
QGIS 13



Classification d'image avec QGIS et OTB

Novembre 2024

























TABLE DES MATIERES

1. INTRODUCTION.....	1
2. INSTALLATION D’OTB	2
2.1. VERIFICATION DE L’INSTALLATION D’OTB SUR LES ORDINATEURS DE LA FAC.....	2
2.2. INSTALLATION D’OTB SUR LES ORDINATEURS	3
3. PREPARATION ET TRAITEMENT DES IMAGES.....	5
3.1. CREATION D’UNE IMAGE MULTISPECTRALE	5
3.2. DIGITALISATION DES ZONES D’ENTRAINEMENT	7
3.3. CLASSIFICATION SUPERVISEE PAR « <i>RANDOM FOREST</i> »	9
3.3.1. <i>Entraînement du modèle de classification</i>	9
3.3.2. <i>Application du modèle de classification</i>	11
4. POUR ALLER PLUS LOIN	13


1. Introduction

- Cet exercice présente un exemple de classification d'images satellitaires réalisée à l'aide de la boîte à outils OTB (Orfeo Toolbox) utilisée dans l'environnement QGIS.
- La méthode principale utilisée dans ce TP repose sur une classification supervisée par pixel, une approche de traitement d'images où chaque pixel est attribué à une classe prédéfinie. Plus précisément, nous utiliserons une approche de modélisation basée sur le « *machine learning* », i.e. la méthode de forêt aléatoire (« *random forest* »). Cette méthode fonctionne en créant plusieurs arbres de décision pour classer les pixels de l'image en fonction de leurs caractéristiques spectrales.
- Pour toute méthode de classification supervisée, il est nécessaire de disposer d'un jeu de données dit « d'entraînement ». Ce jeu de donnée permet de calibrer le modèle de classification. Dans cet exercice, une étape de digitalisation est proposée pour permettre la création de ces données d'entraînement à partir des images disponibles. La digitalisation consiste à numériser manuellement des polygones représentatifs des différences classes d'occupation du sol (ex. végétation ligneuse, végétation herbacée, eaux de surface, zones artificialisées).
- Une fois le jeu de données d'entraînement préparé, il est possible de procéder à la classification des pixels. L'algorithme de forêt aléatoire va analyser les caractéristiques spectrales des pixels de l'image et leur attribuer une classe basée sur les informations extraites du jeu de données d'entraînement.
- Les images utilisées dans cet exercice couvrent la région de Gembloux et sont extraites de scènes Sentinel-2 acquises en 2016 aux cinq dates suivantes : 8 mai, 20 juillet, 28 août, 25 septembre et 4 décembre. Les fichiers sont disponibles sur le lien mis à disposition sur eCampus.

 s2_20160508_B2.tif	 s2_20160720_B2.tif	 s2_20160826_B2.tif	 s2_20160925_B2.tif	 s2_20161204_B2.tif
 s2_20160508_B3.tif	 s2_20160720_B3.tif	 s2_20160826_B3.tif	 s2_20160925_B3.tif	 s2_20161204_B3.tif
 s2_20160508_B4.tif	 s2_20160720_B4.tif	 s2_20160826_B4.tif	 s2_20160925_B4.tif	 s2_20161204_B4.tif
 s2_20160508_B8.tif	 s2_20160720_B8.tif	 s2_20160826_B8.tif	 s2_20160925_B8.tif	 s2_20161204_B8.tif

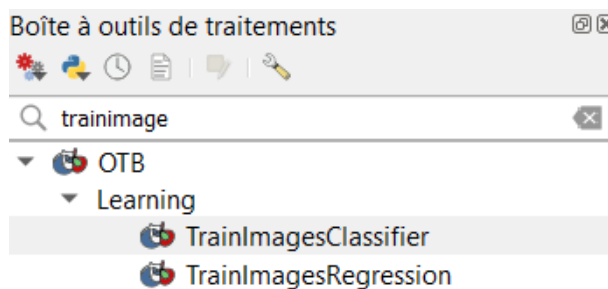
2. Installation d'OTB

2.1. Vérification de l'installation d'OTB sur les ordinateurs de la fac

- Pour vérifier si OTB est installé correctement sur votre ordinateur, cliquer sur l'icône d'option  de la boîte à outils de traitements de QGIS.
- Dans l'onglet « Fournisseurs de services », vérifier que vous disposez bien du service « OTB ».



- Lorsque vous écrivez « *trainimage* » dans l'onglet de recherche de la boîte à outil de traitements, vous devriez voir les options suivantes apparaître.



- Si ce n'est pas le cas, il est nécessaire de procéder à une nouvelle installation d'OTB sur votre ordinateur (cf. point suivant).

2.2. Installation d'OTB sur les ordinateurs

- Si vous souhaitez installer OTB sur vos ordinateurs personnels ou sur ceux du SIG, nous reprenons ci-dessous les principales étapes d'installation expliquées sur le site suivant :

<https://www.sigterritoires.fr/index.php/ajouter-orfeo-toolbox-dans-qgis-3-x/>

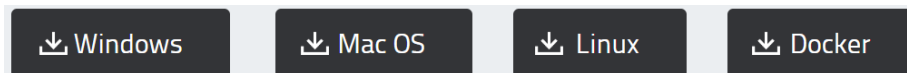


Il faut suivre **TRÈS EXACTEMENT** les étapes décrites ci-dessous pour que l'installation se fasse correctement.



