NextGenCarbon

Monitoring Earth's Carbon Balance

Deliverable 5.1

Customizing land-use data for surface and bookkeeping modelling

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting authority can be held responsible for them.

Funded by the European Union

Funded by
the European Union

Project	NextGenCarbon - Next Generation Modelling of Terrestrial Carbon Cycle by assimilation of in-situ campaigns and Earth Observations
Project Duration	January 2025 – December 2029 (60 months)
Deliverable	Customizing land-use data for surface and bookkeeping modelling
Project website	www.nextgencarbon-project.eu
Work Package	5
Work Package Lead	Julia Pongratz, Guillaume Marie
Dissemination level:	PU
Contributing	n/a
Due date	31.03.2026
Responsible authors	Karina Winkler, Richard Fuchs
Reviewers	Clemens Schwingshackl
Submitted on	19.03.2026

PU = Public
PP = Restricted to other programme participants (including the Commission Services)
RE = Restricted to a group specified by the consortium (including the Commission Services)
CO = Confidential, only for members of the consortium (including the Commission Services)

History of changes

Version	**Date**	**Beneficiary**	**Author**	**Description**

Contents

Figures

Tables

Funded by the European Union

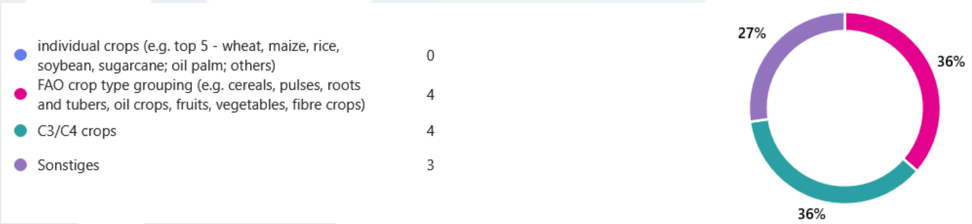
1 Identifying data requirements and prioritisation

1.1 Survey on data requirements

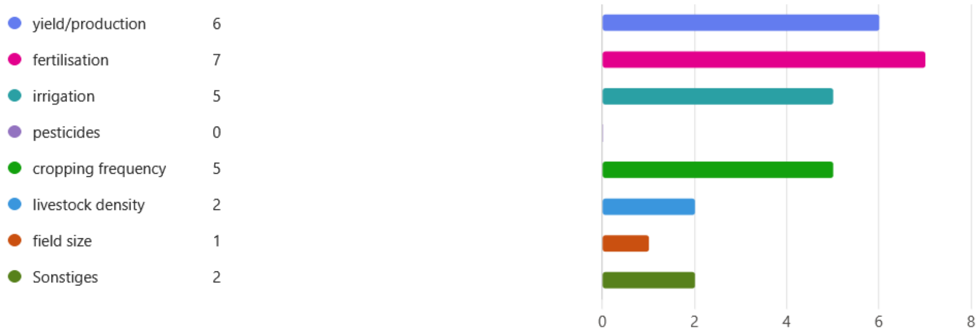
During the **NGC Annual Meeting in Paris** a structured questionnaire was circulated among project partners. The aim was to capture the most urgent data needs for land-use and land-cover modelling and to rank them according to relevance for the consortium's scientific programme. The results of the questionnaire were as follows:

1. Please rank the following information regarding agricultural land use/cover classes according to your priorities.

2. Which crop groupings are most relevant/useful to you?



3. Which of the following management factors are most important to you?



4. Any comments on management factors (e.g. additional or more specific variables and units needed)?
   "soil management, e.g., renewal of grasslands, direct seeding/low/no till"

5. What data type should the new agricultural classes (1x1km grid size) have?
   "Raster file for whole EU", " Netcdf, wgs84, global", " netcdf, wgs84", "netcdf"

6. Any additional comments?
   a. "Really looking forward to the crop type split and information on management, in particular irrigation. With these aspects included I believe we can use it as input to our modeling pipeline with MAgPIE (https://magpiemodel.github.io/)."
   b. "It would be nice to have different forest management practises in Hilda cells and also information if a grid cell is on peatland / drained peatland / mineral soil"
   c. "Don't pay too much attention to my answers, because this is not my field of expertise :-)"
   d. "at the moment JULES uses the fraction of cropland and grazing land, other can use transitions. All the dgvms would want forcing datasets to go back before the satellite era and before the FAO era in 1960. Perhaps one can do as Kees does and scale by population density?"

