EYE-CLIMA
Verifying emissions of climate forcers

# Input driving datasets for process-based models

## DELIVERABLE 2.1


| | |
|---|---|
| **Author(s):** | Philippe Peylin, Tuula Aalto, Tiina Markkanen<br>Almut Arneth, Jianyong Ma |



| | |
|---|---|
| **Date of submission:** | 19-02-2024 |
| **Version:** | 1 |
| **Responsible partner:** | CNRS-LSCE |
| **Deliverable due date:** | 31-08-2023 |
| **Dissemination level:** | Public |

| | |
|---|---|
| **Call:** | HORIZON-CL5-2022-D1-02 |
| **Topic:** | Climate Sciences and Responses |
| **Project Type:** | Research and Innovation Action |
| **Lead Beneficiary:** | NILU - Norsk Institutt for Luftforskning |

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101081395

## Document History

| Version | Date | Comment | Modifications made by |
|---|---|---|---|
| 0.1 | 30-01-2024 | First Draft | Philippe Peylin, Tuula Aalto, Tiina Markkanen, Almut Arneth, Jianyong Ma, Vladislav Bastrikov |
| 0.2 | 06-02-2024 | Internal review | Maria Tenkanen |
| 1.0 | 19-02-2024 | Submitted to the Commission | Rona Thompson |



This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101081395

## Summary

To simulate carbon dioxide ($CO_2$), methane ($CH_4$) and nitrous oxide ($N_2O$) fluxes across Europe using the selected process-based models (ORCHIDEE, LPJ-GUESS and JSBACH), both input datasets and additional data are required. These datasets are needed either to calibrate key model parameters or to validate the simulated fluxes. This deliverable thus consists of the description of the collected data for this modelling effort. Apart from the climate data, it is mostly a collection of existing datasets derived during related projects and community efforts rather than the processing of specific raw data. The main input datasets gathered are:

- Climate data that correspond to the ERA5-land reanalysis from ECMWF (at 11 km resolution) further bias corrected using the CRU monthly data. The product was first derived during the VERIFY project and has been extended in EYE-CLIMA to cover recent years (up to 2022).
- A European subset of the HILDA+ (HIstoric Land Dynamics Assessment+) dataset on land use/land cover (LULC) change. HILDA+ is a global dataset starting in 1960 at 1 km spatial resolution, integrating multiple open data streams (from high-resolution remote sensing, long-term land use reconstructions and statistics).
- Soil organic carbon stocks from the SoilGrids database will be used in ORCHDIEE to initialise the model soil C content. For soil physical properties, we will try to use the Land Use and Cover Area frame Statistical survey (LUCAS) topsoil data, although currently, the LPJ model uses the WISE dataset.
- Cropland management datasets: the MIRCA2000 global dataset with a spatial resolution of 0.083° will be used to provide both irrigated and rain-fed crop harvest areas for all major food crops. For reconstructing the history of anthropogenic nitrogen inputs to the terrestrial biosphere, a comprehensive and synthetic dataset from Tian et al. (2022) will be used.
- Grassland management datasets are still being gathered by the modelling groups in order to derive spatial and temporal information about cutting and grazing; to that end livestock density distribution maps, for different livestock categories, are key and they will be taken from the Gridded Livestock of the World dataset (GLW2; Robinson et al., 2014).
- Forest management and evaluation datasets: different datasets assembled in previous projects, based on National Forest Inventory (NFI) data and remote sensing data, have been gathered to calibrate and evaluate the forest demography of ORCHIDEE and LPJ-GUESS. Recent dataset for Europe from Pucher et al. (2022) are also being used (especially forest age classes and forest height).

The EYE-CLIMA model input datasets are, however, a living structure, growing out of a collaboration between the three modelling groups. As such, the collection of data and this document will continue to be updated as user needs evolve.




This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101081395

# TABLE OF CONTENTS





This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101081395

# 1. Introduction

For the quantification of greenhouse gas (GHG) surface fluxes within the EYE-CLIMA project, three process-based land surface models are going to play a central role. They integrate our physical understanding of the land surface processes controlling these fluxes and are used to extrapolate and interpolate knowledge, obtained from measurements and theory, spatially and temporally. The application of models requires, however, numerous datasets to parameterise the models and validate the results. Climate, soil, management (for cropland, forest and grassland) and land use/land cover are the main driving data that are required for modelling the carbon and nitrogen dynamics (namely $CO_2$, $CH_4$ and $N_2O$ fluxes). Therefore, the first aim of WP2 (Task 2.1) was to collect these data and provide them to the other tasks of WP2. The collection includes data needed to perform the simulations but also to improve and validate model outputs. This deliverable provides details about the actual status of the collected data that defines the first version of the input dataset. Further work will be done during the project to improve and extend specific datasets (especially those related to land management). A subsequent update of this deliverable will thus be provided during the course of the project.



This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101081395

## 2. Meteorological forcing dataset

Two input datasets were considered as potential meteorological forcings for the process-based models used in the EYE-CLIMA project:

- A product based on the ERA5-land climate reanalysis of ECMWF (European Center for Medium-Range Weather Forecast) with an additional bias correction, at the spatial resolution of around 11 km.
- A new reanalysis produced also by ECMWF specifically for Europe: the Copernicus Regional Reanalysis for Europe (CERRA), at the spatial resolution of 5.5 km; https://climate.copernicus.eu/copernicus-regional-reanalysis-europe-cerra.

Given that the CERRA product only starts in 1983 (which is sub-optimal to run the model over the historical period for $CO_2$) and that some preliminary analyses showed precipitation biases at some sites, we decided for the first round of simulations of EYE-CLIMA to use the ERA5-land bias corrected product. Such a product was first built in the VERIFY project and further extended in the CoCO2 project. In 2020, the VERIFY project, through the combined efforts of the University of East Anglia, ECMWF, and LSCE, processed high-resolution meteorological forcing data from the ERA5-Land dataset at 3-hourly resolution across Europe for the historical period: 1901 to 2019. The ERA5-Land has an operational status that guarantees that data for the previous year will be available by April of the current year.

The ERA5-Land reanalysis was then re-aligned with the CRU observational time series dataset. The CRU TS dataset was developed and has been subsequently updated, improved and maintained with support from a number of funders, principally the UK's Natural Environment Research Council (NERC) and the US Department of Energy. This procedure changes the monthly means of each 0.5° pixel to match that of CRU observations; consequently, the regional monthly-averaged climate is that of the CRU, while the sub-monthly and higher spatial resolution come from ERA5-Land. During the CoCO2 project, such a dataset was extended to the year 2021. Within EYE-CLIMA however, we have re-aligned the whole time series (from 1901 onwards) and not only the last year, 2022. The CRU dataset was recently updated to V4.07 on the 19 April 2023, a release that covers the period 1901-2022. More information on the new CRU version can be found at: https://crudata.uea.ac.uk/cru/data/hrg/cru_ts_4.07/Release_Notes_CRU_TS_4.07.txt

The re-aligned dataset is available from the VERIFY THREDDS server:
https://verifydb.lsce.ipsl.fr/thredds/verify/VERIFY_INPUT/CRUERA_V4.0/catalog.html
The most complete set of meteorological forcing are available at sub-daily [3H] time steps but also some variables are available at daily and monthly time steps. The list of available variables is provided in the table below. To access the THREDDS server, you need a login and password:

Login: vdbuser
Password: V3r1fy

In order to illustrate some features of this dataset, we provide below two figures for the annual mean and July mean of the surface air temperature (Figure 1) and the precipitation (Figure 2). For the temperature, we see that the year 2022 was on average slightly warmer than 2021 with a mean temperature across the European domain of 9.65°C for 2022 and 8.99°C for 2021. For July, it was the other way around with July 2021 being warmer than July 2022 because the heat wave was more pronounced in July 2021. For the precipitation, the annual mean was



This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101081395

lower in 2022 than in 2021 (1.77 vs 1.94 mm/day). This was also the case in July but with contrasting spatial patterns: Western Europe (i.e. France) was drier in 2021 but North Eastern Europe was much wetter.

***Table 1:*** *List of the variables and time steps in the dataset.*

| **Variable** | **3-hourly** | **Daily** | **Monthly** |
|---|---|---|---|
| *Tair* | X | X | X |
| *Tmax* | X | X | |
| *Tmin* | X | X | |
| *Wind_N & Wind_E* | X | X | |
| *WS* | X | X | |
| *Psurf* | X | X | |
| *LWdown* | X | X | |
| *SWdown* | X | X | X |
| *Qair* | X | X | |
| *Rainf* | X | | |
| *Snowf* | X | | |
| *Precipitation* | X | X | X |
| *RH* | X | X | |



This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101081395

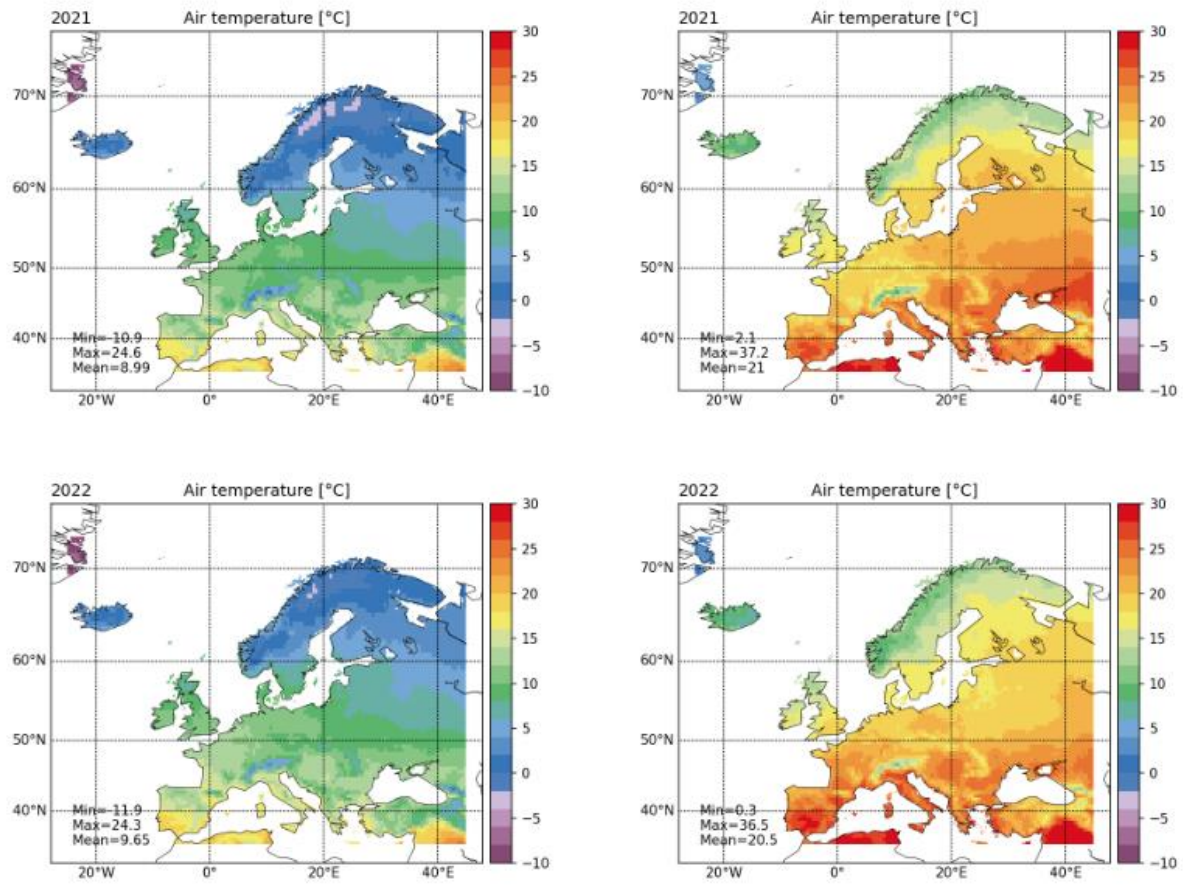


***Figure 1:*** *Illustration of the surface temperature forcing used within EYE-CLIMA (derived from ERA5-Land with CRU monthly bias correction) for two years, 2021 (top) and 2022 (bottom) over Europe. The left column shows the annual mean and the right column the July mean.*



