salsa
small farms
small food businesses and
sustainable food security

29.09.2017
Deliverable 2.2

Map of the estimated distribution of small farms in each reference region

WP2 Estimation of the distribution and production capacity of small farms

Prepared under contract from the European Commission
Project number: 677363
Collaborative project
Horizon2020

Project acronym	**SALSA**
Project full title	Small Farms, Small Food Businesses and Sustainable Food Security
Duration	April 2016 – March 2020
Coordinating organisation	Universidade de Évora (UEvora)
Project coordinator	Teresa Pinto-Correia
Project website:	*www.salsa.uevora.pt*
Deliverable title	Map of the estimated distribution of small farms in each reference region
Deliverable number	2.2
Work package	WP2
Authors	Sérgio Godinho; Nuno Guiomar; Rui Machado; Teresa Pinto Correia; Theodore Tsiligiridis; Katerina Ainali

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 677363

Contents

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 677363

1. Introduction

The Food and Agriculture Organization of the United Nations (FAO) states that there are more than 570 million farms in the world, and that the vast majority of these are small or very small. About 94% of the world's farms are less than 5 hectares in size (FAO, 2014). In many developing countries, farm sizes are becoming even smaller, where small parcels with typically < 2 ha represents approximately 50% of rural populations (Morton, 2007). However, the lack of information on the extent and distribution of small farms still remain uncertain or unknown (Coreletto et al., 2013; Fritz et al., 2010; Holland et al., 2016), because small farms are mostly excluded from the official statistical surveys (Davidova et al., 2013; Fredriksson et al., 2017).

The project Small Farms, Small Food Business and Sustainable Food Security (SALSA) intends to assess the role of small farms and small food business in terms of food production and food security. One important first step in doing this is to test and develop methods and tools able to produce accurate and useful information about small farms. It is as such that SALSA work package 2 (WP2) intends to demonstrate the capabilities and usefulness of Copernicus Sentinel-2A satellite as a data-based method for small farms monitoring, specifically in providing information on the small farms distribution (where are they?), crop types (crop diversity), crop area extent (crop acreage), and yield estimates (crop production) to objectively quantify the crop production capabilities of small farms. By considering a gradient of 30 reference regions in Europe and in Africa, the capabilities of Sentinel-2A will be tested in very differently structured farm landscapes, allowing a better understanding of the accuracy and effectiveness of Sentinel-2A for small farms monitoring.

The WP2 is led by the University of Évora (Portugal) and co-lead by the Agricultural University of Athens (Greece), working in close collaboration with all the SALSA national contacts in Bulgaria, Cape Verde, Croatia, Spain, Czech Republic, France, Ghana, Italy, Kenya, Latvia, Lithuania, Malawi, Norway, Poland, Romania, Tunisia, and Scotland.

WP2 encompasses four main tasks, namely (i) task 2.1 small farms distribution in Europe, (ii) task 2.2 selection of the reference regions; (iii) task 2.3 small farms characterization in the reference regions; and (iv) predictive modelling. This report focuses on the task 2.3., which has two main outputs: a) a map presenting an estimate of the spatial distribution of small farms in each reference region; and b) a crop type map in small farms context in each reference region. According to the deliverable D.2.2, only the first output of the task 2.3 is presented in this report, being posteriorly reported the crop type map (output b) on the deliverable D.2.3.

It is as such that this report aims to show the obtained maps of small farms distribution in each reference region by using Sentinel-2A-derived data, as well as a summary of the main methodological steps and an quantitative assessment of the Sentinel-2A accuracy in estimating small-scale farms distribution.

2. Reference regions under analysis on this report

From the 30 selected reference regions 25 are located in Europe and 5 in Africa (Figure 1). In this report only the European reference regions are considered because the WP2 tasks for the African regions only started in July 2017 – this is due both to the project internal programming and the particularities of the growth season in the African regions included in SALSA. From the European set, four reference regions were excluded: a) two in Scotland (Perth and Kinross, and Stirling; Lochaber, Skye and Lochalsh, Arran and Cumbrae, Argyll and Bute) this exclusion is due to the lack of cloud-free Sentinel-2A images in the whole period considered for the images processing; b) one in France (Ille-et-Villaine), which is a region to be analyseds under a sub-contract with of the SALSA partners, and due to long administrative procedures, was not possible to include in the time framework available for the analysis, and c) one in Portugal (Alentejo Central) - here, small farms are mainly aggregated around villages, and there is available high-resolution regional land cover map, with high enough quality to be used to extract information about small farms in this region. Therefore the results here reported are related to 21 European reference regions (Table 1).

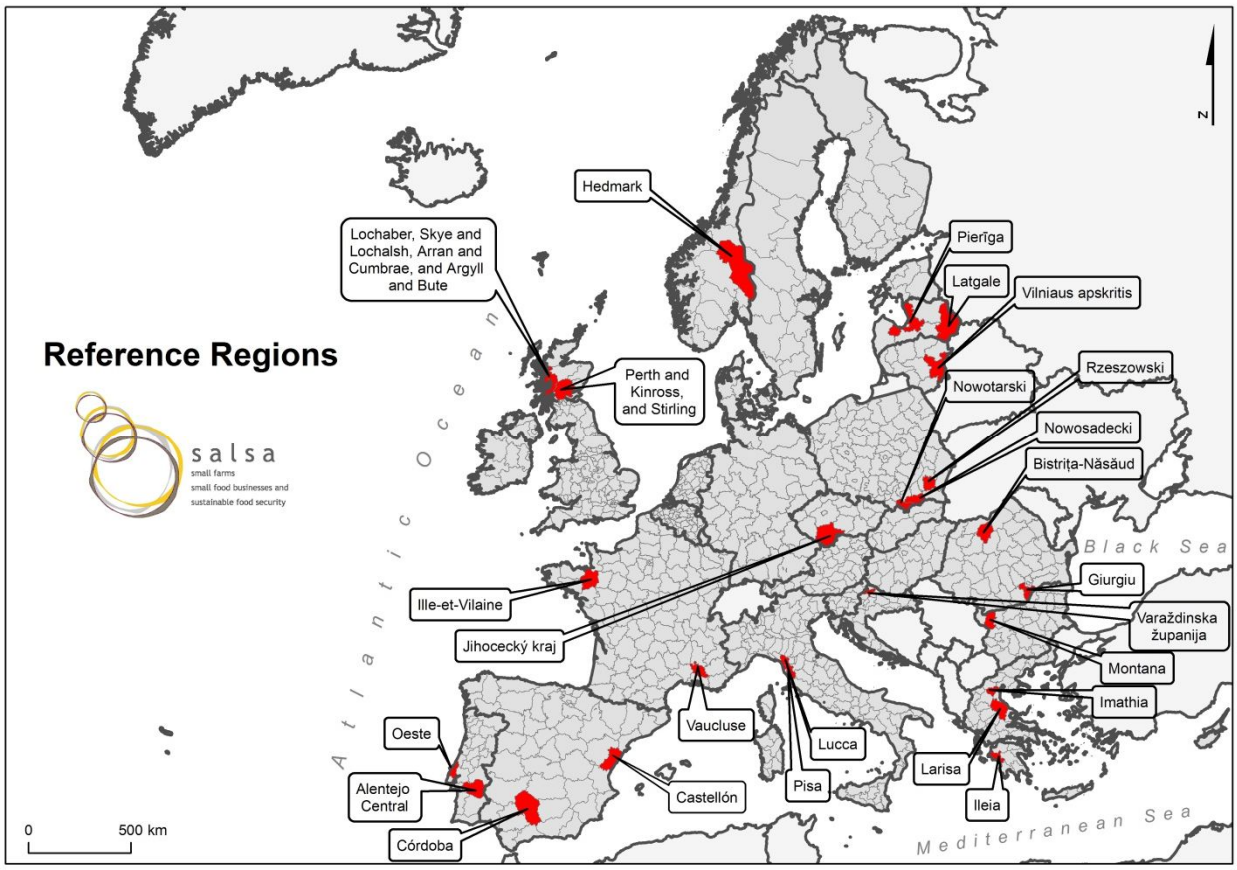


Figure 1 - The geographic distribution of the reference regions

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 677363

Table 1 - European Reference regions reported on the Deliverable 2.2.

CODE	COUNTRY	REFERENCE REGION
R1	Bulgaria	Montana
R3	Croatia	Varaždinska
R4	Czech Rep.	Jihocecký kraj
R6	France	Vaucluse
R8	Greece	Imathia
R9	Greece	Larisa
R10	Greece	Ileia
R11	Italy	Lucca
R12	Italy	Pisa
R14	Latvia	Latgale
R15	Latvia	Pierīga
R16	Lithuania	Vilniaus apskritis
R18	Norway	Hedmark
R19	Poland	Rzeszowski
R20	Poland	Nowosadecki
R21	Poland	Nowotarski
R23	Portugal	Oeste
R24	Romania	Bistrița-Năsăud
R25	Romania	Giurgiu
R26	Spain	Castellón
R27	Spain	Córdoba

3. Material and methods

3.1. Overall approach

The estimated small farms distribution over each reference region involved the implementation of two main stages: i) build an agriculture and non-agriculture mask to exclude for the subsequent analysis all the non-agricultural lands existing in each reference region (land with agriculture-like land cover), and ii) estimation of a surface map presenting the probability of small farms presence in a square grid with 250 x 250 m size. The main steps implemented in each stage are shown on the Figure 2 and explained on the next sections.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 677363

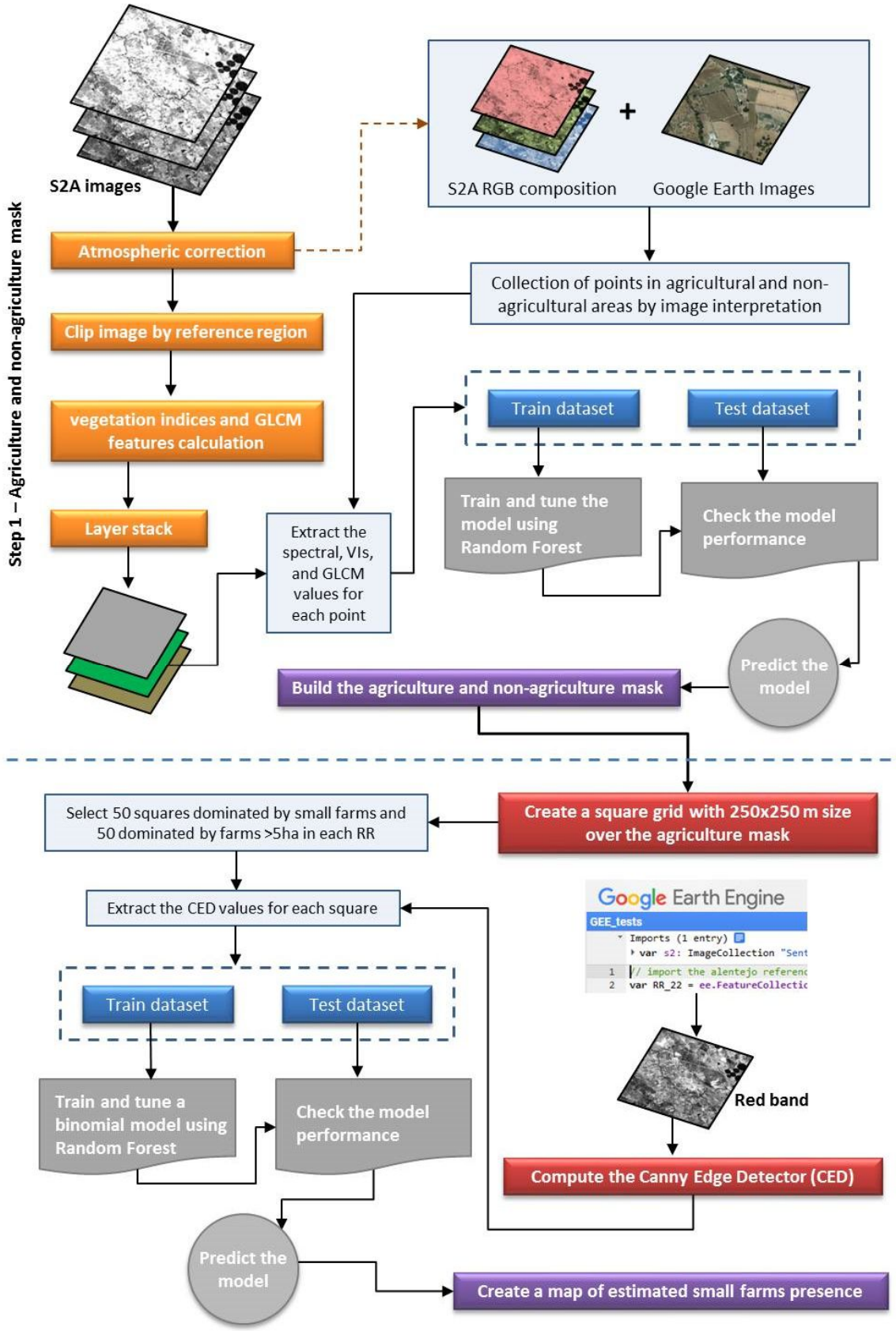


Figure 2 – Processing scheme to estimate small farms distribution maps

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 677363

3.2. Sentinel-2A data, pre-processing and auxiliary variables

Sentinel-2A is a wide-swath and high-resolution satellite with 13 spectral bands with spatial resolution ranging from 10 m to 60 m. Covering a field of view of 290 km, the 13 spectral bands are collecting information in the visible (VIS), near infrared (NIR) and shortwave infrared (SWIR) wavelengths, with four bands at 10 m, six bands at 20 m and three bands at 60 m spatial resolution (Drusch et al., 2012, Immitzer et al., 2016). For this report only the Sentinel-2A bands at 10 m and 20 m spatial resolution were used, namely the B2 (Blue), B3 (Green), B4 (Red), B5 (Red edge 1), B6 (Red edge 2), B7 (Red edge 3), B8 (NIR1), B8a (NIR2), B11 (SWIR1), and B12 (SWIR2). For all of the 21 reference regions a total of 92 Sentinel-2A images (Table 2) were used to produce i) the agriculture and non-agriculture mask , and ii) the estimation of a surface map presenting the probability of small farms presence in a square grid with 250 x 250 m size.

Table 2 – Summary of Sentinel-2A data used for the analysis

CODE	COUNTRY	REFERENCE REGION	Time period	Nº of Images	Data amount
R1	Bulgaria	Montana	July 2016	3	2.54 Gb
R3	Croatia	Varaždinska	September 2016	2	1.20 Gb
R4	Czech Rep.	Jihocecký kraj	July 2015	4	2.43 Gb
R6	France	Vaucluse	April and August 2016	6	3.79 Gb
R8	Greece	Imathia	April – July 2016	20	11.37 Gb
R9	Greece	Larisa			
R10	Greece	Ileia			
R11	Italy	Lucca	July 2016	5	2.92 Gb
R12	Italy	Pisa			
R14	Latvia	Latgale	April, May, August and September 2016	11	6.12 Gb
R15	Latvia	Pierīga			
R16	Lithuania	Vilniaus apskritis	August 2016	4	1.64 Gb
R18	Norway	Hedmark	August 2015 and August 2016	8	4.6 Gb
R19	Poland	Rzeszowski	August 2016	7	4.03 Gb
R20	Poland	Nowosadecki			
R21	Poland	Nowotarski			
R23	Portugal	Oeste	April and July 2016	2	0.87 Gb
R24	Romania	Bistrița-Năsăud	July and September 2016	6	3.58 Gb
R25	Romania	Giurgiu			
R26	Spain	Castellón	January, February, July and August 2016	14	9.07 Gb
R27	Spain	Córdoba			

A summary of the main methodological steps for the pre-processing and auxiliary variables computation are described below:

- Quantification of the number of Sentinel-2A images necessary for each European reference region;

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 677363

- Selection and download of cloud-free (<10%) Sentinel-2A images (Level 1C Top-of-Atmosphere reflectance (ToA)) from the ESA SciHub; Note: images from the spring/summer growth season of 2016 were used in this report, with the exception of Jihoceckýkraj (Czech Republic) and Hedmark (Norway) with images from July and August 2015, respectively.
- Atmospheric correction: conversion of ToA values to surface reflectance using the image-based atmospheric correction Dark-Object Subtraction (DOS1) method;
- Clipping Sentinel-2A images using the border shapefile of each reference region;
- Images mosaicking;
- Calculation of four vegetation indices (Enhanced Vegetation Index (EVI), Normalized Difference Vegetation Index (NDVI), Short Wave Infrared Reflectance 3/2 Ratio (SWIR32), and Plant Senescence Reflectance Index (PSRI)) and three GLCM features (Variance, Mean, and Homogeneity, with 3x3 window size) to use as auxiliary variables in the classification procedure;
- Layer stack to produce one raster layer with all clipped sentinel-2A bands, plus vegetation indices and GLCM features.

