D3.4
Farmland maps for demonstration areas

Deliverable for the Horizon Europe Project BirdWatch

Version 1.0

Funded by
the European Union

Legal Disclaimer

This document reflects only the views of the author(s). Neither the European Global Navigation Satellite Systems Agency (EUSPA) nor the European Commission is in any way responsible for any use that may be made of the information it contains. The information in this document is provided "as is", and no guarantee or warranty is given that the information is fit for any particular purpose. The below referenced consortium members shall have no liability for damages of any kind including without limitation direct, special, indirect, or consequential damages that may result from the use of these materials subject to any liability which is mandatory due to applicable law.
This document and the information contained within may not be copied, used, or disclosed, entirely or partially, outside of the BirdWatch consortium without prior permission of the project partners in written form.

Funded by
the European Union

Document Information

GA Number	101082634		**Type of Action**	Horizon-IA
Full Title	BirdWatch - a Copernicus-based service for the improvement of habitat suitability of farmland birds via satellite-enabled monitoring, evaluation and optimisation of CAP greening measures			
Project Acronym	BirdWatch			
Start Date	February 1st, 2023		**Duration**	36 Months
Project URL	https://birdwatch-europe.org/			
Deliverable	D3.4: Farmland maps for demonstration areas			
Work Package	WP3			
Project Month of Delivery	**Contractual**	M24	**Actual**	M24
Nature	DATA	**Dissemination Level**		PU
Lead Beneficiary	SIN			
Responsible Author	Nika Oman Kadunc – SINERGISE			
Contributions from	Nika Oman Kadunc – SINERGISE Nejc Vesel – SINERGISE Bartolomeo Ventura – EURAC Alexander Jacob – EURAC			

History of Changes

Version	Issue Date	Stage	Description	Comments	Contributor
1.0	31.01.2025		Initial version of D3.4		Nika Oman Kadunc Bartolomeo Ventura Nejc Vesel Alexander Jacob
2.0	31.03.2025		Updated version according to DRS		Nika Oman Kadunc Bartolomeo Ventura Nejc Vesel Alexander Jacob

Table of Contents

1 Introduction

This document describes in detail the geospatial datasets that were prepared and for the BirdWatch project four demonstration regions: the region of Flanders (Belgium), Germany, Lithuania, and the region of South Tyrol (Italy).

In the first stage of the BirdWatch project, the initial dataset selection was made based on the information on the remote sensing parameters needed for the bird habitat assessments and Bird Species Requirements (listed and described Section 2.3 of this document) were taken into account. The datasets were prepared as input for the Species Distribution Model (SDM) and specifications of the model inputs were considered in the preparation of the data layers. The basic technologies used for the vector and raster database structure and data access were set according to SDM input data requirements (described in Section 2.4) and in accordance with the consortium partners managing different phases (WPs) of the BirdWatch project that need the data as input for the activities of their respective WPs.

In Section 3, the dataset acquisition, calculations and postprocessing procedures for the BirdWatch project data layers is described in detail. First, the preparation of the Crop type layers is described in Section 3.1, next, in Section 3.2, the calculation and the results of the Soil moisture data layers are presented. The calculation and postprocessing of the Mowing data layers are described in Section 3.3. The Land cover and the DEM data layers are presented in Sections 3.4 and 4.5 respectively.

Some resulting data layers are visually presented in the respective sub-sections of Section 3, whereas the access to the full dataset for all the produced data layers for all four BirdWatch test regions is explained and given in Section 4 of this document.

All in all, around 30 data layers for each BirdWatch test region (the number of Soil moisture data layers is dependent on the region and ranges from 2 to 6) were prepared and provided as input for the SDM. The data requirements, the preparation procedures, the technologies utilised for the preparation, the calculation and postprocessing methods and results are presented in the following sections of this document.

2 BirdWatch Data Requirements

The BirdWatch datasets were prepared according to the habitat requirements collected for each bird species, along with the EO data and EO data-based measures identified in the activities of WP 2000. The definition of the data requirements and the dataset selection process began early in the project to support all activities of the BirdWatch partners in WP2000. The process involved gathering all available information on parameters relevant to bird habitat assessments for the selected bird species (Bird Species Requirements) and the requirements for the Species Distribution Model (SDM), as well as checking for availability of the data for each of the four BirdWatch regions. Detailed descriptions of the requirements are provided in Sections 2.2 and 2.3 of this document.

In Section 0, we first present the four BirdWatch regions for dataset preparation, and Section 0 outlines the data layers chosen for running the SDM.

To prepare the data into a format suitable for the SDM, the information from the relevant datasets described in this document were be rasterized and combined into a raster stack that was aggregated to EuroStat Grid 200mx200m cells. The raster stack was made available to the species distribution modelling component of WP4000. The overview of the data preparation workflow is depicted in Figure 1.

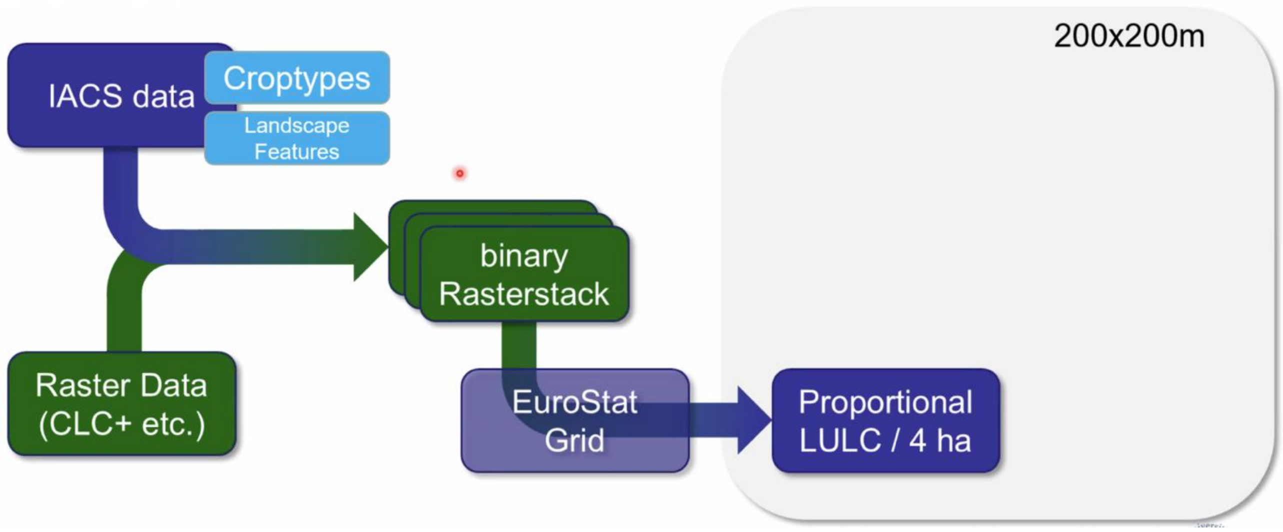


Figure 1: Schematic of the workflow of data preparation in the BirdWatch project.

2.1 BirdWatch test regions Area Of Interest (AOI) and Time Of Interest (TOI)

The four BirdWatch test regions are region of Flanders (Belgium), Germany (whole country), Lithuania (whole country) and region of South Tyrol in Italy. The regions are visually presented in Figure 2.

Figure 2: BirdWatch test regions: Flanders (Belgium), Germany, Lithuania and South Tyrol (Italy)

As the first preparation step, the AOI (Area Of Interest) shape files were gathered for all regions, as well as 200 m x 200 m grid shape files needed for subsequent aggregation for the data to be suitable as input to the SDM model.

The TOIs (Time Of Interest) for the BirdWatch test regions were obtained from the partners of WP4000 upon checking the bird data occurrence availability. Table 1 presents the BirdWatch four test regions together with their respective areas and the year (TOI) for the calculation of data layers.

Table 1: List of BirdWatch test regions with their respective areas

BirdWatch test region	AREA (km^2)	TOI (year)
Flanders	13 678	2018
Germany	353 119	2022
Lithuania	64 849	2022
South Tyrol	13 606	2022

2.2 Species Distribution Model data requirements

The BirdWatch data layers serves as input for the species distribution modelling (SDM), we took into consideration the SDM data requirements as provided by BirdWatch project partners of WP4000. The basic requirements were updated upon testing the initial datasets provided. The initial requirements as provided by partners of WP400, together with additional constraints for the data layers are listed below.

1. General data requirements:

- Continuous raster data are needed for SDMs. For example, early or late mowing as a category is not useful but the day or month in a year when mowing took place could be useful.
- How many parameters can be included in the model depends on the number of occurrence points per species (as a rule of thumb: 10 occurrence points are needed for 1 parameter). We will also have to test for correlation between the parameters. When parameters are correlated the habitat requirement list can help to identify the one parameter out of the correlated ones that will be included in the model. The other correlated parameters will not be included. Statistical outputs could also be used to identify which parameter out of the correlated ones will be included.
- Each environmental data raster layer is aggregated to the same 200 m x 200m grid to be ingested and used by the SDM.
- The SDM model does not work with NA (Not Available) data points. An appropriate value must be chosen to replace NA data values in datasets (this requirement applies to the mowing layers where the mowing was calculated on grassland only).

2. Temporal proximity of environmental data (e.g. EO-based data):

- In general, the most recent bird occurrence data are the most relevant
- If one year has not enough bird occurrence data, data from previous years can be included as well
- The environmental data should be derived from the same time period as the bird occurrence data. The time periods (respective year) for each BirdWatch test region were provided by partners of WP4000 upon reviewing the bird occurrence data availability.

3. Parameter selection for the final habitat models:

- For a lot of parameters, we do not know yet if they will explain the occurrence data well or not. Here, the model output will help. For these cases, the models will be run with each category setup to see which categories explain the occurrence data better.

2.3 Selected EO-parameters – BirdWatch data layers

The selection of parameters and data layers was based on the bird species requirements gathered and identified in the activities of WP 2000. Table 2 lists the types of habitat requirements collected for each bird species, along with the EO data and EO data-based measures to derive the presence or absence of a habitat requirement.

Table 2: Types of habitat requirements collected for each bird species, along with the EO data and EO data-based measures to derive the presence or absence of a habitat requirement

Requirement type	Requirement expressions	EO and EO-based data	Examples for EO-based measures
Structural habitat characteristics	Open (e.g, wide, open fields) Semi-open (e.g., open fields with some vertical structures or landscape elements) Dense (dense growth of e.g., shrubs, trees or bushes) Mosaicked (mix of open and dense structures)	Sentinel 1 and -2 Readily available land cover classifications (e.g., Dynamic World)	Percentage land cover type per habitat Proportional extent of land cover type per habitat Texture measures (e.g., grey-level occurrence) Radar backscatter
Landscape elements	Trees, high shrubberies Low shrubberies, bushes, hedges Water bodies and water edge vegetation Man-made structures	Sentinel 1 and -2 Readily available land cover classifications Digital terrain models & digital elevation models	Percentage cover per land cover type Spectral & radar vegetation indices Radar backscatter Difference calculations
Vegetation types	Crop types Grassland types	Sentinel 1 and -2 Readily available land cover and crop type classifications (e.g., EuroCrops)	Percentage land cover type per habitat Spectral & radar vegetation indices Texture measures (e.g., grey-level occurrence) Time series metrics

Based on the Bird habitat requirements per species (described in the BirdWatch Deliverable D2.2), a summary of the parameters was made and the parameters with most occurrences were chosen for the database preparation for the first iteration of the Species distribution model (SDM) testing. The summary of the habitat requirements was presented in BirdWatch deliverable D3.1 (D3.1, Table 4).

The list of chosen data layers (parameters) grouped according to availability in different dataset types (Landscape features, Crop type, Land cover and Remote Sensing (RS) derived data) is presented in Table 3. Similar crop types were grouped together into subgroups according to relevance for the bird habitat requirements to reduce the number of data layers. This contributes to the efficiency of the SDM as well as facilitates the ingestion of data into the database.

