



OPEN

DATA DESCRIPTOR

European Union crop map 2022: Earth observation's 10-meter dive into Europe's crop tapestry

Babak Ghassemi¹, Emma Izquierdo-Verdiguier¹, Astrid Verhegghen², Momchil Yordanov³, Guido Lemoine², Álvaro Moreno Martínez⁴, Davide De Marchi², Marijn van der Velde², Francesco Vuolo¹ & Raphaël d'Andrimont²

To provide the information needed for a detailed monitoring of crop types across the European Union (EU), we present an advanced 10-metre resolution map for the EU and Ukraine with 19 crop types for 2022, updating the 2018 version. Using Earth Observation (EO) and *in-situ* data from Eurostat's Land Use and Coverage Area Frame Survey (LUCAS) 2022, the methodology included 134,684 LUCAS Copernicus polygons, Sentinel-1 and Sentinel-2 satellite imagery, land surface temperature and a digital elevation model. Based on this data, two classification layers were developed using a Random Forest machine learning approach: a primary map and a gap-filling map to address cloud-covered gaps. The combined maps, covering 27 EU countries, show an overall accuracy of 79.3% for seven major land cover classes and 70.6% for all 19 crop types. The trained model was used to derive the 2022 map for Ukraine, demonstrating its robustness even in regions without labelled samples for model training.

Background & Summary

Land Use and Land Cover (LULC) maps are used for a wide range of purposes, from urban planning and development¹ to environmental monitoring and conservation, such as assessing ecological changes, monitoring biodiversity patterns, and supporting habitat conservation². LULC maps are invaluable for land management, such as monitoring and assessing agricultural activities, conducting forest inventories, and developing effective watershed management plans³. Furthermore, LULC maps provide data to scientists for studying carbon storage dynamics, land-atmosphere interactions⁴, and ecosystem response to climate change⁵. Finally, these maps are helpful in the context of disaster risk management, such as flood and fire analysis and prevention, where they can be used to identify vulnerable areas, analyze hazard-prone regions, and formulate appropriate mitigation strategies^{6,7}.

Approximately 42 per cent of the European Union (EU) land area is devoted to agriculture, supporting the EU in being the world's largest exporter of agri-food products^{4,8}. Policies are being developed to promote sustainable crop production, which requires up-to-date and accurate maps to provide baseline information for assessment and monitoring⁹. Thematically relevant and up-to-date maps are also essential for optimizing resource allocation and increasing crop productivity when using precise agricultural techniques¹⁰. Crop type maps facilitate the monitoring of crop growth, health, and yield estimation. They allow for informed decisions to be made on the timing of planting, fertilization, and pest management^{11–13}. In addition, LULC information can be used to assess the suitability of land for a particular crop or agricultural practice in combination with other factors such as soil quality and hydrological conditions^{14–16}. Besides contributing to agroecosystem modelling, LULC maps can be used to simulate crop growth and assess the effects of different land management practices^{17,18}. Furthermore, these maps contribute to analyzing historical and future land-use changes and evaluating their impact on crop production^{19–21}.

The full, free, and open availability of new Earth Observation (EO) data, together with powerful machine learning (ML) algorithms and cloud computing platforms, has revolutionized the production of accurate and detailed large-scale LULC maps^{22–24}. Thanks to these advancements, a significant number of LULC products

¹Institute of Geomatics, University of Natural Resources and Life Sciences, Vienna, Peter-Jordan-Straße 82, 1190, Vienna, Austria. ²Joint Research Centre (JRC), European Commission, Ispra, Italy. ³SEIDOR Consulting S.L., 08500, Barcelona, Spain. ⁴Image Processing Laboratory (IPL), Universitat de València, Catedrático A. Escardino, 46980, Paterna, València, Spain. ✉e-mail: francesco.vuolo@boku.ac.at; raphael.dandrimont@ec.europa.eu

have been produced with spatial resolutions ranging from 10 m to 1 km in different regions of the world. Datasets derived from different satellite platforms, such as Sentinel^{23–30}, Landsat^{31–34}, MODIS^{35–37}, and AVHRR^{38–40}, have been processed to achieve this goal.

In recent years, cloud computing platforms, particularly Google Earth Engine (GEE), have become increasingly helpful in generating large-scale LULC maps and updating them promptly, providing scalable and distributed computing resources that make it feasible for large volumes of EO data to be processed and analyzed efficiently. In particular, the GEE platform offers an extensive collection of satellite imagery and geospatial datasets, as well as integrated geospatial analysis tools and parallel processing capabilities, allowing for comprehensive mapping of LULC⁴¹. Several studies have used GEE to address this issue. For instance, Miettinen *et al.*⁴² demonstrated the effectiveness of combining MODIS and Sentinel-1 (S1) datasets to produce a comprehensive land cover (LC) map of Southeast Asia, covering 11 countries and classifying 13 different LC types. Xiong *et al.*⁴³ generated a cropland map for Africa using MODIS time-series data. Ghorbanian *et al.*⁴⁴ employed multi-temporal S1 and Sentinel-2 (S2) data to create a detailed LC map of Iran, encompassing 13 distinct classes. Furthermore, Shafizadeh-Moghadam *et al.*⁴⁵ carried out an object-based classification approach comprising spectral, textural, and topographical factors to produce a LULC map of the Tigris-Euphrates Basin with nine classes. Additionally, Mirmazloumi *et al.*²³ successfully generated a high-resolution LULC map of Europe with nine classes, using GEE and integrating S1, S2, and Landsat-8.

Today's processing capabilities and adequate spatial and temporal observations meet the requirements for continental LULC mapping. However, the accurate and thematically detailed mapping of LULC is still challenged by the availability of labelled data for training and validation. In the EU, *in-situ* data collection is organized through the European Land Use and Cover Area Survey (LUCAS)⁴⁶. This survey has been carried out every three years from 2006 to 2018 and in 2022 to collect comprehensive field data in the 28 countries of the EU (EU-28) for EU-wide standardized reporting of LC and LU area statistics. Although the LUCAS protocol was not initially designed for EO applications, several studies used LUCAS data in conjunction with satellite imagery for mapping purposes leading to new advancements in field data availability. For example, Mack *et al.*⁴⁷ used Landsat-7 and Landsat-8 data with LUCAS 2012 data to produce a LULC map for Germany for the year 2014. Recognizing the potential of the LUCAS data, they acknowledged the necessity to enhance the training dataset, focusing on rare classes that typically exhibited considerable uncertainty levels. Close *et al.*⁴⁸ used LUCAS 2015 survey data and S2 imagery to classify the Wallonia region in Belgium to monitor LULC changes and forestry, suggesting further research to effectively use multi-temporal observations and combinations of S1 and S2 data. Pflugmacher *et al.*⁴⁹ developed the first pan-European LC map with 13 different classes based on the LUCAS 2015 data and Landsat-8 imagery. They also emphasized the necessity of expanding training data for smaller classes and integrating validation efforts with the LUCAS sampling strategy to improve compatibility with Copernicus satellite data. Weigand *et al.*⁵⁰ used the LUCAS 2015 data as reference information for high-resolution LC mapping using S2 data in Germany with seven classes and focused on aspects related to the geo-location of LUCAS samples suggesting the use of the originally intended and actually observed locations of the LUCAS sample instead of utilizing only the recorded positioning. The study underscored the necessity for a tailored LUCAS survey that aligns effectively with Copernicus satellite datasets. The feedback provided by previous research brought to the introduction of the LUCAS “Copernicus module” in 2018, tailored explicitly for EO applications²⁴. Combining the LUCAS Copernicus *in-situ* data with S1 radar data, dAndrimont *et al.*²⁴ generated a 10-meter crop type map for the EU-28. Their study employed the random forest (RF) algorithm to classify 19 different crop type classes, plus two broad classes for Woodland and Shrubland, and Grassland. The overall accuracy (OA) achieved was 76.1%. Ghassemi *et al.*²⁵ extended the analysis by including S2 data, resulting in an improved OA of 77.6%. The study also showed that optimal accuracy could be achieved by utilizing just 11% of the total training data, selecting pixels with the greatest spectral diversity. Optimizing information is crucial when deploying algorithms on cloud computing infrastructure for large-scale applications to minimize resource usage and costs. In addition, Ghassemi *et al.*⁵¹ evaluated the effectiveness of combining the time series data from both S1 and S2 to produce a LULC map with small improvements in the accuracy but better spatial coverage.

On the contrary Venter *et al.*²⁹ produced a LC map of Europe at a 10-meter resolution and eight LC classes finding that the fusion of S1 and S2 data can lead to an increase in classification accuracy of 3–10%. The study confirmed the positive impact of balancing class representativeness in training samples with supplementary data (e.g., standard LUCAS points), underscoring the importance for future LUCAS “Copernicus module” samples to accurately reflect class area proportions. Witjes *et al.*⁵² produced LULC time-series maps for Europe (2000–2019) based on LUCAS and other ground truth data using Landsat imagery. However, employing LUCAS data for training resulted in reduced classification accuracy compared to using solely Corine Land Cover points. Further investigation was suggested by the authors.

The dataset from the “Copernicus module” of the LUCAS 2022 survey, encompassing 134,684 polygons across the 27 European Union countries (EU-27), is now accessible. We utilized this data to generate a new 10-meter spatial resolution LULC map for the entire EU-27, focusing particularly on crop type mapping for 2022. This study introduces the new LULC map, conducts a product quality assessment, and discusses the application of LUCAS 2022 “Copernicus module” data for crop type classification. Specific objectives include:

- Investigate the optimal set of input features by considering different temporal aggregations of S1, S2, and climate data to obtain a wall-to-wall map with no spatial gaps and the best possible thematic accuracy.
- Study a balanced number of training data to allow the training and application of the classification algorithm for the entire EU-27 while optimizing GEE computing resources.
- Produce the resulting LULC map for the year 2022 and perform an extensive validation exercise using independent labelled sample sets, ensuring consistency with the previous map²³.

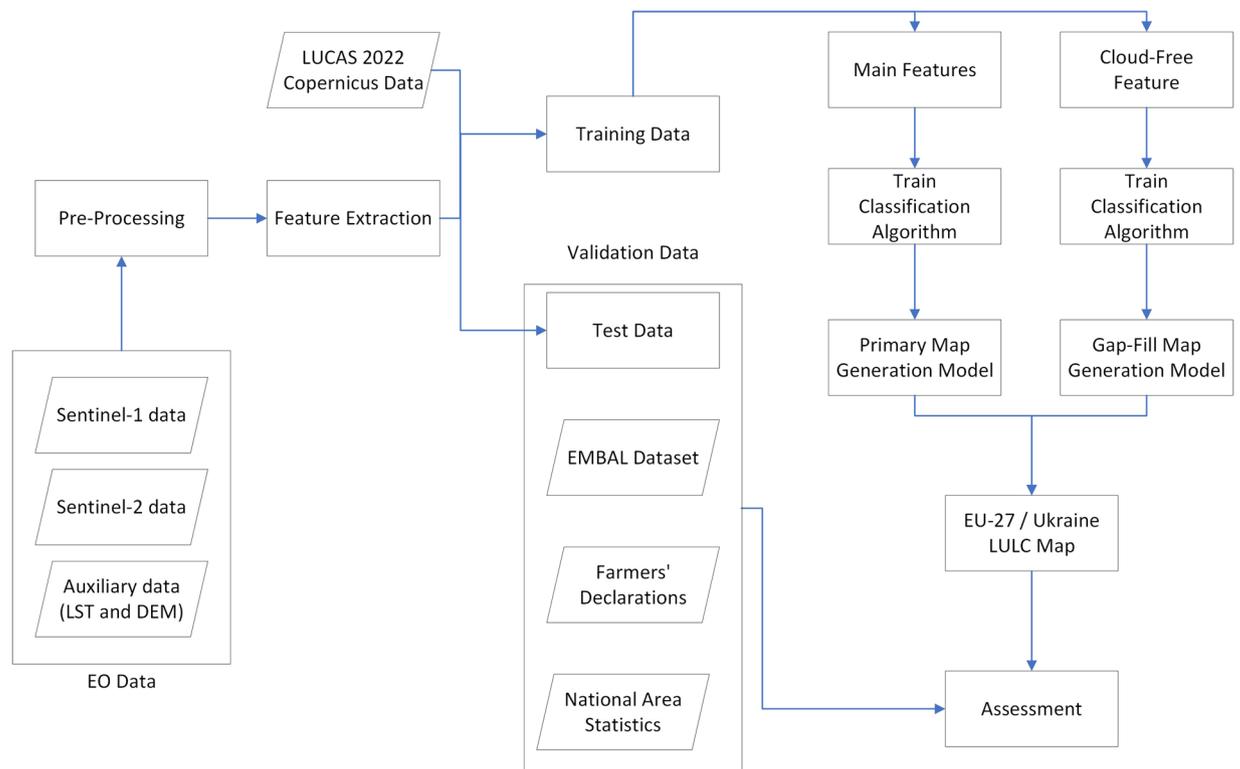


Fig. 1 General overview and main steps of the study.

- Evaluate the classification model's inference performance in Ukraine, where no labelled data was used in the model training stage, and validate the results using third-party independent classification data.

Methods

Figure 1 illustrates the key steps undertaken in this research. The first step in this process involves the extraction of temporal features from EO data for the LUCAS 2022 Copernicus *in-situ* field data locations. This feature set was divided into training and test samples. Following this, two distinct classification models were developed based on the training data: (1) the Primary map generation model was tasked with creating the principal map based on carefully selected effective features; (2) the Gap-Fill map generation model excludes the S2 monthly data, in order to address gaps caused by cloud effects on the S2 monthly data. The LULC map for the EU-27 region was generated using these models and validated using various independent data and statistics. Furthermore, these aforementioned models were applied to infer the LULC map for Ukraine from the equivalent feature sets. The absence of training data in Ukraine makes this an interesting test case for assessing the viability and adaptability of the classification model to other geographic regions.

Field data. *European Land Use/ Cover Area frame Survey (LUCAS) 2022 data.* LUCAS is a triannual survey of LULC that has been carried out since 2006 in the EU. There have been 1,351,293 observations at 651,780 unique locations, containing 106 variables, accompanied by 5.4 million landscape photographs from five editions of the LUCAS surveys⁴⁶.

In 2018, the LUCAS collection strategy was enhanced by adding the “Copernicus module”, which introduced 58,462 polygon geometries representing homogeneous LC patches of approximately 0.5 hectares⁴⁶. Specifically, these polygons were designed to improve accessibility and facilitate data extraction, particularly for satellite imagery such as S1 and S2 with a 10-meter spatial resolution. This module provides detailed information on LC (66 classes, including crop type) and land use (LU) (38 classes) for the mentioned polygons. The dataset offers unique opportunities for applications requiring a higher level of thematic resolution, such as the mapping of crop types.

In 2022, 134,684 polygons were generated, representing 90.1% of the original 149,429 Copernicus module points. These polygons contain 75 LC and 40 LU classes. LUCAS Copernicus 2022 polygons are generated according to a new protocol, and on average, they are 0.2 hectares in size^{53,54}.

