Earth Observation Data Analysis

HOMEWORK02 - Surface mapping from MSI SENTINEL-2 data

Faculty of Information Engineering, Computer Science and Statistics,

Sapienza University, Rome, Italy

Prof. Frank S. Marzano

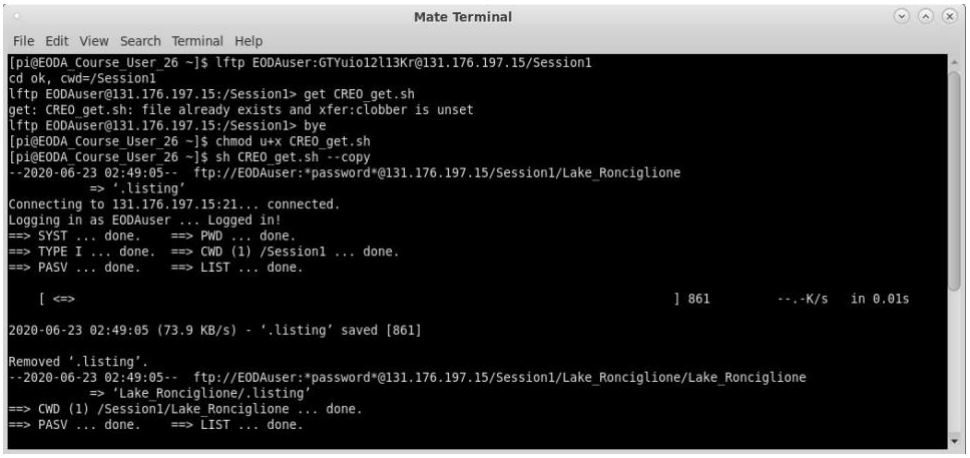
REZA POURRAHIM - 1859334

UWAMAHIRWE BONAVENTURE – 1850306

HAFIZ MUHAMMAD HASSAN - 1873829

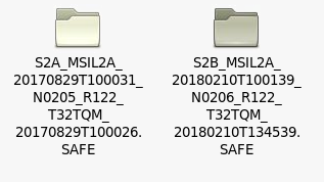
July 2020

1. First we start by downloading the script from shell connected to the ftp:



2. After download 5 files, we focus on 2 of them:

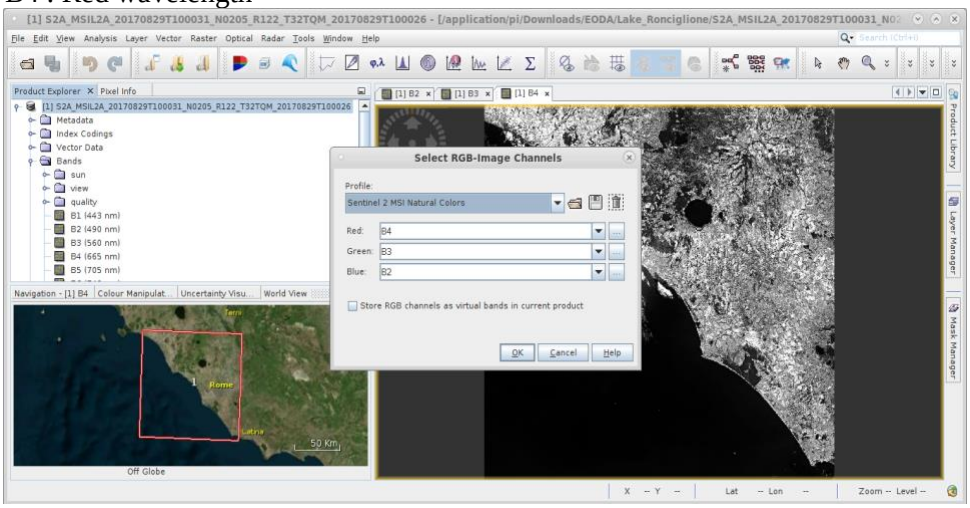
   1- **Summer(August)**:
   S2A_MSIL2A_20170829T100031_N0205_R122_T32TQM_20170829T100026
   2- **Winter(February):**
   S2B_MSIL2A_20180210T100139_N0206_R122_T32TQM_20180210T134539

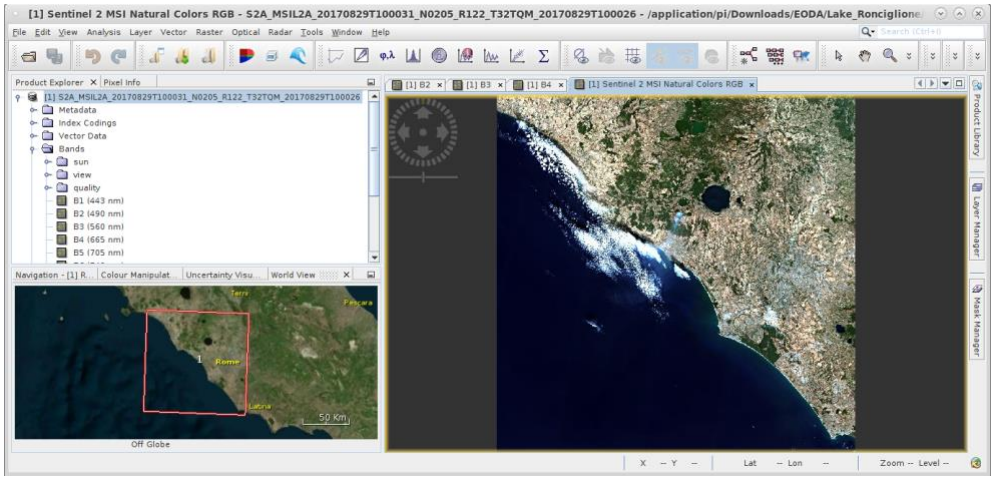


1. DATA EXPLORATION

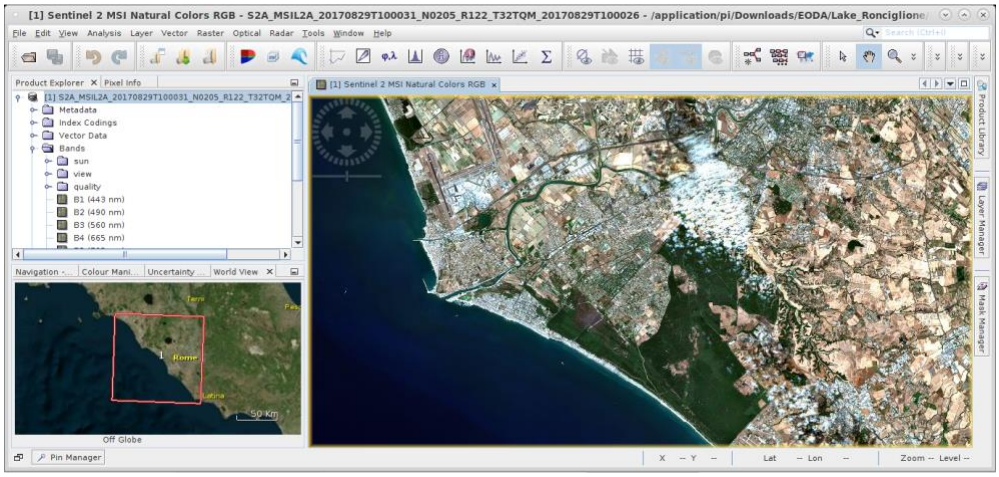
a) For both summer and winter MSI images, combine B2, B3, B4 channels to create an RGB image and zoom in to observe visible details.

1. For the **summer** image we have:
   B2 : Blue wavelength
   B3 : Green wavelength
   B4 : Red wavelength

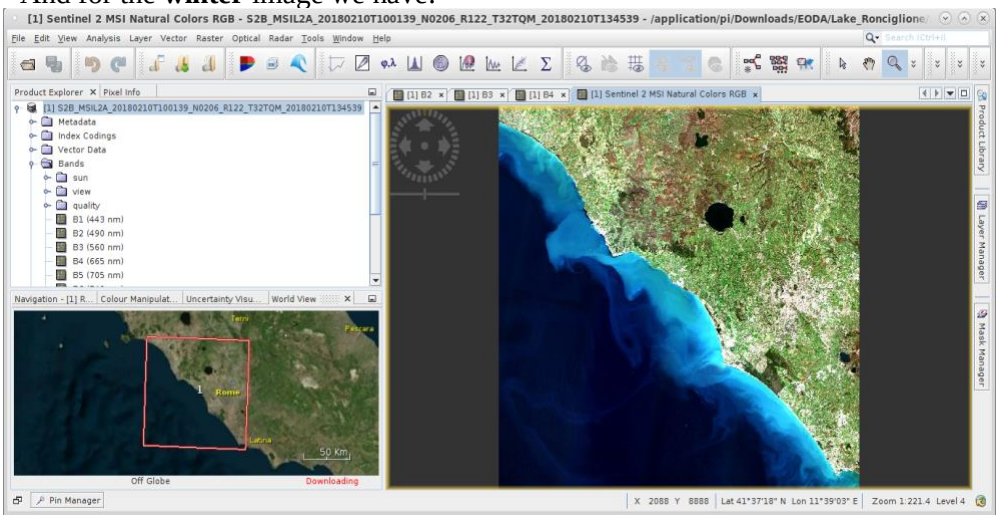




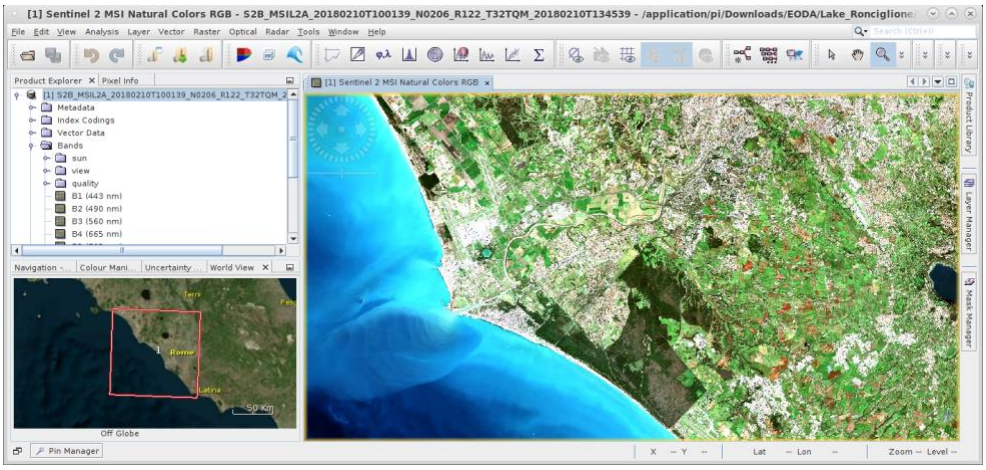
2. Zoom in to observe visible details:



3. And for the **winter** image we have:



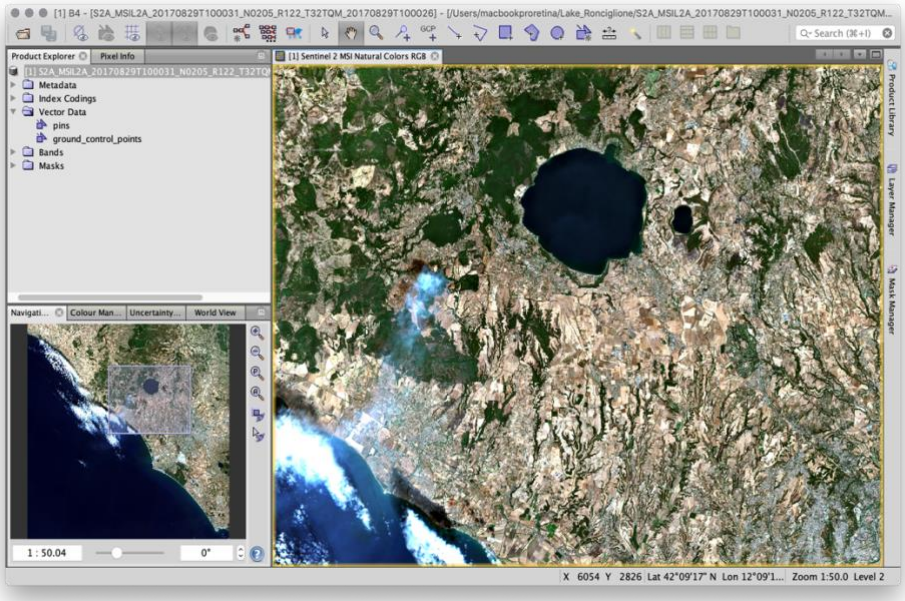
4. And more zoom in to observe visible details:



b) For both summer and winter MSI images, use L2 MSI products for scene classification masks (cloud shadows, vegetation, bare soil, water, snow and ice, …) and build/visualize an RGB using bare soil, vegetation and water respectively.

For summer image:

1. We opened the summer file and created RGB Image.



2. Then we have define image in different classes. For that we have to define training data set. We have to define new vector data containers and we have define vegetation, bare soil and water as vectors for data exploration.

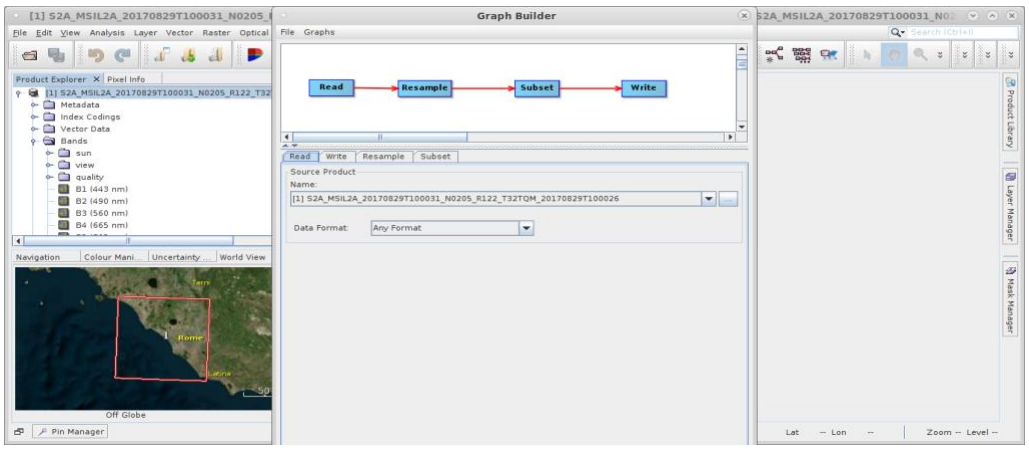
3. Then after reprojection towards the specific area we got this result.

a. Summer: soil, vegetation and water

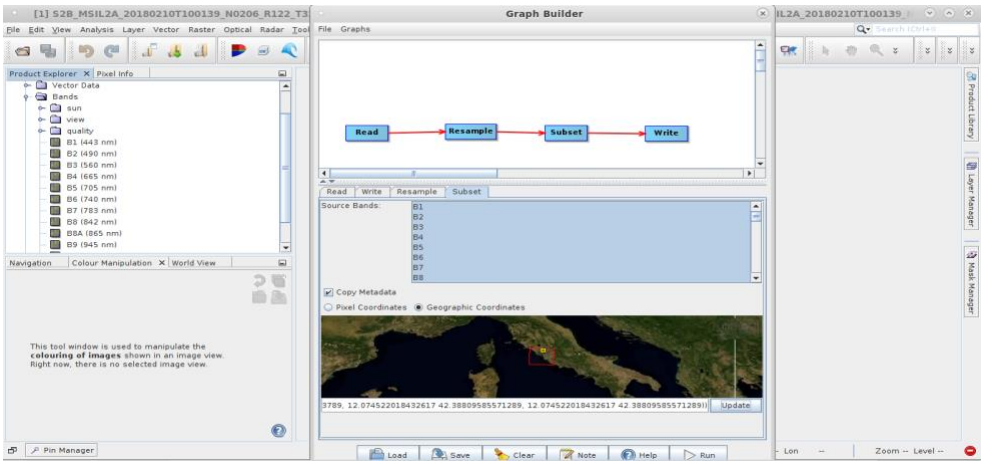
b. Winter: soil, vegetation and water

c) For both summer and winter MSI images, using SNAP Graph Builder tool, **subset around a lake/river** and resample all B2, B3, B4 channels at 10 m and visualize them.

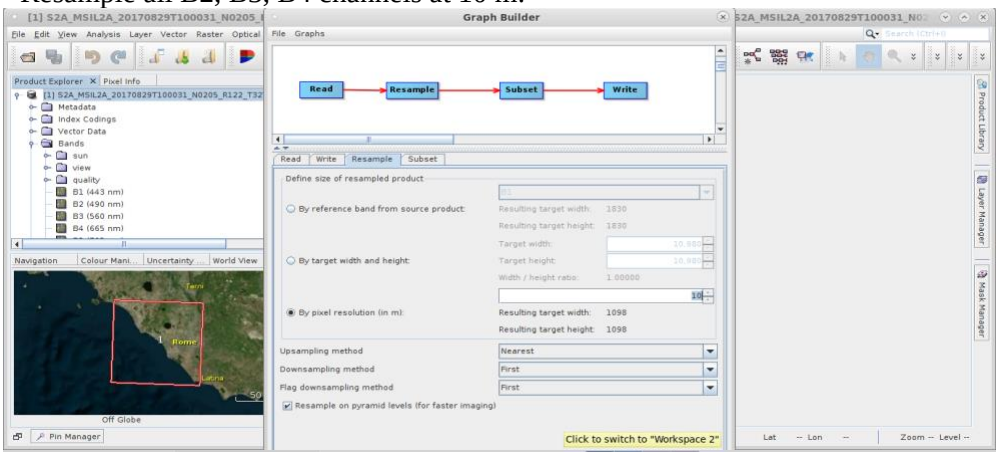
1. For **summer** image we have:



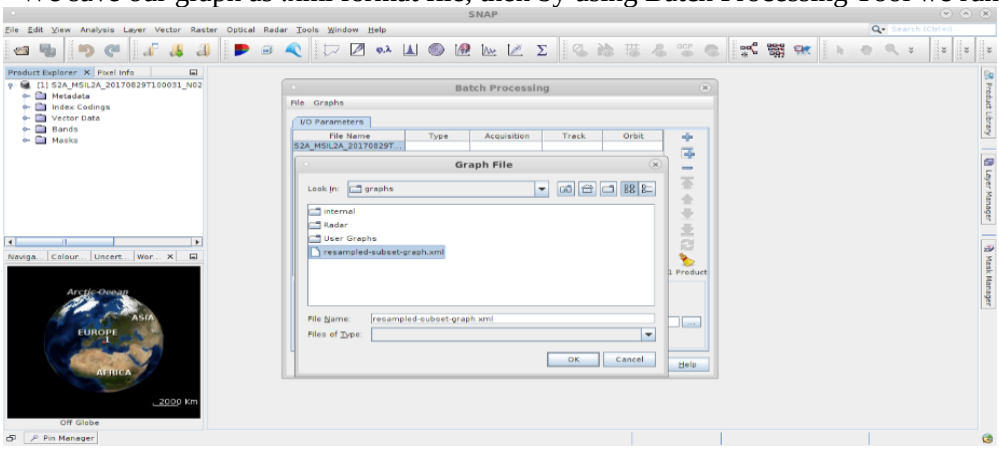
2. Subset around the a lake Ronciglione with predefined geometry WKT:
POLYGON ((12.074522018432617 42.38809585571289, 12.264036178588867
42.38809585571289, 12.264036178588867 42.22330093383789,
12.074522018432617 42.22330093383789, 12.074522018432617
42.38809585571289, 12.074522018432617 42.38809585571289))



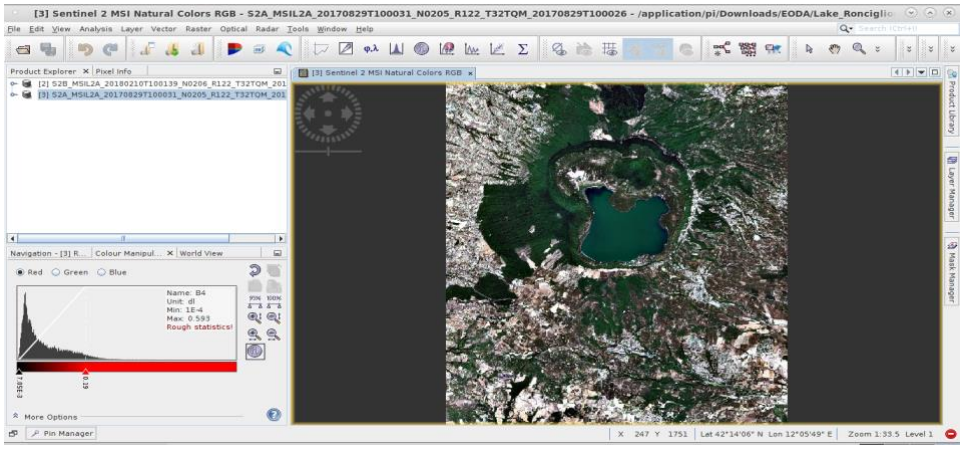
3. Resample all B2, B3, B4 channels at 10 m:



