EODA

Homework 1

May 10/2020
Earth Observation for data Science

Professor:

- *Frank S. Marzano*

Estudents:

- *Nguyen Phu Hien* 1836702
- *Chandraprakash Poompavai* 1853976
- *Riposati Massimiliano* 1785279

Intro:

The purpose of this Homework is to explore MODIS satellite data, to obtain visual results over Italy imagery performing data analysis using the SNAP classification tools and comparing the vegetation index results between them.

Procedure Performed:

1. Download SNAP

From: http://step.esa.int/main/download/snap-download/
We proceed to the download and subsequent installation of SNAP tool

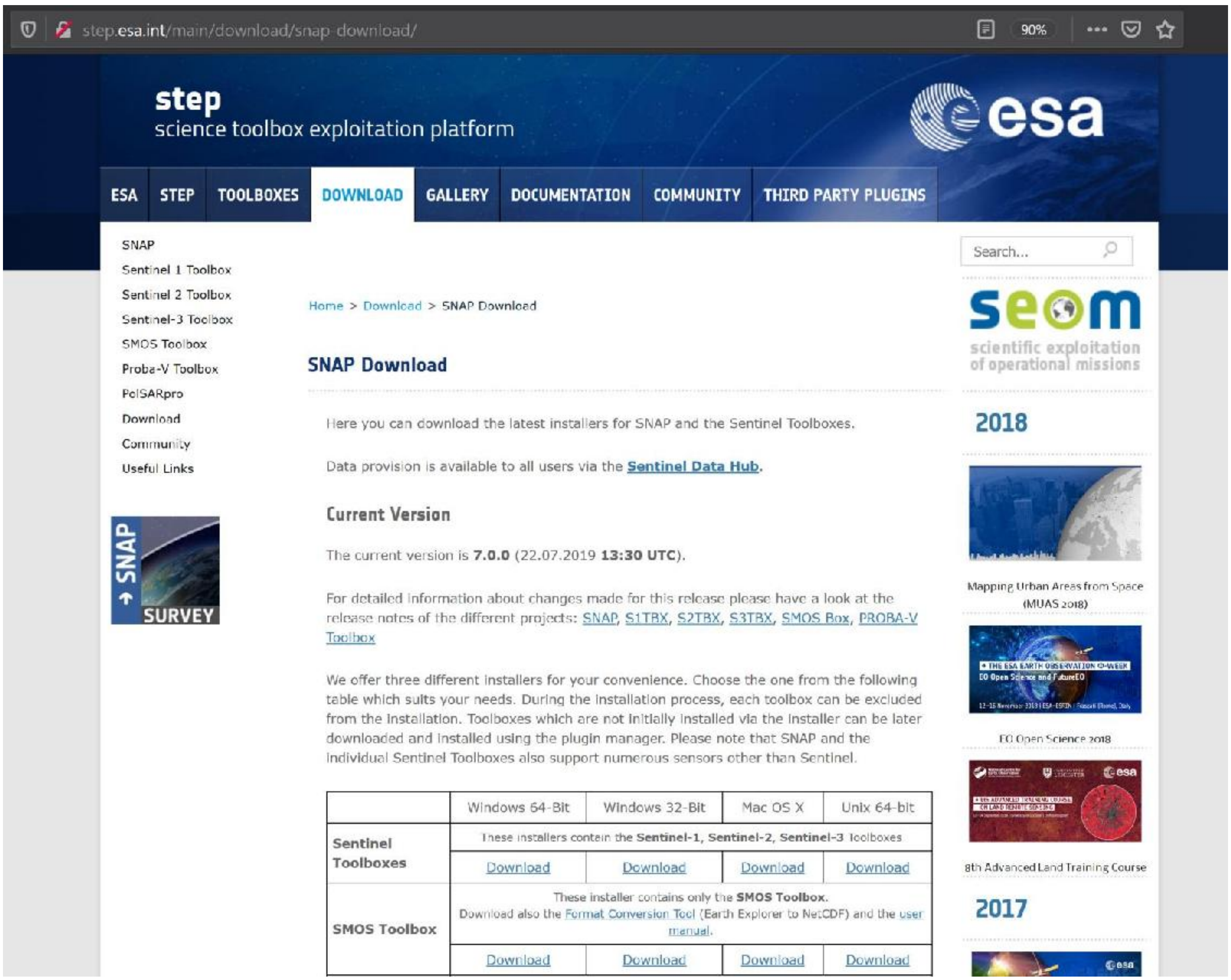


Figure 1: ESA Site for SNAP download

2. Download MODIS Images

From MODIS page, we download the Winter and Summer images for the same region of Italy.

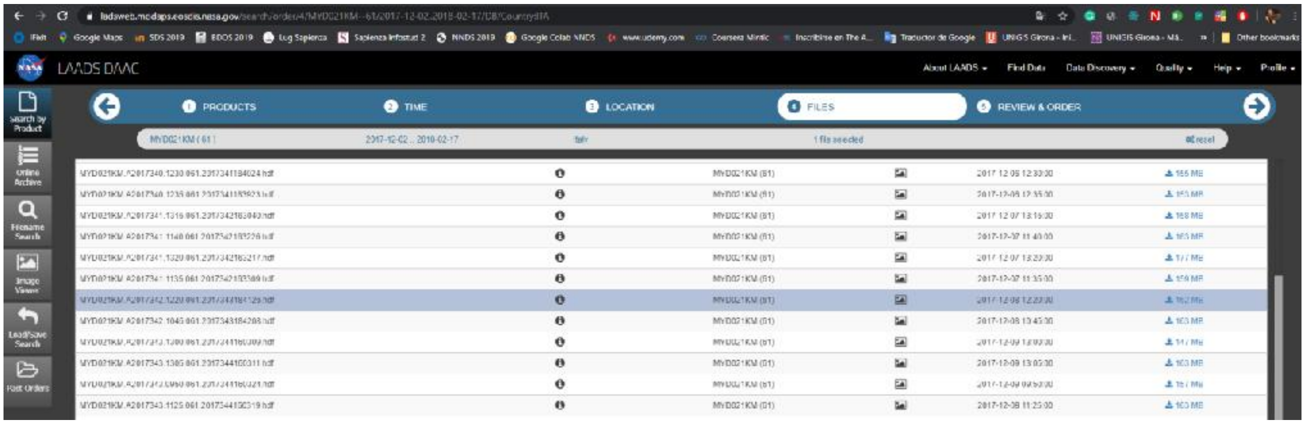


Figure 2.1: Images selection

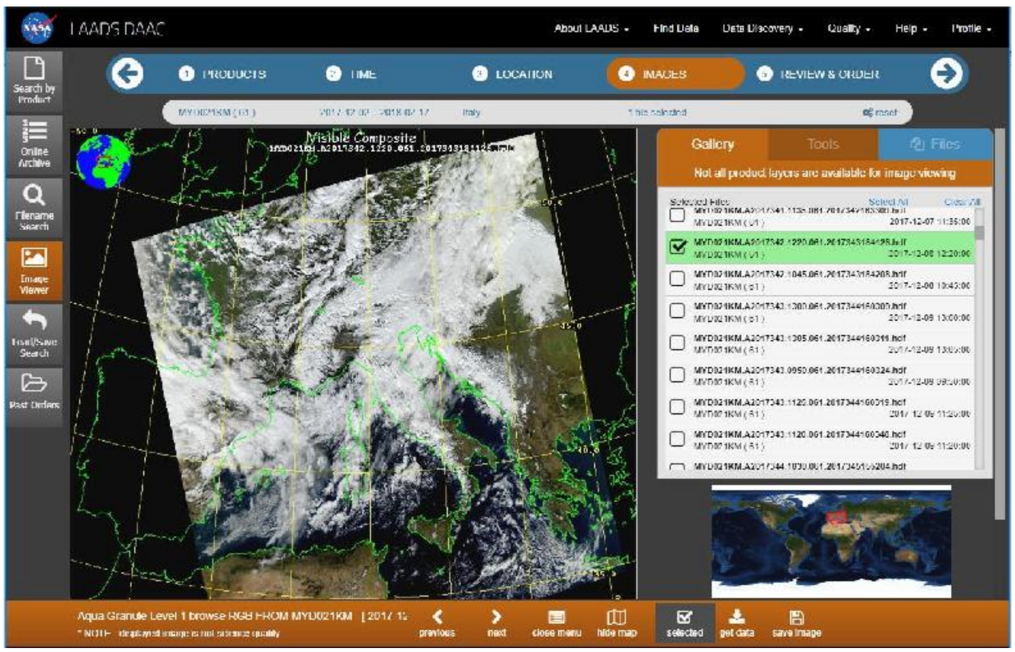


Figure 2.2: Preview image review

The product selected were:

Summer		Winter	
File MYD021KM.A2017185.1250.061.2018031151135.hdf		File MYD021KM.A2017342.1220.061.2017343184126.hdf	
Layers	Band_1KM_Emissive	Layers	Band_1KM_Emissive
ESDT	MYD021KM	ESDT	MYD021KM
Collection	61	Collection	61
File Name	MYD021KM.A2017185.1250.061.2018031151135.hdf	File Name	MYD021KM.A2017342.1220.061.2017343184126.hdf
File Size	143.44 MB	File Size	154.67 MB
Checksum	2856922060	Checksum	2104660346
Date/Time Sampled	2017-07-04 12:50:00	Date/Time Sampled	2017-12-08 12:20:00

Figure 2.3: Information table of the selected images

The format of these files is the widely used Hierarchical Data Format (HDF). HDF is the standard data storage format selected by the Earth Observing System Data and Information System (EOSDIS) Core System (ECS).

Summary of MODIS Level 1 Products

MODIS Level 1B 1KM Earth View Data Product contains calibrated Earth View observations from MODIS bands 8 through 36, at 1-kilometer resolution in scientific units. It also contains calibrated data from MODIS bands 1 through 7, each aggregated to 1km resolution.

ECS Short Name		Product Contents
MODIS/Terra	MODIS/Aqua	
MOD02QKM	MYD02QKM	Calibrated Earth View data at 250m resolution
MOD02HKM	MYD02HKM	Calibrated Earth View data at 500m resolution, including the 250m resolution bands aggregated to 500m resolution.
MOD021KM	MYD021KM	Calibrated Earth View data at 1km resolution, including the 250m and 500m resolution bands aggregated to 1km resolution.
MOD02OBC	MYD02OBC	On Board Calibrator (OBC) and Engineering Data

Figure 2.4: Summary of MODIS Level 1 Products

3. Perform data quality check

Once the images were downloaded, we used SNAP to visualize each of the bands and determine if they can be used or not in the project. Below, some selected images at random.

