Trans-Atlantic Training 2022

Prague, Monday the 5th of September 2022

Vegetation classification with Sentinel-1 data

P.-L. Frison, C. Lardeux,

pierre-louis.frison@univ-eiffel.fr

Data and context

The aim is to use radar (Sentinel-1) and optical (Sentinel-2) remote sensing data together or separately to produce the cartography of the different vegetation types. The study area is the Cayenne region, in French Guiana. The first step will be to pre-process the satellite data. Then we will perform a supervised classification based on the Random Forest algorithm to produce vegetation maps.

The data used for learning and validating the classifications were obtained during *in situ* inventories. The Sentinel data used are detailed in the following table:

	Acquisition dates	Spatial resolution	wavelengths	Polarisation	Product Level
Sentinel-1	2017/04/25 2017/07/06 2017/10/10 2018/01/14	~ 20 m	5.6 cm (C Band)	VV/VH	GRDH
Sentinel-2	2018/02/06	4 bands: 10 m	430 nm – 1600 nm		L1C

In the first part, the data are processed to enable them to be used by the classification algorithm.

The second part is dedicated to the classification for the production of the vegetation map.

I. *DATA PRE-PROCESSING*

The Sentinel-1 radar data were acquired in 2 distinct polarisations: VV and VH. From these two polarizations, we will produce three bands per acquisition: VV, VH, and VH/VV, filtered to reduce speckle noise.

The Sentinel-2 optical data have 13 bands acquired in different wavelengths. We will use the 4 bands acquired at 10 m to which we will add the NDVI.

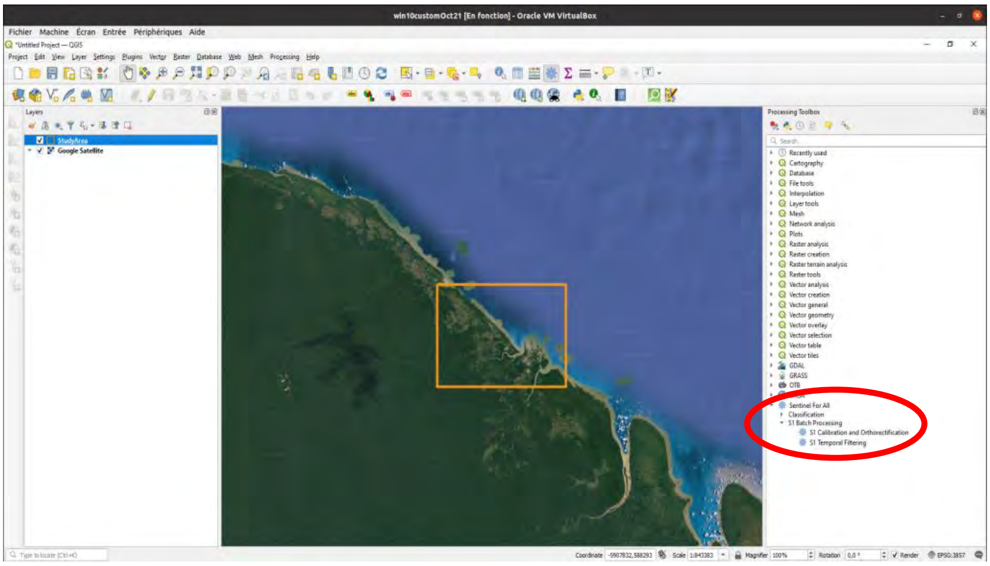
I. 1. Sentinel-1:

The Sentinel-1B data were acquired in IW (Interferometric Wide) mode and were downloaded from https://search.asf.alaska.edu/ in GRDH format.

They need to be orthorectified (in the UTM zone 32N geographic projection, EPSG : 32622) and calibrated (to convert the digital number DN into a physical quantity: the radar backscatter coefficient σ^0) over the area of interest defined by the ***study_area.shp*** vector layer. The resulting images are then stacked (each pixel of each image corresponds to the same position on the Earth).

Then, a spatio-temporal filter[1] is applied to reduce the speckle noise inherent in radar images.

These pre-processings are carried out under an adapted version of the QGIS software, integrating the OTB software (Orfeo Toolbox) and python libraries (see Figure below). OTB is a free software available here https://www.orfeo-toolbox.org/. It is specially dedicated to image processing for remote sensing data. It has the advantage of allowing large data processing without any particular machine configuration (4 GB of RAM is sufficient)



*The QGIS software integrating **OTB** and the **Sentinel for all** plugin (circled in red) dedicated to the processing of Sentinel-1 data for classification. Also the Googlearth background and the polygon delimiting the study area.*

[1] S. Quegan, J. J. Yu, 2001: Filtering of Multichannel SAR images, IEEE Transactions on Geosciences and Remote Sensing, vol. 39, N° 11, 2373-2379.

