Main characteristics of French farms adopting cereal–legume intercropping: A quantitative exploration at the national and local levels

Elodie Yan, Philippe Martin, Marco Carozzi

To cite this version:

Elodie Yan, Philippe Martin, Marco Carozzi. Main characteristics of French farms adopting cereal–legume intercropping: A quantitative exploration at the national and local levels. Agricultural Systems, 2025, 223, pp.104196. 10.1016/j.agsy.2024.104196 . hal-04801626

HAL Id: hal-04801626
https://hal.inrae.fr/hal-04801626v1

Submitted on 25 Nov 2024

Agricultural Systems 223 (2025) 104196

Contents lists available at ScienceDirect

Agricultural Systems

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

Main characteristics of French farms adopting cereal–legume intercropping: A quantitative exploration at the national and local levels

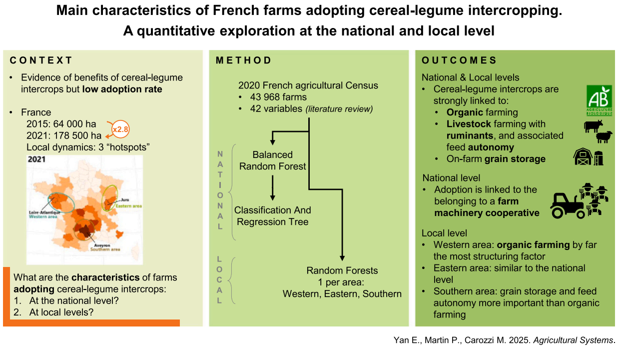
Elodie Yan [*], Philippe Martin, Marco Carozzi

Université Paris-Saclay, INRAE, AgroParisTech, UMR SADAPT, F-91120 Palaiseau, France

HIGHLIGHTS

- A sample of 43,968 French farms was analysed from the French Agricultural Census.
- Cereal–legume intercrops are mainly found on organic and/or feed-autonomous farms.
- Belonging to a farm machinery cooperative was also linked to C–L intercropping.
- The methodology can apply to other agricultural practices at various scales.
- Results suggest ways to promote intercropping, e.g. peer exchanges, market access.

GRAPHICAL ABSTRACT



Main characteristics of French farms adopting cereal-legume intercropping.
A quantitative exploration at the national and local level

CONTEXT

- Evidence of benefits of cereal-legume intercrops but **low adoption rate**
- France
  2015: 64 000 ha
  2021: 178 500 ha
  Local dynamics: 3 "hotspots"

What are the **characteristics** of farms **adopting** cereal-legume intercrops:
1. At the national level?
2. At local levels?

METHOD

2020 French agricultural Census
- 43 968 farms
- 42 variables *(literature review)*

OUTCOMES

National & Local levels
- Cereal-legume intercrops are strongly linked to:
  - **Organic** farming
  - **Livestock** farming with **ruminants**, and associated feed **autonomy**
  - On-farm **grain storage**

National level
- Adoption is linked to the belonging to a **farm machinery cooperative**

Local level
- Western area: **organic farming** by far the most structuring factor
- Eastern area: similar to the national level
- Southern area: grain storage and feed autonomy more important than organic farming

Yan E., Martin P., Carozzi M. 2025. *Agricultural Systems*.

ARTICLE INFO

Editor Name: Leonard Rusinamhodzi

Keywords:
Crop mixtures
Crop diversification
Farming systems
Balanced random forest
Classification and regression tree

ABSTRACT

CONTEXT: Cereal–legume intercropping is a diversification practice that offers many advantages, especially in low-input systems. However, its adoption remains low on European farms, as technical and economic barriers hinder its development. In recent years, an increase in the proportion of arable land cultivated with cereal–legume intercrops has been observed in France. Three areas in particular – in Western, Eastern and Southern France – seem to be particularly dynamic.

OBJECTIVE: This study aimed (i) to identify the main farm characteristics associated with the presence of cereal–legume intercrops at the national level in France and (ii) to highlight more specific characteristics that could explain the particular dynamics observed in each focus region.

METHODS: We analysed data from the 2020 French Agricultural Census for 43,968 farms representative of the French arable crop, livestock, and mixed crop–livestock farming systems. Through a literature review, we identified key factors linked to the presence of cereal–legume intercrops and related them to 42 variables in the census. At the national level, the most important of these variables were identified and interpreted using a balanced random forest and a classification and regression tree (CART). We tested the CART obtained at the national level in the Western, Eastern, and Southern areas and conducted a random forest analysis for each area to identify local particularities.

RESULTS AND CONCLUSION: At the national level, the presence of cereal–legume intercropping was strongly linked to organic farming and the presence of livestock, especially ruminants. These intercrops were prevalent on farms with high feed autonomy for the cattle and sheep. Additionally, they were commonly observed on farms

---

[*] Corresponding author.
E-mail address: elodie.yan@inrae.fr (E. Yan).

https://doi.org/10.1016/j.agsy.2024.104196
Received 13 May 2024; Received in revised form 11 November 2024; Accepted 12 November 2024

with grain storage, possibly indicating feed autonomy, on-farm transformation, or marketing outside of agricultural cooperatives. The belonging to a farm machinery cooperative was also strongly associated with cereal–legume intercropping, likely because these cooperatives give farmers access to specific machinery and provide opportunities for knowledge exchange regarding their practices. Similar characteristics were identified at the local level; organic farming was pivotal in the Western and Eastern areas, followed by feed autonomy for cattle. In the Southern area, however, on-farm grain storage capacity was dominant, likely due to longstanding efforts to achieve feed autonomy.

SIGNIFICANCE: This exhaustive study on French farms identified key farm characteristics strongly linked to cereal–legume intercrops adoption. This insight is critical for promoting this practice, whether through national public policies or local farming support services. The methodology proposed can be easily reproduced to investigate other farming practices at different spatial scales.

1. Introduction

Agriculture has the dual challenge of producing enough output to meet the demands of a growing population while maintaining temporal stability (Foley et al., 2011). Additionally, the increasing emphasis on the sustainability of production – namely its environmental, economic, and social aspects – demands the adoption of diverse and effective farming practices at both the parcel and farm scales (Wezel et al., 2020). Crop diversification is one of the leading sustainable actions promoted within the framework of the European Union's agricultural policy (Galioto and Nino, 2023) and can be implemented using, e.g., cover crops, crop rotations, intercropping, and agroforestry. Although cover crops and crop rotations often come with trade-offs between environmental and economic benefits (Rosa-Schleich et al., 2019), intercropping practices appear to offer more mutually beneficial solutions (Nie et al., 2016; Pelzer et al., 2012).

Intercropping is the practice of growing two or more crop species simultaneously on the same field (Willey, 1979). Plot and field experiments have demonstrated the effectiveness of intercropping in controlling the spread of diseases, insects, and weeds. Cereal–legume intercrops are particularly beneficial: cereals can act as stakes for legumes, which tend to lodge easily (Kontturi et al., 2011), and legumes fix nitrogen from the atmosphere, leaving soil nitrogen available for cereals. This arrangement can result in better harvest quality; specifically, cereal grains grown in intercrops exhibit a higher protein content compared with pure cereal crops (Bedoussac et al., 2015; Li et al., 2023; Pelzer et al., 2012). Intercropping cereals and legumes is also more effective in controlling weeds than sole cereals or sole legumes thanks to the higher soil coverage, even though legumes are weak competitors (Carton et al., 2020; Corre-Hellou et al., 2011; Leoni et al., 2022). Moreover, cereals and legumes protect each other from diseases and pests by creating natural barriers due to the height differences between the plants, which limits the spread of spores, or by "covering the tracks" for insects, resulting in insects failing to locate their host plant due to non-host plants odours masking the host plants' (Lopes et al., 2016; Mansion-Vaquié et al., 2020). Altogether, these advantages can result in better yield stability in intercrops compared with sole crops (e.g., Raseduzzaman and Jensen, 2017) and can lead to decreased usage of chemical inputs, such as pesticides and fertilisers, on farms (Jensen et al., 2020; Yan et al., 2024).

The European Union aims to encourage the implementation of diversification practices in agricultural production with the cross-compliance requirements of the Common Agricultural Policy (CAP). However, spatial and temporal diversification, supported by the EU since the 2013 CAP reform, have had varying and sometimes inconsistent effects across the member states (Galioto and Nino, 2023). In France, the area cultivated with cereal–legume intercrops has more than doubled in recent years, increasing from approximately 64,000 ha in 2015 to approximately 178,500 ha in 2021, according to the French Land Parcel Identification System (IGN, 2023). Nevertheless, this increase remains modest, representing only 1 % of arable land in 2021. By way of comparison, winter wheat, the leading cereal grown in France, represented 29 % of arable land in 2021, whereas peas, the leading legume grown in France, represented 1 % (IGN, 2023). The farmers' low adoption of cereal–legume intercropping can be explained by technical obstacles related to sowing, harvesting, and sorting equipment as well as economic constraints; for example, some agricultural cooperatives do not collect intercrops, the farmers may struggle to find sufficient outlets, and the crops must comply with standards for breadmaking processes (Fares and Mamine, 2023; Magrini et al., 2016; Verret et al., 2020). Together with farmers' risk aversion, these difficulties are strong blocking factors to intercrop adoption that cannot be solved solely at the farm level (Bonke and Musshoff, 2020; Timaeus et al., 2022). However, by tracking on-farm legume-based intercrops in France, Verret et al. (2020) showed that some farmers have successfully implemented cereal–legume intercrops on their farms and are experiencing many of the advantages identified by field experiments. Verret et al. (2020) also highlighted the motivations and technical conditions for successful intercropping, indicating that cereal–legume intercrops are commonly adopted on organic and/or mixed crop–livestock farms. To enhance support for extension services aimed at assisting farmers in adopting this practice, it is essential to identify the factors conducive to its adoption on farms. In France, the adoption of cereal–legume intercropping varies geographically, as shown in Fig. 1. Three areas, specifically those surrounding the departments[1] of Loire-Atlantique in Western France, Aveyron in Southern France, and Jura in Eastern France, have shown a notable rate of cereal–legume intercropping. This variation suggests that local factors, potentially specific to these areas, may also play a role in the adoption of the practice.

