



HAL
open science

DIVERCROP project - Deliverable 1.1: Report on the comprehensive database building

Johanna Fusco, Ricardo Villani, Michel Moulery, Tiziana Sabbatini, Leonith Hinojosa-Valencia, Claude Napoleone, Alberte Bondeau

► **To cite this version:**

Johanna Fusco, Ricardo Villani, Michel Moulery, Tiziana Sabbatini, Leonith Hinojosa-Valencia, et al.. DIVERCROP project - Deliverable 1.1: Report on the comprehensive database building. [Technical Report] Institut national de la recherche agronomique (INRA); IMBE Institut Méditerranéen de Biodiversité et d'Ecologie marine et continentale; Scuola Superiore Sant'Anna, Pisa; Aix-Marseille Sciences Economiques (AMSE). 2018. hal-01907477

HAL Id: hal-01907477

<https://hal.science/hal-01907477>

Submitted on 29 Oct 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Land system dynamics in the Mediterranean basin across scales as relevant indicator for species diversity and local food systems

Deliverable 1.1: Report on the comprehensive
database building

Johanna Fusco¹, Ricardo Villani², Michel Moulery¹, Tiziana Sabbatini², Leonith Hinojosa-Valencia³, Claude Napoleone¹, Alberte Bondeau⁴

¹ UR INRA Ecodéveloppement, Avignon

² Scuola Universitaria Superiore Sant'Anna, Pisa

³ Université Aix-Marseille AMSE

⁴ UMR CNRS IMBE, Aix-en-Provence



Table of contents

Introduction.....	3
Database structure	3
Spatial resolution.....	3
Types of information gathered and original databases	5
Temporal resolution	5
Detailed information on DIVERCROP database variables	6
Land use / Land cover.....	6
Agricultural production systems	7
MAPSPAM.....	7
Livestock	9
National Censuses	10
Bio-physical	17
Terrain	17
Bioclimatic	17
Soil type and quality	18
Soils.....	18
Population	20
Landscan data.....	20
Hyde data	20
Protected areas	20
References.....	22
Appendix.....	23
List of tables	31
List of figures	31
List of appendices.....	31

Introduction

Building a comprehensive and extensive spatial database on the various characteristics of land use, food production, socio-economy and many other dimensions of the Mediterranean basin was the first task of DIVERCROP project. The objective of this work package was to gather huge amounts of scattered data and information in a single harmonized framework, in order to provide the other work packages a solid base to perform in-depth analyses of Mediterranean land system structures and dynamics.

If many precise and interesting information are available for small parts of the Mediterranean basin, we chose to favor homogeneity by focusing on datasets covering the whole studied area. This choice intends providing a better control on the accuracy and the reliability of the results obtained after data processing, as well as building a common discourse on Mediterranean land system structures and dynamics. In order to ensure this homogenous coverage of the Mediterranean basin, global scale datasets with a fine resolution have been privileged.

Various types of data have been collected, gathered and harmonized in a postgres/postgis spatial database in order to fit each spatial resolution:

- Land use
- Agricultural production systems
- Population
- Bio-physical information

This report details all the aspects of the DIVERCROP database, including the characteristics of the original datasets and the various upscaling or downscaling processes required to harmonize them in a common spatial framework.

Database structure

Spatial resolution

After reviewing the various spatial and temporal resolutions of all available datasets including the whole extent of the Mediterranean basin, and the spatial resolutions chosen by previous Mediterranean land systems assessments (Malek, Verburg, 2017) we chose DIVERCROP database to be provided at three spatial resolutions, so that the user can choose the appropriate one following its needs (Figure 1, Table 1):

- 2 km grid (2km x 2km cells): this grid covers the Mediterranean part of the countries as chosen by Malek and Verburg (2017)
- 5 arc-minutes grid (9.25 x 9.25 km at the equator, approximately 9 x 7 km for the Mediterranean basin): this grid covers the whole countries of the Mediterranean basin, and matches the grid built by the MAPSPAM program (see p7).
- Aggregated data at regional scale: for the European part of the Mediterranean countries, we chose to aggregate data at Nuts2 scale¹; for the non-European countries, we referred to the governorates² administrative level.

¹ Spatial data on nuts2 borders downloaded on Eurostat website : <https://ec.europa.eu/eurostat/web/main/home>

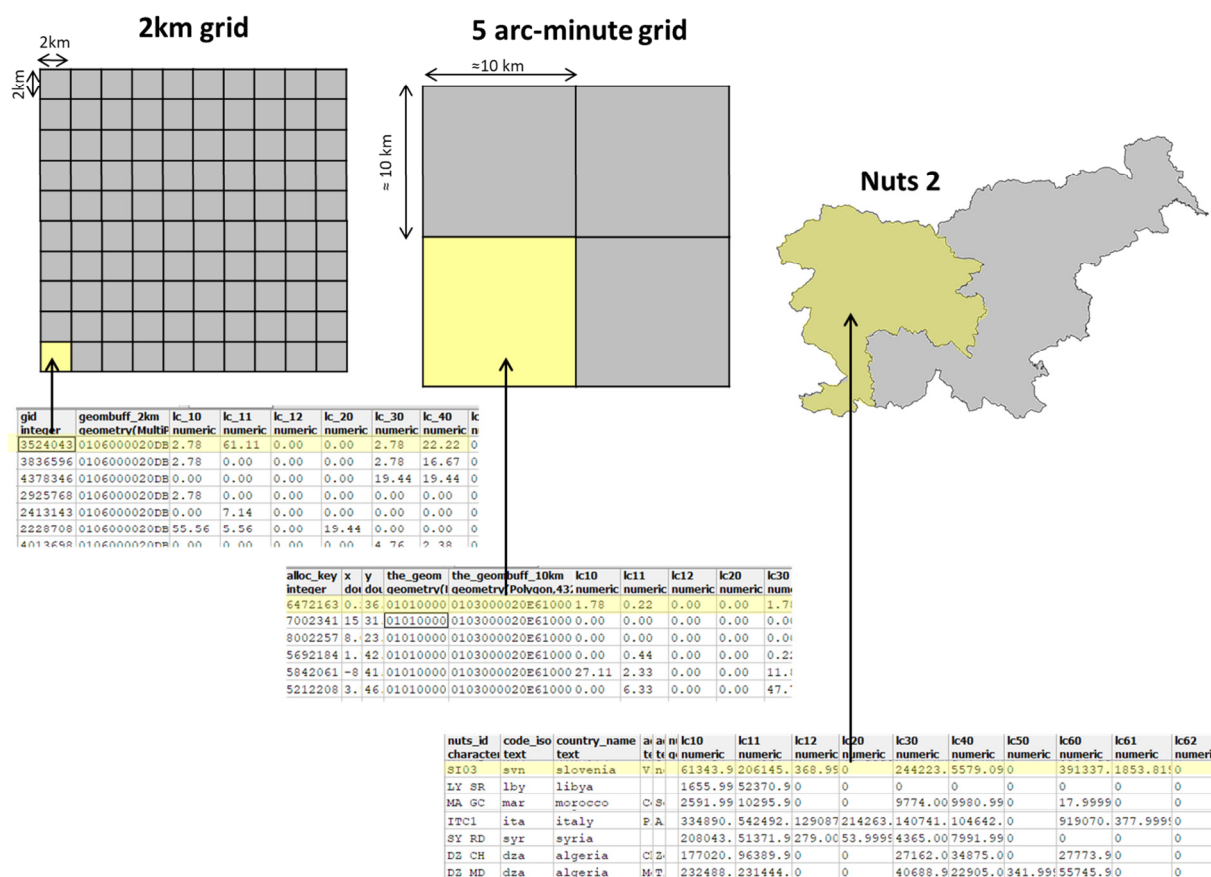


Figure 1 - The three scales and spatial units of DIVERCROP database

Table 1 - Spatial parameters of DIVERCROP multiscale database

	WKID	COORDINATE SYSTEM	PROJECTION	LINEAR UNIT	EXTENT
2 KM GRID	3035	GCS ETRS 1989	Lambert Azimuthal Equal Area	Meter	
5 ARC-MIN GRID					
NUTS 2 SPATIAL SCALE	4326	GCS WGS 1984		Degree	

² Spatial data on administrative borders for non European countries downloaded on <http://www.diva-gis.org/>

Types of information gathered and original databases

The DIVERCROP database gathers various types of information from many different sources, mostly stemming from global-scale databases. The different types of variables and their original sources are listed in Table 2.

Table 2 - Variables of DIVERCROP database and their sources

VARIABLE	ORIGINAL DATABASE	ORIGINAL DATA LINK
LAND USE / LAND COVER	ESA/CCI database	https://maps.elie.ucl.ac.be/CCI/viewer/
CROP PRODUCTION SYSTEMS	MAPSPAM	http://mapspam.info/
NUMBER OF HOLDINGS, CULTIVATED AREAS & LIVESTOCK	National censuses	See Annex 1 and the section "Construction of the Harmonized Census Database" in page 10.
LIVESTOCK	Gridded Livestock of the world	http://www.fao.org/ag/againfo/resources/en/glw/home.html
TERRAIN	Global Digital Elevation Model	https://webmap.ornl.gov/ogc/wcsdown.jsp?dg_id=10003_1
BIOCLIMATIC	WORLDCLIM	http://worldclim.org/version2
SOILS TYPES & QUALITY	Harmonized world soil database	http://webarchive.iiasa.ac.at/Research/LUC/External-World-soil-database/HTML/SoilQuality.html?sb=10
POPULATION	LANDSCAN	https://landscan.ornl.gov/
RURAL & URBAN POPULATION	HYDE	http://themasites.pbl.nl/tridion/en/themasites/hyde/index.html
PROTECTED AREAS	World database on protected areas	https://protectedplanet.net/

Temporal resolution

In order to allow the assessment of land system dynamics, two temporal spans have been chosen, according to the available data: 2005 and 2015, our intention being to add progressively other time spans in the database. Not all the information being available for these two dates, Table 3 indicates the time spans corresponding to each variable of the database.