Funded by the European Union

e. "I guess the file type depends on what is most efficient to store data at this high spatial resolution. One option you could also consider is the Zarr format."

1.2 Meetings with LMU

On 15th October 2025 NGC project partners from Ludwig-Maximilians-Universität Munich - LMU (present: Aparnna Ravi Panangattuparambil, and Clemens Schwingshackl) and Karlsruhe Institute of Technology – KIT (present Karina Winkler and Richard Fuchs) met in Garmisch-Partenkirchen at the KIT Campus Alpine and exchanged information on modelling features of their two models HILDA+ v3 and BLUE. In work package 5, HILDA+ v3 is meant to provide land cover/use and management information for various model types including Earth System Models and BLUE.

To facilitate a smooth data integration into the BLUE model, a list of priorities was discussed for HILDA+ v3:

- Provision of wood harvest data, ideally as area
- Sub-grid fractions for land-use transitions at 0.01° (1 km) resolution
- Distinction between primary and secondary vegetation

For potential usage in Global Carbon Budgets:

- Expansion back in time (realistic)
- Starting in 1800 (or ideally 1700) necessary for estimating cumulative ELUC
- Yearly updates, e.g., for GCB2025 we use LUH2 data up to and including 2024

Funded by the European Union

For the purpose of deliverable 5.1. three priorities where selected:

1. Sub-grid fractions for land-use transitions at 0.01° (1 km) resolution
2. Distinction between primary and secondary vegetation
3. Yearly updates

The remaining priorities will be handled later in the project, if possible or tackled in a partner project.

Funded by
the European Union

2 HILDA+ global land use change v3

This deliverable builds upon the recently published Milestone 6 // 3.11 in WP3: Land use/cover modelling (incl. HILDA+ land management and forest management) (published on 18th October 2025 and shared with the NGC consortium). The data can be downloaded in different formats here:

https://thredds.imk-ifu.kit.edu/thredds/catalog/projects/hildaplus/catalog.html

(either download files individually or entire folders as ZIP archives by accessing them, then clicking on HTTPServer)

A short documentation, which lists the data specifications, content and underlying input data, can be found on the sharepoint: HILDAplus_LandUseChange_version3_NGC_milestone6.

2.1 What is HILDA+ v3?

1. HILDA+ is a synergy product of harmonising high- to medium-resolution remote sensing-based datasets with long-term FAO statistics
2. annual land use/cover maps and land use transitions (gross change) at global scale, with 1x1 km resolution from 1960 to 2023 and also from 1900 to 1960 (based on extrapolated trends)
3. contains discrete classes: 6 generic land use/cover categories: urban, cropland, pasture/rangeland, forest, unmanaged grass/shrubland, sparse/no vegetation

with additional sub-classes

1. 3 cropland categories: annual crops, tree crops and agroforestry
2. 6 forest types: evergreen needle leaf forest, evergreen broad leaf forest, deciduous needle leaf forest, deciduous broad leaf forest, mixed forest, unknown/other forest
4. additional maps of the managed forest fraction (forest fraction with signs of management)

2.2 Updates compared to HILDA+ v2[1, 2] - What is new?

1. extended time period from 2020 to 2023 (last available FAO year)
2. integration of recent high-resolution remote sensing-based data, e.g. Esri global land cover 10m 2017-2024, esa WorldCover 2020-2021, MapBiomas Brazil etc.
3. improved representation of pasture/rangeland dynamics by using Global Pasture Watch livestock densities 2000-2022 and grassland classes 2000-2024

[1] *Karina Winkler et al 2025 Environ. Res. Commun. 7 055013*

[2] *https://doi.pangaea.de/10.1594/PANGAEA.974335.*

Funded by the European Union

Table 1: Observational and modelled datasets used for generating HILDA+ v3. **New and updated data sets in HILDA v3 are indicated in bold.**

