



UNIVERSITÉ PAUL SABATIER TOULOUSE III MASTER 2 EEA - ELECTRONIQUE, ENERGIE ELECTRIQUE, AUTOMATIQUE

Parcours SIA-AMS : Signal Imagerie et Applications Audio-vidéo, Médicales et Spatiales

Rapport de stage de **Martin PASQUIER**

Discipline : Télédétection
Spécialité : Traitement du Signal

Contribution à l'élaboration d'un outil d'inventaire des friches agricoles par télédétection et drone

Sous la direction de :

Flavie BERGOUNIOUX (DDT82) chargée de projet
Damien TULLOT (CNES) chargé de projet

Soutenance le 01/10/2021 devant le jury composé de :

Hervé CARFANTAN	Tuteur pédagogique
Jean-François TROUILHET	Examineur
Yannick DEVILLE	Examineur

Résumé

Mots-clés : friches agricoles, télédétection, algorithme de classification

Les friches agricoles sont des zones naturelles laissées à l'abandon qui méritent d'être revalorisées notamment pour une transition vers une agriculture solidaire et écologique. L'identité du paysage du territoire et l'activité agricole locale dépendent donc de la réappropriation de ces zones. C'est pourquoi la LAAF (loi pour l'Avenir de l'Agriculture et de la Forêt) charge, depuis 2014, la commission départementale de la préservation des espaces naturels, agricoles et forestiers de procéder à un inventaire des terres considérées comme friches, qui pourraient être réhabilitées pour l'exercice d'une activité agricole ou forestière. Il en est fait état dans l'article L.112-1-1 du code rural et de la pêche maritime.

En Occitanie, la DDT82 (Direction Départementale du Territoire de Tarn-et-Garonne) collabore avec le CNES (Centre National d'Etudes Spatiales) et la SAFER (Les Sociétés d'Aménagement Foncier et d'Etablissement Rural), dans l'optique de répondre au besoin identifié. Des vérités terrain issues de l'application mobile collaborative Vigifriche ont permis de mettre en place un outil de classification supervisée. Développé par le Labo'OT (Observation de la Terre) du CNES, un algorithme d'apprentissage semi-automatique combine des données topographiques et d'imagerie satellitaire pour permettre un inventaire généralisable et reproductible.

Mon travail durant ce stage a été d'optimiser la méthode utilisée pour effectuer l'inventaire sur le département du Tarn-et-Garonne. J'ai réalisé une étude sur les différentes données utilisées pour améliorer les résultats en sortie de l'algorithme de classification supervisée Random Forest. J'ai également pu coordonner et participer à plusieurs missions pour collecter de nouvelles vérités terrain et en fournir le modèle.

Remerciements

Ma bonne intégration au sein du projet n'a été possible que grâce à la supervision et au soutien indéfectible de nombreuses personnes. Je tiens ici à tous les en remercier.

Tout d'abord, j'ai sincèrement apprécié l'encadrement que j'ai reçu. Merci Flavie pour ton pragmatisme et Damien pour ta patience. Vous avez été de bons conseils et de véritables émulateurs durant ces six mois. Vous m'avez accordé votre confiance pour coconstruire ce stage, je vous en suis très reconnaissant. Votre accueil chaleureux m'a permis de partager de belles idées, gagner en compétences et je suis heureux d'avoir pu participer à ce travail.

A l'équipe du Lab'OT, cette immersion dans la structure colossale du CNES a été courte mais intense. Mes remerciements tout particuliers vont à Vincent qui a su répondre à mes questions concernant l'algorithme et l'environnement python.

Aux membres de l'équipe du BIGTI, Laurent, Francis et Gilles, merci pour votre temps et votre capital sympathie. Pour tous ces trajets jusqu'à la ville d'Ingres, Morgane, covoitreuse joyeuse, apprentie géomaticienne hors-pair, merci. Pour sa supervision technique et sa bonne humeur, je remercie Christian Simon, chef de l'équipe, qui a su répondre à la moindre de mes sollicitations.

Merci à toute la team drone d'Occitanie pour notre mission rondement menée et au maire de L'Honor-de-Cos, Michel Lamolinairie pour son accueil exceptionnel.

Merci également aux partenaires de la SAFER, Maud, Gilles, Florent et Mathias qui ont suivi de près ce travail et ont pu apporter leur expertise technique pour mes missions sur le terrain.

Je tiens à remercier tous les agents de la DDT82 avec qui j'ai pu échanger sur ce projet, leurs expertises techniques m'ont permis de mieux élaborer mon sujet de stage. Ils m'ont convaincu de l'intérêt de mon travail.

Merci également à toute l'équipe pédagogique du Master SIA-AMS et à Hervé Carfantan, mon tuteur de stage, pour le suivi de ce travail.

Pour finir, merci à ma famille, mes proches ainsi que mes colocataires qui ont su me reconforter dans les moments de doutes et être présents pour partager ceux de joie.

Table des matières

1	Introduction	2
2	Organisation de la structure projet et objectifs	4
3	Données disponibles pour la détection de friches agricoles	6
3.1	Enjeux des friches agricoles :	6
3.2	Apport de la Télédétection :	7
3.3	Vigifriche, l'application mobile de la Safer :	8
3.4	Données d'entrées de l'algorithme de classification supervisée :	8
4	L'algorithme de classification Random Forest au service de l'inventaire des friches agricoles	12
4.1	Utilisation de Random Forest (RF) :	12
4.1.1	Objectif et choix de l'algorithme :	13
4.1.2	Principe de la classification :	14
4.1.3	Paramétrage RF :	15
4.1.4	Apprentissage supervisé :	15
4.1.5	Méthode d'optimisation :	16
4.1.6	Limites de l'algorithme :	17
4.2	Utilisation des scripts :	17
4.2.1	Fonction des scripts :	18
4.2.2	Apprentissage d'un premier modèle multi-communes :	19
5	Optimisation de l'algorithme de classification Random Forest	21
5.1	Prise en main de l'algorithme :	21
5.1.1	Utilisation d'un nouveau modèle HFN :	21
5.1.2	Etude de la généricité du modèle HFN :	21
5.1.3	Etude du score de fiabilité :	22
5.1.4	Ajout de nouveaux descripteurs au modèle HFN :	22
5.2	Optimisations successives du modèle HFN :	23
5.2.1	Modification de l'indice NDVI en MSAVI et TSAVI :	23
5.2.2	Tests par exclusion de données SIG :	24
5.2.3	Utilisation d'images trimestrielles :	24
5.2.4	Modification des hyper-paramètres :	25
5.2.5	Comparaison gain avec optimisation hyper-paramètres :	25
5.3	Optimisation suite aux missions terrain :	26
5.3.1	Nouvelles données de références :	26

5.3.2	Comparaison modèles 6C et 3Cs (HFN et CPG) :	26
5.3.3	Utilisation du modèle générique 6C sur les plus grosses communes :	27
5.3.4	Optimisations non-effectuées :	27
6	Connaissance et vérifications terrain	29
6.1	Réflexions sur les limites de la classification :	29
6.2	Missions de vérification sur L'Honor-de-Cos :	31
6.3	Missions d'inventaire sur de nouvelles communes :	32
6.4	Compléments de vérification sur L'Honor-de-Cos avec des images drones :	34
6.5	Conclusion missions terrain :	35
7	Inventaire comparé sur tout le département de Tarn-et-Garonne	36
7.1	Inventaire complet des friches parcellaires agricoles sur le Tarn-et-Garonne :	36
7.2	Méthodologie poursuite d'étude :	37
8	Conclusion	39
	Annexes	41
A	Déroulé de mon stage	42
B	Tests d'optimisation sur l'algorithme de classification supervisée	43
C	Logistique et itinéraires mission repérage sur trois nouvelles communes	45
D	Cartes d'inventaire des parcelles cadastrales agricoles en Tarn-et-Garonne	47

Table des figures

1.1	Photo d'une friche arborée basse sur la commune de L'Honor-de-Cos, présence de ligneux, ©Martin Pasquier	3
3.1	Orthophotographie d'une parcelle agricole en friche sur la commune de L'Honor-de-Cos, ©DDT82	8
3.2	Interface de déclaration Vigifriche, ©SAFER	9
3.3	Sentinel2, un instrument de mesure multi-spectral, ©site de l'ESA	10
4.1	Classification supervisée des parcelles cadastrales	12
4.2	Arbre de décision à trois niveaux et deux classes	13
4.3	Random Forest, assemblage aléatoire de données dans des arbres de décisions aux descripteurs choisis aléatoirement	14
4.4	Explication des sorties de prédiction	15
4.5	Analyse des contributions des descripteurs dans la prédiction sur l'ensemble des pixels d'une parcelle cadastrale	20
5.1	Fiabilité des prédictions sur L'Honor-de-Cos à partir du modèle HFN	23
6.1	Carte des prédictions sur base test de L'Honor-de-Cos	30
6.2	23 zones de survols pour la Team Drone d'Occitanie, ©DDT82	34
7.1	Inventaire complet des friches parcellaires agricoles du Tarn-et-Garonne	37
7.2	Production participative du projet d'inventaire des friches agricoles du 82	38
B.1	Architecture globale des différents scripts de l'algorithme de classification	43
C.1	Itinéraires 2 premiers groupes de repérage sur la commune de Goudourville	45
C.2	Itinéraires 2 derniers groupes de repérage sur la commune de Goudourville	46
D.1	Comparaison inventaires du Tarn-et-Garonne avec l'étude effectuée en 2017	48
D.2	Comparaison inventaires sur la commune de L'Honor-De-Cos	49

Liste des tableaux

3.1	Tableau récapitulatif des données utilisées pour la classification	11
4.1	Prédictions sur L'Honor-de-Cos avec le modèle 3 communes HFN	19
5.1	Prédictions avec le modèle HFN sur différentes communes	22
5.2	Performances des 6 modèles sur la base test des 3 communes connues . . .	22
5.3	Exclusion de certaines données SIG du modèle HFN	24
5.4	Performances avec utilisation d'images satellites trimestrielles	25
5.5	Modification de l'hyper-paramètre de profondeur des arbres de décision . .	25
5.6	Comparaison gain après optimisation hyper-paramètres	26
5.7	Vérités terrain collectées sur 3 nouvelles communes	26
5.8	Modèle 6C, plus générique, meilleur rappel mais précision plus faible	27
5.9	Modèle 6C_RPG2019_trimestre appliqué aux communes tests	27
5.10	Modèle HFN mensuel appliqué aux communes tests	28
6.1	Analyse des parcelles vérifiées lors de la première mission	31
A.1	Déroulé de mon stage	42
B.1	Tableau récapitulatif des tests et conclusions pour l'optimisation du modèle	44

1

Introduction

Une sensibilisation généralisée du devenir des friches agricoles est amorcée avec au centre des débats, l'humain, les collectivités et l'avenir du paysage de notre ruralité. Plusieurs initiatives d'inventaires des friches parcellaires agricoles (Partie 2), facilités par des bases de données, ont été menées en France. Le conseil régional d'Occitanie a inscrit dans son « Plan d'action régional sur le foncier » l'action suivante : préserver les terres agricoles et faciliter l'accès à la mobilisation des terrains vers du foncier agricole. Cet enjeu se traduit par une reconquête des friches agricoles en travaillant notamment sur leur identification, sur la restructuration du foncier et sur l'animation foncière.

L'inventaire des friches agricoles est une obligation réglementaire issue de la LAAF - Loi d'avenir pour l'Agriculture, l'Alimentation et la Forêt (article L.112-1-1 du code rural et de la pêche maritime). Dans le Tarn-et-Garonne, cet enjeu mobilise les acteurs locaux autour de préoccupations fortes :

- L'état sanitaire des vergers et des vignes en friche, vecteurs de maladies ;
- La gestion des espèces nuisibles, dont les friches constituent un habitat ;
- La reconquête et valorisation de ces espaces ;
- Les enjeux paysagers ;
- La gestion des risques incendies.

En novembre 2020, la DDT82, en collaboration avec le CNES et la SAFER, propose un projet de partenariat innovant en recherche appliquée, reprenant la méthodologie utilisée dans le département de l'Hérault [1] en 2019. Pour répondre au besoin identifié, le Labo'OT (Observation de la Terre) du CNES a développé un algorithme de classification basé sur de l'apprentissage semi-automatique utilisant l'imagerie spatiale. Cet algorithme est alimenté par des repérages sur quelques communes des friches parcellaires agricoles en présence. Ces "vérités terrain" sont déclarées et géolocalisées sur l'application mobile Vigifriche développée par la SAFER (partenaire du projet). La détection des friches agricoles est complexe. La sensibilité et l'expérience de chacun jouent beaucoup dans l'interprétation. Les partenaires se sont mis en accord sur une définition commune :

« Une friche agricole se définit comme une zone ou un terrain sans occupant humain actif qui n'est en conséquence pas ou plus cultivé, productif ni même entretenu. Elle résulte de la déprise agricole des terres (abandon définitif ou sur une longue période) contrairement à la jachère qui n'est qu'un temps provisoire de repos du sol. »



FIGURE 1.1 – . Photo d'une friche arborée basse sur la commune de L'Honor-de-Cos, présence de ligneux, ©Martin Pasquier

Mon travail dans ce contexte : optimiser la méthode utilisée pour effectuer l'inventaire sur le département du Tarn-et-Garonne, mettre au point un outil opérationnel qui sera partagé aux partenaires de la CREFAO (Communauté Régionale des Friches Agricoles d'Occitanie) et par la suite avec les autres DDTs. Il permettra la sensibilisation des élus et des collectivités pour la revalorisation des friches agricoles sur le territoire. Pour débiter, j'ai réalisé une étude sur les différentes données utilisées : topographie open source (Partie 3) et télédétection. Ensuite, j'ai fait un état des lieux (Partie 4) de l'utilisation et l'étude effectuée par le Lab'OT sur l'algorithme de télédétection (classification supervisée Random Forest). J'ai ainsi pu affiner les résultats en sortie de l'algorithme (Partie 5). Les objectifs de cette optimisation ont été de vérifier la généralisation du modèle, en ajustant paramètres et données d'apprentissage, afin d'améliorer les résultats obtenus lors de la télédétection. La méthode de validation définie a permis d'effectuer un inventaire sur tout le département de Tarn-et-Garonne (Partie 7). En complément plusieurs missions sur le terrain (Partie 6) ont été menées afin de valider certaines vérités terrain de Vigifriche. Des données drone ont été obtenues et pourront être intégrées pour analyser ensuite leur pertinence.

Les délais de la validation des résultats ne m'ont pas permis de participer à la phase d'animation auprès des élus et des collectivités suivant l'inventaire. J'ai tout de même pu construire des guides méthodologiques pour les futurs utilisateurs de l'algorithme et sensibiliser les agents des DDTs aux enjeux d'un tel projet. La production participative ou "crowdsourcing" est une évidence pour l'inventaire. L'outil à mettre en place dans la continuité du travail réalisé devra être une interface multifonctionnelle et évolutive. De nombreux acteurs pourront s'en servir mais aussi la compléter et la transformer pour des usages divers. Les perspectives d'évolution de cet outil sont vastes.

2

Organisation de la structure projet et objectifs

Mon stage est co-dirigé par les deux structures que sont la DDT82 (Direction Départementale des Territoires de Tarn-et-Garonne) et le CNES (Centre National d'Etudes Spatiales). J'ai également travaillé en collaboration avec la SAFER (Les Sociétés d'Aménagement Foncier et d'Etablissement Rural).

Les DDTs sont des services déconcentrés de l'Etat français qui assurent le relais des décisions prises par l'administration centrale et qui gèrent les services de l'Etat au niveau local. Ces services déconcentrés sont de niveau départemental donc mais également régional (avec notamment les DRFiP). La DDT de Tarn-et-Garonne s'organise autour de deux sites. Le principal se situe à Montauban au 2 quai de Verdun, au "Carmel", lieu emblématique de la ville d'Ingres. Le deuxième site est à Castelsarrasin à la maison de l'état (MDE) où travaille une partie du Service d'Aménagement Territorial. En plus du cabinet de Direction (CAB) il existe 5 services :

- SAT – Service d'Aménagement Territorial –
- SCR – Service Connaissance et Risques –
- SEB – Service Eau et Biodiversité –
- SEA – Service Economie Agricole –
- SH – Service Habitat –

A mon arrivée à Montauban, j'ai reçu un livret d'accueil expliquant l'organisation de la structure ainsi qu'un badge pour accéder aux locaux. Comme j'ai pu en discuter, chaque service travaille en collaboration, plusieurs équipes sont dédiées à la gestion de tâches spécifiques. J'ai été intégré au SCR, bureau 227 équipé d'un écran. Christian Simon, chef d'équipe BIGTI (Bureau d'Information Géographique et Technologies Innovantes) m'a mis à disposition un téléphone et un ordinateur professionnels. J'ai ainsi pu mettre en place avec lui les outils nécessaires à mon travail comme mon courriel. De plus, j'ai été amené à contacter le SIDSIC (Service Interministériel Départemental des Systèmes d'Information et de Communication) pour mettre en place mon environnement de travail.

Le Centre National d'Etudes Spatiales (CNES) est également un établissement public. Il est chargé d'élaborer et de proposer au gouvernement français le programme spatial français et de le mettre en oeuvre. L'innovation dans des projets orientés vers le développement durable fait partie des 5 grands domaines d'études de la structure. J'ai ainsi suivi une formation nommée Ambassadeurs [2] pour apprendre à connaître les projets de demain.

Dans ce cadre, les équipes du Lab'OT (observation de la Terre) oeuvrent pour l'élaboration de projets innovants pour une meilleure connaissance de l'environnement à l'aide des images satellites. Elles répondent dans le cadre de ce projet à une question simple : les images satellites permettent-elles de mieux inventorier les friches agricoles ? J'ai mis en place mon environnement de travail sur le cluster HPC du CNES grâce à Damien TULLOT et à Vincent POULAIN, le référent technique en ce qui concerne l'algorithme.

Les SAFER (Sociétés d'aménagement foncier et d'établissement rural) départementale et régionale sont parti prenante dans ce projet. Créées par la loi d'orientation agricole du 5 août 1960, ces sociétés privées, anonymes et à but non lucratif présentent des missions d'intérêt général, sous la direction des ministères de l'Agriculture et des Finances. Elles permettent aux agriculteurs et tout porteur de projet artisanal, résidentiel ou environnemental de s'installer en milieu rural. C'est pourquoi, leur expertise terrain est essentielle pour rassembler des données de références utiles à l'ensemble des partenaires. J'ai pu échanger avec les agents terrain et les gestionnaires de l'application de géolocaliation des friches agricoles, Vigifriche. Un inventaire complet et de qualité sur le Tarn-et-Garonne conforterait leur travail sur le foncier pour réhabiliter ces zones d'intérêts.

