Ref. Ares(2022)1960718 - 17/03/2022

MIND STEP

MODELLING INDIVIDUAL DECISIONS TO SUPPORT THE EUROPEAN POLICIES RELATED TO AGRICULTURE

Deliverable D2.8: Common handbook for maintenance and continuation of interfaces for economic and biophysical data handling for farm modelling at IDM unit levels

AUTHORS	Maria Bielza Diaz-Caneja (freelance consultant)
APPROVED BY WP MANAGER:	Gocht, Alexander
DATE OF APPROVAL:	7-03-2022 (version 2)
APPROVED BY PROJECT COORDINATOR:	Hans van Meijl (WR)
DATE OF APPROVAL:	17-03-2022 (version 2)
WP NUMBER & TITLE	WP2 Data requirements for indicators on European policies related to agriculture and data management
DISSIMINATION LEVEL	**Public**
PROJECT WEB SITE:	https://mind-step.eu

This document was produced under the terms and conditions of Grant Agreement No. 817566 for the European Commission. It does not necessary reflect the view of the European Union and in no way anticipates the Commission's future policy in this area.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 817566.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 773901.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 817566.

TABLE OF CONTENTS

EXECUTIVE SUMMARY 5

1. INTRODUCTION 7

2. LITERATURE REVIEW 8

3. DATA 10

3.1. FARM DATA 10
3.2. SPATIAL INFORMATION 11

4. THE ALLOCATION METHOD 11

4.1. THE ALLOCATION-APPROACH 11
4.2. CONSTRUCTION OF FMU BASED ON THE OLD SPATIAL UNIT HSU 13
4.3. PRIOR INFORMATION ON TYPE OF FARMING 14

5. DESCRIPTION OF THE NEW FARM STRUCTURE SPATIAL UNITS (FSU) AND DIFFERENCES WITH THE PREVIOUS HOMOGENEOUS SPATIAL UNITS (HSU) . 15

5.1. MOTIVATION OF THE FSU UNITS 15
5.2. DELINEATION OF FSU 15
5.3. COMPARISON OF FSU AND HSU 16
5.4. METHODOLOGY OF CALCULATION OF LAND USE AND YIELD PRIORS FOR FSUS 18

6. REQUIRED UPDATES FROM MOVING FROM HSU TO FSU 19

6.1. FARM MAPPING UNITS (FMU): DESCRIPTION, RELATION TO HSU, NEED AND RELATION TO FSU 19
6.1.1. CREATION OF FMU: “CREATE HSU” AND UPDATE OF LFA TO ANC 19
6.1.2. DATA PREPARATION: CALCULATION OF FMU AREA AND YIELD 21
6.1.3. DATA PREPARATION: HARMONIZE ALTITUDE ZONES AND LESS FAVOURED AREAS 21
6.2. USE OF SPATIAL DATA ON TYPE OF FARMING AT A RESOLUTION OF 10 KM^2 GRIDS 23
6.2.1. SOURCE DATA 23
6.2.2. ALLOCATION OF FSS DATA TO FMUS 24
6.3. MATCHING AND MAPPING BETWEEN FSU, FADN REGIONS AND NUTS2 REGIONS 27
6.3.1. SELECTION OF FSU ASSOCIATED TO EACH FADN REGION 27
6.3.2. MATCHING OF FSU AND FADN FOR IMPORTING CAPDIS XOBS FILES 27
6.3.3. MAPPINGS WITH NUTS2 REGIONS IN FADN DATASET 27
6.3.4. MAPPINGS WITH NUTS2 REGIONS IN FSS-EUROSTAT DATA 28

7. CONCLUDING REMARKS, CAVEATS AND IMPROVEMENTS 29

7.1. CONCLUDING REMARKS 29
7.2. CAVEATS AND IMPROVEMENTS 29

8. ACKNOWLEDGEMENTS 30

9. REFERENCES 30

10. APPENDIX 32

10.1. HANDBOOK FOR THE ESTIMATION OF THE SPATIAL DISTRIBUTION OF FADN FARMS 32
10.1.1. INITIALISATION OF THE GUI 32

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement N° 817566.

10.1.1.1. Software installation 32
10.1.1.2. Basic layout of the GUI 32
10.1.1.3. Updating the GUI 36
10.1.1.4. Starting GAMS from GUI 37
10.1.1.5. Option: Estimation from the batch utility 38
10.1.2. FADN DATA FORMATTING: "CONVERT FADN CSV TO GDX" 39
10.1.2.1. Files with variables names 40
10.1.2.2. Launching the conversion 41
10.1.3. CREATION OF FMU: "CREATE FSU" 42
10.1.4. ESTIMATION: RUN FARM ALLOCATION 43
10.1.4.1. Data preparation - Calculation of FMU area and yield 46
10.1.4.2. Data preparation - Harmonize altitude zones and less favoured areas 46
10.1.4.3. Data preparation - Prior information on type of farming 48
10.1.4.4. Data preparation - Output 48
10.1.4.5. Final allocation - Split optimisation by type of farming and altitude/less favoured area classification 49
10.1.4.6. Final allocation - Solve without prior information 49
10.1.4.7. Final allocation - Output 50
10.1.5. ANALYSE FARM ALLOCATION RESULTS 51
10.1.5.1. Analyse farm allocation results 51
10.1.5.2. Exploit results 51
10.1.5.3. Convert GRID to FSU data 54
MIND STEP WP2 TEAM 54
CONSORTIUM DESCRIPTION 54

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 817566.

LIST OF FIGURES

FIGURE 1 CONSTRUCTION OF FARM MAPPING UNITS 13

FIGURE 2 COMBING 10 KM^2 GRID AND FMU DATA 14

FIGURE 3 SKETCH OF THE FSU DELINEATION. FSUS ARE OBTAINED BY THE INTERSECTION OF THE OF 10 KM^2 GRID, ADMINISTRATIVE REGIONS, SOIL MAPPING UNITS AND “NOGO” AREAS. ALL BORDERS FOLLOW THE OUTLINE OF THE MINIMUM MAPPING UNIT OF 1 KM^2 16

FIGURE 4 SIZE DISTRIBUTION OF THE FSU IN EU28. TOTAL NUMBER OF FSU IN EU28: 217884 16

FIGURE 5 CONSTRUCTION OF FARM MAPPING UNITS 20

FIGURE 6 OVERVIEW OF CONTENT OF THE GDX FILE "TRUNK\DATA\EUROSTAT_AND_SHAPE\UAAPRIOR.GDX" 24

FIGURE 7 COMBING 10 KM^2 GRID AND FMU DATA 26

FIGURE 8 HISTORY OF NUTS 28

LIST OF TABLES

TABLE 1 DEFINITION OF ALTZLESF 13

TABLE 2 CALCULATION OF UAA IN FMU 14

TABLE 3 CALCULATE UAA PER FARM TYPE 15

TABLE 4 NEW FSU VERSUS OLD HSU: DIFFERENCES IN DEFINITION 17

TABLE 5 NEW FSU VERSUS OLD HSU: DIFFERENCES IN NUMBER AND SIZE 18

TABLE 6 SETS FROM "\CAPDISHSU\HSU_GRID_MAPS" 24

TABLE 7 PARAMETER/SETS FROM "\CAPDISHSU\P_HSU_GRID10N23" 25

TABLE 8 CALCULATE UAA PER FARM TYPE WITH FSU-FMU 26

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement N° 817566.

ACRONYMS

ALTZ	Altitude Zones
ANC	Areas with Natural Constraints
AROPAj	French optimisation model
CAP	Common Agricultural Policy
CLC	Corine Land Cover
CM	Matching method
CSV	Comma Separated Values
CO	Constraint optimization approach
EU	European Union
FADN	Farm Accountancy Data Network
FAOSTSAT	FAO Statistics
FMU	Farm Mapping Units
FSS	Farm Structure Survey
FSU	Farm Spatial Unit
GAMS	General Algebraic Modelling System
GDX	GAMS Data exchange
GGIG	GAMS Graphical User Interface Generator
GUI	Graphical User Interface
HSMU	Homogenous Mapping Units
HSU	Homogeneous Spatial Units
HTML	Hypertext Markup Language
IACS	Integrated Administration and Control System
INSPIRE	Infrastructure for Spatial Information in the European Community
JRC	Joint Research Centre
LFA	Less Favoured Areas
LPIS	Land Parcel Identification System
LUCAS	Land Use and Coverage Area frame Survey'
MINDSTEP	Project acronym
MLN	Multinomial logit model
N2O	Nitrous Oxide
NUTS	Nomenclature of Territorial Units
UAA	Utilized Agricultural Area

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 817566.

EXECUTIVE SUMMARY

This report summarizes the work carried out for the update of a methodological approach for linking economic and bio-physical data. Although often collected, spatial location of the farms in underlying databases are not available due to confidentiality regulations. This is also the case for the data base FADN used in MINDSTEP to develop single farm models and modules. To overcome this shortcoming researchers have developed in the past different strategies to adjust their models to address spatially relevant topics. The literature review showed that, for EU wide approaches, the locations of farms represented by the FADN sample are estimated as a probability estimate to belonging to a spatial unit with homogenous conditions. The allocation in the Constraint Optimization approach is based on a set of variables (altitude zone, less favoured areas classification, chop/land use shares, crop yields, farm type area shares). Land use shares and expected yields in the spatial unit are calculated by a statistical procedure in the CAPRI model. However, the CAPRI spatial unit definition has been modified so the allocation procedures from previous studies require a full revision and update. This report explains and collects the updates performed and includes as an annex a practical manual or handbook on how to use the allocation software.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 817566.

1. INTRODUCTION

The Common Agricultural Policy (CAP) has increasingly been adapted to integrate environmental concerns and one of the core objectives of the CAP is to ensure a sustainable way of farming and the provision of environmentally beneficial public goods and services. One important lesson from previous CAP evaluations is that some policy effects are difficult to assess at national or even regional levels. Moreover, recent CAP reforms have introduced a set of farm-specific measures whose uptake and economic effects differ significantly between individual farms. Consequently, there is an increasing demand for micro level assessment to fully understand farmer responses to CAP instruments and market signals and to better grasp the net effect of policy measures.

To assess such effects, individual farm models have been developed which require detailed input data. For the European Union (EU), the Farm Structure Survey (FSS) collects information on the whole population of farms each 2nd or 3rd year and publishes results for administrative regions. The Farm Accountancy Data Network (FADN) contains currently around 80,000 farms, representing a population of about 5,000.000 farms in the EU and about 90% of the total agricultural production. Most farm level models for the EU represent the farm population using a sample of individual farms recorded in FADN to enhance the capability in providing scientific support for CAP impact analyses at farm micro level (Offerman et al., 2005; Kellermann et al., 2008; OECD, 2010; De Cara and Jayet, 2011; Gocht and Britz, 2011; Gocht et al., 2013; Louhichi et al., 2015; Louhichi et al., 2018; Ciaian et al., 2020).

Besides capturing economic impacts, those models also aim to contribute to assessing the environmental impacts of the CAP. Therefore, a set of agri-environmental indicators have been developed to enable the environmental assessment of policy measures. While for some agri-environmental indicators the location is not an issue (e.g. energy use), for some others, accurate information of bio-physical endowments of the farm is necessary (e.g. soil erosion, landscape diversity, or biodiversity or GHG emissions). For some indicators, such as N2O emissions from cultivated soils, a strong dependence on environmental conditions such as soil type exist. However robust data bases to develop emission factors by soil types are not yet available. Process-based models introduce further data demand and uncertainties, so that generally simple methods are preferred (Leip et al., 2011a,b). A general limitation for agricultural models is the non-availability of spatially explicit farm data, particularly for models that simulate spatially dependent ecological-economic relationships or try to capture decision-making of actors in a spatial context (Uthes and Kiesel, 2020). Although in the monitoring activities of the EU member states spatially explicit farm data are collected, they are not publicly available due to confidentiality regulations (Schmit et al., 2006).

Task 2.5 of the MINDSTEP project aims at establishing a spatial match between economic statistics and bio-physical data, between socio-economic entities (farms clusters) and environmental conditions. The spatial address of farms, via the location-probability, allows to link economic behaviour to climatic, soil and landscapes information. This is crucial for understanding farmers' behaviour with respect to GHG mitigation and other policy and economic constraints. It also allows to identify at the EU level environmental sensitive areas at high resolution, which were invisible given the assumption of homogenous soil, climate and altitude conditions at NUTS2 or NUTS3. The spatial location and the link to the HSU serve not only for improving the environmental indicators for the EU wide single farm models such as IFM-CAP (Louhichi et al., 2015), but also for FADN based statistical analysis and other impact assessment tools based on FADN.

FADN farms actual location must be kept unknown due to confidentiality agreements. Therefore, probabilistic (Bayesian) approaches have been developed to determine the location of EU farm population represented by FADN farms. Authors have developed different techniques for downscaling economic model results to lower spatial scales, for larger regions such as the entire EU or for smaller regions such as specific NUTS 3 regions (see section 2 Summary of the literature review).

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 817566.

Kempen et al. (2011) developed a method to link the farms in the FADN sample to their environmental endowment (e.g. climate, soil attributes) at the EU-wide scale using a constraint optimization approach (CO). The locations of farms represented by the FADN sample are estimated using small-scale spatial units with homogenous conditions for farming. The resulting spatial allocation of FADN-represented holdings extend the analytical capabilities to agri-environmental evaluation and allows the aggregation of the results to more representative environmental zones (e.g. Nitrate Vulnerable Zones, Areas with Natural Constraints). The estimation results for the whole EU were compared with data from FSS to validate the allocation procedure. Results showed that the suitability of prior information seems to depend on the characteristics of the farm as the prior information on land use shares improves the allocation results for arable and dairy systems, which have a strong land dependence and land use share. However, for farming systems with low or no link to land-use (e.g. pigs, poultry) or farm types with low UAA per farm (horticulture, permanent crops), results were quite weak.

For this reason, a second study was done by Rieger et al. (forthcoming), aiming at improving the allocation fit for low UAA-intensive farm types. They extended the constraint optimization approach of Kempen et al. (2011) by using EU-wide spatial data from the Farm Structure Survey (FSS) provided by Eurostat. These data contains information about the share of UAA per farm type on a 10 km² grid level and was used as additional prior information. In addition, they used Homogeneous Spatial Units (HSU) (Leip et al. 2016) as the initial unit to define farm mapping units (FMU). They conducted the analysis for the EU in 2012 and compared the allocation results with the according representative Farm Structure Survey data (FSS data).

Rieger et al. (forthcoming) were allocating farms to Homogeneous Spatial Units (HSU). The allocation routine is based, among other, on areas and yields by HSU provided by the CAPRI model. As the CAPRI model does not work anymore with HSU but with “Farm Structure Units” or FSU, it is necessary to adapt the allocation routine to these new units. Additional changes are required in the allocation procedure: update from of Less Favoured Areas to new Areas with Natural Constraints, changes and updates in the FADN regions, inclusion of Croatia for which data is now available, etc.

This document is structured as follows: Section 2 summarises the findings made in the literature review deliverable of the MINDSTEP project; next, Sections 3 and 4 describe the data and the methodology selected for farm allocation; section 5 describes the reasons for the shift to the new FSU and their characteristics; section 6 goes through each of the steps of the farm allocation procedure indicating the changes done in each of them. Afterward Section 9 with concluding remarks and caveats and section 10 with the references, the manual or handbook for using the allocation software in included as an annex.

2. LITERATURE REVIEW

In the literature, several approaches were developed to link farming systems and agricultural holdings to the spatial location. Especially relevant in the method developed by Kempen et al. (2011) for the EU. This method links the farms in the FADN sample to their environmental endowment (climate, soil attributes …) at the EU-wide scale using a constraint optimization approach (CO). The locations of farms from the FADN are estimated using small-scale spatial units with homogenous conditions for farming also known as Farm Mapping units (FMU). The spatial unit was defined as an aggregation of the so-called Homogenous Mapping Units (HSMU) defined as areas within an administrative unit with homogeneous location factors (Leip et al. 2008). These are based on the Homogeneous Spatial Mapping Units (HSMUs), defined using a Geographical Information System (Kempen et al., 2007 and Leip et al., 2008). Land use shares and expected yields were assigned to each HSMU by a statistical procedure combining grid observations on land use with available aggregate information at regional level. Information on less favoured areas (LFA) and altitude zones (ALTZ) were added by overlaying

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement N° 817566.

HSMU boundaries with specific thematic maps. The estimation allocates farms with the same ALTZ and LFA status. They estimated a matrix, indicating the percentage of a farm located in a FMU. As a single farm in the FADN sample represents many similar farms, this percentage can also be understood as the share of these farms being allocated to a specific FMU. The farm production mix and yields in FADN should match with the highest possible consistency with those of the spatial units using an optimisation approach by maximizing the probability. The authors of the study concluded that the used prior information was insufficient to allocate certain farm types and proposed to further develop the spatial unit such that it represents homogenous regions of farming systems, instead of single production systems. Results showed that the suitability of prior information seems to depend on the characteristics of the farm as the prior information on land use shares improves the allocation results for arable and dairy systems, which have a strong land dependence and land use share. However, for farming systems with low or no link to land-use (e.g. pigs, poultry) or farm types with low UAA per farm (horticulture, permanent crops), results were quite weak.

The shortcomings in Kempen et al. (2011) have been addressed in several projects for the European Commission conducted by the JRC, Eurocare and Thuenen-Institute to further improve the allocation mechanism (Rieger et al., forthcoming). Major improvements to the study of Kempen et al. 2011 include the usage of the statistical representation factor attached to each FADN farm to allocate them to the spatial units instead of allocation of a particular FADN farm exclusively to the one spatial unit. Instead of HSMU an new spatial unit HSU (Homogeneous Spatial Unit) is used to define farm mapping unit (FMU). In addition, data from Eurostat containing information about the share of UAA (utilised agricultural area) per farm type on a 10kmx10km grid level is integrated as an additional constraint in the CO model to improve the allocation for “land-independent” farm types with low UAA per farm.

Cantelaube et al. (2012) use geographical downscaling to map outputs provided by an economic optimization model AROPAj (Galko and Layet, 2011; Jayet 2020). They do it by estimating the probabilities of FADN farm-groups being in 100 x 100 m grid cells within EU-15 FADN regions. The definition of farm groups is based on altitude level (3 classes), farm type (14 types) and economic size unit. The method by Cantelaube et al. (2012) is rather similar to the one by Kempen et al. (2020). However, while the farm-groups allocation is based on the altitude zone and crop (category) area shares, Kempen et al. (2011) based the allocation of farms (weigthed by their representation factor) on altitude zones, less favoured areas, crop area shares and crops yields, and the revised CO method adds also farm type area share. The CO method is homogeneous for all EU while the farm-group method has different crop types for the different FADN regions (from the AROPAj model). Last, the approach by Cantelaube et al. (2012) has not been evaluated with regard to the actual distribution of farm types using the FSS database.

In the literature, there are other spatial allocation studies not aiming at allocating farms in the EU. Temme and Verburg (2011) proposed a disaggregation approach to spatially distribute agricultural land use intensity, for changes in the CAP between 2000 and 2025. In this study, the NUTS2 nitrogen inputs per crop type from agricultural census are related to LUCAS crop type data, as a first step. Another procedure based on animal density is carried out to estimate nitrogeen use intensity for different types of grasslands. Through multinomial logistic regression, the probabilities of different N intensity levels are estimated using 49 biophysical and socioeconomical location factors. Finally, N intensity areas are spatially disaggregated at 1 km × 1 km through the use of an allocation algorithm.

In a study of Guiomar et al. (2018) a map of Europe has been developed showing regions where small farms have different degrees of importance, in relation to the regional context of agriculture and the territorial characteristics on a NUTS-3 level. In contrast to previous studies that estimate the distribution of different farm types in Europe (e.g. Kempen et al., 2011; Andersen, 2017) this study aims at better considering the particular context of each region for small farms in the EU.

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement N° 817566.

Uthes and Kiesel (2020) state that the usefulness of different approaches depends on the focus and the geographical scale of analysis. They argue that from an EU perspective, it is tolerable to create homogenous entities (e.g. Kempen et al. 2011; Andersen 2017). However, for lower scale spatial analyses, the differences within these entities become increasingly relevant. Particularly for smaller regions, studies use a synthetic landscape approach (Saura and Martinez-Millan 2000; Li et al. 2004; Kellermann et al., 2008). Farms are placed on a grid of the area under study by ensuring that that the grid matches the share of a specific land type[1] (e.g. grassland) (Happe et al., 2006; Happe et al., 2008). Some synthetic landscape approaches in the literature used a 1ha resolution and simple techniques by considering only the grassland share in the total area of farms or farm types (e.g. Uthes et al., 2010, Uthes et al. 2011). In a recent study of Uthes and Kiesel (2020) the authors aim at improving the synthetic landscape approach in terms of resolution (25m x 25m), by considering landscape parameters in the allocation of farms, and use allocation quality indicators that allow for an assessment of the overall allocation result. The main data source is the Integrated Administration and Control System (IACS). This dataset includes an identification system for all farms in an administrative area, covering information on the type of farming, type of business, aggregated spatial information (such as total area, arable area, grassland area, individual crop areas etc.). In addition, it contains the Land Parcel Identification System (LPIS) covering all agricultural land parcels managed by these farms. For the spatial allocation procedure various variables are used such as the total arable area, total grassland area, average arable land quality (measured by the German Ackerzahl), average grassland quality (measured by the German Grünlandzahl), number of hectares located in protected areas for meadow birds and Natura 2000, and the Landbaugebiet to which a farm belongs which is a larger area with similar conditions for farming. The overall allocation quality was relatively high for the considered German case study region Ostprignitz-Ruppin (NUT3 level). The authors conclude that this approach is well suited for smaller regions with sufficient data quality. The approach is suitable to link farm data and spatial data to generate a more realistic synthetic landscape of farm locations, compared to other studies that used simpler spatial allocation procedures. However, the computational time of this approach is high and it has not been tested yet in other regions.

