



**PRÉFET  
DE TARN-ET-GARONNE**

*Liberté  
Égalité  
Fraternité*

**GUIDE MÉTHODOLOGIQUE :  
L'algorithme de classification supervisée  
Random Forest utilisant la télédétection  
au service de l'inventaire des friches agricoles**



**Mots clés : Fiches agricoles, télédétection, algorithme de classification**

**À qui s'adresse ce document ?**

Ce guide présente la méthode du projet, vulgarise son développement et présente nos conclusions. Il complète le guide technique qui détaille l'algorithme. Il se compose d'annexes et de 5 parties :

- 1) Principe de l'algorithme**
- 2) Données d'entrée**
- 3) Optimisation**

## Préliminaire au projet :

La loi pour l'Avenir de l'Agriculture et de la Forêt (LAAF) de 2014 charge, tous les cinq ans, 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 (article L.112-1-1 du code rural et de la pêche maritime).

Plusieurs initiatives d'inventaires ont été menées à l'échelle du territoire. 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 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.

En Occitanie, en 2020, la Chambre de l'agriculture et les Sociétés d'aménagement foncier et d'établissement rural (SAFER) du Tarn-et-Garonne ont candidaté à un appel à projet « reconquête du foncier agricole sur la commune de L'Honor-de-Cos ». Il se déroule en deux phases : une animation foncière conduite par la SAFER et une animation agricole menée par la Chambre de l'agriculture.

C'est à l'automne dernier que la Direction départementale des territoires du Tarn-et-Garonne (DDT82), le Laboratoire d'Observation de la Terre du Centre National d'Études Spatiales (Lab'OT/CNES) et la SAFER ont croisé leurs efforts reprenant une méthodologie utilisée en 2019 par la DDT de l'Hérault. L'objectif visé par le consortium consiste à développer un algorithme Python, utilisant des images satellites – on parle alors de télédétection – pour « automatiser » l'inventaire et étendre l'expérience à toute la région après validation.

Grâce à plusieurs repérages sur le terrain, Vigifriche, l'application mobile de la SAFER qui géolocalise les friches, permet de construire une base d'apprentissage pour créer un modèle utilisé par l'algorithme. Le Lab'OT a créé cette Intelligence artificielle (I.A.) qui utilise plusieurs types de données. Cependant, il n'est pas simple de définir ce qu'est une friche agricole. La sensibilité et l'expérience de chacun jouent beaucoup dans l'interprétation. Ces zones en déprise mal-méconnues peuvent être revalorisées sous de nombreuses formes : réhabilitation à l'agriculture, prévention incendie, spéculation foncière, création de zones naturelles, etc.

## Définition friches :

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. »



Fig.1 : Terrain non entretenu depuis une quinzaine d'années

## Apport de la Télédétection :



Fig.2 : Orthophoto sur L'Honor-de-Cos

La diversité des friches agricoles tant par l'âge que par la présence de végétaux typiques à chaque territoire nécessite une étude terrain fastidieuse.

Les images satellites apportent une source d'informations supplémentaire grâce aux 13 longueurs d'ondes de Sentinel 2-L3A. (<https://labo.obs-mip.fr/multitemp/les-syntheses-mensuelles-de-niveau-3a-de-theia/>) Elle permet de calculer les indices de réflectance des végétaux permettant leur caractérisation.

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 palier à ces imprécisions, on couple les images à des données terrains issues de couches SIG (Système d'information géographique).

## L'application mobile Vigifriche comme source de vérités terrains (Guide d'utilisation en annexe)

Il est possible de déclarer et de confirmer qu'une parcelle cadastrale présente une portion importante en friches. Les utilisateurs peuvent également modifier leurs choix.

On peut renseigner 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.