In this paper, we present a quantitative study that aimed to identify the main characteristics of the French farms adopting cereal–legume intercrops. We built upon and extended previous qualitative studies based on a limited number of farm surveys, which suggested that **organic farming** and **animal husbandry** are significant factors influencing the adoption of cereal–legume intercropping in France (e.g., Timaeus et al., 2022; Verret et al., 2020). However, the literature suggests that there are more factors than just these two and that local differences may explain different diffusion dynamics. Indeed, as cereal–legume intercrops are considered an agroecological practice (Altieri et al., 2015; Duru et al., 2015), we hypothesise that they are more likely to be adopted on farms involved in **agroecological initiatives** (e.g. Ha et al., 2023). Studies showing the ability of cereal–legume intercrops to compete with weeds also suggest that the practice might be linked to **conservation agriculture** practices (e.g. Peigné et al., 2016). As stated in the previous paragraph, cereal–legume intercrops currently lack outlets, especially concerning long channel and agricultural cooperatives that mainly collect crops for further industrial transformation. Therefore, we hypothesise that cereal–legume intercrops are more likely to be grown on farms for sale through alternative channels, such as **local distribution networks** (e.g. Casagrande et al., 2017), and possibly with previous **on-farm processing** to increase their added value (e.g. Thomopoulos et al., 2018), or even used for other purposes

---

[1] French departments are the equivalent of Level 3 of the EU Nomenclature of Territorial Units for Statistics (NUTS3)

such as the **production of biogas** on farm (e.g. Himanen et al., 2016). Our last hypothesis regarding the factors linked to the adoption of cereal–legume intercropping is that the exchange of knowledge, advice and experience between farmers can favour the adoption of innovative agriculture practices (e.g. Chantre and Cardona, 2014), so the **belonging to groups of farmers** may also enhance the adoption of cereal–legume intercropping.

The objectives of this study were (i) to identify the main farm characteristics associated with the adoption of cereal–legume intercrops at the national level in France and (ii) to highlight more specific characteristics that could explain the particular dynamics observed locally, based on our hypotheses regarding the links between the presence of cereal–legume intercrops and organic farming, the presence of livestock, the involvement in agroecological schemes and practices, conservation agriculture, the types of distribution networks, on-farm transformation, the production of biogas and the belonging to group of farmers,

2. Materials and methods

2.1. Dataset

Information on the composition of French farms was obtained from the 2020 French Agricultural Census data. The census is conducted once every 10 years to provide an exhaustive photograph of the national agricultural situation. In 2020, the survey was conducted using two questionnaires. The first was administrated to all 450,000 farms of Metropolitan France and French overseas departments and regions. The second questionnaire, which was more detailed, was administered to a representative sample of 70,000 farms. For our analysis, we focused on data from the second questionnaire, which provided information on farm structure, crop and livestock production, fertilisation and animal dejection management, diversification activities, marketing channels, and workforce.

Using these data, we studied the absence or presence of cereal–legume intercrops on farms located in Metropolitan France. Farms where arable crops were not cultivated, such as those specialised in viticulture, perennial crops, market gardening, or horticulture, were excluded. We ultimately furnished our database with information on 43,968 farms specialised in arable crops, mixed crop–livestock, or livestock farming. From this sample, we found cereal–legume intercrops on 2548 farms (6 %).

Based on a literature review, we identified factors that may be linked to the adoption of cereal–legume intercrops at the farm level. We then selected 42 variables from our database to represent those factors (Table 1). At the farm level, we identified that the type of production, particularly the presence of livestock, can play a role in whether a farm adopts cereal–legume intercropping. Fodder intercrops are easy to implement (Himanen et al., 2016), and intercrops in general can be used as animal feed on farms (Thomopoulos et al., 2018). In our database, variables 3 and 14–28 (Table 1) represent the type of production and the presence of livestock. According to Navarrete et al. (2015), organic farming can also be linked to the adoption of intercropping because it encourages crop diversification (variables 4 and 5). We determined that conservation agriculture is likely to involve intercropping to achieve higher soil coverage and increase competition with weeds, thereby reducing tillage (Casagrande et al., 2017; Lemken et al., 2017; Peigné et al., 2016). Variables 6–9 represent conservation agriculture practices. Other agroecological practices should also be considered, including reducing chemical inputs and relying more on the services plants provide (e.g., Casagrande et al., 2017; Verret et al., 2020). Involvement in "private certificate schemes" can also reinforce farmers' intention to adopt intercropping (Ha et al., 2023). Variables 9–13 represent agroecological practices and schemes. The production of biogas (variable 42) from anaerobic digestion of agricultural sources may also be linked to the adoption of intercropping, as the energy cover crops used are usually crop mixtures. Regarding the relationship between the farm and its global environment, the literature indicates that the post-harvest process and marketing channels can influence the adoption of intercropping. On-farm sorting and transformation, as well as direct selling, are potential levers, whereas relying on traditional channels with agricultural cooperatives is more of a barrier (Casagrande et al., 2017; Himanen et al., 2016; Thomopoulos et al., 2018). Variables 29–38 represent the different marketing channels and on-farm transformations in our database. Finally, involvement in groups of farmers or farm machinery co-operatives and having access to training sessions (variables 39–41) support the adoption of new agricultural practices (Chantre and Cardona, 2014; Ha et al., 2023; Himanen et al., 2016). Table 1 provides further details on all 42 variables considered in this study, which will help in interpreting the results in Section 3.

2.2. Variable selection: Balanced random Forest

To characterise the farms on which cereal–legume intercrops are grown, we first aimed to identify which of the 42 selected variables best-separated farms with intercrops ("Presence" group) from those without ("Absence" group). To do this, we used a random forest algorithm (Breiman, 2001), which is a machine-learning classification method. Random forests are based on multiple decision trees. Each decision tree parts the data according to a set of variables from the given 42. At each node, one variable is used to split the data into two subsamples, and each

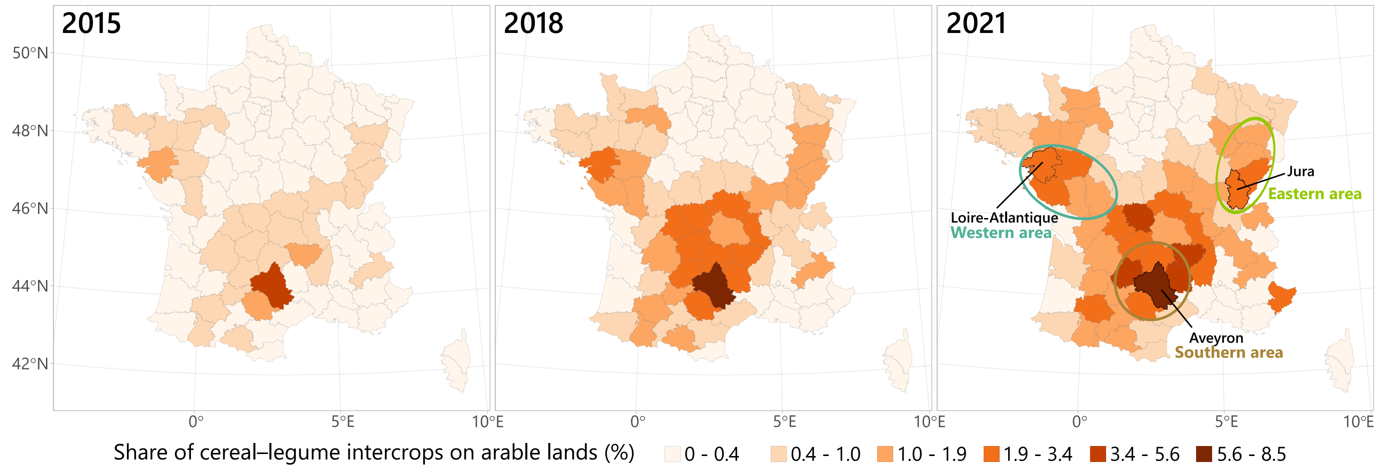


Fig. 1. Share of cereal–legume intercrops on total departmental arable lands in 2015, 2018, and 2021 in French departments, based on data from the French Land Parcel Identification System.

Table 1

Description of the 42 variables selected from the 2020 French Agricultural Census and the related hypotheses tested in the analysis.

