RESEARCH ARTICLE

Predictive modeling of vulnerability level of French irrigated farms

F. Ben Brahim[1,2] · I. Kedidi[2] · M-R. Randriamarolo[2] · H. Randrianosolo[2,3]

Abstract

Climate change has significantly increased the complexity of farm management. This study investigates the vulnerability of French irrigated farms to agricultural risks by developing predictive models using a dataset of 631 farms in metropolitan France from 2020 to 2022, sourced from the Farm Accountancy Data Network (FADN). The vulnerability level is categorized into three classes: low, moderate, and high, based on four indicators that measure economic stability and resilience. We employed several machine learning algorithms to identify the most accurate predictive model for farm vulnerability, with Random Forests achieving the highest accuracy. The analysis revealed that irrigation water costs per irrigated area, farm technical-economic orientation, and geographical location are the most critical determinants of vulnerability. Partial Dependence Plots further highlighted the marginal effects of these factors on the predicted high vulnerability level. Our findings indicate that the farm's location, technical-economic orientation, and irrigation parameters play a role in determining vulnerability levels. These results provide valuable insights for policymakers and farmers in developing strategies to enhance the resilience of agricultural systems in the face of climate change.

Keywords Climate change . Vulnerability level . French irrigated farms . Predictive Modeling . Random Forest . Farm management

---

✉ Ferdaous Ben Brahim
ferdaous.benbrahim@agroparistech.fr

Extended author information available on the last page of the article

Introduction

Making choices to managing the variety of environmental challenges is not an easy objective to achieve. In the agricultural sector, farm management has become increasingly complex due to climate change.

According to the sixth Intergovernmental Panel On Climate Change (IPCC) report published in March 2023, Socio-economic scenarios indicate that the global warming level of 1.5°C above pre-industrial levels will be reached by the early 2030s, regardless of immediate global CO2 emission reduction efforts. In agriculture, climate change has already decreased productivity over the past 50 years. As global warming progresses, climate change impacts will intensify, affecting temperature extremes, precipitation intensity, drought severity, and the frequency and intensity of rare climatic events.

In Europe, agriculture bears the brunt of the region's drought-related losses with a 53% share, representing 1.3 billion euros in annual economic losses (European Commission. Joint Research Center. 2020). Nevertheless, in the event of an extreme weather event, as was the case with the 2003 drought, which cost the European Union 13 billion euros in economic losses, of which 4 billion euros went to France (Létard, Flandre, and Lepeltier 2004). The question of how to manage climate risk and adapt agriculture is brought up by these extraordinary events. In January 2024, the Chair of Risk Management in Agriculture conducted a survey titled "Risk Management in Agriculture in a Context of Vulnerability, with a Focus on Water" among French farmers. The responses to the open-ended questions revealed a significant concern about the threat of drought (Ben Brahim et al., 2024).

The degree to which a farming system is susceptible, or unable to cope with all the climatic risks which, according to Komarek (2020), are closely linked with other agricultural risks (Price, institutional, financial, human …) and can positively or negatively influence them is its vulnerability that differs from farm to farm (IPCC, 2001). In all formulations, vulnerability is defined by three key parameters: the stress to which a system is exposed, its sensitivity, and its adaptive capacity (Adger, 2006). According to Urruty et al. (2016), the level of exposure corresponds to the frequency, intensity, and duration of perturbations affecting the studied systems. The level of sensitivity corresponds to the degree to which the studied system is affected by exposure to perturbations and the adaptive capacity corresponds to the ability of agricultural systems to transform their nature or structure to cope with an ever-changing environment (Milestad et al., 2012). When determining the level of vulnerability in a farming system, the primary objective is to gather information and create policies that enable the identification of potential sources of vulnerability, their mitigation, and the long-term preservation of farms (Sneessens et al., 2019).

An increasing number of publications have studied vulnerability to climate change, and the agricultural sector is often described as one of the most vulnerable ones (Neset et al., 2019). According to M. G. Debesai (2020), various socioeconomic, biophysical and environmental factors influence vulnerability levels in Eritrea, due to climate change. In his paper, some variables such as dependency ratio, profession, and access to clean water were found to have a negative relationship with the vulnerability level indicating that the vulnerability is highly likely to occur as these variables decrease. On the other hand, variables including gender, low level of education and income diversity, poverty level, low access to credit and market were found to have a positive relationship with the high level of vulnerability. Differences in socio-demographic factors of households such as level of education, gender and profession exhibited different levels of vulnerability. Income diversity and level of poverty were also significantly determining the level of vulnerability in farming households. Location was also a factor in determining the level of vulnerability as the different zones indicate significant variability to the degree of vulnerability. According to P. Marie Chimi et al. (2023), multiple factors including socio-economic and infrastructure conditions, crop types, and climate variables such as temperature and rainfall influence vulnerability levels of smallholder farmers to climate change in the northern part of the Centre Region of Cameroon. Finally, geographically speaking and according to Mirza (2003), farms located in arid or semi-arid areas, low-altitude coastal zones, flood-prone areas, or small islands are particularly the most vulnerable.

The perception of risk by the farmer plays an important role in the choice of the strategies adopted by him to manage the risk faced by his farm. Studies have shown that farmers' awareness of risks has increased in recent years (OECD, 2012). When the farmer perceives a risk to their farm, they will attempt to adjust it to make it manageable. At this stage, the farmer begins with prevention using irrigation techniques for example and has two fundamental techniques to alter their level of risk which are diversification or insurance. The measures adopted by the farmer usually depend on their risk aversion, level of equity, and the characteristics of their land (Kapsambelis, 2022). Studies have shown that the more farmers choose to implement preventive measures such as irrigation, the less likely they are to subscribe to insurance contracts (Enjorlas and Sentis, 2011; Blank and mcDonald, 1995; Smith and Goodwin, 1996).