Dataset and reference	Used thematic coverage	Spatial coverage	Used temporal coverage	Spatial resolution	Data type
Global datasets					
Copernicus LC100 https://land.copernicus.eu/global/products/lc	LCCS 22 classes	global	2015-2019	100 m	raster
ESA CCI Land Cover http://maps.elie.ucl.ac.be/CCI/viewer/download.php	LCCS 22 classes	global	1992-2019	300 m	raster
ESA WorldCover https://esa-worldcover.org/en/data-access	**LCCS 11 classes**	**global**	**2020-2021**	**10 m**	**raster**
Esri Land cover: Sentinel-2 10-Meter Land Use/Land Cover https://livingatlas.arcgis.com/landcover/	**LCCs 9 classes**	**global**	**2017-2024**	**10 m**	**raster**
GLAD UMD VCF https://glad.umd.edu/dataset/long-term-global-land-change	tree canopy, bare ground, short vegetation	global	1982-2016	0.05 deg	raster
GLC2000 http://forobs.jrc.ec.europa.eu/products/glc2000/glc2000.php	LCCS 22 classes	global	2000	1 km	raster
GLCNMO https://globalmaps.github.io/glcnmo.html	LCCS 22 classes	global	2003 2008, 2013	30 arc sec 15 arc sec	raster
Global Human Settlement Layer (GHSL)	built-up area (fractional)	global	1975, 1990, 2000, 2014	1 km	raster

Funded by the European Union

https://ghslsys.jrc.ec.europa.eu/datasets.php					
Global Urban Footprint (GUF) https://www.dlr.de/eoc/en/desktopdefault.aspx/tabid-11725/20508_read-47944/	built-up area (fractional)	global	2011/12	2.8 arc sec	raster
GlobCover http://due.esrin.esa.int/page_globcover.php	LCCS 22 classes	global	2005/2006, 2009	300 m	raster
Global Pasture Watch 2024 https://zenodo.org/records/13890401 https://zenodo.org/records/14933660 https://zenodo.org/records/14933641 https://zenodo.org/records/14933653	**grassland classes (cultivated, natural, open shrubland); cattle, sheep and goat density**	**global**	**2000-2022**	**30m**	**raster**
Globeland30 http://www.globeland30.org	10 LULC classes	global	2000, 2010	30 m	raster
Gridded Livestock World v3 (GLW) http://www.fao.org/livestock-systems/en/	density of ruminants	global	2010	5 arc min	raster
Hansen GFC https://storage.googleapis.com/earthenginepartners-hansen/GFC-2024-v1.12/download.html	**tree cover (fractional) loss and gain year**	**global**	**2000-2024 (loss) 2000-2012 (gain)**	**30 m**	**raster**
HYDE 3.3 (preliminary version, requested from authors 2022) https://www.pbl.nl/en/image/links/hyde	cropland, grazing land	global	1900-2019	5 arc min	raster

MODIS MCD12Q1 https://lpdaac.usgs.gov/products/mcd12q1v006/	**IGBP 17 classes**	**global**	**2001-2024 (yearly)**	**500 m**	**raster**
Ramankutty cropland (update) http://www.ramankuttylab.com/data.html	cropland	global	1900-2011	5 arc min	raster
Regional datasets					
AAFC Land Use Canada https://open.canada.ca/data/en/dataset/18e3ef1a-497c-40c6-8326-aac1a34a0dec	15 LULC classes	Canada	1990, 2000, 2010	30 m	raster
Australia DLCD V2.1 https://ecat.ga.gov.au/geonetwork/srv/eng/catalog.search#/metadata/83868	LCCS 22 classes	Australia	2002-2014	500 m	raster
CORINE https://land.copernicus.eu/pan-european/corine-land-cover/view	44 LULC classes with change layers	Europe (changing extent)	1990, 2000, 2006, 2012, 2018	100 m	raster
LULC classification of India https://daac.ornl.gov/cgi-bin/dsviewer.pl?ds_id=1336	11 LULC classes (IGBP scheme)	India	1985, 1995, 2005	100 m	raster
Mapbiomas Brazil https://brasil.mapbiomas.org/colecoes-mapbiomas/	**6+21+9 LULC classes**	**Brazil**	**1985-2024**	**30 m**	**raster**
MoEF Indonesia http://webgis.dephut.go.id:8080/kemenhut/index.php/en/feature/download	22 LULC classes	Indonesia	2000, 2003, 2006, 2009	300 m	raster

NLCD Land Cover (CONUS) https://www.mrlc.gov/data?f%5B0%5D=category%3Aland%20cover	16 LULC classes	U.S.	2001, 2006, 2011	30 m	raster
RCMRD Land Cover http://opendata.rcmrd.org/search?tags=land%20cover	6 LULC classes with country-specific sub-classes	Botswana, Ethiopia, Lesotho. Malawi, Namibia, Rwanda, Tanzania, Uganda, Zambia	different years between 2000 and 2014	30 m	raster
South Africa Land Cover http://www.sasdi.net/results.aspx?text=Land+cover&offset=0&f_text=land%20cover	35/72 LULC classes	South Africa	1990, 2013-14	30 m	raster

Funded by the European Union

3 Generation of fractional land-use coverage from HILDA+ v3

One part of the deliverable consists of a set of global, annual fractional land-use maps (0 – 1) for the six HILDA+ v3 classes (urban, cropland, pasture/rangeland, forest, unmanaged grass-/shrubland, sparse/no vegetation) covering the years 1900-2023. The fractions are provided as NetCDF files on a 0.01° × 0.01° WGS-84 grid and are ready for direct ingestion by climate-impact, carbon-budget and biodiversity models used within the EU **NextGenCarbon** project.