For each of the parameters, a dataset containing the information was either selected from the list of available datasets (the list was provided in BirdWatch deliverable D3.1, Chapter 2.1) or calculated from EO-data. The chosen data layers listed in Table 3 were provided for all four BirdWatch test areas, each for the year where the most bird data occurrences were identified by partners of WP4000. The acquisition, processing or calculations made for each of the layers / layer groups are further described in detail in the Section 0 of this document.

Table 3: List of BirdWatch EO parameters (EO data layers)

Parameter group	Source	Layer	Aggregation metric
Landscape features	**LPIS / Thuenen Croptypes (Germany)**	shrubs / swf (small woody features)	% area coverage within cell
Crop type	**LPIS / Thuenen Croptypes (Germany)**	Cultivated Grassland	% area coverage within cell
		Summer/Spring Cereals	
		Other Cereals	
		Maize	
		Root Crops (sugar beet, potatoes etc.	
		Vegetables (salad, cabbage etc.)	
		Rapeseed	
		Legumes (soy, peas, beans etc.)	
		Permanent Grassland	
		Winter Cereals	
		Fallow	
		Orchards and Berries	
		Grapevine	
		Hops	
		Sunflower	
		Forest Trees / SRF	
		Miscellaneous	
Land cover	**Impact Observatory - 10m Annual Land Use Land Cover (LULC) map**	Forest	% area coverage within cell
		Water	% area coverage within cell
		Urban Areas / Streets	% area coverage within cell
Mowing	**Sinergise Mowing marker / Thuenen Mowing Raster (Germany)**	Mowing_counts	mean
		First mowing doy	minimum
Soil moisture	**EURAC SMC calculation**	SMC-early-breeding-season	
		SMC-full-breeding-season	
Topography	**Copernicus DEM**	Elevation	mean - elevation in metres
		Slope Angle	mean - slope in degrees

Funded by
the European Union

2.4 Data acquisition and processing tools

For the acquisition and the preparation of the different BirdWatch data layers, different technologies and tools were utilised to facilitate the workflows and data processing. The different tools available are listed and described in detail in BirdWatch deliverables D3.1 and D3.2, in this section we will briefly highlight the most important technologies used in data preparation and showcase their usage for the BirdWatch project.

2.4.1 Sentinel Hub, eo-learn and eo-grow

Whenever possible, Sentinel Hub APIs and services were used for data acquisition, storage and processing of the data. Figure 3 shows the Sentinel Hub custom script ("evalscript") for acquisition of the Copernicus DEM layer for BirdWatch project from Sentinel Hub, where the data collection is available.

DEM Download

```
config = SHConfig()
```

```
BUCKET_NAME = 'sinergise-birdwatch'
OUTPUT_PATH = "s3://sinergise-birdwatch/projects/Lithuania/dem/batch_output"
```

```
dem_evalscript = """
//VERSION=3

function setup() {
  return {
    input: [{
      bands: [
        "DEM", "dataMask"
      ]
    }],
    output: [
      {
        id: "dem",
        bands: ["dem"],
        sampleType: SampleType.FLOAT32
      },
      {
        id: "dataMask",
        bands: 1
      }
    ],
  }
}

function evaluatePixel(samples) {

  return {dem: [samples.DEM],
          dataMask: [samples.dataMask]
          }
}

"""
```

Figure 3: example of a Sentinel Hub evalscript for download of the Copernicus DEM data layer

Libraries eo-learn and eo-grow were used in processing of the data. Figure 4 shows the code examples for BirdWatch Soil moisture content (SMC) layers conversion to Cloud Optimized GeoTIFF (COG) using eo-grow functionalities to be stored as BYOC as "smc-collection" on Sentinel Hub and processed through Sentinel Hub Batch API.

```python
import os
import sys
sys.path.append(os.path.abspath(".."))

import subprocess

import numpy as np
import s3fs
import tqdm.auto as tqdm
from eogrow.utils.map import cogify
from utils import check_min_max, multiprocess, parse_custom_name
```

```python
bucket = "sinergise-birdwatch"

raw_path = f"byoc/soil_moisture_content/raw"
cog_path = f"byoc/soil_moisture_content/cog"  # will continue in YYYY/MM/DD/...
out_root = "/home/ubuntu/temp"

if not os.path.exists(out_root):
    os.mkdir(out_root)

s3 = s3fs.core.S3FileSystem()  # profile=profile_name)
```

```python
tiles = [t for t in s3.glob(f"s3://{bucket}/{raw_path}/*.tif")]
tiles[-5:]
```

```python
parse_custom_name(tiles[0])
```

```python
# make batch request
batch = SentinelHubBatch(config=config)


sentinelhub_request = SentinelHubRequest(
    evalscript=smc_evalscript,
    input_data=[
        SentinelHubRequest.input_data(
            data_collection=smc_collection,
            time_interval=('2022-01-01', '2022-01-02'),
        )
    ],
    responses=[
        SentinelHubRequest.output_response("smc", MimeType.TIFF),
        SentinelHubRequest.output_response("dataMask", MimeType.TIFF),
    ],
    geometry=aoi_sh_geom,
    config=config,
)
```

Figure 4: example code from the BirdWatch SMC data layers preparation using of the eo-grow Python library, Sentinel Hub BYOC functionality and the Sentinel Hub Batch API

2.4.2 Postgres Database

Vector data collected and/or shared by the partners responsible have been organized in a common database (DB). The DB is a POSTGRES open-source object-relational database system. It is an open-source, highly stable database system that provides support to different functions of SQL, like foreign keys, subqueries, triggers, and different user-defined types and functions. Direct connection to the DB is possible using different clients (R, Python, PgAdmin, etc.) and/or by a basic, but powerful, web interface for data preview and queries. The connection will be possible for all project partners using specific credentials and connection parameters to be shared. An example of the shared credentials is shown below.

Host	birdwatch_host
Name	birdwatch (schema: hydrology)
Port	XXXX
Username	birdwatch_user
Password	write to bartolomeo.ventura@eurac.edu

The managed datasets consist of various information about different parameters such as trees, high shrubberies, waterbodies, etc., and they will belong to different areas (Flanders, South Tyrol, Germany, Lithuania). With all the permissions to access the DB set, a generic request can be formulated as follows:

```
hostname/TABLENAME/OPTIONALFILTERS
```

2.4.3 SFTP server

In the early phase of the project, data collected and/or shared by the partners responsible for each parameter have been collected in a sFTP server. This sFTP server has been created by EURAC and hosted by eodc. All the partners can access it using shared specific credentials.

3 BirdWatch datasets

In this section, the datasets listed in Section 2, that were prepared for the BirdWatch four test regions are presented in detail, with data sources, postprocessing methods and the metrics for the aggregation to the 200m x 200m grid required by the SDM. The sub-sections are organised according to parameter groups that require a similar workflow of data acquisition and postprocessing methodologies.

All the resulting data layers were provided in the GeoTIFF format. The GeoTIFF standard defines the Geographic Tagged Image File Format by specifying requirements and encoding rules for using the Tagged Image File Format (TIFF) for the exchange of georeferenced or geocoded imagery. The GeoTIFF format is used throughout the geospatial and earth science communities to share geographic image data.

3.1 Crop type datasets

The Crop type datasets require several steps to be prepared for ingestion of the SDM. The first important step is the Harmonization of the different regions' LPIS data sources that were acquired for the purpose of creating BirdWatch datasets.

The BirdWatch project test regions Vector datasets, namely LPIS were provided by the BirdWatch consortium partners. LPIS data were gathered for each test region and are presented in Table 4.

Table 4: List of LPIS vector datasets potentially relevant for bird habitat assessment for different test regions of the BirdWatch project

BirdWatch test Region	LPIS dataset	Link to download
Flanders	Landscape elements	https://landbouwcijfers.vlaanderen.be/open-geodata-kaart-landschapselementen
	Agricultural parcels	https://landbouwcijfers.vlaanderen.be/open-geodata-landbouwgebruikspercelen
Germany - Brandenburg	Agricultural parcels	https://geobroker.geobasis-bb.de/gbss.php?MODE=GetProductInformation&PRODUCTID=9e95f21f-4ecf-4682-9a44-e5f7609f6fa0
Lithuania	Land cover type layer, EFA Layer	https://www.geoportal.lt/metadata-catalog/catalog/search/resource/details.page?uuid=%7B05B2E0DB-31DC-7382-EBB7-7EC5531858A8%7D
South Tyrol	LISS (Land use layer) LAFIS (Agriculture parcels)	Not open-source data, thus shared internally through database with partners.

The dataset available for the South Tyrol area is called LAFIS (sistema informativo agricolo forestale provincia autonoma di Bolzano). It is periodically generated by the agriculture and forest services of

Funded by the European Union

the Province of Bolzano. In the absence of any LPIS dataset present for the region, it serves as a good substitute. The recent version of the data was generated in 2020 and before that, there is a version from 2015. The latest version includes 100 classes including various fruits, vegetables and types of pastures and meadows. Based on the requirements of the birds, features can be extracted from the LAFIS layer.

3.2 Harmonization of BirdWatch test regions' LPIS datasets

Harmonization of LPIS data is the process of integrating and standardizing LPIS datasets from different sources or regions to ensure consistency, compatibility, and interoperability.

We based the target crop type categories on the newly developed hierarchical crop and agriculture taxonomy (HCAT), which contains over 300 crop classes. By associating crop types with different farming practices which are relevant for farmland bird, we extracted 17 crop type categories that correspond to the bird species requirements (presented in Section 2, Table 3 of this document).

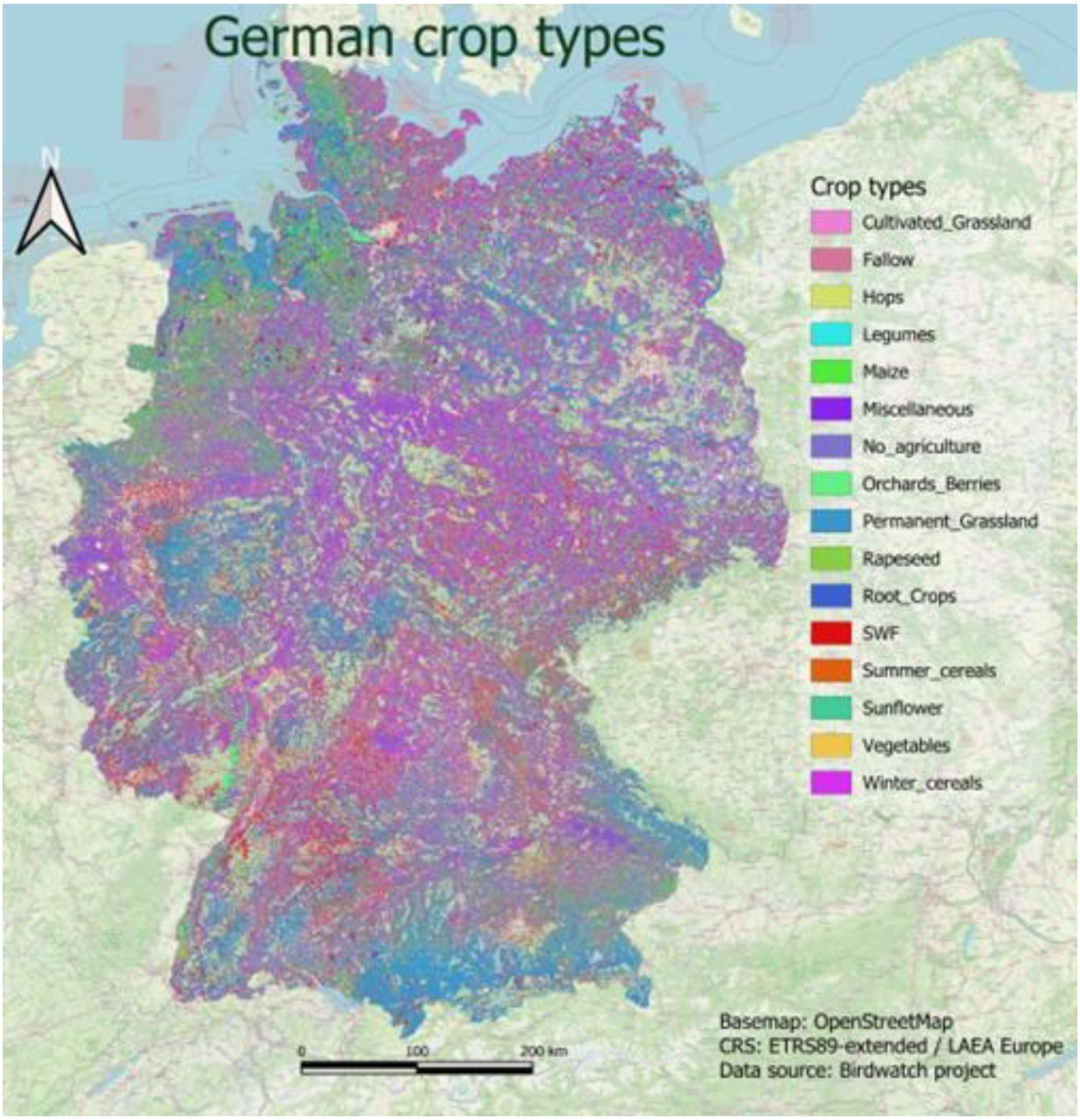
The subsequent harmonization of LPIS data involved several steps. The collected LPIS data was standardized to common (English) language, common data formats, coordinate reference systems (CRS), attribute and coding schemes. The most important step was attribute mapping to match similar attributes across different LPIS datasets to create a consistent set of attributes as defined in the identified crop types for BirdWatch.