Table 3 - Temporal resolutions on DIVERCROP database

VARIABLE	2005	2015
LAND USE / LAND COVER	x	x
CROP PRODUCTION SYSTEMS	x	
NUMBER OF HOLDINGS, CULTIVATED AREAS & LIVESTOCK	Various time spans following the available data. See National censuses part.	
LIVESTOCK	x	
BIOCLIMATIC	Averages for years 1970 - 2000	
POPULATION	x	x

Detailed information on DIVERCROP database variables

Land use / Land cover

The Climate Change Initiative - Land Cover (CCI-LC) project delivers consistent global land cover maps at 300 m spatial resolution on an annual basis from 1992 to 2015. The typology, presented in Table 4, was defined using the Land Cover Classification System (LCCS) developed by the United Nations (UN) Food and Agriculture Organization (FAO), with the view to be as much as possible compatible with the GLC2000, GlobCover 2005 and 2009 products. In addition, the UNLCCS was found quite compatible with the Plant Functional Types (PFTs) used in climate models [RD.1]³.

The 300 m maps have been upscaled at the three spatial resolutions of DIVERCROP database by calculating the percentage of each land cover variable present in each 2 km cell, 5 arc-minute cell and in each nuts-2. The upscaling process presented on Figure 2 for 2km grid is the same for the 5arc-minute and nuts2 scale.

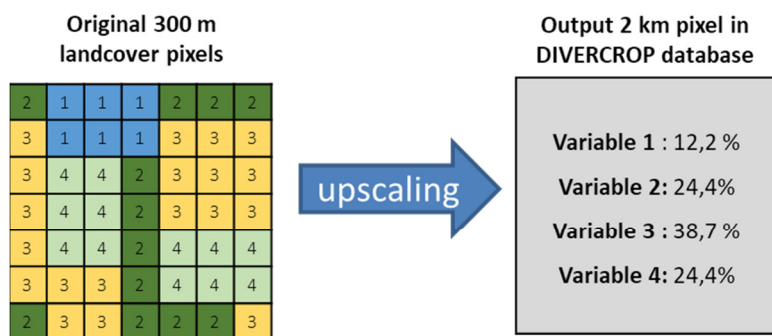


Figure 2 - Upscaling 300m land cover information to 2km grid

³ Source: Land Cover CCI PRODUCT USER GUIDE VERSION 2.0 (2017) https://maps.elie.ucl.ac.be/CCI/viewer/download/ESACCI-LC-Ph2-PUGv2_2.0.pdf

Table 4 - Land use / land cover variables in DIVERCROP database

LAND USE/LAND COVER VARIABLES
• Cropland rainfed
• Cropland rainfed - Herbaceous cover
• Cropland rainfed - Tree or shrub cover
• Cropland irrigated or post-flooding
• Mosaic cropland (>50%) / natural vegetation (tree/shrub/herbaceous cover) (<50%)
• Mosaic natural vegetation (tree/shrub/herbaceous cover) (>50%) / cropland (<50%)
• Tree cover broadleaved evergreen closed to open (>15%)
• Tree cover broadleaved deciduous closed to open (>15%)
• Tree cover broadleaved deciduous closed (>40%)
• Tree cover broadleaved deciduous open (15-40%)
• Tree cover needleleaved evergreen closed to open (>15%)
• Tree cover needleleaved evergreen closed (>40%)
• Tree cover needleleaved evergreen open (15-40%)
• Tree cover needleleaved deciduous closed to open (>15%)
• Tree cover needleleaved deciduous closed (>40%)
• Tree cover needleleaved deciduous open (15-40%)
• Tree cover mixed leaf type (broadleaved and needleleaved)
• Mosaic tree and shrub (>50%) / herbaceous cover (<50%)
• Mosaic herbaceous cover (>50%) / tree and shrub (<50%)
• Shrubland
• Shrubland evergreen
• Shrubland deciduous
• Grassland
• Sparse vegetation (tree/shrub/herbaceous cover) (<15%)
• Sparse tree (<15%)
• Sparse shrub (<15%)
• Sparse herbaceous cover (<15%)
• Tree cover flooded fresh or brakish water
• Tree cover flooded saline water
• Shrub or herbaceous cover flooded fresh/saline/brakish water
• Urban areas
• Bare areas
• Consolidated bare areas
• Unconsolidated bare areas
• Water bodies

Agricultural production systems

Three sources of data have been chosen in order to describe Mediterranean agricultural production systems: MAPSPAM, Gridded Livestock of the World, and National Censuses.

MAPSPAM

The Spatial Production Allocation Model (SPAM) 2005 endeavors to disaggregate crop statistics identified at national and sub-national units for the year 2005 to 5 arc-minute grid cells while taking account of different farming practices. SPAM generates four major output variables: physical area, harvested area, production quantity and yield for each of 42 crops distinguished by four production systems (i.e., irrigated – high input, rainfed – high input, rainfed – low-input and rainfed – subsistence production). The allocation model uses a cross-entropy optimization approach informed by major inputs such as cropland surface, location of irrigated areas, crop suitability and potential yields, rural population densities, production systems characteristics and crop prices to disaggregate crop

statistics (Wood-Sichra et al, 2016). Table 5 presents the MAPSPAM variables available in DIVERCROP database.

Table 5 - MAPSPAM variables available in DIVERCROP database

AGRICULTURAL PRODUCTION		
<p><u>CROP SURFACES</u></p> <p>CEREALS</p> <ul style="list-style-type: none"> • Wheat • Rice • Maize • Barley • Pearl Millet • Small Millet • Sorghum • Other Cereals <p>ROOTS & TUBERS</p> <ul style="list-style-type: none"> • Potato • Sweet Potato • Other Roots <p>PULSES</p> <ul style="list-style-type: none"> • Bean • Chickpea • Cowpea • Lentil • Other Pulses <p>VEGETABLES</p>	<p>OIL CROPS</p> <ul style="list-style-type: none"> • Soybean • Groundnut • Sunflower • Rapeseed • Sesame Seed • Other Oil Crops <p>SUGAR CROPS</p> <ul style="list-style-type: none"> • Sugarcane • Sugarbeet <p>FIBRES</p> <ul style="list-style-type: none"> • Cotton • Other Fibre Crops <p>STIMULATES</p> <ul style="list-style-type: none"> • Tea • Tobacco <p>FRUITS</p> <ul style="list-style-type: none"> • Banana • Tropical Fruit • Temperate Fruit <p>REST OF CROPS</p>	<p><u>VARIABLES</u></p> <ul style="list-style-type: none"> • physical area • harvested area • production • yield • value of production, aggregated to all crops, food and non-food <p><u>TECHNOLOGIES</u></p> <ul style="list-style-type: none"> • all technologies together, ie complete crop • irrigated portion of crop • rainfed high inputs portion of crop • rainfed low inputs portion of crop • rainfed subsistence portion of crop • rainfed portion of crop (= TA - TI, or TH + TL + TS)

As the DIVERCROP 5 arc-minute grid has been chosen to match spatially MAPSPAM grid, transferring the information from one database to another did not require any preliminary treatment.

Upscaling the information from MAPSPAM grid to Nuts2 required an aggregation of the 5 arc-minutes data, by summing crop surfaces for each Nuts2. This aggregation required to relate each 5 arc minutes pixel to the Nuts2 it belongs to. If attributing each pixel to a Nuts2 does not encounter any specific problem if the 5 arc-minute pixel is totally included in a Nuts2 polygon, a choice had to be made at the borders, as shown in Figure 3. We thus decided to consider each arc-minute pixel centroid's position in order to determine the correspondence between each 5 arc-minute pixel and each Nuts2.

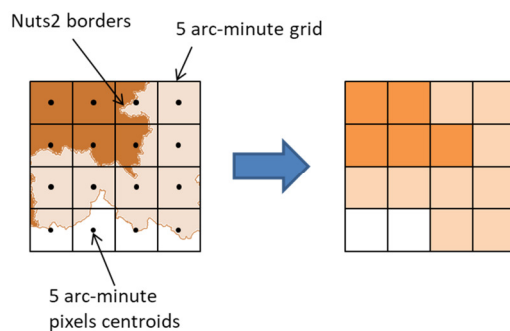


Figure 3 - Matching 5 arc-minutes grid and nuts 2 with 5 arc-minutes pixels centroids

Downscaling 5 arc-minutes grid information to 2 km grid required some data processing as well. We chose to redistribute MAPSPAM's physical areas following the percentage of agricultural land use

contained in each 2 km pixel. The total agricultural land use for each 2km pixel has been first calculated by summing the values of the following land cover variables:

- Cropland rainfed
- Cropland rainfed - Herbaceous cover
- Cropland rainfed - Tree or shrub cover
- Cropland irrigated or post-flooding
- Mosaic cropland (>50%) / natural vegetation (tree/shrub/herbaceous cover) (<50%)

The downscaling process from 5 arc-minute MAPSPAM pixels to 2km pixels is described in Figure 4.

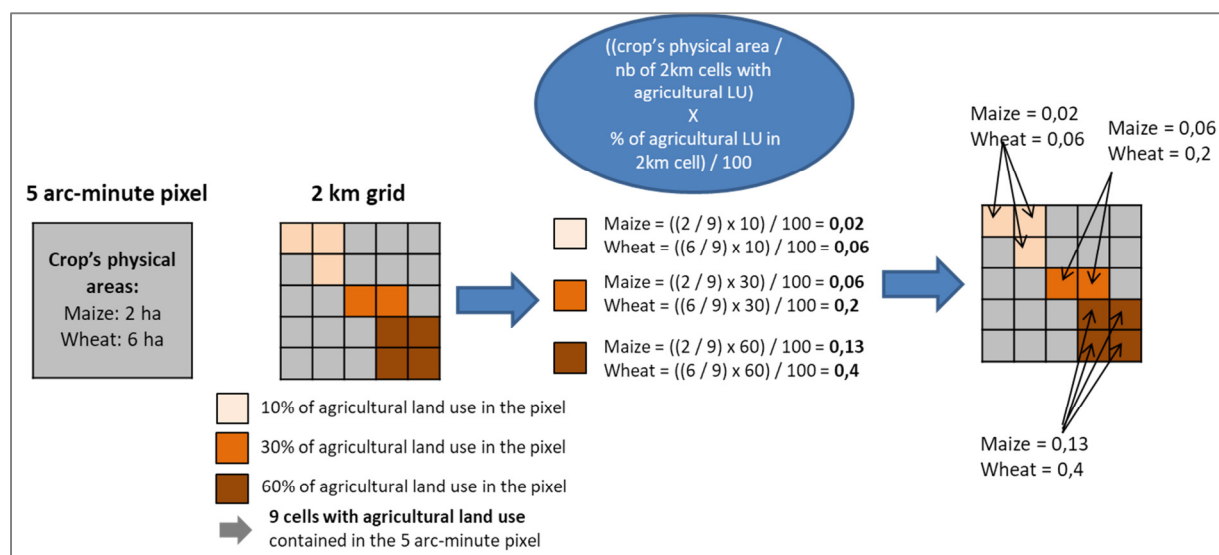


Figure 4 - Downscaling MAPSPAM information from 5 arc-minute pixels to 2km pixels

Livestock

FAO in collaboration with the Environmental Research Group Oxford (ERGO) developed the "Gridded Livestock of the World" (GLW). The GLW's main objective is to produce and distribute contemporary global distribution maps for the main species of livestock. These maps are created through the spatial disaggregation of sub-national statistical data based on empirical relationships with environmental variables in similar agro-ecological zones.

