Supplement of Earth Syst. Sci. Data, 18, 4075–4095, 2026
https://doi.org/10.5194/essd-18-4075-2026-supplement

Open Access Earth System Science Data

Supplement of

EuroCrops v2.0: multi-annual harmonized parcel level crop type data linked to European Union-wide survey, statistical, and Earth Observation products

Martin Claverie et al.

Correspondence to: Martin Claverie (martin.claverie@ec.europa.eu) and Ayshah Chan (ayshah.chan@tum.de)

S1 Methodology

S1.1 Data set structure

The data set is shared in geoparquet format, with one geoparquet per country/region and year. In total, 153 geoparquet files are shared. Each file includes one table with the following attributes:

- geom: geometry in EPSG:3035
- cropfield: unique identifier for the given table
- Original_code: crop code as defined in the original data set. In case the crop code was added (because it was missing) or modified (because of temporal inconsistency), the original_code corresponds to the final code. The crop name can be retrieved from the "eurocrops" table.
- Off_area: when provided in the original data set, the area reported for each parcel is included in this attribute.
- Area_ha: computed area of parcel (in hectares) as it is in the data set.

The data set is completed by the stack layers, i.e., the intersection of the yearly GSA. It includes 18 geoparquet files, one for each country/region. Each file includes one table with the following attributes:

- geom: geometry in EPSG:3035
- cropfield: unique identifier for the given table
- cf<year>: corresponding cropfield in the yearly GSA
- c<year>: corresponding original_code in the yearly GSA
- Area_ha: computed area of parcel (in hectares) as it is in the data set.

In addition to the geoparquet files, some mapping tables are provided. The tables are shared through the EuroCropsV2 github repository in folder /data/cropcodemapping/. It includes the following tables shared as csv files:

- hcat4.csv, with columns: hcat4_code, hcat4_name, seasonality_code, seasonality_name
- Eurocrops.csv, with columns: nuts, original_code, original_name, translated_name, hcat4_code, hcat4_name, usage_code, usage_name
- hcat4_agriprod_mapping.csv, with columns: hcat4_code, hcat4_name, usage_code, usage_name, link, agriprod_code, agriprod_name
- hcat4_hrl_mapping.csv, with columns: hcat4_code, hcat4_name, hrl_code, hrl_name.

S1.2 SQL command examples

The following SQL command examples will help in working with the data in an efficient way.

1. Dynamically creates a view by uniting all GSA layers for 2020, with crop code mapped as HCAT4.

```
DO $do$
DECLARE
    sql TEXT;
BEGIN
    SELECT string_agg(
        format(
            'SELECT %L AS nuts, g.cropfield,
```

```
                e.hcat4_name, g.area_ha, g.geom
            FROM gsa.%I g
            LEFT JOIN cropmapping.eurocrops e
                ON g.original_code = e.original_code
                AND e.nuts = %L',
            split_part(tablename, '_', 1),  -- NUTS prefix (used twice)
            tablename,
            split_part(tablename, '_', 1)
        ),
        E'\nUNION ALL\n'
    )
    INTO sql
    FROM pg_tables
    WHERE schemaname = 'gsa'
      AND tablename LIKE '%\_2020' ESCAPE '\';

    IF sql IS NULL THEN
        RAISE NOTICE 'No tables found matching pattern *_2020';
    ELSE
        EXECUTE format('CREATE OR REPLACE VIEW gsa.all_2020 AS %s;', sql);
    END IF;
END $do$;
```

2. Extracts parcels from the NL 2020 GSA dataset, mapped to the AGRIPROD classification system.

```
CREATE VIEW gsa.nl_2020_agriprod AS
SELECT
    g.cropfield,
    m.agriprod_code,
    g.area_ha,
    g.geom
FROM gsa.nl_2020 g
    LEFT JOIN cropmapping.eurocrops e
        ON g.original_code = e.original_code
        AND e.nuts = 'nl'
    LEFT JOIN cropmapping.hcat4_agriprod_mapping m
        ON e.hcat4_code = m.hcat4_code
        AND e.usage_code = m.usage_code;
```

