Crops classification with Sentinel-1, PALSAR and Sentinel-2 data

P.-L. Frison, C. Lardeux, A. Castro

DATA & BACKGROUND INFORMATION

The goal is to use radar (Sentinel-1 and ALOS-PALSAR) and optical (Sentinel-2) data, combined or separately to classify different crop types over an area located near Hradek Kràlové, in the Czech Republic. First, data pre-processing will be perfomed, an then the crop types will be estimated by classification of the data, based on the Random Forest algorithm.

The data that are used for validation and training are derived from the CZECH AGRI project (collaboration of JRC, ESA and SZIF). This project used Sentinel-1 and Sentinel-2 data to map the different crop types over the whole country.

The different data have been selected for the period extending from October 2015 – September 2016, as detailed in the Table hereafter:

	Acquisition dates	Spatial resolution	wavelengths	Polarisation	Product Level
Sentinel-1	Year 2016: 03/09 – 04/14 – 05/20 – 07/07 – 08/24	~ 20 m	5.6 cm (C Band)	VV/VH	GRDH
PALSAR	2016/08/08	25 m	24 cm L Band	HH/VH	25 m Mosaic
Sentinel-2	2015: 09/19 2016: 03/17 – 03/27 – 08/04	4 bands: 10 m 6 bands: 20 m	430 nm – 1600 nm		L1C

In a first part, the data are processed in order to be suitable to be used as input for classification algorithm.

The second part is dedicated to the classification of the study area.

I. PRE-PROCESSING OF THE DATA

Each radar acquisition is realized in dual polarization. Three bands per acquisition are used for the classification: VV (or HH for PALSAR), VH, and their ratio VV/VH (or HH/VH for PALSAR).

Among the 13 bands per acquisition of Sentinel-2 data, only the 10 at 10 m or 20 m of spatial resolution are used. A Resampling at 10m is then required for the 20 m spatial resolution bands.

I. 1. Sentinel-1:

The Sentinel-1A data have been acquired in IW mode, and have been downloaded from the ESA Scientific Data Hub https://scihub.copernicus.eu/ in GRDH products.

They have to be calibrated (from DN to Radar Backscattering Coefficient s0) and orthorectified (from the image geometry to a geographical projection) over a geographical subset defined by the ***study_area.shp*** vector layer. The resulting images are stacked (the same pixel of each image will correspond to the same location).

Then, a filter is applied in order to reduce the Speckle noise, inherent to radar images.

These pre-processing can be perfomed either with SNAP software, or with a customized version of QGIS, integrating the Orfeo Toolbox (OTB) software. Orfeo Toolbox is an OpenSource software, available at https://www.orfeo-toolbox.org/, especially dedicated to remote sensing data processing. It has the advantage to allow the processing of big amount of data without any specific computers configuration (4 Gb of RAM is enough).

I. 1. 1: with SNAP

I. 1.1.1: Calibration and orthorectification over a subset

Open the five S1 data files (.zip) located in ***data/S1/zip*** (*File → Open Product*)

Open the *Graph Builder* (*Tools → Graph Builder*) and load (File → Load Graph) the ***OrthoCal_Subset_Czech.xml*** graph. Look at the different tabs to see the different parameters needed to Calibrate, Orthorectify, and extract a subset area. In particular, verify that in the ***Terrain Correction*** tab, the ***external DEM*** is selected, pointing on ***data/DEM/tif/SRTM.tif***.

This DEM is derived from simultaneous interferometric radar data acquired during the SRTM mission. Its spatial resolution is 3 arc sec (~90 m). Tiles downloaded from http://step.esa.int/auxdata/dem/SRTMGL1/ are stored in the folder ***data/DEM/orig/3sec*** and have been aggregated in one mosaic (tif format) under QGIS.

Open the *Batch Processing* graph (*Tool → Batch Processing*), then open the ***code/snap/OrthoCal_Subset_Czech.xml*** graph (*File → Load Graph*).

In the ***I/O Parameters*** tab:

Add the S1 opened File (***Add opened*** icon on the right part)
Specify the Output directory ***data/S1/Snap_Processing/OrthoCal_Snap*** (bottom part)

Close the 5 first opened .zip files.