This study uses the LUCAS “Copernicus module” 2022 polygons as *in-situ* field observations to obtain a wall-to-wall LULC map of the EU-27 land. The distribution of these polygons, with the main eight-class scheme, is shown in Figure 2.

Classification scheme based on LUCAS 2022 data. According to LUCAS 2022 data, there are eight main Level-1 LC classes: A-Artificial Land, B-Cropland, C-Woodland, D-Shrubland, E-Grassland, F-Bare Land, G-Water, and

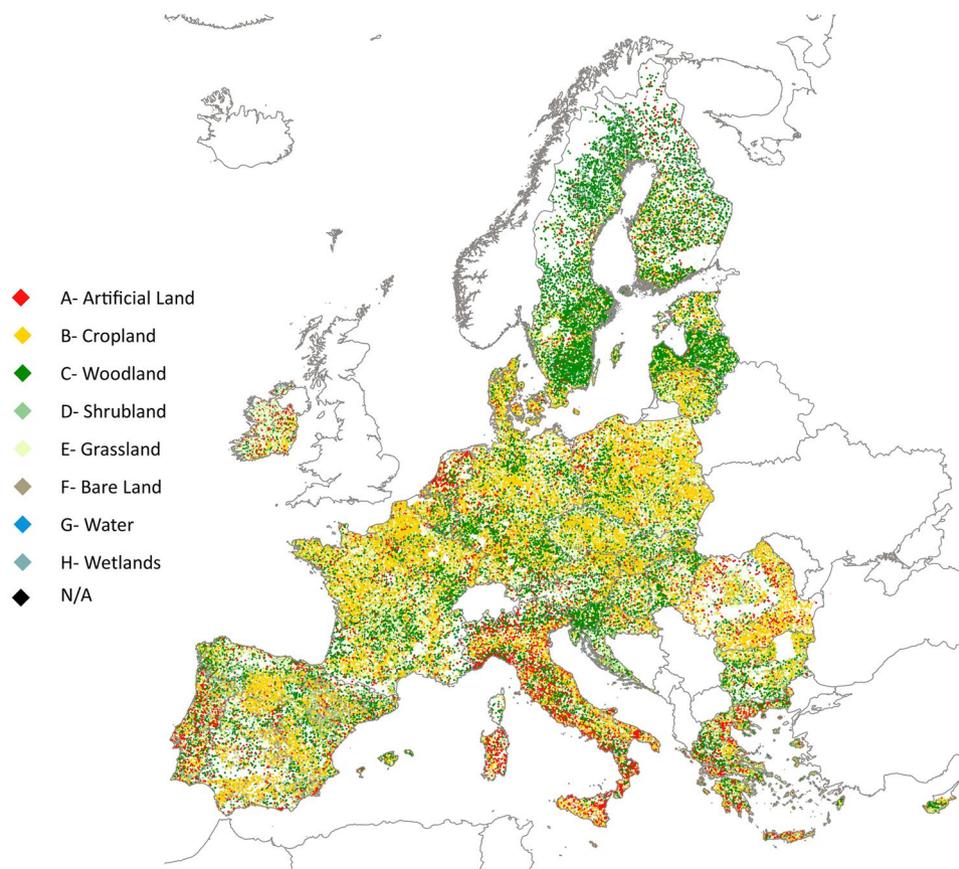


Fig. 2 Distribution of LUCAS 2022 polygons over EU-27 with main land cover categories.

H-Wetlands. To align with the objectives of the study, the legend for the LUCAS survey was recoded based on the scheme proposed by d'Andrimont *et al.* in²⁴. The research aims to classify the main crop types in the EU-27 using LUCAS *in-situ* data, resulting in slight changes in the class definitions from the original LUCAS classification scheme. The four vegetation classes of LUCAS (B-cropland, C-woodland, D-shrubland, and E-grassland) are consolidated into three main categories: Arable land, Woodlands and Shrubland, and Grassland. Arable land has 19 classes representing different crop types or crop groups. The final classes used in this study are shown in Table 1, which includes seven broad categories for Level-1 and 25 detailed categories for Level-2.

Some differences exist between the modified classes used in this work and the original LUCAS. Similar to d'Andrimont *et al.*²⁴, first, the Woodland and Shrubland classes were merged into a single one, acknowledging their resemblance in radiometric vegetation characteristics. Second, the LUCAS B-Cropland class was reorganized to maintain a clear distinction between Arable land and other categories. Specifically, this involves grouping temporary grasslands (B55) with the E-Grassland class, combining permanent crops (B70 and B80) with the woody vegetation class, and aggregating Bare arable land (F40) with agricultural LU (U111/112/113) within the Arable land class. Additionally, the Grassland now contains both natural and agricultural land uses, allowing for a more comprehensive representation. Furthermore, this research added Water and Wetland classes to the investigation procedure, in addition to the classes defined in scheme²⁴.

Earth Observation data. Multiple EO datasets were used to generate the LULC map for the entire desired area. The employed EO data comprises high resolution, frequent revisit reflectance data from S2 and backscattering coefficients from S1, as well as auxiliary data, including Land Surface Temperature (LST) and Digital Elevation Model (DEM). Besides, ancillary data for masking, containing the Digital Surface Model (DSM) and lower resolution LC were utilized. The following sections provide a detailed description of each of these categories.

Sentinel-2 data. The S2 satellite mission was developed by the EU as part of the Copernicus program. The optical imagery constellation consists of two identical satellites, Sentinel-2A (S2A) and Sentinel-2B (S2B), which capture high-resolution images of the Earth's surface. The satellites are equipped with a multispectral sensor that captures data in 13 spectral bands, ranging from visible to infrared wavelengths (WL) and 10 to 60 m pixel sizes. This extensive spectral coverage enables detailed analysis of land and coastal areas, providing valuable information about vegetation, LC, water, and more. With global coverage and a revisit time of five days, S2 data is freely available, allowing regular monitoring of various environmental and agricultural applications⁵⁵.

In this work, Harmonized Sentinel-2 MSI: MultiSpectral Instrument, Level-2A (S2-L2A) products, which are available in GEE, were used. The S2-L2A collection contains atmospherically corrected surface reflectance

Level-1	Level-2		
EU- map code	EU- map code	Classes or categories	LUCAS code
100		Artificial land	A11 A12 A13 A21 A22
200		Arable land	See below
	211	Common wheat	B11
	212	Durum wheat	B12
	213	Barley	B13
	214	Rye	B14
	215	Oats	B15
	216	Maize	B16
	217	Rice	B17
	218	Triticale	B18
	219	Other cereals	B19
	221	Potatoes	B21
	222	Sugar beet	B22
	223	Other root crops	B23
	230	Other non-permanent industrial crops	B34 B35 B36 B37
	231	Sunflower	B31
	232	Rape and turnip rape	B32
	233	Soya	B33
	240	Dry pulses, vegetables, and flowers	B41 B43 B44 B45
	250	Fodder crops	B51 B52 B53 B54
	290	Bare arable land	F40 (LU: U111/112/113)
300		Woodland and Shrubland	B71- B77 B81-B84 C10 C21 C22 C23 C31 C32 C33 D10 D20
500		Grassland	B55 E10 E20 E30
600		Bare land and lichens/moss	F10 F20 F30 F40 (LU: other than U111/112/113)
700		Water	G11 G12 G21 G22 G30 G40
800		Wetlands	H11 H12 H21 H22 H23

Table 1. The EU crop map categories. The classification scheme contains two levels: Level-1, with seven broad land cover classes, and Level-2, with 25 detailed classes, including 19 crop types.

values. The “harmonized” aspect of the collection is related to the seamless correction for the reflectance offsets introduced by ESA on 21 January 2022. S2-L2A stores scene classifier information in the so-called Scene Classification (SCL) band, which includes flagged values for cloud, cloud shadow, and haze on a per-pixel basis.

S2 temporal series from 1st January to 31st December 2022 were used in this work. Only the images with a cloud fraction cover lower than 50% were selected. Pixels with a cloud probability higher than 75% and labeled as Saturated or Defective, Clouds High Probability, Cirrus, and Snow / Ice in the SCL information were masked. It should be noted that thresholds used were determined using a visual, trial-and-error approach. To homogenize the spatial resolution of the S2 images with a common spatial resolution, the 20 m bands were resampled into 10 m using the nearest neighbour method.

In this work, the S2 spectral bands utilized include B02-B08, B8A, B11, and B12. These bands were employed to calculate a set of 15 spectral indices and a biophysical parameter called Leaf Area Index (LAI)⁵⁶, providing additional features for the analysis. Five spectral indices focused on vegetation differences: Enhanced Vegetation Index 2 (EVI2)⁵⁷, Green Normalized Difference Vegetation Index (GNDVI)⁵⁸, Leaf Area Index green (LAIg)⁵⁹, Leaf Chlorophyll Content Index (LCCI)⁶⁰ and Normalized Difference Vegetation Index (NDVI)⁶¹. Four of them focused on Soil differences: Bare Soil Index (BSI)⁶², Modified Soil Adjusted Vegetation Index (MSAVI)⁶³, Normalized Difference Tillage Index (NDTI)⁶⁴, and Soil Adjusted Vegetation Index (SAVI)⁶⁵. Two of them take advantage of providing good separability of the components from urban systems: Built-up Land Features Extraction Index (BLFEI)⁶⁶ and Normalized Difference Built-up Index (NDBI)⁶⁷. Modified Normalized Difference Water Index (MNDWI)⁶⁸ and Normalized Difference Water Index (NDWI)⁶⁹ were employed as water indices. Also, a spectral band difference and ratio were calculated: the difference between Red and SWIR1 bands (DIRESWIR)⁷⁰ and the ratio between NIR and Red bands (SRNIRR)⁷¹; both were used in previous studies⁷². Table 2 summarizes this work’s spectral (i.e., bands and indices) information.

The monthly median values of each spectral band and each index were calculated for the months of April to October, supplying 26 features per month. Our previous studies^{25,51} determined that a high number of pixels with missing values are observed in the composite images for the months of January, February, March, November, and December due to cloud coverage. Therefore, the median of the first three months was calculated and referred to as the winter median. No monthly features were calculated in the spectral field for November or December. In addition, the annual 5th, 50th, and 98th percentiles per band and index were also computed. Consequently, a total number of 286 features ($26 \times (8_{\text{monthly median}} + 3_{\text{yearly percentile}})$) are available from S2 data for the year 2022.

Feature Name	Description	
Spectral Bands	B2: Blue (WL: 496.6 nm (S2A) / 492.1 nm (S2B))	
	B3: Green (WL: 560 nm (S2A) / 559 nm (S2B))	
	B4: Red (WL: 664.5 nm (S2A) / 665 nm (S2B))	
	B5: Red Edge 1 (WL: 703.9 nm (S2A) / 703.8 nm (S2B))	
	B6: Red Edge 2 (WL: 740.2 nm (S2A) / 739.1 nm (S2B))	
	B7: Red Edge 3 (WL: 782.5 nm (S2A) / 779.7 nm (S2B))	
	B8: NIR (WL: 835.1 nm (S2A) / 833 nm (S2B))	
	B8A: NIR narrow (WL: 864.8 nm (S2A) / 864 nm (S2B))	
	B11: SWIR 1 (WL: 1613.7 nm (S2A) / 1610.4 nm (S2B))	
	B12: SWIR 2 (WL: 2202.4 nm (S2A) / 2185.7 nm (S2B))	
	Spectral Indices and biophysical parameter	BLFEI: $((B3 + B4 + B12)/3) - B11 / (((B3 + B4 + B12)/3) + B11)$
		BSI: $((B11 + B4) - (B8 + B2)) / ((B11 + B4) + (B8 + B2))$
DIRESWIR: $B4 - B11$		
EVI2: $((B8 - B4)) / ((B8 + B4 + 1)) * 2.4$		
GNDVI: $((B8 - B3)) / ((B8 + B3))$		
LAI: $3.618 * (2.5 * ((B8 - B4) / (B8 + 6 * B4 - 7.5 * B2 + 1))) - 0.118$		
LAIg: $5.405 * ((B8A - B5) / (B8A + B5)) - 0.114$		
LCCI: $B7/B5$		
MNDWI: $((B3 - B11)) / ((B3 + B11))$		
MSAVI: $0.5 * (2 * B8 + 1 - \sqrt{(2 * B8 + 1)^2 - 8 * (B8 - B4)})$		
NDBI: $((B11 - B8)) / ((B11 + B8))$		
NDTI: $((B11 - B12)) / ((B11 + B12))$		
NDVI: $((B8 - B4)) / ((B8 + B4))$		
NDWI: $((B8 - B11)) / ((B8 + B11))$		
SAVI: $((B8 - B4)) / ((B8 + B4 + 0.5)) * 1.5$		
SRNIRR: $B8/B4$		

Table 2. Spectral bands and indices, as well as a biophysical parameter derived from S2 [Notation: NIR = Near Infrared; SWIR = Shortwave Infrared].

Sentinel-1 Data. The S1 satellite mission is another dual-sensor constellation deployed under the EU's Copernicus program, with Sentinel-1A (S1A) and Sentinel-1B (S1B). The S1 constellation provides global coverage and a revisit time of 6 days. However, S1B failed in December 2021, reducing the revisit to 12 days.

S1 is a synthetic aperture radar (SAR) operating at the C-band frequency (5.4 GHz). SAR is an active remote sensing method that transmits microwave signals in vertical polarization (V) towards the surface of the Earth and receives the backscattered signals in both vertical and horizontal polarization (H). Radar backscatter is influenced by the geometry and dielectric characteristics of the irradiated surface elements. For agricultural land, crop canopy (leaf structure and density, canopy water content) and soil surface characteristics (soil moisture, soil roughness) determine backscatter intensity. Due to the active nature of the sensor and the relative insensitivity of the C-band to atmospheric conditions, S1 can acquire both day and night under all weather conditions⁷³. Combining ascending (local evening passes) with descending (local morning passes) acquisitions results in a revisit of more than 6 (12) days, though with different incidence configurations. Over Europe, all ascending and descending orbits are acquired. Although S1 can acquire in different beam modes, the interferometric wide (IW) mode, with 10 m sampling spacing, is the default mode over global land.

Copernicus S1 data are full, free, and openly accessible. However, they are provided in Level-1 formats (GRD and SLC), which are not application-ready data formats. S1 Level-1 data needs to be processed to geocode and calibrate backscattering coefficients. In GEE, each GRD scene is preprocessed following a standard recipe scripted in the S1 SNAP Toolbox (<http://step.esa.int>) and using the SRTM 90 m DEM for geocoding. The resulting sigma naught backscatter coefficients (σ^0) are made available in the COPERNICUS/S1_GRD_FLOAT collection. For IW mode, both VV and VH bands are used.

The focal median method, a speckle filter with a circular kernel with a radius of 30 m, was applied to reduce noise in the images⁷⁴.