3. For each 10 km grid cell, calculates the total area of parcels classified as ‘maize_corn_popcorn’ in all three consecutive years (2021–2023) within the French GSA.

```
SELECT
    g.grd_id,
    SUM(ST_Area(ST_Intersection(g.geom, s.geom))) AS area_ha
FROM stack.fr_stack s
    LEFT JOIN cropmapping.eurocrops e2021
        ON s.c2021 = e2021.original_code
        AND e2021.nuts = 'fr'
    LEFT JOIN cropmapping.eurocrops e2022
        ON s.c2022 = e2022.original_code
        AND e2022.nuts = 'fr'
    LEFT JOIN cropmapping.eurocrops e2023
        ON s.c2023 = e2023.original_code
        AND e2023.nuts = 'fr'
    LEFT JOIN grid.grid_10k g
```

```
        ON st_intersects(g.geom,s.geom )
WHERE   e2021.hcat4_name = 'maize_corn_popcorn'
    AND e2022.hcat4_name = 'maize_corn_popcorn'
    AND e2023.hcat4_name = 'maize_corn_popcorn'
GROUP BY g.grd_id;
```

4. Retrieves the GSA stack for the Czech Republic (`cz`) across three consecutive years (2021–2023) mapped according to HCAT4 taxonomy.

```
SELECT
    g.cropfield, g.area_ha, g.geom,
    e21.hcat4_code, e22.hcat4_code, e23.hcat4_code
FROM stack.cz_stack g
    LEFT JOIN cropmapping.eurocrops e21
        ON e21.original_code = g.c2021 AND e21.nuts = 'cz'
    LEFT JOIN cropmapping.eurocrops e22
        ON e22.original_code = g.c2022 AND e22.nuts = 'cz'
    LEFT JOIN cropmapping.eurocrops e23
        ON e23.original_code = g.c2023 AND e23.nuts = 'cz';
```

S1.3 Example of overlapping polygons and removal of area

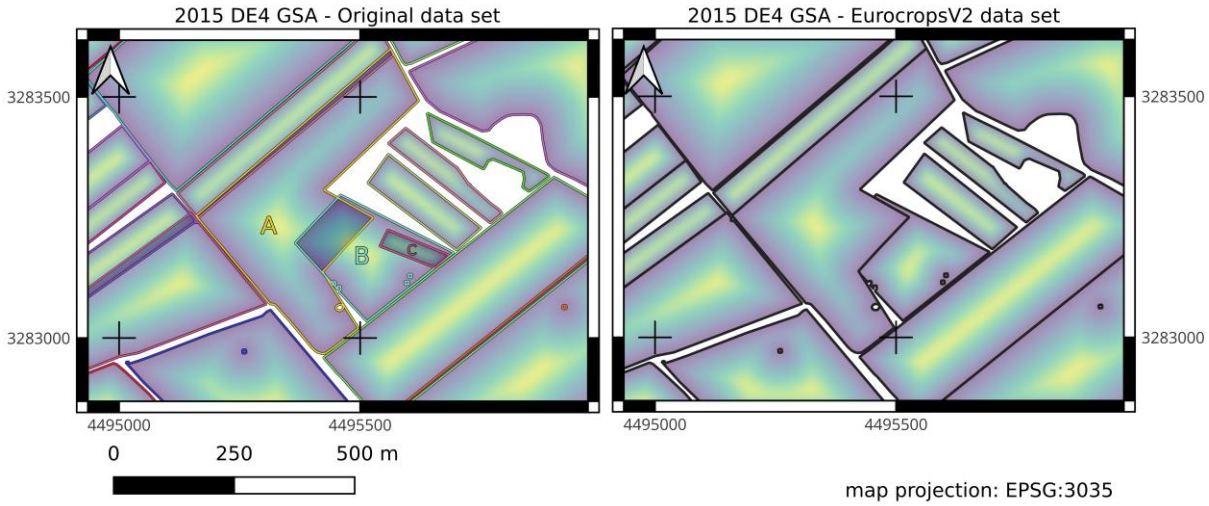

Figure S1. An example illustrating overlapping polygons in the original dataset (on the left) and the final cleaned polygons in EuroCrops V2.0 (on the right). The example is taken from the 2015 Brandenburg (DE4) GSA. In the map on the right hand side, parcel outlines are shown in randomly assigned colors, with shapeburst shading is used for polygon fills. In the central area, parcel B overlaps both parcels A and C. The A/B overlap area was assigned to parcel A because parcel A has a larger total area than parcel B. Parcel C was removed, and its overlap area was assigned to parcel B for the same reason.

