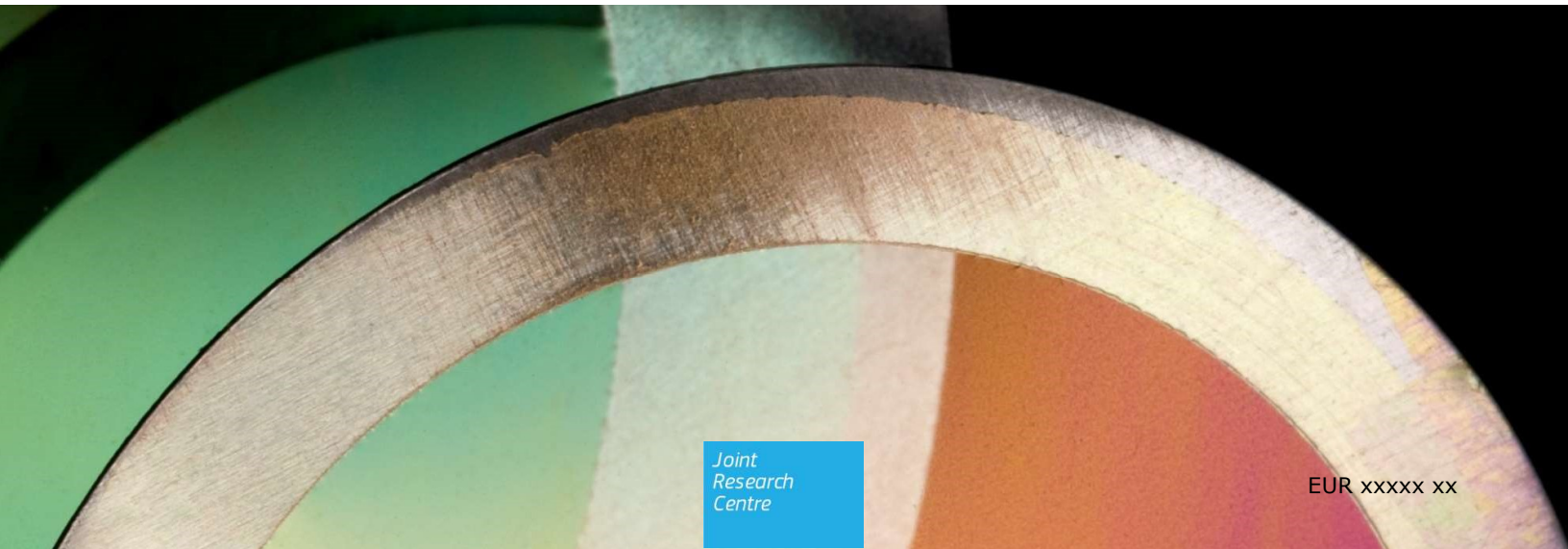
European Commission

JRC TECHNICAL REPORTS

Algorithm for the dissagregation of crop area statistics in the MARS crop yield forecasting system

Cerrani, I.
López Lozano, R.

2017



EUR xxxxx xx

Contents

Foreword

This report describes the algorithm designed to disaggregate regional and area statistics in Europe at all NUTS and GAUL administrative levels and onto a regular grid of 25 km. That algorithm is necessary to aggregate of crop indicators in the MARS Crop Yield Forecasting System (MCYFS) from the original resolution of the BioMA-WOFOST crop model resolution at EMU and 25 km grid cells to national and sub-national scale when using this indicators operationally for crop yield forecasting (in the MARS Bulletins).

The algorithm uses as input crop area statistics from Eurostat and national statistical services delivered form the different Member States, and land cover map. A set of rules and mathematical operations are set to construct, combining the mentioned sources, the full hierarchical system of regional weights per crop and year, expressing the proportion of crop area contained in every single spatial unit (region or grid) of the total national crop area. These regional weights provide, when multiplied by the national figure, the distribution of crop areas over Europe at grid and at any administrative level. Moreover, it can be used in the spatial aggregation or disaggregation of any crop indicator based on crop area.

In the first chapter is an introduction, describing the spatial framework and the main principles of indicators aggregation in the MCYFS. In the second chapter, the processing steps of the algorithm of aggregation are explained in depth. Finally, Chapter 3 describes the outputs of this algorithm that can be downloaded from the MARS data portal.

1 Introduction

In the MARS Crop Yield Forecasting System (MCYFS) the crop indicators are originally produced by the WOFOST crop growth model implemented in the BioMA modelling platform at elementary mapping units (EMUs). The EMUs are generated from the intersection between each of the 25 km grid cells constituting the reference spatial grid used to produce meteorological data in the system, and the soil mapping units (SMUs) which are known spatial units necessary to establish soil suitability for the different crops and simulate the soil water balance in BioMA-WOFOST. SMUs, extracted from the Soil Geographical Database of Europe (version 4.0[1]) are also composed of on one or more soil typological units (or STUs), which contain information of the soil physical properties. A schematic representation of the different spatial units is given in Figure 1.

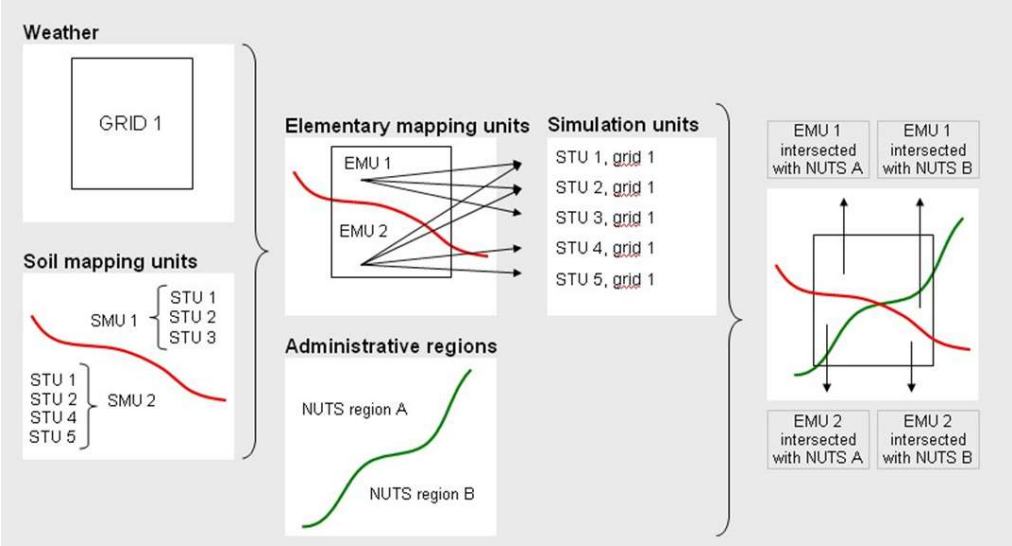


Figure 1. Graphic representation of the intersections between the different spatial units of the MCYFS: soil typological units (STU), soil mapping units (SMU), elementary mapping units (EMU), 25 km grids (GRID) and administrative regions (NUTS).