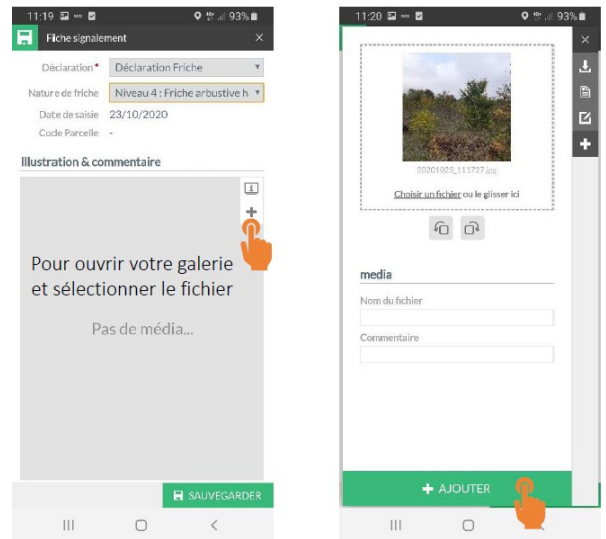


Fig. 3 : Interface de déclaration Vigifriche

### Utilisation d'un Algorithme de classification Random Forest (RF) :

<https://www.ibm.com/cloud/learn/random-forest>

L'algorithme de classification fonctionne par apprentissage supervisé, autrement dit de manière semi-automatique. L'algorithme construit une fonction de prédiction – une boîte noire – à partir d'exemples annotés. Dans notre cas il s'agit d'apprendre des vérités terrains pour construire un modèle de classification en croisant des données. Les données d'entrée utilisées sont les parcelles cadastrales vues sur les images satellites et combinées à d'autres descripteurs. Nous obtenons ainsi en sortie les prédictions sur de nouvelles données réparties en deux classes : friche et non-friche.

La méthode consiste à calculer différentes caractéristiques pour chaque parcelle cadastrale, apprendre au modèle les caractéristiques des vérités terrains puis classer les parcelles selon ces mêmes caractéristiques. Les vérités terrain fournies consistent en un fichier regroupant les parcelles cadastrales en friche pour différentes communes. Les autres parcelles cadastrales sont alors toutes considérées comme n'étant pas en friche.

Afin d'optimiser l'algorithme, nous pouvons consolider le modèle, modifier les jeux de données et les paramètres de classification. Une analyse des résultats de tests d'optimisation (cf Annexes) a été effectuée.

## I) Principe de l'algorithme :

**Objectif :** combiner les données disponibles pour permettre une classification suivant le cadastre parcellaire (parcelle en friches ou non-friche). L'optimisation doit permettre de réaliser une classification répétable à l'ensemble du territoire.

**Choix de Random Forest :** La demande d'inventorier au mieux les parcelles en friches est d'importance pour une meilleure re-valorisation de ces terrains en déprise. Vigifriche nous fournit des vérités terrains sur certaines communes. Ces vérités terrains connues permettent grâce à l'algorithme de classification d'effectuer un apprentissage semi-automatique (supervisé).

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 en réduisant leur variance.

Les méthodes actuelles reposant sur des couches SIG en dur, l'apprentissage supervisé de la méthode proposée permet une meilleure utilisation des caractéristiques des données.

**Prédictions :**

Lorsqu'on utilise l'algorithme sur des zones apprises par le modèle, on obtient en sortie les résultats correspondant aux prédictions sur la base d'apprentissage (80 % des parcelles) et la base de test (20 %, Fig.4). La base d'apprentissage de l'algorithme donne en sortie un indice de fiabilité global, le FScore (calculé à partir du rappel et de la précision des prédictions), des résultats suivants :

True <b>positive</b>	TP	parcelle bien prédite en tant que friche
True <b>negative</b>	TN	parcelle bien prédite en tant que <b>non</b> -friche
False positive	FP	parcelle <b>mal</b> prédite en tant que friche (normalement non-friche selon les VT)
False negative	FN	parcelle <b>mal</b> prédite en tant que <b>non</b> -friche (normalement friche selon les VT)