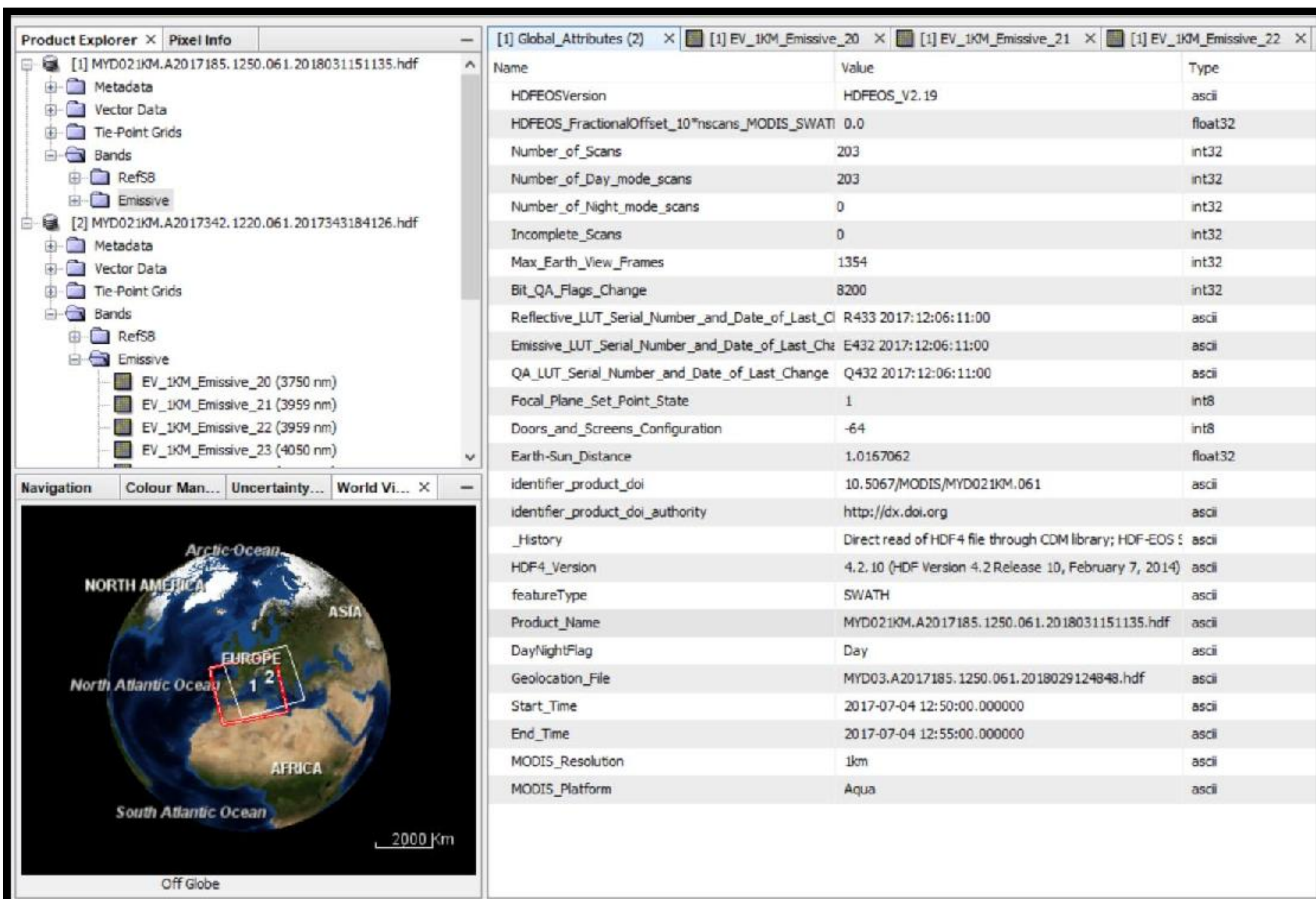


Figure 3.1: Metadata Review

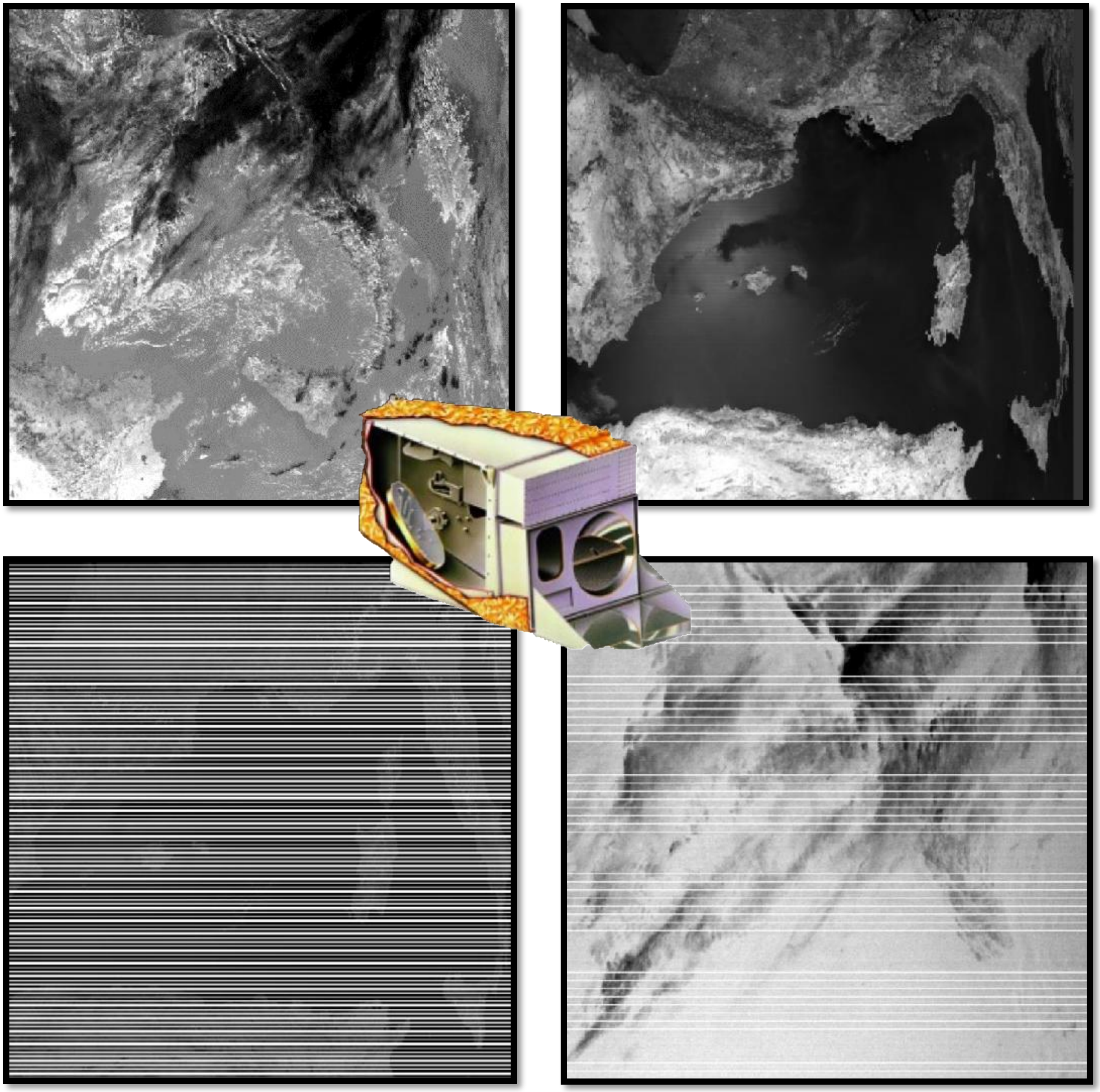


Figure 3.2.: Comparison of good and bad quality images

QUALITY CHECK TABLE

Reflective Solar Bands	Wavelength	Img. Quality	Observations
EV_1KM_RefSB_8	412 nm	Good	
EV_1KM_RefSB_9	443 nm	Good	
EV_1KM_RefSB_10	488 nm	Good	
EV_1KM_RefSB_11	531 nm	Good	
EV_1KM_RefSB_12	547 nm	Good	
EV_1KM_RefSB_13lo	667 nm	Good	
EV_1KM_RefSB_13hi	667 nm	Good	
EV_1KM_RefSB_14lo	678 nm	Good	
EV_1KM_RefSB_14hi	678 nm	Good	
EV_1KM_RefSB_15	748 nm	Good	
EV_1KM_RefSB_16	869 nm	Good	
EV_1KM_RefSB_17	905 nm	Good	
EV_1KM_RefSB_18	936 nm	Good	
EV_1KM_RefSB_19	940 nm	Good	
EV_1KM_RefSB_26	1375 nm	Good	
EV_250_Aggr1km_RefSB_1	645 nm	Good	
EV_250_Aggr1km_RefSB_2	859 nm	Good	
EV_500_Aggr1km_RefSB_3	469 nm	Good	
EV_500_Aggr1km_RefSB_4	555 nm	Good	
EV_500_Aggr1km_RefSB_5	1240 nm	Good	
EV_500_Aggr1km_RefSB_6	1640 nm	Bad	Imperceptible Image - Stripes Present
EV_500_Aggr1km_RefSB_7	2130 nm	Good	
Earth View Emissive Bands			
EV_1KM_Emissive_20	3750 nm	Good	
EV_1KM_Emissive_21	3959 nm	Medium	Minor Stripes Present
EV_1KM_Emissive_22	3959 nm	Good	
EV_1KM_Emissive_23	4050 nm	Good	
EV_1KM_Emissive_24	4465 nm	Good	
EV_1KM_Emissive_25	4515 nm	Good	
EV_1KM_Emissive_27	6715 nm	Good	
EV_1KM_Emissive_28	7325 nm	Good	
EV_1KM_Emissive_29	8550 nm	Good	
EV_1KM_Emissive_30	9730 nm	Good	
EV_1KM_Emissive_31	11030 nm	Good	
EV_1KM_Emissive_32	12020 nm	Good	
EV_1KM_Emissive_33	13335 nm	Good	
EV_1KM_Emissive_34	13635 nm	Good	
EV_1KM_Emissive_35	13935 nm	Good	
EV_1KM_Emissive_36	14235 nm	Medium	Imperceptible Image - Stripes Present

RGB Band combination for true color image composite:

- **R** EV_250_Aggr1km_RefSB_1 (645 nm)
- **G** EV_500_Aggr1km_RefSB_4 (555 nm)
- **B** EV_500_Aggr1km_RefSB_3 (469 nm)

Figure 3.3.: MODIS Composite.

4. Perform and display data analysis by spectrum histogram and profile tools.

To carry out the spectral analysis, two bands of good quality were selected, one in the emissive (B22) and the other in the reflective (B4), corresponding to the summer image collection.

Later, using google maps to determine the coordinates, the sampling points were selected and inserted into the pin manager.
Figure 4.1 below shows the results with pins and figure 4.2 shows spectral analysis result