The BioMA-WOFOST model runs at the STU level with 25km grid meteorological data, and then indicators at the STU level are automatically aggregated at the SMU level based on a constant weight describing the relative area of a STU in the SMU it belongs to. Actually, STUs do not have an explicit spatial representation in the MCYFS, and therefore this automatic aggregation assumes that the weight of a given STU in a SMU is constant over the space.

The aggregation of the BioMA-WOFOST model indicators from EMU to 25 km grid cells and the different administrative levels (NUTS system of EU, *nomenclature d'unités territoriales statistiques, and* GAUL system of FAO for the countries in the EU neighbourhood within the European Window of the MCYFS) has to be computed assigning to the different EMUs a specific weight. That weight represents its expected contribution to the crop area of the higher administrative unit (either grid or NUTS region) and is established differently depending to which administrative level the indicator is aggregated at (Figure 2).

When aggregating crop indicators from EMU to grid the criteria used to establish the weights is based exclusively on land cover area:

$$W_{i,g} = \frac{A_{i \cap s \cap l}}{A_{g \cap s \cap l}} \qquad \text{Eq. 1}$$

$$Y_{g,y,c} = \sum_{i}^{n} W_{i,g} * Y_{c,y,i} \; \forall \; i \in g \qquad \text{Eq. 2}$$

[1] http://eusoils.jrc.ec.europa.eu/ESDB_Archive/ESDBv2/index.htm

Where $Y_{g,y,c}$ is the indicator value aggregated for the crop c year y at the grid g; $W_{i,g}$ is the relative weight of EMU i on grid g; $Y_{c,y,i}$ is the original indicator value simulated by BioMA-WOFOST at EMU I; $A_{i\cap s\cap l}$ is the area intersection of EMU i with land cover class l and suitable soil s; and $A_{g\cap s\cap l}$ is the area intersection of grid g. Suitable soil area s is directly calculated from the STUs considered suitable for a specific crop group, according to soil physical properties. Land cover class l represents arable land in the MCYFS European window, and the spatial layer used to calculate the mentioned area intersections has been generated from the union of the classes "Non-Irrigated arable land" and "Permanently irrigated arable land" from the Corine Land Cover map (Buttner et al., 2004) –classes number 211 and 212, respectively– within EU countries. For the EU neighbourhood arable land class l has been generated from the class "Cultivated and managed areas" –class number 116– from the GLC2000 Global land cover map (Bartholomé and Belward, 2005).

The same approach based on land cover data and soil suitability (Eq. 1 and Eq. 2) is used to aggregate from EMU to NUTS 3 (or GAUL 2) administrative level, calculating analogously $W_{i,r}$ the relative weight of EMU on NUTS 3 region $Y_{r,3}$. These relative weights $W_{i,g}$ and $W_{i,r}$ are easy to calculate having a land cover map and a soil suitability map. Moreover, their value for a given grid or region are the same for many crops, as the arable land cover classes on Corine Land Cover map are used for all crops except rice, and soil suitability is assigned per crop group (cereals/pulses, root crops, and maize crops).

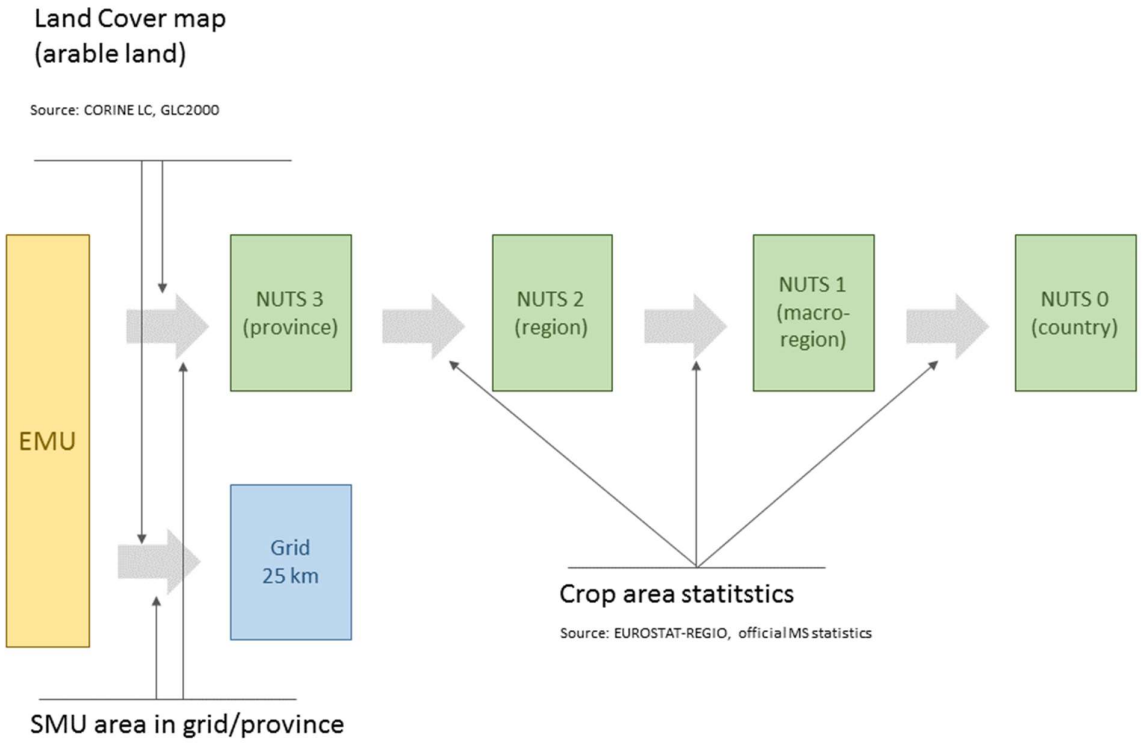


Figure 2. Input data necessary for the aggregation of the CGMS model indicators from their original resolution (EMU, elementary mapping unit) to the different spatial units of the MCYFS: grid, and administrative regions (from NUTS 3 to NUTS 0 level).