The data are produced in ESRI grid format at a spatial resolution of 3 minutes of arc (roughly 5 km at the equator) for cattle, buffalo, sheep, goats, pigs and poultry/chickens and are freely available for download via FAO's GeoNetwork data repository.

The Gridded Livestock of the World (GLW) database produced in 2007 provides modeled livestock densities of the world, adjusted to match official (FAOSTAT) national estimates for the reference year 2005, at a spatial resolution of 3 minutes of arc (about 5 x 5 km at the equator).⁴

Table 6 presents the Gridded Livestock of the World variables available in DIVERCROP database (FAO, 2007).

⁴ Source: <http://www.fao.org/ag/againfo/resources/en/glw/home.html>

Table 6 - Livestock variables in DIVERCROP database

LIVESTOCK
<ul style="list-style-type: none"> • Cattle • Goats • Sheep • Poultry • Pigs

National Censuses

DIVERCROP participant 9 gathered into DIVERCROP database the most relevant information from the national agricultural censuses, regarding number of holdings, cultivated areas and livestock. This task was carried out through a close and constant collaboration of Participants P1, P2, P5, P6, P7 and P8, particularly regarding data gathering and pre-processing phases, and in some cases (P1, P2, P7) regarding the furnishing of geographical datasets of the administrative-units.

Agricultural Censuses are held in European nations every 10 years. The last agricultural census (2009/2010) was the first synchronized agricultural census across all EU-27 Member States. Although the official census data can be derived from several sources, no complete set of data with the necessary criteria was readily available, and as a consequence it had to be built.

The criteria for the database construction were the following: (1) to cover the Mediterranean areas of seven countries, namely Algeria (DZ), France (FR), Italy (IT), Malta (MT), Portugal (PT), Spain (ES) and Tunisia (TN); (2) to contain data referred to the most disaggregated level of administrative units for which data is available in each country; (3) to contain data referred to at least two time points, including the latest available data, in each country; (4) to contain data on number of farm holdings, on physical areas covered by the main annual and permanent crops and on livestock (number of heads); (5) to contain a primary key that allows joining the census and surveys database to a geographical dataset of administrative units covering the entire area; (6) to have an associated complete geographical dataset of administrative units.

Construction of the Harmonized National Agricultural Censuses Database

The official data source for homogeneous data on agricultural issues referred to the European part of the Mediterranean is Eurostat. Nevertheless, data could not be collected directly from Eurostat since the level of detail needed, namely the highest level of disaggregation of administrative units, was unavailable within the Eurostat databases. Indeed, Eurostat stores data aggregated at NUTS 2 level or higher, which was insufficient for the objectives of the Project. As for the two North African countries, a centralized source of data is not available, therefore, for both European and African countries it was necessary to gather data directly from each country's individual official sources. As a consequence of the above agricultural censuses data had to be collected directly from the local partners of the Project.

Although data from national agricultural censuses and surveys were submitted by local members of the Divercrop consortium directly to Participant 9 for their elaboration (except for data of Italy, which was gathered by Participant 9 itself), official web sites of the national statistics authorities were also consulted during the first screening phase of the data gathering activity. In Annex 1, the complete list of links to these web resources is presented.

The Harmonized National Agricultural Censuses Database was built on an open source object-relational database management system (PostgreSQL), and it consists in a product that can be easily integrated into the Comprehensive Database, foreseen as output of WP1. The database 'divercrop_census_db' contains one single table element (dz_es_fr_it_mt_pt_tn_census) exportable from Postgres into the csv format (readable by most worksheets software) using the following SQL string: COPY dz_es_fr_it_mt_pt_tn_census to 'C:\dz_es_fr_it_mt_pt_tn_census.csv' delimiters',' CSV HEADER;

Data from the national agricultural censuses (France, Italy, Malta and Portugal) and from agricultural surveys (Algeria and Tunisia) were collected for several time points. In order to standardize administrative units within the area of interest of the Project, the European geocode standard for referencing the subdivisions of countries for statistical purposes was used, introducing adaptations of this standard where necessary (Algeria, Tunisia).

In most countries, data referred to the highest level of disaggregation of administrative units, namely LAU2, which corresponds to the municipality or equivalent. In fact, data of 4 out of 7 countries, namely, Italy (municipality), Malta (cities), Portugal (parishes) and Spain (municipalities) referred to the most disaggregated administrative level, which is defined by the European geocode standard as *Local Administrative Unit Level 2 (LAU 2)*⁵ while data of France needed aggregation at canton level, and the information provided by two countries (Algeria and Tunisia), contained in agricultural surveys was available at a higher level of geographical-administrative aggregation, namely Wilaya or governorates, as shown in Figure 5.

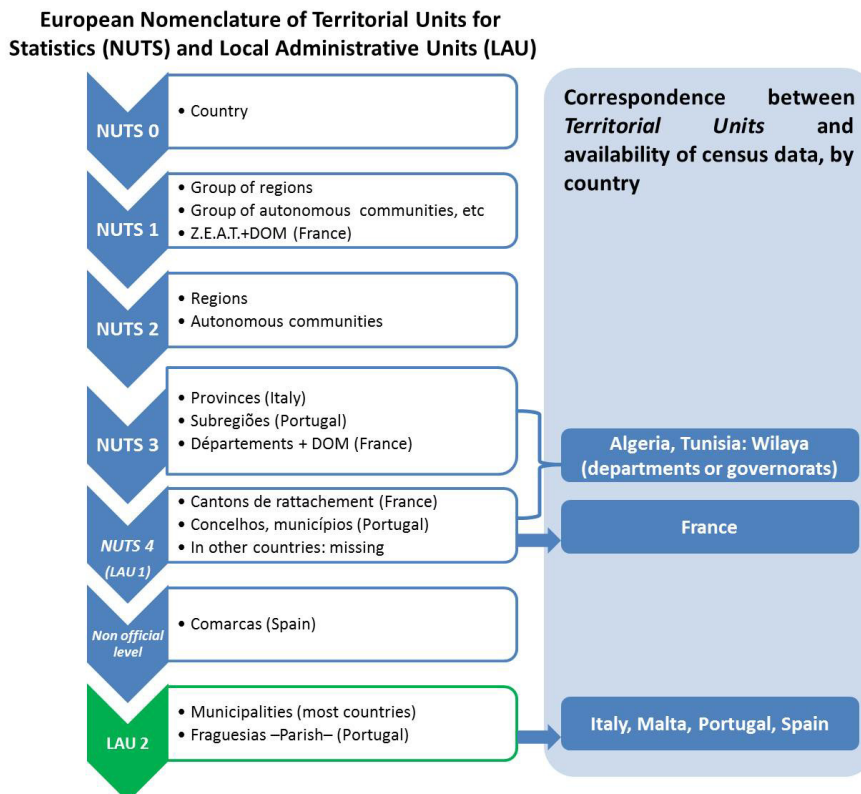


Figure 5 - Hierarchy of the European Nomenclature of Territorial Units (left) and correspondence between the Territorial Unit levels and data availability, by country (right)

⁵ To meet the demand for statistics at a local level, Eurostat maintains a system of Local Administrative Units (LAUs) compatible with NUTS. These LAUs are the building blocks of the NUTS, and comprise the municipalities and communes of the European Union. (<http://ec.europa.eu/eurostat/web/nuts/local-administrative-units>)

Structure of the Harmonized Agricultural Censuses Database

Table 7 shows the hierarchy of the aggregators and the primary key in the relational database. Within the pre-processing phases, each Participant prepared the data keeping as many administrative levels, use for data aggregation (identifiers of NUTS levels) as possible. In such manner, the higher level of aggregation of data is the Country for all participant nations, while disaggregation is possible at LAU2 level for ES, IT, MT and PT, LAU 1 level for FR and Wilaya for DZ and TN (highlighted in grey in the table).

Regarding the latter two countries, the only aggregator is the identifier of Wilaya. For the purpose of this Project, the Wilaya was assimilated to an intermediate level between NUTS 3 and NUTS 4, as it is also shown in Figure 5, nevertheless, to simplify the use of the Database, the Wilaya identifier was classified under NUTS 3 so as to avoid creating a further category in the hierarchy of administrative units levels.

Intermediate administrative levels are, as shown in Table 7, NUTS 2 (regions, autonomous communities) and NUTS 3 (provinces, departments).

Although Eurostat, in agreement with each EU-Member State, established the sub-divisions of NUTS and LAU levels and the corresponding codification systems, in all cases submission of statistical data carried non-standardized sets of LAU identifiers (which is obvious for Algeria and Tunisia, but should not be necessarily the case for the EU Member States).

Therefore, it was necessary to produce a unique set of codes referencing the most detailed Territorial Unit; this set of codes is referred to as “*join_field*”, which is:

- 1) an unambiguous codification, generated by combining the identifier of the country (ISO ALPHA 2-code) and the code of the most disaggregated Territorial Unit, provided by each country,
- 2) the field used for joining the data in the Agricultural Censuses Database with the geographical datasets of Territorial Units.