ID	Variable name	Variable description	Variable nature	Tested factors and hypotheses	References
1	Total_area	Total useable agricultural area	Numeric (hectares)	**Farm characteristics**	
2	Eco_dimension	Economic dimension	Factor		
3	Farm_type	Type of farming	Factor		
4	Prop_organic	Proportion of area certified in organic agriculture	Numeric	**Organic farming** Organic farming fosters crop diversification, and case studies highlight that cereal–legume intercrops are mainly grown in organic farming conditions.	Casagrande et al., 2017; Ha et al., 2023; Verret et al., 2020
5	Prop_organicConv	Proportion of area under conversion to organic agriculture	Numeric		
6	Prop_plough	Proportion of area ploughed	Numeric	**Conservation agriculture** Some studies have linked intercropping to conservation agriculture (e.g., for weed control).	Casagrande et al., 2017; Lemken et al., 2017; Peigné et al., 2016
7	Prop_conservation	Proportion of area under conservation agriculture	Numeric		
8	Prop_direct	Proportion of area with direct seeding	Numeric		
9	HVE	High Environmental Value certification (in French: HVE certification)	Boolean	**Agroecological practices** In low-input systems, intercropping can result in better yields than sole cropping. Case studies often cite reducing chemical inputs as one objective of intercropping; participation in “private certificate schemes” strengthens the intention to adopt intercropping.	Bedoussac et al., 2015; Casagrande et al., 2017; Ha et al., 2023; Verret et al., 2020
10	Other_QualEnv	The farm is engaged in a quality or environmental approach other than DEPHY (network of demonstration farms committed to reducing pesticides), HVE (High Environmental Value), or GIEE (Environmental and Economic Interest Group)	Boolean		
11	Prop_agroforestry	Proportion of area with agroforestry	Numeric		
12	GIEE	Environmental and Economic Interest Group	Boolean		
13	DEPHY	DEPHY Farm	Boolean		
14	Dens_cattle	Stocking density of cattle	Numeric	**Livestock** Fodder intercrops are easy to implement, and sorting is not always necessary before feeding intercrops to animals. In case studies, intercrops are mainly found in livestock or mixed crop–livestock farming systems.	Casagrande et al., 2017; Himanen et al., 2016; Lemken et al., 2017; Thomopoulos et al., 2018; Verret et al., 2020
15	Dens_sheep	Stocking density of sheep	Numeric		
16	Dens_goats	Stocking density of goats	Numeric		
17	Dens_pigs	Stocking density of pigs	Numeric		
18	Dens_poultry	Stocking density of poultry	Numeric		
19	AutoFodder_cattle	Fodder autonomy for cattle	Factor		
20	AutoConcen_cattle	Concentrate autonomy for cattle	Factor		
21	AutoFodder_sheep	Fodder autonomy for sheep	Factor		
22	AutoConcen_sheep	Concentrate autonomy for sheep	Factor		
23	AutoFodder_goats	Fodder autonomy for goats	Factor		
24	AutoConcen_goats	Concentrate autonomy for goats	Factor		
25	AutoConcen_pigs	Concentrate autonomy for pigs	Factor		
26	AutoConcen_hen	Concentrate autonomy for laying hens and pullets	Factor		
27	AutoConcen_poultry	Concentrate autonomy for other poultry	Factor		
28	AutoConcen_duck	Concentrate autonomy for ducks and geese	Factor		
29	Prop_coop	Proportion of sales to an agricultural cooperative	Numeric	**Marketing channels** Sorting intercrops is a central issue, and agricultural cooperatives that are not equipped to sort the intercrops refuse to collect them. In case studies, very few farmers sell their intercrops to agricultural cooperatives.	Casagrande et al., 2017; Himanen et al., 2016; Thomopoulos et al., 2018
30	Local_network	Marketing in local distribution network	Boolean		
31	Prop_directSale	Proportion of direct sales to consumer	Numeric		
32	Prop_oneInterm	Proportion of sales to consumer with one intermediary	Numeric		
33	Prop_longChannel	Proportion of sales to the private sector (long channel)	Numeric		
34	Prop_saleService	Proportion of sales of services as part of integration or boarding of animals	Numeric		
35	Stock_farm	Grain stocking capacity on farm	Numeric (tonnes)		
36	Transfo	Transformation of agricultural products	Boolean	**On-farm transformation** On-farm transformation, along with local distribution networks, can be more profitable for farmers and enable them to sell diversification crops, including intercrops that may not be collected by agricultural cooperatives.	Casagrande et al., 2017
37	Transfo_cer	Transformation of cereals	Boolean		
38	Transfo_olea	Transformation of oleaginous	Boolean		
39	Exchange_group	Group for exchanges on experiences, results, or trainings	Boolean	**Involvement in groups of farmers, exchanges with other farmers** Sharing knowledge and past experiences is important for wider adoption of intercropping. Networks facilitate the diffusion of techniques (references, help, advice), and trainings can foster the adoption of new agricultural practices. Farmers who often interact with other farmers are more likely to adopt intercropping.	Casagrande et al. (2017); Chantre and Cardona (2014); Ha et al. (2023); Himanen et al. (2016); Mamine and Farès (2020)
40	Machinery_group	Farm machinery cooperative	Boolean		
41	Work_otherFarm	Work on other farms or other non-agricultural work	Boolean		

(*continued on next page*)

Table 1 (*continued*)

ID	Variable name	Variable description	Variable nature	Tested factors and hypotheses	References
42	Sale_biogas	Sale of biogas	Boolean	**Production of biogas** Intercropping can produce feedstock to produce biogas.	Himanen et al. (2016)

subsample is then further divided based on another variable, and so on. In the end, the decision tree provides a representation of the variables capable of parting the Presence and Absence groups. The random forest algorithm then averages the results of all the computed decision trees and ranks the variables according to their importance in discriminating between the two groups. The importance of a variable is given by the mean decrease in impurity, a measure of how effectively the variable helps to split the data. The higher the mean decrease in impurity, the more important the variable.

The Presence cases in our dataset accounted for only 6 % of the observations, resulting in an unbalanced classes problem in which the class of interest (Presence) was the minority. In such cases, conventional classification algorithms are unlikely to perform well because they try to indifferently minimise the error rate (Fernández et al., 2018, p. vii). In our case, an error rate of 6 % may seem low in absolute terms – as it would mean that 94 % of the data are correctly classified – but it could hide the fact that none of the Presence observations have been correctly classified. Indeed, in order to minimise the error rate, the algorithm could classify all the observations in the Absence category, and have an error rate of 6 % only in total, but it would represent an error rate of 100 % for the Presence observations. Classification methods such as random forests can be adapted to manage these unbalanced data. In balanced random forests, for example, in each iteration of the random forest algorithm, a sample is taken from the minority class and a sample of the same size is randomly drawn from the majority class (Chen et al., 2004).

As the Presence cases only represent a minority in the dataset (6 %), we used the balanced random forest method with undersampling. We first created a training sample and a test sample by randomly drawing 80 % (n = 35,174) and 20 % (n = 8794) of the dataset, respectively. Then, in the training set, we randomly drew as much Absence data as we had Presence data (i.e., n = 2085, Table 2). We used a 10-fold cross-validation, which means that we repeated this operation 10 times, and each time over 1500 trees. We thus obtained a final random forest model trained on a balanced dataset. We then applied this final model to the test sample to measure the model's performance using the area under the receiver operating characteristic (ROC) curve. The ROC curve represents the true-positive rate (probability of a farm being classified in the Presence group when cereal–legume intercrops are indeed present on the farm) as a function of the false-positive rate (probability of a farm being classified in the Presence group when there are no cereal–legume intercrops on the farm). An area under the ROC curve of 0.5 means the classifier makes random guesses. An area under the ROC curve of 1 means the classifier makes perfect predictions.

2.3. *Roles of the features for the adoption of cereal–legume intercrops*

Based on the variable importance measured by the random forest, we selected the variables that were most important in discriminating the Presence and Absence groups to deepen our understanding of their role in the adoption of intercropping. These variables were used in a classification and regression tree (CART, Breiman et al., 1984), another decision tree algorithm, to classify our data into the two groups and provide a more detailed characterisation of the farms belonging to each group. CART constructs a binary tree structure by recursively partitioning the input data based on feature values. For our classification problem, the algorithm identifies the optimal split at each node to minimise the class impurity. To limit the risks of overfitting, we chose to prune the tree using the complexity parameter value that would give a balance between the complexity of the tree (i.e. the number of splits and, therefore, the interpretability of the tree) and its predictive ability, based on the cross-validation error. From our tests on different complexity parameter values, we determined that the cross-validation error did not decrease significantly after the seventh split, which coincided with a complexity parameter of 0.009.

As with the random forest, we first needed to solve the issue of the imbalanced dataset. Although the subsampling method worked well with the random forest, that resulted in an average tree over 1500 trees, it is not possible to obtain an average CART. For this analysis, we chose to oversample our data, so we penalised the CART more heavily when one observation was classified incorrectly. This way, we gave more weight to the Presence data, multiplying it 15 times to obtain a dataset with 38,220 Presence cases and 41,420 Absence cases, i.e., a ratio of 48 % to 52 %. We used a training sample (80 %, n = 65,750, Table 2) and a test sample (20 %, n = 16,438, Table 2). The CART was trained on the training sample, and the obtained model was tested on the test sample to measure its predictive performance by measuring the area under the ROC curve.

2.4. *Local factors*

Using data from the French Land Parcel Identification System from 2015 to 2021, we mapped the evolution of the area cultivated with cereal–legume intercrops. We identified three regions in which cereal–legume intercropping had spread from hotspots (Fig. 1). The first hotspot was located around the Loire-Atlantique department, and we designated this first area the Western area. The second hotspot was located around the Jura department, and the third around the Aveyron department. We designated these the Eastern and Southern areas,

Table 2
Summary of the number of cases (total, Presence, and Absence) at each step of the analysis.

Step	Level	Total	Whole dataset		Training set		Test set	
			Presence	Absence	Presence	Absence	Presence	Absence
Original dataset	France	43,968	2548	41,420	/	/	/	/
Balanced random forest (undersampling[a])	France	43,968	2548	41,420	2085	2,085[a]	463	8331
CART with (oversampling[b])	France	79,640[b]	38,220[b]	41,420	32,591	33,159	8177	8261
Random forests per area	Western	4056	433	3623	335	2909	98	714
	Eastern	1027	84	943	69	752	15	191
	Southern	2125	512	1613	424	1276	88	337

CART, classification and regression tree.

[a] Undersampling by randomly selecting as many Absence cases as Presence cases in the training set.

