



European
Commission

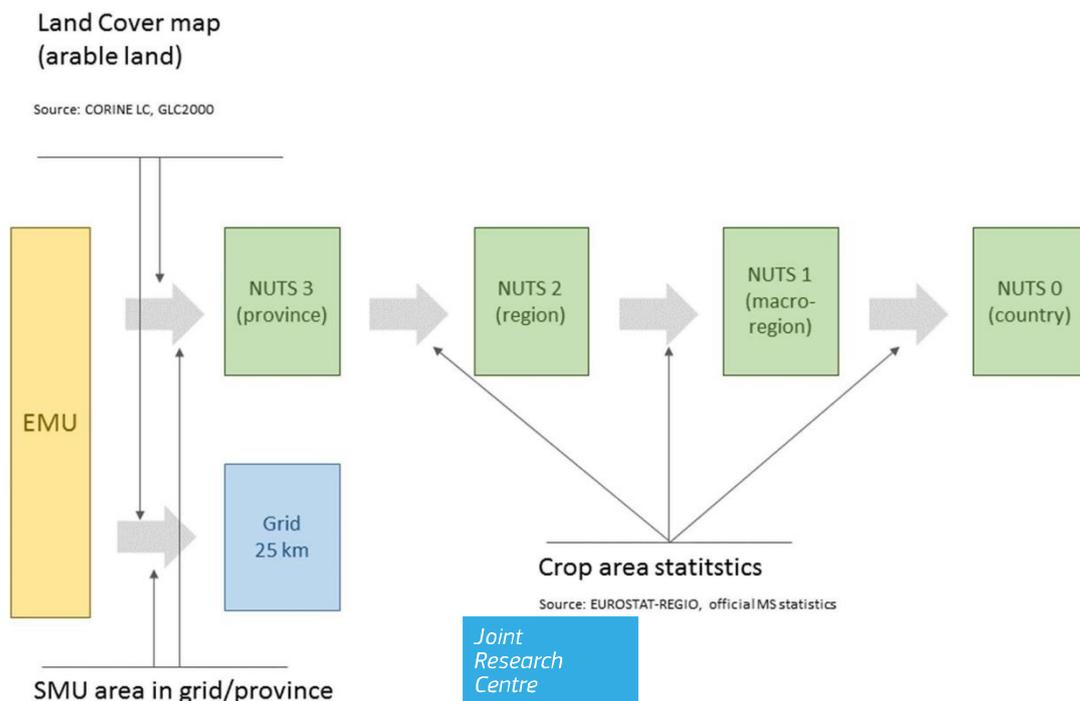
JRC TECHNICAL REPORT

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

Update 2022

Cerrani, I., Ronchetti, G., Nisini
Scacchiafichi, L., López Lozano, R., Van der
Velde, M.

2022



This publication is a Technical report by the Joint Research Centre (JRC), the European Commission's science and knowledge service. It aims to provide evidence-based scientific support to the European policymaking process. The scientific output expressed does not imply a policy position of the European Commission. Neither the European Commission nor any person acting on behalf of the Commission is responsible for the use that might be made of this publication. For information on the methodology and quality underlying the data used in this publication for which the source is neither Eurostat nor other Commission services, users should contact the referenced source. The designations employed and the presentation of material on the maps do not imply the expression of any opinion whatsoever on the part of the European Union concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries.

EU Science Hub
<https://ec.europa.eu/jrc>

JRCXXXXXX

PDF	ISBN XXX-XX-XX-XXXXX-X	ISSN XXXX-XXXX	doi:XX.XXXX/XXXXXX
Print	ISBN XXX-XX-XX-XXXXX-X	ISSN XXXX-XXXX	doi:XX.XXXX/XXXXXX

Ispra: European Commission, 2022
© European Union, 2022



The reuse policy of the European Commission is implemented by the Commission Decision 2011/833/EU of 12 December 2011 on the reuse of Commission documents (OJ L 330, 14.12.2011, p. 39). Except otherwise noted, the reuse of this document is authorised under the Creative Commons Attribution 4.0 International (CC BY 4.0) licence (<https://creativecommons.org/licenses/by/4.0/>). This means that reuse is allowed provided appropriate credit is given and any changes are indicated.

For any use or reproduction of photos or other material that is not owned by the EU, permission must be sought directly from the copyright holders.

