Italian Journal of Agronomy 2023; volume 18:2199

Dynamics of agricultural land systems in western Mediterranean areas: a clustering approach based on the self-organizing map

Marya Cristina Rabelo,[1] Marj Tonini,[2] Nicola Silvestri[1]

[1]Department of Agriculture, Food and Environment, University of Pisa, Italy; [2]Institute of Earth Surface Dynamics, University of Lausanne, Géopolis Building UNIL Mouline, Lausanne, Switzerland

Highlights

- We propose an data driven procedure to extract useful information from a huge volume of multivariate agricultural census data (2000 and 2010).
- The resulting main clusters have been interpreted in terms of agricultural land systems characterizing the Mediterranean areas.
- Transitions reveal a reduction in agricultural land use, increase in utilized agricultural and irrigated area.
- The spatial distribution of agricultural land systems typologies in the geographical space has finally been mapped and and discussed.

Abstract

In the present study, we implemented an unsupervised learning procedure, a self-organizing map (SOM), for characterizing the main agricultural land systems (ALS) in western Mediterranean areas. Input data derived from national agricultural censuses of two periods (2000 and 2010) at the municipality level. The SOM allowed us to aggregate the items into clusters based on the proximity between the associated input variables. The main clusters were then mapped back to the geographical space and interpreted in terms of ASL typologies. The main ALS from the census 2000 included one permanent grassland system with extensive farming; two arable land systems, corresponding to winter and summer crops; and two permanent cropland systems, relatable to intensively cultivated or marginal areas. The ALS from the census 2010 included only one arable land system with a non-intensive use of irrigation; two permanent cropland systems similar to those found in 2000; one more extensive permanent grassland system; and a mixed system characterized by permanent grassland and arable land. In summary, the main trends emerging from the transitions between the two censuses periods were: i) a reduction in agricultural land use; ii) an increase in utilized agricultural and irrigated area; iii) a contraction in arable land and permanent grassland. Using a data-driven approach such as SOM allowed us to discover hidden patterns in the input census data. Therefore, the prevalent agricultural typologies characterising the ALS in the two analysed periods resulted to be shaped by the reality of the surveyed area solely, with regard to its agronomic assessment.

Correspondence: Nicola Silvestri, Department of Agriculture, Food and Environment, University of Pisa, via del Borghetto 80, 56124 Pisa, Italy.
E-mail: nicola.silvestri@unipi.it

Key words: unsupervised learning, agricultural censuses, land-use, land-cover, change analysis, clustering.

Contributions: the authors contributed equally.

Conflict of interest: the authors declare no potential conflict of interest.

Funding: none.

Availability of data and material: Data and materials are available by the authors.

Received: 14 February 2023.
Accepted: 10 September 2023.

Italian Journal of Agronomy 2023; 18:2199
doi:10.4081/ija.2023.2199

Introduction

Land systems are the result of interactions between humans and the natural environment (Verburg *et al.*, 2015), regarded as key tools for providing solutions to sustainable development and to global changes (Verburg *et al.*, 2013). In particular, agricultural land systems (ALS) represent agricultural systems with land as the core, involving all the humans' activities and outcomes of the agricultural land (Li *et al.*, 2023). Such ALS are dynamic and complex social systems whose core function is to ensure human livelihood and food security (Viana *et al.*, 2022b). During the last decades, the interest in ALS dynamics has been rising in Europe (Rega *et al.*, 2020) and especially in Mediterranean areas (Bajocco *et al.*, 2012; Debolini *et al.*, 2013; Marraccini *et al.*, 2015; Malek and Verburg, 2017; Tonini *et al.*, 2018). The capability of representing and interpreting changes in ALS is important to: i) identify the driving forces acting there (Bonet, 2004; Sluiter and De Jong, 2007; Levers *et al.*, 2018); ii) assess changes in land degradation and ecosystem services provisioning (Schröter *et al.*, 2005; Symeonakis *et al.*, 2007; Young *et al.*, 2007; Debolini *et al.*, 2013; DrakeDrake NA, 2014); iii) evaluate the effects of the Common Agriculture Policy (CAP) (van Vliet *et al.*, 2015; Huber *et al.*, 2018; Rega *et al.*, 2020).

Analyses of changes in European ALS revealed two main trends (Silvestri *et al.*, 2012; Marraccini *et al.*, 2015; Levers *et al.*, 2018). Firstly, we could observe a reduction in agricultural land use with a contraction of cultivated land under the action of different driving forces such as urbanization, farmland abandonment, and afforestation. Secondly, we assisted to a general trend towards the adoption of intensive farming practices, as demonstrated by the biodiversity loss in different agroecosystems and by the environmental contamination due to agrochemical use. These two dynamics involve agricultural districts characterized by different pedoclimatic and agronomic conditions, but also by distinctive population settlement patterns (Bajocco *et al.*, 2012). For instance, agriculture declined in hilly lands and in internal marginal areas (inland depopulation), while cultivated lands in the coastal and suburban areas tended to increase their productivity to satisfy a growing demand for food (littoralisation and megalopolis). All these changes often coexist within quite large areas and determine a complex set of patterns and trajectories that can be better understood by assuming some simplification in ALS definition (Marraccini *et al*. 2015; Debolini *et al.*, 2018).

ALS are complex systems resulting from the action of many different scale-dependent factors that interact in space and in time determining the conditions of land cover (LC) and land use (LU) detectable over cultivated lands (van Asselen and Verburg, 2012; Levers *et al.*, 2018; Viana *et al.*, 2022). Specifically, LC refers to the surface cover on the ground (crops, natural vegetation, hydrography, and human structures) and can be detected and mapped by using different data sources (remote sensing, aerial imagery, census data, field scouting) (Jansen and Gregorio, 2002). Conversely, LU concerns the purposes pursued by humans in the exploitation of land (Lambin *et al.*, 2006) and its detection involves both baseline mapping and subsequent surveys. The spatial pattern and the intensity of the LU are the two poles on which the analysis of ALS is based at a broad-scale level (Verburg and Overmars, 2009; Lambin and Meyfroidt, 2011), allowing discrimination between different cropping production models.

In the agricultural context, LU evaluation should also include the intensity of farming practices (Verburg and Overmars, 2009), defined as a multidimensional factor involving three different aspects: i) use of agricultural inputs (tillage, fertilisation, irrigation); ii) outputs produced (yields); and iii) changes in land system properties (non-marketed ecosystem services) (Erb *et al.*, 2013; Václavík *et al.*, 2013). In the absence of this information, attempts to integrate different data sources can produce inaccuracies due to their inconsistencies (Malek and Verburg, 2017). Therefore, although several authors underlined the importance of including LU intensity in ALS evaluation (Erb *et al.*, 2013; Murray-Rust *et al.*, 2014; Jepsen *et al.*, 2015; Estel *et al.*, 2016), many studies simply consider the LC changes as a valid proxy (Hill *et al.*, 2008; Hansen *et al.*, 2013; Stellmes *et al.*, 2013; Estel *et al.*, 2015; Kuemmerle *et al.*, 2015). Moreover, the few existing papers that agree to adopt a holistic approach, are often restricted in terms of the spatial (Estel *et al.*, 2015) or temporal dimensions (van Asselen and Verburg, 2012; Václavík *et al.*, 2013; Levers *et al.*, 2018).

To analyze land system dynamics, it is thus necessary to reduce the complexity of the input information to a few typologies easier to manage. The definition of these typologies can be accomplished by using techniques of clustering able to group similar observations into the same cluster and to maximise the among-clusters dissimilarity (van der Zanden *et al.*, 2016). The cluster analysis methods are based on two types of approaches: bottom-up (*e.g.*, hierarchical agglomerative clustering) and top-down (*e.g.*, hierarchical divisive clustering). In both cases, a threshold, which is normally based on expert evaluations, needs to be chosen to cut the dendrogram and fix the final number of clusters. Among the most popular data-driven clustering techniques, K-means is a simple and fast algorithm aiming to partition data points into K clusters by minimizing the sum of squared distances between each item and its assigned centroid (MacQueen, 1967; Lloyd, 1982). The drawback of K-means clustering is that it is highly sensitive to the initial seed selection of the cluster centers (Khan, 2012).

In a recent study, a new typology was defined to represent the composition, spatial structure, and management intensity of agricultural landscapes in Europe, based on a data driven algorithm relying on machine learning (van der Zanden *et al.*, 2016). Although these sophisticated methodological approaches have been available for some years, their use in ALS analysis is still scarce. Machine learning algorithms (such as artificial neural networks, support vector machine, and decision trees) allow researchers to overcome some constraints posed by the use of more classical deterministic approaches to discover pattern in environmental datasets (Kanevski *et al.*, 2009). In particular, the self-organizing map (SOM) (Kohonen, 2001) performs well in identifying clusters from a multivariate dataset based only on the distance between the input variables. The algorithm SOM is designed as an unsupervised competitive learning neural network allowing the representation of a high-dimensional dataset as a two-dimensional discretized pattern. Overall, SOM is very efficient in terms of data visualization (Vesanto, 1999) as it can provide additional output-maps to support the interpretation of the results.

To detect ALS structures and explore the complex interactions between the variables involved, new tools need to be adopted. To this end, in the present study, we implemented a data-driven procedure based on SOM followed by hierarchical clustering to aggregate the mapped units into a lower number of well-defined pattern structures, representing the main clusters. Input data came from the official national agricultural censuses of Portugal, Spain, France, and Italy, and all information used was referred to the municipality level. The obtained clusters have been interpreted in terms of ALS characterizing the Western Mediterranean areas over two census periods (2000 and 2010). Finally, the changes that occurred between the two investigated periods have been analyzed and discussed in terms of emerging trend-lines in response to driving forces acting at large scale (*e.g.*, CAP, human migration, climate changes).