3. DATA

3.1. Farm data

The utilized farm data is obtained via the FADN, a European system of sample surveys conducted every year to collect structural and accountancy data on farms, with the aim of monitoring the income and business activities of agricultural holdings and evaluating the economic impact of the measures taken under the Common Agricultural Policy. The FADN is the only source of micro-economic data harmonized across the EU, i.e. the same book keeping principles apply in each member country. FADN data are collected in all FADN regions, which are not always equal to a particular NUTS level. Exact natural conditions and/or the location of the holdings cannot be derived from the data set mainly for confidentiality reasons. However, some elements of the FADN data represent spatial characteristics relevant for our analysis. For each sample farm, FADN records report whether it is located in a specific altitude zone and in a Less Favoured Area (LFA). Furthermore, many farms are assigned to sub-region

[1] "AgriPoliS models space in a stylistic way by implementing some, but not explicit, spatial relationships, such as the share of land of a particular type. Space is represented by a set of equally sized cells/plots assembled into a chessboard–like pattern". "we randomly allocate each farm agent in the region. (…) However, as mentioned before, we introduce some true statistical relationships such as the share of land types in the virtual region" (Happe et al. 2006)

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 817566.

codes, which can identify lower levels of administrative units (typically NUTS 2 or NUTS 3). Additionally, the land use patterns and crop yields recorded give hints for the spatial location of the farm. Under the FADN methodology, farms are selected for the database according to a sampling plan aiming at representativity of the sample for the population of farms in a FADN region with respect to a classification by type of farming, economic size and region. To allow for corrections of deviations from a perfect stratified sampling, an individual weight (statistical representation factor) is provided for each farm in the sample calculated as the ratio between the total number of holdings in the farm population and the sampled number of holdings in the same classification. (Kempen et al. 2011).

Additionally, we use data received from Eurostat 2010 containing information about the share of UAA (utilised agricultural area) per farm type on a 10 km² grid level. This data will be used as a priori information to improve the allocation procedure of FADN farms and to validate the estimation results for the CO approach. In order to partially compensate the truncation of the Eurostat 2010 data, the adjustment procedure uses also data from the agricultural census (Farm Structure Survey 2010, full survey) at NUTS2 region.

3.2. Spatial information

The most important spatial data in this study are the Homogeneous Spatial Units (HSU). HSUs are defined as clusters of 1 km^2 raster cells within a subnational region (e.g. NUTS2/3) which covers an area of similar characteristics in terms of soil, climate and relief. They are delineated by the intersection of soil mapping units, landform classes and a 0.25 degrees grid (to facilitate the match with meteorological spatial data sets). Land use cover information was excluded as a delineation criterion, except for areas for which a land use change in the near future was considered as highly unlikely (water, ice, barren, built up). The minimum spatial unit is a 1 km^2 raster cell and conform with the European Reference Grid and Coordinate Reference System proposed under INSPIRE (Infrastructure for Spatial Information in the European Community).

4. THE ALLOCATION METHOD

4.1. The allocation-approach

We developed a model that can be applied to each FADN region and year independently. The FADN farms are mapped to spatial units with a homogenous production mix (crop rotation, grassland shares) and a homogenous yield level (tonnes per hectare). The farm production mix and yields in FADN farms should match with the highest possible consistency with the spatial unit using an optimisation approach by maximising the probability. In other words, a dairy farm recorded in FADN with grass and fodder maize should be allocated to the spatial unit with the same production mix and, of course, dairy cows.

The approach is a two-step procedure: first we measure the statistical fit using the approach in Kempen et al. (2011) between similar variables (e.g. yields for wheat ($py_{f,su}$), share of wheat ($ps_{f,su}$)) in FADN and the land mapping unit (equation 1). Although we cannot base prior expectation on an empirical model since the exact location of farms is not published, the farm records include some information limiting the number of HSU where the farm might be allocated. As example from the FADN statistics it can be exactly derived which farms are located in a certain altitude zone and in a LFA area. This information is taken as fixed and given, i.e. if the FADN farm and the HSU do not belong to the same qualification regarding LFA and altitude zone, we do not allow allocation of the farm in that HSU. The prior probabilities in the objective function ($py^0_{f,su}, ps^0_{f,su}$ in Equation 1) can be calculated

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 817566.

from the perspective of the farms (to which FMU should a particular farm be allocated) or from the FMU perspective (which farms should be in a particular FMU). Both views were tested.

The second step ensures consistency by maximising the similarity over all farms and the spatial units. For this purpose, a Bayesian highest posterior density concept (Heckelei et al., 2008) is applied allowing to measure "similarity" with respect to several criteria simultaneously satisfying regional consistency constraints. The final result of our allocation procedure is a matrix $p_{f,su}$ indicating the percentage probability of a farm f located in the spatial unit. As a single farm in the FADN represents many similar farms, this percentage can also be understood as the share of these farms being allocated to a specific spatial unit. An obvious constraint in the allocation procedure is that the share for each weighting over all spatial units must add up to 1 indicated in Equation (3).

Another obvious constraint refers to the utilised agricultural area (UAA). The UAA of a spatial unit should be filled exactly with the UAA represented by the farms assigned to it. This is achieved using the Equation (2), where $UAA_0^{f(FADN)}$ is the utilised agricultural area operated by a FADN farm, weighted with the representativity weight $Weight_f^{(FADN)}$ taken from FADN record, and UAA (spatial unit) the agricultural area in a spatial unit. The use of the representation factor for the allocation procedure means that we assume that all farms represented by the recorded sample farm have the same production mix, yield levels, LFA and altitude information. As FADN data do not fully represent the agricultural area in a region, consistency with the area derived from other sources cannot be expected. The adjustment factor $a^0_{spatial\ unit}$ (a^0_{su}) is a given and fixed correction factor between the land use statistics in FADN and the spatial unit.

Using the logarithm of this, we obtain the final objective function of our problem in the form of a loss function that minimises the sum of the standardised proportional deviations between our prior expectation ($py^0$and ps^0) and the estimates (p):

$$(1) \qquad \min \begin{bmatrix} vec\left(w_y\left(p_{f,su} - py^0_{f,su}\right), w_s\left(p_{f,su} - ps^0_{f,su}\right)\right)' \\ \times \sum_{total}^{-1} vec\left(w_y\left(p_{f,su} - py^0_{f,su}\right), w_s\left(p_{f,su} - ps^0_{f,su}\right)\right) \end{bmatrix}$$

The objective function consists of the true parameters ($p_{f,su}$), where p is the probability of a farm *f* (with f=1 ... N, N being the number of FADN farms) to be allocated to the spatial unit (su), in this study the Farm Mapping Unit (FMU). The prior information enters the objective function in the form of the derived prior probability ($py^0_{f,su}, ps^0_{f,su}$). The standard deviation defines the diagonal elements of the covariance matrix *Σ*. The weighting factors *wy* and *ws* must be set a priori. In our validation, various settings have been tested and compared to find out which setting might produce the best overall results. After the *p* vector by FMU is estimated, the values are mapped back to the HSU vector.

The objective function is subject to the following data constraint

$$(2) \qquad UAA^0_{su} = \alpha^0_{su} \sum_f^c p_{f,su}\, UAA_f^{0(FADN)} Weight_f^{(FADN)},$$

where *c* are the crops.

Another constraint in the allocation procedure is the condition that the probability of a farm *f* sum up to unity over the spatial unit.

$$(3) \qquad \sum_{su} p_{f,su} = 1$$

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 817566.

4.2. Construction of FMU based on the old spatial unit HSU

The farm mapping units (FMU) for the estimation are a subdivision of HSU by altitude and less favoured area class. The FMUs are not created as spatial explicit units, but the data on 1 km^2 grid cells is used to calculate the share of the HSU that belongs to a certain combination of altitude zone and less favoured area class (see Figure 1).

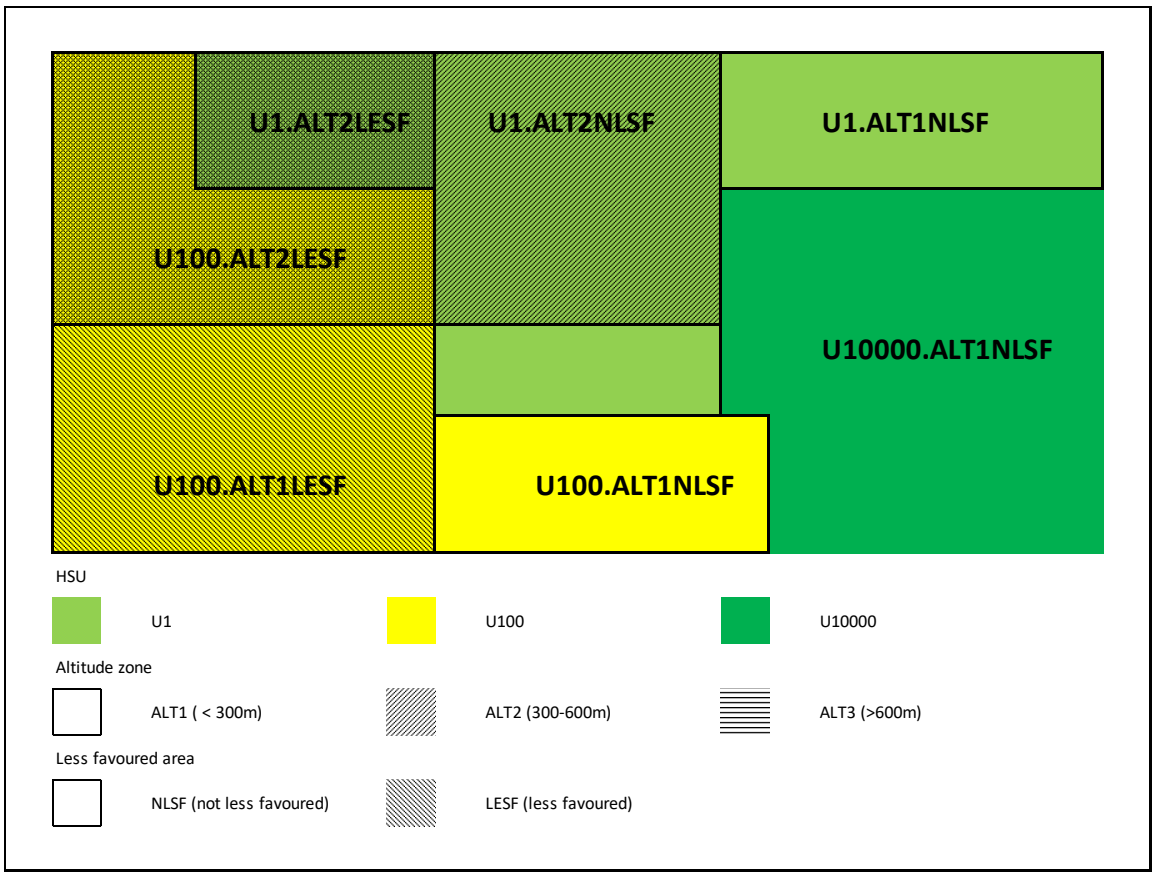


Figure 1 Construction of Farm Mapping Units

These combinations will be referred to as ALTZLESF, and according to their definition there are 6 types (see Table 1)

Table 1 Definition of ALTZLESF

Acronym	Definition
ALT1LESF	altitude < 300 m and less favored zone
ALT2LESF	altitude 300-600m and less favored zone
ALT3LESF	> 600 m and less favored zone
ALT1NLSF	altitude < 300 m and not less favored zone
ALT2NLSF	altitude 300-600m and not less favored zone
ALT3NLSF	altitude > 600 m and not less favored zone

It is calculated how many 1 km^2 USCIE grids belong to the HSU (total) and how many belong to a specific altitude zone and less favoured area combination. Then the share of each ALTZLESF class is calculated. The UAA and yield of FMUs are calculated by assuming that the UAA assigned to the HSU is evenly distributed over the ALTZLESF classes and that the yield is identical in all ALTZLESF classes (see Table 2).

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 817566.

Table 2 Calculation of UAA in FMU

FMU or HSU	km²	% of total	UAA (1000 ha)	Yield (kg)
U100.ALT2LESF	33	35%	133,4	6750
U100.ALT1LESF	40	43%	161,7	6750
U100.ALT1NLSF	21	22%	84,9	6750
U100 (Total)	94			
U100 (from capdis)			380	6750

4.3. Prior information on type of farming

To further improve the allocation mechanism particularly for those farm types with lower land dependency (see Kempen et al 2011), we used data received from Eurostat containing information about the share of UAA (utilised agricultural area) per farm type on a 10 km² grid level. Due to confidentiality manipulation of the Eurostat data an adjustment procedure is implemented to minimise the error accounting for the truncation of the UAA of Eurostat.

Therefore the officially available Eurostat data at NUTS2 level containing information about UAA by farm types was used. We compared the high-resolution data and the official NUTS2 data from Eurostat and implemented the mentioned adjustment procedure. The share of an HSU overlapping with a certain 10 km² grid cell is taken from CAPRI/CAPDIS. The principle is illustrated in Figure 2.

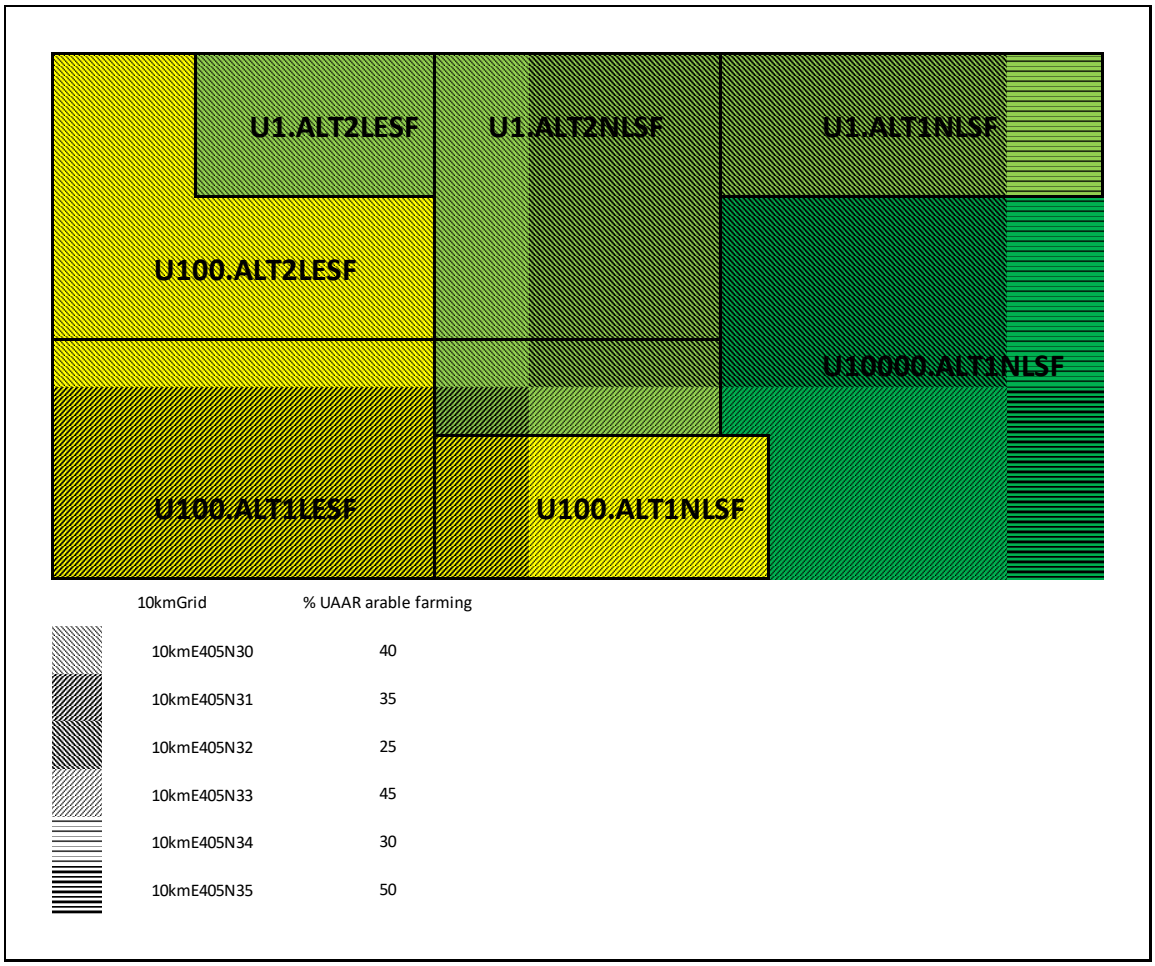


Figure 2 Combing 10 km² grid and FMU data

The calculation of UAA that can be assigned to a certain farm type is visualised in Table 3. The share of area covered by a farm type in a FMU is calculated as the weighted mean of area shares reported

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 817566.

in the 10 km^2 grid data. The UAA covered by a certain farm type is then calculated by multiplying the average share of the farm type with the total UAA assigned to the FMU.

Table 3 Calculate UAA per farm type

FMU	10 km^2 grid	km^2	% of FMU	% UAA arable farming	average % UAA arable	Area UAA	FMU
U100.ALT1LESF	10kmE405N30	8	20%	40%	36%	161,7	58,2
U100.ALT1LESF	10kmE405N31	32	80%	35%			
Total		40			36%	161,7	58,2

5. DESCRIPTION OF THE NEW FARM STRUCTURE SPATIAL UNITS (FSU) AND DIFFERENCES WITH THE PREVIOUS HOMOGENEOUS SPATIAL UNITS (HSU)

5.1. Motivation of the FSU units

A new high-resolution spatial layer was created in the CAPRI model moving from 'Homogeneous Spatial Units' (HSU) to 'Farm Structure Spatial Units' (FSU). The main reason for this was to improve the quality of the downscaled to high resolution mapping of CAPRI outputs. The Farm Structure Survey data, which constitutes one of the most reliable sources of data on agriculture, was made available by Eurostat at high resolution (10 km x 10 km grid). These data provide high quality priors to downscale outputs from agro-economic models. Given that converting from the Eurostat 10 km^2 grid to the HSU units resulted in a loss of accuracy, it was decided to define new spatial units based on the Eurostat grid. Given that the definition of these units and the priors for crop areas, yields, etc. are based on the Farm Structure Survey, they are called Farm Structure Units, from now on FSU.

However, due to confidentiality issues, some of the high resolution FSS data were removed or modified. A gap-filling procedure was implemented for the estimation of these values (see section 5.3).

5.2. Delineation of FSU

Figure 3 outlines the FSU delineation. By intersecting the input parameters, a total of 217,884 FSUs with a median size of 11 km^2 are obtained for the EU28 area (see Figure 4 and Table 5). FSU have a maximum size of 100 km^2 and a minimum size of 1 km^2. The minimum mapping unit (=minimum size of an FSU) is 1 km^2.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 817566.

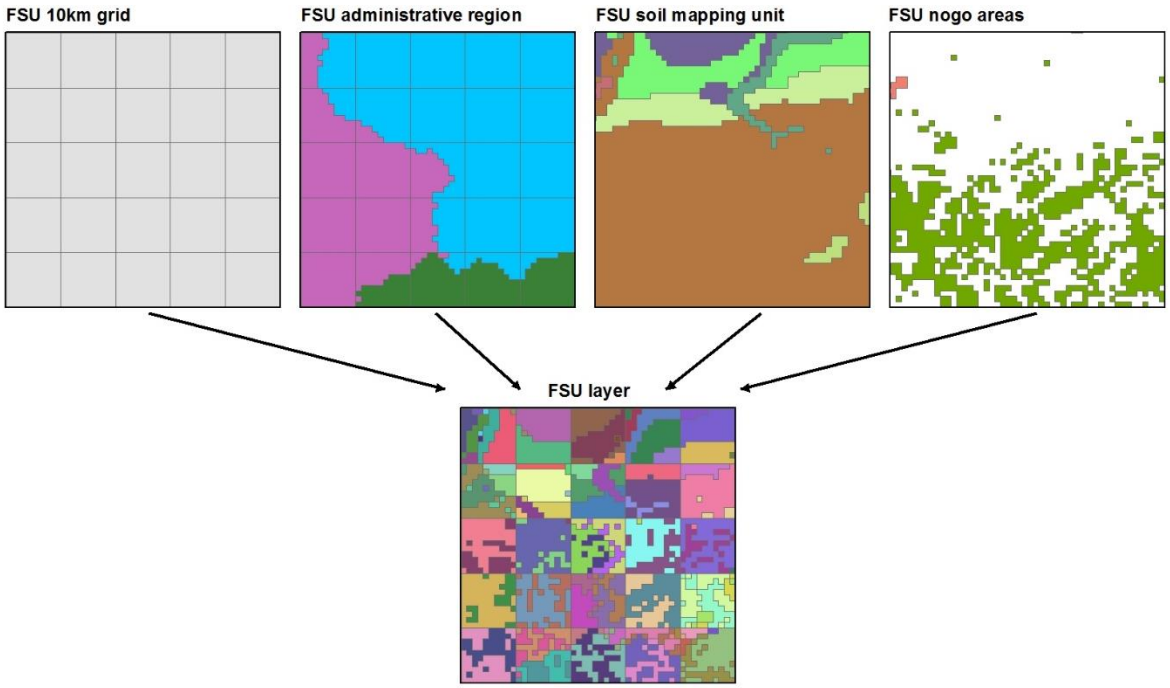


Figure 3 Sketch of the FSU delineation. FSUs are obtained by the intersection of the of 10 km² grid, administrative regions, soil mapping units and "nogo" areas. All borders follow the outline of the minimum mapping unit of 1 km²

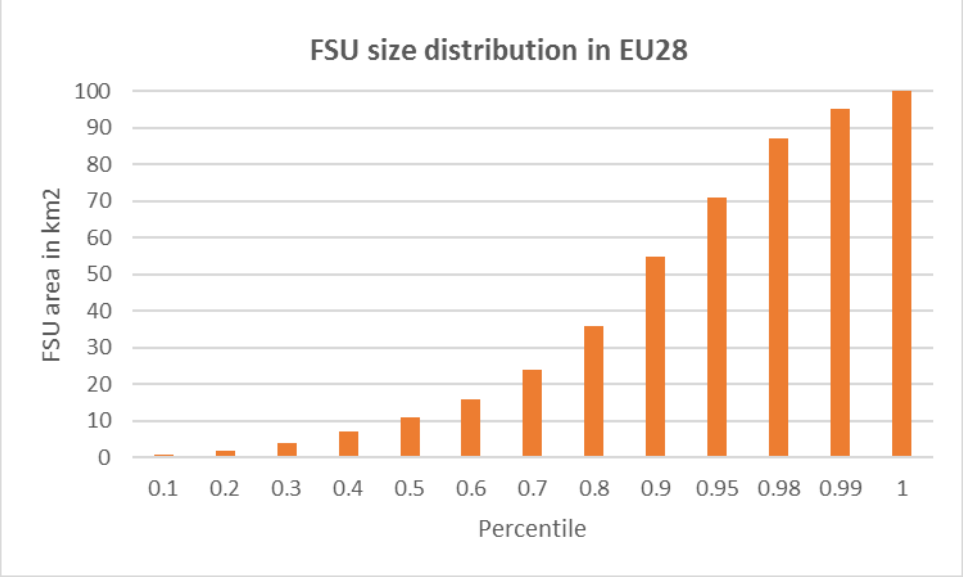


Figure 4 Size distribution of the FSU in EU28. Total number of FSU in EU28: 217884

5.3. Comparison of FSU and HSU

The main differences between HSU and FSU units are (see also Table 4):

- Introducing a 10 km x 10 km grid as delineating factor. This ensures that data exchange with Eurostat is facilitated and statistical information maintained at the highest available resolution. The good results obtained with the gap filling procedure under the

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 817566.