How to cite this report: Cerrani, I., Ronchetti, G., Nisini Scacchiafichi, L., López Lozano, R., Van der Velde, M. *Algorithm for the disaggregation of crop area statistics in the MARS crop yield forecasting system*. Ispra: European Commission, 2022, JRCXXXXXX.

Contents

- Abstract 1
- 1 Introduction..... 2
- 2 Principles and procedures of the aggregation areas algorithm 5
 - 2.1 Input data..... 5
 - 2.1.1 Crop statistics..... 5
 - 2.1.2 Land cover data..... 6
 - 2.2 Description of the algorithm..... 7
 - 2.2.1 Harmonization of administrative units..... 9
 - 2.2.2 Detection of null values..... 9
 - 2.2.3 Crop groups filling..... 9
 - 2.2.4 Spatial filling..... 10
 - 2.2.5 Time-series filling..... 10
 - 2.2.6 Weights generation..... 10
 - 2.2.7 Statistical sources merging..... 11
- 3 Outputs from the aggregation areas algorithm..... 12
 - 3.1 Regional crop areas..... 12
 - 3.2 Continuous distribution maps of harvested area..... 13
- 4 Conclusions..... 17
- References..... 18
- List of abbreviations and definitions..... 19
- List of figures 20
- List of tables 21

Abstract

This report describes the algorithm to disaggregate regional crop area statistics in Europe, at all NUTS and GAUL administrative levels and onto a regular grid of 25 km. This algorithm is necessary to aggregate 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 these indicators operationally for crop yield forecasting (e.g. MARS Bulletins).

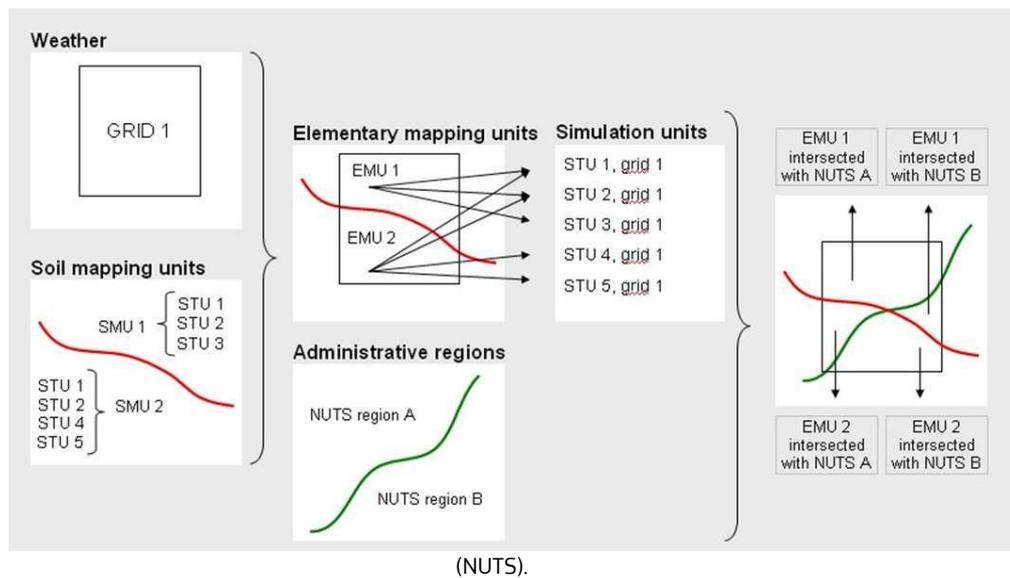
The algorithm uses as input crop area statistics from Eurostat and national statistical services delivered from the Member States, and land cover. Rules and operations are set to construct, combining the mentioned sources, the full hierarchical system of regional weights per crop and year, expressing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

The introduction (Section 1) describes the spatial framework and the main principles of indicators aggregation in use in the MCYFS. In Section 2, the processing steps of the aggregation are explained in depth. Section 3 reports the outputs of this algorithm that can be downloaded from the Agri4cast Resources Portal.

1 Introduction

In the MARS Crop Yield Forecasting System (MCYFS, Van der Velde et al., 2019) 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 4.0¹) 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



Source: JRC.

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).