3.1 Scope and Objectives

The objective of this part of the deliverable is to deliver **harmonised, spatially-consistent fractional land-use layers** that:

- reflect the mean-average probabilities derived from the HILDA+ v3 probability maps (based on underlying observational datasets),
- respect the official HILDA+ state classification for each year,
- guarantee that the sum of fractions does not exceed 1 per pixel, and
- are supplied in an open, self-describing NetCDF format adhering to FAIR data-management principles.

These products form the basis for downstream ecosystem-service and greenhouse-gas accounting in the EU Horizon 2020 / Horizon Europe framework.

3.2 Data Sources

The data sources used for generating the land use fractions from HILDA+ v3 are listed in Table 2.

Table 2: Sources for generating fractional land-use coverage from HILDA+ v3.

Source	**Content**	**Temporal coverage**	**Spatial resolution**
HILDA+ v3 states	Harmonised land use/land cover (LULC) database, compiled from FAO statistics and multiple remote-sensing products (see Table 1)	Year-specific (1900-2023)	Native grid (≈ 0.01°)
“Probability” maps	Average class-specific fractions generated during HILDA+ pre-processing	Year-specific (1900-2023)	Native grid (≈ 0.01°)

The class identifiers used throughout the dataset are listed in Table 3.

Table 3: Mapping of land use class names to integer identifiers.

Class name	**ID**
Urban	1
Cropland	2
Pasture/rangeland	3
Forest	4
Unmanaged grass-/shrubland	5
Sparse/no vegetation	6

3.3 Methodology

The production chain follows four deterministic steps for each target year. Figure 1 provides a visual overview.

Funded by the European Union

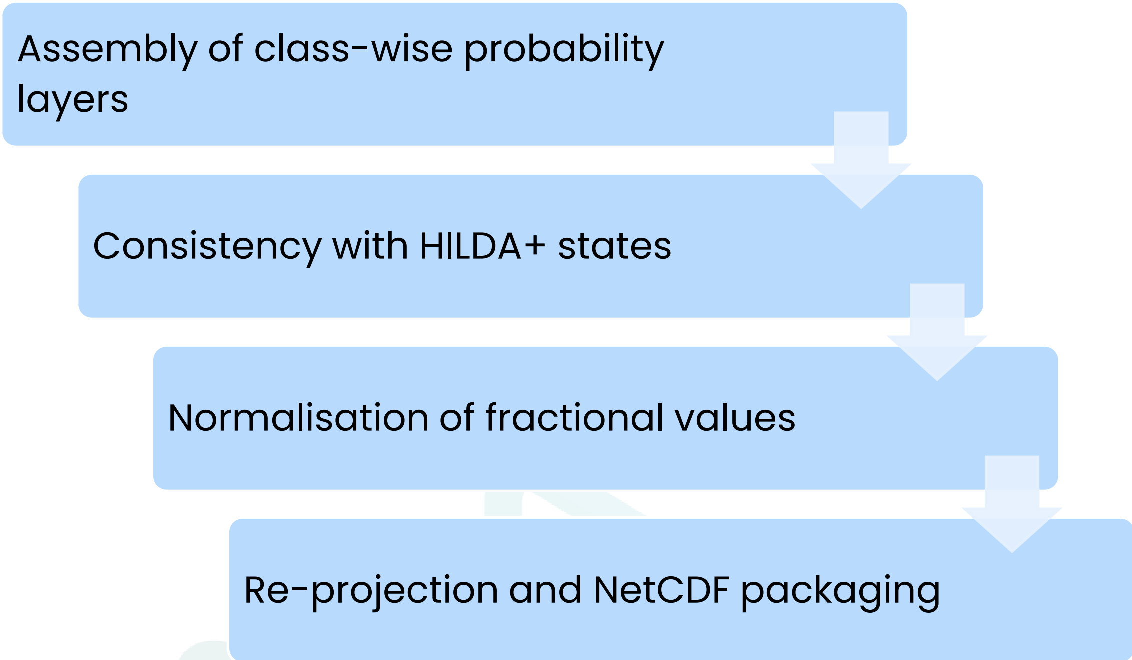


Figure 1: Workflow diagram for generating fractional land use coverage from HILDA+ v3.