En terme d'organisation personnelle, j'ai pu faire le point avec mes référents de manière hebdomadaire puis mensuellement avec l'arrivée des congés. Mon stage s'est effectué à moitié en télétravail en accord avec mes tuteurs. Quand mon environnement de travail a été mis en place, j'ai travaillé exclusivement sur l'algorithme de classification les mardi et jeudi. Les autres journées étaient consacrées aux recherches ainsi qu'à la préparation et synthèses de différents documents (retours d'expériences, présentations et rapports).

Afin de travailler avec l'ensemble des partenaires du projet, j'ai pu échanger régulièrement sur l'avancement de mon travail. En mettant à jour l'agenda de mon courriel, mes référents internes à la DDT82 ont pu suivre mon travail. En annexe A se trouve un déroulé synthétique de mon stage. Ce stage co-dirigé et aux nombreux partenaires a fait l'objet de plusieurs réunions avec les acteurs du projet et avec la direction de la DDT82. J'ai pu y présenter régulièrement mon avancement.

Difficultés rencontrées :

- mise en place de mon environnement de travail avec cluster HPC du CNES
- accès au serveurs de la DDT82 en télétravail ou au bureau
- accès au Gitlab du CNES pour obtenir les protocoles utiles

3

Données disponibles pour la détection de friches agricoles

La définition d'une friche agricole (voir Partie 1 : Introduction) acceptée par l'ensemble des partenaires se base sur l'étude bibliographique des différents projets d'inventaire ayant eu lieu [1][3][4][5][6]. Pour comprendre en détails, les tenants et aboutissement du projet, un état de l'art s'impose. J'ai ainsi pu étudier les enjeux du projet, l'apport de la télédétection, l'application Vigifriche ainsi que les données topographiques dites SIG (Systèmes d'Information Géographiques) utilisées dans la classification.

3.1 Enjeux des friches agricoles :

La reconnaissance de terrain en friches agricoles est ardue [1]. Ce sont des terres agricoles, auparavant exploitées qui sont aujourd'hui à l'état d'abandon depuis au moins trois ans. Elles sont dans un état transitoire où une végétation spontanée se développe et tend à évoluer vers un milieu forestier. Elles ne doivent pas être confondues avec les jachères, qui sont déclarées à la PAC (Politique Agricole Commune). Cela implique l'intérêt d'utiliser le RPG (Registre Parcellaire Graphique qui prend en compte ces déclarations) pour réduire les mauvaises interprétations et notamment permettre la distinction entre friches agricoles et jachères (repos du sol).

Les méthodologies d'inventaires que j'ai étudié prennent également en compte la réflectance des végétaux calculée sur des images satellites mais uniquement pour l'appliquer en tant que masque sur d'autres couches SIG d'occupation du sol. L'utilisation des données MAJIC (Mise À Jour des Informations Cadastres fournies par la DGFIP) est probante. Les surfaces de plus d'un hectare présentent l'intérêt d'être plus facilement décrites [3]. La validation et la confirmation terrain des résultats obtenus par analyse SIG puis par photo-interprétation (Figure 3.1) sont essentielles [4] car l'interprétation est faussée par les pentes notamment. L'étude présentée sur la Haute-Garonne [5] met en avant les limites de l'utilisation du RPG qui ne soustrait pas les parcelles cultivées non déclarées à la PAC. La solution alors établie est d'utiliser davantage d'images satellites sur une année pour mieux définir l'évolution de la végétation prise en compte sur 3 années dans l'inventaire de Lozère [6].

L'utilisation du mot friche ne se limite pas aux friches agricoles [7] : "Une friche est un terrain précédemment exploité (champ, prairie, verger, vigne, jardin...), abandonné par l'Homme et colonisé par une végétation spontanée (espèces héliophiles, rudérales). Cette zone sans occupant humain actif n'est plus cultivée, productive ni même entretenue, mais des activités marginales peuvent s'y étendre si ses parties restent accessibles : pâturage, cueillette, braconnage, chasse ou pêche et autres activités."

Il existe une grande diversité de friches. On peut les classer par surface, temps de vacance (durée de non occupation) et nature du terrain (dernière activité officielle exercée). Plusieurs catégories de friches ont ainsi été créées [7] :

- Les friches industrielles dans lesquelles les activités industrielles ont été démantelées, le terrain ayant été totalement débarrassé ou éventuellement conservant des vestiges d'installation.

- Les friches ferroviaires : 4 000 km de voies sont inutilisés en France à l'heure actuelle d'après le Réseau Ferré de France auxquelles s'ajoutent d'autres bâtiments liés à cette activité.

- Les friches portuaires : elles sont apparues à partir des années 1970-1980 à la suite de l'effondrement de l'industrie lourde.

- Les friches administratives et d'équipements publics (écoles, stades sportifs, hôpitaux, etc.)

- Les friches d'habitat : elles sont apparues en premier lieu dans les bassins industriels en déclin, mais peuvent aussi se retrouver dans tout type de ville.

- Les friches commerciales et tertiaires (ZAC à l'abandon, complexes hôteliers désaffectés, etc.)

- La friche agricole résulte de la déprise agricole (ou abandon) des terres. Lorsqu'une terre agricole abandonnée commence à se reboiser naturellement, on parle alors d'accrus. La friche doit être distinguée de la jachère, qui est une préparation et un repos du sol."

3.2 Apport de la Télédétection :

La diversité des friches agricoles tant par l'âge que par la présence de végétaux typiques à chaque territoire nécessite une étude terrain fastidieuse. Après une formation sur QGIS [8] j'ai pu les observer. Les photos aériennes (Figure 3.1) présentent une résolution très élevée mais l'occultation par des arbres et la multitude de définitions d'occupation du sol ne suffit pas pour "trouver" des friches agricoles. Les images satellites apportent une source d'informations supplémentaire grâce aux 12 longueurs d'ondes de Sentinel 2-L3A [9]. La caractérisation d'une parcelle cadastrale passe alors par le calcul d'un indice de réflectance des végétaux sur les pixels des images.

La Télédétection a plusieurs limites : la résolution faible (10-20-60 m par pixel en fonction de la bande spectrale) et l'interprétation parfois difficile des parcelles cadastrales. Pour pallier à ces imprécisions, on couple les images à des données terrains issues de couches SIG (Système d'information géographique).



FIGURE 3.1 – Orthophotographie d'une parcelle agricole en friche sur la commune de L'Honor-de-Cos, ©DDT82

3.3 Vigifriche, l'application mobile de la Safer :

L'application mobile Vigifriche est notre source de vérités terrains. Elle géolocalise les parcelles cadastrales en permettant d'y déclarer et de confirmer qu'une parcelle cadastrale présente une friche. Les utilisateurs peuvent également modifier leurs choix.

On peut renseigner (Figure 3.2) l'âge symbolique selon 4 niveaux de friches :

- 1) friche herbacée
- 2) friche herbacée avec présence de ligneux
- 3) friche arbustive basse
- 4) friche arbustive haute

Une photo ainsi qu'une description peuvent être ajoutées à la déclaration. Après validation par un opérateur, les confirmations Vigifriche sont utilisées pour fournir l'algorithme bi-classe en vérités terrains : friche mais aussi non-friche.

3.4 Données d'entrées de l'algorithme de classification supervisée :

- Cadastre [10] : plan des propriétés foncières servant de base à la classification qui s'effectue sur chaque parcelle cadastrale.

- RPG (registre parcellaire graphique) [11] : référence à l'instruction des aides de la Politique Agricole Commune (PAC), permettant d'éliminer les îlots qui y sont déclarés de la classification.

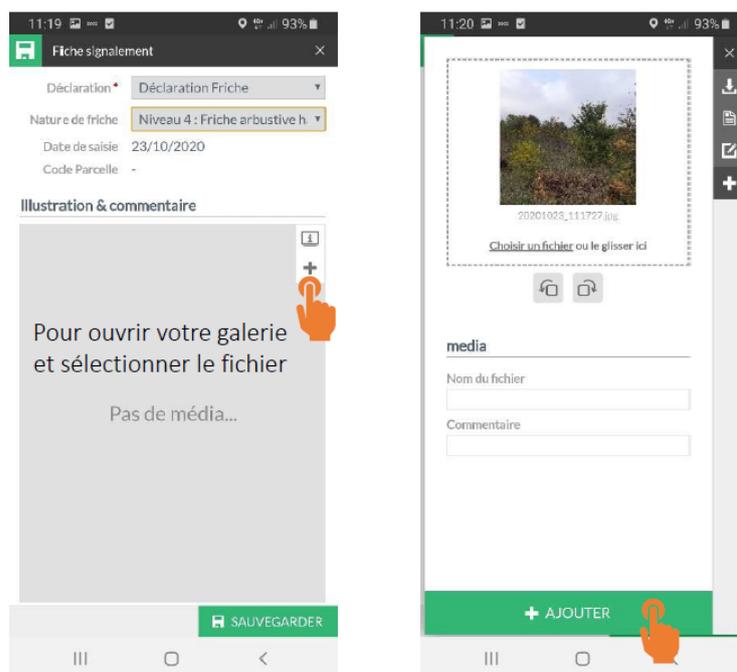


FIGURE 3.2 – . Interface de déclaration Vigifriche, ©SAFER

- BD_Topo [12] : comprend différentes descriptions du territoire (administratif, bâti, hydrographie, occupation des sols, zones réglementées, etc). L'IGN (Institut National de l'Information géographique et Forestière) a mis en place le RGE (Référentiel à Grande Echelle) qui fournit de nombreuses données terrains grâce à cette couche.
- BD_Forêt (inventaire forestier) [13] : permet d'éliminer les îlots de forêt de plus de 5000 m² de la classification.
- OCSGE (occupation du sol à grande échelle) [14] : cette couche est construite à partir du RPG, de la BD_Topo et de la BD_Forêt ; la validation de ces données est effectuée par photo-interprétation.
- CVI (casier viticole informatisé) [15] : permet d'éliminer les parcelles cadastrales de vignes déclarées à la DGFIP.

Remarque : Les données suivantes sont libres d'utilisation (licence ouverte version 2.0) et peuvent être exploitées avec un outil libre tel que QGIS.

Remarque : Suite aux différents tests effectués pour l'optimisation de l'algorithme (Annexe B.2), certaines données ont été écartées, car elles présentaient un biais trop important (RPG2017, RPG2020). De plus, le CVI apporte une plus-value faible sur des territoires sans viticultures et fait l'objet d'une demande directe, il faudra donc enlever cette couche si elle n'est pas donnée par la DGFIP.

• Images satellites (Sentinel2 L3A) : la constellation Sentinel fournit des observations fréquentes (tous les 5 jours en Europe) et répétitives sous forme de tuiles de taille 110*110 km². Celle de Toulouse est T31TCJ [16] pour la zone UTM T31 (60 zones en longitudes au total), la lettre T correspondant à zone en latitude (80 zones au total) et les dernières lettres CJ correspondant aux coordonnées axes ouest-est et nord-sud. En plus des corrections géométriques, les produits de niveau 3A contiennent des images améliorées (il existe des données niveau 1 et 2 avec moins de corrections), sans nuages, basées sur les données acquises pendant une période mensuelle. Pour chaque pixel, le pixel correspondant à la date la plus favorable est conservé. Les pixels contiennent 12 longueurs d'ondes différentes de taille allant de 10m à 50m (Figure 3.3). La médiane et l'écart-type d'un indice de réflectance des végétaux constituent de nouveaux descripteurs. Ils sont calculés sur les pixels de chaque parcelle cadastrale pour chaque image utilisée. L'indice utilisé en début d'étude est le NDVI pour Normalized Difference Vegetation Index (indice de végétation normalisé) qui est généralement utilisé pour décrire la végétation par télédétection. Il permet de compenser les changements de conditions d'éclairage, de pente de surface et d'exposition. Il se calcule [17] par le rapport : $(PIR-R)/(PIR+R)$ où PIR est la bande proche infrarouge et R la bande rouge d'une image. Cependant, il n'est pas souvent l'indice optimal pour généraliser toutes les végétations d'un territoire.

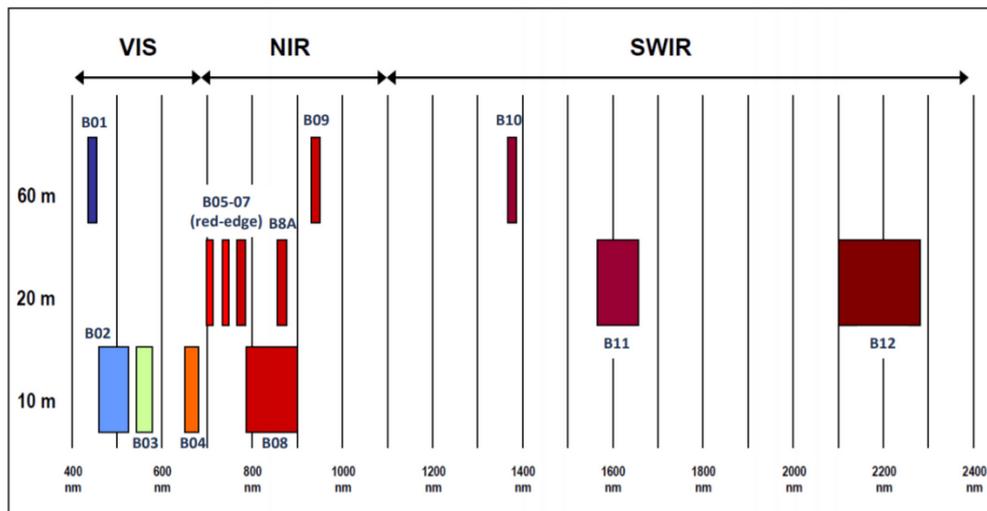


FIGURE 3.3 – Sentinel2, un instrument de mesure multi-spectral, ©site de l'ESA

Les différentes données ne présentant pas un maillage suivant le cadastre (certaines sont en îlots et non en parcellaire), elles sont combinées sous forme de données vectorielles (Table 3.1) pour donner jusqu'à 14 descripteurs d'occupation du sol. Les images satellites utilisées correspondent à 41 dates différentes ce qui fait 82 descripteurs puisqu'on y calcule la médiane et l'écart-type du NDVI.

L'ajout de certaines données n'a pas été jugé pertinent car apportant trop peu d'information à l'échelle du territoire : données GéoMCE (mesures compensatoires environnementales), CEN (conservatoire d'espaces naturels), APB (aire de protection des biotopes). Cependant l'exclusion de certaines couches a montré des résultats probants (RPG). Enlever certaines sous-couches du RPG pour ne conserver que celles utiles à la

<i>Données</i>		<i>Fournisseur</i>	<i>Millésime</i>	<i>Fréquence</i>	<i>Accès</i>
Données images :	Satellites (Sentinel2 L3A)	Theia	du 07/2017 au 12/2020	Synthèse mensuelle	Open Source
Données vectorielles :	Cadastre	IGN	2019	3 ans ou plus	Open Source
	RPG (Registre parcellaire graphique)	IGN	2017-2019-2020 (82)	Annuelle	Open Source
	OCSGE (Occupation du sol grande échelle)	IGN	2019	3 ans ou plus	Open Source
	BD_Forêt	IGN	2018	3 ans ou plus	Open Source
	BD_TOPO	IGN	2019	Trimestre	Open Source
	CVI (Casier viticole informatisé)	DGFIP	2019	3 ans ou plus	Partenaire
	Vérités terrains (parcelles cadastrales en friches sur Vigifriche)	SAFER	« Temps réel »	« Temps réel »	Partenaire

TABLE 3.1 – Tableau récapitulatif des données utilisées pour la classification

classification semble également être une piste intéressante mais qui n'a pas été explorée dans ce projet par manque de temps.

L'optimisation et les tests complémentaires ont été progressivement fournis avec pour vérités terrains (VT) issues de Vigifriche. Ces données ont été considérées comme suffisamment exhaustives pour faire l'apprentissage. En reprenant le code, j'y ai eu accès sur les 3 communes de :

- Nyls Ponteilla (66) : 359 parcelles en friche (sur 3535 parcelles cadastrales)
- Figeac (46) : 406 parcelles en friche (sur 12 026 parcelles cadastrales)
- L'Honor de Cos (82) : 244 parcelles en friche (sur 6669 parcelles cadastrales)

Suite à mes missions de repérage, nous avons récoltés des VT sur 3 nouvelles communes :

- Coutures (82) : 2 parcelles en friche (sur 268 parcelles cadastrales)
- Le Pin (82) : 6 parcelles en friche (sur 635 parcelles cadastrales)
- Goudourville (82) : 36 parcelles en friche (sur 2450 parcelles cadastrales)

Suite à l'étude de son PLU, nous avons accédé aux ressources de la commune de :

- Meauzac (82) : 31 parcelles en friche (sur 3191 parcelles cadastrales)

La classification s'effectue avec les frontières communales comme référence. C'est une simplification de la définition des zones à prédire. Dans l'algorithme, on renseigne le code INSEE des communes afin d'avoir accès directement à toutes les données de ses parcelles cadastrales. Les vérités terrain sur ces 7 communes présentent 1084 parcelles cadastrales en friches pour 27708 non-friches, soit 3.765% en friches. Le dernier export Vigifriche contient 1652 parcelles cadastrales en friches sur toute l'Occitanie. Un apprentissage sur la totalité de ces vérités terrain "friches" de la région et seulement sur une partie en "non-friches" permettrait d'améliorer d'équilibrer le nombre d'échantillons de chacune des deux classes.