This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101081395

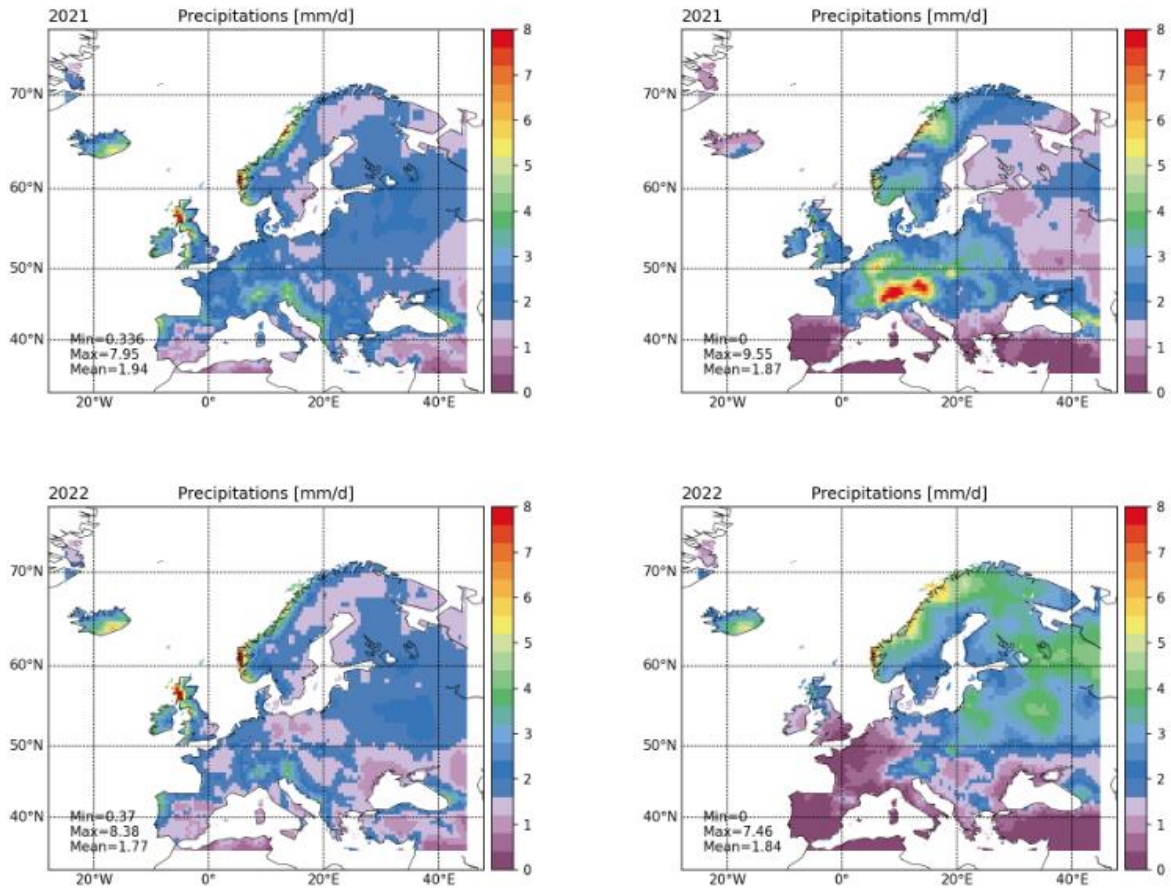


***Figure 2:*** *Illustration of the precipitation forcing used within EYE-CLIMA (derived from ERA5-Land with CRU monthly bias correction) for two years, 2021 (top) and 2022 (bottom) over Europe. The left column shows the annual mean and the right column the July mean.*

# 3. High-resolution land cover and land use data

## 3.1 HILDA+ land use/cover dataset

We proposed to use a European subset of the HILDA+ (HIstoric Land Dynamics Assessment+) global dataset on land use/land cover (LULC) change (Wrinkler et al., 2019). HILDA+ is a global dataset of land use/cover change starting in 1960 at 1 km spatial resolution and annual temporal resolution. It is based on a data-driven reconstruction approach and integrates multiple open data streams (from high-resolution remote sensing, long-term land use reconstructions and statistics). It covers six generic land use/cover categories: 1: Urban areas, 2: Cropland, 3: Pasture/rangeland, 4: Forest, 5: Unmanaged grass/shrubland, 6: Sparse/no vegetation. Forest generic type is further refined into different plant functional types (see Figure 3).

Starting with a FAO-calibrated base map (derived from ESA Copernicus LC100 2015), HILDA+ allocates land use/land cover transitions iteratively for each time step (annually) and for each country along a backwards-looking time loop on a 1x1 km grid. Net change magnitudes are based on national FAO land use and population statistics. Gross change magnitudes are calculated from mean transition matrices, which are extracted from a time series of satellite-



This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101081395

derived land use/land cover maps. The change allocation depends on class probability maps (mean class fractions) generated from year- and region-specific remote sensing-based land use/land cover maps.

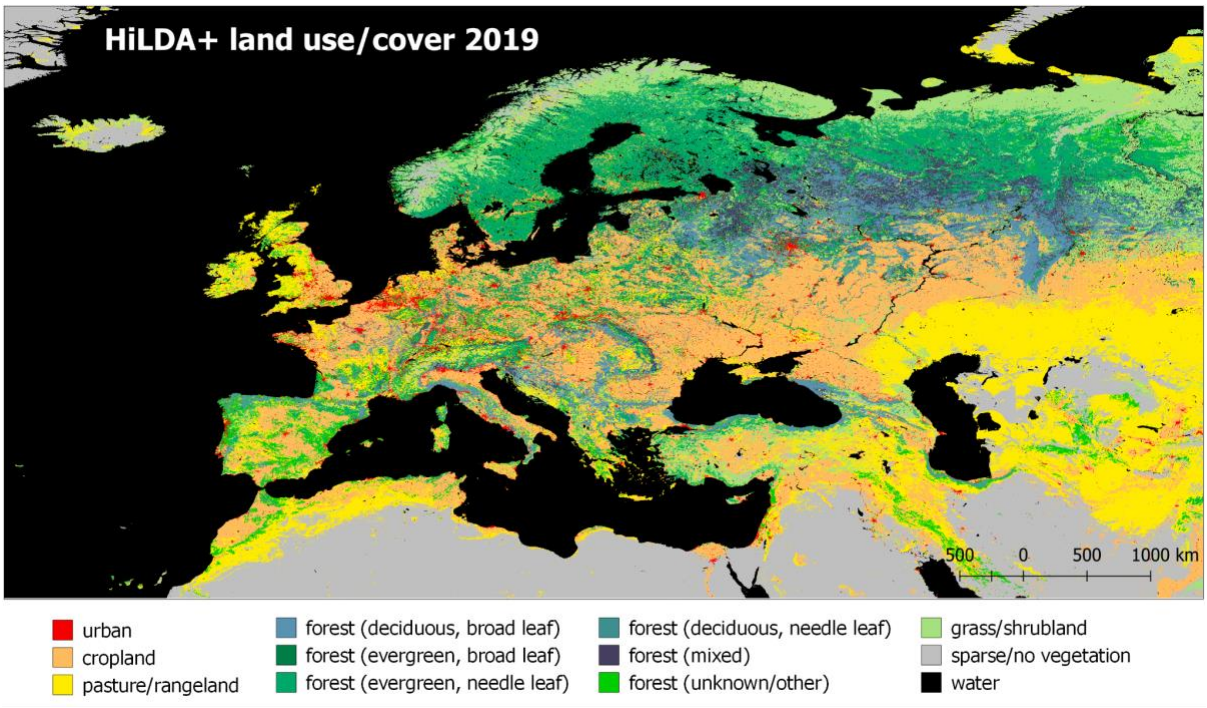


***Figure 3:*** *Land use/cover from HILDA+ to be used within EYE-CLIMA.*

## 3.2 Mapping HILDA+ to ORCHIDEE Plant Functional Types

From the HILDA+ product, we had to map the generic land cover into the specific Plant Functional Types (15 PFTs) of the ORCHIDEE land surface model. The approach consists of defining a cross-walking table (CWT) between the generic land use/land cover classes of HILDA+ and the 15 PFTs of ORCHIDEE, using also additional information such as: i) a climate zone definition from Koppen Geiger, ii) the C4 grassland fraction from Still et al. (2018) and the C4 crop fraction from LUH2 historical dataset. The figure below illustrates how such CWT is defined with the “Still” and “LUH2” additional maps.