Nous n'avons, à ce jour, pas trouvé de solution pour faire tourner OTB sur Mac OS. Pour utiliser OTB, nous vous recommandons de travailler sur les ordinateurs du SIG.

- À partir de l'adresse <https://www.orfeo-toolbox.org/download/>, télécharger la version correspondant à votre système d'exploitation (*a priori*, Windows).



- Dézipper le fichier .zip téléchargé et copier le répertoire décompressé (OTB-8.1.2-Win64) directement sur le disque C:/. Renommer-le « OTB ». Vous aurez donc un répertoire **C:/OTB**.

Ce PC > OS (C:)				
Nom	Modifié le	Type	Taille	
Microsoft	07-05-21 18:34	Dossier de fichiers		
OTB	26-10-23 11:02	Dossier de fichiers		
PerfLogs	07-05-22 07:24	Dossier de fichiers		
ProgramData	18-09-23 13:48	Dossier de fichiers		
Programmes	22-09-23 16:22	Dossier de fichiers		
Programmes (x86)	24-07-23 18:12	Dossier de fichiers		

- Il faut ensuite télécharger un second dossier, le « qgis-otb-plugin », via l'adresse suivante : <https://gitlab.orfeo-toolbox.org/orfeotoolbox/qgis-otb-plugin>. Comme dans la capture d'écran reprise ci-dessous, cliquer sur l'icône « Code » sur la droite **Code** et choisir « Télécharger le code source .zip ».

Nom	Dernière validation	Dernière mise à jour
otb	release for 1.4.2	il y a 5 ans
.gitignore	Initial commit	il y a 7 ans
COPYING	first version	il y a 7 ans
LICENSE	Initial commit	il y a 7 ans
README.md	Update README.md wit...	il y a 3 ans
mock_extended...	pending qgis PR6739	il y a 6 ans

- Télécharger le répertoire et décompresser le fichier .zip résultant. Copier le contenu du répertoire « qgis-otb-plugin-master » dans un nouveau répertoire que vous devez créer sur votre disque C:/ et que vous nommerez **OBLIGATOIREMENT** :

C:/qgis-plugins/qgis-otb-plugin

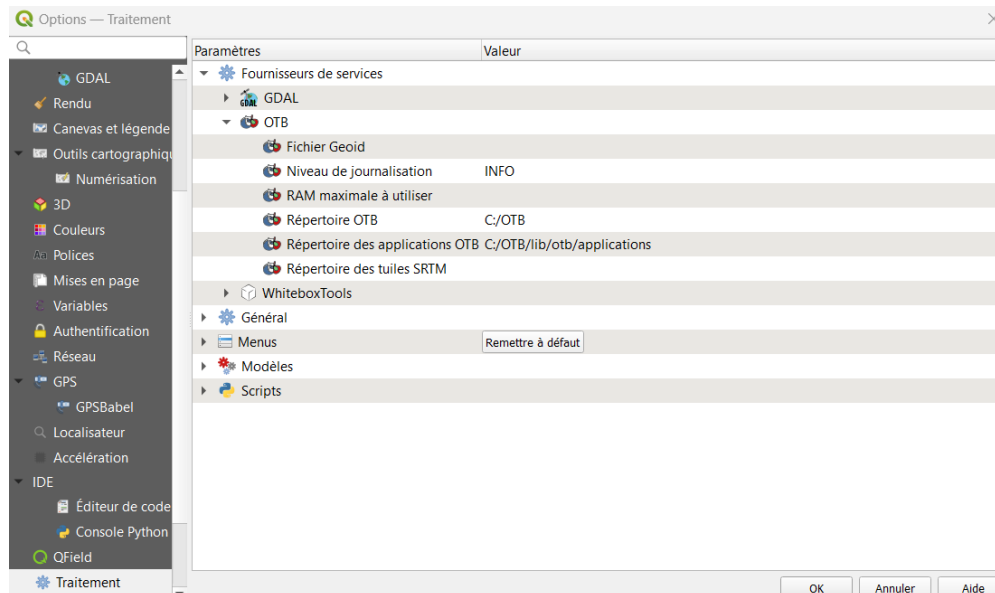
- Vous devez obtenir la structure de dossier suivante :

📁 > Ce PC > OS (C:) > qgis-plugins > qgis-otb-plugin

- L'installation est terminée. Il reste maintenant à configurer le plugin. Dans QGIS, activer l'extension « OrfeoToolbox provider » via la commande [Extensions] → [Installer/Gérer les extensions].

- Retourner ensuite dans l'onglet « Fournisseurs de services » via l'icône d'option de la boîte à outils de traitements de QGIS.
- Vérifier que vous disposez bien du service « OTB ». Encoder les chemins d'accès pour le répertoire OTB et pour le répertoire des applications OTB.