There is a long list of adapting production systems and preventive measures in response to new challenges, especially those imposed by climate change among which irrigation that acts as insurance against the risk of drought and can be integrated as a measure to anticipate the risk of water shortage (Amigues et al., 2006). The strengthening of irrigation appears to be the primary response of farmers to climate change, especially for farms in the south of France (Ayphassorho et al., 2020). Thus, irrigation often appears as the first response to the risk of drought, but it is not necessarily the most desirable solution. Indeed, this preventive measure is criticized by a group of stakeholders advocating for the preservation of water resources and natural areas (Kapsambelis, 2022). This technique involves several parameters such as irrigation costs which are defined according to fixed costs (depreciation, equipment maintenance, membership fees, etc.), variable costs (water and electricity consumption), labor costs. Those costs depend on the type of the farm, its size, the equipment used, and the method of water withdrawal (Salmon, 2020). The other parameter are water resources which are divided into two networks, individual water networks include various sources tailored to local needs. Hill reservoirs, ponds, and water tanks not connected to a watercourse allow for the storage of rainwater for irrigation and domestic use, especially in rural areas. Groundwater, obtained through drilling or wells, provides a continuous water supply but requires pumping infrastructure. Surface water, such as rivers, canals, and lakes, is also used, though its availability varies according to local regulations. Collective networks encompass public infrastructure for large-scale distribution. Alternative sources like rainwater harvesting and greywater reuse are encouraged for sustainable management (FAO, 2021). Among these parameters we have also irrigation method. Micro-irrigation, for example, distributes water in a planned manner at low pressure and close to the plants, primarily used in orchards and market gardening. Sprinkler irrigation, on the other hand, sprays water like artificial rain, making it easy to implement depending on the crops and terrain. Lastly, gravity irrigation allows water to flow on the surface through small channels, mainly used on sloped terrain (SDES, 2024). Each method has its specific characteristics and use conditions.

In this study, we consider, according to the literature, determinants of vulnerability such as the farm’s geographical location and its technical-economic orientation. However, we test if the farmer’s choices regarding irrigation contribute to the farm’s vulnerability level. Also, we aim to fit a predictive model that quantifies the contribution of each determinant on the overall vulnerability.

The paper is organized into five sections. The « Data & descriptive statistics » section presents our data source and describes the variables that we used in this study. The « Exploratory Data Analysis » section to visualize our data. The « Method » section describes our methodology to test our variables importance and fit the predictive model in detail. The « Results » section presents our results. Finally, we discuss these results in the last section of the paper.

Data & descriptive statistics

In our study, we will examine 631 irrigated farms in metropolitan France during a three-year period from 2020 to 2022. Data is sourced from the Farm Accountancy Data Network (FADN) which was implemented in France since 1968 and is based on an annual survey conducted in all Member States of The European Union according to common rules. Information on the status of the farm, economic data, and cultivated crops is provided for each farm. The French FADN is designed to ensure that the sample is representative of a set of farms and consists of almost 7000 farms per year. The data is anonymized to prevent the identification of any specific farm within the network. Therefore, in terms of the farm's location, the finest scale provided is the administrative region.

Study variables are presented in Table 1 below. The outcome variable is vulnerability level, categorized in three levels ranging from low, moderate, high.

To calculate the vulnerability level of agricultural farms, we adopted the method by Sneessens et al. (2019). This method relies on four indicators designed to assess vulnerability according to the definition proposed by the Intergovernmental Panel on Climate Change (IPCC) (Figure 1).

The first indicator is the relative standard deviation of annual pre-tax operating income per worker over a long period. This indicator provides insights into the sensitivity and exposure of each farm to risks. The mathematical formula for this indicator is written as follows:

$$RSD^{CR.LU}\ (\%) = \left| \frac{SD^{CR.LU}}{\mu^{CR.LU}} \right| * 100$$

The second indicator is the average relative distance of annual consolidated pre-tax operating income to the minimum wage. This indicator adds a social dimension to vulnerability measurement.

The mathematical formula for this indicator is written as follows:

$$RD^{CR.LU}\ (\%) = \text{mean}\left(\frac{CR.LU - MIN}{MIN}\right) * 100$$

The third indicator is the number of economic disruptions, which corresponds to the instances where annual consolidated pre-tax operating income per worker decreases by more than 25% from one year to the next. This measures the adaptive capacity of farms.

Finally, the fourth indicator is the economic recovery time, which corresponds to the number of years required to restore annual consolidated pre-tax operating income per worker to pre-disruption levels. This indicator qualifies the resilience of a farm.

Combining the results obtained for the four vulnerability indicators through Hierarchical Cluster Analysis allows us to identify three groups of agricultural systems. Statistical analysis of these groups subsequently defines the three levels of vulnerability for each identified group.

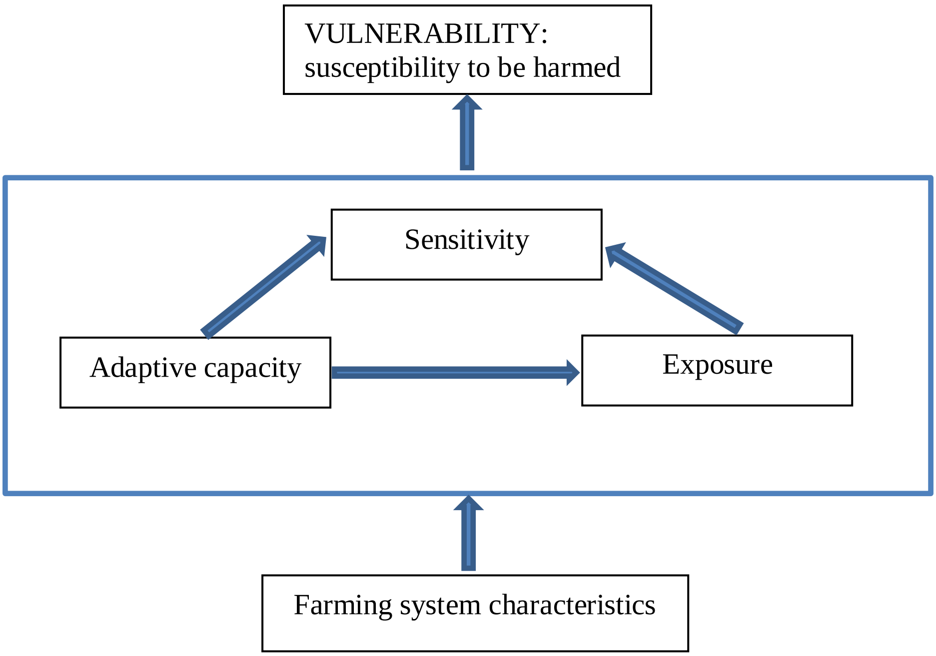

Figure 1. Components of vulnerability (Source: IPCC, 2001).