VERIFY project encourages EUROSTAT to provide gridded data also for the upcoming Farm Structure Surveys, thus contributing to a long-lasting legacy of the project

- Removing the 0.25° meteorological grid as delineating factor. The meteorological grid was used in the HSU spatial layer to facilitate the exchange of data with other models, in particular climate models. However, as the meteorological grid does not coincide with any relevant boundary for agriculture, it increased artificially the number of spatial units (burdening the disaggregation process), while a post-processing re-mapping can provide a much faster & flexible solution for data exchange.
- Removing landform classes, as they also increase the number of spatial units and are not strictly necessary.
- Defining strict 'no go' spatial units which are characterized by the dominance of soil-free land such as 'bare soils', 'glaciers', 'water bodies' or 'urban areas'. No agricultural activities are allowed in these units.
- Defining 'no go' spatial units which are areas with >95% of forest according the CORINE 2018 land cover map. In these units, agricultural activities are allowed only if they cannot be allocated elsewhere.
- An update was made on the layer of administrative regions.
- Soil mapping units as delineation parameter remain unchanged

Table 4 New FSU versus old HSU: differences in definition

	Farm Structure Units	Homogeneous Spatial Units
Delimiters / boundaries	(Updated) administrative regions (CAPRI NUTS - NUTS2 or NUTS3 or NUTS1)	Administrative regions (CAPRI NUTS - NUTS2 or NUTS3 or NUTS1)
	Soil mapping units	Soil mapping units
	10 km x 10 km grid	-
	-[2]	0.25° meteorological grid as delineating factor
	-[3]	Landform classes

[2] to exchange with climate models post-processing remapping is used

[3] Landform classes are not used anymore but there is the altitude and slope characterization of the FSU. The file p_fsudem.gdx under CAPRImodel/epnf/dat/capdishsu/ has the distribution of altitude and slope for each FSU. Including the landform would be positive, but it was more important to introduce the 10 km grid, and it became necessary to keep small the number of small 'snippets', which make the whole process very slow without adding real benefit. Some checks done on landform, but also on altitude and slope, showed that within a 10 km grid the distribution is rather random. However, if we know from FSS that wheat is cultivated in a certain cell, then any information coming from large-scale regression or simulation models are not applicable any more. In the creation of the FSU, it was just necessary then to eliminate the land on which no wheat can grow (dense forest, water etc). This was the option by which the best results can be obtained. Altitude and slope are still used, though, when it is necessary to shift crop areas in CAPRI model, choosing FSUs which are more or less steep / high, depending on the crop.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 817566.

	Farm Structure Units	Homogeneous Spatial Units
	(Updated) Strict 'no go' spatial units dominated by soil free land: 'bare soils', 'glaciers', 'water bodies', 'urban areas'	Strict 'no go' spatial units dominated by soil free land: 'bare soils', 'glaciers', 'water bodies', 'urban areas'
	'no go' spatial units with >95% of forest according the CORINE 2018 land cover map. In these units, agricultural activities are allowed only if they can not be allocated elsewhere.	-
Resolution	1 km^2 raster cell (not necessarily contiguous)	1 km^2 raster cell (not necessarily contiguous)

Sources: FSU: VERIFY Deliverable 4.2; HSU: Lamboni et al., 2016.

Table 5 New FSU versus old HSU: differences in number and size

	Farm Structure Units	Homogeneous Spatial Units
Total number in EU28	217,884	150,000 approx. (excluding evident non-agricultural areas)
Average size (EU28)		22 km^2
Median size (EU28)	11 km^2	11 km^2
Minimum size	1 km^2	1 km^2
Maximum size	100 km^2	566 km^2

Sources: FSU: VERIFY Deliverable 4.2; HSU: Lamboni et al., 2016.

5.4. Methodology of calculation of land use and yield priors for FSUs

Land use priors are calculated through an improved Land Area Prediction model (LAPM), based on the methods of Lamboni et al. (2016), and assessing its accuracy (Leip et al., 2017). They include newly available data of crop areas from the Farm Structure Survey 2010 at a 10 km x 10 km grid (FSS2010-10k, EUROSTAT, 2017, personal communication), developing a methodology to gap-fill missing data which was filtered out by confidentiality rules (Bujnowska et al., 2019).

Yields priors include new data on potential and water-limited yields and an updated distribution module for yield and irrigation shares. The nitrogen distribution model was updated. The disaggregation algorithms were improved to enhance performance and speed. For further information see the VERIFY project deliverable 4.2.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 817566.

6. REQUIRED UPDATES FROM MOVING FROM HSU TO FSU

In this chapter the required updates are presented. The more detailed steps and how to apply this in the GUI are explained in the appendix, especially 10.1.3 and 10.1.4.

6.1. Farm Mapping Units (FMU): Description, relation to HSU, need and relation to FSU

The FADN farm spatial allocation procedure takes place at the level of Farm Mapping Units (FMU). However, their definition has changed in time. In the oldest version of the Constraint Optimization method, the old Homogeneous Spatial Mapping Units (HSMU) were used. The implementation of the Homogeneous Spatial Units (HSU) as the final spatial unit in the 2018-19 update required to redefine farm mapping units (FMU), and the same happens for the new FSUs. We next review the different steps in the calculation of FMUs and indicate in red the things that have been modified or updated.

6.1.1. Creation of FMU: “Create HSU” and update of LFA to ANC

The HSU-related farm mapping units (FMU) are a subdivision of HSU by altitude and less favoured area class. The FMUs are not created as spatial explicit units, but the data on 1 km^2 USCIE grid cells is used to calculate the share of the HSU that belongs to a certain combination of altitude zone and less favoured area class (see Figure 5). These combinations will be referred to as ALTZLESF, and according to their definition there are 6 types:

- ALT1LSF: altitude < 300 m and less favoured zone
- ALT2LSF: altitude 300-600m and less favoured zone
- ALT3LSF: altitude > 600 m and less favoured zone
- ALT1NLSF: altitude < 300 m and not less favoured zone
- ALT2NLSF: altitude 300-600m and not less favoured zone
- ALT3NLSF: altitude > 600 m and not less favoured zone

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement N° 817566.

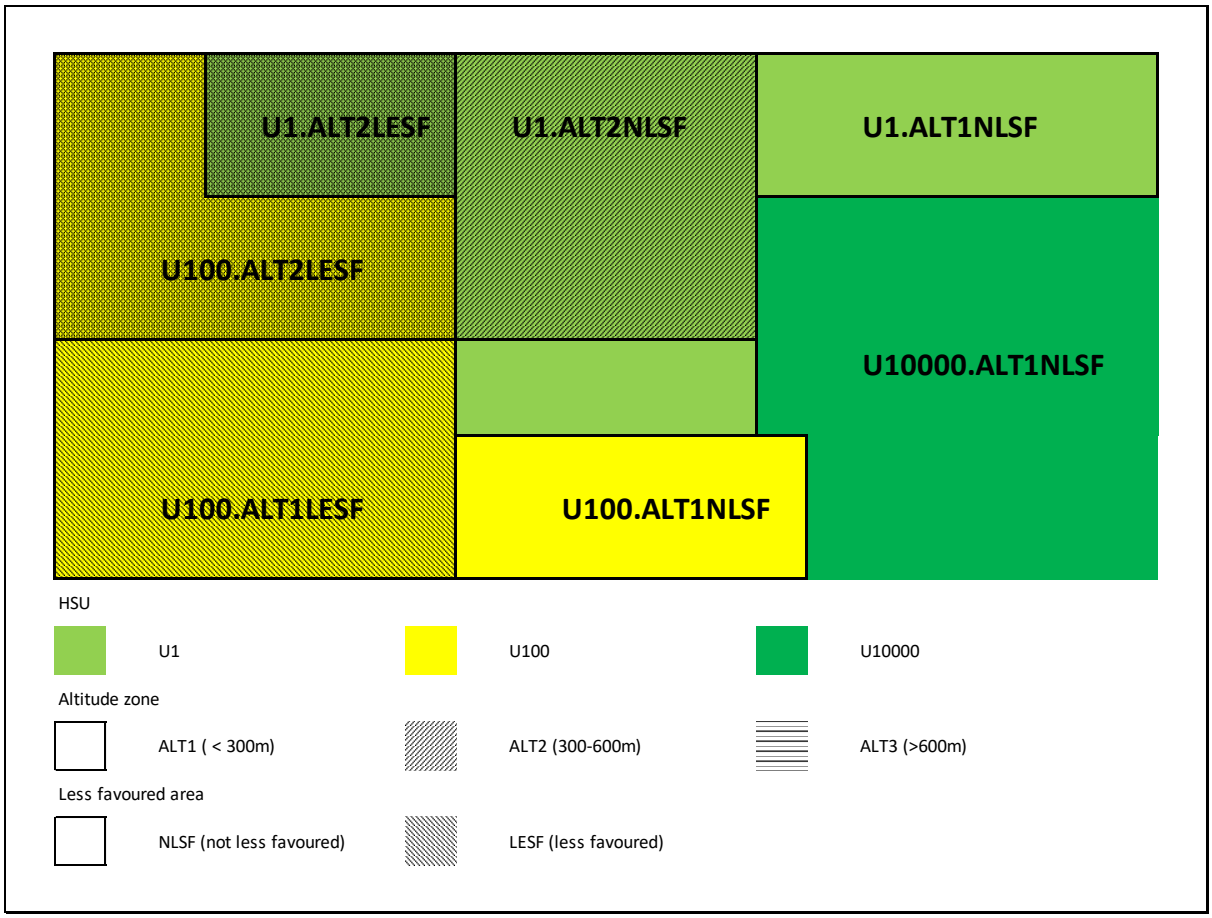


Figure 5 Construction of Farm Mapping Units

It is calculated how many 1 km^2 USCIE grids belong to the HSU (total) and how many belong to a specific altitude zone and less favoured area combination (set ALTZLESF). As the new FSU are also based on the 1 km^2 grid cells, the procedure for the FMU seems fully compatible with FSUs. The calculations are prepared under task "Create HSU" (suggestion: the name should be changed to "Create FMU") in the gams file "Trunk\gams\createHSU_EU_sets_data.gms". The datasets used are:

- The data from 1 km^2 USCIE is stored in gdx format in "data\HSU\USCIE_DEM.gdx" (digital elevation model) and "Trunk\data\HSU\inputdata\USCIE_LFA.gdx". While the digital elevation model does not seem to have suffered changes, "Less Favoured Areas" (LFA) have been replaced in 2013 [4] by "Areas facing natural and other specific constraints" (ANC). FADN data from 2014 already collect this variable: the most similar FADN variable to LFA is ANC3, that is, the grouping of ANC areas in three classes. For a proper match, also the USCIE_LFA map was updated to a more recent USCIE_ANC map. However, this shift from LFA to ANC has been done at different speed in different countries, so that, depending on the year, some countries might be under ANC and other countries still under the old LFA. At the beginning of June 2021, the DG AGRI GIS unit (AGRI C3) provided a Geodatabase with the ANC datasets. However, this map needs an update, as Bulgaria and Italy are still under the former LFA delimitation. They

[4] Regulation (EU) No 1305/2013 of the European Parliament and of the Council of 17 December 2013 on support for rural development by the European Agricultural Fund for Rural Development (EAFRD) and repealing Council Regulation (EC) No 1698/2005.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 817566.

have finalised their ANC delimitation only in 2019/2020, but the work still has to be finalised. The link to the latest version of the ANC map is: https://agridata.ec.europa.eu/extensions/IndicatorsEnvironmental/LessFavouredAreas.html.

- HSU set and mapping of HSU and CAPRI NUTS regions (parameter p_hsu_srnuts2; set hsu(*), m_srnuts2_hsu(*,*)). The first two (set and parameter) have been updated, while m_srnuts2_hsu is calculated from p_hsu_srnuts2.
- Mapping of HSU with USCIE cells %datdir%\hsu\USCIE_HSU2.gms. This was also updated.

6.1.2. Data preparation: Calculation of FMU area and yield

The FMUs are produced as a combination of HSU with ALTZLESF zones[5] prepared before.[6] During the data preparation phase, the share of each ALTZLESF class in the HSU is calculated. For the calculation of the FMU area and yield, it is assumed that the UAA assigned to the HSU in CAPDIS is evenly distributed over the ALTZLESF classes. The yield is assumed to be identical in all ALTZLESF classes.

These calculations are done in the gams files:

- "Trunk\gams\PREP_HSU_DATA.gms. Loads HSU areas and yields
- "Trunk\gams\PREP_FMUfromHSU.gms. Calculates FMU areas and yields.

This section only required the update of the area and yield by FSU, which is provided in the XOBS files (CAPRI-CAPDIS output) but no other dataset or mappings (apart from the HSU sets names).

6.1.3. Data preparation: Harmonize altitude zones and less favoured areas

The availabe data allows to assign FMU (HSU shares) and FADN farms to 6 ALTZLESF zones. In most FADN regions not all 6 zones occur and quite often, less frequent zones in a region are not occuring in both - FADN and FMU data. This potential inconsistency has to be overcome before the independent allocation models are started since the allocation expects for each FADN farm suitable FMU are availabe and vice versa. This harmonisation of data is done in gams file "Trunk\gams\adjust_ALTZLESF.gms".

The code identifies ALTZLESF zones with inconsistent data among FADN and FMU:

- ALTZLESF zones missing in FADN data but found in FMU data (set ALTZLESF_missFADN_foundFMU)
- ALTZLESF zones found in FADN data but missing in FMU data (set ALTZLESF_foundFADN_MissFMU)

If an inconsistency is detected in a FADN region, a remapping procedure is started, that changes the attributes in FADN and/or FMU data until consistency is achieved. The basic idea is to remap FADN farms and/or FMU areas to a zone that is "as similar as possible" to the original zone.

[5] It is calculated how many 1km USCIE grids belong to the HSU (total) and how many belong to a specific altitude zone and less favoured area combination (set ALTZLESF).

[6] The results (in parameter p_HSU_ALTZLESF) are loaded by the file "Trunk\data\FMU\sets_FMUfromHSU.gms", which is called directly by the parent file "Trunk\gams\farm_alloc_final.gms" before run_data_preparation.gms.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 817566.

Here, the principle remains unchanged with the new FSU. The matching between FMU and FADN farms is done by FADN region. This is the first time in the allocation procedure that there is matching between FMU-FSU and FADN data/regions. How the matching between FSU and FADN regions takes place is next checked. In the file Trunk\gams\adjust_ALTZLESF.gms we have the following sets:

- ALTZLESF found for current FMU (Cur_FMU, apply only to FMU in the CAPRINUTS SRG that match the current FADN region SFREG):
  ALTZLESF_foundFMU(ALTZLESF) = YES$SUM(Cur_FMU, EXIST(CUR_FMU,ALTZLESF));
- ALTZLESF found for current farms (SFANR, probably in the FADN region?):
  ALTZLESF_foundFADN(ALTZLESF)=YES$SUM(SFANR,FANR_ALTZLESF(SFANR,ALTZLESF)) ;

We need to check then how EXIST(CUR_FMU,ALTZLESF) is calculated:

- It is calculated in other files[7] as:
  EXIST(CUR_FMU,ALTZLESF) = YES $ p_DataFMU(CUR_FMU,"UAAR","LEVL",ALTZLESF);
- Being:
  - *p_DataFMU(CUR_FMU,"UAAR","LEVL",ALTZLESF) = p_HSU_ALTZLESF(CUR_FMU,ALTZLESF,"Shar")*OBS(CUR_FMU,"UAAR","LEVL");*
  - *p_HSU_ALTZLESF(CUR_FMU,ALTZLESF,"Shar")$ p_HSU_ALTZLESF(CUR_FMU," Total","CountUSCIE") = p_HSU_ALTZLESF(CUR_FMU,ALTZLESF,"CountUSCIE") / p_HSU_ALTZLESF(CUR_FMU,"Total","CountUSCIE");*
  - *p_HSU_ALTZLESF(CUR_FMU,"Total","CountUSCIE") = SUM(ALTZLESF,p_HSU_ALTZLESF(CUR_FMU,ALTZLESF,"CountUSCIE"));*
- *p_HSU_ALTZLESF(HSU_%countries_fst2%,ALTZ LESF,"CountUSCIE") = SUM((USCIE_HSU_%countries_fst2%(USCIE_%countries_fst2%,HSU_%countries_fst2%) ,ALTZLESF_ALTZ_LESF(ALTZLESF,ALTZ,LESF))$ (USCIE_ALTZ(USCIE_%countries_fst2%,A LTZ)* and
  USCIE_LESF(USCIE_%countries_fst2%,LESF)), 1);
  --> These are the number of USCIE cells (and so the area) by FMU, and includes all the FMU of each country. USCIE_HSU_%countries_fst2% is calculated from All_USCIE_HSU and this is provided by %datdir%\hsu\USCIE_HSU2.gms. The map of USCIE x FSU was needed for this.
- *CUR_FMU(Cur_HSU) = YES;*
  CUR_HSU(hsu) = YES $ SUM(%m_spatunitsrnuts2%(SRG,hsu), 1);
  - *$setglobal m_spatunitsrnuts2 m_srnuts2_hsu;*
  - *m_srnuts2_hsu(srnuts2,%spatunit%_all) = YES;*
    $ p_hsu_srnuts2(%spatunit%_all,srnuts2);
    - *$setglobal spatunit hsu*
    - *$gdxIn %datdir%\capdishsu\p_hsu_grid10n23;*
      $load %spatunit%_all=s_hsu
      --> (all HSU 203,797 are included)
  - SRG is the current CAPRI region(s) that match the FADN region (therefore it is actually using CAPRI regions and not FADN regions):
    SRG(RG) = YES $ SUM(FADN_REG_TO_NUTS(SFREG,RG), 1);

[7] Trunk\gams\Prep_fmuFromHSU.gms called by Trunk\gams\rundatapreparation.gms; The other files are not used anymore: Trunk\gams\PrepFMUFromHSU.gms; Trunk\gams\prep_fmu.gms; Trunk\gams\prep_fmu2.gms.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 817566.

SET FADN_REG_TO_NUTS
--> Cross set that maps FADN regions (SFREG) to CAPRI regions (RG). As FADN regions have changed this set needs to be updated.

- FANR_ALTZLESF(SFANR,ALTZLESF)) would not change, even though FADN regions have changed, because it is calculated from the FADN dataset: Set FANR_ALTZLESF(SFANR,ALTZLESF) is calculated from FANR_LFAZ(SFANR,ALTZ,LESF), this from set FANR_ZONE(SFANR,LFAZ) and this from the FADN DATA parameter.

6.2. Use of spatial data on type of farming at a resolution of 10 km^2 grids

The last major methodological updates done in 2018-19 were:

- the use of HSU as the initial spatial unit to define farm mapping units (FMU) and
- the use of spatial data on type of farming at a resolution of 10 km^2 grids. Spatial data on type of farming is included to improve the allocation mechanism particularly for those farm types with weak results in the predecessor project ("Feasibility study on spatial allocation of FADN farms" under the framework contract "Provision of thematic studies in the field of Agri-Environment"). The 10 km^2 grid data is used to calculate the UAA that is expected to be covered by a certain farm type in each FMU

6.2.1. Source data

Due to confidentiality manipulation of the Eurostat data, an adjustment procedure was implemented to minimise the error accounting for the truncation of the UAA of Eurostat. The output of this is "Trunk\data\Eurostat_and_shape\Adjust_Eurostat_EU.gdx" (see Figure 6), variable v_UAAprior, by the file "Trunk\gams\prep_grid.gms", called by "Trunk\gams\runfinalallocation.gms". There is a specific set defined here region_FREG for matching FSS regions (region) to FADN regions (FREG).

Given that the FSS new data (corresponding to 2020) will not be available before the end of the MINDSTEP project, this will not be updated within the project. However, as FADN regions have changed, the set region_FREG was updated. Also the name was changed to fssreg_FREG for consistency with the new name of FSS regions (fssregion).

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 817566.

Adjust_Eurostat_EU.gdx

Entry	Symbol	Type	Dim	Nr Elem
5	c	Set	1	2
6	cat	Set	1	3
18	consistenc	Equ	2	2,112
4	dummy	Set	1	2
12	ef_m_farmleg_EU_Data	Par	5	2,007
16	err	Var	0	1
1	grid	Set	1	44,498
9	grid_to_region	Set	2	48,076
11	p_prior_info_on_UAA_Allocation	Par	3	266,299
13	p_prior_info_on_UAA_Allocation_diff	Par	5	0
21	p_res	Par	3	4,014
10	p_UAAprior	Par	6	257,288
2	region	Set	1	264
19	regionsolved	Par	2	3,306
14	runreg	Set	1	1
20	SameAs	Set	2	44,792
17	target	Equ	0	1
3	type	Set	1	8
15	v_UAAprior	Var	3	384,608
7	value	Set	1	1
8	years	Set	1	1

p_prior_info_on_UAA_Allocation(region, *, *)

AT11	10kmE477N271	1	270
		3	7
		4	0
		5	5
		7	3 9
		8	0
	10kmE478N265	1	0
		3	0
		4	0
		5	0
		6	11 6
		7	0
		8	0
	10kmE478N266	1	0
		3	0
		4	0
		5	0
		6	114.286

Figure 6 Overview of content of the gdx file "Trunk\data\Eurostat_and_shape\UAAprior.gdx"

6.2.2. Allocation of FSS data to FMUs

The share of a HSU-FMU overlapping with a certain 10 km^2 grid cell is taken from CAPRI/CAPDIS. The following sets are imported (see Table 6 and Table 7):

Table 6 Sets from "\capdishsu\hsu_grid_maps"

Set	Description	Example
grid10n23(*)	NUTS3_Grid	AT111_10kmE480N274 HU221_10kmE481N274 AT112_10kmE482N276
grid10n23_grid(*,*)	NUTS3_Grid x Grid	AT111_10kmE480N274 10kmE480N274 HU221_10kmE481N274 10kmE481N274 AT112_10kmE482N276 10kmE482N276
hsu_grid(*,*)	HSU x Grid	U2129 10kmE457N283 U2129 10kmE457N282 U2134 10kmE457N283 U2134 10kmE458N283

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 817566.