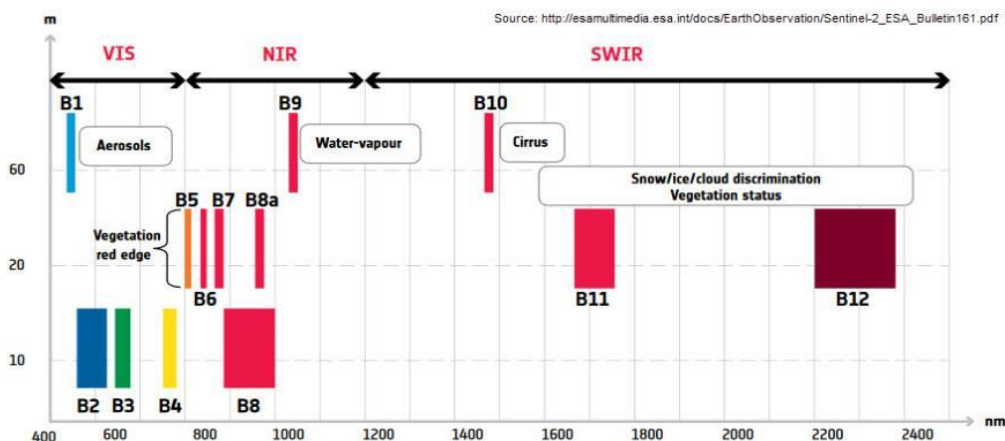
- Répertoire OTB : **C:/OTB**
- Répertoire des applications OTB : **C:/OTB/lib/otb/applications**
- Ces chemins d'accès sont repris dans la capture d'écran ci-dessous. Pour encoder ces chemins d'accès, il est nécessaire de double-cliquer à l'extrême droite de la ligne que vous souhaitez modifier. Les trois petits points s'affichent et vous pouvez encoder les chemins d'accès.



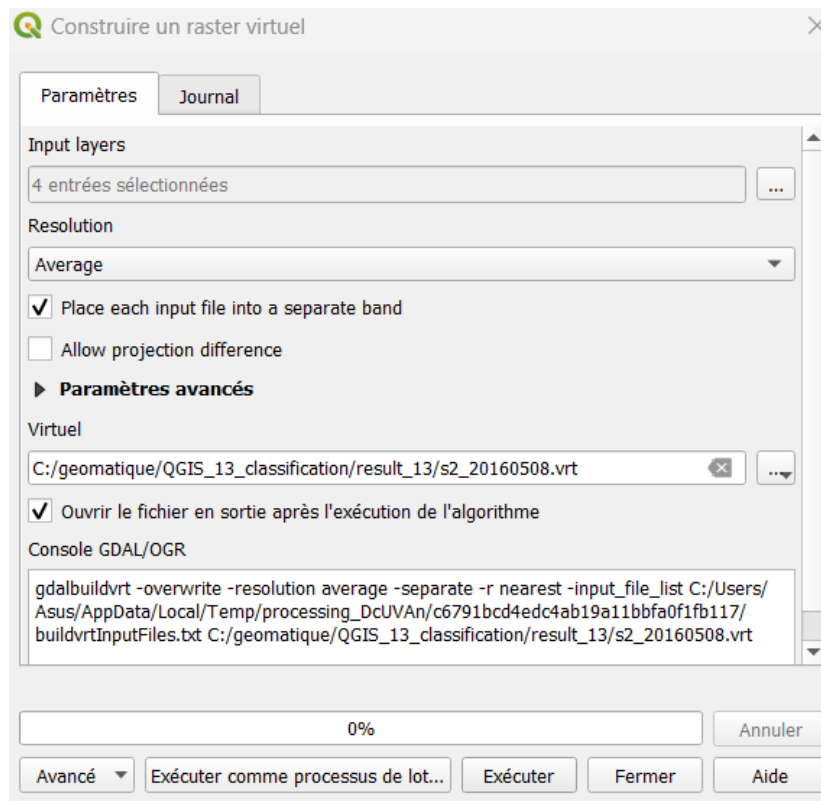
3. Préparation et traitement des images

3.1. Création d'une image multispectrale

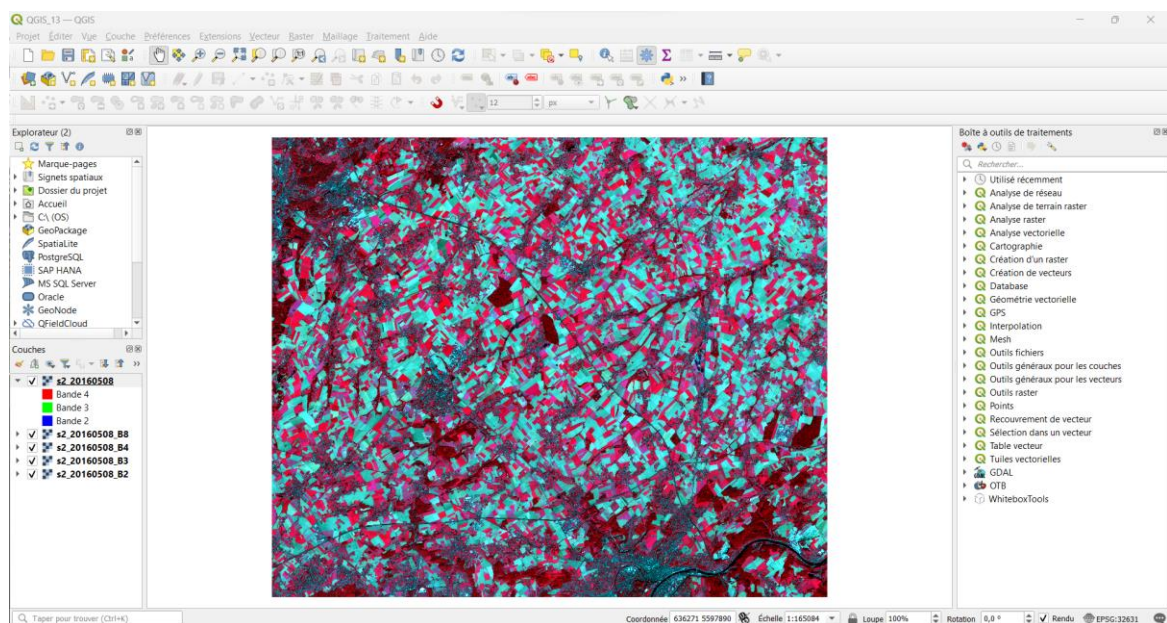
- Pour rappel, les images Sentinel-2 comportent 13 bandes spectrales présentant des résolutions allant de 10 m à 60 m. Ces bandes couvrent une large gamme de longueurs d'onde, ce qui permet de capturer des informations détaillées sur différents types de surface terrestre, tels que la végétation, l'eau, les zones urbaines et agricoles. Les bandes B1, B9 et B10 sont dédiées aux corrections atmosphériques permettant d'améliorer la qualité des images. Elles ne sont pas utilisées dans cet exercice.