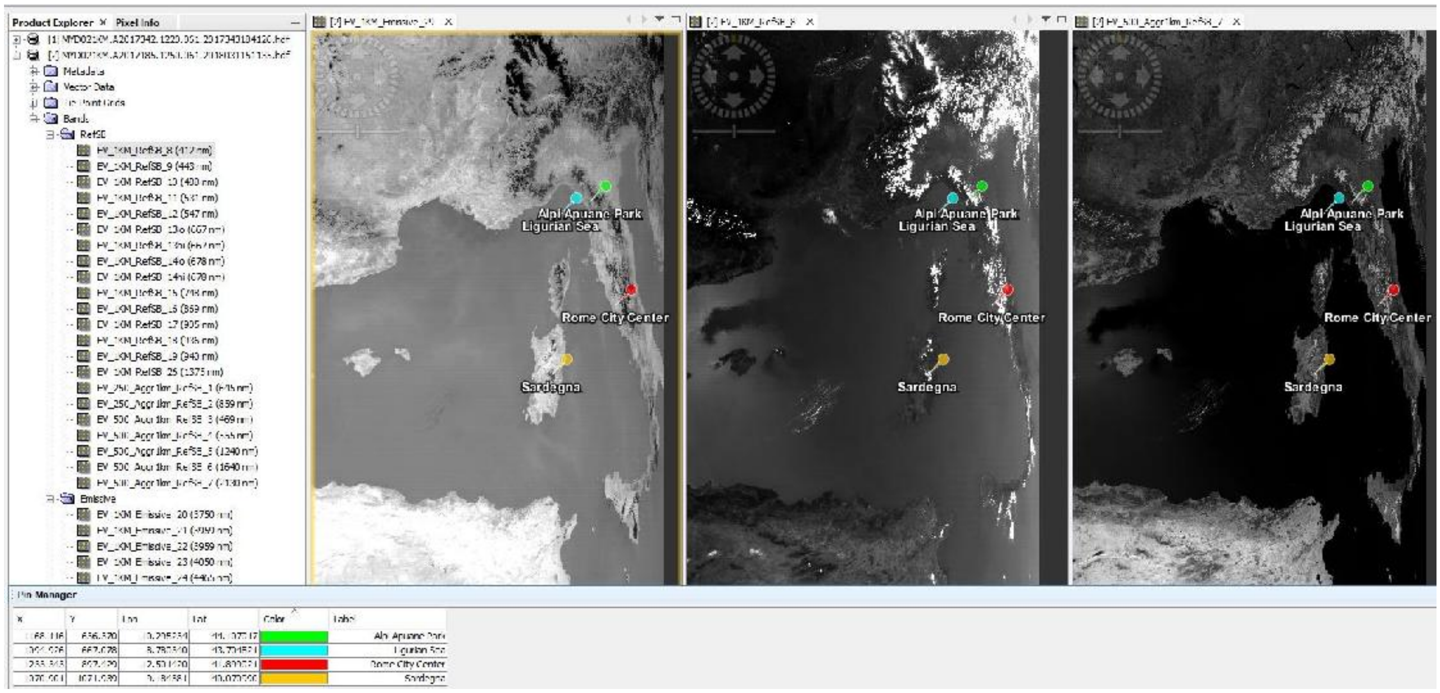


Figure 4.1

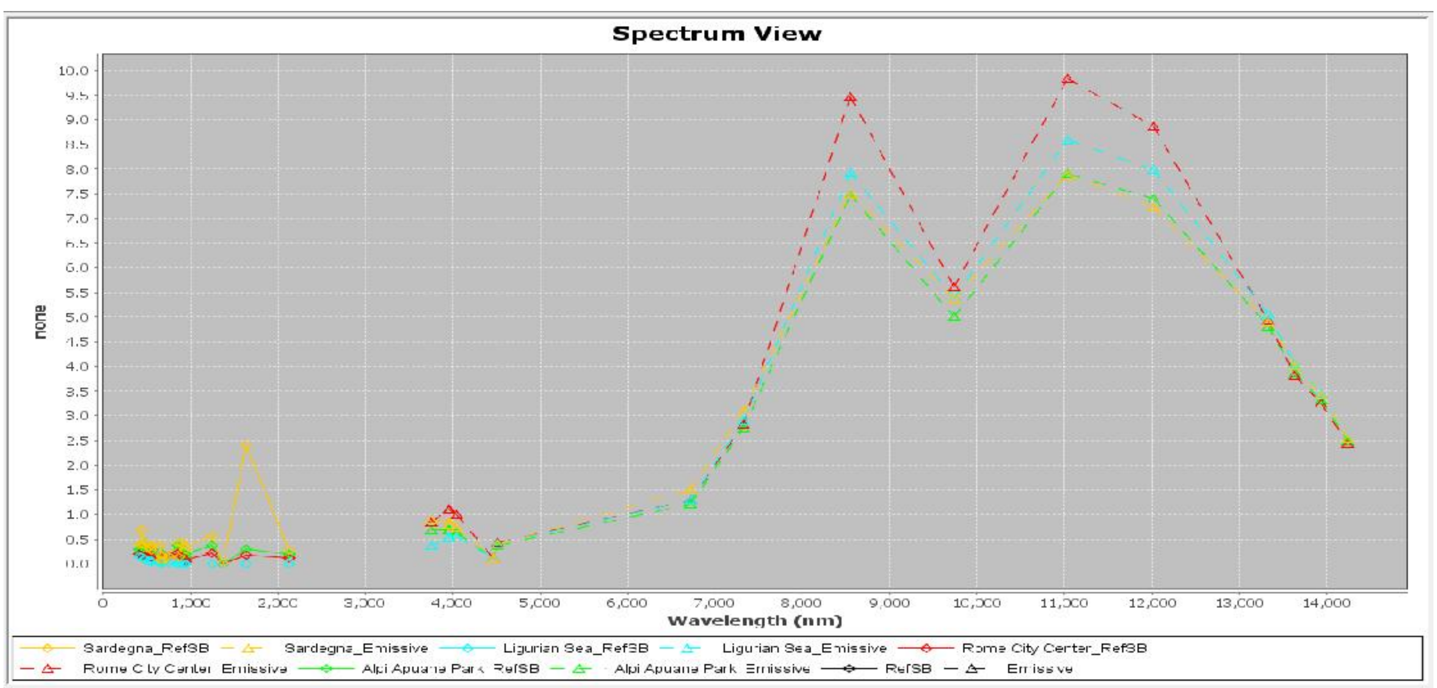
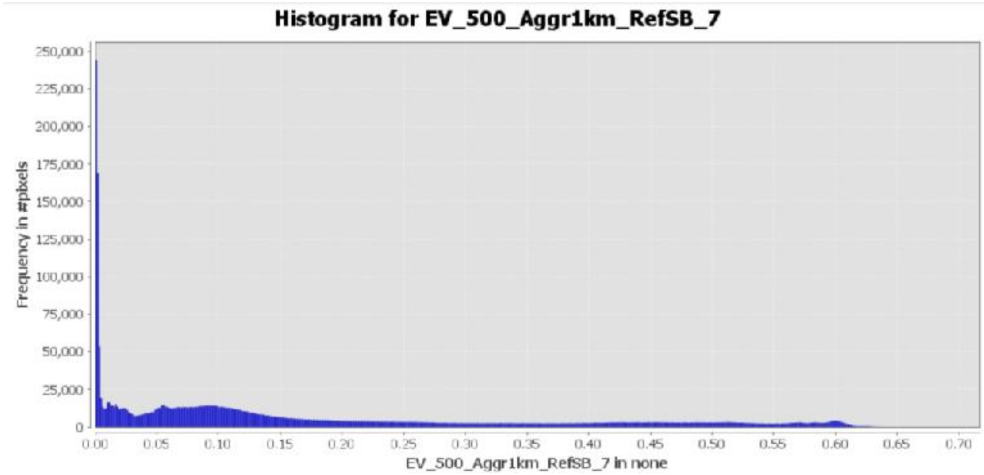


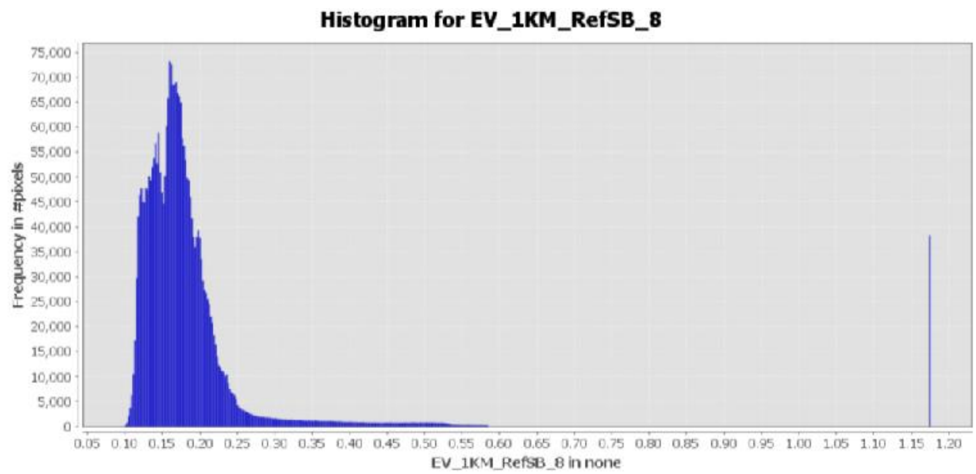
Figure 4.2

Looking at the spectral analysis graph in figure 4.2, it is evident that Pin 4 in orange (Ovodda - Serdegna) for reflectivity and Pin 3 in red (City Center – Rome) for emissivity have the highest values respectively.

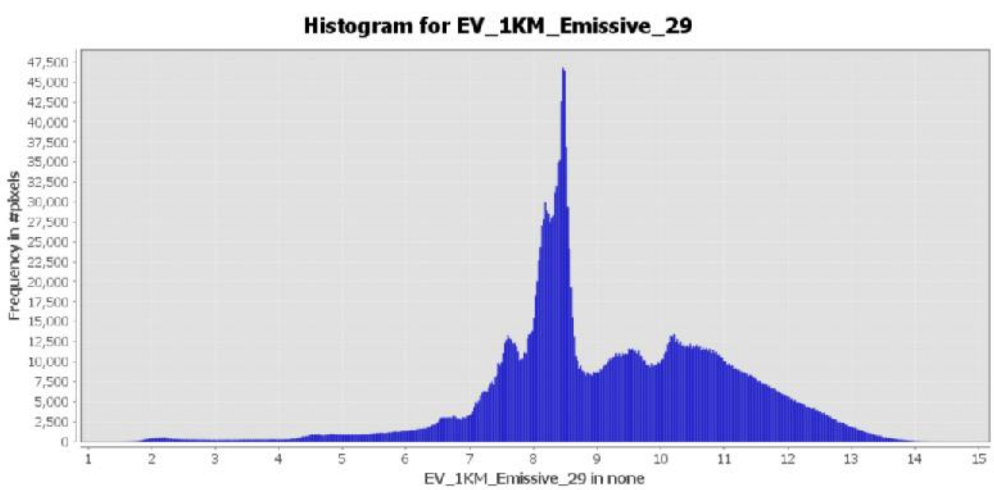
To check intensity of pixels, we perform an Histogram analysis.



