Agricultural Systems 191 (2021) 103140

Contents lists available at ScienceDirect

Agricultural Systems

journal homepage: www.elsevier.com/locate/agsy

Explaining farming systems spatial patterns: A farm-level choice model based on socioeconomic and biophysical drivers

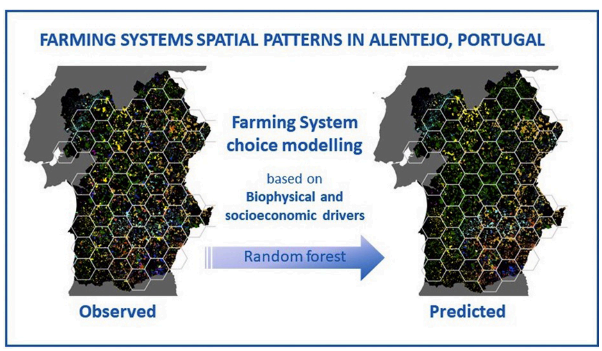
Paulo Pacheco de Castro Flores Ribeiro [a,*], José Manuel Osório de Barros de Lima e Santos [a], Maria João Prudêncio Rafael Canadas [a], Ana Maria Contente de Vinha Novais [a], Francisco Manuel Ribeiro Ferraria Moreira [b,d], Ângela Cristina de Araújo Rodrigues Lomba [c]

[a] *CEF – Centro de Estudos Florestais, Instituto Superior de Agronomia, Universidade de Lisboa, Tapada da Ajuda, Lisboa, Portugal*
[b] *CIBIO-InBIO, Universidade do Porto, Porto, Portugal*
[c] *CIBIO-InBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos, Universidade do Porto, Porto, Portugal*
[d] *CIBIO-InBIO, Instituto Superior de Agronomia, Universidade de Lisboa, Lisboa, Portugal*

HIGHLIGHTS

- Farmland landscapes can show high spatial variations, resulting from farmers' adaptive responses to the environmental context.
- Understand how biophysical and socio-economic factors influence farming system choice to anticipate farm management decisions.
- Farms' biophysical and socioeconomic features constrain farmers' decisions, driving the choice of the farming system.
- Considering farm-level drivers of farming systems choice is crucial to an informed agricultural landscape planning.
- An innovative tool to assess policy or climate scenarios is presented, based on a farm-level farming systems approach.

GRAPHICAL ABSTRACT



ARTICLE INFO

Editor: Dr Jagadish Timsina

Keywords:
Farming systems
Territorial context
Choice modelling
Spatial analysis
Random
Forest

ABSTRACT

CONTEXT: Efforts to bring together landscape analysis and farming systems have failed to explain the drivers behind their spatial distribution. Since agricultural landscapes are an outcome of farmers' decisions, understanding the role of socioeconomic and biophysical drivers of such decisions is essential for policy-making targeting landscape-level provision of public goods and ecosystem services from agriculture.
OBJECTIVE: Aiming to better understand the role of these drivers, we focused on a region dominated by agricultural use, with extensive variability in biophysical and socioeconomic conditions. A typology of farming systems was derived from spatially explicit farm-level data provided by the Portuguese agency responsible for Common Agricultural Policy payments, for 2017. Farms were thoroughly characterized through relevant biophysical and socioeconomic variables considered as potential drivers of farming systems.

* Corresponding author at: Instituto Superior de Agronomia, Tapada da Ajuda, 1349-017 Lisboa, Portugal.
E-mail address: pfribeiro@isa.ulisboa.pt (P. Pacheco de Castro Flores Ribeiro).

https://doi.org/10.1016/j.agsy.2021.103140
Received 24 September 2020; Received in revised form 21 March 2021; Accepted 24 March 2021
Available online 10 April 2021

METHODS: A random forest approach was used to develop a farming system choice-model, dependent on those biophysical and socioeconomic variables. Variable importance measures and partial dependence plots were used to explore the role of these variables in explaining the spatial distribution of farming systems and to predict spatial patterns at the landscape scale.
RESULTS AND CONCLUSIONS: Results showed that both biophysical and socioeconomic drivers play a significant role in the spatial distribution of most agricultural systems. Its importance, however, varies significantly across farming systems, being crucial for some and almost irrelevant for others. Farm size and climate have proved to be the most relevant drivers for most farming systems. Overall, our approach proved to be quite accurate in predicting patterns of farming systems at the landscape scale.
SIGNIFICANCE: The proposed framework has shown great potential as a tool to support information-based policy design to improve agricultural landscape planning, by linking farm-level management decisions with the provision of socially valued public goods from agriculture, perceived at the landscape-level.

1. Introduction

Agriculture is a dominant land use in many parts of the world, resulting from human interaction with nature over time. This interaction is mostly regulated by two main types of drivers: biophysical (climate, soil, topography…) and socioeconomic (farm structure, characteristics of farmers, markets, policies…). The way each of these drivers affects agricultural landscapes has attracted the interest of researchers (Grigg, 2005; Hazell and Wood, 2008; Kristensen et al., 2016; Plieninger et al., 2016; van Vliet et al., 2015), but many unanswered questions still persist (Plieninger et al., 2016; Wilson, 2009). Advancing knowledge about the role played by each of these factors in shaping agricultural landscapes can thus improve our understanding of human/environment interactions, allowing to anticipate farm management decisions and supporting evidence-based public intervention (Levers et al., 2016; van de Steeg et al., 2010).

Such issues have recently been raised in the context of the provision of public goods and agroecosystem services in general, including biodiversity conservation (Landis, 2017; Schaller et al., 2018). Much literature resort to aggregated data concerning land use or to agriculture intensification or specialization indicators, privileging landscape-dynamics analysis over landscape regional differentiation, and seldom take the farm as the unity of inquiry (Debolini et al., 2018; Ruiz-Martinez et al., 2015). There has been, however, a pressing need to the development of approaches linking landscape analysis and farming systems (FS) to understand agricultural landscapes, which are able to establish the FS geography but do not go into explaining the drivers behind their spatial distribution (Andersen, 2017; Benoît et al., 2012; Martel et al., 2019; Rizzo et al., 2013; van de Steeg et al., 2010). Indeed, considering the mismatch between the farm-scale, where management decisions take place, and the landscape-scale, where ecosystem services are perceived, landscape analysis can greatly benefit from a deeper

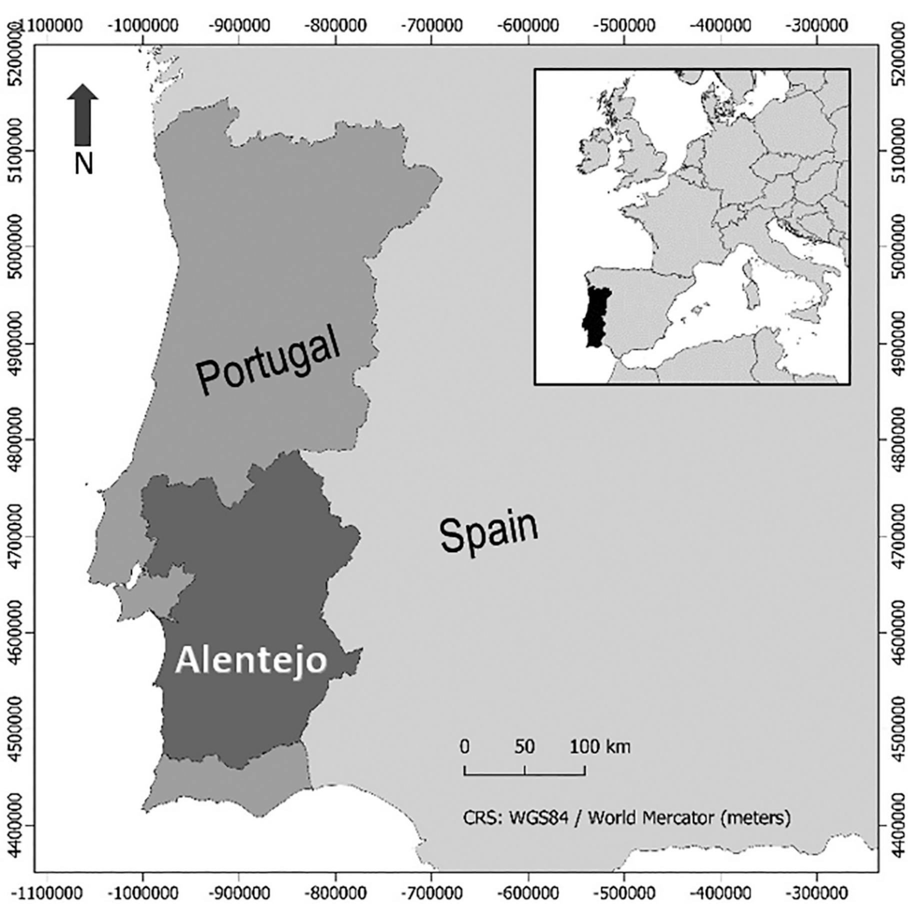


Fig. 1. Location of the study area in the Alentejo region (NUTS2), Portugal.

understanding of the factors that influence farm management decisions. Thus, understanding the multiple production decisions of adjacent farmers and combining these decisions at the landscape-scale is key to explain the landscape mosaic and the ecological disturbance regimes (fire, grazing, ploughing…) that shape the habitats of wild species and the provision of diverse ecosystem services.

The FS concept used in this study follows that proposed by Santos et al. (2020), according to which a FS can be defined as a set of farms roughly practicing the same crops and agricultural activities, using similar technological processes and input endowments. A key aspect in this concept is that only variables resulting from farm management decisions are considered, when defining a FS; all variables that may influence these decisions but do not result from them, at least in the short run (e.g. farm size or fragmentation level, climate, slopes, market or policy), should be considered as exogenous to the FS and, therefore, as potential drivers of the FS choice (Silva et al., 2020).

To explain the spatial distribution of FS, distinct groups of drivers can be considered according to distinct disciplinary perspectives or theoretical approaches. The analysis of farm biophysical endowments to explain spatial patterns of FS has largely been explored by geography and geo-agronomy (Deffontaines, 2004; Deffontaines et al., 1995; Grigg, 2005; Lacoste et al., 2018). Climate, soil, and slope are often considered to establish a range of restrictions to the choice of the farming system. But FS are also dependent on farmland structure and social context. Farmland structure covers an ensemble of constraints such as farm size, fragmentation and spatial composition which potentially restrict farmers decisions (Grigg, 2005; Latruffe and Piet, 2014; Reboul, 1976; Ribeiro et al., 2018). The influence of territorial socioeconomic context on FS location may be grounded in the notion of local embeddedness, supported by local sociocultural, demographic and economic structures (Canadas and Novais, 2014; Debolini et al., 2018).

Using farm-level data collected in 2017 in a large-scale study area, we developed an innovative methodological approach to: 1) derive a spatially-explicit FS typology; 2) assess the role of socioeconomic and biophysical factors in explaining the spatial distribution of those FS; 3) assess the extent to which we can predict FS patterns based on biophysical and socioeconomic variables. Results were used to discuss the role of these drivers on the choice of the FS and their potential to predict landscape patterns, seeking to draw conclusions to better inform policy design for landscape-level provision of public goods from agriculture and prediction of landscape patterns in face of biophysical or socio-economic changes.