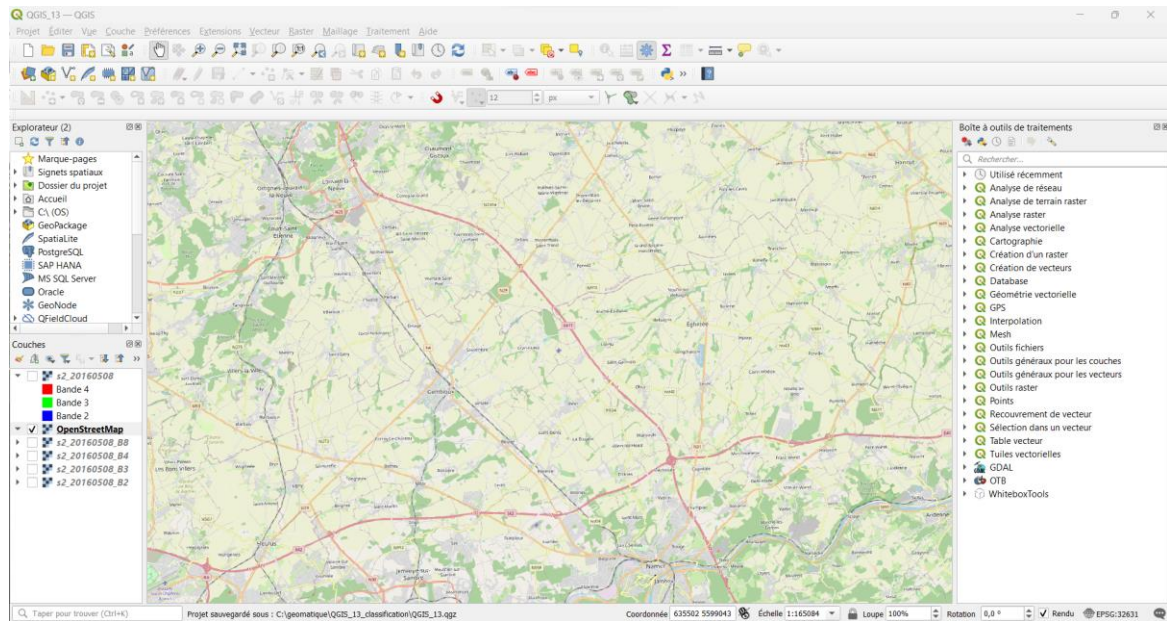
- Empiler les 4 bandes (B2, B3, B4, B8) de l'image Sentinel-2 du 8 mai 2016 dans un raster virtuel à l'aide de l'outil « **Construire un raster virtuel** ». Nommer le **s2_20160508.vrt**.





- Afficher cette image avec une composition colorée « infra-rouge fausses couleurs » (bandes B8, B4 et B3).
- L'image devrait apparaître comme dans la figure suivante.




- Afficher également la couche OpenStreetMap depuis le serveur XYZ Tiles situé dans l'explorateur de QGIS. Cette couche permet de localiser la zone d'étude qui est centrée sur la ville de Gembloux.



3.2. Digitalisation des zones d'entraînement

- Pour réaliser une classification supervisée, il est nécessaire d'identifier un sous-ensemble de pixels appartenant à différentes classes d'occupation du sol. Ces pixels seront utilisés comme données d'entraînement du modèle.
- Dans le cadre de ce TP, nous allons réaliser une classification de base, reprenant les cinq catégories d'utilisation du sol suivantes :
 - Les zones artificialisées (bâties, industries, autoroutes...) - classe 1
 - La végétation herbacée (prairies, jardins, cultures...) - classe 2
 - La végétation ligneuse (forêts, bois, parcs...) - classe 3
 - Les eaux de surface (rivières, étangs, lacs...) - classe 4
 - Les zones d'extraction (mines...) - classe 5
- Pour digitaliser ces catégories d'utilisation du sol, nous vous recommandons de revoir le TP de digitalisation (QGIS_06). Les principales étapes sont tout de même reprises ci-dessous.
- Dans le menu « Couche », sélectionner « Créer une couche » et ensuite « Nouvelle couche shapefile ».
- Définir l'emplacement et le nom de cette couche (ex. train_qgis13.shp). Sélectionner « polygone » comme type de géométrie. Cliquer sur « OK ». Un nouveau shapefile nommé **train_qgis13.shp**, vide, s'ajoute dans la liste des couches dans QGIS.
- Pour créer des polygones, activer votre nouveau shapefile en cliquant une fois dessus dans l'onglet « Couches » et activer le mode édition en cliquant sur l'icône . Cliquer ensuite sur l'icône  pour ajouter des entités polygonales.