IMPORTANT PRELIMINARY COMMENTS :

1) To avoid execution errors, the paths and directories that you define on your computer must not contain special characters, in particular accents, spaces, operation signs.....
   For example, you should transform ***C:\My documents\Frédéric-Sultan\données Sentinel-1 i***n ***C:\Mes_documents\Frederic_Sultan\donnees_Sentinel1***
2) It is better to put all the data in a directory not too far from the root C:\, D:\ or another disk on your system. For example in D:\FrenchGuiana
   Avoid leaving downloaded files or files that you create in bulk anywhere on your computer.

I. 1.1: Calibration, orthorectification and clip over the study area

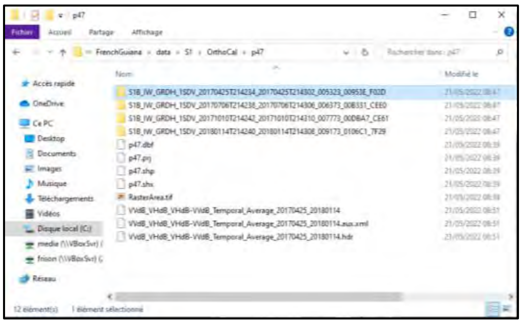
First of all, load in QGIS the ***study_area.shp*** layer (*Layer* → *Add Layer* → *Add Vector Layer* ***/data/vector/study_area.shp)***.
Then, in the panel ***Processing*** of QGIS, click on ***Scripts*** → ***Sentinel For All*** → ***S1 Batch Processing*** → ***S1 Calibration and Orthorectification***
FIll in the fields as follow:

Input folder: data\S1\zip (!!! Must contain only S1 acquisitions in ***.zip*** format)
Shape File: ***data\vectors\study_area.shp***
Field Contaning the orbit number: must contain ***« rel_orb »***, the name of the corresponding field of the « study_area » layer
DEM Folder: ***data\DEM*** (!!! This repertory must contain only the DEM file in ***.tif*** format).
This DEM file is derived from simultaneous interferometric radar data acquired during the SRTM mission. Its spatial resolution is 1 arcsec (~ 30 m). The tiles can be downloaded from https://earthexplorer.usgs.gov/.

Orbit Relative Number to process: must contain the value of the field ***« rel_orb »*** of the « study_area » layer.
Output Data Folder: ***data\S1\Orthocal*** (!!! Must be an empty repertory)
RAM : 1000[2]

Une fois le traitement réalisé, le répertoire ***data\S1\Orthocal\p47*** (cf l'orbite relative) est créé, comprenant l'arborescence suivante:

[2] It is recommanded to enter the ¼th of the RAM of your computer (1000, *i.e.* 1 Gb if your computer gets 4 Gb of RAM)

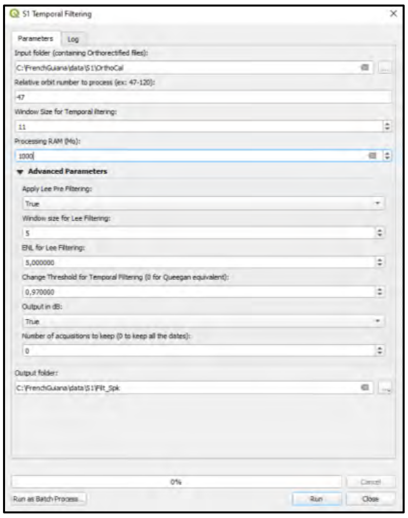


Each of the 4 sub-folders contains 2 tif files (VV, VH), calibrated, orthorectified and cropped to the area of interest (***study_area.shp***). Each tif file consists of the same number of rows and columns i.e. ready to be overlaid on each other. In addition, the file ***VVdB_VHdB_VHdB-VVdB_Temporal_Average_20170425_20180114*** has also been created. You can visualise it: it corresponds to the temporal average of each pixel according to the colour composition :

R : polarisation ratio VV/VH
V : VH Polarisation
B : VV Polarisation

I.1.2: Speckle Filtering

In the ***Processing*** panel:
Scripts → S1 Batch Processing → S1 Temporal Filtering.
Fill in the fields as follow:



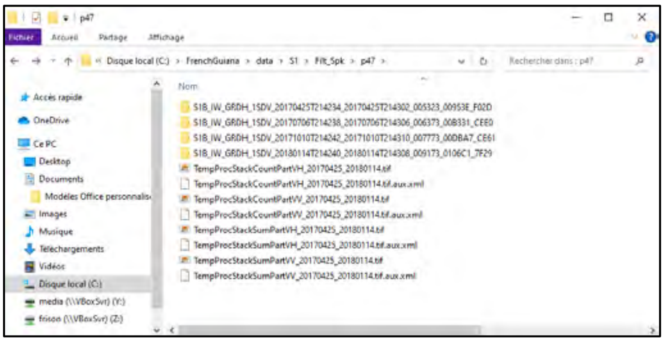
Input folder: data\S1\OrthoCal
Output Folder: data\S1\Filt_Spk (!!! Must be an empty folder)

As there are only 4 radar acquisitions, an additional pre-filtering of each data is performed. This is the Lee filter[3],[4] applied only in the spatial domain. Then the spatio-temporal filter is applied. Lee's pre-filtering is not necessary when you have a large number of acquisitions to process. Indeed,