2. Methods

2.1. Study area

The study focused on the Alentejo region, in southern Portugal (Fig. 1), corresponding to the EU statistical region PT18, at the NUTS2 level (Nomenclature of Territorial Units for Statistics). Covering about 31,551 km^2 (ca. 1/3 of Portugal), the region has a Mediterranean climate, with hot dry summers and mild rainy winters. The annual average temperature is about 16.3 °C, ranging from 9.9 °C to 23.4 °C in January and August, respectively, and the total annual rainfall is about 619 mm, largely concentrated in the rainy season (approx. October to March). The relief is predominantly smooth (47% of the land with slope $<$ 5%; but 14% with slope $>$ 15%), with few mountain areas (average altitude is 176 m a.s.l., ranging from 0 to 1020 m).

According to the latest agricultural census in Portugal (2009), the utilized agricultural area (UAA) in Alentejo (NUT 2) was then ca. 2.2 million hectares, covering almost 70% of the region and making it the dominant land use. Official statistics report that in 2016 the utilized agricultural area (UAA) was dominated by permanent pastures (64%), followed by annual crops (24%) and permanent crops (11%). Cereals, forages and olive groves were the main crops, with roughly equal shares of 8% in total UAA (making ca. 70% of the UAA excluding permanent pastures). Nearly 40% of the UAA is under the canopy of scattered trees, mainly cork and holm oaks (*Quercus suber* and *Q. rotundifolia* respectively), originating an agroforestry system locally named “montado”, which is largely acknowledge for its high nature value (Ferraz-de-Oliveira et al., 2016). Cropland in these undercover areas are mainly permanent pastures (70%) and annual crops (30%). Most of the UAA is rainfed (ca. 90%) and irrigated areas are mostly located within state-promoted irrigation systems, often depending on large dams. The region is dominated by large holdings, with almost 90% of the UAA in farms with more than 50 ha.

2.2. Farming systems identification

To build a farming systems typology for the study area we used data from the EU Integrated Administration and Control System (IACS) for 2017, associated with spatially explicit farm parcel data from the Land Parcel Identification System (LPIS), provided by the Portuguese agency responsible for Common Agricultural Policy (CAP) payments. These data are collected on a yearly basis from farmers declarations when applying for CAP payments and its usefulness for FS research has been demonstrated by previous studies (Lomba et al., 2017; Ribeiro et al., 2014, 2016, 2018).

The raw dataset identified 26,648 CAP beneficiaries in the study area, covering a total of 2,221,816 ha distributed over 208,338 parcels which, in turn, included 560,213 subparcels for which land use/crop cover was described. Livestock declared by each beneficiary was also provided, describing species composition, gender, age groups and an indication of whether they were kept in stables or grazing.

First, all parcels declared by the same CAP beneficiary were taken as a single farm. However, we found that some beneficiaries reported very scattered parcels, sometimes separated by hundreds of kilometres, where the farm concept (as an agricultural management unit) would not

Table 1
Summary statistics for the land use/cover and livestock farm characterization variables (n = 24,313 farms).

Variable	Mean	SD
Land use/cover variables (proportion of total UAA)		
Rice (both Indica and Japonica)	0.012	0.1
Cereals Irrigated (corn, wheat, oats, barley, triticale)	0.018	0.104
Cereals rainfed (wheat, corn, oats, barley, rye and triticale)	0.056	0.165
Orchards (orange, apple, plum, fig, loquat, cherries, blackberry, raspberry)	0.013	0.078
Forages Irrigated (ryegrass, lucerne, silage maize, sorghum, vetch)	0.006	0.051
Forages Rainfed (ryegrass, oats, corn, sorghum, lupine)	0.049	0.153
Horticultural (potatoes, carrots, onions, cabbages, beans, chickpeas)	0.017	0.089
Industrial horticulture (tomato and pepper)	0.011	0.092
Oilseeds (sunflower and rapeseed)	0.01	0.067
Pastures (temporary grass and permanent grasslands)	0.511	0.41
Fallows	0.043	0.146
Olive groves Irrigated	0.034	0.156
Olive groves Rainfed	0.171	0.291
Vineyards	0.034	0.145
Walnuts and almond trees	0.003	0.048
Stone pine	0.009	0.079
Other dry fruits (hazelnut, chestnut, pistachios, carob)	0.001	0.019
Cork oak cover	0.149	0.265
Holm oak cover	0.111	0.229
Livestock variables (proportion in total LU)		
Cattle grazing	0.168	0.34
Cattle stabled	0.003	0.04
Fattening cattle grazing	0.018	0.054
Fattening cattle stabled	0.002	0.037
Sheep grazing	0.205	0.386
Goat grazing	0.024	0.131
Dairy cows	0.004	0.047
Pigs grazing	0.008	0.076
Livestock density (LU/ha UAA) (includes all farm animals, added-up in LU)	0.526	3.506

apply. In these cases, we decided to regroup these parcels into new (sub) farms by forcing the distance between them not to exceed 25 km, which increased the total number of farms to 28,739. This decision also helped to narrow down the range of biophysical variability within each farm, and thus to better link farm units to their biophysical context, described in the next section. We also discarded farms with total area equal or below 2 ha (4409 farms, representing less than 1% of total UAA) because the land use in smaller farms is likely to be highly sensitive to crop rotations, which cannot be properly captured with one-year data.

The raw data included 129 land use/cover categories, which were simplified by aggregation into broader categories, while maintaining the distinction between irrigated and rainfed crops, when applicable (e.g., irrigated and rainfed cereals). We also included two variables describing the proportion of the UAA under the cover of cork and holm oaks, respectively, because their presence is prone to influence farm management, as the first is a major source of income for farmers (cork production) and the later provides shade and food (acorns) to livestock grazing, in addition to valuable firewood. These two variables were computed on a geographical information system (GIS) environment by intersecting the farms map (derived from the LPIS spatial data) with digital information on cork and holm oak distribution and computing, for each farm, the share of the UAA covered with both land cover classes.

Livestock numbers were converted into livestock units (LU) using EU standard conversion factors, and these were used to describe the percentage composition of livestock by species, as well as livestock density in each farm. Thus, a set of 28 variables was defined to characterize the land use/cover and livestock patterns for each farm (Table 1).

A principal component analysis (PCA) was performed on a correlation matrix of these 28 variables to reduce variable redundancy and the principal axes with eigenvalues above 1 entered a hierarchical cluster analysis (Ward method) to derive the FS typology. The number of clusters to retain was decided based on a visual analysis of the dendrogram and on expert knowledge of the study area.

To help interpreting the resulting FS, we calculated three variables indicating the level of agricultural intensity, specialization and dependence on labour. The intensity variable was calculated following the EU "standard output" approach (Commission Regulation (EC) No 1242/2008 of 8 December 2008) by estimating the total gross product per land unit (in €/ha UAA) for each farm. The specialization variable was computed as the highest proportion of standard output from a single farm activity. The labour indicator aims to differentiate the FS based on their specific labour needs, in annual work units per land unit (AWU/ha UAA). Due to data limitations, we had to resort to official statistics on the "EU farm typology by economic size and type of farming" (in the sense of the above-mentioned legal text), at NUT2 level (Alentejo) for the year of 2013, from which we extracted the number of annual work units per hectare for each farm type, to be directly associated to each of the resultant FS on a similarity base. Thereby, this indicator was not computed at farm level, but directly at FS level.

2.3. Socioeconomic and biophysical drivers

Potential socioeconomic and biophysical drivers of farming system choice were screened from literature (e.g. Grigg, 2005; Hazell and Wood, 2008; Kristensen et al., 2016; Martel et al., 2019; Plieninger et al., 2016; Reboul, 1989; van Vliet et al., 2015) and the authors' experience from previous studies where similar approaches were applied (Ribeiro et al., 2014, 2018; Silva et al., 2020). Subsequently, each farm was characterized according to a set of socioeconomic and biophysical variables thus identified, considered as potential drivers of FS spatial patterns (Table 2). These variables vary spatially but are mostly constant over time (at least for the time scale of most farm management decisions).

Socioeconomic variables included seven farm structure variables (farm and block size, farm fragmentation and dispersion, access to public and private water sources for irrigation, nature conservation

Table 2
Summary statistics for the socioeconomic and biophysical drivers ($n = 23{,}416$ farms).

Variable	Description	Mean	SD	Min	Max
Socioeconomic variables – farm structure variables					
FSIZE	Farm size – Total UAA (ha) (1)	84.09	184.46	2.01	7191.16
BLKSIZE	Average farm-block size (ha) (1)	23.15	45.37	0.20	1109.93
JANUS	Januszewski index (adimensional) (1) (2)	0.65	0.23	0.13	1.00
BLKDIST	Average area-weighted block distances to farm centroids (m) (1)	1571.00	2128.00	0.00	56,951.00
WPRIVATE	Access to water from private ponds or small streams (yes = 1; no = 0) (5)	0.16	0.37	0.00	1.00
WPUBLIC	Proportion of UAA in public irrigation systems (6)	0.15	0.31	0.00	1.00
NATURE	Proportion of UAA included in areas classified for nature conservation (7)	0.22	0.39	0.00	1.00
Socioeconomic variables – local socioeconomic variables					
INCAGRI	Proportion of farms where agriculture is the main household income source (3)	0.23	0.14	0.00	0.84
INCOTH	Proportion of farms where household income is mostly from outside the farm, but not pensions (3)	0.26	0.07	0.00	0.67
PDENS	Population density (inhabitants/km^2) (4)	32.5	77.8	0.89	1084.24
AWU	Number of annual work units (AWU) per km^2 of total parish area (3)	1.96	1.70	0.21	17.95
AWU hired	Proportion of hired work in total labour (3)	0.26	0.15	0.00	0.93
RENT	Proportion of rented land in total UAA (3)	0.18	0.12	0.00	1.00
Biophysical variables					
TMIN	Average minimum temperature in the coldest month 1970–2000 (°C) (8)	4.71	0.59	3.01	8.40
TMAX	Average maximum temperature in the warmest month 1970–2000 (°C) (8)	31.56	1.95	20.24	35.68
PREC	Average annual rainfall 1970–2000 (mm) (8)	592.89	107.28	376.83	1195.51
SDEPTH	Soil depth (cm) (5)	52.74	29. 80	0.00	150.00
SMOOTH	Proportion of UAA with smooth slopes (<5%) (5)	0.51	0.32	0.00	1.00
MODERATE	Proportion of UAA with moderate slopes (5–16%) (5)	0.38	0.24	0.00	1.00

(continued on next page)

Table 2 (*continued*)

Variable	Description	Mean	SD	Min	Max
STEEP	Proportion of UAA with steep slopes (>16%) (5)	0.11	0.19	0.00	1.00
HEAVY_S	Proportion of UAA with heavy texture soils (5)	0.33	0.37	0.00	1.00
MEDIUM_S	Proportion of UAA with medium texture soils (5)	0.42	0.38	0.00	1.00
LIGHT_S	Proportion of UAA with light texture soils (5)	0.24	0.36	0.00	1.00
VERYACID	Proportion of UAA with very acid soils (pH < 5) (5)	0.27	0.33	0.00	1.00
ACID	Proportion of UAA with acid soils (5 < pH < 6) (5)	0.41	0.38	0.00	1.00
NEUTRAL	Proportion of UAA with pH neutral soils (6 < pH < 7) (5)	0.21	0.30	0.00	1.00
ALKALINE	Proportion of UAA with alkaline soils (pH > 7) (5)	0.11	0.24	0.00	1.00

Sources: (1) Computed from LPIS data; (2) Farm spatial fragmentation index, varying from 0 to 1 with higher values indicating a higher degree of farmland consolidation (Januszewski, 1968); (3) Agricultural census 2009 - parish level; (4) Population census 2011 - parish level; (5) EPIC WebGIS Portugal (http://epic-webgis-portugal.isa.ulisboa.pt/); (6) DGADR - Direção-Geral de Agricultura e Desenvolvimento Rural (http://sir.dgadr.gov.pt/expl-alentejo); (7) ICNF – Instituto de Conservação da Natureza e das Florestas (http://www2.icnf.pt/portal/pn/ap); (8) IPMA - Instituto Português do Mar e da Atmosfera (https://www.ipma.pt/pt/oclima/normais.clima/).

constraints on farm use), and six local context variables computed from official statistics at the administrative parish level (one demographic variable, population density, and five agricultural variables, e.g. AWU availability or the share of rented UAA); all farms in the same parish where assigned the same value in these variables; when farms had areas in more than one parish, these variables were computed through average-weighting by farm-area shares in each parish. Biophysical variables included three climatic variables (describing temperature and precipitation), eight soil quality variables (describing soil depth, texture and pH) and three topographic variables (slope categories). (Table 2).