A further harmonization step has been added to align the LAFIS classification from South Tyrol to the LPIS one. This has ensured that the 1 to 1 comparison can be conducted among the various classes belonging to the four test regions' AOIs.

For all the 17 classes of crops, the vector data was used to calculate the area of all polygons for each of the crop types within the 200 m x 200 m grid that was used to later convert to a raster file. Other vector data apart from the crops was Hedges which was also extracted from the LPIS data or other regional data sources for the regions of Flanders, Lithuania and South Tyrol.

In the absence of the full LPIS database for Germany, Thuenen Institute crop type data layer derived from time series of Sentinel-1, Sentinel-2 and Landsat 8 data, was used to extract relevant classes. Their methodology is based on a random forest classifier and is described in detail in their publication. The dataset was obtained from Zenodo by partners LUP.

In Figure 5, examples of all the extracted vector crop types for Flanders and Germany are shown. Examples of a detailed view of some crop types for the same areas are shown in **Fehler! Verweisquelle konnte nicht gefunden werden.**.


German crop types
N
Crop types
Cultivated_Grassland
Fallow
Hops
Legumes
Maize
Miscellaneous
No_agriculture
Orchards_Berries
Permanent_Grassland
Rapeseed
Root_Crops
SWF
Summer_cereals
Sunflower
Vegetables
Winter_cereals
0
100
200 km
Basemap: OpenStreetMap
CRS: ETRS89-extended / LAEA Europe
Data source: Birdwatch project

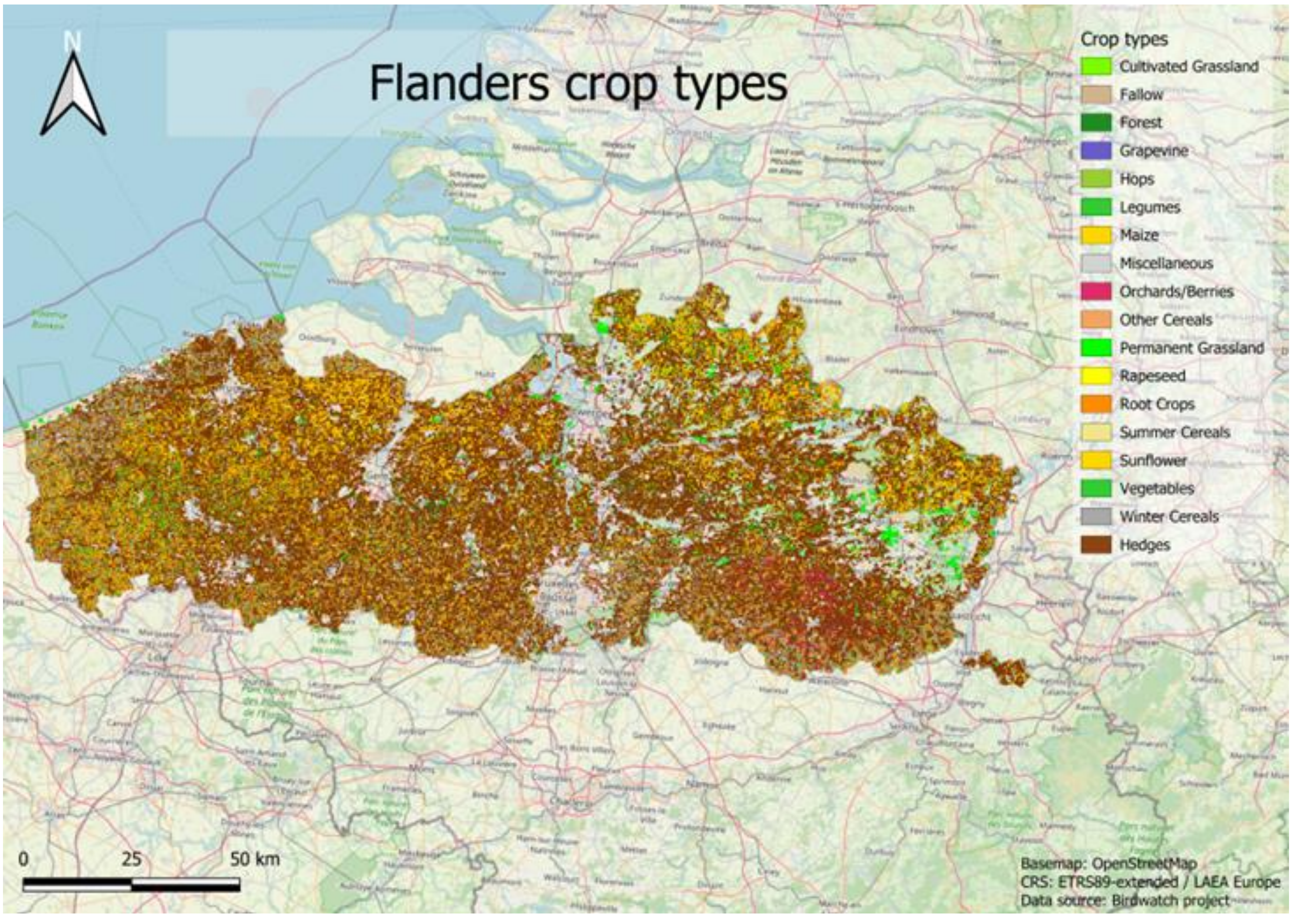
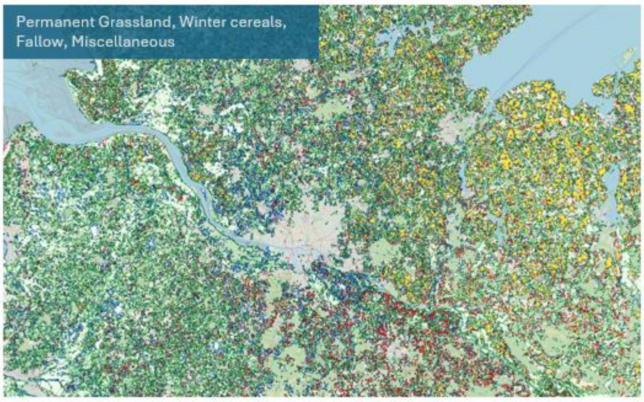
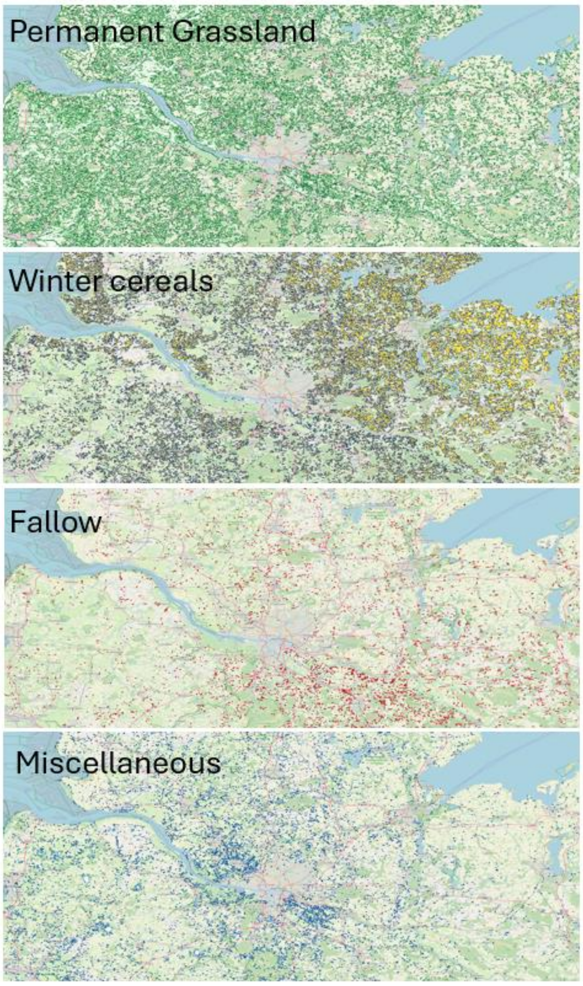


Figure 5 (this and previous page): Examples of all extracted crop types for BirdWatch test regions of Germany and Flanders and Germany.


