https://doi.org/10.5194/essd-2023-494
Preprint. Discussion started: 18 January 2024

Open Access Earth System Science Data Discussions

Advancements in LUCAS Copernicus 2022: Enhancing Earth Observation with Comprehensive In-Situ Data on EU Land Cover and Use

Raphaël d'Andrimont[1], Momchil Yordanov[2], Fernando Sedano[1], Astrid Verhegghen[1], Peter Strobl[1], Savvas Zachariadis[3], Flavia Camilleri[3], Alessandra Palmieri[3], Beatrice Eiselt[3], Jose Miguel Rubio Iglesias[4], and Marijn van der Velde[1]

[1]European Commission Joint Research Centre (JRC), Ispra, Italy
[2]SEIDOR Consulting S.L., 08500 Barcelona, Spain
[3]European Commission, Eurostat (ESTAT), Luxembourg, Luxembourg
[4]European Environment Agency, (EEA), Copenhagen, Denmark

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

Abstract. The Land Use/Cover Area frame Survey (LUCAS) of the European Union (EU) presents a rich resource for detailed understanding of land cover and use, making it invaluable for Earth Observation (EO) applications. This manuscript discusses the recent advancements and improvements in the LUCAS Copernicus module, particularly the data collection process of 2022, its protocol simplifications, and geometry definitions compared to the 2018 survey and data. With approximately 150,000 polygons collected in 2022, an increase from 60,000 in 2018, the LUCAS Copernicus 2022 data provides a unique and comprehensive in-situ dataset for EO applications. The protocol simplification also facilitates a faster and more efficient data collection process. In 2022, there are 137,966 polygons generated, out of the original 149,408 LUCAS Copernicus points, which means 92.3% of the points were actually surveyed. The data holds 82 land cover classes for the Copernicus module LUCAS level 3 legend (88 classes). For land use the data holds 40 classes, along with 18 classes of land use types. The dataset is available here for download (PID: http://data.europa.eu/89h/e3fe3cd0-44db-470e-8769-172a8b9e8874). The paper further elaborates on the implications of these enhancements and the need for continuous harmonisation to ensure semantic consistency and temporal usability of data across different periods. Moreover, it calls for additional studies exploring the potential of the collected data, especially in the context of remote sensing and computer vision. The manuscript ends with a discussion on future data usage and dissemination strategies.

1 Introduction

The importance of in-situ data for Earth Observation (EO) applications cannot be overstated. In-situ observations provide ground-based reference data, crucial for the production, validation and calibration of remote sensing products derived from satellite or airborn observations. The two largest constraints to satellite-based model performance are training data and imagery (Burke et al., 2021). While imagery has become abundant, the scarcity and frequent unreliability of ground-based reference observations data make both training and validation of satellite-based models difficult. Particularly, they are pivotal in the assimilation practices to better inform Earth surface modeling and other EO endeavors (Balsamo et al., 2018). The Copernicus component of the EU's space program, known for its Earth Observation capabilities, heavily relies on a vast array of in-situ data. The Cross cutting coordination of Copernicus access to in-situ data[1] provides support to Entrusted Entities in accessing such data for both the production and validation of Copernicus products.

Despite the significant value, the collection of low-uncertainty in-situ data presents a myriad of challenges. The systematic collection of such data by humans is very resource-intensive, and ensuring the necessary quality and representativity for effective use in Earth Observation applications further exacerbates the challenge (Teucher et al., 2022; Andries et al., 2022).

In the EU, a regular surveyed sample of land cover and land use has been collected since 2006 under the Land Use/Cover Area frame Survey (LUCAS) (d'Andrimont et al., 2020). The data collected under this survey, especially the new LUCAS Copernicus module introduced in 2018, offers a remarkable resource of in-situ data. In 2018, this new LUCAS module (the Copernicus module) specifically tailored to EO was introduced, see d'Andrimont et al. (2021). A specific protocol was designed to collect *in-situ* information with specific characteristics fitting EO processing requirements. As a result, a total of 58,428 polygons are provided with a level-3 land cover (66 specific classes including crop types) and land use (38 classes) information. This represent a unique set of in-situ data, opening the possibility for applications with higher thematic detail as compared to previous LUCAS surveys, such as crop type mapping. This dataset has been used to generate continental mapping of crop with sentinel-1 (d'Andrimont et al., 2021), with Sentinel-2 (Ghassemi et al., 2022a; Luo et al., 2022) but also for Land cover (Venter and Sydenham, 2021; Ghassemi et al., 2022b; Witjes et al., 2022).

In 2022, a new survey and protocol was carried out. The advancements in the LUCAS Copernicus module, particularly the data collection process of 2022, its protocol simplifications, and geometry definitions compared to the 2018 survey and data, provide a substantial enhancement in the in-situ dataset available for EO applications. This manuscript delves into these improvements and the implications thereof, alongside discussing the future data usage and dissemination strategies.

2 LUCAS Survey

LUCAS is a two phase sample survey. The LUCAS first phase sample is a systematic selection of points on a grid with a 2 km spacing in Eastings and Northings covering the whole of the EU's territory Gallego and Bamps (2008). Currently, it includes around 1.1 million points (Figure 1) and is referred to as the master sample. Each point of the first phase sample is classified in one of ten land-cover classes via visual interpretation of ortho-photos or satellite images ESTAT (2018).