Materials and Methods

The study domain

In the Mediterranean area, environmental (soil nature, orography, climate), biological (high level of plant biodiversity and suitable crops) and historical (traditional food production, eating habits) conditions interact with each other to determine a large variety of agricultural systems with a fast-changing dynamics (Debolini *et al.*, 2018). In the present work, the studied domain has been limited to the four western Mediterranean countries that have similar socio-economic conditions: Portugal, Spain, France, and Italy (Abu Hammad and Tumeizi, 2012). In this way, we aim to detect and characterize ALS shaped mainly by biophysical factors, and to stem the effect of heterogeneous anthropogenic factors.

The boundary of the study area was selected according to the Natura 2000-Biogeographical Mediterranean Region, and included almost the entire territory of Portugal (95% of the national surface) and Spain (90%), about half of Italy (56%), and a small portion of France (11%) (Figure 1). The total surveyed area amounted to 78.151 million of hectares and included about 80 million inhabitants (just over 40% of the total population of the four countries). We restricted the analysis to the two last agriculture censuses available at the time, which are 1999 and 2009 for Spain and Portugal, 2000 and 2010 for France and Italy (hereinafter all referred to as "census 2000" and "census 2010").

The agricultural census datasets

The variables analyzed in the present study (Table 1) come from the national agricultural censuses datasets. All data used are referred to the lowest level of local administrative unit (LAU2, according to the Eurostat classification) corresponding to the municipality. By this way, we intended to preserve, as far as possible, the variety of the input information and the regional characterization of the entire area.

Three of the selected agronomic variables are closely related to the land use [total agricultural area (TAA), utilized agricultural area (UAA), and irrigated area (IA)], while the other three are related to the land cover [arable lands (AL), permanent crops (PC), and permanent grassland, (PG)]. All these variables are in accordance with the definitions given by the Eurostat Agriculture Glossary (Eurostat, 2023).

The final database consisted of 16,580 records (total number of municipalities) and 6 fields (corresponding to the six variables considered). Its consistency was checked by verifying the fulfilment of a series of irrefutable conditions such as: TAA≥UAA, UAA≥IA, UAA≥(AL+PC+PG).

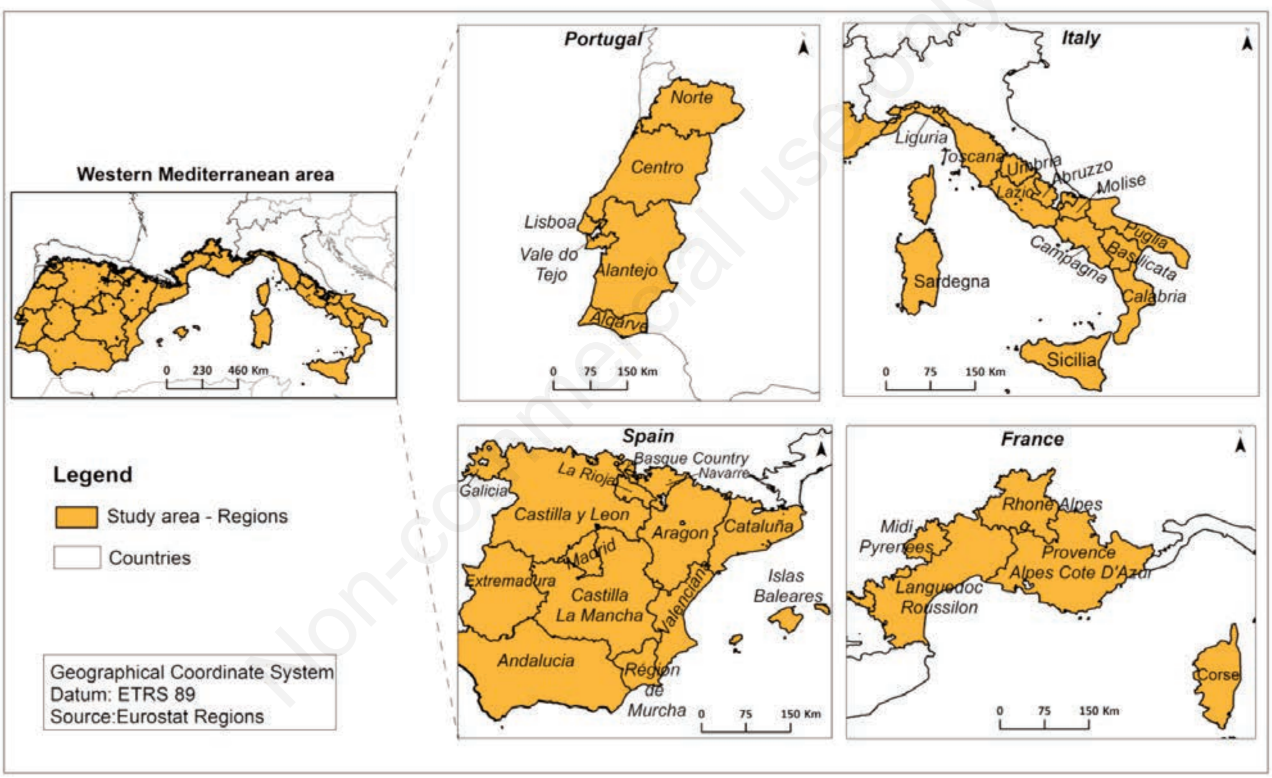

Figure 1. The western Mediterranean areas analysed in the present study (Eurostat).

Table 1. The variables used in the present study. "$Code_0$" indicates the original values coming from national agricultural census data; "Code" indicates the values after normalization.

Variable	$Code_0$	Normalization	Code
Total area of municipality	TA_0	-	-
Total agricultural area	TAA_0	TAA_0/TA_0	TAA
Utilized agricultural area	UAA_0	UAA_0/TAA_0	UAA
Irrigated area	IA_0	IA_0/UAA_0	IA
Arable land	AL_0	AL_0/UAA_0	AL
Permanent crops	PC_0	PC_0/UAA_0	PC
Permanent grassland	PG_0	PG_0/UAA_0	PG

Discrepancies in the size of the municipalities for the four countries could affect the homogeneity of ground resolution in the analysis, but these differences have been considered acceptable for the aim the present study (Rabelo *et al.*, 2021). The agronomic variables were normalized to the range [0, 1] by dividing their original value by the value of the variable they belong to (Table 1). Finally, as no homogeneous information was available from the national agricultural census regarding the use of agricultural inputs, the outputs produced, or changes in land system properties, we considered the first three variables (TAA, UAA, and IA) as a proxy for the evaluation of the LU intensity. Indeed, these variables provide information about the expansion reached by agricultural land use (TAA), the rate of exploitation of farmland (UAA), and the input level used (IA). The main descriptive statistics calculated for the six variables are reported in supplementary materials (*Supplementary Tables 1 and 2*). Further details on the database structure can be find in two previous works (Villani *et al.*, 2019; Rabelo *et al.*, 2021).

Data analysis: two-step clustering

A SOM (Kohonen, 1982) is an unsupervised machine learning algorithm based on an artificial neural network. Self-organization is defined as the process of partitioning the output layer neurons to correspond to individual patterns or clusters, based on the heterogeneous observations in the input layers (McCulloch and Pitts, 1943). In other words, a self-organizing network is a competitive learning neural network with an input layer of observations and an output layer of neurons (called units) resulting from non-linear functions.

In SOM, the proximity of the output clusters reflects the similarity of the corresponding input observations. The main idea is to project the high-dimensional input space, where each dimension corresponds to an input variable, in a lower-dimensional output space. The output space is organized on a grid made up of regular units with a fixed number of neighborhoods: four in a rectangular grid or six in a hexagonal grid. Each unit contains a vector of weights of the same dimension as the input vectors. At the end of the process, each observation from the input space is associated with a unit of the SOM grid. Depending on the size of the grid, one unit can include one or more input observations. An important characteristic of SOM is the preservation of the topological structure (*i.e.*, the preservation of the proximity between the input observations). Finally, hierarchical clustering can be performed to isolate groups of input vectors characterized by similar values of the input variables. This second step takes the SOM-units as input and group them into the desired number of clusters that, in our case, will finally represent the ALS typologies derived from the census data analysis.

Self-organizing maps

The training process associated with SOM consists of several steps and many iterations (Kohonen, 1982). First, a so-called codebook vector (W_i) is assigned to every unit of the SOM grid, randomly initializing the unit's weights ($w_{i,j}$). Then an input vector (I_i) is chosen at random from the input dataset and its distance to every unit is computed to evaluate the closest one, defined as the "best-matching unit" (BMU). The BMU's neighboring units included within a certain distance radius are then selected. The radius starts large and diminishes at each iteration, so that in the end only the BMU are selected. The topological structure of the SOM grid is preserved thanks to the fact that BMU's neighbouring units result in similar codebook vectors. This is achieved by updating in the same way the BMU together with its neighbouring units, as follows:

$$W_{(t+1)} = W_t + \alpha\,(I - W_t) \tag{1}$$

This means that the updated weighted vector ($W_{(t+1)}$) for a given output unit gains at each iteration α-times the value of the difference between the input vector (I) and the previous weight vector (W_t), where the learning rate α indicates the amount of change. These training's steps are repeated for several iterations. The number of iterations must be of sufficient size to ensure that the SOM grid eventually settles into a map of stable zones. At the end of the SOM process, each unit of the grid is associated to a codebook vector, representing the local means of the input vectors. Similar observations end up being mapped close together.