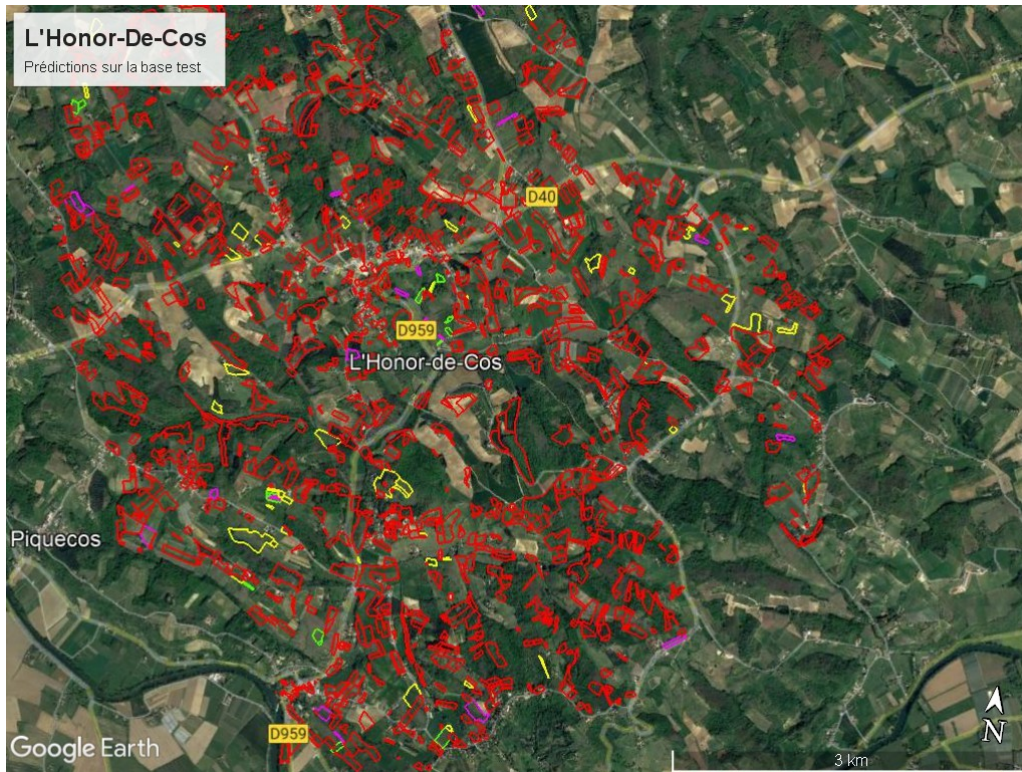


Fig. 4 : prédictions sur base test - L'Honor-De-Cos (TP : vert ; FP : jaune ; FN : rose ; TN : rouge)

Pour chaque parcelle cadastrale, l'algorithme effectue une prédiction en respectant un seuil de décision correspondant à la fiabilité de la classification. Ce score est compris entre 0 et 100 (Fig.5). Le seuil de décision est fixé à 50, au-dessus l'algorithme classe la parcelle en friche.

Toutes les prédictions sont ensuite comparées aux vérités terrains pour calculer trois indicateurs prenant des valeurs entre 0 et 1 :

- **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**

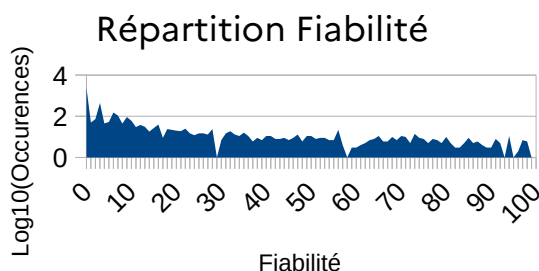


Fig. 5 : Visualisation du nombre de parcelles et leur fiabilité de classification

Remarque : Nous souhaitons assurer un modèle le plus générique possible qui puisse reconnaître des parcelles en friches sur différentes topologies, le rappel est donc à privilégier. Le modèle peut devenir beaucoup trop influencé par les données d'entraînements, ce qui engendre un biais. 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 faites des données d'entraînement, ce qui est à notre avantage.