4

L'algorithme de classification Random Forest au service de l'inventaire des friches agricoles

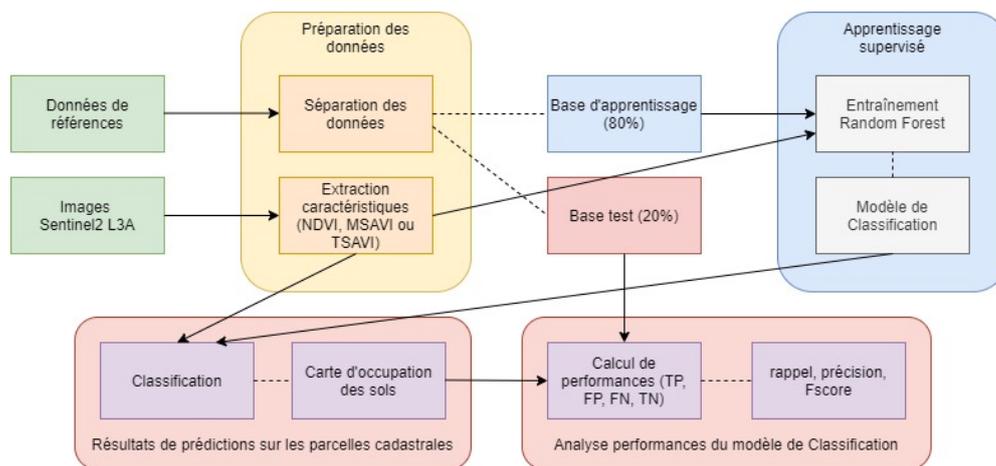


FIGURE 4.1 – Classification supervisée des parcelles cadastrales

4.1 Utilisation de Random Forest (RF) :

L'algorithme de classification utilisé fonctionne par apprentissage supervisé. Pour résumer, il construit une fonction de prédiction à partir d'exemples annotés. Dans les détails, il calcule des descripteurs (médiane et écart-type d'un indice de réflectance des végétaux ; pourcentage d'intersection avec les couches SIG telle que OCSGE, BD_TOPO, etc) afin de caractériser les parcelles cadastrales correspondant aux vérités terrains sur les communes tests. Ces descripteurs sont fusionnés pour construire un modèle de référence. Pour chaque base de donnée, on obtient en sortie les prédictions sur chaque parcelle cadastrale réparties en deux classes : friche et non-friche. Ces prédictions peuvent faire ensuite l'objet de post-traitement sur des logiciels de SIG, tel que QGIS, afin de produire des cartes spécifiques (comme en annexes D) pour diverses applications.

Pour effectuer des tests, il faut analyser les performances en sortie de l'algorithme (Figure 4.1). Pour ce faire, j'ai pu préparer de nouvelles données et réaliser l'apprentissage de nouveaux modèles. Ils sont fournis par des données de références (vérités terrain), séparés en 2 base, une de 20% où on effectue exclusivement les tests avec le modèle construit grâce aux autres 80%.

4.1.1 Objectif et choix de l'algorithme :

L'algorithme de classification mis en place combine les données disponibles pour permettre une classification suivant le cadastre parcellaire (parcelle en friches ou non-friche). Il est essentiel qu'il soit généralisable à l'ensemble du territoire et puisse être utilisé à l'avenir pour de nouveaux inventaires.

Random Forest est un algorithme de classification simple à mettre en place et qui ne demande pas une expérience importante en Deep learning pour être développé. Nécessitant peu de prétraitement des données, il améliore les prédictions d'un arbre de décision basique (Figure 4.2) en réduisant leur variance.

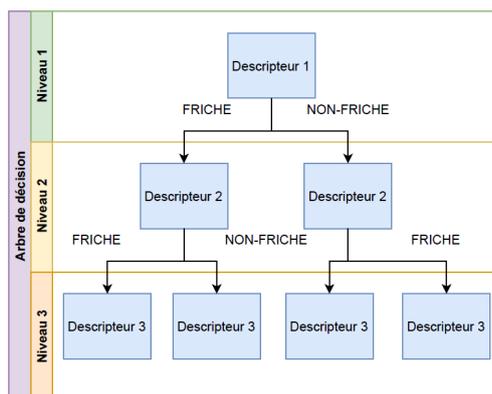


FIGURE 4.2 – Arbre de décision à trois niveaux et deux classes

Les méthodes actuelles reposant uniquement sur des couches SIG en dur, l'apprentissage supervisé de la méthode proposée (Figure 4.1) permet une meilleure utilisation des descripteurs des données en les combinant. Random Forest combine la simplicité des arbres de décision. Cet algorithme assure une généralisation de ses prédictions puisqu'il moyenne la décision de plusieurs arbres de décision. Par son processus aléatoire, il permet également un paramétrage flexible. Cela donne un modèle de bien meilleure qualité qu'un arbre de décision seul.

La performance d'un algorithme par arbre de décision basique est très dépendante de la qualité et de la quantité des données de références. La base d'apprentissage modifiée par quelques données change parfois radicalement le modèle et donc les résultats. L'assemblage de plusieurs arbres de décision, une « forêt », permet donc de moyenniser les prédictions. Plus on dispose d'arbres, plus la forêt sera fiable.

4.1.2 Principe de la classification :

Chaque arbre de la forêt aléatoire est entraîné en fractionnant aléatoirement les données d'origine selon le principe de « ensachage », avec un sous-ensemble de descripteurs (caractéristiques variables des données) selon le principe des « projections aléatoires ». Chaque arbre dispose d'une vision parcelaire du problème du fait d'un double tirage aléatoire : d'ensachage (tirage aléatoire avec remplacement sur les observations) et d'échantillonnage des descripteurs (tirage aléatoire sur les descripteurs). Ces arbres de décision sont ensuite assemblés afin de produire des résultats généralisables. Il reste difficile d'interpréter certains résultats.

Nous utilisons deux classes dans cet algorithme afin de généraliser sa prise de décision. La décision finale sur une parcelle cadastrale, moyennée sur l'ensemble des arbres sera friche ou non-friche.

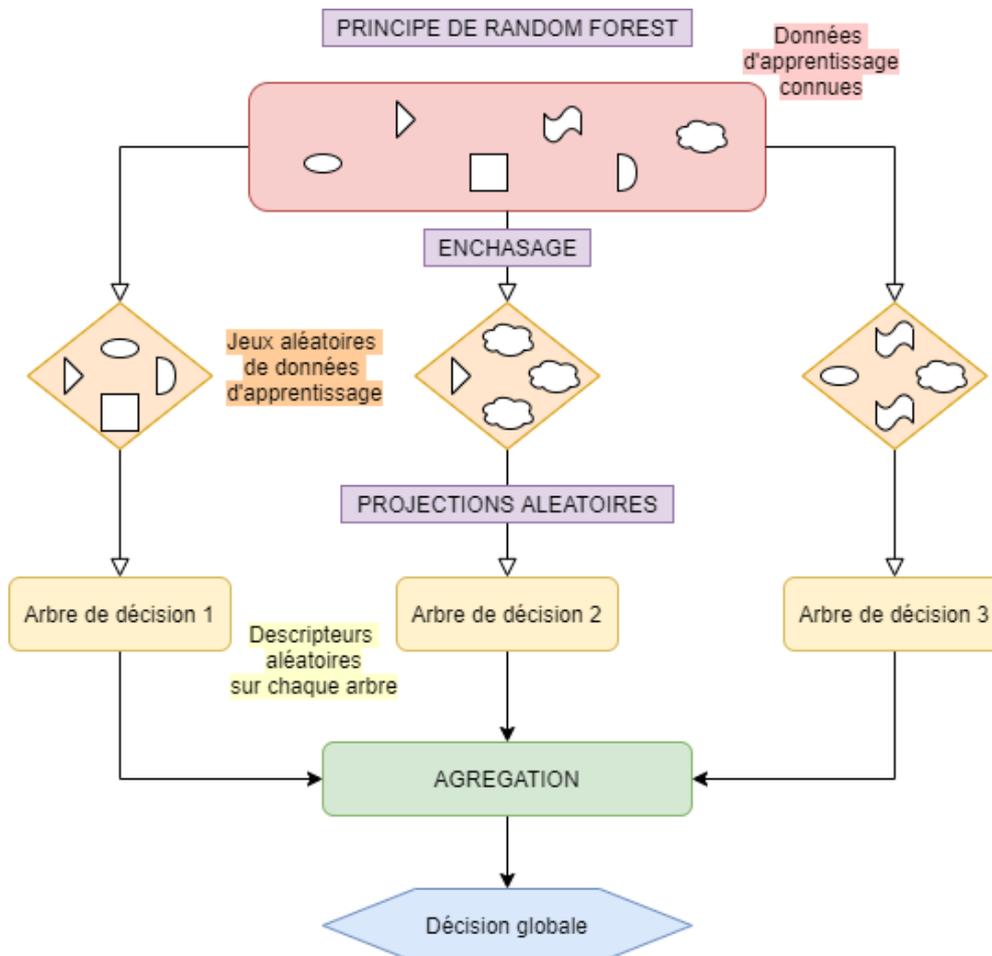


FIGURE 4.3 – Random Forest, assemblage aléatoire de données dans des arbres de décisions aux descripteurs choisis aléatoirement

4.1.3 Paramétrage RF :

Il présente trois paramètres d'entrées : la profondeur maximale, le nombre de descripteurs par arbre et le nombre de projections aléatoires. Lors du paramétrage d'un Random Forest, le choix du nombre d'arbres et du nombre de descripteurs est essentiel. Le nombre d'arbres correspond au nombre de fractionnement aléatoire (projections aléatoires) des données d'origine et la profondeur correspond au nombre de niveaux de branches (un par descripteur) utilisés dans chaque arbre de décision. Je n'ai pas utilisé de "grid search" pour tester une série de paramètres et identifier les plus utiles plus facilement, j'ai réalisé cette étude par dichotomie.

4.1.4 Apprentissage supervisé :

L'apprentissage est effectué en parallèle sur des sous-ensembles différents de données selon les principes d'enchasement et de projections aléatoires (Features sampling).

Enchassage :

- construction de n arbres de décisions en tirant aléatoirement n descripteurs
- entraînement de chaque arbre
- pour faire une prédiction sur de nouvelles données, il faut appliquer chacun des n arbres et prendre la décision majoritaire

Echantillonnage des descripteurs : tirage aléatoire de Racine n descripteurs pour un problème à n descripteurs au total

L'"unité" utilisée étant la parcelle cadastrale, certaines communes présentant beaucoup plus de parcelles que d'autres, l'équilibre est respecté au pro-rata dans la construction du modèle.

Prédictions :

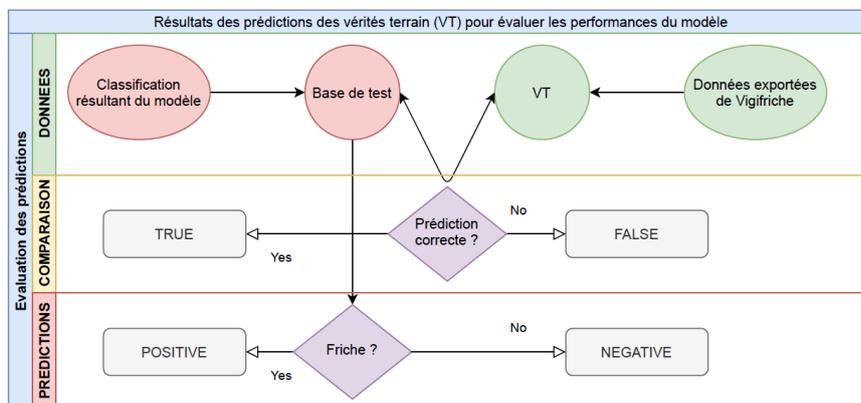


FIGURE 4.4 – Explication des sorties de prédiction

Lorsqu'on utilise l'algorithme sur des zones apprises par le modèle (vérités terrain : VT), on obtient en sortie les résultats correspondant aux prédictions sur la base de test (20%) qui donne en sortie un indice de performance, le Fscore (calculé à partir du rappel et de la précision des prédictions) des résultats suivants :

True positive	TP	parcelle bien prédite en tant que friche
True negative	TN	parcelle bien prédite en tant que non-friche
False positive	FP	parcelle mal prédite en tant que friche (VT : non-friche)
False negative	FN	parcelle mal prédite en tant que non-friche (VT : friche)

Si la prédiction d'une parcelle de la base test (Figure 4.4) correspond aux vérités terrain le résultat est TRUE (T), sinon FALSE (F). Si la même prédiction d'une parcelle dit qu'il s'agit d'une parcelle en friche alors le résultat est POSITIVE (P), sinon NEGATIVE (N). Le nombre de chacun de ces résultats de prédictions permet de construire une matrice de confusion qui sert à calculer les scores suivants :

- précision = $TP / (TP + FP)$ → pourcentage d'identifications positives
- rappel = $TP / (TP + FN)$ → pourcentage de résultats positifs
- FScore = $2 * \text{précision} * \text{rappel} / (\text{précision} + \text{rappel})$ → fiabilité globale

En ce qui concerne l'entraînement, il faut éviter le sur-apprentissage qui ne permet pas de généraliser avec assurance le modèle. Nous souhaitons donc assurer un modèle le plus générique possible qui puisse reconnaître des parcelles en friches sur différentes topologies. Le modèle peut devenir beaucoup trop influencé par les données d'entraînements, ce qui engendre un biais. Dans les algorithmes à réseaux de neurones profonds, il existe un dilemme entre efficacité et interprétabilité. Ce sont souvent les plus efficaces qui sont les moins interprétables. Random Forest permet d'avoir une bonne transparence sur l'utilisation faite des données d'entraînement, ce qui est à notre avantage.

4.1.5 Méthode d'optimisation :

- 1) Analyse des vérités terrains (VT)
- 2) Séparation des vérités terrain (VT) en base d'apprentissage/train (80%) et base test (20%)
- 3) Pour chaque parcelle cadastrale, construction d'un vecteur de descripteurs, qui sera mis en entrée de l'apprentissage supervisé. Ce vecteur contient :
 - le pourcentage d'intersection avec chaque base de données (BD) vectorielle
 - la médiane et l'écart-type d'un indice de réflectance des végétaux (NDVI, MSAVI ou TSAVI) pour chaque produit L3A
- 4) Entraînement d'un modèle Random Forest sur base d'apprentissage
- 5) Analyse des résultats sur base test

4.1.6 Limites de l'algorithme :

- Le modèle d'apprentissage, encore perfectible, présente plusieurs failles :
- l'exhaustivité des vérités terrains n'est pas assurée sur toutes les zones (en l'occurrence aujourd'hui des communes) qui servent à l'entraînement du modèle
 - la généralisation du modèle n'est pas garantie sur tout le territoire. Il faudrait pourvoir le modèle de vérités terrains de zones aux typologies représentatives de sa diversité
 - la fiabilité des tags Vigifriche, autrement dit la sensibilité des tagueurs, pose problème. L'interprétation sur le terrain est variable et sujet à débat dans beaucoup de cas. Les vérités terrains peuvent donc présenter des erreurs humaines
 - la temporalité n'est pas pris en compte dans le modèle. Une mise à jour des données par un opérateur est donc nécessaire pour rendre plus robuste le modèle

4.2 Utilisation des scripts :

L'algorithme codé en python comporte plusieurs scripts dont 6 principaux qui font appel aux scripts des différents fichiers :

- Fichier de performances des prédictions (callbacks) :
 - performances.py
- Fichier de configuration des données d'entrée (config) :
 - config_input_data.py (configuration appelée)
 - config_input_data_...py (configurations préparées à appeler dans config_input_data.py)
- Fichier calculs (modules) :
 - data_utils.py
 - s2datamodule.py
 - vectordatamodule.py
- Fichier modèle Random Forest :
 - model_RF.py
 - data_preparation_RF.py
 - merge_data.py
 - prediction.py
 - prediction_commune.py
 - prediction_tuile.py
 - theia_download.py
 - train_RF.py

L'algorithme a tourné sur le cluster HPC du CNES sur un environnement conda/4.9.0 en langage Python. Les modules nécessaires sont listés dans le fichier requirements.yml.

La méthode consiste à calculer un vecteur de caractéristiques pour chaque parcelle cadastrale, et d'entraîner puis classifier les parcelles selon leur vecteur de caractéristiques.

Les vérités terrain fournies consistent en un fichier vectoriel regroupant les parcelles cadastrales en friche pour différentes communes.

Pré-requis : récupérer les données vectorielles disponibles sur geoservices [18] et les produits Sentinel2 L3A [19] requis sur le département de travail; mettre en place l'environnement python :

- paramétrage proxy
- création environnement python (env_friches) avec installation modules :
'conda env create -name env_friches -file=requirements.yml'
'conda activate env_friches'

Remarque : la gestion des répertoires de sortie est importante. S'ils ne sont pas bien renseignés dans le fichier de configuration confi_input_data.py, on écrase d'autres répertoires.

4.2.1 Fonction des scripts :

Les différents scripts principaux permettent de :

- * theia_download.py : télécharger les tuiles de données Sentinel2 L3A correspondant à la zone à classifier [20].
- * data_preparation.py : préparer des données d'apprentissage et de test pour la détection de friches (construction du vecteur de caractéristiques pour chaque parcelle cadastrale), en utilisant des données vectorielles et satellites
- * merge_data.py : concaténer les données d'entrée de plusieurs communes (fusion de données)
- * train_RF.py : réaliser l'apprentissage d'un modèle avec Random Forest
- * prediction.py : réaliser l'inférence
- * prediction_commune.py : réaliser la classification sur une commune en renseignant son code INSEE
- * prediction_tuile.py : réaliser la classification sur une liste de commune en renseignant leurs codes INSEE

Les différents scripts qui sont appelés par les principaux permettent de :

- * config_input_data.py : configuration du jeu de données d'entrée
- * performances.py : calcul performances sur les prédictions (précision, rappel et Fscore) en comparant vérités terrains et prédictions
- * s2data_module.py : calcul indice de végétation sur les images satellites Sentinelle 2 L3A
- * data_utils.py : recherche données images selon configuration
- * vectordatamodule.py : concaténation des descripteurs vectoriels
- * model_RF.py : création modèle de type Random Forest, réalise le tree bagging et le feature sampling

4.2.2 Apprentissage d'un premier modèle multi-communes :

Cette partie présente le développement et les premiers tests de l'algorithme ont été réalisés par Vincent Poulain (Thales Group) avant mon arrivée dans le projet.

Pour analyser la généralité du premier modèle appris uniquement sur l'Honor-de-Cos, et plus généralement pour étudier la variabilité des friches et la faisabilité d'appliquer à grande échelle un modèle appris localement, les modèles de chaque commune ont été testés indépendamment pour les appliquer aux autres communes connaissant leur vérités terrain. Il se révèle que les prédictions d'un modèle monocommune ne présentent pas du tout de bons résultats en étant appliqué aux autres communes. Il est donc nécessaire de créer un modèle multi-communes qui prenne en compte différents types de friches.

Les jeux de données d'apprentissage sur les communes de LHonorDeCos, Figeac et NylsPonteilla (HFN) sont combinés (par fusion de données). Une base de données de référence est obtenue pour 1009 friches sur 22230 parcelles cadastrales de ces 3 premières communes. Pour cela il a été nécessaire de récupérer les tuiles S2 correspondant à ces communes (T31TDK et T31TDH) ainsi que les données topographiques (BD TOPO, OCSGE, BD Foret, etc) des départements 46 et 66(Lot et Pyrénées-Orientales).