The further aggregation of model indicators at higher administrative levels is performed in successive steps from the indicators values at NUTS 3 level to NUTS 2, NUTS 1 and NUTS 0 (Figure 2). In these steps, additional information on crop area based on official crop area statistics is ingested, with the purpose of weighting more those administrative units where the presence of a given crop is higher in the successive aggregation of the crop model indicators:

$$Y_{r,N,c,y} = \sum_{i}^{n} \frac{A_{c,y,i,N+1}}{A_{c,y,r,N}} \; \forall \; (i, N+1) \; \in \; (r, N) \qquad \text{Eq. 3}$$

where $Y_{r,N,c,y}$ is the model indicator for crop c aggregated for the region r at NUTS (or GAUL) level N and $A_{c,y,i,N+1,c}$ is the area of crop c in year y for region i at NUTS level $N+1$ belonging to region r, obtained from official statistics. The weight of each region i in the aggregation is given by the term $\frac{A_{c,y,i,N+1}}{A_{c,y,r,N}}$, and is established on a yearly basis, according to the crop area statistics available.

Calculating the term $\frac{A_{c,y,i,N+1}}{A_{c,y,r,N}}$ is not starightfoward as it is primarily derived from official crop area statistics. Crop area statistics can be available in Europe from different sources: Eurostat data, statistical services from member states, provisional figures from DG AGRI, etc. These different sources provide data with different degree of completeness and detail: time period covered, lowest administrative level (NUTS) at which the data is available, missing data, etc. This generates a large heterogeneity in the quality of statistical data among sources, countries and crops, which has to be properly managed in order to have a unique, consistent and complete of regional crop area figures that satisfy the requirements of the scaling method (Eq. 3). These requirements are:

1. The weights $\frac{A_{c,y,i,N+1}}{A_{c,y,r,N}}$ must be available for all existing administrative units and crops in the NUTS and GAUL systems included in the European window of the MCYFS.
2. These weights have to be available for all the historical series of the BioMA-WOFOST indicators, from 1975 to the current year.
3. They have to represent, as accurately as possible, the actual distribution of crops in the European window of the MCYFS.

Once all relative weights are established, the aggregation method described in Eq. 1 to 3 can be applied to fully disaggregate official area statistics at country level to to NUTS 1, NUTS 2, NUTS 3 level and grids.

In this document, the main principles of the new aggregation areas algorithm are presented: input data, processing steps, rules for combining the different data sources.

2 Principles and procedures of the aggregation areas algorithm

2.1 Input data

2.1.1 Crop statistics

Three statistical sources are used as input for the aggregation areas algorithm, collected, where available, for the following crops: soft wheat, durum wheat, total wheat, winter barley, spring barley, total barley, triticale, rye, oats, soybean, grain maize, silage maize, rice, sugar beet, potato, rapeseed and sunflower.

i. **Regional area statistics from Eurostat (Table REGIO). Eurostat delivers to AGRI4CAST periodically** regional and country-level statistics coming from two different databases: the country-level database[2], providing figures for a list of agricultural commodities available at EU-28; and the regional database[3], which contains data at subnational level (NUTS 1 and NUTS 2). In the regional database, nevertheless, missing data can be largely found for some crops and regions. These statistics are stored in the MCYFS DB in a specific table called REGIO and cover the period 1975-current year. The availability of historical area statistics varies largely from one country to another.

ii. **Regional statistics from national statistical services (table NATIONAL_STATS).** An additional database of agricultural time series, covering in most of the countries the period from 1998 to 2012. This statistics were collected contacting the different national statistics services (Table 1), which provided crop area statistics at the lowest administrative level available (e.g. NUTS3) for these crops: soft wheat, durum wheat, spring barley, winter barley, grain maize, rye, oats, triticale, potato, sugar beets, rape and turnip rape, sunflower, rice soybean. This dataset was used to complete the Eurostat statistics, which in most cases only report yield and area figures up to NUTS 1 or NUTS 2 level, presenting sometimes numerous missing values. Regional statistical were also collected from some of the EU neighbourhood (Ukraine, Maghreb countries, Turkey). All these statistics are stored in a separate table of the MCYFS named NATIONAL_STATS.

iii. **Country-level statistics provided by DG-AGRI (table CHRONOS).** DG-AGRI provides to AGRI4CAST during the growing season provisional figures coming from member states. These statistics, stored in the table CHRONO, are used only for the current season as a back-up when REGIO statistics are not yet available (e.g. the growing season has not yet finished, statistics are not consolidated, etc.).

Table 1. National source and administrative level provided of the statistics collected in the EU-28 members and neighbouring countries.

	Country	Administra-tive level	National source
EU member states	Cyprus	NUTS3	Statistical Service of Cyprus
	Czech Republic	NUTS3	Czech Statistical Office
	Denmark	NUTS3	Statistics Denmark
	Estonia	NUTS3	Statistics Estonia
	Finland	NUTS3	Ministry of Agriculture and Forestry
	France	NUTS3	Ministère de l'Agriculture, de l'Agroalimentaire et de la Forêt
	Germany	NUTS3	Statistical offices of the Länder and the Federal Statistical Office
	Greece	NUTS3	National Statistical Service of Greece

[2] Database: Agriculture/Agricultural production/Crop products/Crop Statistics (apro_acs)
[3] Database: Agriculture/Regional agriculture statistics/Crop statistics by NUTS 2 regions (agr_r_acs)