Clustering outputs

The main SOM outputs, which are generally mapped for visualization purposes, are the node counts, the neighborhood distances, and the heatmaps (Wehrens and Buydens, 2007). The node counts map informs about the number of input vectors falling inside each output unit. The neighborhood distance plot shows the distance between each unit and its neighborhoods. The heatmaps show the distribution of each input variable, associated with each input vector, across the SOM grid; heatmaps are a useful tool to visually explore the relationship between the input variables.

In the second clustering step, the codebooks resulting from SOM can be grouped using a hierarchical clustering method to form the final partitioning, where similar units are aggregated into single clusters. The process starts by assigning each observation to a single cluster; the most similar clusters are then joined together based on the Euclidean distance between them. This step is repeated until a pre-fixed number of clusters is reached. At the end of the entire process, including SOM and hierarchical clustering, clusters were assigned back to the single observations in the original dataset and mapped using a geographical information system (GIS), allowing the detection and visualization of the main clusters over the geographic space.

Quality assessment and parameters

The quality of the SOM grid can be assessed as the mean distance value of each input vector, mapped to a particular unit, to the codebook vector of that unit: the smaller the distances, the better the input data are represented by the codebook vectors. The average value over the entire grid, called quantization error (QE), is considered an indicator of the overall quality of the SOM grid. SOM's parameters were estimated via a trial-and-error process, seeking to minimize the QE. These resulted in the following values: 500 iterations; α declining linearly from 0.05 to 0.01; starting radius covering 2/3 of all unit-to-unit distances; size of the SOM grid equals to 50 by 30 (horizontal / vertical) units with hexagonal topology. The number of the final main clusters was set to five, established by examining the plot of the within cluster sum of squares obtained using k-means. This choice is supported by the data, including six variables, and by the aim of the present study, that is to discover clusters characterized by similar agricultural systems in terms of variable distribution.

Computations were performed using the R free software environment; SOM was performed using the Kohonen package for supervised and unsupervised SOM (Wehrens and Kruisselbrink, 2018). The GIS software ArcMap (v. 10.8 - ESRI) was used to elaborate the final maps.

Quantitative assessment of the main clusters

To characterize the five final clusters in terms of ALS, and verify their consistency at the global level, we evaluated the distribution of each variable within the revealed clusters using box plots. These graphics are routinely used to display a dataset based on the five-number summary statistics: minimum, maximum, sample median, first quartile (which splits off the lowest 25% of data from the highest 75%), and third quartiles (which splits off the highest 25% of data from the lowest 75%). The statistical descriptive values summarised are available in *Supplementary Tables 3 and 4*.

In addition, for every cluster, its centroid was computed as the average of the values of each unit (*i.e.*, the municipalities) belonging to that cluster for each of the six variables. The dispersion of the units belonging to each cluster was calculated as an average of the Euclidean distances of each municipality from its centroid [within distance (WD)]. Moreover, we evaluated the Euclidean distance between clusters detected in the two census periods to quantify the changes that had occurred during the decade time-span 2000-2010 [decade distance (DD)].

Results and Discussion

Self-organizing map outputs and clustering detection

Several parameters and configurations of the SOM grid were implemented and compared. The grid size of 50 by 30 units with hexagonal topology gave the smallest QE, equals to 0.010 and 0.012 for the 2000 and 2010 censuses, respectively. The resulting neighborhood distance maps are reported in *Supplementary Figure 1*. The node counts maps, which plot the number of input vectors falling inside each output unit, were relatively uniform for both periods, with most of the units including between 5 and 20 observations and no empty units (*Supplementary Figure 2*). This indicates that the size of the SOM grid allowed a proper representation of the input data.

The heatmaps are the most meaningful visualization tool of SOM. A heatmap displays the pattern distribution of each variable: how it is distributed along the SOM grid and how values change in space. Visualized side by side, heatmaps show a picture of the different areas and their characteristics. Indeed, the position of the individual units in the SOM grid is the same for all the heatmaps arising from the same input dataset; what changes is the represented variable. This way, it is possible to explore the level of correlation that links one or more variables. Heatmaps for both the census periods (Figures 2 and 3) showed a negative correlation in the distribution of the LC-related variables: AL, PC, and PG. Less strict, but still noticeable, were the positive correlations between some LU-related variables, such as the TAA and the UAA, or between the UAA and the IA. A negative correlation can be observed between the IA and the PG.

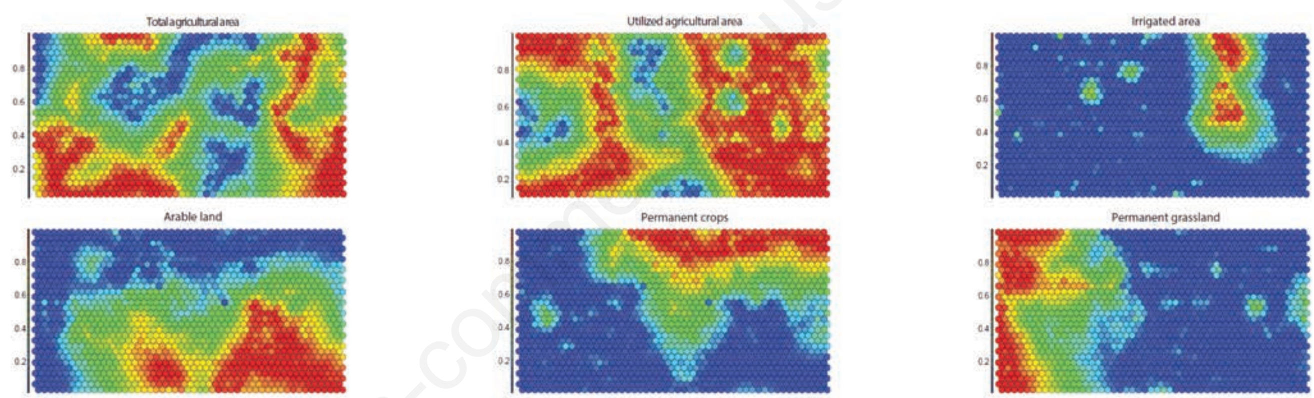

Figure 2. Heatmaps showing the pattern distribution of the agronomic variables within the self-organizing map grid for census 2000. The gradient color scale represents the values of the normalized input variables (blue=low values, green=medium values, red=high values).

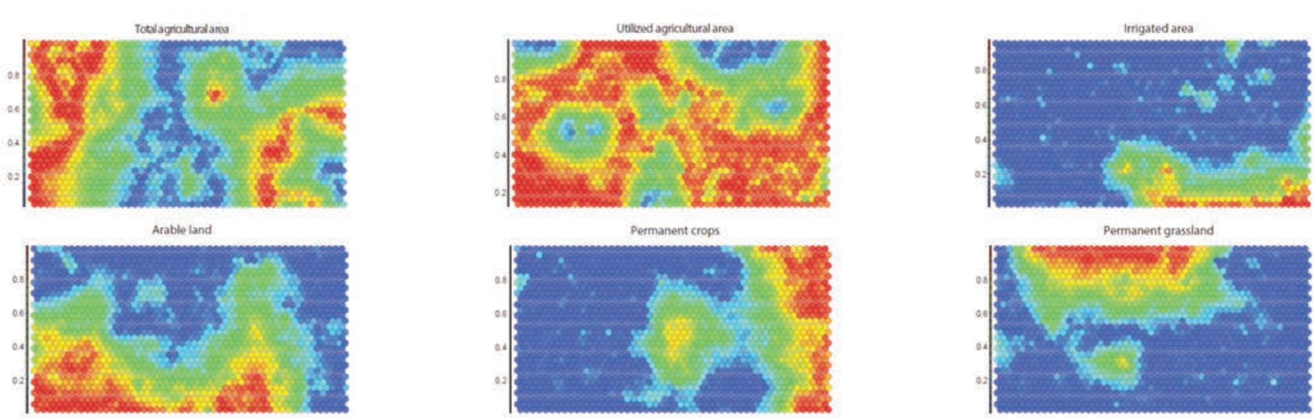

Figure 3. Heatmaps showing the pattern distribution of the agronomic variables within the self-organizing map grid for census 2010. The gradient color scale represents the values of the normalized input variables (blue=low values, green=medium values, red=high values).

The codebook map seeks to visualize all the variables on the same diagram. For a high-dimensional space, especially with many units as it is the case in the present study, this representation is quite unsuitable. Nevertheless, the codebook map turned out to be useful in exploring the uniformity of the final main clusters performed using hierarchical clustering to aggregate together similar SOM units in the second step of the analysis (*Supplementary Figures 3 and 4*).