3.3. Agriculture and non-agriculture mask

A mask to exclude non-agricultural areas from the estimated small farms distribution process was created. The spatial information used as well as the geo-processing steps to build the mask is listed below:

- Collection of Corine Land cover maps for each reference region;
- Creation of a randomly stratified sample points (~1000; 500 points for non-agriculture fields and 500 for agriculture fields) over the reference region (first the land cover categories were defined using the Corine nomenclature). This procedure ensures that the 1000 points will cover the diversity of the main land cover types existing in each reference region. For this procedure the Sampling Design Tool for ArcGIS 10 was used (http://www.arcgis.com/home/item.html?id=ecbe1fc44f35465f9dea42ef9b63e785);
- Codification of each point as “n_agri” for non-agriculture and “agri” for agriculture by visual-interpretation of high-resolution Google Earth imagery and Sentinel-2A true colour composition;
- Split the main dataset in training (80%) and test (20%) datasets using the createDataPartition function from the caret R package (Khun, 2014);
- Preparation of an R script for image classification process, accuracy assessment, and prediction, using Random Forest as a classifier;
- Classification based on Random Forest machine learning algorithm;
- Accuracy assessment based on the test dataset (overall accuracy, kappa index);
- Prediction to create the “agriculture non-agriculture mask” using the best Random Forest model;

Afterwards, the created non-agricultural raster was converted to vector format and used to clip all the pre-processed S2A images to guarantee that the procedure for the estimation of small farms distribution is only performed in agricultural lands – meaning land under an agriculture-like land cover.

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 677363

3.4. Estimated distribution of small farms

It is recognized that information about field size can be used as a proxy of farm size because there is a positive correlation between field size and farm size (Fritz et al., 2015; Levin et al., 2006). Moreover, landscape heterogeneity, in particular the configurational heterogeneity, can be used to measure the degree of spatial complexity of the landscape pattern (Fahrig and Nuttle, 2005). In the context of farmland, Fahrig et al. (2015) stated that farmland with higher configurational heterogeneity have smaller crop fields and a greater total length of field edges. Therefore, the link between total length of field edges and the farm size can be established and used as a proxy to infer about farm size. The higher the total length of agricultural field edges, the smaller the crop fields, and thus, the smaller the farms. Figure 3 illustrates two examples of farming systems (from Giurgiu, Romania), the first represents the dominance of large farms (upper left square, 2x2 km), and the second one demonstrates the dominance of small farms (upper right square). The configurational heterogeneity is much higher in the small farm context than in the large farms, presenting a huge difference in terms of total edge length.

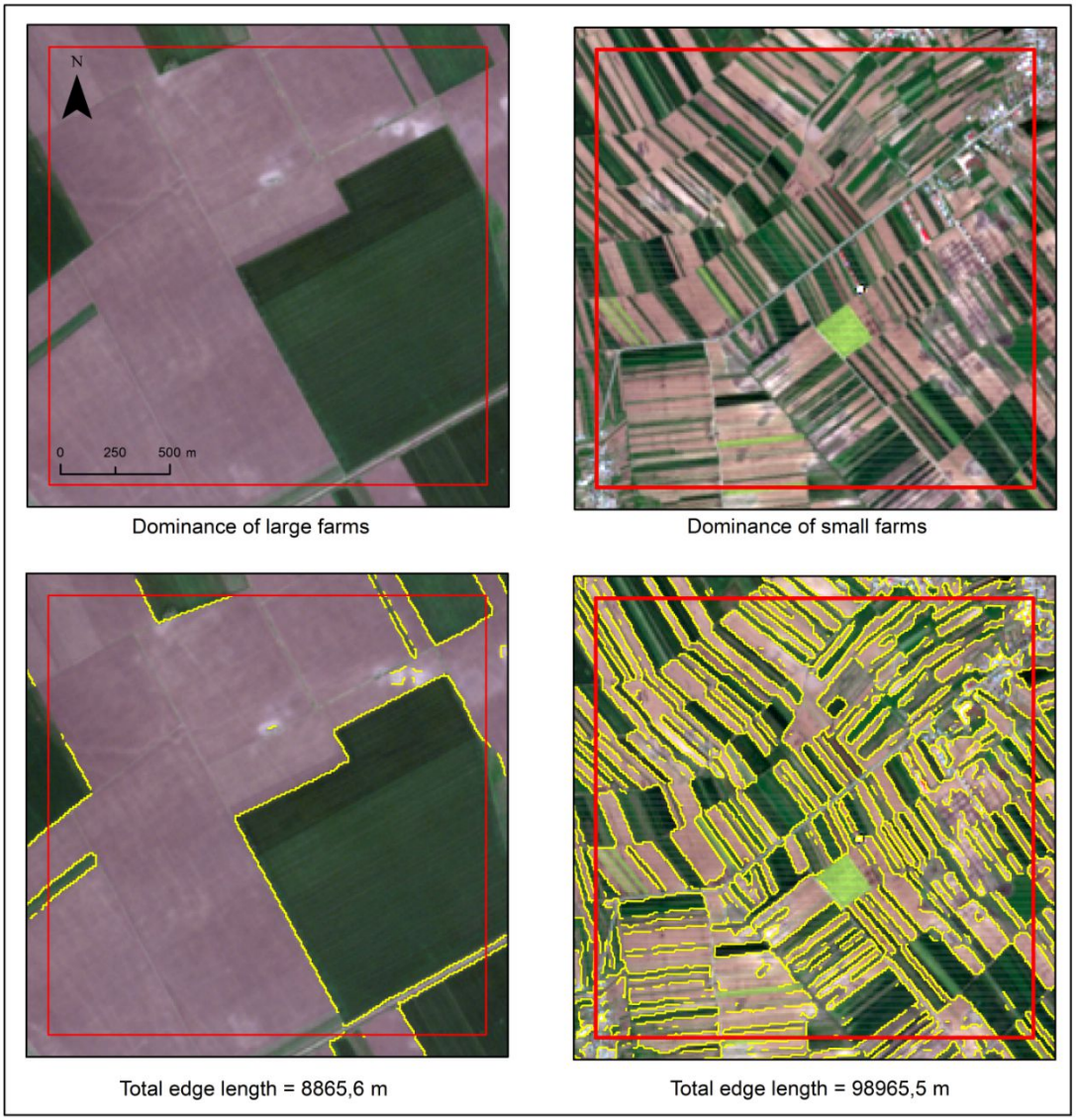


Figure 3 – Two examples of contrasting agricultural farm plot sizes and respective total edge length computed using the Canny Edge Detector algorithm.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 677363

To produce an estimate about the extent of small farms over each reference region, a probabilistic model was developed using the edge length as the predictor variable and the dominance (1) and non-dominance of small farms (0) as dependent variable. This was developed through the following steps:

1. Small and non-small farms data collection
2. Edge length computation using Canny Edge Detector algorithm (CED);
3. Probabilistic model for small farms prediction using machine learning algorithm

To implement such approach it was assumed that small agricultural parcels are mostly related with small-scale farming system, even though acknowledging that the small farm parcels can be grouped in different patterns to constitute one farm unit.

3.4.1. Small and non-small farms data collection

To ensure a representative and feasible reference dataset about the dominance and non-dominance of small farms over each reference region, a square grid of 250x250 m was created. From that grid, and through the intersection with the agricultural mask shapefile, only the squares with more than 80% of its area covered by agricultural land were selected for the analysis. Next, an average of 50 squares with dominance of small farms (coded as 1) and 50 dominantly covered by large farms (coded as 0) was selected in each reference region. These data were collected using visual-interpretation of high-resolution Google Earth imagery and the true color composition of the Sentinel-2 images acquired for each region. An example of the described procedure is illustrated in Figure 3, where the first image should be coded as 0 (large farm dominance) and the second coded as 1 due to the dominance of small agricultural parcels, thus high probability to be small farms. Among the 21 reference regions a dataset with 1942 squares were produced and used for the predictive model building.

3.4.2. Edge length computation using Canny Edge Detector (CED)

The Canny Edge Detector (CED) (Canny, 1986) is widely considered to be the standard and an effective edge detection algorithm in image processing. CED is a numerical optimization criterion to derive several common image features, such as the step edges. Overall, CED computation encompasses four main steps:

1. image smoothing by using Gaussian convolution filter to remove high-frequency noise from the image under analysis;
2. computation of two-dimensional first derivatives: the gradient magnitude (edge strength) and gradient direction (edge orientation). This process shows changes in intensity, which indicates the presence of edges;
3. non-maximal suppression process, which is applied to the gradient magnitude image to identify the local maxima. Edges will occur at pixels where the gradient is at a maximum. All pixels presenting non-maximum gradient values are suppressed. For this step the magnitude and direction of the gradient is computed at each pixel;

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 677363

4. edge thresholding, where CED algorithm makes use of both a high and low threshold. All pixels with a value above the high threshold are classified as edge pixels. Pixels with a value above the low threshold and neighbour of edge pixels are considered as edge pixels as well. If a pixel has a value above the low threshold but is not the neighbour of an edge pixel, it is not set as an edge pixel. If a pixel has a value below the low threshold, it is never set as an edge pixel.

To extract information about the total edge length existing in each reference region, CED algorithm was computed using the red channel (Band 4, 10-meters spatial resolution) of the Sentinel-2A image. During the tests phase of the WP2, the CED algorithm was computed testing the four 10-m spatial resolution Sentinel-2A bands (B2, B3, B4, and B8). Among all the tests the red band was the one with the best performance in detecting edge pixels. The computation of the CED algorithm was carried out through the use of Google Earth Engine (GEE), which is a cloud-computing platform for processing satellite images and geospatial datasets. For this a small java script were prepared and executed on the code editor of the GEE platform. After CED computation, the total edge length (in meters) existing in each 250x250m square with more than 80% of agricultural area was calculated for each reference region.

3.4.3. Probabilistic model for small farms prediction and accuracy assessment

Random Forest (RF) algorithm (Breiman 2001) was used to generate the probabilistic model to predict the presence of small farms. The effectiveness of RF algorithm for remote sensing applications has been demonstrated in several studies (e.g. Freeman et al., 2015; Rodriguez-Galiano et al., 2012; Wang et al., 2016). According to Rodriguez-Galiano et al (2012) this machine-learning algorithm presents many advantages:

I. It runs efficiently on large data bases;
II. It can handle thousands of input variables without variable deletion;
III. It gives estimates of what variables are important in the model;
IV. It generates an internal unbiased estimate of the generalization error (oob error).
V. It computes proximities between pairs of cases that can be used in locating outliers.
VI. It is relatively robust to outliers and noise.
VII. It is computationally lighter than other tree ensemble methods (e.g. Boosting).

For the RF modelling procedure the 1942 selected squares (coded as 0 and 1) were split into training (80%) and test (20%) datasets using the *createDataPartition* function from the *caret* R package (Khun, 2016). This function is useful for creating balanced splits in the data. It ensures that random sampling occurs within each class (in this case 0 and 1) while also preserving the overall class distribution over the dataset (Kuhn and Johnson, 2013). The training dataset (n= 1554 squares) was used to train RF models by applying a repeated (five times) 10-fold cross validation resampling method. The test dataset (n= 388 squares) was used to evaluate the model performance through the computation of the Receiver Operating Characteristic (ROC) curve and its derived accuracy index Area Under the Curve (AUC). The maximum AUC=1 means that the model is perfect in differentiating between squares with dominance of small farm plots and squares coded as non-dominated by small farm plots. The threshold above which a square should be considered as dominantly occupied by small farm plots was also computed from the ROC curve. To implement all the modelling and predictions steps an R script was created.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 677363

4. Preliminary Results

4.1. Agriculture and non-agriculture mask

The RF classifications performed for the selected Sentinel-2A spectral bands, vegetation indices and textural features showed a strong overall agreement and good accuracy among all the reference regions studied (Overall Accuracy > 0.90; Kappa index > 0.80) (Figure 4). Figure 5 illustrates the agricultural mask obtained for the Varaždinska (Croatia) reference region, where the first image is a false colour composition (S2A bands: B8, B3 and B2) highlighting forest patches, agricultural plots, roads and villages, while the second image shows the agricultural land area obtained by the RF classification model (blue patches). Maps of the Agricultural areas for each reference region are presented in the ANNEX I.

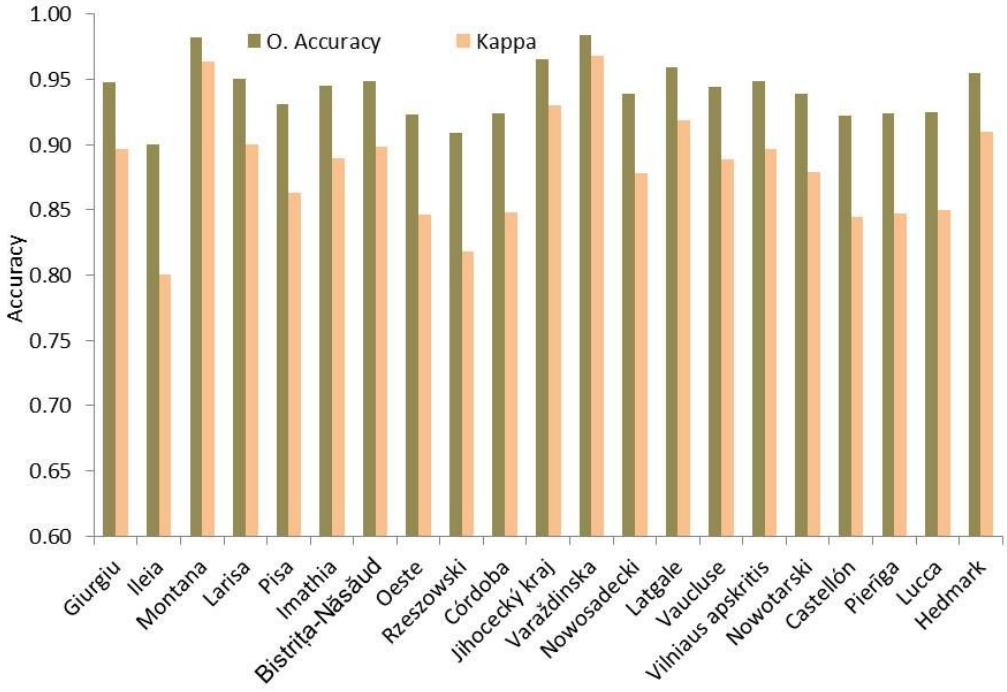


Figure 4 – Overall accuracy and Kappa index values obtained for each reference region

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 677363

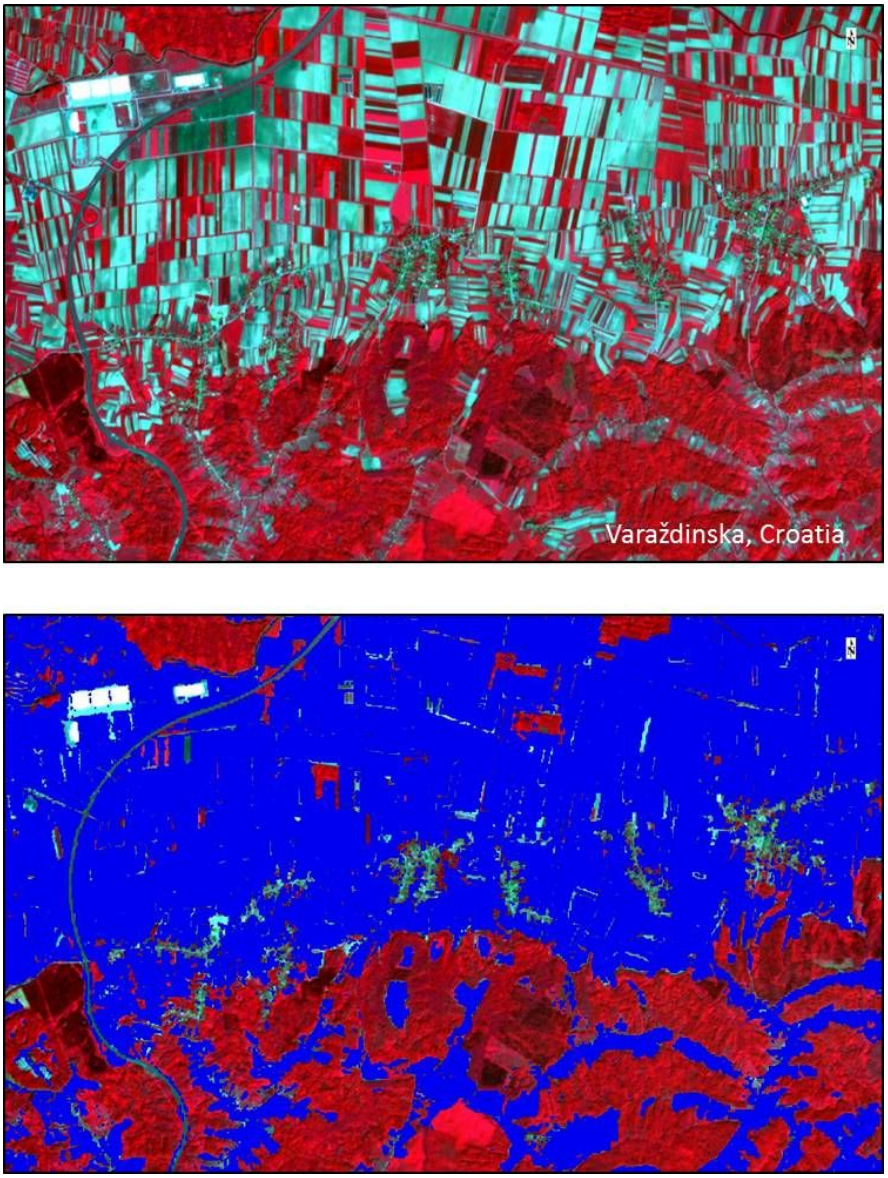


Figure 5 – Agricultural mask (in blue) obtained by applying Random Forest classifier to the Sentinel-2A spectral bands, vegetation indices and textural variables.