L'application du modèle Random Forest à la base de test (Table 4.1) présente des scores faibles. Une optimisation de ces scores est donc primordiale pour fiabiliser l'utilisation d'un tel algorithme de classification.

TP	TN	FP	FN	précision	rappel	Fscore
155 / 216	4061 / 4230	169 / 4230	61 / 216	0.48	0.72	0.57

TABLE 4.1 – Prédications sur L'Honor-de-Cos avec le modèle 3 communes HFN

Random Forest utilisé via sklearn (bibliothèque python opensource) permet de facilement mettre en place des outils d'explicabilité des décisions. En premier lieu, les composantes du vecteur de caractéristiques les plus importantes dans la prise de décision sont mises en évidence. Le module shap est utilisé pour pousser plus loin l'analyse (Figure 4.5). On s'aperçoit que le OCSGE et le RPG sont essentiels. En effet l'OCSGE permet d'éliminer toutes les parcelles construites et le RPG permet d'éliminer les parcelles déclarées à la PAC. En termes de données image, les valeurs médianes sont plus importantes que les écarts-types calculés sur les indices de réflectance des végétaux. Tout en bas de ce classement, on retrouve les différentes classes de la BD TOPO, sauf la classe « Forêt fermée de feuillus » qui semble être la seule importante.

Un score de fiabilité (entre -1 et 1 sur la figure 4.5) est calculé pour chaque donnée vectorielle. Il correspond à son pourcentage d'intersection avec la parcelle cadastrale. Dans l'algorithme il est calculé entre 0 et 100 avec un seuil de décision à 50. Au-dessus, la classification sera en friche et en-dessous en non-friche. L'analyse de ce score permet de conclure sur le "poids" de chaque descripteur, ainsi pour la parcelle (Figure 4.5) :

- L'OCSGE et le RPG ont un impact faible sur la sortie du modèle (points

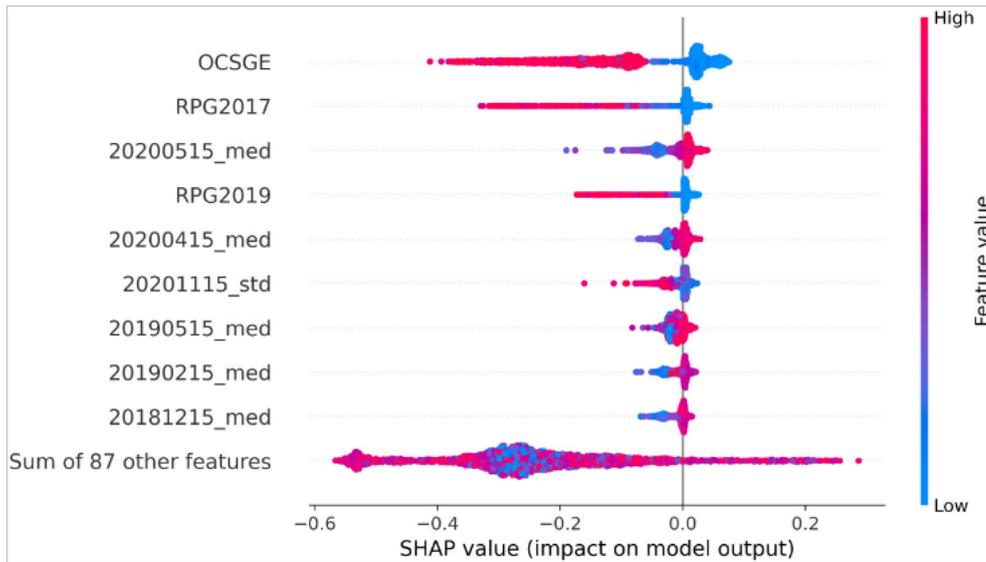


FIGURE 4.5 – Analyse des contributions des descripteurs dans la prédiction sur l’ensemble des pixels d’une parcelle cadastrale

bleu), contrairement aux fortes intersections (en rouge). En effet, le fait qu’une parcelle appartienne à OCSGE indique avec quasi-certitude que ce n’est pas une friche, par contre le fait qu’il n’y ait pas d’intersection ne nous permet pas de conclure.

- Pour les cas sans intersections avec données vectorielles, il est nécessaire de s’appuyer sur l’imagerie S2. Des valeurs élevées de NDVI sont plutôt favorables au classement en friche.

Cette interprétabilité globale permet de mettre en évidence les descripteurs majoritaires, et ceux moins utiles dans la prise de décision. Combinée à une étude par photo-interprétation, l’étude des descripteurs majoritaires sur les données de références permet de mieux comprendre les décisions sujettes à interprétation pour généraliser les résultats.

5

Optimisation de l'algorithme de classification Random Forest

Le coeur de mon travail a été de trouver des solutions, de les appliquer pour ensuite les analyser et conclure sur leur pertinence dans l'amélioration des prédictions. Les scores de rappel, précision et Fscore m'ont permis de l'évaluer. J'ai également réalisé une étude plus précise des vérités terrain par photo-interprétation pour analyser leur fiabilité.

5.1 Prise en main de l'algorithme :

J'ai donc repris l'algorithme avec des vérités terrain sur les trois communes de L'Honor-de-Cos (H), Figeac (F) et Nyls-Ponteilla (N). Après avoir mis en place mon environnement de travail python et suivi les instructions du gitlab du CNES pour utiliser le cluster HPC, j'ai pu construire un nouveau modèle HFN en y ajoutant certaines données (CVI et RPG2020 sur le 82). Celui-ci contient 22 149 parcelles dont 948 en friches (soit 4.28%).

5.1.1 Utilisation d'un nouveau modèle HFN :

J'ai ainsi effectué les prédictions avec le modèle HFN sur 11 communes du 82 et du 11 (Table 5.1). Celà m'a permis de tester le modèle et comprendre les limites des prédictions d'un modèle à 3 communes sur des communes de tailles différentes, aux friches parcellaires agricoles différentes et au nombre de friches différents.

5.1.2 Etude de la généralité du modèle HFN :

J'ai également effectué des entraînements mono-communes sur L'Honor-de-Cos (H) et Figeac (F), bi-communes (HF, HN et FN) pour valider la généralité du modèle HFN. Les résultats prédits sur les bases tests des 3 communes (Table 5.2) démontrent que le modèle présente un des meilleurs rappels (0.633 contre 0.761 pour le modèle FN) mais le sur-apprentissage joue de manière trop importante pour analyser la précision et donc le Fscore des modèles. HN et FN montrent les meilleurs Fscore.

Modèle HFN	Parcelles	Friches	Non-Friches
Coutures	268	1 (0.37%)	267
Le Pin	635	11 (1.73%)	624
Merles	1108	3 (0.27%)	1105
Lavalette (11)	1412	1	1411
Goudourville	2450	79 (3.22%)	2371
Castelmayran	2441	4	2437
Saint-Nicolas-de-La-Grave	3597	2	3595
Villemoustaussou (11)	4243	4	4239
Beaumont-de-Lomagne	5435	2	5433
Castelsarrasin	16350	97 (0.59%)	16253
Moissac	25429	327 (1.28%)	25102
TOTAUX	37939	531 (1.4%)	37735

TABLE 5.1 – Prédications avec le modèle HFN sur différentes communes

L'Honor de Cos-Figeac-Nyls Ponteilla	Précision	Rappel	Fscore
Modèle H	0.471	0.222	0.302
Modèle F	0.413	0.411	0.412
Modèle HN	0.625	0.528	0.572
Modèle FN	0.548	0.761	0.637
Modèle HF	0.347	0.522	0.417
Modèle HFN	0.354	0.633	0.454

TABLE 5.2 – Performances des 6 modèles sur la base test des 3 communes connues

Le modèle HFN montre globalement le meilleur rappel (pourcentage de friches identifiées sur toutes les friches réelles). Cependant, la précision (pourcentage de friches sur toutes les friches identifiées) reste plus faible que pour les modèles moins génériques avec une forte exhaustivité sur leurs communes. À noter que l'utilisation des données de références Nyls-Ponteilla dans l'apprentissage présente les meilleurs résultats ce qui peut directement être relié au fait que les validations terrain ont été exhaustives sur cette commune.

5.1.3 Étude du score de fiabilité :

Le score de fiabilité (Figure 5.1) variant sur les prédictions peut être ajusté pour les optimiser. Deux seuils sont présents à 30 et 58 et permettent d'enlever de nombreuses fausses prédictions. Ce score présente des valeurs de 0 à 100 avec un seuil de décision à 50. Cependant, trop peu de données de référence en friche sont présentes pour affirmer que la modification du seuil de décision (de 50 à 30 ou 58) soit concluante.

5.1.4 Ajout de nouveaux descripteurs au modèle HFN :

L'ajout des données CVI améliore très légèrement les performances (on passe d'un Fscore de 0.57 lors des tests précédemment effectués - Table 4.1 - à un Fscore de 0.606 ; la

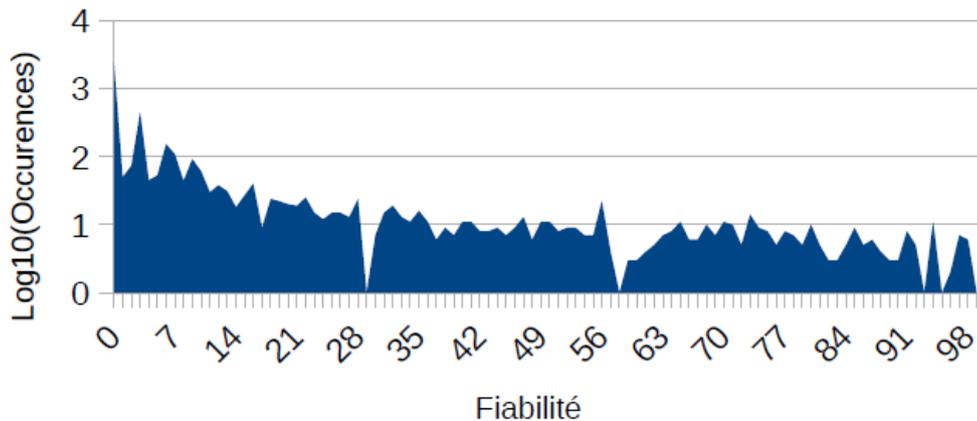


FIGURE 5.1 – Fiabilité des prédictions sur L'Honor-de-Cos à partir du modèle HFN

précision a diminué de 2.29% et le rappel augmentant de 19%). Cependant, le RPG2020 n'étant disponible que sur le Tarn-et-Garonne, il est impossible de créer un modèle HFN avec, puisqu'il n'est pas fourni sur les départements de Figeac et Nyls-Ponteilla.

5.2 Optimisations successives du modèle HFN :

5.2.1 Modification de l'indice NDVI en MSAVI et TSAVI :

Durant mes tests, j'ai pu comparer les performances du NDVI avec deux nouveaux indices de réflectance des végétaux. Ils sont tous les deux tirés de l'indice SAVI (indice de végétation ajusté au sol) qui est une combinaison du NDVI et du PVI qui permet de mesurer la couverture végétale en minimisant les influences de la luminosité du sol. SAVI permet de prendre en compte l'interaction entre le feuillage et le sol qui manquait au NDVI pour être généralisable :

- TSAVI [21] : Transformed Soil-Adjusted Vegetation Index (indice de végétation transformé ajusté au sol) suppose que la ligne du sol présente une pente et une interception arbitraires. Il se calcule par le rapport :

$(s * (NIR - s * R - a)) / (a * NIR + R - a * s + X * (1 + s^2))$ où s est la pente de la ligne de sol, a , l'intersection de ligne de sol et X le facteur d'ajustement défini pour minimiser le bruit du sol

- MSAVI [22] : Modified Soil-Adjusted Vegetation Index (indice modifié de végétation ajusté au sol) qui minimise l'effet de sol nu de SAVI. Il se calcule par le rapport : $(1/2) * (2 * (NIR + 1) - \sqrt{(2 * NIR + 1)^2 - 8(NIR - R)})$

J'ai analysé les différences entre l'utilisation des indices de réflectance des végétaux (TSAVI, MSAVI et NDVI, précédemment utilisé) renseignés dans `s2data_module.py`. L'amélioration est très légère (0.25% sur le Fscore) pour l'indice TSAVI. J'ai donc choisi de conserver cet indice pour la suite des tests.

5.2.2 Tests par exclusion de données SIG :

La fiabilité des données SIG ayant été remise en cause dans le projet, j'ai effectué des tests par exclusion. La Table 5.3, présente les proportions de chaque résultat par rapport au nombre de friches prédites pour les TP et FN, de non-friches prédites pour les FP et TN. En effet, elles peuvent entraîner un sur-apprentissage néfaste aux prédictions. J'ai donc entraîné de nouveaux modèles HFN en excluant les données OCSGE et RPG qui pourraient être source de "confusion".

L'Honor de Cos-Figeac-Nyls Ponteilla	Précision	Rappel	Fscore
Sans OCSGE	0.393	0.633	0.485
TP : 114/180	FP : 176/4250	FN : 66/180	TN : 4074/4250
Sans RPG	0.317	0.611	0.417
TP : 110/180	FP : 237/4250	FN : 70/180	TN : 4013/4250
RPG2017 uniquement	0.398	0.639	0.49
TP : 115/180	FP : 174/4250	FN : 65/180	TN : 4076/4250
RPG2019 uniquement	0.434	0.622	0.511
TP : 112/180	FP : 146/4250	FN : 68/180	TN : 4104/4250
Modèle de référence (sans exclusion)	0.354	0.633	0.454
TP : 114/180	FP : 174/4250	FN : 65/180	TN : 4076/4250

TABLE 5.3 – Exclusion de certaines données SIG du modèle HFN

On constate bien que l'exclusion totale du RPG ne présente pas d'intérêt. Cependant celle d'une partie des RPG présente de meilleurs résultats, notamment en utilisant uniquement le RPG2019 (le plus récent utilisé sur les 3 communes). L'exclusion de l'OCSGE présente également des améliorations sur la précision. L'exclusion complémentaire de l'OCSGE aux RPGs ne présente pas de résultats probants. La précision régresse. J'ai donc choisi de n'utiliser pour la suite de mes tests que le modèle contenant uniquement le RPG2019. Il présente une amélioration de 12.5% du Fscore par rapport au modèle de référence (0.454 à 0.511) avec une légère baisse du rappel (1.8%) et une augmentation conséquente de la précision (22.6%).

5.2.3 Utilisation d'images trimestrielles :

Mes tests ont ensuite consisté à exclure une partie des images satellites (Table 5.4) pour évaluer la différence de performances entre exploiter une seule image satellite par mois et en utiliser une par trimestre (ce qui amène à 11 images au lieu de 41 sur lesquelles est calculé l'indice TSAVI). La différence se révèle être significative, le gain de temps d'exécution de l'algorithme est important cependant. Les images trimestrielles présentent l'intérêt d'être moins longues à traiter mais baissent de manière très importante la précision (32.3%) sans modifier le rappel. J'ai donc choisi d'utiliser les images trimestrielles pour la suite des tests.

L'Honor de Cos-Figeac-Précision Nyls Ponteilla (RPG2019)		Rappel	Fscore
Images mensuelles	0.434	0.622	0.511
TP : 112/180	FP : 146/4250	FN : 68/180	TN : 4104/4250
Images trimestrielles (Avril/Août/Décembre)	0.328	0.622	0.43
TP : 112/180	FP : 229/4250	FN : 68/180	TN : 4076/4250

TABLE 5.4 – Performances avec utilisation d'images satellites trimestrielles

5.2.4 Modification des hyper-paramètres :

Profondeur des arbres de décision RPG2019-trimestre	Précision	Rappel	Fscore
6/30/5	0.146	0.839	0.248
TP : 151/180	FP : 886/4250	FN : 29/180	TN : 3364/4250
12/30/5	0.246	0.694	0.363
TP : 125/180	FP : 384/4250	FN : 55/180	TN : 3866/4250
16/30/5	0.328	0.622	0.43
TP : 112/180	FP : 229/4250	FN : 68/180	TN : 4076/4250
20/30/5	0.44	0.572	0.498
TP : 103/180	FP : 131/4250	FN : 77/180	TN : 4119/4250
24/30/5	0.492	0.517	0.504
TP : 93/180	FP : 96/4250	FN : 87/180	TN : 4154/4250

TABLE 5.5 – Modification de l'hyper-paramètre de profondeur des arbres de décision

Plus j'ai augmenté la profondeur des arbres de décisions (Table 5.5), plus le Fscore est important, car la précision augmente. Mais en contrepartie, le rappel baisse. Toute l'optimisation de classification tourne autour de l'équilibre entre rappel et précision. Une profondeur de 20 semble être un bon compromis du fait qu'elle augmente significativement la précision (34%) sans diminuer énormément le rappel (-8%) d'une profondeur à 16, pour un Fscore quasi égal à 0.5 ici. La profondeur a une incidence faible.

Très peu de différence avec une profondeur d'arbres de 20, le nombre d'arbres de décisions ne change qu'à quelques détails près. On peut choisir le meilleur Fscore pour 20/50/5.

De la même manière que pour le nombre d'arbres, le nombre de descripteurs pour l'ensachage n'importe pas beaucoup. On trouve le meilleur résultat pour 20/50/30. De plus chaque modification du modèle entraîne un paramétrage optimal spécifique. Il faudra donc modifier ces paramètres à chaque nouvel ajout de VT.

5.2.5 Comparaison gain avec optimisation hyper-paramètres :

J'obtiens en sortie de classification le même Fscore (Table 5.6) avec des images trimestrielles (20/50/30) qu'avec des images mensuelles (16/30/5). La progression importante sur la précision (36% images par trimestre, 18% images mensuelles) induit une faible régression sur le rappel (5.4%).

L'Honor de Cos-Figeac-Précision		Rappel	Fscore
Nyls Ponteilla (RPG2019)			
Images mensuelles 20/50/30	0.512	0.59	0.548
TP : 106/180	FP : 101/4250	FN : 74/180	TN : 4149/4250
Images trimestrielles 20/50/30 (Avril/Août/Décembre)	0.446	0.594	0.51
TP : 107/180	FP : 133/4250	FN : 73/180	TN : 4117/4250
Images mensuelles - 16/30/5	0.434	0.622	0.511
TP : 112/180	FP : 146/4250	FN : 68/180	TN : 4104/4250
Images trimestrielles - 16/30/5 (Avril/Août/Décembre)	0.328	0.622	0.43
TP : 112/180	FP : 229/4250	FN : 68/180	TN : 4076/4250

TABLE 5.6 – Comparaison gain après optimisation hyper-paramètres

5.3 Optimisation suite aux missions terrain :

5.3.1 Nouvelles données de références :

Les missions de repérage sur l'Honor-de-Cos m'ont permis de modifier les vérités terrain "erronées" pour fiabiliser celles présentes dans le modèle d'apprentissage. Suite à la mission terrain effectuée avec les élèves du lycée agricole de Moissac, nous avons à disposition de nouvelles vérités terrains (Table 5.7) utilisables pour construire un modèle légèrement plus générique. J'ai ainsi pu, grâce à un nouvel export Vigifriche, les intégrer.