The monthly median and the yearly 5th, 50th, and 98th percentiles were also calculated from the following bands and indices: VV, VH, VV/VH, VV-VH, Ratio Vegetation Index (RVI)⁷⁵, Normalized Difference Polarization Index (NDPI)⁷⁶, and modified Dual Polarimetric SAR Vegetation Index (DPSVIm)⁷⁷, which are described in Table 3. As microwave data is not affected by cloud coverage issues, monthly medians were calculated for all the months. Subsequently, a total number of 105 features ($7 \times (12_{\text{monthly median}} + 3_{\text{yearly percentile}})$) are available from S1 data.

Auxiliary temperature and masking data. Along with the monthly and yearly S1 and S2 features, the study also incorporated data on LST and elevation to account for the study area's climatic, topographic, and ecological

Feature Name	Description
Microwave features	VV: Single co-polarization, vertical transmit/vertical receive
	VH: Dual-band cross-polarization, vertical transmit/horizontal receive
	VV/VH: The ratio between the VV polarization and the VH polarization
	VV-VH: The difference between the VV and the VH polarizations
	RVI: $\frac{4\sigma_{VH}^0}{\sigma_{VV}^0 + \sigma_{VH}^0}$
	NDPI: $\frac{\sigma_{VV}^0 - \sigma_{VH}^0}{\sigma_{VV}^0 + \sigma_{VH}^0}$
	DPSVIm: $\frac{\sigma_{VV}^0 * \sigma_{VV}^0 + \sigma_{VV}^0 * \sigma_{VH}^0}{.2}$

Table 3. Microwave bands and indices derived from S1 and used for LULC classification.

factors. When combined with spectral and microwave features, temperature data can help to enhance the discriminative power of the classification algorithm, particularly in regions where similarities exist among features, but are shifted due to climatic differences. MOD21C3 is a monthly composite LST product derived from MODIS data from the MODIS sensor mounted on Terra and Aqua satellites⁷⁸. Using multiple daily measurements, this dataset produced representative LST values for every month, enabling the assessment of changes in temperature over time at a spatial resolution of 1 km. The LST_Day feature containing the monthly average daytime LST for 2022 was used in this study.

Furthermore, the investigation utilized the Global Digital Elevation Model (GDEM) v3, which was extracted from the ASTER instrument aboard the Terra satellite⁷⁹. Providing a 30 m grid-spaced terrain model of the surface of the Earth, this elevation dataset captures variations in the height and shape of the terrain.

Considering 12 monthly features for LST data and one for DEM data, a total of 13 auxiliary features were used in this study.

In order to apply thematic masks on the final produced map, Global ALOS DSM (Digital Surface Model) data representing the Earth's surface, encompassing both natural and artificial surfaces, stored in GEE as the JAXA/ALOS/AW3D30/V3_2 collection⁸⁰ at a spatial resolution of 30 m was utilized. In addition, data from the CORINE Land Cover (CLC) 2018 inventory⁸¹, a pan-European inventory that categorizes land cover into 44 thematic classes for the reference year 2018, was used to mask the final map. The CLC dataset provides information on LU and LC across Europe at a spatial resolution of 100 m.

Prior to application, using the nearest neighbour algorithm, all auxiliary data was resampled to a spatial resolution of 10 m.

Classification process. *Classification.* The Random Forest (RF) classifier was employed for this work based on its strength, speed, and robustness to the noise of the target data⁷². First introduced by Breiman⁸², the RF classifier is a robust ML algorithm capable of yielding high classification accuracy and quantifying the importance of features. As an ensemble technique, RF uses bagging (bootstrap + aggregation) and builds an ensemble of random decision trees. Rather than training all data, bootstrap trains a subset of samples and features. Bags are created from selected samples, while remnant samples are called out-of-bag samples (OOB). The output of all trained trees is combined, typically using the majority vote, which reduces variance and increases classification accuracy⁸³.

This research conducted a comprehensive classification process based on a hierarchical classification particularly focused on crops. A first classification of the seven major landcover classes (Level-1 in Table 1) was followed by a detailed classification of the Arable land class into 19 specific crop types or crop groups (Level-2 in Table 1), providing the primary map.

However, due to cloud contamination in certain areas, gaps appeared in the primary map, impacting its completeness. Therefore, a second process was performed to produce the gap-fill map used to fill the gaps in the primary map and obtain the final classification map. In this case, the S2 monthly features, which are sensitive to cloud cover, were intentionally excluded, and additional training samples from the affected areas were added.

A feature selection process and parameter tuning of RF were performed to optimize the model efficiency of GEE. The feature selection was done locally using the Scikit-learn package⁸⁴ in Python. The features were selected based on their importance provided by the RF model; only those features presenting importance more than their standard deviation were retained for the classification. The number of features was reduced from 404 (286 from S2, 105 from S1, and 13 from auxiliary data) to 184 for the primary classification and from 196 (78 from S2, 105 from S1, and 13 from auxiliary data) to 109 for the gap-fill classification. The utilized features for both primary and gap-fill classification are shown in Supplementary Table 1.

The RF optimization was based on⁵¹, where the RF had a proper performance for EU-wide LULC classification with 200 trees, and setting the minimum number of samples required to be at a leaf node and the feature numbers per split by default (i.e., one and the square root of total number of features, respectively). Considering these parameters, a trade-off between the number of trees and the minimum number of samples in a leaf node was found to achieve the appropriate outcome in both classification phases, which is shown in Table 4. It is worth mentioning that the other parameters were set to their default values.

Train and test data. In preparation for the training and testing of the classification model, the feature data values were extracted at the locations of the LUCAS 2022 field data, specifically for the centroid point of the 134,670

Parameter	Optimized value for classification Level-1	Optimized value for classification Level-2
Number of trees	130 for primary map 115 for gap-filled process	150 for primary map 130 for gap-filled process
Min_Leaf Population	2	2

Table 4. RF tuned parameters settings in GEE for the first (7 major classes) and second (19 arable land classes) classification phases, valid for both primary and supplementary classification stages.

available LUCAS 2022 Copernicus polygons. Subsequently, Level-1 and Level-2 class labels were assigned following Table 1. By removing points whose LC labels were unavailable in Table 1, the total number of samples resulted in 134,260, with a further reduction to 133,829 after deleting samples containing null values in their LST indicators.

The study area encompasses eight distinct climate regions: Alpine, Atlantic, Black Sea, Boreal, Continental, Mediterranean, Pannonian, and Steppic. These climatic regions were merged into Mediterranean and Non-Mediterranean according to the Biogeographical regions map in Europe and used as stratification before splitting the samples into train and test data⁸⁵. In total, 29,826 samples were in the Mediterranean region, whereas 104,003 samples belonged to the non-Mediterranean region. This stratification was done to ensure an equal portion share of train and test data in both zones.

Based on the LUCAS code in Table 1 (excluding Water, Wetlands, and Bare land classes, which followed the Level-2 scheme), the samples in both regions were randomly divided into train and test datasets using a 75% and 25% split ratio, respectively. Subsequently, the regional train and test datasets were merged pairwise, resulting in final train and test datasets with 100,360 and 33,456 samples. Thirteen samples were excluded from the process due to an inadequate number of available subclass samples, which was insufficient for a proper division.

The 18,933 samples that fall in the gaps of the primary classification were only used in the gap-filled classification procedure. Therefore, the available training samples for the primary classification were 81,427. Figure 3 depicts the EU-27 area divided into two climatic zones. It illustrates the distribution of train and test samples across the entire area while highlighting the distinction between primary and additional gap-filled samples. The additional gap-filled train samples are distinguished by their distinct yellow colours, while the primary train samples are depicted in green. Additionally, the test points are displayed in black.

Train data balancing. In addition to the central point for each of the LUCAS Copernicus module polygons, labelled data could potentially be extracted for other points within the polygons. While incorporating additional data samples for training and testing purposes can expand the dataset, it is essential to consider the potential drawbacks. Increasing the number of samples can lead to longer computational times for training the models without guaranteeing a substantial improvement in classification accuracy. Extracting points from homogeneous polygons may result in redundant data with similar characteristics, which can lead to overfitting of the trained model²⁵. However, to address the imbalanced distribution of samples amongst the different classes in the primary training dataset, efforts were made to achieve a more balanced distribution by reducing samples in specific classes and oversampling others. The abundance of certain classes was reduced by eliminating similar samples; meanwhile, increasing samples for sparse classes was increased by including additional samples from within polygons, not limited to centroids.

In balancing the training data, various choices were made. To assess the similarity between samples within each polygon (excluding the centroid), the Euclidean Distance (ED) metric in feature space was calculated. The samples were then sorted according to the ED values (most dissimilar first), and only certain proportions were retained or excluded. The process was run incrementally/decrementally, and at each step, an RF classifier was trained using the existing samples, and the results were evaluated using the test samples. A good balance between the distribution of samples and classification accuracy was sought. The classes with a substantial number of samples (i.e., Woodland and Shrubland and Grassland) were reduced, considering only 35% of the dissimilar samples, specifically for samples belonging to the C10 (Broadleaved Woodland) and E20 (Grassland without tree/shrub cover) LUCAS classes. The balanced classes selected up to a minimum of 1500 samples (if available) for the Arable land in Level-2, and 4000 for Water, Wetland, and Bare land classes in Level-1 (see Table 1).

Post-classification process. *Spatial filtering.* The produced map underwent a spatial post-processing (applied in GEE) to reduce spatial noise and small patches in the classified map, resulting in a smoother and more consistent map. The process begins by evaluating the number of interconnected identical pixels between each pixel on the map. It is restricted to a maximum of 30 pixels, as well as horizontal and vertical orientations (4-connected), and findings are given the name “patch size”. A refinement was then applied to pixels with patch sizes smaller than 20. It consists of applying a filtering approach centred around a 10-pixel radius and a square kernel. It is important to note that all post-processing values were assigned practically.

Thematic masking. Given the focus of the LUCAS dataset on vegetation cover and a lack of training samples in mountainous (or rocky) or snow-covered areas, these areas are mostly classified incorrectly as Artificial land and Water, respectively. To alleviate this issue, two external layers were utilized for masking. First, using Global ALOS DSM data, pixels identified as Artificial land with elevations exceeding 1000 meters and slopes greater than 10 degrees were masked. In addition, using CLC 2018 data, pixels where their CLC value were 322 (Moors and heathland), 331 (Beaches, dunes, sands), 332 (Bare rocks), 333 (Sparsely vegetated areas), 335 (Glaciers and perpetual snow) were masked.

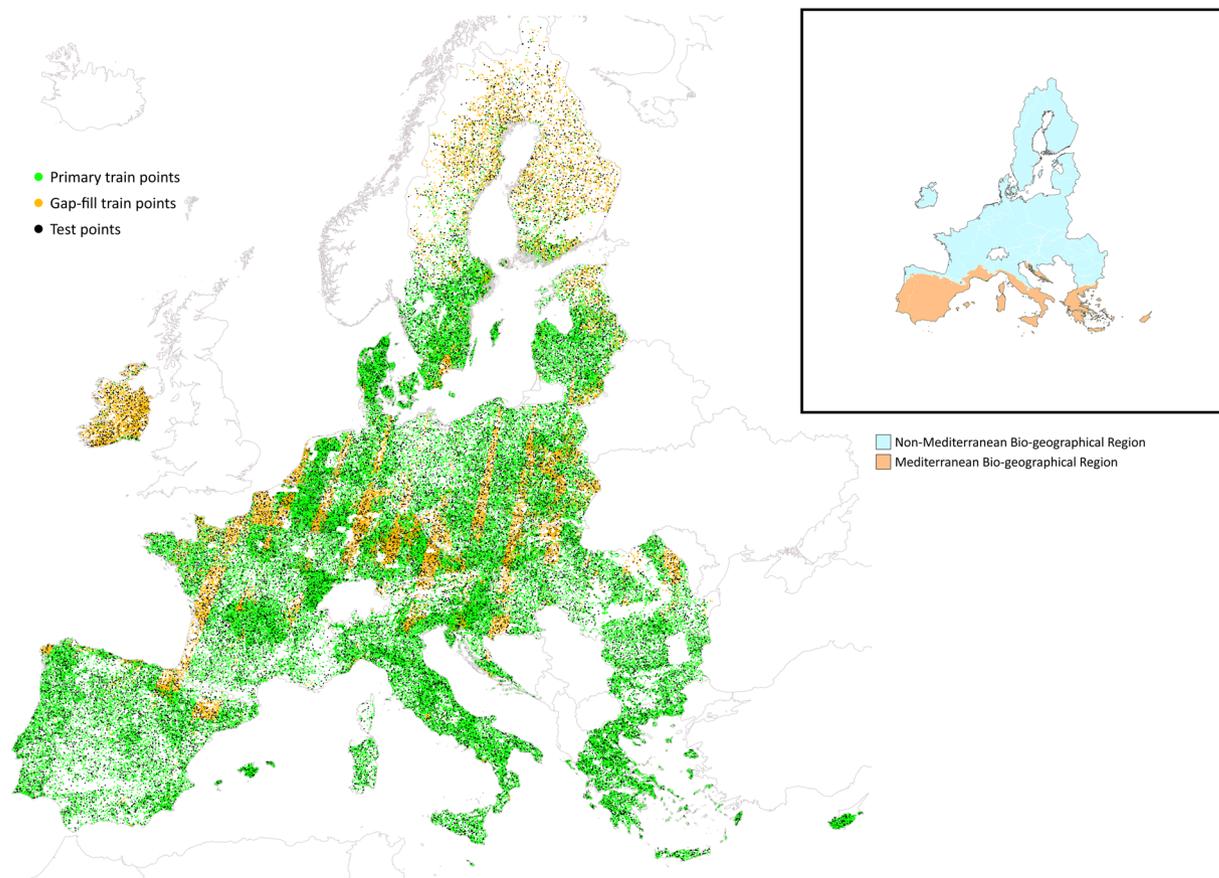


Fig. 3 Distribution of train and test samples in EU-27 with distinguishing primary and additional gap-fill data.

Accuracy assessment and spatial analysis. To ensure the reliability and precision of the EU LULC map, a comprehensive evaluation was conducted using both accuracy assessment and spatial analysis methods. In the accuracy assessment, the classification results were validated against independent reference datasets, ensuring that the map accurately represented land cover types. Additionally, the spatial analysis examined the spatial distribution and consistency of classified land cover across different regions. By combining these approaches, a holistic view of the map's performance is provided that not only verifies overall accuracy but also highlights regional variations.

Four accuracy assessment and spatial analysis approaches were employed to evaluate the final EU LULC map. The final map was evaluated with LUCAS test data as well as additional field samples from the European Monitoring of Biodiversity in Agricultural Landscapes (EMBAL) dataset. A second validation approach was based on a regional subset of Geospatial Aid Application (GSA) data, which is based on the 2022 farmer declaration data. Finally, aggregated area sums for several main crops obtained from the EU LULC map are compared to the corresponding official Eurostat national statistics.

LUCAS test data evaluation. The proposed classification workflow was evaluated by the confusion matrix (CM) assessing the 33,456 test samples. Five assessment metrics were determined from the CM: User's Accuracy (UA), Producer's Accuracy (PA), F1-score (a weighted average of UA and PA), Overall Accuracy (OA, representing the ratio of correctly predicted samples to the total samples), and Kappa coefficient (κ , quantitative measure of reliability the classification).