Regarding the point 2 above, it is worth pointing out that although Eurostat provides ready-to-use geographical datasets at LAU 2 level, carrying the official LAU 2 codes, as mentioned before, agricultural censuses data were submitted with codifications which do not comply with the EU standard. Furthermore, censuses data often referred to the local territorial units as they were in the year of the last agricultural census, and in some cases there were changes (i.e. between two censuses two or more municipalities may have been merged together) making it necessary the acquisition of the geography as it was in the year when the census was held. To be able to preserve the whole time-series without interruption in any single territorial unit, it was necessary to have all local territorial units matching throughout the time series, working on a case-by-case basis. This was done also through direct support from the Participants (this is particularly so in the case of Portugal).

For the above reasons, it was necessary to obtain also the geographical datasets of Local Territorial Units from each country, and then to transform the codification contained in these datasets so that it would match the “*join_field*” values in the Agricultural Censuses Database.

Finally, due to the inevitable diversity of sources of the geographical datasets, it was necessary to align all projections to fit the 3035 ETRS-LAEA European projection and to merge all datasets to obtain a unique coverage, which allows joining alphanumeric data from the Agricultural Censuses Database to the geographical dataset.

Table 7 - Primary key and aggregators of the Agricultural Censuses Database, corresponding to different levels of the hierarchy of administrative units

Country	country	nuts2	nuts3 (including equivalent Wilaya)	lau1	nuts3_lau2 (LAU 2)	Primary key: <i>join_field</i> (showing number of administrative units)
DZ-Algeria	X	N.A.	X			48
ES-Spain	X	X	X		X	8,096
FR-France	X			X		1,163
IT-Italy	X	X	X		X	8,092
MT-Malta	X	N.A.	N.A.		X	68
PT-Portugal	X	X	X		X	4,077
TN-Tunisia	X	N.A.	X			24
Total number of administrative units						21,568

in **bold**: names of the aggregator fields in the database

Data insertion into the Harmonized Agricultural Census Database

In the following paragraphs, general features of the data obtained from all participant countries are presented along with the formats used by members of the Divercrop consortium for data submission.

DZ-Algeria

Data from national agricultural surveys (at Wilaya - governorat level) were available for 2012 through 2016, as spreadsheet. The Agricultural Censuses Database comprises data referred to the first (2012) and the last (2016) available surveys.

ES-Spain

Data at LAU 2 level were available for the latest two national agricultural censuses, 1999 and 2009, as SPSS files.

The case of Spain required specific elaborations (carried out by Participant 8) for data harmonization at municipal level since the census universe was considerably restricted in 2009 compared to the previous censuses in Spain. In fact, in the previous census (1999), the minimum threshold of Utilised Agricultural Area (UAA) for a farm to be registered was 0.1 ha while for the census of 2009 it has been raised to 1 ha (with exceptions in case of intensive orientations⁶). The result is that both sets of data are completely comparable, by means of the application of constraints that excluded from the 1999 census universe all data referred to farms falling outside the 2009 universe. The ready-to-use data available from INE (Spanish National Statistics Institute) do not allow to obtain a proper time-series.

⁶ More specifically, the farms are eligible when falling in one of the following independent criteria: Agricultural holdings with at least 1 ha of utilised agricultural area (UAA).

Agricultural holdings with at least 0.2 ha of UAA used for fresh vegetables, melons and strawberries, flowers and ornamental plants (outdoors or under low protective cover) or irrigated fruit and berry plantations and citrus plantations, or nurseries, or woody crops under greenhouse.

- Agricultural holdings with at least 0.1 ha of UAA used for under-glass fresh vegetables, melons and strawberries.
- Agricultural holdings with at least 0.1 ha of UAA used for under-glass flowers and ornamental plants.
- Agricultural holdings with at least 0.5 ha of UAA used for tobacco.
- Agricultural holdings with at least 0.5 ha of UAA used for hops.
- Agricultural holdings with at least 0.5 ha of UAA used for cotton
- Agricultural holdings with one or more livestock units and a total standard output equal to or above 0.75 of the economic size of holdings.

FR-France

Data at LAU 1 level was submitted in spreadsheets referred to the latest two national agricultural censuses, 2000 and 2010. An agricultural census previous to that of the year 2000 exists, but data structure is different to the point that it is impossible to elaborate a time-series including that early census. Data acquisition, carried out by Participant 1, was supported by a public grant overseen by the French National Research Agency (ANR) as part of the “Investissements d’Avenir” program (reference: ANR-10-EQPX-17 – Centre d’accès sécurisé aux données – CASD).

It is worth pointing out that the French data were collected at canton level, which is an aggregation of municipality clusters (groups of neighbouring municipalities). This was necessary in order to keep the confidentiality of farms. In fact, the French authority for statistics does not disclose information at municipal level if in a given municipality one of the following conditions is met:

- there are less than 3 holdings;
- a single farm covers more than 85% of the total Utilized Agricultural Area of the municipality;
- the Standard Output of a single farm exceeds 85% of the total Standard Output of the farms within the municipality.

According to results of the data pre-elaboration phase, carried out by Participant 1, in the Mediterranean area of France a total number of 1,027 municipalities met at least one of the above conditions; this figure corresponds to 12% of the total number of municipalities of the area (8,571 municipalities). Therefore, while aggregations of data means reducing the geographical resolution of the data, on the other hand, keeping the information at municipality level would have led to a massive loss of information.

IT-Italy

Data at LAU 2 level, matching the 2010 list of Italian municipalities was available ready-to-use from the ISTAT (Italian National Institute for Statistics) website for the latest four national agricultural censuses, 1982, 1990, 2000 and 2010.

MT-Malta

Data at LAU 2 level from agricultural censuses 2001 and 2010 were collected by Participant 7 and elaborated into a ready-to-use geographical database. Only general information on areas (UAA, arable lands, set-aside, vineyards) were available for 2001 while a more detailed set of data for 2010. An important voice in the national agricultural census of Malta are areas dedicated to kitchen garden (12% of the total arable lands and an area equivalent to that dedicated to permanent crops), so the variable kitchen gardens was introduced for the case of Malta only.

PT-Portugal

Data at LAU 2 level from the latest three national agricultural censuses, namely 1989, 1999, 2009. Data gathering and elaboration was carried out by Participant 2. This was necessary particularly because INE-Statistics Portugal provides data at LAU 2 level only referred to 2009.

TN-Tunisia

Data from national agricultural surveys (at Wilaya - governorat level) were available for 1995 (hardcopy) and 2005 (html). Both sets of data were transferred to spreadsheet by Participant 9 before their inclusion in the database.

Table 8 shows, in relation with each Divercrop country, the years for which national agricultural datasets were incorporated into the database.

Table 8 - Years for which national agricultural datasets were elaborated. Column headings indicate the year codification in the database (y0 through y4).

	y0	y1	y2	y3	y4
DZ				2012	2016
ES			1999	2009	
FR			2000	2010	
IT	1982	1990	2000	2010	
MT			2001	2010	
PT		1989	1999	2009	
TN		1995	2005		

In Table 9, variables are classified according to the number of Countries for which they are available in a given time point. Only variables available in at least 4 Countries are shown, for the complete set of variables, with indications of their availability per country and per year, please see Appendix 2.

In the table, values available in 4 Countries are highlighted in orange, those available in 5 Countries are shown with a light-green background, while the background is dark green when variables are available for 6 countries.

Variable codes were defined similarly to another georeferenced database used in the Project (Mapspam). Therefore physical crop areas are composed of the letter **a** followed by the abbreviated name of the crop/crop type/land use (the latter case refers to UAA and TAA). A difference with the Mapspam codification consists in the use of the letter **r** to indicate irrigated areas and of the letter **s** to indicate non irrigated areas. The absence of any of these two letters in the physical areas codes indicates total values (*irrigated+rained areas*).

Figure 6 shows an elaboration of the geographically explicit data from the Harmonized Agricultural Censuses Database at the most disaggregated level of the administrative units for the area of interest of the Project.

Table 9 - Variables of the Agricultural Censuses Database available for most Divercrop countries. backgrounds show: 4 countries (orange), 5 countries (light green), 6 countries (dark green).

	variable code	legend	country							
			DZ	ES	FR	IT	MT	PT	TN	
Year 2 = 1999 (ES, PT), 2000 (FR, IT), 2001 (MT), 2005 (TN)	num_hold_y2	number of holdings		X	X	X				X
	bovine_y2	heads of cows		X	X	X			X	X
	ovine_y2	heads of sheep		X	X	X			X	X
	caprine_y2	heads of goats		X	X	X			X	X
	a_barl_y2	barley		X	X	X				X
	a_cer_y2	cereals (total area)		X	X	X			X	X
	a_puls_y2	pulses		X	X				X	X
	a_fodder_y2	fodder crops (total area)		X	X	X			X	X
	a_ind_crops_y2	industrial crops (total area)		X	X	X			X	X
	a_pota_y2	potato		X	X	X			X	
	a_vege_y2	vegetables		X	X	X			X	X
	a_flowers_y2	flowers		X	X	X			X	
	a_setaside_y2	set aside		X	X	X	X			
	a_arable_y2	arable lands (total area)		X			X	X		X
	a_olive_y2	olive		X	X	X			X	X
	a_fruit_olive_vine_y2	total area under fruit plantation, olive, vineyards		X	X	X				X
	a_viney_y2	vineyards		X	X	X	X		X	X
	a_uaa_y2	UAA		X	X	X	X		X	X
Year 3 = 2012 (DZ), 2009 (ES, PT), 2010 (FR, IT, MT)	bovine_y3	heads of cows	X	X	X	X	X		X	
	ovine_y3	heads of sheep	X	X	X	X	X		X	
	caprine_y3	heads of goats	X	X	X	X	X		X	
	a_whea_y3	durum wheat	X	X	X	X				
	a_c_whea_y3	common wheat	X	X	X	X				
	a_barl_y3	barley	X	X	X	X				
	a_cer_y3	cereals (total area)	X	X	X	X			X	
	a_puls_y3	pulses	X	X	X	X			X	
	a_fodder_y3	fodder crops (total area)	X	X	X	X	X		X	
	a_ind_crops_y3	industrial crops	X	X	X	X			X	
	a_pota_y3	potato	X	X	X	X	X		X	
	a_vege_y3	vegetables	X	X	X	X	X		X	
	a_flowers_y3	flowers		X	X	X	X		X	
	a_setaside_y3	set aside	X	X	X	X	X			
	a_arable_y3	arable lands (total area)	X	X		X	X		X	
	a_citrus_y3	citrus plantations	X	X		X	X		X	
	a_temf_y3	temperate fruits	X	X		X	X		X	
	a_olive_y3	olive	X	X	X	X	X		X	
	a_fruit_olive_vine_y3	total area under fruit plantation, olive, vineyards	X	X	X	X	X			
	a_viney_y3	vineyards	X	X	X	X	X		X	
a_uaa_y3	UAA	X	X	X	X	X		X		
a_meadow_y3	meadows	X	X	X	X					
a_taa_y3	TAA	X	X	X	X					