The process with the different maps that are used as well as the resulting PFTs is illustrated on a dedicated webpage of the ORCHIDEE model development: https://orchidas.lsce.ipsl.fr/dev/verify/hilda.php

| KG CLASS | 11 Urban | 22 Cropland | 33 Pasture | 40 Forest Unknown/Other | 41 Forest EvNe | 42 Forest EvBr | 43 Forest DeNe | 44 Forest DeBr | 45 Forest Mixed | 55 Grass/shrubland | 66 Other land | 77 Water | 0 No data |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Tropical | 80% → PFT1<br>20% → PFT14 | PFT12/13 (using LUH2) | PFT14/11 (using Still) | 50% → PFT2<br>50% → PFT3 | PFT2 | PFT2 | PFT3 | PFT3 | 50% → PFT2<br>50% → PFT3 | 40% → PFT2<br>40% → PFT3<br>20% → PFT14/11 (using Still) | PFT1 | PFT1 | masked at ocean<br>dillute between given PFTs at coast |
| Arid Warm | 80% → PFT1<br>20% → PFT10 | | PFT10/11 using Still | 33% → PFT4<br>33% → PFT5<br>34% → PFT6 | PFT4 | PFT5 | PFT6 | PFT6 | 33% → PFT4<br>33% → PFT5<br>34% → PFT6 | 26% → PFT4<br>27% → PFT5<br>27% → PFT6<br>20% → PFT10/11 (using Still) | | | |
| Arid Cool | | | | | | | | | | | | | |
| Temperate Warm | | | | | | | | | | | | | |
| Temperate Cool | | | | | | | | | | | | | |
| Boreal Warm | 80% → PFT1<br>20% → PFT15 | | PFT15/11 (using Still) | 33% → PFT7<br>33% → PFT8<br>34% → PFT9 | PFT7 | PFT7 | PFT9 | PFT8 | 33% → PFT7<br>33% → PFT8<br>34% → PFT9 | 26% → PFT7<br>27% → PFT8<br>27% → PFT9<br>20% → PFT15/11 (using Still) | | | |
| Boreal Cool | | | | | | | | | | | | | |

***Figure 4:*** *Cross-walking table from HILDA+ classes to ORCHIDEE PFTs.*



This project has received funding from the European Union’s Horizon Europe research and innovation programme under grant agreement No 101081395

The figure 5 illustrates the spatial distribution and temporal evolution of one PFT of ORCHIDEE (Temperate Broadleaf Summer-green forest, PFT6) following the use of HILDA+ classes. We notice a significant increase in this forest PFT at the expense of crops and natural grassland.

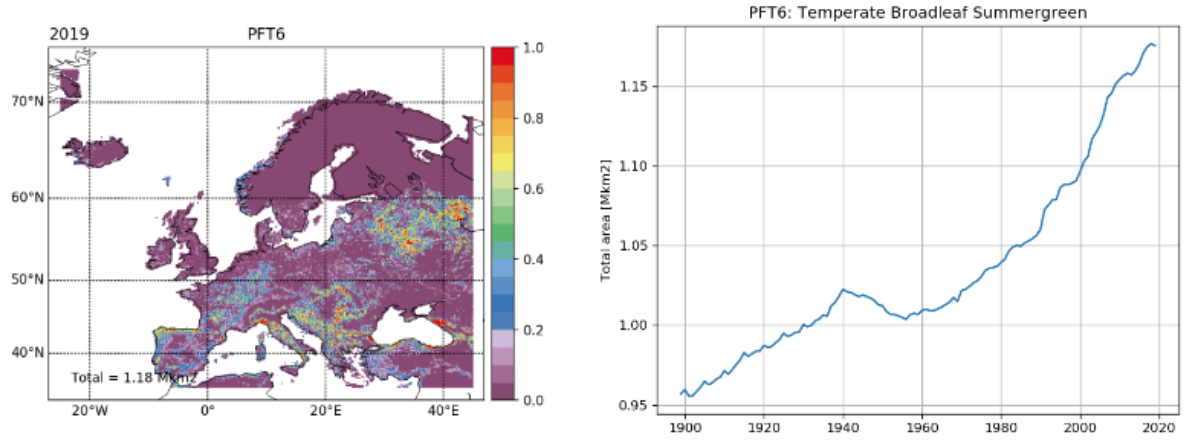


***Figure 5:*** *Spatial distribution and temporal evolution of the Temperate Broadleaf Summer-green forest (PFT6) of ORCHIDEE following the use of the HILDA+ land use/land cover classes and the cross-walking approach defined above.*

## 3.3 Mapping HILDA+ to LPJ-GUESS Plant Functional Types

The LULC information required in LPJ-GUESS primarily includes three vegetation types (cropland, pasture, and natural vegetation). Thus, we aggregated the HILDA+ classes forest (for the initial simulations including the managed forest) and unmanaged grass/shrubland to "natural vegetation" category, and regridded from 0.01° to 0.1° resolution across Europe. Additionally, the classes of pasture/rangeland and cropland in HILDA+ are aggregated to "pasture" and "cropland" categories in LPJ-GUESS, respectively, at the same resolution (see Table 2 below for details).

***Table 2*:** Categorising land use/cover from HILDA+ to LPJ-GUESS at 0.1° resolution

| **Land use/cover categories in HILDA+ (from)** | **Land use/cover categories in LPJ-GUESS (to)** |
|---|---|
| Urban | Urban |
| Cropland | Cropland |
| Pasture/Rangeland | Pasture |
| Evergreen needle leaf forest (unmanaged) | Natural Vegetation |
| Evergreen broad leaf forest (unmanaged) | |
| Deciduous needle leaf forest (unmanaged) | |
| Deciduous broad leaf forest (unmanaged) | |
| Mixed forest (unmanaged) | |
| Other forest (managed) | |
| Grass/Shrubland (unmanaged) | |
| Sparse/no vegetation | Barren |
| Water | N/A |



This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101081395

As most forests in Europe are managed, we will explore the impacts of managed forests in the future as well, but this will require more investigation into what HILDA+ classifies as managed and discussion of how (in the absence of gridded information) forest management would be best implemented. We expect that we can cooperate here with other EU projects such as ForestPaths or ClimbForest.

Figure 6 illustrates the spatial patterns of three vegetation types in LPJ-GUESS following the use of HILDA+ classes at 0.1° resolution (a-c). We compare the total areas of each land use/cover class between LUH2 and HILDA+ across Europe (d) and notice a significant area difference between these two datasets in natural vegetation, cropland, and pasture from 1901-2020.

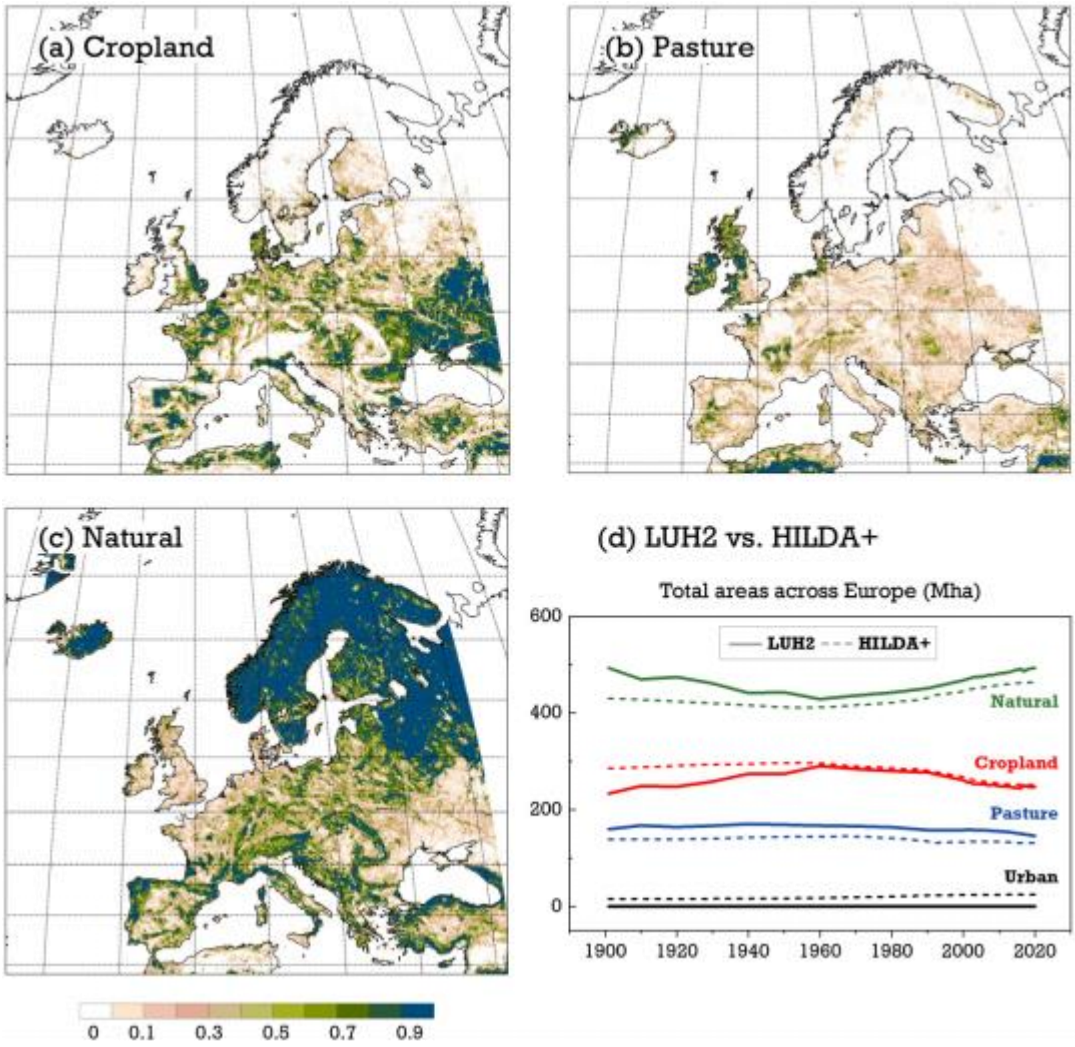


***Figure 6:*** *Maps of land use/cover fractions (a: cropland; b: pasture; c: natural vegetation) of the remapped HILDA+ datasets in LPJ-GUESS (averaged over 2011-2020), and the comparison of total areas in each land use/cover class between LUH2 and HILDA+ datasets across Europe from 1901-2020 (d).*



This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101081395

### 3.3 Mapping HILDA+ to JSBACH Plant Functional Types

Similarly to ORCHIDEE, for JSBACH we will map HILDA+ types into non-peatland PFTs. The cross-walking table of Figure 4 can be adopted for JSBACH surface fields as such. The status of HILDA+ in 2020 can be used to represent the current-day non-peatland PFT distribution.

### 3.4 Wetland spatial and temporal extent

For the current-day peatland extent, we use various region-specific data sources. For the first submission, the EU-CORINE land cover (CLC, https://land.copernicus.eu/en/products/corine-land-cover) data is used to derive peatland cover in the EU region. The CLC classes “bogs” and “inland marshes” are interpreted as peatland.

For the remainder of the rectangular model domain, which is not covered by CLC, the peatland distribution is based on the fraction of Histosols in the FAO Harmonised World Soil Database (HWSD).

In addition to the constant fractional distribution of peatlands, we use the monthly inundation fraction data WAD2M version 2.0 (Zhang et al. 2021) to simulate methane emissions from inundated mineral soil. The areal fraction of peatland (as defined above) in each grid cell is first subtracted from the monthly inundated fraction and the remainder is assigned as inundated mineral soil fraction. The land area after removing the peatland and the varying inundated fraction have been assigned for upland mineral soils.