⁽¹⁾ http://eusoils.jrc.ec.europa.eu/ESDB_Archive/ESDBv2/index.htm

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

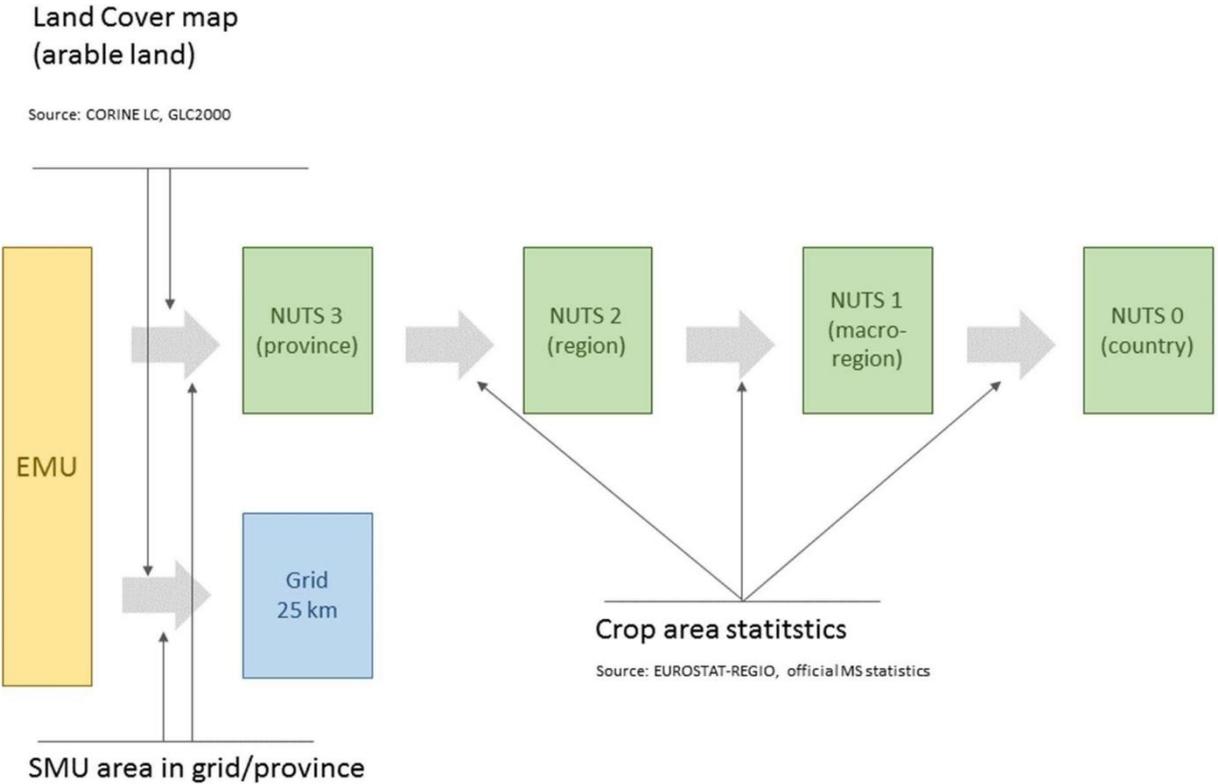
$$W_{i,g} = \frac{A_{insnl}}{A_{gnsnl}} \tag{Eq. 1}$$

$$Y_{g,y,c} = \sum_i^n W_{i,g} * Y_{c,y,i} \forall i \in g \tag{Eq. 2}$$

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_{insnl} is the area intersection of EMU i with land cover class l and suitable soil s ; and A_{gnsnl} 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 (Soukup et al., 2016) –classes number 211 and 212, respectively– within EU countries. For the EU neighbouring countries, 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 administrative level, calculating analogously $W_{i,r}$ the relative weight of EMU on NUTS 3 region r . 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).



Source: JRC.

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:

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

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 straightforward 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:

- 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.
- These weights have to be available for all the historical series of the BioMA-WOFOST indicators, from 1980 to the current year.
- 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 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, soybean, grain maize, silage maize, rice, sugar beet, potato, rapeseed and sunflower (Eurostat, 2020).

- **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 database², providing figures for a list of agricultural commodities available at EU27; and the regional database³, 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 1980-current year. The availability of historical area statistics varies largely from one country to another.
- **Regional statistics from national statistical services (table NATIONAL_STATS).** An additional database of agricultural time series, covering in not a homogeneous way the period from 1980 to 2021. 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, , 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 neighbourhoods (Ukraine, Maghreb countries, Turkey). All these statistics are stored in a separate table of the MCYFS named NATIONAL_STATS.
- **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.).