The final five clusters detected for each census period are referred to as Cn, with n ranging from 1 to 5 (C1, C2, ...) for the census 2000, and from 6 to 10 (C6, C7, …) for the census 2010. It is worth specifying that SOM was computed independently for the two censuses. Comparison between the clusters was made *a posteriori*, highlighting similarities and differences that naturally emerged between them. The main structures revealed by each cluster can be analysed by using the box plots of the distribution of the variables (Figures 4 and 5). The median values, expressed as a percentage, are considered here to characterize the clusters. Looking

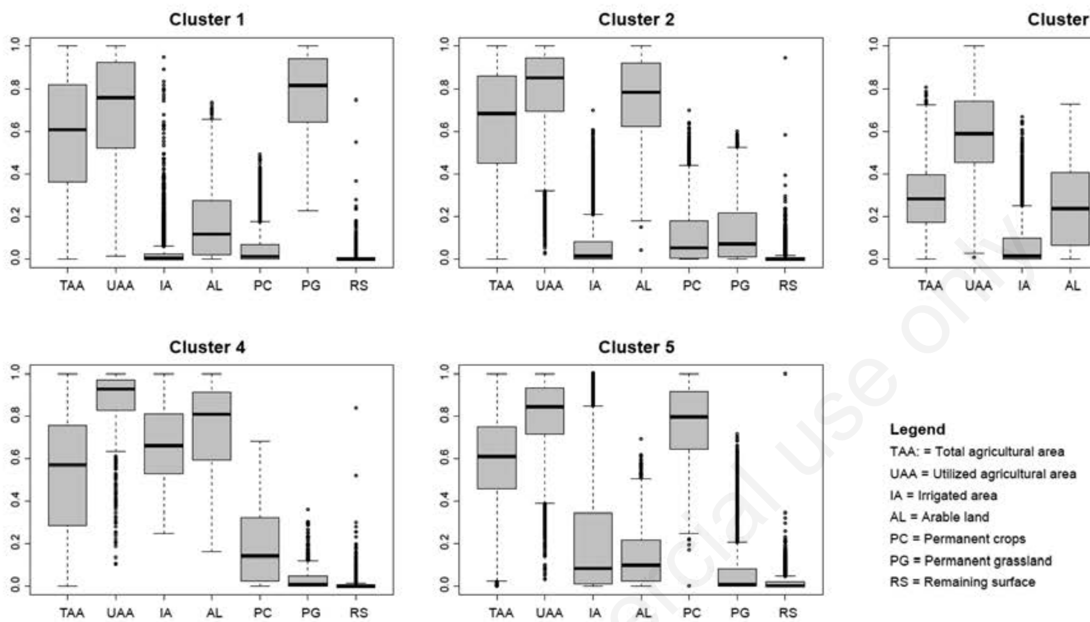

Figure 4. Box plot of variables distribution, based on summary statistic, for each cluster from census 2000. The median value is represented by the line in the middle of the box.

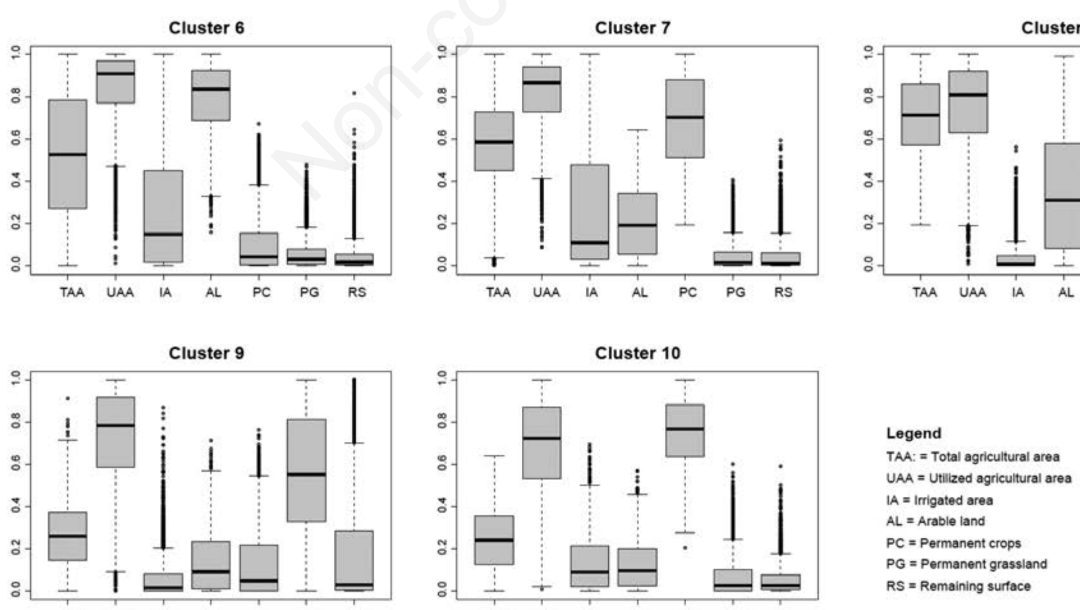

Figure 5. Box plot of variables distribution, based on summary statistic, for each cluster from census 2010. The median value is represented by the line in the middle of the box.

at the clusters from census 2000 (Figure 4), C1 was mainly devoted to permanent grass cultivation (median value of PG equal to 81%). The value of total agricultural area (TAA=61%) was similar for all the clusters, but the incidence of the cultivated area (UAA=76%) was slightly lower. The percentage of irrigated area was the lowest of all clusters (IA~0), and the dispersion of the municipalities was rather contained here (WD=0.43) (Table 2).

The following two clusters (C2 and C4) were both dominated by arable land (AL=78% and 81%, respectively), but C2 showed a lower value for the irrigated land (IA=1%), whereas C4 registered the highest value among all clusters (IA=66%). The incidence of total and utilized agricultural area were similar for these two clusters (TAA=69% and 57% for C2 and C4, respectively, and UAA=85% and 93%, respectively). On average, the municipalities belonging to these clusters were close to their centroids (WD=0.48 and 0.58 for C2 and C4, respectively).

The remaining two clusters (C3 and C5) were both devoted to permanent crop cultivation, but with a different level of incidence (PC=54% and 80%, respectively). These two clusters differed remarkably in terms of total agricultural area (TAA=29% and 61% for C3 and C5, respectively), utilized agricultural area (UAA=59% and 84%, respectively) and average distance from their centroids (WD=1.20 and 0.45, respectively). Finally, the use of irrigation resulted higher in C5 that in C3 (IA = 8% *vs.* 1%). As regards the global assessment of the municipalities involved in each cluster (Table 2), C2 was the cluster that grouped the highest number of administrative units (6237, equal to 38% of the total) and covered the largest surface (32.776 M ha, equal to 42% of the entire studied domain), followed by C1 with 27% of municipalities and 32% of surface. The sizes of C3 and C5 were similar (about 15-17% of municipalities, and 10-13% of surface), whereas C4 was the smallest cluster with just 3% of municipalities and surface.

Turning to the results from census 2010 (Figure 5), arable land predominated over the other LC-related variables only in one cluster (C6, AL=84%). This cluster showed an intermediate value of total agricultural area (TAA=53%) and the highest value of utilized agricultural area (UAA=91%). The use of irrigation was limited (IA=15%) and the dispersion of municipalities within this cluster was quite contained (WD=0.58) (Table 3). The largest land cover variable for the following cluster (C7) was PC (71%), while the values of other variables (TAA=59%, UAA=87%, IA=11%, and WD=0.46) have a similar distribution as in C6. The cluster C8 was characterized by the highest value of TAA (71%) and the lowest value of IA (1%), whereas the value of WD was close to those of the previous clusters (0.48). Arable land and permanent grassland have a similar distribution here (AL=34%, PG=29%), while permanent crops cover a small area (PC=3%).

The last two clusters (C9 and C10) were both associated with a very low level of land devoted to agricultural use (TAA=26% and 24%, respectively) and a limited rate of cultivated surfaces (UAA=78% and 72%, respectively). However, they showed a different type and level of crop specialization. C9 was mainly devoted to permanent grass cultivation (PG=55%), whereas permanent cropland prevailed in C10 (PC=77%). Finally, the dispersion of municipalities around their respective centroids was very different in these two clusters (WD=1.24 and 0.38 for C9 and C10, respectively). As regards the size of the clusters (Table 3), we found a different pattern in 2010 compared to the previous census period. Globally, the differences among clusters were limited in terms of the municipality's number, ranging from 13% (C10) to 27% (C6 and C8). Instead, a major variability was observed at the level of the occupied surface, whose values passed from 9% (C10) to 39% (C8). For instance, C8 was the most widespread cluster (30.115 M ha), but C6, showed the highest number of municipalities (4462) although it covered only 18.458 M ha.

From clusters to agricultural land systems

On the basis of the results of the clustering process, we propose an agronomic interpretation of final cluster in terms of ALS typologies by using the median values of the six variables consid-

Table 2. Within distance (WD), number of municipalities (n) and relative percentage (%), total surface (as millions of hectares, M ha) and relative percentage (%) relatable to each cluster in census 2000 (C1 to C5).

Clusters	WD	Number		Surface	
		n	%	M ha	(%)
C1	0.43	4329	26.6	24.737	31.7
C2	0.48	6237	38.3	32.776	41.9
C3	1.20	2418	14.9	8.022	10.3
C4	0.58	550	3.4	2.187	2.8
C5	0.45	2735	16.8	10.430	13.3

Table 3. Within distance (WD), number of municipalities (n) and relative percentage (%), total surface (as millions of hectares, M ha) and relative percentage (%) relatable to each cluster in census 2010 (C6 to C10).

Clusters	WD	Number		Surface	
		n	%	M ha	(%)
C6	0.58	4462	27.4	18.458	23.6
C7	0.46	2366	14.5	10.850	13.9
C8	0.48	4434	27.3	30.115	38.5
C9	1.24	2867	17.6	12.058	15.4
C10	0.38	2140	13.2	6.670	8.5

ered in this study. The correspondence between each cluster and the ALS typologies has been summarized in Table 4.

The ALS resulting from census 2000 was characterized by a diversified productive vocation: permanent grassland (C1), arable land (C2 and C4), and permanent cropland (C3 and C5). The first cluster (C1) can be related to extensive farming, as revealed by the quite high value of the TAA, but with a rather contained rate of UAA and by the low value of IA. The two systems mainly specialized in arable land (C2 and C4) differed from each other in terms of the use of irrigation, denoting a different level of cropping intensity. Indeed, the low value of IA in C2 can be mainly related to the cultivation of winter cereals, whereas C4 can be ascribable to specialized summer cereals and industrial crops, which usually require a higher use of inputs. The second cluster (C2) showed a higher value of TAA, but a lower of UAA, which indicates less profitability in land cultivation.