I. 1.1.2 Speckle filtering

a) In order to apply a spatio-temporal speckle filtering, it is necessary to stack the data: ***Radar → Coregistration***.
Then, you can close the five first products already open to keep only the ***....._stack*** file open.

b) To apply a spatio-temporal speckle filtering: ***Radar → Speckle Filtering → Multi-temporal Speckle Filter***

In the ***Processing parameters*** tab, set the Number of looks to 4. Then ***run***.

c) Destack the filtered data: ***Radar → Coregistration → Stack Tools → Stack Split***

In the Stack Split tab, enter as *Target folder* ***data/S1/Snap_Processing*** and as File *Name* ***Destack***. Then *Run*

Close the ***....._stack file*** and open the *five* **....dim** files located in ***data/S1/Snap_Processing/Destack*** folder.

d) Then in order to produce the three bands (VV, VH, VV/VH) for each acquisition, open the *Batch Processing graph (Tools → Batch rocessing).* Load (*File → Load Graph*) the graph ***codes/3bands_dB_graph.xml***.

In the ***I/O Parameters*** tab:

Add the opened File (***Add opened*** icon on the right part)

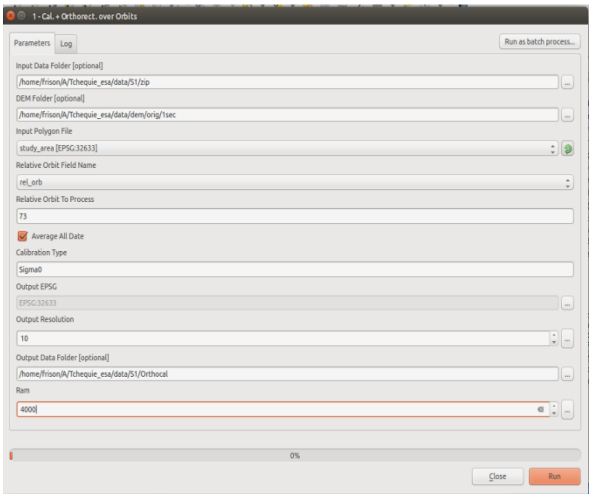
Specify the Output directory ***data/S1/Snap_Processing/3Bands_Snap*** (bottom part)

The five resulting files contain each the three desired bands, ready to be processed for classification algorithms.

I. 1. 2: with QGIS/OTB scripts

I. 1.2.1: Calibration and orthorectification over a subset

First, open in QGIS the vector layer (*Layer → Add Layer → Add Vector Layer*) ***/data/vector/study_area.shp***.
Then, in the *Processing Toolbox* panel *click on Scripts → Sentinel-1 IW GRD Batch Processing →* ***1 – Cal. + Orthorect. Over orbits.***
Fulfil the fields as follow:

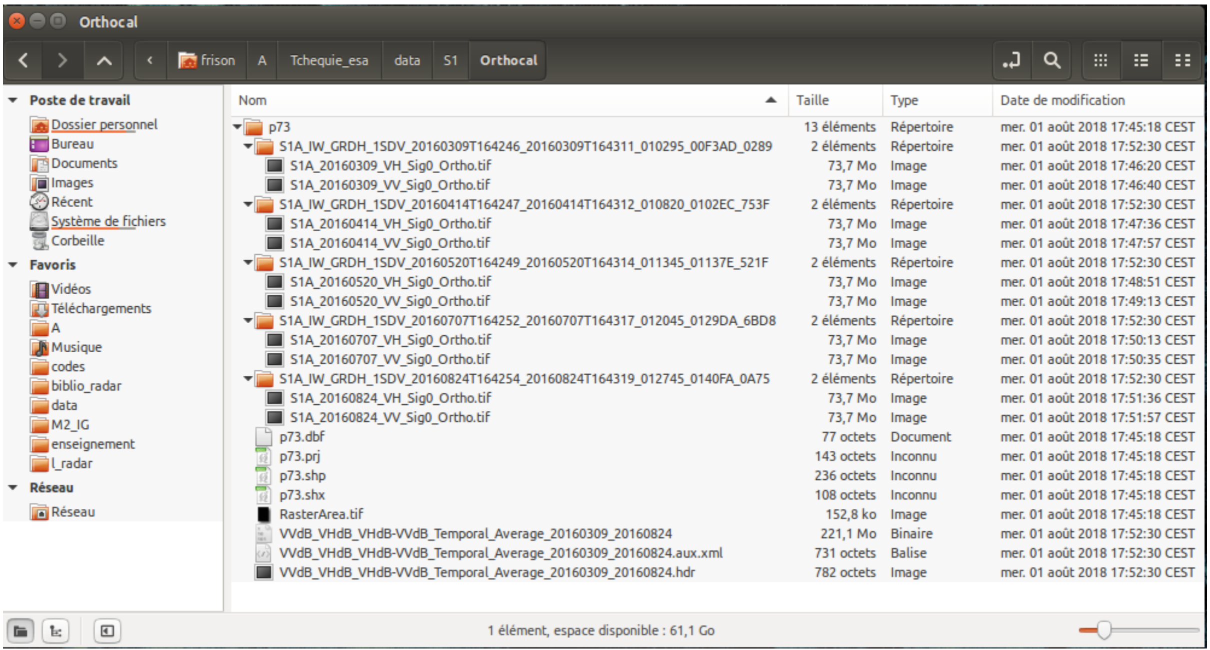


Input folder: data/S1/zip (!!! Must contain only the 5 .zip S1 files)
DEM Folder: data/DEM/orig/1sec (!!! Must contain only the .tif srtm files. Better te be in geographic projection EPSG 4326)

This DEM is derived from simultaneous interferometric radar data acquired during the SRTM mission. Its spatial resolution is 1 arc sec (~ 30 m). Tiles downloaded from https://earthexplorer.usgs.gov/ are stored in the folder ***data/DEM/orig/1sec***. They don't need to have been preliminary aggregated as a mosaic.

Input Polygon File: data/vectors/study_area.shp
Relative Orbit to Process: must contain the value of the field ***« rel_orb »*** of the « study_area » vector layer attribute table
Output Data Folder: data/S1/Orthocal (!!! Must be an empty folder)

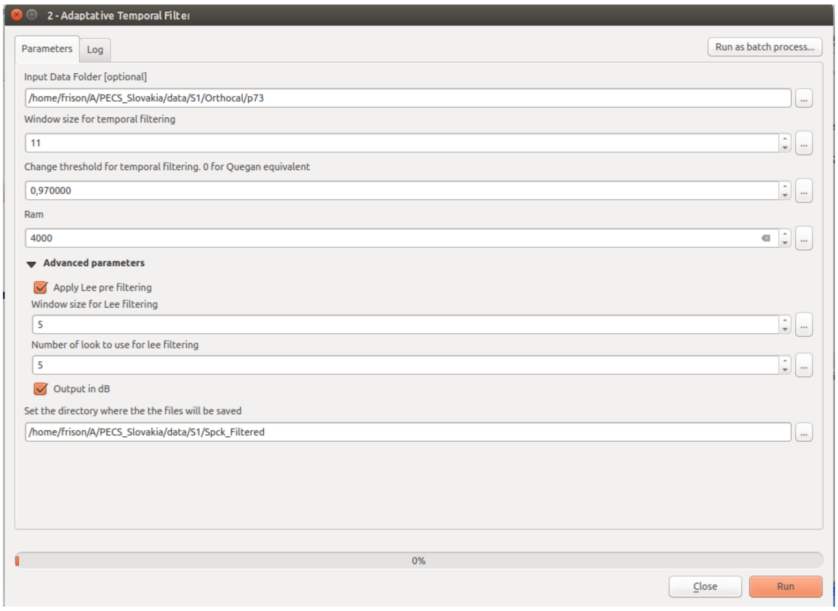
The arborescence of the output folder ***data/S1/Orthocal*** is as follow:



Each subfolder contains 3 ***tif files*** (VV, VH, and VH/VV), calibrated, orthorectified, and cropped over the region of interest (study_area.shp). All the tif files have exactly the same number of lines and columns, *i.e.* ready to be overlaid. In addition, a ***.vrt file*** allowing their representation in color composite image in QGIS.

I.1.2.2: Speckle Filtering

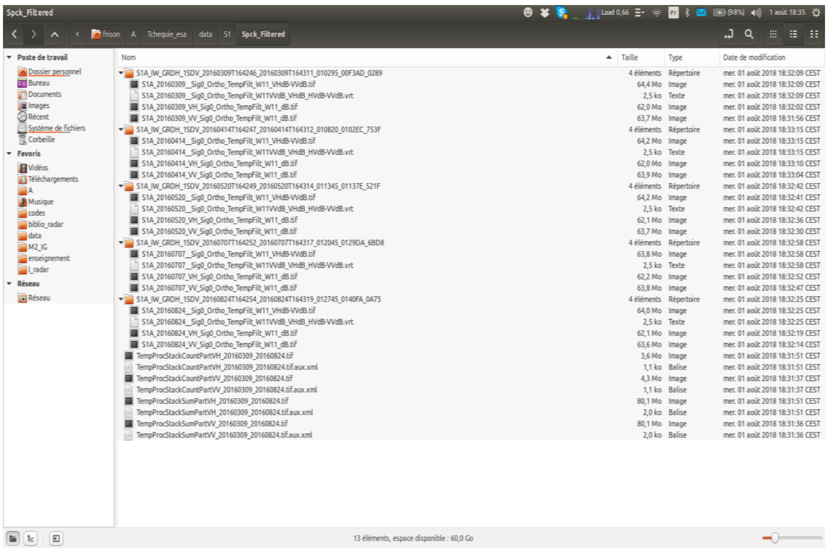
In the *Processing Toolbox* panel, click on
Scripts → Sentinel-1 IW GRD Batch Processing → ***2 – Adaptative Temporal Filter.***
Fulfil the fields as follow:



Input folder: data/S1/OrthoCal
Output Folder: data/S1/Spck_Filtered (!!! Must be an empty folder)

As there are only 5 S1 acquisitions, an additional filtering is performed. It is the Lee filter, operating only in the spatial domain, which is applied before the spatio-temporal filtering. This pre-filtering step is not necessary when large amount of acquisitions are processed (such as 60 over one year, when S1A and S1B are availavble).

The arborescence of the output folder ***data/S1/Spck_Filtered*** is similar to the one of ***Orthocal*** folder:



I. 2. ALOS PALSAR

The PALSAR sensor onboard the ALOS Japanese satellite acquires radar data at L Band (λ = 24 cm). The JAXA have processed yearly global mosaic for the years 2007-2010 and 2015-2017 at HH and HV polarizations. They are freely available at http://www.eorc.jaxa.jp/ALOS/en/palsar_fnf/fnf_index.htm. The study area of this training is at the intersection of four tiles which are stored in the folder ***data/PALSAR/zip***. They are already orthorectified, and need to be:

- Agregated and cropped over the study area
- Calibrated (according the relation σ^0 (dB) = 10 . $\log_{10}$(DN^2) -83)
- Filtered to reduce the speckle.

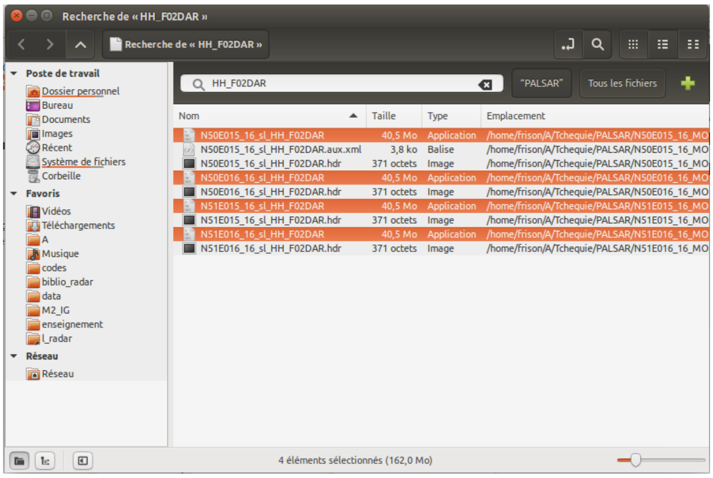
As for Sentinel-1 data, 3 bands will be derived: HH, VH, and the ration HH/VH. These processing will be done with QGIS.

They have to be calibrated (from DN to Radar Backscattering Coefficient s0) and orthorectified (from the image geometry to a geographical projection) over a geographical subset defined by the ***study_area.shp*** vector layer. The resulting images are stacked (the same pixel of each image will correspond to the same location).

I. 1.2.1: Agregating the tiles and copping over the study area

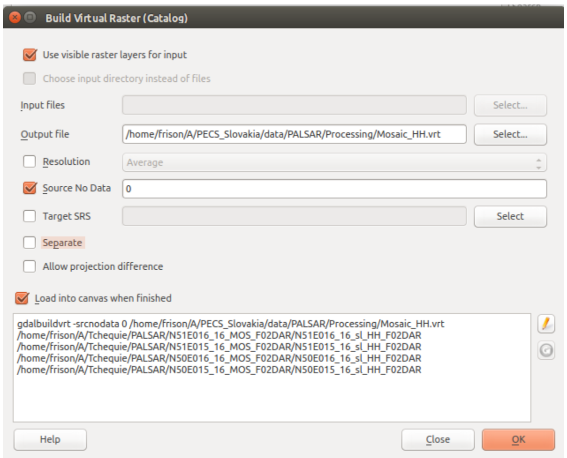
Dezip the four .zip files within the folder ***data/PALSAR/zip*** in the folder ***data/PALSAR/processing***