The results included in Figure 6 shows the per-region agricultural area (in percentage) estimated from the agricultural mask. One may observe that the region with the highest agricultural area is Giurgiu (Romania) with 77.2% covered by agricultural land, while the lowest agricultural area percentage was found in Hedmark (Norway) presenting 19.8%. From the Agricultural Census and Farm Structure Survey (FSS) (2007-2010) Giurgiu and Hedmark are indeed considered as the highest and lowest regions in terms of agricultural land area percentage, respectively. However, significant differences were found when compared the percentage values of agricultural areas estimated from the Sentinel-2A and the agricultural census. A clear example of such differences can be observed in the Montana (Bulgaria) reference region, where the official statistics only estimated 24.3% of agricultural areas, while the results here reported estimates that 62.5% of this region is covered by agricultural land, meaning a very significative difference of 38.2%. The 62.5% of agricultural areas reported with the present analysis, for

This project has received funding from the European Union's Horizon 2020
research and innovation programme under grant agreement No 677363

Montana, is in agreement with the values obtained when using the Corine Land Cover map, which determines that Montana has in fact ~63% of agricultural areas. Moreover, by visual interpretation of the Sentinel-2A images and the very high-resolution Google Earth Images, is easy to conclude that Montana is mostly covered by agricultural plots in activity. In other regions also disagreements were found, though less high than in the Bulgarian region. The above mentioned disagreements with the official statistics, and considering only differences above 10%, were detected in 12 of the 21 reference regions here described. The agricultural area reported by the agricultural census and FSS datasets refers to the Utilised Agricultural Area (UAA) which includes arable lands (irrigated and non-irrigated), permanent grasslands, permanent crops (olive groves, vineyards and orchards), and other agricultural lands such as kitchen gardens. UAA does not include unused agricultural land, and this can explain some of the differences between the area estimated by Sentinel-2A and agricultural statistics. However, and considering the Montana case, a report from the Institute for European Environmental Policy (Keenleyside and Tucker, 2010) focused on farmland abandonment in Europe stated that non-utilized agricultural land in Bulgaria only represents 4.1% of the territory. Over the next stages of the SALSA project, these issues will be analysed and discussed towards a better understanding of the main causes of such differences.

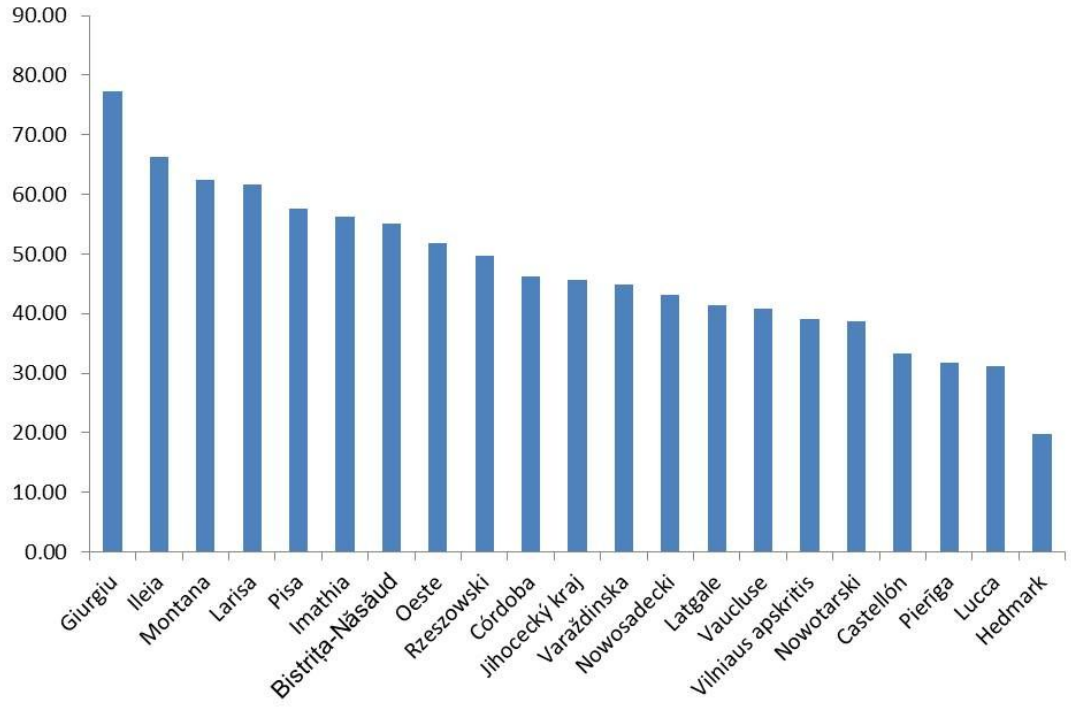


Figure 6 – Percentage of agricultural area estimated for each reference region.

4.2. Estimated distribution of small farms

The images represented in the Figure 7 reveal that the first results about the use of Canny Edge Detector algorithm seem to be promising in detecting and delineating the borders of agricultural plots. It can be seen that the majority of the edges between the agricultural plots existing in this area (Varaždinska, Croatia) was correctly extracted.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 677363

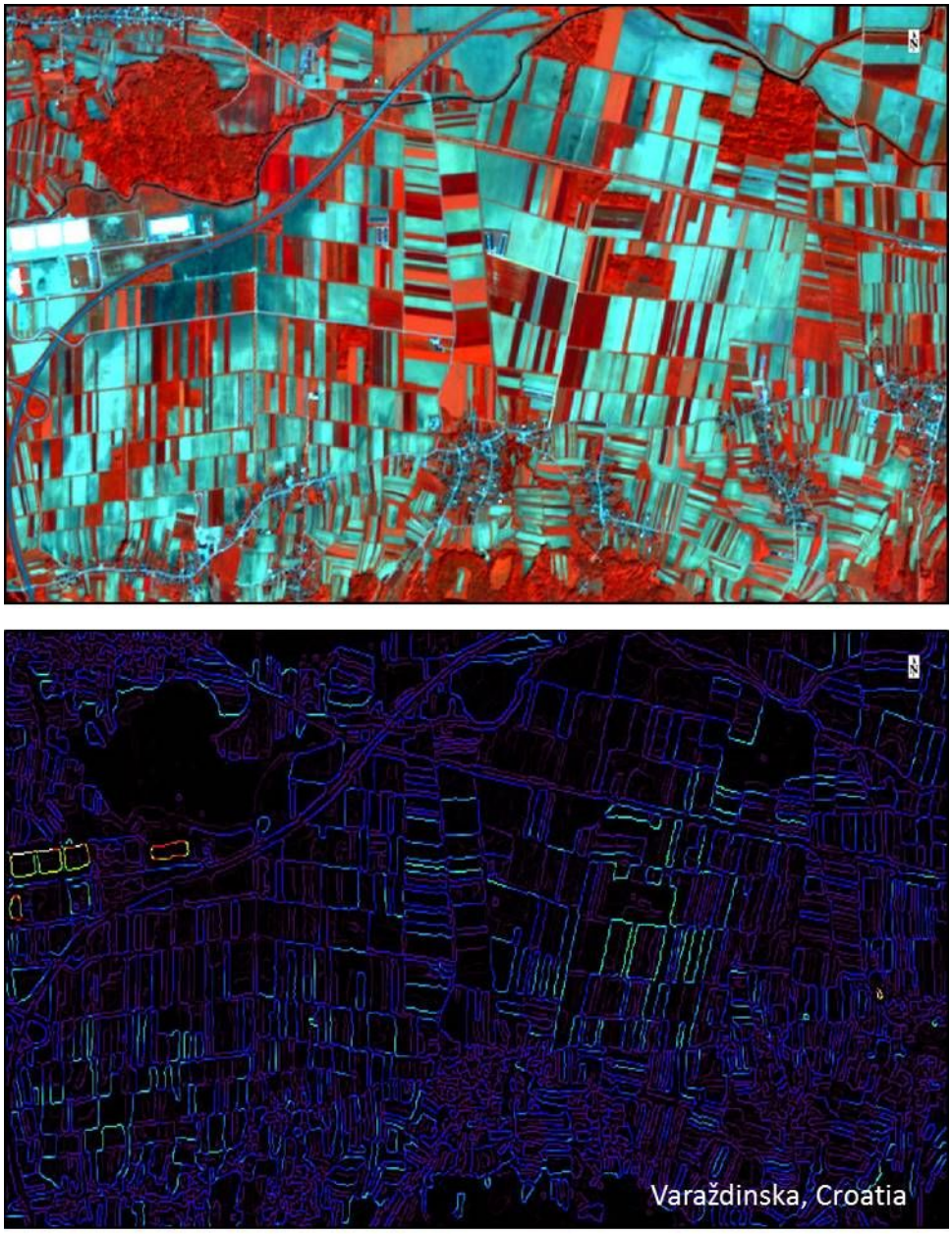


Figure 7 – Canny Edge Detector output. Upper image refers to the Sentinel-2A false colour composition using the B8, B3 and B2 bands.

In order to evaluate the effectiveness in using plots edge information as a proxy of the small farm presence, a regression analysis was performed between edge density (total edge length in the agricultural mask divided by its area) and the Mean Farm Size (MFS) in each reference region. Information about MFS was extracted from the agricultural census and farm statistical survey (2007 – 2010). Using all the 21 reference regions, the obtained results revealed a significant and moderate inverse relationship (R^2 = 0.52, p > 0.001) between the edge density and the MFS, demonstrating its usefulness for estimating farm size (Figure 8 (a)). However, it can be seen that Castellon (Spain) and Vaucluse (France) reference regions presents an atypical behaviour regarding the relationship between its MFS and the edge density. Castellon was the one with the highest edge density values (ED = 269.45

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 677363

m/ha) but is not the region with the lowest mean farm size (MFS = 5.12 ha). All the regions with farm size less than 5 ha presents an average edge density values of 202.06 m/ha.

Plots with citrus and olive trees are dominant over the Castellón agricultural landscape. In these areas, the tree cover density is generally sparse (mainly in olive groves), and therefore, the reflectance of soils not covered by the canopy trees and the reflectance of the trees itself creates a mixed spectral environment leading to different levels of reflectance intensities, which in turn will confuse the CED algorithm in detecting the real plots boundaries. In these spectral conditions the CED algorithm will compute much more edge length than what exists in reality. Therefore, in order to avoid such results and be able to use this algorithm in those areas, the CED image smoothing and edge thresholding steps (see section 3.4.2) should be performed in a more conservative way, it means, decreasing as much as possible the image noise levels and increasing the threshold from which one pixel should be coded as edge pixel.

Regarding the deviance of the Vaucluse observation to the fitted model curve the explanation is much more related to agricultural land planning processes than with the CED algorithm performance. From the spatial heterogeneity point of view this region presents an extremely fragmented agricultural landscape, where small and medium agricultural plots dominate the landscape. However, due to land consolidation process in France mean farm size has been increasing (Boinon, 2011). Vaucluse presents a MFS of 20.2 ha and a similar edge density (199.39 m/ha) with the regions with MFS less than 5 ha. Actually this edge density exists because the agricultural land in Vaucluse is extremely fragmented, therefore the use of this metric as a direct proxy to estimate farm size in these situations should be performed with caution. Inevitably, these two cases (Castellon and Vaucluse) will affect the model performance, and this can be observed when both regions were excluded from the analysis, leading to an R^2 of 0.65 (Figure 8 (b)).

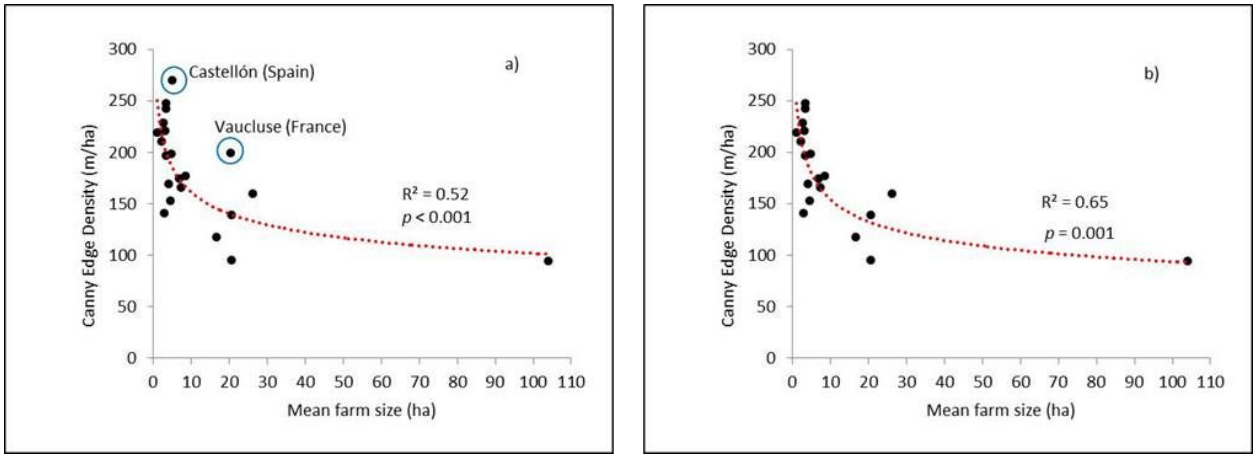


Figure 8 – Statistical relationship between edge density and the mean farm size (MFS).

The statistical relationship between edge length and farm size was generally demonstrated in the previous paragraphs. This constitutes the base line for the development of a probabilistic model to estimate the presence of small farm plots using edge length information as a proxy variable. It is as such that a random forest model was adjusted using a training dataset with 1554 squares (see section 3.4.3). After testing the potential of edge length for predicting the presence of small farm plots based on

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 677363

Sentinel-2A images it can be observed that derived model produced plausible results when applied to an independent dataset (n=388): AUC of 87.1%. From the ROC curve a threshold value of 0.709 was obtained (Figure 9), meaning that after model prediction all squares with probability values greater than 0.709 should be considered as dominantly covered by small farm plots.

A map showing the surface probability values for the presence of small farm plots was computed for each reference region using the adjusted random forest model (ANNEX I).