EMBAL data analysis. EMBAL is a program aimed at assessing and monitoring biodiversity in agricultural landscapes throughout the EU, contributing to a variety of environmental policies. Within the frame of EMBAL, field surveys and sampling are conducted to collect valuable data on various aspects of biodiversity in agricultural areas, including plant and animal species and their habitats. The EMBAL 2022 dataset contains 3000 square plots with 500×500 m dimensions, each having various parcels with the relevant LC classes⁸⁶.

To extract the validation set from EMBAL data, firstly, sensitive and irrelevant attributes to the study, like user ID and geometries of EMBAL parcels, were removed. Secondly, label recoding was performed for certain classes to allow legend matching with LUCAS (Supplementary Table 2). From this set, all records from LC classes Arable land (eA), Fallow (eB), Grassland (eC), Shrubland (eD), and Non-agricultural elements (eN), which have a valid lc1 attribute and with an area of more than or equal to 50 square meters were selected, amounting to 34,370 instances.

Region	Description	Area (Mha)	The area ratio of the total (%)	Parcels
bewa2022	Wallonia (Belgium)	0.801	94.1	341,968
dk2022	Denmark	2.658	89.0	576,544
ee2022	Estonia	0.973	92.7	175,968
nl2022	Netherlands	1.850	89.8	758,504
sk2022	Slovakia	1.820	94.1	263,067
Total/Avg	—	8.102	91.9	2,116,051

Table 5. Results of the evaluation based on the GSA data.

This research used the centroids from these parcels, representing 18 different LCs, to evaluate the final produced map. Notably, the extracted dataset focuses solely on vegetation and does not include classes such as Artificial land, Bare land and lichens/moss, Water, and Wetlands. Additionally, within the Arable land category, Common wheat, Durum wheat, and Triticale were merged into a single Wheat class (210 in Level-2 code), while the Other cereals class had no samples available. Supplementary Table 3 displays the LC classes of the EMBAL 2022 dataset alongside their equivalent EU map codes and the number of validation points for each class.

Comparing with GSA. In the second approach for the 2022 crop map assessment, we expanded our evaluation and compared our classes with vector parcels derived from the Geospatial Aid Application (GSA) across several EU regions. The GSA refers to the annual crop declarations made by EU farmers for Common Agricultural Policy (CAP) area-aid support measures. Each GSA record represents a polygon of an agricultural parcel with one crop (or a single crop group with the same payment eligibility), providing a valuable dataset for comparison.

For the 2022 assessment, we incorporated GSA data from bewa2022 (Wallonia in Belgium, BE), dk2022 (Denmark, DK), ee2022 (Estonia, EE), nl2022 (Netherlands, NL), and sk2022 (Slovakia, SK). While the GSA system is implemented across the EU, the design and operation of each region's data set vary. Each GSA has its own nomenclature for crop types and crop practices eligible for regional support schemes.

Maintaining a focus on crop classes covering at least 0.5% of the cumulative GSA area in each selected region, we ensured that our assessments targeted agriculturally significant areas and crop types. Following this principle, we achieved a total area coverage ranging from 89.8% (lowest coverage country) to 97.1% (highest coverage country) of the total parcels for each respective GSA. We extracted a total of 2,116,051 parcels spanning over 8.102 million hectares.

We then compared the declared crops, mapped to the LUCAS legend, to the predicted classes within each GSA parcel from our pixel-based classification. This comparison provided a comprehensive assessment of how practically applicable our classification algorithms are. Table 5 summarizes the result of the GSA data used in the study.

Comparing with national statistics. The official national statistical data available through Eurostat for the main crops in the EU-27 is the final database used to evaluate the EU-LULC map. The dataset (cultivation/harvested/production) (1000 ha) is available for download on (https://ec.europa.eu/eurostat/databrowser/view/tag00103/default/table?lang=en&category=agr.apro.apro_crop.apro_cp.apro_cpsh). In this case, the two available EU-LULC years were analyzed (i.e., 2018 and 2022). In order to facilitate this comparison, national area data was taken from the dataset “Area (cultivation/harvest/production) (1000 ha) by NUTS-1 regions”. A semantic comparison was conducted between the Eurostat crop classes and the EU crop map legends (see Table S17 in²⁴) to ensure they are compatible.

By comparing the frequency distribution of values in one dataset within classes of another dataset, the estimated areas of EU crop maps were compared to the corresponding Eurostat data. For each of the nine main crops in the EU, the Pearson correlation coefficient was calculated in order to assess the strength of the correlation. The distribution of these variances is subsequently aggregated to deduce conclusions at an individual crop level, incorporating insights from both 2018 and 2022 data in the analysis.

Expansion of the classification model to an external territory. This study aims to develop the LULC map for the EU-27 area by utilizing training and test data from inside the region. However, the research scope extends beyond this region to include Ukraine, which is adjacent to the EU-27. Contrary to the EU-27 area, no label features were available for the Ukraine territory for use in model training.

Ukraine's climate is characterized by a Continental climate in the western part and a Steppic climate in the eastern part. This places it in the non-Mediterranean climatical region as defined in the Classification Process section. Agriculture is Ukraine's largest export industry, with over 41 million hectares of agricultural land covering 70 percent of the country (<https://www.usaid.gov/ukraine/news/private-sector-frontlines-land-reform-unlock-ukraines-investment-potential>). The most cultivated crops in Ukraine include wheat, sunflower, maize, barley, potatoes, soybeans, and rapeseed, all of which are included in the classification schemes of the trained model (<https://www.statista.com/statistics/1381043/ukraine-planted-area-by-crop/>).

As part of this investigation, the efficacy and transferability of the classification model trained in one area and validated in a different external region was assessed. By analyzing the results of this study, a better understanding of the reliability of the classification model and its potential broader applicability across external territories will be gained.

The analysis involved a comprehensive comparison between our generated map and the independently generated high-resolution crop map produced by the Kyiv Polytechnic Institute (KPI) team for the year 2022 (<https://ukraine-cropmaps.com>). This step was pivotal in understanding any deviations, anomalies, or similarities between the two mapping results, providing a robust assessment of our model's performance against an established standard.

Data Records

The dataset is available in the **JRC data catalogue**⁸⁷ (<https://doi.org/10.2905/555e5d1d-1aae-4320-a716-2e6d18aa1e7c>) where users can find the direct FTP access to the map (<https://jeodpp.jrc.ec.europa.eu/ftp/jrc-opendata/EUCROP/2022/>) and the Web Map Service (WMS) (<https://jeodpp.jrc.ec.europa.eu/jeodpp/services/ows/wms/landcover/eucropmap>).

The dataset is reprojected to the Lambert Azimuthal Equal-Area projection (ETRS89-LAEA, EPSG:3035) and it includes individual maps for the EU-27 and Ukraine (GeoTiff), as well as a collection of tiles of 327 km (in width) and 327 km (in height). The tile index is available in ESRI Shapefile format.

The maps are also available as ImageCollection in the GEE catalogue (https://developers.google.com/earth-engine/datasets/catalog/JRC_D5_EUCROP/2022) which contain required data for EU27 and Ukraine.

To facilitate access to the EU crop map dataset to a large audience, a web application has been created that simplifies consultation and guides users to deepen the analysis. This is a dashboard created in the JupyterLab environment, the most used web application for data science tasks, through a Jupyter Notebook. The open-source Voilà software (<https://voila.readthedocs.io/en/stable/>) takes care of displaying notebooks without showing the code cells, rendering only the output produced by each of the cells. In this way, even very complex applications that are focused on data can be created, allowing users to interact with the data themselves through dedicated graphical interfaces. The EUcropmap dashboard (available here <https://jeodpp.jrc.ec.europa.eu/eu/dashboard/voila/render/REFOCUS/EUcropmap/EUcropmapExplorer.ipynb>) and in Supplementary Figure 1) was created using Python programming and a series of libraries that help build the page, present information, and manage user interactions. Among these libraries is VOIS⁸⁸, a library developed by the JRC and available as open source in the code.europa.eu repository, which supports the developer by simplifying the activities of building captivating visualizations within the Voilà environment. The EU crop map dashboard allows users to quickly compare the 2018 and 2022 versions of the dataset and highlight a single crop through a customizable legend in which the user can turn each class on or off. During zoom and panning operations on the map, the consistencies of the classes in the area visible on the map are calculated in real-time and shown in the legend itself in the form of a horizontal bar graph. Finally, it provides an on-the-fly computation of crop diversity, allowing for quick and efficient insights⁸⁹.

Technical Validation

Training data balancing result. Table 6 presents the number of samples per class and the selected number after the balancing process. Additionally, the F1-score using an RF classifier for the two levels is reported. The procedure focused on Woodland and Shrubland, Grassland, Bare land, Water, and Wetland classes in the Level-1 scheme and the Arable land class in the Level-2 scheme.

The total number of primary training samples increased from 81,427 to 86,831 after the balancing. The impact of the process is noticeable in the results. The F1-scores increase for classes that expand in the number of samples. However, no significant change was observed in the F1-scores for the Woodland and Shrubland, and Grassland classes, which underwent sample reduction. Notably, no modifications were made to the Artificial land samples.

In terms of accuracy, the OA slightly decreased from 78.9% to 78.4% in Level-1, whereas it increased from 70.9% to 71.1% in Level-2. The κ also showed a marginal change from 0.71 to 0.70 in Level-1 and from 0.62 to 0.63 in Level-2. Overall, there was a slight improvement in the Level-2 scheme validation results.

Crop map over EU. *Insights From EU Map.* For analysis purposes, the EU crop map is re-projected to the Lambert azimuthal equal-area (ETRS89-LAEA, EPSG:3035) projection. The final LULC map presented in Figure 4 comprehensively depicts 25 distinct classes, encompassing around 3.973 million square kilometres (Mkm²) of land area. This value was calculated after performing the masking process in post-classification.

The resulting map was visually inspected and spatially consistent with high spatial resolution base maps and existing products such as the EU crop map 2018, particularly in terms of the major LC categories. For each class, a simple area estimation based on pixel counting is presented in Supplementary Table 4.

Figure 5 illustrates parcels of varying sizes and crop types across different regions of EU-27. These classifications were achieved without prior knowledge, relying solely on non-parametric classification methods. Parcel boundaries are clearly distinguishable even when using pixel-based classification methods.

In Spain - Castile and León (Figure 5a), the major crops are Common wheat, Sunflower, and Rye, as well as some parcels of fodder crops and oats. Figure 5b (France - Centre-Val de Loire) shows the classifier's ability to distinguish various crops in a region where crops such as barley, common wheat, dry pulses, maize, potatoes, rape and turnip rape, sugar beet, and sunflower are well differentiated. Figure 5c (Austria - Lower Austria) displays the various crop types (Common wheat, Dry pulses, Maize, Potatoes, Rape and turnip rape, Sugar beet) in different parcel sizes along with the Woodland and Shrubland covered National Park Donau-Auen and Danube River. In Romania - Sud-Muntenia, (Figure 5d) very large parcels along with very small parcels are detectable which are mostly covered in Common wheat, Maize, and Rape and turnip rape.

Level-1 (7 major classes)					Level-2 (25 minor classes)									
Level-1 code	Initial samples	Balanced samples	Initial F1-score	Balanced F1-score	Level-2 code	Initial samples	Balanced samples	Initial F1-score	Balanced F1-score					
100	9063	9063	57.3%	58.6%	100	9063	9063	57.6%	58.7%					
					211	4677	4677	64.2%	65.6%					
					212	629	1500	25.5%	36.9%					
					213	2315	2315	51.0%	51.2%					
					214	597	1500	11.7%	30.9%					
					215	652	1500	22.1%	31.6%					
					216	3136	3136	78.8%	80.4%					
					217	16	73	0.0%	0.0%					
					218	530	1500	4.4%	21.6%					
					219	135	771	4.4%	18.5%					
					200	20827	31824	87.5%	86.3%	221	393	1500	65.0%	65.4%
										222	383	1500	80.0%	80.6%
										223	147	884	0.0%	11.1%
										230	395	1500	28.4%	39.0%
										231	914	1500	70.9%	71.3%
										232	1497	1500	73.9%	73.8%
										233	270	1500	37.9%	60.4%
										240	673	1500	24.7%	29.6%
										250	1635	1635	11.2%	23.0%
290	1833	1833	37.4%	34.9%										
300	29116	21431	84.5%	84.5%	300	29116	21431	84.5%	84.4%					
500	21419	12513	72.7%	71.1%	500	21419	12513	71.3%	71.9%					
600	561	4000	6.1%	18.1%	600	561	4000	5.1%	17.6%					
700	238	4000	0.0%	34.7%	700	203	4000	18.2%	32.4%					
800	203	4000	18.2%	31.1%	800	238	4000	0.0%	34.0%					

Table 6. The number of primary training samples per class used in the training step for both levels before and after balancing, as well as the performance of F1-score.

Accuracy Assessment and Spatial Analysis Results. *LUCAS test points.* Based on the evaluation of 32,628 out of 33,456 test samples (828 samples were unavailable due to the masking process), a confusion matrix was generated to assess the classification accuracy of seven major LC classes (Table 7). The three main vegetation classes, namely Arable Land, Woodlands and Shrubland, and Grassland, demonstrated excellent F1-scores ranging from 73% to 87%. These high scores indicate a robust capability to accurately classify these specific LC classes. On the other hand, discriminating between the Bare Land, Water, and Wetlands classes proved to be more challenging, conceivably due to the limited number of training samples available for these classes. Nineteen test samples could not be evaluated due to possible small remaining gaps in the final LULC map. The OA was determined to be 79.3%, indicating a reasonably high level of correctness in the classification results. The κ was 0.71, indicating substantial agreement beyond chance. Besides, Arable land is well captured, achieving 81.9% and 93.4% of UA and PA, respectively.

Concerning analyzing crop types, the confusion matrix was expanded to Level-2, which contains 25 classes, six majors, and 19 crop types. The results are summarized in Table 7, where the OA and κ are reported as 70.6% and 0.63, respectively. It is worth noting that the weighted confusion matrix based on the estimated area coverage of each class (described in Supplementary Table 4) and using the instructions in⁹⁰ can be found in Supplementary Table 5.

Focusing on the 19 crop classes, four classes: Maize, Sugar beet, Sunflower, and Rape and turnip rape, achieved F1-scores above 69%, indicating a high level of UA and PA. However, for Rye, Rice, Other cereals, Other root crops, the F1-scores were notably lower, falling below 30%. This suggests challenges in distinguishing these LC classes from others. The remaining 11 classes showed F1-scores ranging between 30% and 69%.

EMBAL validation dataset. The classification accuracy of available 33,308 out of 34,370 EMBAL validation samples containing 18 LC classes was evaluated. Notably, the number of 1062 samples were located in masked areas. A confusion matrix (Table 8) was calculated, revealing an OA of 64.4% and a κ of 0.54.