For the second submission, we plan to use the newly released global wetland and wetland loss time series data by Fluet-Chouinard et al. (2023) (https://doi.org/10.5281/zenodo.7293597). We use the coverage of the natural wetland class in 2020 to represent peatlands. Alternatively, higher resolution GLWDv.2 (update of Lehner and Döll, 2004) wetland data can be used if available and better suited for the purpose.

For Finland, where intense drainage of pristine peatlands for forestry and agriculture took place during the 20th century, we plan to use the detailed data from GTK (Geological Survey of Finland, https://tupa.gtk.fi/paikkatieto/meta/suotyypit_ja_turvekankaat.html), which was recently published (2023) and contain remaining peat stocks both in managed and pristine peatlands. The land cover classes are detailed by various, mainly plant species distribution based, peatland type characteristics and state of drainage and current land use at.

## 4. Soil organic carbon stocks and soil properties

### 4.1 Soil properties

Although the three models (ORCHIDEE, LPJ-GUESS and JSBACH) are using different soil properties datasets in their current settings, we will try to harmonise them in the course of the project towards the use of a common dataset. We will use the top soil physical properties for Europe based on the Land Use and Cover Area frame Statistical survey (LUCAS) topsoil data. These data are downloaded from the following website: https://esdac.jrc.ec.europa.eu/

LUCAS aimed at collecting harmonised data about the state of land use/cover over the European Union (EU). Among these 200k land use/cover observations selected for validation, a topsoil survey was conducted at about 10% of these sites. Topsoil sampling locations were selected to be representative of the European landscape using a Latin hypercube stratified random sampling, taking into account CORINE land cover 2000, the Shuttle Radar Topography Mission (SRTM) DEM and its derived slope, aspect and curvature. Several soil properties were predicted using hybrid approaches like regression kriging. For those datasets,



This project has received funding from the European Union’s Horizon Europe research and innovation programme under grant agreement No 101081395

topsoil texture and related derived physical properties were predicted. Regression models were fitted using, along with other variables, remotely sensed data coming from the MODIS sensor. The high temporal resolution of MODIS allowed the detection of changes in the vegetative response due to soil properties, which can then be used to map soil feature distribution. The prediction of intrinsically co-linear variables like soil texture required the use of models capable of dealing with multivariate constrained dependent variables like Multivariate Adaptive Regression Splines (MARS). Cross-validation of the fitted models showed that the LUCAS dataset constitutes a good sample for mapping purposes leading to cross-validation R2 between 0.47 and 0.50 for soil texture and normalised errors between 4 and 10%.

This dataset provides the following soil properties at 500 m resolution, for the geographical coverage: European Union (EU) plus Balkan countries, Switzerland and Norway:

- Clay content (%) in topsoil (0-20cm) modelled by Multivariate Additive Regression Splines
- Silt content (%) in topsoil modelled by Multivariate Additive Regression Splines
- Sand content (%) in topsoil modelled by Multivariate Additive Regression Splines
- Coarse fragments (%) content in topsoil modelled by Multivariate Additive Regression Splines
- Bulk density derived from soil texture datasets (obtained from the packing density and the mapped clay content following the equation of Jones et al. 2003)

With the ORCHIDEE model, although we are using the global USDA soil texture map as the standard, a test with the LUCAS dataset will be made to see the impact of different soil properties on the soil water holding capacity and consequently on the different GHG fluxes.

For the LPJ–GUESS model, we are currently using the WISE databases for Europe (https://www.isric.org/explore/wise-databases). Like with ORCHIDEE, a test will be made with the LUCAS dataset to see the impact, especially for the N transformations in the soil.

With the JSBACH model, we use the grid-based data from FAO Harmonized World Soil Database (WHSD) as the soil texture map. Further, we use the PEATMAP by Xu et al. (2018) to adjust the soil hydrology parameter values in accordance with the peat content. Other maps (such as Tanneberger et al., 2017) will be tested as well. Peatland parameters from Hagemann and Stacke (2015) are used to describe the soil properties, e.g. soil porosity, saturated hydraulic conductivity, field capacity and wilting points and saturated moisture potential.

## 4.2 Soil organic carbon stocks

With the ORCHIDEE model, we use the SoilGrids database (see https://www.isric.org/explore/soilgrids; Ribeiro and Batjes (2019) ; Poggio et al. (20121)) to initialise the model soil organic carbon content (and its vertical distribution) and then we let the model equilibrate with a long spin-up simulation of thousands of years, recycling a 10-year climate forcing. Figure 7 illustrates the soil organic carbon content for the upper 30 cm of soil as estimated by the SoilGrids product. Note that we also have optimised key soil organic matter decomposition parameters from the CENTURY module (used in ORCHIDEE) so that the SOC content after the spin-up remains close to that of the SoilGrids database. This SOC content is also used to weigh soil dry and solid thermal conductivities, thermal capacity and to correct the value of the porosity used to calculate the saturated conductivity and the saturation



This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101081395

ratio. With ORCHIDEE, the SoilGrids data are thus used both for the initialisation and the evaluation after the spin-up.

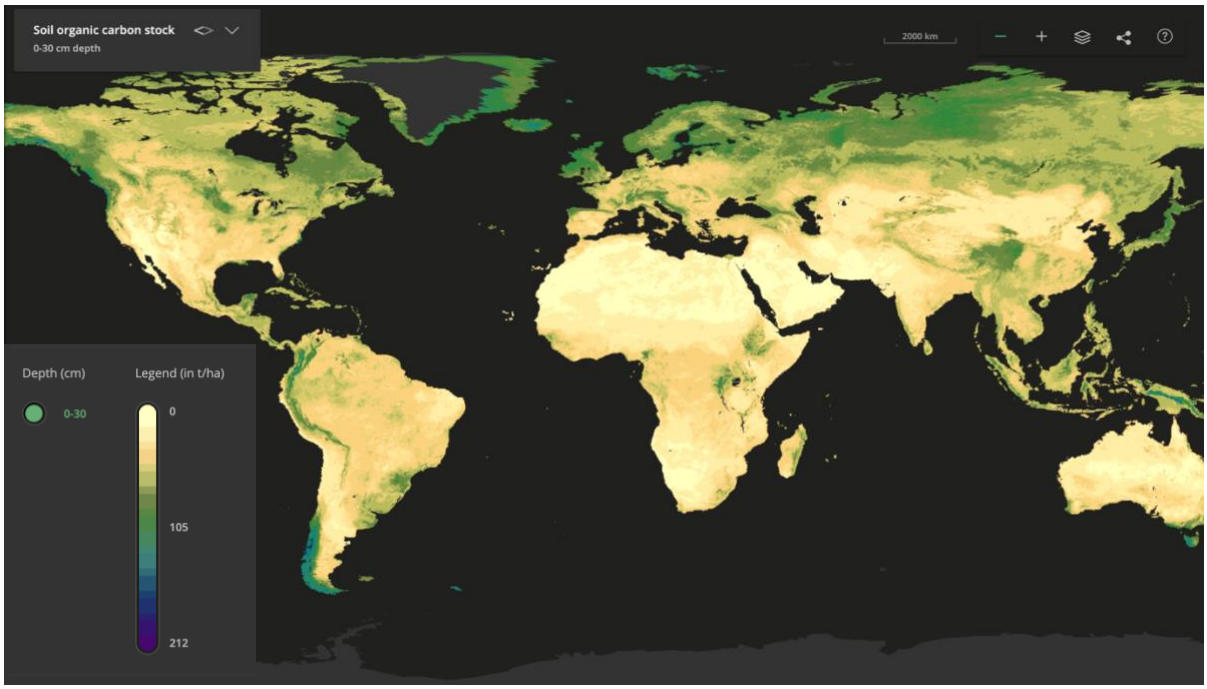


***Figure 7:*** *Maps of soil organic carbon stock (in t/ha) from the SoilGrids dataset.*

With the JSBACH model, soil organic carbon stocks are accumulated in a long spin-up simulation of thousands of years using as an input a period of climate drivers in a loop, and corresponding simulation of NPP and water table levels. The peat accumulation is a continuous process in water-logged conditions, and thus the simulation is not attempted to run until equilibrium, rather the simulation is ended when peat depths reach current-day levels.

Similar to the JSBACH model, soil organic carbon stock in LPJ-GUESS is computed as the long-term accumulation under the potential natural vegetation with a 1000-year spin-up simulation.

# 5. Cropland management datasets

We describe below the cropland management datasets that are currently used by the LPJ-GUESS model. These datasets will also be used with the ORCHIDEE model as much as possible. Currently, the ORCHIDEE model team is integrating the cropland specificities developed in a separate branch, ORCHIDEE-CROP (Wu et al., 2016) into the main version of the model that includes the nitrogen cycle and that will be used to simulate the three GHG fluxes. The first set of ORCHIDEE simulations for EYE-CLIMA will thus be made with a version that only accounts for standard C3 and C4 crops and thus using the FAO dataset (FAOSTAT, 2023) for the spatial distribution of these two “photosynthetic-pathways” of crops. The next round of simulation with ORCHIDEE will include, similar to LPJ-GUESS, major crop functional types for Europe and use the datasets described below.

## 5.1 Crop growth distribution dataset

MIRCA2000 (Portmann et al., 2010) is a global dataset with a spatial resolution of 5 arc minutes (=0.083°) which provides both irrigated and rain-fed crop harvest areas of 26 crop



This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101081395

classes around the year 2000. The dataset includes all major food crops (wheat, maize, rice, barley, rye, millet, sorghum, soybean, sunflower, potato, cassava, sugarcane, sugar beet, oil palm, canola, groundnut, pulses, citrus, date palm, grape, cocoa, coffee, other perennials, fodder grasses, other annuals) as well as cotton.

At present, cropland in LPJ-GUESS is characterised by six "crop functional types" (CFTs): two temperate C3 crops with spring and autumn sowing dates, a tropical C3 crop representing rice, a C4 crop representing maize, and two N-fixing grain legumes representing soybean and pulses. Considering the importance of barley, pulses (e.g., beans and peas), rapeseed, and maize in the overall agriculture in Europe (e.g., harvest areas and total production; FAOSTAT, 2023), we thus aggregated these main food crops from MIRCA2000 to CFTs in LPJ-GUESS from 0.083° to 0.1° resolution for better accounting for agricultural production in this region. The details on how we mapped the MIRCA2000 crops to the LPJ-GUESS CFTs are given in Table 3 and Figure 8 below.

Our estimated total areas of six CFTs in Europe are in general higher than the statistics from FAO (Figure 9), most likely due to the inclusion of parts of areas in Turkey and Russia across the European domain in our estimation (Figure 8). The overestimation is expected to be largely diminished when these two countries are removed from the comparison.