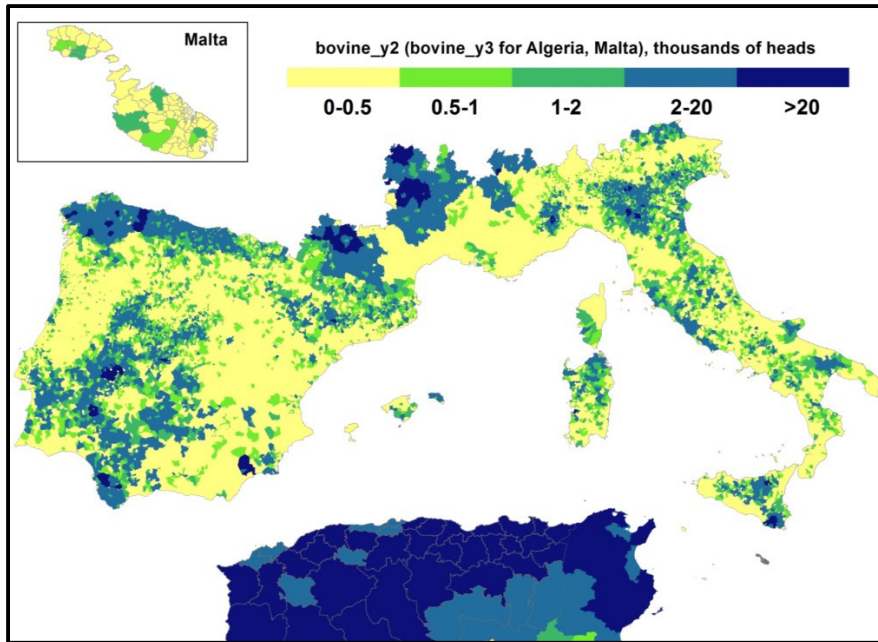


Figure 6 - Number of heads of cows in y2 (y3 in DZ and MT).

Bio-physical

Terrain

Elevation and slope data stem from the 30 arc-seconds (approximately 1km) Global Digital Elevation Model V2 from The Ministry of Economy, Trade, and Industry (METI) of Japan and the United States National Aeronautics and Space Administration (NASA). For each 2km, 5 arc-min and nuts 2 spatial units, the minimum and maximum slope and elevation values have been extracted, and the mean of the values have been calculated as well (Table 10).

Table 10 - Terrain variables in DIVERCROP database

TERRAIN	
<u>ELEVATION</u>	<u>SLOPE</u>
<ul style="list-style-type: none"> • Min. • Max. • Average 	<ul style="list-style-type: none"> • Min. • Max. • Average

Bioclimatic

Bioclimatic values stem from the Worldclim database (Fick, Hijmans, 2017), available in both 5 arc-minutes and 30 arc-seconds. *Worldclim bioclimatic variables are an average for the years 1970-2000. Bioclimatic variables are derived from the monthly temperature and rainfall values in order to generate more biologically meaningful variables. The bioclimatic variables represent annual trends (e.g., mean annual temperature, annual precipitation) seasonality (e.g., annual range in temperature and precipitation) and extreme or limiting environmental factors (e.g., temperature of the coldest and warmest month, and precipitation of the wet and dry quarters). A quarter is a period of three months (1/4 of the year).*⁷

⁷ Source : <http://worldclim.org>

As 5 arc-minutes data matched spatially DIVERCROP 5 arc-minutes grid, no data processing was required to transfer this data from one grid to another. The aggregation at nuts2 level was performed by calculating the averages of each variable for each nuts2 (see Figure 3). For the 2km database, 30 arc-seconds WORLDCLIM data has been considered. Each 2km pixel was thus given the average value of each 30 arc-seconds pixel contained (considering that each 30 arc-seconds pixel is represented by its centroid, as shown in Figure 3).

Table 11 - Bioclimatic variables in DIVERCROP database

BIOCLIMATIC	
<p>TEMPERATURE (Average for years 1970 to 2000)</p> <ul style="list-style-type: none"> • Annual Mean Temp. • Mean Diurnal Range (Mean of monthly (max temp - min temp)) • Isothermality (BIO2/BIO7) (* 100) • Temp. seasonality (stand. dev. *100) • Max Temp. of Warmest Month • Min Temp. of Coldest Month • Temp. Annual Range (BIO5-BIO6) • Mean Temp. of Wettest Quarter • Mean Temp. of Driest Quarter • Mean Temp. of Warmest Quarter • Mean Temp. of Coldest Quarter 	<p>PRECIPITATION (Average for years 1970 to 2000)</p> <ul style="list-style-type: none"> • Annual Precip. • Precip. of Wettest Month • Precip. of Driest Month • Precip. Seasonality (Coef. of Variation) • Precip. of Wettest Quarter • Precip. of Driest Quarter • Precip. of Warmest Quarter • Precip. of Coldest Quarter

Soils

Soil data for DIVERCROP database stem from the 30 arc-seconds Harmonized World Soil Database. They include a lot of pedological variables (soil types) and an assessment of soil qualities for crop production. *On the basis of soil parameters provided by HWSD seven key soil qualities important for crop production have been derived, namely: nutrient availability, nutrient retention capacity, rooting conditions, oxygen availability to roots, excess salts, toxicities, and workability. Soil qualities are related to the agricultural use of the soil and more specifically to specific crop requirements and tolerances. For the illustration of soil qualities, maize was selected as reference crop because of its global importance and wide geographical distribution. Remember that classes are qualitative not quantitative. Only classes 1 to 4 are corresponding to an assessment of soil limitations for plant growth. Class 1 is generally rated between 80 and 100% of the growth potential, class 2 between 60 and 80%, class 3 between 40 and 60%, and class 4 less than 40%⁸.*

In DIVERCROP database, each 2km and 5 arc-minutes pixel and each nuts2 was given the most represented soil quality index, and the percentage of each type of soils it contains. Table 12 lists all the soil types and soil quality types available in DIVERCROP database. For extended information about soil quality index elaboration, see Nachtergaele et al, 2009.

⁸ Source : <http://webarchive.iiasa.ac.at/Research/LUC/External-World-soil-database/HTML/SoilQuality.html?sb=10>

Table 12 - Soils variables in DIVERCROP database

SOILS	
<p><u>SOIL TYPE</u></p> <ul style="list-style-type: none"> • ACRISOLS: Soils with subsurface accumulation of low activity clays and low base saturation • ALISOLS: Soils with sub-surface accumulation of high activity clays, rich in exchangeable aluminium • ANDOSOLS: Young soils formed from volcanic deposits • ARENOSOLS: Sandy soils featuring very weak or no soil development • ANTHROSOLS: Soils in which human activities have resulted in profound modification of their properties • CHERNOZEMS: Soils with a thick, dark topsoil, rich in organic matter with a calcareous subsoil • CALCISOLS: Soils with accumulation of secondary calcium carbonates • CAMBISOLS: Weakly to moderately developed soils • FLUVISOLS: Young soils in alluvial deposits • FERRALSOLS: Deep, strongly weathered soils with a chemically poor, but physically stable subsoil • GLEYSOLS: Soils with permanent or temporary wetness near the surface • GREYZEMS: Acid soils with a thick, dark topsoil rich in organic matter • GYPISOLS: Soils with accumulation of secondary gypsum • HISTOSOLS: Soils which are composed of organic materials • KASTANOZEMS: Soils with a thick, dark brown topsoil, rich in organic matter and a calcareous or gypsum-rich subsoil • LEPTOSOLS: Very shallow soils over hard rock or in unconsolidated very gravelly material • LUVISOLS: Soils with subsurface accumulation of high activity clays and high base saturation • LIXISOLS: Soils with subsurface accumulation of low activity clays and high base saturation • NITISOLS: Deep, dark red, brown or yellow clayey soils having a pronounced shiny, nut-shaped structure • PODZOLUVISOLS: Acid soils with a bleached horizon penetrating into a clay-rich subsurface horizon • PHAEZEMS: Soils with a thick, dark topsoil rich in organic matter and evidence of removal of carbonates • PLANOSOLS: Soils with a bleached, temporarily water-saturated topsoil on a slowly permeable subsoil • PLINTHOSOLS: Wet soils with an irreversibly hardening mixture of iron, clay and quartz in the subsoil • PODZOLS: Acid soils with a subsurface accumulation of iron-aluminum-organic compounds • REGOSOLS: Soils with very limited soil development • SOLONCHAKS: Strongly saline soils • SOLONETZ: Soils with subsurface clay accumulation, rich in sodium • VERTISOLS: Dark-coloured cracking and swelling clays 	<p><u>SOIL QUALITY</u></p> <ul style="list-style-type: none"> • Nutrient availability: Soil texture, soil organic carbon, soil pH, total exchangeable bases • Nutrient retention capacity: Soil Organic carbon, Soil texture, base saturation, cation exchange capacity of soil and of clay fraction • Rooting conditions: Soil textures, bulk density, coarse fragments, vertic soil properties and soil phases affecting root penetration and soil depth and soil volume • Oxygen availability to roots: Soil drainage and soil phases affecting soil drainage • Excess salts: Soil salinity, soil sodicity and soil phases influencing salt conditions • Toxicity: Calcium carbonate and gypsum • Workability (constraining field management): Soil texture, effective soil depth/volume, and soil phases constraining soil management (soil depth, rock outcrop, stoniness, gravel/concretions and hardpans)

Population

DIVERCROP database integrates two population data for 2005 and 2015 each: LANDSCAN⁹ and HYDE models.