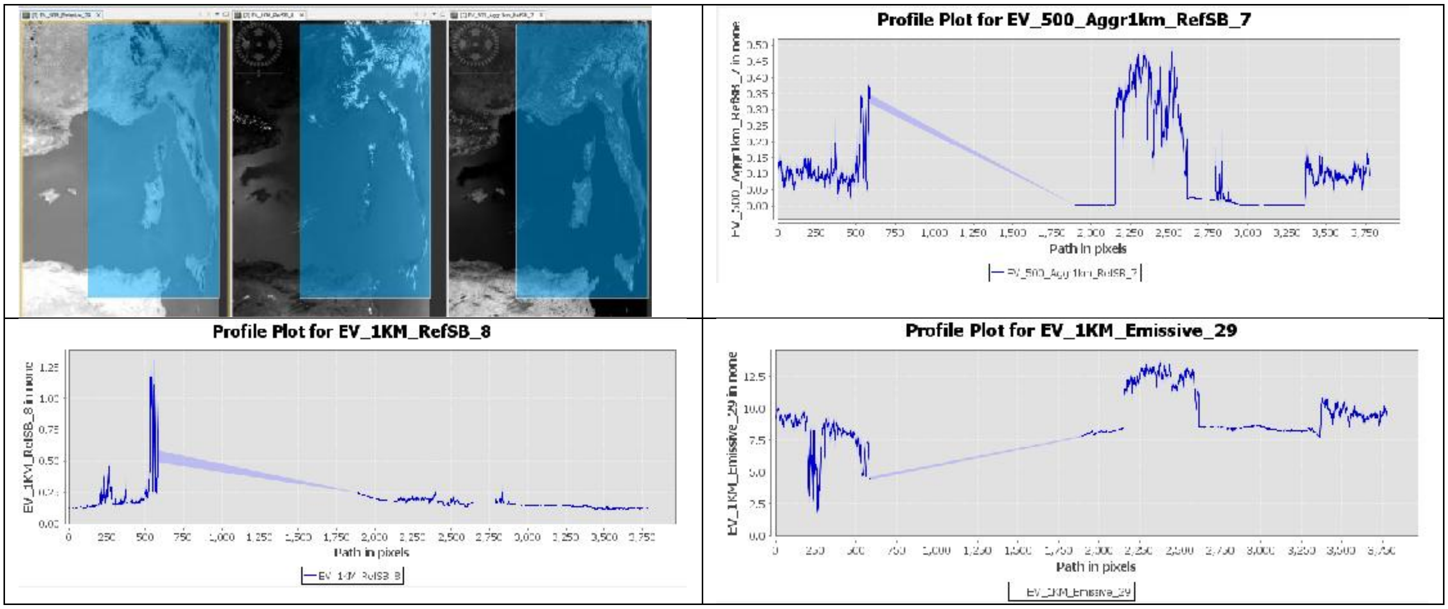
Intensity of pixels in this channel is extremely low



The highest observed frequency in this ref. channel is around 0.175

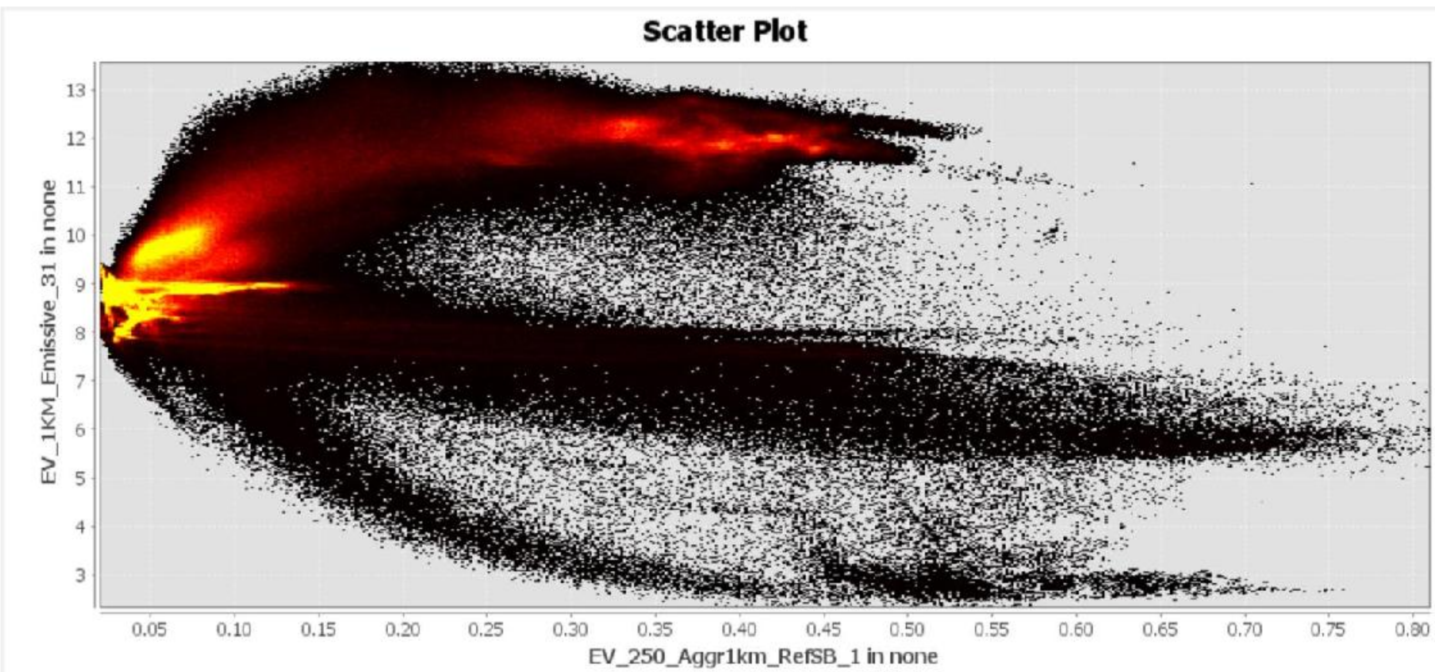


The highest observed frequency for this emissive channel is around 8.6



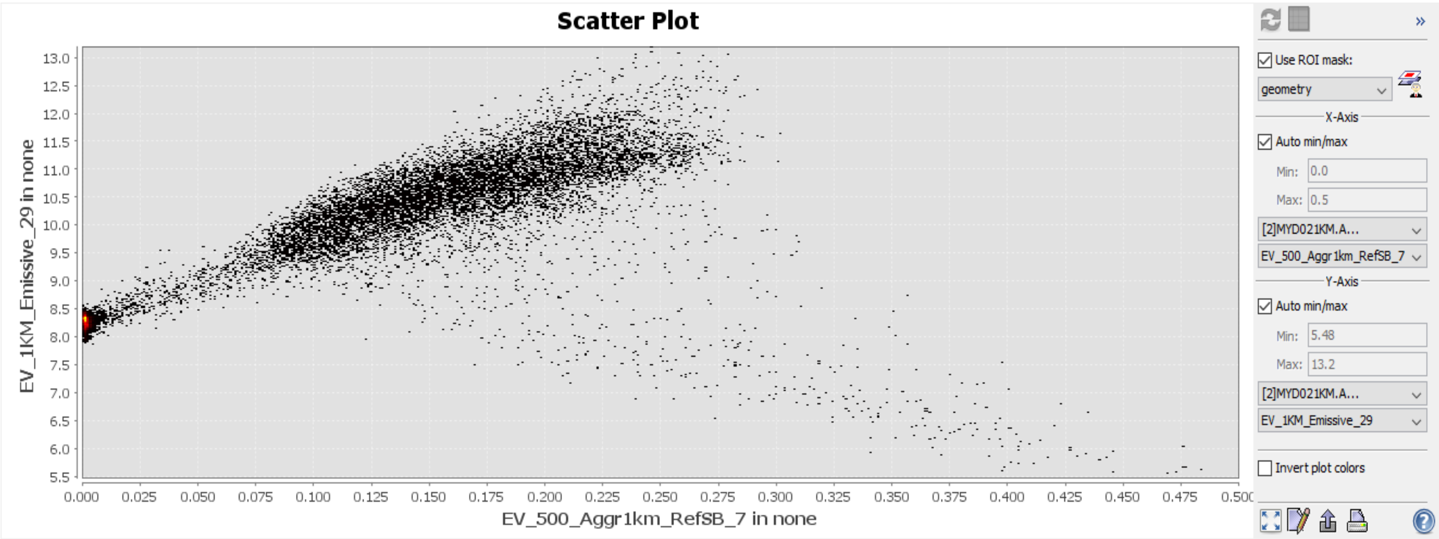
5. Perform and display channel data correlation of the whole image

