 5 for label correspondences.

2. *Tables.* (EXIF and LUCAS-harmonised) https://jeodpp.jrc.ec.europa.eu/ftp/jrc-opendata/LUCAS/LUCAS_COVER/tables

- *LUCAS cover EXIF table.* (*lucas_cover_exif.csv*). The table contains 82 variables described in Table 1.
- *LUCAS-harmonised cover table.* (*lucas_harmo_cover_attr.csv*). The table contains 121 variables described in Table 1.

In addition to photos' availability on this FTP, the photos are available on the GISCO platform (https://gisco-services.ec.europa.eu/lucas/photos/*/*/*/*C.jpg (last access: 29 September 2022), where “*” is variable and represents the subdirectory and name of the photos).

11 Conclusions

The LUCAS surveys have resulted in the most comprehensive in situ database on land cover and land use in the EU. While close-up photos of the land cover had been collected for most of the in situ points, they had never been published. This data paper represents an effort to organise, document, curate, and publish this dataset following FAIR principles. This resulted in 874 646 geo-located photos along with surveyed information on land cover and land use following LUCAS legend level 2, inherited from the attribute information

of the LUCAS-harmonised product. The LUCAS cover photos and dataset can feed various applications and developments relying on recent advances in geo-spatial analysis and statistical learning.

Author contributions. All the authors processed and analysed the data, wrote the paper, and provided comments and suggestions on the manuscript. ESTAT designed the survey methodology. EF-TAS and Imaging Consulting carried out the anonymisation process of the LUCAS cover photos. Rd'A, MY, LMS, and MvdV processed and analysed the data and wrote the paper. All authors provided suggestions to the manuscript. BE and colleagues at ESTAT designed the survey methodology. PH, OB, and CH carried out the anonymisation process of the LUCAS cover photos.

Competing interests. The contact author has declared that none of the authors has any competing interests.

Disclaimer. Publisher's note: Copernicus Publications remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Acknowledgements. The authors would like to express their appreciation to all that have been involved in the LUCAS surveys and especially to the LUCAS surveyors for their careful observations that underpin the value of the resulting LUCAS datasets.

Review statement. This paper was edited by Francesco N. Tubiello and reviewed by two anonymous referees.

References

- d'Andrimont, R., Yordanov, M., Martinez-Sanchez, L., Eiselt, B., Palmieri, A., Dominici, P., Gallego, J., Reuter, H. I., Joebges, C., Lemoine, G., and van der Velde, M.: Harmonised LUCAS in-situ land cover and use database for field surveys from 2006 to 2018 in the European Union, *Sci. Data*, 7, 1–15, <https://doi.org/10.1038/s41597-020-00675-z>, 2020.
- d'Andrimont, R., Yordanov, M., Martinez Sanchez, L., Haub, P., Buck, O., Haub, C., Eiselt, B., and Van Der Velde, M.: LUCAS Cover 2006–2018, European Commission [data set], <http://data.europa.eu/89h/c83906d7-1817-495f-b4ba-f2e4b8859d48> (last access: 27 September 2022), 2022a.
- d'Andrimont, R., Yordanov, M., Martinez-Sanchez, L., and van der Velde, M.: Monitoring crop phenology with street-level imagery using computer vision, *Comput. Electron. Agr.*, 196, 106866, <https://doi.org/10.1016/j.compag.2022.106866>, 2022b.
- ESTAT: Technical reference document S1: Stratification Guidelines, https://ec.europa.eu/eurostat/documents/205002/7329820/LUCAS2018_S1-StratificationGuidelines_20160523.pdf (last access: 27 September 2022), 2018.
- European Commission, Joint Research Centre (JRC): LUCAS Cover 2006–2018, European Commission, Joint Research Centre (JRC) [data set], <http://data.europa.eu/89h/c83906d7-1817-495f-b4ba-f2e4b8859d48> (last access: 29 September 2022), 2022.
- European Commission, Eurostat, Ballin, M., Barcaroli, G., Maselli, M., and Scarno, M.: Redesign sample for Land Use/Cover Area frame Survey (LUCAS) 2018, 2018 edition, Publications Office, <https://doi.org/10.2785/132365>, 2018.
- Eurostat: Technical reference document C-1: Instructions for surveyors, https://jeodpp.jrc.ec.europa.eu/ftp/jrc-opendata/LUCAS/LUCAS_harmonised/3_supporting/LUCAS2018-C1-Instructions.pdf (last access: 29 September 2022), 2018.
- Gallego, J. and Bamps, C.: Using CORINE land cover and the point survey LUCAS for area estimation, *Int. J. Appl. Earth Obs.*, 10, 467–475, 2008.
- Gallego, J. and Delincé, J.: The European land use and cover area-frame statistical survey, *Agricultural survey methods*, 149–168, <https://doi.org/10.1002/9780470665480.ch10>, 2010.
- Goëau, H., Bonnet, P., Joly, A., Bakić, V., Barbe, J., Yahiaoui, I., Selmi, S., Carré, J., Barthélémy, D., Boujemaa, N., Molino, J.-F., Duché, G., and Péronnet, A.: Pl@ntNet mobile app, in: *Proceedings of the 21st ACM international conference on Multimedia*, Barcelona, Spain, 21–25 October 2013, 423–424, <https://doi.org/10.1145/2502081.2502251>, 2013.
- Goëau, H., Joly, A., Bonnet, P., Lasseck, M., Šulc, M., and Hang, S. T.: Deep learning for plant identification: how the web can compete with human experts, *Biodiversity Information Science and Standards*, 2, e25637, <https://doi.org/10.3897/biss.2.25637>, 2018.
- Harvey, P. and Körtner, G.: ExifTool, Kingston, Ontario, Canada [code], <http://owl.phy.queensu.ca/~phil/exiftool/> (last access: 27 September 2022), 2016.
- Mäder, P., Boho, D., Rzanny, M., Seeland, M., Wittich, H. C., Deggelmann, A., and Wäldchen, J.: The Flora Incognita app—interactive plant species identification, *Methods Ecol. Evol.*, 12, 1335–1342, 2021.
- Miller-Rushing, A., Primack, R., and Bonney, R.: The history of public participation in ecological research, *Front. Ecol. Environ.*, 10, 285–290, 2012.
- Nugent, J.: iNaturalist, *Science Scope*, 41, 12–13, 2018.
- Redmon, J. and Farhadi, A.: YOLOv3: An Incremental Improvement, *arXiv* [preprint], <https://doi.org/10.48550/arxiv.1804.02767>, 8 April 2018.
- Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., Gonzalez-Beltran, A., Gray, A. J. G., Groth, P., Goble, C., Grethe, J. S., Heringa, J., Hoen, P. A. C.'t, Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S. J., Martone, M. E., Mons, A., Packer, A. L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S. A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M. A., Thompson, M., van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J., and Mons, B.: The FAIR Guiding Principles for scientific data management and stewardship, *Sci. Data*, 3, 1–9, <https://doi.org/10.1038/sdata.2016.18>, 2016.