3.3.1 Assembly of class-wise probability layers

For every year, the six probability rasters, based on the common agreement of the data in Table 1 and used for the allocation procedure, are stacked into a three-dimensional array (depth, row, column). A dummy first band (index 0) is inserted to simplify later argmax calculations. All arrays are stored as 32-bit floating-point values to preserve the original probability precision.

3.3.2 Consistency with HILDA+ states

The HILDA+ state raster encodes the dominant land-use class in the tens digit (e.g., 40 → forest). By integer-dividing the raster values by 10 we retrieve the class identifier. Where the state raster indicates a class that does not match the locally maximal probability, the corresponding probability value is substituted by the original maximal value (1). This step guarantees that the dominant class in the final fractions is identical to the official HILDA+ classification for the given year.

3.3.3 Normalisation of fractional values

Raw average fractions from the underlying input data to HILDA+ (probabilities) may sum to values greater than 1 in some pixels due to overlapping class confidence. A pixel-wise scaling factor **s** is applied:

$$s(i,j) = \frac{1}{\sum_c p_c(i,j)}$$

where

- $p_c(i,j)$ – the raw probability for class *c* at pixel *(i, j)*,
- $\sum_c p_c(i,j)$ – the per-pixel sum of all class probabilities,
- $s(i,j)$ – the scaling factor.

The scaled fractions are then clipped to the interval [0, 1] and rounded to four decimal places. After this operation the per-pixel sum statistic lies in the range 0.0 – 1.0 (see validation section).

3.3.4 Re-projection and NetCDF packaging

Each class fraction is re-projected from the native HILDA+ grid to the global WGS-84 grid using nearest-neighbour interpolation (GDAL's GRA_NearestNeighbour). The re-projected arrays are written directly into a four-dimensional NetCDF variable (landuse_fraction) with the dimensions *(time, class, y, x)*. The NetCDF files are created with the following characteristics:

- **Format:** NETCDF4_CLASSIC (CF-compliant, portable)
- **Compression:** zlib=True, complevel=6, shuffle=True

Funded by
the European Union

- **Chunking:** 1 × 6 × 256 × 256 (time × class × y × x) – optimised for efficient sub-setting
- **Metadata:**
    - time – integer year, long_name = "year"
    - class – zero-based index, long_name = "land-use class index"
    - global attribute class_names = "settlement,crop,past,forest,shrub,other"
    - coordinate variables x (longitude, degrees east) and y (latitude, degrees north)

The resulting files are named hilda_landuse_fractions_wgs84_<year>.nc and stored on a THREDDS server (link is shared with all project members):

3.4 Validation and quality assurance

After normalisation, the minimum, maximum and mean of the summed fractions are computed for each year. All values must satisfy $0 \leq \text{sum} \leq 1$ and the class with the maximum fractional value must match the class in the HILDA+ states layer (Figure 2).
As the fractional coverage values are not validated against observational data, they represent the mean fractional coverage of the underlying HILDA+ input data (see Table 1), which differ by year, and can therefore be affected by data artefacts and inconsistencies (misclassifications, varying datasets or spatial resolutions in time). Country-specific totals are not calibrated against FAO values.