Permanent Grassland, Winter cereals,
Fallow, Miscellaneous


Permanent Grassland
Winter cereals
Fallow
Miscellaneous

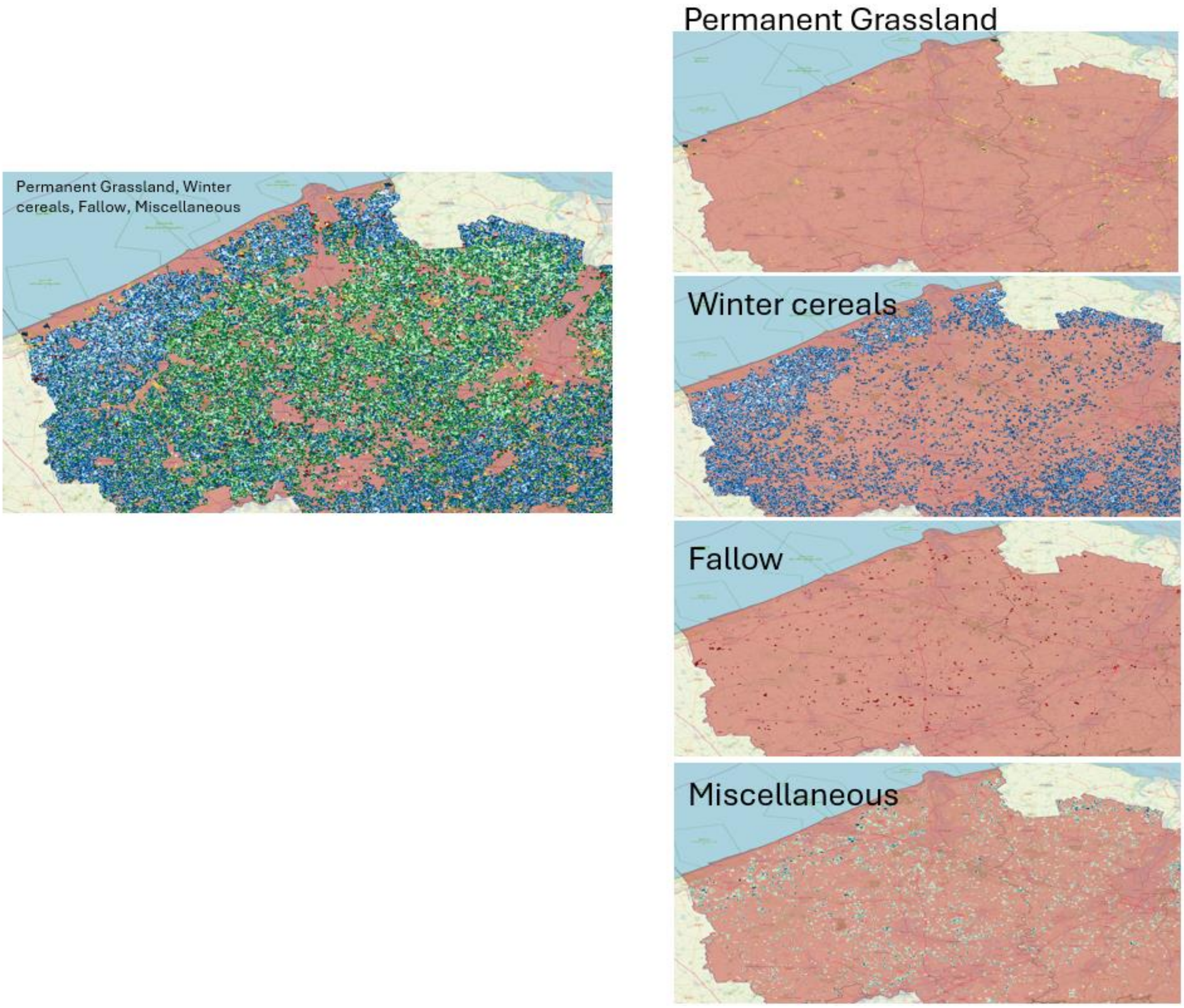


Figure 6 (this and previous page): A detailed view of some crop types for the same areas as in Figure 5: Germany and Flanders.

3.2.1 Quality checks of BirdWatch test regions' data layers

The LPIS data was chosen as the basis for crop type data layers as it is regarded as a very reliable geospatial dataset for mapping and analysing agricultural land use within the EU. Its reliability stems from the fact that it is based on high resolution orthophoto imagery and undergoes strict legal requirements placed on EU member states to maintain up-to-date and accurate information, which is verified through regular audits and quality checks under the LPIS Quality Assurance (QA) system: https://wikis.ec.europa.eu/spaces/GUIDANCEANDTOOLSFORCAP/pages/86968631/LPIS+QA+-+LPIS+Quality+Assurance+framework

The quality checks assess spatial accuracy, thematic correctness and completeness in order to standardize data quality across different countries.

With the preparation of the BirdWatch data layers, quality checks and validation of the processing steps were done for each layer on a sample of the data to validate and verify the harmonization of the crop type classes. A visual as well as statistical analysis of the data was made to check the correctness of the processing procedures. All layers were checked for consistency and correctness of the CRS (Coordinate Reference System) when aligning and aggregating to the 200 x 200m grid. For each layer, the area of the coverage of the resulting raster data layer was checked to match the area of the specific class in the source data to ensure the correctness of the extraction and aggregation procedure.

In the absence of LPIS data in the case of Germany, checks for consistency between different data layers were done by comparing the overlap of different crop type classes with the land cover layers (Forest / Trees, Built-up, Water). We expected a small level of inconsistency, because the different data sources with different resolutions and different definitions of classes, which might result in slight overlap of the pixels in source data and thus in some of the resulting aggregated data cells. The overlap was found to be small for a large majority of the cells, with the area exceeding 100 % by less than 1%. A larger overlap was found for 128 grid cells out of total 13.911.966 grid cells, which is less than 0.001% of all grid cells. For these 128 cells, we checked the overlapping classes and found overlap between “Miscellaneous” and / or “Hedges” classes and “Forest / Trees” land cover. To mitigate these overlaps, in discussion with partners of WP4000, it was decided that the cells will be normalised to 100% area within the WP4000 SDM workflow in a way that the ratio of the presence of classes will be retained.

3.3 Soil moisture layers

The estimation of high spatial resolution (up to 20 m) Surface Soil Moisture is based on a combination of SAR and optical imagery with a machine-learning approach (Greifender et al.) to download or preprocess the input datasets. The source code is available online at https://zenodo.org/records/4552813 (accessed on 23 March 2024), and the related documentation is available at https://pysmm.readthedocs.io/en/latest/ (accessed on 23 March 2024). This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 641762 (Website: http://www.ecopotential-project.eu/).

Based on different areas the breeding season varies as well as the periods vary for different bird species. The breeding season periods, for which the soil moisture content was temporally aggregated, were provided by BirdWatch project partners of WP2000.

The early breeding season SMC layers should include the first month of the breeding season and the prior month.

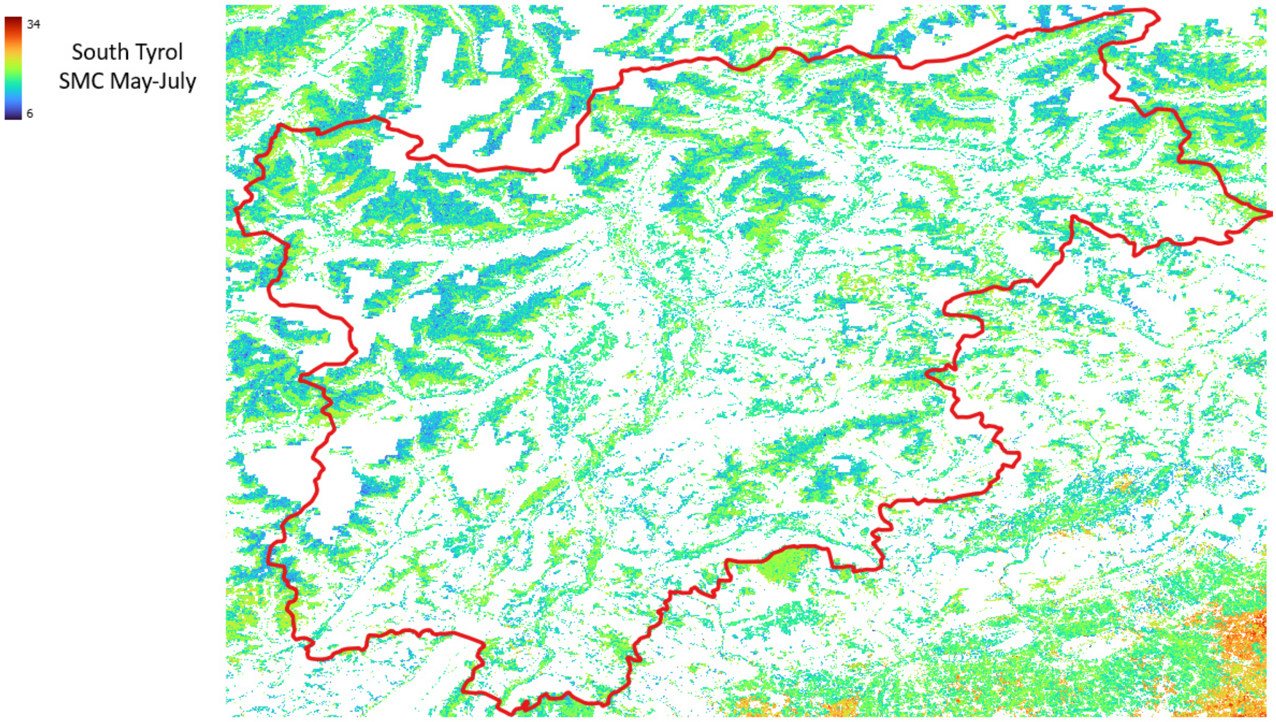
1. For Flanders this is for all birds: February - March
2. For Germany this is for all birds: February - March

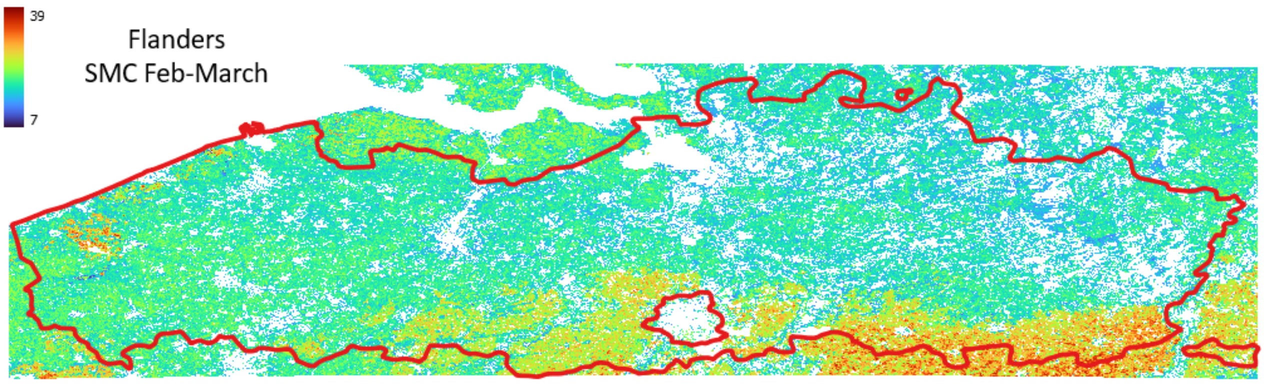
3. For South Tyrol this includes the periods:
    - March – April
    - April – May
    - May – June
4. For Lithuania this includes the periods:
    - February – March
    - March – April
    - April – May

The whole breeding season should include soil moisture over the entire breeding season.

1. For Flanders this is for all birds: March - June
2. For Germany this is for all birds: March - June
3. For South Tyrol this includes the periods:
    - April – July
    - June – July
    - May – July
4. For Lithuania this includes the periods:
    - March – June
    - April – June
    - April – July
    - May – July

Figure 7 shows examples of temporally aggregated whole breeding season SMC data layer for all the AOIs. The temporal aggregation was obtained calculating the average of the images in the relative time span. In the legend, you can find the interval corresponding to the values of the Soil Moisture Content (SMC).