To open in QGIS the four acquisitions in HH polarization, open a ***nautilus*** file manager (in a terminal, type nautilus). Then in the nautilus window, go to the ***data/PALSAR/processing*** folder. Then click on the magnifying glass icon and type *HH_F02DAR*.



a) *Mosaicing the four tiles*

The different ***...HH_F02DAR....*** located in the subfolders appear. Select the four files without any extension. Drag and drop into the *QGIS layers panel*. To aggregate them into a mosaic, *Raster → Miscellaneous → Build Virtual Raster.*

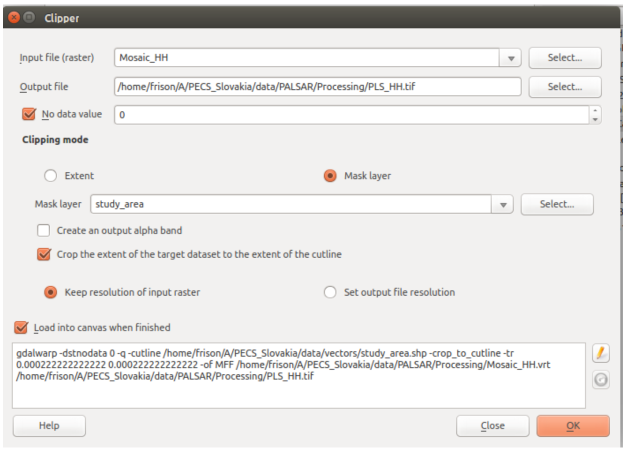


For the output file, select ***data/PALSAR/Processing/Mosaic_HH.vrt***, and fulfil as above. Note the GDAL command in the bottom frame that this window allows to construct, and that is run to build the mosaic. Copy and paste this command in a text editor.

In the text editor change the "HH" string by "HV" in the whole command. Then copy and paste the new command in a terminal. The ***data/PALSAR/Processing/Mosaic_HV.vrt*** has been created.

b) Crop the mosaic on the study area

To crop the ***Mosaic_HH.vrt*** on the ***study_area.shp*** extent: *Raster* → *Extraction* → *Clipper*, and fulfil the different fields as follow:



Note the GDAL command in the bottom frame that this window allows to construct, and that is run to build the mosaic. Copy and paste this command in the same text editor than the last step.

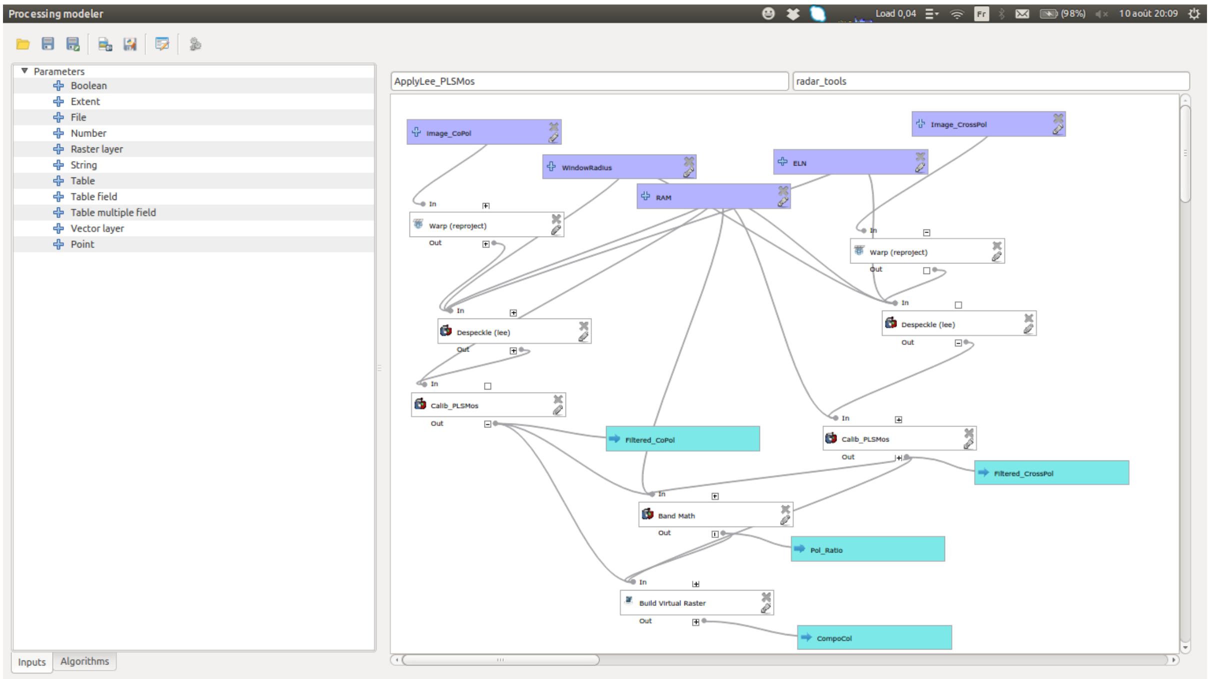
To crop the ***Mosaic_HV.vrt***, replace in the text editor the "HH" string by "HV" in the whole command. Then copy and paste the new command in a terminal. The ***data/PALSAR/Processing/PLS_HV.tif*** has been created.