[b] Oversampling by duplicating Presence cases to get a balanced ratio between Absence and Presence cases before constituting the training and test sets.

respectively. The Western area was dominated by arable crop farming, with 30 % of the useable agricultural area (UAA) dedicated to this type of farming. The main product was winter wheat, which represented 15 % of the UAA. Mixed crop–livestock, dairy cattle, and meat cattle farms represented 16 %, 16 %, and 15 % of the UAA, respectively (Table 3). The distribution of farm types in the Western area was comparable to the national distribution but with larger farms (+20 ha on average) and lower winter wheat yields (−14 q ha^{-1} on average) in 2020. Additionally, the Western area exhibited a larger proportion of organic farmland at 13 % of the UAA versus 9.5 % of the national UAA (Table 3). The Eastern area was largely dominated by livestock farming systems, with dairy cattle farms accounting for 53 % of the UAA (Table 3). The next most prevalent systems were arable crop, mixed crop–livestock, mixed dairy and meat cattle, and dairy cattle farms, which accounted for 13 %, 10 %, 10 %, and 9 % of the UAA, respectively (Table 3). Hence, the Eastern area was mainly covered by permanent grassland (60 % of the UAA), and the main crop, winter wheat, accounted for 9 % of the UAA (Table 3). Compared with the national statistics, the Eastern area's farms were larger (+28 ha on average), with a slightly higher mean winter yield and proportion of organic farmland in 2020 (Table 3). The Southern area was also dominated by livestock farming systems, with meat cattle farms representing 38 % of the UAA and sheep and goat farms representing 25 % of the UAA. Arable crop farms only accounted for 10 % of this area's UAA (Table 3); grasslands mainly covered the agricultural area, with 59 % of the UAA dedicated to permanent grassland and 14 % dedicated to temporary grassland. The most commonly grown crop was winter wheat, making up 3.8 % of the UAA. The mean farm size was similar to that on the national scale (−1 ha), and the proportion of organic farmland was higher (12 % of the UAA; Table 3). Of the three areas studied, the Southern area had the lowest mean winter wheat yield in 2020 (47.3 q ha^{-1}, i.e., −21.2 q ha^{-1} compared with the national mean; Table 3).

To better understand the reasons behind the observed local diffusion of intercropping, we used the CART obtained from nationwide data to classify the cases in the three regions. If the CART was able to classify the local data correctly, it would mean that the farm characteristics associated with the presence of cereal–legume intercrops are not specific to local contexts. Otherwise, we would attempt to identify those local features by running random forests on the local data.

2.5. Calculations and statistical analysis

To handle our classification problem, we chose to use a random forest algorithm, as this method is robust to noise and variance, not prone to overfitting and can handle high-dimensional datasets with both categorical and numerical data (Breiman, 2001). Also, random forests are an ensemble method, based on the aggregation of multiple decision trees, which makes them more reliable than single decision trees. However, random forests can be difficult to interpret, and therefore, we chose to perform a decision tree using a set of features selected based on the feature importance given by the random forest. We chose to use a Classification and Regression Tree model as it also handles large datasets and is easy to interpret and implement, which was an important criterion as we wanted to propose a replicable methodology.

Statistical analyses were carried out using R version 4.3.1 (R Core Team, 2023). We used the package caret, version 6.0–94, for the random forests and the balanced random forest (Kuhn, 2008). The CART was built with the package rpart, version 4.1.19 (Therneau et al., 2023). The ROC validations were computed with the package ROCR (Sing et al., 2005). Access to the French Agricultural Census data was made possible within a secure environment offered by the Centre d'accès sécurisé aux données (CASD; Ref. 10.34724/CASD).

3. Results

3.1. Key variables associated with the presence of cereal–legume intercropping

The balanced random forest resulted in an area under the ROC curve of 0.915, thereby validating our model, in accordance with Mandrekar (2010) who considers that an AUC of 0.8 to 0.9 is "excellent" and more than 0.9 is "outstanding". Fig. 2 shows the importance of the top 20 selected variables. The proportion of land under organic farming (*Prop_organic*) on the farm was the most important variable in discriminating Presence cases from Absence cases. The second most important variable was the farm's grain storage capacity (*Stock_farm*). The top 10 most important variables include those linked to livestock, such as feed autonomy for cattle (*AutoConcen_cattle*, third position) and sheep (*AutoConcen_sheep*, ninth position) and stocking density (*Dens_cattle*, seventh position). Variables pertaining to the farm structure were also significant, including the total agricultural area in use (*Total_area*, fourth position) and the type of farming, such as arable, mixed crop–livestock, or

Table 3
Principal agricultural characteristics of the Western, Eastern, and Southern areas compared with the national scale (France), based on data from the 2020 French Agricultural Census and the 2020 annual agricultural statistics (Agreste, 2020), which provide information on the mean winter yield in 2020.

Area	Western	Eastern	Southern	France
Climate	Temperate, oceanic to altered oceanic	Mountain/semi-continental	Mountain/altered oceanic	/
Useable agricultural area (UAA, ha)	2,243,952	869,662	1,605,158	26,745,875
Types of farming (% of UAA)	30 % arable crop 16 % mixed crop–livestock 16 % dairy cattle 15 % meat cattle 8 % pig and poultry	53 % dairy cattle 13 % arable crop 10 % mixed crop–livestock 10 % mixed dairy and meat cattle 9 % dairy cattle	38 % meat cattle 25 % sheep and goats 10 % arable crop 10 % dairy cattle 7 % mixed dairy and meat cattle	36 % arable crop 15 % meat cattle 14 % mixed crop–livestock 14 % dairy cattle 6.5 % sheep and goats
Main production (% of UAA)	*23 % permanent grassland* 15 % winter wheat 14 % temporary grassland 8.3 % maize (fodder and silage) 8.2 % maize (grain) 6.1 % sunflower 3.8 % rapeseed	*60 % permanent grassland* 9.0 % winter wheat 7.0 % temporary grassland 4.6 % maize (fodder and silage) 4.3 % winter barley 3.3 % rapeseed 3.0 % maize (grain)	*59 % permanent grassland* 14 % temporary grassland 3.8 % winter wheat 3.2 % winter barley 2.7 % alfalfa	*29 % permanent grassland* 16 % winter wheat 6.8 % temporary grassland 6.5 % maize (grain) 5.0 % maize (fodder and silage) 4.4 % winter barley 4.2 % rapeseed
Mean farm size (ha)	89	97	68	69
Mean winter wheat yield in 2020 (q ha^{-1})	54.2	70.0	47.3	68.5
Organic farming (% of UAA)	13	11	12	9.5

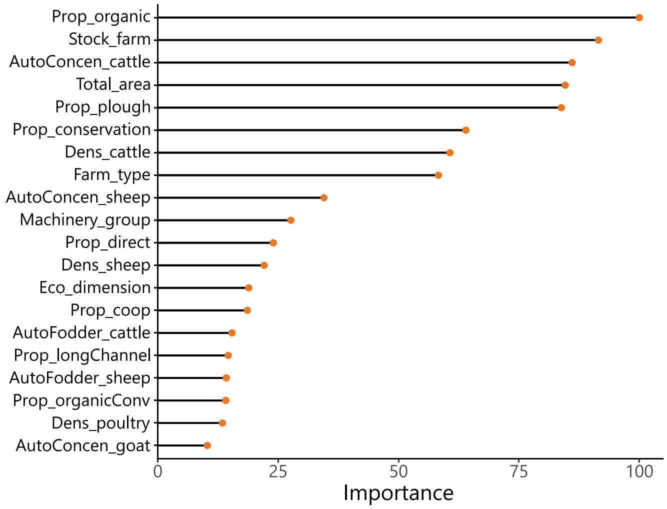


Fig. 2. National-level balanced random forest variable importance plot. The importance is scaled from 0 % to 100 % relative to the first most important variable.

livestock farming (*Farm_type*, eighth position). Furthermore, we identified two variables associated with soil tillage: the proportion of ploughing (*Prop_plough*) and conservation farming (*Prop_conservation*), which ranked fifth and sixth, respectively. Finally, belonging to a farm machinery cooperative (*Machinery_group*, tenth position) was also discriminant in terms of presence or absence of cereal–legume intercrops. In the top 20 most important variables were those associated with marketing channels as well as other variables linked to livestock, farm structure (economic dimension), and tillage (direct seeding). The variables associated with diversification activities did not play a strong role in the Presence/Absence classification, nor did those related to the farmers' environmental or quality approaches other than organic farming. The balanced random forest results show the relative importance of the variables in explaining the presence or absence of cereal–legume intercrops on farms. However, it is unclear which values of these variables actually dictate the Presence/Absence classification. Consequently, we chose to further investigate the values of the 10 most important variables with the CART model, according to the results of the balanced random forest shown in Fig. 2.

The CART model was validated based on its area under the ROC curve value of 0.83. The classification tree separated the Presence and Absence cases based on six of the top 10 variables identified by the random forest. The proportion of the farm's land under organic farming was the first variable in the classification tree, with a threshold of 43 % of the total farm UAA (Fig. 3). However, the majority of farms (82 %) in the sample had between 0 % and 10 % of their land under organic farming versus 80 %–100 % of the land on the remaining farms (17 %). Therefore, we can consider the split as operated between two categories: less than 10 % or more than 80 % of UAA under organic farming. The presence of cattle or sheep farming, even with minimal feed autonomy, were also strongly linked to the presence of cereal–legume intercropping. For grain storage capacity, another significant factor, the results show a split at 8.5 or 9.5 t, although the majority (95 %) of the French farms below this split did not have any grain storage capacity. Given that the average on-farm grain storage capacity in France is 579 t (CERESCO and Systra, 2020), the split could be considered to represent binary yes/no responses to the existence of on-farm grain storage. Furthermore, the classification tree indicated that cereal–legume intercropping was more adopted in livestock farming with cattle, sheep, goat, or other herbivore production than in arable farming, mixed crop and/or mixed livestock farming, and livestock farming with pig or poultry production. Belonging to a farm machinery cooperative also appears to be associated with the adoption of cereal–legume intercropping.