34
6
South Tyrol
SMC May-July


39
7
Flanders
SMC Feb-March

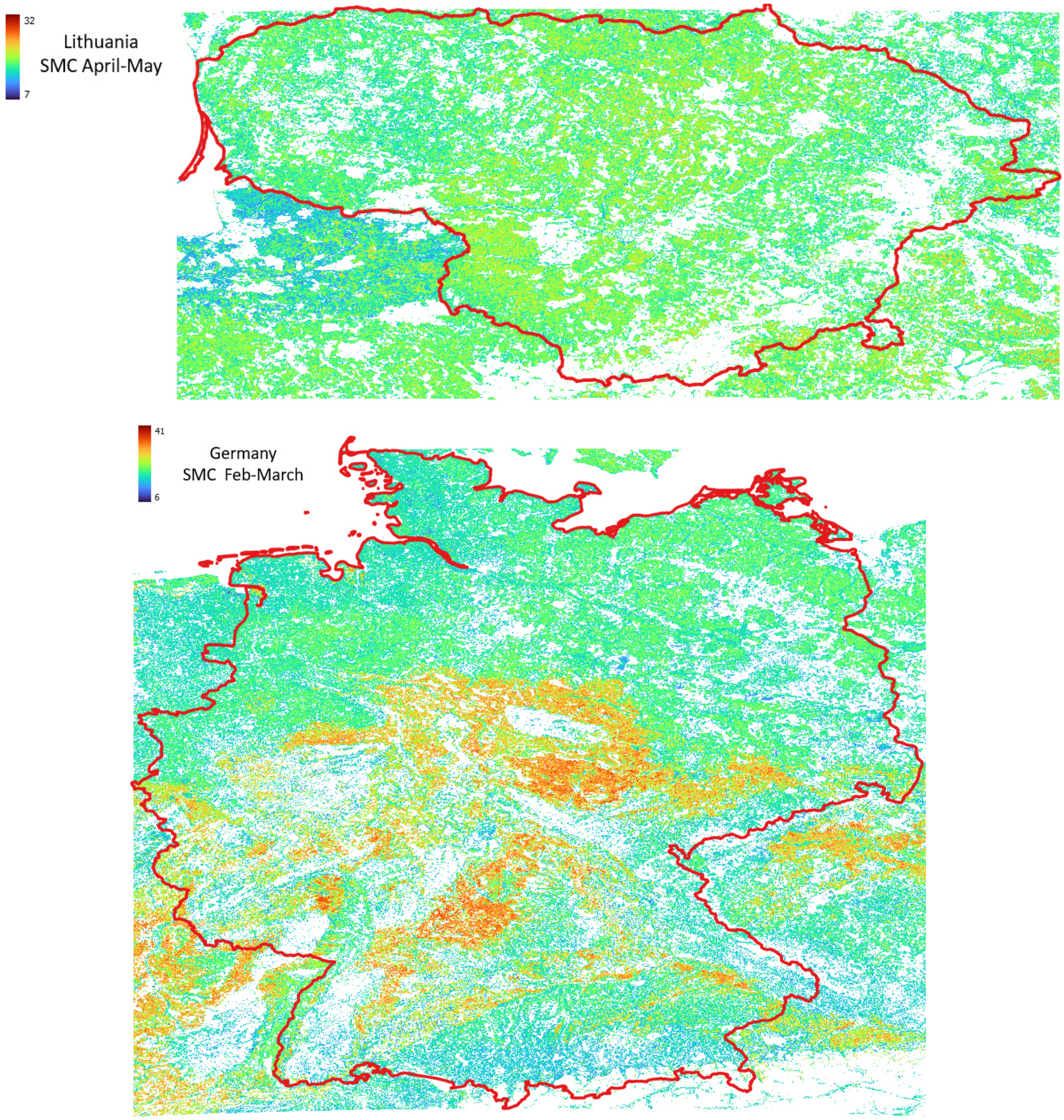


Figure 6: Examples of temporally aggregated whole breeding season SMC data layers for South Tyrol, Lithuania, Flanders, and Germany.

The temporally aggregated SMC layers for each breeding time period were then spatially aggregated to the 200m x 200m grids to obtain the final SMC data layers to be used by the SDM. Figure 7 shows examples of final SMC data layers for Flanders.

Flanders SMC layers – year 2018

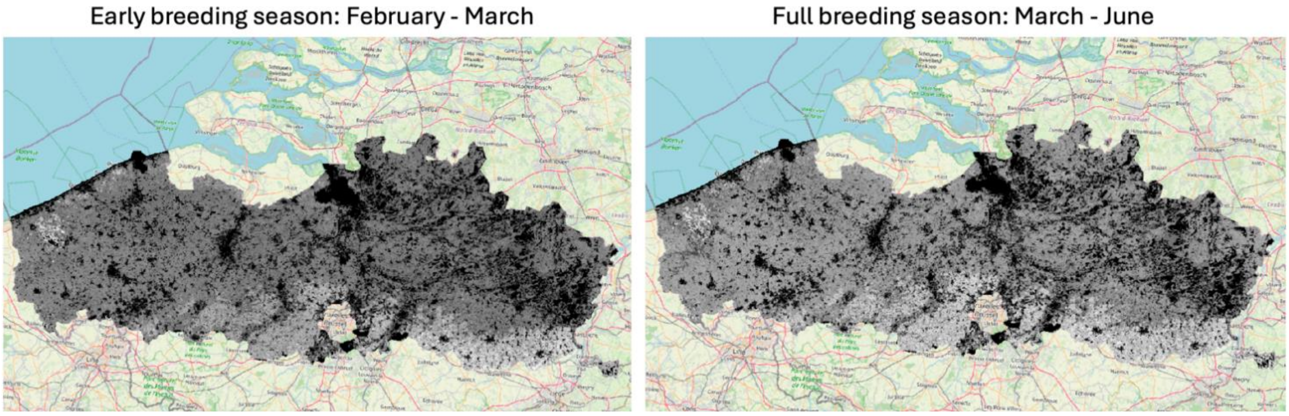


Figure 7: The final, spatially aggregated SMC data layers for the region of Flanders

3.4 Mowing data layers

The Mowing counts and the First mowing DOY (Day Of Year) parameters are mapped by calculating the Sinergise's parcel-based mowing marker from the Sentinel-2 L2A timeseries.

The parcels for all grassland (Permanent grassland, Cultivated grassland and Fallow) were obtained from the LPIS datasets prepared for the Crop types for respective BirdWatch test areas of Flanders, Lithuania and South Tyrol. The obtained parcels were used as input for the calculations of the Sinergise mowing marker for respective regions.

For Germany, a ready raster dataset mapping mowing events in Germany in 2022 from Theunen Institute was obtained from Zenodo. The layer contains both, the mowing counts and the mowing dates, and the First mowing date was extracted and converted to DOY to obtain First mowing DOY data layer. The rasters already had CRS, and we additionally transformed the extent to be aligned with the BirdWatch Germany AOI and aggregated the values to the Germany 200m x 200m grid.

3.4.1 Calculation of the mowing marker

The Sinergise's mowing marker identifies all observations where mowing events lead to a drop in the vegetation index values (NDVI). The marker identifies mowing events by identifying drops in the

NDVI profile of the input parcels. The algorithm used by the mowing marker searches for variations in NDVI to identify candidate mowing events. The algorithm doesn't depend on the neighbouring data and solely uses the NDVI time-series of the target parcel.

The mowing event detection is performed in the following steps:

1. Perform cloudy observation filtering using the s2cloudless cloud detector.
2. Perform other outlier-like filtering using observation outlier detector.
3. Search for pairs of local maxima and minima from the remaining valid observations.
4. Keep only pairs where:
   NDVI drops below threshold value and,
   NDVI recovers above threshold value.
5. Extract information like duration, number of observations, etc. and construct the marker.

An example of detected mowing events on a Lithuanian parcel in the time interval from March 1st to Oct 31st, 2022, is shown in Figure 8.

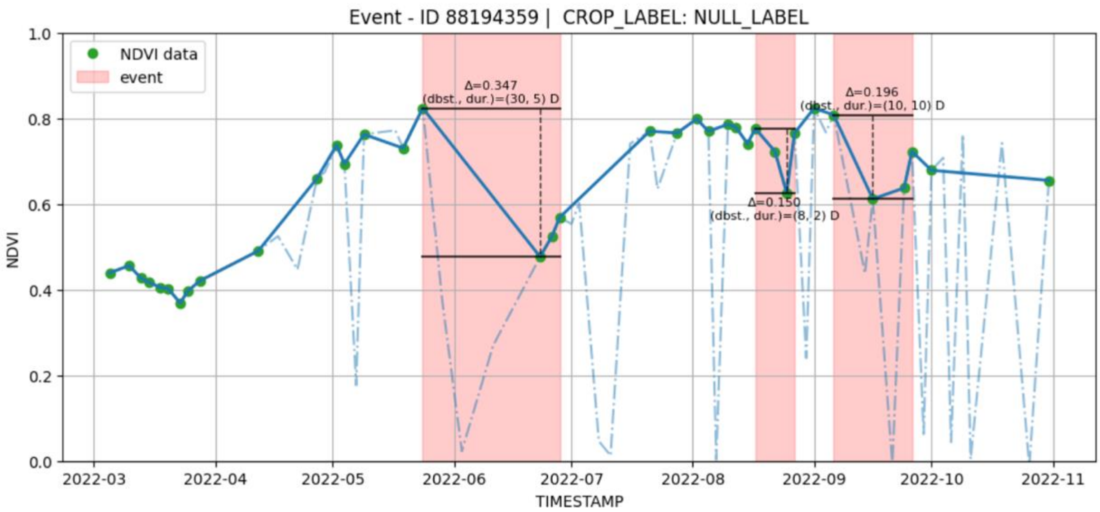


Figure 8: An example of detected mowing events on a Lithuanian parcel. The dashed line represents the whole downloaded NDVI timeseries before filtering steps

The thresholds for NDVI drops and recovery are different, which is achieved by applying different scaling factors to the estimated value of fluctuation in the time-series. The rate of fluctuation in the NDVI time-series is estimated via the differences between the consecutive values in the time-series and averaged over a large sample of parcels beforehand.

3.4.2 Postprocessing of Mowing Marker Results

From the results of the mowing marker for the time period from March 1st – October 31st, the mowing count for each input parcel was extracted and rasterized. The layer was additionally normalised with respect to the number of satellite observations to account for the overlapping satellite swaths artefact. Example of mowing layers before and after normalization are shown in Figure 9.

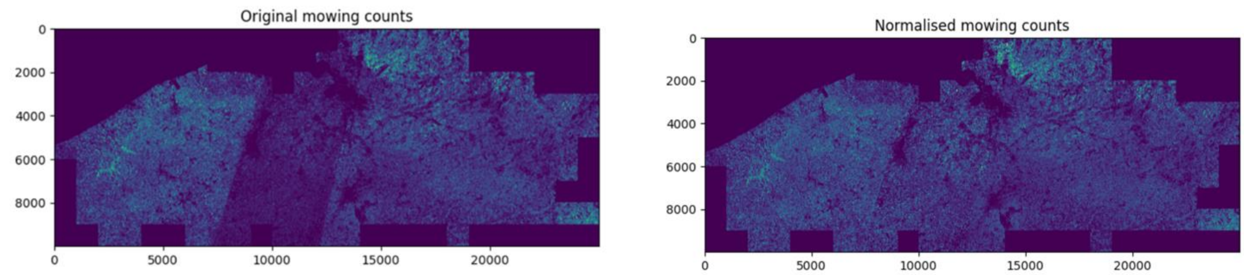


Figure 9: Flanders mowing counts before and after applying the normalisation step

The output was then aggregated with the “mean” metric over the respective area 200mx 200m grid and rasterised to obtain the “Mowing counts” data layer.

For the “First mowing DOY” layer, the first detected mowing event date was extracted from the mowing marker results and converted to day-of-year. The metric for the aggregation to the grid was “minimum”, as it is important to capture the earliest mowing event within each grid cell. An example of the “minimum” metrics aggregation from 10m x 10m resolution to the 200m x 200 m grid of the First mowing DOY results is shown in Figure 10.