	Country	Administra-tive level	National source
	Hungary	NUTS3	Hungarian Central Statistical Office
	Ireland	NUTS3	Central Statistics Office
	Italy	NUTS3	Italian National Statistical Institute
	Latvia	NUTS3	Central Statistics Authority database
	Lithuania	NUTS3	Agriculture and Environment. Statistics Lithuania
	Romania	NUTS3	National Institute of Statistics
	Slovakia	NUTS3	Statistical Office of the Slovak Republic
	Spain	NUTS3	Ministry of Agriculture, Food and Environment
	Sweden	NUTS3	Official Statistics of Sweden
	UK-England	NUTS3	Ministerial Department for Environment Food & Rural Affairs
	UK-Scotland	NUTS3	Rural & Environment Science & Analytical Services. Scottish Government
	UK-rest of UK	NUTS1	Ministerial Department for Environment Food & Rural Affairs
	Croatia	NUTS2	Croatian Bureau of Statistics
	Netherlands	NUTS2	Statistics Netherlands
	Poland	NUTS2	Central Statistical Office
	Portugal	NUTS2	Statistics Portugal, Instituto Nacional de Estadistica
	Austria	NUTS2	Statistics Austria
Others	Morocco	GAUL2	L'Institut National de la Recherche Agronomique (INRA)
	Ukraine	GAUL2	State Statistics Service of Ukraine
	Algeria	GAUL1	La Direction des statistiques Agricoles et des Systèmes d'Information
	Turkey	GAUL1	Turkish Statistical Institute

2.1.2 Land cover data

In addition to crop statistics, land cover data is used as back-up to calculate regional weights where any statistical source is available. Two land cover maps are used (see Section **Error! Reference source not found.**):

- Corine Land Cover map (Buttner et al., 2004) used to generate the land cover class 'Arable land', which is generated as the union of the original Corine classes number 211 and 212, corresponding to rain-fed and permanently irrigated arable land. The 'Arable land' class was used to establish the regional weights, when necessary, for all crops in the EU countries except rice, for which the original class number 213 names 'Rice fields' in the Corine map was used instead.
- For the EU neighbourhood weights were established for all crops using "Cultivated and managed areas" –class number 116– from the GLC2000 Global land cover map (Bartholomé and Belward, 2005).

2.2 Description of the algorithm

As mentioned before, the regional data on crop area collected from Eurostat and national statistical services (Section 2.1.1) are very heterogeneous. The administrative level at which statistics are provided (NUTS 1, NUTS 2 or NUTS 3) varies among the different sources, and

even within the same source, as the statistical offices from the member states often produce regional statistics disaggregated at a different administrative level. Moreover, the statistics often have missing values that have to be properly filled in order to have a consistent and complete statistical time-series. The algorithm implemented tries to solve the mentioned issues in different steps to generate a coherent, complete dataset of regional weights ($\frac{A_{c,y,i,N+1}}{A_{c,y,r,N}}$in Eq. 3), following the workflow shown in Figure 3.

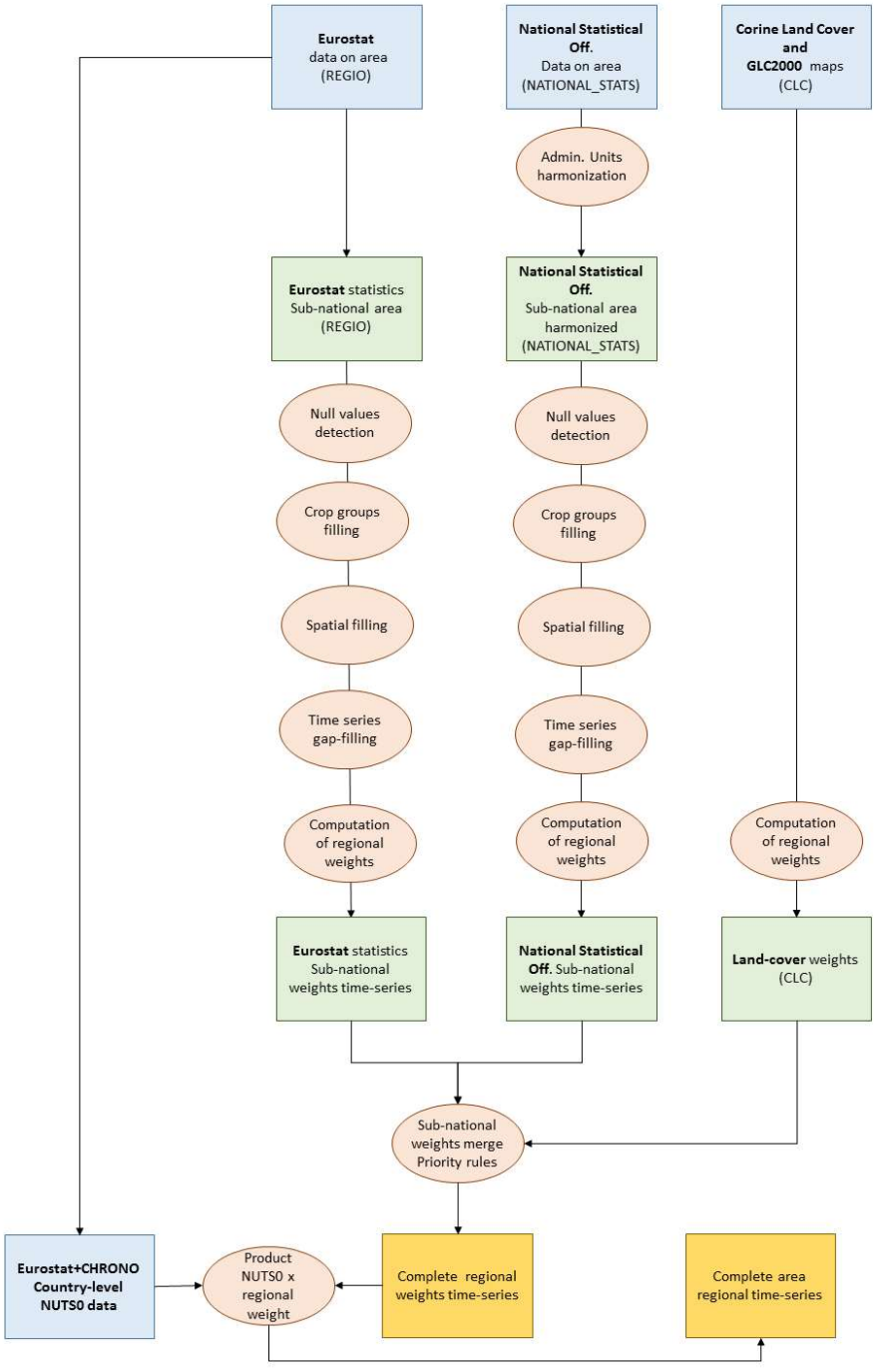


Figure 3. Aggregation algorithm workflow to generate a complete dataset of regional weights and absolute area figures from statistical data and land cover maps.