2.1 LUCAS 2018 survey

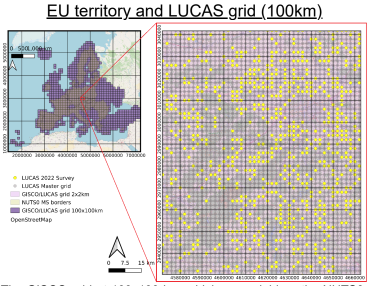
The LUCAS 2018 survey collected 97 variables at 337,854 points. Most of the points surveyed fall in a homogeneous area for which the minimum mapping unit is about 7 m^2(a circle of 1.5 m radius). When the land cover is not homogeneous, for example when it is composed of trees or shrubs interspersed with grass, the scale of observation is extended to classify

[1] https://insitu.copernicus.eu/

https://doi.org/10.5194/essd-2023-494
Preprint. Discussion started: 18 January 2024

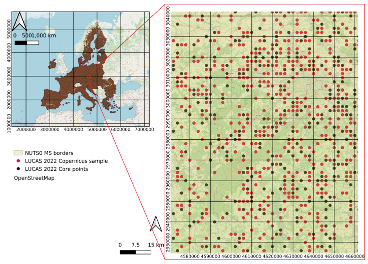
Open Access Earth System Science Data Discussions

EU territory and LUCAS grid (100km)



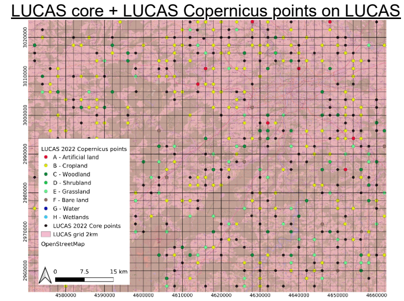
The GISCO grid at 100x100 km gridsize, overlaid on the NUTS0 Member state boundaries. The inline map shows a zoom on the Prague countryside with the same grid at 2x2 km cellsize (essentially the LUCAS Master grid) and with the fraction of all LUCAS points from the vertices of the grid (gray points) that were selected for the 2022 survey (yellow points).

EU territory and LUCAS core + Copernicus points



LUCAS core (399,648) and the fraction of them that makes up the LUCAS Copernicus (137,966) points with a valid geometry overlaid on the NUTS0 MS boundaries. The inline map shows a zoom of the same region. We can visually see the excellent coverage of Copernicus points (red) as a sample from the 2022 survey (black).

LUCAS core + LUCAS Copernicus points on LUCAS



Detailed zoom on the same region. Copernicus points are shown broken down into the major land cover types, while the core points are shown in black, overlaid on the LUCAS Master grid.

LUCAS 2022/23 Core snapshot	
N of points	399,648
N of points with field survey	199,080
N of attributes	113
N of Copernicus points	149,408
N of LF points	92,382
N of Grassland points	19,889
N of Extended Grassland	39,842
N of Soil points	40,881

LUCAS 2023/23 Copernicus snapshot	
N of Copernicus polygons	137,966
N of LC1 classes	88
N of LC1_spec classes	90
N of LU1 classes	40
N of LU1_type classes	18
N of crop types (B class)	43

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. (The background map is obtained from © OpenStreetMap contributors 2023. Distributed under the Open Data Commons Open Database License (ODbL) v1.0).

it. In these cases, a systematic observation of the "environment" in the vicinity of the point, which in LUCAS is called the extended window of observation, has to be adopted. The extended window of observation expands to a radius of 20 meters from the point (representing an area of 0.13 ha) for forest and shrublands. Detailed information about the survey can be found in Eurostat (2018a). The land cover surveyed is classified according to a harmonized 3-level legend system Eurostat (2018b). Additionally to the *core* variables collected, other specific modules were carried out on demand on a subset of points, such as (i) the topsoil module and (iii) the grassland module. The LUCAS 2018 *core* data are available in an harmonised open database in d'Andrimont et al. (2020).

2.2 LUCAS Copernicus module

The LUCAS 2018 Copernicus module was applied to a subset of points from the 2018 survey to collect land cover information up to an extent of 51 meters until the land cover changes in four cardinal directions around a point of observation, offering *in-situ* data compatible with the spatial resolution of high-resolution sensors (specifically Sentinel 1 and 2; see d'Andrimont et al. (2021) for the open ready-to-use dataset). The LUCAS Copernicus dataset contains 63,287 polygons that represent the

https://doi.org/10.5194/essd-2023-494
Preprint. Discussion started: 18 January 2024

pure land cover at level-2 (genus). When filtering the data for which a level-3 (species) legend is available, 58,426 polygons with a level-3 land cover are available. The minimum mapping unit (MMU) of the in-situ data is 78.53 m^2 (i.e. a circle of 5-m radius) as the Copernicus module survey is not executed for areas smaller than 25 square meters.

3 LUCAS Copernicus 2022

In 2022, the Copernicus module was simplified as illustrated by the field form (Figure A4). By removing some variables that were sampled in the 2018 protocol but that did not prove to be useful, the survey cost at each point could be reduced. Therefore, the total number of points could be increased to 150k. The protocol requires the surveyor to register the LUCAS LC level 3 at the position they have reached. In contrast to the 2018 Copernicus module, this means that detailed information on e.g. permanent tree crops is also collected. The position reached can coincide with the LUCAS point or not. Even when a surveyor cannot reach the LUCAS point, they normally are able to collect Copernicus-relevant information, except when on a linear feature narrower than 3 m (e.g. tracks, grass margins or similar). For most cases when doing Photo-Interpretation (PI) in the field it was generally possible to find a suitable location to carry out the survey.

The protocol also allows for special cases, such as the possibility to move a few meters in order to position the Copernicus polygon better in the landscape, marking hedges and other linear elements, etc.

All LUCAS points	Copernicus point = YES	CANDO = YES	With geometry
399,648	149,408	138,024	137,966

Table 1. The number of LUCAS points at each step. Column one shows all LUCAS 2022 points to be surveyed. The second, whether the point was originally selected as a LUCAS Copernicus point. The third, whether the Copernicus sample was actually taken (CANDO is the suffix of the column name). The final column shows the number of points, for which it was possible to generate the radial geometry.

https://doi.org/10.5194/essd-2023-494
Preprint. Discussion started: 18 January 2024