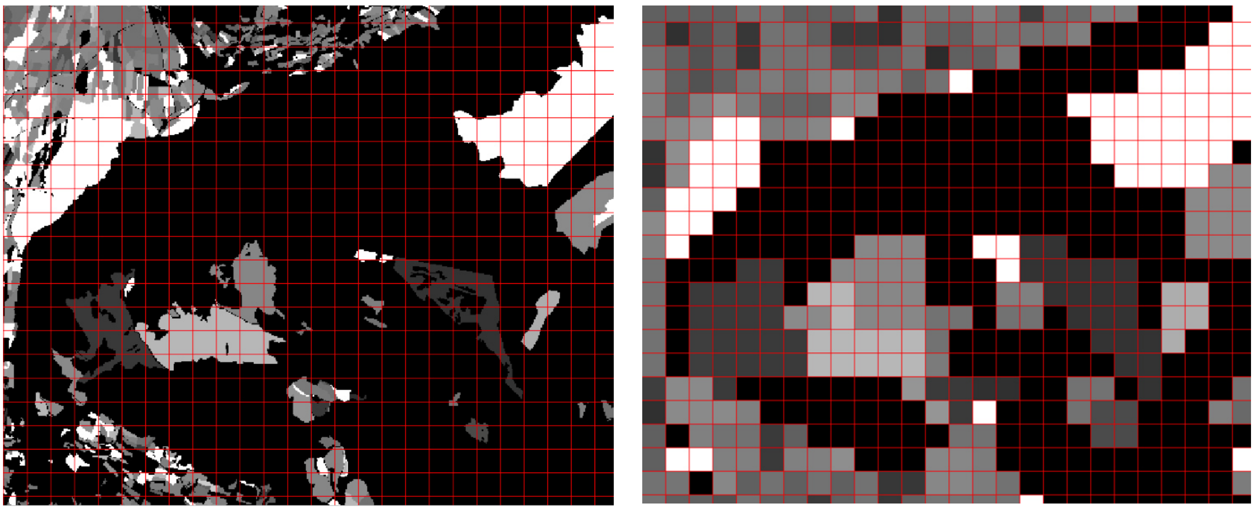
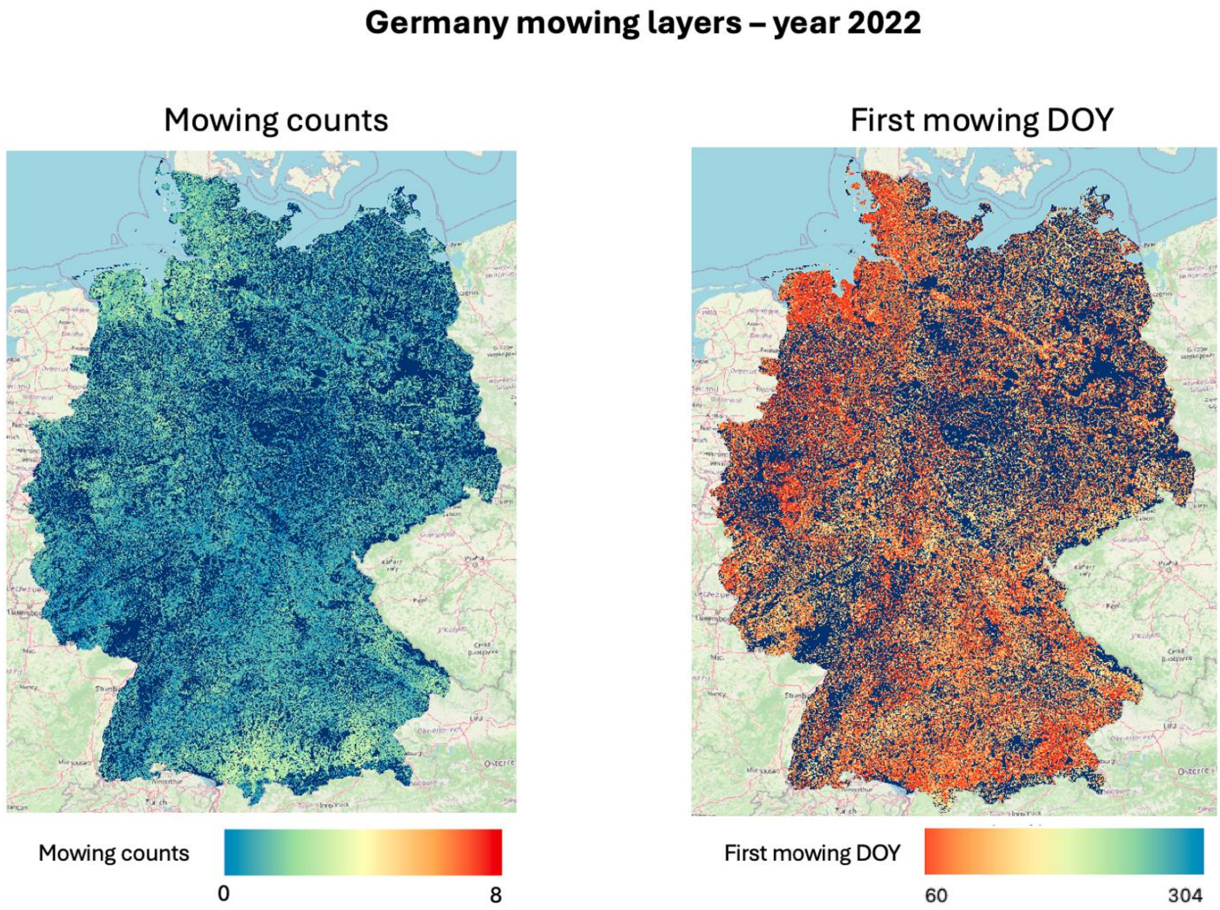
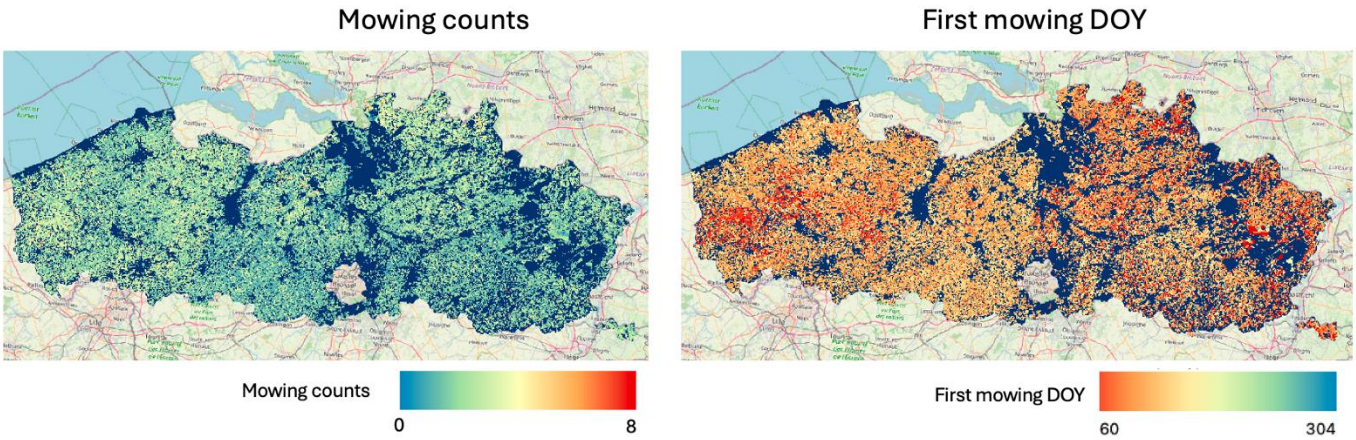


Figure 10: Example of South Tyrol First mowing DOY results before (left) and after (right) aggregation to the200m x 200m grid

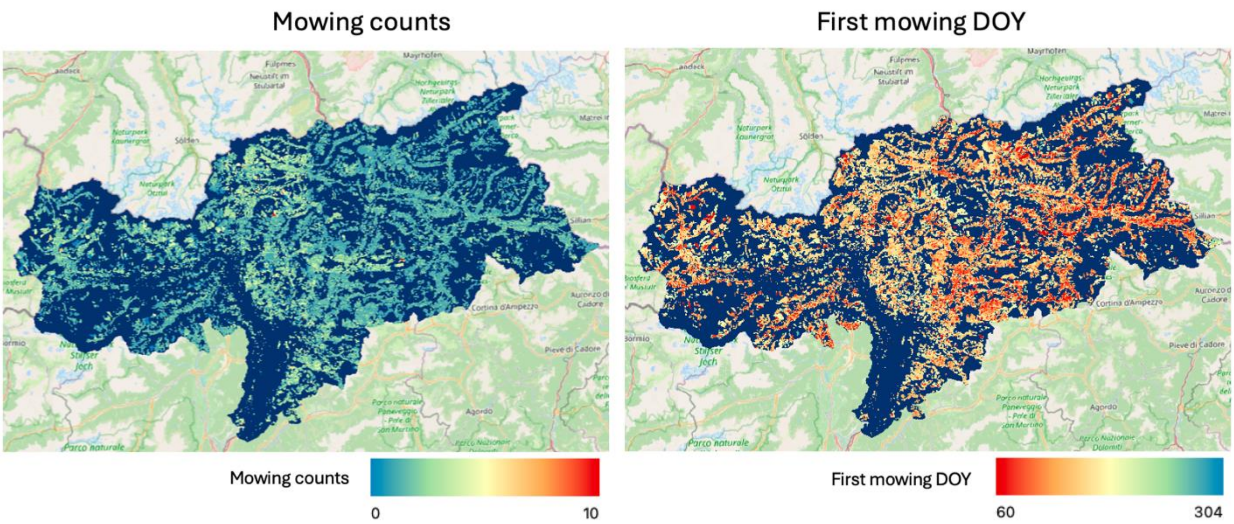
The “Mowing counts” and “First mowing DOY” data layers for BirdWatch test regions are shown in Figure 11.



Flanders mowing layers – year 2018



South Tyrol mowing layers – year 2022



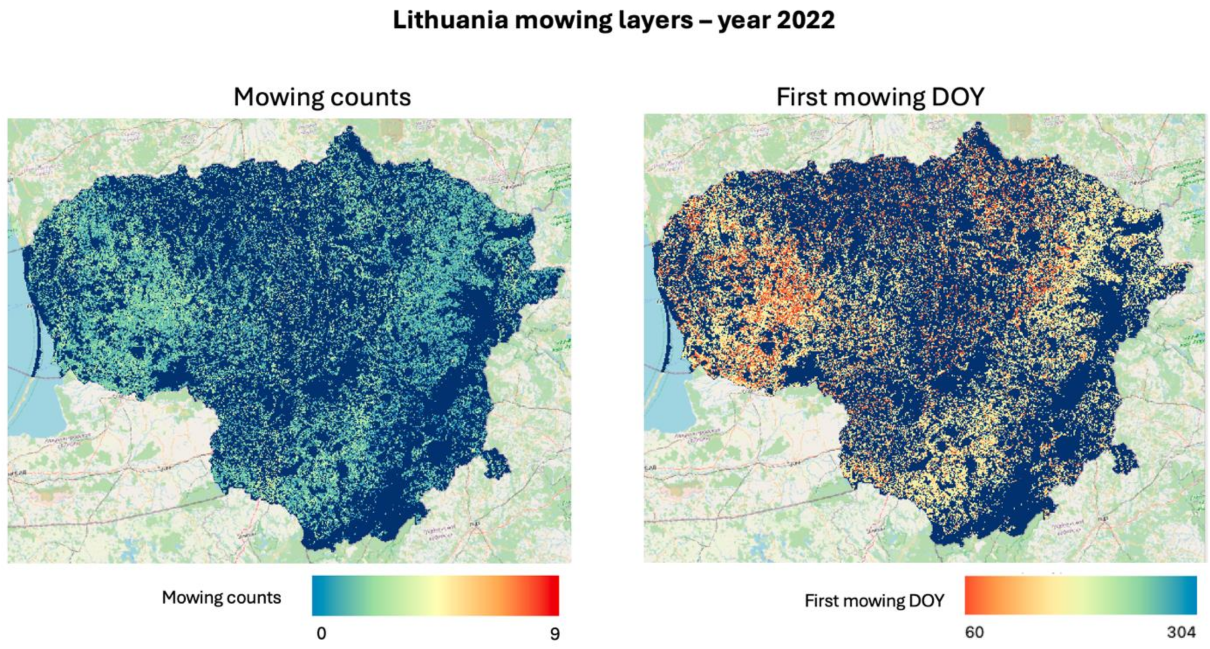


Figure 11: Mowing layers provided for BirdWatch test regions

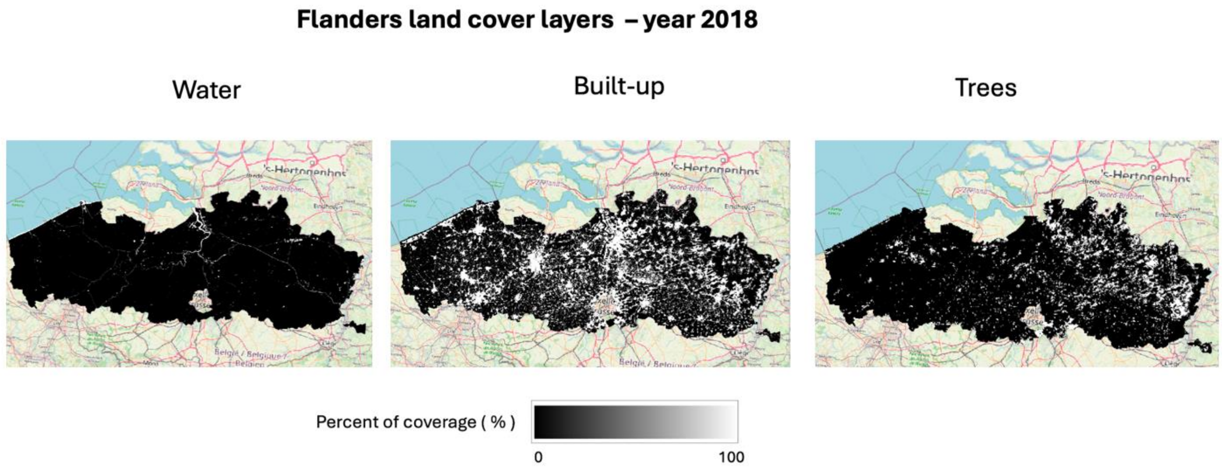
3.5 Land Use Land Cover (LULC) Data Layers