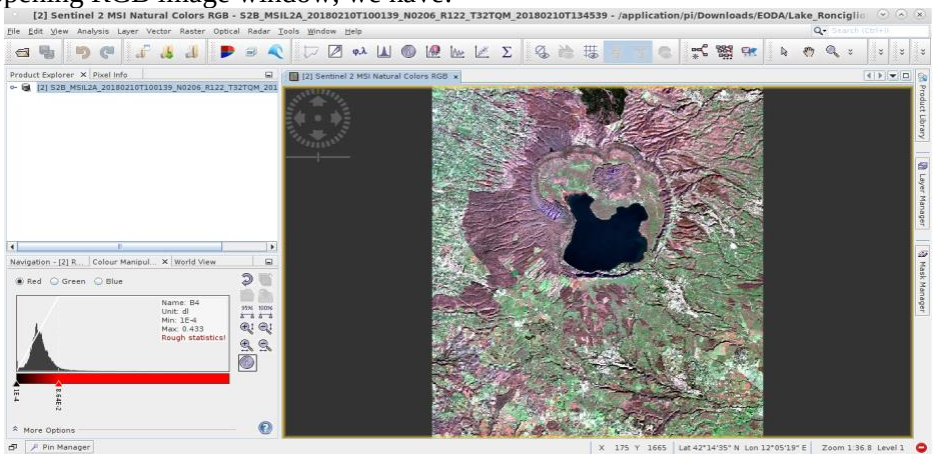
4. We save our graph as .xml format file, then by using Batch Processing Tool we run the graph:



5. And if we open RGB image window we see:



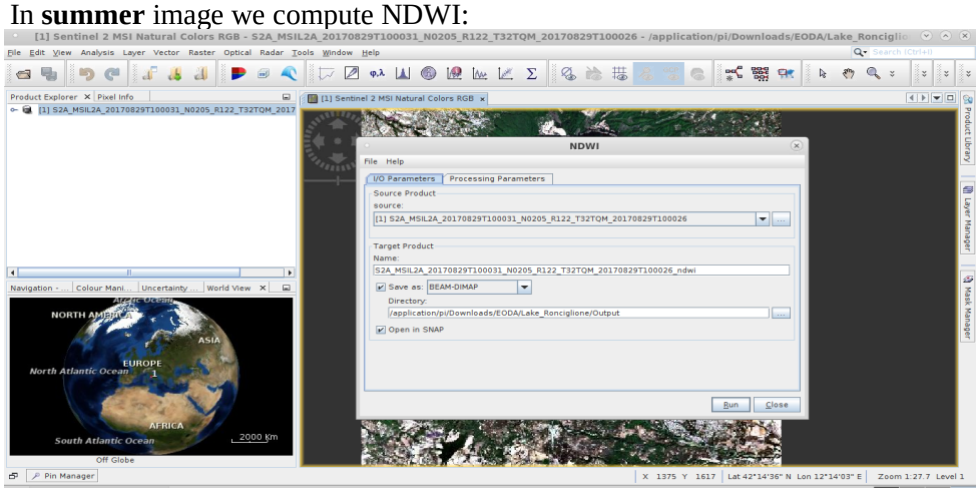
6. In the same way after making the graph and loading it with new **winter** image and then opening RGB image window, we have:

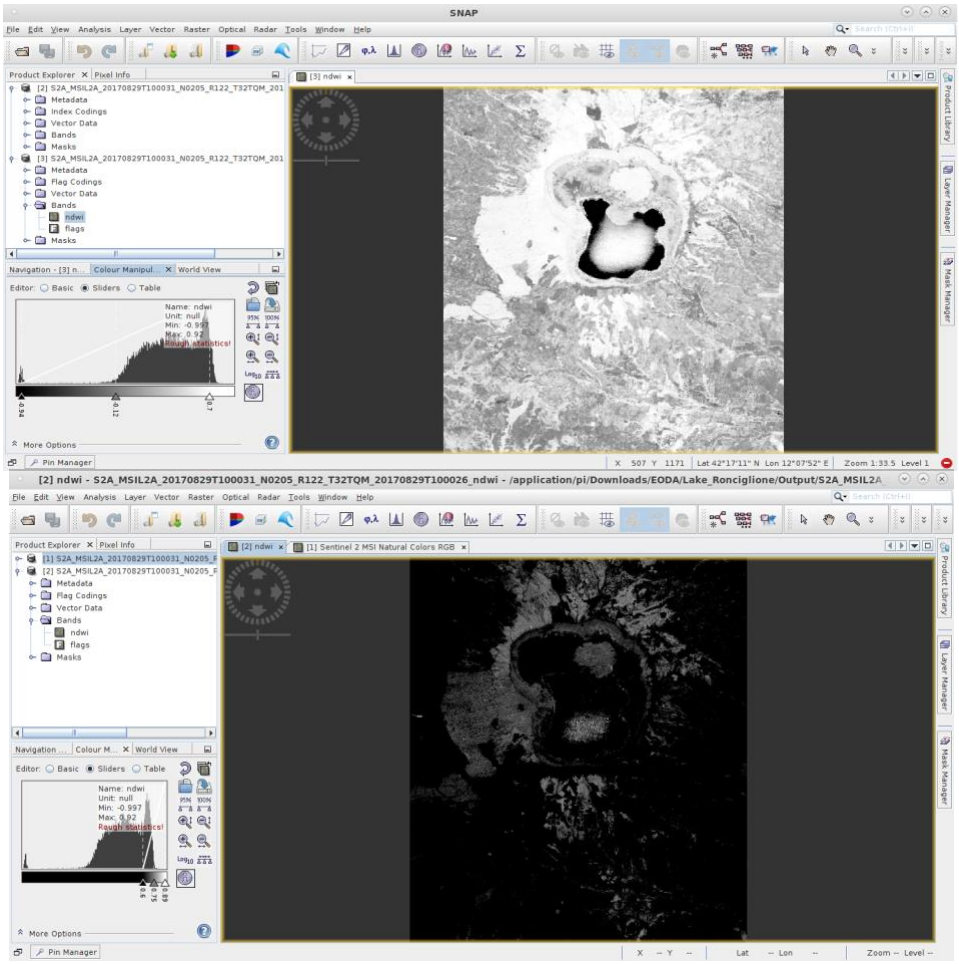


2. WATER AND VEGETATION NORMALIZED INDEX

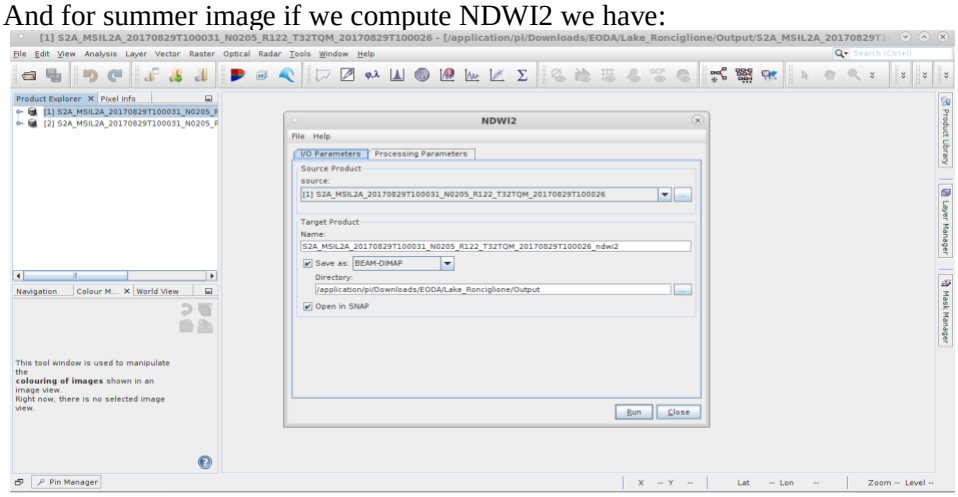
a) For both summer and winter MSI resampled images, compute NDWI on winter product and compute NDWI2 showing that NDWI puts in evidence water content in the leaves while NDWI2 can be used to build a "water mask'' estimate chlorophyll-a (Chl-a) and total suspended sediments (TSS).

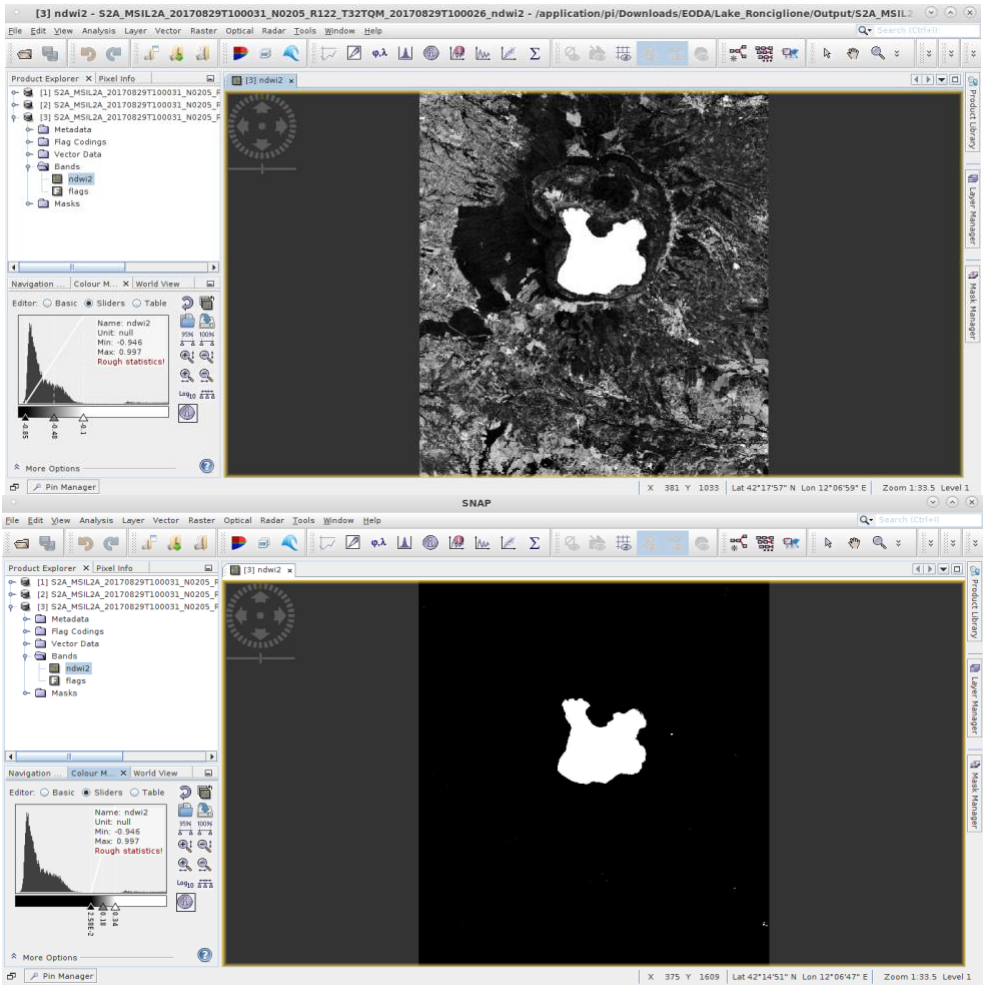
1. In **summer** image we compute NDWI:





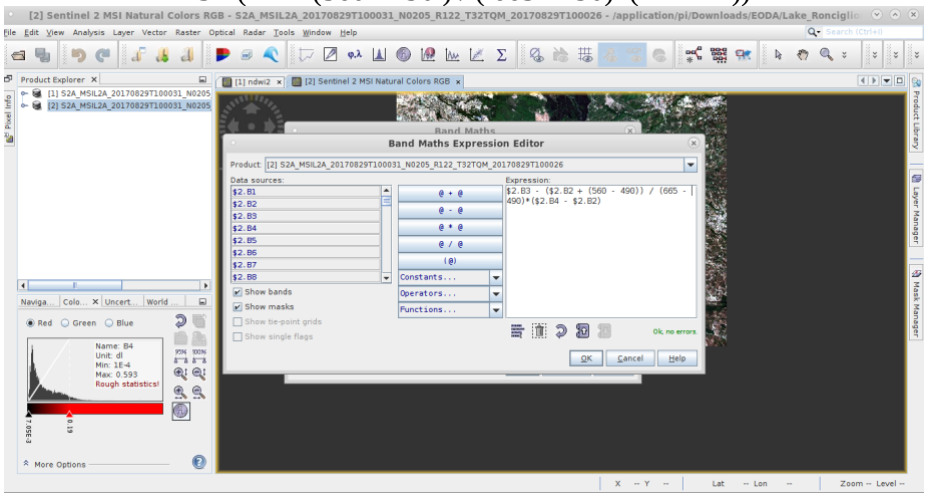
2. And for summer image if we compute NDWI2 we have:

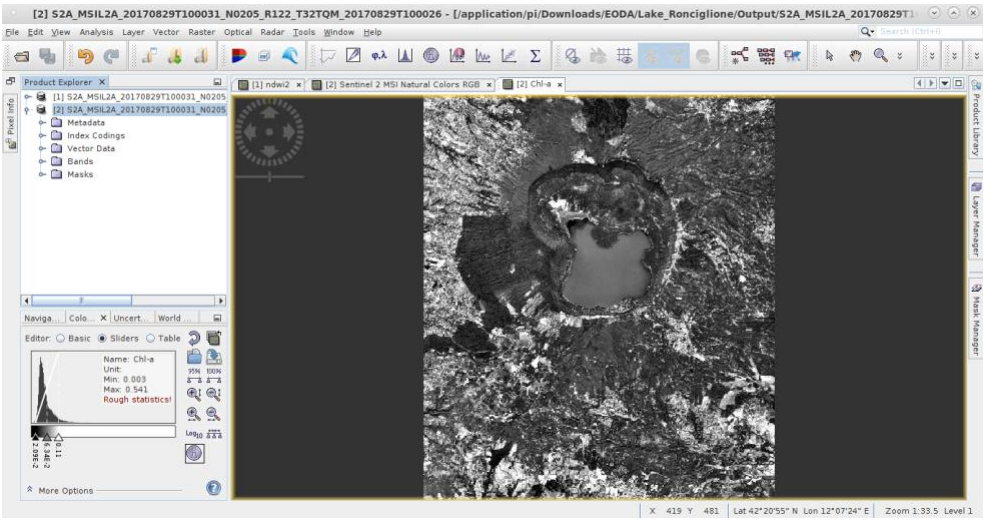




3. For estimating chlorophyll-a (Chl-a) we apply the following expression in Raster > Band Maths:

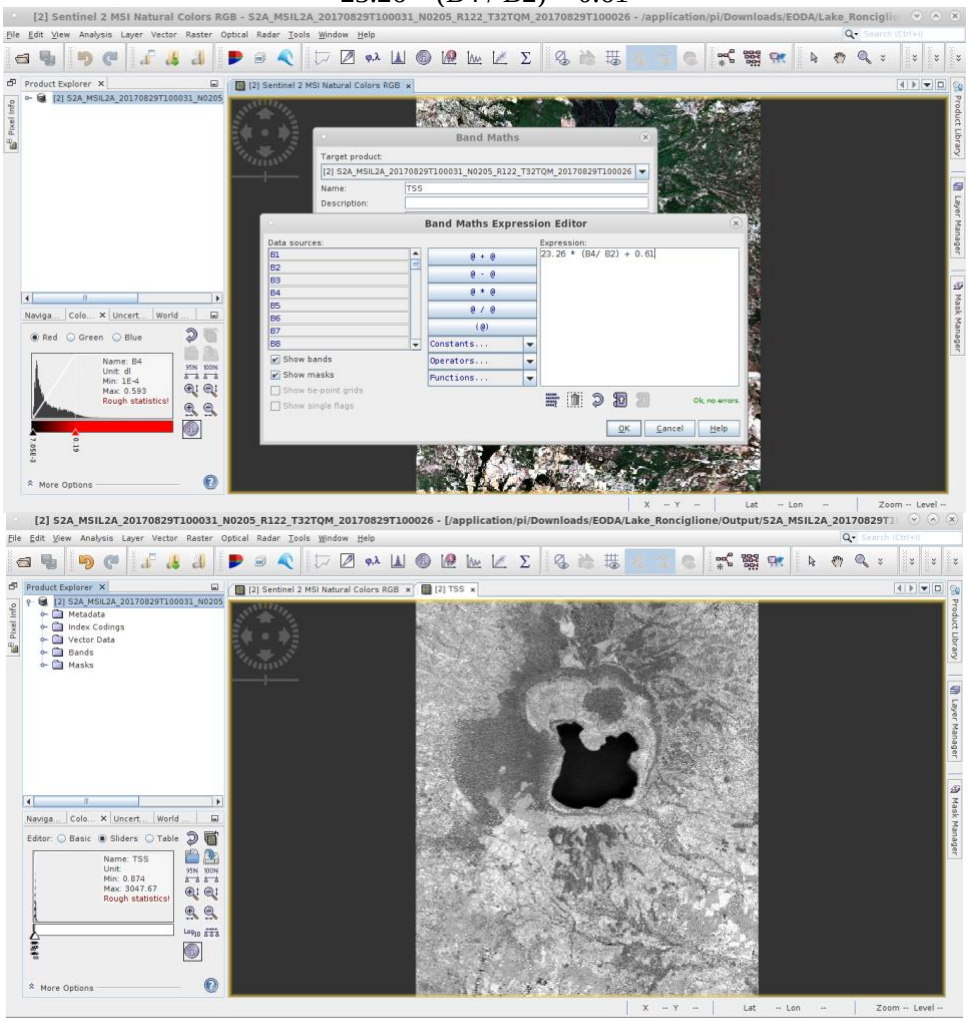
$$B3 - (B2 + (560 - 490) / (665 - 490)*(B4 - B2))$$



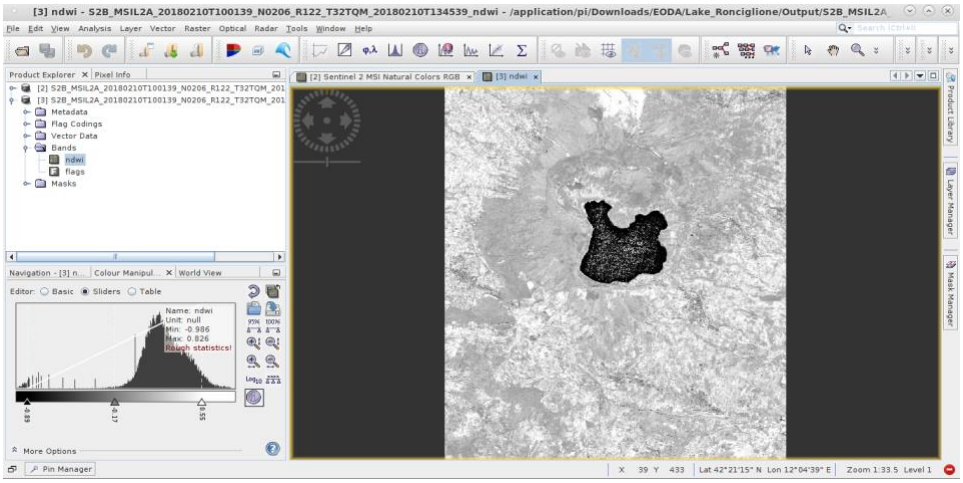


4. And in order to estimate total suspended sediments (TSS) we apply following expression in Raster > Band Maths:

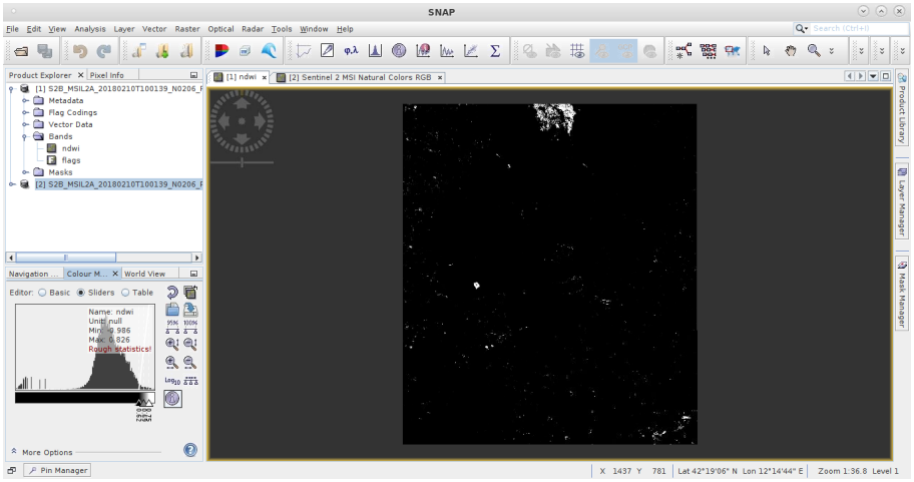
$$23.26 * (B4 / B2) + 0.61$$



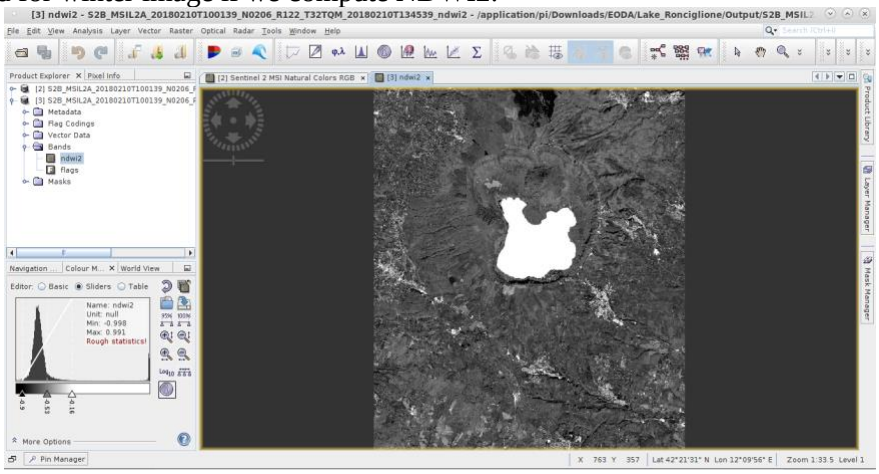
5. And in the same way for **winter** image if we compute NDWI we have:



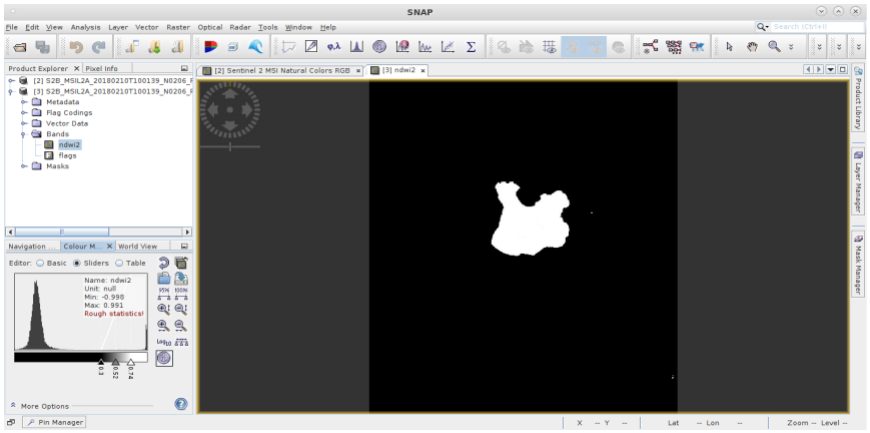
6. Here if we change thresholds in the histogram X-axis to high values, the image will be quite dark and only some white spots remain in the image. Now if we compare it with RGB image we can see that those white spots are vegetation area. Therefore this index can be used to monitor water content in the leaves



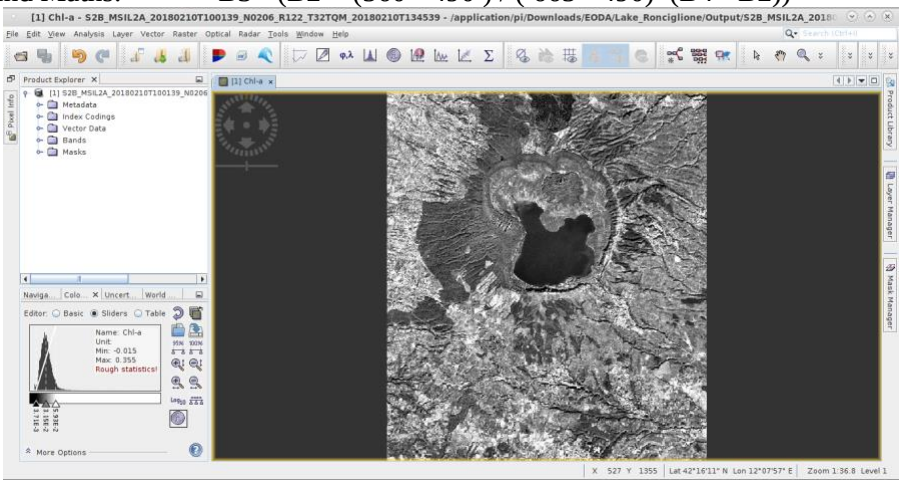
7. And for winter image if we compute NDWI2:



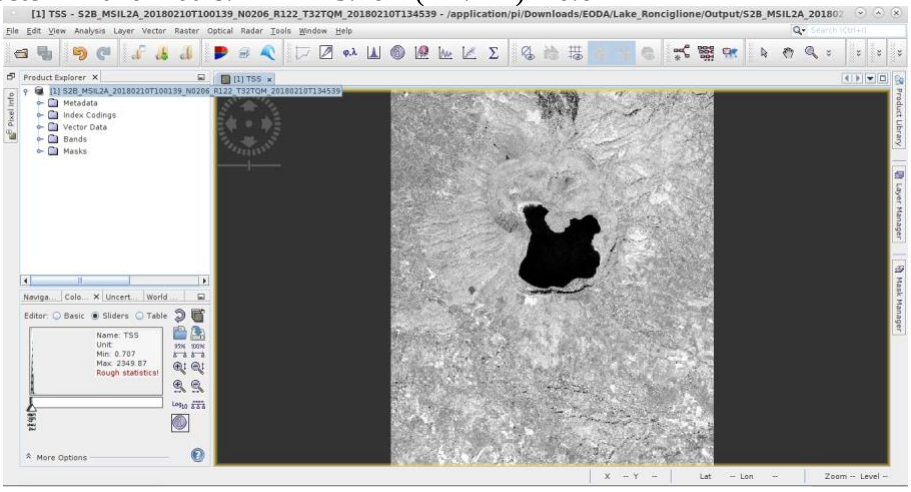
8. In this image we can see the values of the histogram are very high for the water in the lake and it is white, and for better show this we can change thresholds in the X-axis to isolate the water in the lake:



9. As we can see in this image we can conclude that NDWI2 can be used to build a "water mask''.
10. Now, for estimating chlorophyll-a (Chl-a) we apply the following expression in Raster > Band Maths: B3 - (B2 + (560 - 490) / (665 - 490)*(B4 - B2))

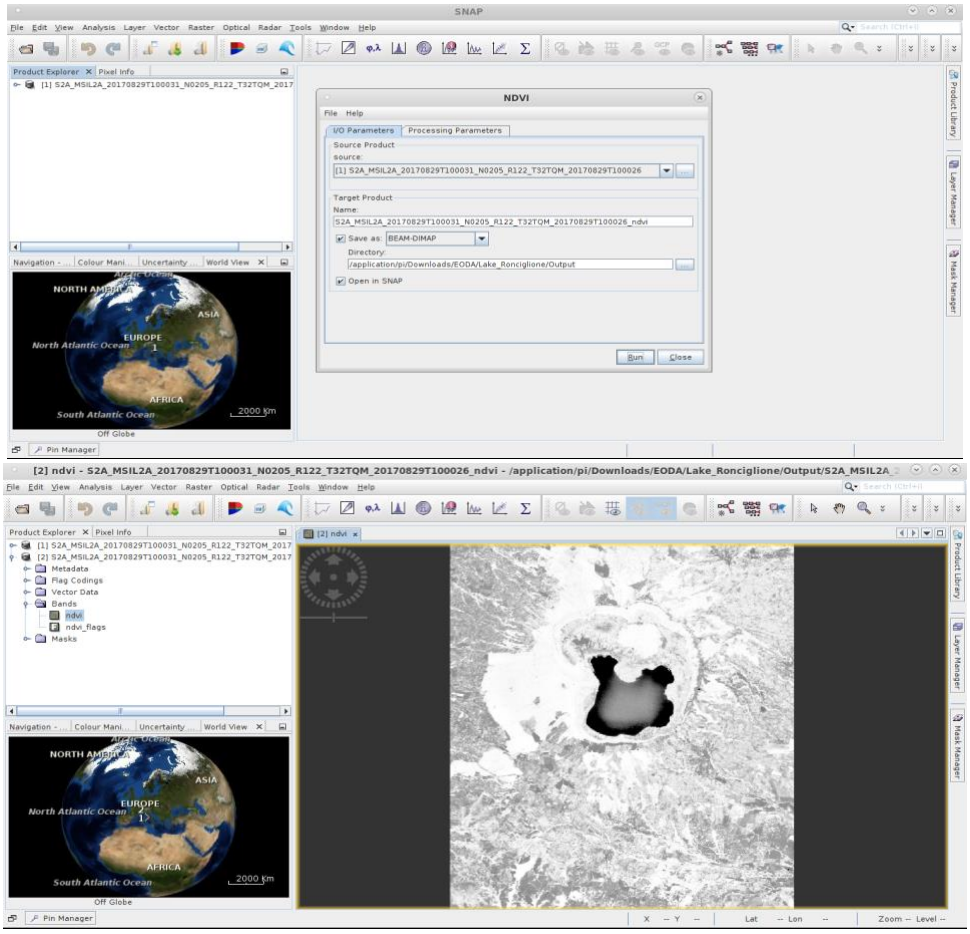


11. And in order to estimate total suspended sediments (TSS) we apply following expression in Raster > Band Maths: 23.26 * (B4 / B2) + 0.61

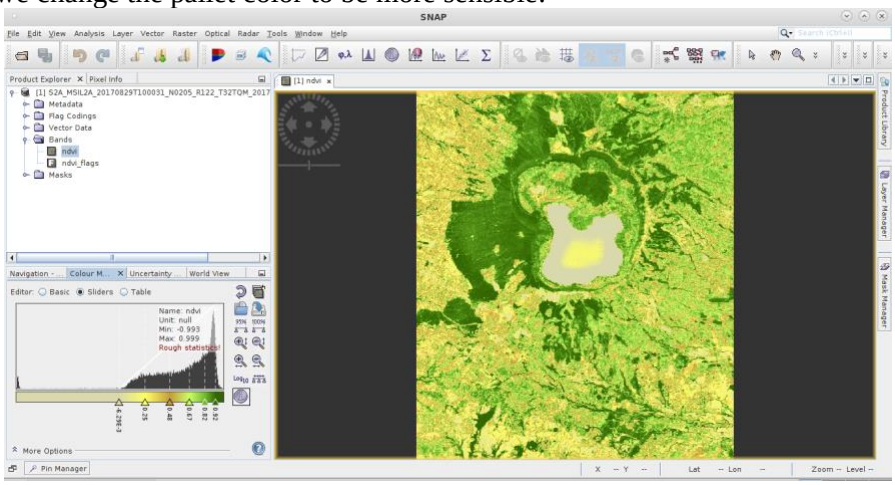


b) Compute the NDVI on winter and summer resampled MSI products showing the difference between winter and summer NDVI products in terms of histograms.