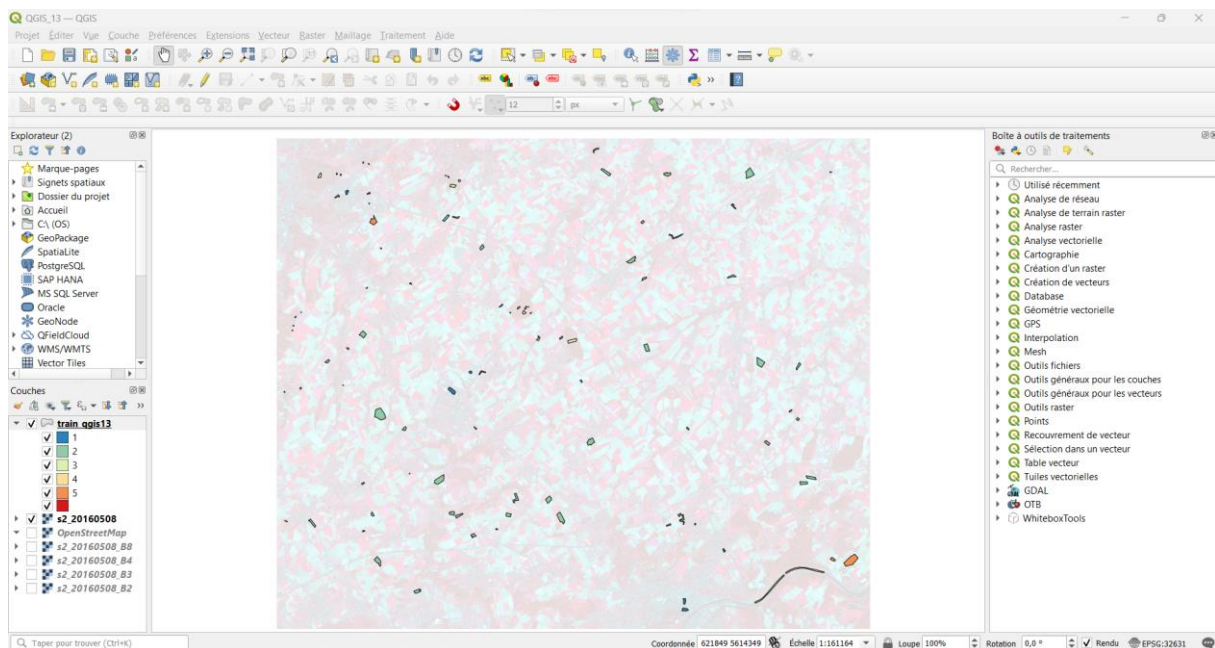
- La construction d'un polygone se fait par une succession de « clic-gauche » et se clôture par un « clic-droit ». Pour chaque polygone créé, on vous demande d'enregistrer une valeur numérique. Choisir une valeur comprise entre 1 et 5, correspondant aux cinq classes d'occupation du sol mentionnées précédemment.
- Une fois votre digitalisation terminée, effectuer une sauvegarde en cliquant sur  et sortir du mode édition.



Pour identifier correctement vos classes d'occupation du sol, il vous est recommandé de charger une orthophoto de la période correspondante. Ces orthophotos sont disponibles pour visualisation sur WalOnMap via une connexion WMS (cf. QGIS_11).