S1.4 Translations of crop names

Before the crop names were translated from their native language to English, a pre-processing step was undertaken to align the crop names across the years because countries provide an updated crop code list on an annual basis and the EuroCropsv2.0 process has been ongoing since 2021. The github now includes a table with the original_code and how these are related to multiple instances of original_name. The process for consolidating the dataset (i.e., having unique pairs of original_code/original_name) involved the use of an LLM (gpt-4o and nous-hermes-2-mixtral-8x7b-dpo), or a manual process. The source of the consolidated original_name is provided in this table here: https://github.com/Martincccc/EuroCropsV2/blob/main/data/processing/name_mistakes.csv.

The aligned crop names were then translated using three machine learning packages. We have uploaded one table per country (XX_trans_stats.csv) with the original codes and their translations from Google Translate (g_tname), DeepL (dpl_name) and OPUS-MT (opus_tname) to github, placed in this folder: https://github.com/Martincccc/EuroCropsV2/tree/main/data/eurocrop_trans, where XX in the filename is the two letter country code or the three letter regional code. Note that these files have been updated in 2026 so that they are reproducible since this process was done over several years (since 2021) as outlined above.

We attempted to provide the best translations possible, which is why a third machine translator was introduced as part of EuroCropsv2.0. At the beginning of EuroCropsv2.0, we discovered some wrong translations by DeepL and Google Translate. Therefore, we added OPUS-MT, which is not as good as the other two translators, accounting for 63-92% of the translation errors, but it served as an additional check. Without the use of OPUS-MT, the number of manual checks would have been reduced significantly, but we ran the risk of wrong translations that would not have been picked up using only DeepL and Google Translate.

The translation accuracy varies by country with the lowest in Spain at 40% to the highest in Belgium at around 72%. Estonia provides English translations of most labels on https://klassifikaatorid.stat.ee/, so a simple search there resolved most disagreements. Other countries do not have such convenient portals. Slovenia releases their labels along with the scientific name of the crops while Ireland's labels are in English. Spelling mistakes were rare except for the Spanish, Bulgarian, Czech, Portuguese and Italian GSA, which led to direct failures of translation and no results were provided by Google Translate. In the German GSA, there were many abbreviations and short forms of names used, which hindered the automatic translation. In the GSA of Czechia, there were extremely detailed species level crop descriptions of grasses and legumes as well as the use of colloquial terms. After machine translation, colloquial terms often led to non-crop-like descriptions such as diabetes, quiet, etc. Subsequent Google searches of these terms sometimes turned up empty, whereas for other niche translations, they showed up in Wikipedia. In other cases, we sent the colloquial terms to native speakers to review but this was not possible for all languages.

Table S1 summarizes the number and percentage of multiple instances of original_name per country and the translation statistics and errors. The accuracy refers to the percentage of time when all the machine translations were in agreement. Where there was disagreement in the remaining translations, these were all manually checked. Of these manually checked translations, five different cases may have occurred. In the majority of cases, at least one machine translator was correct. Often OPUS-MT was the poorest performer, but other times, e.g., in Slovakia, Google Translate and DeepL failed but OPUS-MT was correct. The second type of case included the occurrence of niche crop names, e.g., species level trees or legumes, which were confirmed by Wikipedia, or where the translator chose a more frequently used homonym, which were then checked using Wikipedia's disambiguation pages. A third case was the occurrence of colloquial or regional names, which were

also checked using Wikipedia. Finally, spelling errors (case 4) and crop names that were abbreviated or truncated (case 5) were also identified through this manual checking process.