⁽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)

Table 1. National source and administrative level provided of the statistics collected in the EU27 members.

Country	Administrative level	National source
Austria	Nuts2	Statistics Austria
Belgium	NUTS2	Belgian statistical office
Bulgaria	NUTS2	National statistical institute
Cyprus	NUTS3	National statistical service
Croatia	NUTS2	Croatian Bureau of Statistics
Czech Republic	NUTS3	Czech statistical office
Germany	NUTS3	Statistical offices of the Länder and the Federal Statistical Office
Denmark	NUTS3	National statistical office
Eesti	NUTS3	National Statistical office
Greece	NUTS3	Hellenic Statistical Authority
Spain	NUTS3	Ministry of Agriculture, Food and Environment
Finland	NUTS3	Natural resource institute
France	NUTS3	Ministry of Agriculture
Hungary	NUTS3	Hungarian Central Statistical
Ireland	NUTS2	Central statistics office
Latvia	NUTS3	Ministry of Agriculture
Netherlands	NUTS2	Statistics Netherlands
Poland	NUTS2	Statistics Poland
Portugal	NUTS2	Statistics Portugal, Instituto Nacional de Estadística
Romania	NUTS3	National institute for statistics
Sweden	NUTS3	Statistics Sweden
Slovenia	NUTS2	Statistical office
Slovakia	NUTS3	Statistical Office of the Slovak Republic

Source: JRC.

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. The land cover map is Corine Land Cover map (Buttner et al., 2012) 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.

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.

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

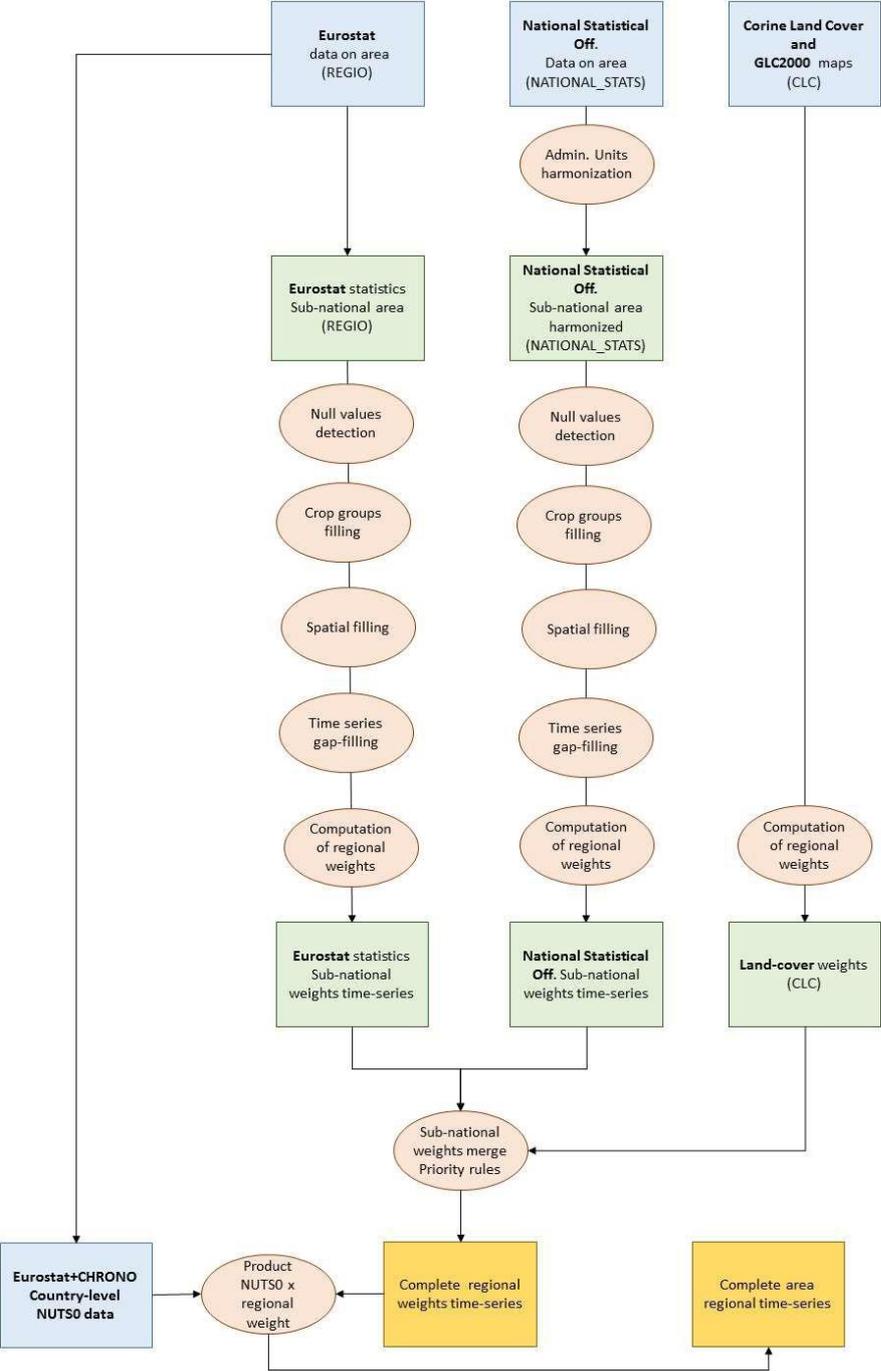
1. a group of regions belonging to the same region of the immediate higher hierarchical level (e.g. NUTS regions ES241, ES242 and ES243, all belonging to ES24) constitutes an indivisible group of regions in which all regional weights are computed from a single statistical source and must sum 1.
2. The mentioned rule applies to every administrative level (province, region, macro-region). 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.
3. From previous rules, 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.
4. 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.
5. 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 rules permit the 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.

