Poste : Variable Numériques en appui aux métiers de police de l’OFB

Auteur : Hugo Senges

Date de poste : Octobre – Décembre 2021

Objet : Routine d’entraîneme,nt et d’expansion d’un modèle de classification des parcelles par télédétection- scope CIPAN

Documents cadres : « Note télédétection»

**Contexte et existant :**

Afin de pouvoir rapidement implémenter un nouveau modèle de classification, et d’en comparer les résultats avec les précédentes itérations, deux conditions doivent être réunies :

Premièrement la vérité terrain acquise doit respecter (dans les grandes lignes) les nomenclatures de couverture végétale et de qualité du couvert indiquées dans la documentation.

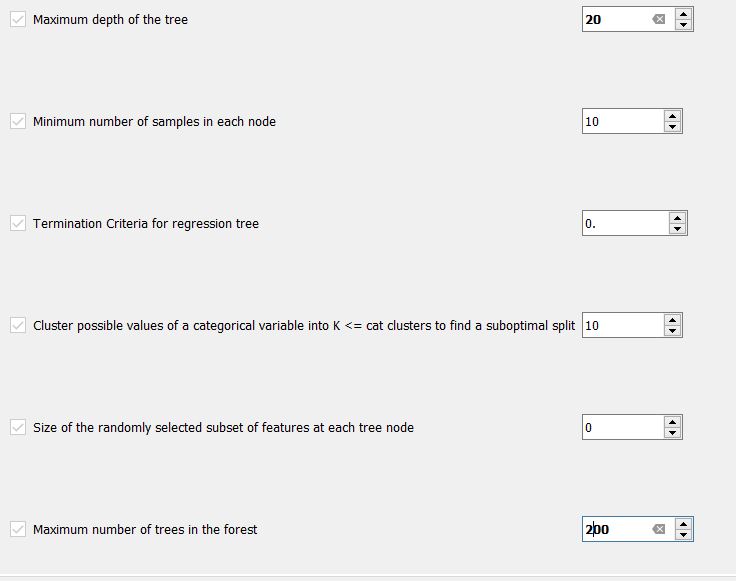
Deuxièmement, les images satellites sélectionnées sur les zones à classifier doivent être comparables à celles utilisées dans la méthode (en termes de résolution spatiale, de canaux radiométriques convoqués et de saison représentée).

Pour rappel les caractéristiques images utilisées jusqu’ici sont les suivantes :

* Image issues du programme Sentinel 2, traitées par la chaîne MAJA de désenuagement (produit fourni par le pôle Theia), de résolution spatiale 10m sur les bandes suivantes   
  [2 : Bleu, 3 : Vert, 4 : Rouge, 5 : PIR] entre les mois de Novembre et de Janvier

Les traitements sont standardisés et automatisés via les scripts fourni dans l’archive (***Images\_Classification\Modele\_classif\_prez***). Deux scripts sont disponibles : ***Prepa\_img.py*** et ***Entrainement.py***

* Le programme ***Prepa\_img.py*** effectue les différents traitements permettant de générer le « cube-image », c’est-à-dire la ressource image (***.tif***) sur laquelle faire tourner l’entraînement du modèle puis sont expansion. Il procède par mosaïquage, masquage, calculs des dérivées (NDVI et paramètres de texture d’Haralick) et enfin concaténation des bandes images. L’ordre final des bandes images doit être le suivant : [ Bleu, Vert, Rouge, Infrarouge, NDVI, Haralick-0 à Haralick-8]
* Le programme ***Entrainement.py*** permet de produire le fichier paramètre du modèle de classification supervisé, la matrice de confusion associée et la couche ***.shp*** résultante, sur la base de l’échantillon d’entrainement acquis en vérité-terrain. Les paramètres de l’algorithme de classification (outil ***TrainVectorClassifier*** de la librairie OTB, option ***Random Forest*** selon le paramétrage suivant :