3.2. *Variables for local adoption of cereal–legume intercropping*

The CART trained on national data and tested on data from the three regions (presented in Section 2.4) yielded satisfactory predictions, with areas under the ROC curves of 0.80, 0.79, and 0.73 for the Western, Eastern, and Southern areas, respectively. However, the tree's performance was not as robust as on the national data, prompting further investigation into the local factors that are linked to the presence of cereal–legume intercrops on farms. A classic random forest was trained

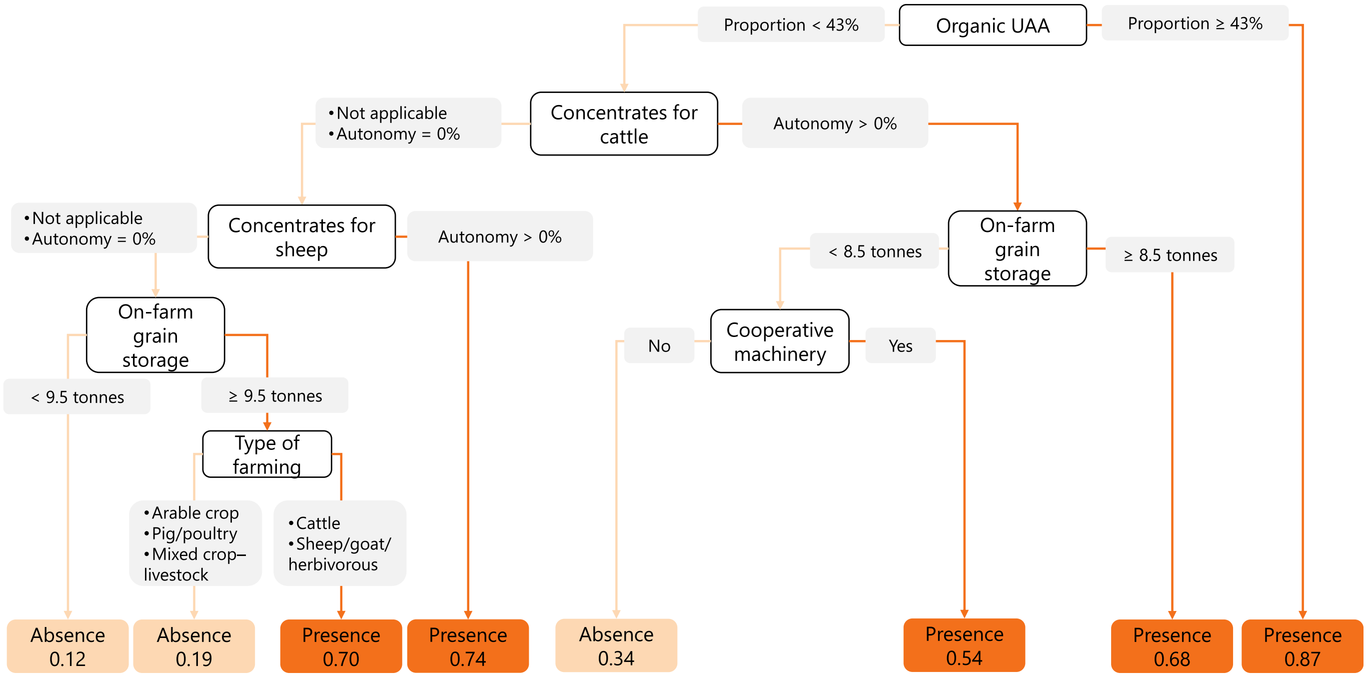


Fig. 3. Classification tree obtained with the CART algorithm applied at the national level. The numbers in the Absence and Presence boxes give the probability of one observation being classified as Presence following the path. UAA, useable agricultural area.

for each area, resulting in areas under the ROC curves of 0.85, 0.85, and 0.86 for the Western, Eastern, and Southern areas, respectively.

The top 10 most important variables in all three areas were similar to those at the national scale (Fig. 4). In the Western area, nine of the top 10 variables were the same as those at the national level; only the belonging to a farm machinery cooperative (*Machinery_group*) was absent. In this area, the proportion of land under organic farming (*Prop_organic*) was predominant compared with the other variables; the second most important variable (*AutoConcen_cattle*) only contributed approximately 34.8 % of the *Prop_organic* variable contribution.

In the Eastern area, eight of the top 10 variables were the same as those at the national level; feed autonomy for sheep (*AutoConcen_sheep*) and belonging to a farm machinery cooperative (*Machinery_group*) were absent. In this area, the proportion of land under direct seeding (*Prop_direct*) and the selling to an agricultural cooperative (*Prop_coop*) were more important than at the national level (Fig. 4). The proportion of land dedicated to organic farming (*Prop_organic*) remained the most important variable. Finally, in terms of variable importance, this area had the ranking closest to that of the national level.

In the Southern area, eight of the top 10 variables were the same as those at the national level. Sheep stocking density (*Dens_sheep*) and direct seeding proportion (*Prop_direct*) were more important than the type of farming (*Farm_type*), which was absent from this area's top 10 along with the belonging to a farm machinery cooperative (*Machinery_group*; Fig. 4). In this area, the proportion of land cultivated under organic farming (*Prop_organic*) fell to the fifth position.

4. Discussion

4.1. Factors associated with the adoption of cereal–legume intercropping at the national scale

The results of this study show that organic farming is strongly associated with the presence of cereal–legume intercrops in French farms, which is in line with the hypotheses drawn from previous case studies in the literature that organic farming is a major factor fostering the adoption of cereal–legume intercrops. This result can be linked to various features of organic farming. Restrictions on the chemical inputs allowed in organic farming oblige farmers to seek alternative solutions to controlling weeds, pests, and diseases, including crop diversification and intercrops in particular. Crop diversification offers acceptable performance in terms of yield and quality of production (Lithourgidis et al., 2011; Pelzer et al., 2012). Furthermore, in France, organic agricultural cooperatives collect and sort binary intercrops (e.g., wheat–faba bean, triticale–pea) to sell through their channels (Corre-Hellou et al., 2013). Such market opportunities for organic cereal–legume intercrops can contribute significantly to facilitating their adoption in organic farming. Cereal–legume intercrops are uncommon outside of organic farming systems, and they are mainly observed in livestock farming situations (Lemken et al., 2017; Verret et al., 2020), often ruminant farming with cattle and goats rather than monogastric farming. Pig and poultry systems are primarily driven by cost-efficiency, with precise feed rations that vary according to the animals' physiological stage. Therefore, in monogastric systems, breeders need to evaluate the composition of the intercrops, so sorting is necessary and the use of intercrops for animal feed becomes more challenging (Casagrande et al., 2017). As a result, intercrops are less likely to be used as monogastric feed. In conventional pig and poultry farming, which represent 98 % and 94 % of the pig and broiler stock, respectively, according to the French Agricultural Census, the feed mainly consists of high–nutritive value and low–water content material that is easy to transport and stock. Our data indicate that 51 % of the conventional pig farms and 89 % of the conventional broiler farms have a concentrate autonomy of less than 25 %, indicating that this type of feed is easily available on the market.

In ruminant farms, on the other hand, feed rations are more flexible, particularly in the context of meat production – in dairy production, the milk protein and fat rates must be respected, which necessitates more precise feeding. Our findings indicate that a significant proportion of cattle (90 %) and sheep farms (89 %) growing cereal–legume intercrops had a fodder autonomy of more than 90 %, compared with 83 % of the cattle farms and 76 % of the sheep farms that did not grow cereal–legume intercrops. Moreover, farms growing cereal–legume intercrops demonstrated higher concentrate autonomy: 33 % of the cattle farms and 26 % of the sheep farms with cereal–legume intercrops had over 90 % concentrate autonomy versus only 13 % of the cattle farms and 18 % of the sheep farms without cereal–legume intercrops. Indeed, one of the main advantages of cereal–legume intercrops for breeders is to enhance feed autonomy (Verret et al., 2020) because they help secure the fodder stock and constitute a balanced alternative to soya meal, which is mainly imported in France.