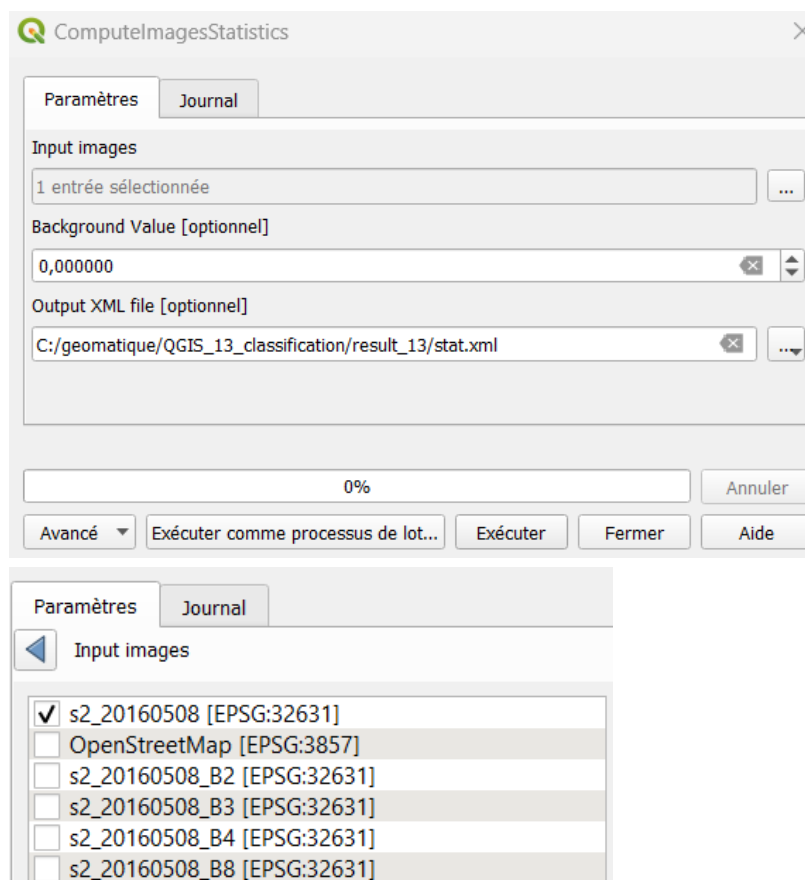


Il est nécessaire de digitaliser vos polygones de manière à recouvrir à chaque fois plusieurs dizaines de pixels de votre image satellite (si possible, sélectionner des zones homogènes d'un point de vue spectral). De plus, il est nécessaire de digitaliser plusieurs polygones pour chaque classe (5 par classe pour l'exercice proposé) et de bien les répartir sur votre zone d'étude. Au final, votre shapefile « train_qgis13.shp » devrait ressembler à la capture d'écran reprise ci-après.



3.3. Classification supervisée par « *Random Forest* »

- Avant de créer le modèle de classification, il convient de calculer les statistiques de base des différentes bandes utilisées par ce modèle. Cela permet à l’algorithme de création du modèle de normaliser les données spectrales et ainsi d’éviter qu’une bande présentant une variabilité plus importante que les autres ait un poids exagéré dans la classification.
- Le calcul des statistiques de l’image multi-spectrale est réalisé avec la fonction « **ComputImagesStatistics** » de la boîte à outils OTB. Il convient de renseigner le nom du fichier contenant l’image multi-spectrale (rubrique « Input images »), ainsi que le nom du fichier dans lequel seront sauvegardées les statistiques (moyenne et écart-type) des différentes bandes (rubrique « Output XML file »). Nommer ce fichier **stat.xml**.



3.3.1. Entraînement du modèle de classification

- L’entraînement (la création) du modèle de classification est réalisé avec la commande « **TrainImagesClassifier** » de la boîte à outils OTB. La boîte de dialogue de cet outil comporte un très grand nombre de paramètres dont la plupart peuvent conserver leur valeur par défaut.
- Suivre les indications fournies dans la figure suivante pour définir correctement les paramètres.

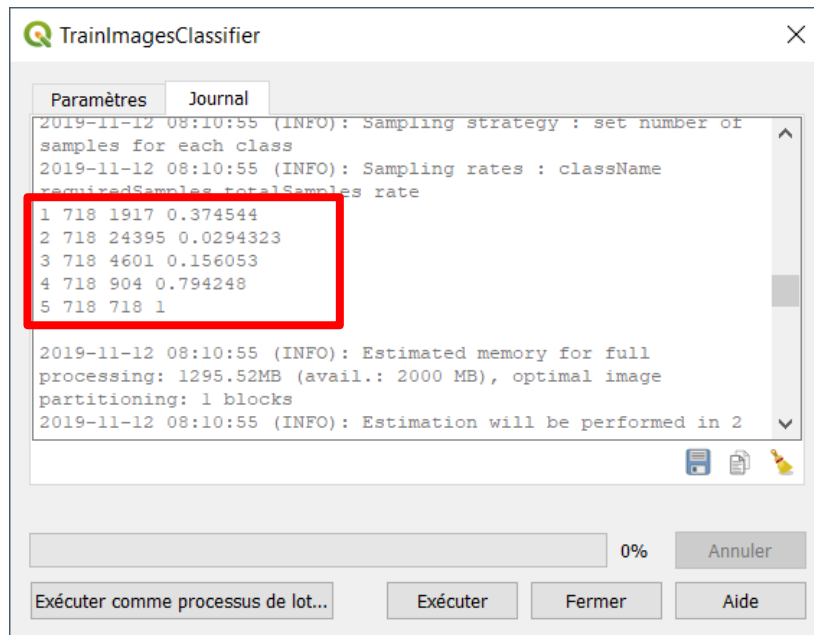
- 1° Sélectionner l'image à traiter. Choisir le fichier .vrt contenant toutes les bandes spectrales pour le 8 mai 2016 (s2_20160508.vrt).
- 2° Sélectionner le fichier contenant les données d'entraînement. Il s'agit du fichier train_qgis13.shp que vous avez digitalisé.
- 3° Renseigner le fichier contenant les statistiques des images (stat.xml).
- 4° Sélectionner le champ contenant les codes de classes. Si le champ n'apparaît pas dans la liste, il faut l'ajouter avec le clavier.
- 5° Sélectionner l'algorithme de classification « Random Forest » (rf).
- 6° Fixer le paramètre « Maximum depth of the tree » à 5.
- 7° Fixer le paramètre « Minimum number of samples in each node » à 10.
- 8° En-dessous, dans « Output model », définir ensuite le nom et l'emplacement du fichier contenant le modèle de classification (fichier .xml). Nommer celui-ci **model1.xml** pour la classification de niveau 1.
- 9° Exécuter la commande avec le bouton « Exécuter ».



Il est possible que QGIS vous demande également de définir le « User defined Input Centroid », pour lequel vous pouvez choisir votre shapefile d'input « train_qgis13.shp ».

- L'onglet « Journal » de la boîte de dialogue contient les informations générées par OTB lors de l'exécution de la commande. On peut notamment y lire le nombre de pixels par classe utilisé pour entraîner le modèle. OTB équilibre en effet le nombre de pixels utilisés entre les classes définies.