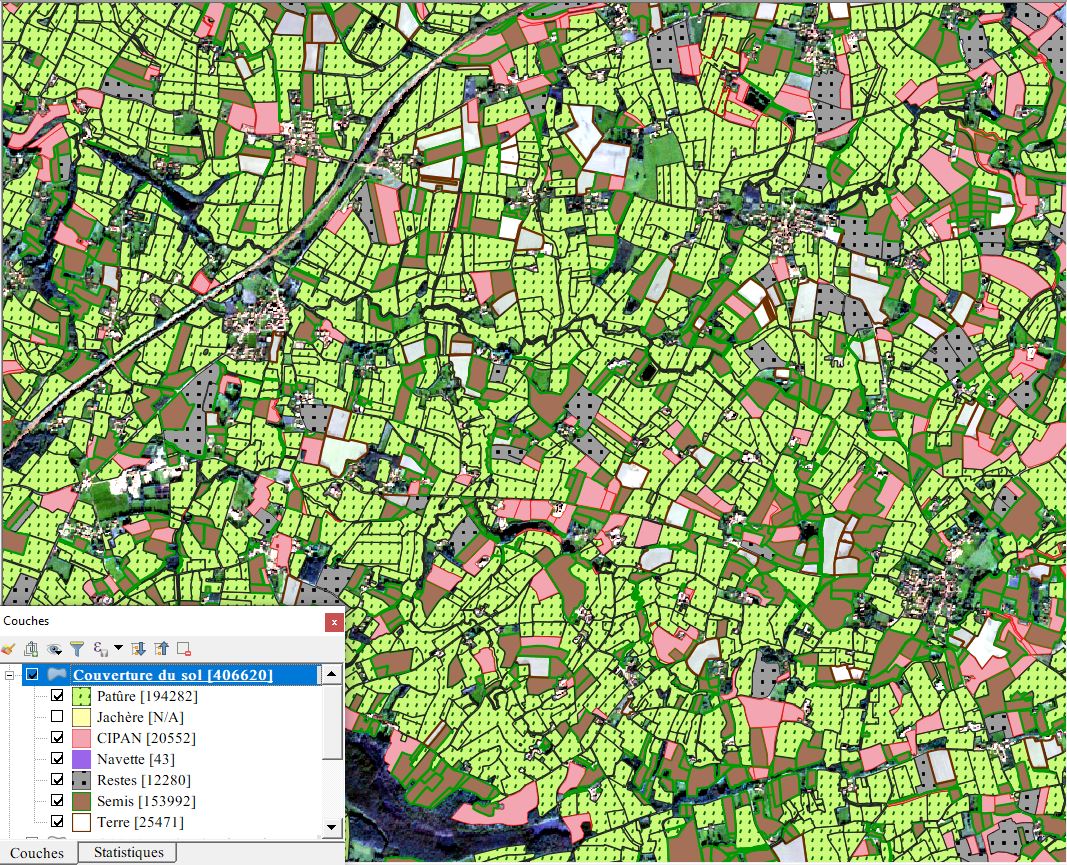
**Tâche type : production d’une couche géographique de couverture du sol :**

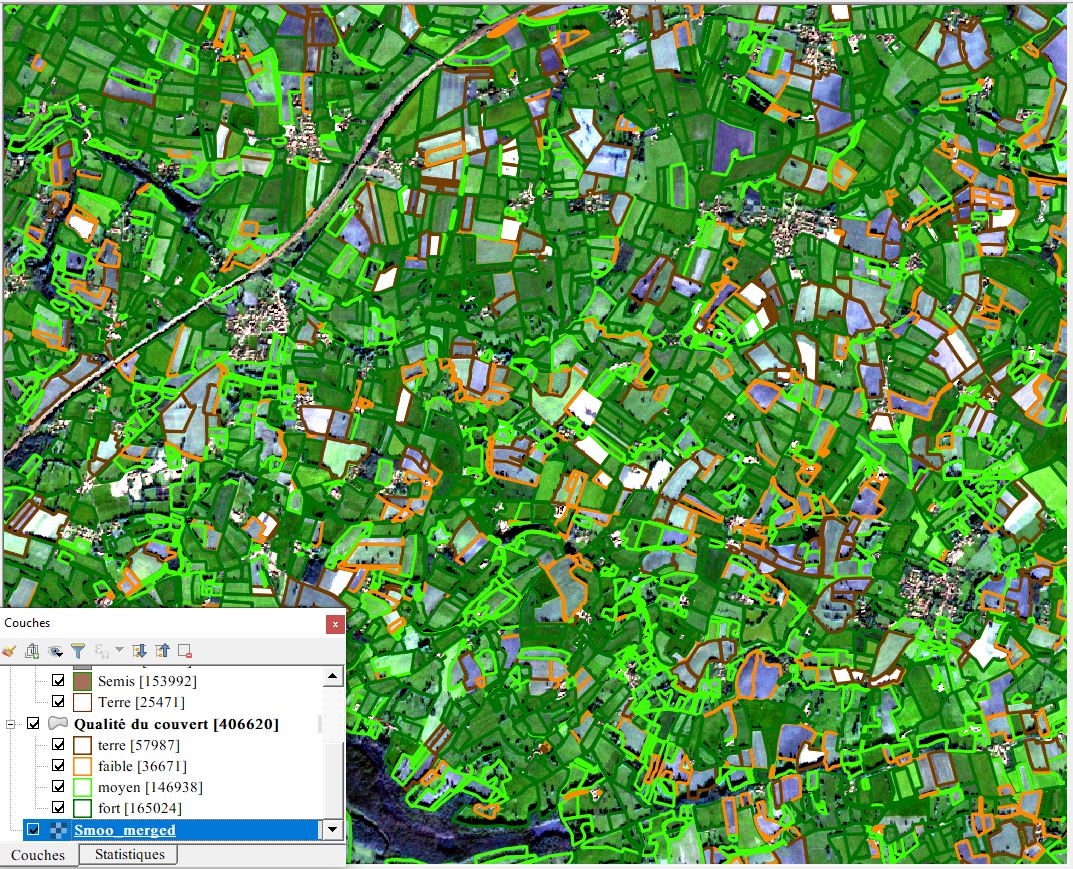
* Acquisition d’une VT sous la forme d’une couche ***.shp*** associant aux parcelles du RPG la nomenclature de couverture du sol approprié (entre 4 et 8 postes *a priori*).
* Acquisition des images Sentinel-2, level 2A sur la plateforme Theia (<https://catalogue.theia-land.fr> ) selon la date la plus proche de l’acquisition de la vérité-terrain et le meilleur taux de couverture de l’aire d’intérêt.
* Préparation du « cube-image » via le programme ***Prepa\_img.py***. (inputs : les images satellite couvrant l’aire d’intérêt)
* Production du modèle entraîné via le programme ***Entrainement.py***. (inputs : l’image produite à l’étape précédente et la couche RPG d’entraînement)
* Vérification de la conformité des résultats sur la Matrice de confusion produite (OA > 70%)
* Extrapolation du modèle à l’ensemble de l’aire d’intérêt (inputs : la couche RPG de la zone renseignée en ***Zonal Statistics*** sur le « cube-image », le modèle de classification produit aux étapes précédentes).