Open Access Earth System Science Data Discussions

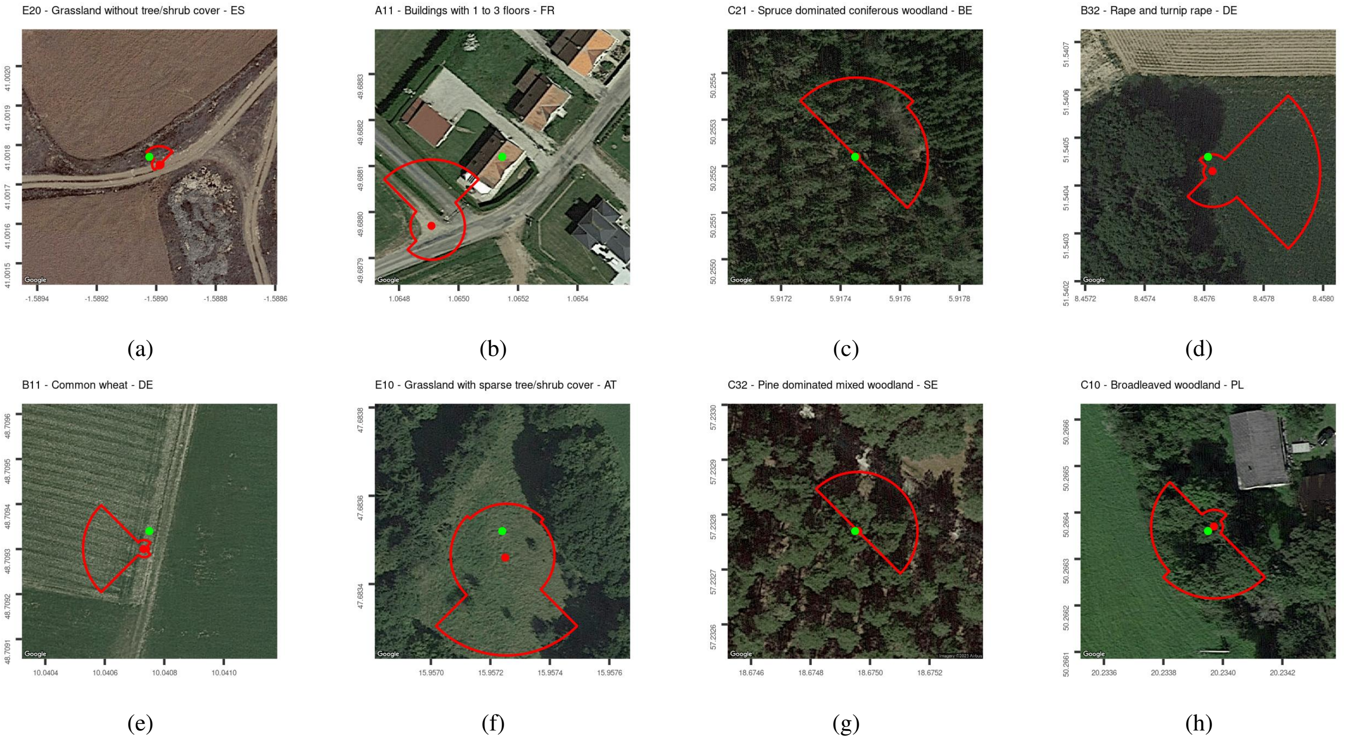


Figure 2. Examples of LUCAS Copernicus built polygons. The green point is the theoretical LUCAS point. The red point is the GPS location of the Copernicus surveyor. Polygons are built using distances in the N, E, S, and W directions collected on the ground. The background RGB imagery is obtained from map data ©2023 Google Maps.

4 Public data and usage note

In the 2022 survey, 137,966 polygons were generated from validly collected survey data over EU-27 as shown in the map Figure 3. They were produced in varying amounts for 88 LC classes as illustrated on Figure 4.

The mean size of the LUCAS Copernicus polygons is 0.34 Ha with the full distribution visible on Figure 5. There are two tails at either end of the distribution - before the first quantile, and after the third. There is no obvious reason behind the tails, meaning no LCLU class or NUTS0 is overly represented within them, although many of the polygons larger than 0.8 Ha are in the Arable land class, meaning used for agriculture.

The crop samples of the survey are mapped on Figure A2 and their distribution is on Figure A3.

https://doi.org/10.5194/essd-2023-494
Preprint. Discussion started: 18 January 2024

Open Access Earth System Science Data Discussions

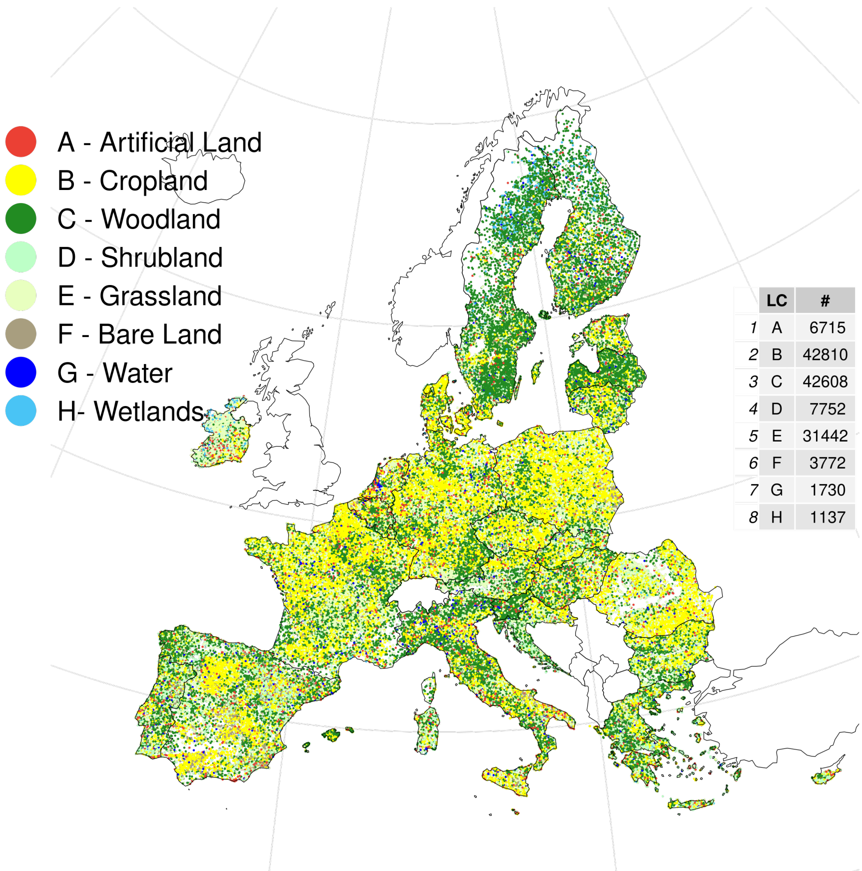


Figure 3. Map of the LUCAS Copernicus polygon collected in 2022.