***Table 3*:** Categorizing crop types from MIRCA2000 to LPJ-GUESS at 0.1° resolution

| **Crop classes in MIRCA2000 (from)** | **Crop functional types (CFTs) in LPJ-GUESS (to)** |
|---|---|
| Potatoes, Sugar beet, Sunflowers | C3 crops sown in spring (representing spring wheat) |
| Barley, Rapeseed, Rye, Wheat | C3 crops sown in autumn (representing winter wheat) |
| Maize, Millet, Sorghum | C4 crops (representing maize) |
| Rice | Rice |
| Soybean | Soybean |
| Pulses | Pulses (representing faba bean) |
| Others (e.g., Groundnut, Oil palm, Sugarcane, Date palm, Citrus, Cocoa, Coffee, Cassava) | N/A |



This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101081395

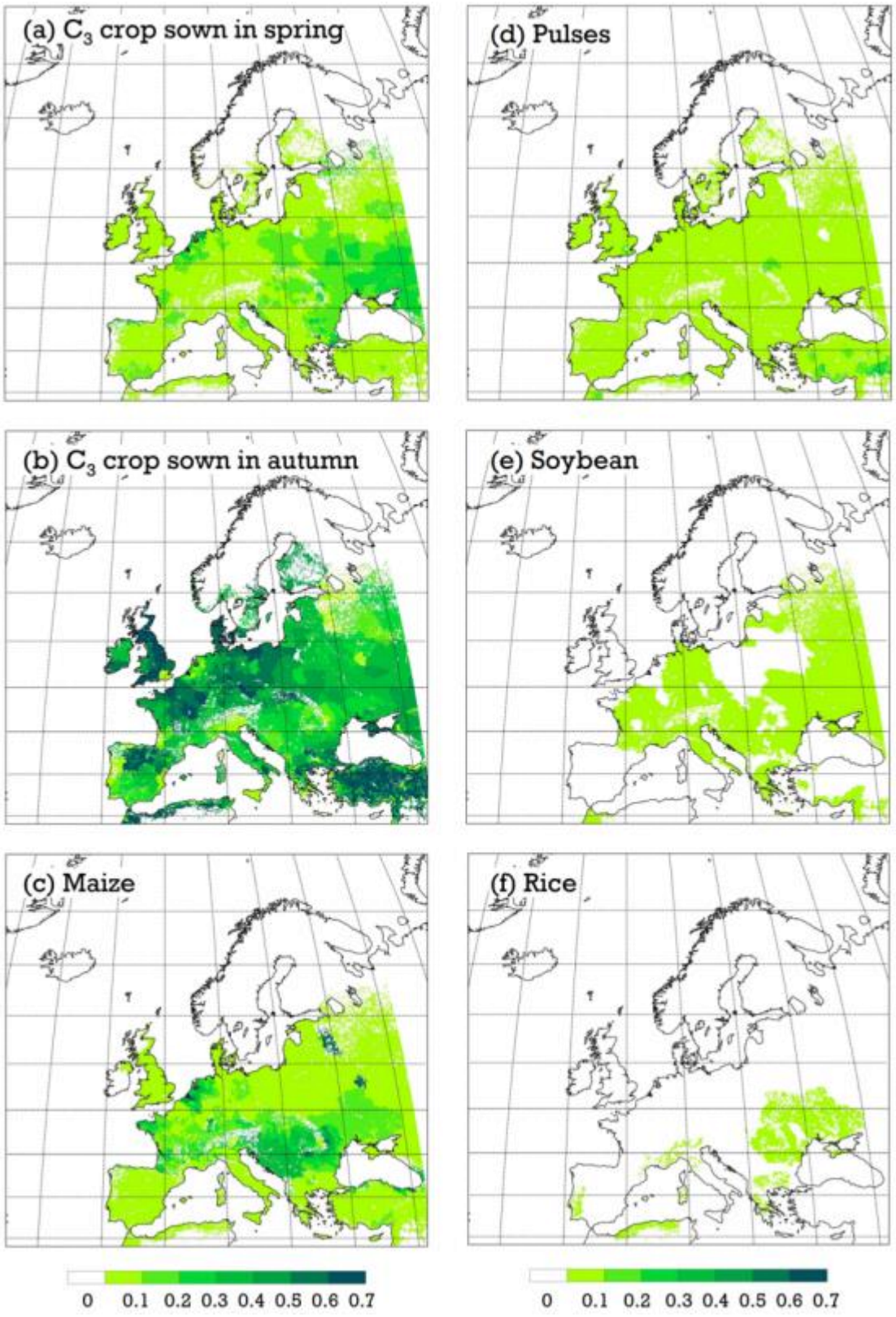


***Figure 8:*** *Maps of the rain-fed crop fractions, aggregated from MIRCA2000 crops to CFTs in LPJ-GUESS across Europe at 0.1° resolution.*



This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101081395

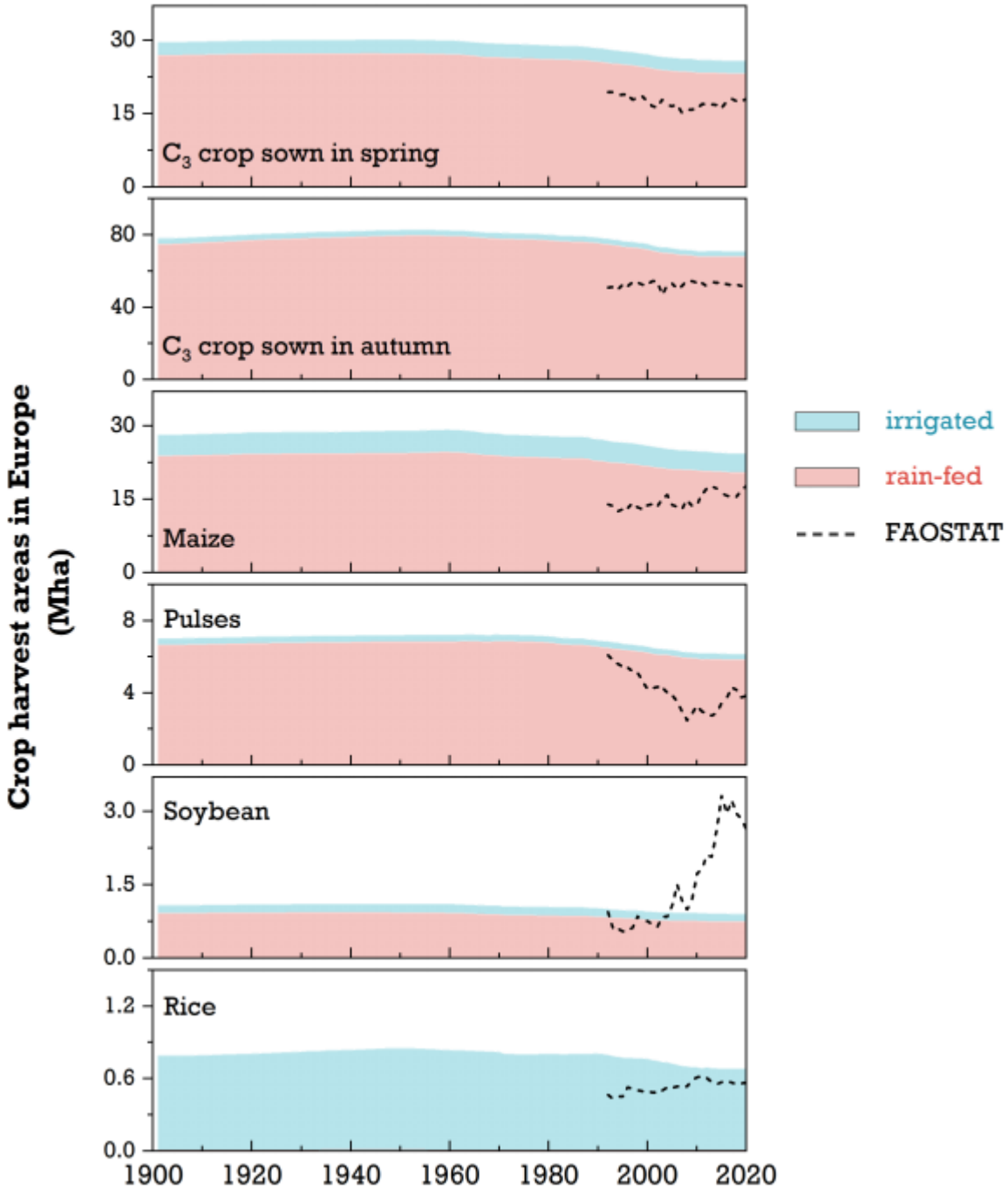


***Figure 9***: *Simulated crop-specific total areas (rain-fed and irrigated; Mha) by combining HILDA+ with MIRCA2000 datasets over the historical period across Europe in LPJ-GUESS. The statistics from FAO between 1992-2020 used for comparison are shown in dashed lines.*

## 5.2 Nitrogen fertilisation dataset

Tian et al. (2022) developed a comprehensive and synthetic dataset for reconstructing the History of anthropogenic Nitrogen inputs (HaNi) to the terrestrial biosphere. The HaNi dataset takes advantage of different data sources in a spatiotemporally consistent way to generate a set of gridded high-resolution N input products from the preindustrial period to the present (1860-2019). The HaNi dataset includes annual rates of synthetic N fertilizer, manure



This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101081395

application/deposition, and atmospheric N deposition on cropland, pasture, and rangeland at a spatial resolution of 0.083°.

We first aggregated the HaNi classes $NH_4^+$ and $NO_3^-$ N fertilizer to the "synthetic N fertilizer" category from 0.083° to 0.1° resolution across European cropland and pasture (Table 4). Since the HaNi does not provide the crop-specific N input rates, we subsequently separated the total N application rates in each grid cell to each crop (Figure 10) using the crop fraction information from the MIRCA2000 dataset. Due to the unavailable information for the timing of N fertilizer application on a regional scale, we assumed with LPJ-GUESS that synthetic fertilizer application takes place at three crop development stages — sowing, halfway through the vegetative phase, and flowering — with different application rates depending on crop type. All manure is applied to crops at the time of sowing as a single application to reflect real-world practices that account for the time required for manure N to be made available to plants.

We compared the total N inputs from the HaNi with the statistics from FAO. Preliminary results indicate that our estimates of N fertilizer rates and manure application to agricultural soils in Europe are in general agreement with the FAO-based records in terms of both magnitude and long-term trends (Figure 11).