Table 7 Parameter/sets from "\capdishsu\p_hsu_grid10n23"

Parameter/Set	Description	Example (with value if parameter)
p_hsu_grid10n23(*,*,*)	"Mapping between FSS 10 km^2 grid at NUTS2 level" The real definition is under set fssGridPars: "Parameter with mapping between HSU and 10kmNUTS23 grids"	U1 AT111_10kmE480N274 area 1 U1 AT111_10kmE480N274 gridarea 74 U1 AT111_10kmE480N274 fracHSU 1 U10 AT111_10kmE480N274 area 1 U10 AT111_10kmE480N274 gridarea 74 U10 AT111_10kmE480N274 fracHSU 0.5 U10 HU221_10kmE481N274 area 1 U10 HU221_10kmE481N274 gridarea 46 U10 HU221_10kmE481N274 fracHSU 0.5
fssGridPars(*)	Area = unit area; fracHSU = fraction of HSU in gridcell (gridcellarea).	nogo area gridarea fracHSU
s_hsu(*)	HSU	U1 U10 U100

The matching between the 10 km^2 grid and the FMUs is illustrated in Figure 7.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 817566.

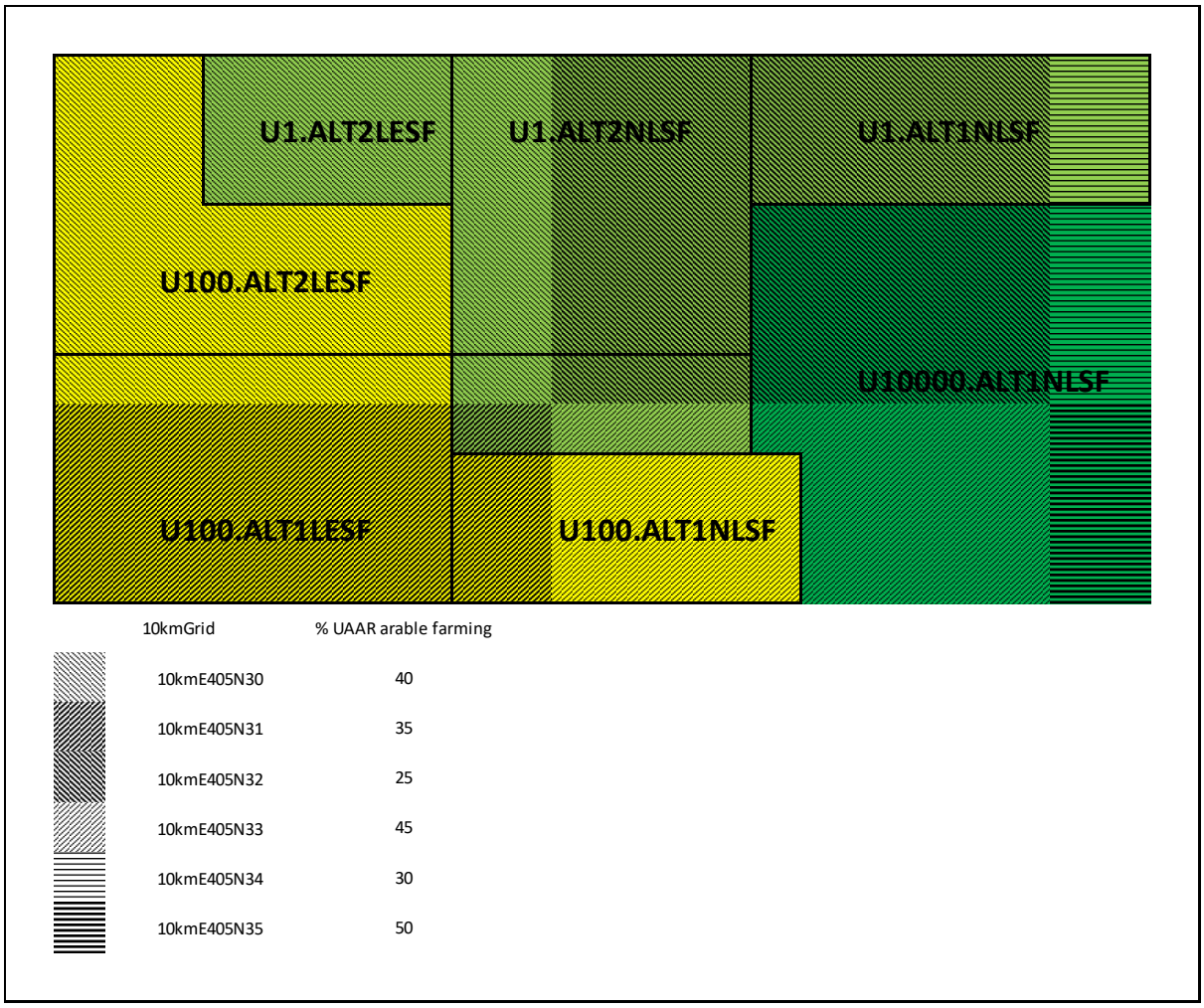


Figure 7 Combing 10 km² grid and FMU data

The calculation of UAA that can be assigned to a certain farm type has been visualised in Table 3. The share of area covered by a farm type in a FMU is calculated as weighted mean of area shares reported in the 10 km² grid data. The UAA covered by a certain farm type is then calculated by multiplying the average share of the farm type with the total UAA assigned to the FMU.

With the new FSU, the intersection of FSU with 10 km² grid is direct as the 10 km² grid is defining FSU, but as FMU are limited to altitude zones and less favoured area types, the matching needs still to be done.

Table 8 Calculate UAA per farm type with FSU-FMU

FMU	10 km² grid	km²	% of FMU	% UAA arable farming	average % UAA arable	Area UAA	FMU
U100.ALT1LESF	10kmE405N30	0	0%	0%	35%	161,7	56,6
U100.ALT1LESF	10kmE405N31	40	100%	35%			
Total		40			35%	161,7	56,6

According to this method we get the same values for all the FMU in the FSU. So this step could be simplified (see Table 8).

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 817566.

Sets and parameters that have been updated are:

- hsu_grid(*,*) in %Datdir%\capdishsu\hsu_grid_maps
- p_hsu_grid10n23(*,*,*)Parameter with mapping between HSU and 10 km^2 NUTS23 grids with:
  - area: seems it is not used
  - Gridarea: seems it is not used
  - fracHSU (= fraction of HSU in gridcell): it would be always 1.

  ➔ Therefore parameter p_hsu_grid10n23 is not needed anymore.

6.3. Matching and mapping between FSU, FADN regions and NUTS2 regions

6.3.1. Selection of FSU associated to each FADN region

The matching between HSU and the FADN regions is done though SET FADN_REG_TO_NUTS cross set that maps FADN regions (SFREG) to CAPRI regions (RG) (and CAPRI provides the mapping between CAPRI regions and HSU). As new FSU are mapped to CAPRI regions (CAPRINUTS), there is nothing to update from this side.

However, as some changes in the FADN regions, therefore FADN_REG_TO_NUTS and some other sets need to be modified accordingly (see Annex 9 for more details).

6.3.2. Matching of FSU and FADN for importing CAPDIS Xobs files

For the importing of the crop areas and yields by FSU (CAPDIS Xobs files) some changes have taken place:

- In the old HSU version, Xobs files (…capdis\xobs_2_%MS%_1212.gdx) were provided by country, therefore the country xobs file was imported based on the matching between MS and FADN region from files setglobalMS.gms and setglobalMS_res.gms.
- In the new FSU version, Xobs files (…capdis\xobs_2_%splitregions%_1212.gdx) are provided by new CAPDIS CAPRI split regions.

As the split regions do not match FADN regions it is necessary in any case to aggregate the different split-regions Xobs files. However, as the aggregation to MS can give very large files, the option taken has been to aggregate into new Xobs file not per county but by CAPRI regions or aggregates of those in such away they best match the FADN regions. This aggregation of split-region Xobs files into new aggregate regions Xobs files has been done outside the allocation procedure.[8] Then, we have added a new file similar to setglobalMS.gms linking the FADN regions to the mixed MSorNUTS regions (Gams\setglobalMSorNUTS.gms).

6.3.3. Mappings with NUTS2 regions in FADN dataset

It must be noted that NUTS2 regions have also changed in the FADN dataset. Information on historic changes in NUTS regions can be found in: History of NUTS - NUTS - Nomenclature of territorial units for statistics - Eurostat (europa.eu) (https://ec.europa.eu/eurostat/web/nuts/history) – see also Figure 8.

[8] The gams file Reaggregating_CAPDIS-Xobs.gms has been used for this.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 817566.

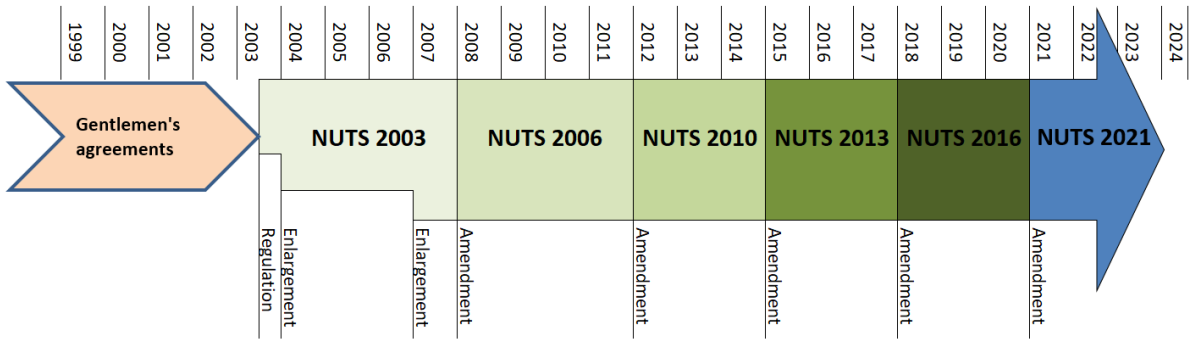


Figure 8 History of NUTS

The FADN data received now use NUTS2016 names while previous datasets most probably used an older version (no way to verify now as the data has been deleted according to confidentiality agreement). However, as we are mostly using FADN regions and CAPRI and FSS regions, but not NUTS2/NUTS3 regions from FADN dataset, these changes do not need an update, as shown here:

- The set key_to_nuts2_csv((allFD,SelNuts2) is linked to 4 digits NUTS2 regions, comes directly from Java importing of FADN data and was used only for statistics tests and is not used anymore.
- The set Key_to_Nuts2(Key,RG) linked to CAPRI regions, comes also from Java importing of FADN data, but it is only used:
    - to calculate RESULT_NUTS but this is not used later.
    - to calculate FANR_REG(Key,FREG_N2) through also FADN_REG_TO_NUTS(FREG_N2,RG), but it is also not used, only FANR_REG(Key,FREG) is used

6.3.4. Mappings with NUTS2 regions in FSS-Eurostat data

NUTS2 regions are used for the Eurostat FSS priors, however, these have not been updated. In the current 2010 dataset, the Code 2010 was used (under data\Eurostat and shape, Eurostat –FSS NUTS2 regions and the mappings), but also codes 2013 were included in some sets and in the data with no values (e.g. Eurostat_EU.gms).

For example, we find:

EL53 from code 2013 under set:

- regions in file data\Eurostat_and_shape\combine_FADN_Eurostat_Eu

EL13 from code 2010 under sets:

- regions, nuts2, mappingNUTS2_NUTS1(NUTS2,NUTS1) in file data\Eurostat_and_shape\combine_FADN_Eurostat_Eu
- region (singular) in file region.gms called by data\Adjust_Eurostat_differences_EU.gms

For the future, it can be interesting to know the past NUTS2 code changes. They are reported in under sheet "Correspondence NUTS-2" in the excel tables that can be found at https://ec.europa.eu/eurostat/web/nuts/history.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 817566.

7. CONCLUDING REMARKS, CAVEATS AND IMPROVEMENTS

7.1. Concluding remarks

The work objective was to update and develop methodologies to merge economic (full population data estimated from FADN survey data) and biophysical data sets of high spatial and temporal resolution. The methodology builds on biophysical data bases provided at the level of Spatial Units, which have shifted from HSU to FSU, and socio-economic data by geo-referencing farm clusters. The result is a location-probability of a sample farm or farms represented by it, which allows to link economic behaviour to climatic, soil and landscapes information.

The following categories of data, parameters, sets and code of the FADN spatial allocation routine have been updated.

- HSU/FSU sets, parameters, data and their reading by the programme
- LFA/ANC area maps and their adaptations to different mappings depending on the years of the FADN data
- FADN regions sets

The updates of the software have been implemented and the tests performed by the Thuenen institute are positive although still need more exhaustive quality checking. Running times are still long but run smoothly, without errors on infeasibilities.

7.2. Caveats and improvements

The data on Areas with Natural Constraints (ANC) is ready for use in the code for running until 2020 included. The Geodatabase with the new ANC datasets was provided by the DG AGRI GIS unit (AGRI C3) at the beginning of June 2021. However, ANC areas have been progressively implemented by the MS (this has been considered in the routine) and the ANC map provided did not include BG Bulgaria and Italy, as they were still under the former LFA delimitation. These two countries have drafted their ANC delimitation only in 2019/2020, but the work still has to be finalised. Therefore, in order to run the spatial allocation routine for FADN data from 2021 on, it should be updated including also ANC for Italy and Bulgaria. The link to the latest version of the ANC map is: https://agridata.ec.europa.eu/extensions/IndicatorsEnvironmental/LessFavouredAreas.html.

There is a new variable in FADN which was not considered in the allocation software but could be used for a more precise allocation. This variable is _TA, which is defined as total area _A plus follow-up crop areas. This variable could be used for a more accurate calculation of yields. However, it contains some errors or inconsistencies that would need to be treated/corrected through the use of specific conditions.

For the next Eurostat FSS data update, it will be necessary to revise the NUTS2 codes, as they have changed in the different years, and the sets/mappings include the code of 2010.[9]

[9] It is advisable, when this will be done, to change the name “region” (in singular) to “fssregion” not to create confusion with the new FADN regions which are also called “region”.

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement N° 817566.

8. ACKNOWLEDGEMENTS

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 817566. Neither the European Commission nor any person acting on behalf of the Commission is responsible for how the following information is used. The views expressed in this publication are the sole responsibility of the author and do not necessarily reflect the views of the European Commission.

Reproduction and translation for non-commercial purposes are authorised, provided the source is acknowledged and the publisher is given prior notice and sent a copy.

9. REFERENCES

Andersen, E. (2017). The farming system component of European agricultural landscapes. European Journal of Agronomy, 82, 282-291.

Cantelaube, P., Jayet, P.A., Carré, F., Bamps, C., Zakharov, P. (2012). Geographical downscaling of outputs provided by an economic farm model calibrated at the regional level. Land Use Policy, 29, 35–44.

Ciaian, P., Espinosa, M., Louhichi, K., Perni, A., Publique, I. U. E., & Thiverval-Grignon, F. (2020). Farm Level Impacts of Trade Liberalisation and CAP Removal Across EU: An Assessment using the IFM-CAP Model. The German Journal of Agricultural Economics, 69(2), 108–126

De Cara, S. and Jayet, P.A. (2011). Marginal abatement costs of greenhouse gas emissions from European agriculture, cost effectiveness, and the EU non -ETS burden sharing agreement. Ecological Economics, 70, 1680–1690.

Galko, E., & Jayet, P. A. (2011). Economic and environmental effects of decoupled agricultural support in the EU. Agricultural Economics, 42(5), 605-618.

Gocht, A. and Britz, W. (2011). EU-wide farm type supply models in CAPRI – how to consistently disaggregate sector models into farm type models. Journal of Policy Modeling, 33, 146–167.

Gocht, A., Britz, W., Ciaian, P. and Gomez y Paloma, S. (2013). Farm type effects of an EU-wide direct payment harmonisation. Journal of Agricultural Economics, 64, 1–32.

Guiomar, N., Godinho, S., Pinto-Correia, T., Almeida, M., Bartolini, F., Bezak, P., ... & Wästfelt, A. (2018). Typology and distribution of small farms in Europe: Towards a better picture. Land use policy, 75, 784-798.

Happe, K., Kellermann, K., & Balmann, A. (2006). Agent-based analysis of agricultural policies: an illustration of the agricultural policy simulator AgriPoliS, its adaptation and behavior. Ecology and society, 11(1).

Heckelei, T., Mittelhammer R., Jansson T. (2008). A Bayesian Alternative to Generalized Cross Entropy Solutions for Underdetermined Econometric Models. Agricultural and Resource Economics, Discussion Paper, Institute for Food and Resource Economics, Bonn. https://ideas.repec.org/p/ags/ubfred/56973.html#download

Jayet, P.-A., Petsakos, A., Chakir, R., Lungarska, A., De Cara, S., Petel, E., Humblot, P., Godard, C., Leclère, D., Cantelaube, P., Bourgeois, C., Clodic, M., Bamière, L., Ben Fradj, N., Aghajanzadeh-Darzi, P., Dumollard, G., Isbăşoiu, A., Adrian, J., Pilchak, G., Bounaffaa, M., Barberis, D., Assaiante, C., Ollier, M., Henry, L., Florio, A., Chidami, I., Gossiaux, E. (2020). The European agro-economic AROPAj model.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 817566.

INRA, UMR Economie Publique, Thiverval-Grignon. https://www6.versailles-grignon.inra.fr/economie_publique_eng/Research-work.

Kellermann, K., Happe, K., Sahrbacher, C., Balmann, A., Brady, M., Schnicke, H. and Osuch, A. (2008). AgriPoliS 2.1 – Model Documentation. Technical Report, Leibniz Institute of Agricultural Development in Transition Economies (IAMO), Halle.

Kempen M., Elbersen, B. S., Staritsky, I., Andersen, E., and Heckelei, T. (2011). Spatial allocation of farming systems and farming indicators in Europe. Agriculture, Ecosystems & Environment, 142 (1-2), 51-62.

Lamboni, M. R. Koeble and A. Leip (2016). Multi-scale land-use disaggregation modelling: concept and application to EU countries. Environmental Modelling & Software 82, 183-217.

Leip, A., Busto, M., Corazza, M., Bergamaschi, P., Koeble, R., Dechow, R., Monni, S., de Vries, W. (2011a). Estimation of N2O fluxes at the regional scale: data, models, challenges. Curr. Opin. Environ. Sustain. 3, 328–338. https://doi.org/10.1016/j.cosust.2011.07.002

Leip, A., Busto, M., Winiwarter, W. (2011b). Developing spatially stratified N2O emission factors for Europe. Environ. Pollut. 159, 3223–3232. https://doi.org/10.1016/j.envpol.2010.11.024

Leip, Adrian; Koeble, Renate; Reuter, Hannes; Lamboni, Matieyendou (2016): Homogeneous Spatial Units (HSU) - a Pan-European geographical basis for environmental and socio-economic modelling. PANGAEA, https://doi.org/10.1594/PANGAEA.860284, Supplement to: Leip, A et al. (in prep.): Homogeneous Spatial Units (HSU) - a Pan-European geographical basis for environmental and socio-economic modelling. Scientific Data

Louhichi, K.; Ciaian, P.; Espinosa, M.; Colen, L.; Perni, A. and Gomez y Paloma, S. (2015). An EU-wide individual farm model for Common Agricultural Policy Analysis (IFM-CAP): first application to crop diversification policy. JRC Science and Policy Reports EUR 20910 EN, European Commission. http://publications.jrc.ec.europa.eu/repository/bitstream/JRC92574/jrcreport_jrc92574.pdf

Louhichi, K., Ciaian, P., Espinosa, M., Perni, A., & Gomez y Paloma, S. (2018). Economic impacts of CAP greening: application of an EU-wide individual farm model for CAP analysis (IFM-CAP). European Review of Agricultural Economics, 45(2), 205-238.

OECD (Organisation for Economic Co-operation and Development) (2010). Linkages between Agricultural Policies and Environmental Effects: Using the OECD Stylised Agri-environmental Policy Impact Model. OECD Publishing .Available online: http://dx.doi.org/10.1787/9789264095700-en

Offermann, F., Kleinhanss, W., Huettel, S. and Kuepker, B. (2005). Assessing the 2003 CAP reform impacts on German agriculture, using the farm group model FARMIS. In: Arfini, F. (ed.), Modelling Agriculture Policies: State of the Art and New Challenges, proceedings of the 89th European Seminar of the European Association of Agricultural Economists (EAAE), Parma, Italy, February 3-5, 2005. Parma: Monte Universita Parma Editore, 546–564.

Rieger, J., A. Gocht, M. Kempen, M. Bielza, J.-M. Terres, S. Neuenfeldt (forthcoming). Linkage of economic and bio-physical data for a more targeted agricultural and environmental policy in the EU. Work document.

Schmit, C., Rounsevell, M.D.A., La Jeunesse, I. (2006). The limitations of spatial land use data in environmental analysis. Environmental Science & Policy, 9, 174–188.

Temme, A. J. A. M., & Verburg, P. H. (2011). Mapping and modelling of changes in agricultural intensity in Europe. Agriculture, Ecosystems & Environment, 140(1-2), 46-56.

Uthes, S., Kiesel, J. (2020). Creating a synthetic landscape: Spatial allocation of non-spatial single farm data. Agricultural Systems, 177, 102740.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 817566.

10. APPENDIX

10.1. HANDBOOK FOR THE ESTIMATION OF THE SPATIAL DISTRIBUTION OF FADN FARMS

10.1.1. Initialisation of the GUI

10.1.1.1. Software installation

As a first step the TortoiseSVN software has to be installed which helps different users to manage different versions of the source code for their programs (version, revision and source control). Afterwards the software for the farm allocation procedure can be downloaded from the Thuenen repository (with user permission and a password) from: "https://svn1.agp.uni-bonn.de/svn/FADN_spatial/". The software is contained in the Trunk folder or directory ("https://svn1.agp.uni-bonn.de/svn/fadn_spatial/trunk"). To open the program, the Graphical User Interface (GUI), double click in the file "\Trunk\GUI\ start_FADNspatial.bat".

10.1.1.2. Basic layout of the GUI

The manual is included in "Trunk \doc" which can be accessed via the GUI (see Figure 1).