According to Table 8, similar to the results of the LUCAS, classes of Maize, Sunflower, Rape and turnip rape, Woodlands and Shrubland, and Grassland, demonstrated notable F1-scores exceeding 69%. However, the performance of the Barley, Potatoes, Dry pulses, vegetables, and flowers, and Bare arable land classes was comparatively weaker, showing a decline of more than 20% in their F1-scores Table 9.

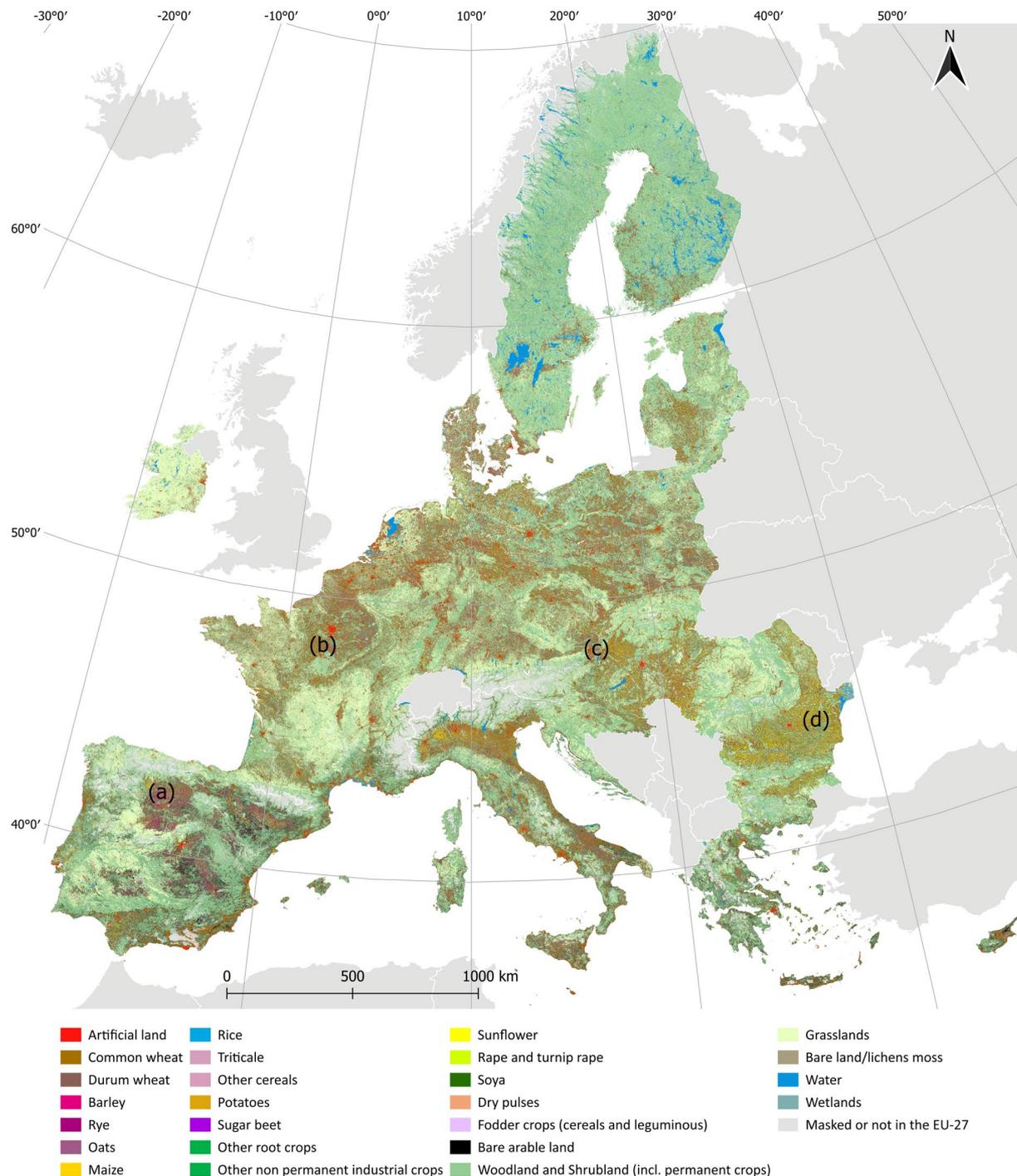


Fig. 4 EU-27 Land Use and Land Cover map at 10 m pixel size for the year 2022. The letters refer to zoom views in Figure 5.

Assessment with EU farmers' declarations (GSA). An alternative evaluation was conducted utilizing an independent data source: farmers' declarations. As detailed in the methodology section, parcel information and their associated crop types were derived from these declarations for crop classes representing at least 0.5% of the total area in the GSA. The selected features were then transformed to align with the grid of the EU crop map, and the GSA crop classes were translated into the corresponding EU LULC codes. Confusion matrices were derived for each region, and accuracies were then computed.

OA varies across the regional data sets, with 91.5% in the Netherlands (nl) being the highest (Supplementary Table 6), followed by Belgium Wallonia (bewa) with 89.6% (Supplementary Table 7) and Denmark (dk) with 84.2% (Supplementary Table 8), respectively. Estonia (ee) (Supplementary Table 9) and Slovakia (sk) (Supplementary Table 10) have significantly lower OA of 73.6% and 56.0%, respectively.

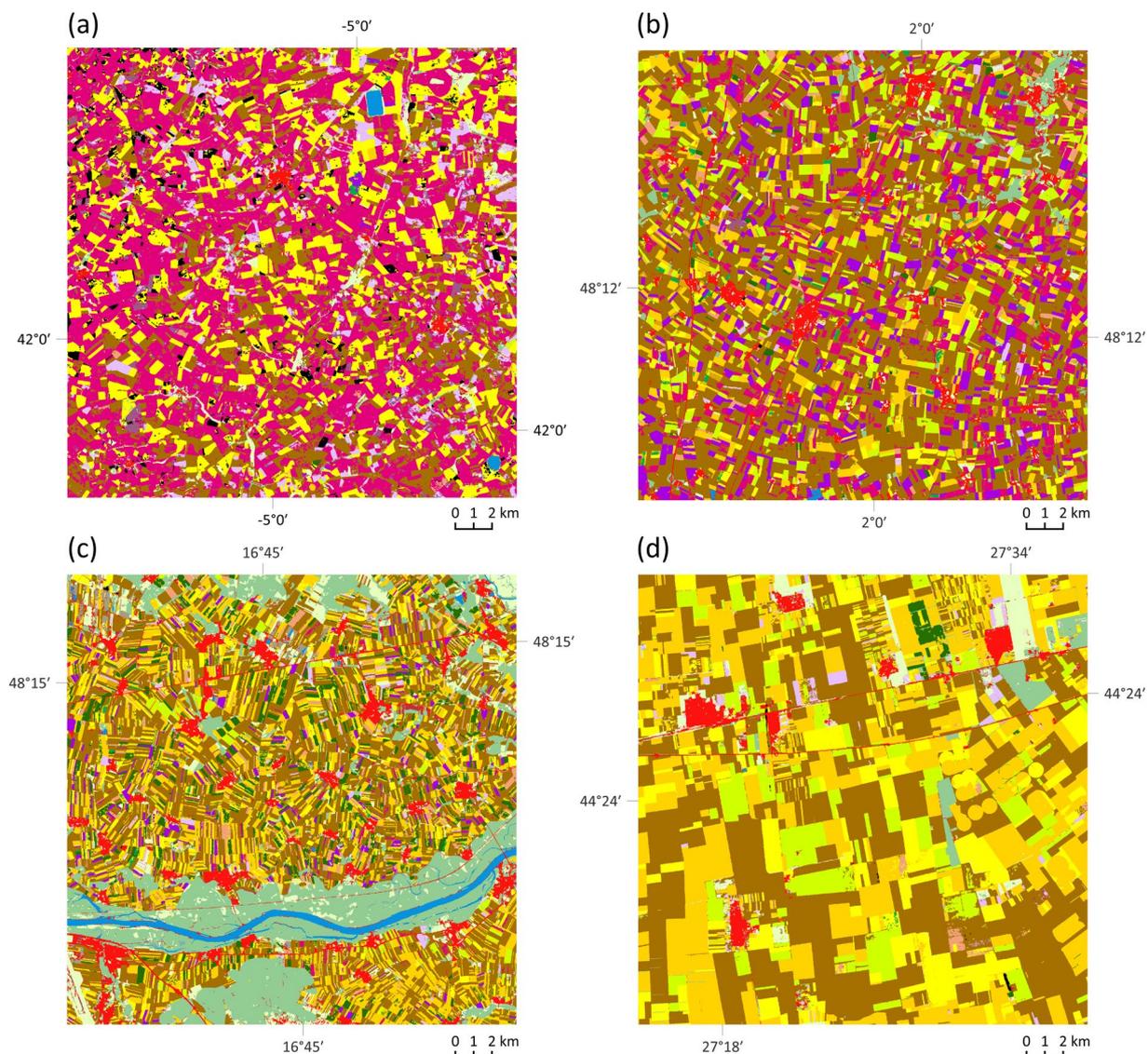


Fig. 5 EU-27 Land Use and Land Cover map for a subset of regions in (a) Castile and León (Spain), (b) Centre-Val de Loire (France), (c) Lower Austria (Austria), (d) Sud-Muntenia (Romania). Legend is presented in Figure 4.

While the overall accuracies give only limited understanding, as classes from GSA are not the same as from the crop map, PAs are the most interesting to investigate. The PAs for crops are derived from the confusion matrices (Supplementary Tables 6–10) and illustrated in Figure 6. Supplementary Figure 2 represents the proportion of pixels assigned to a particular GSA crop type that were accurately classified. In the bewa22 region, the crops Sugar beet, Rape and turnip rape, and Maize show PAs of approximately 90.5%, 94.3%, and 88.9%, respectively, indicating a strong correlation with independent data. However, the category “Other cereals” is significantly lower at a PA of around 20.6%, suggesting discrepancies.

For the ee22 region, Rape and turnip rape is notable with a PA of 97.7%, while Barley, Maize, and the category “Dry pulses, vegetables and flowers” have moderate accuracies ranging from 60.1% to 66.2%. Crops Rye, Oats, and “Fodder crops” have PAs below 50%, pointing to potential misalignments with the actual data. In dk22, the accuracies for Maize, Sugar beet, Rape and turnip rape, and Potatoes are impressive, all exceeding 90%, demonstrating reliable crop mapping. Barley also shows a commendable PA of 89.7%. However, Oats and Rye present lower accuracies, indicating room for improvement. The Netherlands region nl22 exhibits high accuracies for Potatoes, Sugar beet, and “Other root crops”, all with PAs above 90%. However, the “Dry pulses, vegetables and flowers” category shows a lower accuracy, signalling an area where the crop map might be refined. Finally, in Slovakia, sk22, Maize shows a moderate PA, and Sugar beet has a slightly higher accuracy, yet Barley is lower, suggesting some level of misclassification. Rye, Oats, and “Fodder crops” are even less accurate, indicating significant discrepancies from the independent data.

The EU crop map 2022 displays strong classification accuracies for crops like Sugar beet, Rape and turnip rape, and Maize, with PAs frequently surpassing 90%, indicative of a reliable algorithm for these crops. In contrast, crops such as Other cereals, Rye, and Oats show lower accuracies, hinting at potential issues with spectral

Level-1 Code	100	200	300	500	600	700	800	Total	UA	F1-Score
100	1990	123	515	738	60	10	6	3442	57.8%	58.1%
200	333	8065	240	1157	37	8	5	9845	81.9%	87.3%
300	770	84	9901	1153	49	19	69	12045	82.2%	85.0%
500	272	350	547	5809	8	7	29	7022	82.7%	73.0%
600	33	12	13	11	27	1	1	98	27.6%	18.7%
700	12	1	4	11	10	19	5	62	30.6%	29.7%
800	2	0	20	17	0	2	73	114	64.0%	48.3%
Total	3412	8635	11240	8896	191	66	188	32628	OA = 79.3%	
PA	58.3%	93.4%	88.1%	65.3%	14.1%	28.8%	38.8%		$\kappa = 0.71$	

Table 7. The confusion matrix obtained from the classification for Level-1 classes. [Notation: UA = User's Accuracy, PA = Producer's Accuracy, OA = Overall Accuracy, Kappa coefficient = κ].

overlap, training data quality, or image resolution. The observed regional variability in accuracy also suggests that localized conditions affect classification success, pointing to a need for region-specific algorithm adjustments and improved data handling to enhance overall mapping precision.

National Statistics. The crop area data for the EU crop map 2018 and 2022 is extracted and compared across the EU-27 countries. As shown in Figure 7, the calculated Pearson's correlations (R) range from 0.79 in 2022 (0.78 in 2018) for maize to 1 for Rape and turnip rape. While the robustness of the estimation for Rape and turnip rape, Sunflower, and Sugar beet is confirmed in 2022, we observe a larger overestimation of Rye, Potatoes, and Durum wheat compared to 2018. Barley is slightly underestimated in 2022 compared to 2018. Last, there is an overestimation for common wheat both in 2018 and 2022, which can potentially be attributed to commission errors originating for other cereal classes to the most common cereal class.

External territory. Figure 8 displays the Ukraine LULC map at 10 m pixel size for the year 2022. For the Ukraine region, the KPI crop type map is used as a validated product suitable for evaluating our generated map. As shown in Figure 6 and Supplementary Table 11, in the ua22 region, rape and turnip rape stands out with a high PA of 96.8%, while sunflower and "Bare arable land" show good accuracies. However, common wheat and barley are less accurate, and maize shows a notable decline in PA, suggesting substantial misclassifications for these crops in this area.

Confusion between wheat and barley is considerable, as in most of the EU regions. The OA is in the order of 66%.

Usage Notes

Potential improvements. The LULC classification produced in this study is based on monthly median feature values, and it presents some considerable limitations. The primary constraint of the input features used is its temporal resolution. Despite their ability to provide a broad overview of LC dynamics, monthly medians are unable to capture finer temporal nuances. Rapid changes, such as seasonal transitions or short-term disturbances, may not be adequately reflected or may even be overlooked altogether in crops with similar growing behaviour. This problem is more accentuated when mapping areas that span over a great gradient of longitudes and latitudes. Variations in different areas can lead to substantial seasonal variations that affect growing patterns and, therefore, LULC classification accuracy.

Latitude significantly impacts crop development patterns as the angle and intensity of sunlight depend on it. Areas at lower latitudes receive intense sunlight throughout the year compared to regions at higher latitudes. Due to the availability of ample energy for photosynthesis, crops exhibit a relatively stable and higher NDVI, which indicates productive and healthy vegetation. On the other hand, southern regions also exhibit more frequent and variable water stress and drought conditions that limit crop growth and influence phenological patterns.