The presence of grain storage on the farm was also identified as one of the main characteristics of the farms adopting cereal–legume intercropping. Grain storage is likely to be associated with livestock farming, as harvested grains are stored to feed the animals, thereby ensuring the farm's self-sufficiency in concentrate feeds. However, on-farm grain storage can also reveal a particular marketing strategy in arable farming outside of the traditional agricultural cooperative channels. On average, the storage capacity on arable crop farms is higher than on cattle and sheep farms: 700 t versus 109–192 t, respectively. Harvested crops stored on arable farms are likely sold through channels other than

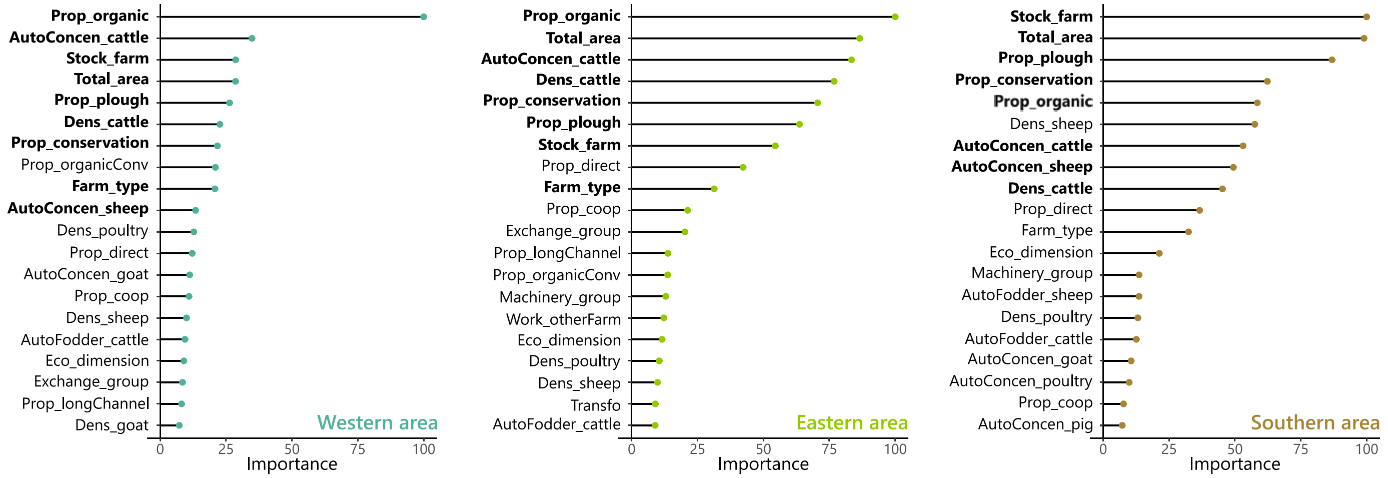


Fig. 4. Random forest variable importance plots for the Western, Eastern, and Southern areas. In each plot, the importance is scaled from 0 % to 100 % relative to the first most important variable. Variables in bold are the common variables between the top 10 at the national level and the top 10 at the area level.

cooperatives, which enables arable farmers to avoid the problem of intercrops going uncollected by storage organisations that are not equipped to sort them. Furthermore, we assume that the storage capacity on these arable crop farms exists so the intercrops can be sorted on the farm before being sold (for food or feed). However, this hypothesis could not be verified with the available data.

The belonging to a farm machinery cooperative also appears in our results to be associated with the adoption of cereal–legume intercropping. This association is likely present because the machinery cooperatives provide access to specific and costly agricultural equipment through joint investments. Therefore, farmers can overcome technical obstacles related to sowing, harvesting, and even sorting (Mamine and Farès, 2020). The farm machinery cooperative can also act as a vector for innovation by providing space for farmers to share their practices and experiences (Darnhofer et al., 2010; Himanen et al., 2016). This way, farmers can draw inspiration from the experience of their colleagues who are growing intercrops and adopt the practice themselves (Ha et al., 2023).

The CART did not use the variables related to ploughing, soil conservation, or the UAA. The French Agricultural Census questionnaire defined ploughing as deep tillage of at least 15 cm and turning over with a plough. Soil conservation was defined as deep tillage without turning over (using a chisel or decompactor) or with reduced turning (using disc or tine stubble cultivators or a rotavator). On the farms with cereal–legume intercrops, we noted more areas using soil conservation techniques (median of 10 % of areas cultivated using these techniques for the Presence group vs. 3 % for the Absence group) and slightly fewer areas using ploughing techniques (median of 20 % of areas using ploughing for the Presence group vs. 24 % for the Absence group). These differences may not be clear-cut enough for the classification tree to use them to separate the Presence and Absence groups. Nevertheless, this suggests that farmers growing intercrops are more likely to be involved in other agroecological approaches such as soil conservation techniques. A perspective to improve our method could be using a C4.5 algorithm (Quinlan, 1992) that is a classification tree algorithm based on entropy measures and that allows multiple partitioning, and not only binary partitioning, as does the CART algorithm. Therefore, the C4.5 algorithm may be helpful in interpreting the roles of variables related to ploughing, soil conservation or the UAA, that were not handled by the CART.

4.2. Local factors

In all three areas, the UAA was ranked as the second most important variable separating the Absence and Presence groups (*Total_area*, Fig. 4). Looking more closely at the data, we noted that farms growing cereal–legume intercrops tended to have a greater UAA than those that were not. Specifically, farms with intercrops in the Western, Eastern, and Southern areas were on average 11, 65, and 51 ha larger, respectively, than those without. Based on the crop types grown in the three areas, it is evident that farms with cereal–legume intercrops have higher forage and grass coverage than farms without (on average, +28, +48, and + 47 ha in farms with intercrops in the Western, Eastern, and Southern areas, respectively). Therefore, we suggest that the observed difference in UAA between farms with and without cereal–legume intercrops is partly due to the variations in forage and grass areas, which reflect the presence of animals or at least crops for animal feed.

In the Western and Eastern areas, just as observed at the national level, the proportion of farm area under organic farming (*Prop_organic*, Fig. 4) was the most discriminating factor between the Presence and Absence groups. However, we noted that the proportion of farm area in the process of conversion to organic farming was also among the top 10 variables in the Western area. Adoption of cereal–legume intercropping began earlier in the Western area than in the Eastern area (Fig. 1), so knowledge of the practice may have diffused more widely, leading farmers to adopt cereal–legume intercropping earlier in the process of conversion to organic farming than in the Eastern area.

In the Eastern area, the proportion of direct seeding (*Prop_direct*, Fig. 4) emerged as the eighth most important variable for discriminating between the Presence and Absence groups. In the Presence group, the area under direct seeding was on average + 27 ha larger than in the Absence group (36 ha vs. 9 ha, respectively). However, this factor was less helpful in discriminating between the groups at the national level, as the average difference was only +3 ha for the Presence group. In the Eastern area, marketing through agricultural cooperatives (*Coop*, Fig. 4) was also one of the top 10 factors for classifying the Absence and Presence groups. However, it is not possible to conclude why this factor seems more discriminating in this area than in the others or at the national level as the rate of marketing via cooperatives does not differ significantly between the Presence and the Absence groups. Yet the variable sits in the tenth position, so it is not surprising that we cannot identify significant differences when considering only this variable (without interactions with the first nine variables).

In the Southern area, the proportion of farmland cultivated under organic farming was only the fifth most important variable, whereas on-farm grain storage was ranked first. In this area, which is dominated by cattle and sheep farming and has low-potential land compared to the national average (Table 1), cereal–legume intercropping has been practised for many years with the aims of improving feed autonomy for the livestock sector and securing production against climate hazards and diseases (Clouet et al., 1986). On-farm storage could be an indicator of this self-sufficiency, and livestock farming currently seems to be a more discriminating factor than conversion to organic farming in this area. However, via the diffusion of agriculture innovation, the presence of cereal–legume intercrops in livestock farms in the area may have influenced their adoption on other farms and even fostered the conversion to organic farming.

Besides the local factors, our investigation revealed that the first eight factors identified on a national scale were present in each of the three focus areas. These factors included organic farming, on-farm grain storage, and the presence of livestock with a certain degree of feed autonomy. However, given the structure of our study, this finding was not unexpected; the three zones we studied in more detail accounted for 38 % of the total area cultivated with cereal–legume intercrops in France in 2020.

4.3. Contributions of the study for the adoption of cereal–legume intercrops

In this paper, we present an exhaustive analysis on 43,968 farms representative of the French arable crops, mixed crop-livestock and livestock farming systems in 2020. Therefore, our study complements the case studies that have been conducted so far and that helped formulate hypotheses on the barriers and levers to the adoption of cereal–legume intercrops. Although our methods and results do not allow us to conclude on causal links between the farm characteristics highlighted and the adoption of the practice, they still give insights on some levers that could be tested to foster the adoption. Indeed, we showed that the belonging to a farm machinery cooperative is one of the main characteristics of the farms adopting cereal–legume intercrops. In the French context, farm machinery cooperatives are local groups of farmers that invest together and share machinery and/or workforce. Those farm machinery cooperatives also provide spaces for the farmers to exchange on their practices. Therefore, at the farm and farmer's scale, it could be relevant to initiate local collective actions to help farmers overcome technical or material barriers (through farm machinery cooperatives or other groups of farmers, for example). The association we showed between the presence of cereal–legume intercrops and organic farming, feed autonomy for the livestock and on-farm grain storage can refer to the use of cereal–legume intercrops. In organic farming, outlets exist in long distribution networks for binary cereal–legume intercrops. Such outlets are still lacking in conventional farming, so intercrops are mainly used on farms for animal feed, or possibly on-farm transformation. At the supply chain scale, for cereal–legume intercropping to diffuse beyond organic and/or livestock farms, it seems necessary to secure the outlets and develop market opportunities.

4.4. Relevance of our method and limits of the study

Our study had some limitations and may raise questions about our method choices. First, our study is based on a random forest approach to identify the most important variables in discriminating the Presence from the Absence cases. However, one limit of the random forest approach is that we could not directly interpret the contribution of each variable to the classification in the Presence or Absence groups. An alternative method for our aim would be to use a logistic regression, which allows a binary classification of the data and is easier to interpret as the estimates given for each predictor variable can reflect its positive or negative impact on the probability of finding cereal–legume intercrops on farms. However, logistic regressions assume independence between the predictor variables (42 variables detailed in Table 1), and a linear relationship between those predictor variables and the logit of the outcome (in our case, the probability of presence of cereal–legume intercrops). Those assumptions were not met in our dataset, and we used variables that might be correlated, as we wanted to rank the 42 predictor variables and also understand their potential interactions to characterise the farms that adopt cereal–legume intercrops. Therefore, the logistic regression does not seem ideal to address our research questions. Additionally, Couronné et al. (2018) conducted a comparative study between logistic regression and random forest and concluded that random forest "copes better with large numbers of features" and "performs better than [logistic regressions] in the presence of a non-linear dependence pattern between features and response".