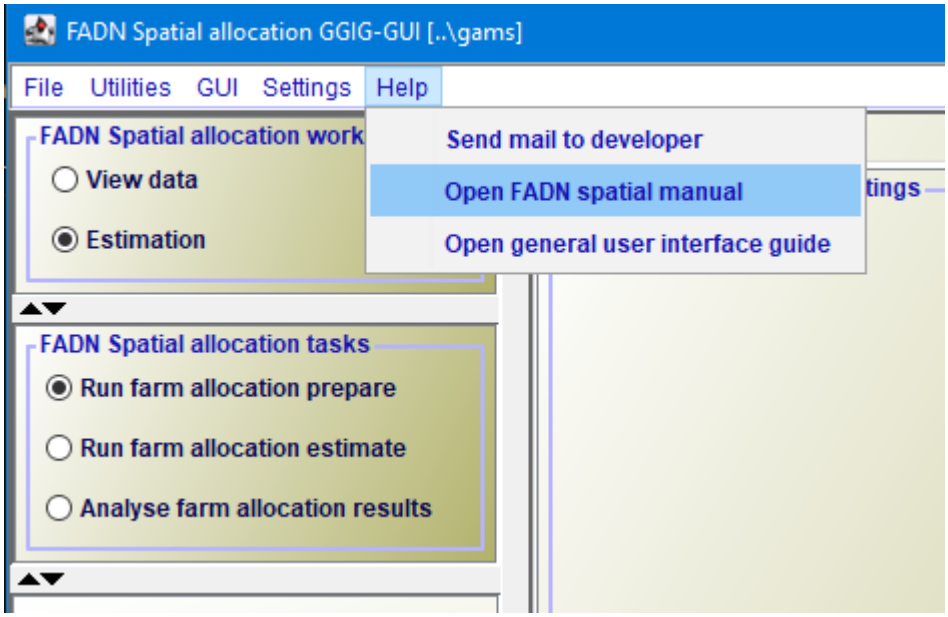


Figure A 1 Link to the manual and the project report in the GUI

The GUI is generally structured as seen in Figure 2. The left upper hand panel allows the selection of the different FADN Spatial work steps. The left lower hand panel lists the tasks belonging to the work steps. In both cases, only one button will be active. The right-hand side offers controls depending on the properties of the task, grouped on different panes. There are buttons allowing starting the task, and a window which collects information at runtime. The footer lists the user name and type, and comprises a progress bar. For tasks linked to a GAMS program, the buttons as shown below will be active:

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 817566.

- Compile GAMS: starts the GAMS compiler, but does not execute the program. A listing file will be generated. Used to test if a program compiles without errors (more information in section 3.4)
- Start GAMS: tries to execute the GAMS program. A listing file will be generated where possible compilation or run-time errors are reported
- Stop GAMS: sends a "signal interrupt" to the GAMS engine. It may take a while until GAMS reacts and stops with an error message after running its finalization routines
- Hide/Unhide controls: allows hiding the control panels so that only the generated listing file is visible
- Exploit results: open the scenario exploiter

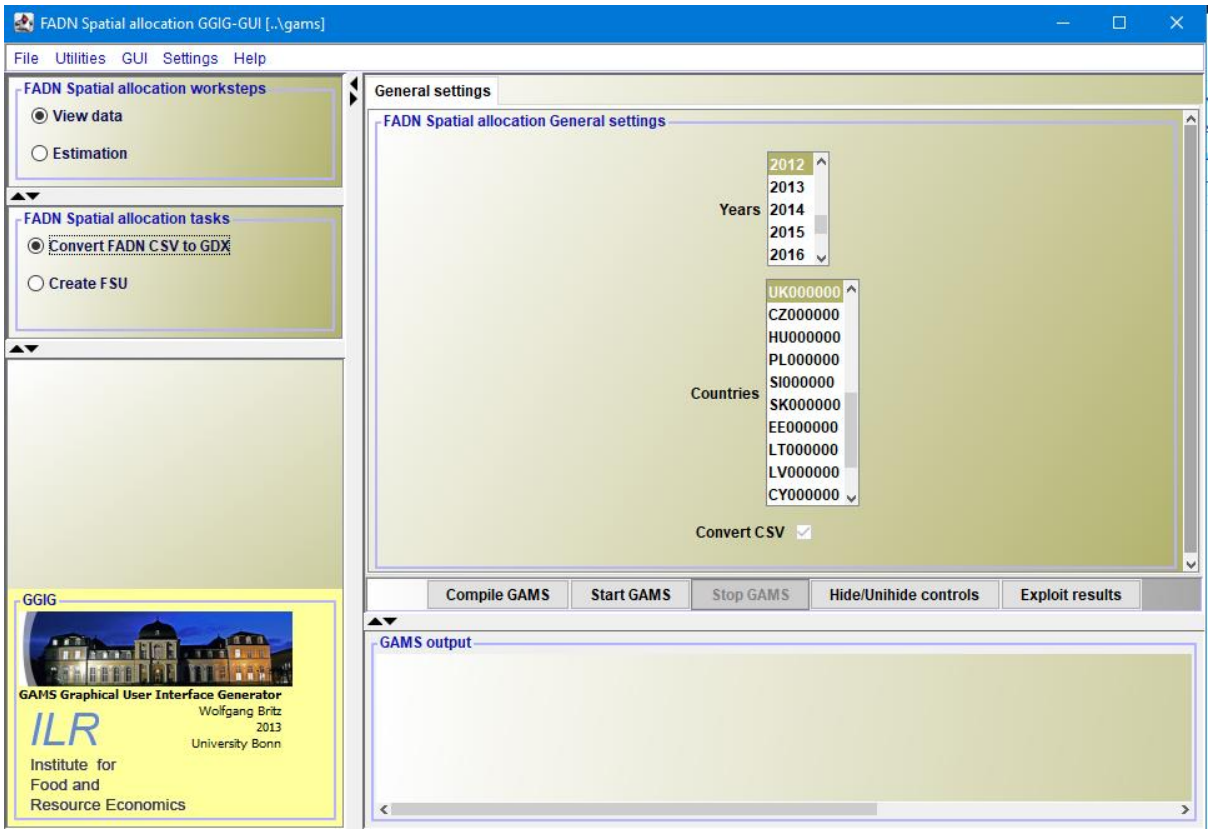


Figure A 2 Basic layout of the GUI

The interface has a few standard settings which can also be accessed via the "edit settings" dialogue.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 817566.

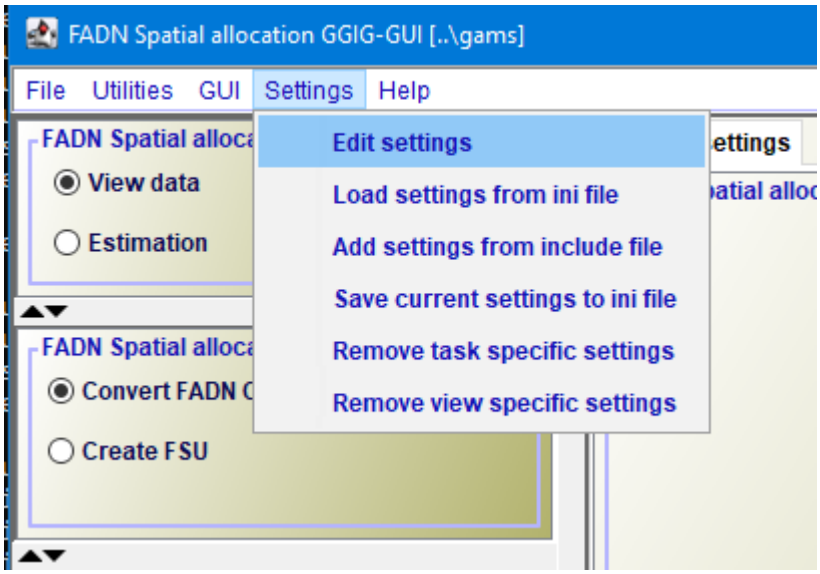


Figure A 3 Edit settings in the GUI

These are (see Figure A 4):

Username: Put your preferred username: e.g. Maria

User type: Runner

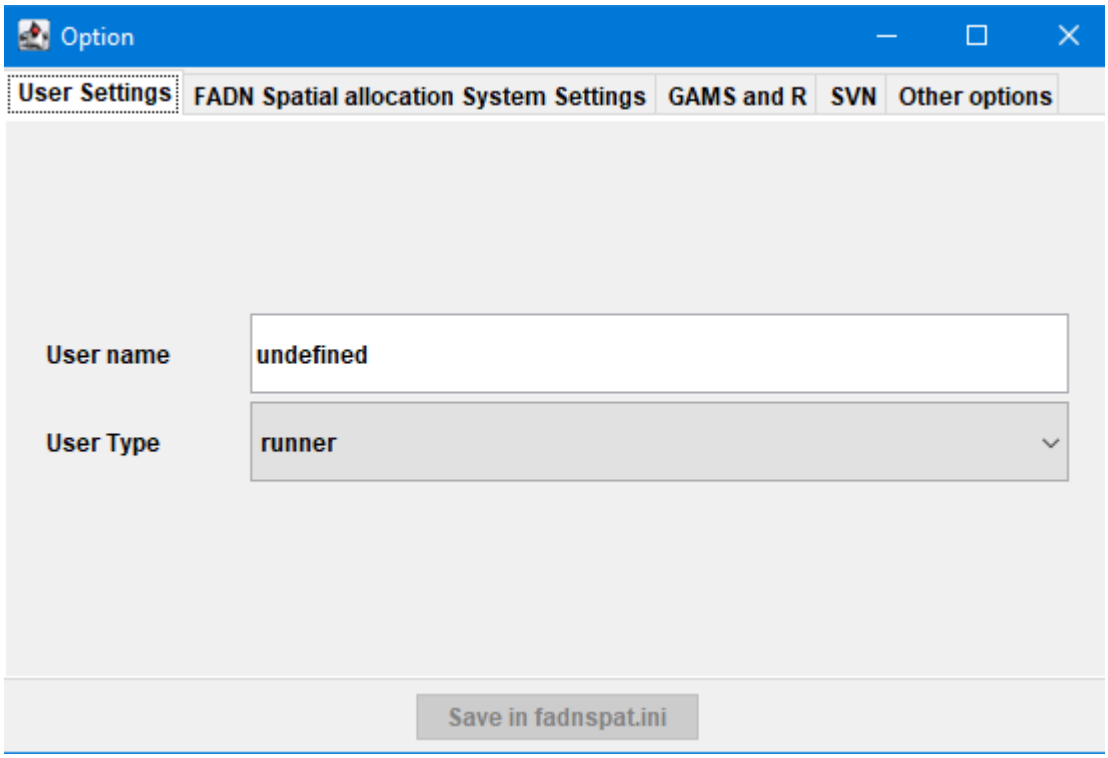


Figure A 4 User settings in the GUI

In the "System Settings" relevant file locations are shown (see Figure A 5):

- Result directory: the directory where GDX files for results are assumed to be stored (corresponds to GGIG name resDir): "..\Results"

The other three directories can be used to adjust the specific model application:

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 817566.

- Model files directory: the root of the GAMS file (modelDir in GGIG)
- Restart directory: a directory for restart files (restartDir)
- Data files: directory for data files "..\data or path\data" (datDir)

If it appears in red it means that it is not found. Thus the directory needs to be created or another directory path has to be indicated. If the program is not under a secured zone (for data confidentiality), at least the data and results directory should be put in a secured area so that they can only be accessed from the pcs with permission, and the path to this area should be indicated in front of the relevant directories.

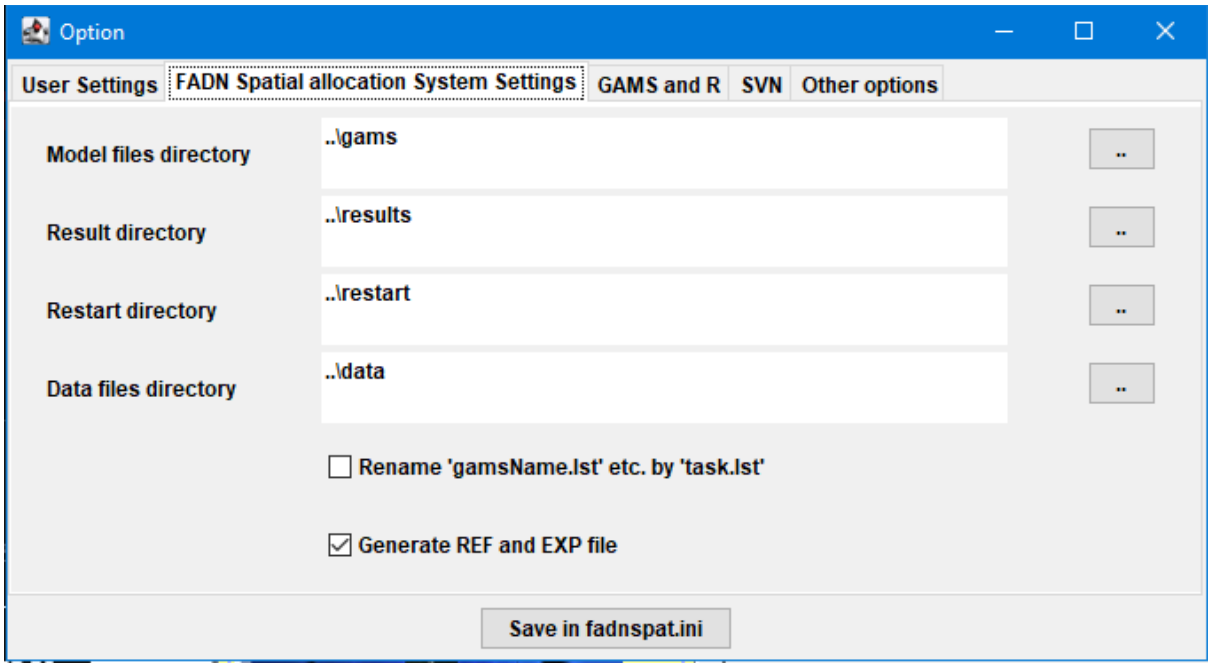


Figure A 5 GUI system settings

These file locations are passed to GAMS and can be used in the GAMS code to read / include files from the correct locations on disk. In order to make an initialisation file portable, locations can be defined relative to the GUI directory (..\).

Under the "GAMS and R" settings (see Figure A 6) various options are available and the relevant ones are explained below:

- Path to GAMS.exe: Indicates the path where the gams.exe program is located on your computer
- GAMS options: Threads 0 is ok
- Number of processors used in GAMS: an advisable number can be calculated by clicking on "Get the number of processors" and substracting 2
- Processor speed relative: 100%
- Allow parallel processing of tasks in interactive mode: Yes

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 817566.

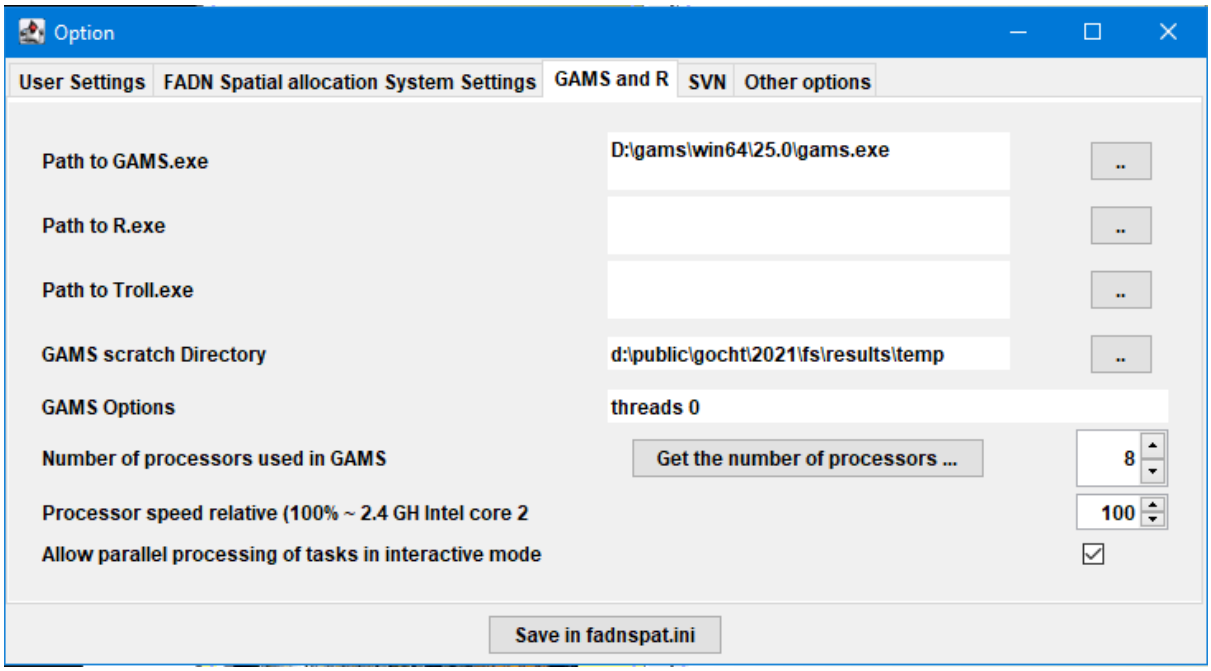


Figure A 6 GUI GAMS and R settings

10.1.1.3. Updating the GUI

To add/change years or countries to the GUI, it can be done in file "Trunk\GUI\FADNSpatial_default.xml".

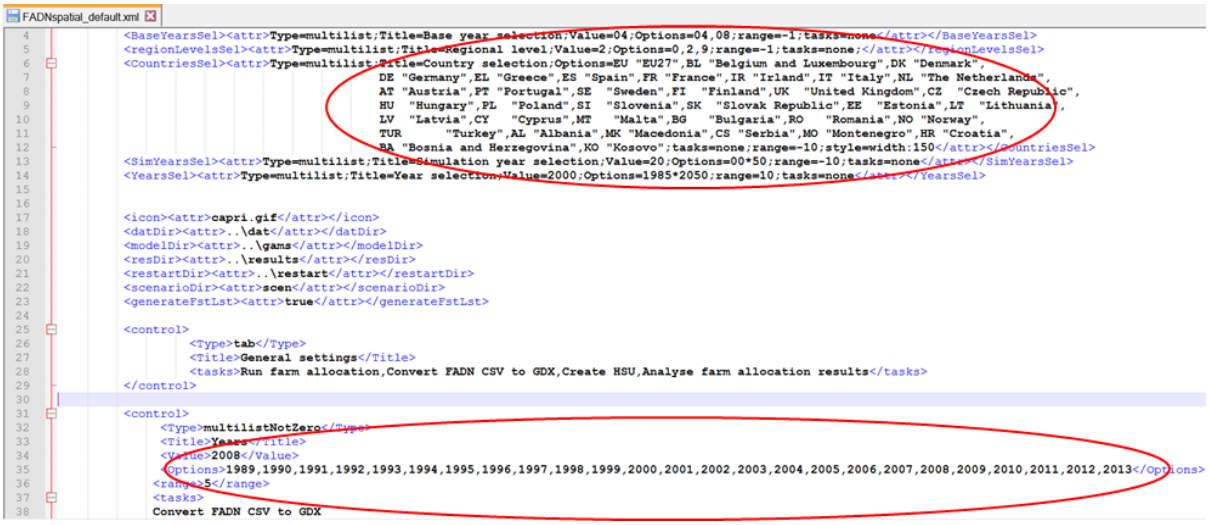


Figure A 7 Adding or removing countries and years to the GUI

The same file is to be used in order to add more or delete certain FADN regions in the GUI. As seen in the screenshot (Figure A 8) the user can add new or delete existing NUTS1 regions in the paragraph "convert FADN CSV to GDX". If the user intends to allocate the FADN regions to the corresponding countries the code has to be modified as shown in Figure 8 in the paragraph "Run farm allocation; Analyse farm allocation results".

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 817566.

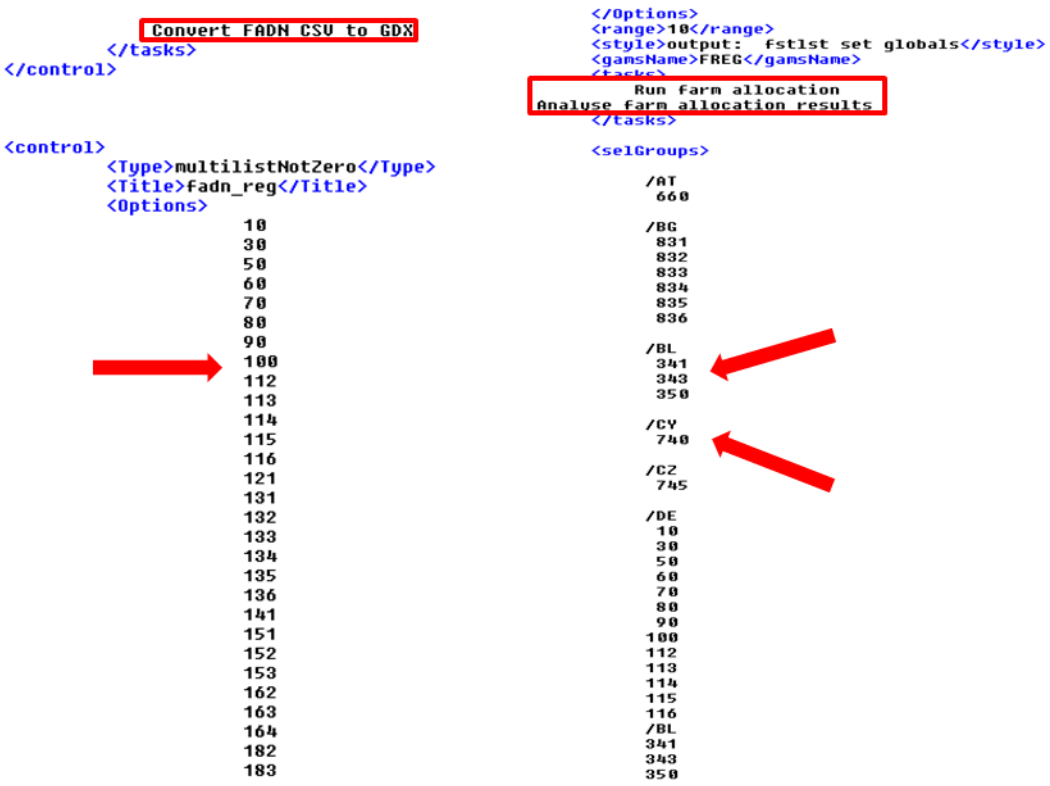


Figure A 8 Modifying the number of FADN regions in the GUI

10.1.1.4. Starting GAMS from GUI

The GUI allows starting the GAMS project directly from the interface, either in compile or run mode. A break request can also be sent to GAMS ("stop GAMS"):

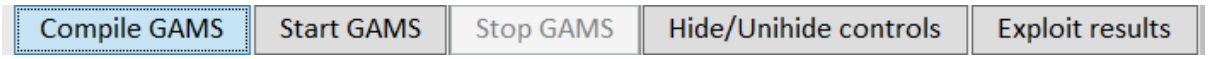


Figure A 9 Different modes to interact with

Once started, the GAMS project routes its output to the console (the DOS prompt) to the lower right pane of the interface, such that the user can follow the progress:

Figure A 10 GUI console output

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 817566.