Table S1: Summary of instances of multiple crop names and translation statistics.

Country or Region	# of multiple original_ name [%]	Accuracy	Manual Correction	At Least 1 Machine Translation Correct	Niche Crop/ Homonymy	Colloquial/ Regional name	Spelling errors	Abbreviation/ Truncation
AT	13 [5%]	60.98	39.02	82.29	6.25	1.04	0.00	10.42
BE2	52 [14%]	71.82	28.18	92.16	6.86	0.98	0.00	0.00
BE3	75 [36%]	71.29	28.71	88.33	11.67	0.00	0.00	0.00
BG	-	58.30	41.70	72.57	24.78	1.77	0.88	0.00
CZ	146 [35%]	45.08	54.92	62.56	33.04	0.88	2.64	0.00
DE4	89 [30%]	67.69	32.31	77.13	12.77	0.00	0.00	10.11
DEA	51 [18%]							
DK	286 [56%]	63.78	36.22	66.30	27.72	0.00	0.00	5.98
EE	-	43.77	55.23	76.14	16.30	0.00	0.00	0.00
ES	-	40.29	59.71	69.14	26.34	2.47	1.65	0.41
FI	-	57.20	42.80	85.71	14.29	0.00	0.00	0.00
FR	-	64.91	35.09	84.87	15.13	0.00	0.00	0.00
IE	-	-	-	-	-	-	-	-
ITI1	1 [0%]	59.35	40.65	75.51	23.81	0.00	0.68	0.00
NL	91 [18%]	76.42	23.58	87.93	8.62	0.00	0.00	3.45
PT	-	56.73	43.27	77.78	14.81	1.48	5.19	0.74
SI	8 [4%]	-	-	-	-	-	-	-
SK	-	63.48	36.52	88.35	9.71	1.94	0.00	0.00