### 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 **généricité** 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
- le **suivi temporel** 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

## II) Données d'entrées :

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

\*Vérités terrains (parcelles cadastrales en friche) sur les communes suivantes :

- Nyls Ponteilla (66): 359 parcelles en friche (sur 3535 parcelles cadastrales dans la commune)
- Figeac (46): 406 parcelles en friche (sur 12 026 parcelles cadastrales dans la commune)
- L'Honor de Cos (82): 244 parcelles en friche (sur 6669 parcelles cadastrales dans la commune)
- Coutures (82) : 2 parcelles en friche (sur 268 parcelles cadastrales dans la commune)
- Le Pin (82) : 6 parcelles en friche (sur 635 parcelles cadastrales dans la commune)
- Goudourville (82) : 36 parcelles en friche (sur 2450 parcelles cadastrales dans la commune)
- Meuzac (82) : 94 parcelles en friche (sur 3125 parcelles cadastrales dans la commune)

## Détail des données :

- **Images satellites** (Sentinel2 L3A) : la constellation Sentinel fournit des observations fréquentes et répétitives. En plus des corrections géométriques, les produits de niveau 3A contiennent des images, 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.
- **Cadastre** : plan des propriétés foncières servant de base à la classification qui s'effectue sur chaque parcelle cadastrale. <https://cadastre.data.gouv.fr/>
- **RPG** (registre parcellaire graphique) : référence à l'instruction des aides de la politique agricole commune (PAC), permettant d'éliminer les îlots qui y sont déclarés de la classification. <https://www.data.gouv.fr/fr/datasets/registre-parcellaire-graphique-rpg-contours-des-parcelles-et-ilots-cultureaux-et-leur-groupe-de-cultures-majoritaire/>
- **BD\_Topo** : 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 : <https://artificialisation.biodiversitetousvivants.fr/bases-donnees/bd-topo>
- **BD\_Forêt** (inventaire forestier) : élimine les îlots de forêt de plus de 5000m<sup>2</sup> de la classification.
- **OCSGE** (occupation du sol à grande échelle) : 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. <https://artificialisation.biodiversitetousvivants.fr/bases-donnees/ocs-ge>
- **CVI** (casier viticole informatisé) : permet d'éliminer les parcelles cadastrales de vignes déclarées à la MSA (Mutualité Sociale Agricole) de la classification.

*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 (cf Annexes), 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 MSA*

*Ajout données GéoMCE (mesures compensatoires environnementales), CEN (conservatoire d'espaces naturels), APB (aire de protection des biotopes)*

## III) Optimisation :

L'algorithme a tourné avec le cluster HPC du CNES sur un environnement conda/4.9.0 en langage Python. La méthode utilisée sur les communes possédant des vérités terrain est la suivante :

### 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, les caractéristiques suivantes sont mises en entrée du modèle - boîte noire :

- le pourcentage d'intersection avec chaque BD vectorielle (vecteur de 14 composantes, car on sépare les différentes classes de la BD TOPO zone\_vegetation.shp)

- la moyenne, la médiane et l'écart-type d'un indice de réflectance des végétaux (NDVI, MSAVI ou TSAVI) pour chaque image satellite de Sentinel2 L3A

### 4) Entraînement d'un modèle Random Forest sur base train

### 5) Analyse des résultats sur base test

Une méthode de crowdsourcing est à envisager pour rendre le modèle plus robuste. Pour améliorer la précision des modèles, c'est-à-dire enlever les faux-positifs (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** (Fig.6) 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 palier aux problèmes d'**exhaustivité**, de **sensibilité** des tagueurs et de **temporalité**.