Table 1. Description of study variables

Variable	Description
Outcome variable	
Vulnerability level	1: Slightly vulnerable
	2: Moderately vulnerable
	3: Highly vulnerable
Independent variables	
Disadvantaged zone code	1: Majority of the farm is not located in a disadvantaged zone
	21: Majority of the farm is located in an area subject to natural constraints
	22: Majority of the farm is located in an area subject to specific constraints
	3: Majority of the farm is located in a mountainous area
Code for main source for water irrigation	1: Individual network (hill reservoirs, ponds, water reservoirs, not connected to a watercourse)
	2: Individual network (groundwater: wells, boreholes)
	3: Individual network (surface water: streams, canals, lakes)
	4: Collective Networks
	5: Individual network (other sources)
Irrigation water costs per irrigated area	Continuous variable
Code for main irrigation method	1: Surface irrigation
	2: Sprinkler irrigation
	3: Micro-irrigation

Table 1 (Continued). Description of study variables

Variable	Description
Independent variables	
Region Code	Categorical variable with categories representing the 13 regional codes of metropolitan France
Technical-economic orientation	1500: Cereal cultivation and cultivation of oilseed and protein crops 1600: Other major crops 2800: Vegetable and mushroom cultivation 2900: Flower and diverse horticulture cultivation 3500: Viticulture (grape cultivation) 3900: Fruit cultivation and other permanent crops 4500: Dairy cattle farms 4600: Beef cattle farms 4700: Dairy, beef, and meat production 4813: Sheep and goat farms 4840: Sheep, goats, and other herbivores 5100: Pig farms 5200: Poultry farms 5374: Various combinations of grain-eating animals 6184: Mixed crop-livestock farming

To provide a comprehensive overview of the dataset, we present summary statistics for our study variables.

Table 2 provides a summary of the categorical variables. The total number of farms were 631 out of which a majority of them, 32%, in Normandy. Nearly half of the farms, 49.13%, are classified as highly vulnerable. Additionally, 49.92% of the farms are not situated in disadvantaged zones. In terms of irrigation practices, 50.24% use micro-irrigation, while 55.94% rely on collective networks as their irrigation water source. Regarding specialization, the majority of the sample, 27.73%, focus on fruit cultivation and other permanent crops.

Summary statistics of the continuous variable "Irrigation costs per irrigated area" are presented in the table 3 down below.

Table 3. Summary statistics of continuous variables

Variable	**Min/Max**	**Mean (SD)**
Irrigation costs per irrigated area	0/17 767	860.9 (1407.883)

The minimum amount of irrigation costs per irrigated area is 0. This may be due to the fact that some farms might have access to free water sources, such as rivers and lakes. The average amount is 860.9 with a standard deviation of 1407.883 which suggests that there is a wide variation in irrigation costs among different farms.

Exploratory Data Analysis (EDA)

Exploratory Data Analysis (EDA) serves as a crucial initial step in data science projects, involving the scrutiny and visualization of data to grasp its fundamental characteristics, uncover patterns, and discern relationships between variables. It encompasses studying and exploring datasets to comprehend their primary traits, reveal patterns, pinpoint outliers, and recognize connections among variables, typically preceding more formal statistical analyses or modeling endeavors. EDA holds considerable significance for several reasons, including facilitating familiarity with the dataset, identifying patterns and relationships, detecting anomalies and outliers, informing feature selection and engineering, optimizing model design, facilitating data cleaning, and enhancing communication of findings (Joshi, Bhargava, & Aggarwal, 2020).

In our analysis, we conducted univariate analysis, focusing on the distribution of variables like vulnerability levels, main source of irrigation water, main irrigation method and disadvantaged zone code. These visualizations (Figure 2) offer a detailed overview of the farm cohort, reflecting the complex interplay of variables. Each chart provides a distinct perspective on the diversity of farm characteristics within the dataset.

Table 2. Summary statistics of categorical variables

Variables		
Categorical Variables	Frequency	Percentage
Vulnerability level		
1	64	10.14%
2	257	40.73%
3	310	49.13%
Region Code		
76	202	32%
75	119	18.86
93	91	14.42%
84	60	9.5%
94	40	6.34%
52	38	6.02%
(Other)	81	12.83%
Disadvantaged zone code		
1	315	49.92%
3	66	10.46%
21	108	17.11%
22	142	22.5%

Table 2 (Continued). Summary statistics of categorical variables

Variable		
Categorical Variables	Frequency	Percentage
Code for main irrigation method		
1	39	6.18%
2	275	43.58%
3	317	50.24%
Code for main source of irrigation water		
1	48	7.6%
2	153	24.25%
3	72	11.41%
4	353	55.94%
5	5	0.79%
Technical-economic orientation		
3900	175	27.73%
3500	96	15.21
1500	82	13%
2800	63	10%
6184	60	9.5%
2900	58	9.2%
(Other)	97	15.37%