Funded by the European Union

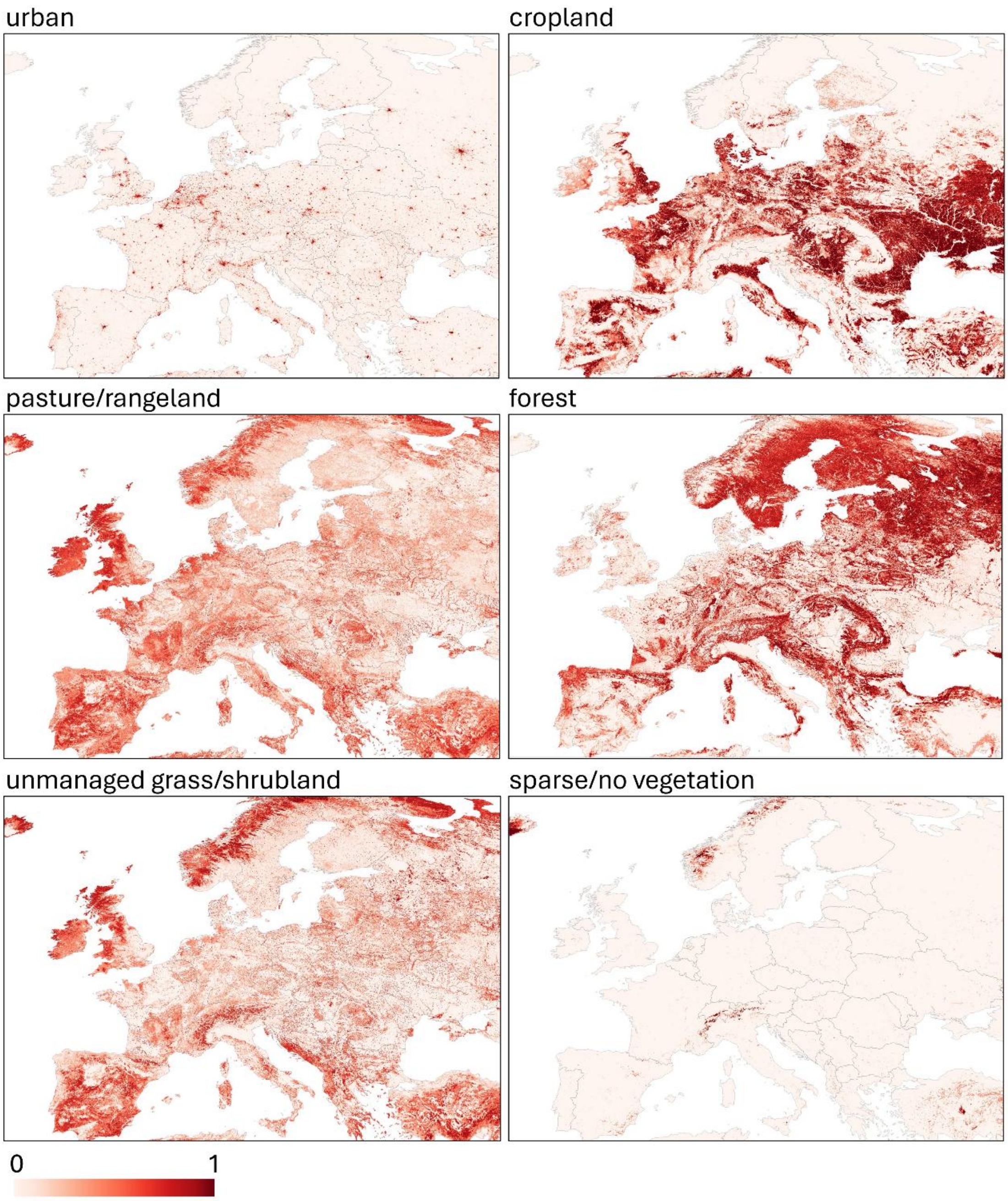


Figure 2: Fractional coverage of the six land use categories across Europe in 2023.

3.5 Usage guidelines

- **Software requirements** – Any CF-compliant NetCDF reader (e.g., Python netCDF4, R ncdf4, MATLAB, CDO, NCO) can read the files. No proprietary software is needed.
- **File structure** – Each NetCDF contains a single time slice (time = <year>). The landuse_fraction variable can be sliced as [: , class_index , :, :] to retrieve the full-coverage raster for a particular class.
- **Memory considerations** – Because the data are chunked, users can load a single class (≈ 150 MB per year) without exceeding modest RAM budgets.

4 Primary/secondary forest layers

Another part of the deliverable consists of a set of global, annual land-use maps for primary and secondary forests based on HILDA+ v3 covering the years 1960-2023. The maps have three classes: no forest (0), primary forest (1) and secondary forest (2) and are provided as NetCDF files on a 0.01° × 0.01° WGS-84 grid and are ready for direct ingestion by climate-impact, carbon-budget and biodiversity models used within the EU **NextGenCarbon** project.

4.1 Scope and Objectives

The objective of this part of the deliverable is to deliver **harmonised, spatially-consistent primary and secondary forest layers** that:

- reflect the state of primary and secondary forests derived from the HILDA+ v3 land use classes,
- are complementary and consistent with the official HILDA+ state classification for each year,
- are supplied in an open, self-describing NetCDF format adhering to FAIR data-management principles.

These products form the basis for downstream ecosystem-service and greenhouse-gas accounting in the EU Horizon 2020 / Horizon Europe framework.