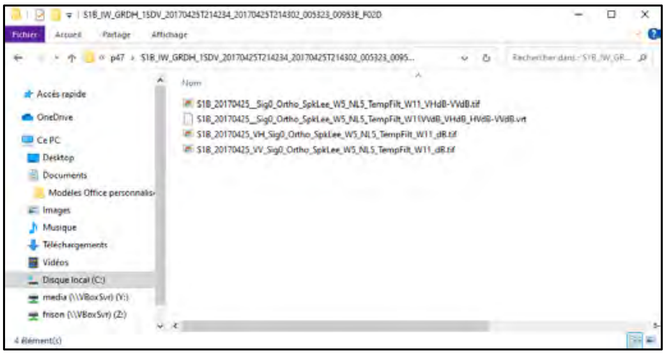
[3] J. S. Lee, "Speckle suppression and analysis for synthetic aperture radar images," Opt. Eng., vol. 25, no. 5, pp. 636–643, May 1986.
[4] J. S. Lee, J. H. Wen, T. I. Ainsworth, K. S. Chen, A. J. Chen, "Improved Sigma Filter for SpeckleFiltering of SAR Imagery", IEEE Trans. Geosc. Rem. Sens., vol. 37, n°1, pp 202-213, Jan. 2009.

beyond 30 radar acquisitions over the same area, the spatio-teporal filter alone is more than sufficient.

Once the processing is done, the directory ***data\S1\ Filt_Spck\p47*** is created (see relative orbit), including the following tree structure:



Each of the 4 folders is composed of similar files :



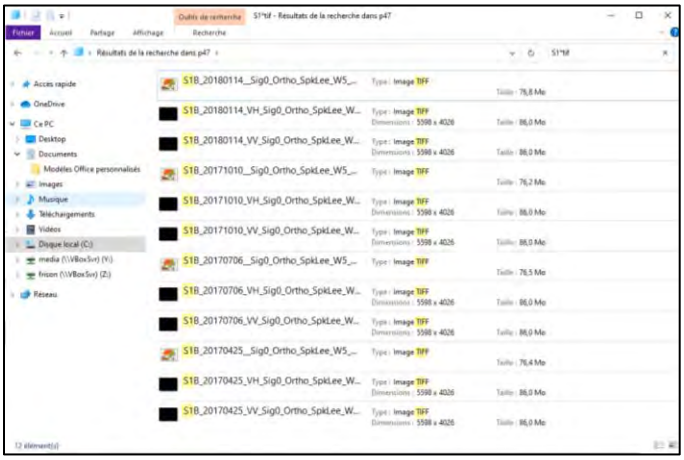
3 ***.tif*** files corresponding to filterd images VV, VH polarisations as well as the polarisation ratio VH/VV.

As well the virtual file
S1B_20170425__Sig0_Ortho_SpkLee_W5_NL5_TempFilt_W11VVdB_VHdB_HVdB-VVdB.vrt
allowing to display the 3 polarizations as a color composite image, with the color composition similar to the one used for the temporal average § I.1.1.

I.1.3: Creating a multi-date, multi-polarization filtered radar image

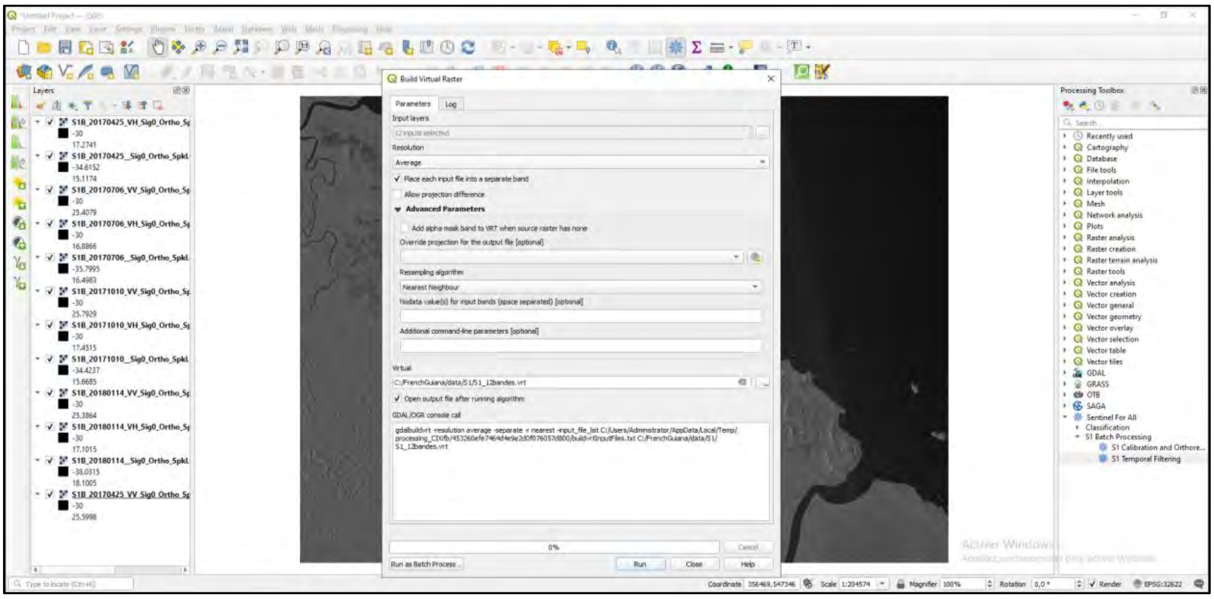
Finally, we need to create the image that will be used to classify the vegetation. This is a 12-bands image, consisting of 4 dates x 3 polarisations. We will create a virtual image:

1) Close all open layers in qgis.
2) In a file manager, go to the directory ***data\S1\ Filt_Spck\p47*** and type in the search bar at the top right ***S1*.tif***.



Select the 12 files and drag and drop them into QGIS

Then ***Raster → miscellaneous → Build Virtual Raster*** and fill in the fields as follow :



Input layers : Select the 12***.tif*** files

Virtual : ***data\S1\S1_12bandes.vrt***

Don't forget to check the box ***place each input file in a separate band.***

You have just created a Sentinel 1 virtual radar image containing 12 bands which you will use for vegetation classification.

I. 2. SENTINEL-2

The Sentine -2 data
data\S2\zip\S2B_MSIL1C_20180206T140049_N0206_R067_T22NCL_20180206T172306.zip
has been doanload from https://earthexplorer.usgs.gov/. It corresponds to the level 1C product. These products are already orthorectified and contain the following 13 spectral bands

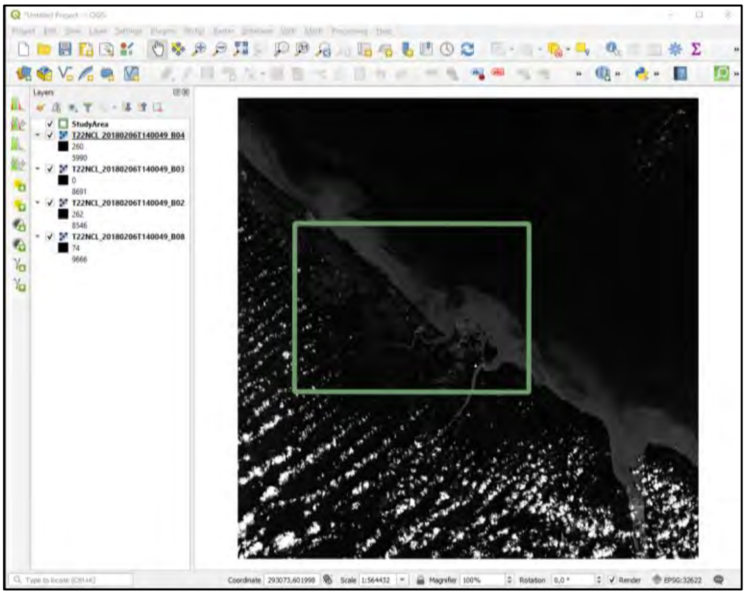
Band	Band range (nm) Band Center (nm)	Spatial resolution (m)	Purpose in L2 processing context
B1	433-453 / 443	20	Atmospheric Correction
B2	458-523 / 490	10	Blue-Sensitive to Vegetation Aerosol Scattering
B3	543-578 / 560	10	Gree-Green peak, sensitive to total chlorophyll in vegetation
B4	650-680 / 665	10	Red-Max chlorophyll absorption
B5	698-713 / 705	20	Vegetation classification
B6	734-748 / 740	20	Vegetation classification
B7	765-785 / 783	20	Vegetation classification
B8	785-900 / 842	10	Vegetation classification
B8a	855-875 / 865	20	NIR-Used for water vapour absorption reference
B9	930-950 / 945	60	Water vapour absorption atmospheric correction
B10	1365-1385 / 1375	60	Detection of thin cirrus for atmospheric correction
B11	1565-1655 / 1610	20	MIR-Snow / ice / cloud discrimination / Soils detection
B12	2100-2280 / 2190	20	AOT determination

We will only process the 4 bands at 10m spatial resolution (B2-4 +B8) to which we will add an NDVI channel.

1.2.1 Creating a 5 bands image over the study area

a) Loading the images into QGIS

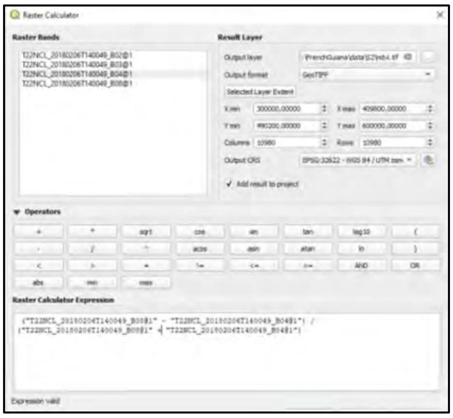
Décompress the file
S2B_MSIL1C_20180206T140049_N0206_R067_T22NCL_20180206T172306.zip in the same folder
Then in a file manager, go to the repertory
data\S2\zip\S2B_MSIL1C_20180206T140049_N0206_R067_T22NCL_20180206T172306.SAFE\GRANULE\L1C_T22NCL_A004812_20180206T140045\IMG_DATA then drag and drop the 4 bands ***T22NCL_20180206T140049_BXX.jp2*** (***XX*** for ***02***, ***03***, ***04*** et ***08***) into QGIS.