Modèle HFN	Parcelles	Friches	Non-Friches
Coutures	268	1 (0.37%)	267
Le Pin	635	11 (1.73%)	624
Goudourville	2450	79 (3.22%)	2371
TOTAUX	3353	91 (2.71%)	3262
Ajout Vérités Terrain	13.15 %	8.76 %	13.3 %

TABLE 5.7 – Vérités terrain collectées sur 3 nouvelles communes

Plusieurs remarques importantes sont à faire suite à ces missions concernant l'algorithme. Il y a trop peu de friches dans les petites communes. Le modèle utilisé est trop peu générique. Et s'il n'y a pas de friches, aucun fichier ne peut s'exporter lors d'une prédiction (Le Pin et Coutures).

5.3.2 Comparaison modèles 6C et 3Cs (HFN et CPG) :

Le modèle CPG (Coutures-LePin-Goudourville) présente trop peu de données pour que l'analyse soit concluante sur sa base test. On remarque néanmoins (Table 5.8) un seul FN sur 663 parcelles ce qui montre une bonne exhaustivité. La précision n'est pas aussi significative que le rappel dans le calcul du Fscore, il faudrait utiliser des poids différents.

RPG2019-trimestre- 20/50/30	Précision	Rappel	Fscore
Coutures-LePin- Goudourville	0.375	0.75	0.5
TP : 3/8	FP : 5/663	FN : 1/8	TN : 662/663
6C	0.321	0.613	0.422
TP : 117/191	FP : 247/4910	FN : 74/191	TN : 4663/4910
L'Honor de Cos-Figeac-0.512		0.59	0.548
Nyls Ponteilla			
TP : 106/180	FP : 101/4250	FN : 74/180	TN : 4149/4250

TABLE 5.8 – Modèle 6C, plus générique, meilleur rappel mais précision plus faible

Le modèle 6C présente le meilleur rappel interprétable comme attendu (0.613) mais de trop nombreux FP (249, plus du double) se rajoutent et devraient être vérifiés sur le terrain (précision diminuée de presque 60% par rapport au modèle HFN). L'ajout des vérités terrain des trois communes dans l'apprentissage entraîne donc une diminution importante de la précision.

5.3.3 Utilisation du modèle générique 6C sur les plus grosses communes :

Modèle 6C RPG2019- trimestre	Précision	Rappel	Fscore
Goudourville	0.207	0.857	0.333
TP : 6/7	FP : 23/483	FN : 1/7	TN : 460/483
L'Honor de Cos	0.313	0.886	0.463
TP : 31/35	FP : 68/1283	FN : 4/35	TN : 1215/1283
Figeac	0.431	0.942	0.591
TP : 81/86	FP : 107/2320	FN : 5/86	TN : 2213/2320
Nyls-Ponteilla	0.698	0.949	0.804
TP : 74/78	FP : 32/629	FN : 4/78	TN : 597/629

TABLE 5.9 – Modèle 6C_RPG2019_trimestre appliqué aux communes tests

En appliquant le modèle 6C aux données, une à une (et non fusionnées HFN comme en Table 5.8), de la base test sur les 4 plus grosses communes, on obtient (Table 5.9) bien de meilleurs résultats sur le rappel (entre 0.857 pour G et 0.949 pour N) que le modèle HFN initial (entre 0.606 pour F et 0.822 pour N). Cependant la précision diminue (de 0.469 à 0.313 pour H ; de 0.441 à 0.431 pour F ; de 0.733 à 0.698 pour N). Le rappel est plus important que la précision dans une optique d'un inventaire complet sur un grand territoire. La généricité du modèle 6C est donc approuvée.

A titre de comparaison, les résultats du modèle HFN mensuel sans exclusion (Table 5.10) présentent une meilleure précision (grâce aux images mensuelles) mais un rappel plus faible.

5.3.4 Optimisations non-effectuées :

J'ai rajouté de nouvelles vérités terrain sur la commune de Meauzac qui présentait une étude sur les friches agricoles dans son PLU (Plan Local d'Urbanisme) de 2020 [23].

Modèle 3C HFN-mensuel	Précision	Rappel	Fscore
L'Honor de Cos	0.469	0.857	0.606
TP : 30/35	FP : 34/1283	FN : 5/35	TN : 1249/1283
Figeac	0.441	0.965	0.606
TP : 83/86	FP : 105/2320	FN : 3/86	TN : 2215/2320
Nyls-Ponteilla	0.735	0.923	0.818
TP : 72/78	FP : 26/629	FN : 6/78	TN : 603/629

TABLE 5.10 – Modèle HFN mensuel appliqué aux communes tests

J'ai demandé un nouvel export à la SAFER Occitanie qui a été fourni courant août. Un modèle 7C peut donc être construit et devrait présenter une meilleure généralité que celui 6C. L'utilisation de ce dernier sur la commune de Meuzac prédit 96 friches pour 31 friches annotées soit une précision faible de 0.244. Des vérifications terrain sont donc souhaitables pour renforcer ces données de références.

Pour compléter mon étude, l'optimisation de l'algorithme est secondaire car elle permet une consolidation certaine mais faible du modèle. Voici une liste des tests non effectués qui pourraient permettre de meilleures prédictions :

- exclusion des écarts-type ou médianes des indices de végétations calculés sur le parcellaire des images satellites
- modification du seuil de décision (56 au lieu de 50)
- exclusion des parcelles trop petites (surfaces inférieures à 1000 m²) ou trop fines (largeur inférieures à 20 m²)
- ajout de nouvelles couches SIG utiles
- ouverture des images parcellaires pour éviter les problèmes de bords qui faussent les prédictions
- utilisation du modèle 7C_RPG2019_mensuel
- utilisation d'un autre algorithme de classification type Temporal CNN (Convolutional Neuron Network)

6

Connaissance et vérifications terrain

6.1 Réflexions sur les limites de la classification :

J'ai eu la chance de pouvoir organiser ou participer à plusieurs missions sur le terrain pour répondre aux différentes problématiques soulevées par la mise en pratique de l'algorithme. Ce dernier présente en effet plusieurs limites en lien avec l'apprentissage des vérités terrain (issues de Vigifriche je le rappelle) et des couches SIG :

- le modèle apprend-t-il une variété suffisamment large de friches agricoles pour être appliqué à l'ensemble du territoire? La pratique montre que le modèle d'apprentissage gagne en généralité en ajoutant de nouvelles vérités terrain ou en les "améliorant" avec des validations sur le terrain. L'analyse du score de rappel de mon étude en est la caractéristique, puisque le nombre de faux-négatifs (parcelle mal prédite en tant que non-friche ayant donc pour VT friche) diminue bien pour tendre vers 0. Autrement dit, avec un modèle générique comportant davantage de VT, les prédictions "oublieront" moins de parcelles en friches.

- les vérités terrain considèrent-elles toutes les parcelles cadastrales sur la zone d'apprentissage? L'inventaire effectué sur le terrain de certaines communes ne prend pas en compte certaines parcelles inaccessibles qui sont laissées à l'interprétation par photo-aérienne ou images satellitaire. L'analyse du score de précision de mon étude en est la caractéristique, puisque le nombre de faux-positifs (parcelle mal prédite en tant que friche ayant donc pour VT non-friche) dépend en partie de la validation. En effet, les scores de précision sont moins importants sur les communes où l'inventaire est complexe. L'utilisation de frontières communales pour réaliser l'apprentissage sur toutes ses parcelles, n'est pas optimal. Ce choix est avant tout pratique, il serait néanmoins plus juste d'utiliser des zones plus petites par îlots sur lesquelles les fausses interprétations seraient moins nombreuses. De plus, cela permettrait de palier au déséquilibre entre les deux classes qui fausse notre interprétation quant au calcul de la précision (il y a environ 4% de friches annotées dans le modèle le plus générique). Ce dernier point rejoint la dernière question d'importance.

- les vérités terrain sont-elles fiables? Et comment? En effet, deux facteurs viennent influencer les résultats sur nos vérités terrain. Le premier réside dans le choix de classifier des parcelles cadastrales dont l'occupation du sol peut évoluer constamment (les millésimes des données d'apprentissage et des déclarations doivent être adaptés) et qui peuvent présenter sur une partie de leur surface des friches. Il a été choisi dans cette étude de ne considérer que celles déclarables sur Vigifriche dont plus de la moitié de la surface était en friche sachant que certaines occupations du sol ne sont pas excluantes sur l'application mais le sont lors des prédictions. Cela m'amène au deuxième facteur, l'humain. La sensibilité des agents lors des déclarations Vigifriche est à entretenir pour garantir leur fiabilité. Les déclarations doivent être systématiquement vérifiées à l'aide d'un logiciel comme QGIS, en superposant les couches SIG excluantes qui ne sont pas intégrées et mises à jour sur Vigifriche. Quand il s'agit de friches agricoles, il y aura toujours certaines indéisions qui méritent de retourner sur le terrain. Le Fscore présente un bon moyen d'évaluer l'efficacité d'un modèle de prédiction en faisant un compromis entre sensibilité (rappel) et valeur ajoutée (précision). C'est dans cette optique là que j'ai commencé à préparer ma première mission sur la commune test du projet : L'Honor-de-Cos (Figure 6.1 : TN rouge, FN rose, FP jaune, VP vert).

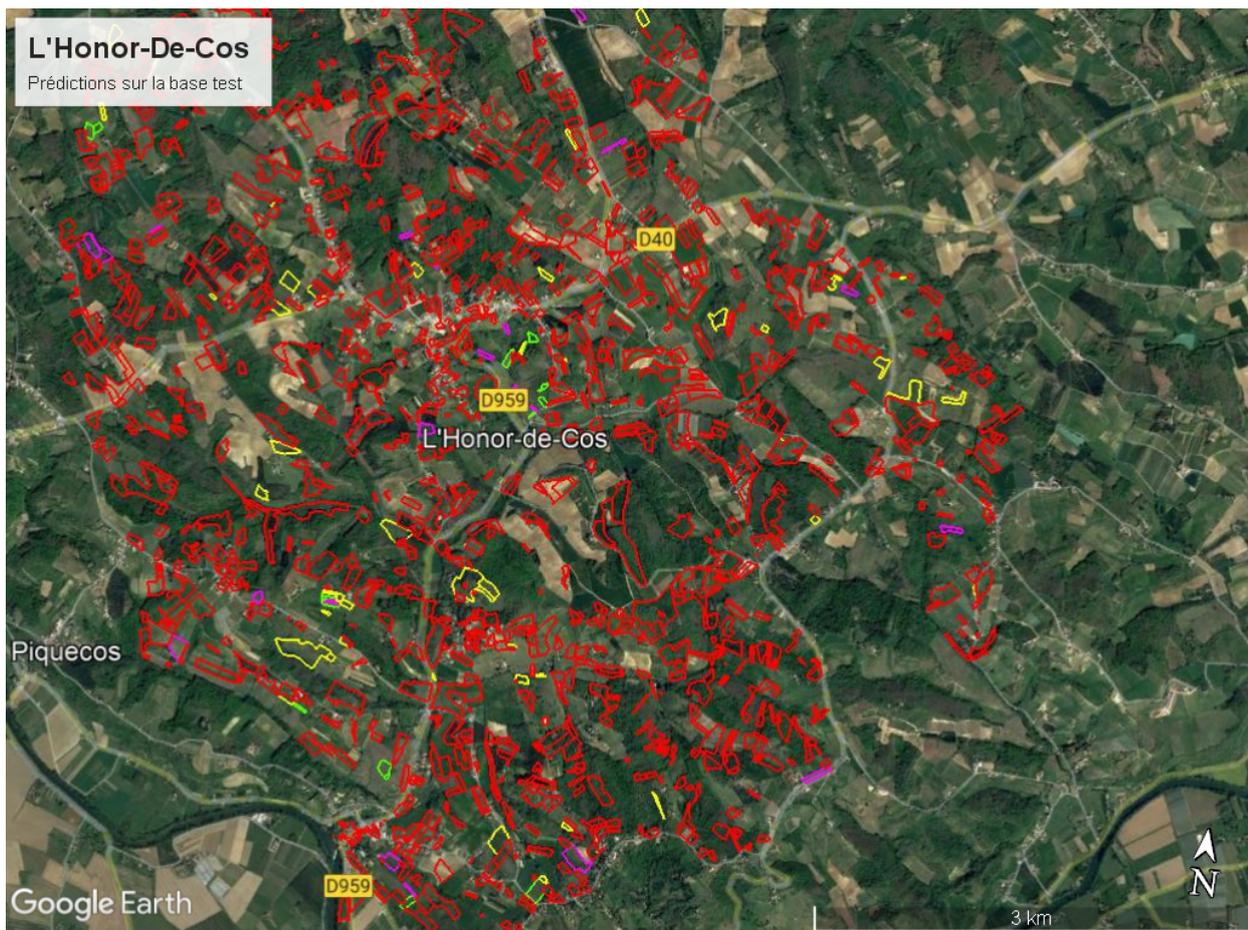


FIGURE 6.1 – Carte des prédictions sur base test de L'Honor-de-Cos

6.2 Missions de vérification sur L'Honor-de-Cos :

Suite à mes premiers tests avec l'algorithme sur L'Honor-de-Cos avec le modèle à 3 communes, j'ai analysé les différentes prédictions en me penchant notamment sur les fausses prédictions. Certaines parcelles de la base test prêtent à réflexion. C'est dans la démarche de comprendre et vérifier les données Vigifriche que je me suis donc rendu sur le terrain avec une voiture de service les lundis 7/04 et 19/04.

Après plusieurs journées de préparation des itinéraires et des zones à vérifier, la première mission ne m'a pas permis de faire le tour des parcelles d'intérêts sur le terrain. J'ai pu analyser 63 parcelles ce jour là (Table 6.1). Beaucoup de prédictions de l'algorithme y sont fausses au vu de mes repérages et des vérités terrain. La deuxième mission a consolidé mes vérifications sur certaines zones et a finalisé celles qui étaient isolées des cartes que j'ai utilisées pour me repérer. J'ai ainsi validé les 115 parcelles prédites (TP, FN et FP) sur la base test qui présentaient un intérêt à vérifier selon mes analyses.

Analyse de 63 parcelles (20 en orthophotos) dont 29 peu accessibles dont 9 nécessitant le drone												
			MAUVAISE s Prédictions	2 TP, 18 FP	3 FN	% ERREURS prédictions (par rapport au repérage plus aux vérités terrains)	69 % (20/29)	8.8 % (3/34)	14.3 % (2/14)	48.7 % (18/37)	25 % (3/12)	
	Parcelles validées	Dont peu accessi- bles	Dont accessi- bles	Dont orthopho- tos	2ième validation drone	FRICHE OU PAS FRICHE ?	PAS FRICHE	FRICHE	TP (bien classée friche)	FP (mal classée friche)	FN (mal classée non- friche)	
ZONE 1 (NORD-EST)	13	5	8	0	5		5 (3 FP, 2 FN)	8 (3 TP, 5 FP)	3	8	2	
ZONE 2 (CENTRE)	20	3	17	0	3		9 (1 TP, 3 FP, 5 FN)	11 (5 TP, 4 FP, 1 FN)	7	7	6	
ZONE 3 (SUD-OUEST)	10	8	2	5	3		7 (6 FP, 1 FN)	3 (1 TP, 2 FP)	1	8	1	
ZONE 4 (SUD-EST)	15	13	2	15	0		7 (1 TP, 5 FP, 1 FN)	8 (1 TP, 7 FP)	2	12	1	
ZONE 5 (NORD-OUEST)	5	0	5	0	0		1 FP	4 (1 TP, 1 FP, 2 FN)	1	2	2	
TOTAUX	63	29	34	20	11		26 (5 TP, 15 FP, 9 FN)	34 (12 TP, 19 FP, 3 FN)	14	37	12	
						% ERREURS vérités terrains par rapport au repérage	38 % (11/29)	65 % (22/34)				

TABLE 6.1 – Analyse des parcelles vérifiées lors de la première mission sur l'Honor-de-Cos

Afin de valider l'exhaustivité sur la commune-test de L'Honor-de-Cos, des compléments sont nécessaires suite à ces deux repérages. Une mission utilisant l'imagerie drone est alors envisagée pour ajouter de l'information sur les parcelles qui ne m'étaient pas accessibles depuis la route. De plus, les friches étant souvent sujettes à interprétation (notamment du fait de leur définition parcellaire) une stratégie de « tag and go » nécessitant une deuxième étape de validation (comparaison avec photo-interprétation et analyse des couches SIG) a été utilisée. Ainsi se pose la question centrale : "Taguer davantage ou taguer mieux ?" Il est essentiel d'avoir des données de références de qualité mais également de d'obtenir un modèle généralisable. J'ai donc été amené à préparer de nouvelles missions dans ce sens.

6.3 Missions d'inventaire sur de nouvelles communes :

Afin de collecter davantage de vérités terrains, il convient de fournir Vigifriche, sur des communes aux typologies différentes (céréalière et montagnaise).

Pour ce faire, le lycée d'enseignement agricole (LEPA) de Moissac a été contacté afin que ses élèves soient initiés à l'application et effectuent un taguage encadré sur plusieurs communes choisies spécifiquement.

Les critères pour cibler ces communes sont les suivants :

- accessibilité (visibilité depuis la route)
- commune céréalière
- petite taille
- peu ou pas de relief
- proche de Moissac

Durant l'élaboration de la mission, j'ai analysé et préparé les itinéraires sur les 5 communes pour optimiser les deux journées terrain durant lesquelles nous avons pu parcourir 3 d'entre elles :

Le Pin : 4.71 km² ; Coutures : 6.9 km² ; Goudourville : 11.27 km² ;

Suite à la prise de contact du lycée, j'ai été mis en relation avec Arnaud BLONDEL, enseignant en mathématiques et informatique, pour coordonner la mission durant tout le mois de mai. Nous avons ainsi mis au point l'organisation aussi bien administrative (convention, contact mairies) et logistique (minibus, pique-nique, gilets jaunes, planning), que technique (demande création de comptes Vigifriche, installation, bonne utilisation) et fonctionnelle (itinéraires, suivi WhatsApp, choix des communes). Cette expérience que j'ai coordonnée, a fait l'objet d'un rapport pour les différents partenaires du projet. L'objectif a été de valider la méthodologie, l'encadrement et la formation d'une sortie terrain avec ces lycéens pour collecter exhaustivement (toutes les parcelles des communes ciblées) de nouvelles vérités terrains. J'ai mis en place une conversation Whatsapp pour échanger et suivre le bon déroulé de la mission. Il a été d'une grande utilité pour créer une émulation rapide avec les élèves, les motiver.