4.2 Data sources

Annual HILDA+ v3 data from Milestone 6, see section 2 of this deliverable, in Eckert 4 projection from 1900 to 2023 was used to derive primary and secondary forest layers.

4.3 Methodology

In order to distinguish different carbon densities, for primary and secondary forests, the forest classes in HILDA+ v3 were separated into primary and secondary forest, with the following definition:

- Primary vegetation: HILDA+ forest area that remained a forest class since 1900
- Secondary vegetation: HILDA+ forest area that has been afforested since 1900 and not deforested again thereafter.
- The remaining HILDA+ classes were non-forest.

4.3.1 Step-wise modelling procedure

- In a 1st step, all data was reclassified into either forest or non-forest.
- In a 2nd step, forest gain transitions were derived for each iteration between one year to another (Figure 3).
- In a 3rd step, all forest gain transitions were accumulated for each year, equal to the data period it covered, e.g. from 1900 to year X (Figure 3).
- In a 4th step, all accumulated forest gains from 1900 up to a specific year of the modelling period was reclassified into secondary forests (2). All remaining forest into primary forest (1)
- In a 5th step, all primary (1) and secondary forest (2) data were merged onto data of step 1 for each year. The condition was as follows: For each forest area class from step 1, the data of step 4 was used instead. This enabled to use only those primary and secondary forest information that has not been deforested since 1900, meaning it had to stay forest after afforestation (Figure 3).
- In a 6th step, all primary/secondary data layers were converted from Eckert4 TIFFs into WGS84 and NetCDF file formats.

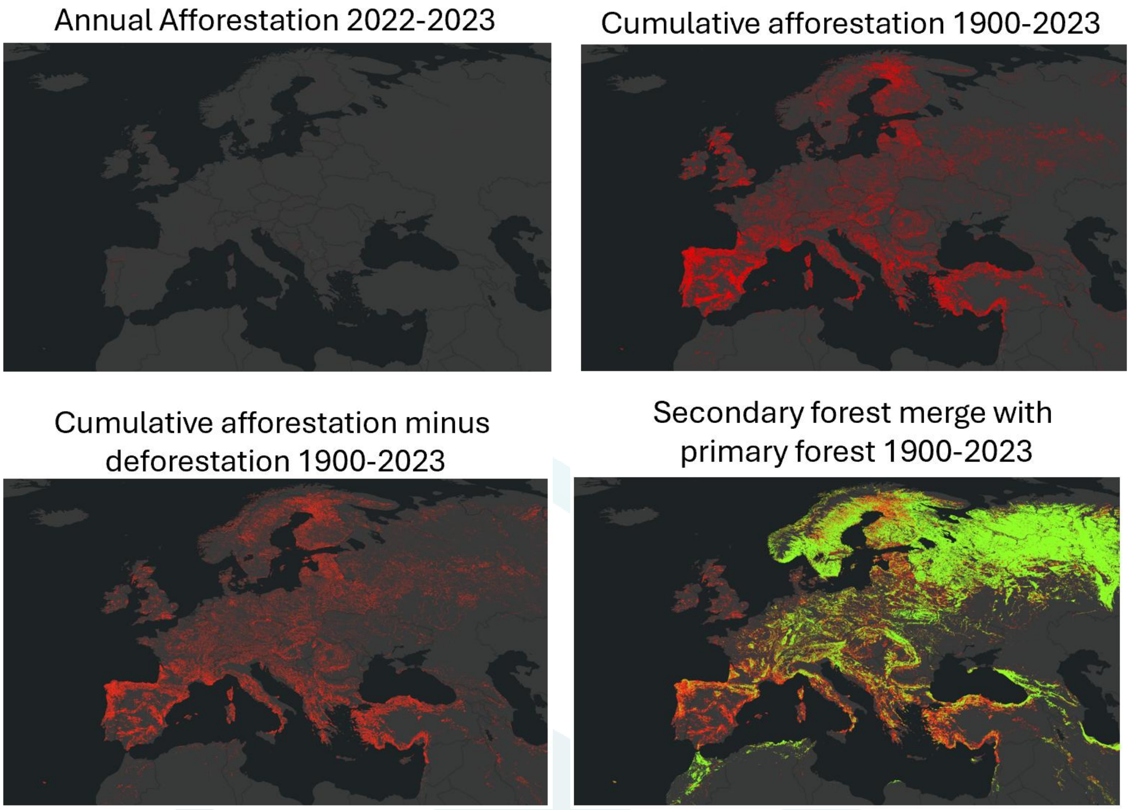


Figure 3: (top left) Annual secondary forests – step 2, (top right) Accumulated secondary forests – step 3, (bottom left) Accumulated secondary forests minus accumulated deforestation during same period – step 5, (bottom right) Primary and secondary vegetation combined – step 5.