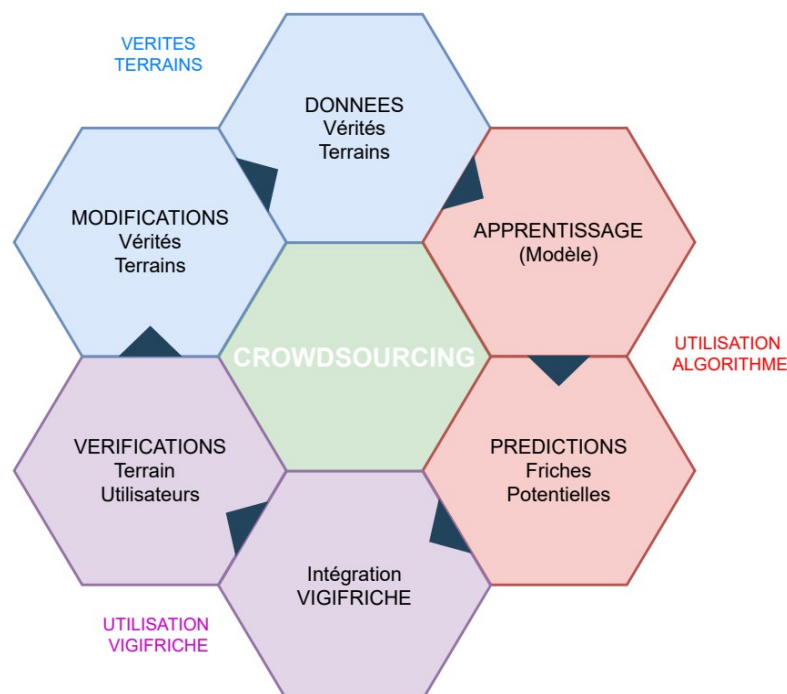


Fig.6 : Production participative/Crowdsourcing du projet d'inventaire des friches agricoles du 82

L'importance de cette production participative est prépondérante au bon déroulement du projet et sera profitable à l'ensemble des utilisateurs.

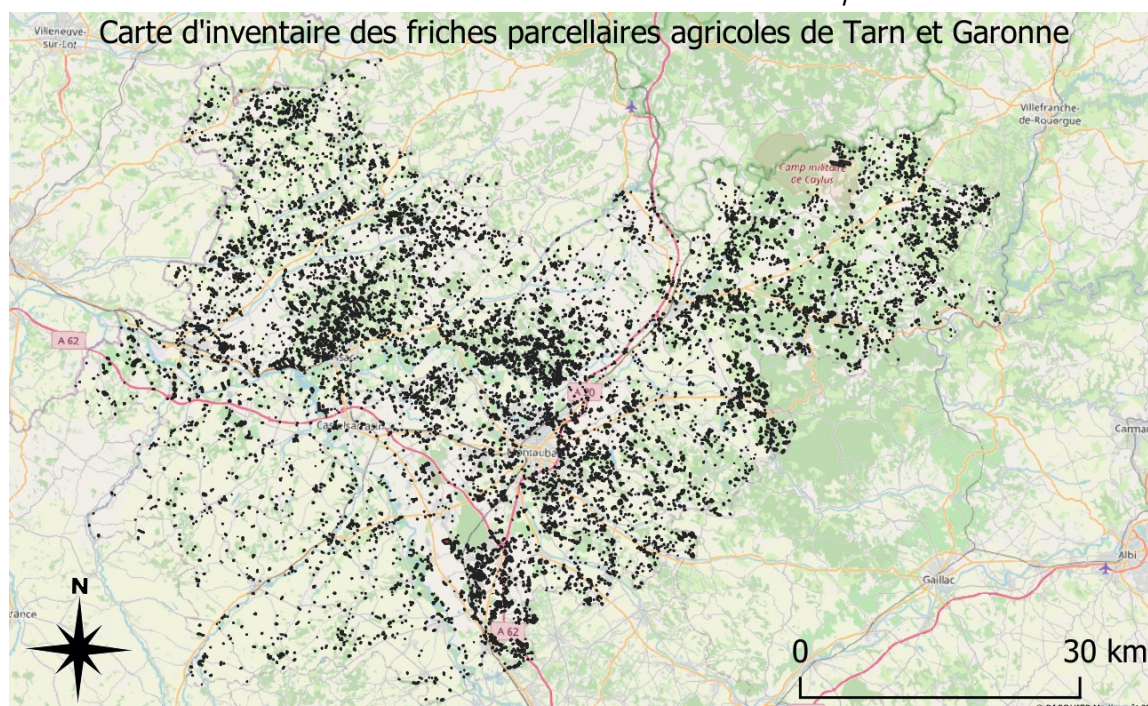
Suite aux tests d'optimisation effectués sur cet algorithme, un inventaire (Annexes) a été effectué. Il reste perfectible et présente un état des lieux pour de futures applications.

Le département présente des images satellites sur 4 tuiles Sentinel 2 (T31TDK, T31TCJ, T31TDJ et T31TCK). Le site *labo.obs-mip.fr* fournit les informations utiles sur la dénomination des tuiles. Le temps d'exécution du code a été de 42h et a nécessité que les communes de la tuile TCJ soient séparées en plusieurs sous-listes.