https://doi.org/10.5194/essd-2023-494
Preprint. Discussion started: 18 January 2024

Open Access Earth System Science Data Discussions

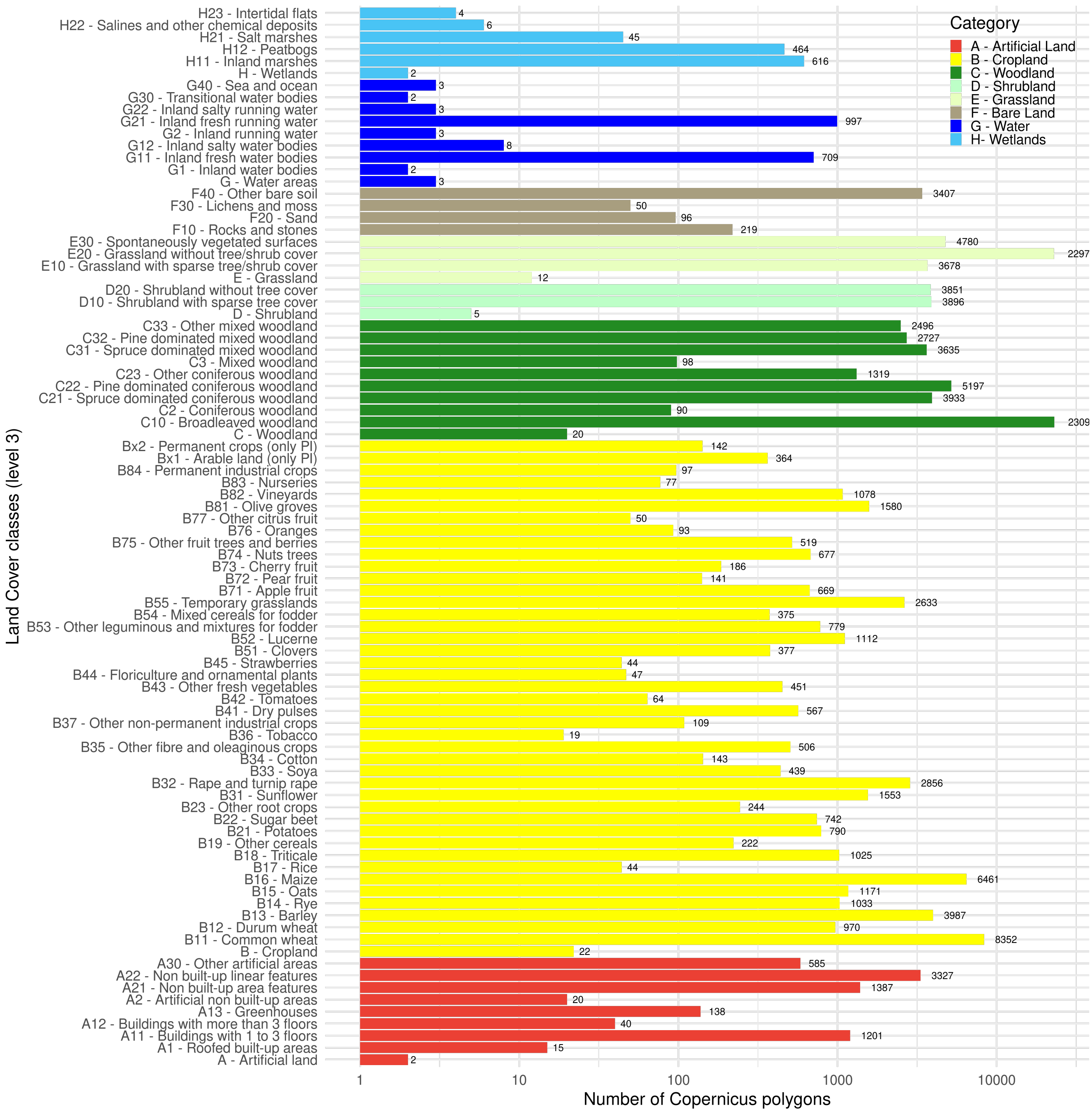


Figure 4. Land cover class distribution of the LUCAS Copernicus polygon collected in 2022.

https://doi.org/10.5194/essd-2023-494
Preprint. Discussion started: 18 January 2024

Open Access Earth System Science Data Discussions

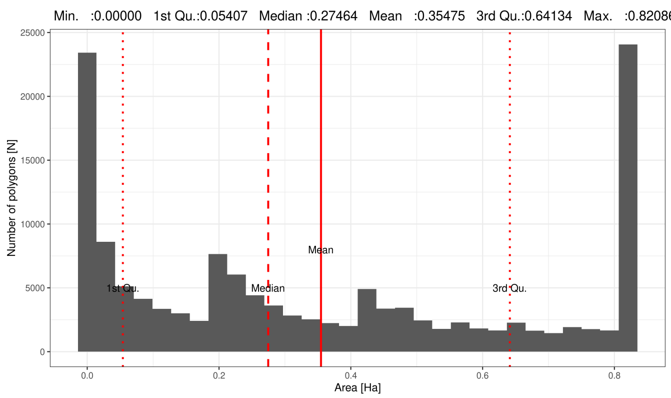


Figure 5. Area of polygons collected in 2022.