Another element we wanted to discuss is the oversampling method (presented in Section 2.3) that was used to penalise the CART model by giving more weight to the minority observations, i.e., those in the Presence group. To do this, we chose to replicate these observations, which carried a risk of overfitting as the CART model could potentially learn the data too well and give a biased result. To counter this risk, we added noise to two randomly selected variables for each replicate. For instance, we modified the value of the UAA and the value of the proportion of ploughing while keeping the values within the actual range of the data. In the end, the CART obtained was very similar to the tree shown in Fig. 3, but with an additional node on cattle density. However, after analysis, we noticed that this node was empty, meaning that no real data was classified by this node. The node was created because of the applied noise, which ended up giving weight to a variable with inconsistent values (Supplementary material 1).

To gain a better understanding of the relationships between the variables identified by the random forest and the presence of cereal–legume intercrops at the national level, we chose to use the first 10 variables to construct a CART as we noted a "gap" in importance values in Fig. 2. However, we were aware that such choice might have biased the results obtained with the CART. Therefore, a further test was conducted to validate the robustness of the approach, which involved building the CART with 36 variables selected using a recursive selection algorithm developed by Kuhn (2008). Similar results were obtained, with only the node for the type of farming missing (Supplementary material 2). This result provides further evidence that our approach is robust for this application.

Furthermore, we strove to accurately translate the factors identified in the literature into variables for analysis, but our database did not allow us to translate them exactly. For example, sorting was frequently cited as an obstacle to the adoption of cereal–legume intercrops, but we conveyed it using variables related to cooperatives, on-farm grain storage, and on-farm transformation because the 2020 French Agricultural Census did not specifically address crop sorting (whether it be on-farm sorting, with a shared sorter in a farm machinery cooperative, or by a professional sorter). Finally, it is difficult to see how each of the identified variables operates and interacts with the others, especially for those not covered by the CART. In this respect, our results should be supplemented by more in-depth investigations into the rationale behind farmers' actions, perhaps in the three areas we identified in Western, Eastern, and Southern France.

5. Conclusion and perspectives

In this study, we demonstrated existing links between cereal–legume intercropping and various agricultural or farm characteristics on French farms, aiming to be as exhaustive as possible by using data from the French Agricultural Census. The most important characteristics were organic farming, the presence of livestock, on-farm grain storage capacity, and belonging to a farm machinery cooperative. Our analysis confirmed the hypothesis that organic farming and livestock production were strongly associated with the adoption of cereal–legume intercrops both at the national level and in smaller, dynamic areas.

Our findings raise questions about transferring the practice to other farm models: should cereal–legume intercropping expand to other types of farming than organic farming and livestock farming? If so, how? Our study shows that cereal–legume intercrops are particularly beneficial for feed self-sufficiency, which supports their development in conjunction with crop–livestock interactions.

In France, some agricultural cooperatives and agricultural and food industries are beginning to study this practice in conventional farming circuits. Identifying the farm characteristics associated with the presence of cereal–legume intercrops can improve our understanding of their relevance for the farms' activities and thus allow us to promote them efficiently, whether at the national level (e.g., through public policies) or local levels, e.g., with the involvement of agricultural or farming support services. However, it is not possible to conclude with certainty that these characteristics can be used as levers to foster the adoption of cereal–legume intercrops. In this context, case studies on areas with observed dynamics – such as the Western, Eastern, and Southern areas identified in this study – would complement our results and prove highly valuable for a more in-depth study of the farmers' pathways that led to the adoption of the practice as well as the roles of the various factors highlighted here in these pathways.

The present study focused on a single practice, namely cereal–legume intercropping, to gain a better understanding of the apparent contradiction between the benefits it provides in the context of agroecological transition and its low adoption rate in France. We believe that the methodology proposed here can be reproduced for other practices to identify links with other practices or farm characteristics.

CRediT authorship contribution statement

Elodie Yan: Writing – review & editing, Writing – original draft, Methodology, Formal analysis, Conceptualization. **Philippe Martin:** Writing – review & editing, Conceptualization. **Marco Carozzi:** Writing – review & editing, Conceptualization.

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.

Data availability

The authors do not have permission to share data.

Acknowledgements

The authors would like to thank Pr. Antoine Cornuéjols for his advice on the models and their validation. Access to some confidential data, on which is based this work, has been made possible within a secure environment offered by CASD – Centre d'accès sécurisé aux données (Ref. 10.34724/CASD). This study is part of the MoBiDiv project (www.inrae.fr/mobidiv_eng/), funded by the ANR program PPR-CPA (ANR-20-PCPA-0006).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.agsy.2024.104196.

References

Agreste, 2020. Cultures développées (hors fourrage, prairies, fruits, fleurs et vigne). https://agreste.agriculture.gouv.fr/agreste-saiku/?plugin=true&query=query/open/SAANR_DEVELOPPE_2#query/open/SAANR_DEVELOPPE_2.

Altieri, M.A., Nicholls, C.I., Henao, A., Lana, M.A., 2015. Agroecology and the design of climate change-resilient farming systems. Agron. Sustain. Dev. 35 (3), 869–890. https://doi.org/10.1007/s13593-015-0285-2.

Bedoussac, L., Journet, E.-P., Hauggaard-Nielsen, H., Naudin, C., Corre-Hellou, G., Jensen, E.S., Prieur, L., Justes, E., 2015. Ecological principles underlying the increase of productivity achieved by cereal-grain legume intercrops in organic farming. A review. Agron. Sustain. Dev. 35 (3), 911–935. https://doi.org/10.1007/s13593-014-0277-7.

Bonke, V., Musshoff, O., 2020. Understanding German farmer's intention to adopt mixed cropping using the theory of planned behavior. Agron. Sustain. Dev. 40 (6), 48. https://doi.org/10.1007/s13593-020-00653-0.

Breiman, L., 2001. Random forests. Mach. Learn. 45 (1), 5–32. https://doi.org/10.1023/A:1010933404324.

Breiman, L., Friedman, J., Olshen, R.A., Stone, C.J., 1984. Classification and Regression Trees. Routledge. https://doi.org/10.1201/9781315139470.

Carton, N., Naudin, C., Piva, G., Corre-Hellou, G., 2020. Intercropping winter lupin and triticale increases weed suppression and total yield. Agriculture 10 (8). https://doi.org/10.3390/agriculture10080316. Article 8.

Casagrande, M., Alletto, L., Naudin, C., Lenoir, A., Siah, A., Celette, F., 2017. Enhancing planned and associated biodiversity in French farming systems. Agron. Sustain. Dev. 37 (6), 57. https://doi.org/10.1007/s13593-017-0463-5.

CERESCO, Systra, 2020. Réalisation d'une étude sur l'évaluation des coûts de la chaîne logistique céréalière française.

Chantre, E., Cardona, A., 2014. Trajectories of French field crop farmers moving toward sustainable farming practices : change, learning, and links with the advisory services. Agroecol. Sustain. Food Syst. 38 (5), 573–602. https://doi.org/10.1080/21683565.2013.876483.

Chen, C., Liaw, A., Breiman, L., 2004. Using Random Forest to Learn Imbalanced Data, vol. 110(24). University of California, Berkeley, pp. 1–12.

Clouet, Y., Guilloneau, A., Ruf, T., 1986. Diagnostic du système agraire et des systèmes de production en Ségala Aveyronnais.

Corre-Hellou, G., Dibet, A., Hauggaard-Nielsen, H., Crozat, Y., Gooding, M., Ambus, P., Dahlmann, C., von Fragstein, P., Pristeri, A., Monti, M., Jensen, E.S., 2011. The competitive ability of pea–barley intercrops against weeds and the interactions with crop productivity and soil N availability. Field Crop Res. 122 (3), 264–272. https://doi.org/10.1016/j.fcr.2011.04.004.

Corre-Hellou, G., Bedoussac, L., Bousseau, D., Chaigne, G., Chataignier, C., Celette, F., Cohan, J.-P., Coutard, J.-P., Emile, J.-C., Floriot, M., Foissy, D., Guilbert, S., Hemptinne, J.-L., Le Breton, M., Lecompte, C., Marceau, C., Mazoué, F., Mérot, E., Métivier, T., Tauvel, O., 2013. Associations céréale-légumineuses multi-services.

Couronné, R., Probst, P., Boulesteix, A.-L., 2018. Random forest versus logistic regression : a large-scale benchmark experiment. BMC Bioinform. 19 (1), 270. https://doi.org/10.1186/s12859-018-2264-5.

Darnhofer, I., Bellon, S., Dedieu, B., Milestad, R., 2010. Adaptiveness to enhance the sustainability of farming systems. A review. Agron. Sustain. Dev. 30 (3), 545–555. https://doi.org/10.1051/agro/2009053.

Duru, M., Therond, O., Fares, M., 2015. Designing agroecological transitions; A review. Agron. Sustain. Dev. 35 (4), 1237–1257. https://doi.org/10.1007/s13593-015-0318-x.

Fares, M., Mamine, F., 2023. Relative importance of barriers and levers to intercropping systems adoption : a comparison of farms and co-operatives. Sustainability 15 (8). https://doi.org/10.3390/su15086652. Article 8.

Fernández, A., García, S., Galar, M., Prati, R.C., Krawczyk, B., Herrera, F., 2018. Learning from Imbalanced Data Sets. Springer Cham. https://doi.org/10.1007/978-3-319-98074-4.