Le 27 mai, j'ai pu présenter, aux côtés de Thomas ROUMAGNAC (SAFER - directeur départemental 82), de Gilles LEFRANCOIS (SAFER - attaché régional) et Flavie BERGOUNIOUX, la mission aux élèves dans leur lycée. Cette phase d'animation, a permis d'explicitier le contexte des friches agricoles, présenter les structures de la Safer et de la DDT82, le déroulé du projet ainsi que l'organisation pratique sur le terrain. La journée s'est terminée par un jeu « Friche ou pas friche ? » très bien reçu par tous. Suite à cette animation, le taguage s'est effectué sur deux journées : Lundi 31 mai avec 11 élèves et Lundi 7 juin avec 15 élèves. Ils ont été répartis en quatre groupes (G1 -rose-, G2 -jaune-, G3 -vert- et G4 -bleu) de 3 à 4, encadrés par 4 enseignants. Chaque groupe d'élèves est en possession d'itinéraires (Annexe C), de 3 cartes de suivi de mission, de batteries externes et de smartphones.

Bilan première journée : l'exhaustivité a été assurée sur les deux communes de Coutures (matin) et Le Pin (après-midi). Peu de friches étaient présentes mais l'exhaustivité va permettre de compléter de manière efficace et fiable les vérités terrain utilisées pour l'apprentissage du modèle générique de l'algorithme. L'utilisation de Vigifriche ne s'est pas faite sur le terrain, les élèves n'avaient pas assez été cadrés et sensibilisés. De légers retards ont eu lieu par rapport au planning prévisionnel.

Bilan deuxième journée : l'exhaustivité a été assurée à 95% sur la commune de Goudourville. Beaucoup de friches (70) compléteront avec efficacité les vérités terrain utilisées pour l'apprentissage du modèle générique de l'algorithme. L'organisation logistique a été améliorée par rapport à la première sortie et de nombreuses améliorations ont été apportées : vérification utilisation Vigifriche et briefing sur les itinéraires avant de partir en mission, délais du planning respectés, cadrage plus rigoureux des élèves sur le terrain. L'organisation et les itinéraires (Annexes) ont été efficaces mais restent perfectibles.

Les objectifs de la mission ont été atteints sur les deux journées terrain. Après vérification des déclarations par Maud CHEVIGNON (Safer), les nouvelles vérités terrains sont utilisables pour rendre le modèle d'apprentissage de l'algorithme plus générique. Aussi, l'exhaustivité de l'inventaire des friches agricoles a été assurée dans la mesure du possible. La mission s'est déroulée dans de bonnes conditions. De nombreuses améliorations à cette première ont été relevées (décrites à travers le retour d'expérience que j'ai élaboré) et pourront permettre de réitérer une telle expérience.

« Que peut-on améliorer dans le futur ? » : l'expérience utilisateur de l'application Vigifriche a été satisfaisante sur la deuxième journée après correction suite à la première journée ; les délais étaient trop courts entre l'animation et la première mission terrain ; l'installation de l'application mobile n'est pas possible avec IOS ; une connexion directe via internet permet de contourner ce problème. Maud CHEVIGNON (SAFER) a fourni aux élèves les accès (identifiant et mot de passe) pour utiliser l'outil que ce soit par l'application ou par internet. L'application fournit un moyen efficace, rapide et précis, pour taguer. Un message d'erreur permet d'avoir connaissance des déclarations à la PAC. Les appareils photos ne sont pas toujours paramétrables pour permettre de taguer avec image à l'appui (cela dépend du smartphone). Il semble pertinent de passer plus de temps à former les élèves (également leurs encadrants sur le terrain, qui n'étaient pas tous présents lors de l'animation) et à vérifier qu'ils aient tous bien installé et compris comment utiliser l'application. Visualiser les déclarations à la PAC et de la BD_Forêt pourrait permettre d'optimiser les déclarations et les validations.

L'encadrement des élèves et les réglages techniques sont importants. Deux demi-journées d'intervention supplémentaires ont été prévues pour limiter les risques d'intempéries. Le parcellaire est compliqué à interpréter. Pour certaines parcelles étant en partie seulement en friches, une estimation à 50% de la surface a été considérée pour faire une déclaration.

6.4 Compléments de vérification sur L'Honor-de-Cos avec des images drones :

L'objectif de cette dernière mission a été de valider les parcelles qui n'ont pu l'être lors du repérage terrain sur L'Honor-de-Cos à l'aide d'images prises par les drones. Les zones de survols ont été définies par Gilles AGRECH (Figure 6.2).

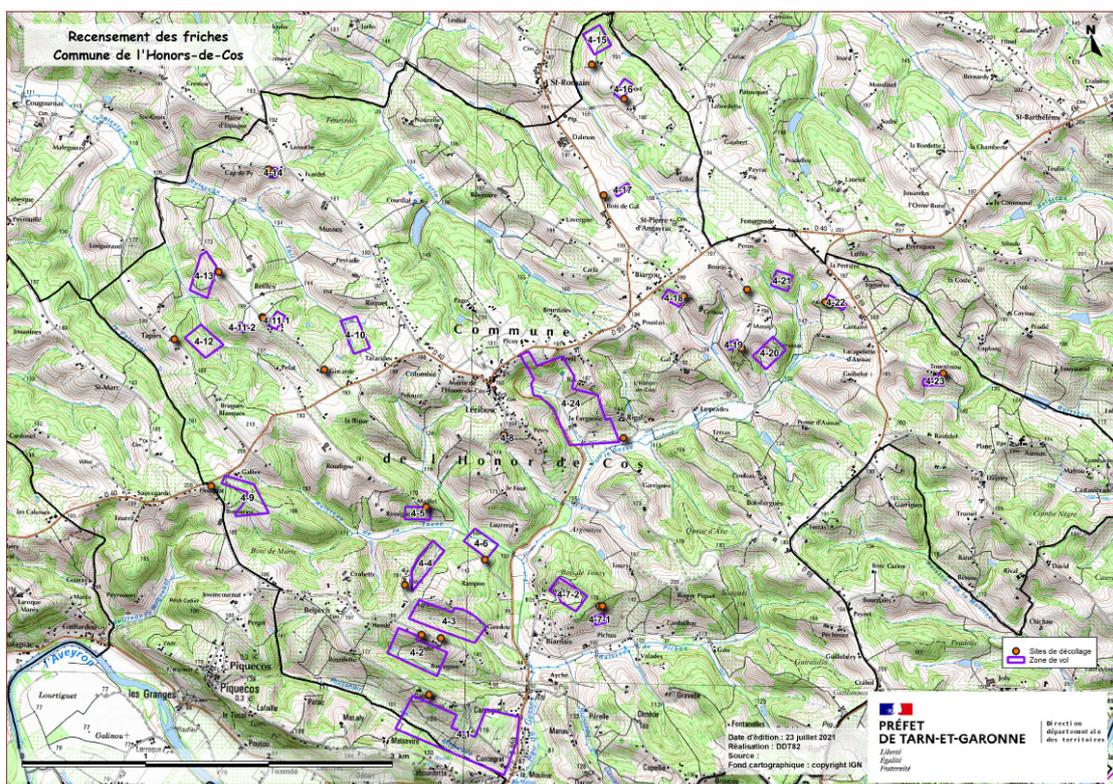


FIGURE 6.2 – 23 zones de survols pour la Team Drone d'Occitanie, ©DDT82

Les images RGB drone vont permettre de récupérer des données 2D et 3D. Elles permettront également de fournir des couches MNS, MNF et MNH sur les parcelles étudiées.

Cette mission est la première réunissant 6 pilotes de 5 départements d'Occitanie (12,31,32,81 et 82). Cinq drones quadri-rotors ont été utilisés, un drone multispectral aurait pu l'être également. L'échange des compétences, savoir-faire de pilotes expérimentés et des données autour de cette mission sont primordiaux. Par la suite, l'analyse de la reproductibilité, de la généricité et des résultats permettra de connaître l'efficacité d'une telle méthode.

La mission s'est déroulée sur la journée du 27 août après avoir été repoussée plusieurs fois du fait de l'indisponibilité de certains télépilotes. Gilles AGRECH, carthographe et géomaticien de l'équipe du BIGTI s'est chargé de préparer la mission avec ses collègues de la Team Drone.

Un télépilote doit être accompagné d'un assistant pour assurer sa sécurité celle du drone. La mission a donc mobilisé 10 personnes. Il a fallu mettre en place une logistique importante compte tenu du nombre de participants et de la complexité de la mise en oeuvre. Le plus important étant de vérifier que tout le monde est bien véhiculé et que les batteries des drones sont suffisamment nombreuses pour que les agents puissent rester sur le terrain. Après avoir communiqué aux propriétaires le déroulé de la mission, des feuilles de missions ont été distribuées à chaque télépilote. Nous avons reçu un accueil très enthousiaste du maire de la commune, M. Lamolinairie. Une vidéo de communication à la suite de cette mission est envisagée pour promouvoir le travail d'équipe des télépilotes.

6.5 Conclusion missions terrain :

Ces différentes missions terrain m'ont permis de mieux comprendre les enjeux relatifs aux friches agricoles et valider des méthodologies pour les vérifier ou pour les recenser. J'ai pu y découvrir l'engagement nécessaire à la préparation de telles missions pour une bonne exécution sur le terrain. Les problèmes rencontrés sont les suivants :

- les zones en pente sont souvent inaccessibles et posent problèmes dans le calcul du NDVI
- les pâtûres peu broutées sont parfois difficilement interprétables, certains animaux ne mangeant pas toutes les plantes
- les arbres et les haies occultent certaines parcelles
- la reconnaissance des jeunes friches (3 ans) est ardue

7

Inventaire comparé sur tout le département de Tarn-et-Garonne

7.1 Inventaire complet des friches parcellaires agricoles sur le Tarn-et-Garonne :

Pour finaliser mon travail, j'ai effectué un inventaire complet (Figure 7.1) sur le département du Tarn-et-Garonne. J'ai donc utilisé des listes de communes ainsi qu'un nouveau script (`prediction_tuile.py`) permettant de fusionner les données sur chacune des 4 tuiles Sentinel (T31(TDK,TCJ,TDJ et TCK)). J'ai choisi d'utiliser le modèle `6C_RPG2019_trimestre` qui présente les meilleurs résultats et m'a permis d'obtenir plus rapidement (presque 42 heures tout de même) un inventaire complet. La taille des images, plusieurs Terra de données à traiter, est un facteur limitant pour effectuer des runs rapidement sur un territoire. J'ai eu plusieurs déconnexions intempestives du cluster qui m'ont demandé de séquencer plusieurs runs pour la tuile TCJ (la plus grosse avec 158 sur les 195 communes).

Cet inventaire est loin d'être parfait mais il présente l'intérêt de mettre à disposition un premier état des lieux qui permettra aux agents sur le terrain d'avoir une bonne approximation des zones en friches. La fiabilité des résultats reste cependant à prouver mais mes tests permettent de supposer une généricité confortable avec un rappel supérieur à 0.9. Il pourra être repris en utilisant le meilleur modèle selon mes analyses (`7C_RPG2019_mensuel`) pour obtenir des résultats plus précis.

Ainsi, 16651 friches parcellaires agricoles ont été prédites. Elles représentent une surface de 8617.4206 hectares soit 2.32% de la surface totale des 195 communes du département. Certaines revalorisations des friches (l'agriculture par exemple) excluant les parcelles trop petites, il est intéressant d'étudier la surface de ces prédictions. La surface moyenne de ces friches parcellaires agricoles prédites est de 5000 m² soit 0.5 ha ; 70% d'entre elles sont plus petites que 0.5 ha et 35% d'entre elles sont plus petites que 0.2 ha.

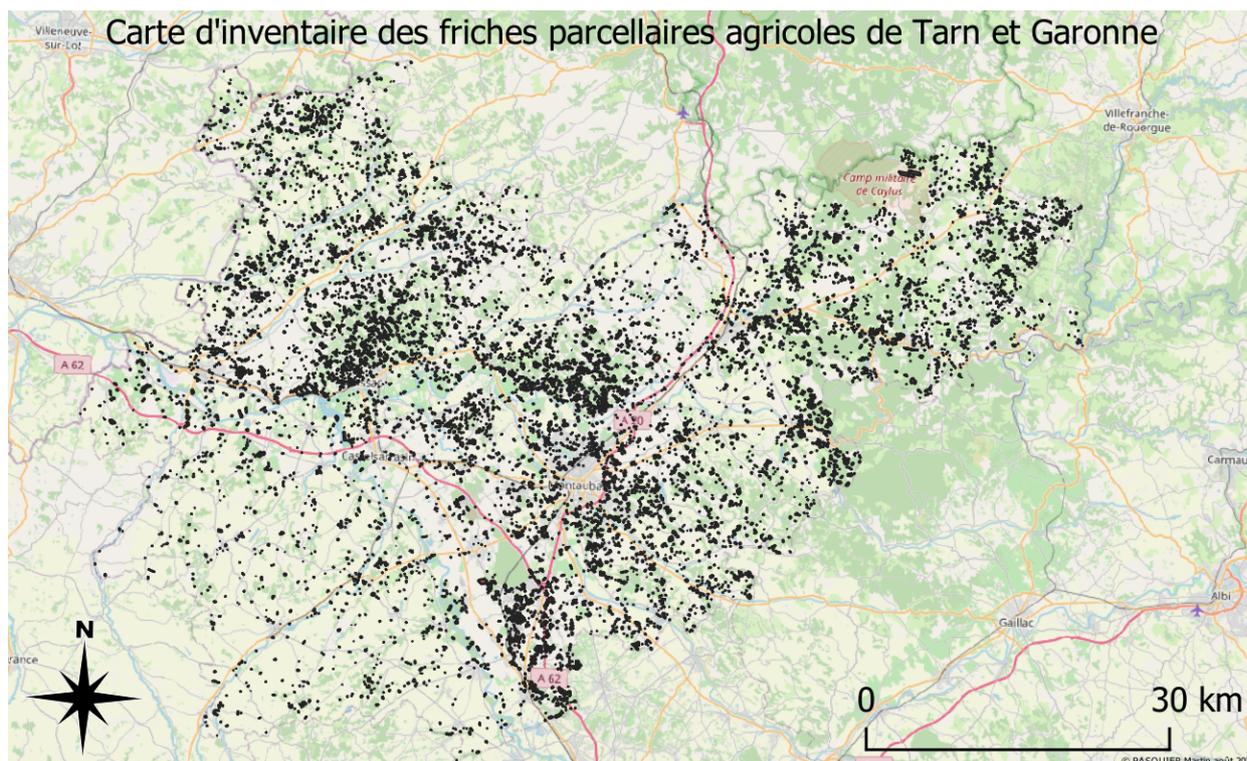


FIGURE 7.1 – Inventaire complet des friches parcellaires agricoles du Tarn-et-Garonne

Cependant l'algorithme présente l'intérêt de calculer un score de fiabilité. Les utilisateurs peuvent donc s'assurer une meilleure précision en augmentant le seuil de fiabilité, cependant cela augmente le nombre de friches "oubliées". Les prédictions avec un score supérieur ou égal à 56 représentent 5649.0354 hectares pour 1.52% du territoire.

J'ai réalisé une comparaison avec l'inventaire effectué en 2017 (Annexes D.1) ne prenant en compte que les parcelles d'une surface de plus d'un hectare. Mes résultats présentent moins de parcelles, moins de la moitié, mais ce n'est toujours pas l'estimation attendue par les différentes structures. Nous pouvons nous attendre à ce que la répartition des friches sur tout le département soit représentative (peu d'oublis, moins de 5% sur les tests) mais il reste beaucoup trop de fausses prédictions (environ 65%). La deuxième carte en annexe est une étude plus précise sur la commune de L'Honor-De-Cos (Annexe D.2) qui est intégrée en données de référence et a fait l'objet de missions terrain.

7.2 Méthodologie poursuite d'étude :

Un meilleur inventaire passe d'abord par une vérification terrain globale des friches sur les zones d'apprentissage. L'inventaire que j'ai effectué permet d'alléger ce travail. Les vérités terrain utilisées dans le modèle ne devraient pas autant prêter à confusion. Une méthode de production participative [24] est à envisager pour rendre le modèle plus robuste.

Pour améliorer la précision des modèles, c'est-à-dire enlever les faux-positifs

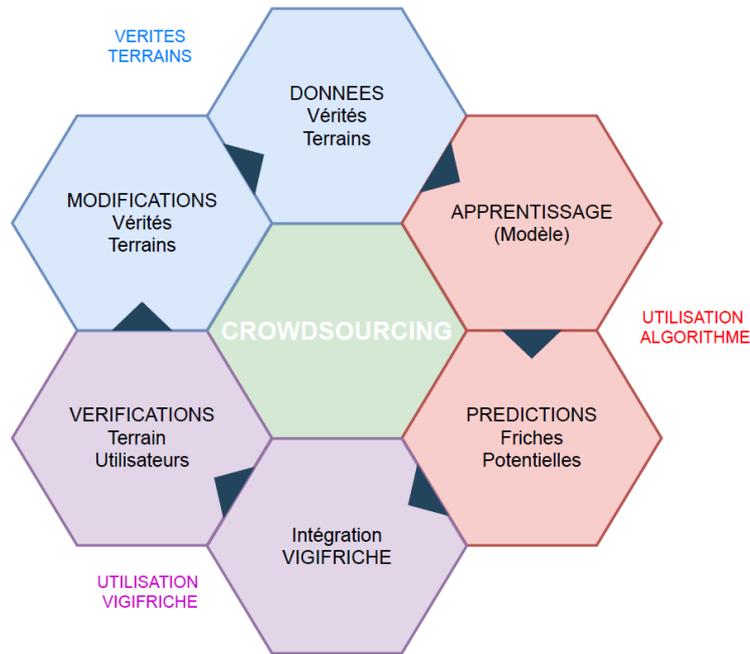


FIGURE 7.2 – Production participative du projet d’inventaire des friches agricoles du 82

(mauvaises prédictions de parcelles considérées comme friches par l’algorithme), une méthode itérative est optimale. Il s’agit d’une production participative, aussi appelée crowdsourcing. L’optimisation des vérités terrains passe par une mise à jour cyclique des vérités terrains (Figure 7.2) effectuées par plusieurs partenaires. Après avoir anticipé les résultats avec l’algorithme, des vérifications terrains sont faites pour être ensuite intégrées au modèle précédemment utilisé. Cette méthode présente l’avantage de pallier aux problèmes d’exhaustivité, de sensibilité/expérience (liée à la définition de friches utilisée lors des tags Vigifriche) et de suivi temporel.