The 4 Sentinel-2 bands loaded into QGIS as well as the polygon delimiting the study area. They are already in the geographic projection UTM zone 22 N - EPSG: 32622.

b) ***Creating the NDVI band*** **(**$NDVI = \frac{\rho_{IR}-\rho_R}{\rho_{IR}+\rho_R}$, where ρ_{IR} and ρ_R are the reflectances of Red and InfraRed channels respectively).

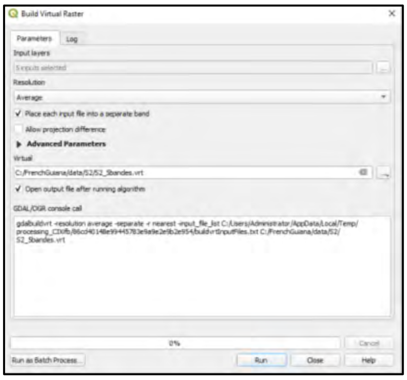
For this ***Raster → Raster Calculator*** then fill in the fields as folow :



In the ***output layer*** field specify : ***data\S2\ndvi.tif***

c) Creating a 5 bands image (B2, B3, B4, B8 and NDVI) cropped over the study area

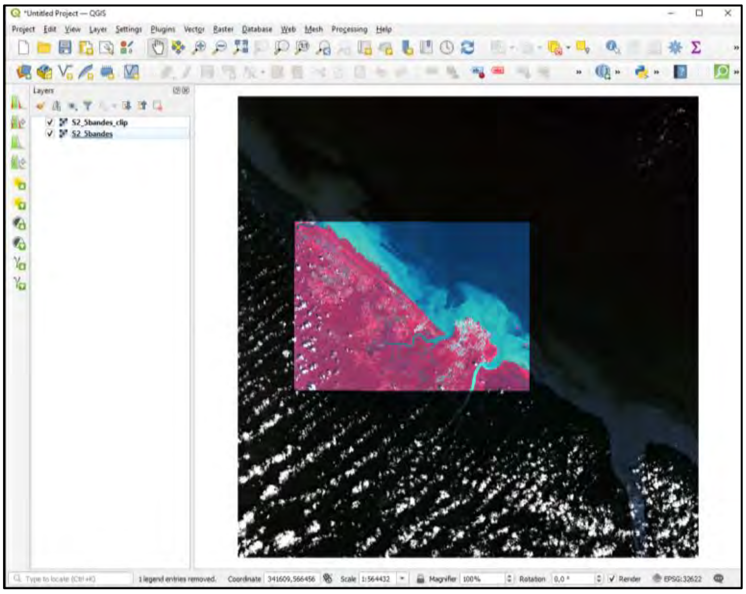
Raster → miscellaneous → Build Virtual Raster



Store this virtual imagein ***data\S2\S2_5bandes.vrt***

Then ***Raster → Extraction → Clip Raster by Mask Layer*** and fill in the fields as follow :

In ***clipped*** (the output image) specify ***data\S2\S2_5bandes_clip.tif***



Color composite of the S2_5bandes_clip.tif image overlapped to the original Sentinel-2 scene.

II. CLASSIFICATIONS OF THE DIFFERENT TYPES OF VEGETATION

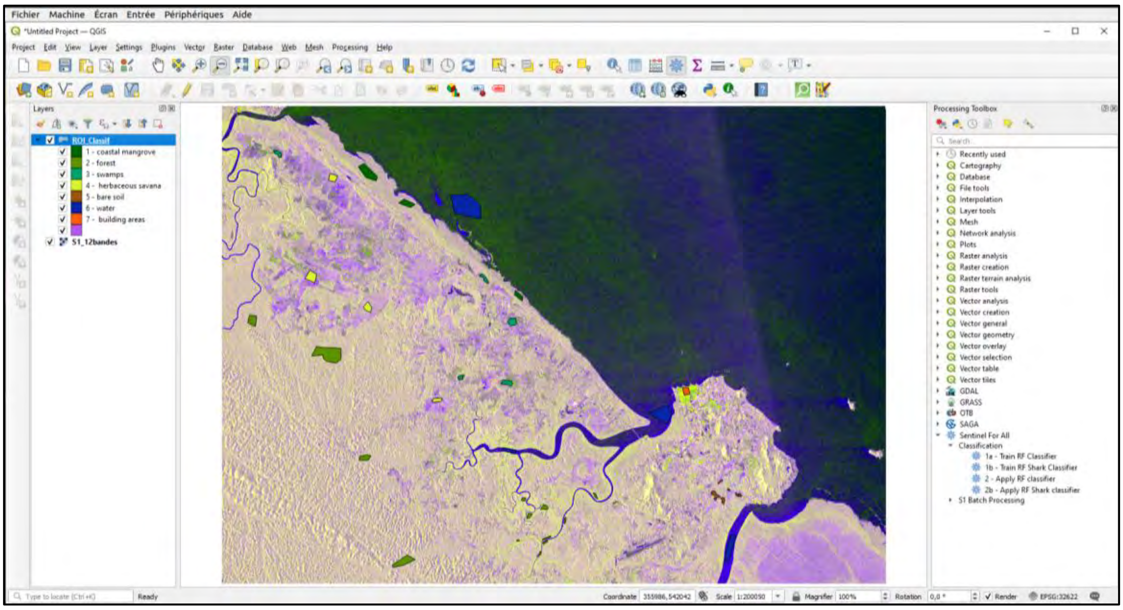
The Sentinel-1 and Sentinel-2 data are now ready to be used in a classification algorithm. We used the RANDOM FOREST supervised algorithm [5,6], as it gives overall good results with limited computing time, even when a large number of bands are involved.