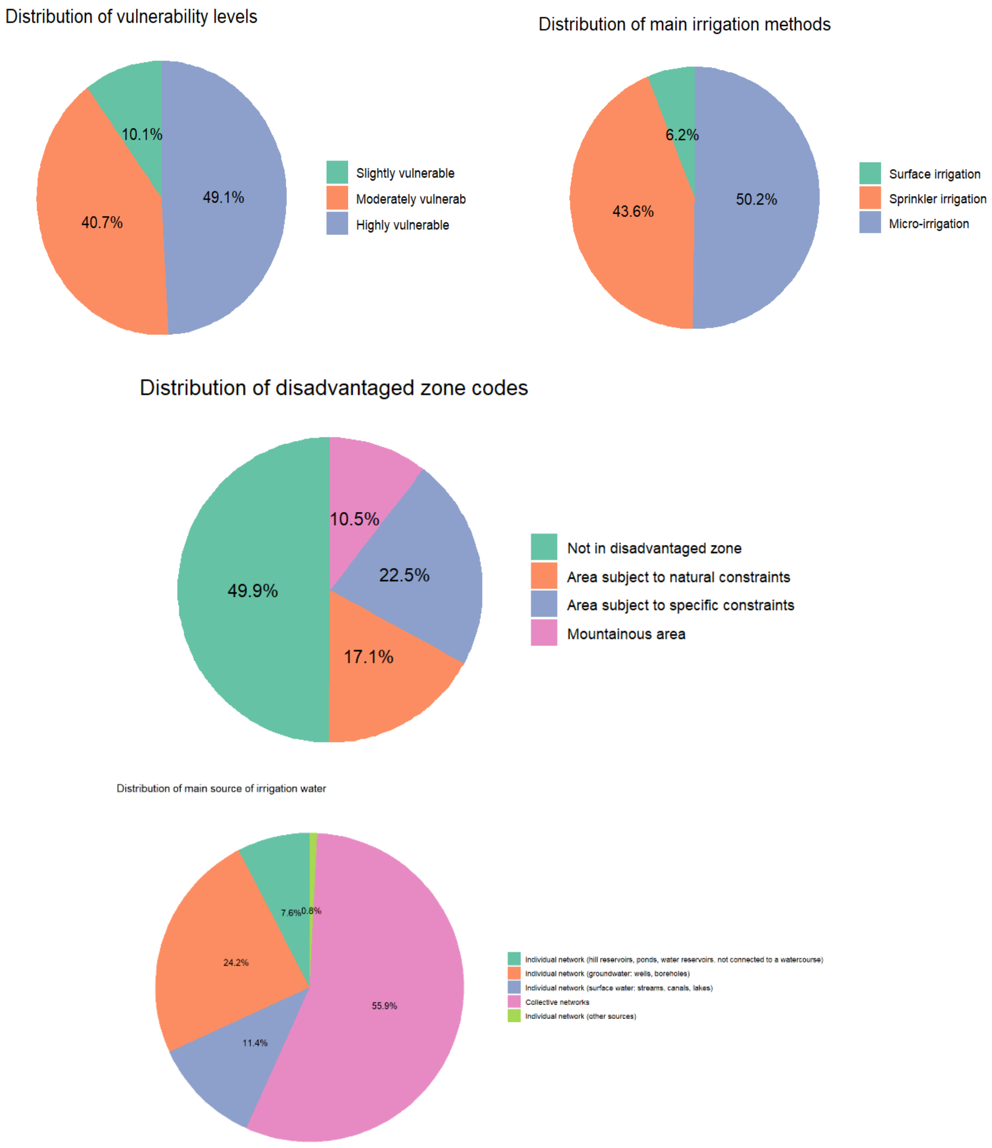

Figure 2. Pie chart distribution of four categorical variables from our dataset.

Next, we employed bivariate analysis to explore and quantify the relationships between pairs of variables (Figure 3). This approach was integral to our exploratory data analysis, enabling us to uncover potential associations that could inform further research.

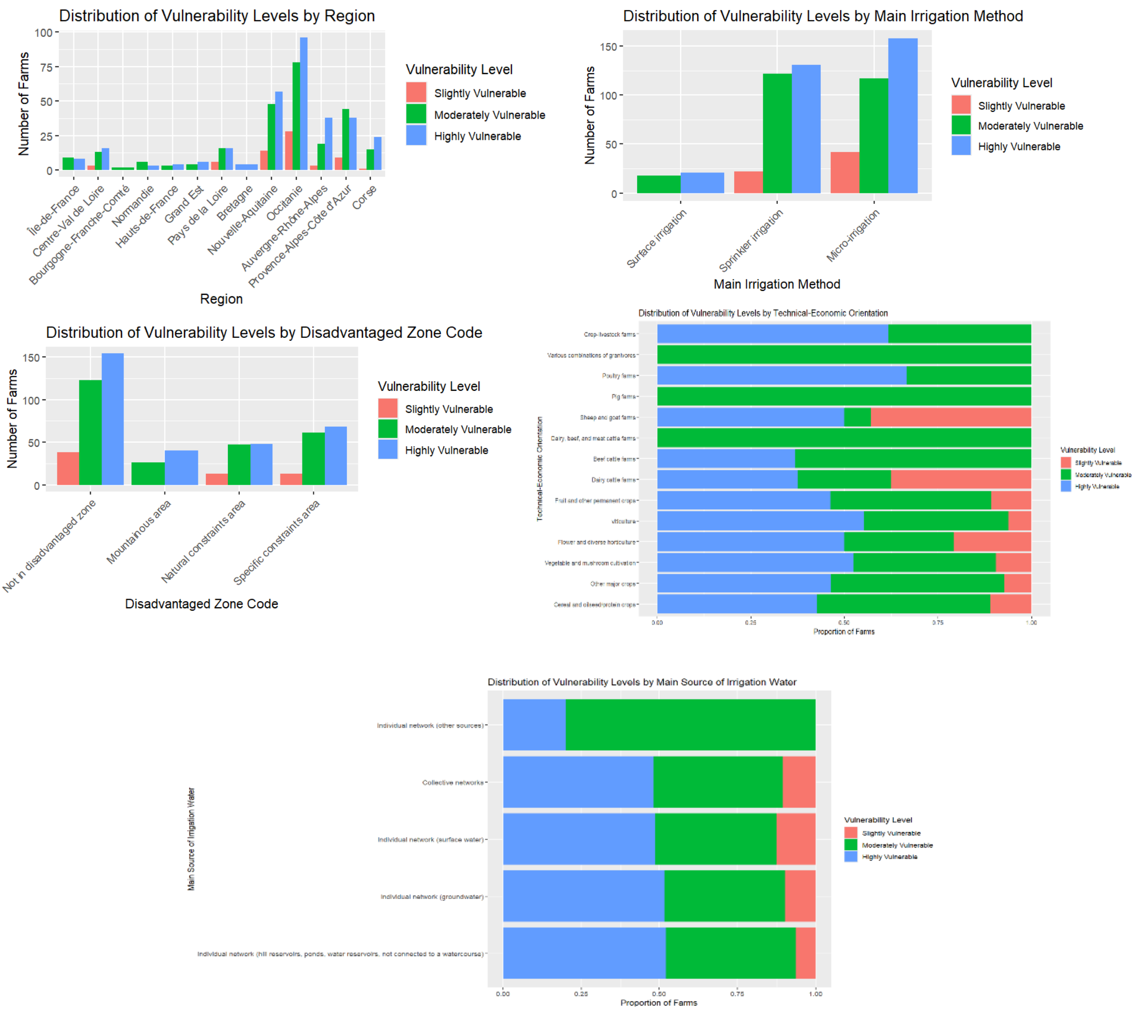

Figure 3. Bar plots to visualize the distributions of categorical variables in our dataset.