The pane with the content can be scrolled by a right mouse click in the pane to open a popup menu. If an editor is added under “Edit settings”-“other options”, the GAMS, the listing file and the include file can be opened as well (right clicking with the mouse on the lower right pane):

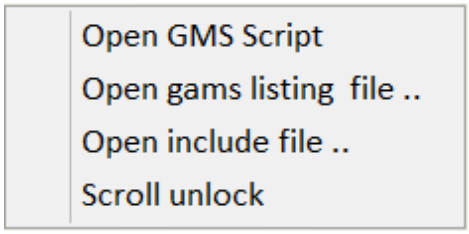


Figure A 11 GUI options

The pane can hence be “frozen” so that e.g. the status of a model solve can be inspected while the project continues to run. In order to successfully start a project, the ini file for GUI must comprise the information where the GAMS executable can be found, but also where the GAMS code of the project to start is stored, see the discussion on settings above.

10.1.1.5. Option: Estimation from the batch utility

The GUI also supports a batch utility to run several tasks (csv to GDX, run data preparation, run final allocation) without interactively using the different work steps and tasks in the GUI and allows inspecting output in a HTML page. In Figure 11 it is shown how to start the batch execution from the GUI.

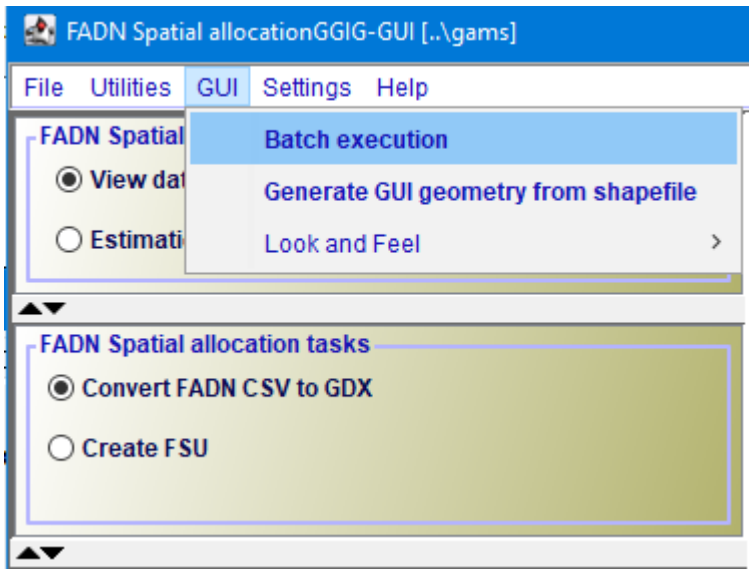


Figure A 12 Starting the batch execution tool

The batch execution tool is depicted in Figure 12. Using Set file the batch file SpatialAllocation.txt can be loaded which is located in “Trunk\GUI\batchfiles”. Activating only compile the GAMS programs starts the GAMS compiler to test if the batch file compiles without errors. Thus deactivating only compile the GAMS programs will execute the GAMS program and run the batch file.

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement N° 817566.

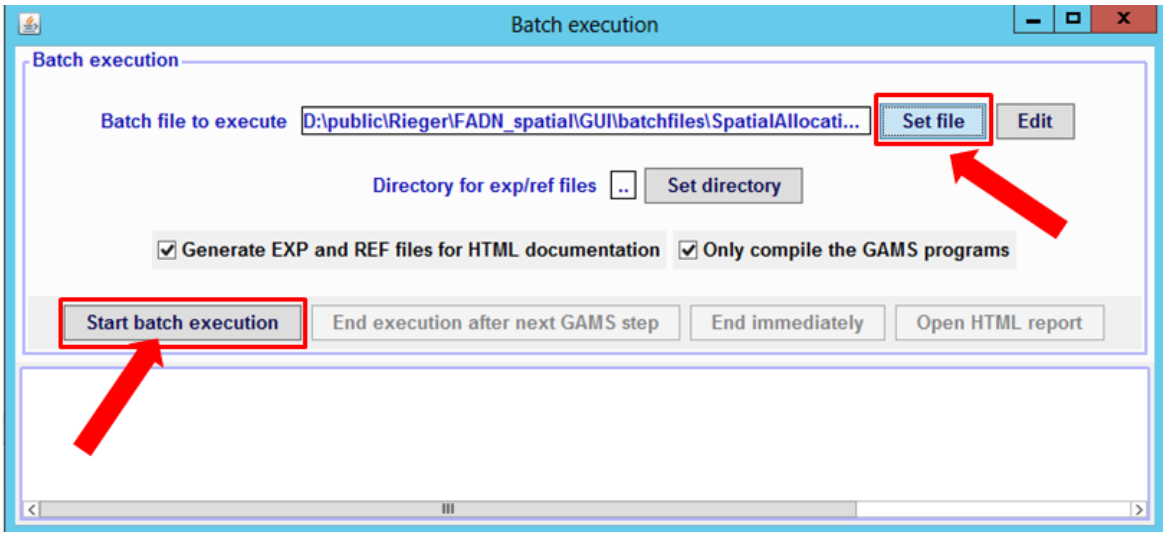


Figure A 13 Batch execution tool

The batch file (SpatialAllocation.txt) does run for all EU countries in 2012 so far but can easily be extended to more countries and years of interest (see Figure 13 for a snapshot of a part of the batch file). In order to add more years MACRO = Years and Years in the task convert FADN CSV to GDX can be modified. For additional FADN regions and countries MACRO = FADNREG and Countries (Countries = DE000000 "Germany") must be changed respectively. In addition, the different calculation steps can be switched on and off. Here only one calculation step is switched on ("Run_data_preparation") and will be executed when starting the batch file (SpatialAllocation.txt).

In order to generate HTML output report pages (see Figure A 14) which include information about the used directories, settings, GAMS options and specific listing files for every work step and task, Generate EXP and REF files for HTML documentation has to be activated (see Figure A 13).

Workstep	*Task*	*User*	*Directories: work* *result* *dat* *restart*	*Settings*
View data	Convert FADN CSV to GDX	Jörg	D:\public\Rieger\FADN_spatial\gams D:\public\Rieger\FADN_spatial\results D:\public\Rieger\FADN_spatial\data D:\public\Rieger\FADN_spatial\restart	**Updated settings**(click to show/hide) **All settings in use**(click to show/hide)

Gams options	*Started* *Ended* *Time used*	*RC*	*Listing* *GdxDiff*
compile --scen=fromGUIforconvertFADN --ggig=on threads 0	Fr., 14 Sep. 18 10:51:52 Fr., 14 Sep. 18 10:51:52 0 min and 0 secs	0	D:\temp\14.09.2018__10.51.52\1.lst
compile --scen=fromjavatofarmalloc_final.gms --ggig=on threads 0	Fr., 14 Sep. 18 10:51:52 Fr., 14 Sep. 18 10:51:56 0 min and 3 secs	0	D:\temp\14.09.2018__10.51.52\4.lst

Figure A 14 HTML output from the batch execution tool

10.1.2. FADN data formatting: "Convert FADN CSV to GDX"

The routine allows an efficient conversion of the CSV formatted FADN data into a GDX. The main program was developed during several projects and since 2013 integrated as part of the GGIG (GAMS Graphical User Interface Generator). The integration so far is performed in form of a graphical user

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 817566.

interface (GUI), which allowed a more intuitively selection (see Figure 2). The CSV data files needs to be located in one folder. For this routine the corresponding csv data files for the EU-countries are located in “Trunk\data\FADN”. As an example, the csv file for Germany for the year 2012 is located in “Trunk\data\FADN\DEU2012.csv”.

10.1.2.1. Files with variables names

The codes behind the variables in the csv files and their meaning (e.g. A1 = Region) can be found in “Trunk\data\FADN\CODE_defs2014.txt” (in Figure A 15). The variable codes that will be used are contained in the file “Trunk\data\FADN\headerrow_xxx.txt” (Figure 18). It consists of one line where all columns, separated by the delimiter of the csv data file, are stored.

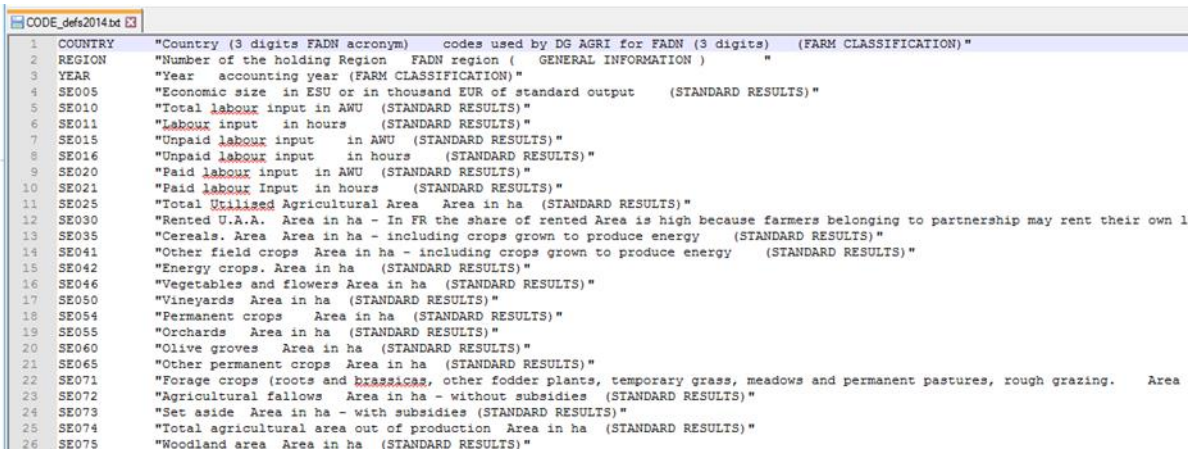


Figure A 15 CODE_defs2014.txt (variables names and their meaning) and original version CODE_dec.txt

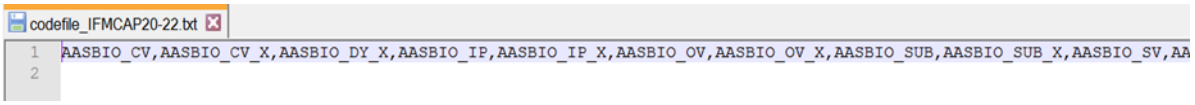


Figure A 16 Headerrow_xxx.txt and original version codefile.txt (variables in csv file)

The codes in the “headerrow_xxx.txt” have to be consistent in terms of labelling and order with the column names and order of the FADN csv data, while “CODE_defs.txt” must have at least the same names/definitions or more and the order does not matter. Thus in case new FADN data will have a different variable structure, “headerrow_xxx.txt” has to be adjusted. A new file can be created so that it can be easy to use one or another depending on the years used (see Table A 1).

Table A 1 Matching between FADN CSV files, headerrow_xxx.txt and CODE_defsxxx.txt

FADN variables years	CODE_defsxxx.txt Variables names and their meaning	FADN Database	headerrow_xxx.txt Variables in the csv file
Until 2016	CODE_dec.txt (3046 variables)	IFM-CAP databases until 2012	codefile_before_2012.txt (3396 variables)
		Farm Spatial Allocation 2012 database	codefile.txt (646 variables)
From 2017	CODE_defs2017.txt	IFMCAP2021 database	codefile_IFMCAP2021.txt
		MINDSTEP database 2012	headerrow_MINDSTEP2012.txt

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement N° 817566.

The txt files to be used need also to be called in the gms file leading the importing and formatting of the data: “Trunk\gams\convertFADN.gms” (see Figure 19). In case of a new column structure of the FADN data, two things have to be adjusted:

1. A new “headerrow_xxx.txt” file has to be generated where all column names are included in the correct order, separated by the delimiter of the FADN csv data file. As depicted in Figure 19, for analysis before 2012 the text file “codefile.txt” would have to be replaced by “codefile_before_2012”.
2. The “keysDim1” code representing the column number of “countryyear” in the csv file has to be adjusted. If the user intends to use the data before 2012 the current column number (“645”) has to be replaced by the old one (“3358”).

10.1.2.2. Launching the conversion

To start the conversion procedure in the GUI the commands view data in the left check box work steps as well as Convert FADN CSV to GDX in the checkbox tasks have to be selected (see Figure A 17). In the right checkbox General settings the respective year and country can be selected which automatically uses the respective FADN CSV file for the corresponding country and year. Also several countries and years at once can be chosen. The generated GDX file will be located in “Trunk\results\fadn” e.g. for Germany for 2012 “Trunk\results\fadn\FADN2012DE”.

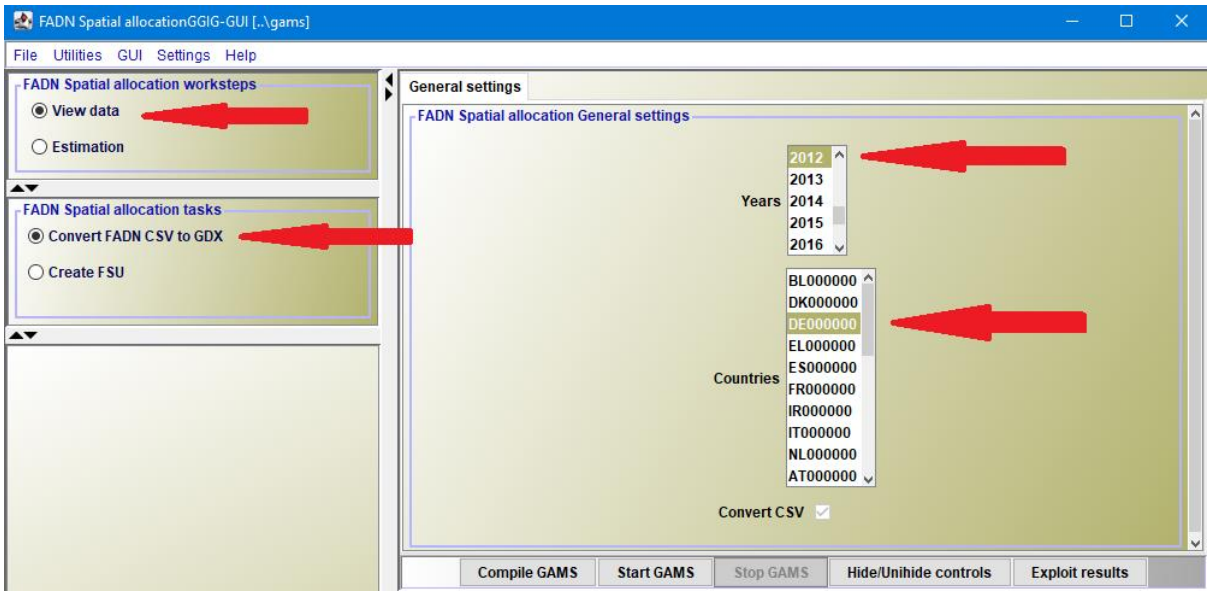


Figure A 17 Conversion of csv to GDX in the GUI

The following figures (Figure 22-Figure 25) show the structure of the GDX output. The first column “Entry” shows the item number:

- Entry 1: DATA is the data parameter of three dimensions.
- Entries 2 and 3: KEY is also saved with the name (to be compatible with old GAMS versions FD), refers to the farm identifier.
- Entries 4-23: are the cross sets or mappings between farms FD and different NUTS regions. The cross sets with the CSV flag consists of the original region name given in the CSV file (e.g. DE24), whereas without the flag the CAPRI conversion is applied (e.g. DE240000).
- Entry 24: The set COLUMN is the definition of the second dimension (FADN variables) from the code description file (codefile.txt) including their description.
- Entry 27: The META set describes input and output file and the date of the GDX file conversion.

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement N° 817566.

10.1.3. Creation of FMU: “Create FSU”

The creation of FMU can be started in the GUI as seen in Figure A 18.

- Important: now a year needs to be selected, which is not the year of the CAPRI disaggregation data but the year of the FADN data to be used.
- Parallel running often gives errors because some files are tried to be used at the same time for different countries.
- Estimated running time: In parallel: 6-8 hours?. Not in parallel:

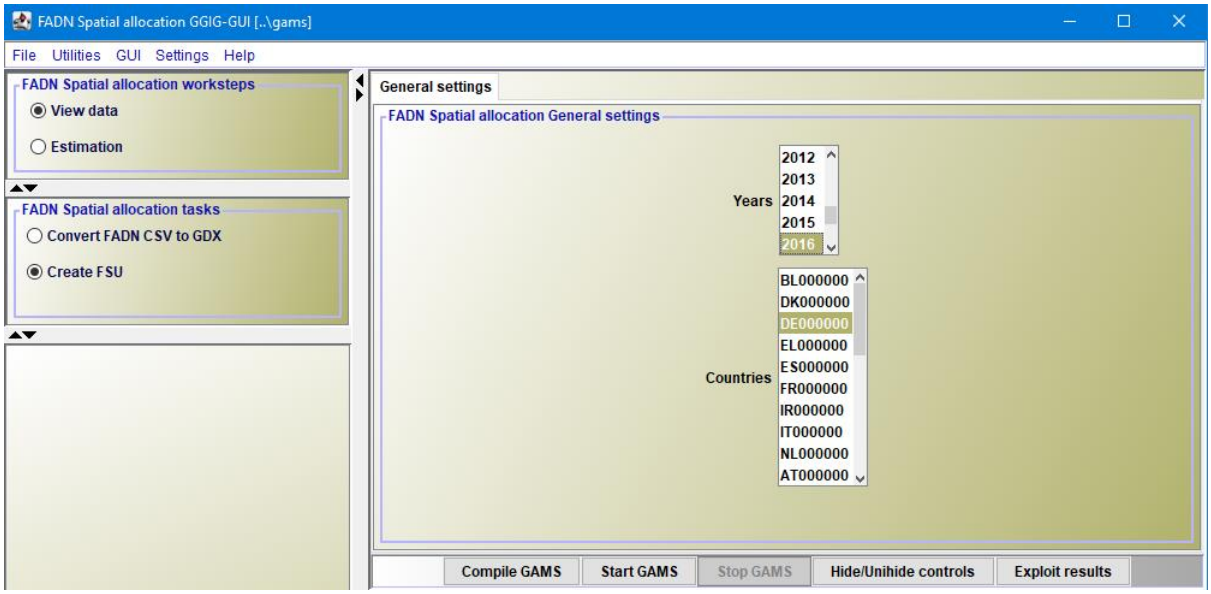
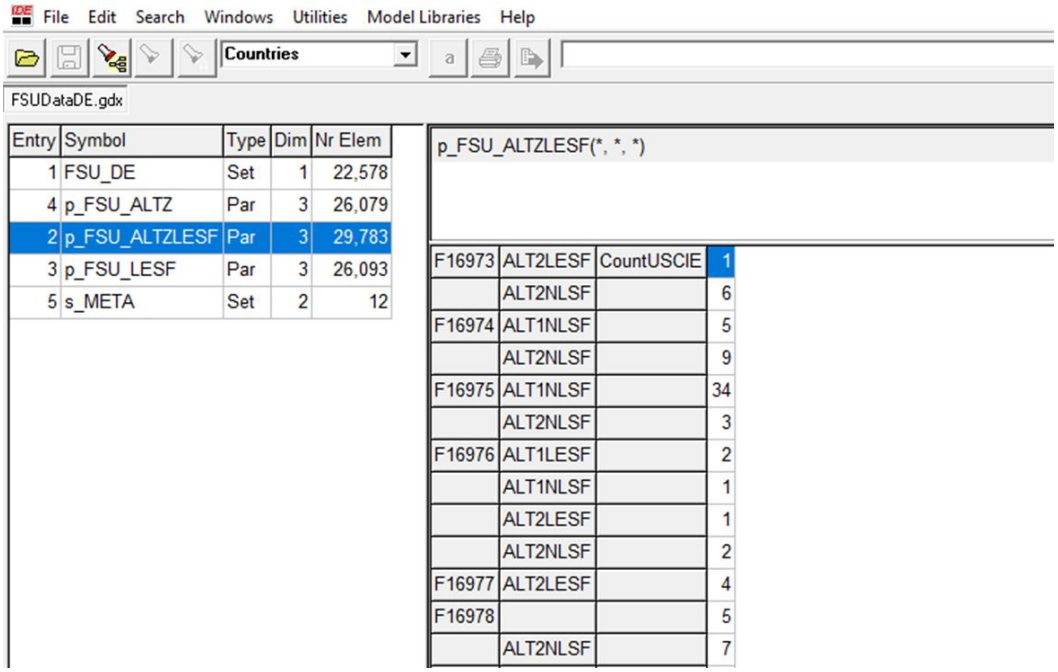


Figure A 18 Creation of FSU in the GUI

In this section, it is calculated how many 1 km^2 USCIE grids belong to the FSU (total) and how many belong to a specific altitude zone and less favoured area combination (set ALTZLESF). Results are stored under “Trunk\results\fsu\FSUDataCN.gdx” (CN=abbreviation of country name). During the “Run data preparation” step of the estimation, the area and yields of these FMU will be calculated.

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement N° 817566.



Entry	Symbol	Type	Dim	Nr Elem
1	FSU_DE	Set	1	22,578
4	p_FSU_ALTZ	Par	3	26,079
2	p_FSU_ALTZLESF	Par	3	29,783
3	p_FSU_LESF	Par	3	26,093
5	s_META	Set	2	12

F16973	ALT2LESF	CountUSCIE	1
	ALT2NLSF		6
F16974	ALT1NLSF		5
	ALT2NLSF		9
F16975	ALT1NLSF		34
	ALT2NLSF		3
F16976	ALT1LESF		2
	ALT1NLSF		1
	ALT2LESF		1
	ALT2NLSF		2
F16977	ALT2LESF		4
F16978			5
	ALT2NLSF		7

Figure A 19 Results of task "Create FSU"(FMU): results\fsu\FSUDataCN.gdx

10.1.4. Estimation: run farm allocation

As already mentioned in the introduction the CO method estimates the probability in which FMU each FADN farm is allocated.