The overall principle of the algorithm relies on these basic rules:

i. a group of regions belonging to the same region of the immediate higher hierarchical level (e.g. NUTS regions ES241, ES242 and ES243, all belonging to ES24) constitute an indivisible group of regions in which all regional weights are computed from a single statistical source and must sum 1. The mentioned rule applies to every administrative level (province, region, macro-region).
ii. The source of the area statistics used to compute the regional weights for a given crop on a given year must be the same for every group of regions, to ensure the consistency of the regional weights calculated.
iii. From rule ii, a statistical source (Eurostat, national statistics) must provide a value different from null for all the regions in the same group to be considered as candidate to construct regional weights for that group for a given crop and year.
iv. The possible null values of the original statistical time-series for any group of regions can be filled using indirect methods (e.g. temporal gap-filling, aggregation of regions, etc.). Gap-filling methods have to be rather conservative, and involve data only coming from a single statistical source, avoiding to generate inconsistencies in the regional weights linked to the possible discrepancies between statistical data sources.
v. When there are different sources candidate to establish regional weights for a given group of regions in a given year and crop (e.g. Eurostat, national stats, land cover data) a single one will be selected based on different quality criteria (merging rule). A source for regional weight is guaranteed, as land cover data is always available.

These principles permit to coherence coherence in the regional weights for every group of regions, as they are established from the same source and method. At the same time, these rules allow to combine different statistical sources to generate the complete network of regional weights in a robust way –given that regional weights are independent from one group or the other– avoiding possible inconsistencies in the absolute data among different sources.

When relative weights are established for all administrative levels NUTS 1 to NUTS 3 across all years and all crops, computing absolute area data for all regions is straightforward multiplying successively NUTS 0 level area statistics (in most of cases available and reliable) by the relative weights of NUTS 1 regions, the NUTS 2 and finally NUTS 3. In such way, the resulting absolute regional crop area values will be always coherent between the different hierarchical levels and also with the country-level statistics. The processing steps of the aggregation areas algorithm (Figure 3) are explained in the rest of this Section.

2.2.1 Harmonization of administrative units

The base layer for the administrative sub-division of EU member states selected for this assessment was the Nomenclature of Territorial Units for Statistics (NUTS, Eurostat). In some EU countries there were changes in the boundaries, and also the names of some regions in the last 10 years have changed, producing a mismatch between the NUTS classification and some administrative units used by the member states when providing statistical data. In order to make them match with the European official standards considering for the project, a harmonization is necessary. When two or more different administrative regions included in national offices statistics appear joined in the NUTS classification, their values are aggregated summing areas of the individual regions affected, and a weighted sum of yields based on areas.

That aggregation is applied, for instance, to several administrative regions in Germany and Sweden, which were grouped between 2003 and 2008, but also to two groups of administrative units in Italy and Romania.

By contrast, statistics for some regions have to be disaggregated to smaller administrative units. When disaggregating area statistics, the crop area for the new, smaller region is calculated multiplying the crop area of the old region by the proportion of the arable land area from that old region –from the Corine land cover map– contained in the new small one.

Disaggregation is necessary in some administrative units of Ireland and the UK from 1998 to 2015, and also in Finland since the crops statistics are provided in Ely-Center regions, which are Finnish administrative units sometimes including more than one NUTS3 region.

2.2.2 Detection of null values

One of the critical points when processing crop statistics is differencing between zeroes and null values. By default, when processing numerically the statistical data all the values where crop area or production is not higher than 0 are considered 0, however some of them are actually null data, which have to be identified. This distinction is important because only null values – data not available– on crop area and production will be filled by the algorithm in successive steps. In both sources of statistical data (Eurostat and national statistical services) the differentiation between zeroes and null is not systematic, and some rules have to be applied. Two rules are applied at all the regional hierarchical levels to mark as null:

$$A_{c,y,i,N}\ is\ null\ if\ R_{i,N} = 0\ and\ \sum_{j}^{n} A_{c,y,j,N_k} > 0\ \forall\ (j, N_k = N+1, \dots, 3)\ \in (i, N) \qquad \text{Eq. 4}$$

$$\forall\ (j, N_k = N+1, \dots, 3)\ \in (i, N)\ are\ null\ \ if\ \sum_{j}^{n} A_{c,y,j,N_k} = 0\ \ and\ A_{c,y,i,N} > 0 \qquad \text{Eq. 5}$$

In Eq. 4 the area A for crop c, year y and given region i at NUTS level N is null if its value is 0 and the sum of all the regions j in a group of regions at a given lower NUTS level $(N+1, \dots, 3)$ belonging to region i is higher than 0. Eq. 5 is the opposite to Eq. 4: the administrative regions j of the same group of regions at a given lower NUTS level $(N+1, \dots, 3)$ belonging to region i at level N are all null if they sum 0 and the value for region i is higher than 0.

2.2.3 Crop groups filling

Total wheat and total barley are composed crop classes, that are generated from other ones. Total wheat is composed as the sum of durum wheat and soft wheat; total barley is composed summing spring and winter barley. In this processing step, applied separately for each statistical source, if values for a single crop in any of the two groups are null, they are deduced from the available ones.

For instance, if total barley area is null for a given region and year and statistical source, but winter barley and spring barley are available then total barley is filled summing the other two. Similarly, if total barley and spring barley are available, winter barley is obtained as the difference of them.

2.2.4 Spatial filling

This is applied when a null exists in $A_{c,y,i,N}$ and the area for a crop c and year y is known for all the administrative regions of the immediate lower administrative level $(N+1)$ belonging to it . If that is the case, the area for the higher region is calculated aggregating the values of the smaller regions:

$$A_{c,y,i,N} = \sum_{j}^{n} A_{c,y,j,N+1}\ \forall\ (j, N+1)\ \in (i, N) \qquad \text{Eq. 6}$$

2.2.5 Time series filling

when at least for a given crop/region one value in the historical series is not null. If this is the case, the method handles two different cases:

Crop/regions in which the number of available and non-zero values in the historical set is greater than 60%, a linear regression using year y and area as predicted value is used to fill the null values.

$$A_{c,y,i,N} = ay + b \qquad \text{Eq. 7}$$

This method permits to capture possible trends in crop area or production when filling gaps in time-series (a and b are the regression empirical constants). If there are years with null value before first valid one of the historical series, the latter is used to fill missing data. Similarly, null values posterior to the last available one are filled with it, too. This prevents from retrieving values that are out of range (e.g. area < 0) when extrapolating $A_{c,y,i,N}$.