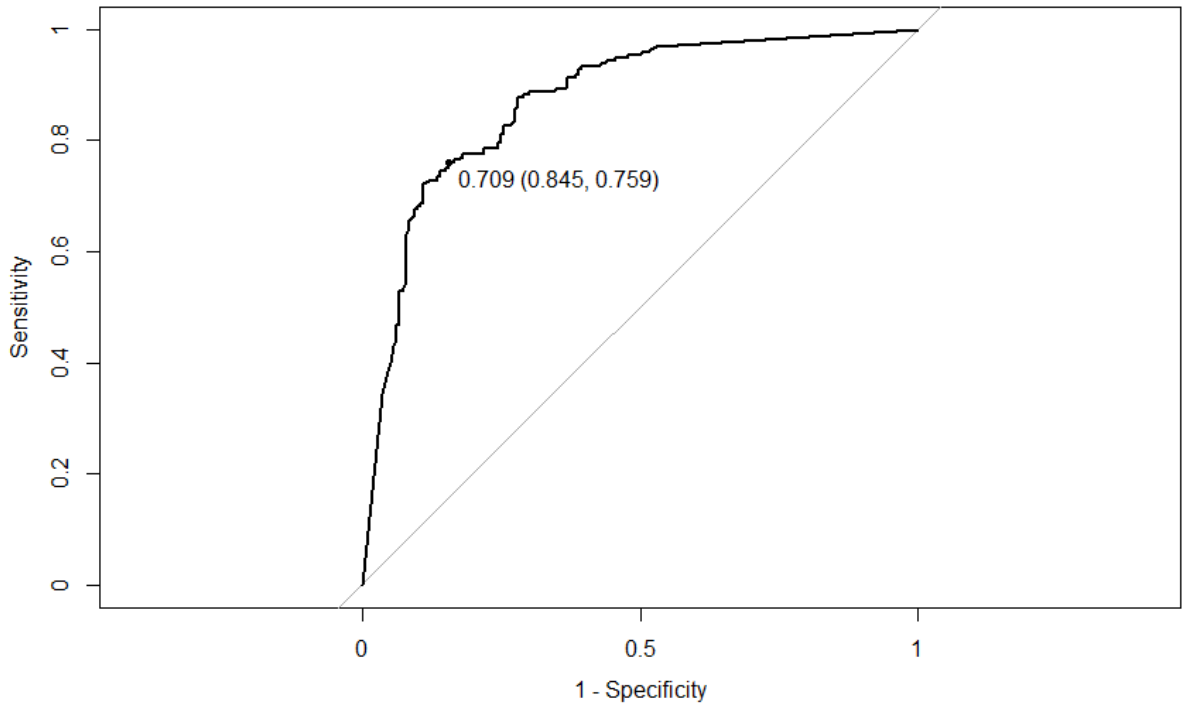


Figure 9 – ROC curve obtained from the test dataset presenting a threshold of 0.909 and an area under the ROC curve of 87.1%.

From these maps, the percentage of the agricultural area occupied by small farm plots was estimated for each reference region. According with the graph represented in Figure 10, showing the percentage of agricultural area classified as high probability for the presence of small farm plots, excluding Castellon and Vaucluse due to the reasons already mentioned, small farms are predominant (> 50% of agricultural area occupied by small farm plots) in Nowosadecki (Poland), Ileia (Greece), Imathia (Greece), Varaždinska (Croatia), Lucca (Italy), Rzeszowski (Poland), Bistrița-Năsăud (Romania), and Oeste (Portugal).

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 677363

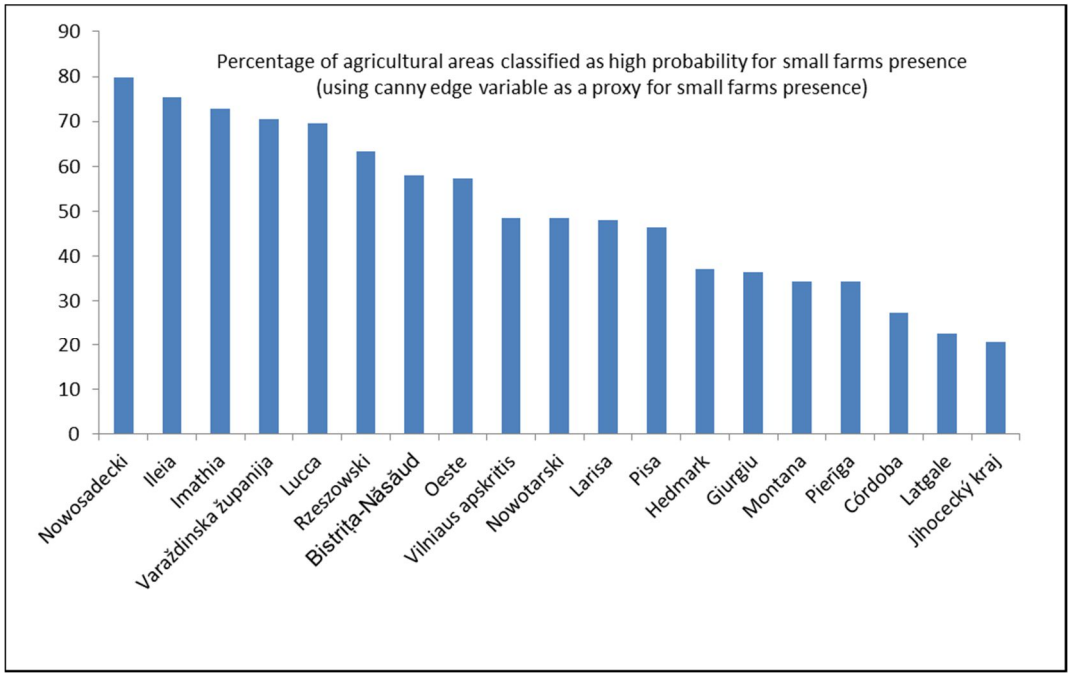


Figure 10 – Percentage of agricultural area classified as high probability for the presence of small farm plots estimated for each reference region.

5. Final remarks

This report presented the preliminary results regarding the assessment of the capabilities and usefulness of Sentinel-2A satellite as a data-based method for small farms monitoring, in particular to estimate the spatial distribution of small-scale farm systems. The main findings achieved constitute probably the first remote sensing-based small farm distribution map developed by using Sentinel-2A imagery. As a first step, a set of agricultural and non-agricultural maps were produced for 21 reference regions with good levels of accuracy (OA > 0.90 and Kappa>0.80), demonstrating that this sensor is suitable for generating agricultural maps with high accuracy for different European environmental and territorial conditions.

Results from the first step highlighted the fact that not all the agricultural areas (used and unused) are being considered in the official agricultural statistics. The main difference between the results here reported and the existing statistics is related to the fact that UAA excludes non-utilized agricultural lands. Nevertheless, this alone does not explain the huge difference detected in some regions, such the one verified in Montana. It is not clear whether public or communal lands are included or not in the UAA estimation in some countries, for that reason the UAA may not reflect the real agricultural land area extent. Thus UAA statistics should be used carefully mainly when small farms are the main topic to be addressed.

This work hereby reported also provide an interesting and promising methodological approach to estimate small farm plots distribution based on Canny Edge Detector algorithm and high-resolution Sentinel-2A bands (in this case only red band) towards the operational use of Sentinel-2A images for

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 677363

small farms monitoring. Although it has successfully demonstrated that the use of edge length/density as a proxy variable to infer about farm size and small farm presence probability, the results also highlighted two main limitations in using this methodological approach; i) the use of this method in regions where scattered trees in co-occurrence with soils presenting high reflectance (e.g Castellon) can overestimate the edge length over the agricultural plots, and ii) the use of CED algorithm in agricultural landscapes with high levels of configurational heterogeneity, but where land planning issues such as land consolidation determines the land ownership without altering the spatial configuration of agricultural plots. In summary, these technical and methodological issues have not been solved and will require further development.

The analysis in the WP2 is under progress towards the analysis of the spatial distribution and estimated production potential of small farms in each reference region of SALSA. Further results are expected to be communicated after this next phase.

6. References

Breiman, L., 2001. Random Forests. Machine Learning, 45:5-32.

Boinon, J.P., 2011. Les politiques foncières agricoles en France depuis 1945. Economie et Statistique 444–445, 19–37.

Canny, J., 1986. A computational approach to edge detection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 8(6): 679 - 698.

Carletto, C., Jolliffe, D., and Banerjee, R., 2013. The emperor has no data! Agricultural statistics in Sub–Saharan Africa (Technical Report). The World Bank.

Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., Hoersch, B., Isola, C., Laberinti, P.,Martimort, P., Meygret, A., Spoto, F., Sy, O., F., Marchese, & Bargellini, P., 2012. Sentinel-2: ESA's Optical High-Resolution Mission for GMES Operational Services. Remote Sensing of Environment, 120: 25–36.

Fahrig, L., Girarda, J., Duro, D., Pasher, J., Smith, A., Javorek, S., King, D., Lindsay, K.F., Mitchell, S., Tischendorf, L., 2015. Farmlands with smaller crop fields have higher within-field biodiversity. Agriculture, Ecossystems and Environment, 200: 219 – 234.

Fahrig, L., Nuttle, W.K., 2005. Population ecology in spatially heterogeneous environments. In: Lovett, G.M., Jones, C.G., Turner, M.G., Weathers, K.C. (Eds.), Ecosystem Function in Heterogeneous Landscapes. Springer-Verlag, New York, pp. 95–118.

FAO, 2014. Towards stronger family farms: Voices in the International Year of Family Farming, Rome: Food and Agriculture Organization of the United Nations

Fredriksson, L., Bailey, A., Davidova, S. Gorton, M., Traikova, D., 2017. The commercialisation of subsistence farms: Evidence from the new member states of the EU. Land Use Policy, 60: 37 – 47.

Freeman, E. A., Moisen, G. G., Coulston, J. W. and Wilson, B. T. 2015. Random forests and stochastic gradient boosting for predicting tree canopy cover: comparing tuning processes and model performance. Canadian Journal of Forest Research, 45: 1–17.

Fritz, S., See, L., and Rembold, F., 2010. Comparison of global and regional land cover maps with statistical information for the agricultural domain in Africa. International Journal of Remote Sensing, 31: 2237–2256.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 677363

Fritz, S., See, L., Mccallum, I., et al., 2015. Mapping global cropland and field size. Global Change Biology, 21, 1980–1992.

Holland, M.B., Shamer, S.Z., Imbach, P., Zamora, J. C., Moreno, C.M., Hidalgo, E.J.L., Donatti, C. I., Martínez-Rodríguez, M.R., Harvey, C.A., 2016. Mapping adaptive capacity and smallholder agriculture: applying expert knowledge at the landscape scale. Climatic Change, 141: 139 – 153.

Immitzer, M., Vuolo, F. and Atzberger, C., 2016. First Experience with Sentinel-2 Data for Crop and Tree Species Classifications in Central Europe. *Remote Sensing,* 8:166.

Keenleyside, C., and Tucker, G.M., 2010. Farmland Abandonment in the EU: an Assessment of Trends and Prospects. Report prepared for WWF. Institute for European Environmental Policy, London.

Kuhn M. 2016. caret: Classification and Regression Training. R package version 6.0-73. URL http://CRAN.R-project.org/package=caret, accessed January 2017.

Kuhn, M. and Johnson, K., 2013. Applied Predictive Modelling. Springer,New York.

Levin, G., 2006. Farm size and landscape composition in relation to landscape changes in Denmark. Danish Journal of Geography, 106, 45–59.

Morton, J.F. 2007. The impact of climate change on smallholder and subsistence agriculture. Proceedings of the National Academy of Sciences of the United States of America, 104: 19680–19685.

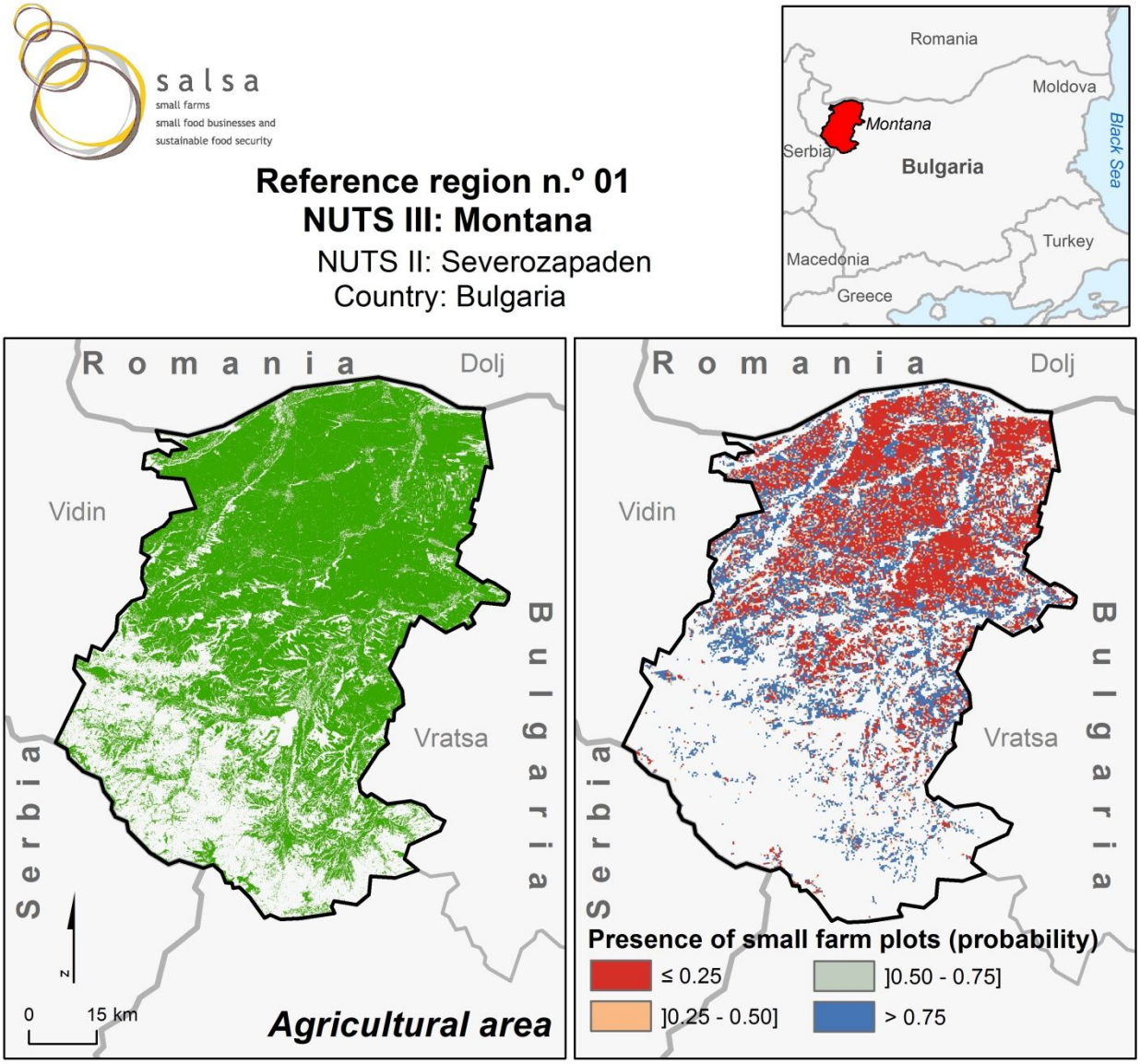
Rodriguez-Galiano, V.F., Ghimire, B., Rogan, J., Chica-Olmo, M., and Rigol-Sanchez, J.P., 2012. An assessment of the effectiveness of a random forest classifier for land-cover classification. ISPRS Journal of Photogrammetry and Remote Sensing, 67: 93-104.