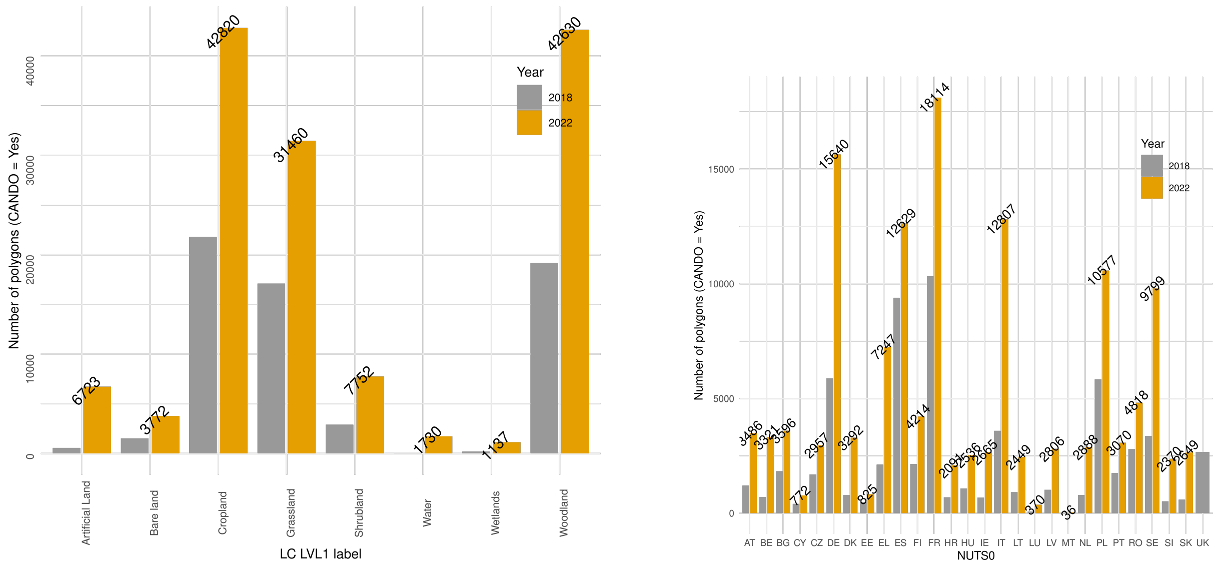
https://doi.org/10.5194/essd-2023-494
Preprint. Discussion started: 18 January 2024

Open Access Earth System Science Data Discussions

5 Discussion

5.1 Comparison 2018-2022

Evolution between the LUCAS Copernicus 2018 and 2022 by Land cover (Figure 6a) and by country (Figure 6b). From Figure 6a we see that for every LUCAS LC1 the number of samples has at least doubled. For some classes, like Artificial land (A) and Water (G), the samples are much more - 91.6% and 98.6% respectively.



(a) Evolution by LC.

(b) Evolution by country.

Figure 6. Change between the two LUCAS Copernicus surveys 2018 - 2022/23 in terms of representativeness in general LUCAS Land Cover classes and by NUTS0. The number at the top of the bar shows the value for the 2022 survey.

A similar trend is observed for each country (NUTS0), where the number of samples has at least doubled, in some cases tripled (Belgium, Germany, Denmark, Greece, Ireland, Italy, Netherlands, Slovenia, Slovakia), safe for the UK, which is no longer part of the survey. In essence, because of BREXIT, the quota of points available to the UK have now been re-distributed to the other Member States.

5.2 Quadrilateral vs radial polygon geometry

As shown in Figure A1, in 2018 the polygons were generated by using the distances noted by surveyors as the lengths of the line segments that make up the shape of the irregular quadrilateral Copernicus polygons. In essence, this was a simplification of the actual survey design, which required the generation of radial quarter-arc slices in each cardinal direction, which are to be merged together to form another irregular quasi-circular polygon shape. The difference between the two shapes obviously has an impact on the area of the polygon and hence - on the amount of Sentinel pixels that fall within it.

https://doi.org/10.5194/essd-2023-494
Preprint. Discussion started: 18 January 2024

The difference in area between the two types of polygon definition (quadrilateral and radial) is shown on Figure 7. The total area of quadrilateral LUCAS Copernicus 2022 polygons is 274.6 km2, while for the radial definition the area is 489.4 km2, or an increase of 78.2% for the entire EU-27. The maximum increase in area is registered in Malta (95.5%) and the minimum in Latvia (68.5%).

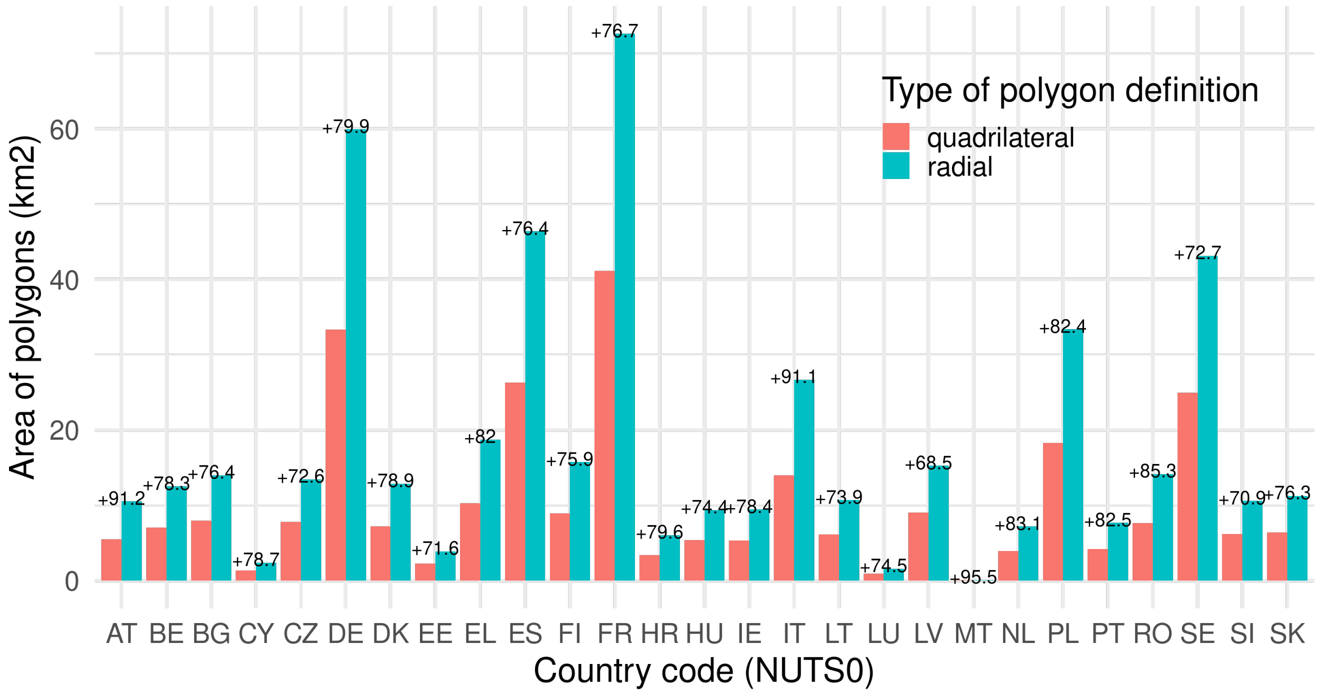


Figure 7. Area of LUCAS Copernicus polygons, aggregated to NUTS0 region, as derived based on the two types of polygon definition. The number at the top of each bar group represents the increase (in percent) between the two.