1. In **summer** image we compute NDVI:

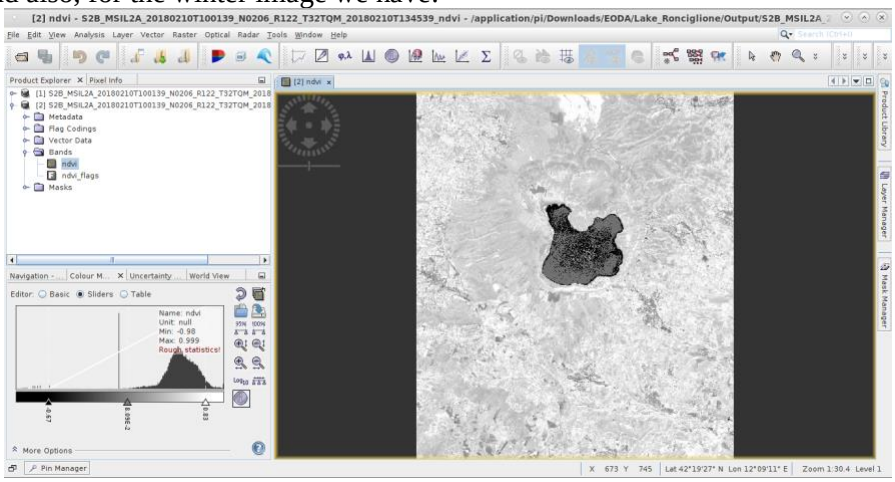


2. If we change the pallet color to be more sensible:

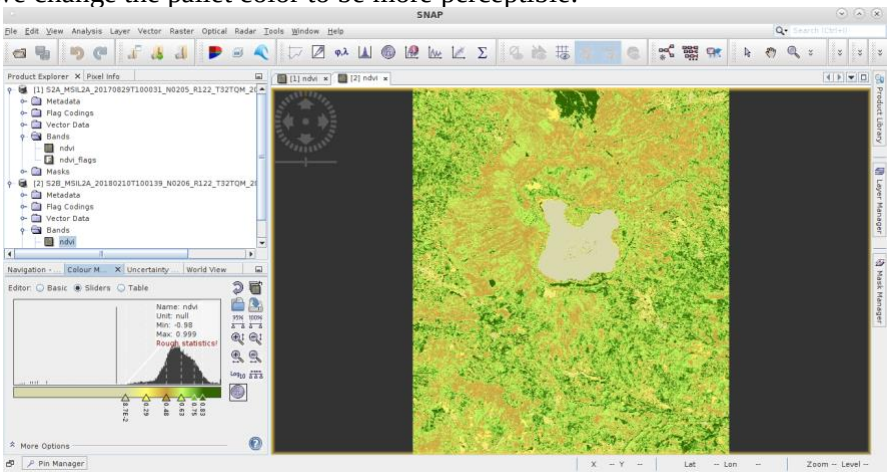


3. Green is associated to large values and brown is associated to small values.

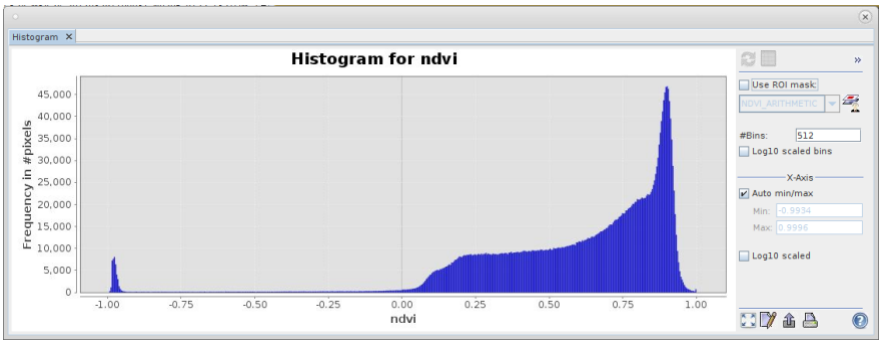
4. And also, for the winter image we have:



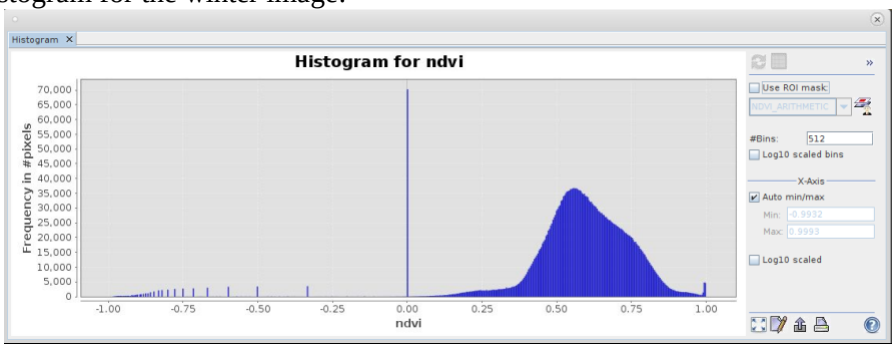
5. If we change the pallet color to be more perceptible:



6. Now if we want to see the distribution of NDVI values we can use histogram tool. Histogram for the summer image:



7. Histogram for the winter image:



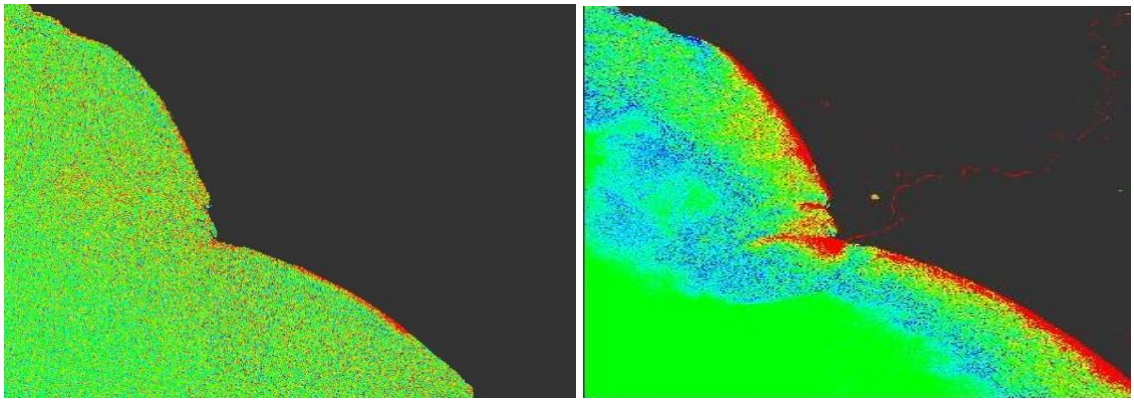
8. As we can see in histogram distribution, the summer distribution is more than winter image distribution skewed to high values(most are near 1) and the high values associated to green parts of the image(vegetation), and this is reasonable, because in the summer we have more vegetation than winter season(more green parts in the image).

3. SEA CHLOROPHYLL-A AND SEDIMENT ESTIMATION

a) Select a subset ROI within the summer and winter MSI image subsetting a coastline near a river estuary/delta.

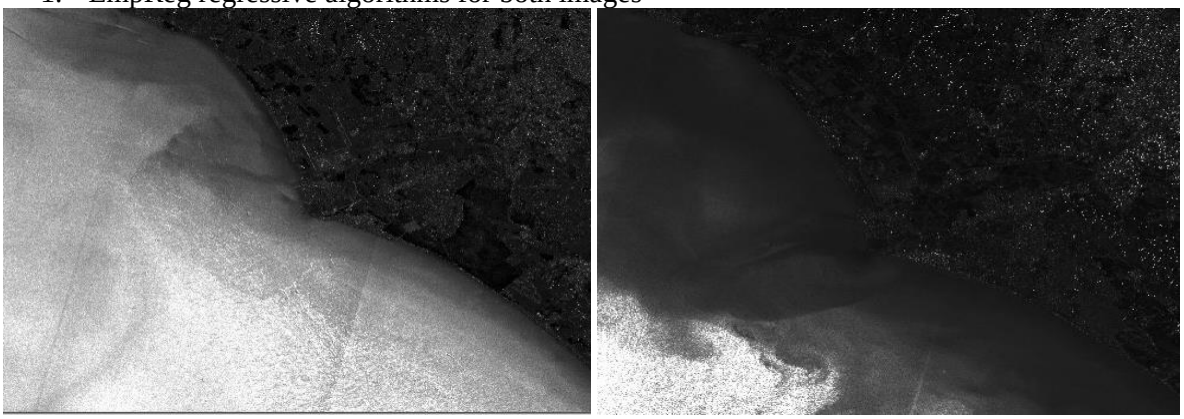
1. ROI in summer and winter (20/07/2017 and 10/02/2018)

2. Chlorophyll-a (Chl-a) and total suspended sediments (TSS) in color



b) Implement at the EmpReg regressive algorithms (see below) to estimate chlorophyll-a (Chl-a) and total suspended sediments (TSS) using SNAP formula processing tool

1. EmpReg regressive algorithms for both images

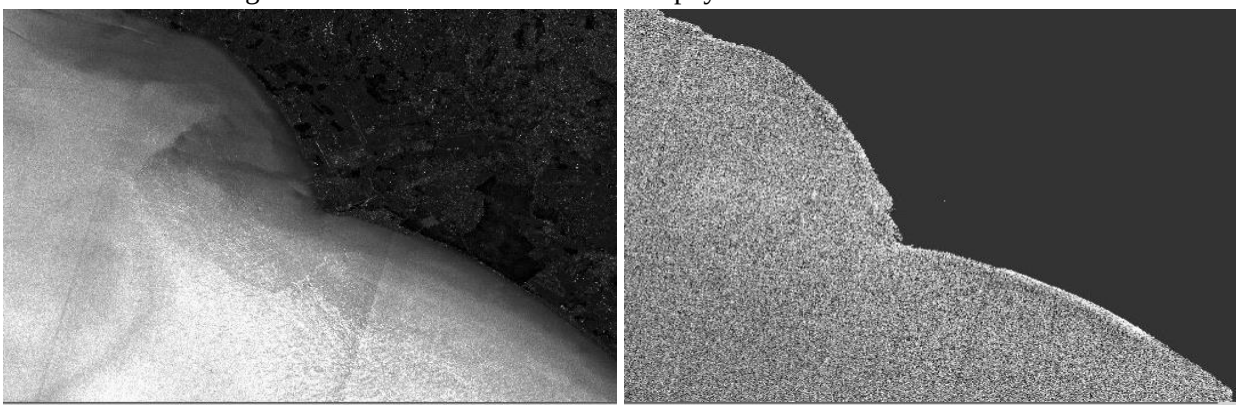


c) Compare estimated chlorophyll-a (Chl-a) and total suspended sediments (TSS) for a winter and summer cases by making the image differences.

EmpReg regressive algorithms image difference

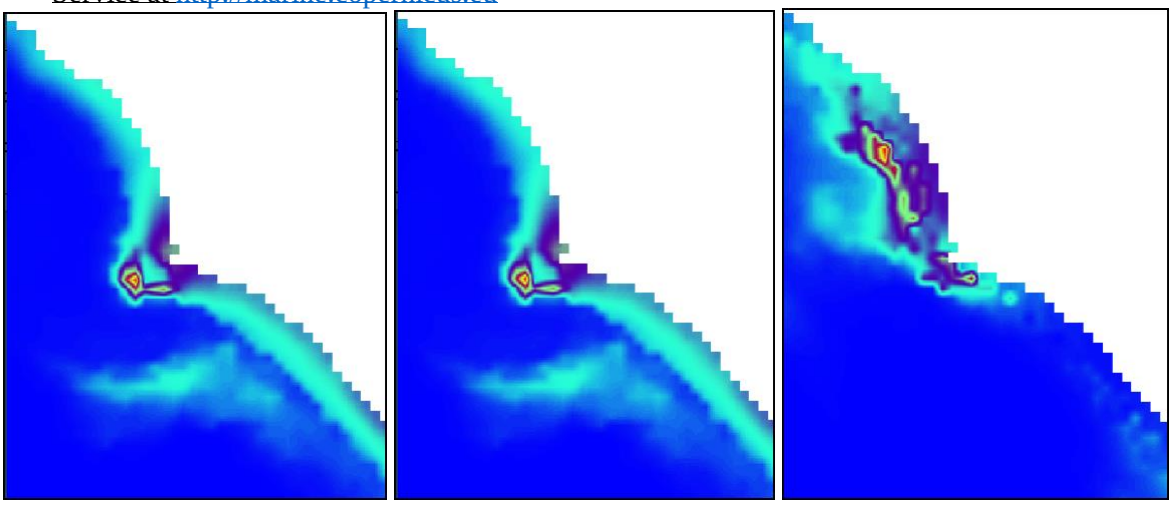
d) Compare estimated chlorophyll-a (Chl-a) with the one from SNAP C2RCC neural-network and MCI (Maximum Chlorophyll Index) plugged-in algorithms.