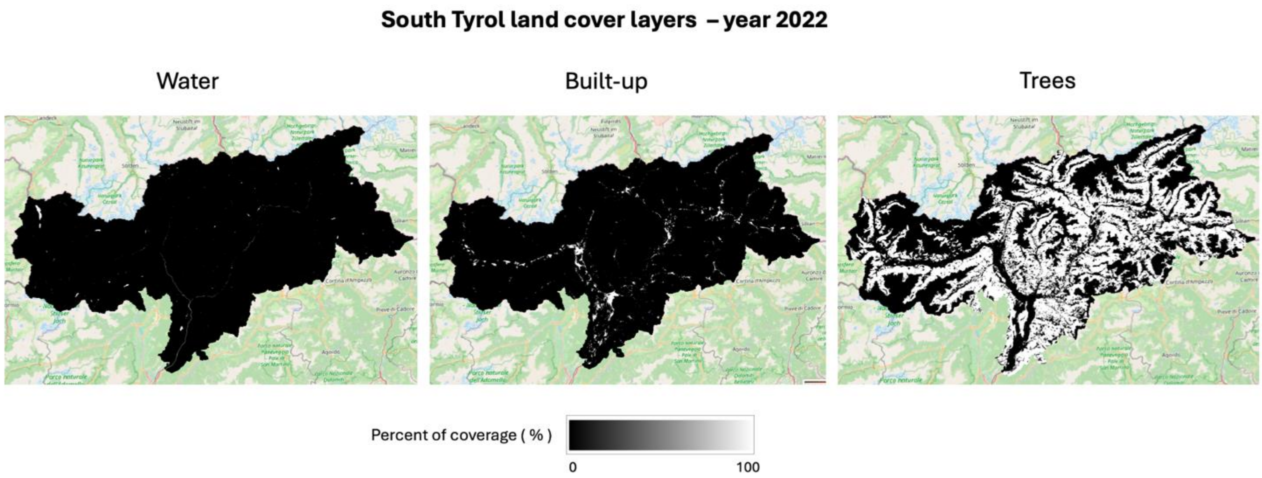
The 10m Annual Land Use Land Cover (LULC) is a land cover dataset produced by Impact Observatory, Microsoft, and Esri collaboratively. It was chosen based on the annual availability, which is required to fulfil the temporal proximity of the bird occurrence data requirement of the SDM. The data collection is derived from ESA Sentinel-2 imagery at 10m resolution using a deep learning AI land classification model trained by billions of human-labelled image pixels. It is produced on annual basis worldwide, so it was chosen as a suitable dataset for the BirdWatch. There are 9 LULC classes generated by the dataset, 3 have been identified as relevant but for BirdWatch:

- **Forest – „Trees“ dataset class**: Any significant clustering of tall (~5 m or higher) dense vegetation, typically with a closed or dense canopy; examples: wooded vegetation, clusters of dense tall vegetation within savannas, plantations, swamp or mangroves (dense/tall vegetation with ephemeral water or canopy too thick to detect water underneath).
- **Water – „Water“ dataset class:** Areas where water was predominantly present throughout the year; may not cover areas with sporadic or ephemeral water; contains little to no sparse vegetation, no rock outcrop nor built up features like docks; examples: rivers, ponds, lakes, oceans, flooded salt plains.
- **Urban Areas / Streets – „Built-up area“ dataset class**: Human made structures; major road and rail networks; large homogenous impervious surfaces including parking structures, office buildings and residential housing; examples: houses, dense villages / towns / cities, paved roads, asphalt.

The three classes were downloaded from Sentinel Hub and aggregated from the original 10m resolution as per the “percent of cell coverage” metric over the 200m x 200m grid for each respective BirdWatch test region. Each LULC class was saved as a separate layer to be used as input for the SDM.

Figure 12 shows the Water, Built-up and Trees land cover data layers prepared for BirdWatch test regions.





Funded by
the European Union

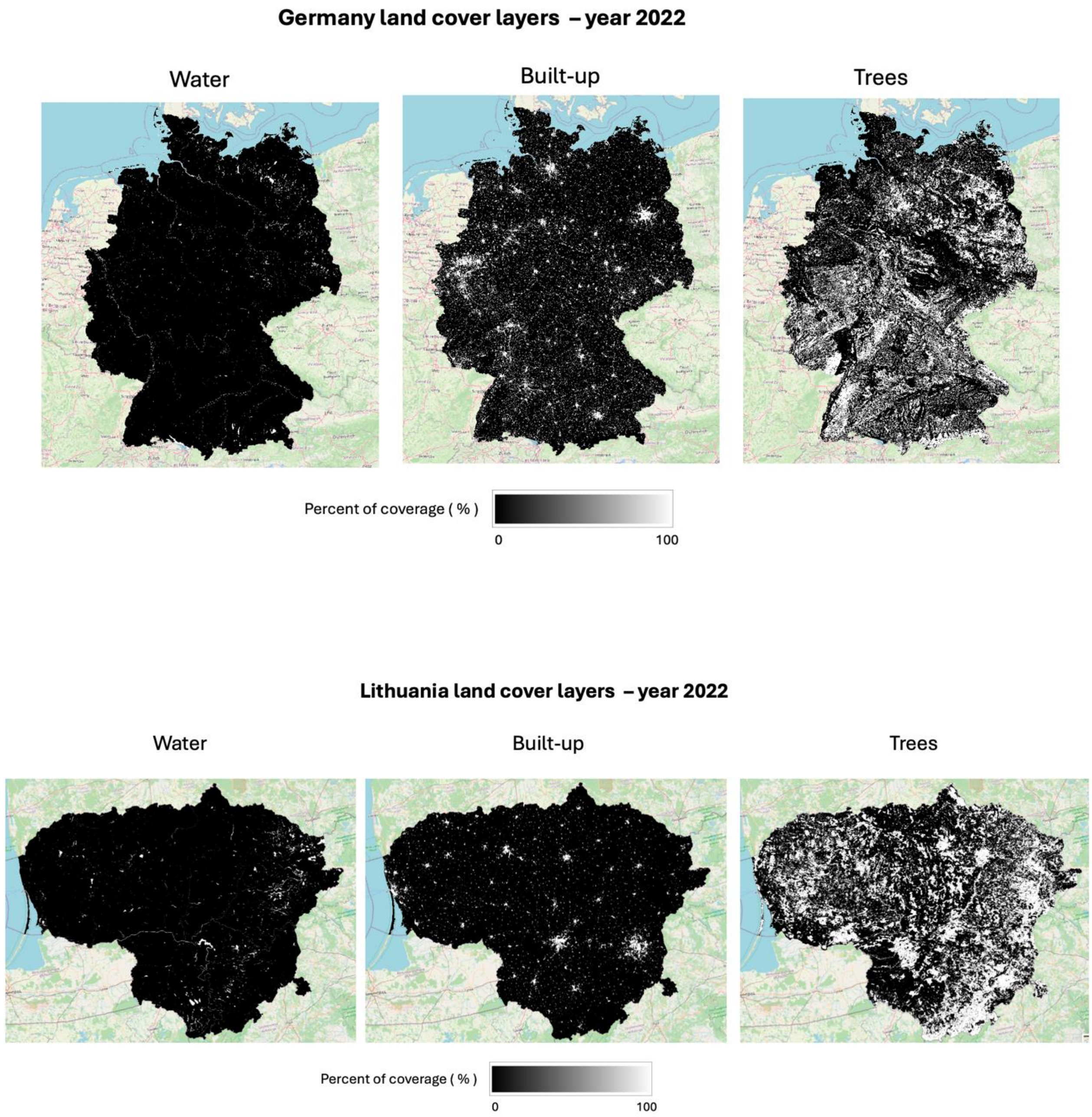


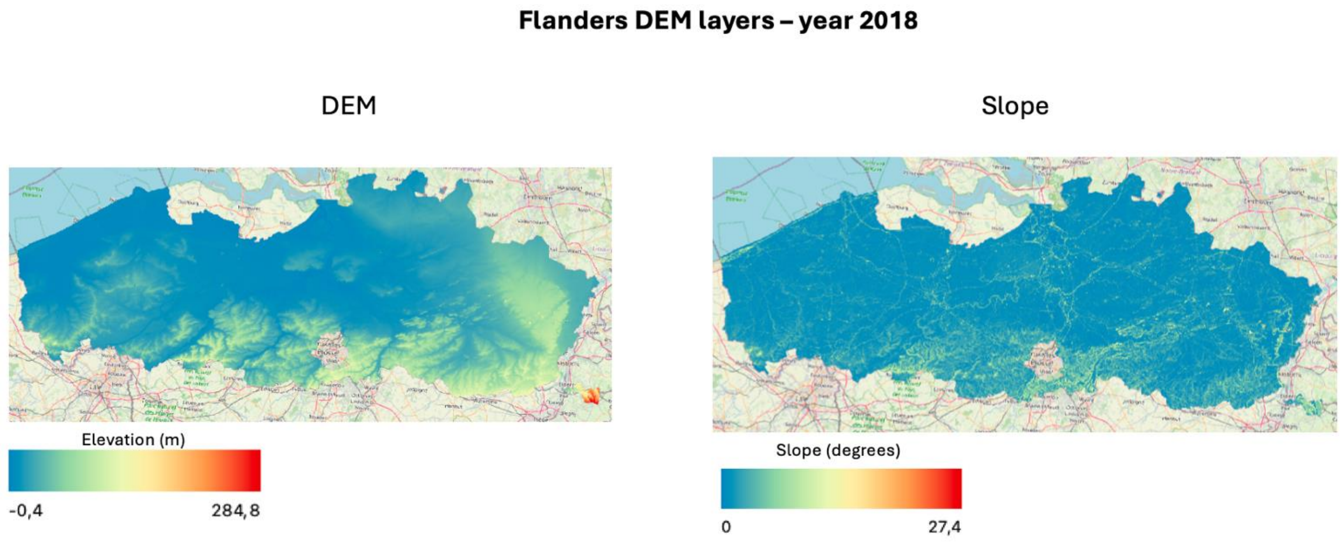
Figure 12 (this and previous page): Water, Built-up and Trees land cover data layers for BirdWatch test regions

3.6 DEM and Slope data layers

A DEM is a digital model or 3D representation of a terrain's surface. With a DEM we can obtain and analyse heights within our areas of interest, which has an impact on bird habitat suitability.