4.4 Validation and quality assurance

Data was checked for consistency with original data from HILDA+ v3. For example, primary and secondary forest layers were overlayed with the original forest classes of the HILDA+ v3 data for complementary years to check misalignments or misclassification (Figure 4).

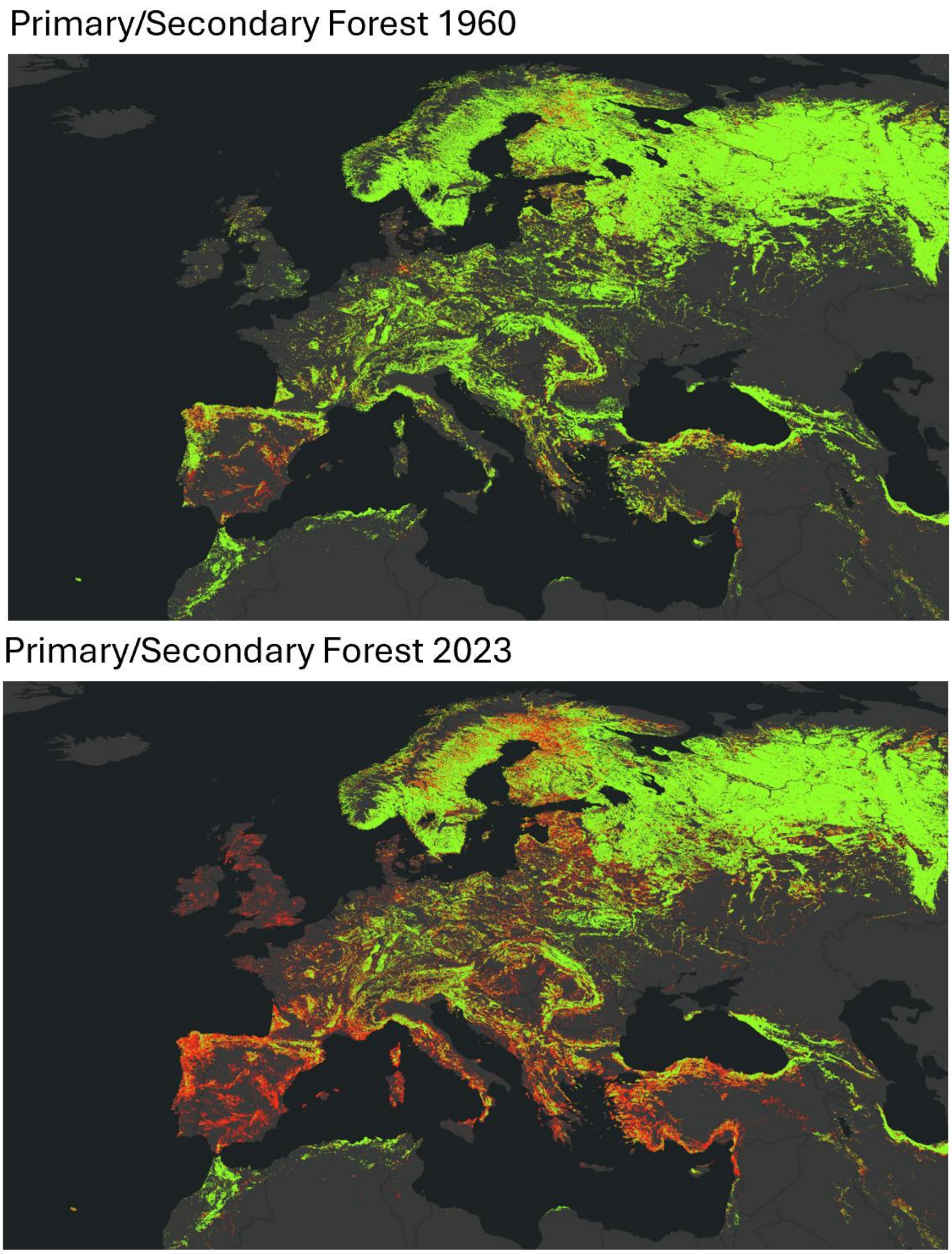


Figure 4: (top) Primary (green) and secondary (red) forests in 1960, (bottom) Primary (green) and secondary forests (red) in 2023.

Funded by
the European Union