5.3 Assessment

A preliminary assessment of the dataset highlights the following key insights:

- LUCAS Copernicus data is a valuable source for training and/or validation data for various operational programs, such as the Copernicus Land Monitoring Service (CLMS), Horizon 2020, Horizon Europe, and other associated activities.
- LUCAS point data can be used as a sampling and/or stratification scheme for selecting training/validation data that originates from elsewhere.
- Sampling grid density, spatio-temporal coverage, temporal asynchrony from other CLMS products, geo-location precision issues during data collection, and legend-matching issues still prevent the easy and straightforward use of LUCAS data for many applications (Schweitzer et al., 2023).

https://doi.org/10.5194/essd-2023-494
Preprint. Discussion started: 18 January 2024

6 Conclusions

A new simplified Copernicus protocol has been defined. Some notable complications have been removed such as collecting the land cover directly in the Copernicus polygon. In part because of the improved procedure, the LUCAS Copernicus 2022 is almost completed. With 150K polygons collected in 2022, compared to 60K in 2018, the dataset provides unique in-situ data for Earth Observation applications, after the quality has been assessed. Further harmonisation is needed in order to guarantee the semantic consistency of the coding and legend, as well as the temporal inter-usability of both the 2018 and 2022 data. In terms of analysis there is much to be done with the polygons themselves in the remote sensing context, as well as, the collection of ground photos that can be leveraged for further computer vision work. From a data collection point of view, most issues with the surveys (FAQ from surveyor) have been solved. Further discussion about the data use and diffusion still needed.

7 Code and data availability

To produce the processing pipeline the authors used both the R programming language, version 4.2.1, and PostgreSQL (13.0) with the PostGIS plugin (3.0.2). The code for producing the dataset from the raw LUCAS data and all the figures shown in the manuscript consists of four commented scripts - pre-processing, generating quadrilateral LUCAS polygons, generating radial LUCAS polygons, and producing figures and tables. The order in which these are executed is important and they are numbered accordingly. The pre-processing includes the download of the micro data from Eurostat website[2].. These, alongside a ReadMe file can be accessed here: https://jeodpp.jrc.ec.europa.eu/ftp/jrc-opendata/LUCAS/LUCAS_2022_Copernicus/.

The produced dataset is provided in geopackage format and hosts 117 relevant LUCAS attribute columns, plus the radial geometry of the Copernicus polygons. The dataset is available here for download (PID: http://data.europa.eu/89h/e3fe3cd0-44db-470e-8769-172a8b9e8874, European Commission (2022)). The link between this dataset and previous LUCAS surveys, or with the harmonised product d'Andrimont et al. (2020), is accomplished via the pointid column. Although all columns come from the data, the authors have added two additional columns - "survey_year" and "poly_area_sqm"; the first one in order to track which records come from which year of survey; the second is the calculated area in square meters.

[2]The micro data were obtained from https://ec.europa.eu/eurostat/web/lucas/data/primary-data/2022 accessed on 1/11/2023. For the latest version of the data, please refer to Eurostat website.

https://doi.org/10.5194/essd-2023-494
Preprint. Discussion started: 18 January 2024

Open Access Earth System Science Data Discussions

Appendix A

A1

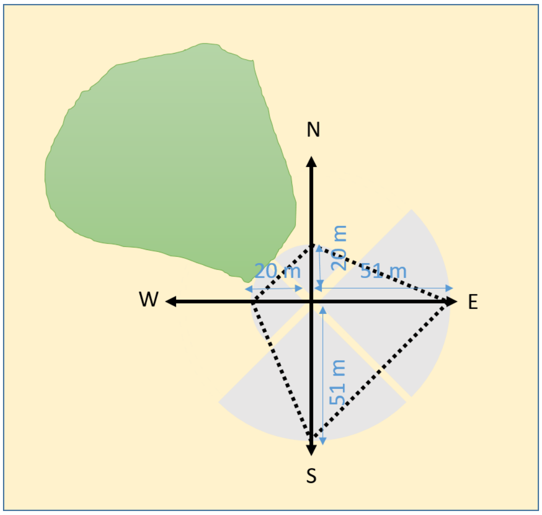


Figure A1. Overview of how polygons were built based on the radial distance collected on the ground. A distinction can be made between constructing a polygon based on quadrilateral distance (2018) or by generating the radial areas corresponding to the field of view the observation should represent.

Competing interests. The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements. The authors would like to acknowledge the invaluable efforts of the LUCAS field surveyors and the teams involved in the LUCAS database management at Eurostat. Their dedication and hard work have made the LUCAS Copernicus 2022 dataset possible. We would also like to thank the various project partners for their continuous support and feedback.

https://doi.org/10.5194/essd-2023-494
Preprint. Discussion started: 18 January 2024