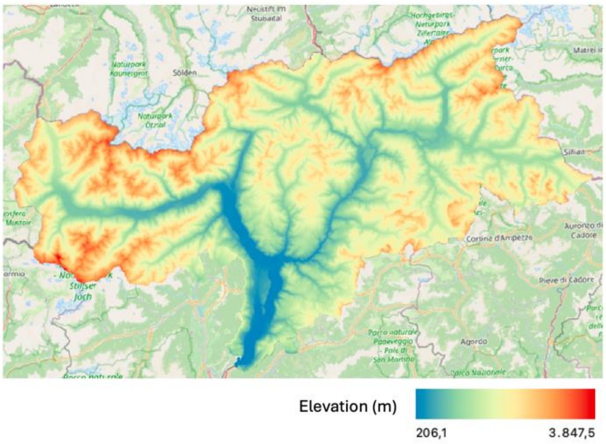
The Digital Elevation Model (DEM) layer was obtained from Copernicus DEM dataset available as a data collection on Sentinel Hub. The slope layer was calculated using “gdal” library functionalities from the obtained DEM layer for all four BirdWatch test regions. Both layers were aggregated with “mean” metric to the 200m x 200m grid.

Figure 13 shows DEM and Slope layers for the BirdWatch test regions.

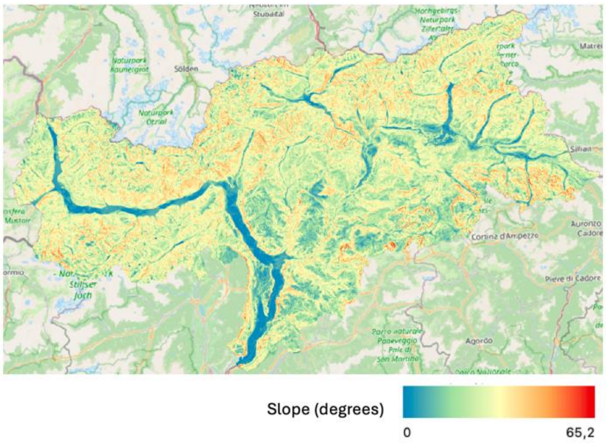


South Tyrol DEM layers – year 2022

DEM



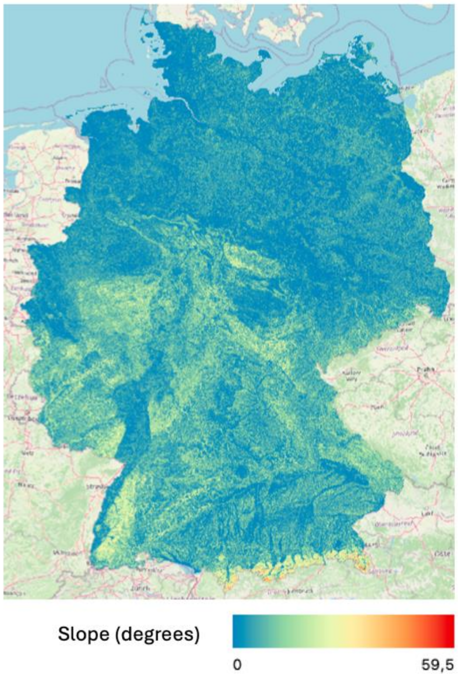
Slope



Germany DEM layers – year 2022

DEM

Slope



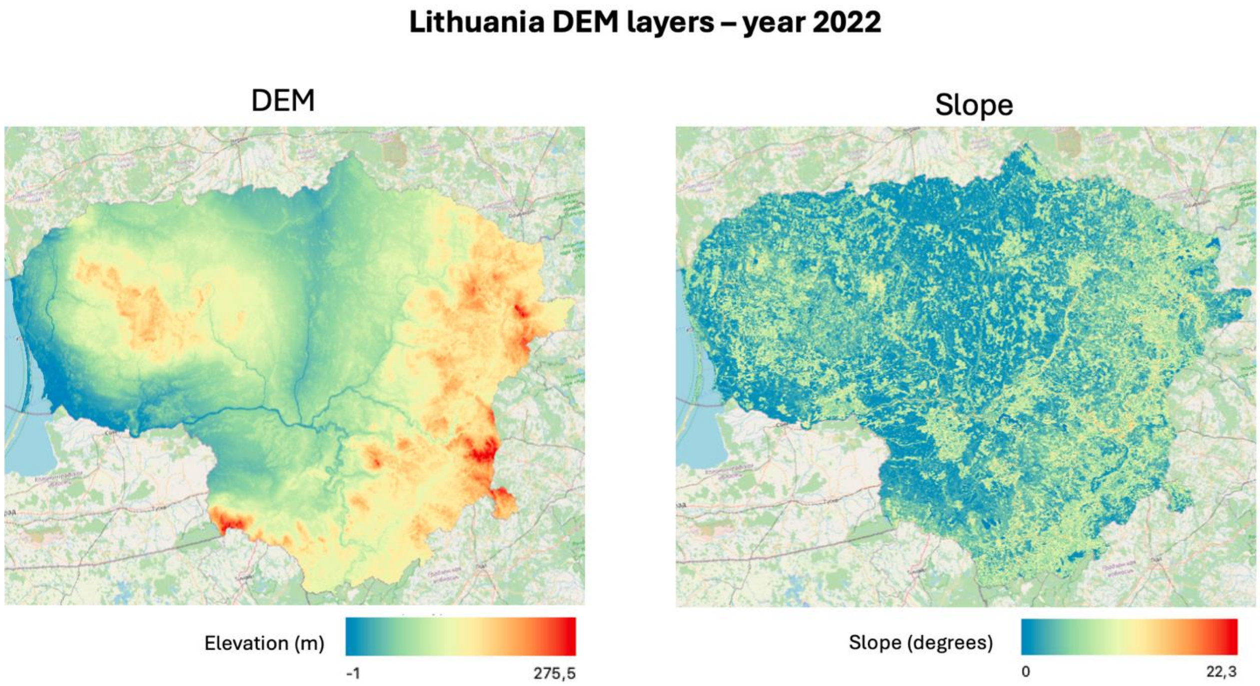


Figure 13 DEM and Slope layers for the BirdWatch test regions

4 Access to BirdWatch datasets

4.1 Access to Crop type data (WMS service)

To offer the chance to view both, all the raw crop type data and the final maps, a WMS service has been set up for BirdWatch project: EDP-MAPS. The service is provided to the BirdWatch consortium as a means to explore and quality check the data. EDP-MAPS is an open-source platform for sharing geospatial data and maps. Here, you can upload the datasets and set permissions for viewing. The Platform is one component of EDP managed by the Center for Sensing Solutions of Eurac Research. All the datasets are accessible via a specified user with its credentials. In Figure 14, we can see EDP-MAPS with all the BirdWatch test regions with some of their crop types layers.

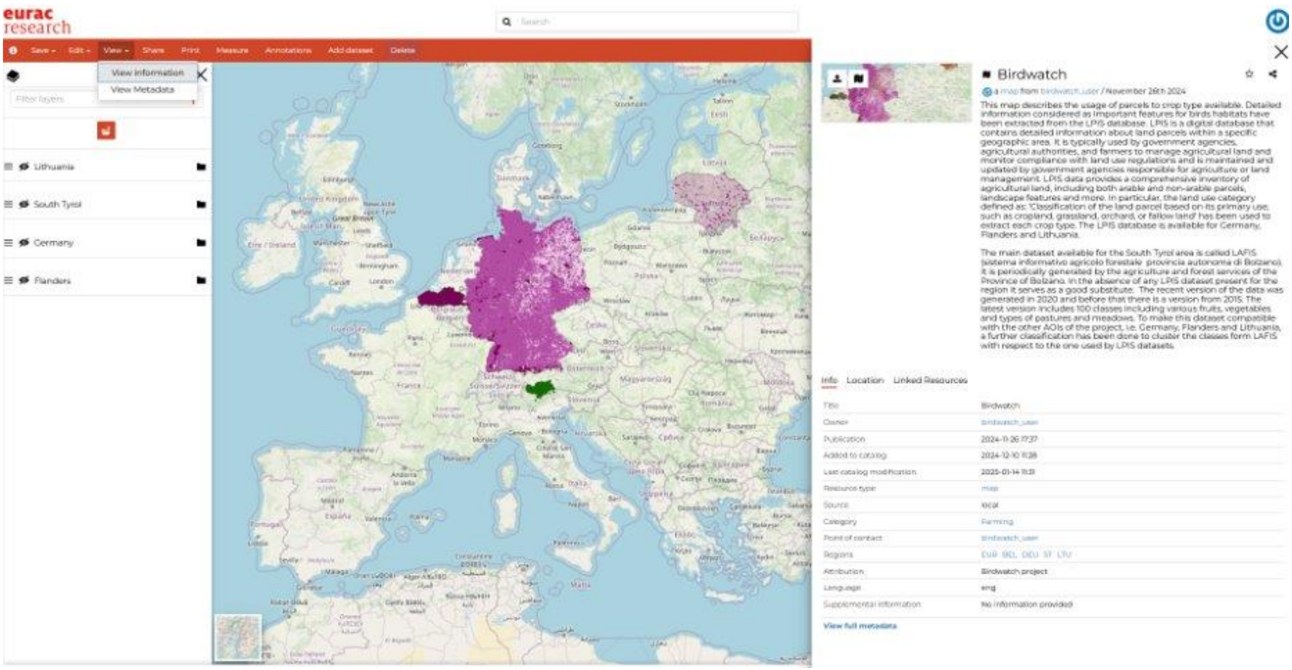


Figure 14: EDP-MAPS with all the BirdWatch test regions with some of their crop types layers

The WMS service was setup for internal use only because all the maps will be publicly available in the BIRDWATCH website that will be published in the next few months.

4.2 Access to raster data layers (Sentinel Hub BYOC)

Sentinel Hub is a cloud-based platform developed by Sinergise that provides access to raster EO data. Some raster data for further processing are available through the Sentinel Hub platform via the “Bring your own Data” functionality. Access to the rasters is provided with a Sentinel Hub account. All the final raster layers can be made available through Sentinel Hub BYOC through agreement with the BirdWatch consortium partners.

4.3 LUP cloud

For quick and easy data access and exchange throughout the process of data layers creation, we also provided the intermediate as well as the final resulting data layers on the cloud storage, specially provided for the BirdWatch project by lead partners LUP. This exchange enabled quick feedback and discussions on some aspects of the data layers and provided an efficient way of sharing the data and communicating within the consortium. As per agreement within the consortium, all the data layers are accessible on LUP Cloud, with access permissions being granted by BirdWatch partners LUP.

5 Conclusion

The document detailed the data acquisition, calculations, and postprocessing procedures for the BirdWatch data layers, including Crop Type, Soil Moisture, Mowing, Land Cover, and DEM. Each dataset was processed to ensure compatibility with SDM, and we provided visual representations of the results.

The document has provided a comprehensive overview of the geospatial datasets prepared for the BirdWatch project across the four demonstration regions: Flanders (Belgium), Germany, Lithuania, and South Tyrol (Italy). The dataset selection was guided by the requirements of bird habitat assessments and the Species Distribution Model (SDM), ensuring that all relevant remote sensing parameters were considered. The data preparation process was structured to meet the SDM input specifications, with careful attention to vector and raster database structures, data access protocols, and the specific needs of different project phases. This coordinated effort among consortium partners ensured that all necessary data layers were available to support the respective Work Packages (WPs) of the BirdWatch project.

Ultimately, more than 120 data layers were generated as input for the SDM of WP4000. The full dataset, covering all four regions, is accessible as outlined in Section 4. This structured approach to data preparation ensures a solid geospatial data foundation to support further progress of the BirdWatch project.