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



Source: JRC.

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 (Eurostat, 2018). In some EU countries, there were changes in the boundaries, and also the names of some regions in the last 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 considered 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.

For example, disaggregation is necessary in some administrative units of Ireland and also in Finland, since the crops statistics are provided in ELY Centre 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 numerically processing 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, i.e. data not available, on crop area and production will be filled by the algorithm in successive steps. In both sources of statistical data (i.e. Eurostat and national statistical services) the differentiation between zeros 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:

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

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

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, ... ,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, ... ,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 by 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 step 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:

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

2.2.5 Time-series filling

This step is applied 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:

- (a) 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.
- $$A_{c,y,i,N} = ay + b \quad \text{Eq. 7}$$

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 $A_{c,y,i,N}$.

- (b) Crop/regions in which the number of available and non-zero values in the historical set is less than 60%, the value for a given year is calculated as the average of the existing values for that region/crop combination in the period 1998-2021.

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:

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

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 steps– 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 levels 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:

$$W'_{c,y,j,N} = W_{c,y,j,N} * W_{c,y,j,N-1} * \dots * W_{c,y,j,1} \quad \text{Eq. 9}$$

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 backup 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 algorithm uses National statistics and vice versa. These permit 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) shall privilege original than 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 original data is selected.
3. If both 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

Two different outputs can be derived from the described algorithm. The first output includes crop area statistics distributed on a regional basis (i.e. regional crop areas), while the second output includes crop area statistics distributed in a spatially explicit grid with 25 km x 25 km cells. This second output, in particular, is derived from the regional crop areas at NUTS 3 level by means of specific weights.

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 Agri4cast Resources Portal.

Those countries not reporting, for a given crop and year, an area figure 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. The structure of the .csv file is given in Table 2.

Table 2. Structure of the .csv file that can be download from the Agri4cast Resources Portal, including the regional crop areas.

Column Name	Type name	Description
IDCROP	NUMBER(4)	MCYFS Crop number, this crop identifier is also part of the table primary key
IDREGION	VARCHAR2(100)	Regional ID, regional code used in the MCYFS and part of the table primary key (NUTS code)
YEAR	NUMBER(4)	Year of the aggregation area value, part of the table primary key
AREA	NUMBER(*,3)	Regional crop area, in hectares

Source: JRC, Agri4cast Resources Portal.

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:

$$A_{c,y,j,3} = W'_{c,y,j,3} * A_{c,y,j,3} \quad J \in \quad \text{Eq. 10}$$

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 CHRONOS 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:

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

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:

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

where A_g is total land area in grid g .

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.