The training and validation polygons have been defined from *in situ* surveys, coupled to Google Earth visual interpretation as well as the Sentinel-1 temporal average image ***VVdB_VHdB_VHdB-VVdB_Temporal_Average_20170425_20180114*** (cf. § I. 1.1).

II.1. classification of the Sentinel-1 data

Within QGIS, load the ***data\S1\S1_12bandes.vrt*** image

Load as well the layer ***classification\vector\ROI_Classif.shp*** containing the training / validation polygons. You can see the 7 vegetation types that we are trying to map.



a) Training the model

In the ***Processing*** panel : ***Sentinel for All → Classification → Train RF Classifier***

---

[5] https://fr.wikipedia.org/wiki/Forêt_d'arbres_décisionnels
[6] https://en.wikipedia.org/wiki/Random_forest

The output model is a text file that you will name ***classification\model_S1_12bandes.txt***

You will analyse in the *Log* tab of the *Train RF classifier* window the statistics describing the performance of each class :

$$Precision = \frac{Number\ of\ pixels\ correctly\ affected\ to\ the\ class}{Total\ number\ of\ pixels\ affected\ to\ the\ class}$$

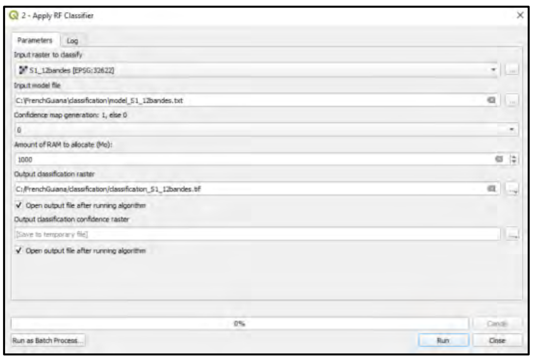
$$Recall = \frac{Number\ of\ pixels\ correctly\ affected\ to\ the\ class}{Total\ Number\ of\ effective\ pixels\ of\ the\ class}$$

$$F - score = \left(\frac{Precision^{-1} + Recall^{-1}}{2}\right)^{-1}$$

The lowest statistics (around 78%) are observed for the Accuracy of class 1 (Mangroves) and the Recall of class 2 (Forests), leading to an F-score of around 85%. If the classification were to be improved, the training polygons for these two classes would have to be refined. The highest statistics are, as is often the case with radar, those of classes 6 (Water) and 7 (Built-up areas) with F-scores of 99%. Generally speaking, (this is obviously relative and depends on the expectations of the problem) we can estimate that the results are not acceptable (training polygons not representative of the class sought) when one of the statistics < 70%.

b) Application of the model to the whole image

The model ***classification\model_S1_12bandes.txt*** was created from the training polygons, so we will apply it to the whole image. For this, ***Sentinel for All → Classification → Apply RF Classifier***



The output classification image is the file ***classification/classification_S1_12bandes.tif***.