If available data is less than 60% of the historical series, the value for a given year is calculated as the average of the existing values for that region/crop combination in the period 1998-2015.

2.2.6 Weights generation

This step is applied separately for each statistical source. For every crop, year, indicator (area or production) and group of regions, when all the values for $A_{c,y,j,N}$ are known, their weights are calculated following the expression:

$$W_{c,y,j,N} = \frac{A_{c,y,j,N}}{A_{c,y,i,N-1}} \; \forall \; (j,N) \; \in (i,N-1) \qquad \text{Eq. 8}$$

where $W_{c,y,j,N+1}$ is the calculated weight, relative to $A_{i,N}$ for crop c in year y. If, for a given source of statistics, one or more regions j have a null value –not filled in the previous step– the weights cannot be calculated and thus all $W_{c,y,j,N+1}$ are set to null. When the relative weight of all administrative units at all hierarchical NUTS level are calculated, the weight of every individual region with respect to the country figures can be obtained by multiplying its weight by the weights of the administrative units above in the hierarchy:

$$W'_{c,y,j,N} = W_{c,y,j,N} * W_{c,y,i,N-1} * \ldots * W_{c,y,a,N=1} \qquad \text{Eq. 9}$$

In addition to statistical data from Eurostat and national statistical offices data, weights are also established based on land cover data from Corine Land Cover map (CLC) as a back-up where any of the statistical sources is available or a group of regions. Land cover weights represent the area weight of a given land cover L in a specific region i, at hierarchical level N to the total area of L in the country region i belongs to.

2.2.7 Statistical sources merging

In this final step, a source of relative weights is selected from the three sources available (Eurostat data, national statistical services data, and land cover data) for every crop, year and group of regions (e.g. the weights for all regions in the same group are established with the same source).

Among the three sources, two of them –Eurostat and national statistical services data– contain time-dependent values reporting the yearly crop area, whereas land cover is invariant over time. The current version of the algorithm considers the time-dependent sources are more accurate and uses the land cover information as backup for filling the null values of the two statistical datasets.

Currently, the sources merging procedure is straightforward, and can be further developed in the future, and consists in the following rules:

1. If the sum of weights for a given group of regions is equal to 0 in Eurostat, the use National statistics and vice versa. These permits to reject a source where possible gaps in specific groups of regions were not filled in the temporal, spatial and crop groups filling procedures.
2. Data availability from the two statistical sources (Eurostat and National statistics) and privilege original against gap-filled data. This rule evaluates, for a group of regions, the number of original data (including original non-null statistical values but also those filled by the spatial and crop group filling procedures) against those filled by the temporal filling. The source with the highest number of regions in the group filled with original data.
3. If both the sources contain the same amount of original data, the Eurostat is selected for that crop, year and group of regions.
4. If data from both statistical sources is null, the land cover weights are used.

3 Outputs from the aggregation areas algorithm

3.1 Regional crop areas

A .csv file with the regional crop areas for all administrative units, calculated multiplying country-level statistics reported by DG-AGRI (source: CHRONOS, see 2.1.1) by the regional weights calculated using the aggregation areas algorithm can be downloaded from the MARS data portal.

Those countries not reporting, for a given crop and year, an area figure will not appear will not appear. Please, consider that some countries may not produce separately soft wheat and durum wheat figures, or winter barley and spring barley, and hence only regional areas for total wheat and total barley will be available for these countries. For instance, soft wheat areas are not available for Ukraine, only total wheat areas, because the statistical data services from Ukraine do not distinguish between soft and durum. For the same reason, only total barley figures are available for Maghreb countries. The structure of the downloadable .csv file is given in Table 2.

Table 2. Structure of the .csv table donwloadable from the MARS data portal containing the regional crop areas.

COLUMN NAME	TYPE NAME	DESCRIPTION
CROP_NO	NUMBER(4)	MCYFS Crop number, this crop identifier is also part of the table primary key
CRP_NAME	VARCHAR2(100)	Crop name
REG_CODE	NUMBER(10)	Regional ID, regional code used in the MCYFS and part of the table primary key
REG_NAME	VARCHAR2(100)	Region name
YEAR	NUMBER(4)	Year of the aggregation area value, part of the table primary key
AREA_CULTIVATED	NUMBER(*,3)	Regional crop area, in hectares
SOURCE	VARCHAR2(100)	Source of the regional are value: Eurostat, national statistics or land cover
TREATMENT	VARCHAR2(255)	Operations applied to the source data (original, or filled by the aggregation algorithm)

3.2 Continuous distribution maps of harvested area

The distribution maps show the expected proportion of total land area occupied by a given crop, and have a spatial distribution of 25 km, as they are generated using the grid layer of the MCYFS. To compute that expected proportion for a given crop and year the harvested area of all the NUTS 3 regions in Europe have to be calculated, multiplying the absolute weight of every NUTS 3 region in the total country area ($W'_{c,y,j,3}$ from Eq. 9) by the crop harvested area of the country that NUTS 3 region belongs to:

$$A_{c,y,j,3} = W'_{c,y,j,3} * A_{c,y,i,0} \text{ where } j \in i \qquad \text{Eq. 10}$$

where $A_{j,3}$ is the NUTS 3 area for a given crop c and year y, calculated after the aggregation areas algorithm. If the regional weights calculated by aggregation areas algorithm, the value for $A_{c,y,j,3}$ should be very close to the actual harvested area in that region. $A_{c,y,i,0}$ country area figures are always coming from CHRONO statistics, delivered by DG-AGRI (see 2.1.1). The expected

area on a given grid is disaggregated from NUTS 3 areas and land cover data following the expression:

$$A_{c,y,g} = \sum_{j}^{n} \frac{A_{c,y,j,3} * A_{g \cap l}}{A_{j,3 \cap l}} \; \forall \, j \; where \; A_{j \cap g} > 0 \qquad \text{Eq. 11}$$

Where $A_{g \cap l}$ is the area of intersection between grid g and land cover l (arable land, for all crops, except rice), $A_{j,3 \cap l}$ is the area of the intersection between NUTS 3 region j and land cover class l. Finally, the expected proportion of crop area in grid g is calculated as:

$$R_{c,y,g} = \frac{A_{c,y,g}}{A_g} \qquad \text{Eq. 12}$$

where A_g is total land area in grid g.