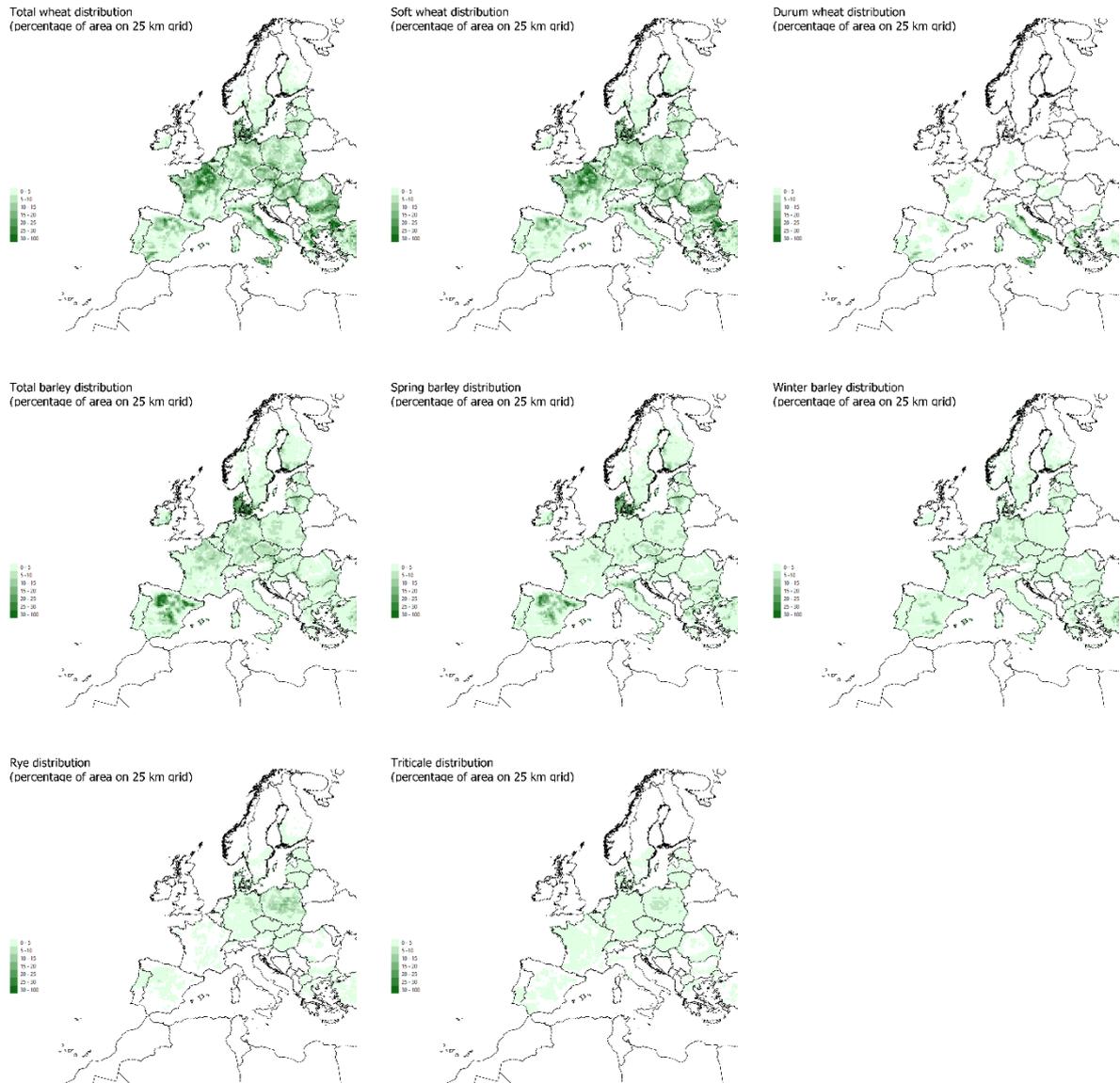
A .csv file is generated by the data portal containing A_g , $A_{c,y,g}$ and $R_{c,y,g}$ (Table 3). Linking the data from this file to the MARS grid shapefile (.shp) with the year crop distribution maps can be generated for EU countries. Figure 4 and Figure 5 show the harvested area distribution maps 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, north-west of Spain, south of Italy, Germany, the Black Sea 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 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 file that can be download from the Agri4cast Resources Portal, including the 25 km cell grid data of crop areas.