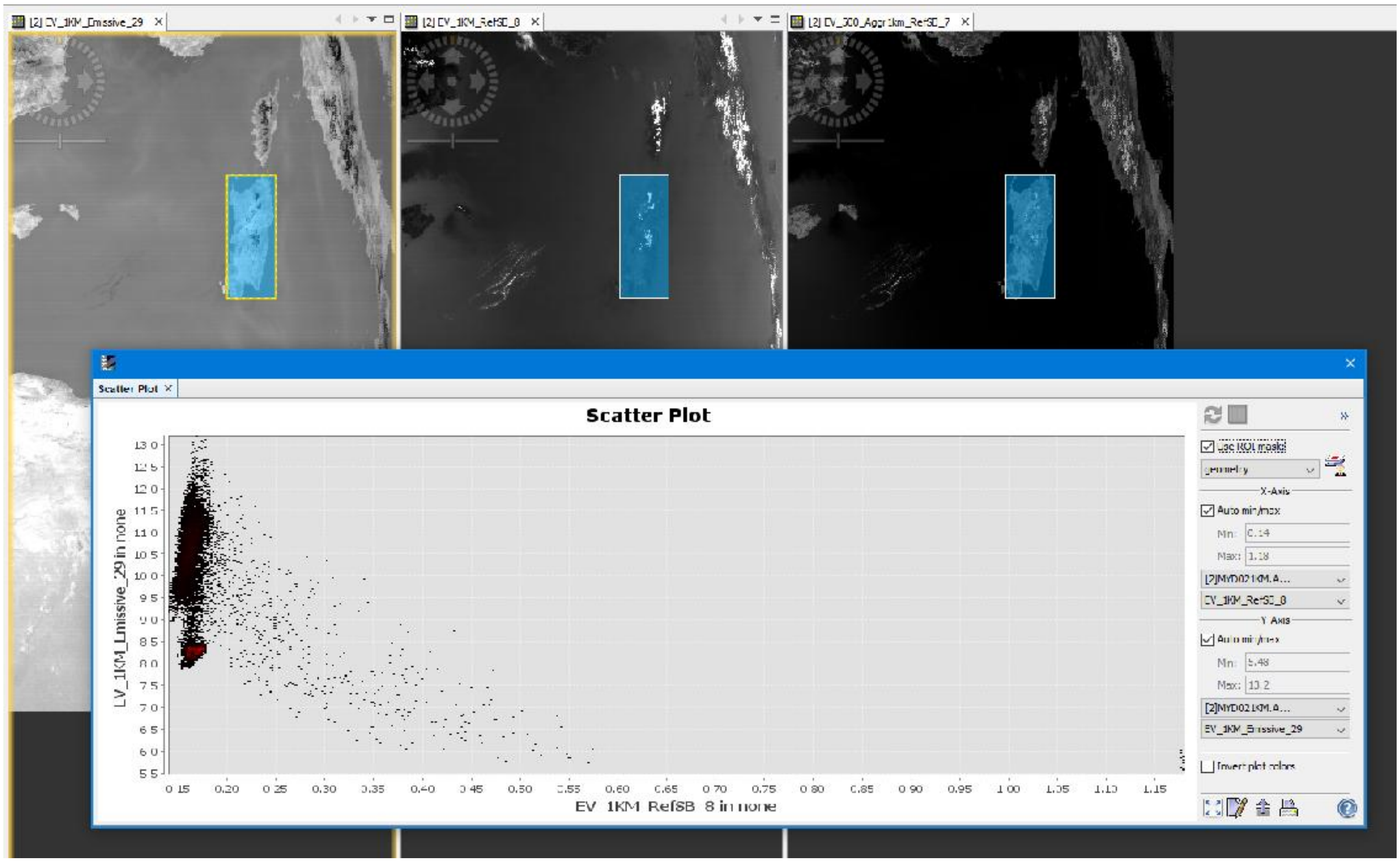
Correlation between reflective image and emissive image on the Scatter Plot shows the correlation between those channels. Pixel distributions for the two specified image channels display in the scatter plot using one channel as the x-axis and the other as the y-axis. They allow you to see where most pixels are concentrated.



- Land: low reflective at 06 micron and highly emitting at 11 microns
- Sea: low reflective at 0.6 micron and medium-to-high emitting at 11 microns
- Clouds: highly reflective at 0.6 micron and low-to--medium emitting at 11 microns

6. Perform and display channel data correlation of selected ROI (Region of Interest)





We obtain the corresponding scatter plot of selected ROI (Sardegna):

7. Perform and display principal component analysis

The PCA consists of remapping the information of the input co-registered images into a set of new images. For doing PCA, we select the first seven reflective bands.

After run PCA, as a result 7 components were generated, the components 3, 4, 5, 6 and 7 carry most of the information, whereas, the component 1 is almost all noise, and the components 2 has some information, but presents some noise features.

PCA results are showed on Figure 7.1 bellow.

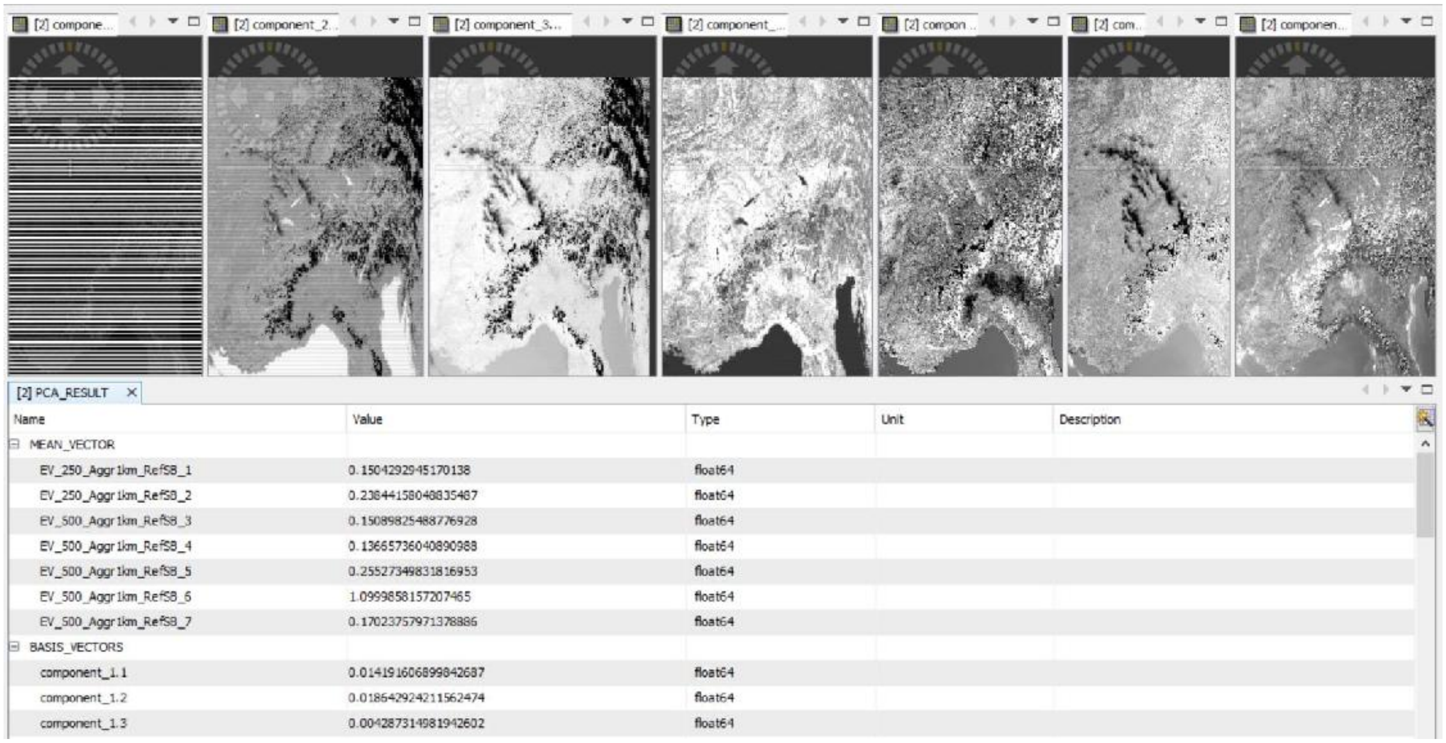


Figure 7.1 – PCA Results

8. Perform display and interpret unsupervised classification with at least 3 classes (sea, land, cloud)

An unsupervised K-means group analysis was performed with 3 groups. after some trials with a random selection of bands as feature sets, but no satisfactory results were obtained. In contrast, with fewer band parameters, discarding the noisy bands obtained an acceptable classification of land, sea and cloud. (Fig.8.1)