c) Calibration, Speckle Filtering, and creation of the 3 filtered PALSAR BANDS (HH, HV, HH/HV)

In the *Processing toolbox* panel, right click on *Models* → *radar tools* → *ApplyLee_PLSMos* and have a look on the whole processing chain that is applied to the HH and HV images.

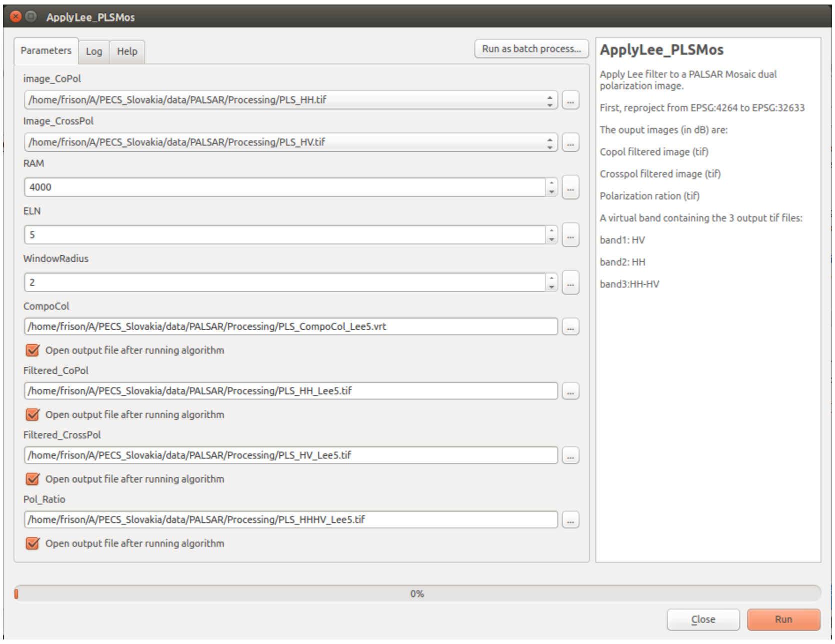
1) Reprojection in the EPSG 32633 geographical projection (usefull to be combined with S1 and S2 data)
2) Apply a Lee Filter

3) Calibrate the filtered data in dB
4) Creation of the HH/HV band
5) Build a virtual Band allowing to vizualize a Color composite of these 3 bands
6) Change the 25 m resolution of the three bands to 10 m (allowing the combination with S1 an S2)



Close this window, and click on *Models* → *radar tools* → *ApplyLee_PLSMos*

Fulfil the different fields as follow:



Be careful, in the CompoCol field, don't forget to specify vrt format

Change the 25 m resolution to 10m: in the *Layers Panel* right click on ***PLS_CompoCol_Lee5.vrt***.

→ *Save as*...

Output file: data/PALSAR/Output/PLS_CompoCol_Lee5_10m.tif
change ***Resolution Horizontal*** (and ***Vertical***) to 10.0

The last step is to split the 3 bands of the file ***PLS_CompoCol_Lee5_10m.tif***. To do so:
In the *Processing Toolbox* Panel, *Scripts* → *Split Raster bands in vrt format*.