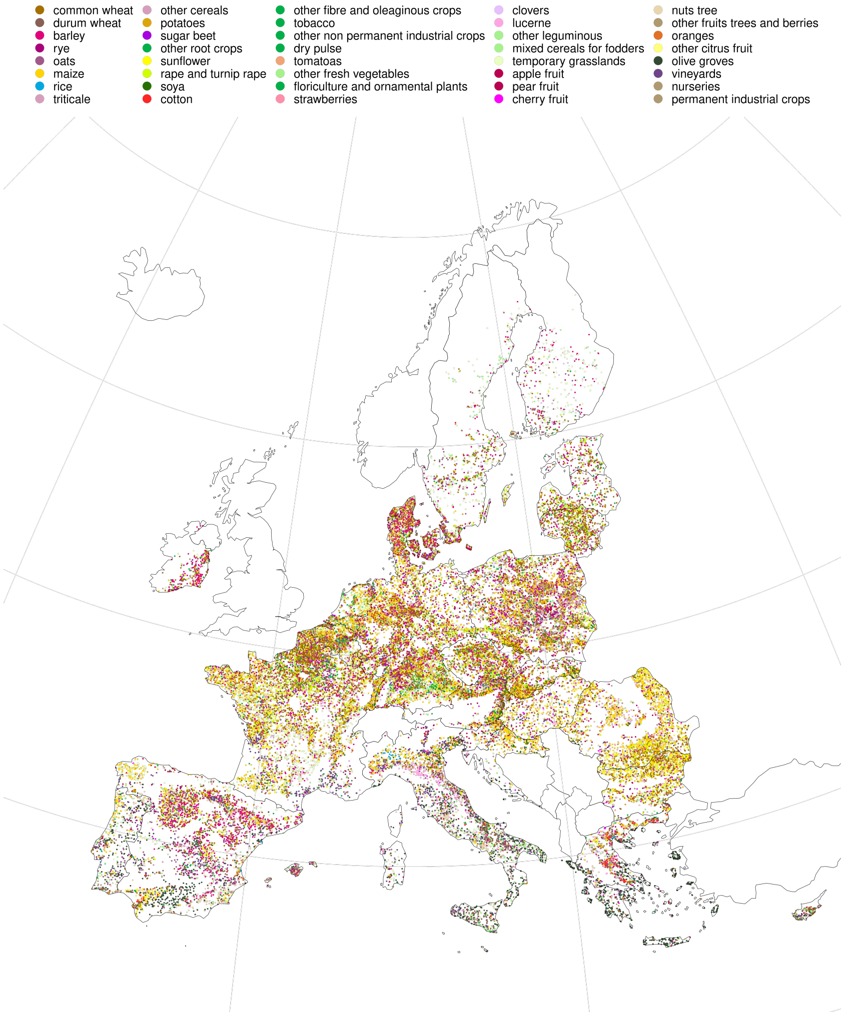


Figure A2. Map of the LUCAS Copernicus polygon collected in 2022, only showing the crops.

References

Andries, A., Morse, S., Murphy, R. J., Lynch, J., and Woolliams, E. R.: Using data from earth observation to support sustainable development indicators: An analysis of the literature and challenges for the future, Sustainability, 14, 1191, 2022.

Balsamo, G., Agusti-Parareda, A., Albergel, C., Arduini, G., Beljaars, A., Bidlot, J., Blyth, E., Bousserez, N., Boussetta, S., Brown, A., et al.: Satellite and in situ observations for advancing global Earth surface modelling: A Review, Remote Sensing, 10, 2038, 2018.

Burke, M., Driscoll, A., Lobell, D. B., and Ermon, S.: Using satellite imagery to understand and promote sustainable development, Science, 371, eabe8628, 2021.

https://doi.org/10.5194/essd-2023-494
Preprint. Discussion started: 18 January 2024