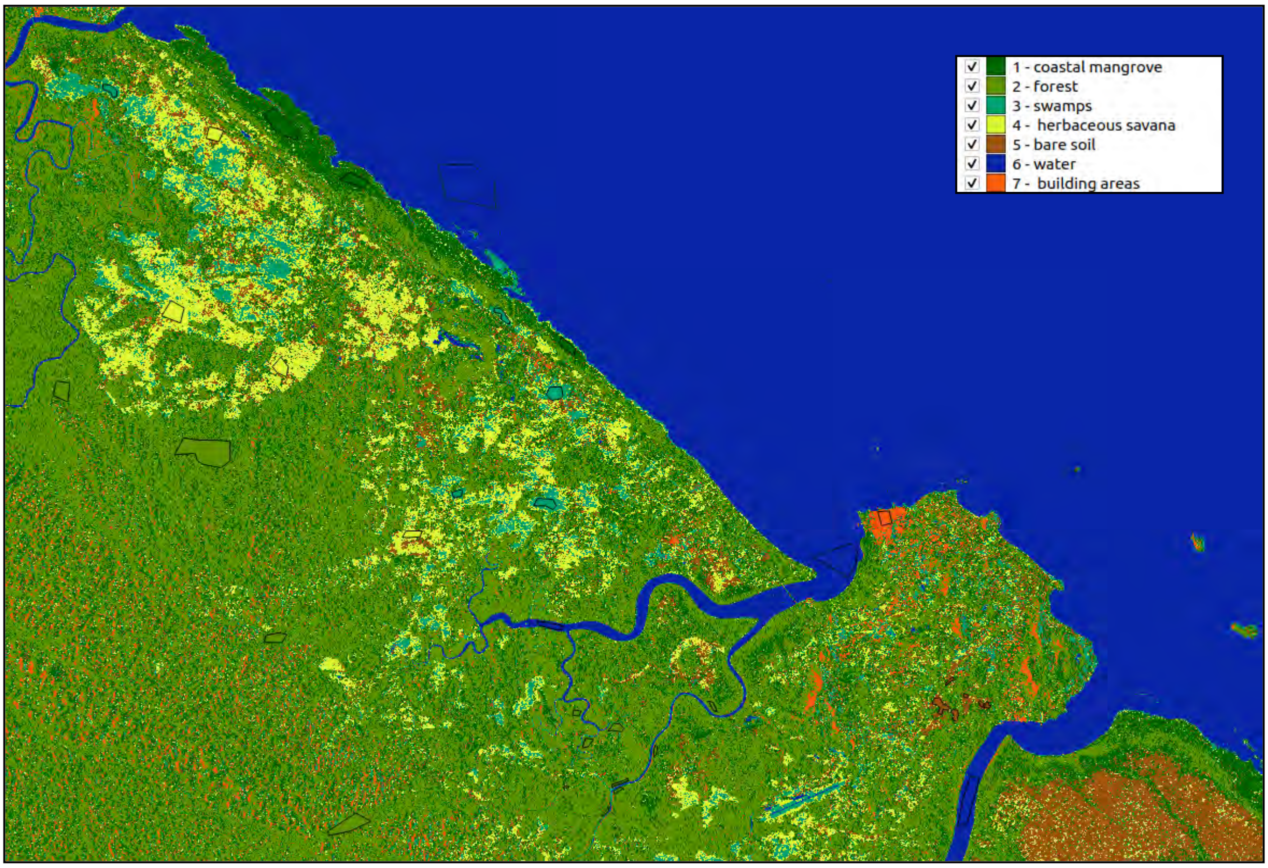
Loa dit into QGIS.

Chargez le dans QGIS. It remains to apply the same colours to the classes as those of the training polygons. To do this, right click on the ***classification_S1_12bandes***, then ***Propriétés*** and click on the *Symbology* tab in the left-hand panel. Click on the *Style* button at the bottom left of the window then load the style ***classification/colours_classif_7cl_tif.qml***.

Comment :

We can remove the isolated points using ***Raster → Analysis → Sieve…*** that we apply to the image ***classification_S1_12bandes*** with a threshold value of *20*, and by using ***8-connectedness***.

You will notice that this classification is far from perfect. There are many pixels of the forest class that are classified as mangroves. It is therefore necessary to refine the choice of mangrove polygons and redo the classification model. You can also see the effects of terrain, as the signal is high into the 3 polarisations for the slopes facing the radar. This is the reason why they are classified as built-up areas. This effect can be remedied by applying a slope mask.

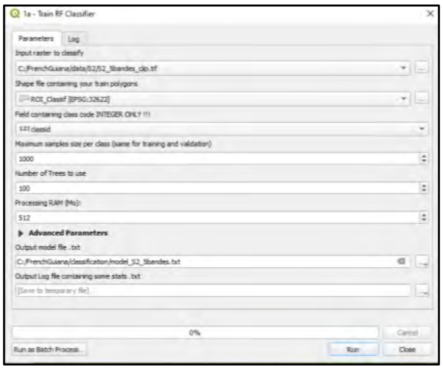


Classification result od Sentinel-1 images (4 dates x 3 polarizations).

II.2 classification of Sentinel-2 data

a) Training the model

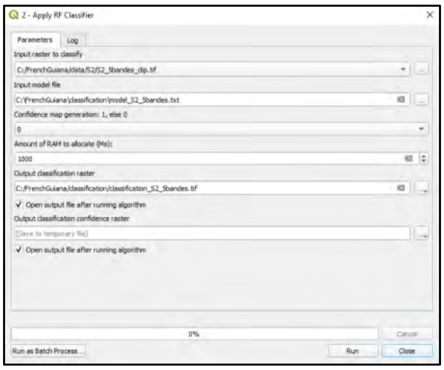
Same as step II.1.a) by replacing the file ***S1_12bandes.vrt*** by ***data/S2/S2_5bandes_clip.tif*** (cf § I.2.1) and by specifying for the output file **classification/model_S2_5bandes.txt**