A .csv file is generated by the data portal containing A_g, $A_{c,y,g}$ and $R_{c,y,g}$ (Table 3). Linking the data from this file to the MARS grid shapefile (.shp) with the , year crop distribution maps can be generated for EU and neighbouring countries. Figure 4 and Figure 5 show the harvested area distribution maps for the following crops: soft wheat, durum wheat, total wheat (durum + soft), winter barley, spring barley, total barley (winter + spring), triticale, rye, rapeseed, grain maize, green maize, rice, sunflower, soybean, sugar beet and potato. Especially in the case of winter cereals, the relevance of the main European producing areas can be appreciated in the maps: centre-north of France, the east of the UK, north-west of Spain, south of Italy, Germany, the Black Sea countries, north of Maghreb countries, etc. The maps permit as well to identify areas of high concentration of a given crop at the EU level, e.g. south of Italy and eastern Algeria for durum wheat, Romania, north Italy and south of France for grain maize, or the main rice areas of Europe.

Table 3. Structure of the .csv table donwloadable from the MARS data portal containing the 25 km grid data on crop areas.

COLUMN NAME	TYPE NAME	DESCRIPTION
CROP_NO	NUMBER(4)	MCYFS Crop number, this crop identifier is also part of the table primary key
CRP_NAME	VARCHAR2(100)	Crop name
GRID_CODE	NUMBER(10)	Regional ID, regional code used in the MCYFS and part of the table primary key
GRID_AREA	NUMBER(*,3)	Total land area of the grid (A_g), expressed in hectares
REG_NAME	VARCHAR2(100)	Region name
YEAR	NUMBER(4)	Year of the aggregation area value, part of the table primary key
CROP_AREA_EXPECTED	NUMBER(*,3)	Expected crop area in the grid after disaggregating regional crop area using land cover ($A_{c,y,g}$) calculated from Eq. 11 and expressed in hectares
CROP_RATIO_EXPECTED	NUMBER(*,3)	Ratio of expected crop area and total grid area ($R_{c,y,g}$), calculated from Eq. 12.
LANDCOVER	VARCHAR2(100)	Land cover class used in Eq. 11 to disaggregate regional areas to grid areas

TREATMENT	VARCHAR2(255)	Operations applied to the source data (original, or filled by the aggregation algorithm)

The proportion of crop harvested area, $R_{c,y,g}$, is an estimation based on regional statistics and land cover data. Eq. 11 assumes that distribution of the area for a given crop is homogeneous over the land cover class l in NUTS 3 regions. Therefore, the reliability of the estimated $R_{c,y,g}$ depends on how much that assumption is valid across the different NUTS 3 regions in Europe. Rice is the most favourable case, as a specific land cover class is used in the disaggregation, and the distribution of actual rice area within that class can be considered homogeneous. In the rest of crops, this hypothesis has to be verified. In the case of southern Europe, where summer crops (e.g. grain maize, sugar beet, potato) are permanently irrigated, assuming they are distributed equally across all arable land constitutes a limitation.

The accuracy of the disaggregation depends also greatly from the detail of the statistical data available. For those crops/country combinations where statistics are available at NUTS 3 level the reliability of the grid areas would be much higher than those where statistics are only available at NUTS 1 or NUTS 2 and the regional weights at NUTS 3 have been derived using land cover.