The necessary steps to start the allocation procedure are shown in Figure A 20 below. In the left toolbox work steps the command estimation and in the lower toolbox tasks the "Run farm allocation …" procedure has to be selected. In the right toolbox General settings the year and the respective FADN region (fadn_reg) can be chosen. By right-clicking inside the fadn_reg pane country specific NUTS1 regions can be selected. By using the command add to selection it is possible to analyse different country-specific NUTS1 region at once.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 817566.

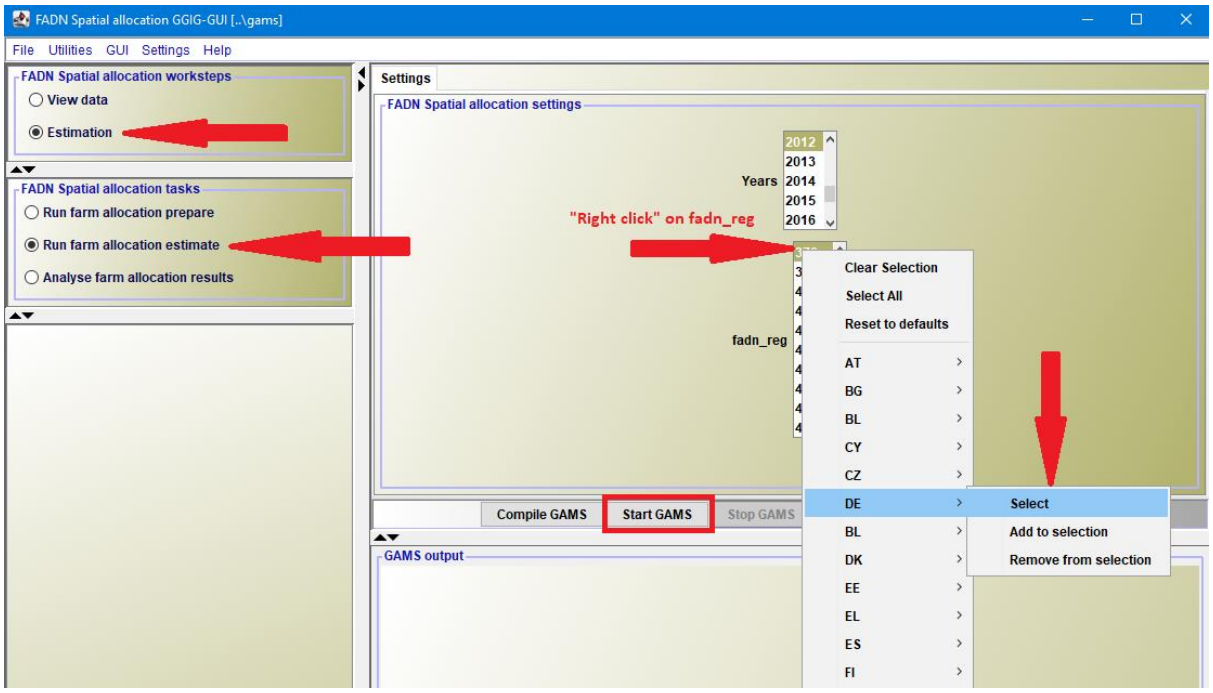


Figure A 20 Estimation of farm allocation in the GUI

The estimation main file is: "Trunk\gams\farm_alloc_final.gms". The allocation procedure uses two differentiated steps:

- Step 1- Data preparation ("Trunk\gams\runDataPreparation.gms): The data preparation step reads all relevant input data from the database and calculates the prior probabilities of FADN farms to be in specific FSUs. This preprocessed data is stored in "Trunk\results\data_region\dataregion_%FREG%_%YEAR%.gdx" (e.g. DataRegion_370_2012.gdx).
- Step 2- Final allocation ("Trunk\gams\runfinalAllocation.gms): Once the above file is existing, the final allocation, i.e. the complex optimisation problem, can be started independent from the data preparation. The final allocation results are then stored in "Trunk\results\allocation_FSU\probFarmInFSU_%FREG%_%YEAR%.gdx".

Allocating n farms to m spatial units gives n*m potential combinations. Since in many FADN regions the number of farms and FSU is more than 1000, far over a million combinations have to be evaluated. This is (1) time consuming in the data preparation when calculating the prior probability of a farm to be in a FSU and (2) creates a complex optimization problem in the final allocation. In order to avoid that a problem in step (2) requires every time to rerun step (1), both steps can be started separately by selecting the right tasks in the GUI.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 817566.

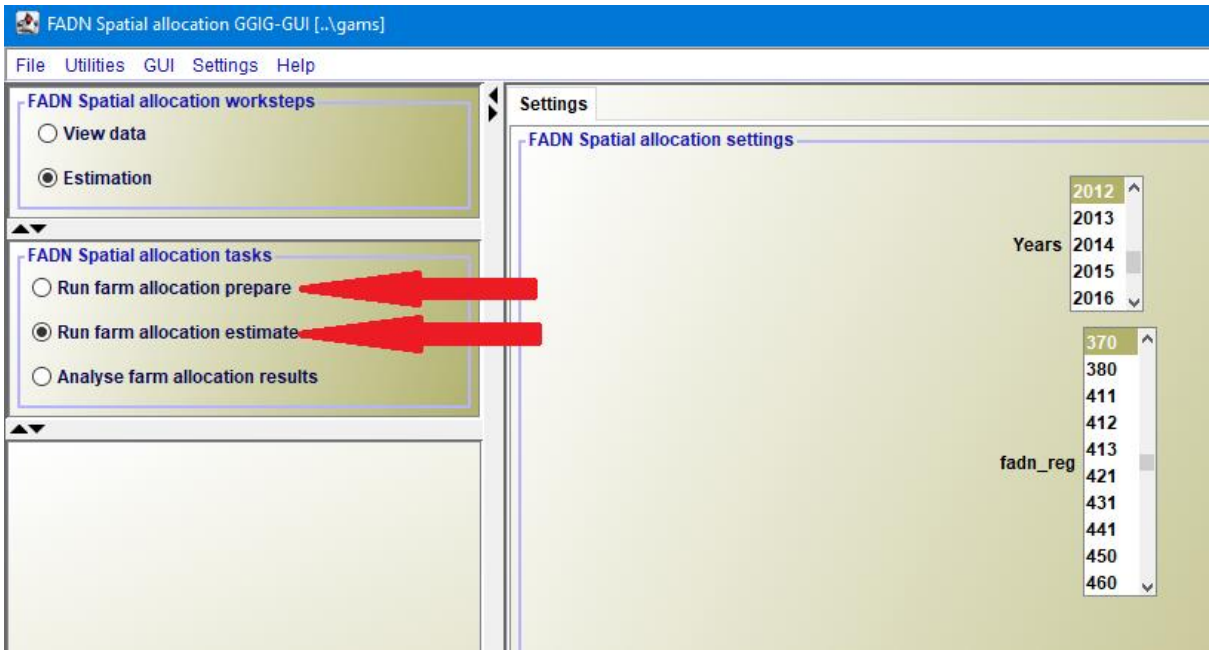


Figure A 21 Selecting work steps data preparation, estimation and/or analyse final allocation

As the model run takes long we can run the FADN regions in parallel using the setting allow parallel processing in task in interactive mode. As the results are anyway stored by region and years, no additional data merging is required in parallel mode. The mode can be done for both check box options as described above.

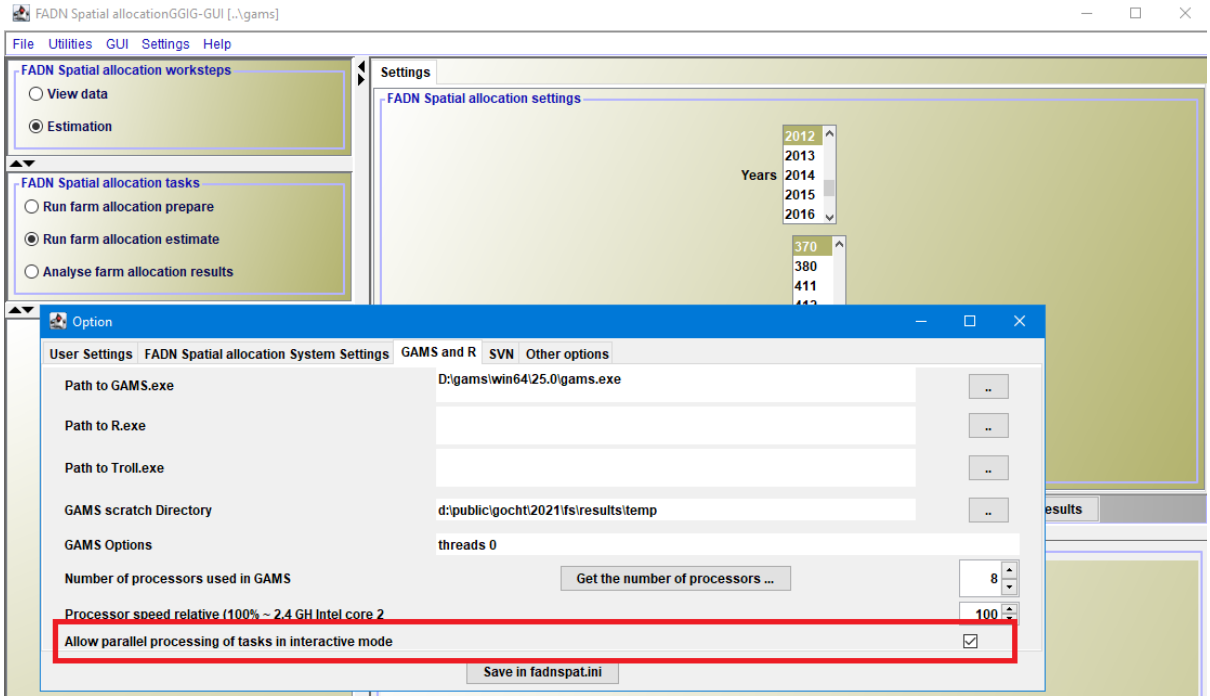


Figure A 22 Parallel processing in interactive mode

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 817566.

10.1.4.1. Data preparation - Calculation of FMU area and yield

The FMUs as a combination of FSU with ALTZLESF zones[10] were prepared in section 6 Creation of FMU[11]. During the data preparation, the share of each ALTZLESF class is calculated. Following it is assumed that the UAA assigned to the FSU in CAPDIS is evenly distributed over the ALTZLESF classes. The yield is assumed to be identical in all ALTZLESF classes.

Table A 2 Calculation of UAAR in FMU

FMU or FSU	km²	% of total	UAA (1000 ha)	Yield (kg)
U100.ALT2LESF	33	35%	133,4	6750
U100.ALT1LESF	40	43%	161,7	6750
U100.ALT1NLSF	21	22%	84,9	6750
U100 (Total)	94			
U100 (from capdis)			380	6750

These calculations are done in the gams files:

- "Trunk\gams\PREP_FSU_DATA.gms. Loads FSU areas and yields
- "Trunk\gams\PREP_FMUfromFSU.gms. Calculates FMU areas and yields.

10.1.4.2. Data preparation - Harmonize altitude zones and less favoured areas

The availabe data allows to assign FMU (FSU shares) and FADN farms to 6 ALTZLESF zones. In most FADN regions not all 6 zones occur and quite often less frequent zones in a region are not occuring in both - FADN and FMU data. This potential inconsistency has to be overcome before the independent allocation models are started since the allocation expects for each FADN farm suitable FMU are availabe and vice versa. This harmonisation of data is done in gams file "Trunk\gams\adjust_ALTZLESF.gms".

The code identifies ALTZLESF zones with inconsistent data among FADN and FMU:

1. ALTZLESF zones missing in FADN data but found in FMU data (set ALTZLESF_missFADN_foundFMU)
2. ALTZLESF zones found in FADN data but missing in FMU data (set ALTZLESF_foundFADN_MissFMU)

If an inconsistency is detected in a FADN region, a remapping procedure is started, that changes the attributes in FADN and/or FMU data until consistency is achieved. The basic idea is to remap FADN farms and/or FMU areas to a zone that is "as similar as possible" to the original zone.

The process starts with the FADN farms (see also). If a farm's zone does not match any available FMU zone, then it is tested, if it is sufficient to change the LFA specification in order to find an existing FMU zone. If this is not successful, it is tested to move the specification to a higher altitude zone. If this does not help, it is tested to remap the farm to a lower altitude zone. If after this no existing FMU is found,

[10] It is calculated how many 1 km² USCIE grids belong to the FSU (total) and how many belong to a specific altitude zone and less favoured area combination (set ALTZLESF).

[11] The results (in parameter p_FSU_ALTZLESF) are loaded by the file "Trunk\data\FMU\sets_FMUfromFSU.gms", which is called directly by the parent file "Trunk\gams\farm_alloc_final.gms" before run_data_preparation.gms-

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 817566.

the last remaining altitude zone is testet. Finally, as ultimatio ratio, the most frequent zone occuring in the FMU would be selected.

Table A 3 Scheme of remapping attributes of FADN farms

Original FADN attribute	Remap until existing FADN attribute is found				
	remap FADN1 remap to other LFA specification	remap FADN2 remap to next higher altitude zone	remap FADN3 remap to next lower altitude zone	remap FADN4 remap to other remaining altitude zone	assign most frequent FMU attribute
ALT1LESF	ALT1NLSF	ALT2LESF	n.a.	ALT3LESF	specific for FADN region
ALT1NLSF	ALT1LESF	ALT2NLSF	n.a.	ALT3NLSF	
ALT2LESF	ALT2NLSF	ALT3LESF	ALT1LESF	n.a.	
ALT2NLSF	ALT2LESF	ALT3NLSF	ALT1NLSF	n.a.	
ALT3LESF	ALT3NLSF	n.a.	ALT2LESF	ALT1LESF	
ALT3NLSF	ALT3LESF	n.a.	ALT2NLSF	ALT1NLSF	

After all FADN farms are assigend to existing zones in the FMU data, there might still be zones occuring in FMU, but without corresponding FADN farms.

Table A 4 Scheme of remapping attributes of FMU

Original FMU attribute	Remap until existing FADN attribute is found					
	"delete" if other ALTZLESF is observed in FSU	remap FMU1 remap to other LFA specification	remap FMU2 remap to next lower altitude zone	remap FMU3 remap to next higher altitude zone	remap FMU3 remap to other remaining altitude zone	assign most frequent FMU attribute
ALT1LESF	distribute UAA proportianal to remaining ALTZLESF	ALT1NLSF		ALT2LESF	ALT3LESF	specific for FADN region
ALT1NLSF		ALT1LESF		ALT2NLSF	ALT3NLSF	
ALT2LESF		ALT2NLSF	ALT1LESF	ALT3LESF		
ALT2NLSF		ALT2LESF	ALT1NLSF	ALT3NLSF		
ALT3LESF		ALT3NLSF	ALT2LESF		ALT1LESF	
ALT3NLSF		ALT3LESF	ALT2NLSF		ALT1NLSF	

Quite often a FSU is split in 2 or 3 FMU because of different ALTZLESF zones. If there are ALTZLESF zones observed in the FSU that are consistent with FADN farms, the non-consistent ALTZLESF zones are "deleted", i.e. the FMU are reconstructed (as described in section 2.3) and the non-existing ALTZLESF are not considered when calculating the UAA shares. Table 5 shows the result of UAA calculation of the example given in Table 2, assuming that no farm in the FADN sample has the attribute ALT2LESF.

Table A 5 Calculation of UAA when ALT2LESF attribute is not in FADN sample

FMU or FSU	USCII Data	Calculation when no FADN farm in ALT2LESF			
	km^2	km^2	% of total	UAA (1000 ha)	Yield (kg)

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 817566.

U100.ALT2LESF	33				
U100.ALT1LESF	40	40	66%	249	6750
U100.ALT1NLSF	21	21	34%	131	6750
U100 (Total)	94	61			
U100 (from capdis)				380	6750

If the UAA of the inconsistent ALTZLESF cannot be distributed proportionally to other zones, the procedure tries step by step to allocate the full UAA to specific zones as indicated in the columns to the right. The following steps of the remapping procedure are mainly analogous to the remapping of FADN farms. Since it was first tried to move FADN farms upwards in altitude, it is done in the opposite direction here.

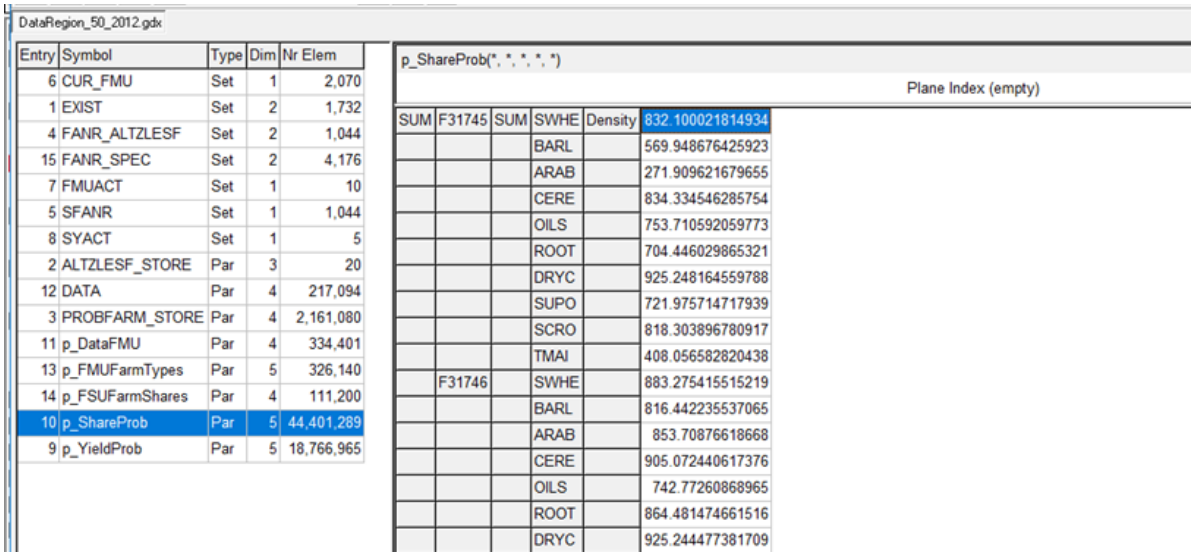
This sequence is of course to some extent arbitrary. However, when analysing the inconsistensies in various FADN regions, it was found that this procedure comes up with reasonable adjustments. The remapping is usually only necessary for zones that are not frequently occuring. For example in many FADN regions the altitude zone 3 (ALT3) is very unusual. But often a very small area is above 600m and following a FMU is assigend the ALT3 attribute. The chance that a FADN farm in the sample "happens" to be in this small ALT3 zone is very low. In this case the procedure would assign the FMU in ALT3 most probably to ALT2.

10.1.4.3. Data preparation - Prior information on type of farming

This has been shown in chapter 6.2.

10.1.4.4. Data preparation - Output

The data preparation step calculates the prior probabilities of FADN farms to be in specific FSUs. This preprocessed data is stored in "Trunk\results\data_region\dataregion_%FREG%_%YEAR%.gdx" (e.g. DataRegion_370_2012.gdx). The format of this file can be seen in the following figure.


DataRegion_50_2012.gdx

Entry	Symbol	Type	Dim	Nr Elem
6	CUR_FMU	Set	1	2,070
1	EXIST	Set	2	1,732
4	FANR_ALTZLESF	Set	2	1,044
15	FANR_SPEC	Set	2	4,176
7	FMUACT	Set	1	10
5	SFANR	Set	1	1,044
8	SYACT	Set	1	5
2	ALTZLESF_STORE	Par	3	20
12	DATA	Par	4	217,094
3	PROBFARM_STORE	Par	4	2,161,080
11	p_DataFMU	Par	4	334,401
13	p_FMUFarmTypes	Par	5	326,140
14	p_FSUFarmShares	Par	4	111,200
10	p_ShareProb	Par	5	44,401,289
9	p_YieldProb	Par	5	18,766,965

p_ShareProb(*, *, *, *, *)

Plane Index (empty)

SUM	F31745	SUM	SWHE	Density	832.100021814934
			BARL		569.948676425923
			ARAB		271.909621679655
			CERE		834.334546285754
			OILS		753.710592059773
			ROOT		704.446029865321
			DRYC		925.248164559788
			SUPO		721.975714717939
			SCRO		818.303896780917
			TMAI		408.056582820438
	F31746		SWHE		883.275415515219
			BARL		816.442235537065
			ARAB		853.70876618668
			CERE		905.072440617376
			OILS		742.77260868965
			ROOT		864.481474661516
			DRYC		925.244477381709

Figure A 23 Results of task "Run data preparation": results\data_region\dataregion_%FREG%_%YEAR%.gdx

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 817566.

10.1.4.5. Final allocation - Split optimisation by type of farming and altitude/less favoured area classification

The number of FSUs is significantly higher than the number of old FMUs used based on the old units HSMU. Since the probability of every potential combination of FADN farm and FSU has to be evaluated in the optimisation, a high number of FSUs in a (big) FADN region can lead to computational problems (solution time, no optimal solution). Solving independently for 8 farm types * 6 ALTZLESF classes makes the optimisation problem manageable in all big FADN regions tested so far. This requires that FADN farms and FMUs are assigned to a unique ALTZLESF class in a consistent way.