Wang, L., Zhou, X., Zhu, X., Dong, Z. and Guo, W. 2016. Estimation of biomass in wheat using random forest regression algorithm and remote sensing data. The Crop Journal 4: 212–219.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 677363

ANNEX I

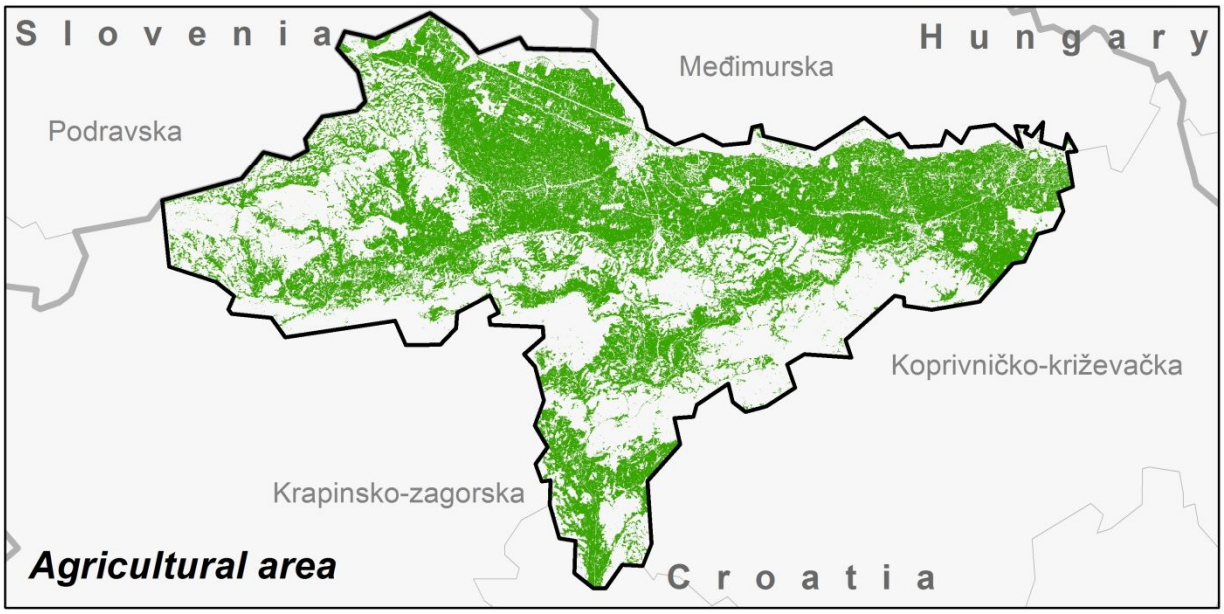
Agricultural masks and small farms distribution maps

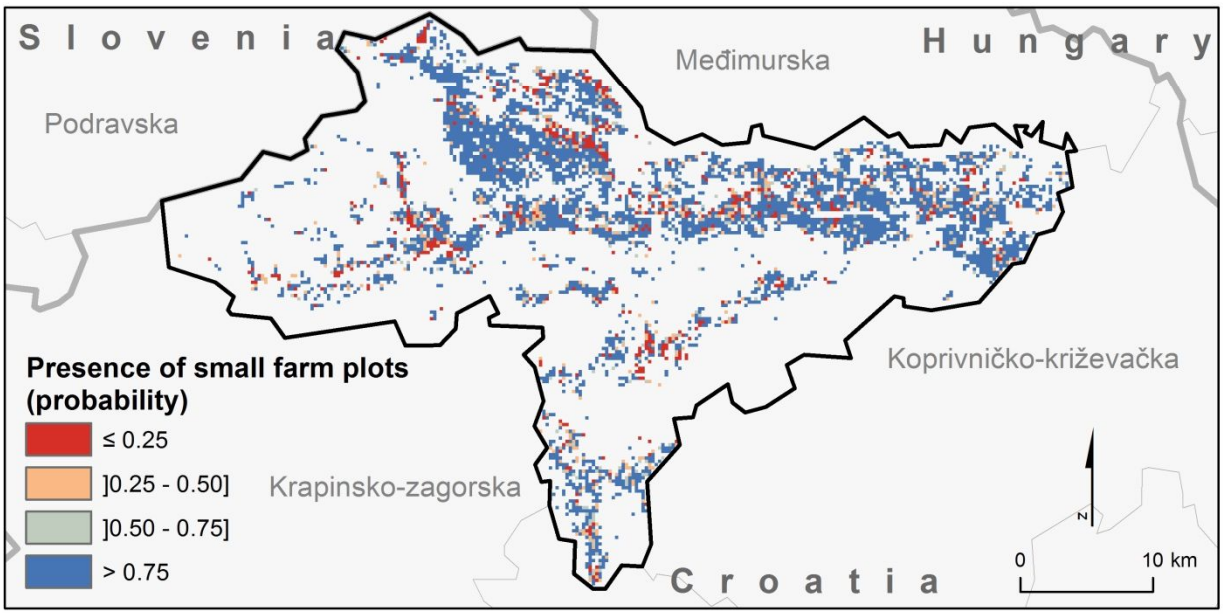


This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 677363

salsa
small farms
small food businesses and
sustainable food security

Reference region n.º 03
NUTS III: Varaždinska
NUTS II: Kontinentalna Hrvatska
Country: Croatia

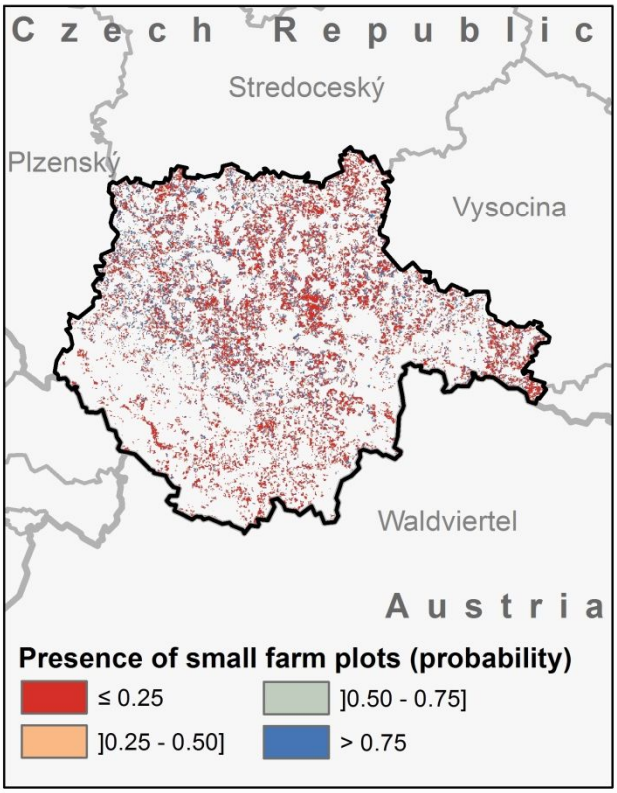




This project has received funding from the European Union's Horizon 2020
research and innovation programme under grant agreement No 677363

salsa
small farms
small food businesses and
sustainable food security

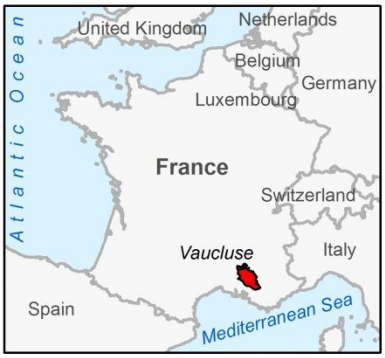
Reference region n.º 04
NUTS III: Jihocecký
NUTS II: Jihozápad
Country: Czech Republic

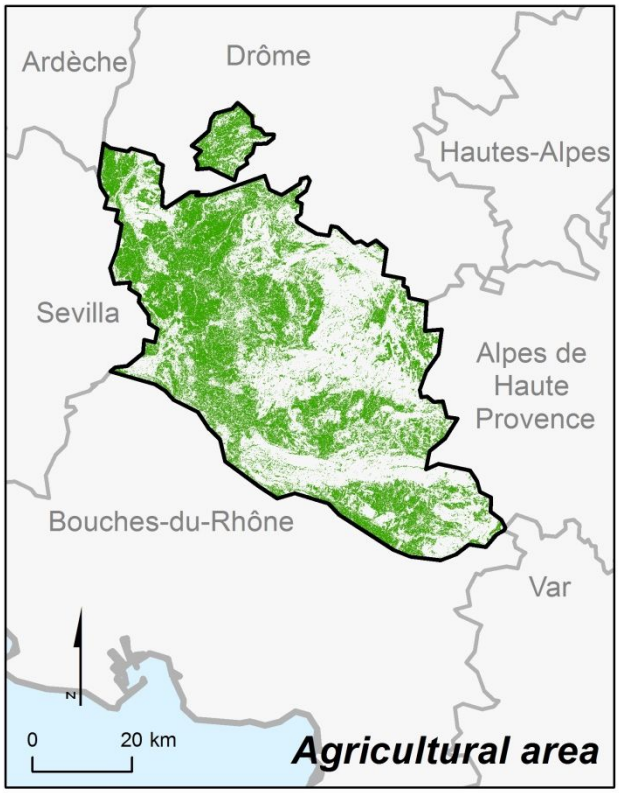


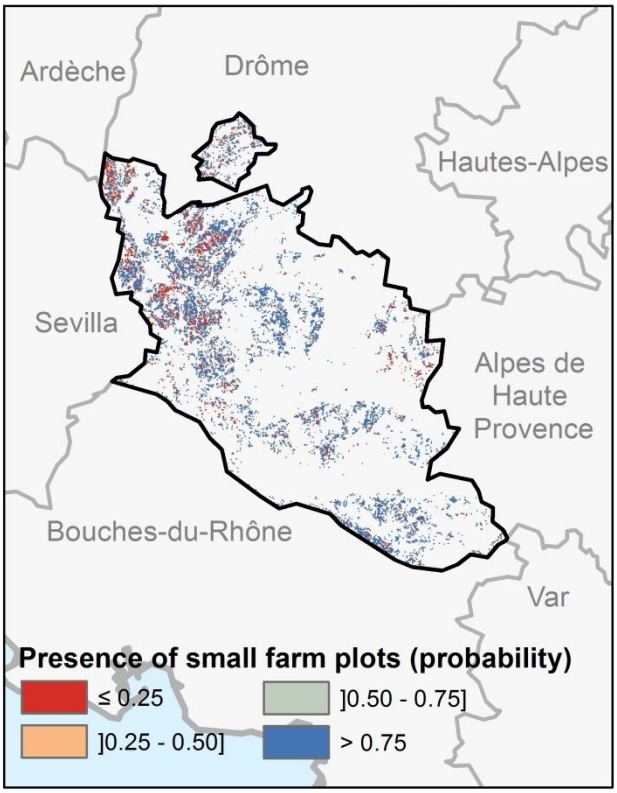
This project has received funding from the European Union's Horizon 2020
research and innovation programme under grant agreement No 677363

salsa
small farms
small food businesses and
sustainable food security

Reference region n.º 06
NUTS III: Vaucluse
NUTS II: Provence-Alpes-Côte d'Azur
Country: France



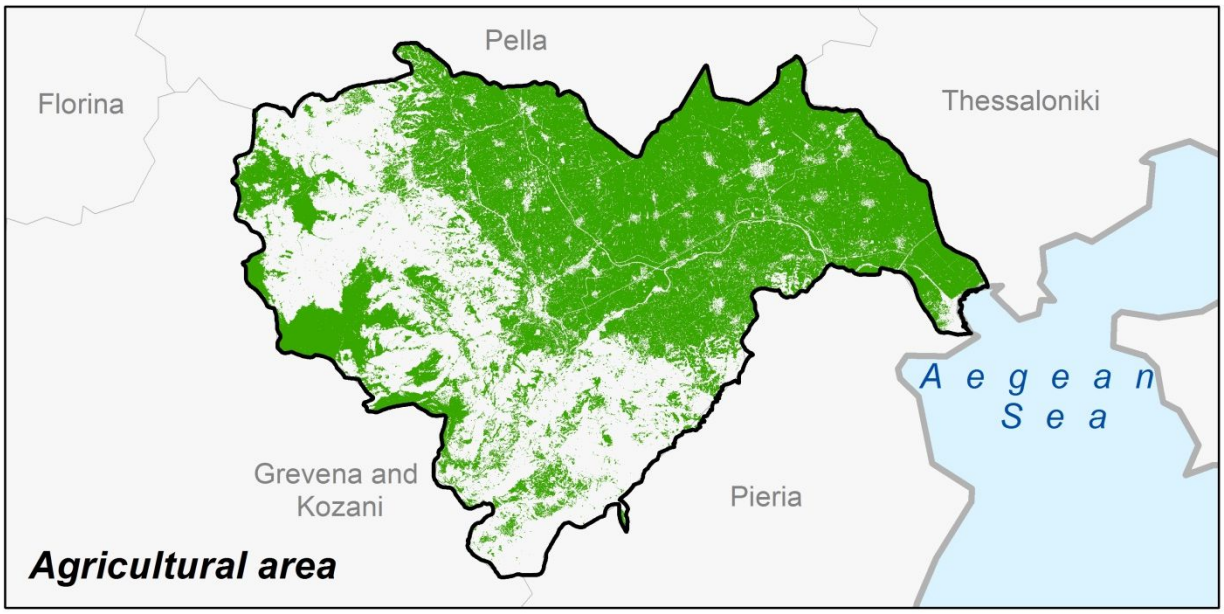


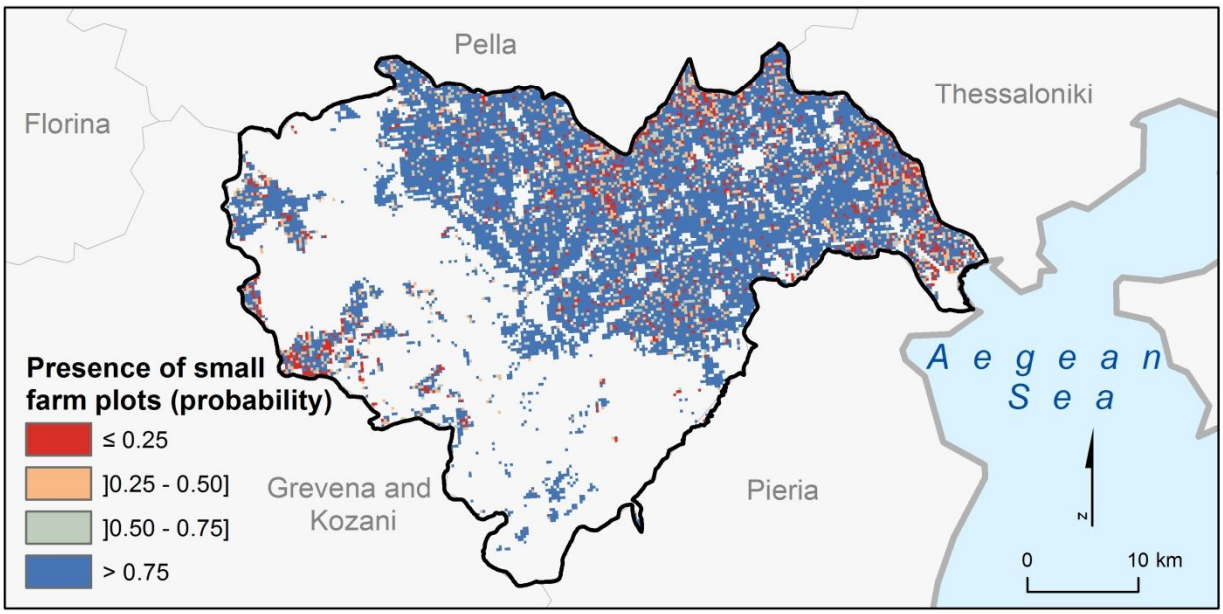


This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 677363

salsa
small farms
small food businesses and
sustainable food security

Reference region n.º 08
NUTS III: Imathia
NUTS II: Kentriki Makedonia
Country: Greece

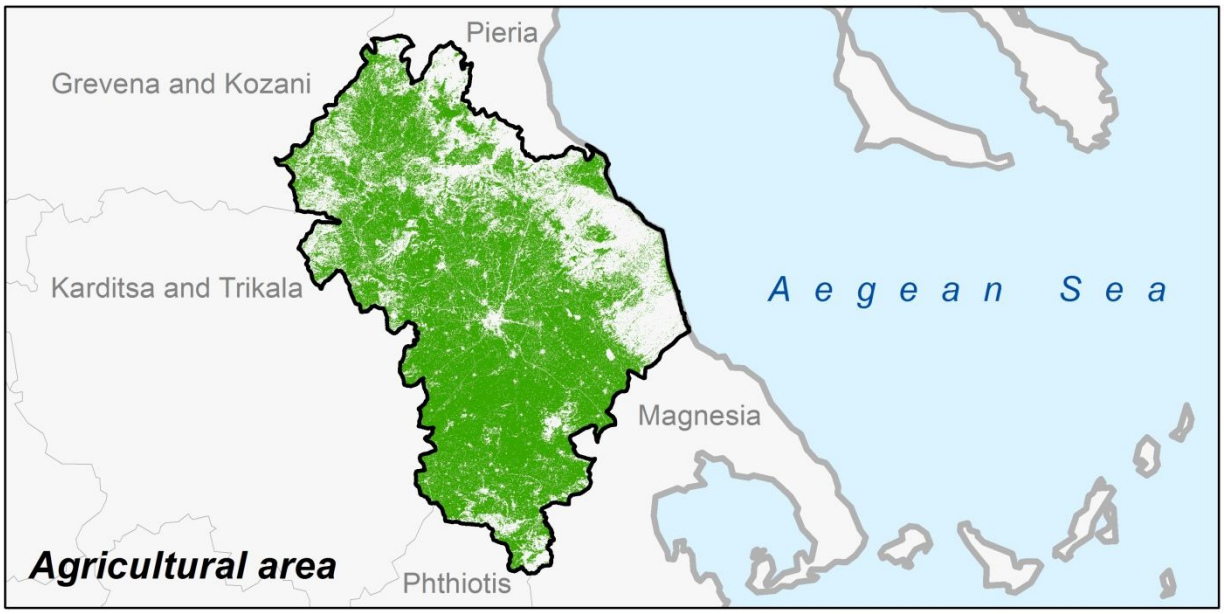


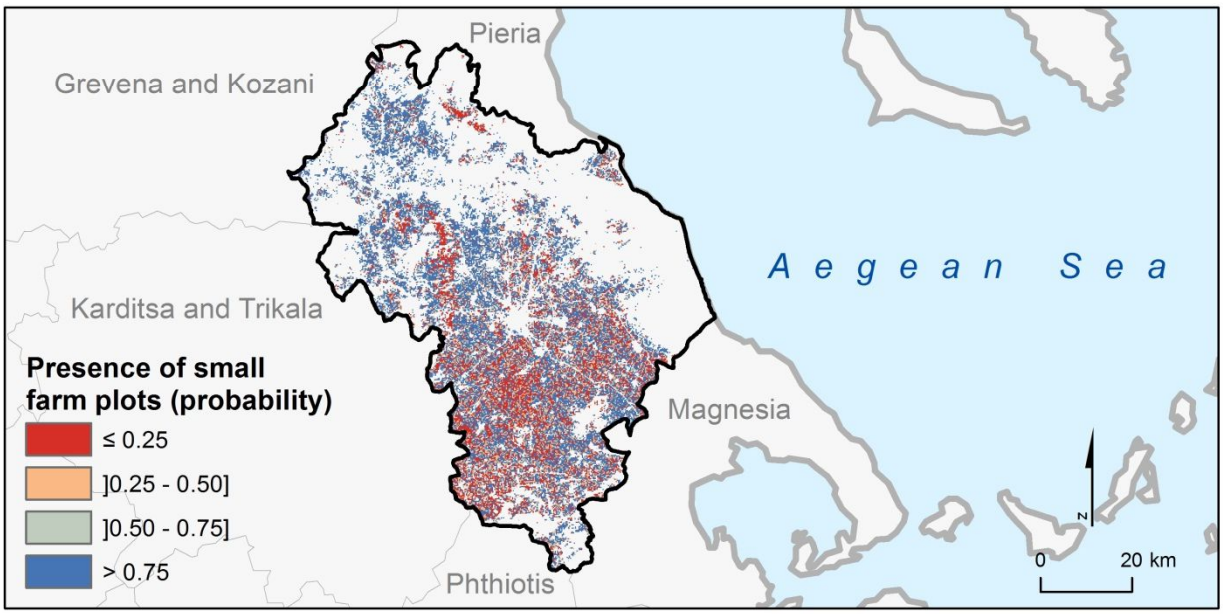


This project has received funding from the European Union's Horizon 2020
research and innovation programme under grant agreement No 677363