In regard to the two permanent crop-based ALS, distinctions can be drawn in relation to: i) the different level of specialization (PC value was lower in C3 than in C5); ii) the incidence of agricultural use of land (TAA in C3 was about half than in C5, and UAA was the lowest of all cluster); iii) the different use of irrigation (much lower in C3 than in C5). Therefore, C5 can be linked to high-specialized vineyards, olive groves and orchards located in intensively cultivated areas. On the other hand, C3 can be related to mixed systems where vineyards, olive and other fruit trees are coupled with herbaceous annual crops, characteristic of marginal areas where agriculture is declining, and the use of irrigation is very limited. Moreover, the remaining surface [RS=UAA - (AL+PG+PC)], which can be assimilated to temporary non-cultivated areas, was very small and almost negligible for all 2000 clusters.

In 2010, the productive vocation of ALS was changed. Only one ALS was specialized in arable lands (C6) and another one in permanent grassland (C9), two ASL (C7 and C10), in PC, and the last one (C8) showed an almost equal interest in AL and PG. To draw a parallel with clusters detected in the previous census period, C1 is close to C9, as both represent an ALS typology based on permanent grassland. In the latest period, C9 showed a remarkable reduction in TAA and a more moderate level of specialization (lower PG), whereas the value of UAA was comparable with C1. This situation could indicate the abandonment of agricultural lands in favor of other land uses (such as afforestation, re-naturalization, or urbanization). Only one cluster from census 2010, C6, showed a clear prevalence of arable land with non-intensive water use, which defines a specialized middle-intensive ALS suffering from the competition with non-agricultural land uses, as also attested by a quite low value of TAA and the high rate of cultivated land. In this case, it was not possible to distinguish between a prevalence of autumn or spring crops, while the intermediate value of IA suggested that this typology of ALS included both crop types.

The two permanent crop-based ALS revealed in 2010 (C7 and C10) were characterized by similar high levels of incidence of PC. One cluster, C10, also showed a very limited agricultural land use (low value of TAA), and a less intensive cultivation rate (UAA) compared to C7. Therefore, we can relate C10 to marginal areas, where the lack of profitability determined a partial abandonment of agricultural lands and low utilization of the available agricultural land. The C7 was characterized by more favorable conditions both for the total and for the utilized agricultural area. In addition, the non-negligible rate of arable land suggests the adoption by farmers of a more diversified strategy of production.

The balanced incidence of agricultural land and permanent grassland characterized in C8 probably resulted from the conversion of arable land-based systems towards less intensive cultivation practices. The limited use of irrigation here could confirm this hypothesis. This ALS does not seem to suffer the competition of other land uses, and showed a fairly high rate of cultivated land (UAA), but the transition towards less intensive cropping systems could be the signal of de-structuring processes anticipating potential abandonment. Two 2010 clusters (C8 and C9) showed a non-negligible rate of non-classified land use (RS) that could be used for self-consumption (kitchen garden) or be temporarily uncultivated. In any case, the expansion of RS represents a signal of a decreasing of professional engagement of farmers in agricultural land use which, when coupled with the low value of the agricultural land (TAA) as in C9, can indicate a regression of agricultural activity.

Agronomic interpretation of the transitions

In 2010, most of the municipalities have transitioned towards the clusters more similar to the starting ones, *i.e.*, those showing the lowest value of DD, as reported in Table 5.

The main transitions recorded (primary and secondary) are drawn in the Figure 6, while the details relating to all the transitions that occurred between the clusters of 2000 and those of 2010 are reported in *Supplementary Tables 5 and 6*.

The extensive grassland systems (C1) moved towards two dif-

Table 4. The correspondence between each cluster and the agricultural land system typologies for the two census periods (C1 to C5 for census 2000; C6 to C10 for census 2010).

Cluster	Agricultural land system
C1	Permanent grassland, extensive (medium TAA and UAA and very low IA)
C2	Arable land (winter cereals), less intensive than C4 (low IA)
C3	Mixed systems (PC and AL), non-intensive (very low TAA and UAA)
C4	Arable land (summer cereals and industrial crops), more intensive than C2 (high IA)
C5	Specialised permanent crops, rather intensive (high UAA and medium IA)
C6	Arable land (summer and winter crops), medium-intensive (low TAA, high UAA, medium IA)
C7	Prevalence of permanent crop, rather intensive (high TAA, UAA and IA)
C8	Mixed systems (PG and AL), quite intensive (high TAA but low IA)
C9	Permanent grassland, abandonment (very low TAA and quite low UAA)
C10	Prevalence of permanent crops, marginal areas (very low TAA and quite low UAA)

TAA, total agricultural area; UAA, utilized agricultural area; IA, irrigated area; PC, permanent crops; AL, arable lands; PG, permanent grassland.

ferent ALS that can represent two successive stages in marginalisation and land abandonment processes. The majority of C1 (61% in surface) passed to a PG-AL mix systems (C8) with a high incidence of RS, but with a still large agricultural use of land. About one-third of the surface passed to a more extensive PG-based system (C9), largely characterized by the land abandonment. As regards the less intensive arable land systems (C2), part of the municipalities (involving 46% of the surface) passed to the AL-based system (C6) characterized by a lower land agricultural use, but a higher level of cultivation intensity. A significant number of municipalities (involving 42% of the surface) changed towards marginalization process, contributing to the establishment of the already mentioned C8 system. The highly intensive arable-land-based systems (C4) lost their specificity in 2010 and converged essentially to C6 systems (75% in surface). The less specialized PC-based system (C3) mostly (45% in surface) passed to a bit more intensive forms of cultivation (C10) while maintaining its original productive vocation, whereas the remaining surface of C3 resulted fragmented among all the other clusters. Finally, the more specialized PC-based systems (C5) largely migrated (67% in surface) towards the new PC system (C7) characterized by the same intensity of cultivation, but a lower level of specialization, due to the contribution of AL coming from C4.

These transitions allowed us to identify the following major temporal trends: i) a reduction of agricultural land use (attested by a global decrease in the mean TAA vale); ii) an increase in UAA and IA; iii) a contraction of arable land (decrease in AL) and permanent grassland (decrease in PG). These results are the consequences of two different and partially contraposed tendencies. In the coastal and suburban areas, the competition with different land uses reduced the availability of surface for agriculture, because of urbanization, littoralization and megalopolis developments. On the other side, the loss of profitability eroded the agricultural surface from more marginal areas in favor of afforestation, re-naturalization or land abandonment. Where agriculture was not undermined by other land use practices, and where an economical convenience was retained, the rate of cultivated land increased, as attested by increasing values of UAA, partially balancing the effects of the reduction of TAA. The increase of IA could be related to the contrasting effects of global warming, but it was also due to the aim of pursuing profitability by intensifying farming practices, where possible.

Although the examination of the changes in ALS with respect to the CAP is not the subject of this study, a few considerations can be made. As a matter of fact, some of the dynamics highlighted here risk being worsened by the measures recently proposed by the European Union (EU). For instance, the decision to convert 10% of UAA in favor of non-productive areas (EU Biodiversity strategy for 2030, 2021) may further increase the shortage of agricultural areas and arable lands. Moreover, the goal pursued to limit the use of fertilizers (-20%) and pesticides (-50%) (A Farm to Fork strategy, 2020) could reduce the crop yields and exacerbate the food dependence of EU. Conversely, measures that implemented measures for grassland conservation could reverse the trend of PG reduction observed in our study.

Spatial distribution of agricultural land systems typologies in western Mediterranean areas

The five ALS typologies resulting from the two-step clustering process were finally assigned back to the single entities (municipalities) in the original dataset and mapped under a GIS environment. This allowed us to visualize the ALS patterns in the two census periods and to evaluate the coherence of the results obtained with the regional geo-morphological and climatic conditions.

The spatial pattern distribution of ALS in the first census period (Figure 7) was characterized by rather large patches slightly fragmented. The permanent grassland-based system (C1) corresponds mainly to the mountainous and hilly areas of Castilla y Leon, Estremadura and Andalucia (Spain), the sub-alpine areas of Provence Alpes - Côte d'Azur, Auvergne Rhone Alpes, the sub-Pyrenean areas of Languedoc and most of them are in Corse (France), and in Sardegna, as well as to some internal areas of Apennines (Lazio, Abruzzo and Umbria) and the reliefs of Liguria (Italy). The less intensive arable land system (C2) covered the main alluvial plains created by the deposition of sediments from the rivers within the study area (Duero, Guadiana in Spain and Portugal; Segura, Guadalquivir and Ebro in Spain; Arno, Tevere, Basento in Italy). Conversely, the most intensive arable land system (C4) was concentred in relatively small districts characterized

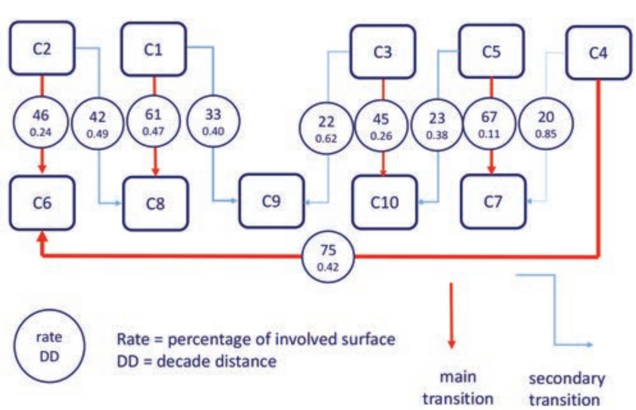

Figure 6. The scheme of transition among clusters from the two census periods (C1-C5 for 2010 and C6-C10 for 2010). The thickness of the manifolds is correlated to the rate of surface involved.