Values for explanatory variables were derived for each farm using a GIS (maps for explanatory variables are provided in supplementary information, Annex I). Farms with missing values resulting from map mismatches were discarded, dropping the number of valid observations to 23,416 farms.

2.4. Model design

We developed a random forest FS choice model to explore the farm-level relationships between the typologies of FS derived from cluster analysis and the socioeconomic and biophysical variables. Random forest is a popular machine learning method that can be used both for regression and classification, and is well-suited for high dimensional data (Strobl et al., 2009). Random forest use bootstrap and aggregation (bagging), building multiple decision trees based on random subsets of the data and using a random subset of predictor variables candidates for each node, in each decision tree (Liaw and Wiener, 2002). On a classification problem, each observation is assigned to a class according to the majority of votes from all trees. Both the number of trees and the number of predictor variables sampled for each node are user-defined and can be used to tune the model. The mean out-of-bag (OOB) error rate computed across all trees provides a measure of model prediction accuracy (Breiman, 2001). Random forests have been widely used in many scientific fields and have proved to be one of the best machine learning techniques currently available, including for predictive modelling of spatial and spatio-temporal data (Hengl et al., 2018).

2.4.1. Explaining spatial distribution of farming systems

Since we were firstly interested in exploring causal theories on the main drivers of FS spatial distribution, rather than using the model to make predictions on new data (e.g. to assess scenarios of policy or climate change), we tuned the model to optimize its average prediction accuracy across FS, rather than maximizing the overall prediction power, by testing different stratified sampling approaches to deal with anticipated unbalanced data (high variance in group sizes) (see details of model parametrization in supplementary information – Annex II). At this stage, model overfitting should not be an issue, since the focus was on explaining our training data, rather than the generalization of the model (Shmueli, 2010).

With this modelling outset, all FS are assumed to be competing simultaneously for each farm and the choice is made dependent only on variables that vary in space, while keeping constant the effect of temporal variables (such as prices or policies). The effect of these temporal variables on the choices observed in the study year cannot be estimated, as we only have one observation on FS choice for each farm, that is: the choice observed in the study year 2017.

We used variable importance measures to assess the relevance of each predictor variable in the model and their marginal effect on each FS was examined using partial dependence plots (Friedman, 2001) (supplementary information – Annex II). We investigated the shape of the partial dependence plots fitted functions for each class of the dependent variable (that is, for each FS) to infer their role as drivers or constraints for each FS. In addition, we computed the correlation coefficient between the level of farming intensity characterizing each FS with the corresponding prediction accuracy rate obtained by the model, to test the hypothesis of a positive relationship between the levels of this indicator and the degree of FS dependence on socioeconomic and biophysical drivers.

All statistical analyses were carried out in R 3.4.1 (R Development Core Team, 2017).

2.4.2. Predicting spatial patterns of farming systems

On a following step, we focused on exploring the predictive capacity of the model in the choice of the FS, based on the socioeconomic and biophysical variables described above. Since we were mostly interested in predicting FS choice at the landscape-scale rather than at farm-scale, taking into account the importance of landscape patterns for biodiversity and public goods delivery, we focused the analysis on the model's ability to predict FS spatial patterns at a scale comparable to that of the landscape (Andersen, 2017). For this purpose, the study area was divided into a random network of hexagons of about 54,125 ha each, corresponding to a hexagon apothem of 12.5 km which was chosen with reference to the 25 km threshold used to define the farms. These hexagons were then used as analysis units to compare, for each hexagon, the percentage distribution of the UAA by FS in the observed situation with that predicted by the model. A hexagonal grid was preferred over a square grid because it is less subject to bias from the edge effects when computing landscape metrics (Birch et al., 2007). We rejected all hexagons with more than 66% of the area outside the LPIS data, due to low significance for this purpose. In each hexagon, we calculated the difference between the observed and predicted UAA shares for each FS and computed the half-sum of their absolute values. The average of these results across all hexagons was interpreted as an estimate of the percentage of accuracy obtained in model predictions, that is, the capacity of the model to predict spatial patterns of FS composition at the landscape-scale. In addition, we also computed the determination coefficient (r^2) between the observed and predicted values in each hexagon, taking its mean as a measure of the quality of fit of the model. Model predictions were obtained by running the model on a random

Table 3.a
Farming system description – Land cover composition (average values in proportion to the total UAA; values under 0.01 are omitted; values above 0.5 are in bold).

Farming system	Rice	Cereals Irrigated	Cereals rainfed	Forages Irrigated	Forages Rainfed	Horticultural	Industrial horticulture	Oilseeds	Fallows	Pastures	Fruit trees	Olive groves Irrigated	Olive groves Rainfed	Vineyards	Walnuts and almond trees	Stone pine	Other dry fruits	UAA under Cork oak cover	UAA under Holm oak cover
Cattle grazing – CO*					0.039					**0.884**			0.026					0.329	0.082
Cattle grazing – HO*			0.020		0.035					**0.906**			0.020					0.048	**0.590**
Cattle grazing – forages		0.011	0.143		0.239	0.015		0.020	0.027	0.388		0.045	0.093					0.049	0.086
Grazing goats					0.027					**0.923**			0.023					0.314	0.207
Mixed Cattle and sheep - Irrigated forages		0.034	0.037	0.444	0.112	0.019		0.016	0.018	0.219	0.018	0.015	0.049					0.066	0.028
Sheep grazing – CO*					0.012					**0.936**			0.022			0.012		**0.686**	0.022
Sheep grazing – HO*			0.029		0.025				0.013	**0.891**			0.032					0.051	**0.641**
Sheep grazing - pastures			0.014		0.021					**0.852**			0.085					0.141	0.048
Sheep grazing - pastures and forages			0.169		0.139	0.015			0.027	0.437			0.156	0.030				0.056	0.036
Sheep grazing - forages			0.041		**0.650**				0.033	0.127			0.107					0.068	0.053
Rainfed olive groves with sheep					0.016					0.291			**0.660**	0.010				0.039	
Rainfed olive groves									0.016	0.099	0.013		**0.823**	0.018				0.011	
Irrigated olive groves			0.012						0.022	0.052		**0.770**	0.083		0.021			0.016	0.024
Vineyards		0.010			0.013				0.054	0.066	0.014	0.029	0.093	**0.697**				0.025	0.010
Fruit trees					0.014				0.025	0.243	**0.548**	0.012	0.083	0.020			0.025	0.142	0.040
Stone pine	0.038								0.019	0.189			0.023			**0.713**	0.000	0.249	0.010
Rice	**0.850**	0.012							0.039	0.068								0.024	
Irrigated cereals and horticultural crops		0.300	0.038			0.241	0.242		0.049	0.077								0.018	0.012
Rainfed cereals and oilseeds		0.087	0.300			0.038		0.430	0.063	0.030		0.011	0.026						0.018
Rainfed cereals			0.463		0.016	0.015		0.010	0.171	0.150			0.147					0.031	0.039
Pastures without livestock										**0.785**		0.015	0.152	0.012				0.082	0.088
Fallows			0.079		0.011	0.016			**0.752**	0.037	0.011		0.077					0.048	0.141

* CO – Under cover of cork oak; HO – Under cover of holm oak.

Table 3.b
Farming system description – Livestock composition in livestock-oriented farming systems (average values in proportion to total LU; values under 0.01 are omitted; proportions above 0.5 are in bold) and livestock density.

Farming system	Cattle grazing	Cattle stabled	Fattening steers grazing	Fattening steers stabled	Sheep grazing	Goat grazing	Dairy cows	Pigs grazing	Livestock density (LU/ha)
Cattle grazing – CO*	**0.872**		0.084		0.028				0.799
Cattle grazing – HO*	**0.865**		0.080		0.044				0.677
Cattle grazing – forages	**0.859**		0.083		0.041				0.967
Grazing goats					0.082	**0.910**			1.039
Mixed Cattle and sheep - Irrigated forages	0.480		0.073		0.300	0.069	0.077		0.618
Sheep grazing – CO*	0.144		0.010		**0.799**	0.043			0.245
Sheep grazing – HO*					**0.963**	0.028			0.387
Sheep grazing - pastures					**0.973**	0.017			1.004
Sheep grazing - pastures and forages					**0.792**	0.206			0.711
Sheep grazing - forages	0.049				**0.709**	0.236			0.378
Rainfed olive groves with sheep					**0.974**	0.024			1.212

* CO – Under cover of cork oak; HO – Under cover of holm oak.

test-set of the data with ca. 1/3 of the observations (farms), after estimating it in a train-set with the remaining 2/3.

3. Results

3.1. Farming systems typology

A solution of 30 groups, representing farming systems, was selected from the cluster analysis. As some groups included only a very small number of observations (farms), we anticipated potential problems in the estimation of the predictive model and so we decided to eliminate groups with less than 0.7% of the total number of observations, an arbitrary threshold mostly based on expert judgement. This led to the removal of 8 non-representative FS, comprising 613 farms accounting for 3.1% of total UAA, which were discarded for further analysis. Consequently, the final number of FS was set at 22 (Tables 3a and 3b).

By chance, these FS resulted equally divided into livestock-oriented systems and crop-oriented systems. Both groups include similar shares in number of farms (51.5% and 48.5%, respectively), although farms in livestock-oriented systems cover a much larger share of total UAA (78.2%) denoting they are larger farms, on average.