8

Conclusion

Ce projet collaboratif m'a amené à travailler avec différentes structures ayant de multiples intérêts dans l'élaboration d'un outil d'inventaire précis et exhaustif sur un large territoire. Les perspectives d'évolution de cette étude sont nombreuses, les projets qui en découleront feront l'objet d'animations foncières auprès des élus, collectivités et propriétaires. Pour la DDT82, cette étude apporte aux différents services une nouvelle information sur le territoire et permettra de définir les zones où la procédure terres incultes pourrait être appliquée. Pour le CNES, ce projet démontre l'utilité des images satellites dans un cadre innovant. Quant à la SAFER, elle pourra utiliser les résultats de la classification afin de cibler ces terrains avec un enjeu foncier d'importance et compléter Vigifriche après vérification les parcelles en friches.

L'intérêt de l'algorithme de classification utilisé réside dans l'explicabilité des résultats représentée par le score de fiabilité (Partie 4) et les résultats de prédictions (rappel, précision et Fscore). Le rappel est représentatif de la généricité du modèle et la précision de la qualité des données. Dans ce projet, la précision est fortement réduite au détriment d'un meilleur rappel en construisant un modèle plus global du territoire. Le Fscore subit le déséquilibre entre les deux classes (friches et non-friches). Cependant l'analyse de ces performances prouve aujourd'hui que la généricité et la fiabilité des données de références, n'est aujourd'hui pas suffisante. Une collecte plus importante, variée et précise de ces vérités terrain est nécessaire. L'exhaustivité et la temporalité des données SIG est à prendre en compte dans la continuité du projet. En effet, le découpage en îlots de certaines données SIG induit des erreurs car la classification suit le cadastre parcellaire et qu'elles ne présentent pas de mises à jour à la même fréquence.

Suite à mes tests (Partie 5), il est clair que l'optimisation de l'algorithme est complémentaire aux méthodologies des études précédemment réalisées. Ainsi, le modèle doit être le plus représentatif du territoire sur lequel s'effectue les prédictions. L'ajout de données donne une information supplémentaire mais elles ne peuvent être intégrées que si elles sont suffisamment exhaustives. De plus, certaines données SIG (RPG, OCSGE) présentent des défauts. L'utilisation d'images trimestrielles présentent l'avantage d'une exécution plus rapide (42h) à l'échelle du département. Cependant, la qualité (synthèses mensuelles) et le nombre d'images utilisées augmentent les performances. Le paramétrage optimal des entrées de Random Forest reste à effectuer pour chaque nouvel apprentissage.

Il me semble important d'assurer en priorité la qualité des données. C'est dans cette optique là que j'ai développé des méthodologies avec l'ensemble des partenaires après avoir effectué plusieurs missions sur le terrain (Partie 6). De manière effective, elles permettent de renforcer les résultats ce qui permet de pallier aux limites de l'algorithme actuel. Suite à ce travail, les friches agricoles identifiées sur le Tarn-et-Garonne (Partie 7) feront l'objet d'une valorisation auprès des élus et représentants locaux.

Au travers des missions terrain que j'ai pu réaliser, la complexité est d'appliquer la définition de friche agricole de manière exhaustive et précise, particulièrement en prenant les parcelles définies par le cadastre comme unité de classification. Les problèmes de photos-interprétation et l'inaccessibilité des parcelles pour contrôle sont autant de problèmes rencontrés. De plus la sensibilité des agents terrain ne garantit pas la fiabilité des vérités terrain récoltées. L'utilisation de solutions innovantes comme les drones pourrait permettre une meilleure analyse de ces terrains.

Ce travail m'a permis d'obtenir une connaissance globale des friches agricoles au département de Tarn-et-Garonne. Le croisement d'images satellite et de données topographiques nous donne une meilleure précision sur la classification que les études précédemment réalisées (Partie 2).

Annexes

Annexe A

Déroulé de mon stage

Semaines	Agenda
07/03	Réunion partenaires du projet
15/03-27/03	Début de stage : Intégration structures Formation CNES Formation QGIS
29/03-09/04	Prise en main Cluster HPC Formation Python Formation QGIS Recherches bibliographiques Mission terrain 1 de vérification sur L'Honor-de-Cos
12/04-23/04	Prise en main Algorithme Formation Python Synthèse analyse des résultats mission terrain Mission terrain 2 de vérification sur L'Honor-de-Cos
26/04-07/05	Premiers Tests sur l'algorithme Valorisation projet (articles/vidéo) Présentation réunion direction Présentation réunion partenaires
10/05-20/05	Point stage encadrant pédagogique Optimisation algorithme de classification Préparation missions repérage LEPA Moissac
24/05-04/06	Animation et mission 1 repérage LEPA Moissac Optimisation algorithme de classification Point stage encadrant pédagogique
07/06-18/06	Mission 2 repérage LEPA Moissac Retour d'expérience Optimisation algorithme de classification Présentation réunion partenaires Présentation réunion direction
21/06-02/07	Optimisation algorithme de classification
05/07-30/07	Optimisation algorithme de classification Rapport de stage Guides méthodologiques
02/08-13/08	Point stage encadrant pédagogique Optimisation algorithme de classification Rapport de stage Guides méthodologiques
16/08-27/08	Mission drone sur L'Honor-de-Cos Optimisation algorithme de classification Rapport de stage Guides méthodologiques
30/08-10/09	Optimisation algorithme de classification Préparation présentations DDT82/CNES
13/09-24/09	Préparation soutenance Présentation projet DDT82 Présentation technique CNES
01/09	Soutenance de stage

TABLE A.1 – Les dates et travaux clés de mon stage

Annexe B

Tests d'optimisation sur l'algorithme de classification supervisée

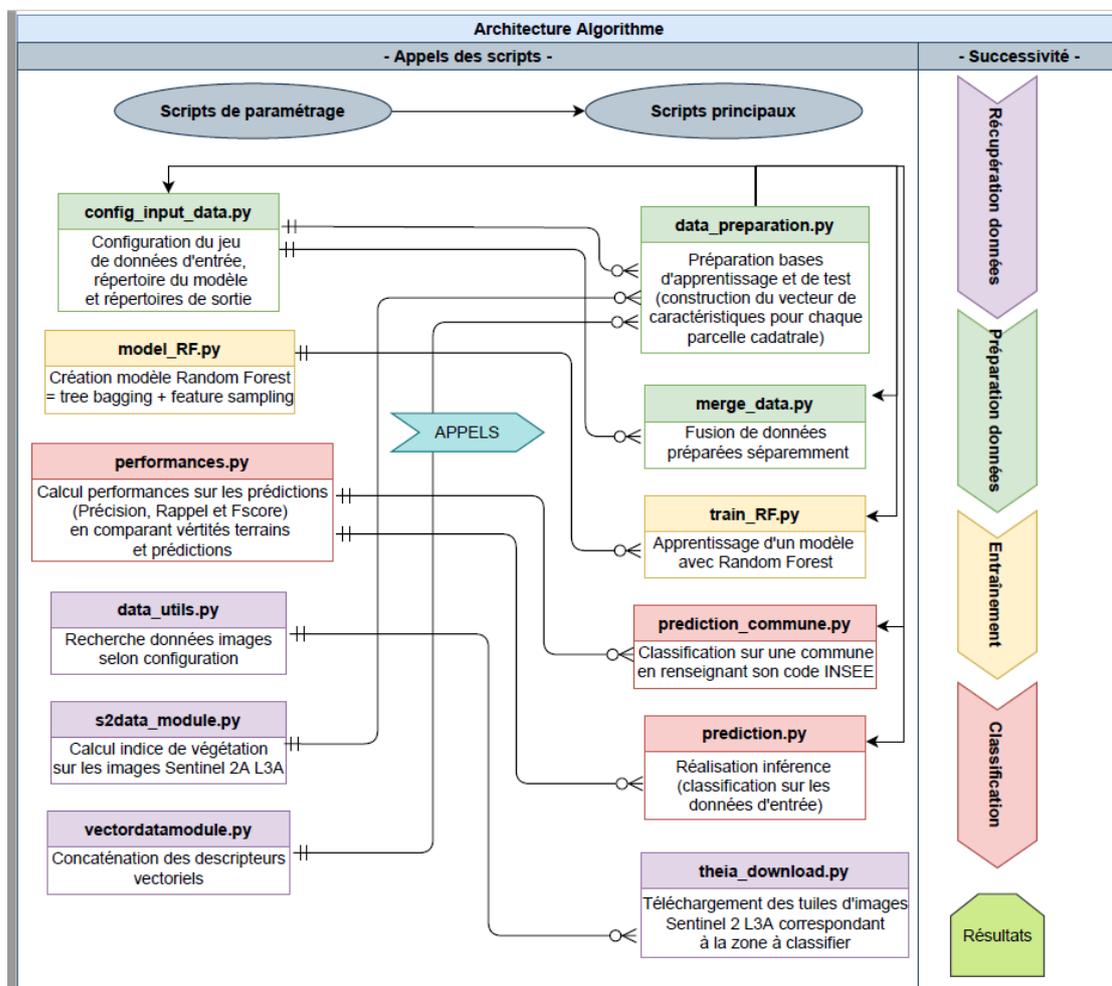


FIGURE B.1 – Architecture globale des différents scripts de l'algorithme de classification

Groupe Tests	Scripts modifiés			Analyse
Création de plusieurs modèles (mono, bi et tri communes)	merge_data.py			Meilleure généralité du modèle à 3 communes
Conclusion : Utilisation modèle 3 communes				
Ajout des données CVI et RPG2020 (82)	config_input_data.py	data_preparation.py		Impossibilité d'utiliser les données RPG2020 car non disponibles sur tous les départements Le CVI apporte une très légère amélioration des résultats
Conclusion : Exclusion du RPG2020				
Ajout Vérités terrain sur 3 nouvelles communes	merge_data.py			Meilleure généralité du modèle à 6 communes
Conclusion : Utilisation modèle 6 communes				
Changement indice NDVI en MSAVI/TSAVI	s2data_module.py			L'indice TSAVI apporte les meilleures performances
Conclusion : Utilisation de l'indice TSAVI				
Exclusion de certaines données SIG (OCSGE et RPG 2017/2019)	data_preparation.py			Meilleures performances en excluant le RPG2017, en conservant l'OCSGE
Conclusion : Exclusion du RPG2017				
Utilisation d'images trimestrielles au lieu de mensuelles	s2data_module.py			Meilleures performances en prenant en compte toutes les images mensuelles mais gain de temps conséquent avec une seule image par trimestre
Conclusion : Choix 1) rapidité : une image par trimestre 2) performances : une image par mois				
Modification des hyper-paramètres Random Forest	model_RF.py	train_RF.py		La profondeur des arbres de décision est le seul des 3 paramètres qui change significativement les performances, les autres restent tout de même à modifier selon le modèle utilisé
Conclusion : Paramètres conservés 20/50/30 au lieu de 16/30/5				
Ajout Vérités terrain sur 1 nouvelle commune	merge_data.py			Meilleure généralité du modèle à 7 communes

TABLE B.1 – Tableau récapitulatif des tests et conclusions pour l'optimisation du modèle

Annexe C

Logistique et itinéraires mission repérage sur trois nouvelles communes

Itinéraires G1 (rose) et G2 (jaune)
- Commune de Goudourville -



FIGURE C.1 – Itinéraires 2 premiers groupes de repérage sur la commune de Goudourville

Itinéraires G3 (vert) et G4 (bleu)
- Commune de Goudourville -

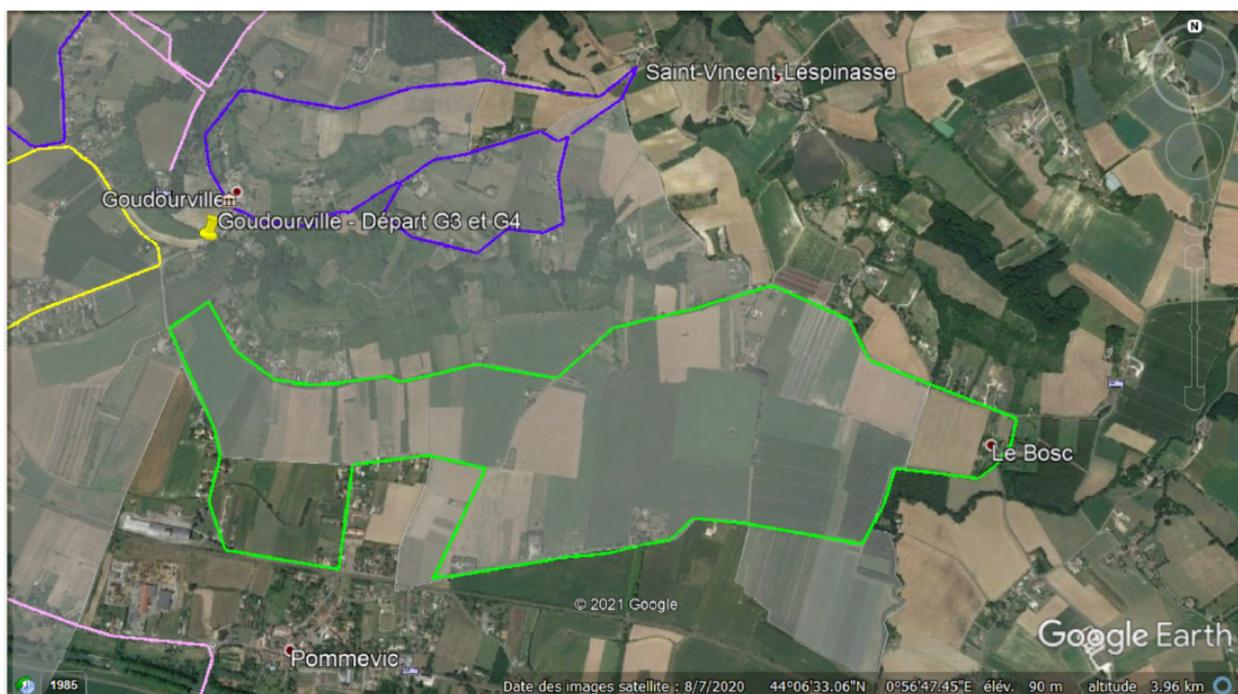


FIGURE C.2 – Itinéraires 2 derniers groupes de repérage sur la commune de Goudourville

Annexe D

Cartes d'inventaire des parcelles cadastrales agricoles en Tarn-et-Garonne

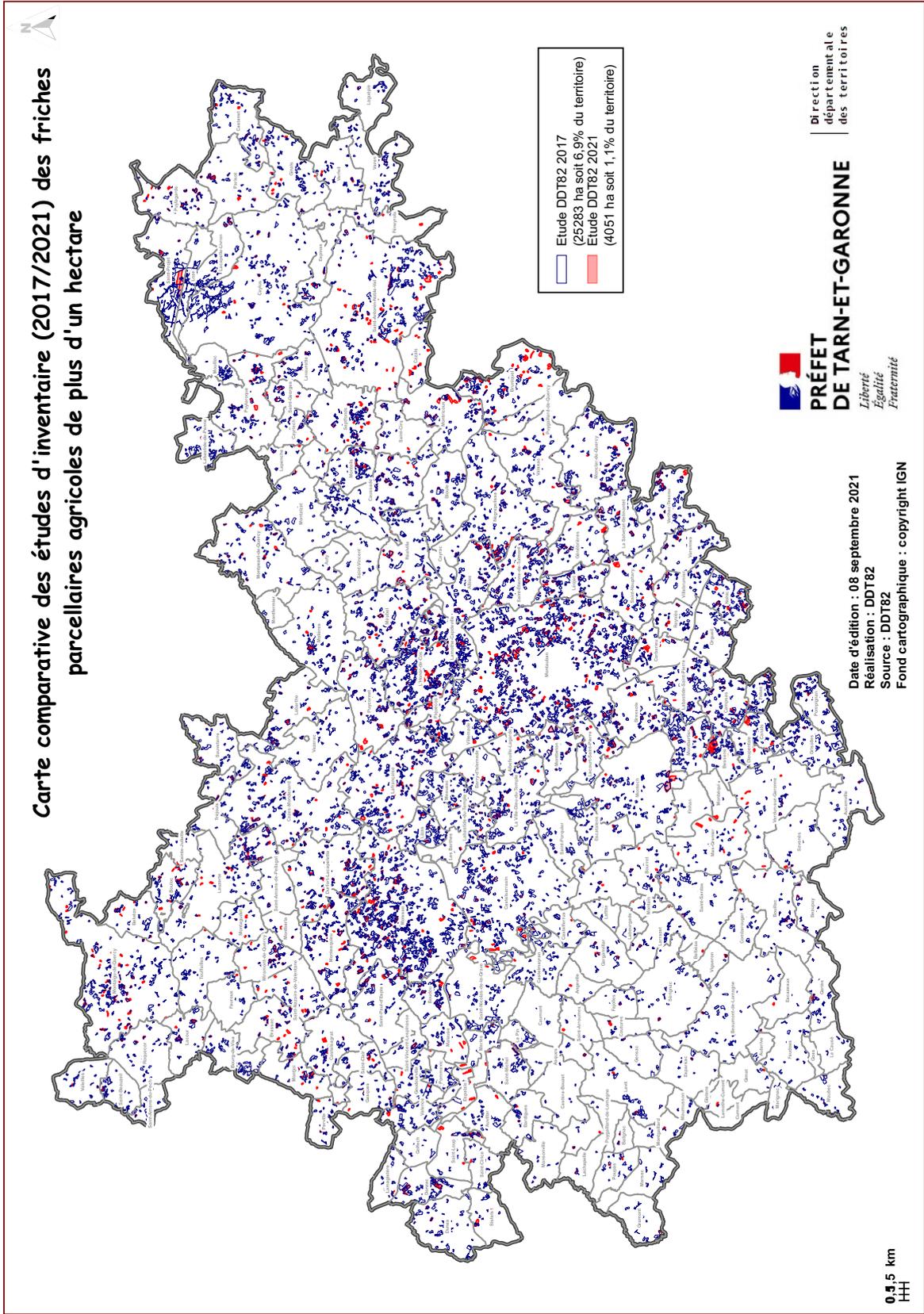


FIGURE D.1 – Comparaison inventaires du Tarn-et-Garonne avec l'étude effectuée en 2017