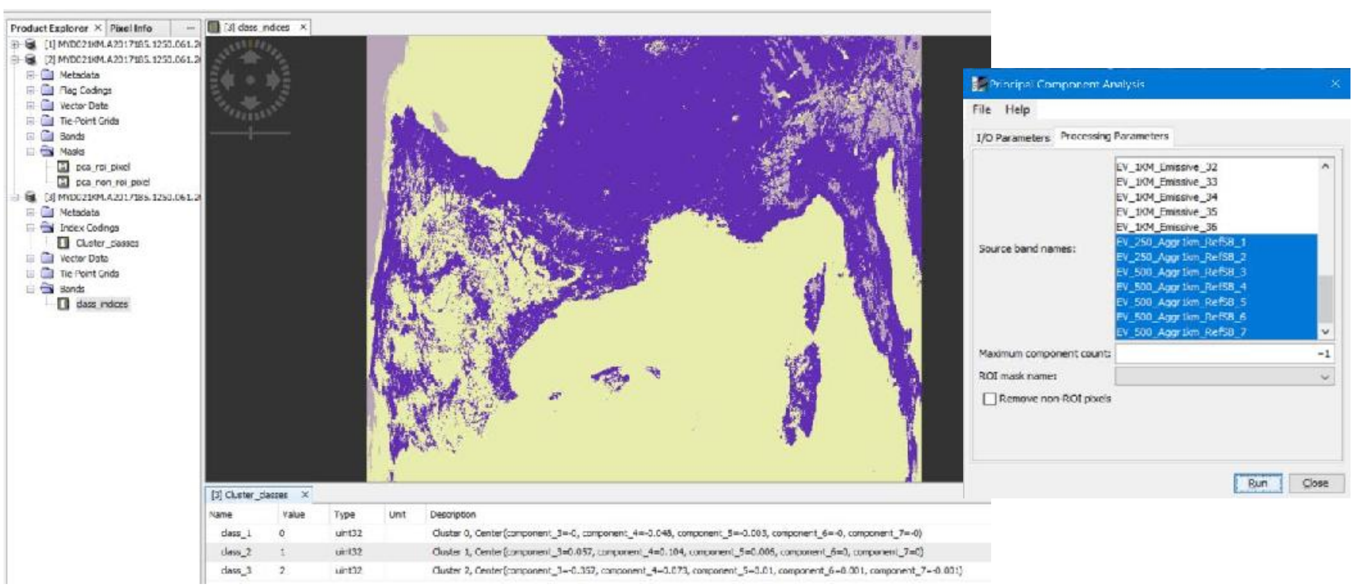


Figure 8.1 Unsupervised Classification Result

The process tool appropriately classified the section into different groups:

- sea (yellow)
- land (purple)
- cloud (gray)

9. Perform display and interpret supervised classification- 4 classes (sea, land, cloud)

To perform supervised classification, we first create an RGB visible composite True Color Image and False Color Image using the following parameters: (Figure 9.1)

TRUE COLOR IMAGE

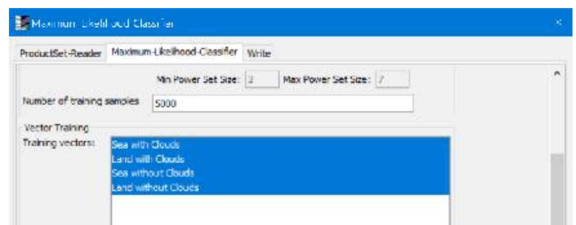
- Red: Open EV_250_Aggr1km_RefSB_1 (645 nm) – MODIS Band 1 [620-670 nm]
- Green: Open EV_250_Aggr1km_RefSB_4 (555 nm) – MODIS Band 4 [545-565 nm]
- Blue: Open EV_250_Aggr1km_RefSB_3 (469 nm) – MODIS Band 3 [459-479 nm]

FALSE COLOR IMAGE

Red: Open EV_250_Aggr1km_RefSB_2 (859 nm) – MODIS Band 2 (NIR)
Green: Open EV_250_Aggr1km_RefSB_1 (645 nm) – MODIS Band 4 (VISIBLE RED)
Blue: Open EV_250_Aggr1km_RefSB_4 (555 nm) – MODIS Band 3 (BLUE)

Then we create various data containers like:

- Land without Clouds
- Sea without Clouds
- Clouds over Water
- Clouds over Land