S2 Results

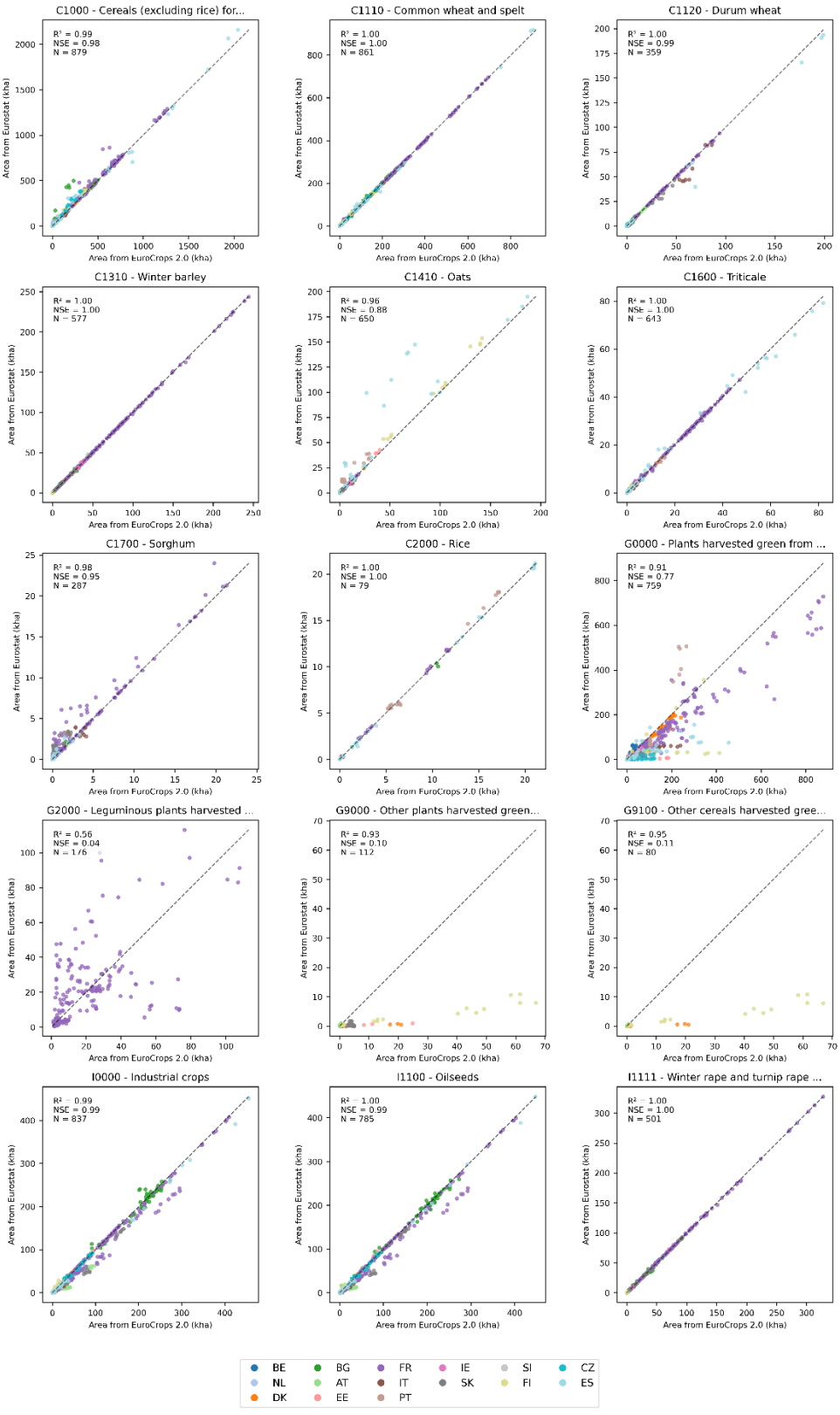

Figure S2: Area comparison between crops in the EuroCrops v2.0 database and Eurostat at NUTS2 level and linked by AGRIPROD code (C1000 to I1111) for all years available