Landscan data

*Sensing, Oak Ridge National Laboratory's LandScan™ is the community standard for global population distribution. At approximately 1 km resolution (30" X 30"), LandScan is the finest resolution global population distribution data available and represents an ambient population (average over 24 hours). The LandScan algorithm, an R&D 100 Award Winner, uses spatial data and imagery analysis technologies and a multi-variable dasymetric modeling approach to disaggregate census counts within an administrative boundary. Since no single population distribution model can account for the differences in spatial data availability, quality, scale, and accuracy as well as the differences in cultural settlement practices, LandScan population distribution models are tailored to match the data conditions and geographical nature of each individual country and region.*¹⁰

In DIVERCROP database, each 2km, 5 arc minute pixel and each nuts 2 was given the average value of LANDSCAN 30 arc-seconds unit.

Hyde data

*The HYDE database is developed under the authority of the Netherlands Environmental Assessment Agency. HYDE presents (gridded) time series of population and land use for the last 12,000 years. It also presents various other indicators such as GDP, value added, livestock, agricultural areas and yields, private consumption, greenhouse gas emissions and industrial production data, but only for the last century.*¹¹ Historical population, cropland and pasture statistics are combined with satellite information and specific allocation algorithms (which change over time) to create spatially explicit maps, which are fully consistent on a 5' longitude/ latitude grid resolution, and cover the period 10,000 bc to ad 2000 (Goldewijk et al, 2011).

In DIVERCROP database, only 5 arc-minute and nuts2 levels were given the HYDE model data. As 5 arc-minutes HYDE and DIVERCROP grids matched spatially, no data processing was required to transfer the data. Each nuts2 was given the average value of the HYDE 5 arc-minutes pixels it contains (referring to the centroids of each 5 arc-minute pixels, as explained in Figure 3).

Protected areas

The World Database on Protected Areas (WDPA)¹² is the most comprehensive global database on terrestrial and marine protected areas. It is a joint project between the United Nations Environment Programme (UNEP) and the International Union for Conservation of Nature (IUCN), managed by UNEP World Conservation Monitoring Centre (UNEP-WCMC). ProtectedPlanet.net is the online interface for the WDPA. ProtectedPlanet.net lets you discover the protected areas of the world

⁹ This product was made utilizing the LandScan (insert dataset year)™ High Resolution global Population Data Set copyrighted by UT-Battelle, LLC, operator of Oak Ridge National Laboratory under Contract No. DE-AC05-00OR22725 with the United States Department of Energy. The United States Government has certain rights in this Data Set. Neither UT-BATTELLE, LLC NOR THE UNITED STATES DEPARTMENT OF ENERGY, NOR ANY OF THEIR EMPLOYEES, MAKES ANY WARRANTY, EXPRESS OR IMPLIED, OR ASSUMES ANY LEGAL LIABILITY OR RESPONSIBILITY FOR THE ACCURACY, COMPLETENESS, OR USEFULNESS OF THE DATA SET.

¹⁰ Source : <https://landscan.ornl.gov/>

¹¹ Source : <http://themasites.pbl.nl/tridion/en/themasites/hyde/>

¹² UNEP-WCMC and IUCN (2018), Protected Planet: [The World Database on Protected Areas (WDPA)/The Global Database on Protected Areas Management Effectiveness (GD-PAME)] [On-line], [01/2018], Cambridge, UK: UNEP-WCMC and IUCN. Available at: www.protectedplanet.net.

*through exploring the maps and intuitive searching, feeding you information from the WDPA, photos from Panoramio and text descriptions from Wikipedia.*¹³

In DIVERCROP database, each 2km and 5 arc-minutes pixel was given the attributes of the protected area they intersect (Table 13).

Table 13 - Attributes of the protected areas in DIVERCROP database

PROTECTED AREAS	
•	Name
•	Type
•	Year of creation

¹³ Source : <https://protectedplanet.net/>

References

Malek, Z., Verburg, P. H. (2017). Mediterranean land systems: Representing diversity and intensity of complex land systems in a dynamic region. *Landscape and Urban Planning*, 165, 102-116. DOI: 10.1016/j.landurbplan.2017.05.012

LAND COVER CCI

Land Cover CCI PRODUCT USER GUIDE VERSION 2.0 (2017)

https://maps.elie.ucl.ac.be/CCI/viewer/download/ESACCI-LC-Ph2-PUGv2_2.0.pdf

MAPSPAM

You, L., U. Wood-Sichra, S. Fritz, Z. Guo, L. See, and J. Koo. (2017). Spatial Production Allocation Model (SPAM) 2005 v3.2. October 17, 2018. Available from <http://mapspam.info>.

Wood-Sichra, U., A.B. Joglekar and L. You. 2016. "Spatial Production Allocation Model (SPAM) 2005: Technical Documentation". *HarvestChoice Working Paper*. Washington, D.C.: International Food Policy Research Institute (IFPRI) and St. Paul: International Science and Technology Practice and Policy (InSTePP) Center, University of Minnesota

HARMONIZED WORLD SOIL DATABASE

Nachtergaele, F., van Velthuisen, H., Verelst L. (2008). Harmonized World Soil Database Version 1.1

FAO/IIASA/ISRIC/ISS-CAS/JRC (2009). Harmonized World Soil Database (version 1.1). FAO, Rome, Italy and IIASA, Laxenburg, Austria.

WORLDCLIM DATABASE

Fick, S.E., Hijmans, R.J. (2017). Worldclim 2: New 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*

HYDE MODEL

Goldewijk, K.K., Beusen, A., van Drecht, G., de Vos, M. (2011). The HYDE 3.1 spatially explicit database of human-induced global land-use change over the past 12,000 years *Global Ecology and Biogeography*, 20, 73–86.

GRIDDED LIVESTOCK OF THE WORLD

FAO. 2007. Gridded livestock of the world 2007, by G.R.W. Wint and T.P. Robinson. <http://www.fao.org/3/a-a1259e.pdf>

Appendix

Appendix 1 - List of the national statistics authorities of Divercrop Countries

DZ-Algeria	http://www.ons.dz/-Donnees-Statistiques-.html http://www.ons.dz/-Productions-vegetales-2005-2009,219-.html
ES-Spain	http://www.ine.es/dyngs/INEbase/es/operacion.htm?c=Estadistica_C&cid=1254736176851&menu=resultados&idp=1254735727106
FR-France	http://agreste.agriculture.gouv.fr/page-d-accueil/article/donnees-en-ligne
IT-Italy	http://dati.istat.it/?lang=en
MT-Malta	no web resources were available at the time of the data screening
PT-Portugal	https://www.ine.pt/xportal/xmain?xpid=INE&xpgid=ine_bdc_tree&contexto=bd&selTab=tab2
TN-Tunisia	http://dataportal.ins.tn/fr/DataAnalysis?DRllk3Kak0irBSijS5Jkg

Appendix 2 - Variables of the Agricultural Censuses Database. Orange background: 4 countries, light green background: 5 countries, dark green background: 6 countries.

	variable code	variable legend	country						
			DZ	ES	FR	IT	MT	PT	TN
Year 0 = 1982 (IT)	num_hold_y0	number of holdings				X			
	bovine_y0	heads of cows				X			
	ovine_y0	heads of sheep				X			
	caprine_y0	heads of goats				X			
	a_wood_plant_y0	wood plantations				X			
	a_whea_y0	durum wheat				X			
	a_c_whea_y0	common wheat				X			
	a_barl_y0	barley				X			
	a_rice_y0	rice				X			
	a_maiz_r_y0	maize, irrigated				X			
	a_maiz_y0	maize				X			
	a_puls_y0	pulses				X			
	a_cer_y0	cereals (total area)				X			
	a_oil_y0	annual oil crops (total area)				X			
	a_fodder_r_y0	fodder crops, irrigated				X			
	a_fodder_y0	fodder crops (total area)				X			
	a_toba_y0	tobacco				X			
	a_sugb_y0	sugarbeet				X			
	a_ind_crops_y0	industrial crops (total area)				X			
	a_vege_r_y0	vegetables, irrigated				X			
	a_vege_y0	vegetables (total area)				X			
	a_flowers_y1	flowers				X			
	a_setaside_y0	set aside				X			
	a_arable_y0	arable lands (total area)				X			
	a_citrus_r_y0	citrus, irrigated				X			
	a_citrus_y0	citrus plantations				X			
	a_temf_r_y0	temperate fruits, irrigated				X			
	a_temf_y0	temperate fruits (total area)				X			
	a_olive_y0	olive				X			
	a_fruit_olive_vine_y0	total area under fruit plantation, olive, vineyards				X			
	a_viney_r_y0	vineyards, irrigated				X			
	a_viney_y0	vineyards				X			
	a_uaa_y0	UAA				X			
	a_meadow_y0	meadows				X			
a_uaa_r_y0	UAA, irrigated				X				
a_taa_y0	TAA				X				
Year 1 = 1990 (IT), 1989 (PT), 1995 (TN)	num_hold_y1	number of holdings				X			X
	bovine_y1	heads of cows				X		X	X
	ovine_y1	heads of sheep				X		X	X
	caprine_y1	heads of goats				X		X	X
	a_wood_plant_y1	wood plantations				X			
	a_whea_r_y1	durum wheat, irrigated							X
	a_whea_y1	durum wheat				X			X
	a_c_whea_r_y1	common wheat, irrigated							X
	a_c_whea_y1	common wheat				X			X
	a_barl_y1	barley				X			X
	a_rice_y1	rice				X			
	a_maiz_r_y1	maize, irrigated				X			
	a_maiz_y1	maize				X			
	a_ocer_r_y1	other cereals, irrigated							X
	a_ocer_y1	other cereals							X
	a_cer_r_y1	cereals, irrigated							X
	a_cer_y1	cereals (total area)				X		X	X
	a_chic_r_y1	chickpea, irrigated							X