salsa
small farms
small food businesses and
sustainable food security

Reference region n.º 09
NUTS III: Larisa
NUTS II: Thessalia
Country: Greece

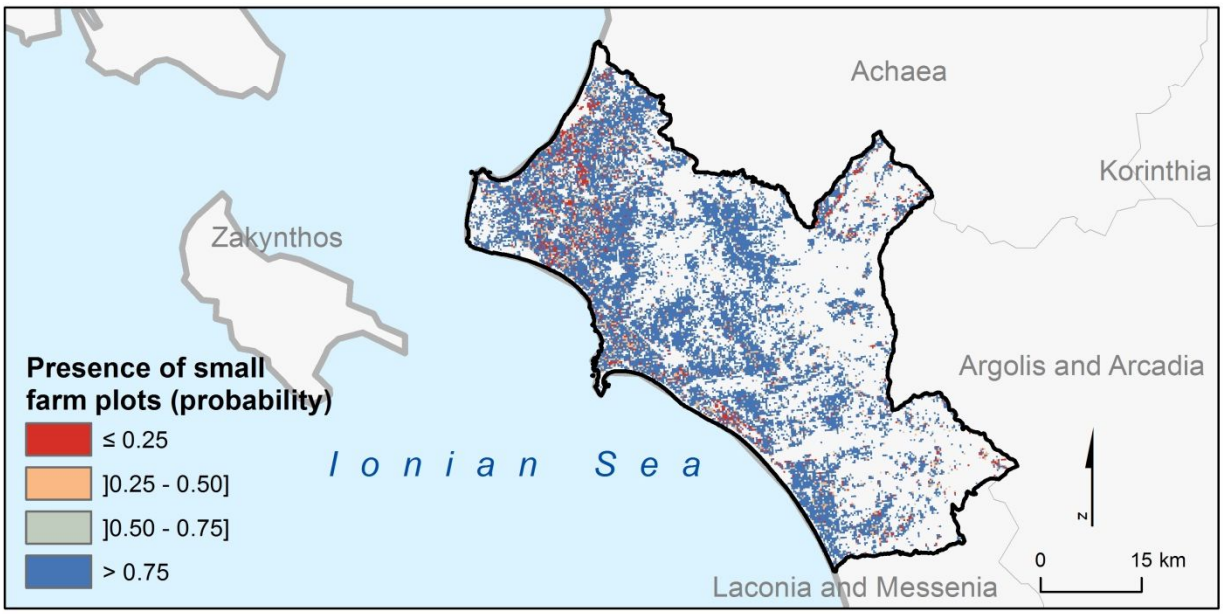




This project has received funding from the European Union's Horizon 2020
research and innovation programme under grant agreement No 677363

salsa
small farms
small food businesses and
sustainable food security

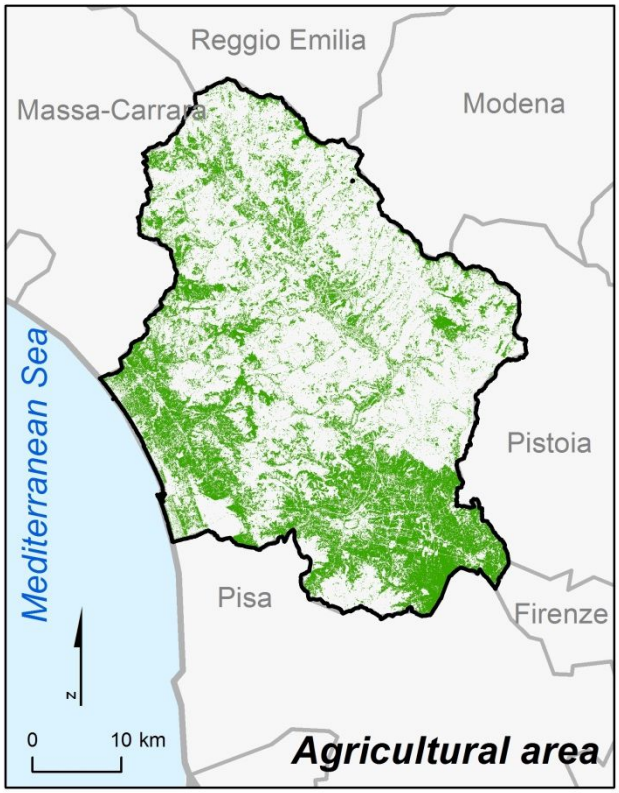
Reference region n.º 10
NUTS III: Ileia
NUTS II: Dytiki Ellada
Country: Greece

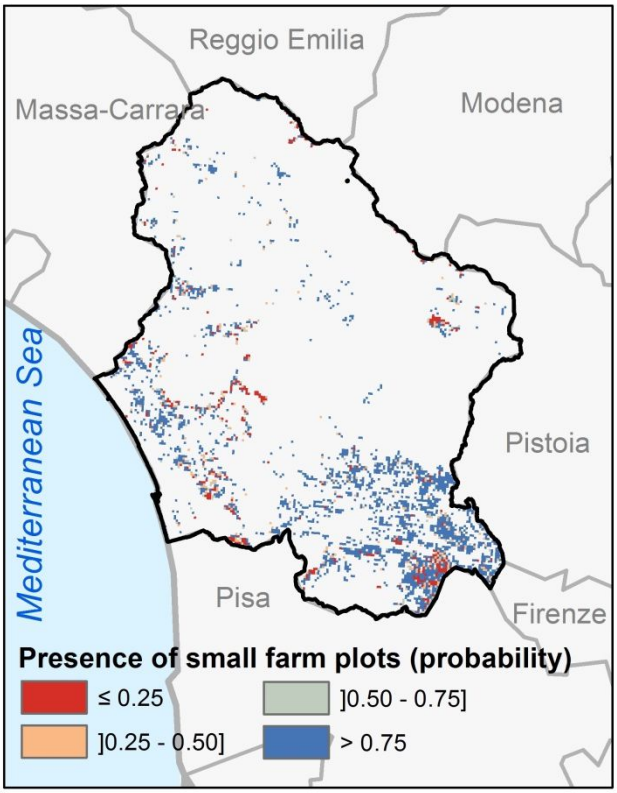


This project has received funding from the European Union's Horizon 2020
research and innovation programme under grant agreement No 677363

salsa
small farms
small food businesses and
sustainable food security

Reference region n.º 11
NUTS III: Lucca
NUTS II: Toscana
Country: Italy

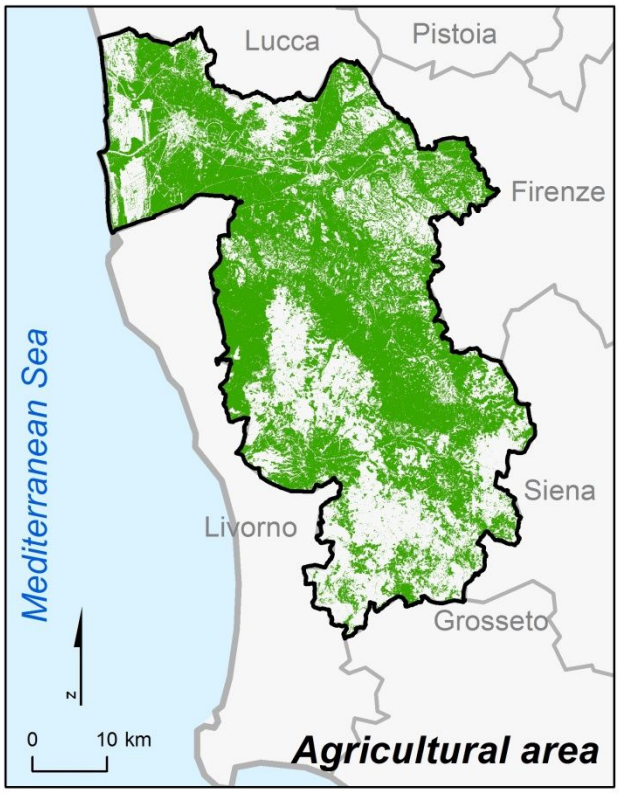


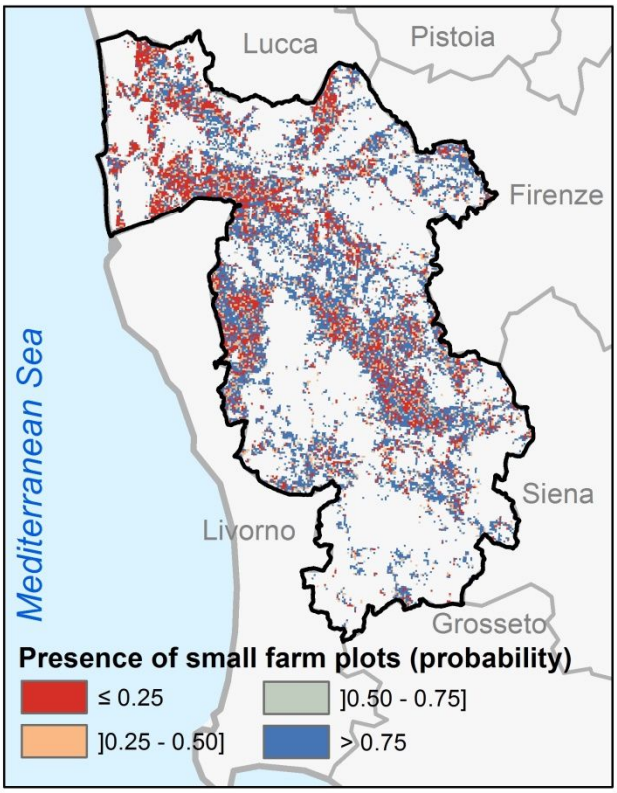


This project has received funding from the European Union's Horizon 2020
research and innovation programme under grant agreement No 677363

salsa
small farms
small food businesses and
sustainable food security

Reference region n.º 12
NUTS III: Pisa
NUTS II: Toscana
Country: Italy

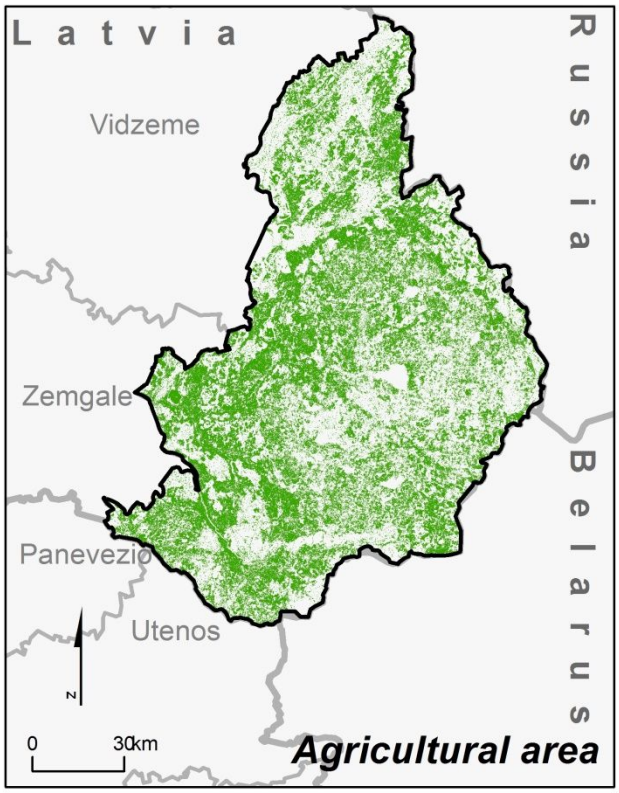


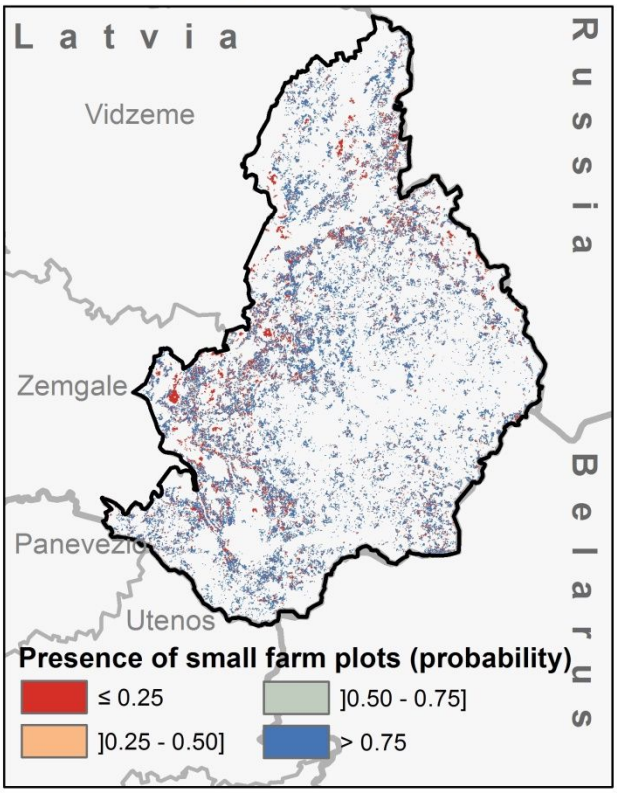


This project has received funding from the European Union's Horizon 2020
research and innovation programme under grant agreement No 677363

salsa
small farms
small food businesses and
sustainable food security

Reference region n.º 14
NUTS III: Latgale
NUTS II: Latvia
Country: Latvia

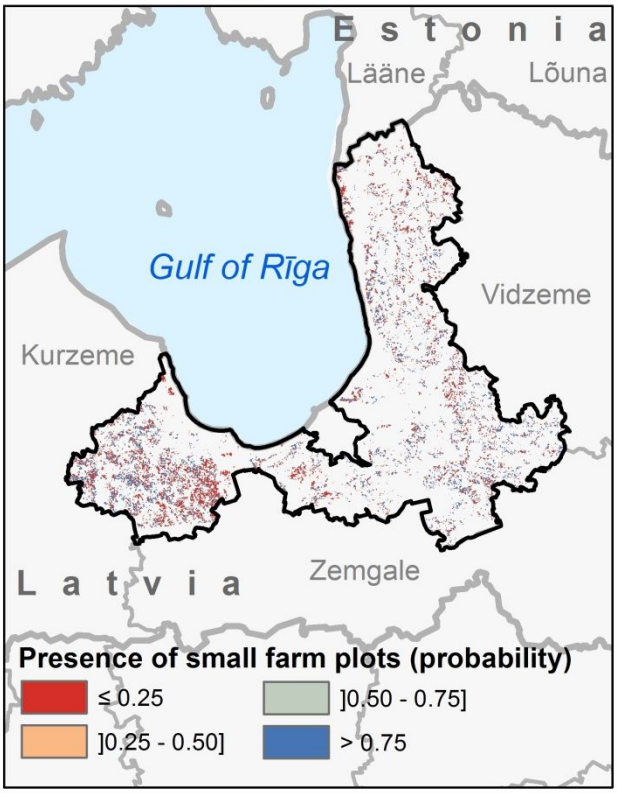




This project has received funding from the European Union's Horizon 2020
research and innovation programme under grant agreement No 677363

salsa
small farms
small food businesses and
sustainable food security

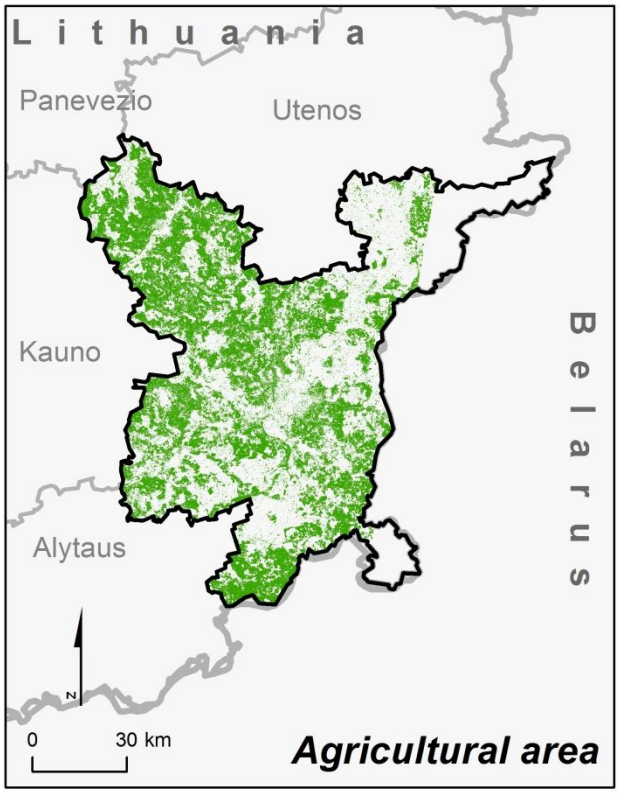
Reference region n.º 15
NUTS III: Pierīga
NUTS II: Latvia
Country: Latvia