***Table 4:*** Categorizing N inputs types from HaNi to LPJ-GUESS at 0.1° resolution in Europe

| **N input classes in HaNi (from)** | **N fertilization types in LPJ-GUESS (to)** |
|---|---|
| $NH_4^+$ N fertilizer in cropland | Synthetic N fertilizer in cropland |
| $NO_3^-$ N fertilizer in cropland | |
| Manure application in cropland | Manure application in cropland |
| $NH_4^+$ N fertilizer in pasture | Synthetic N fertilizer in pasture |
| $NO_3^-$ N fertilizer in pasture | |
| Manure application in pasture | N/A |
| Manure deposition in pasture | In the model, 25% of N in the harvested AGB on pasture is returned to the soils to simply account for the manure deposition from grazing animals |
| Manure deposition in rangeland | N/A |



This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101081395

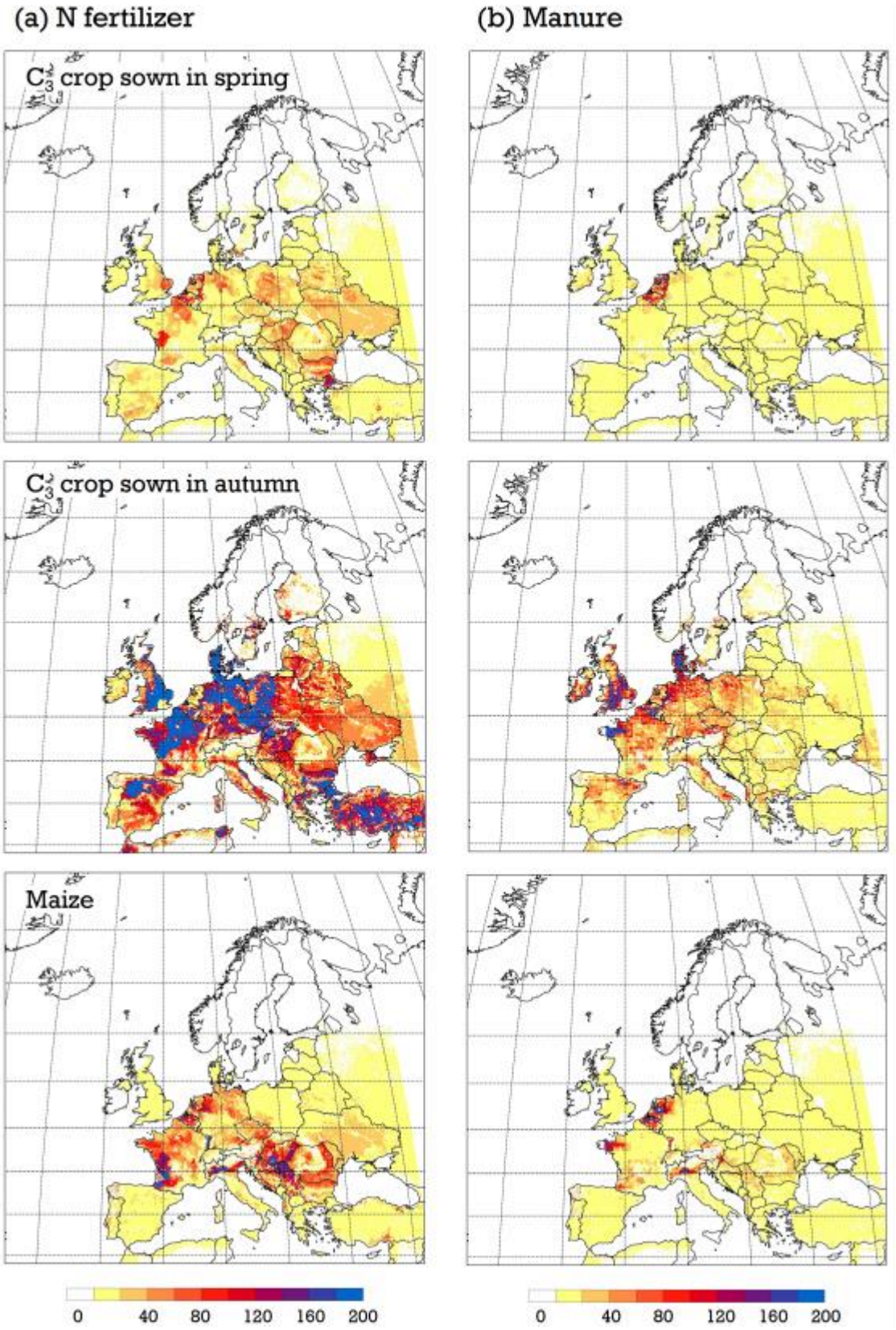


***Figure 10:*** *Maps of N fertilizer and manure applied to the three main crop types in Europe (tons N per grid cell averaged over 2010-2019), remapped from the HaNi dataset at 0.1° resolution for LPJ-GUESS.*



This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101081395

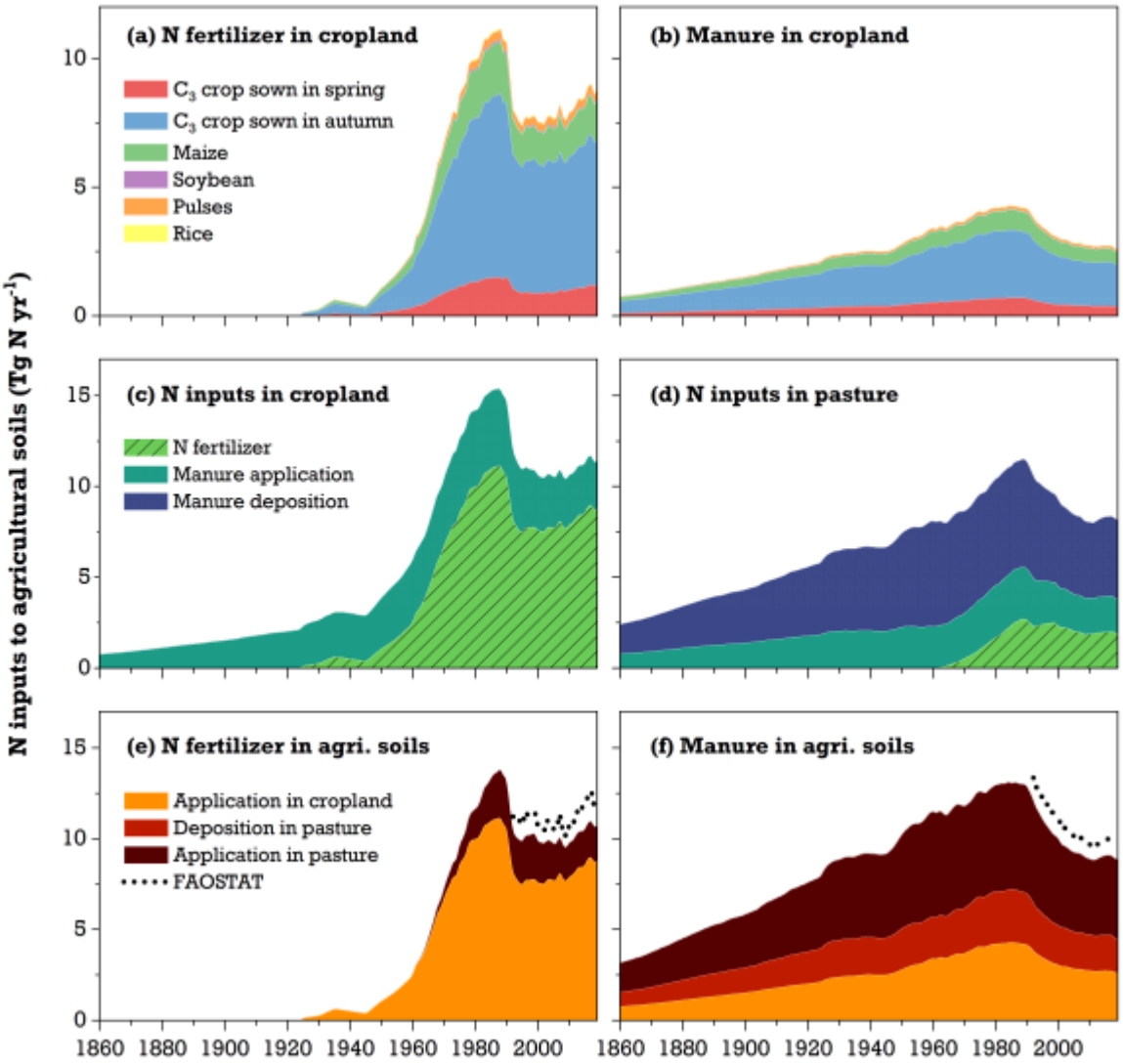


***Figure 11:*** *The estimated total N inputs to cropland and pasture from the HaNi dataset between 1860 and 2019 across Europe. The statistics from FAO between 1992-2019 used for comparison are shown in dots.*

It should be noted that harmonisation between different datasets needs to be done before running the simulation regionally. More specifically, all the management datasets on agricultural soils (i.e., MIRCA2000 and HaNi) should be gap-filled spatially to match the crop grid cells in the HILDA+ using 3×3 moving window with nearest neighbour modal method, which is adapted to produce the crop calendar dataset by Jägermeyr et al. (2021).

## 6. Grassland management datasets

In ORCHIDEE, similar to that for cropland, there is a separate branch for grassland, ORCHIDEE-GM (Grassland Management, Chang et al., 2013), that describes the impact of two grassland management practices (cutting and grazing) on grassland ecosystem dynamics and in particular on the exchange of carbon and water with the atmosphere. These developments were inspired (and partly taken) by a grassland model (PaSim, version 5.0).



This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101081395

Current efforts are ongoing to integrate the grassland branch back into the main version of ORCHIDEE that includes the nitrogen cycle. In the first set of simulations, we will thus not use any grassland management datasets, while for the next round, the GM module will be activated. For that module, we need the livestock distribution in Europe as livestock feeding and bedding needs are calculated within each grid cell from livestock density distribution maps, for different livestock categories. The distribution of each livestock category will be taken from the Gridded Livestock of the World dataset (GLW2; Robinson et al., 2014). This dataset can be accessed from the FAO website: https://www.fao.org/livestock-systems/en/. As an illustration, Figure 12 provides the distribution of cattle over the world, while 8 categories of livestock are reported in this dataset (Goats, Ducks, Buffaloes, Sheeps, Horses, Cattle, Pigs, and Chickens). From the livestock density in each grid cell, the grassland but also the cropland NPP will be used to provide the feeding and bedding needs (see Beaudor et al., 2022). More information and details on the exact use of these livestock densities will be provided in the update deliverable for the input datasets.