Table 5. Cross-tabulation showing the decade distance between centroids belonging to the two census periods (C1-C5 for 2000, and C6-C10 for 2010).

Clusters	C6	C7	C8	C9	C10
C1	1.00	1.00	0.47*	0.40*	1.06
C2	0.24*	0.83	0.49	0.84	1.00
C3	0.82	0.44	0.71	0.62	0.26*
C4	0.42*	0.85	0.85	1.04	1.05
C5	0.93	0.11*	0.81	0.88	0.38

*The cells with the lower value represent the main transition.

by very specific and narrow conditions, such as Lleida and Almeria (Spain), Salerno (Italy). The more specialized permanent cropland system (C5) was dislocated in the traditional districts of olive trees and vineyards such as Norte (Portugal), Andalusia, Région de Murcha, Valenciana and Cataluña (Spain), Languedoc in France, Toscana, Puglia and Sicilia (Italy). Finally, the less specialized permanent cropland system (C3) was found in Centro (Portugal), Galicia (Spain), Provence Alpes - Côte d'Azur (France) and in some neighbouring areas to C5.

Regarding the second period of investigation (Figure 8), the distribution of the spatial pattern of the clusters was generally more fragmented. It is evident that many municipalities characterized by low-intensive permanent-grass (C1) and arable land (C2) systems converged in low-intensive mixed crop systems (C8), especially in Extremadura, Castilla La Mancha and Castilla y Leon (Spain), in Alantejo (Portugal), and in Toscana, Sicilia, Sardegna and Basilicata (Italy). The permanent-grass systems (C9) remained only in Galizia (Spain), Corse and Provence Alpes Côte d'Azur

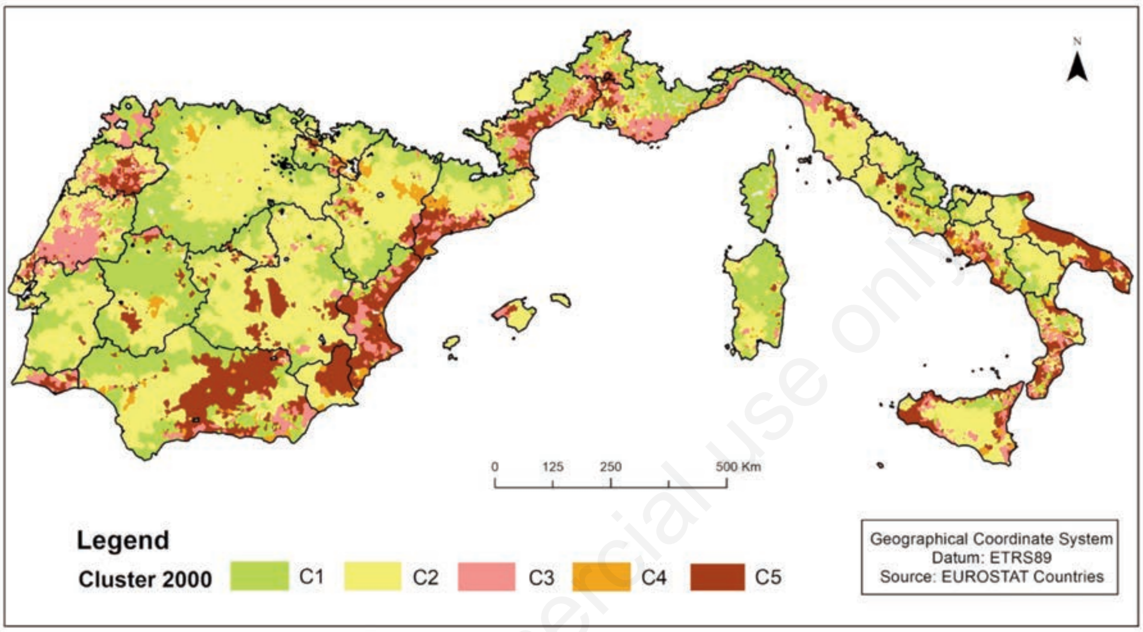

Figure 7. Spatial pattern of agricultural land systems derived from census 2000 (see the text for the description of the clusters).

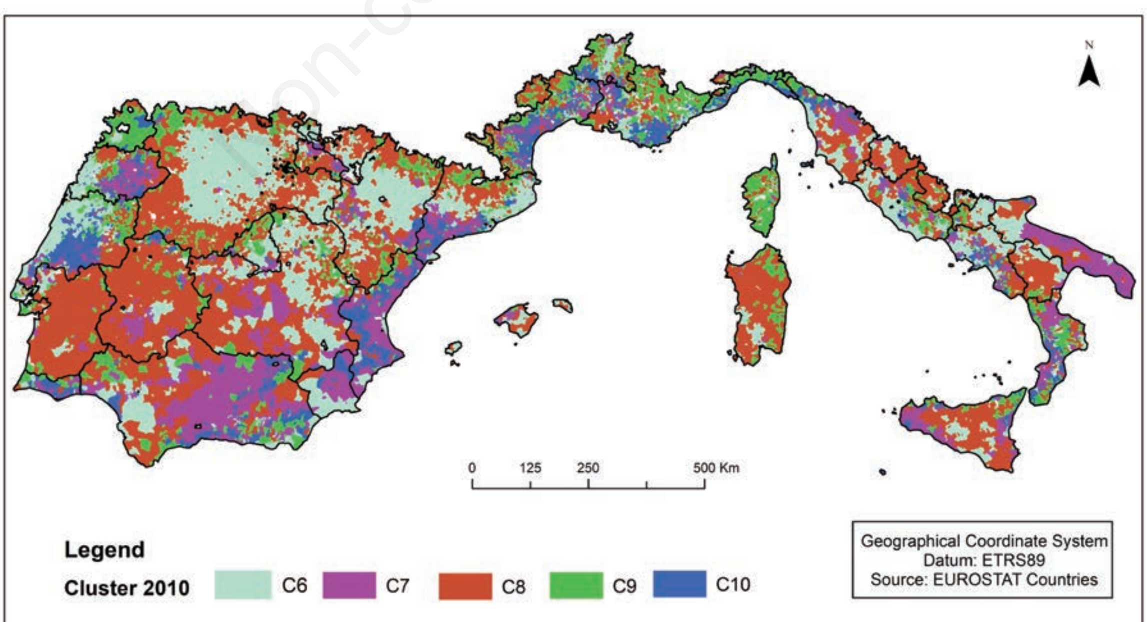

Figure 8. Spatial pattern of agricultural land systems derived from census 2010 (see the text for the description of the clusters).

(France), and Liguria and Calabria (Italy).

The arable land-system (C6) maintained its importance in many regions of Spain (Cataluña, Castilla y Leon and Aragon) while in Italy its distribution in more fragmented and intermingled with permanent grassland, characteristic of C8. Most of the areas devoted to the specialized permanent crop systems in 2000 (C5) transitioned unaltered to C7 (Norte in Portugal; Andalucia in Spain; Puglia and some parts of Toscana and Sicilia in Italy). In some cases, we observed a migration to C10, that is a less intensive and specialized PC system (Région de Murcia and Valenciana in Spain; Languedoc in France). A low-intensive permanent crop systems remain unaltered in different area, like Centro (Portugal), Provence (France), and Calabria (Italy), where C3 passed to C10.

The quantitative assessment of these transitions revealed that the observed changes were contained, but not negligible (Table 6). About 50% of municipalities and the surveyed surface presented very low or low changes (DD lower than 0.50) and with a limited dispersion of the municipalities within their cluster (WD equal to 0.38, 0.46, and 0.48 for C10, C7, and C8, respectively). Greater changes (DD>0.68) were reached only by few municipalities (13%) and involved limited surfaces (11%). Intermediate values of decade distance (between 0.49 and 0.68) represented the most widespread class by surface involved (39%) and second as number of municipalities (29%).

Geographically, changes were mainly concentrated in the southern part of Portugal, in the western part of Spain, in Auvergne and Rhone Alpes (France), in Sardegna, and the internal areas of Italy (Figure 9). The areas with higher values of DD were very small and fragmented, and these dynamics are probably the result of driving forces acting at a local level. For instance, their proximity to urbanized areas (such as in the regions of Cordoba, Jaen, and Tarragona in Spain, or Siracusa and Palermo in Italy) or to alluvial plains (as the Mouth of Guadalquivir in Spain or the plain of Teria in Italy) could cause the observed changes.

Our findings are consistent with those of several authors. For example, Levers *et al.* (2018) and Rega *et al*. (2020) found patterns in ALS with a similar geographical distribution and characterisation than ours. In particular, Levers *et al.* (2018) identified four

Table 6. Classification of the municipalities (number and surface involved, plus the relative percentages) in relation to the decade distance. Decade distance was calculated for each municipality as the Euclidean distance between the centroids of clusters detected in the two census periods.

Decade distance	Municipalities		Surface	
	n	%	M ha	(%)
Very low (0.21-0.30)	6235	38.3	25.745	32.9
Low (0.30-0.49)	3248	20.0	13.302	17.0
Intermediate (0.49-0.68)	4666	28.7	30.711	39.3
High (0.68-0.87)	1472	9.0	5.959	7.6
Very high (0.87-1.06)	648	4.0	2.434	3.1
Total	16269	100.0	78.151	100.0

M ha, millions of hectares.