1. For the figures below, the one on the left is from EmpReg regressive algorithm while the one on the right is from SNAP Maximum Chlorophyll Index.



EmpReg regressive algorithms and SNAP MCI

e) Download all MSI/S2 Chl-a and TSS products around the target area within the Summer selected period (the closest in time) from international Copernicus Marine Service at http://marine.copernicus.eu



2018 2019 2020

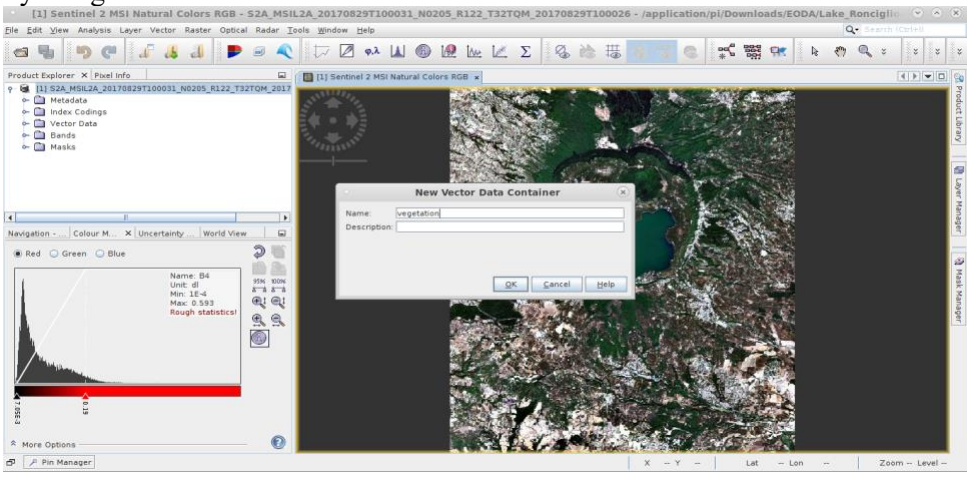
f) Compare your EmpReg Chl-a and TSS retrievals with official Copernicus S2 MSI products within the selected areas.

Despite the Copernicus products offer for a wide range of dates as well as a vast cover area, the spatial resolution is way worse than the images from SNAP. Looking at the above images over our ROI, we can hardly draw any conclusions except for the fact that the concentration of chlorophyll-a seems to increase over the year. There are also limited options while editing the low resolution images downloaded from Copernicus Marine Service.

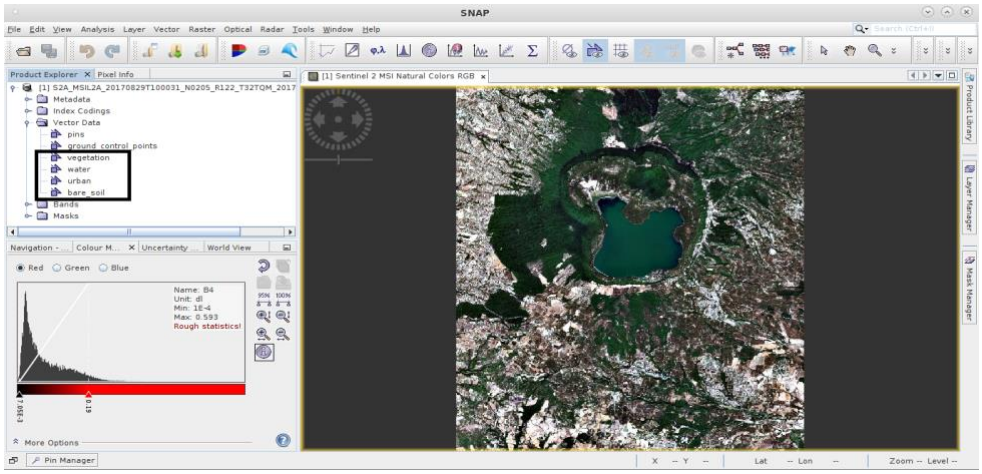
4. SCENE CLASSIFICATION

a) Using L2 MSI summer image, select scene classification masks, open a subset image and define 4 different polygons over vegetation, (ii) water, (iii) urban and (iv) bare soil. These vectors will be used as training areas for automatic classification.

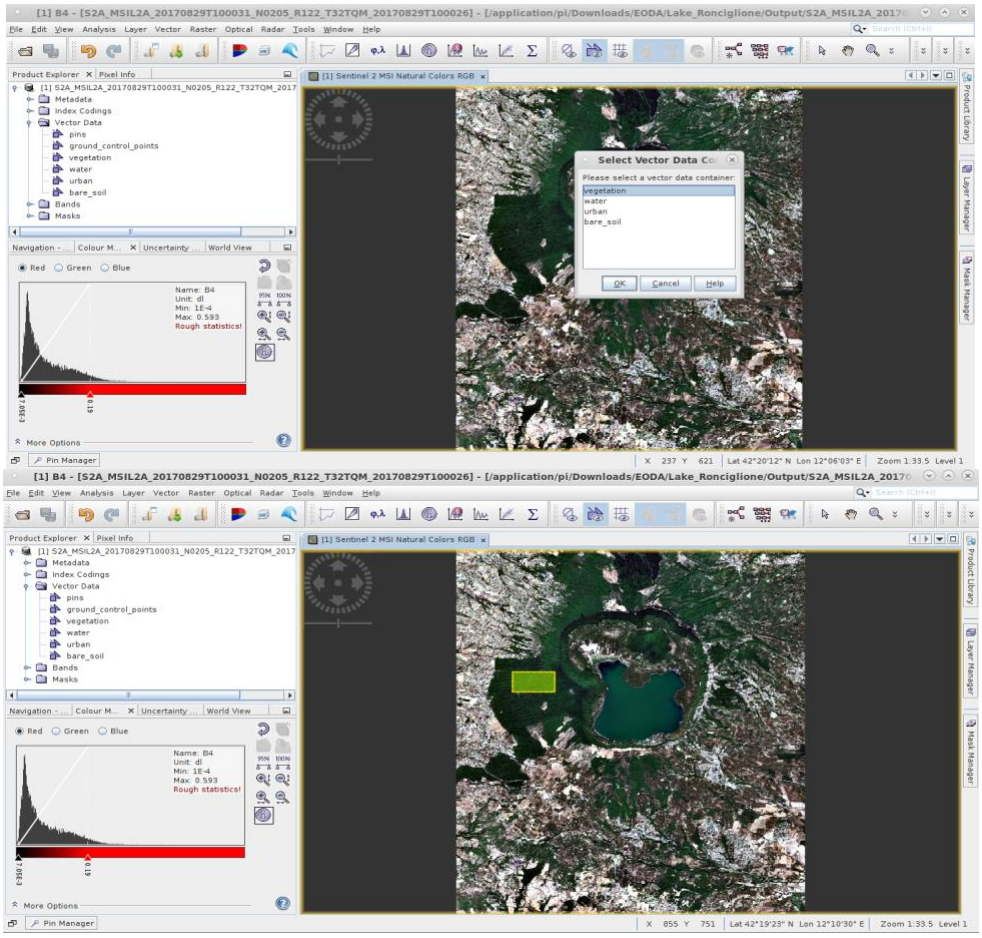
1. As we want to classify our summer image in 4 classes, we define a train areas for each class by using “new vector data container”.



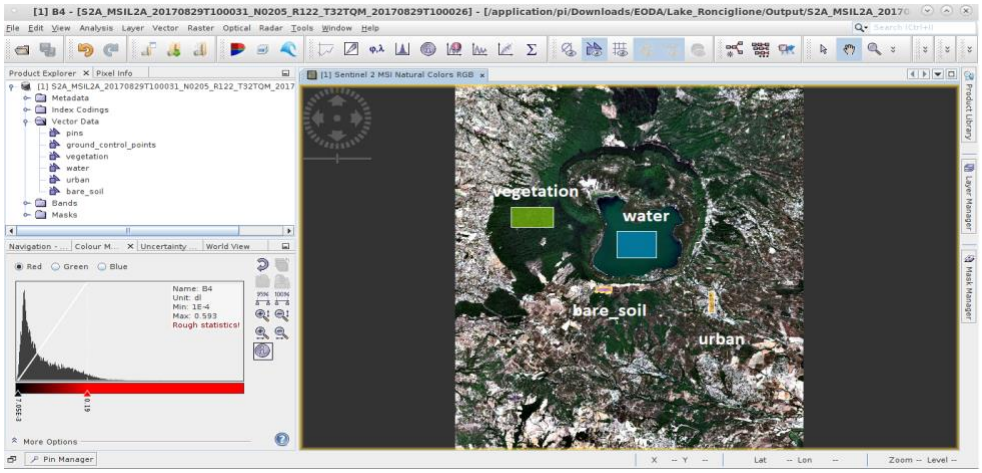
2. After creating 4 different polygons we can see them in product explorer:



3. Now we associate related pixels in image to its vector data container. For example for vegetation we have:

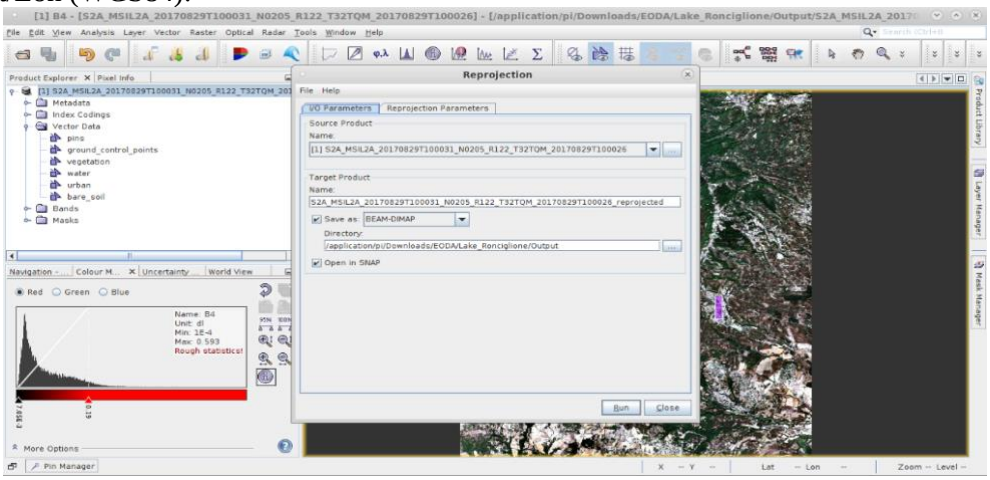


4. We do the same for water, urban and bare soil:

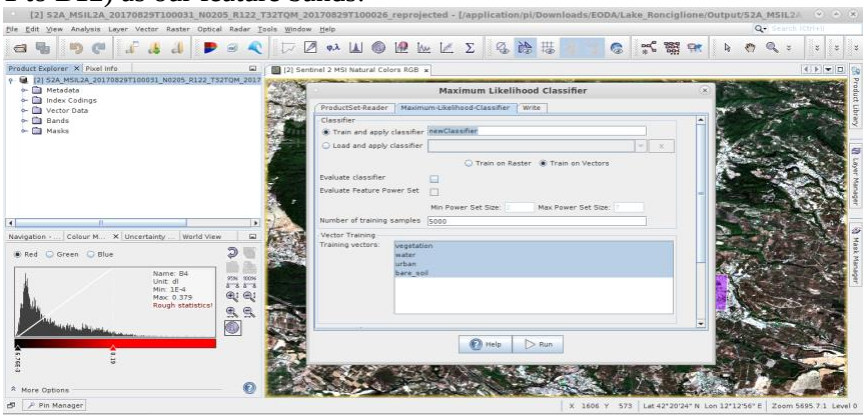


b) Reproject the subset image Lat/Lon WGS84 geographiy coordinate system. Use the SNAP Supervised Classification algorithms, such as Maximum Likelihood or Random Forest, to compute a classification.