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 5000m<sup>2</sup> soit 0,5 ha ; 70% d'entres-elles sont plus petites que 0,5 ha et 35% d'entres-elles sont plus petites que 0,2 ha.

## ANNEXES

Annexe 1 : Visualisation des résultats de l'inventaire sur l'ensemble du département



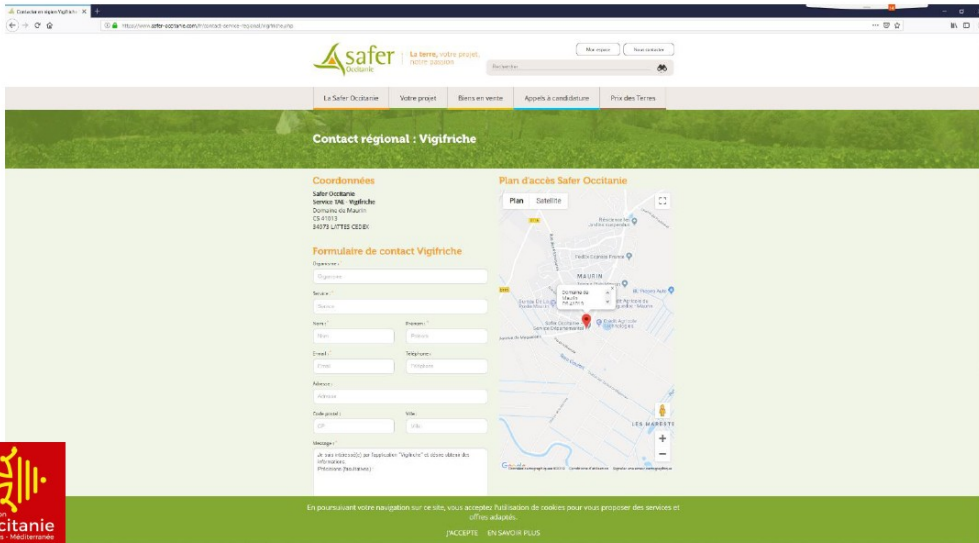
Annexe 2 : Tableau récapitulatif des tests d'optimisation avec l'algorithme de classification :

<u>Groupe Tests</u>	<u>Analyse</u>
Création de plusieurs modèles (mono, bi et tri communes)	Meilleure généricité du modèle à 3 communes
<b>Conclusion : Utilisation modèle 3 communes</b>	
Ajout des données CVI et RPG2020 (82)	Impossibilité d'utiliser les données RPG2020 car non disponibles sur tous les départements Le CVI apporte une très légère amélioration des résultats
<b>Conclusion : Exclusion du RPG2020</b>	
Ajout VÉRITÉS terrain sur 3 nouvelles communes	Meilleure généricité du modèle à 6 communes
<b>Conclusion : Utilisation modèle 6 communes</b>	
Changement indice NDVI en MSAVI/TSAVI	L'indice TSAVI apporte les meilleures performances
<b>Conclusion : Utilisation de l'indice TSAVI</b>	
Exclusion de certaines données SIG (OCSGE et RPG 2017/2019)	Meilleures performances en excluant le RPG2017, en conservant l'OCSGE
<b>Conclusion : Exclusion du RPG2017</b>	
Utilisation d'images trimestrielles au lieu de mensuelles	Meilleures performances en prenant en compte toutes les images mensuelles mais gain de temps conséquent avec une seule image par trimestre
<b>Conclusion : Choix 1) rapidité : une image par trimestre 2) performances : une image par mois</b>	
Modification des hyper-paramètres Random Forest	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
<b>Conclusion : Paramètres conservés 20/50/30 au lieu de 16/30/5</b>	
Ajout VÉRITÉS terrain sur 1 nouvelle	Meilleure généricité du modèle à 7 communes