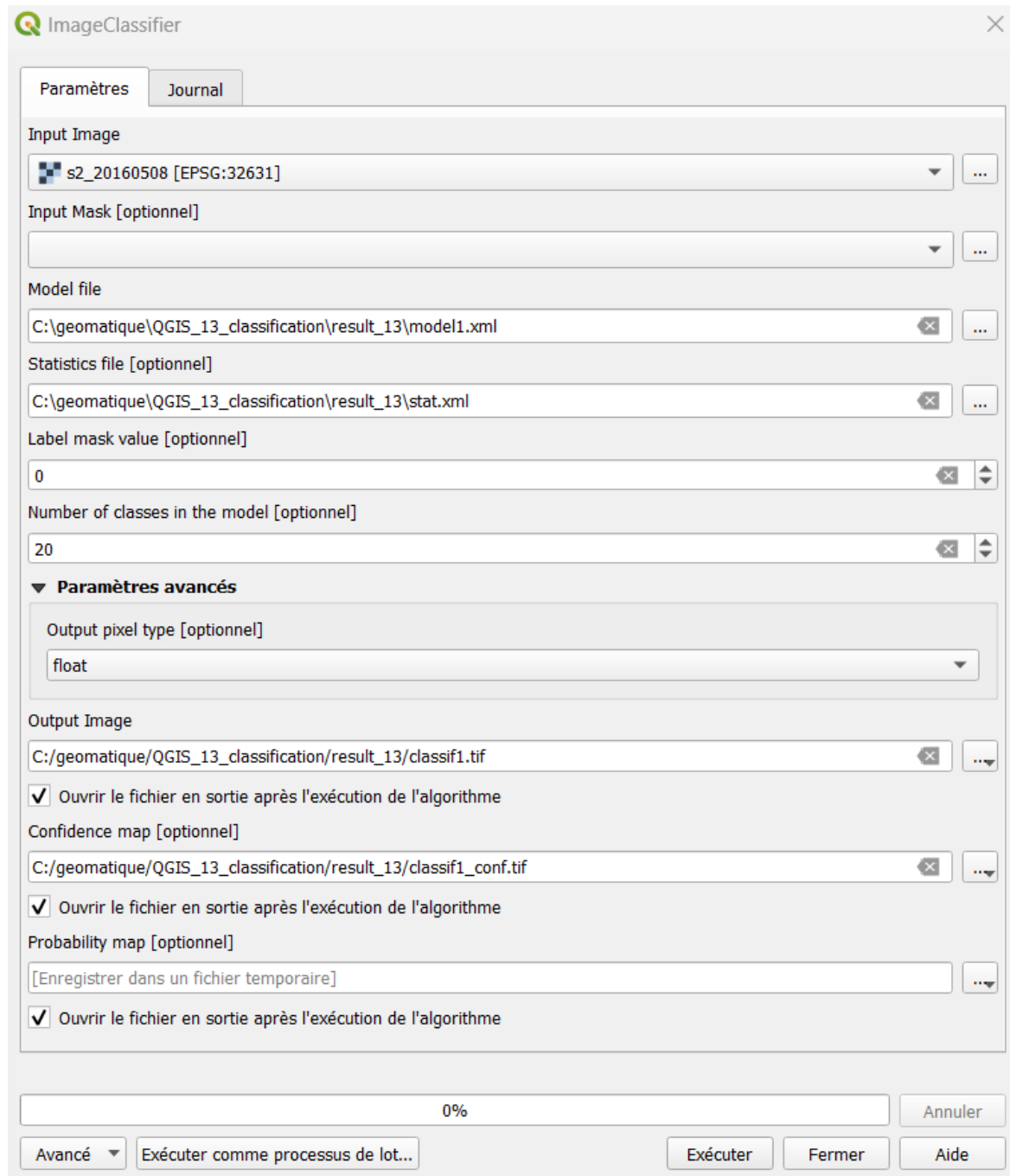
Dans notre cas, on constate que c'est la classe la moins représentée dans l'échantillon (classe 5, avec 718 pixels et un rate de 1) qui conditionne le nombre de pixels utilisé afin de respecter un échantillonnage équilibré.