We observe that Auvergne-Rhône-Alpes and Occitanie have the highest number of highly vulnerable farms compared to other regions. We can also note that specialized in viticulture and mixed crop-livestock farms have the highest number of highly vulnerable farms compared to other orientations. Furthermore, for all three irrigation methods, the number of highly vulnerable farms is high, but those using micro-irrigation have a slightly larger difference between the highly vulnerable and moderately vulnerable categories compared to other methods. Concerning irrigation water source, farms using collective networks or individual networks such as groundwater (wells, boreholes) tend to be more highly vulnerable compared to other irrigation water sources.

Finally, we observe that farms not in disadvantaged zones show a considerable amount of vulnerability, particularly in the highly vulnerable category, possibly due to bias since in our database, nearly 50% of the farms fall into the category of non-disadvantaged zones.

Method

In this study, we are more inclined to initially employ a Random Forest model for testing our variables importance and predicting the contribution of each one on the farm's vulnerability level. This algorithm depends on the parameter number of trees, which the user must choose. In practice, it is essential to ensure that the forest has reached its convergence regime. The evolution of Out-Of-Bag (OOB) errors as a function of the number of trees, the classification error, and the Area Under the Curve (AUC) are presented in the figure below (Figure 4).

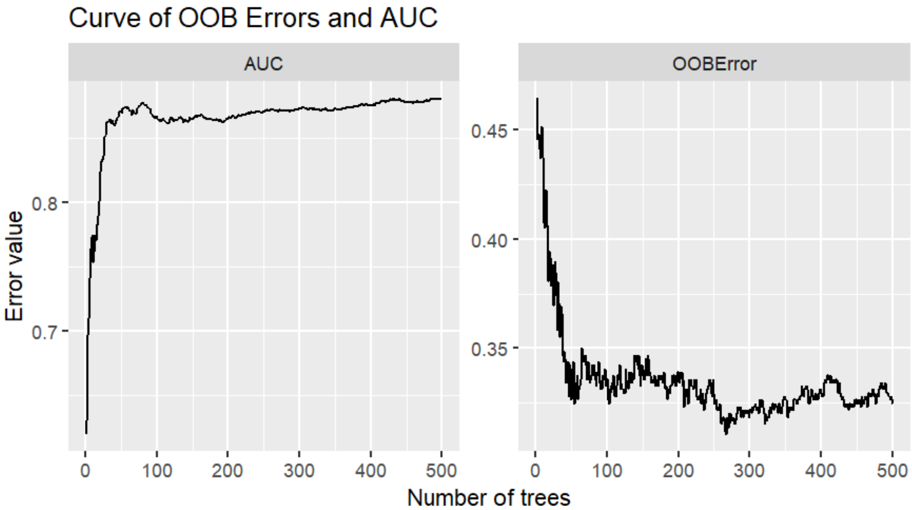

Figure 4. AUC (left) and classification errors (right) as a function of the number of trees.

We observe in Figure 4 that the errors are stable; therefore, we can consider that 500 trees are enough.

Random Forests are selected for their ability to handle both numerical and categorical data, capture complex nonlinear relationships, and provide a feature importance measure for interpretability. However, to ensure a thorough analysis, we explored and compared the performance of multiple models, including Logistic Regression, Naive Bayes and Linear Discriminant Analysis. To identify the most accurate predictive model for vulnerability level, we tested these machine learning algorithms and the accuracy scores for each model are illustrated in this Table down below (Table 3).

Table 3. Comparative Analysis of ML Models for Predicting Vulnerability Level

Model	Accuracy	Performance Summary
Logistic Regression	0.5736926	It achieved an accuracy of 57.36%. This model provides a moderate level of performance indicating that it may capture some patterns in the data but may be limited by the complexity of the factors involved.
Naive Bayes	0.5340729	Performed the worst among the tested models, achieving an accuracy of 53.40%, indicating that it may not capture the complexities of the dataset effectively for predicting vulnerability level.
Linear Discriminant Analysis	0.5768621	Similar to Logistic Regression, this ensemble method suggests that it might not fully capture the intricate relationships within the dataset.
Random Forest	0.9286846	It achieved a high accuracy of 92.86%. This ensemble method leverages multiple decision trees to improve prediction accuracy and handle overfitting effectively.

After choosing the model, its evaluation is a critical step in machine learning workflow. It involves assessing the performance of a trained model using various metrics to determine how well it generalizes to unseen data. In our study, a confusion matrix was used to assess the model's performance which calculates indicators such as accuracy, error rate, specificity and sensitivity. Accuracy is a measure of the correctly predicted observations over the prediction's total number (Table 4).

Table 4. Model Accuracy results

Overall Statistics	
Accuracy	0.9287
95% C.I	(0.9057, 0.9475)
No Information Rate	0.4913
P-Value	<2e-16
Kappa	0.8775
Mcnemar's Test P-Value	0.4218

The matrix helps in showing how the classification has been done for each class. Table 5 shows the confusion matrix for all classes. This classification has the correctly classified observations in the test data set in the diagonal, while off-diagonal elements represent misclassified observations for the classes. Moderately and highly vulnerable classes got the highest number of classifications (289 and 242) in the classification problem.