Foley, J.A., Ramankutty, N., Brauman, K.A., Cassidy, E.S., Gerber, J.S., Johnston, M., Mueller, N.D., O'Connell, C., Ray, D.K., West, P.C., Balzer, C., Bennett, E.M., Carpenter, S.R., Hill, J., Monfreda, C., Polasky, S., Rockström, J., Sheehan, J., Siebert, S., Zaks, D.P.M., 2011. Solutions for a cultivated planet. Nature 478 (7369). https://doi.org/10.1038/nature10452. Article 7369.

Galioto, F., Nino, P., 2023. Investigating the reasons behind the choice to promote crop diversification practices through the new CAP reform in Europe. Land Use Policy 133, 106861. https://doi.org/10.1016/j.landusepol.2023.106861.

Ha, T.M., Manevska-Tasevska, G., Jäck, O., Weih, M., Hansson, H., 2023. Farmers' intention towards intercropping adoption: the role of socioeconomic and behavioural drivers. Int. J. Agric. Sustain. 21 (1), 2270222. https://doi.org/10.1080/14735903.2023.2270222.

Himanen, S.J., Mäkinen, H., Rimhanen, K., Savikko, R., 2016. Engaging farmers in climate change adaptation planning : assessing intercropping as a means to support farm adaptive capacity. Agriculture 6 (3). https://doi.org/10.3390/agriculture6030034. Article 3.

IGN, 2023. Institut National de l'Information Géographique et Forestière—IGN. Base de Données. https://geoservices.ign.fr/telechargement.

Jensen, E.S., Carlsson, G., Hauggaard-Nielsen, H., 2020. Intercropping of grain legumes and cereals improves the use of soil N resources and reduces the requirement for synthetic fertilizer N : a global-scale analysis. Agron. Sustain. Dev. 40 (1), 5. https://doi.org/10.1007/s13593-020-0607-x.

Kontturi, M., Laine, A., Niskanen, M., Hurme, T., Hyövelä, M., Peltonen-Sainio, P., 2011. Pea–oat intercrops to sustain lodging resistance and yield formation in northern European conditions. Acta Agric. Scand. Sect. B — Soil Plant Sci. 61 (7), 612–621. https://doi.org/10.1080/09064710.2010.536780.

Kuhn, M., 2008. Building predictive models in R using the caret package. J. Stat. Softw. 28, 1–26. https://doi.org/10.18637/jss.v028.i05.

Lemken, D., Spiller, A., von Meyer-Höfer, M., 2017. The case of legume-cereal crop mixtures in modern agriculture and the Transtheoretical model of gradual adoption. Ecol. Econ. 137, 20–28. https://doi.org/10.1016/j.ecolecon.2017.02.021.

Leoni, F., Lazzaro, M., Ruggeri, M., Carlesi, S., Meriggi, P., Moonen, A.C., 2022. Relay intercropping can efficiently support weed management in cereal-based cropping systems when appropriate legume species are chosen. Agron. Sustain. Dev. 42 (4), 75. https://doi.org/10.1007/s13593-022-00787-3.

Li, C., Stomph, T.-J., Makowski, D., Li, H., Zhang, C., Zhang, F., van der Werf, W., 2023. The productive performance of intercropping. Proc. Natl. Acad. Sci. 120 (2), e2201886120. https://doi.org/10.1073/pnas.2201886120.

Lithourgidis, A.S., Vlachostergios, D.N., Dordas, C.A., Damalas, C.A., 2011. Dry matter yield, nitrogen content, and competition in pea–cereal intercropping systems. Eur. J. Agron. 34 (4), 287–294. https://doi.org/10.1016/j.eja.2011.02.007.

Lopes, T., Hatt, S., Xu, Q., Chen, J., Liu, Y., Francis, F., 2016. Wheat (Triticum aestivum L.)-based intercropping systems for biological pest control. Pest Manag. Sci. 72 (12), 2193–2202. https://doi.org/10.1002/ps.4332.

Magrini, M.-B., Anton, M., Cholez, C., Corre-Hellou, G., Duc, G., Jeuffroy, M.-H., Meynard, J.-M., Pelzer, E., Voisin, A.-S., Walrand, S., 2016. Why are grain-legumes rarely present in cropping systems despite their environmental and nutritional benefits? Analyzing lock-in in the French agrifood system. Ecol. Econ. 126, 152–162. https://doi.org/10.1016/j.ecolecon.2016.03.024.

Mamine, F., Farès, M., 2020. Barriers and Levers to Developing Wheat–Pea Intercropping in Europe: A Review. Sustainability 12 (17). https://doi.org/10.3390/su12176962. Article 17.

Mandrekar, J.N., 2010. Receiver operating characteristic curve in diagnostic test assessment. J. Thorac. Oncol. 5 (9), 1315–1316. https://doi.org/10.1097/JTO.0b013e3181ec173d.

Mansion-Vaquié, A., Ferrer, A., Ramon-Portugal, F., Wezel, A., Magro, A., 2020. Intercropping impacts the host location behaviour and population growth of aphids. Entomol. Exp. Appl. 168 (1), 41–52. https://doi.org/10.1111/eea.12848.

Navarrete, M., Dupré, L., Lamine, C., 2015. Crop management, labour organization, and marketing : three key issues for improving sustainability in organic vegetable farming. Int. J. Agric. Sustain. 13 (3), 257–274. https://doi.org/10.1080/14735903.2014.959341.

Nie, Z., McLean, T., Clough, A., Tocker, J., Christy, B., Harris, R., Riffkin, P., Clark, S., McCaskill, M., 2016. Benefits, challenges and opportunities of integrated crop-livestock systems and their potential application in the high rainfall zone of southern Australia: a review. Agric. Ecosyst. Environ. 235, 17–31. https://doi.org/10.1016/j.agee.2016.10.002.

Peigné, J., Casagrande, M., Payet, V., David, C., Sans, F.X., Blanco-Moreno, J.M., Cooper, J., Gascoyne, K., Antichi, D., Bàrberi, P., Bigongiali, F., Surböck, A., Kranzler, A., Beeckman, A., Willekens, K., Luik, A., Matt, D., Grosse, M., Heß, J., Mäder, P., 2016. How organic farmers practice conservation agriculture in Europe. Renew. Agric. Food Syst. 31 (1), 72–85. https://doi.org/10.1017/S1742170514000477.

Pelzer, E., Bazot, M., Makowski, D., Corre-Hellou, G., Naudin, C., Al Rifaï, M., Baranger, E., Bedoussac, L., Biarnès, V., Boucheny, P., Carrouée, B., Dorvillez, D., Foissy, D., Gaillard, B., Guichard, L., Mansard, M.-C., Omon, B., Prieur, L., Yvergniaux, M., Jeuffroy, M.-H., 2012. Pea–wheat intercrops in low-input conditions combine high economic performances and low environmental impacts. Eur. J. Agron. 40, 39–53. https://doi.org/10.1016/j.eja.2012.01.010.

Quinlan, J.R., 1992. C4.5: Programs for Machine Learning. Morgan Kaufmann.

R Core Team, 2023. R: A Language and Environment for Statistical Computing [Logiciel]. R Foundation for Statistical Computing. https://www.R-project.org/.

Raseduzzaman, Md., Jensen, E.S., 2017. Does intercropping enhance yield stability in arable crop production? A meta-analysis. Eur. J. Agron. 91, 25–33. https://doi.org/10.1016/j.eja.2017.09.009.

Rosa-Schleich, J., Loos, J., Mußhoff, O., Tscharntke, T., 2019. Ecological-economic trade-offs of diversified farming systems – a review. Ecol. Econ. 160, 251–263. https://doi.org/10.1016/j.ecolecon.2019.03.002.

Sing, T., Sander, O., Beerenwinkel, N., Lengauer, T., 2005. ROCR : visualizing classifier performance in R. Bioinformatics 21 (20), 3940–3941. https://doi.org/10.1093/bioinformatics/bti623.

Therneau, T., Atkinson, B., Port, B. R. (producer of the initial R., & maintainer 1999–2017), 2023. rpart: Recursive Partitioning and Regression Trees (Version 4.1.23) [Logiciel]. https://cran.r-project.org/web/packages/rpart/index.html.

Thomopoulos, R., Moulin, B., Bedoussac, L., 2018. Supporting decision for environment-friendly practices in the Agri-food sector : when argumentation and system dynamics simulation complete each other. Intern. J. Agric. Environ. Inform. Syst. 9, 1–21. https://doi.org/10.4018/IJAEIS.2018070101.

Timaeus, J., Ruigrok, T., Siegmeier, T., Finckh, M.R., 2022. Adoption of food species mixtures from Farmers' perspectives in Germany: managing complexity and harnessing advantages. Agriculture 12 (5). https://doi.org/10.3390/agriculture12050697. Article 5.

Verret, V., Pelzer, E., Bedoussac, L., Jeuffroy, M.-H., 2020. Tracking on-farm innovative practices to support crop mixture design: the case of annual mixtures including a legume crop. Eur. J. Agron. 115, 126018. https://doi.org/10.1016/j.eja.2020.126018.

Wezel, A., Herren, B.G., Kerr, R.B., Barrios, E., Gonçalves, A.L.R., Sinclair, F., 2020. Agroecological principles and elements and their implications for transitioning to sustainable food systems. A review. Agron. Sustain. Dev. 40 (6), 40. https://doi.org/10.1007/s13593-020-00646-z.

Willey, R.W., 1979. Intercropping : its importance and research need. I. Competition and yield advantages. Field Crop Abstracts 32, 1–10.

Yan, E., Munier-Jolain, N., Martin, P., Carozzi, M., 2024. Intercropping on French farms : reducing pesticide and N fertiliser use while maintaining gross margins. Eur. J. Agron. 152, 127036. https://doi.org/10.1016/j.eja.2023.127036.