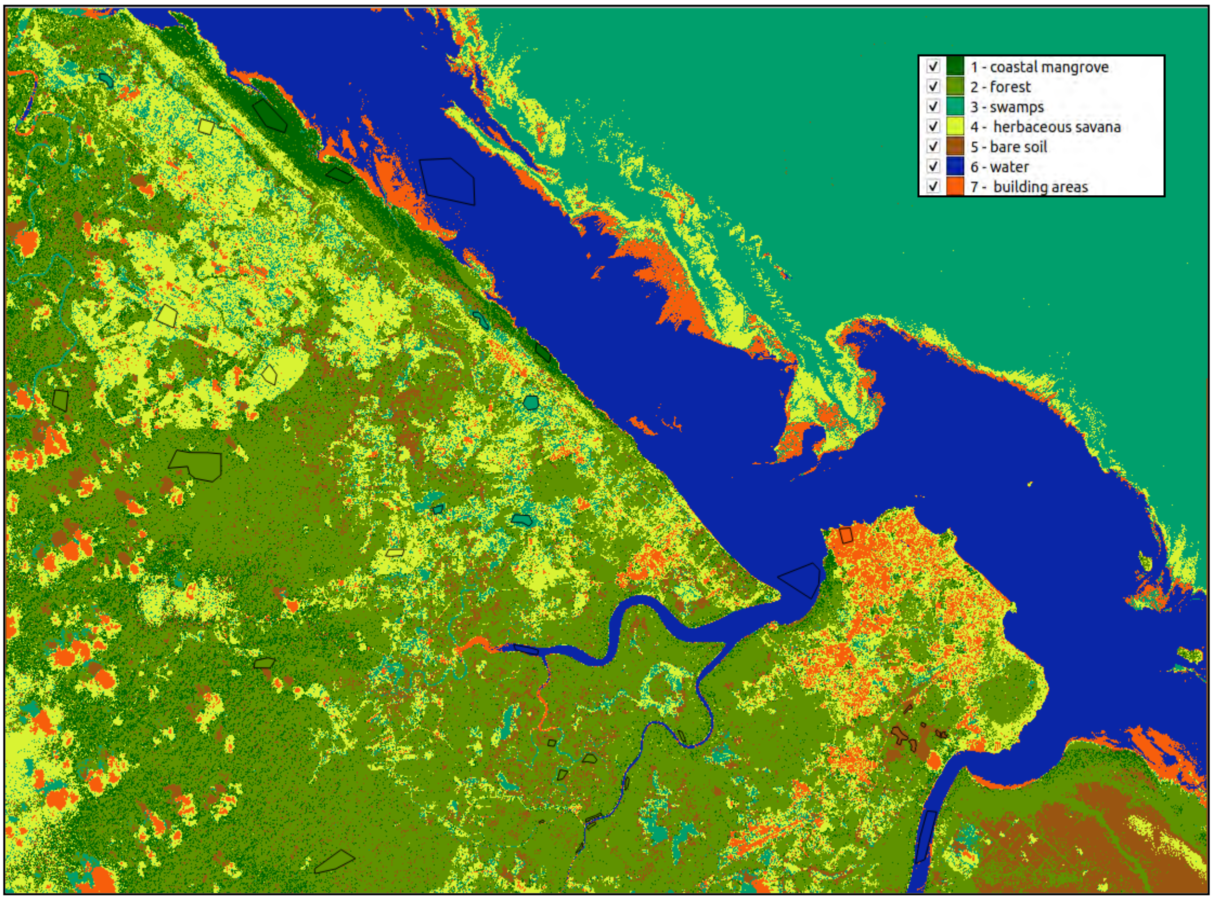
You can note the higher performance statistics (> 80%) than those obtained during the creation of the Sentinel-1 model ***model_S1_12bandes.txt***).

b) Application of the model to the image

Same as step II.1.b) by replacing the input file *S1_12bandes.vrt* by ***data/S2/S2_5bandes_clip.tif***, the classification model *model_S1_12bandes.txt* by ***model_S2_5bandes.txt***, and by specifying for the ouput file **classification/classification_S2_5bandes.tif**



The result (below) clearly shows the differences between the optical and radar data. The effects of terrain on the Sentinel-2 classification are not very noticeable. It also allows better discrimination between mangrove and tropical forest. But.... Obviously, clouds and their shadows have a direct and very disturbing effect on the classification.



Result of the classification from Sentinel-2 data.

III Classification merging Sentinel-1 and Sentinel-2 :

You can repeat the process to create a file containing the 12 (filtered) Sentinel-1 bands and the 5 Sentinel-2 bands. You will have to build it from the 17 bands separately (unfortunately you cannot create this file from *S1_12bands.vrt* and *S2_5bands.tif*).

Vous pouvez recommencer les manipulations pour créer un fichier contenant les 12 bandes Sentinel-1 (filtrées) et les 5 bandes Sentinel-2. Il faudra le construire a partir des bandes de 17 bandes séparément (vous ne pouvez malheureusement pas créer ce fichier à partir de *S1_12bandes.vrt* et *S2_5bandes.tif*.

It's up to you !!!!!