Table 5. Confusion Matrix by Class

	Reference		
Predictions	Slightly vulnerable	Moderately vulnerable	Highly vulnerable
Slightly vulnerable	55	2	5
Moderately vulnerable	6	242	16
Highly vulnerable	3	13	289

To determine the performance of the classification per class, an overall performance by class table was drawn, showing sensitivity and specificity; positive classes for each class classified correctly against other classes, and negative classes classified correctly against other classes respectively for the three classes. Once again, the per class performance ranked highly and moderately vulnerable as with high sensitivity (93.23% and 94.16%) and specificity (95.02% and 94.12%) respectively. The other class, while its specificity was high (98.76%), it means that its classification as not belonging to that class was high. This is shown in Table 6.

Table 6. Statistics by Class

	Slightly vulnerable	Moderately vulnerable	Highly vulnerable
Sensitivity	0.85938	0.9416	0.9323
Specificity	0.98765	0.9412	0.9502
Pos Pred Value	0.88710	0.9167	0.9475
Neg Pred Value	0.98418	0.9591	0.9356
Prevalence	0.10143	0.4073	0.4913
Detection Rate	0.08716	0.3835	0.4580
Detection Prevalence	0.09826	0.4184	0.4834
Balanced Accuracy	0.92351	0.9414	0.9412

Results

In this section, we employed two methods: Variable Importance to assess the contribution of our independent variables to the outcome variable, and Partial Dependence Plots to examine the marginal effect of each independent variable in predicting a certain class of vulnerability.

The Variable importance is a measure of how the variance is reduced, or the impurities reduction on each decision tree brought about by the gain in information or the Gini coefficient index. Variable importance is calculated through a Mean Decrease Impurity which sums up Gini Index decrease of the variables and averages to obtain a list of important variables. It is given as:

$$V_{imp}(x_i) = \frac{1}{n_{trees}}\left[1 - \sum_{j=1}^{n_{trees}} GI\,(i)^{(j)}\right]$$

Figure 5 depicts the variable importance by measuring the decrease in mean Gini. A higher mean decrease in Gini will imply a higher importance. The results in Figure 5 indicate that the variable Irrigation water charges per irrigated area was the variable with the highest importance, followed by farm's technical-economic orientation. Irrigation method is the least important but its non-zero mean decrease in Gini value indicates that it still provides some predictive power.

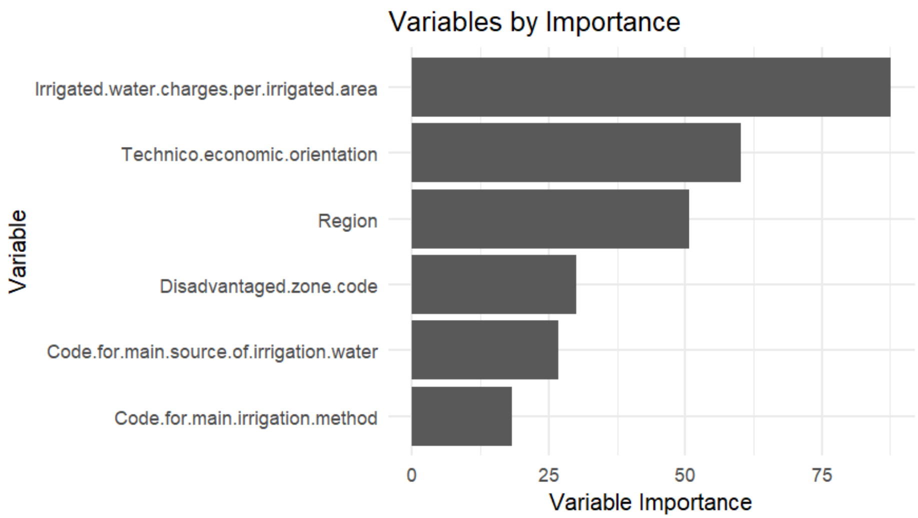

Figure 5. Variables by importance

The second method is Partial Dependence Plots (PDPs) which illustrate the marginal effect of a feature or a set of features on the predicted outcome of our machine learning algorithm. They provide insights into how the model's predictions change when the feature values vary, while keeping all other features constant. Partial dependence plots for irrigation method, irrigation water source, farm's technical-economic orientation, disadvantaged zone code, irrigation charges per irrigated area and region code are presented in Figure 6.