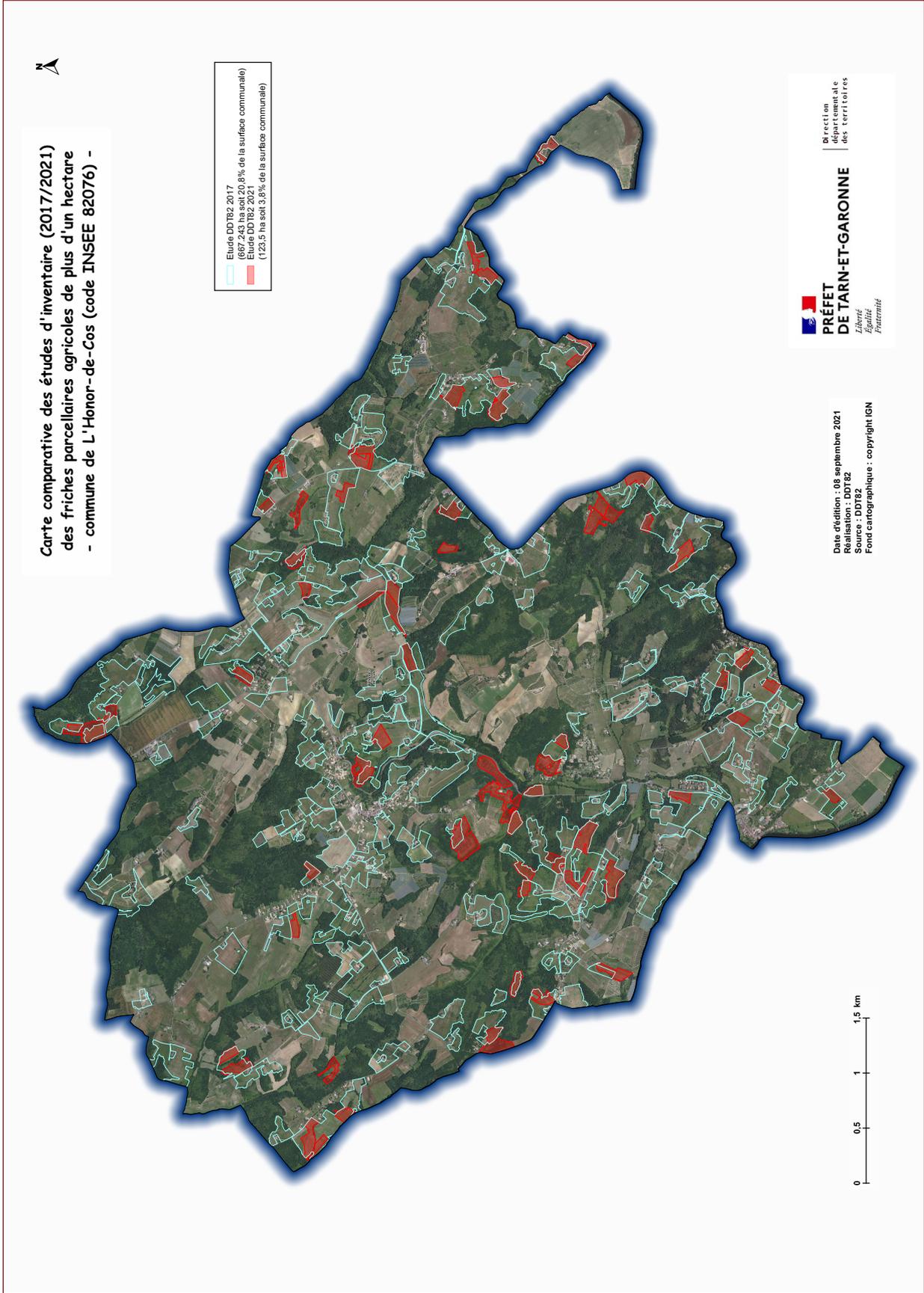


FIGURE D.2 – Comparaison inventaires sur la commune de L'Honor-De-Cos

Glossaire

- CAB : cabinet de direction de la DDT82 "en appui dans le pilotage, le management global de la structure et la communication. Depuis 2012, il coordonne au sein de la DDT les missions relatives à la gestion de crise."
- CEREMA : "Le Centre d'études et d'Expertise sur les Risques, l'Environnement, la Mobilité et l'Aménagement est un établissement public à caractère administratif placé sous la tutelle conjointe du ministre de la transition écologique et solidaire, et du ministre de la cohésion des territoires."
- Chambre de l'Agriculture : "Les Chambres d'agriculture sont des organismes consulaires (chambre consulaire). Elles ont été fondées en France par la Loi du 3 janvier 1924. Elles sont chargées de : représenter l'ensemble des différents agents économiques de l'agriculture : exploitants agricoles, mais aussi propriétaires, salariés, et organisations agricoles telles que les mutualités, coopératives, crédits et syndicats à vocation générale ; d'accompagner les exploitants agricoles dans leur développement. Elles peuvent également être saisies par les collectivités territoriales sur une problématique d'aménagement du territoire."
- Classification orientée objet : traitement des pixels par rapport à leurs voisins, leur forme et texture.
- CDPENAF : "Commission Départementale de Préservation des Espaces Naturels Agricoles et Forestiers. Cette commission instaurée par la LAAF de 2014 succède à la Commission Départementale de Consommation des Espaces Agricoles (CDCEA) et voit son rôle renforcé. Elle est présidée et animée par la DDTM34, se réunit une fois par mois et est composée de représentants de l'Etat, des collectivités territoriales ainsi que de la profession agricole, des propriétaires fonciers, des chasseurs, des notaires, d'associations agréées pour la protection de l'environnement, de l'INAO, de la SAFER et de l'ONF."
- CVRH : Le Centre de Valorisation des Ressources Humaines est un service de formation et de ressources humaines à compétence nationale rattaché à la Direction des Ressources Humaines du Ministère de la Transition Ecologique et Solidaire et du Ministère de la Cohésion des Territoires et des Relations avec les Collectivités Territoriales.
- Défrichement : "nettoyage provisoire ou intensif d'une friche amenant, ou non, une mise en culture ou en parc surveillé, par exemple la coupe des ronces et broussailles, voire de la strate arborée, s'appelle le défrichage. Le défrichement désignait et désigne encore l'activité de suppression de l'état de forêt. Par analogie, on parle de friche urbaine (propriété laissée à l'abandon) ou de friche industrielle (espace industriel désaffecté)."
- DINUM : "La direction interministérielle du numérique (Dinum) est une direction de l'administration publique française, chargée de coordonner les actions des administrations en matière de systèmes d'information. Elle est considérée comme la direction des systèmes d'information de l'Etat français."
- DRAAF : "En France, les directions régionales de l'Alimentation, de l'Agriculture et de la Forêt (DRAAF ou DAAF dans les départements et régions d'outremer) sont des services déconcentrés du ministère de l'Agriculture et de l'Alimentation qui ont pour mission de mettre en œuvre au niveau de la région la politique nationale en matière d'alimentation, d'agriculture et de forêt."
- DGFIP : "La Direction Générale des Finances Publiques gère la fiscalité de l'État français et des collectivités territoriales qui s'impose aux particuliers et aux entreprises, en définit les règles juridiques, en assure le recouvrement au moyen de son réseau territorial

et en instruit les contentieux fiscaux." – programme EIG : "Le programme Entrepreneurs d'Intérêt Général a pour objectif de faire travailler ensemble des personnes extérieures à l'administration aux compétences numériques pointues et des agents publics engagés dans une démarche d'innovation. Les entrepreneurs d'intérêt général sont répartis en binômes ou trinômes pluridisciplinaires. Avec leurs mentors, ils ont 10 mois pour relever un défi numérique autour des données. Le programme, fort de 3 promotions, comporte 37 défis et une communauté de plus de 100 personnes. Le programme EIG est piloté par la mission Etalab au sein de la DINUM."

– Enfrichement : "passage progressif à l'état de friche selon le principe de succession végétale. Selon l'âge de la friche et son degré de recouvrement, sa hauteur et sa composition floristique changent, ce qui engage les écologues à distinguer, selon le type de végétation qui s'y développe, la friche herbacée ou herbeuse qui peut être entretenue à l'état ouvert ou semi-ouvert et peut être pâturée, la friche arbustive (stade de friche armée où se développent en masse des épineux ou selon, le taux d'embroussaillage par les fourrés, stade de lande ou de fruticée) et la friche arborée (reboisée naturellement ou artificiellement)."

- Jachère : "historiquement, l'ensemble des pratiques culturelles de printemps et d'été des terres arables préparant l'ensemencement d'une céréale d'automne. Le terme désigne aussi, par métonymie, la période où l'on effectue ces façons, et la terre qui les reçoit. Cette préparation consiste en plusieurs labours dont le but est de détruire les adventices (mauvaises herbes), ensevelir la fumure et accélérer la décomposition de la matière organique. [...] La jachère a souvent été appelée repos de la terre, terme doublement ambigu : parce qu'il s'appliquait aussi à la friche enherbée, et parce que cette terre au repos était intensément travaillée."

– Ligneux : "Se dit d'une plante contenant suffisamment de faisceaux lignifiés pour que ses tiges soient résistantes." – MAAF : "Le Ministère de l'agriculture, dénommé depuis 2012 Ministère de l'Agriculture, de l'Agroalimentaire et de la Forêt, est chargé de mener la politique agricole, forestière et alimentaire du pays."

– MAJIC : Mise À Jour des Informations Cadastreales, sont des fichiers fonciers issus des déclarations faite aux impôts et créés par la Direction Générale des Finances."

– Orthophotos : images aériennes ou satellitaires de la surface terrestre corrigées géométriquement et égalisées radiométriquement. Ces images se présentant sous forme de tuiles et servent de fonds cartographiques dans les Systèmes d'Information Géographique.

– PAC : "Politique Agricole Commune menée à l'échelle de l'Union européenne. À l'origine, elle est fondée principalement sur des mesures de contrôle des prix et de subventionnement, visant à moderniser et développer l'agriculture."

– PLU(i) : Plan Local d'Urbanisme (intercommunal)

– QGIS : est un logiciel de SIG libre.

– RPG : "Le Registre Parcellaire Graphique est un SIG permettant l'identification des parcelles agricoles, il est construit sur la base des déclarations faites à la PAC par les agriculteurs."

– SAT : Service d'Aménagement Territorial de la DDT82 "regroupant les missions liées à l'application du droit des sols (instruction, animation et expertise vis-à-vis des centres instructeurs des collectivités), la fiscalité de l'urbanisme, la planification urbaine (conseil et association de l'Etat) et le conseil global en aménagement (appui technique de proximité aux collectivités territoriales, communautés de communes prioritairement, le plus en amont possible, sur des projets complexes ou à forts enjeux)."

-
- SCR : Service Connaissance et Risques de la DDT82 "en charge de la mise en oeuvre des politiques publiques dans les domaines des énergies, des déplacements, de la prévention des risques, de l'éducation routière et de la coordination interministérielle en matière de sécurité routière. Il assure également, en lien avec les autres services, le pilotage des démarches en matière d'études générales, de suivi des grands projets et de connaissance des territoires à travers l'animation du SIG."
 - SDIS : Service Départemental d'Incendie et de Secours
 - SEA : Service Economie Agricole de la DDT82 "chargé de la mise en oeuvre de la politique agricole commune, des politiques agricoles nationales et du programme de développement rural. Il est le guichet unique pour les aides destinées aux exploitations agricoles (aides PAC et conjoncturelles) et service instructeur pour certaines aides entrant dans le cadre du document régional de développement rural, piloté par le Conseil régional. Il est également en charge du contrôle des structures agricoles et formes juridiques spécifiques (GAEC), de la coordination des contrôles liés à la PAC, ainsi que de la gestion des indemnisations au titre du Fonds national de garantie des calamités agricoles."
 - SEB : Service Eau et Biodiversité de la DDT82 qui s'occupe de la "mise en oeuvre de la politique de la police de l'eau et des milieux aquatiques, de l'animation de la mission inter-services de l'eau et de la nature (MISEN), de la biodiversité (chasse, pêche, forêt, milieux naturels), de la police de la navigation, de la gestion du domaine public fluvial (DPF) et de la gestion qualitative de l'eau."
 - Sentinel 2 : est une constellation de satellites (Sentinel 2A et 2B) d'observation de la Terre de l'Agence spatiale européenne. Ils fournissent des images multi-spectrales à grande résolution disponible gratuitement sur le site du Centre National d'Etudes Spatiales (CNES).
 - Service Déconcentré : Les services déconcentrés de l'Etat français sont les services qui assurent le relais, sur le plan local, des décisions prises par l'administration centrale et qui gèrent les services de l'Etat au niveau local. Ces services déconcentrés sont habituellement de niveau régional (avec notamment les DRFiP, ARS ou Rectorats) ou départemental (avec notamment les directions départementales interministérielles), mais il existe également pour certains ministères ou certains services un niveau zonal, interrégional ou interdépartemental.
 - SH : Service Habitat de la DDT82 "en charge des études et de la définition des politiques dans le domaine de l'habitat, du renouvellement urbain et du financement du logement, des politiques concernant le logement des personnes défavorisées, notamment par l'application au droit au logement opposable, de l'urbanisme, de l'aménagement opérationnel et foncier et du contentieux relevant de l'ensemble des activités de la DDT, et enfin de l'animation de la mission inter-services pour le logement."
 - SIG : Un Système d'Information Géographique est un système d'information conçu pour recueillir, stocker, traiter, analyser, gérer et présenter tous les types de données spatiales et géographiques.
 - Télédétection : mesure ou acquisition d'informations sur un objet ou un phénomène, par l'intermédiaire d'un instrument n'ayant pas de contact avec l'objet étudié.
 - Terres incultes : "procédure du Code rural, relativement lourde à mettre en oeuvre, qui vise à remettre en culture des parcelles agricoles non exploitées depuis plus de trois ans."

Bibliographie

[1] Louis CRETIN - Rapport de stage de Master 2 I.G.P.E. - septembre 2019 - Hérault - état des lieux et éléments prospectifs sur les friches agricoles

[2] Connect by CNES - Ambassadeurs - Formation Spatial - 2020-2021 - URL : www.youtube.com/watch?v=ormkUCTX3vQ&list=PLlooRuVXa-leI1zXPWFDLQ41VYhH5eLv5.

[3] CDPENAF Lozère - Présentation aux DDTs - 16 mai 2019 - Inventaire des friches agricoles en Lozère

[4] DDTM Landes - Présentation DRAAF Nouvelle-Aquitaine - 12 avril 2017 - Mise au point d'une méthodologie pour l'inventaire des terres considérées comme friches

[5] DDT31 - Présentation réunion nationale des correspondants foncier du MAAF - 16 novembre 2016 - Détection des friches agricoles

[6] DDTM Calvados - Présentation Journée nationale des correspondants fonciers - 07 novembre 2018 - Inventaire des potentielles friches agricoles du Calvados

[7] Annik Schnitzler, Jean-Claude Génot - La France des friches. De la ruralité à la féralité - Editions Quae - 2013 - p. 88-110

[8] DDT82 - Formation QGIS - 2020-2021

[9] CESBIO - Séries Temporelles - Les synthèses mensuelles de niveau 3A de Theia pour Sentinel-2 - URL : <https://labo.obs-mip.fr/multitemp/les-syntheses-mensuelles-de-niveau-3a-de-theia/>

[10] Données cadastrales ouvertes - URL : <https://cadastre.data.gouv.fr/>

[11] Informations et téléchargement RPG national - URL : <https://www.data.gouv.fr/fr/datasets/registre-parcellaire-graphique-rpg-contours-des-parcelles-et-ilots-cultureaux-et-leur-groupe-de-cultures-majoritaire/>

[12] Portail de l'artificialisation des sols - Informations générales - BD TOPO - URL : <https://artificialisation.biodiversitetousvivants.fr/bases-donnees/bd-topo>

[13] IGN (Institut National de l'Information Géographique et forestière) - BD Forêt version 2 - URL : <https://inventaire-forestier.ign.fr/spip.php?article646>

[14] Portail de l'artificialisation des sols - Informations générales - OCSGE - URL : <https://artificialisation.biodiversitetousvivants.fr/bases-donnees/ocs-ge>

[16] Douanes et droits indirects - Démarche : Immatriculer votre entreprise vitivinicole au CVI - URL :
<https://www.douane.gouv.fr/demarche/immatriculer-votre-entreprise-vitivinicole-au-cvi>

[16] CESBIO - Séries Temporelles - Les tuiles de Sentinel-2, comment ça marche? - URL :
<https://labo.obs-mip.fr/multitemp/les-tuiles-de-sentinel-2-comment-ca-marche/>

[17] ArcGIS - Galeries d'indices - URL :
<https://pro.arcgis.com/fr/pro-app/latest/help/data/imagery/indices-gallery.htm>

[18] IGN - Géoservices - téléchargement de données - URL :
<https://geoservices.ign.fr/telechargement>

[19] Theia - Sentinel-2 N2A et N3A - URL :
theia.cnes.fr/atdistrib/rocket/#/search?page=1&collection=SENTINEL2&processingLevel=LEVEL3A.

[20] Gitlab - détails de l'utilisation - URL :
https://github.com/olivierhagolle/theia_download pour les détails de l'utilisation

[21] ArcGISPro - Galeries d'indices - URL :
<https://pro.arcgis.com/fr/pro-app/latest/help/data/imagery/indices-gallery.htm>

[22] ArcGISPro - MSAVI - URL :
<https://pro.arcgis.com/fr/pro-app/latest/arcpy/image-analyst/msavi.htm>

[23] PLU immo.fr - Meauzac (82) : PLU, cadastre et risques naturels - URL :
<https://plu-immo.fr/meauzac-82290/>

[24] 360-webmarketing - Le crowdsourcing : définition, enjeux, typologie - URL :
<https://www.360-webmarketing.fr/crowdsourcing/>

[25] Python - Machine Learning Tutorial - URL :
https://python-course.eu/machine_learning.php

[26] Marc Lang - Cours de Python pour la Télédétection - Approfondissement - SIGMA - 2020-2021 - URL :
<https://mlang.frama.io/sigma-teledec-avancee/>.