Within the livestock systems, six are oriented to sheep, three to cattle, one to goats and one is mixed with cattle and sheep. Among the six sheep-oriented systems, two are agroforestry grazing systems, one associated with cork oak and the other with holm oak, a third one is related with open land pastures, a fourth sheep system is mainly dependent on forage crops, the fifth depends both on permanent pastures and forage crops, and the last is mostly a mixed-system combining rainfed olive groves with sheep grazing. The three cattle systems also include two agroforestry grazing systems with permanent pastures under the canopy of cork and holm oaks, respectively, and a third one depending mainly on forage crops. The mixed cattle-sheep system is highly dependent on irrigated forages and the last livestock-oriented system is the goat system, which is also a pasture-dependent grazing

Table 4
Characterization of farming systems according to the levels of farming intensity, specialization and labour needs, average farm size (in hectares of UAA and number of LU) and representativeness (in number of farms, UAA and LU).

	Characterization of farming systems			Average farm size		Representativeness					
						Number of farms		UAA		Livestock Units	
	Intensity (10^3 €/ha)	Speciali-zation (%)	Labour needs (AWU/ha)	UAA (ha)	LU (n.)	Total	%	Total (10^3 ha)	(%)	Total (10^3 LU)	(%)
Cattle grazing – CO*	0.46	84.3	0.005	248.5	128.7	2515	10.6	625	31.1	323.8	36.5
Cattle grazing – HO*	0.31	84.0	0.005	288.2	157.0	1245	5.3	359	17.9	195.5	22.0
Cattle grazing – forages	0.75	68.3	0.005	186.1	88.1	463	2.0	86	4.3	40.8	4.6
Grazing goats	0.94	88.6	0.004	52.6	19.7	251	1.1	13	0.7	4.9	0.6
Mixed Cattle and sheep - Irrigated forages	1.14	75.0	0.005	58.8	24.3	171	0.7	10	0.5	4.1	0.5
Sheep grazing – CO*	0.24	54.6	0.004	89.0	19.0	2346	9.9	209	10.4	44.6	5.0
Sheep grazing – HO*	0.28	64.9	0.004	84.9	23.5	1391	5.9	118	5.9	32.6	3.7
Sheep grazing - pastures	0.71	84.1	0.004	54.5	22.4	1461	6.2	80	4.0	32.7	3.7
Sheep grazing - pastures and forages	0.79	70.6	0.004	52.8	18.7	745	3.1	39	2.0	13.9	1.6
Sheep grazing - forages	0.46	69.9	0.004	25.3	4.3	848	3.6	21	1.1	3.7	0.4
Rainfed olive groves with sheep	0.91	74.1	0.006	12.0	9.6	774	3.3	9	0.5	7.4	0.8
Rainfed olive groves	0.30	92.3	0.010	10.6	0.1	2626	11.1	28	1.4	0.3	0.0
Irrigated olive groves	1.45	93.0	0.023	82.4	0.8	864	3.6	71	3.5	0.7	0.1
Vineyards	1.84	90.3	0.050	24.3	0.5	928	3.9	23	1.1	0.4	0.0
Fruit trees	12.59	89.9	0.036	25.7	1.5	325	1.4	8	0.4	0.5	0.1
Stone pine	4.63	97.6	0.009	68.0	1.4	221	0.9	15	0.7	0.3	0.0
Rice	1.70	93.8	0.018	52.8	4.0	314	1.3	17	0.8	1.3	0.1
Irrigated cereals and horticultural crops	4.66	90.9	0.259	53.7	1.8	1070	4.5	57	2.9	1.9	0.2
Rainfed cereals and oilseeds	0.98	88.2	0.006	65.7	0.9	421	1.8	28	1.4	0.4	0.0
Rainfed cereals	0.42	82.0	0.010	33.8	0.4	1537	6.5	52	2.6	0.6	0.1
Pastures without livestock	0.20	61.5	0.003	48.5	8.5	2602	11.0	126	6.3	22.2	2.5
Fallows	0.59	54.1	0.003	20.8	0.0	582	2.5	12	0.6	0.0	0.0
Total	–	–	–	–	–	23,700	100	2007	100	733	–

* CO – Under cover of cork oak; HO – Under cover of holm oak.

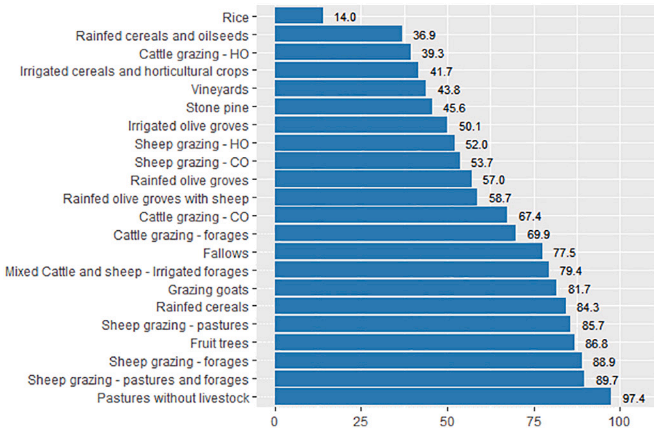


Fig. 2. Classification error rates for the 22 farming systems (values in %).

system (Tables 3a and 3b).

Among the crop-oriented systems, five are dedicated to permanent crops, four to annual crops and the last two refer to special situations, one including farms without livestock but with almost all UAA under pasture, probably yearly rented to neighbours with cattle, and the other encompassing farms with almost all UAA set to fallow. The permanent crops systems included two systems dedicated to olive groves, one of which was irrigated and the other rainfed, one to vineyards, another to fruit trees and the last one to stone pines (for pine nut production). The annual crops systems included two rainfed systems, one dedicated to cereals and the other to cereals and oilseeds, one dedicated to irrigated cereals and horticultural crops, and the last one to rice production (Tables 3a and 3b).

The average farming intensity across the 22 FS is about 1650 €/ha, with the Fruit trees system as the most intensive, reaching ca. 12,600 €/ha, and 15 systems below 1000 €/ha. Agricultural specialization is relatively high, with more than half of the FS earning more than 80% of their standard output from a single activity. Average farm specialization is higher in crop systems than in livestock systems (85% and 74%, respectively), where most systems earn more than 90% from a single activity. Average labour needs are also higher in crop systems than in livestock systems (0.039 and 0.004 AWU/ha, respectively, i.e. nearly 10 times more), with a maximum of 0.259 AWU/ha found in the Irrigated cereals and horticultural crops system and a minimum of 0.003 AWU/ha found in systems Pastures without livestock and Fallows (Table 4).

Average farm size varies significantly across FS, with values going from ca. 11 ha in both rainfed olive grove systems (with and without sheep) until over 200 ha, in cattle grazing – HO and – CO systems (288 and 249 ha, respectively) (Table 4).

Almost 1/3 of all farms are included in only three FS, all with more than 2500 farms (systems Rainfed olive groves, Pastures without livestock and Cattle grazing – CO). However, nearly 1/3 of total UAA is concentrated in one single FS, the Cattle grazing – CO. The three cattle-oriented FS comprise more than half of the total UAA (53.3%) (Table 4).

3.2. *Spatial determinants of farming system choice*

The tuning of the random forest model led to a 500 trees model, with 5 variables randomly sampled as candidates at each split and using the "sampsize" option to correct size differences across the FS categories (see details in Annex II – supplementary information). The classification error rates for each of the 22 FS ranged from 14.0% in the Rice system to 97.4% in the Pastures without livestock system (Fig. 2), with an average of 63.7% across all FS, a value that should be evaluated positively considering the high number of classes in the dependent variable (22 FS, for which the random error rate would be about 95.4% with balanced data).

The relative importance of socioeconomic and biophysical variables was very similar, and among the top ten variables, in terms of mean decrease accuracy (Annex II – supplementary information), six are socioeconomic and four are biophysical. The farm physical dimension variables (FSIZE and BLKSIZE) and a local context of high dependence on family income in agriculture (INCAGRI) proved to be the most relevant socioeconomic factors influencing the choice of FS, while in the biophysical variables the most important were the climatic variables (Fig. 3). The variable indicating access to surface water sources (WPRIVATE) was found to be the least important, either in the global model or in most of the class-specific models.

Farm size (FSIZE) and average farm-block size (BLKSIZE) were the most relevant variables for the choice of Cattle grazing FS, positively influencing its choice (Fig. 3). The same variables also have a relevant effect on most sheep systems but, in this case, predominantly on the opposite direction (Fig. 3). The choice of the Cattle grazing – CO system is positively influenced by the increase of the average annual rainfall (PREC) and negatively by high summer temperatures (TMAX), which has a positive effect on the choice of the Cattle grazing – HO system. The Cattle grazing – forages system is distinguished by a preference for warmer winters (Fig. 3).

The Grazing goats system is positively related to sloping terrain; its choice is favoured by increasing the slope (STEEP), while avoiding flat land (SMOOTH). This system is also characterized by avoiding public irrigation areas (WPUBLIC) and deep soils (SDEPTH). The choice of the Mixed Cattle and sheep - Irrigated forages system is favoured by deeper soils, public irrigation systems (WPUBLIC) and high local labour availability (AWU), which is probably related to the irrigated forages component of this FS or with the labour needs associated with grazing herds. The average annual rainfall (PREC) has opposite effects in Sheep grazing – HO and – CO systems, with the first system being favoured by lower rainfall values, and the other way around in the later system. Sheep grazing – CO is also favoured by areas with steeper slopes (STEEP) and light soils (LIGHT_S), while the choice of Sheep grazing – HO decreases with deeper soils, smoother terrain and public irrigation structures. Lower values of local labour availability (AWU) seem to promote the choice of the Sheep grazing – pastures system, while the choice of Sheep grazing – forages system is negatively influenced as the local values of agricultural income dependence (INCAGRI) raises (Fig. 3).

Both Rainfed olive groves systems (with and without sheep) are strongly related to smaller farm sizes, as these are also the two systems with lower average UAA (Table 4). Both are positively related to high summer temperatures (TMAX) and negatively to higher regional values of agricultural income dependence. The Rainfed olive groves with sheep system is favoured when average annual rainfall increases, and the Rainfed olive groves system is positively related to neutral pH soils (NEUTRAL) (Fig. 3).

The Irrigated olive groves system is positively related to high summer temperatures, public irrigation systems, high local labour availability and high average farm-block size. It is negatively related to high average annual rainfall. The choice of the Vineyards system tends to increase with higher values of regional labour availability, public irrigation systems and population density (PDENS). The Fruit trees system is positively associated with average annual rainfall and negatively with high population density and warmer winters (TMIN). The choice of the Stone pine system is favoured by light soils (LIGHT_S) and discouraged by high summer temperatures and population density (Fig. 3).