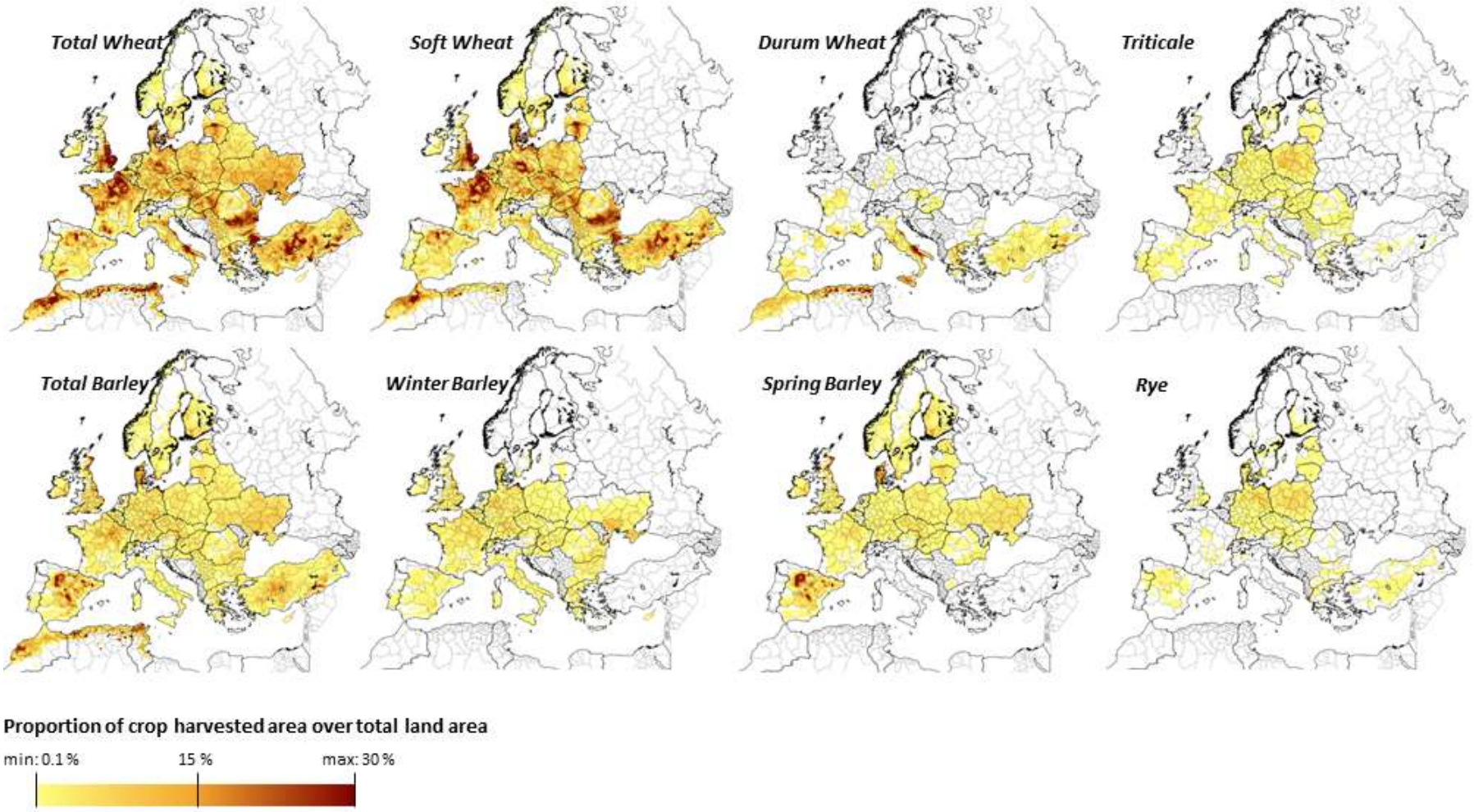


Figure 4. Distribution of the harvested area of wheat (durum, soft and total), barley (winter, spring and total), rye and triticale in the EU and neighbouring countries calculated as an output of the aggregation areas algorithm disaggregated over 25 km grid.

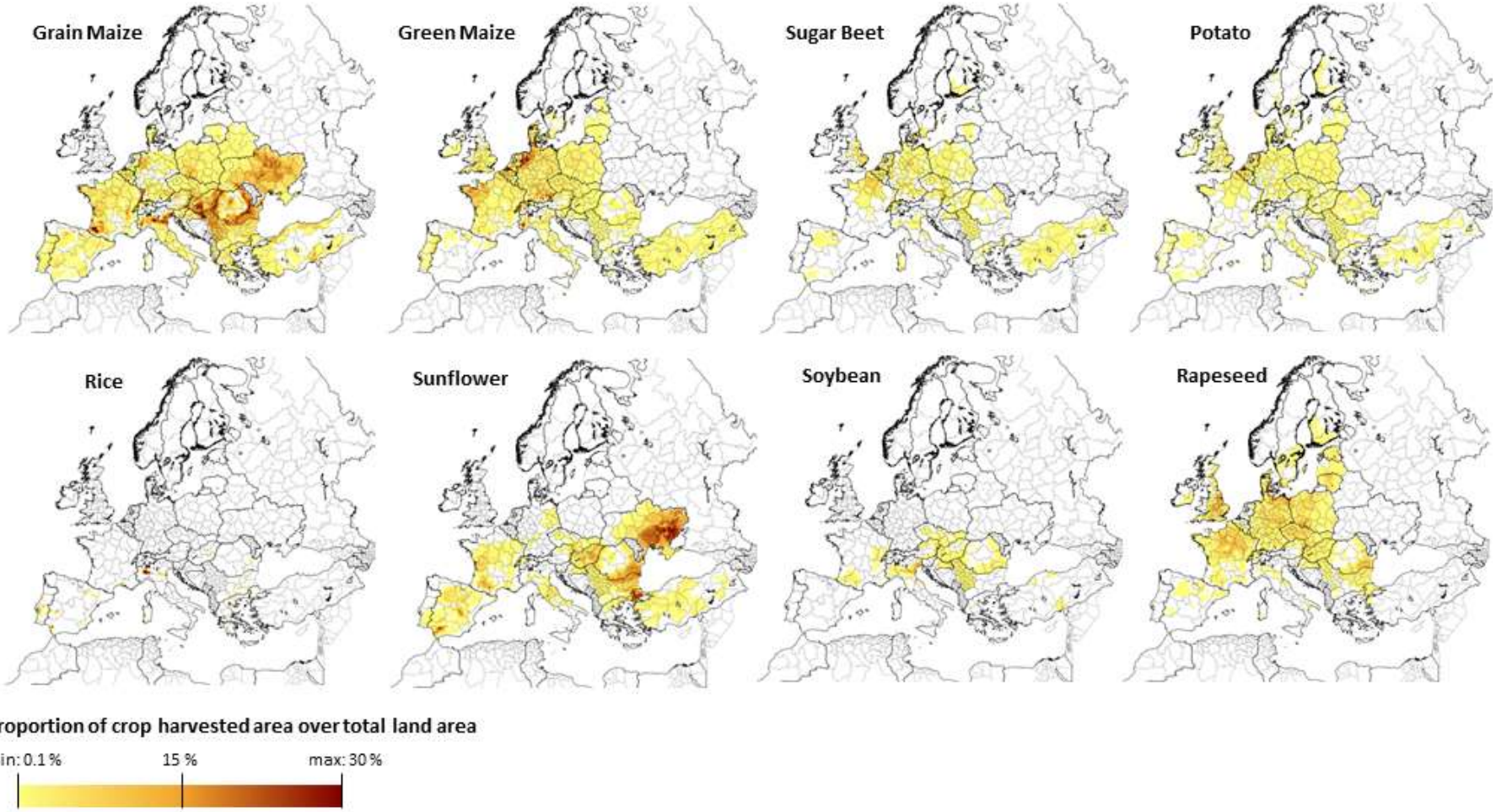


Figure 5. Distribution of the harvested area of maize (grain and green), sugar beet, potato, rice, sunflower, soybean and rapeseed in the EU and neighbouring countries calculated as an output of the aggregation areas algorithm disaggregated over 25 km grid.

List of figures

List of tables

EN-C

JRC Mission

As the science and knowledge service of the European Commission, the Joint Research Centre's mission is to support EU policies with independent evidence throughout the whole policy cycle.

EU Science Hub
ec.europa.eu/jrc

@EU_ScienceHub

EU Science Hub - Joint Research Centre

Joint Research Centre

EU Science Hub

doi:xx.xxxx/xxxx

ISBN xxx-xx-xx-xxxxx-x

Publications Office