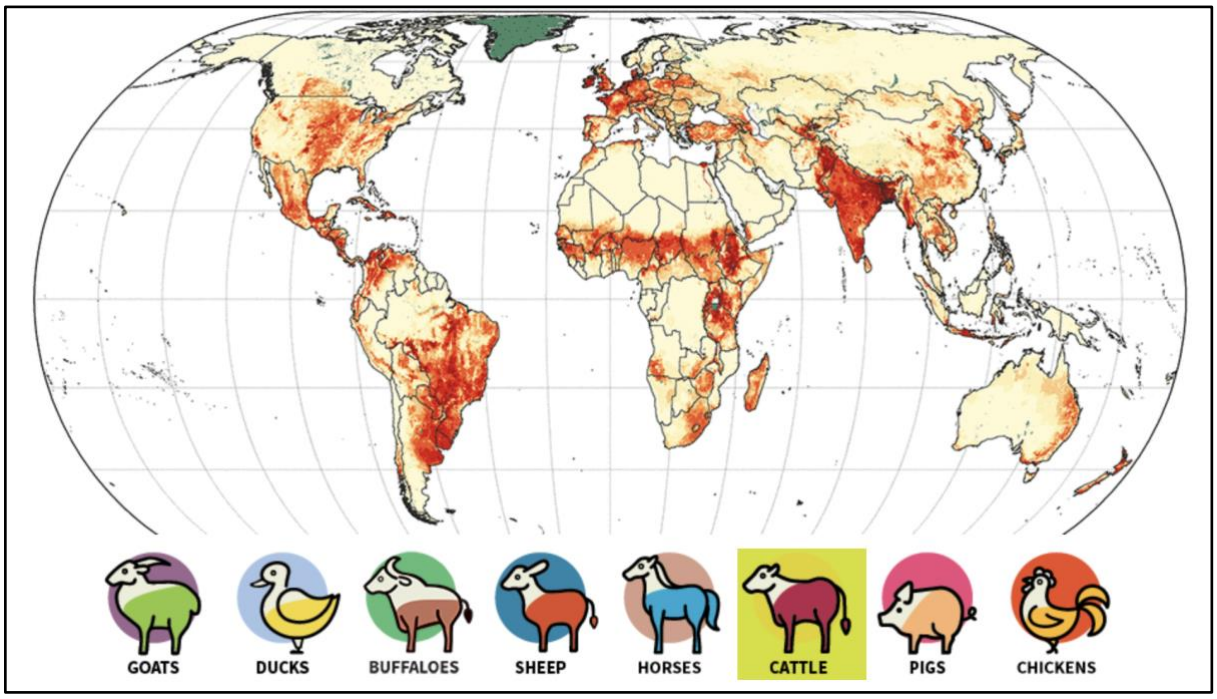


***Figure 12:*** *Illustration of the Cattle global distribution from the FAO website:* https://www.fao.org/livestock-systems/en/.

In LPJ-GUESS, we currently assume a fixed amount of biomass removed each year from pasture. Updates will be made in collaboration with ORCHIDEE to derive common gridded data on which grasslands are grazed based on livestock densities and types. Note that the pasture can also be fertilised in LPJ-GUESS. For JSBACH, there is currently no specific need of grassland management data for the simulation of wetland $CH_4$ emissions.



This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101081395

# 7. Forest management datasets

For forests, ORCHIDEE and LPJ-GUESS are at different stages in terms of using forest management data. For ORCHIDEE, the first set of simulations planned for early 2024, will not use the forest management module (diameter and age classes as well as tree height dynamics, described in Naudts et al., 2015) that has been only recently included in the Trunk version of ORCHIDEE and coupled with the nitrogen cycle, given that some parameterisations are still under adjustment. For LPJ-GUESS, we have started to reproduce present-day forest age products (such as the one described in Pucher et al. 2022). This is in a testing phase, as clear-cuts seen in the HILDA+ land cover dataset may not necessarily provide the age derived in the dataset by Pucher et al. (2022). Note also that forest management in Europe significantly varies between countries and its description is thus difficult to parameterise given that many countries do not have clear-felling but rather complex thinning practices. In this context, we have gathered different datasets that are currently used for model calibration and evaluation or that will be used in the next round of simulation directly as inputs. First, we are trying to valorize data that were produced and collected during the VERIFY precursor project through a dedicated activity on forest data collection. Second, we are currently trying to use the spatial products, combining in situ and remote sensing observations, that are described in Pucher et al. (2022).

## 7.1 Datasets assembled in the VERIFY project

We choose to valorise the data collected during the previous EU project, VERIFY, that had also a specific focus on European greenhouse gas budgets. The data have so far not been transferred to the EYE-CLIMA dataset but they are rather directly used from the VERIFY database (https://webportals.ipsl.fr/VERIFY/Ressources_2D.html). For a detailed description of the different datasets, we refer to specific VERIFY deliverable that are accessible from: https://verify.lsce.ipsl.fr/index.php/repository/public-deliverables/wp3-verification-methods-for-terrestrial-co2-sources-and-sinks-and-carbon-stock. We thus provide below a brief note on these products while more information will be given in the next release of this deliverable based on the final strategy used in each model.

### In-situ National Forest Inventory (NFI) data

More and more countries in Europe are setting up NFI programs, where especially Central/Eastern European countries are changing from inventories based on Forest Management Planning systems towards statistically based NFI programs. However, NFI data are considered politically sensitive, and in-situ plot data are not always shared. Gradually, this attitude is changing with more and more countries publishing their raw data on the internet or making them available on request for specific purposes and projects. Figure 13 illustrates the location of the different NFI data that were collected; there is a clear gradient of decreasing availability of raw NFI data from West to East in Europe. The collected raw data are thus available to both ORCHIDEE and LPJ-GUESS teams to evaluate, for example, the simulated height, diameter, and volume increments.



This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101081395


***Figure 13:*** *Availability of raw NFI data over Europe gathered in Verify project (see* https://verify.lsce.ipsl.fr/index.php/repository/public-deliverables/wp3-verification-methods-for-terrestrial-co2-sources-and-sinks-and-carbon-stock/d3-14-national-forest-inventory-and-high-resolution-forest-cover-for-eastern-europe*).*

## Observed management and mortality data

Observed mortality due to management or natural causes based on NFI data is provided in Schelhaas et al. (2018a,b). The most important species are treated individually, while the remainder are combined into one group. Data are presented by 5 cm classes, separately for management (HarvestProbability) and natural causes (DeadProbability) (see Figure 14). Note that harvested trees will include trees that died for natural reasons and that were subsequently extracted. These data are being used by the modelling groups to define spatially explicit forest management intensity.

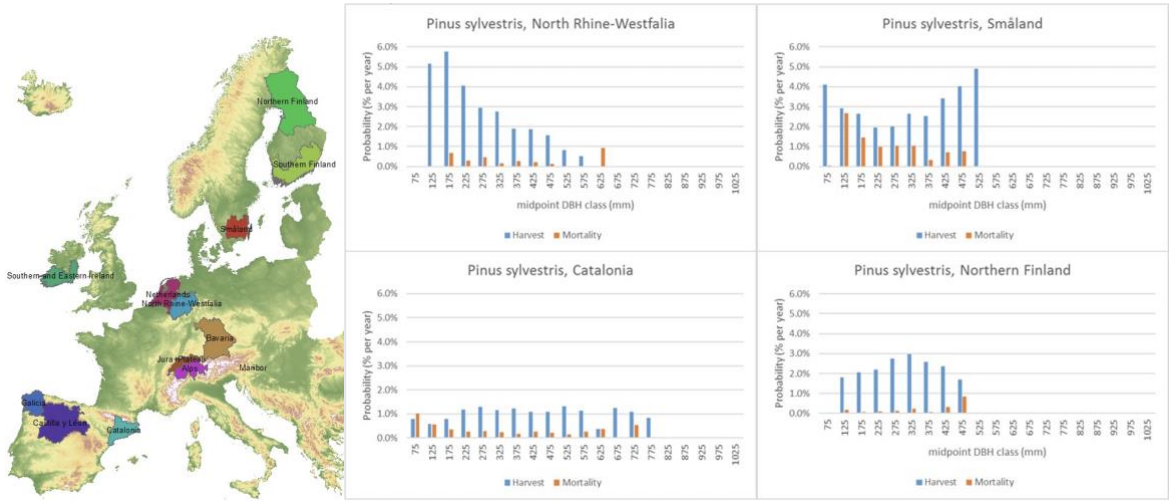


***Figure 14:*** *Left: regions for which management and mortality data are provided; Right: Example of regionally different patterns of management and mortality in Pinus sylvestris (Schelhaas et al. 2018b)*



This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101081395

### Remote sensing based biomass products

Estimates of forest above-ground biomass and carbon changes (ΔAGB/C) using remote sensing have progressed due to the increasing demand from the scientific community and countries, in line with the increasing availability of satellite data. Recently, several methods and maps of forest carbon fluxes were published using the "spatial approach" of mapping ΔAGB/C in multiple periods. Harris et al., 2021 mapped carbon fluxes using forest "gain-loss" pixels (Hansen et al., 2013), where the gained or lost carbon is estimated separately depending on the nature of changes from the IPCC activity data. Other recent methods used spaceborne radar and LiDAR data to map AGB/C in multiple periods (ESA-CCI product, Santoro and Cartus, 2021) and in time series (JPL product, Xu et al., 2021).

The data from these studies were used to assess carbon fluxes for Europe from 2010 to 2018 during the VERIFY project. All products were adjusted for bias using independent reference biomass datasets following the uncertainty assessment framework in Araza et al., 2022. In total, six map-based estimates of forest carbon fluxes were derived that were also compared with the estimates from the Forest Resource Assessment (FRA). FRA data are available for the years 2010, 2015, and 2020 so they averaged 2015 and 2020 to obtain a 2018 proxy.

More details on these products are available in the report: https://verify.lsce.ipsl.fr/index.php/repository/public-deliverables/wp3-verification-methods-for-terrestrial-co2-sources-and-sinks-and-carbon-stock/d3-14-national-forest-inventory-and-high-resolution-forest-cover-for-eastern-europe. These data are currently being used for the evaluation of the ORCHIDEE model outputs.

## 7.2 Pucher dataset

Pucher et al. (2022) released a recent dataset to provide an improved forest structure for Europe. Harmonized inventory data from 16 European countries were used in combination with remote sensing data and a gap-filling algorithm to produce a consistent and comparable forest structure dataset across European forests. They showed how land cover data can be used to scale inventory data to a higher resolution, which in turn ensures a consistent data structure across sub-national, country and European forest assessments. Cross-validation and comparison with published country statistics of the Food and Agriculture Organization (FAO) indicated that their methodology was able to produce robust and accurate forest structure data across Europe, even for areas where no inventory data were available.

Such dataset is available from the BOKU university in Vienna: https://boku.ac.at/en/wabo/waldbau/wir-ueber-uns/daten and will be used separately in the LPJ-GUESS and ORCHIDEE teams. The dataset comprises several variables of interest for both groups, including volume, carbon content, biomass by compartment, height, diameter at breast height, stem number, basal area, stand density index, age class and tree species group. Figure 15 illustrates the two main variables (age class and tree height) that we are currently using to constraint our model. Indeed, one key challenge is to have some coherence between the land cover change/use products, the management rule implemented in the models and the age specified in the Pucher dataset. As this is mostly work in progress, more information will be provided on the use of the Pucher data in the next release of this deliverable.