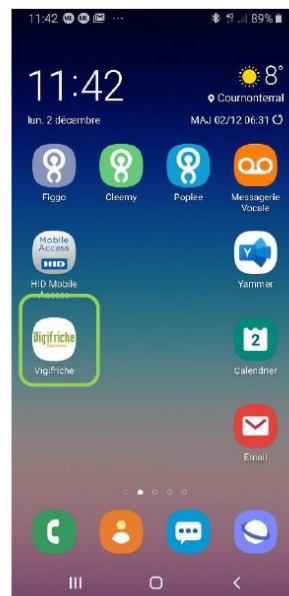
## Remplir la fiche contact

<https://www.safer-occitanie.com/fr/contact-service-regional/vigifriche.php>





## Télécharger l'application

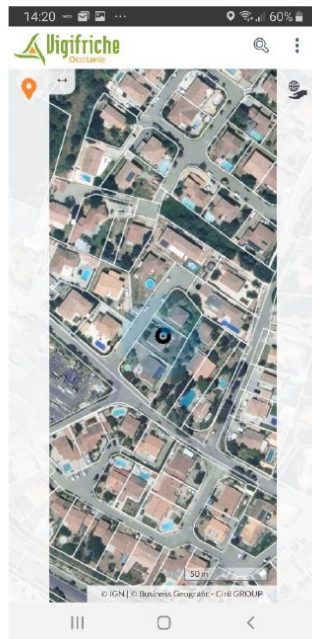
Télécharger et installer l'application Vigifriche avec les accès communiqués



## Ouvrir l'application


 Nécessité de cliquer sur ce pictogramme pour lancer la saisie de friche sur le terrain

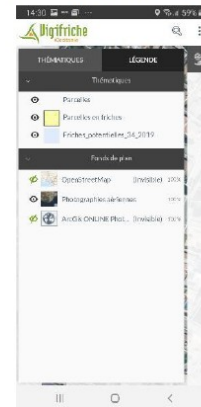
 L'application s'ouvre en se localisant automatiquement sur le lieu de situation du téléphone




## Vigifriche Occitanie

 Permet de rechercher un lieu à partir d'un nom de commune ou d'une adresse

 Permet d'afficher la légende des couches présentes dans Vigifriche



## Repérer une friche

 Le clic sur ce picto ouvre la possibilité de taguer une parcelle en friche grâce à la cible

Les parcelles en friche déjà repérées avec Vigifriche sont visibles à l'écran (en jaune)



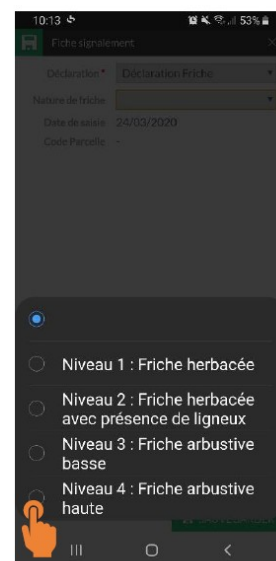
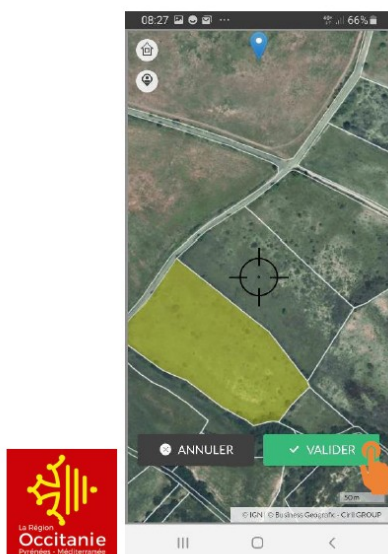
## Vigifriche Occitanie

Les possibilités de saisies sont les suivantes :

- « taguer » et qualifier une nouvelle friche
- confirmer une friche
- changer une occupation du sol renseignée



## Déclarer et qualifier la friche



## Illustrer et commenter la friche

Possibilité d'intégrer une photo caractérisant la parcelle taguer. Il est nécessaire de prendre d'abord la photo ; ensuite, il faut la sélectionner dans la galerie de photos pour l'affecter à la parcelle sélectionnée.

*(Attention à la taille du fichier, utiliser le format le moins volumineux que propose le smartphone)*



# Sauvegarder la friche