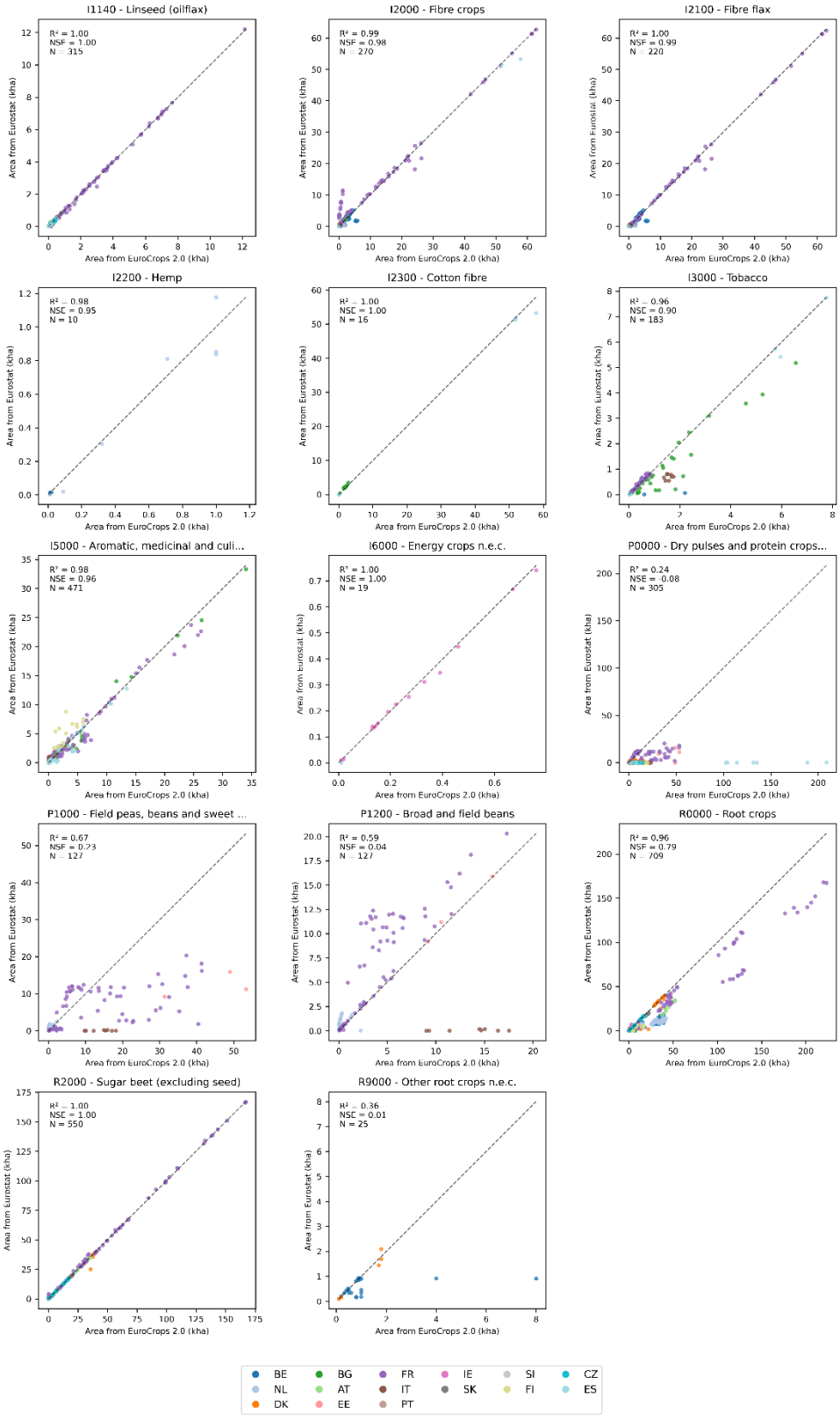

Figure S3: Area comparison between crops in the EuroCrops v2.0 database and Eurostat at NUTS2 level and linked by AGRIPROD code (for I1140 to R9000) for all years available

Table S2: Evaluation scores for all crops in Figure 11, Figure S2 and Figure S3. NSE is the Nash Sutcliffe Efficiency and MSE is the Mean Squared Error in ha.

AGRIPROD code	AGRIPROD name	R2	NSE	MSE	N
C0000	Cereals for the production of grain (including seed)	0.9965	0.9917	538	879
C1000	Cereals (excluding rice) for the production …	0.9937	0.9844	977	879
C1110	Common wheat and spelt	0.9998	0.9996	6.0312	861
C1111	Common winter wheat and spelt	1.0000	0.9999	2.1143	628
C1120	Durum wheat	0.9966	0.9917	5.7657	359
C1200	Rye and winter cereal mixtures (maslin)	0.9861	0.9627	3.1068	710
C1300	Barley	0.9987	0.9975	19	869
C1310	Winter barley	0.9999	0.9999	0.2123	577
C1410	Oats	0.9640	0.8849	47	650
C1500	Grain maize and corn-cob-mix	0.9372	0.8244	452	772
C1600	Triticale	0.9987	0.9974	0.3857	643
C1700	Sorghum	0.9816	0.9510	0.6755	287
C2000	Rice	0.9990	0.9977	0.0969	79
G0000	Plants harvested green from arable land	0.9297	0.8434	2559	848
G1000	Temporary grasses and grazings	0.7646	0.5642	3018	628
G2000	Leguminous plants harvested green	0.8295	0.2432	387	681
G3000	Green maize	0.9959	0.9882	43	323
G9000	Other plants harvested green from arable land	0.8267	0.1346	171	112
G9100	Other cereals harvested green (excluding green maize)	0.8372	0.1460	229	80
I0000	Industrial crops	0.9989	0.9978	13	837
I1100	Oilseeds	0.9990	0.9980	11	785
I1110	Rape and turnip rape seeds	0.9999	0.9997	0.5432	746
I1111	Winter rape and turnip rape seeds	0.9999	0.9998	0.5792	501
I1120	Sunflower seed	0.9994	0.9986	5.8294	468
I1130	Soya	0.9996	0.9991	0.1018	357
I1140	Linseed (oilflax)	0.9995	0.9989	0.0027	315
I2000	Fibre crops	0.9892	0.9778	2.8233	270
I2100	Fibre flax	0.9972	0.9942	0.6876	220
I2200	Hemp	0.9757	0.9495	0.0095	10
I2300	Cotton fibre	0.9991	0.9966	1.4402	16
I3000	Tobacco	0.9582	0.8990	0.1351	183
I5000	Aromatic, medicinal and culinary plants	0.9820	0.9633	0.5818	471
I6000	Energy crops n.e.c.	0.9988	0.9964	0.0002	19
P0000	Dry pulses and protein crops for the …	0.9500	0.8625	38	772
P1000	Field peas, beans and sweet lupins	0.9748	0.9413	6.6009	571
P1100	Field peas	0.9818	0.9619	3.4871	434
P1200	Broad and field beans	0.9191	0.7397	2.7534	532
P1300	Sweet lupins	0.9811	0.8969	0.0536	220
R0000	Root crops	0.9945	0.9887	8.4466	810
R1000	Potatoes (including seed potatoes)	0.9950	0.9896	1.1781	757
R2000	Sugar beet (excluding seed)	0.9996	0.9991	0.4645	550