In the annual crops, the Rice system is mostly favoured by the presence of public irrigation systems, also by higher regional values of agricultural income dependence (INCAGRI) and soil depth, while negatively influenced by high summer temperatures. The Irrigated cereals and horticultural crops system is positively related to soil depth, regional labour availability and smooth slope terrain. The choice of the Rainfed cereals and oilseeds system is encouraged with public irrigation systems and higher values of soil depth, neutral pH and high summer temperatures. The Rainfed cereals system is negatively related to bigger

	MODEL OVERALL	Cattle grazing - CO	Cattle grazing - HO	Cattle grazing - forages	Grazing goats	Mixed Cattle and sheep - Irrigated forages	Sheep grazing - CO	Sheep grazing - HO	Sheep grazing - pastures	Sheep grazing - pastures and forages	Sheep grazing - forages
FSIZE	65,6	↑ 59,5	↑ 55,2	↑ 41,5	↕ 5,8	↓ 4,7	↑ -0,7	↑ 6,0	↕ 12,6	↕ 12,4	↓ 15,6
BLKSIZE	50,9	↑ 37,9	↑ 32,9	↑ 19,3	↕ 6,4	↑ 3,1	↑ 28,5	↑ 11,3	↓ 15,5	↓ 14,2	↓ 15,2
JANUS	32,5	↑ 3,3	↑ 6,0	↓ 13,9	↑ 4,3	↕ 1,3	↑ 7,0	↑ 5,8	↕ 7,4	↓ 10,6	↑ 4,8
BLKDIST	30,9	↕ 6,7	↑ 7,0	↕ 5,6	↕ 0,1	↕ 1,1	↑ 3,6	↕ 5,5	↓ 6,3	↕ 2,1	↓ 4,8
INCAGRI	52,7	↕ 18,4	↑ 5,4	↓ 13,6	↓ 1,4	↕ 4,4	↑ 14,0	↕ 26,8	↓ 14,3	↓ 11,6	↓ 12,6
RENT	47,3	↑ 12,6	↑ 6,1	↑ 7,9	↕ 2,1	↕ 3,8	↑ 8,7	↑ 17,8	↕ 10,3	↕ 5,3	↕ 11,7
INCOTH	47,3	↓ 14,8	↕ 8,7	↑ 5,4	↕ 4,8	↕ 4,6	↓ 14,6	↕ 15,8	↕ 12,3	↕ 6,2	↑ 11,3
PDENS	46,0	↕ 16,3	↓ 6,4	↓ 5,4	↓ 1,1	↑ 5,0	↓ 14,8	↓ 23,5	↕ 11,7	↓ 5,8	↕ 12,1
AWU	45,7	↓ 15,7	↓ 18,4	↓ 10,3	↓ 4,9	↑ 5,9	↓ 20,6	↓ 21,7	↓ 19,2	↓ 10,5	↕ 10,2
WPUBLIC	43,7	↓ 28,2	↓ 20,0	↓ 5,3	↓ 14,6	↑ 6,4	↓ 24,1	↓ 27,0	↓ 20,6	↓ 3,0	↕ 8,2
AWU_hired	43,6	↑ 11,9	↕ 11,2	↓ 9,0	↕ 3,1	↕ 3,7	↑ 14,1	↕ 17,3	↕ 7,8	↓ 7,8	↓ 8,8
NATURE	27,0	↑ 5,5	↕ 8,8	↑ 11,8	↕ 4,0	↕ 6,2	↕ 6,4	↓ 6,0	↓ 6,0	↕ 8,4	↓ 6,8
WPRIVATE	8,7	↑ 0,8	↑ 3,0	↑ -0,7	↓ 4,0	↑ 0,9	↓ -1,5	↑ 1,3	↓ 1,7	↓ 1,2	↓ 1,8
TMAX	56,4	↓ 32,0	↑ 21,0	↕ 16,5	↓ 12,2	↕ 6,8	↓ 31,7	↑ 25,6	↕ 11,7	↑ 9,7	↓ 13,6
TMIN	56,2	↑ 18,7	↑ 9,8	↑ 15,6	↕ 3,5	↓ 5,4	↕ 20,0	↕ 26,0	↕ 10,4	↕ 8,0	↕ 10,1
PREC	48,1	↑ 25,7	↓ 23,7	↓ 20,1	↑ 11,5	↑ 5,4	↑ 32,2	↓ 33,3	↑ 14,7	↕ 12,9	↑ 13,2
SDEPTH	47,3	↓ 32,2	↓ 13,8	↓ 10,5	↓ 15,9	↑ 6,2	↓ 23,4	↓ 33,2	↓ 9,1	↓ 9,0	↓ 5,3
LIGHT_S	41,5	↑ 15,8	↕ 19,0	↓ 14,2	↑ 7,0	↑ 5,8	↑ 26,4	↓ 18,5	↓ 8,2	↓ 9,2	↓ 2,4
ACID	41,2	↑ 20,8	↑ 12,5	↓ 7,7	↑ 5,9	↓ -0,4	↑ 13,5	↑ 12,0	↕ 6,0	↕ 4,0	↓ 4,3
NEUTRAL	40,9	↓ 14,2	↓ 10,8	↓ 10,9	↓ 9,7	↑ 2,0	↓ 15,9	↓ 24,3	↓ 8,9	↓ 5,0	↓ 4,8
MEDIUM_S	40,1	↕ 13,1	↑ 14,4	↕ 7,3	↕ 4,1	↓ 3,5	↕ 12,5	↑ 16,6	↑ 5,7	↕ 6,8	↕ 1,3
VERYACID	38,1	↑ 13,5	↑ 7,9	↕ 5,9	↑ 2,9	↑ 4,7	↑ 17,7	↑ 16,6	↕ 6,8	↕ 6,8	↓ 5,3
SMOOTH	37,7	↓ 20,3	↓ 13,4	↓ 10,5	↓ 14,6	↑ 3,6	↓ 22,7	↓ 26,7	↓ 10,6	↓ 6,4	↕ 5,9
HEAVY_S	34,2	↓ 14,3	↓ 10,5	↕ 5,7	↓ 4,9	↑ 5,9	↓ 17,2	↕ 15,3	↓ 10,4	↕ 6,1	↕ 5,8
STEEP	31,8	↑ 19,2	↑ 10,7	↓ 9,4	↑ 19,3	↓ 0,8	↑ 27,4	↑ 25,5	↑ 2,2	↕ 6,8	↓ 4,0
ALCALINE	31,7	↓ 12,1	↓ 11,5	↓ 8,6	↓ 3,5	↑ 4,2	↓ 18,1	↓ 16,6	↓ 9,4	↓ 1,1	↓ 8,1
MODERATE	29,9	↑ 18,0	↑ 12,0	↑ 6,1	↑ 5,2	↓ 2,9	↑ 18,0	↑ 18,4	↑ 11,2	↑ 3,0	↕ 4,6

	Rainfed olive groves with sheep	Rainfed olive groves	Irrigated olive groves	Vineyards	Fruit trees	Stone pine	Rice	Irrigated cereals and horticultural crops	Rainfed cereals and oilseeds	Rainfed cereals	Pastures without livestock	Fallows
FSIZE	↓ 29,9	↓ 54,8	↓ 19,6	↓ 23,5	↓ 13,0	↕ 9,0	↑ 15,7	↓ 12,2	↑ 13,9	↓ 14,6	↕ 9,7	↓ 29,1
BLKSIZE	↓ 42,5	↓ 50,2	↑ 25,9	↓ 27,1	↓ 7,1	↑ 19,4	↕ 19,4	↓ 14,3	↓ 19,8	↓ 29,0	↕ 6,9	↓ 18,2
JANUS	↕ 5,3	↑ 7,8	↑ 6,3	↕ 12,5	↑ 5,3	↑ 3,5	↓ 8,5	↕ 12,5	↓ 21,4	↕ 3,5	↑ 2,4	↑ 12,8
BLKDIST	↓ 7,1	↓ 11,6	↓ 9,5	↕ 14,0	↓ 6,5	↑ 1,5	↑ 2,3	↑ 12,4	↑ 14,8	↕ 3,1	↕ 4,5	↓ 9,3
INCAGRI	↓ 23,9	↓ 30,2	↓ 17,9	↓ 27,0	↓ 10,3	↕ 29,6	↑ 27,3	↑ 16,6	↕ 24,7	↕ 18,3	↓ 5,8	↓ 26,2
RENT	↕ 11,6	↕ 18,5	↑ 17,7	↓ 23,1	↓ 8,9	↓ 11,8	↕ 15,5	↓ 16,7	↕ 14,3	↕ 11,6	↕ 7,0	↓ 17,1
INCOTH	↕ 10,6	↕ 18,8	↑ 13,8	↕ 20,6	↑ 7,6	↕ 19,1	↓ 16,1	↕ 14,9	↑ 16,0	↕ 10,3	↕ 5,6	↑ 19,6
PDENS	↑ 20,4	↕ 19,6	↓ 16,4	↑ 27,3	↓ 10,7	↓ 20,4	↕ 13,3	↑ 20,4	↑ 21,6	↓ 15,2	↓ 5,3	↓ 25,4
AWU	↑ 14,8	↑ 27,2	↑ 27,0	↑ 42,0	↕ 9,6	↓ 23,3	↕ 17,0	↑ 27,6	↑ 18,1	↑ 20,7	↓ 11,2	↑ 23,5
WPUBLIC	↕ 20,4	↑ 17,2	↑ 43,4	↑ 29,2	↕ 9,6	↓ 18,6	↑ 32,2	↑ 18,8	↑ 42,5	↕ 8,2	↓ 10,3	↕ 4,7
AWU_hired	↕ 9,3	↓ 18,2	↓ 14,4	↓ 19,7	↓ 5,4	↑ 12,4	↑ 12,0	↓ 18,2	↕ 19,0	↓ 14,7	↓ 6,8	↓ 15,0
NATURE	↕ 6,2	↓ 9,8	↓ 12,2	↓ 12,0	↕ 7,6	↕ 6,1	↑ 9,6	↓ 10,5	↓ 12,2	↓ 9,6	↓ 2,4	↓ 11,0
WPRIVATE	↓ 4,6	↓ 8,5	↓ 4,2	↓ 6,2	↓ 2,0	↑ 3,3	↑ 3,4	↑ 0,9	↑ 3,3	↓ 3,5	↓ 1,9	↓ 2,8
TMAX	↑ 22,4	↑ 36,8	↑ 43,7	↑ 28,1	↕ 15,9	↓ 26,2	↓ 31,6	↕ 25,9	↑ 40,2	↑ 28,4	↑ 9,0	↑ 25,6
TMIN	↕ 16,9	↓ 22,2	↓ 19,4	↕ 26,4	↓ 10,3	↕ 14,4	↕ 15,4	↓ 19,7	↓ 19,2	↕ 15,3	↕ 7,3	↕ 18,4
PREC	↑ 22,4	↕ 22,0	↓ 25,7	↑ 24,4	↑ 17,7	↑ 20,5	↑ 22,4	↕ 23,6	↓ 22,3	↓ 19,4	↕ 6,9	↓ 23,3
SDEPTH	↕ 16,7	↕ 20,0	↓ 14,0	↕ 22,7	↓ 5,5	↕ 27,9	↑ 25,7	↑ 35,9	↑ 34,5	↓ 14,9	↓ 8,5	↓ 15,7
LIGHT_S	↕ 13,2	↓ 17,1	↓ 16,9	↓ 14,3	↑ 9,7	↑ 26,8	↑ 22,8	↕ 15,8	↓ 20,6	↓ 16,7	↓ 4,0	↓ 16,6
ACID	↓ 9,4	↓ 24,6	↓ 17,1	↓ 18,8	↕ 3,4	↑ 8,7	↑ 10,6	↓ 12,0	↓ 20,0	↓ 10,7	↕ 2,9	↓ 11,2
NEUTRAL	↑ 11,1	↑ 28,6	↑ 21,9	↑ 19,7	↓ 9,7	↓ 15,6	↓ 19,0	↑ 15,6	↑ 34,3	↕ 12,0	↓ 0,8	↓ 10,9
MEDIUM_S	↓ 10,4	↓ 11,8	↑ 8,2	↓ 17,6	↓ 4,9	↕ 16,0	↕ 13,0	↓ 10,2	↑ 17,3	↑ 9,2	↓ 5,9	↕ 12,3
VERYACID	↓ 9,7	↓ 11,7	↓ 9,8	↓ 10,6	↑ 2,5	↑ 11,2	↓ 10,3	↓ 11,7	↓ 12,3	↕ 4,7	↓ 3,5	↑ 11,8
SMOOTH	↓ 12,9	↓ 14,2	↕ 7,9	↑ 18,9	↓ 7,4	↓ 13,2	↑ 20,1	↑ 24,8	↑ 19,7	↓ 11,9	↓ 3,0	↓ 16,1
HEAVY_S	↑ 11,6	↑ 15,6	↑ 13,0	↑ 16,3	↓ 10,2	↓ 20,6	↕ 20,4	↑ 16,3	↑ 17,3	↑ 15,8	↓ 5,1	↑ 14,6
STEEP	↑ 3,1	↑ 11,8	↓ 14,2	↓ 14,5	↑ 5,8	↑ 9,2	↓ 9,9	↓ 9,4	↓ 15,2	↓ 12,9	↕ 5,9	↓ 13,1
ALCALINE	↕ 9,7	↑ 11,0	↕ 7,9	↕ 12,0	↓ 0,6	↓ 12,6	↑ 15,1	↑ 19,8	↑ 12,3	↓ 12,4	↓ 5,7	↓ 9,2
MODERATE	↑ 10,0	↑ 7,1	↑ 5,2	↕ 14,2	↑ 4,4	↕ 7,7	↓ 17,8	↓ 22,1	↓ 13,4	↑ 4,6	↕ 0,3	↑ 11,2