Column Name	Type name	Description
IDCROP	NUMBER(4)	MCYFS Crop number, this crop identifier is also part of the table primary key
IDGRID	NUMBER(10)	Regional ID, regional code used in the MCYFS and part of the table primary key
IDCOVER	NUMBER(4)	Land cover class used in regional areas to grid areas Eq. 11 to disaggregate regional areas to grid areas
YEAR	NUMBER(4)	Year of the aggregation area value, part of the table primary key
AREA	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
RATIO	NUMBER(*,3)	Ratio of expected crop area and total grid area ($R_{c,y,g}$), calculated from Eq. 12.

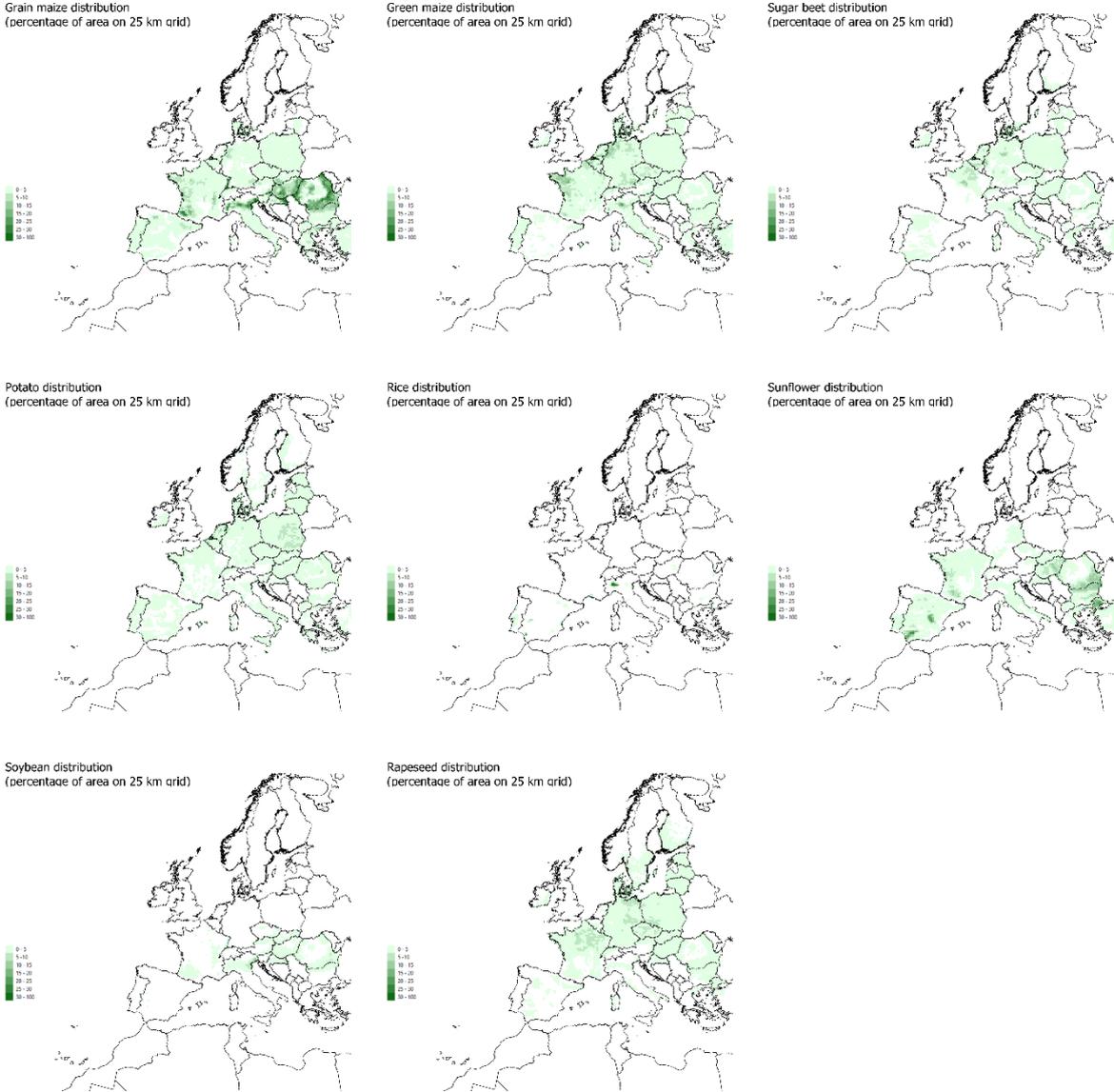
Source: JRC, Agri4cast Resources Portal.