**Piste de développement :**

* L’intertemporalité devrait être pensée en amont du processus de récolte de la VT et ce sur plusieurs plans. Premièrement le choix des dates enquêtées est d’importance car il détermine la nature de la couverture du sol à l’instant (avant ou après les travaux du sol). Deuxièmement, il est peut-être plus facile de lever une ambiguïté de classification sur une date seule si le caractère intertemporel apporte de l’information supplémentaire sur le seul couvert hivernal. Autrement dit une première date de vérité-terrain correspondant au couvert précédent suivi d’une autre en couvert hivernal donnera sans doute de moins bon résultat que deux vérité-terrain ne variant que par le développement du même couvert hivernal. Pour déterminer les dates optimales, se référer à des dires d’experts sur le paysage et les pratiques agronomiques de la région considérée.
* Le Radar apporterait sans doute, pour la qualité du couvert, une information précieuse. Cependant de par la nature du signal il ne peut être utilisé qu’en intertemporalité pour qualifier le développement de la qualité du couvert. La piste polarimétrique (variation différentiel des réponses-signal en fonction de la polarisation de l’onde) pourrait cependant apporter un complément d’information sur la couverture elle-même par rapport à la seule interférométrie.
* L’implémentation de cette méthode sous le standard SISSPEO permettrait, en plus d’accroître son inter-opérabilité de pouvoir capitaliser sur ce travail en développant sur cette base les cartouches inter-temporelles.

**Exemple de production :**

Cartographies obtenues par le biais de la méthode en Bretagne, à l’hiver 2021-2022 :



|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Couverture du sol | | **Données du terrain** | | | | | | | | | | | | | | | | | | | Total des échantillons | | |
| Légende: | | Pâture | | Récoltes | | Repousses | | Maïs/Sorgho | | CIPAN | | Colza | | Tournesol | | Vigne | | Soja | Céréales | Terre nue | | |  |
| **Prédiction du modèle supervisé** | Label : | | 1 | 2 | | 3 | | 4 | | 5 | | 6 | | 7 | | 8 | | 9 | 10 | 11 | | |  |
| Pâture | | 246 | 0 | | 0 | | 0 | | 0 | | 0 | | 0 | | 0 | | 0 | 0 | 3 | | | 249 |
| Récoltes | | 2 | 152 | | 0 | | 3 | | 0 | | 0 | | 0 | | 0 | | 0 | 0 | 2 | | | 159 |
| Repousses | | 7 | 0 | | 100 | | 0 | | 0 | | 0 | | 0 | | 0 | | 0 | 0 | 3 | | | 110 |
| Maïs/Sorgho | | 3 | 3 | | 0 | | 90 | | 0 | | 0 | | 0 | | 0 | | 0 | 0 | 1 | | | 97 |
| CIPAN | | 1 | 1 | | 1 | | 1 | | 1 | | 0 | | 0 | | 0 | | 0 | 0 | 2 | | | 7 |
| Colza | | 3 | 1 | | 1 | | 0 | | 0 | | 29 | | 0 | | 0 | | 0 | 0 | 0 | | | 34 |
| Tournesol | | 1 | 2 | | 2 | | 0 | | 0 | | 0 | | 0 | | 0 | | 0 | 0 | 1 | | | 6 |
| Vigne | | 6 | 0 | | 0 | | 0 | | 0 | | 0 | | 0 | | 14 | | 0 | 0 | 0 | | | 20 |
| Soja | | 1 | 0 | | 0 | | 2 | | 0 | | 0 | | 0 | | 0 | | 0 | 0 | 0 | | | 3 |
| Céréales | | 0 | 2 | | 0 | | 0 | | 0 | | 0 | | 0 | | 0 | | 0 | 5 | 1 | | | 8 |
| Terre nue | | 1 | 0 | | 1 | | 0 | | 0 | | 0 | | 0 | | 0 | | 0 | 0 | 227 | | | 229 |
| Total des échantillons | | 271 | | 161 | | 105 | | 96 | | 1 | | 29 | | 0 | | 14 | | 0 | 5 | 240 | | | 922 |
| Prédictions correctes | | 865 | |  | OA (%) | | 93,82 | |  | |  | |  | |  | | OA (sans label 5-7-9-10) (%) | | | | | 95,66 | |