Fig. 3. Variable importance for the overall model and for each farming system. Socioeconomic farm structure variables in blue; local-socioeconomic variables in orange; biophysical variables in green. Variables ordered by decreasing variable importance in the overall model and within each sub-group. Symbols ↑, ↓ and ↕ indicate whether the marginal effect of the variable in each farming system is mostly positive, negative or non-monotonic, respectively, based on the shape of the fitted function on the partial dependence plots (partial dependence plots are provided in supplementary information, Annex IV). Variable description in Table 2.

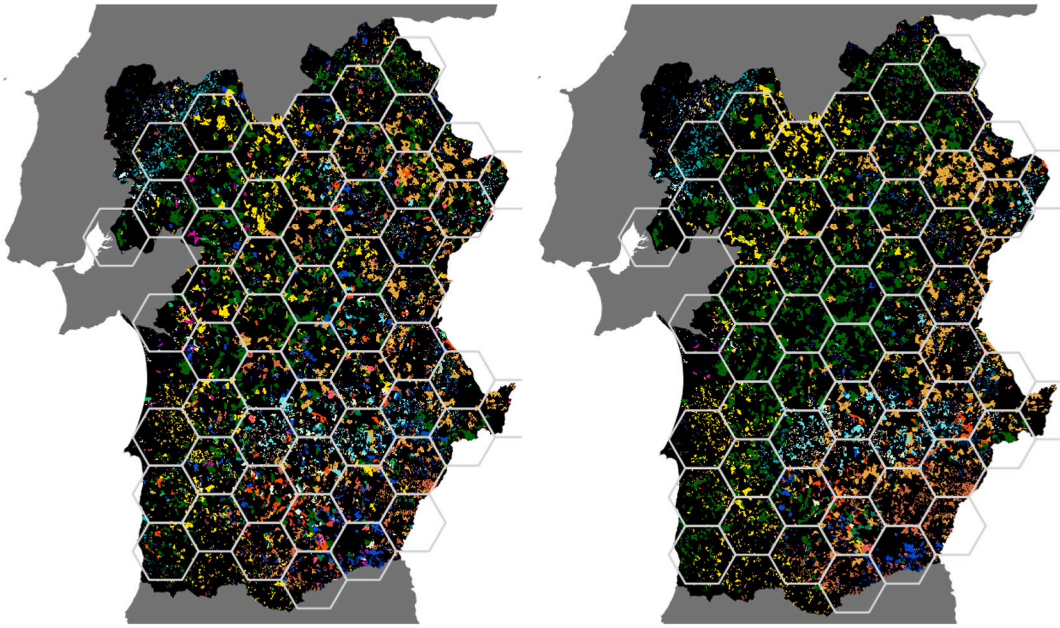


Fig. 4. Observed (left) and predicted (right) FS maps for the 1/3 observations used in the model validation dataset and the hexagons network used to assess model accuracy in FS spatial patterns prediction (different colours identify distinct FS; detailed maps showing the spatial distribution of each farming system are provided in supplementary information, Annex III).

farm-block sizes and average annual rainfall. The Pastures without livestock system seems to be promoted when labour availability is lower and outside public irrigation systems, although this FS presented the highest error rate (Fig. 2). The Fallows system also displays complex relations with the predictors, though it seems to be more positively associated with small farms and areas of low population density (Fig. 3).

Finally, the prediction accuracy for the different farming systems (Fig. 2) showed a modest but positive correlation with the corresponding levels of agricultural specialization and labour needs (Table 4) (correlation coefficients of 0.44 and 0.26, respectively), and a virtually non-existent relationship with the level of agricultural intensity (correlation coefficient − 0.03).

3.3. Spatial patterns of landscape-scale farming systems composition

The hexagonal lattice resulted with 56 usable analysis units, i.e., hexagons with >33% of the area overlapped with LPIS data (Fig. 4). The average error rate in the FS spatial pattern predictions across all hexagons was 28.7% (max. 47.3%; min. 9.2%), which is substantially lower than the error rate obtained with model predictions at the farm-level (67.3%). The average coefficient of determination was 0.89 (max. 1.00; min. 0.28), revealing a good model fit.

4. Discussion

The use of farm-level data (IACS) provided by the national CAP paying agency proved to be a suitable approach to derive the FS typology for the study area, in line with previous studies (Ribeiro et al., 2014, 2016, 2018). The spatial-explicit nature of these data (LPIS) allowed a very fine characterization of farms, including in their biophysical, structural and socioeconomic features. As expected, the extent and heterogeneity of the study area, in both socioeconomic and biophysical features, led to a broad typology of 22 farming systems, which are a direct outcome of distinct farm-management adaptive-responses to a variety of farm features and contexts.

Although the FS typology was balanced in terms of crop- and livestock-oriented systems, the results showed that most of the study area is currently devoted to livestock systems, particularly cattle grazing. Although the present study does not allow this to be confirmed, farmers' preference for these systems may be due to an (at date) ongoing direct payment for suckler cows (and partially to sheep and goats), a national agricultural policy option taken under the 2003 CAP reform that significantly impacted FS dynamics in the region (Ribeiro et al., 2014).

4.1. Farm structure drivers

Many of the effects of structural socioeconomic variables observed here are consistent with those of previous studies. For example, the farm-size was found to positively influence the choice of extensive livestock systems over crop systems, which was also observed in Ribeiro et al. (2018), and also in the choice between cattle grazing over some sheep grazing specialized systems, which was also observed in studies by Ribeiro et al. (2014).

Access to private sources of surface irrigation water showed very little importance in the FS choice-models, which is apparently odd for a region where water is often a limiting factor. This was probably due not only to the type of variable used (dummy variable, with 1 = "yes, the farm has access to surface water sources" and 0 otherwise) but also to the fact of not including access to groundwater from water wells, due to lack of data, which are a common source in parts of the region. In contrast, water availability from public irrigation systems is essential in explaining the spatial location of several irrigated FS (either cereals, oil seeds or intensive olive groves and vineyards) showing the importance of public water management policy over other biophysical constraints (Kahil et al., 2015). Not surprisingly, these farming systems most associated with large public irrigation systems are among the most intensive ones.

Public intervention in nature conservation areas seems to be of little relevance for FS choice since although a considerable share of agriculture area is classified for nature conservation, the corresponding variable (NATURE) was one of the least relevant within a list of dimensions that has farm and block size at the top.

An interesting side-result of our approach was the insight of an overall negative, though moderate, relationship between farm size and the level of agricultural intensity, indicating that larger farms tend to adopt less intensive FS, a finding that goes back to earlier works (Cornia, 1985; Grigg, 2005; Reboul, 1976, 1989). Exceptions, however, can be found when contrasting, e.g., the Rainfed olive groves and the Irrigated olive groves systems, where large investments in fixed capital (including irrigation systems), together with labour availability, seem to provide increasing returns to scale, which was also reported in more recent studies (Deininger et al., 2018; Rada and Fuglie, 2019).

4.2. Socioeconomic context drivers

Regarding the socioeconomic context of the farms, the level of agricultural professionalization (inferred from the INCAGRI variable) and farm labour availability proved to be significant drivers of FS. On one side, higher levels of professionalization, which in Portugal are considerably low in average when comparing to non-South European countries (Arnalte-Alegre and Ortiz-Miranda, 2013), are positively associated with Rice, Stone pine or Rainfed cereals and oilseeds systems. On the other side, Vineyards and Irrigated cereals and horticultural crops, which show the highest levels of labour intensity per hectare and the highest average of labour units per farm, are positively associated with local availability of farm labour. Considering that horticultural crops typically have the highest wage labour ratios compared to other crops (Baptista and Rolo, 2017), it was surprising that it did not show up associated with high local proportion of hired labour. A possible explanation may be the high geographic mobility of hired workers (Baptista and Rolo, 2017), although it may also emerge from the heterogeneity in labour intensity within this FS, since it encompasses irrigated cereals and industrial horticulture, with considerable levels of mechanization, as well as horticultural crops with very high levels of labour needs.

The fact that local labour availability has a more widespread importance as a FS driver than rural population density, which only stands out in the single case of Vineyards, contradicts the idea of permanent crops and horticulture as able of promoting rural population retention (Egea and Pérez y Pérez, 2016), i.e., it points to the dissociation between farm labour dynamics and local demographics (Baptista and Rolo, 2017). While vineyards remain located in higher populated parishes, following deep-rooted institutional constraints by protected designations of origin, olive groves (either irrigated or rainfed) show no relation with local demographics.

Land renting (RENT) did not appear in the top 5 drivers in any FS, suggesting that the size of the land renting market does not appear to have much effect on the choice of FS in the study area. However, the positive relationship observed between land renting and livestock grazing FS, especially cattle, suggests that these systems, which have experienced marked growth in the region in recent years (Ribeiro et al., 2018), expanded in part at the expense of this tenancy regime.

4.3. Biophysical drivers

As anticipated, biophysical factors related to climate, soils and relief, proved to be strong determinants of FS spatial distribution (Grigg, 2005). Summer heat and annual precipitation came up as the main biophysical drivers of FS spatial distribution in the study area. High summer temperatures seem to favour the choice of olive groves, vineyards, rainfed cereals and cattle grazing systems associated to Holms oak, and to discourage livestock systems associated to Cork oak, Stone pine or Rice systems. Winter cold increases the likelihood of fruit tree systems and the opposite with forage systems.