The result is 3 bands ***PLS_CompoCol_Lee5_10m_b1.vrt (..._b2.vrt*** and ***..._b3.vrt***) located in the folder ***data/PALSAR/Output***

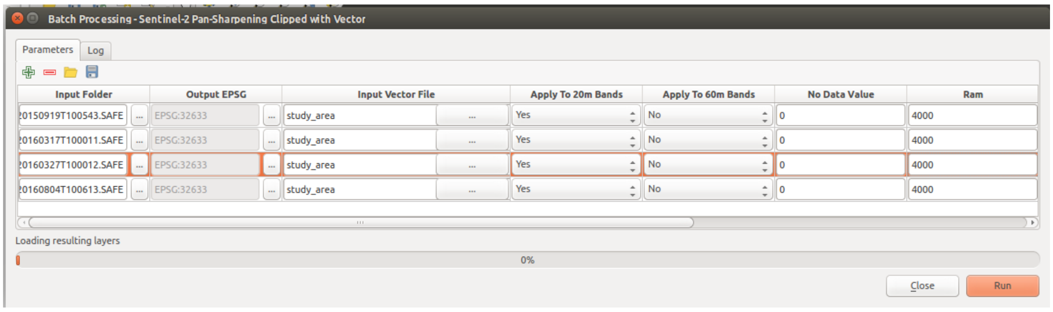
I. 3. SENTINEL-2

A whole Sentinel-2 scene occupies 6 Gb. It is the reason why the world has been divided into tiles, and the Sentinel-2 data has been splitted according to these tiles. As the S2 data used in this training correspond to the beginning of the mission, the corresponding tiles are not available on the ESA Scientific Data Hub. The 4 *.zip* files in the folder ***data/S2*** have been downloaded from the web site https://earthexplorer.usgs.gov/. They correspond to Level 1C products. The data are already orthorectified, and contain 13 bands as detailed hereafter:

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 Vegettion 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

1) unzip these four compressed files within the same folder.

2) The resampling at 10 m of the 20 m bands[1] is made with QGIS/OTB
In the *Processing Toolbox* panel, right click on *Scripts → Sentinel-2 PanSharpening with Clipped with Vector*. Fulfil the different fields as follow:



Input Folder: the 4 folders which name is finishing by ***.SAFE***
Output EPSG: 32633
Input Vector File: Study_area.shp
Output Raster: data/S2/Output/S2A_yyyymmdd where yyyy, mm, and dd are the year, month and day of the corresponding Input Folder.

The resulting ***.tif*** files contain 10 bands, with 10 m of pixel size cropped over the study area. The denominations ***_b1***,... ***_b10*** correspond to the original bands denominated b2,... b8, b8a, b11,b12, respectively.

As for PALSAR, split the 10 bands of each of the four S2 resampled acquisition:
In the *Processing Toolbox* Panel, *Scripts → Split Raster bands in vrt format*.

The result is 10 bands ***data/S2/Output/S2A_yyyymmdd_b1.vrt (..._b2.vrt*** ... ***..._b10.vrt***).

[1] The 3 60 m bands (B1, B9 and B10) are not considered because they are especially dedicated for the characterization of the atmosphere?

II. *Classification with the processed data*

The classification of the crop types is based on the Random Forest algorithm. As it is a supervised classification a preliminary step is to ***create Polygons of Interests***. Pixels are randomly chosen by the algorithm for the training step, and others for the performance estimation (validation step). Once this preliminary stage is performed, the classification is made according the following steps:

1) ***Create a virtual file (.vrt)*** containing the different bands that are taken into account.
2) ***Create a classification model*** based on the training polygons
3) ***Apply the model*** to the whole study area.

0) Preliminary step: Create the polygons of interest

First of all, remove all the layers that can be opened in QGIS.

Open the ***data/vectors/crops_layer.shp*** (*Layer* → *Add Layer* → *Add Vector Layer*). Open the Attribute Table: the class indice corresponds to the filed *CVM*.

The wanted polygon of interests will be the result of the selection of some of these numerous polygons.

Click on the icon *Select Feature*, and select different polygons (multiple selection is obtained with *CTRL+click*). The goal is to select multiple polygons in each class (between 4 to 10 per class), with same order of cumulative areas. Control with satellite data, that for each class, the selected polygons have quite homogeneous radiometric properties.

Save the selected polygons in ***/classif/roi_classif.shp***. To do so, right click on *crop_layer* → *save as..... !!! Don't forget to click on the checkbox:* ***Save only the selected features***

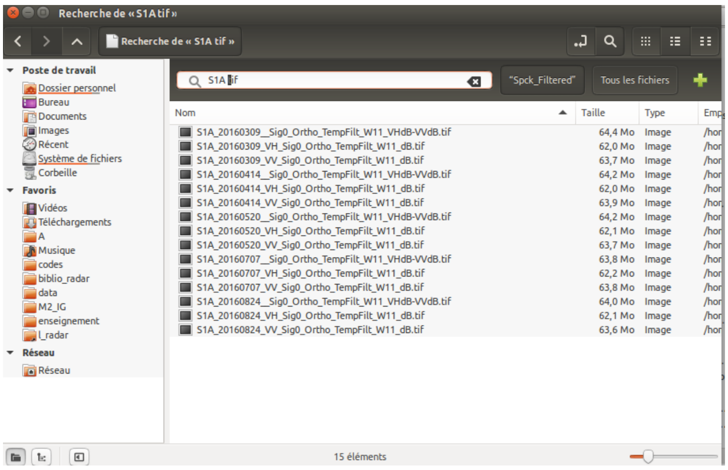
Then, you can add a class named ***forest*** with polygons easy to define.