Latitude is also related to temperature, and locations located at lower latitudes generally experience warmer temperatures, which are generally conducive to the growth of crops. A warmer climate can extend the growing season and enhance the NDVI value of crops since they thrive under these conditions. Conversely, higher latitudes are often characterized by colder winters and shorter growing seasons, resulting in lower NDVI values during periods of dormancy or winter. These factors, along with soil and management conditions, influence the crop growing patterns, and classifying the same crop under variable conditions becomes challenging. In Figure 9, the NDVI trends for the 2022 crop growth of three different crops: Common Wheat, Durum Wheat, and Rapeseed (Rape and Turnip Rape) in two geographically distinct regions, Slovakia (SK) and Spain (ES), which exhibit significant latitudinal and climatic differences, were analyzed. To create this figure, NDVI values were extracted from preprocessed S2 images for each available acquisition. In SK, there were data for 276, 28, and 106 crop samples for Common Wheat, Durum Wheat, and Rapeseed, respectively, while in ES, there were 474, 168, and 49 crop samples for the same crop types. Next, the Whittaker smoother⁹¹ was applied to these NDVI time series data to effectively reduce noise and obtain continuous crop growth trend curves for each individual sample. Subsequently, the median values of these trend curves for each crop type in both countries were computed.

Level-2 Code	100	211	212	213	214	215	216	217	218	219	221	222	223	230	231	232	233	240	250	290	300	500	600	700	800	Total	UA	F1-Score
100	1990	9	8	9	1	3	18	0	1	1	1	1	0	3	4	6	0	22	6	30	515	738	60	10	6	3442	57.8%	58.1%
211	86	1721	87	310	96	65	35	0	108	11	3	1	9	56	5	125	3	24	70	194	45	362	7	5	3	3431	50.2%	63.2%
212	29	27	88	36	0	14	0	0	3	1	0	0	1	2	2	1	0	9	19	41	19	84	4	1	0	381	23.1%	28.9%
213	8	67	19	485	18	36	3	0	21	4	0	1	4	13	3	37	1	8	22	82	7	64	2	0	0	905	53.6%	51.8%
214	10	30	0	19	82	7	5	0	39	2	2	0	3	0	0	1	0	4	5	10	4	51	2	0	0	276	29.7%	31.1%
215	10	37	5	49	11	110	6	0	12	2	0	0	0	1	0	0	3	24	22	11	87	1	0	1	392	28.1%	32.7%	
216	90	30	0	13	5	7	1189	5	2	13	9	7	9	20	23	6	38	47	27	23	63	205	6	0	1	1838	64.7%	75.4%
217	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.0%	0.0%
218	0	5	0	8	18	2	1	0	30	0	0	0	0	1	0	1	0	0	1	4	0	4	0	0	0	75	40.0%	19.7%
219	0	1	0	0	0	1	0	0	0	9	0	0	0	2	0	0	1	1	3	1	1	5	1	0	0	26	34.6%	22.2%
221	6	8	0	0	1	1	3	0	1	1	124	7	1	5	11	5	3	9	2	10	3	13	1	0	0	215	57.7%	63.6%
222	8	4	0	0	0	1	1	0	0	3	6	152	7	6	6	2	0	7	3	3	3	13	1	0	0	226	67.3%	75.2%
223	0	0	0	0	0	0	0	0	0	0	2	0	8	0	1	0	0	0	0	0	0	0	0	0	0	11	72.7%	22.2%
230	12	3	0	1	0	1	0	0	0	0	0	1	1	41	3	2	0	5	3	2	5	9	0	0	0	89	46.1%	31.7%
231	19	2	1	2	0	0	12	0	0	0	13	3	4	5	269	3	2	22	4	16	17	26	3	0	0	423	63.6%	70.1%
232	14	24	1	2	3	1	1	0	2	1	0	1	4	2	5	429	1	7	3	23	2	32	1	0	0	559	76.7%	72.3%
233	3	2	0	0	0	0	8	0	0	1	1	0	0	1	4	0	52	1	3	2	1	5	0	0	84	61.9%	55.9%	
240	4	10	1	1	0	1	1	0	1	4	6	2	6	7	0	7	0	79	5	16	4	17	0	0	172	45.9%	35.6%	
250	17	6	4	5	2	7	3	0	2	0	2	0	1	4	0	0	5	5	187	11	12	129	0	1	0	398	47.0%	35.8%
290	17	1	3	9	1	10	0	0	1	0	2	1	1	0	6	0	0	8	13	168	43	51	8	1	0	344	48.8%	31.9%
300	770	6	2	9	3	2	12	0	1	0	2	0	0	0	2	0	1	4	15	25	9901	1153	49	19	69	12045	82.2%	85.0%
500	272	21	9	10	11	10	15	0	6	2	2	1	1	1	1	3	0	7	232	18	547	5809	8	7	29	7022	82.7%	73.0%
600	33	0	0	0	0	1	1	0	0	0	0	0	1	0	0	0	0	0	0	9	13	11	27	1	1	98	27.6%	18.7%
700	12	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	11	10	19	5	62	30.6%	29.7%
800	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	20	17	0	2	73	114	64.0%	48.3%
Total	3412	2014	228	968	252	281	1314	5	230	55	175	178	61	170	345	628	102	272	647	710	11240	8896	191	66	188	32628	OA = 70.6%	
PA	58.3%	85.5%	38.6%	50.1%	32.5%	39.1%	90.5%	0.0%	13.0%	16.4%	70.9%	85.4%	13.1%	24.1%	78.0%	68.3%	51.0%	29.0%	28.9%	23.7%	88.1%	65.3%	14.1%	28.8%	38.8%		$\kappa = 0.63$	

Table 8. Confusion matrix obtained from the hierarchical classification. [Notation: UA = User's Accuracy, PA = Producer's Accuracy, OA = Overall Accuracy, Kappa coefficient = κ].

Level-2 Code	100	210	213	214	215	216	217	219	221	222	223	230	231	232	233	240	250	290	300	500	600	700	800	Total	UA	F1-Score
100	0	98	37	7	12	56	0	0	36	1	1	0	11	7	3	239	78	129	312	547	0	0	0	1574	0.0%	0.0%
210	0	2426	636	235	222	70	0	0	31	9	0	0	16	85	3	59	360	14	109	364	0	0	0	4639	52.3%	60.7%
213	0	128	268	68	50	1	0	0	1	2	1	0	0	29	1	7	61	0	15	70	0	0	0	702	38.2%	27.3%
214	0	69	37	62	12	12	0	0	5	0	0	0	0	7	0	7	14	0	4	40	0	0	0	269	23.0%	17.6%
215	0	102	66	20	92	4	0	0	7	0	0	0	2	0	0	8	37	0	15	65	0	0	0	418	22.0%	19.0%
216	0	95	36	9	27	2339	38	0	58	5	8	5	51	3	57	118	158	18	57	176	0	0	0	3258	71.8%	78.1%
217	0	0	0	0	0	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	100.0%	20.4%	
219	0	2	0	0	5	1	0	0	2	0	0	0	0	0	0	1	13	1	0	5	0	0	0	30	0.0%	0.0%
221	0	3	3	1	2	10	0	0	134	1	11	0	12	3	4	30	13	0	6	10	0	0	0	243	55.1%	39.5%
222	0	0	1	0	1	3	0	0	2	88	12	1	3	4	0	26	6	0	0	7	0	0	0	154	57.1%	65.2%
223	0	0	0	0	0	0	0	0	0	1	2	0	0	0	0	4	1	0	0	1	0	0	0	9	22.2%	7.3%
230	0	5	4	0	0	0	0	0	1	0	0	27	5	0	0	28	3	0	0	1	0	0	0	74	36.5%	47.0%
231	0	15	4	1	4	37	0	0	21	3	2	5	398	0	17	35	27	2	6	9	0	0	0	586	67.9%	70.9%
232	0	14	0	0	0	10	0	0	0	2	0	0	6	279	0	5	34	0	7	8	0	0	0	365	76.4%	69.0%
233	0	14	1	0	0	20	0	0	3	0	0	0	10	0	84	15	1	0	1	7	0	0	0	156	53.8%	50.8%
240	0	5	5	0	6	3	0	0	9	3	5	2	1	17	1	50	51	1	5	6	0	0	0	170	29.4%	9.7%
250	0	52	26	1	13	14	1	0	7	0	0	0	1	1	0	6	255	1	10	195	0	0	0	583	43.7%	23.0%
290	0	97	13	1	7	4	0	0	5	0	0	1	5	0	0	43	18	3	19	45	0	0	0	261	1.1%	0.9%
300	0	82	25	5	22	68	0	0	32	1	2	0	9	5	0	77	82	133	4978	2375	0	0	0	7896	63.0%	70.1%
500	0	147	101	24	76	83	0	0	81	0	2	0	6	4	4	101	421	119	733	9962	0	0	0	11864	84.0%	77.3%
600	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	12	8	0	0	0	21	0.0%	0.0%
700	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	4	0	0	0	11	0.0%	0.0%
800	0	2	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	12	4	0	0	0	20	0.0%	0.0%
Total	0	3356	1264	434	551	2735	44	0	435	116	46	41	537	444	175	859	1634	421	6307	13909	0	0	0	33308	OA = 64.4%	
PA	0.0%	72.3%	21.2%	14.3%	16.7%	85.5%	11.4%	0.0%	30.8%	75.9%	4.3%	65.9%	74.1%	62.8%	48.0%	5.8%	15.6%	0.7%	78.9%	71.6%	0.0%	72.3%	21.2%		$\kappa = 0.54$	

Table 9. Confusion matrix derived from the final map for the EMBAL dataset. UA = User's Accuracy, PA = Producer's Accuracy, OA = Overall Accuracy, Kappa coefficient = κ .

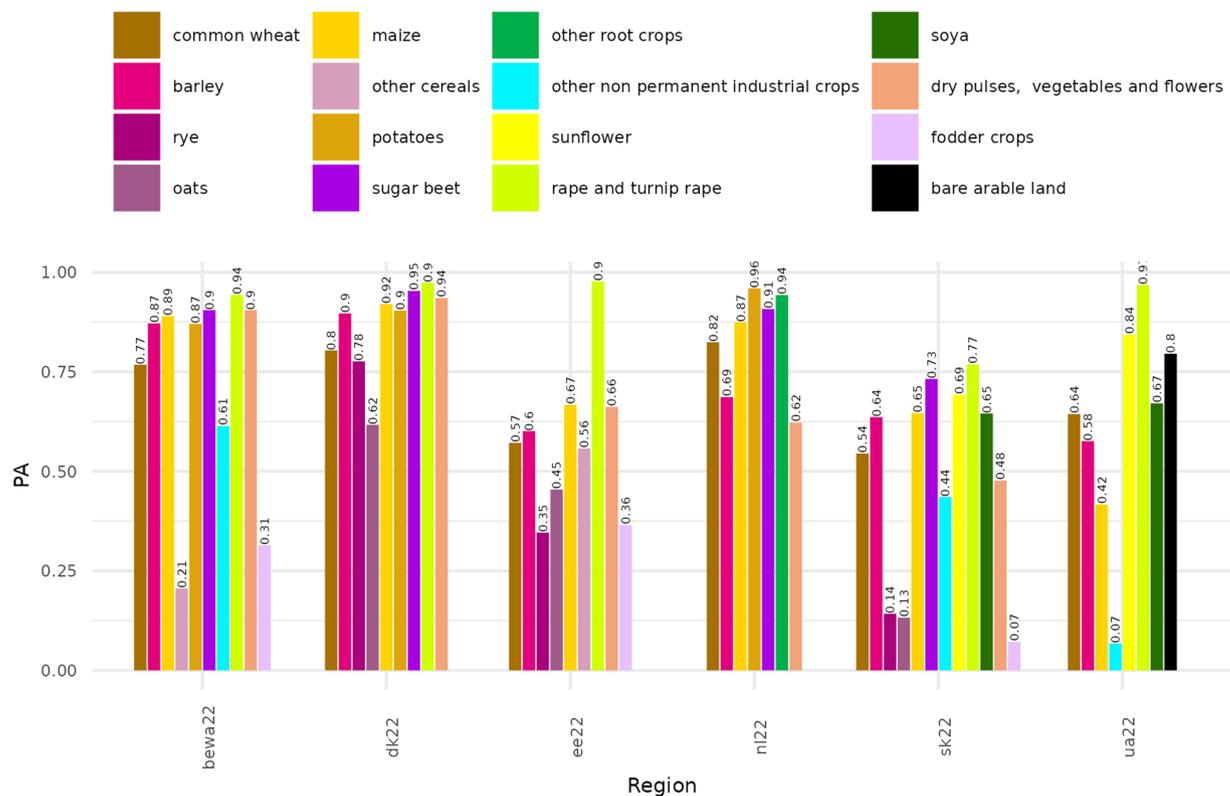


Fig. 6 Producer's accuracies comparing GSA farmer declaration data for EU LULC crop codes and specific country/region.

The median values served as representative trend curves, capturing the overall behaviour of the three crop types used for this example (Common Wheat, Durum Wheat, and Rapeseed).

As depicted in Figure 9, ES experiences a longer growing season, primarily owing to its warmer climate, which results in crops reaching their peak growth earlier than in SK. Despite this difference, the growth behaviour of the three mentioned crops exhibits striking similarities in both regions. Notably, Durum Wheat starts the season and attains its peak growth earlier, while Common Wheat and Rapeseed follow a slightly delayed growth trajectory.

Various other factors, including proximity to large water bodies, landforms, and large-scale climate, can impact the local climate and vegetation patterns. In coastal regions, for instance, temperatures may be milder, and precipitation patterns may differ from those in inland regions at the same latitude, resulting in differences in crop NDVI based on these factors.

According to Figure 10, the growth trends of Common Wheat and Barley crops were analyzed in northern France (FR) and southern Poland (PL), two regions with similar latitudes but varying longitudes. For Common Wheat, 1258 samples in FR and 330 samples in PL were evaluated, while Barley had 730 samples in FR and 255 samples in PL, following the same procedure outlined earlier. Crops in northern FR, situated in the Atlantic climate region, exhibit earlier peak growth compared to southern PL, which falls within the continental climate region. Notably, in both regions, Barley consistently experiences an earlier start and reaches its peak growth before Common Wheat.

Therefore, LULC classification at the continent scale with different climates, based on a single classifier and based on monthly median input features, presents significant challenges. It is important to recognize that the growth behaviour of the same crop types can exhibit shifts influenced by regional variations as well as crops with similar growth patterns. It is often difficult to capture these subtle differences using only monthly medians. The use of more advanced time series analysis or deep learning techniques by integrating all available acquisitions as well as utilizing more suitable auxiliary data, is recommended to overcome these limitations.

Potential applications. Similarly to the EU Crop Map 2018, the EU Crop Map 2022 provides a refined lens for European agricultural and environmental analyses. Various uses can be envisioned. These maps make the interplay between crop diversity and agricultural system resilience more tangible, empowering stakeholders to strategize against potential disruptions across various scales. The map can be used as an input to account for the environmental implications of pesticide use. Highlighting regions close to sensitive and urban areas underscores the need to balance agricultural output with ecological preservation. The map also provides an essential base-layer needed to evaluate the potential effects of agricultural intensification on biodiversity. In geopolitically charged situations, such as the ongoing war between Russia and Ukraine, this resource proves invaluable in assessing agricultural impacts, guiding both national and European strategic responses.