This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101081395

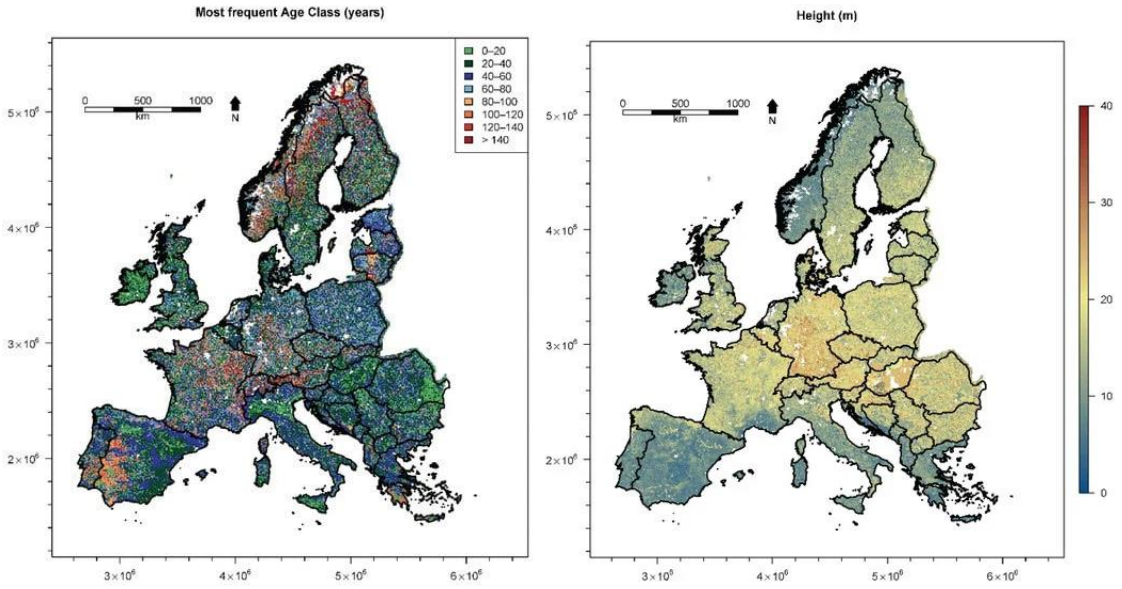


***Figure 15:*** *Illustration of the most frequent age class (left map; in years) and the tree height (right map, in metres) from Pucher et al., (2022) dataset. The data are gridded on an 8 × 8 km cell. Note that within each cell no distinction between forested or non-forested area is made; and a forest area mask is needed to quantify extent of forests.*

## 8. Conclusion

This deliverable presented key datasets that are either currently being used as inputs for the on-going model simulations of $CO_2$, $CH_4$ and $N_2O$ fluxes or used as validation and calibration for the development of the three models involved in EYE-CLIMA: ORCHIDEE, LPJ-GUESS and JSBACH. The EYE-CLIMA model input datasets deliverable is a living document, growing out of a collaboration between the three modelling groups. As such, the collection of data and this document will continue to be updated as user needs evolve. A second release of this deliverable is thus planned at month M20. Note that some of the data-streams described above have not yet been fully taken advantage of, but this is the current focus of some modelling groups and they will thus be fully valorised for the next round of process-based model simulations.

## 9. References

Ballabio C., Panagos P., Montanarella L. Mapping topsoil physical properties at European scale using the LUCAS database (2016) Geoderma, 261 , pp. 110-123.

Beaudor, M., Vuichard, N., Lathière, J., Evangeliou, N., Van Damme, M., Clarisse, L., & Hauglustaine, D. (2022). Global agricultural ammonia emissions simulated with the ORCHIDEE land surface model. *EGUsphere*, *2022*, 1-44.

Chang, J.F., Viovy, N., Vuichard, N., Ciais, P., Wang, T., Cozic, A., Lardy, R., Graux, A.I., Klumpp, K., Martin, R. and Soussana, J.F., 2013. Incorporating grassland management in ORCHIDEE: model description and evaluation at 11 eddy-covariance sites in Europe. *Geoscientific Model Development*, *6*(6), pp.2165-2181.



This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101081395

FAOSTAT. (2023). Production/Crops and livestock products [dataset]. Retrieved from https://www.fao.org/faostat/en/#data

Fluet-Chouinard, E., Stocker, B.D., Zhang, Z. *et al.* Extensive global wetland loss over the past three centuries. *Nature* **614**, 281–286 (2023). https://doi.org/10.1038/s41586-022-05572-6

Hagemann, S., Stacke, T., 2015. Impact of the soil hydrology scheme on simulated soil moisture memory. Clim Dyn 44, 1731–1750. https://doi.org/10.1007/s00382-014-2221-6

Harris, N. L., Gibbs, D. A., Baccini, A., Birdsey, R. A., De Bruin, S., Farina, M., ... & Tyukavina, A. (2021). Global maps of twenty-first century forest carbon fluxes. Nature Climate Change, 11(3), 234-240.

Hansen, M. C., Potapov, P. V., Moore, R., Hancher, M., Turubanova, S. A., Tyukavina, A., ... & Townshend, J. (2013). High-resolution global maps of 21st-century forest cover change. *science*, *342*(6160), 850-853.

Jägermeyr, J., Müller, C., Ruane, A. C., Elliott, J., Balkovic, J., Castillo, O., et al. (2021). Climate impacts on global agriculture emerge earlier in new generation of climate and crop models. Nature Food, 2(11), 873–885. https://doi.org/10.1038/s43016-021-00400-y

Lehner, B., Döll, P.: Development and validation of a global database of lakes, reservoirs and wetlands, Journal of Hydrology, Volume 296, Issues 1–4, 20 August 2004, Pages 1-22, http://dx.doi.org/10.1016/j.jhydrol.2004.03.028.

Naudts, K., Ryder, J., McGrath, M. J., Otto, J., Chen, Y., Valade, A., ... & Luyssaert, S. (2015). A vertically discretised canopy description for ORCHIDEE (SVN r2290) and the modifications to the energy, water and carbon fluxes. *Geoscientific Model Development*, *8*(7), 2035-2065.

Poggio, L., De Sousa, L. M., Batjes, N. H., Heuvelink, G., Kempen, B., Ribeiro, E., & Rossiter, D. (2021). SoilGrids 2.0: producing soil information for the globe with quantified spatial uncertainty. *Soil*, *7*(1), 217-240.

Portmann, F. T., Siebert, S., & Döll, P. (2010). MIRCA2000-Global monthly irrigated and rainfed crop areas around the year 2000: A new high-resolution dataset for agricultural and hydrological modeling. Global Biogeochemical Cycles, 24(1), 1–24. https://doi.org/10.1029/2008gb003435

Pucher, C., Neumann, M., & Hasenauer, H. (2022). An Improved Forest Structure dataset for Europe. *Remote Sensing*, *14*(2), 395.

Ribeiro, E., & Batjes, N. (2019, January). WoSIS: standardised soil profile data for digital soil mapping at global scale. In *Geophysical Research Abstracts* (Vol. 21).

Robinson TP, Wint GRW, Conchedda G, Van Boeckel TP, Ercoli V, Palamara E, Cinardi G, DâAietti L, Hay SI, and Gilbert M. (2014) Mapping the Global Distribution of Livestock. PLoS ONE 9(5): e96084. doi:10.1371/journal.pone.0096084

Santoro, M., Cartus, O., Carvalhais, N., Rozendaal, D., Avitabile, V., Araza, A., ... & Willcock, S. (2021). The global forest above-ground biomass pool for 2010 estimated from high-resolution satellite observations. Earth System Science Data, 13(8), 3927-3950.

Santoro, M., & Cartus, O. (2021). ESA biomass climate change initiative (Biomass_cci): Global datasets of forest above-ground biomass for the years 2010, 2017 and 2018, v3. Cent. Environ. Data Anal.

Schelhaas MJ, Hengeveld GM, Heidema N, Thürig E, Rohner B, Vacchiano G, Vayreda J, Redmond J, Socha J, Fridman J, Tomter S, Polley H, Barreiro S, Nabuurs GJ, 2018a. Species-specific, pan-European diameter increment models based on data of 2.3 million trees. Forest Ecosystems 5:21 doi.org/10.1186/s40663018-0133-3

Schelhaas MJ, Fridman J, Hengeveld GM, Henttonen H, Lehtonen A, Kies U, Krajnc N, Lerink B, Ní Dhubháin A, Polley H, Pugh TAM, Redmond J, Rohner B, Temperli C, Vayreda J, Nabuurs GJ, 2018b.



This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101081395

Actual European forest management by region, tree species and owner based on 714,000 re-measured trees in national forest inventories. PLoS ONE 13(11): e0207151.

Tanneberger, F., Tegetmeyer, C., Busse, S., Barthelmes, A., and 55 others, 2017. The peatland map of Europe. Mires and Peat 1–17. https://doi.org/10.19189/MaP.2016.OMB.264

Tian, H., Bian, Z., Shi, H., Qin, X., Pan, N., Lu, C., et al. (2022). History of anthropogenic Nitrogen inputs (HaNi) to the terrestrial biosphere: a 5 arcmin resolution annual dataset from 1860 to 2019. Earth System Science Data, 14(10), 4551–4568. https://doi.org/10.5194/essd-14-4551-2022

Winkler, K., Fuchs, R., Rounsevell, M. D., & Herold, M. (2020). HILDA+ Global Land Use Change between 1960 and 2019. *Pangaea https://doi. org/10.1594/PANGAEA*, *921846*.

Wu, X., Vuichard, N., Ciais, P., Viovy, N., de Noblet-Ducoudré, N., Wang, X., ... & Ripoche, D. (2016). ORCHIDEE-CROP (v0), a new process-based agro-land surface model: model description and evaluation over Europe. *Geoscientific Model Development*, *9*(2), 857-873.

Xu, J., Morris, P.J., Liu, J. and Holden, J. (2018). PEATMAP: Refining estimates of global peatland distribution based on a meta-analysis, CATENA, 160, 134-140, https://doi.org/10.1016/j.catena.2017.09.010

Xu, L., Saatchi, S. S., Yang, Y., Yu, Y., Pongratz, J., Bloom, A. A., ... & Schimel, D. (2021). Changes in global terrestrial live biomass over the 21st century. Science Advances, 7(27), eabe9829.

Zhang, Z., Fluet-Chouinard, E., Jensen, K., McDonald, K., Hugelius, G., Gumbricht, T., Carroll, M., Prigent, C., Bartsch, A., and Poulter, B.: Development of the global dataset of Wetland Area and Dynamics for Methane Modeling (WAD2M) , Earth Syst. Sci. Data, 13, 2001–2023, https://doi.org/10.5194/essd-13-2001-2021, 2021.



This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101081395

https://eyeclima.eu

BRUSSELS, 19 02 2024

*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. Neither the European Union nor the granting authority can be held responsible for them.*

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101081395