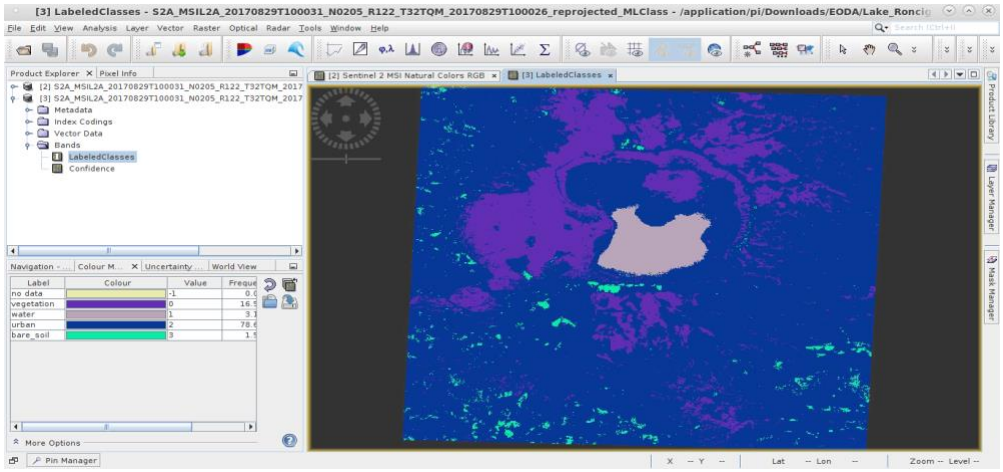
1. Before running the Maximum Likelihood algorithm, we perform reprojection on Geographic Lat/Lon (WGS84):



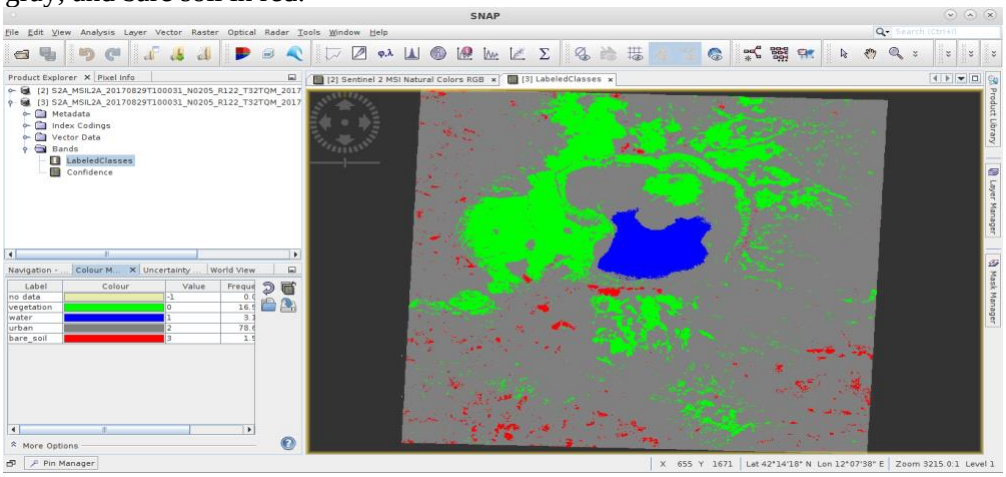
2. Now we perform Maximum Likelihood algorithm, a supervised classification. The classifier is "newClassifier" and train it on our vectors that we've made before, and also we select all the bands(B1 to B12) as our feature bands:



3. The obtained result is:



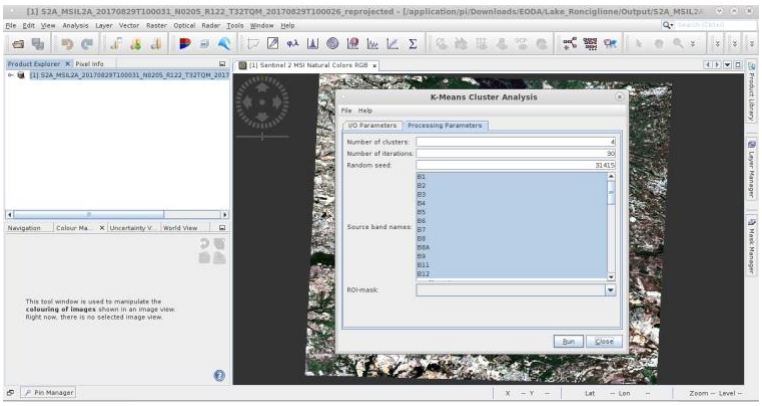
4. And if change the colors to be more sensible, we have. Vegetation in green, water in blue, urban in gray, and bare soil in red:



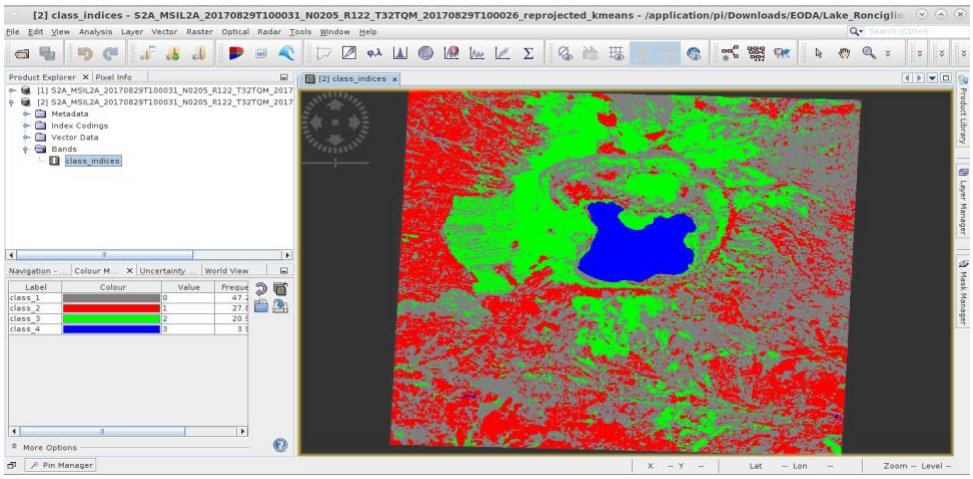
5. And if we check this labeled classes with RGB image, we can find out our classifier has done his job well enough.

c) <u>Compare different Classifiers to assess visually the most reliable using the L2 provided mask.</u>

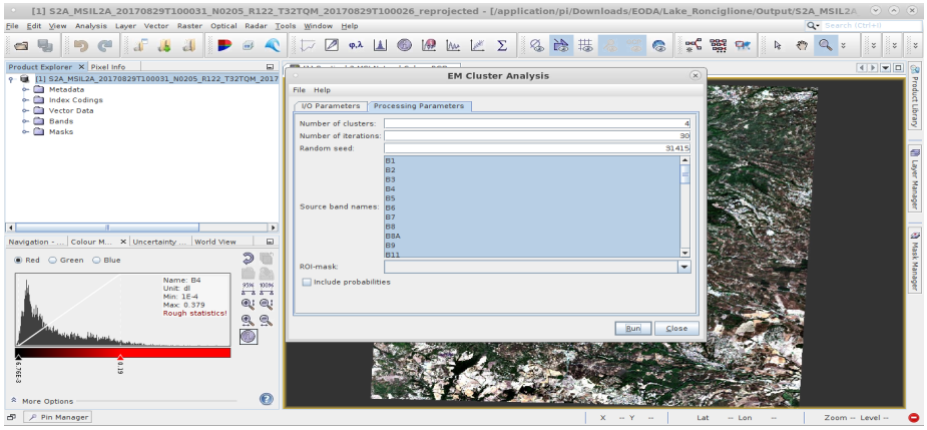
1. We decide to perform K-means Cluster analysis (an unsupervised classification algorithm) for our second approach, we choose 4 as the number of clusters (to obtain vegetation-water-urban-bare soil) and 30 as the number of iteration to and B1 to B12 as the source bands:



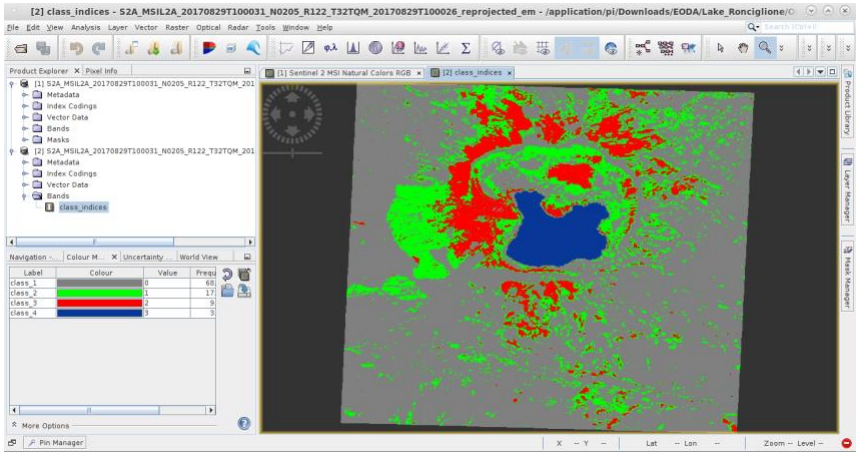
2. The obtained result with changed colors to the same colors that we set in last analysis(Maximum Likelihood) can be seen in the below:



3. We also perform EM Cluster analysis (an unsupervised classification algorithm) with number of clusters equals to 4 and for 30 iterations. We used B1 to B12 as the source bands:



4. The obtained result with changed colors to the same colors that we set in last analysis can be seen in the below:



EM clustering seems was not successful to get a good results, maybe it is due to number of clusters.