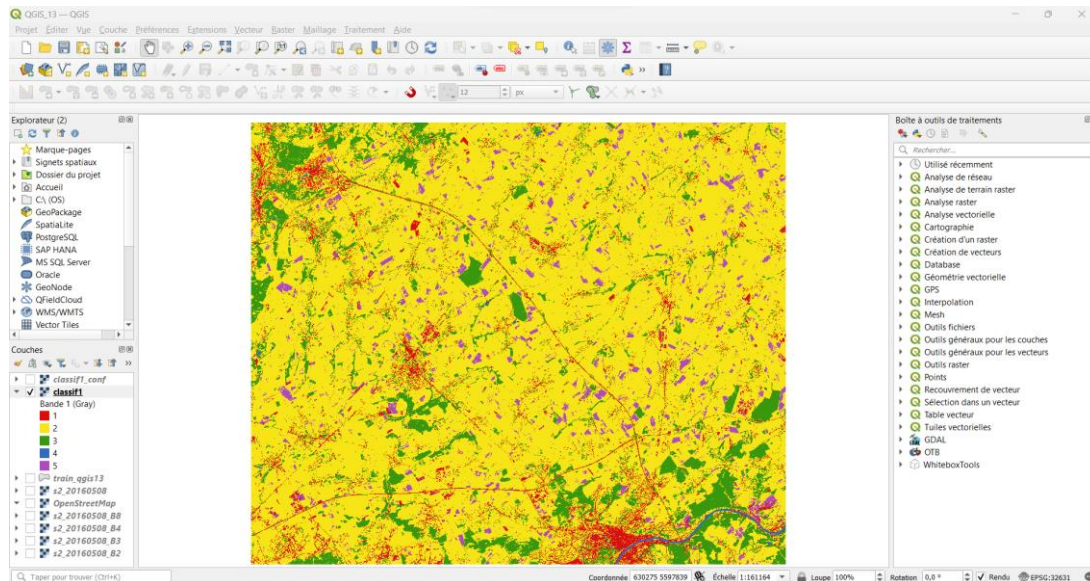
En l'absence de données de validation spécifiques (« Validation vector data list »), les données d'entraînement sont séparées en deux parties égales : une pour l'entraînement et l'autre pour la validation du modèle. Dans ce cas de figure, il est cependant possible que des pixels d'entraînement et de validation soient issus du même polygone. Les niveaux de précision obtenus sont donc plus optimistes que lorsqu'on utilise des données de validation indépendantes. Il existe cependant différents moyens pour valider une classification, nous reviendrons dessus lors de prochains TP.

3.3.2. Application du modèle de classification

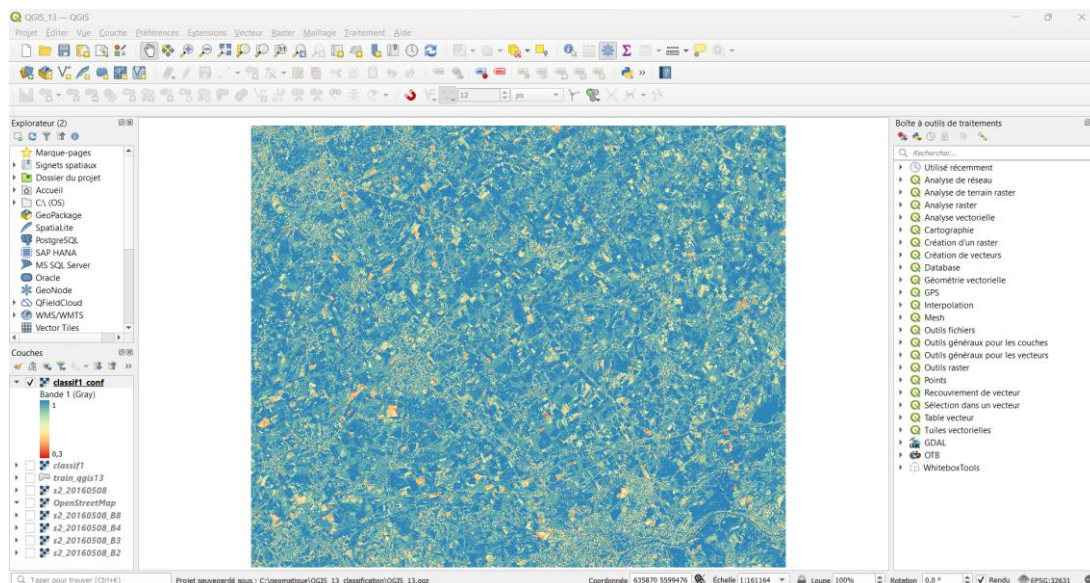
- Une fois le modèle créé, il peut être utilisé pour réaliser la classification de l'image. Cette étape est prise en charge par la commande « **ImageClassifier** » de la boîte à outils OTB.
- Dans les paramètres de cette commande, il convient de définir l'image sur laquelle porte la classification (« Input image »), le fichier contenant le modèle (« Model file »), le fichier de statistiques (« Statistics file »), le mode de codage du résultat (« Output pixel type ») qui est fixé à « float » (avec décimale) dans l'exemple ci-dessous (important pour la carte de confiance !). Il faut également définir les noms et emplacements des fichiers de sortie : le raster contenant la classification (« Output Image »), ainsi que le raster contenant un indicateur du niveau de confiance de la classification (« Confidence map »). Lorsque l'on utilise l'algorithme « Random Forest », cet indicateur contient le pourcentage de votes obtenu par la classe majoritaire dans le processus de forêt aléatoire.



- Visualiser le résultat dans QGIS en utilisant la symbologie contenue dans le fichier de style **landcover_5class.qml**.



- Utiliser le WMS de la couche orthoimage de 2016 pour évaluer visuellement la qualité de la carte.
- Identifier quelques endroits où la classification n'a pas bien fonctionné.
- Afficher également la couche correspondant aux niveaux de confiance de la classification. Utiliser le fichier de style **confidence_level.qml** pour définir la symbologie de cette couche.



4. Pour aller plus loin

- Si vous souhaitez améliorer votre classification, tester la même approche sur base d'un empilement de rasters non pas pour une date, mais pour les cinq dates disponibles (8 mai, 2 juillet, 26 août, 25 septembre, 4 décembre). Comparer visuellement les résultats obtenus et les cartes de confiance.
- Nous n'avons pas vu dans le présent TP de méthode pour évaluer la classification avec des données indépendantes. Cela sera fait dans un TP ultérieur, qui sera dédié à la question.