1) Create a virtual file

Open in QGIS all the individual bands (either ***.tif*** or ***.vrt*** format) you want to consider.

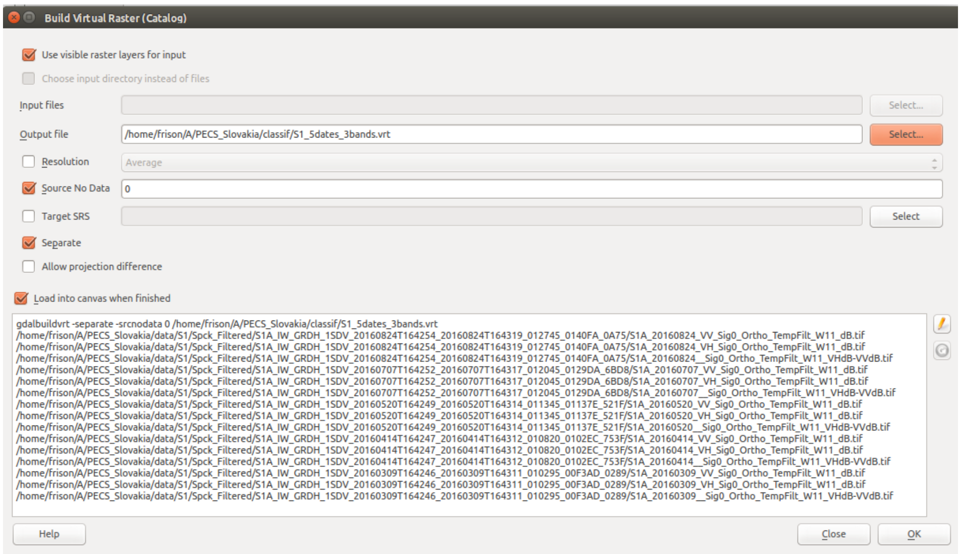
Build a virtual file containing all these bands.

For example, to perform a classification with the five S1 filtered acquisitions (*i. e.* 5 dates * 3 polarisation = 15 bands): with the nautilus file manager, go to data/S1/Spck_Filtered. Then, with the magnifying glass icon, type "S1A   tif".



Select all the 15 files, and drag and drop them into the QGIS *Layers Panel*.

Then *Raster* → *Micsellaneous* → *Build Virtual Raster* and fulfil the fields as follow:



!!!! Don't forget to click on Separate checkbox to differentiate the different bands

2) Create the classification model

Control that the Input Image you want to classify (.vrt) and the *roi_classif.shp* are open in QGIS. In the ***Processing Toolbox*** Panel, *Orfeo Toolbox→ 1 – Classification → 2 - Train Random Forest Image Classifier.* Fulfil the fields as follow:

Input image file: *classif/S1_5dates_3bands.vrt*
Input Region of Interest Vector file: *classif/roi_classif.shp*
Field name containing the classes id: *CVM*
Output Model: *classif/model_S1_5dates.txt*

Then ***Run***

In the ***Log*** tab, you can look at the performance of classification of each training class (Precision, Recall, F-Score) allowing to analyse if the classes you have defined are suitable for classification (> 85% means your class is quite well defined). If not, correct the concerned polygons.

3) Apply the model

In the ***Processing Toolbox*** Panel, *Orfeo Toolbox→ 1 – Classification →* 3 – Create Image Classification*.* Fulfil the fields as follow:

Input image file: *classif/S1_5dates_3bands.vrt*
Input Model FIle: *classif/model_S1_5dates.txt*
Output Image Classification: classif/classif_S1_5dates.tif

Load the adapted style for the classified image :
right clik on ***classif_S1_5dates.tif*** *→ Properties → Style → Load Style → classif/style_im_classif.qml*

You can remove the isolated pixels by appluing a post_processing like the Sieve algorithm (*Raster → Analysis → Sieve)* with a Threshold of 50 pixels.

☞ ***Compare the classifications results you obtain with S1, PALSAR, S2, S1+PALSAR, S1+S2, PALSAR+S2, S1+PALSAR+S2.***