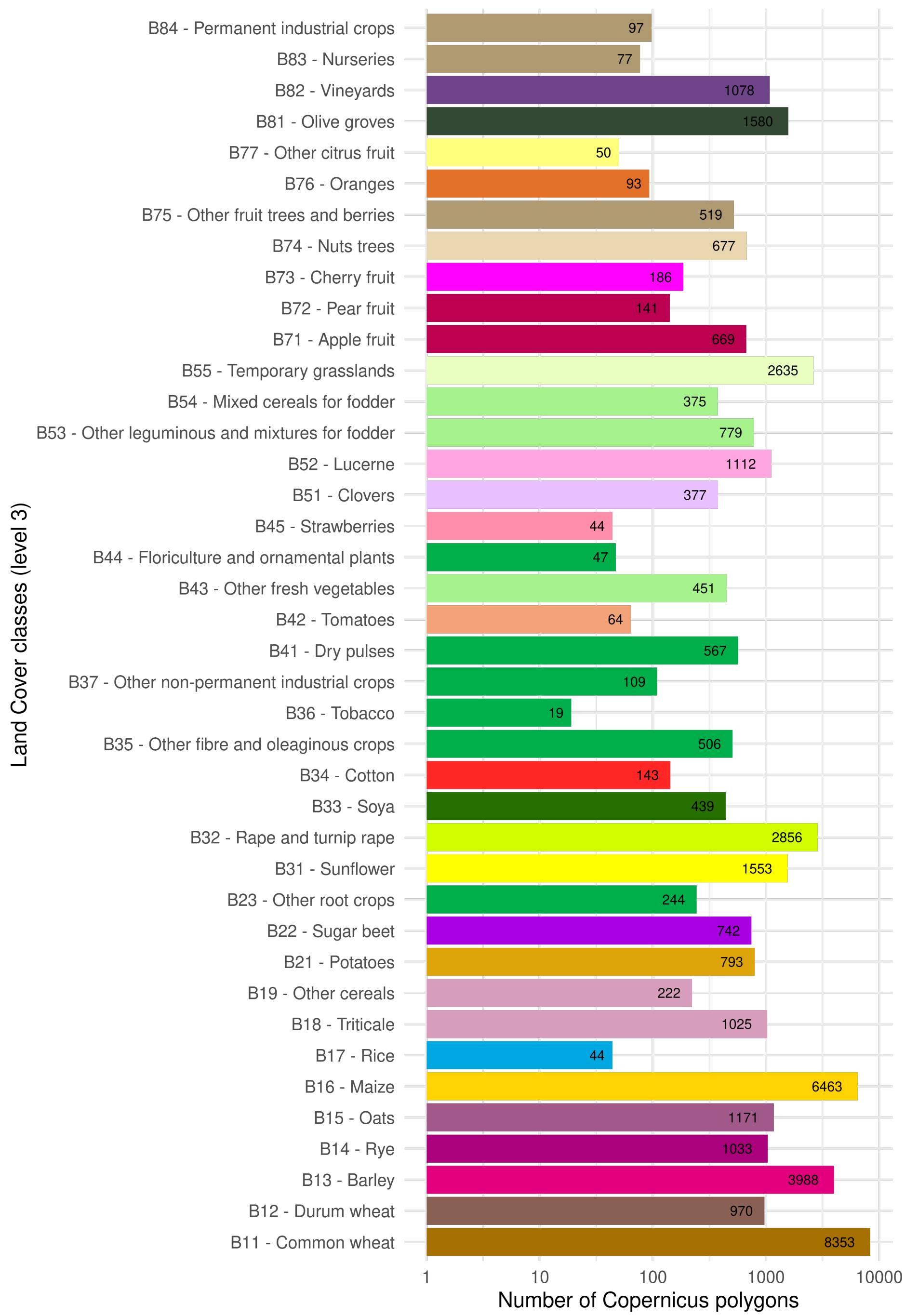


Figure A3. Crop class distribution of the LUCAS Copernicus polygon collected in 2022.

https://doi.org/10.5194/essd-2023-494
Preprint. Discussion started: 18 January 2024

Open Access Earth System Science Data Discussions

11 COPERNICUS

The Copernicus module is only assessed if:

- The point is part of the Copernicus module
- Observation type = 1, 2 or 3

The Copernicus LC is registered at level 3.

* Only collect Copernicus data if point is part of Copernicus module and Field 24 (observation type) is 1

G	**Copernicus point:** 1 ☐ Yes 2 ☐ No		
57	**Can you do the Copernicus survey?** 1 ☐ Yes 2 ☐ No	58	**Copernicus LC:**

	Extension in Copernicus LC in 4 cardinal directions (in m) Values: 0-51 (m) Extension >50: always insert 51	
N	60	
E	61	
S	62	
W	63	

XXX	**Comment about Copernicus [structured comments / other] (PA01, PA02, PA99)** .

	Photo: Take photo of Copernicus and cover: COPERNICUS_LC (No point marker) Take other photos as needed: COPERNICUS_Other (multiple photos « other » possible)

Figure A4. Extract of the Field form of LUCAS 2022 for the Copernicus module .

d'Andrimont, R., Verhegghen, A., Meroni, M., Lemoine, G., Strobl, P., Eiselt, B., Yordanov, M., Martinez-Sanchez, L., and van der Velde, M.: LUCAS Copernicus 2018: Earth-observation-relevant in situ data on land cover and use throughout the European Union, Earth System Science Data, 13, 1119–1133, 2021.

d'Andrimont, R., Yordanov, M., Martinez-Sanchez, L., Eiselt, B., Palmieri, A., Dominici, P., Gallego, J., Reuter, H. I., Joebges, C., Lemoine, G., et al.: Harmonised LUCAS in-situ land cover and use database for field surveys from 2006 to 2018 in the European Union, Scientific Data, 7, 1–15, 2020.

d'Andrimont, R., Verhegghen, A., Lemoine, G., Kempeneers, P., Meroni, M., and Van der Velde, M.: From parcel to continental scale–A first European crop type map based on Sentinel-1 and LUCAS Copernicus in-situ observations, Remote sensing of environment, 266, 112 708, 2021.

ESTAT: Technical reference document S1 : Stratification Guidelines, https://ec.europa.eu/eurostat/documents/205002/7329820/LUCAS2018_S1-StratificationGuidelines_20160523.pdf, 2018.

European Commission, J. R. C. J.: LUCAS Copernicus 2022 [Dataset], http://data.europa.eu/89h/e3fe3cd0-44db-470e-8769-172a8b9e8874, 2022.

https://doi.org/10.5194/essd-2023-494
Preprint. Discussion started: 18 January 2024

Open Access Earth System Science Data Discussions

Eurostat: Technical reference document C-1: Instructions for surveyors, https://ec.europa.eu/eurostat/documents/205002/8072634/LUCAS2018-C1-Instructions.pdf, 2018a.

Eurostat: Technical reference document C-3: Classification, https://ec.europa.eu/eurostat/documents/205002/8072634/LUCAS2018-C3-Classification.pdf, 2018b.

Gallego, J. and Bamps, C.: Using CORINE land cover and the point survey LUCAS for area estimation, International Journal of Applied Earth Observation and Geoinformation, 10, 467–475, 2008.

Ghassemi, B., Dujakovic, A., Żółtak, M., Immitzer, M., Atzberger, C., and Vuolo, F.: Designing a european-wide crop type mapping approach based on machine learning algorithms using LUCAS field survey and sentinel-2 data, Remote sensing, 14, 541, 2022a.

Ghassemi, B., Immitzer, M., Atzberger, C., and Vuolo, F.: Evaluation of Accuracy Enhancement in European-Wide Crop Type Mapping by Combining Optical and Microwave Time Series, Land, 11, 1397, 2022b.

Luo, Y., Zhang, Z., Zhang, L., Han, J., Cao, J., and Zhang, J.: Developing high-resolution crop maps for major crops in the european union based on transductive transfer learning and limited ground data, Remote Sensing, 14, 1809, 2022.

Schweitzer, K., Lindmayer, A., and Sorini, P.: LUCAS Assessment, Task B, Tech. Rep. 1.1, Space program of EU Copernicus, Frescatti, Rome, 2023.

Teucher, M., Thürkow, D., Alb, P., and Conrad, C.: Digital In Situ Data Collection in Earth Observation, Monitoring and Agriculture—Progress towards Digital Agriculture, Remote Sensing, 14, 393, 2022.

Venter, Z. S. and Sydenham, M. A.: Continental-scale land cover mapping at 10 m resolution over Europe (ELC10), Remote Sensing, 13, 2301, 2021.

Witjes, M., Parente, L., van Diemen, C. J., Hengl, T., Landa, M., Brodskỳ, L., Halounova, L., Križan, J., Antonić, L., Ilie, C. M., et al.: A spatiotemporal ensemble machine learning framework for generating land use/land cover time-series maps for Europe (2000–2019) based on LUCAS, CORINE and GLAD Landsat, PeerJ, 10, e13 573, 2022.