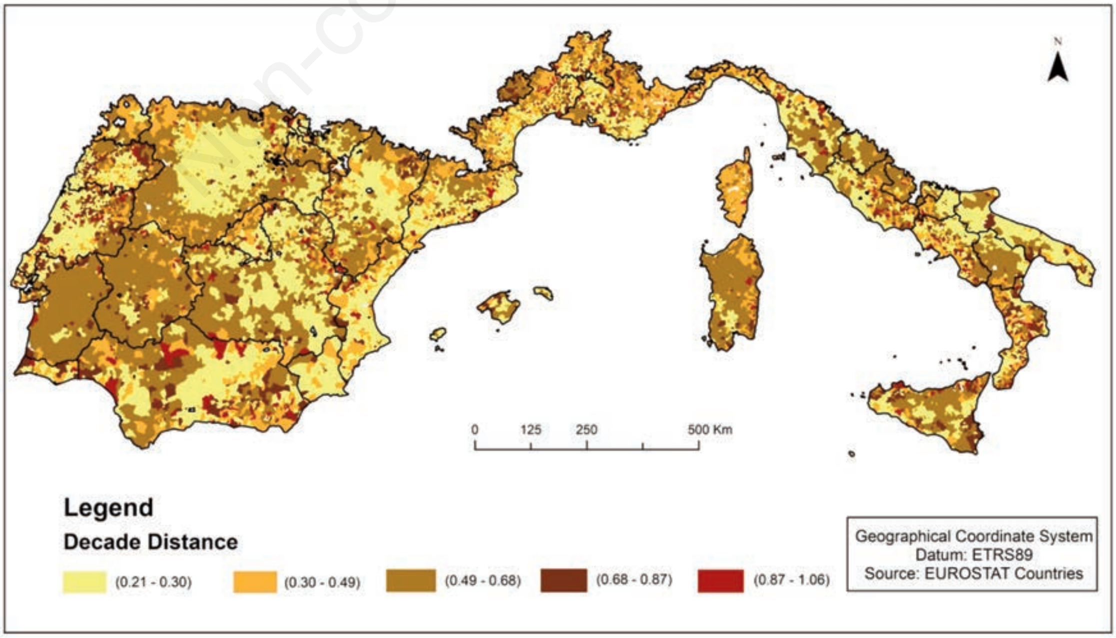

Figure 9. Map representing decade distance in five class intervals. Decade distance was calculated for each municipality as the Euclidean distance between the centroids of clusters detected in the two census periods.

main trends: i) yield increase and cropland restriction; ii) forestry expansion in Mediterranean areas; iii) urbanisation along the cost; and iv) stable land-use patterns in part of Western Europe. The detailed maps reported for Spain and Italy by Malek and Verburg (2017), who used an expert hierarchical classification procedure, show a similar distribution of ALS even if the number of categories is higher than in our case. Some analogies in the distribution of ALS can also be found in the work of van der Zanden *et al.* (2016), despite the authors focused on the agricultural landscape, including additional elements such as the management intensity and available livestock. Marraccini *et al.* (2015) underlined the effect of urbanization around the Mediterranean medium-large cities, with an increase of urban areas higher than 10% over three-decade period (1980-2010). Holman *et al.* (2017) estimated an extensification and abandonment of agriculture in Mediterranean countries (especially Spain and Italy) due to climatic changes (heat stress and drought) as a possible scenario projected for the 2050s. Debolini *et al.* (2018), by analyzing 80 papers published in international journals from 1985 to 2015, confirmed the geographic relationship existing between land abandonment, mountain areas, urbanization, and coastal plains.

Conclusions

In the present study, we proposed an unsupervised learning procedure based on SOM able to extract useful information from a huge volume of multivariate agricultural census data. The implemented clustering approach led to the aggregation of the municipalities into output units on the basis of the proximity among the variables in the input datasets, derived from the national agricultural censuses of Portugal, Spain, France, and Italy. The units of the SOM grid were then grouped into five main clusters to form the final partitioning, which have been interpreted in terms of ALS.

This innovative approach (in the context of this domain of application), the size of the investigated area, and the scale of detail of the used input data enabled us to obtain important outputs. More specifically, the framework adopted allowed us to describe and detect the spatial pattern of the main ALS typologies, to assess the contribution of the agronomic variables related to land use and land cover to each ALS, to identify the main changes that occurred between the two census periods (2000 and 2010), and to relate them to local conditions.

The main advantage of using a data-driven approach is that it allows us to discover hidden patterns in the data, and the resulting outputs derived from hard information instead of from expert-based rules. In the present case study, the prevalent agricultural typologies characterizing the ALS in the two census periods were shaped by the reality of the surveyed area solely, with regard to its agronomic assessment.

The reasons that are at the origin of the observed changes and the impacts that these changes can have on many aspects of human societies, have briefly been discussed. For instance, the CAP measures recently proposed by the EU could enhance some of the observed changes in ALS and mitigate others. The recent trend toward longer dry periods due to climate change can be invocated to explain the more fragmented distribution of ALS due to a reduction of the suitable area for traditional crops and the increase in irrigation use, both occurring in 2010.

The effects of human migration (littoralisation and megacities) are evident by examining the spatial distribution of ALS and their temporal dynamic, even if its range of influence is more limited in comparison with a few decades ago. These trends look set to continue as the pressure of these driving forces will not weaken in the coming years.

The forthcoming publication of the 2020 agricultural census will allow us to verify whether the transitions of ALS described here have persisted or whether new dynamics have occurred. Ultimately, an analysis of the agronomic and environmental consequences of ALS changes (*e.g.*, soil carbon storage, soil erosion control, and soil waterproofing due to urban expansion) could provide valuable information to feed the debate about land-sharing and land-sparing in Mediterranean areas.

References

A Farm to Fork strategy, 2020. Available from: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52020DC0381 (accessed on 7th April 2023)

Abu Hammad A, Tumeizi A, 2012. Land degradation: Socioeconomic and environmental causes and consequences in the eastern mediterranean. Land Degrad. Dev. 23:216-26.

van Asselen S, Verburg PH, 2012. A Land System representation for global assessments and land-use modeling. Glob. Change Biol. 18:3125-48.

Bajocco S, De Angelis A, Perini L, Ferrara A, Salvati L, 2012. The impact of Land Use/Land Cover Changes on land degradation dynamics: A Mediterranean case study. Environ. Manag. 49:980-9.

EU biodiversity strategy for 2030, 2021. Available from: https://op.europa.eu/en/publication-detail/-/publication/31e4609f-b91e-11eb-8aca-01aa75ed71a1 (accessed on 7th April 2023)

Bonet A, 2004. Secondary succession of semi-arid Mediterranean old-fields in south-eastern Spain: Insights for conservation and restoration of degraded lands. J. Arid Environ. 56:213-33.

Debolini M, Marraccini E, Dubeuf JP, Geijzendorffer IR, Guerra C, Simon M, Targetti S, Napoléone C, 2018. Land and farming system dynamics and their drivers in the Mediterranean Basin. Land Use Policy 75:702-10.

Debolini M, Schoorl JM, Temme A, Galli M, Bonari E, 2013. Changes in Agricultural Land Use Affecting Future Soil Redistribution Patterns: A Case Study in Southern Tuscany (Italy). Land Degrad. Dev. 26:574-86.

DrakeDrake NA VA, 2014. Review of spatial and temporal methods for assessing land degradation in the Mediterranean. Advances In Environmental Monitoring And Modelling 1:1-52.

Erb KH, Haberl H, Jepsen MR, Kuemmerle T, Lindner M, Müller D, Verburg PH, Reenberg A, 2013. A conceptual framework for analysing and measuring land-use intensity. Curr. Opin. Environ. Sustain. 5:464-70.

Estel S, Kuemmerle T, Alcántara C, Levers C, Prishchepov A, Hostert P, 2015. Mapping farmland abandonment and recultivation across Europe using MODIS NDVI time series. Remote Sens. Environ. 163:312-25.

Estel S, Kuemmerle T, Levers C, Baumann M, Hostert P, 2016. Mapping cropland-use intensity across Europe using MODIS NDVI time series. Environ. Res. Lett. 11.