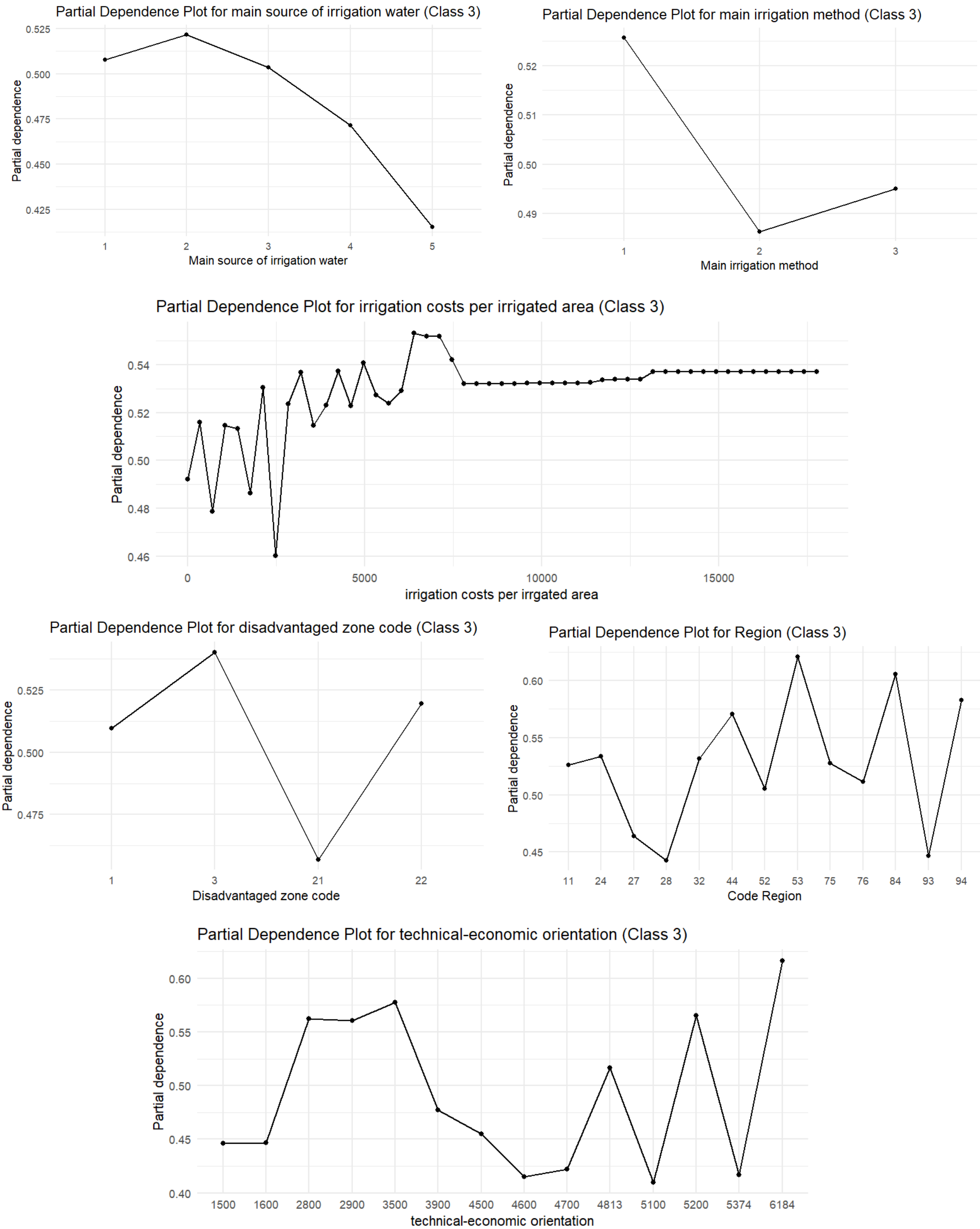

Figure 6. Partial Dependence Plots for Features Influencing Predicted Probability of Class 3 (Highly vulnerable).

The Partial Dependence Plots (PDPs) for various factors reveal significant insights into their marginal effects on classifying farms as highly vulnerable (class 3). For the main irrigation method, surface irrigation (Method 1) exhibits the highest partial dependence (~0.56), indicating a strong positive marginal effect, while sprinkler irrigation (Method 2) shows a marked decrease (~0.44), suggesting a negative marginal effect. Micro-irrigation (Method 3) has a slight increase compared to Method 2 but remains lower than Method 1, indicating that surface irrigation is most favorable for classifying farms as highly vulnerable, whereas sprinkler irrigation is less favorable. Regarding the main irrigation water source, source 2 (individual network such as groundwater: wells, boreholes) exhibits the highest partial dependence (~0.525), indicating a significant positive marginal effect, while source 5 (individual network: other sources) shows the lowest (~0.425), suggesting it is least favorable for classifying farms as highly vulnerable. For the disadvantaged area code, farms located in mountainous areas (code 3) exhibit the highest partial dependence (~0.530), indicating a significant positive marginal effect, whereas farms in areas subject to natural constraints (code 21) show the lowest partial dependence (~0.475), indicating a negative marginal effect, thus making mountainous areas more favorable for high vulnerability classification. The PDP for irrigation water costs per irrigated area shows significant variability for low to moderate costs, with partial dependence reaching a maximum at around 7500 euros/ha and stabilizing around 0.54 beyond this point, suggesting a positive marginal effect up to this threshold. For region codes, pays de la Loire (code 53) exhibits the highest partial dependence (~0.60), suggesting a significant positive marginal effect, while Centre-Val de Loire (code 28) shows the lowest (~0.45), indicating a negative marginal effect. Finally, for the farm's technical-economic orientation, specialized farms in viticulture (code 3500) exhibit the highest partial dependence (~0.60), indicating a strong positive marginal effect, whereas farms combining dairy, beef, and meat production (code 4700) show the lowest (~0.40), suggesting a negative marginal effect. These findings collectively highlight the factors most and least favorable for classifying farms as highly vulnerable.

Discussion — conclusion

This study brings out the application of random forests to predict a complex variable such as farm's vulnerability level. The random forest is a significant improvement from classical regression techniques, although not an exhaustive level, use of both for the suitable problem is appropriate. We proved that strategic choices such as the farm's location, technical-economic orientation and irrigation's parameters play a role that cannot be ignored in the determination of the farm's vulnerability level and should be taken into consideration for the farm's management strategy.

Partial Dependence Plots highlighted the empirical differences between modalities of each determinant of vulnerability. The predicted gradient between highly vulnerable and irrigation methods defines a difference between surface irrigation and sprinkler irrigation. Similarly, for the class highly vulnerable and the source of irrigation water. Individual networks from groundwater and those from other sources show a difference in predicting the highly vulnerable outcome. However, irrigation water charges per irrigated area, which is the most important variable in determining our outcome, show a complex function with several fluctuations at first then a stabilization from a certain value. For the disadvantaged zone code, it seems that the farm's altitude is a factor to consider and probably an indicator of its exposition level. Finally, for the farm's location and technical-economic orientation, we show that certain regions and crops might be contributing to class the farm as highly vulnerable. Although, it is important to mention that these determinants are complex and depend on many factors so even though our approach has several strengths such as the used algorithm which allows us to highlight the complex functions necessary for predicting farm's vulnerability level, it is not without limitations. However, it could have been more pertinent if we worked at the scale of a single region and a single technical-economic orientation, or if we had access to the farms' departments or some more precise information about their locations and production mode.

Our results are relevant for practical decisions but future studies may improve these results by exploring further factors and using a larger sample size.

Acknowledgements We would like to thank Groupama Paris Val de Loire and the Polytechnic Institute UniLaSalle for funding this research study.

References

Adger, W. N. (2006). Vulnerability. *Global Environmental Change*, 16, 268-281. https://doi.org/10.1016/j.gloenvcha.2006.02.006.

Agreste. FADN data. http://agreste.agriculture.gouv.fr.

Amigues, J.-P., Debaeke, P., Itier, B., Lemaire, G., Seguin, B., Tardieu, F., & Thomas, A. (éditeurs). (2006). *Sécheresse et agriculture : Réduire la vulnérabilité de l'agriculture à un risque accru de manque d'eau*. Synthèse du rapport d'expertise réalisé par l'INRA à la demande du Ministère de l'Agriculture et de la Pêche. INRA.