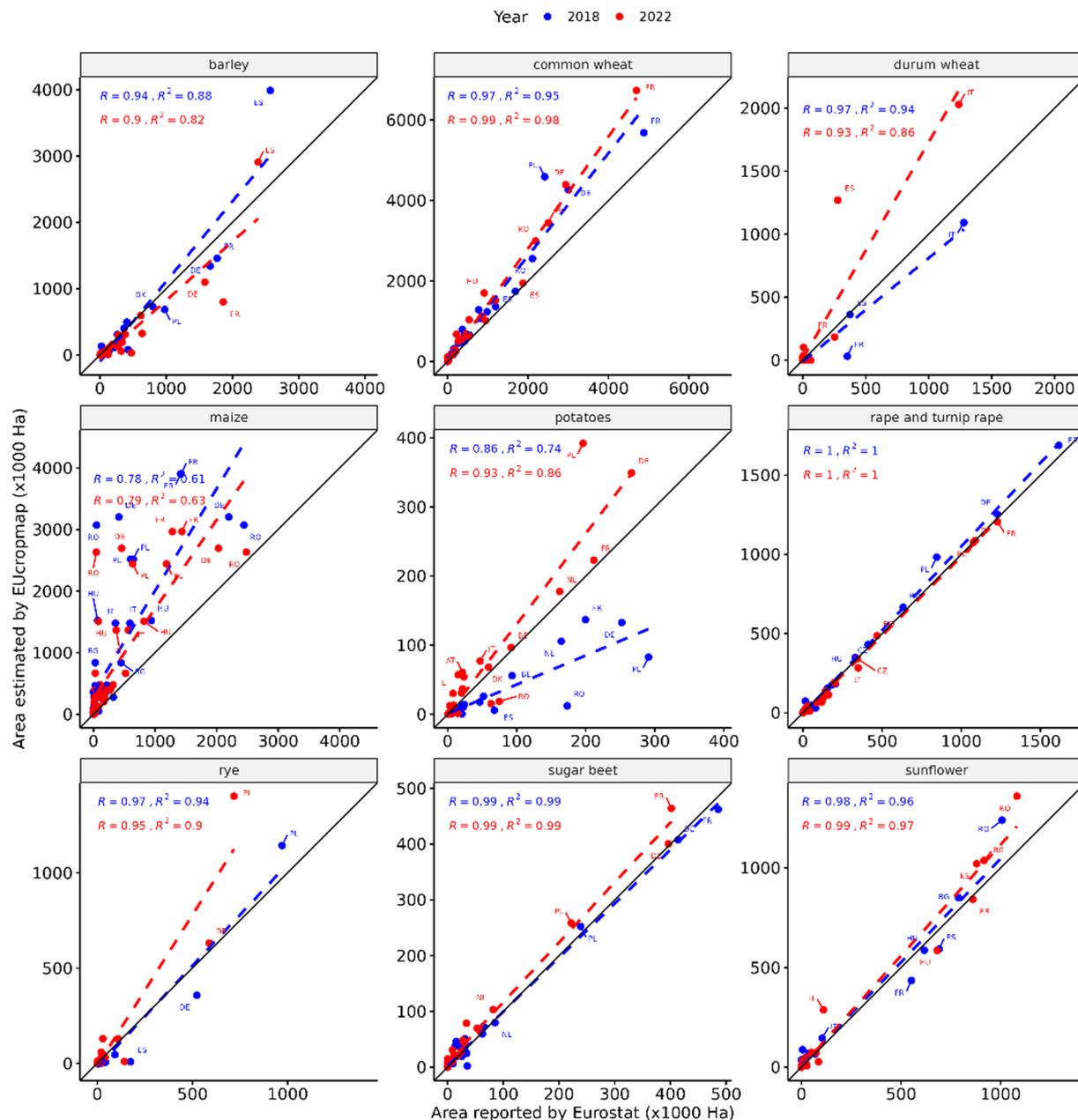


Fig. 7 The areas reported by Eurostat at country level are compared with the area retrieved from the EU 2018 and EU 2022 crop maps. R is the Pearson correlation coefficient, and R^2 is the coefficient of determination.

In summary, the EU Crop Map 2022 has the potential to enhance agricultural research and to play a pivotal role in shaping responsive and sustainable policies across the continent.

The future of EU crop mapping: A forward perspective. In the context of EU-wide crop mapping, it is useful to highlight the new Copernicus High Resolution Layer (HRL) on Crop Types, which is part of the HRL Vegetation Land Cover Component (VLCC) and it will be released in Q3 2024 by the European Environment Agency (EEA). Using Sentinel observations, the HRL Crop Types will cover the EEA 38 member nations encompassing annual maps starting from 2017 with 10-meter resolution. Technical specifics are elaborated in EEA/DIS/R0/21/013 (Tender available here: <https://etendering.ted.europa.eu/cft/cft-display.html?cftId=8630>).

These new European high-resolution and satellite-sourced maps provide an unprecedented view into the continent's agrarian landscape. Through their European-wide coverage and detail, they can underpin higher precision agronomic, agri-environmental, and climate assessments. As such, they have the potential to reshape policy-making approaches, elevate agricultural standards, and fortify the EU's commitment to sustainable food production.

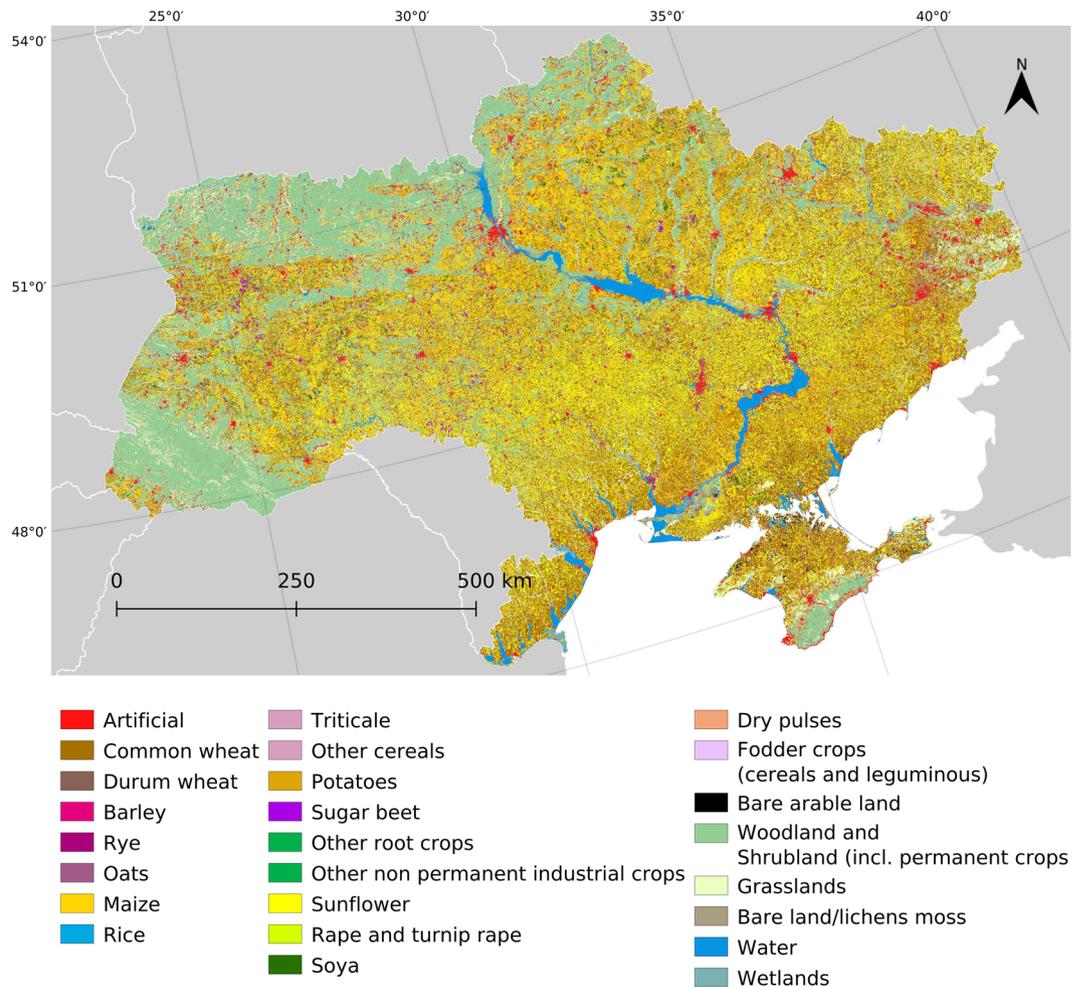


Fig. 8 Ukraine’s Land Use and Land Cover map at 10 m pixel size for the year 2022.

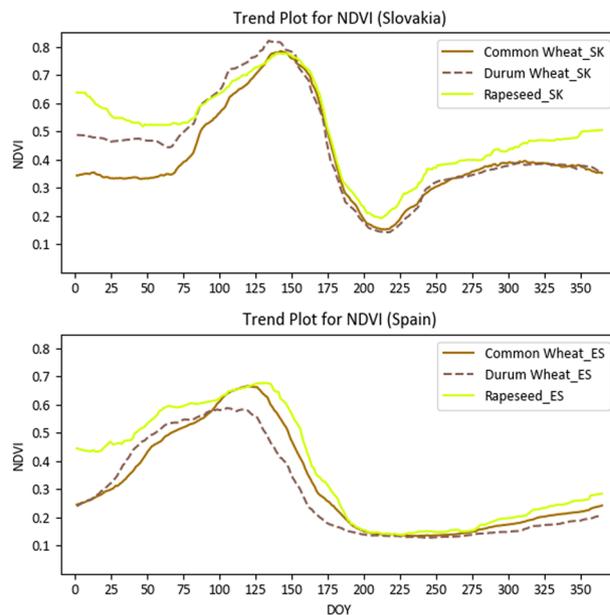


Fig. 9 NDVI trends for the 2022 crop growth of three different crops: Common Wheat, Durum Wheat, and Rapeseed (Rape and Turnip Rape) in Slovakia and Spain.

Code availability

Detailed information regarding the generation of a LULC classification map in GEE can be found in the following repository⁹² (<https://zenodo.org/doi/10.5281/zenodo.10220964>), along with guidelines contained in a README file.

Received: 10 January 2024; Accepted: 13 September 2024;

Published online: 27 September 2024

References

- Lu, D. & Weng, Q. A survey of image classification methods and techniques for improving classification performance. *International Journal of Remote Sensing* **28**, 823–870 (2007).
- Turner, W. *et al.* Remote sensing for biodiversity science and conservation. *Trends in Ecology & Evolution* **18**, 306–314 (2003).
- Thenkabail, P. S. & Lyon, J. G. *Hyperspectral remote sensing of vegetation*. (CRC Press, 2012).
- Directorate-General for Agriculture and Rural Development, Unit G.1. Monitoring EU agri-food trade. *Developments in May 2023* https://agriculture.ec.europa.eu/system/files/2024-01/monitoring-agri-food-trade_may2023_en_1.pdf (2023).
- Shukla, P. R. *et al.* IPCC, 2019: Climate Change and Land: an IPCC special report on climate change, desertification, land degradation, sustainable land management, food security, and greenhouse gas fluxes in terrestrial ecosystems (2019).
- Mouillot, F. *et al.* Ten years of global burned area products from spaceborne remote sensing—A review: Analysis of user needs and recommendations for future developments. *International Journal of Applied Earth Observation and Geoinformation* **26**, 64–79 (2014).
- Munawar, H. S., Hammad, A. W. A. & Waller, S. T. Remote Sensing Methods for Flood Prediction: A Review. *Sensors* **22**, 960 (2022).
- PERPIÑA CASTILLO Carolina, KAVALOV Boyan, DIOGO Vasco, JACOBS Christiaan, BATISTA E SILVA Filipe, BARANZELLI Claudia, LAVALLE Carlo. Trends in the EU Agricultural Land Within 2015–2030. <https://joint-research-centre.ec.europa.eu/system/files/2018-12/jrc113717.pdf> (2018).
- Common Agricultural Policy For 2023–2027. 28 CAP Strategic Plans at a glance https://agriculture.ec.europa.eu/document/download/a435881e-d02b-4b98-b718-104b5a30d1cf_en?filename=csp-at-a-glance-eu-countries_en.pdf (2022).
- Sishodia, R. P., Ray, R. L. & Singh, S. K. Applications of Remote Sensing in Precision Agriculture: A Review. *Remote Sensing* **12**, 3136 (2020).
- Ali, A. M. *et al.* Crop Yield Prediction Using Multi Sensors Remote Sensing (Review Article). *The Egyptian Journal of Remote Sensing and Space Science* **25**, 711–716 (2022).
- Berra, E. F., Gaulton, R. & Barr, S. Commercial Off-the-Shelf Digital Cameras on Unmanned Aerial Vehicles for Multitemporal Monitoring of Vegetation Reflectance and NDVI. *IEEE Trans. Geosci. Remote Sensing*. **55**, 4878–4886 (2017).
- Matese, A. *et al.* Intercomparison of UAV, Aircraft and Satellite Remote Sensing Platforms for Precision Viticulture. *Remote Sensing* **7**, 2971–2990 (2015).
- Binte Mostafiz, R., Noguchi, R. & Ahamed, T. Agricultural Land Suitability Assessment Using Satellite Remote Sensing-Derived Soil-Vegetation Indices. *Land* **10**, 223 (2021).
- Kennedy, C. M. *et al.* Optimizing land use decision-making to sustain Brazilian agricultural profits, biodiversity and ecosystem services. *Biological Conservation* **204**, 221–230 (2016).
- Pimentel, D. & Burgess, M. Soil Erosion Threatens Food Production. *Agriculture* **3**, 443–463 (2013).
- Hoefsloot P *et al.* Combining Crop Models and Remote Sensing for Yield Prediction: Concepts, Applications and Challenges for Heterogeneous Smallholder Environments. 1831–9424 (2012).
- Dorigo, W. A. *et al.* A review on reflective remote sensing and data assimilation techniques for enhanced agroecosystem modeling. *International Journal of Applied Earth Observation and Geoinformation* **9**, 165–193 (2007).
- MohanRajan, S. N., Loganathan, A. & Manoharan, P. Survey on Land Use/Land Cover (LU/LC) change analysis in remote sensing and GIS environment: Techniques and Challenges. *Environ Sci Pollut Res* **27**, 29900–29926 (2020).
- Potapov, P. *et al.* Global maps of cropland extent and change show accelerated cropland expansion in the twenty-first century. *Nat Food* **3**, 19–28 (2022).
- Amol D. Vibhute & Dr. Bharti W. Gawali. Analysis and Modeling of Agricultural Land use using Remote Sensing and Geographic Information System: a Review (2013).
- Gomes, V., Queiroz, G. & Ferreira, K. An Overview of Platforms for Big Earth Observation Data Management and Analysis. *Remote Sensing* **12**, 1253 (2020).
- Mirmazloumi, S. M. *et al.* ELULC-10, a 10 m European Land Use and Land Cover Map Using Sentinel and Landsat Data in Google Earth Engine. *Remote Sensing* **14**, 3041 (2022).
- d’Andrimont, R. *et al.* 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**, 112708 (2021).
- Ghassemi, B. *et al.* 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 (2022).
- Immitzer, M., Vuolo, F. & Atzberger, C. First Experience with Sentinel-2 Data for Crop and Tree Species Classifications in Central Europe. *Remote Sensing* **8**, 166 (2016).
- Malinowski, R. *et al.* Automated Production of a Land Cover/Use Map of Europe Based on Sentinel-2 Imagery. *Remote Sensing* **12**, 3523 (2020).
- Tsendbazar, N. *et al.* Towards operational validation of annual global land cover maps. *Remote Sensing of Environment* **266**, 112686 (2021).
- Venter, Z. S. & Sydenham, M. A. K. Continental-Scale Land Cover Mapping at 10 m Resolution Over Europe (ELC10). *Remote Sensing* **13**, 2301 (2021).
- K Karra, *et al.* Global land use / land cover with Sentinel 2 and deep learning. 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS (2021).
- Fuller, R. M., Groom, G. B. & Jones, A. R. Land cover map of Great Britain. An automated classification of Landsat Thematic Mapper data. *Photogrammetric engineering and remote sensing*. **60** (1994).
- Gong, P. *et al.* Finer resolution observation and monitoring of global land cover: first mapping results with Landsat TM and ETM+ data. *International Journal of Remote Sensing* **34**, 2607–2654 (2013).
- Hansen, M. C. & Loveland, T. R. A review of large area monitoring of land cover change using Landsat data. *Remote Sensing of Environment* **122**, 66–74 (2012).
- Potapov, P. *et al.* Landsat Analysis Ready Data for Global Land Cover and Land Cover Change Mapping. *Remote Sensing* **12**, 426 (2020).
- Friedl, M. *et al.* Global land cover mapping from MODIS: algorithms and early results. *Remote Sensing of Environment* **83**, 287–302 (2002).
- Sulla-Menashe, D., Gray, J. M., Abercrombie, S. P. & Friedl, M. A. Hierarchical mapping of annual global land cover 2001 to present: The MODIS Collection 6 Land Cover product. *Remote Sensing of Environment* **222**, 183–194 (2019).