Comparaison des différentes matrices de confusion obtenues par la méthode sur les département bretons (hiver 2021-2022) et gersois (hiver 2018-2019) :

Tableau 1 : Matrice de confusion obtenue sur le Gers en 2018 : Couverture du sol

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Couverture du sol | | **Données du terrain** | | | | | | Total des échantillons |
| Légende: | | Pature | CIPAN | Navette | Restes | Semis | Terre |
| **Prédiction du modèle supervisé** | Label : | 1 | 3 | 4 | 5 | 6 | 7 |
| Pature | 81 | 0 | 0 | 0 | 0 | 0 | 81 |
| CIPAN | 1 | 30 | 0 | 0 | 0 | 1 | 32 |
| Navette | 4 | 0 | 9 | 0 | 0 | 0 | 13 |
| Restes | 2 | 0 | 0 | 17 | 0 | 0 | 19 |
| Semis | 2 | 0 | 0 | 0 | 57 | 0 | 59 |
| Terre | 0 | 0 | 0 | 0 | 0 | 29 | 29 |
| Total des échantillons | | 90 | 30 | 9 | 17 | 57 | 30 | 233 |
| Prédictions correctes | | 223 | OA (%) | 95,71 |  | OA (patureXjachere) (%) | | 97,42 |

Tableau 2 : Matrice de confusion obtenue sur l’Ille-et-Vilaine et le Morbihan en 2021 : Couverture du sol

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Couverture du sol | | **Données du terrain** | | | | | | Total des échantillons |
| Légende: | | Faible | Moyen | Fort | Résidus Enfouis | Résidus non-enfouis | Terre nue |
| **Prédiction du modèle supervisé** | Label : | 1 | 2 | 3 | 4 | 6 | 7 |
| Faible | 53 | 0 | 1 | 0 | 0 | 2 | 56 |
| Moyen | 0 | 41 | 10 | 0 | 0 | 1 | 52 |
| Fort | 0 | 0 | 387 | 0 | 1 | 2 | 390 |
| Résidus Enfouis | 0 | 0 | 0 | 42 | 0 | 4 | 46 |
| Résidus non-enfouis | 0 | 0 | 10 | 0 | 148 | 1 | 159 |
| Terre nue | 0 | 0 | 2 | 0 | 0 | 217 | 219 |
| Total des échantillons | | 53 | 41 | 410 | 42 | 149 | 227 | 922 |
| Prédictions correctes | | 888 |  |  |  |  | OA (%) | 96,31 |

Tableau 3 : Matrice de confusion obtenue sur le Gers en 2018 : Qualité du Couvert

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Qualité du couvert | | **Données du terrain** | | | | Total des échantillons |
| Légende: | | terre | faible | moyen | fort |
| **Prédiction du modèle supervisé** | Label : | 1 | 2 | 3 | 4 |
| terre | 39 | 0 | 0 | 1 | 40 |
| faible | 0 | 34 | 0 | 0 | 34 |
| moyen | 0 | 0 | 64 | 0 | 64 |
| fort | 0 | 0 | 0 | 95 | 95 |
| Total des échantillons | | 39 | 34 | 64 | 96 | 233 |
| Prédictions correctes | | 232 |  |  | OA (%) | 99,57 |

Tableau 4 : Matrice de confusion obtenue sur l’Ille-et-Vilaine et le Morbihan en 2021 : Qualité du Couvert