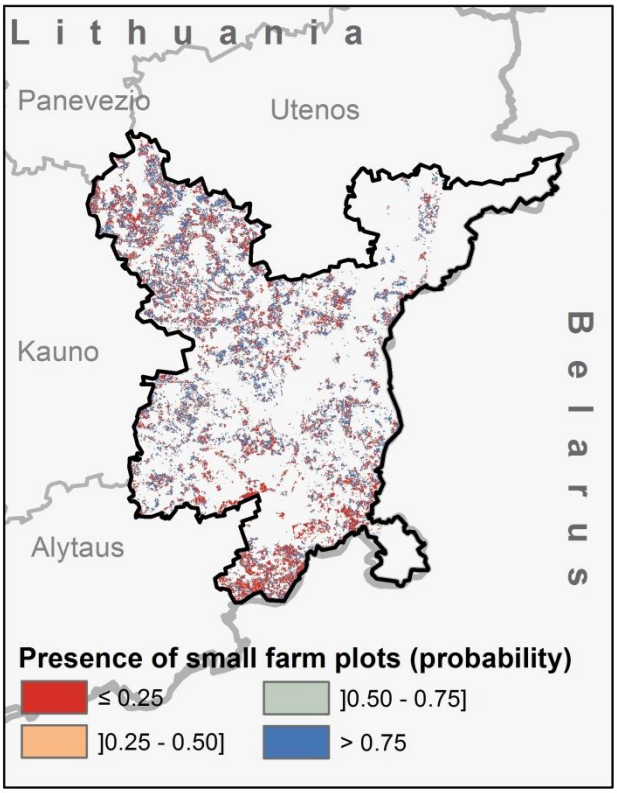


This project has received funding from the European Union's Horizon 2020
research and innovation programme under grant agreement No 677363

salsa
small farms
small food businesses and
sustainable food security

Reference region n.º 16
NUTS III: Vilniaus
NUTS II: Lithuania
Country: Lithuania

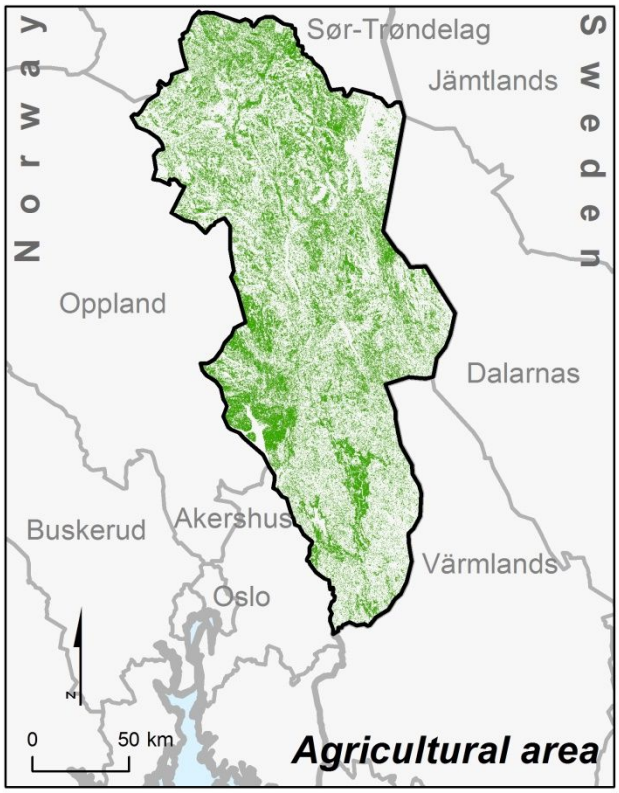


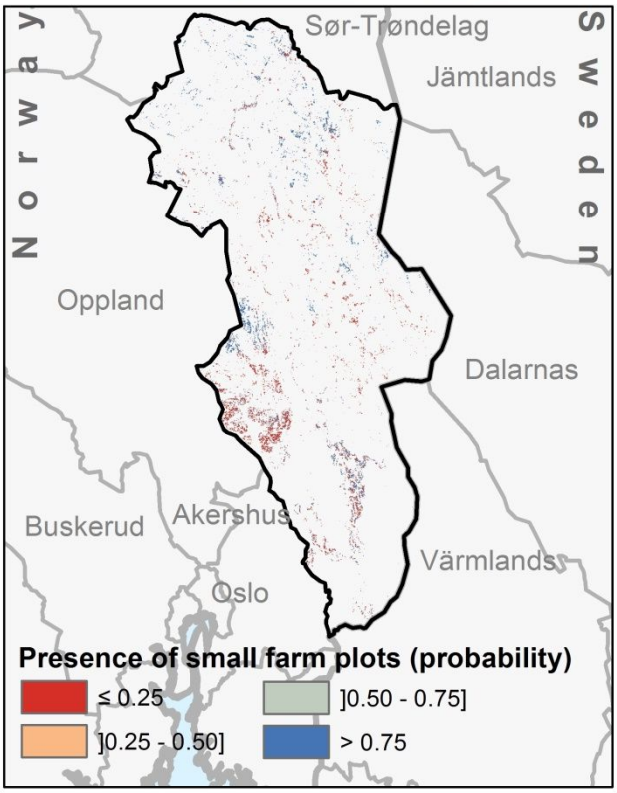


This project has received funding from the European Union's Horizon 2020
research and innovation programme under grant agreement No 677363

salsa
small farms
small food businesses and
sustainable food security

Reference region n.º 18
NUTS III: Hedmark
NUTS II: Hedmark og Oppland
Country: Norway

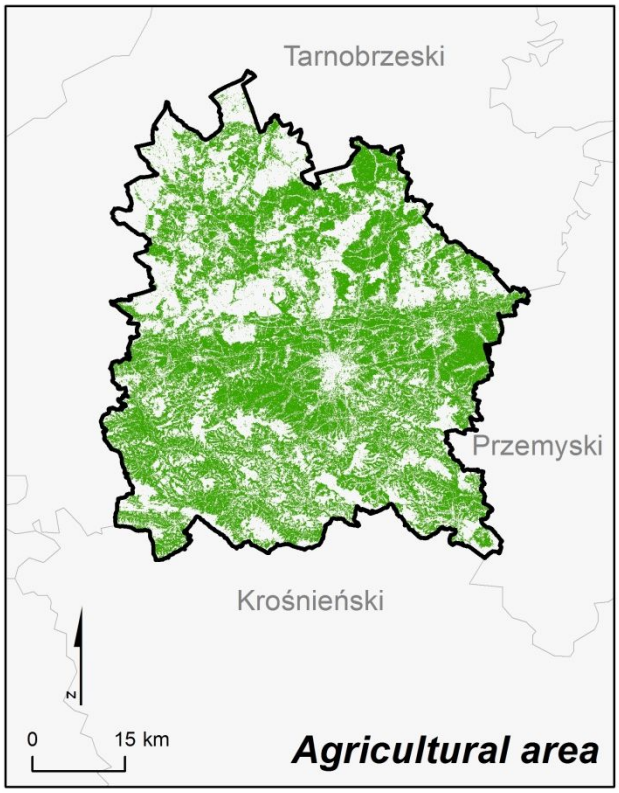


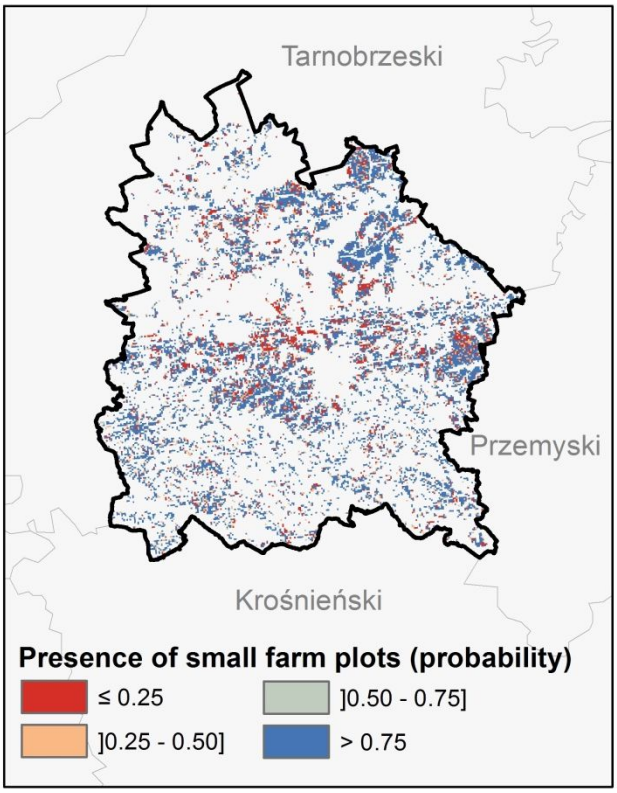


This project has received funding from the European Union's Horizon 2020
research and innovation programme under grant agreement No 677363

salsa
small farms
small food businesses and
sustainable food security

Reference region n.º 19
NUTS III: Rzeszowski
NUTS II: Podkarpackie
Country: Poland

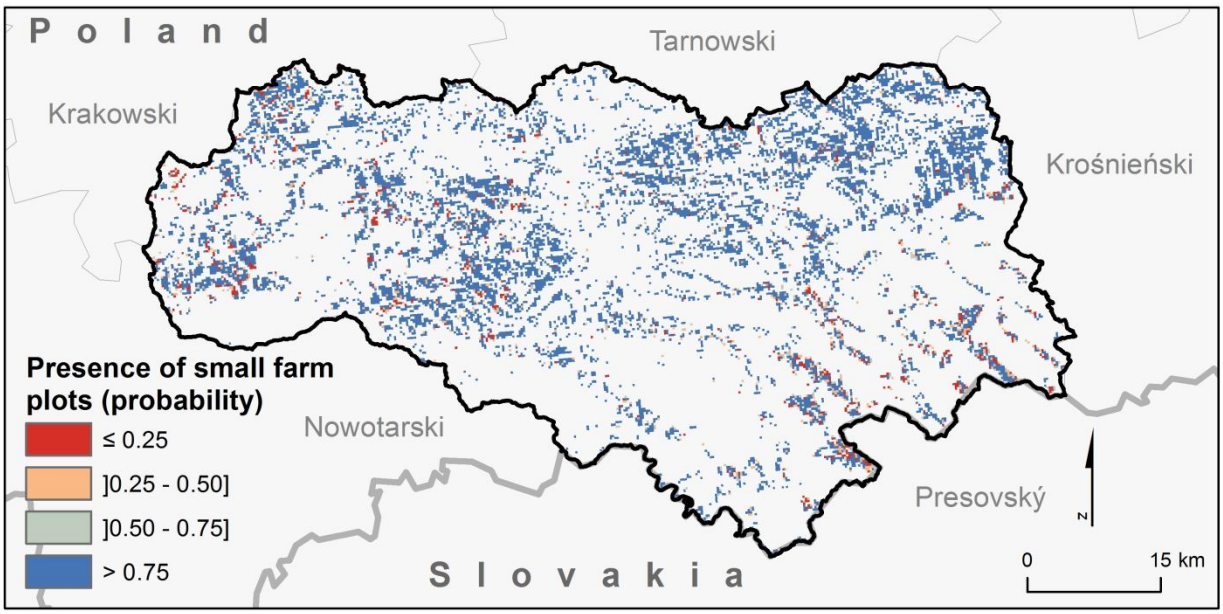




This project has received funding from the European Union's Horizon 2020
research and innovation programme under grant agreement No 677363

salsa
small farms
small food businesses and
sustainable food security

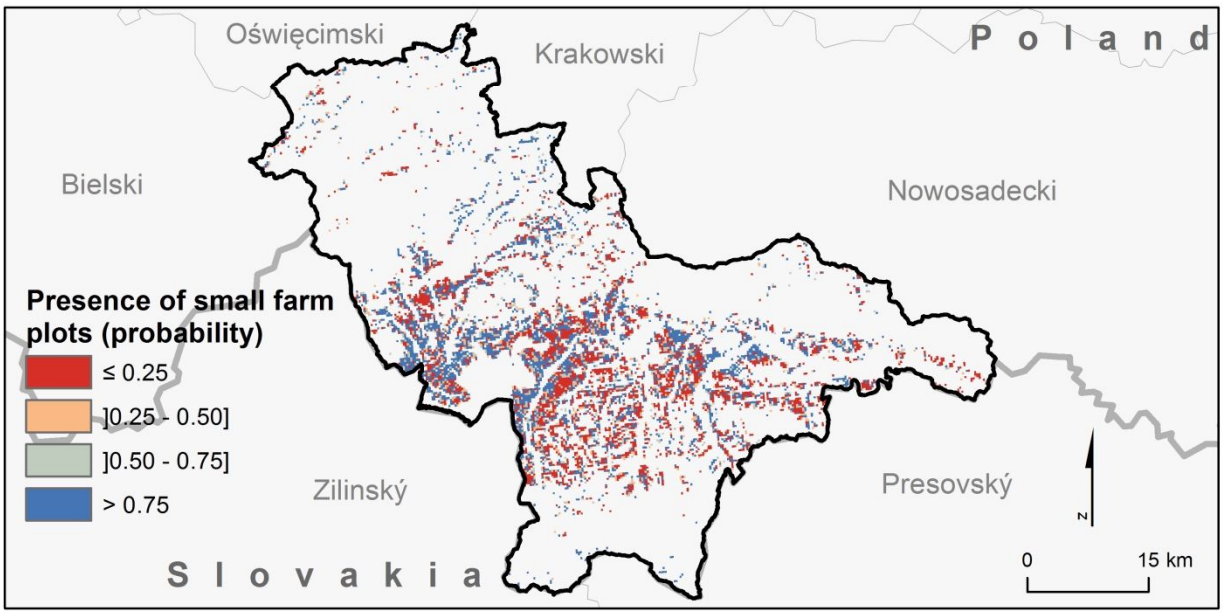
Reference region n.º 20
NUTS III: Nowosadecki
NUTS II: Małopolskie
Country: Poland



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 677363

salsa
small farms
small food businesses and
sustainable food security

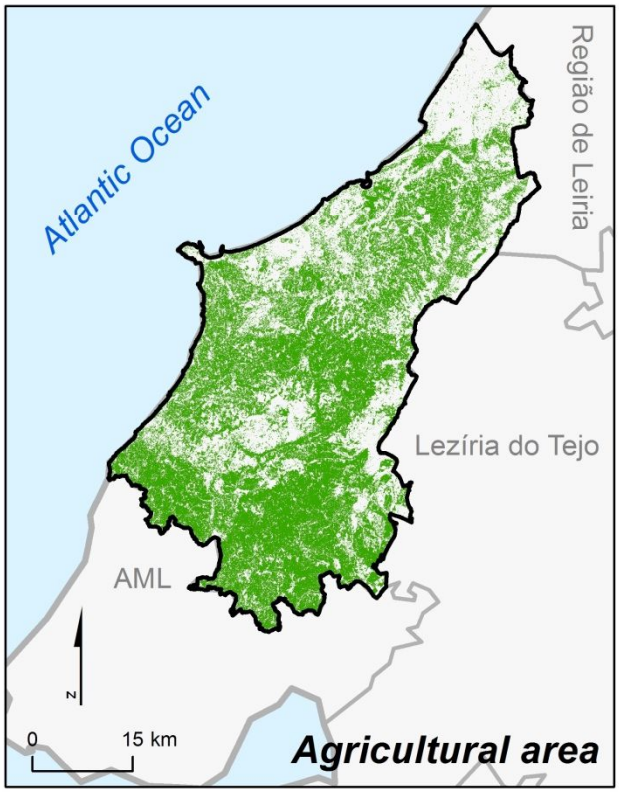
Reference region n.º 21
NUTS III: Nowotarski
NUTS II: Małopolskie
Country: Poland