Deep soils and smooth relief are positive drivers of the Rice and Irrigated cereals and horticultural crops systems. The opposite effect is found towards the Grazing goats system, which is strongly related to stepper slopes. Soil pH did not emerge as a major driver for the distribution of any FS, except for rainfed cereals and olive groves systems which showed a preference for neutral pH soils.

Following Cork and Holm oak distinct preferences for soil and climate (Surová and Pinto-Correia, 2008), livestock systems associated with these two species of oaks were found distributed accordingly: Cork oak-associated systems prevail more to the coast and north of the study area, where summer temperatures are milder, annual rainfall is higher and soils are sandy and light-textured; Holm oak-associated systems are further inland an south, where summers are warmer, annual rainfall is lower and soils are frequently poor and fairly thin.

4.4. Farming system prediction at the farm and landscape levels

Although the model's ability to predict individual FS was quite varied, depending on the FS, when applied to predicting FS patterns at the landscape-level the model revealed a much higher hit rate. The random forest approach applied in the model estimation proved to be a valuable choice, particularly in dealing with such high dimensional data (Strobl et al., 2009). At the landscape level, the model was very effective in predicting farming systems patterns, i.e., the shares of FS composition within hexagon-shaped landscape units. For agricultural landscape planning focused on agroecosystem services provision, this may be the right scale of analysis, since a minimum share of farmland managed under the FS delivering those services should be sufficient to ensure the socially desired level of service, rather than requiring the service to be provided by a specific set of farms over a period of time (Andersen, 2017), as is typically the case with many agri-environment schemes requiring multi-annual contracts with individual farmers.

4.5. Shortcomings of the approach and recommendations for future research

Despite the valuable advantages evidenced by the proposed approach, there is still room for future improvement. Improvements mostly relate to characteristics of the IACS and LPIS datasets and methodological options that are dependent on the geographic context of our study area.

While recognized as having high potential for supporting data driven research, the IACS / LPIS datasets present limitations, such as the lack of information to characterize farmers' socioeconomic profile, or information on complementarity relationships between farms, such as the rental or sale of pastures, which can mislead the computation of farms' stock density. Such information would be valuable to include in the FS choice models.

The fact that the empirical work was carried out in a region where the landscape is largely dominated by agriculture, makes it possible to closely link FS choice with landscape modelling. Where this is not the case, such as many mountain and less favoured regions across the EU, this approach may not deliver the same results, given the smaller share of agriculture in the landscape. Additionally, in such regions a significant part of agriculture is probably outside any CAP support system, so that an approach based on IACS / LPIS data can only partially capture an agricultural reality that is itself marginal at the landscape scale. Paradoxically, these regions often include significant shares of high nature value farmlands at the EU level (Lomba et al., 2014). Nevertheless, it should be worth trying to reproduce the approach in such regions in the future, to test the generalization of the framework.

Because our farm characterization variables report to a single year, the effect of economic or policy variables such as prices or subsidies can only be assumed as underpinning the farmers' choices reflected on the observed 2017 IACS / LPIS data. However, the use of this type of variables in the model, provided that time-series of farm-level data can be made available, would significantly extend the scope of this approach, allowing its use to evaluate policy and price change scenarios. Even without additional temporal data, the framework can take advantage of the wide extension of the study area to perform, e.g., climate-change scenarios assessment, by adopting a space-for-time substitution approach.

The selection of candidate variables to be tested as drivers of FS choice is also a key step in the modelling approach. The misspecification or the absence of key variables can substantially undermine models' performance. The problems observed with variable WPRIVATE may be one such case, as this variable only reported access to small private surface water sources, which are mostly torrential regime in this region, with insufficient water guarantees to encourage investing in irrigation systems, and not taking into account that a significant portion of private irrigation in this region is probably resorting to groundwater sources. This premise, which we could not test due to lack of data, would be worth further investigation, should spatially explicit data on groundwater uptakes becomes available.

Another issue deserving further investigation concerns the dimension of the grid of landscape analysis units. It is possible that the size of these units (i.e. the hexagons, in the current case) influences the accuracy of the model, so future investigation focused on determining its optimal size could prove to be of high value.

Also, one aspect that has not been explored in the present study and should merit further investigation is the occurrence of interaction effects between drivers. Although the way random forests deal with these effects is still subject to discussion (Wright et al., 2016), its likely existence recommends additional analysis.

Finally, the fact that the prediction error rate has shown significant disparities across the FS suggests that the choice of some of these FS may be due to effects not measured by the variables examined, including factors related to farmers' desires, attitudes and motivations, or with their socioeconomic profile which, as mentioned above, cannot be assessed on the basis of IACS data. One such case would be the Pastures without livestock system, whose choice is probably mostly determined by the presence of livestock farms in the nearby, with whom the farm can negotiate grazing land renting, rather than by the biophysical characteristics of the farm or its socioeconomic context. On the other hand, FS with lower error rates in the model were those who most depend on the chosen socioeconomic or biophysical factors, such as the Rice, Irrigated cereals and horticulture or Rainfed cereals and oilseed systems (where cereals are an autumn-winter rainfed crop and oilseeds are grown in spring-summer season, often irrigated) that highly depend on irrigation water provided by public irrigation systems in this region. The same applies to the Vineyards system, whose location is highly dependent on the availability of regional labour supply, to meet peaks of labour needs at certain times of the year, related to certain crop operations (e.g. harvesting or pruning). In the present market, policy and technological context, these FS revealed greater dependence on farm structure and "territorial embeddedness" (sensu Cerceau et al., 2018).

4.6. Concluding remarks

Our framework proved to be a suitable approach to investigate the role of human and physical factors in farmers' decisions regarding the choice of the FS, providing effective contributions to improve our understanding of the spatial distribution of FS when observed at a regional scale.

This research led to a better understanding of how each of the considered socioeconomic and biophysical factors influences the spatial location of a wide range of FS, a subject seldom explored in such detail in the literature. Results showed that both socioeconomic and biophysical factors exert a high influence on the spatial distribution of FS, clearly revealing the shortcomings of planning proposals exclusively confined to the agroecological aptitude perspective (Nguyen et al., 2015; Pirovani et al., 2018). That influence, however, is not comparable across FS, being decisive for the location of some FS and marginal for others.

Contrasting relationships were found between the agricultural intensity level and the degree of dependence on biophysical drivers among the FS, with the simultaneous existence of intensive FS with strong connection to biophysical factors (e.g. Rice system), and others similarly intensive FS but where this relation is much weaker (e.g. Fruit trees system). This finding shows the shortcomings of the assimilation between agricultural intensity and degree of artificialization of the farm's conditions, largely dominant in the literature on the relationship between agriculture and biodiversity/natural resources (Keenleyside et al., 2014). This assimilation ignores the distinction between land and labour productivity and the fact that intensity differences may be due to labour intensity levels rather than higher levels of external outputs. Our results point thus to the need of not reducing farming systems diversity to an intensity gradient, when comparing across distinct productions (Ribeiro et al., 2016).

The use of IACS / LPIS data proved to be an invaluable asset for the research, enabling a high-detailed farm-level analysis, not achievable using official statistics and usually only possible through expensive and time-consuming farm surveys, often unfeasible for research works developed at regional scales like the one used in this study. Therefore, it is worth renewing an appeal previously made (Santos et al., 2020; Tóth and Kučas, 2016), addressed at the EU bodies responsible for maintaining the IACS databases, to make them more accessible to the scientific community, while safeguarding confidentiality duties.

Overall, the model's ability to perform scenario simulations and to predict patterns of farming systems assigns this approach with a high potential to support information-based policy design to improve agricultural landscape planning and ensure the provision of socially valued agroecosystem services.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was funded by project "FARSYD– FARming SYstems as tool to support policies for effective conservation and management of high nature value farmlanDs" – POCI-01-0145-FEDER-016664 (PTDC/AAG-REC/5007/2014), supported by Norte Portugal Regional Operational Programme (NORTE 2020), under the PORTUGAL 2020 Partnership Agreement, through the European Regional Development Fund (ERDF). This research was also supported by the Forest Research Centre, a research unit funded by Fundação para a Ciência e a Tecnologia I.P. (FCT), Portugal (UID/AGR/00239/2019). FM was supported by FCT (contract IF/01053/2015). AL was supported by national funds through FCT – Fundação para a Ciência e a Tecnologia, I.P., in the context of the Transitory Norm - DL57/2016/CP1440/CT0001.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.agsy.2021.103140.

References

Andersen, E., 2017. The farming system component of European agricultural landscapes. Eur. J. Agron. 82, 282–291. https://doi.org/10.1016/j.eja.2016.09.011.

Arnalte-Alegre, E., Ortiz-Miranda, D., 2013. The "southern model" of european agriculture revisited: continuities and dynamics, Research in Rural Sociology and Development. Emerald Group Publishing Limited. https://doi.org/10.1108/S1057-1922(2013)0000019005.

Baptista, F.O., Rolo, J.C., 2017. Trabalho agrícola: percursos e modelos. Cultiv 10, 27–35.

Benoît, M., Rizzo, D., Marraccini, E., Moonen, A.C., Galli, M., Lardon, S., Rapey, H., Thenail, C., Bonari, E., 2012. Landscape agronomy: a new field for addressing agricultural landscape dynamics. Landsc. Ecol. 27, 1385–1394. https://doi.org/10.1007/s10980-012-9802-8.

Birch, C.P.D., Oom, S.P., Beecham, J.A., 2007. Rectangular and hexagonal grids used for observation, experiment and simulation in ecology. Ecol. Model. 206, 347–359. https://doi.org/10.1016/j.ecolmodel.2007.03.041.

Breiman, L., 2001. Random forests. Eur. J. Math. 45, 5–32. https://doi.org/10.1023/A:1010933404324.