37. Vuolo, F. & Atzberger, C. Exploiting the Classification Performance of Support Vector Machines with Multi-Temporal Moderate-Resolution Imaging Spectroradiometer (MODIS) Data in Areas of Agreement and Disagreement of Existing Land Cover Products. *Remote Sensing* **4**, 3143–3167 (2012).
38. Ali, S. *et al.* A time series of land cover maps of South Asia from 2001 to 2015 generated using AVHRR GIMMS NDVI3g data. *Environ Sci Pollut Res* **27**, 20309–20320 (2020).
39. Andres, L., Salas, W. A. & Skole, D. Fourier analysis of multi-temporal AVHRR data applied to a land cover classification. *International Journal of Remote Sensing* **15**, 1115–1121 (1994).
40. Thenkabail, P. S., Gangadhara Rao, P., Biggs, T. W., Krishna, M. & Turrall, H. *Spectral matching techniques to determine historical land use/Land cover (LULC) and irrigated areas using time-series 0.1 degree AVHRR Pathfinder Datasets.* (2007).
41. Gorelick, N. *et al.* Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment* **202**, 18–27 (2017).
42. Miettinen, J., Shi, C. & Liew, S. C. 2015 Land cover map of Southeast Asia at 250 m spatial resolution. *Remote Sensing Letters* **7**, 701–710 (2016).
43. Xiong, J. *et al.* Automated cropland mapping of continental Africa using Google Earth Engine cloud computing. *ISPRS Journal of Photogrammetry and Remote Sensing* **126**, 225–244 (2017).
44. Ghorbanian, A. *et al.* Improved land cover map of Iran using Sentinel imagery within Google Earth Engine and a novel automatic workflow for land cover classification using migrated training samples. *ISPRS Journal of Photogrammetry and Remote Sensing* **167**, 276–288 (2020).
45. Shafizadeh-Moghadam, H., Khazaei, M., Alavipanah, S. K. & Weng, Q. Google Earth Engine for large-scale land use and land cover mapping: an object-based classification approach using spectral, textural and topographical factors. *GIScience & Remote Sensing* **58**, 914–928 (2021).
46. d'Andrimont, R. *et al.* Harmonised LUCAS *in-situ* land cover and use database for field surveys from 2006 to 2018 in the. *European Union. Sci Data*. **7**, 352 (2020).
47. Mack, B., Leinenkugel, P., Kuenzer, C. & Dech, S. A semi-automated approach for the generation of a new land use and land cover product for Germany based on Landsat time-series and Lucas *in-situ* data. *Remote Sensing Letters* **8**, 244–253 (2017).
48. Close, O., Benjamin, B., Petit, S., Fripiat, X. & Hallot, E. Use of Sentinel-2 and LUCAS Database for the Inventory of Land Use, Land Use Change, and Forestry in Wallonia, Belgium. *Land* **7**, 154 (2018).
49. Plugmacher, D., Rabe, A., Peters, M. & Hostert, P. Mapping pan-European land cover using Landsat spectral-temporal metrics and the European LUCAS survey. *Remote Sensing of Environment* **221**, 583–595 (2019).
50. Weigand, M., Staab, J., Wurm, M. & Taubenböck, H. Spatial and semantic effects of LUCAS samples on fully automated land use/land cover classification in high-resolution Sentinel-2 data. *International Journal of Applied Earth Observation and Geoinformation* **88**, 102065 (2020).
51. Ghassemi, B., Immitzer, M., Atzberger, C. & Vuolo, F. Evaluation of Accuracy Enhancement in European-Wide Crop Type Mapping by Combining Optical and Microwave Time Series. *Land* **11**, 1397 (2022).
52. Witjes, 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**, e13573 (2022).
53. European Commission, Joint Research Centre (JRC). LUCAS Copernicus 2022. *European Commission, Joint Research Centre (JRC)* <https://data.jrc.ec.europa.eu/dataset/e3fe3cd0-44db-470e-8769-172a8b9e8874> (2023).
54. d'Andrimont, R. *et al.* Advancements in LUCAS Copernicus 2022: Enhancing Earth Observation with Comprehensive *In-Situ* Data on EU Land Cover and Use. *Earth Syst. Sci. Data Discuss* [preprint], <https://doi.org/10.5194/essd-2023-494> (2024).
55. Drusch, M. *et al.* Sentinel-2: ESA's Optical High-Resolution Mission for GMES Operational Services. *Remote Sensing of Environment* **120**, 25–36 (2012).
56. Boegh, E. *et al.* Airborne multispectral data for quantifying leaf area index, nitrogen concentration, and photosynthetic efficiency in agriculture. *Remote Sensing of Environment* **81**, 179–193 (2002).
57. Jiang, Z., Huete, A., Didan, K. & Miura, T. Development of a two-band enhanced vegetation index without a blue band. *Remote Sensing of Environment* **112**, 3833–3845 (2008).
58. Gitelson, A. A., Kaufman, Y. J. & Merzlyak, M. N. Use of a green channel in remote sensing of global vegetation from EOS-MODIS. *Remote Sensing of Environment* **58**, 289–298 (1996).
59. Pasqualotto, N., Delegido, J., van Wittenberghe, S., Rinaldi, M. & Moreno, J. Multi-Crop Green LAI Estimation with a New Simple Sentinel-2 LAI Index (SeLI). *Sensors* (Basel). **19** (2019).
60. Wulf, H. & Stuhler, S. Sentinel-2: Land Cover, Preliminary User Feedback on Sentinel-2A Data. In Proceedings of the Sentinel-2A Expert Users Technical Meeting, Frascati, Italy, 29–30 September, 29–30 (2015).
61. Kriegler, F. J., Malila, W. A., Nalepka, R. F. & Richardson, W. Preprocessing Transformations and Their Effects on Multispectral Recognition. *Remote Sensing of Environment*, **VI**, 97 (1969).
62. Rikimaru, A., Roy, P. S. & Miyatake, S. Tropical forest cover density mapping. *Tropical ecology* **43**, 39–47 (2002).
63. Qi, J., Chehbouni, A., Huete, A. R., Kerr, Y. H. & Sorooshian, S. A modified soil adjusted vegetation index. *Remote Sensing of Environment* **48**, 119–126 (1994).
64. van Deventer, A. P., Ward, A. D., Gowda, P. H. & Lyon, J. G. Using thematic mapper data to identify contrasting soil plains and tillage practices. *Photogrammetric engineering and remote sensing* **63**, 87–93 (1997).
65. Huete, A. A soil-adjusted vegetation index (SAVI). *Remote Sensing of Environment* **25**, 295–309 (1988).
66. Bouhennache, R., Bouden, T., Taleb-Ahmed, A. & Cheddad, A. A new spectral index for the extraction of built-up land features from Landsat 8 satellite imagery. *Geocarto International* **34**, 1531–1551 (2019).
67. Zha, Y., Gao, J. & Ni, S. Use of normalized difference built-up index in automatically mapping urban areas from TM imagery. *International Journal of Remote Sensing* **24**, 583–594 (2003).
68. Xu, H. Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery. *International Journal of Remote Sensing* **27**, 3025–3033 (2006).
69. McFEETERS, S. K. The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. *International Journal of Remote Sensing* **17**, 1425–1432 (1996).
70. Jacques, D. C., Kergoat, L., Hiernaux, P., Mougou, E. & Defourny, P. Monitoring dry vegetation masses in semi-arid areas with MODIS SWIR bands. *Remote Sensing of Environment* **153**, 40–49 (2014).
71. Blackburn, G. A. Quantifying Chlorophylls and Carotenoids at Leaf and Canopy Scales. *Remote Sensing of Environment* **66**, 273–285 (1998).
72. Izquierdo-Verdiguier, E. & Zurita-Milla, R. An evaluation of Guided Regularized Random Forest for classification and regression tasks in remote sensing. *International Journal of Applied Earth Observation and Geoinformation* **88**, 102051 (2020).
73. Torres, R. *et al.* GMES Sentinel-1 mission. *Remote Sensing of Environment* **120**, 9–24 (2012).
74. Kupidura, P. Comparison of Filters Dedicated to Speckle Suppression in Sar Images. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.* **XLI-B7**, 269–276 (2016).
75. Arai, M., van Zyl, J. J. & Kim, Y. A General Characterization for Polarimetric Scattering From Vegetation Canopies. *IEEE Trans. Geosci. Remote Sensing* **48**, 3349–3357 (2010).
76. Mitchard, E. T. A. *et al.* Mapping tropical forest biomass with radar and spaceborne LiDAR in Lopé National Park, Gabon: overcoming problems of high biomass and persistent cloud. *Biogeosciences* **9**, 179–191 (2012).

77. dos Santos, E. P., Da Silva, D. D. & do Amaral, C. H. Vegetation cover monitoring in tropical regions using SAR-C dual-polarization index: seasonal and spatial influences. *International Journal of Remote Sensing* **42**, 7581–7609 (2021).
78. Hulley, G. & Hook, S. MODIS/Terra Land Surface Temperature/3-Band Emissivity Monthly L3 Global 0.05Deg CMG V061. NASA EOSDIS Land Processes Distributed Active Archive Center (2021).
79. NASA/METI/AIST/Japan Spacesystems and U.S./Japan ASTER Science Team. ASTER Global Digital Elevation Model V003.
80. Tadono, T. *et al.* Precise Global DEM Generation by ALOS PRISM. *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci.* **II-4**, 71–76 (2014).
81. European Environment Agency. CORINE Land Cover 2018 (raster 100 m), Europe, 6-yearly - version 2020_20u1, May 2020. European Environment Agency <https://land.copernicus.eu/en/products/corine-land-cover/clc2018> (2019).
82. Breiman, L. Random Forests. *Machine Learning* **45**, 5–32 (2001).
83. Criminisi, A. Decision Forests: A Unified Framework for Classification, Regression, Density Estimation, Manifold Learning and Semi-Supervised Learning. *FNT in Computer Graphics and Vision* **7**, 81–227 (2011).
84. Pedregosa, F. *et al.* Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* **12**, 2825–2830 (2011).
85. European Environment Agency. Biogeographical regions in Europe. European Environment Agency <https://www.eea.europa.eu/data-and-maps/figures/biogeographical-regions-in-europe-2>.
86. Lindemann, D., Oppermann, R., Sutcliffe, L. M., Moser, D. & Haub, C. European Monitoring of Biodiversity in Agricultural Landscapes (EMBAL). *Sensing-New Insights into Grassland Science and Practice*, 150 (2021).
87. European Commission, Joint Research Centre (JRC). EUCROPMAP 2022. European Commission, Joint Research Centre (JRC) <https://doi.org/10.2905/555e5d1d-1aae-4320-a716-2e6d18aa1e7c> (2022).
88. JRC-BDAP. VOIS library. Code development platform for open source projects from the European Union institutions <https://code.europa.eu/jrc-bdap/vois> (2023).
89. Machefer, M. *et al.* Earth Observation based multi-scale analysis of crop diversity in the European Union: First insights for agro-environmental policies. *Agriculture, Ecosystems & Environment*, **374**, 109143, <https://doi.org/10.1016/j.agee.2024.109143> (2024).
90. Olofsson, P. *et al.* Good practices for estimating area and assessing accuracy of land change. *Remote Sensing of Environment* **148**, 42–57 (2014).
91. Eilers, P. H. C. A perfect smoother. *Anal Chem* **75**, 3631–3636 (2003).
92. Ghassemi, B., Izquierdo-Verdiguier, E. & Vuolo, F. Google Earth Engine Scripts for Land Use Land Cover (LULC) Classification of the European Union (EU27). *Zenodo* <https://zenodo.org/doi/10.5281/zenodo.10220964> (2023).

Acknowledgements

This project was co-funded by the European Union's Expert Contract No CT-EX2023D752321-101 under the EU Crop Map 2022 – a 10-m crop type map of EU based on the Earth Observation satellite and LUCAS Copernicus 2022. This research was also co-funded by the European Union's Horizon Europe Research and Innovation Programme under Grant Agreement No 101060423 LAMASUS project.

Author contributions

Babak Ghassemi: Methodology, Software, Formal analysis, Writing - Original Draft, Writing - Review & Editing. Emma Izquierdo-Verdiguier: Methodology, Software, Formal analysis, Writing - Original Draft, Writing - Review & Editing, Supervision. Astrid Verhegghen: Formal analysis, Writing - Review & Editing. Momchil Yordanov: Formal analysis, Writing - Review & Editing. Guido Lemoine: Formal analysis, Writing - Review & Editing. Alvaro Moreno Martínez: Methodology, Software. Davide De Marchi: Develop the dashboard viewer, Writing - Review & Editing. Marijn van der Velde: Formal analysis, Writing - Review & Editing. Francesco Vuolo: Conceptualization, Formal analysis, Writing - Original Draft, Writing - Review & Editing, Supervision, Funding acquisition. Raphael d'Andrimont: Conceptualization, Formal analysis, Writing - Original Draft, Writing - Review & Editing.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41597-024-03884-y>.

Correspondence and requests for materials should be addressed to F.V. or R.d.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2024