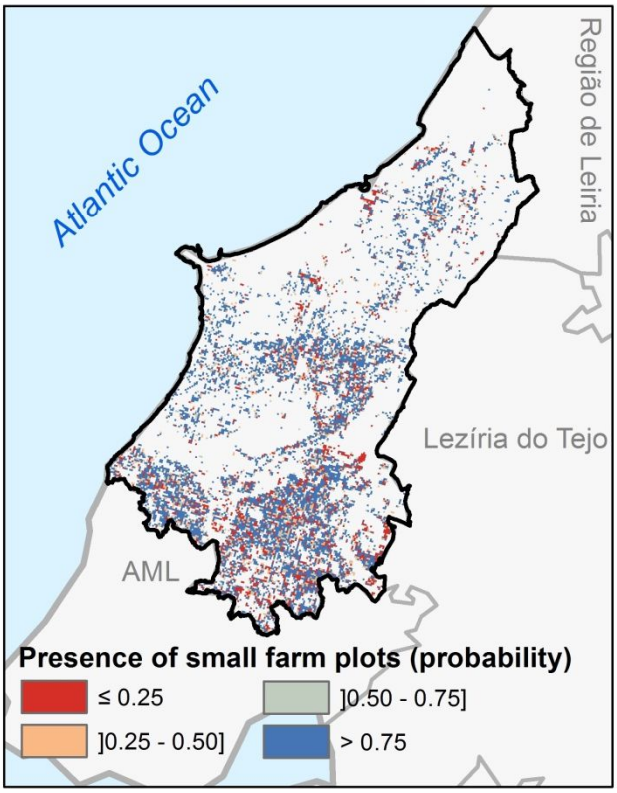


This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 677363

salsa
small farms
small food businesses and
sustainable food security

Reference region n.º 23
NUTS III: Oeste
NUTS II: Centro
Country: Portugal

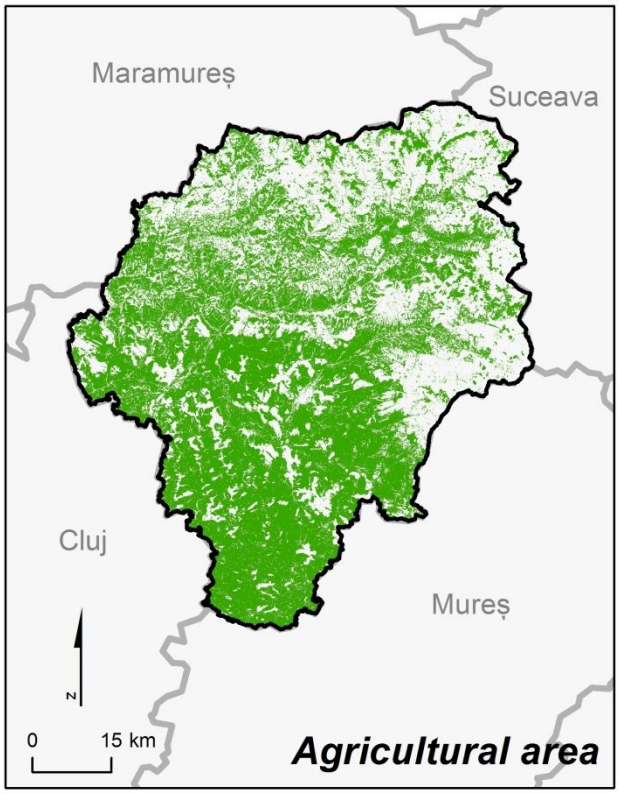


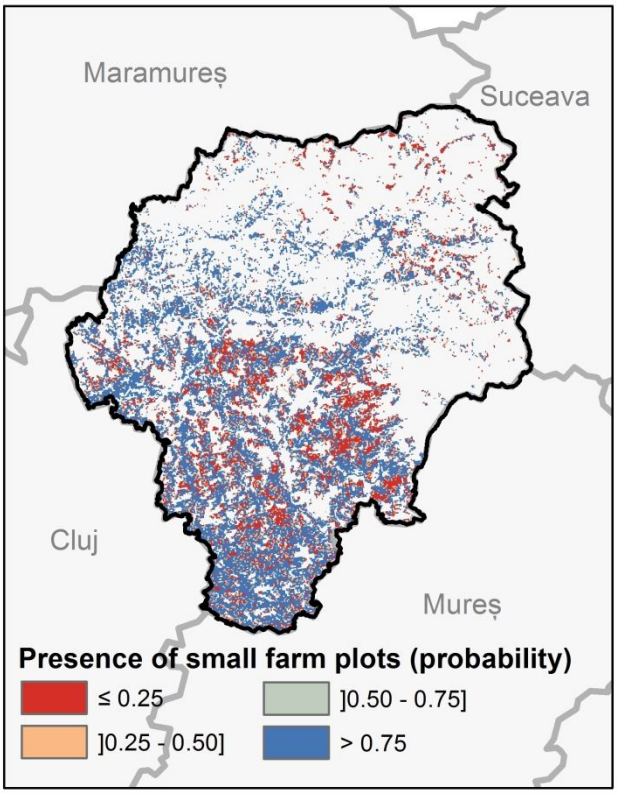


This project has received funding from the European Union's Horizon 2020
research and innovation programme under grant agreement No 677363

salsa
small farms
small food businesses and
sustainable food security

Reference region n.º 24
NUTS III: Bistrița-Năsăudi
NUTS II: Nord-Vest
Country: Romania

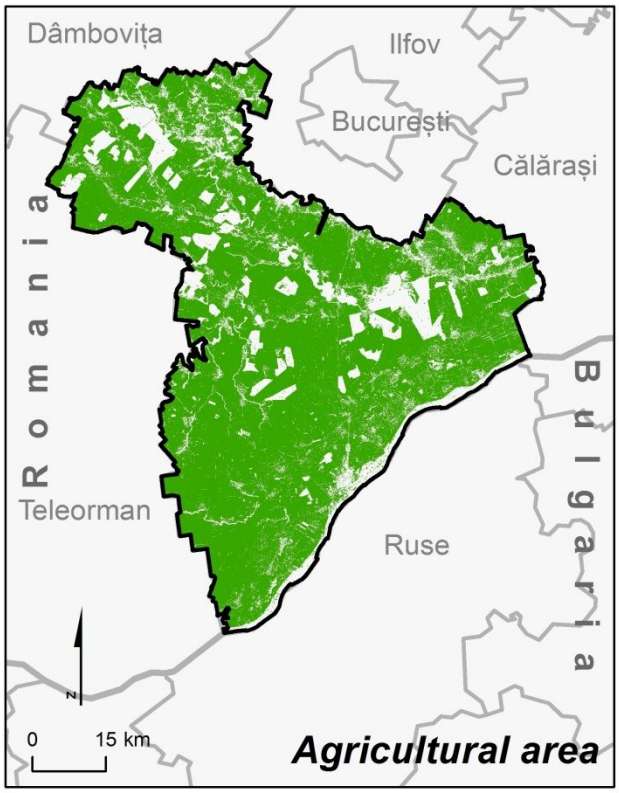


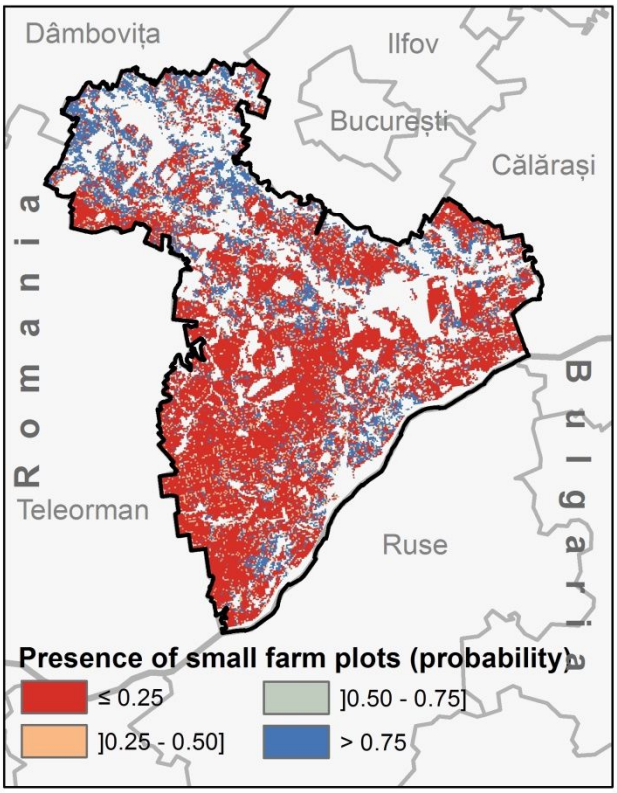


This project has received funding from the European Union's Horizon 2020
research and innovation programme under grant agreement No 677363

salsa
small farms
small food businesses and
sustainable food security

Reference region n.º 25
NUTS III: Giurgiu
NUTS II: Sud-Muntenia
Country: Romania

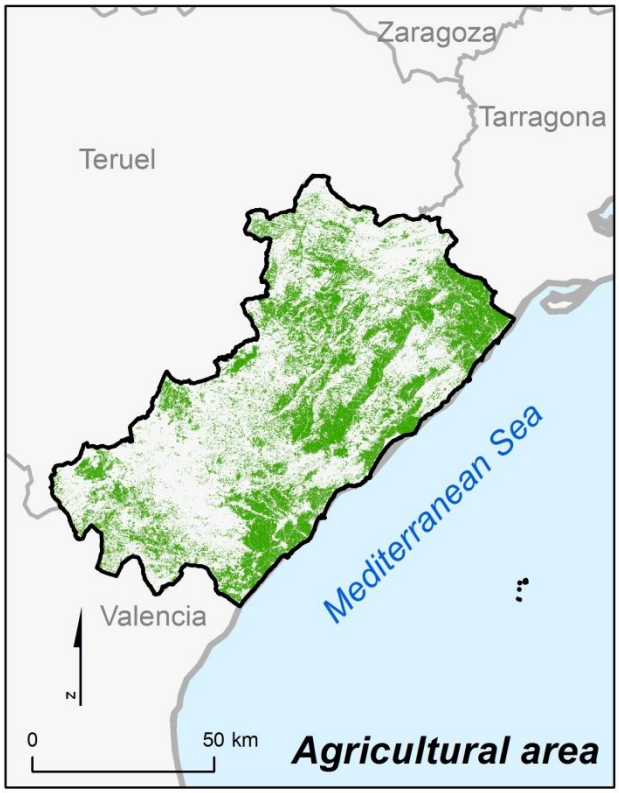


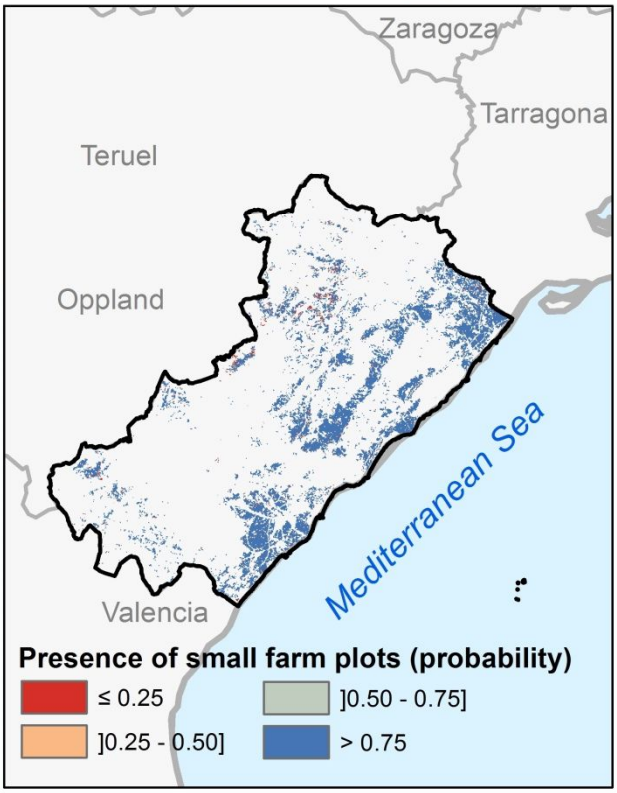


This project has received funding from the European Union's Horizon 2020
research and innovation programme under grant agreement No 677363

salsa
small farms
small food businesses and
sustainable food security

Reference region n.º 26
NUTS III: Castellón
NUTS II: Valencian Community
Country: Spain

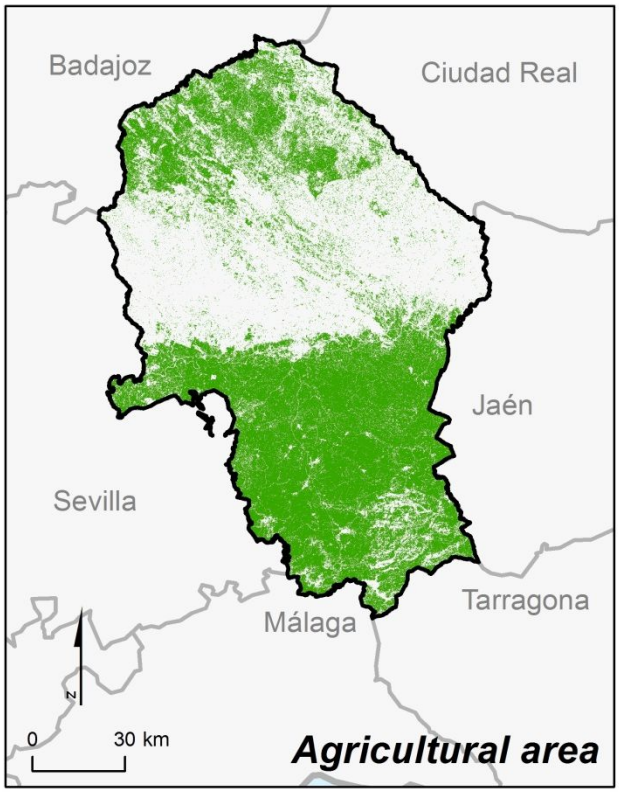


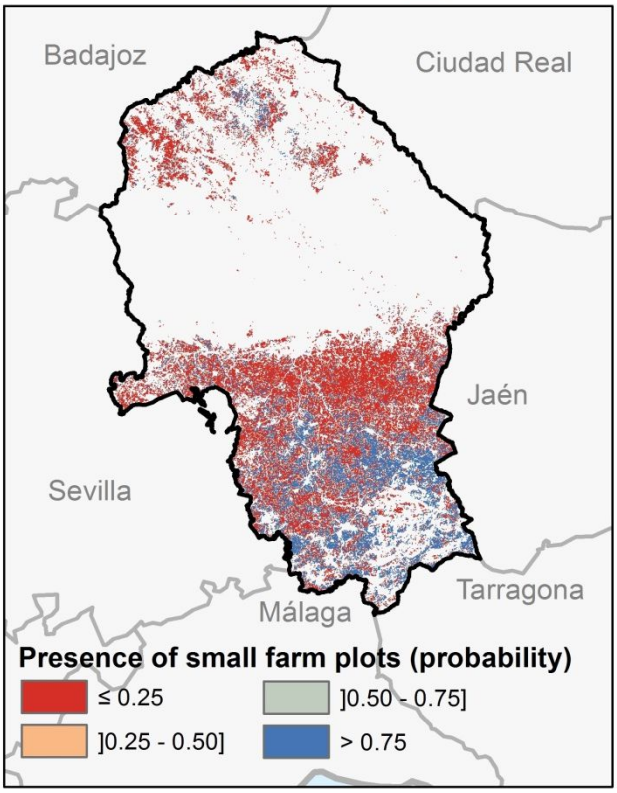


This project has received funding from the European Union's Horizon 2020
research and innovation programme under grant agreement No 677363

salsa
small farms
small food businesses and
sustainable food security

Reference region n.º 27
NUTS III: Córdoba
NUTS II: Andalusia
Country: Spain





This project has received funding from the European Union's Horizon 2020
research and innovation programme under grant agreement No 677363