Year 2 = 1999 (ES, PT), 2000 (FR, IT), 2001 (MT), 2005 (TN)	a_chic_y1	chickpea							X
	a_opul_r_y1	other pulses, irrigated							X
	a_opul_y1	other pulses							X
	a_puls_r_y1	pulses, irrigated							X
	a_puls_y1	pulses			X		X		X
	a_oil_y1	annual oil crops (total area)			X				
	a_temp_grass_y1	temporary grasses						X	
	a_fodder_r_y1	fodder crops, irrigated			X				
	a_fodder_y1	fodder crops (total area)			X		X		X
	a_toba_y1	tobacco			X				
	a_ind_crops_r_y1	industrial crops, irrigated							X
	a_ind_crops_y1	industrial crops (total area)			X		X		X
	a_pota_y1	potato						X	
	a_sugb_y1	sugarbeet			X				
	a_vege_r_y1	vegetables, irrigated			X				X
	a_vege_y1	vegetables (total area)			X		X		X
	a_flowers_y1	flowers			X		X		
	a_other_arable_y1	other arable crops						X	
	a_setaside_y1	set aside			X				
	a_arable_y1	arable lands (total area)			X		X		X
	a_citrus_r_y1	citrus, irrigated			X				
	a_citrus_y1	citrus plantations			X		X		
	a_temf_r_y1	temperate fruits, irrigated			X				
	a_temf_y1	temperate fruits (total area)			X		X		X
	a_nutf_y1	nut fruits (total area)						X	X
	a_table_olive_y1	table olive							X
	a_oil_olive_y1	oil olive							X
	a_olive_r_y1	olive, irrigated							
	a_olive_y1	olive			X		X		X
	a_fruit_olive_vine_y1	total area under fruit plantation, olive, vineyards			X				X
	a_fruit_palm_y1	fruit palms							X
	a_permcrops_y1	permanent crops						X	
a_grape_y1	grapes							X	
a_viney_r_y1	vineyards, irrigated			X					
a_viney_y1	vineyards			X		X		X	
a_uaa_y1	UAA			X		X		X	
a_uaa_r_y1	UAA, irrigated			X					
a_meadow_y1	meadows			X					
a_taa_y1	TAA			X					
	num_hold_y2	number of holdings		X	X	X		X	
	bovine_y2	heads of cows		X	X	X		X X	
	ovine_y2	heads of sheep		X	X	X		X X	
	caprine_y2	heads of goats		X	X	X		X X	
	a_wood_plant_y2	wood plantations					X		
	a_whea_r_y2	durum wheat, irrigated		X	X			X	
	a_whea_s_y2	durum wheat, non irrigated		X					
	a_whea_y2	durum wheat		X	X	X		X	
	a_c_whea_r_y2	common wheat, irrigated		X	X			X	
	a_c_whea_s_y2	common wheat, non irrigated		X					
	a_c_whea_y2	common wheat		X	X	X		X	
	a_barl_r_y2	barley, irrigated		X				X	
	a_barl_s_y2	barley, non irrigated		X					
	a_barl_y2	barley		X	X	X		X	
	a_oat_r_y2	oat, irrigated		X					
	a_oat_s_y2	oat, non irrigated		X					
	a_oat_y2	oat		X	X				
	a_rye_r_y2	rye, irrigated		X					
	a_rye_s_y2	rye, non irrigated		X					
	a_rye_y2	rye		X	X				
	a_rice_y2	rice		X	X	X			

a_maiz_r_y2	maize, irrigated	X	X	X		
a_maiz_s_y2	maize, non irrigated	X				
a_maiz_y2	maize	X	X	X		
a_ocer_r_y2	other cereals, irrigated	X				X
a_ocer_s_y2	other cereals, non irrigated	X				
a_ocer_y2	other cereals	X				X
a_cer_r_y2	cereals, irrigated	X				X
a_cer_s_y2	cereals, non irrigated	X				
a_cer_y2	cereals (total area)	X	X	X		X X
a_chic_r_y2	chickpea, irrigated	X				X
a_chic_s_y2	chickpea, non irrigated	X				
a_chic_y2	chickpea	X				X
a_greenpea_r_y2	greenpea, irrigated	X				
a_greenpea_s_y2	greenpea, non irrigated	X				
a_greenpea_y2	greenpea	X				
a_opul_r_y2	other pulses, irrigated	X				X
a_opul_s_y2	other pulses, non irrigated	X				
a_opul_y2	other pulses	X				
a_puls_r_y2	pulses, irrigated	X				X
a_puls_s_y2	pulses, non irrigated	X				
a_puls_y2	pulses	X	X	X		X X
a_cott_r_y2	cotton, irrigated	X				
a_cott_s_y2	cotton, non irrigated	X				
a_cotton_y2	cotton	X				
a_sunf_r_y2	sunflower, irrigated	X	X			
a_sunf_s_y2	sunflower, non irrigated	X				
a_sunf_y2	sunflower	X	X			
a_soyb_r_y2	soybean, irrigated	X	X			
a_soyb_s_y2	soybean, non irrigated	X				
a_soyb_y2	soybean	X	X			
a_rape_r_y2	rapeseed, irrigated	X				
a_rape_y2	rapeseed	X	X			
a_rape_s_y2	rapeseed, non irrigated	X				
a_oil_y2	annual oil crops (total area)	X	X	X		
a_perm_fodder_s_y2	permanent fodder crops, non irrigated	X				
a_perm_fodder_r_y2	permanent fodder crops, irrigated	X				
a_perm_fodder_y2	permanent fodder crops	X				
a_fodder_maiz_s_y2	fodder maize, non irrigated	X				
a_fodder_maiz_r_y2	fodder maize, irrigated	X	X			
a_fodder_maiz_y2	fodder maize	X	X			
a_fodder_legu_r_y2	fodder legumes, irrigated	X				
a_fodder_legu_s_y2	fodder legumes, non irrigated	X				
a_fodder_legu_y2	fodder legumes	X				
a_alfalfa_r_y2	alfalfa, irrigated	X				
a_alfalfa_s_y2	alfalfa, non irrigated	X				
a_alfalfa_y2	alfalfa	X				
a_temp_grass_y2	temporary grasses					X
a_fodder_r_y2	fodder crops, irrigated			X		
a_fodder_y2	fodder crops (total area)	X	X	X		X X
a_toba_r_y2	tobacco, irrigated	X				
a_toba_s_y2	tobacco, non irrigated	X				
a_toba_y2	tobacco	X		X		
a_hop_r_y2	hop, irrigated	X				
a_hop_s_y2	hop, non irrigated	X				
a_hop_y2	hop	X				
a_arom_r_y2	aromatic plants, irrigated	X				
a_arom_s_y2	aromatic plants, non irrigated	X				
a_arom_y2	aromatic plants	X	X			
a_othind_r_y2	other industrial crops, irrigated	X				
a_othind_s_y2	other industrial crops, non irrigated	X				

Year 3 = 2012 (DZ), 2009 (ES, PT), 2010 (FR, IT, MT)	a_othind_y2	other industrial crops	X					
	a_ind_crops_s_y2	industrial crops, non irrigated	X					
	a_ind_crops_r_y2	industrial crops, irrigated	X					X
	a_ind_crops_y2	industrial crops (total area)	X	X	X		X	X
	a_rts_r_y2	roots, irrigated	X					
	a_rts_s_y2	roots, non irrigated	X					
	a_rts_y2	roots	X					
	a_pota_r_y2	potato, irrigated	X	X				
	a_pota_s_y2	potato, non irrigated	X					
	a_pota_y2	potato	X	X	X		X	
	a_sugb_r_y2	sugarbeet, irrigated	X	X				
	a_sugb_s_y2	sugarbeet, non irrigated	X					
	a_sugb_y2	sugarbeet	X	X	X		X	
	a_vege_r_y2	vegetables, irrigated	X	X	X			X
	a_vege_s_y2	vegetables, non irrigated	X					
	a_vege_y2	vegetables	X	X	X		X	X
	a_flowers_y2	flowers	X	X	X		X	
	a_other_arable_y2	other arable crops						X
	a_setaside_y2	set aside	X	X	X	X		
	a_arable_y2	arable lands (total area)	X		X	X	X	X
	a_citrus_r_y2	citrus, irrigated	X	X	X			
	a_citrus_y2	citrus plantations	X		X		X	
	a_temf_r_y2	temperate fruits, irrigated	X		X			
	a_temf_s_y2	temperate fruits, non irrigated	X					
	a_temf_y2	temperate fruits (total area)	X		X		X	X
	a_nutf_r_y2	nut fruits, irrigated	X		X			
	a_nutf_s_y2	nut fruits, non irrigated	X					
	a_nutf_y2	nut fruits (total area)	X				X	X
	a_table_olive_r_y2	table olive, irrigated	X					
	a_table_olive_s_y2	table olive, non irrigated	X					
	a_table_olive_y2	table olive	X					X
	a_oil_olive_y2	oil olive	X					X
	a_olive_r_y2	olive, irrigated	X					
	a_olive_s_y2	olive, non irrigated	X					
	a_olive_y2	olive	X	X	X		X	X
	a_fruit_olive_vine_y2	total area under fruit plantation, olive, vineyards	X	X	X		X	
	a_fruit_palm_y2	fruit palms						X
	a_permcrops_y2	permanent crops	X				X	
	a_grape_s_y2	grapes, non irrigated	X					
	a_grape_r_y2	grapes, irrigated	X					
	a_grape_y2	grapes	X					X
a_viney_r_y2	vineyards, irrigated	X	X	X				
a_viney_s_y2	vineyards, non irrigated	X						
a_viney_y2	vineyards	X	X	X	X	X	X	
a_uaa_y2	UAA	X	X	X	X	X	X	
a_meadow_y2	meadows	X	X	X				
a_uaa_r_y2	UAA, irrigated			X				
a_taa_y2	TAA	X	X	X				
num_hold_y3	number of holdings	X	X	X	X			
bovine_y3	heads of cows	X	X	X	X	X	X	
ovine_y3	heads of sheep	X	X	X	X	X	X	
caprine_y3	heads of goats	X	X	X	X	X	X	
a_wood_plant_y3	wood plantations				X			
a_whea_r_y3	durum wheat, irrigated	X	X					
a_whea_s_y3	durum wheat, non irrigated	X						
a_whea_y3	durum wheat	X	X	X	X			
a_c_whea_r_y3	common wheat, irrigated	X	X					
a_c_whea_s_y3	common wheat, non irrigated	X						
a_c_whea_y3	common wheat	X	X	X	X			
a_barl_r_y3	barley, irrigated	X	X					