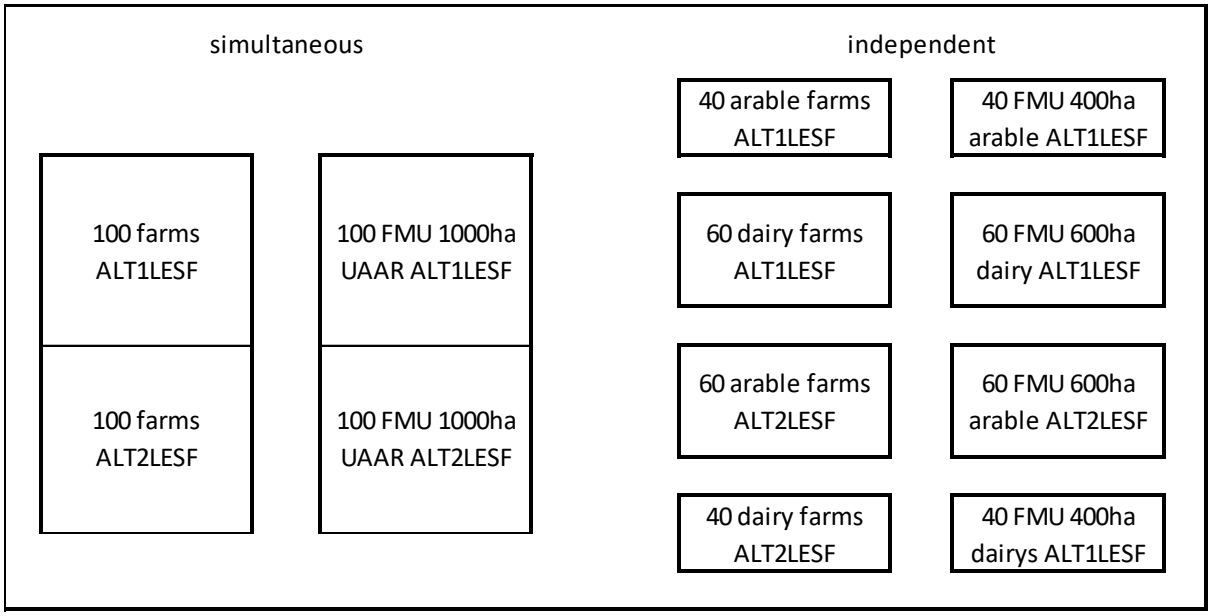


Figure A 24 Scheme of independent optimisation problems

The equations and models[12] used are defined under "Trunk\gams\farm_equ_grid.gms"[13] and solved under "Trunk\gams\runfinalallocation.gms".

The prior information of farm type area is brought into the model as a constraint:

```
e_consArea_(CUR_FMU,Cur_SPECD1,Cur_ALTZLESF) $ EXIST(CUR_FMU,Cur_ALTZLESF) ..

      p_DataFMU(Cur_FMU,Cur_SPECD1,"LEVL" ,Cur_ALTZLESF)  * v_adjustSpecD1Grid(Cur_FMU,Cur_SPECD1,Cur_ALTZLESF) =E=
           SUM(FANR_ALTZLESF(CUR_FANR,Cur_ALTZLESF), v_probFarm(CUR_FANR,CUR_FMU,Cur_ALTZLESF) * DATA(CUR_FANR,"UAAR","LEVL","REPR"));
```

Figure A 25 Code from "Trunk\gams\farm_equ_grid.gms" (l. 73ff rev. 96)

10.1.4.6. Final allocation - Solve without prior information

This option allows to run the allocation without the prior information on farm types area from FSS 10 km^2 grid. It cannot be selected in the GUI but in the code (see below).

This option is not run by default as the number of FMUs is significantly higher than the number of the old FMUs (from HSMU) used before. Since the probability of every potential combination of FADN farm and FSU has to be evaluated in the optimisation, a high number of FMUs in a (big) FADN region can lead to computational problems. When no farm type area from 10 km^2 grid is used, it is not

[12] The models solved are: m_yield, m_share and m_yieldshare.

[13] There is also a file "Trunk\gams\farm_equ.gms"from the old project with the models m_yield_m_share, m_yieldshare and m_yieldsharetypo.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 817566.

possible to solve the farm types independently. Hence it is not reasonable to try this model specification in FADN regions with more than 600 FADN farms. The validation of results showed that using 10 km^2 grid data is highly recommendable, it seems to be acceptable that the model design ignoring spatial farm type information (referred to as "NoGridInfoUsed") is not applicable in all FADN regions.

The regions where the model design "NoGridInfoUsed" is calculated can be "selected" in the code, by changing the maximal number of farms that trigger the execution of the loop:

```
*
*  For comparison: use no information from grids on farm types
*  Note: this is only reasonalbe when there are few farms (e.g. lt 200) in the region
*        otherwise this test is too time consuming or does not reach optimal solution
*
*-------------------------------------------------------------------------------------

Loop(ALTZLESF $ (CARD(SFANR) lt 600),
```

Figure A 26 Code taken from "Trunk\gams\runFinalAllocation.gms" (l. 350ff rev 96)

A maximal number of 600 farms was tested and seems to be feasible. However, 600 farms are already very time consuming. If the results of NoGridInfoUsed are not of interest for validation purposes, it is recommended to choose a much smaller number (e.g. 200 or even 0).

10.1.4.7. Final allocation - Output

The final allocation results are stored in "Trunk\results\allocation_FSU\probFarmInFSU_%FREG%_%YEAR%.gdx". The format of this file can be seen in the following figure.

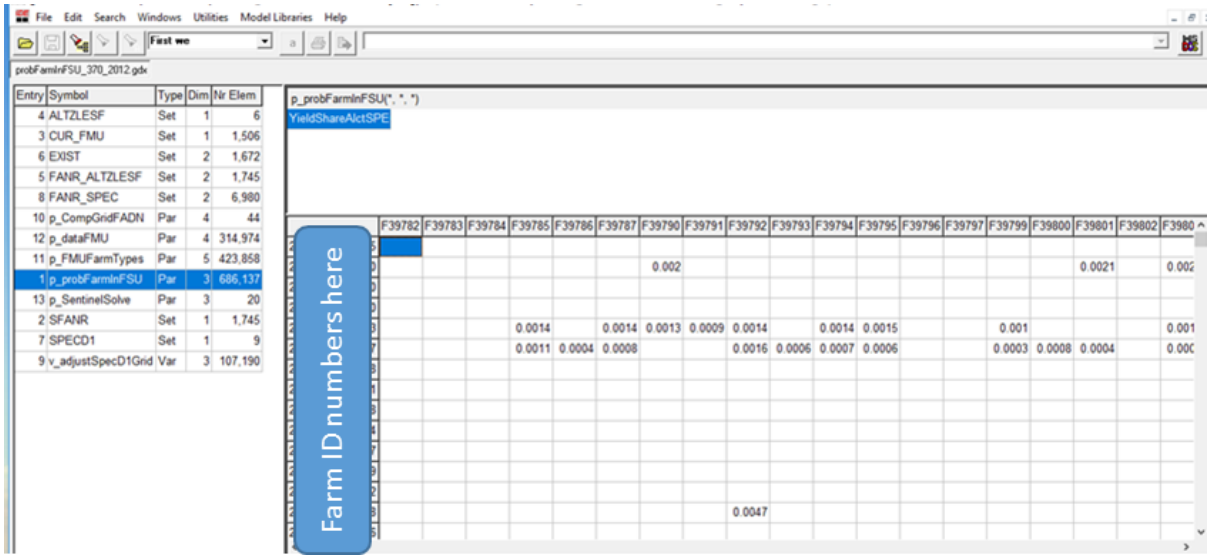


Figure A 27 Results of task "Run farm allocation": results\allocation_FSU\probFarmInFSU_%FREG%_%YEAR%.gdx

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 817566.

10.1.5. Analyse farm allocation results

After the farm allocation is ready, the GUI foresees the possibility to produce maps with the area share for each farm type and the probability share of each farm type. For this, there is a first step for calculating the area and probability share under each farm type: “Analyse farm calculation results”.

10.1.5.1. Analyse farm allocation results

The task “Analyse farm allocation results” calculates for each FSU the area share of each farm type (see figure above). It is done by the file “Trunk\gams\farm_alloc_results.gms”. Results are produced to the file “Trunk\results\Mainresults2012.gdx” (see **Error! Reference source not found.**). It contains two parameters:

- p_res(farm type, FSU, model, year) the probability of farm type located in certain FSU
- p_res1(farm type, FSU, model, year) percentage of the FSU utilisable agricultural area covered by each of the farm types. It is calculated in the same way as p_res but multiplied by the agricultural area (UAAR-LEVL) and the farm weight.

It is not possible to identify if all FADN regions are included. They need to be run all in one go as they do not accumulate.

In order to avoid cancelling old results, it is advised to add a suffix manually to the name of the output file, to identify the run or version: e.g. Mainresults2012_DE04Nov21.gdx (for results only for Germany run on date 04Nov2021); Mainresults2012_EUall20Nov21.gdx; etc.

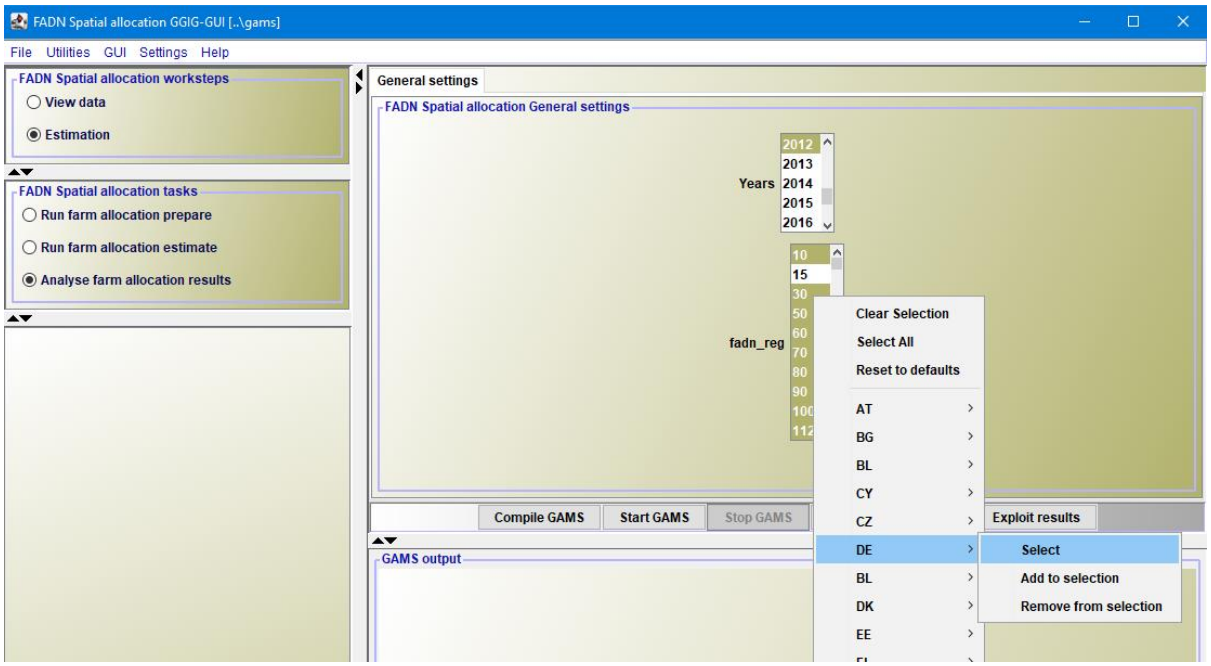


Figure A 28 “Analyse farm allocation results” in the GUI

10.1.5.2. Exploit results

In order to view the results of the analysis “Exploit results” has to be clicked (see Figure A 29). This opens the result exploiter which is shown in Figure A 30. First the file “Mainresults2012.gdx” has to be selected, then “show results”. It takes p_res (probability of farm type located in certain FSU) by default. For the maps, it is important that the FSU are rows and the farm types columns (use the pivot button on the top right corner the file to shift rows and columns). Then select View – Map and the “trunk\GUIcoo\FSU.zip” should be selected.

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement N° 817566.

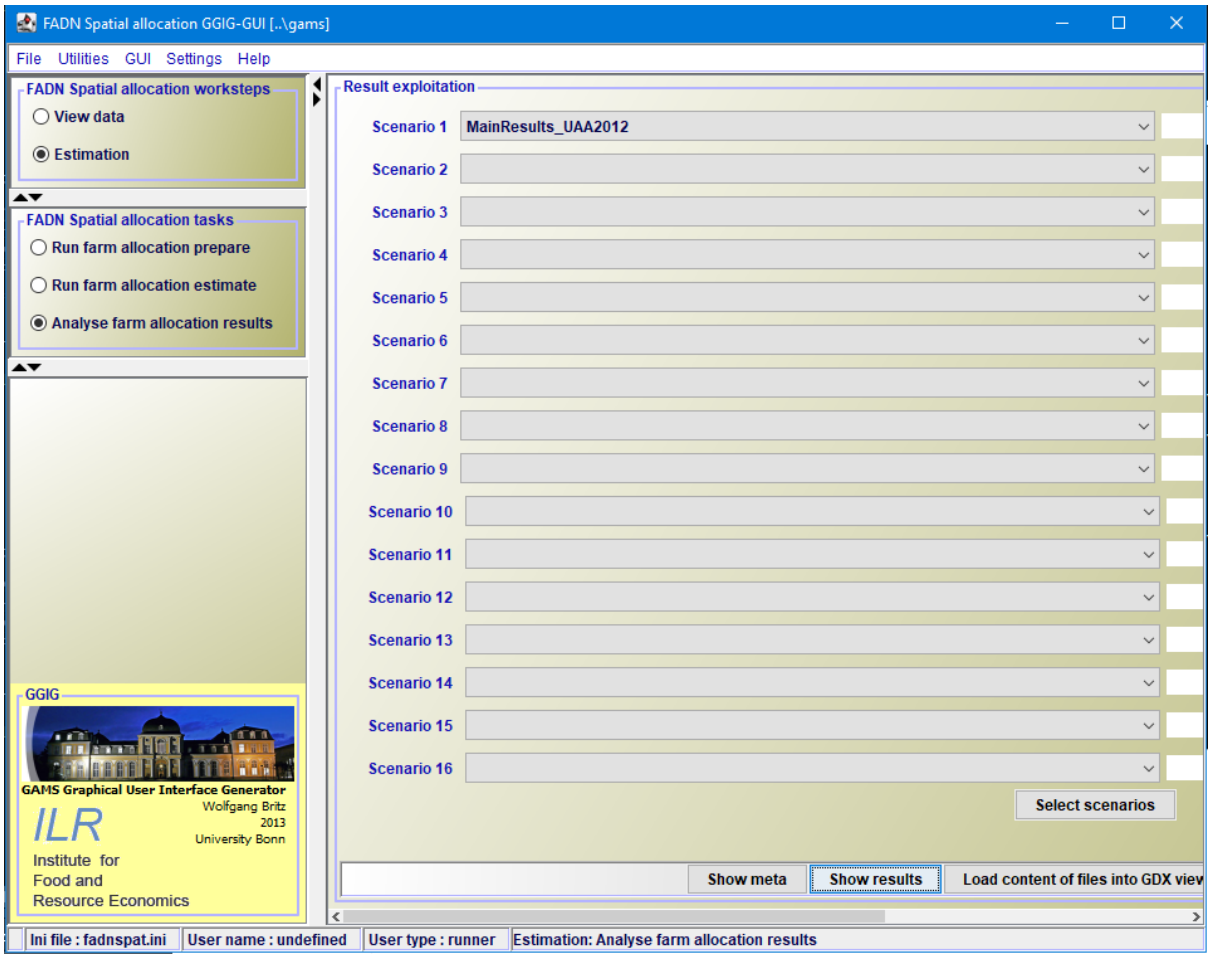


Figure A 29 Result exploiter in the GUI

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 817566.

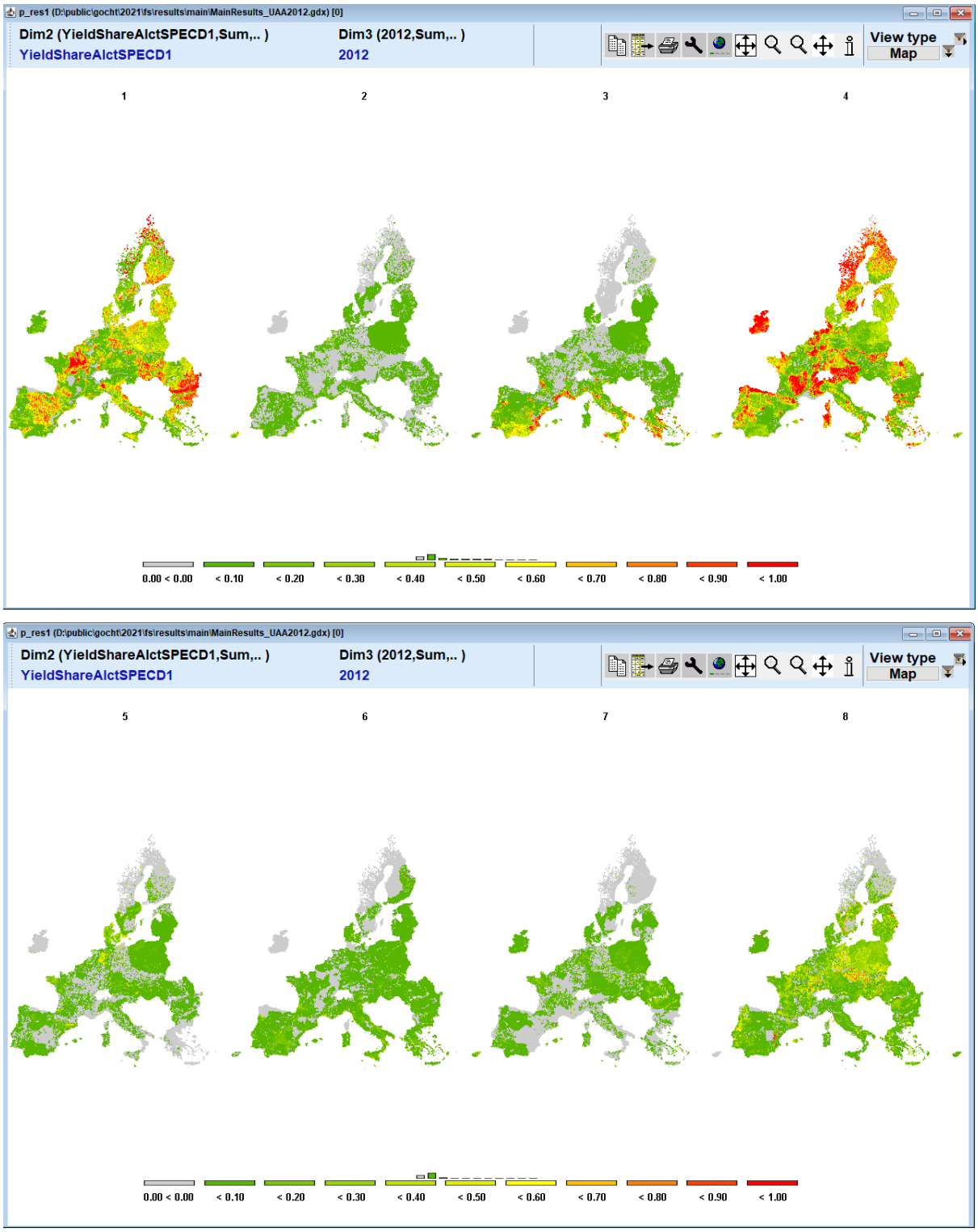


Figure A 30 Results maps: p_res or probability of farm type located in certain FSU

In order to use p_res1 (percentage of a utilisable agricultural area covered by each of the farm types) the GDX viewer should be used (see Figure A 31 and **Error! Reference source not found.**).

Figure A 31 Open GDX Viewer in the GUI

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 817566.

10.1.5.3. Convert GRID to FSU data

For comparison of the farm spatial allocation results with the FSS-grid data, FSS grid data need to be remapped to FSU. This is done embedded in the software but results are not kept. For this reason, ex-post, it can be done with the file ..\ConvertGridtoFSUdata\1-ConvertGRIDtoFSUdata.gms. The output is in FSS2010data_byfsu.gdx.

Then the file ..\ConvertGridtoFSUdata\2-ComparewithFSSFSUdata.gms allows to calculate a gdx file with UAA shares, with the direct comparison between FSS and Spatially allocated UAA shares by farm type: Comparison_FarmAlloc_FSS_%ver%&noAreaCons.gdx.

Modify $setglobal ver with the version identifier of the run, which is the suffix used for the main output file Mainresults%year%.gdx (e.g. $setglobal ver EUall_30Nov2019).

MIND STEP WP2 TEAM

Mr Alexander Gocht (THÜNEN)

Mr Sebastian Neuenfeldt (THÜNEN)

Ms Xinxin Yang (THÜNEN)

Mr Hugo Storm (UBO)

Mr John Helming (WR)

Mr Marc Müller (WR)

Mr Sander Jannsen (WR)

Mr Gerbert Roerink (WR)

Mr Albert Brouwer (IIASA)

Dimitrios Kremmydas (JRC)

Consortium description

The consortium of MIND STEP consists of 11 partners from 7 countries in Europe (the Netherlands, Germany, Austria (IIASA), Italy, France, Spain (JRC-Seville), Norway and Hungary). It includes partners from the private and public sector representing:

- Academia and higher education (UBO, UCSC, WU).
- SME dealing with research consultancy, data collection, strategic advice, normalization and policy in the field of energy, environment and sustainable development. This SME has also a strong track record in the field of communication, stakeholder engagement and exploitation (GEO)
- Public government bodies dealing with agricultural and environmental research and data collection and building agricultural models at different scales (WR, IIASA, IAMO, THÜNEN, INRA, NIBIO, JRC)

The consortium has been carefully constructed in such a way that it is capable of jointly managing all activities and risks involved in all project stages. Each partner contributes its own particular skills, (inter) nationally wide network and expertise, and has a critical role in MIND STEP. Partner expertise smoothly complements each other and all together form the full set of capabilities necessary to lead MIND STEP to a success. Achieving the overall objective is determined by all partners in the consortium as well as their ability to involve other interested stakeholders in the process of developing, validating

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 817566.

and disseminating the IDM models, indicators and methodologies (WR, UBO, IAMO, UCSC, WU, THÜNEN and INRA) and linking IDM models to current agricultural policy models (WR, IIASA, UBO) included in the MIND STEP model toolbox. Dissemination and communication activities are steered by partner GEO who has graphic design, IT and marketing communication teams to deliver out-of-the-box and novel solutions for dissemination and communication and JRC who has a large network with policy makers. GEO has experience in leading comparable activities in H2020 projects as UNISECO and COASTAL. The coordinator WR is part of Stichting Wageningen Research (Wageningen Research Foundation, WR). WR consists of a number specialised institutes for applied research in the domain of healthy food and living environment. WR collaborates with Wageningen University (WU) under the external brand name Wageningen University & Research. One of the strengths of Wageningen University & Research (including WR) is that its structure facilitates and encourages close cooperation between different disciplines. The institutes Wageningen Economic Research (proposed coordinator of MIND STEP, WR) and Wageningen Environmental Research (WR) are involved in this proposal. The One-Wageningen approach will also be applied to MIND STEP. WR has a long-standing reputation of leading large scale EU projects, such as SUPREMA, Foodsecure, SUSFANS, FLINT, SAT-BBE, and SIM4NEXUS.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 817566.