R9000	Other root crops n.e.c.	0.3645	0.0056	2.4879	25

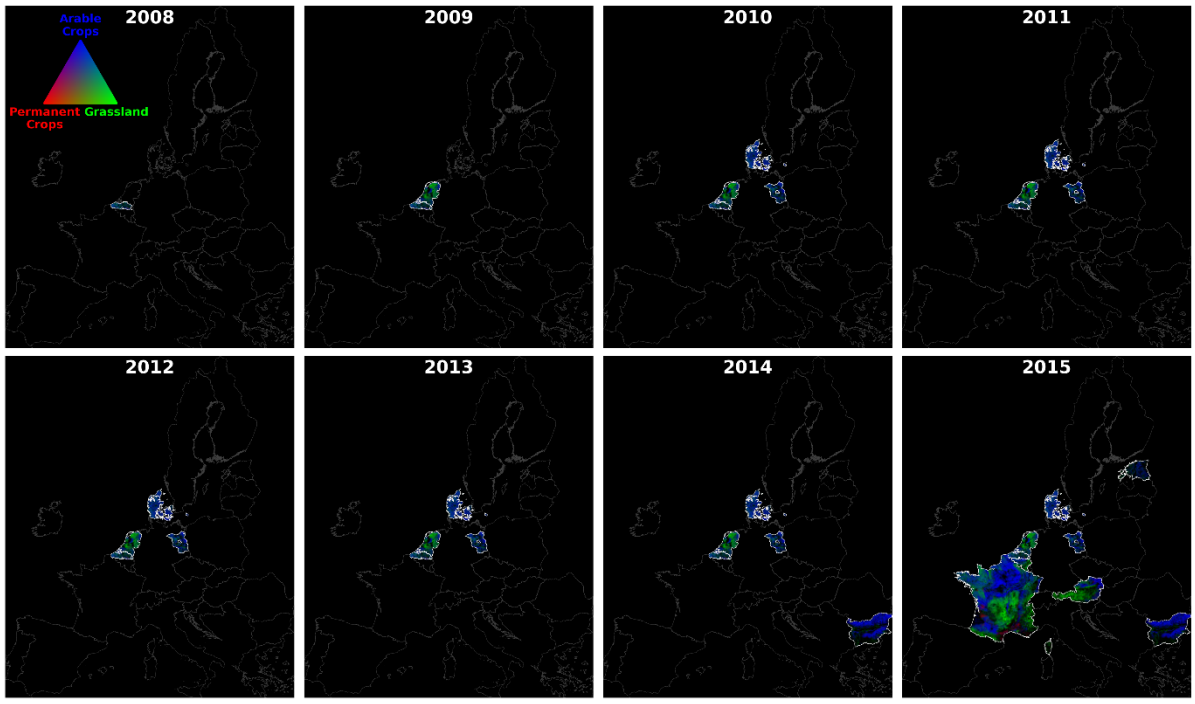

Figure S4: The proportion of arable crops, permanent crops and grassland in the GSA at a 10 km resolution for the years 2008 to 2015. The borders are shown in white.