a_barl_s_y3	barley, non irrigated					X
a_barl_y3	barley	X	X	X	X	
a_oat_r_y3	oat, irrigated			X	X	
a_oat_s_y3	oat, non irrigated			X		
a_oat_y3	oat	X	X	X		
a_rye_r_y3	rye, irrigated			X		
a_rye_s_y3	rye, non irrigated			X		
a_rye_y3	rye			X	X	
a_rice_y3	rice			X	X	X
a_maiz_s_y3	maize, non irrigated			X		
a_maiz_r_y3	maize, irrigated			X	X	X
a_maiz_y3	maize			X	X	X
a_ocer_r_y3	other cereals, irrigated			X		
a_ocer_s_y3	other cereals, non irrigated			X		
a_ocer_y3	other cereals			X		
a_cer_y3	cereals (total area)	X	X	X	X	X
a_chic_r_y3	chickpea, irrigated			X		
a_chic_s_y3	chickpea, non irrigated			X		
a_chic_y3	chickpea	X	X			
a_greenpea_r_y3	greenpea, irrigated			X		
a_greenpea_s_y3	greenpea, non irrigated			X		
a_greenpea_y3	greenpea	X	X			
a_opul_r_y3	other pulses, irrigated			X		
a_opul_s_y3	other pulses, non irrigated			X		
a_opul_y3	other pulses	X	X			
a_puls_y3	pulses	X	X	X	X	X
a_cott_r_y3	cotton, irrigated			X		
a_cott_s_y3	cotton, non irrigated			X		
a_cotton_y3	cotton			X		
a_sunf_r_y3	sunflower, irrigated			X	X	
a_sunf_s_y3	sunflower, non irrigated			X		
a_sunf_y3	sunflower			X	X	
a_soyb_r_y3	soybean, irrigated			X	X	
a_soyb_s_y3	soybean, non irrigated			X		
a_soyb_y3	soybean			X	X	
a_rape_r_y3	rapeseed, irrigated			X	X	
a_rape_s_y3	rapeseed, non irrigated			X		
a_rape_y3	rapeseed			X	X	
a_oil_y3	annual oil crops (total area)			X	X	X
a_perm_fodder_r_y3	permanent fodder crops, irrigated			X		
a_perm_fodder_s_y3	permanent fodder crops, non irrigated			X		
a_perm_fodder_y3	permanent fodder crops			X		
a_fodder_maiz_s_y3	fodder maize, non irrigated			X		
a_fodder_maiz_r_y3	fodder maize, irrigated			X	X	
a_fodder_maiz_y3	fodder maize			X	X	
a_fodder_legu_s_y3	fodder legumes, non irrigate			X		
a_fodder_legu_r_y3	fodder legumes, irrigated			X		
a_fodder_legu_y3	fodder legumes			X		
a_temp_grass_y3	temporary grasses					X
a_fodder_r_y3	fodder crops, irrigated					X
a_fodder_y3	fodder crops (total area)	X	X	X	X	X
a_toba_r_y3	tobacco, irrigated			X		
a_toba_s_y3	tobacco, non irrigated			X		
a_toba_y3	tobacco			X		X
a_hop_r_y3	hop, irrigated			X		
a_hop_s_y3	hop, non irrigated			X		
a_hop_y3	hop			X		
a_arom_r_y3	aromatic plants, irrigated			X	X	
a_arom_s_y3	aromatic plants, non irrigated			X		
a_arom_y3	aromatic plants			X	X	

Year 4 = 2016 (DZ)	a_othind_r_y3	other industrial crops, irrigated		X				
	a_othind_s_y3	other industrial crops, non irrigated		X				
	a_othind_y3	other industrial crops		X				
	a_ind_crops_r_y3	industrial crops, irrigated		X				
	a_ind_crops_y3	industrial crops		X	X	X	X	X
	a_rts_r_y3	roots, irrigated		X				
	a_rts_y3	roots		X				
	a_rts_s_y3	roots, non irrigated		X				
	a_pota_r_y3	potato, irrigated		X	X			
	a_pota_s_y3	potato, non irrigated		X				
	a_pota_y3	potato		X	X	X	X	X
	a_sugb_r_y3	sugarbeet, irrigated		X	X			
	a_sugb_s_y3	sugarbeet, non irrigated		X				
	a_sugb_y3	sugarbeet		X	X	X		
	a_vege_r_y3	vegetables, irrigated			X	X		
	a_vege_y3	vegetables		X	X	X	X	X
	a_flowers_y3	flowers		X	X	X	X	X
	a_other_arable_y3	other arable crops						X
	a_setaside_y3	set aside		X	X	X	X	X
	a_arable_y3	arable lands (total area)		X	X	X	X	X
	a_citrus_r_y3	citrus, irrigated		X	X	X		
	a_citrus_y3	citrus plantations		X	X	X	X	X
	a_temf_r_y3	temperate fruits, irrigated		X	X			
	a_temf_s_y3	temperate fruits, non irrigated		X				
	a_temf_y3	temperate fruits		X	X	X	X	X
	a_trof_y3	tropical fruits		X				
	a_nutf_r_y3	nut fruits, irrigated		X				
	a_nutf_s_y3	nut fruits, non irrigated		X				
	a_nutf_y3	nut fruits		X				X
	a_table_olive_s_y3	table olive, non irrigated		X				
	a_table_olive_r_y3	table olive, irrigated		X				
	a_table_olive_y3	table olive		X				
	a_oil_olive_y3	oil olive		X				
	a_olive_r_y3	olive, irrigated		X	X			
	a_olive_s_y3	olive, non irrigated		X				
	a_olive_y3	olive		X	X	X	X	X
	a_fruit_olive_vine_y3	total area under fruit plantation, olive, vineyards		X	X	X	X	X
	a_fruit_palm_y3	fruit palms		X				
	a_permcrops_y3	permanent crops			X		X	X
	a_grape_s_y3	grapes, non irrigated			X			
	a_grape_r_y3	grapes, irrigated			X	X		
a_grape_y3	grapes		X	X	X			
a_viney_r_y3	vineyards, irrigated		X	X	X			
a_viney_y3	vineyards		X	X	X	X	X	
a_uaa_y3	UAA		X	X	X	X	X	
a_meadow_y3	meadows		X	X	X	X		
a_uaa_r_y3	UAA, irrigated			X	X			
a_taa_y3	TAA		X	X	X	X		
a_kitchgard_y3	kitchen gardens						X	
bovine_y4	heads of cows		X					
ovine_y4	heads of sheep		X					
caprine_y4	heads of goats		X					
a_uaa_y4	UAA		X					
a_taa_y4	TAA		X					
a_arable_y4	arable lands (total area)		X					
a_setaside_y4	set aside		X					
a_whea_y4	durum wheat		X					
a_c_whea_y4	common wheat		X					
a_barl_y4	barley		X					
a_oat_y4	oat		X					

a_cer_y4	cereals (total area)	X
a_fodder_y4	fodder crops (total area)	X
a_meadow_y4	meadows	X
a_ind_crops_y4	industrial crops (total area)	X
a_vege_y4	vegetables	X
a_greenpea_y4	greenpea	X
a_opul_y4	other pulses	X
a_chic_y4	chickpea	X
a_puls_y4	pulses	X
a_pota_y4	potato	X
a_ind_crops_r_y4	industrial crops	X
a_viney_r_y4	vineyards, irrigated	X
a_olive_y4	olive	X
a_fruit_palm_y4	fruit palms	X
a_citrus_y4	citrus plantations	X
a_temf_y4	temperate fruits	X
a_viney_y4	vineyards	X
a_grape_y4	grapes	X
a_fruit_olive_vine_y4	total area under fruit plantation, olive, vineyards	X

List of tables

Table 1 - Spatial parameters of DIVERCROP multiscale database.....	4
Table 2 - Variables of DIVERCROP database and their sources.....	5
Table 3 - Temporal resolutions of DIVERCROP database.....	5
Table 4 - Land use / land cover variables in DIVERCROP database.....	7
Table 5 - MAPSPAM variables available in DIVERCROP database.....	8
Table 6 - Livestock variables in DIVERCROP database	10
Table 7 - Primary key and aggregators of the Agricultural Censuses Database, corresponding to different levels of the hierarchy of administrative units.....	13
Table 8 - Years for which national agricultural datasets were elaborated. Column headings indicate the year codification in the database (y0 through y4).....	15
Table 9 - Variables of the Agricultural Censuses Database available for most Divercrop countries. backgrounds show: 4 countries (orange), 5 countries (light green), 6 countries (dark green).....	16
Table 10 - Terrain variables in DIVERCROP database.....	17
Table 11 - Bioclimatic variables in DIVERCROP database.....	18
Table 12 - Soils variables in DIVERCROP database.....	19
Table 13 - Attributes of the protected areas in DIVERCROP database	21

List of figures

Figure 1 - The three scales and spatial units of DIVERCROP database.....	4
Figure 2 - Upscaling 300m land cover information to 2km grid.....	6
Figure 3 - Matching 5 arc-minutes grid and nuts 2 with 5 arc-minutes pixels centroids	8
Figure 4 - Downscaling MAPSPAM information from 5 arc-minute pixels to 2km pixels.....	9
Figure 5 - Hierarchy of the European Nomenclature of Territorial Units (left) and correspondence between the Territorial Unit levels and data availability, by country (right)	11
Figure 6 - Number of heads of cows in y2 (y3 in DZ and MT).	17

List of appendices

Appendix 1 - List of the national statistics authorities of Divercrop Countries.....	23
Appendix 2 - Variables of the Agricultural Censuses Database. Orange background: 4 countries, light green background: 5 countries, dark green background: 6 countries.	24