Eurostat, 2023. Eurostat Agriculture Glossary. Available from: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Category:Agriculture_glossary

Hansen MC, Potapov PV, Moore R, Hancher M, Turubanova SA, Tyukavina A, Thau D, Stehman S V., Goetz SJ, Loveland TR,

Kommareddy A, Egorov A, Chini L, Justice CO, Townshend JRG, 2013. High-resolution global maps of 21st-century forest cover change. Science 342:850-3.
Hill J, Stellmes M, Udelhoven T, Röder A, Sommer S, 2008. Mediterranean desertification and land degradation. Mapping related land use change syndromes based on satellite observations. Glob. Planet. Change 64:146–57.
Huber R, Bakker M, Balmann A, Berger T, Bithell M, Brown C, Grêt-Regamey A, Xiong H, Le QB, Mack G, Meyfroidt P, Millington J, Müller B, Polhill JG, Sun Z, Seidl R, Troost C, Finger R, 2018. Representation of decision-making in European agricultural agent-based models. Agric. Syst. 167:143-60.
Jansen LJM, Di Gregorio A, 2002. Parametric land cover and land-use classifications as tools for environmental change detection. Agric. Ecosyst. Environ. 91:89-100.
Jepsen MR, Kuemmerle T, Müller D, Erb K, Verburg PH, Haberl H, Vesterager JP, Andrič M, Antrop M, Austrheim G, Björn I, Bondeau A, Bürgi M, Bryson J, Caspar G, Cassar LF, Conrad E, Chromý P, Daugirdas V, Van Eetvelde V, Elena-Rosselló R, Gimmi U, Izakovicova Z, Jančák V, Jansson U, Kladnik D, Kozak J, Konkoly-Gyuró E, Krausmann F, Mander Ü, McDonagh J, Pärn J, Niedertscheider M, Nikodemus O, Ostapowicz K, Pérez-Soba M, Pinto-Correia T, Ribokas G, Rounsevell M, Schistou D, Schmit C, Terkenli TS, Tretvik AM, Trzepacz P, Vadineanu A, Walz A, Zhllima E, Reenberg A, 2015. Transitions in European land-management regimes between 1800 and 2010. Land Use Policy 49:53-64.
Kanevski M, Pozdnoukhov A, Timonin V, 2009. Machine Learning for Spatial Environmental data. EPFL Press, Lausanne, Switzerland.
Khan F, 2012. An initial seed selection algorithm for k-means clustering of georeferenced data to improve replicability of cluster assignments for mapping application. Appl. Soft Comput. J. 12:3698-700.
Kohonen T, 1982. Self-organized formation of topologically correct feature maps. Biol. Cybern. 43:59-69.
Kohonen T, 2001. Self-organizing maps. Springer-Verlag, Berlin, Heidelberg, New York 3rd ed., p. 501.
Kuemmerle T, Kaplan JO, Prishchepov AV, Rylsky I, Chaskovskyy O, Tikunov VS, Müller D, 2015. Forest transitions in Eastern Europe and their effects on carbon budgets. Glob. Change Biol. 21:3049-61.
Lambin EF, Geist H, Rindfuss RR, 2006. Introduction: Local Processes with Global Impacts. Land-Use And Land-Cover Change, pp. 1-8.
Lambin EF, Meyfroidt P, 2011. Global land use change, economic globalization, and the looming land scarcity. Proc. Nat. Acad. Sci. (PNAS). 108:3465-72.
Levers C, Müller D, Erb K, Haberl H, Jepsen MR, Metzger MJ, Meyfroidt P, Plieninger T, Plutzar C, Stürck J, Verburg PH, Verkerk PJ, Kuemmerle T, 2018. Archetypical patterns and trajectories of land systems in Europe. Reg. Environ. Change. 18:715-32.
Li X, Wu K, Liang Y, 2023. A Review of Agricultural Land Functions: Analysis and Visualization Based on Bibliometrics. Land. 12:561.
Lloyd S, 1982. Least squares quantization in PCM. IEEE Trans. Inf. Theory. 28:129-37.
MacQueen J, 1967. Some methods for classification and analysis of multivariate observations. Proceedings Of The Fifth Berkeley Symposium On Mathematical Statistics And Probability, Volume 1. Statistics 5.1:281-98.
Malek Ž, Verburg P, 2017. Mediterranean land systems: Representing diversity and intensity of complex land systems in a dynamic region. Landsc. Urban Plan. 165:102-16.
Marraccini E, Debolini M, Moulery M, Abrantes P, Bouchier A, Chéry JP, Sanz Sanz E, Sabbatini T, Napoleone C, 2015. Common features and different trajectories of land cover changes insix Western Mediterranean urban regions. Appl. Geogr. 62:347-56.
McCulloch WS, Pitts W, 1943. A logical calculus of the ideas immanent in nervous activity. Bull. Math. Biophys. 5:115-33.
Murray-Rust D, Robinson DT, Guillem E, Karali E, Rounsevell M, 2014. An open framework for agent based modelling of agricultural land use change. Environ. Model. Softw. 61:19-38.
Rabelo M, Debolini M, Villani R, Sabbatini T, Silvestri N, 2021. Expansion and specialization of agricultural systems in western mediterranean areas: A global analysis based on the two last census data. Agronomy 11.
Rega C, Short C, Pérez-Soba M, Luisa Paracchini M, 2020. A classification of European agricultural land using an energy-based intensity indicator and detailed crop description. Landsc. Urban Plan. 198.
Schröter D, Cramer W, Leemans R, Prentice IC, Araújo MB, Arnell NW, Bondeau A, Bugmann H, Carter TR, Gracia CA, De La Vega-Leinert AC, Erhard M, Ewert F, Glendining M, House JI, Kankaanpää S, Klein RJT, Lavorel S, Lindner M, Metzger MJ, Meyer J, Mitchell TD, Reginster I, Rounsevell M, Sabaté S, Sitch S, Smith B, Smith J, Smith P, Sykes MT, Thonicke K, Thuiller W, Tuck G, Zaehle S, Zierl B, 2005. Ecology: Ecosystem service supply and vulnerability to global change in Europe. Science 310:1333-7.
Silvestri N, Pistocchi C, Sabbatini T, Rossetto R, Bonari E, 2012. Diachronic analysis of farmers' strategies within a protected area of central Italy. Ital. J. Agron. 7.2:e20-e20.
Sluiter R, De Jong SM, 2007. Spatial patterns of Mediterranean land abandonment and related land cover transitions. Landsc. Ecol. 22:559-76.
Stellmes M, Röder A, Udelhoven T, Hill J, 2013. Mapping syndromes of land change in Spain with remote sensing time series, demographic and climatic data. Land Use Policy 30:685-702.
Symeonakis E, Calvo-Cases A, Arnau-Rosalen E, 2007. Land use change and land degradation in southeastern Mediterranean Spain. Environ. Manag. 40:80-94.
Tonini M, Parente J, Pereira MG, 2018. Global assessment of rural-urban interface in Portugal related to land cover changes. Nat. Hazards Earth Syst. Sci. 18:1647-64.
Václavík T, Lautenbach S, Kuemmerle T, Seppelt R, 2013. Mapping global land system archetypes. Glob. Environ. Change 23:1637-47.
Verburg PH, Crossman N, Ellis EC, Heinimann A, Hostert P, Mertz O, Nagendra H, Sikor T, Erb K-H, Golubiewski N, Grau R, Grove M, Konaté S, Meyfroidt P, Parker DC, Chowdhury RR, Shibata H, Thomson A, Zhen L, 2015. Land system science and sustainable development of the earth system: A global land project perspective. Anthropocene 12:29-41.
Verburg PH, Erb K-H, Mertz O, Espindola G, 2013. Land System Science: between global challenges and local realities. Curr. Opin. Environ. Sustain. 5:433-7.
Verburg PH, Overmars KP, 2009. Combining top-down and bottom-up dynamics in land use modeling: Exploring the future of abandoned farmlands in Europe with the Dyna-CLUE model. Landsc. Ecol. 24:1167-81.
Vesanto J, 1999. SOM-based data visualization methods. Intell.

Data Anal. 3:111-26.
Viana CM, Freire D, Abrantes P, Rocha J, Pereira P, 2022a. Agricultural land systems importance for supporting food security and sustainable development goals: A systematic review. Sci. Total Environ. 806.
Viana CM, Freire D, Abrantes P, Rocha J, Pereira P, 2022b. Agricultural land systems importance for supporting food security and sustainable development goals: A systematic review. Sci. Total Environ. 806:150718.
Villani R, Sabbatini T, Perez OM, Guiomar N, Debolini M, 2019. An open dataset about georeferenced harmonized national agricultural censuses and surveys of seven mediterranean countries. Data In Brief 27:104774.
van Vliet J, de Groot HLF, Rietveld P, Verburg PH, 2015. Manifestations and underlying drivers of agricultural land use change in Europe. Landsc. Urban Plan. 133:24-36.
Wehrens R, Buydens LMC, 2007. Self- and Super-organizing Maps in R: The kohonen Package. J. Stat. Softw. 21:1-19.
Wehrens R, Kruisselbrink J, 2018. Flexible Self-Organizing Maps in kohonen 3.0. J. Stat. Softw. 87:1-18.
Young J, Richards C, Fischer A, Halada L, Kull T, Kuzniar A, Tartes U, Uzunov Y, Watt A, 2007. Conflicts between biodiversity conservation and human activities in the central and eastern European countries. Ambio 36:545-50.
van der Zanden EH, Levers C, Verburg PH, Kuemmerle T, 2016. Representing composition, spatial structure and management intensity of European agricultural landscapes: A new typology. Landsc. Urban Plan. 150:36.49.

Online supplementary material
Table S1. The main descriptive statistics calculated for the 6 considered variables on data of the first census date.
Table S2. The main descriptive statistics calculated for the 6 considered variables on data of the first census date.
Table S3. Descriptive statistics of variables within each cluster (census 2000).
Table S4. Descriptive statistics of variables within each cluster (census 2010).
Table S5. The cross-tabulation showing the transitions, in terms of municipality number and percentage, from clusters in 2000 (C1-C5) to clusters in 2010 (C6-C10). The cells with a value higher than 30% are highlighted in grey.
Table S6. The cross-tabulation showing the transitions, in terms of surface and percentage, from clusters in 2000 (C1-C5) to clusters in 2010 (C6-C10). The cells with a value higher than 30% are highlighted in grey.
Figure S1. Neighborhood distance maps (census 2000 on the left and census 2010 on the right). Gradient color scales: yellow=low, red=high values.
Figure S2. Node counts maps (census 2000 on the left and census 2010 on the right). Gradient color scales: blue=low, green=medium, red=high values.
Figure S3. The final five clusters with codebooks for census data 2000.
Figure S4. The final five clusters with codebooks for census data 2010.