Ayphassorho, H., Sallenave, M., Bertrand, N., Mitteault, F., & Rollin, D. (2022). Quelles perspectives pour l'eau et l'agriculture d'ici à 2050 dans le contexte du changement climatique? *Annales des Mines - Responsabilité & environnement*, 106, 81-84. https://doi.org/10.3917/re1.106.0081.

Ben Brahim, F., Mahamat, H. Y., Kedidi, I., Malavaux, M.-R., & Rakotobe, R. (2024). Perception et gestion des risques liés à l'eau. *Pistes d'action inspirées par des agriculteurs*. Soumis aux Cahiers Costech, pour le numéro 7, mai 2024.

Blank, S. C., & McDonald, J. (1995). How California Agricultural Producers Manage Risk. *California Agriculture*, 49, 9-12. https://doi.org/10.3733/ca.v049n02p9.

Chimi, P. M., Mala, W. A., Fobane, J. L., Abdel, K. N., Nkoué, B. B., Nganmeni, L. F. F., Pokam, E. Y. N., Minfele, S. P. E., Matick, J. H., Tchandjie, F. M., Essouma, F. M., & Bell, J. M. (2024). Factors affecting decision-making to strengthen climate resilience of smallholder farms in the Centre region of Cameroon. *Climate Smart Agriculture*, 1, 100004. https://doi.org/10.1016/j.csag.2024.100004.

Debesai, M. G. (2020). Factors affecting vulnerability level of farming households to climate change in developing countries: evidence from Eritrea. *IOP Conference Series: Materials Science and Engineering*, 1001, 012093. https://doi.org/10.1088/1757-899X/1001/1/012093.

Enjolras, G., & Sentis, P. (2011). Crop insurance policies and purchases in France. *Agricultural Economics*, 42, 475-486. https://doi.org/10.1111/j.1574-0862.2011.00535.x.

European Commission, Joint Research Centre. (2020). *Annual Report 2020*. European Commission's science and knowledge service.

FAO. (2021). *La situation mondiale de l'alimentation et de l'agriculture 2021 : Rendre les systèmes agroalimentaires plus résilients face aux chocs et aux situations de stress*. Rome: FAO. https://doi.org/10.4060/cb4476fr.

Goodwin, B. K., & Smith, V. H. (1996). An Econometric Analysis of the Demand for Multiple Peril Crop Insurance: Comment. *American Journal of Agricultural Economics*, 78(2), 428-438. https://doi.org/10.2307/1243719.

IPCC. (2001). *Climate Change 2001: The Scientific Basis*. Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change [Houghton, J.T., Ding, Y., Griggs, D.J., Noguer, M., van der Linden, P.J., Dai, X., Maskell, K., & Johnson, C.A. (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 881pp.

IPCC. (2023). *Summary for Policymakers. In Climate Change 2023: Synthesis Report*. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, H. Lee and J. Romero (eds.)], 1-34. Geneva, Switzerland: IPCC. doi: 10.59327/IPCC/AR6-9789291691647.001.

Joshi, A., Bhargava, R., & Aggarwal, P. (2020). Exploratory Data Analysis: A Comprehensive Review. *Journal of Big Data*. ResearchGate.

Kapsambelis, D. (2022). Modélisation d'événements climatiques extrêmes sur les productions agricoles à horizon 2050 : Application à la gestion économique du risque. Thèse de doctorat, Institut Agro Rennes-Angers, COMUE Université Bretagne Loire. https://theses.hal.science/tel-03953483.

Komarek, A. M., De Pinto, A., & Smith, V. H. (2020). A Review of Types of Risks in Agriculture: What We Know and What We Need to Know. *Agricultural Systems*, 178 (février), 102738. https://doi.org/10.1016/j.agsy.2019.102738.

Létard, V., Flandre, H., & Lepeltier, S. (2004). Rapport d'information fait au nom de la mission commune d'information 'La France et les Français face à la canicule: les leçons d'une crise'. *Sénat, session ordinaire de 2003-2004. Annexe au procès-verbal de la séance du 3 février 2004*.

Milestad, R., Dedieu, B., Darnhofer, I., & Bellon, S. (2012). Farms and farmers facing change: The adaptive approach. In *Farming Systems Research into the 21st Century: The New Dynamic* (pp. 365-385). Springer Netherlands.

Mirza, M. M. Q. (2003). Climate Change and Extreme Weather Events: Can Developing Countries Adapt? *Climate Policy*, 3, 233-248. http://dx.doi.org/10.3763/cpol.2003.0330.

OECD. (2012). *Education at a Glance 2012: OECD Indicators*. OECD Publishing. http://dx.doi.org/10.1787/eag-2012-en.

Salmon, C. (2020). *Agro-Écologie et Besoins en Eau*. Rapport d'étude.

Service des données et études statistiques (SDES). (2024, février). L'irrigation des surfaces agricoles : évolution entre 2010 et 2020. Commissariat général au développement durable. https://www.statistiques.developpement-durable.gouv.fr/lirrigation-des-surfaces-agricoles-evolution-entre-2010-et-2020

Sneessens, I., Sauvée, L., Randrianasolo-Rakotobe, H., & Ingrand, S. (2019). A framework to assess the economic vulnerability of farming systems: Application to mixed crop-livestock systems. *Agricultural Systems*, 176, 102658. https://doi.org/10.1016/j.agsy.2019.102658.

Urruty, N., Tailliez-Lefebvre, D., & Huyghe, C. (2016). Stability, robustness, vulnerability and resilience of agricultural systems: A review. *Agronomy for Sustainable Development*, 36, 15. https://doi.org/10.1007/s13593-015-0347-5.

Authors and Affiliations

F. Ben Brahim[1,2] · I. Kedidi[2] · M-R. Randriamarolo[2] · H. Randrianosolo[2,3]

[1] AgroParisTech, 22 place de l'Agronomie, 91120 Palaiseau, France

[2] InterACT Research Unit, UniLaSalle, 60000 Beauvais, France

[3] IDEST, Paris - Saclay University, 92330 Sceaux, France