Canadas, M.J., Novais, A., 2014. Bringing local socioeconomic context to the analysis of forest owners' management. Land Use Policy 41, 397–407. https://doi.org/10.1016/J.LANDUSEPOL.2014.06.017.
Cerceau, J., Mat, N., Junqua, G., 2018. Territorial embeddedness of natural resource management: a perspective through the implementation of industrial ecology. Geoforum 89, 29–42. https://doi.org/10.1016/j.geoforum.2018.01.001.
Cornia, G.A., 1985. Farm size, land yields and the agricultural production function: an analysis for fifteen developing countries. World Dev. 13, 513–534. https://doi.org/10.1016/0305-750X(85)90054-3.
Debolini, M., Marraccini, E., Dubeuf, J.P., Geijzendorffer, I.R., 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, 702–710. https://doi.org/10.1016/j.landusepol.2017.07.010.
Deffontaines, J.-P., 2004. L'objet dans l'espace agricole. Le regard d'un géoagronome. Nat. Sci. Soc. 12, 299–304. https://doi.org/10.1051/nss:2004041.
Deffontaines, J.P.P., Thenail, C., Baudry, J., 1995. Agricultural systems and landscape patterns: how can we build a relationship? Landsc. Urban Plan. 31, 3–10. https://doi.org/10.1016/0169-2046(94)01031-3.
Deininger, K., Jin, S., Liu, Y., Singh, S.K., 2018. Can labor-market imperfections explain changes in the inverse farm size–productivity relationship? Longitudinal evidence from rural India. Land Econ. 94, 239–258. https://doi.org/10.3368/le.94.2.239.
Egea, P., Pérez y Pérez, L., 2016. Sustainability and multifunctionality of protected designations of origin of olive oil in Spain. Land Use Policy 58, 264–275. https://doi.org/10.1016/j.landusepol.2016.07.017.
Ferraz-de-Oliveira, M.I., Azeda, C., Pinto-Correia, T., 2016. Management of Montados and Dehesas for high nature value: an interdisciplinary pathway. Agrofor. Syst. 90, 1–6. https://doi.org/10.1007/s10457-016-9900-8.
Friedman, J.H., 2001. Greedy function approximation: a gradient boosting machine. Ann. Stat. 29, 1189–1232. https://doi.org/10.1214/009053606000000795.
Grigg, D., 2005. An Introduction to Agricultural Geography, Second edition. Second edi. ed. Routledge, London and New York.
Hazell, P., Wood, S., 2008. Drivers of change in global agriculture. Philos. Trans. R. Soc. B Biol. Sci. 363, 495–515. https://doi.org/10.1098/rstb.2007.2166.
Hengl, T., Nussbaum, M., Wright, M.N., Heuvelink, G.B.M., Gräler, B., 2018. Random forest as a generic framework for predictive modeling of spatial and spatio-temporal variables. PeerJ 6, e5518. https://doi.org/10.7717/peerj.5518.
Januszewski, J., 1968. Index of land consolidation as a criterion of the degree of concentration. Geogr. Plonica 14, 291–296.
Kahil, M.T., Connor, J.D., Albiac, J., 2015. Efficient water management policies for irrigation adaptation to climate change in Southern Europe. Ecol. Econ. 120, 226–233. https://doi.org/10.1016/j.ecolecon.2015.11.004.
Keenleyside, C., Beaufoy, G., Tucker, G., Jones, G., 2014. High Nature Value farming throughout EU-27 and its financial support under the CAP. Report Prepared for DG Environment, Contract No ENV B.1/ETU/2012/0035. Institute for European Environmental Policy, London. https://doi.org/10.2779/91086.
Kristensen, S.B.P., Busck, A.G., van der Sluis, T., Gaube, V., 2016. Patterns and drivers of farm-level land use change in selected European rural landscapes. Land Use Policy 57, 786–799. https://doi.org/10.1016/j.landusepol.2015.07.014.
Lacoste, M., Lawes, R., Ducourtieux, O., Flower, K., 2018. Assessing regional farming system diversity using a mixed methods typology: the value of comparative agriculture tested in broadacre Australia. Geoforum 90, 183–205. https://doi.org/10.1016/j.geoforum.2018.01.017.
Landis, D.A., 2017. Designing agricultural landscapes for biodiversity-based ecosystem services. Basic Appl. Ecol. 18, 1–12. https://doi.org/10.1016/j.baae.2016.07.005.
Latruffe, L., Piet, L., 2014. Does land fragmentation affect farm performance? A case study from Brittany, France. Agric. Syst. 129, 68–80. https://doi.org/10.1016/j.agsy.2014.05.005.
Levers, C., Butsic, V., Verburg, P.H., Müller, D., Kuemmerle, T., 2016. Drivers of changes in agricultural intensity in Europe. Land Use Policy 58, 380–393. https://doi.org/10.1016/j.landusepol.2016.08.013.
Liaw, A., Wiener, M., 2002. Classification and regression by random. Forest. R News 2, 18–22.
Lomba, A., Guerra, C., Alonso, J., Honrado, J.P., Jongman, R., McCracken, D., 2014. Mapping and monitoring High Nature Value farmlands: challenges in European landscapes. J. Environ. Manag. 143, 140–150. https://doi.org/10.1016/j.jenvman.2014.04.029.
Lomba, A., Strohbach, M., Jerrentrup, J.S., Dauber, J., Klimek, S., McCracken, D.I., 2017. Making the best of both worlds: can high-resolution agricultural administrative data support the assessment of High Nature Value farmlands across Europe? Ecol. Indic. 72, 118–130. https://doi.org/10.1016/j.ecolind.2016.08.008.
Martel, G., Aviron, S., Joannon, A., Lalechère, E., Roche, B., Boussard, H., 2019. Impact of farming systems on agricultural landscapes and biodiversity: from plot to farm and landscape scales. Eur. J. Agron. 107, 53–62. https://doi.org/10.1016/j.eja.2017.07.014.
Nguyen, T.T., Verdoodt, A., Van, Y.T., Delbecque, N., Tran, T.C., Van Ranst, E., 2015. Design of a GIS and multi-criteria based land evaluation procedure for sustainable land-use planning at the regional level. Agric. Ecosyst. Environ. 200, 1–11. https://doi.org/10.1016/j.agee.2014.10.015.
Pirovani, D.B., Pezzopane, J.E.M., Xavier, A.C., Pezzopane, J.R.M., de Jesus Júnior, W.C., Machuca, M.A.H., dos Santos, G.M.A.D.A., da Silva, S.F., de Almeida, S.L.H., de Oliveira Peluzio, T.M., Eugenio, F.C., Moreira, T.R., Alexandre, R.S., dos Santos, A. R., 2018. Climate change impacts on the aptitude area of forest species. Ecol. Indic. 95, 405–416. https://doi.org/10.1016/j.ecolind.2018.08.002.
Plieninger, T., Draux, H., Fagerholm, N., Bieling, C., Bürgi, M., Kizos, T., Kuemmerle, T., Primdahl, J., Verburg, P.H., 2016. The driving forces of landscape change in Europe: a systematic review of the evidence. Land Use Policy 57, 204–214. https://doi.org/10.1016/j.landusepol.2016.04.040.
R Development Core Team, 2017. R: A language and environment for statistical computing. [WWW Document]. R Found. Stat. Comput. URL. http://www.r-project.org (accessed 11.12.18).
Rada, N.E., Fuglie, K.O., 2019. New perspectives on farm size and productivity. Food Policy 84, 147–152. https://doi.org/10.1016/j.foodpol.2018.03.015.
Reboul, C., 1976. Mode de production et systèmes de culture et d'élevage. Écon. Rural. 112, 55–65. https://doi.org/10.3406/ecoru.1976.2413.
Reboul, C., 1989. Monsieur le capital et madame la terre - Fertilité agronomique et fertilité économique, p. 1989.
Ribeiro, P.F., Santos, J.L., Bugalho, M.N., Santana, J., Reino, L., Beja, P., Moreira, F., 2014. Modelling farming system dynamics in High Nature Value Farmland under policy change. Agric. Ecosyst. Environ. 183, 138–144. https://doi.org/10.1016/j.agee.2013.11.002.
Ribeiro, P.F., Santos, J.L., Santana, J., Reino, L., Leitão, P.J., Beja, P., Moreira, F., 2016. Landscape makers and landscape takers: links between farming systems and landscape patterns along an intensification gradient. Landsc. Ecol. 31, 791–803. https://doi.org/10.1007/s10980-015-0287-0.
Ribeiro, P.F., Nunes, L.C., Beja, P., Reino, L., Santana, J., Moreira, F., Santos, J.L., 2018. A spatially explicit choice model to assess the impact of conservation policy on high nature value farming systems. Ecol. Econ. 145, 331–338. https://doi.org/10.1016/j.ecolecon.2017.11.011.
Rizzo, D., Marraccini, E., Lardon, S., Rapey, H., Debolini, M., Benoît, M., Thenail, C., 2013. Farming systems designing landscapes: land management units at the interface between agronomy and geography. Geogr. Tidsskr. J. Geogr. 113, 71–86. https://doi.org/10.1080/00167223.2013.849391.
Ruiz-Martinez, I., Marraccini, E., Debolini, M., Bonari, E., 2015. Indicators of agricultural intensity and intensification: a review of the literature. Ital. J. Agron. 10, 74. https://doi.org/10.4081/ija.2015.656.
Santos, J.L., Moreira, F., Ribeiro, P.F., Canadas, M.J., Novais, A., Lomba, A., 2020. A farming systems approach to linking agricultural policies with biodiversity and ecosystem services. Front. Ecol. Environ. https://doi.org/10.1002/fee.2292 in press, fee.2292.
Schaller, L., Targetti, S., Villanueva, A.J., Zasada, I., Kantelhardt, J., Arriaza, M., Bal, T., Fedrigotti, V.B., Giray, F.H., Häfner, K., Majewski, E., Malak-Rawlikowska, A., Nikolov, D., Paoli, J.-C., Piorr, A., Rodríguez-Entrena, M., Ungaro, F., Verburg, P.H., van Zanten, B., Viaggi, D., 2018. Agricultural landscapes, ecosystem services and regional competitiveness—assessing drivers and mechanisms in nine European case study areas. Land Use Policy 76, 735–745. https://doi.org/10.1016/j.landusepol.2018.03.001.
Shmueli, G., 2010. To explain or to predict? Stat. Sci. 25, 289–310. https://doi.org/10.1214/10-STS330.
Silva, J.F., Santos, J.L., Ribeiro, P.F., Canadas, M.J., Novais, A., Lomba, A., Magalhães, M.R., Moreira, F., 2020. Identifying and explaining the farming system composition of agricultural landscapes: the role of socioeconomic drivers under strong biophysical gradients. Landsc. Urban Plan. 202, 103879. https://doi.org/10.1016/j.landurbplan.2020.103879.
Strobl, C., Malley, J., Tutz, G., 2009. An introduction to recursive partitioning: rationale, application, and characteristics of classification and regression trees, bagging, and random forests. Psychol. Methods 14, 323–348. https://doi.org/10.1037/a0016973.
Surová, D., Pinto-Correia, T., 2008. Landscape preferences in the cork oak Montado region of Alentejo, southern Portugal: searching for valuable landscape characteristics for different user groups. Landsc. Res. 33, 311–330. https://doi.org/10.1080/01426390802045962.
Tóth, K., Kučas, A., 2016. Spatial information in European agricultural data management. Requirements and interoperability supported by a domain model. Land Use Policy 57, 64–79. https://doi.org/10.1016/j.landusepol.2016.05.023.
van de Steeg, J.A., Verburg, P.H., Baltenweck, I., Staal, S.J., 2010. Characterization of the spatial distribution of farming systems in the Kenyan Highlands. Appl. Geogr. 30, 239–253. https://doi.org/10.1016/j.apgeog.2009.05.005.
van Vliet, J., de Groot, H.L.F., Rietveld, P., Verburg, P.H., 2015. Manifestations and underlying drivers of agricultural land use change in Europe. Landsc. Urban Plan. 133, 24–36. https://doi.org/10.1016/j.landurbplan.2014.09.001.
Wilson, G.A., 2009. The spatiality of multifunctional agriculture: a human geography perspective. Geoforum 40, 269–280. https://doi.org/10.1016/j.geoforum.2008.12.007.
Wright, M.N., Ziegler, A., König, I.R., 2016. Do little interactions get lost in dark random forests? BMC Bioinform. 17, 145. https://doi.org/10.1186/s12859-016-0995-8.