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



Source: JRC, Agri4cast Resources Portal.

Figure 5. Distribution of the harvested area of maize (grain and green maize), 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.



Source: JRC, Agri4cast Resources Portal.

4 Conclusions

This report describes the algorithm used in the creation of a dataset containing disaggregate area statistics at 25 km cells grid for the EU countries.

Two main limitations can be identified in this algorithm:

- In case statistics are not available for a NUTS level, land cover data are used. The more this level is at high level (NUTS 1 for example) the more the generated weights tend to distribute according to the land cover, hence losing the actual crop distribution.
- The disaggregation of NUTS 3 crop area to grid crop area is always evaluated using the land cover data. This source does not change year by year as statistical values do, and in some cases there is no coherency between the two sources. Sometimes statistical sources report area values greater than the arable land area available at regional level and this leads to assign (in few cases) crop grid area greater than the total grid area.

References

Bartholomé, E. and Belward, A.S. 'GLC2000: a new approach to global land cover mapping from Earth observation data'. *International Journal of Remote Sensing*, 26(9), pp.1959-1977, 2005.

Eurostat. *Regions in the European Union, edition 2018, Nomenclature of territorial units for statistics - NUTS 2016/EU-28*. Publications Office of the European Union, Luxembourg, 2018. doi: 10.2785/475524.

Eurostat. *Annual crop statistics handbook*. Crop statistics working group, 2020.

Soukup, T., Büttner, G., Feranec, J., Hazeu, G., Jaffrain, G., Jindrova, M., Kopecky, M. and Orlitova, E. 'CORINE Land Cover 2012 (CLC2012): Analysis and Assessment'. *European Landscape Dynamics* (pp. 123-128). CRC Press, 2016

Van der Velde, M., Van Diepen, C., and Baruth, B. 'The European crop monitoring and yield forecasting system: celebrating 25 years of MARS bulletins'. *Agricultural Systems*, 168, 56-57, 2019.

List of abbreviations and definitions

BioMA	Biophysical Models Applications
CLC	Corine Land Cover
EMU	Elementary Mapping Unit
EU	European Union
GAUL	Global Administrative Unit Layers
GLC	Global Land Cover
MARS	Monitoring Agricultural ResourceS
MCYFS	MARS Crop Yield Forecasting System
NUTS	Nomenclature of territorial units for statistics
SMU	Soil Mapping Unit
STU	Soil Typological Unit
WOFOST	WOrld FOod STudies

List of figures

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). 2

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). 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. 8

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.15

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.15

List of tables

Table 1. National source and administrative level provided of the statistics collected in the EU27 members.	6
Table 2. Structure of the .csv file that can be download from the Agri4cast Resources Portal, including the regional crop areas.	12
Table 3. Structure of the .csv file that can be download from the Agri4cast Resources Portal, including the 25 km cell grid data of crop areas.	14

GETTING IN TOUCH WITH THE EU

In person

All over the European Union there are hundreds of Europe Direct centres. You can find the address of the centre nearest you online (european-union.europa.eu/contact-eu/meet-us_en).

On the phone or in writing

Europe Direct is a service that answers your questions about the European Union. You can contact this service:

- by freephone: 00 800 6 7 8 9 10 11 (certain operators may charge for these calls),
- at the following standard number: +32 22999696,
- via the following form: european-union.europa.eu/contact-eu/write-us_en.

FINDING INFORMATION ABOUT THE EU

Online

Information about the European Union in all the official languages of the EU is available on the Europa website (european-union.europa.eu).

EU publications

You can view or order EU publications at op.europa.eu/en/publications. Multiple copies of free publications can be obtained by contacting Europe Direct or your local documentation centre (european-union.europa.eu/contact-eu/meet-us_en).

EU law and related documents

For access to legal information from the EU, including all EU law since 1951 in all the official language versions, go to EUR-Lex (eur-lex.europa.eu).

Open data from the EU

The portal data.europa.eu provides access to open datasets from the EU institutions, bodies and agencies. These can be downloaded and reused for free, for both commercial and non-commercial purposes. The portal also provides access to a wealth of datasets from European countries.

The European Commission's science and knowledge service

Joint Research Centre

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



EU Science, Research and Innovation



EU Science Hub



Publications Office
of the European Union

doi:XX.XXXX/XXXXXX

ISBN XXX-XX-XX-XXXXX-X