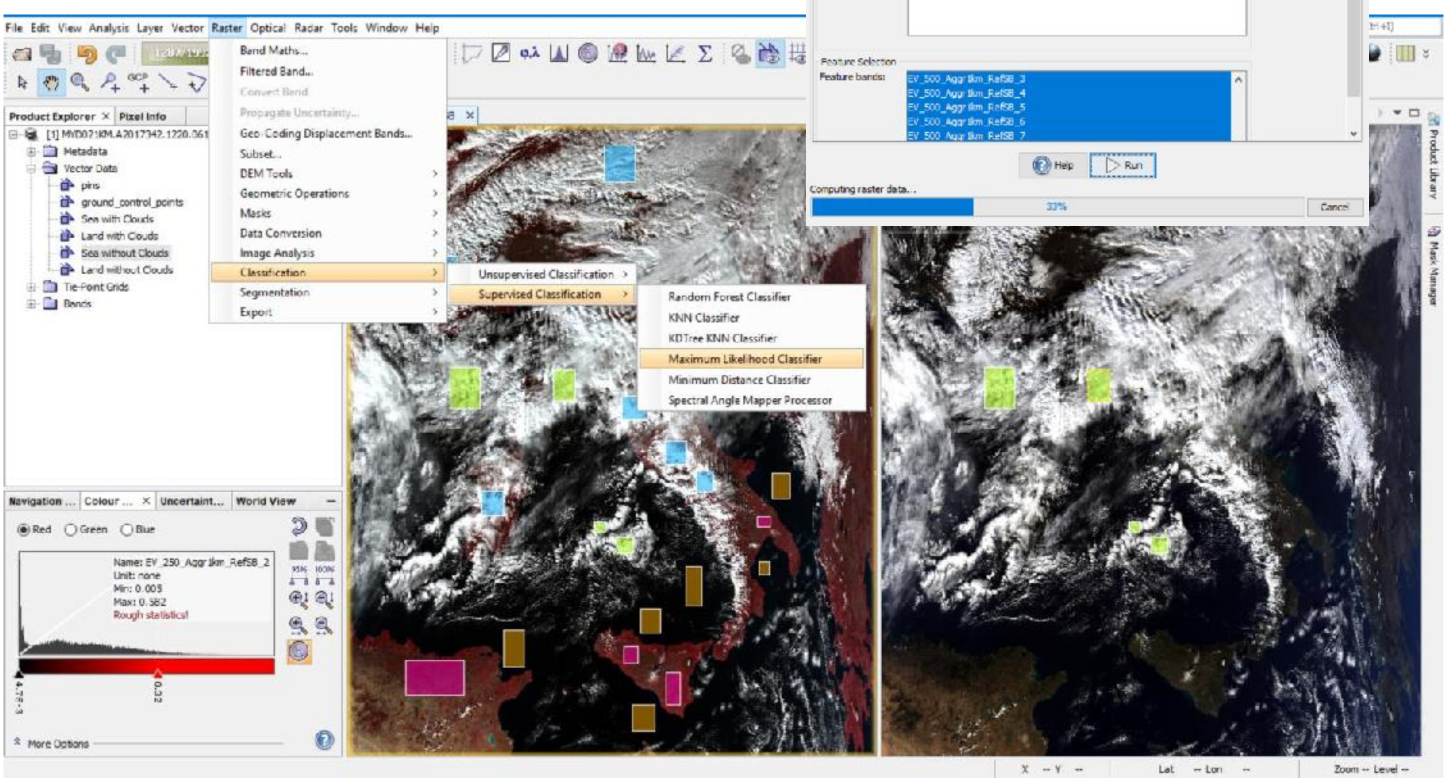


Figure 9.1 – Maximum Likelihood Classifier

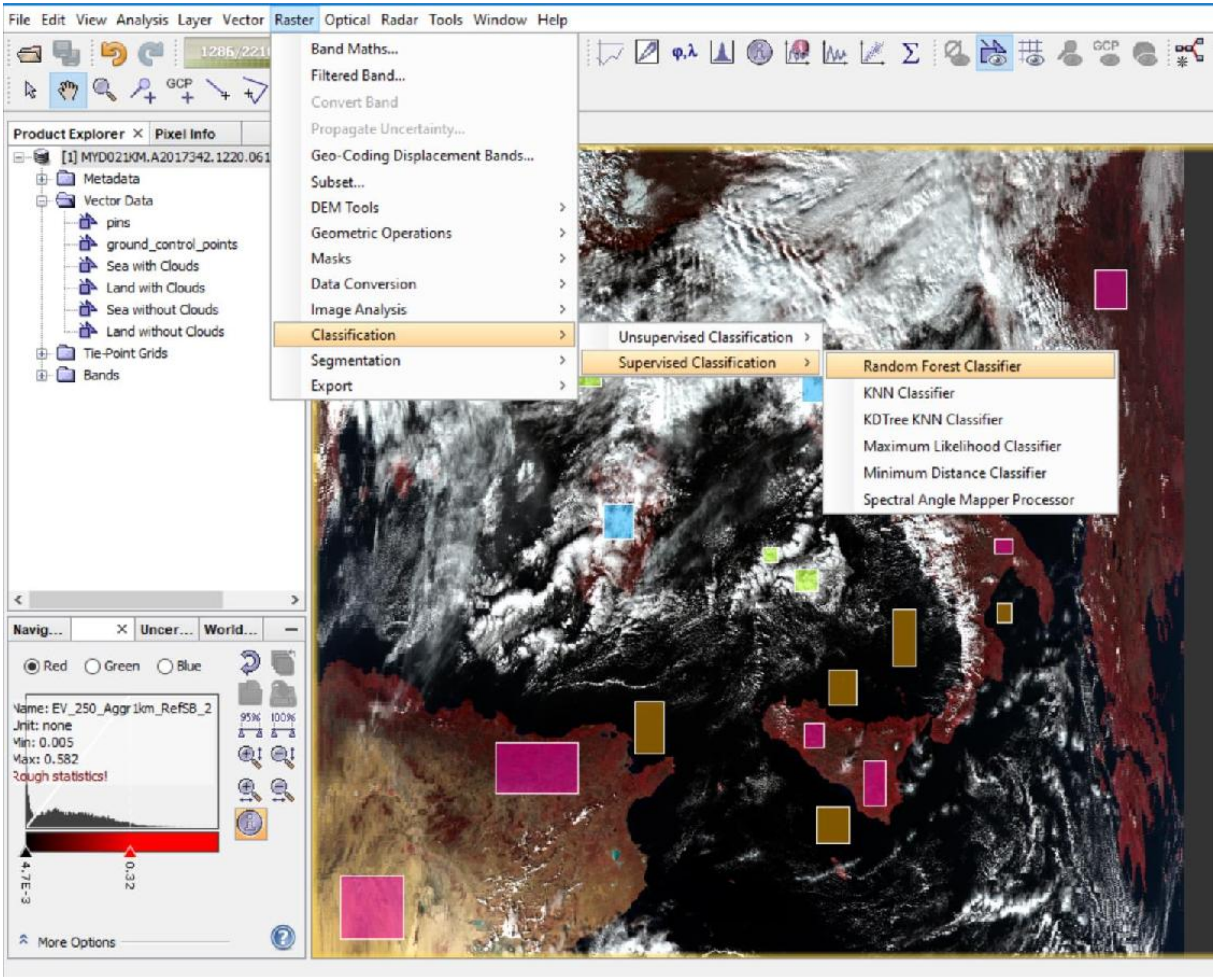


Figure 9.2 – Random Forest Classifier

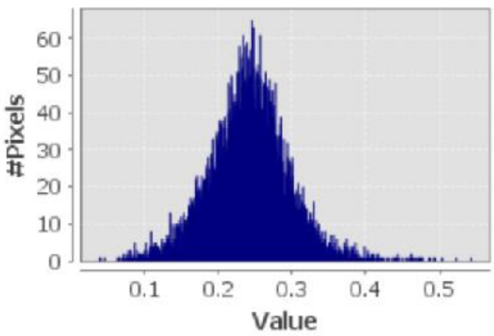
Comparatively:

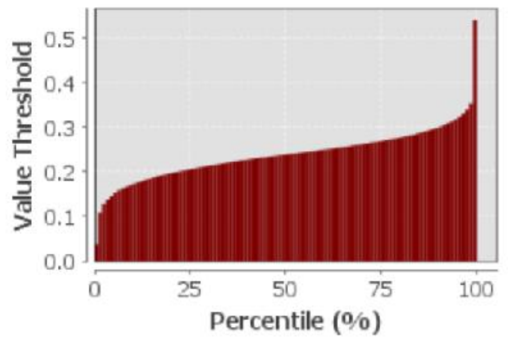
- The land is identified better by Random Forest
- The water is identified better by Random Forest
- The clouds are better identified by Random Forest

Overall, the most efficient classifier among these would be the **Random Forest Classifier**. (Figure 9.4)

EV_250_Aggr1km_RefSB_2 with ROI-mask **Sea with Clouds**

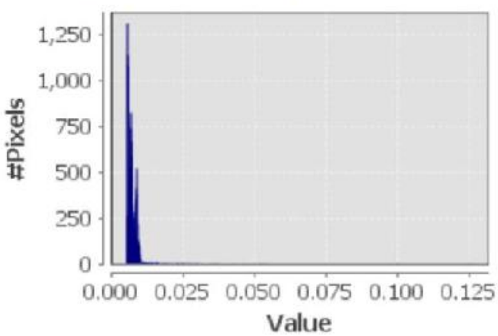
#Pixels total:	10987
Minimum:	0.0373
Maximum:	0.5420
Mean:	0.2398
Sigma:	0.0530
Median:	0.2400
Coef Variation:	0.4394
ENL:	5.1784
P75 threshold:	0.2710
P80 threshold:	0.2786
P85 threshold:	0.2892
P90 threshold:	0.3023
Max error:	5.0468E-4

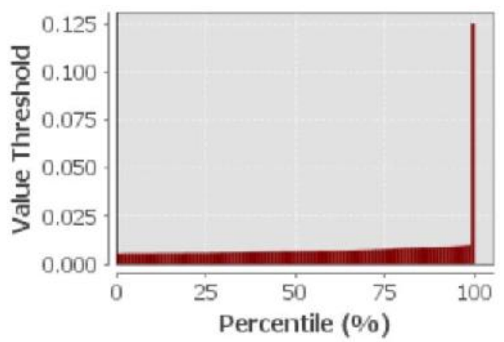




EV_250_Aggr1km_RefSB_2 with ROI-mask **Sea without Clouds**

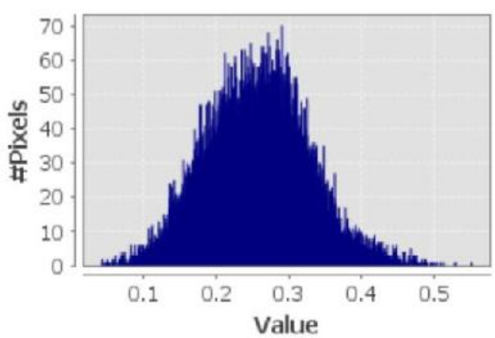
#Pixels total:	14684
Minimum:	0.0055
Maximum:	0.1253
Mean:	0.0072
Sigma:	0.0024
Median:	0.0067
Coef Variation:	2.9599
ENL:	0.1141
P75 threshold:	0.0079
P80 threshold:	0.0085
P85 threshold:	0.0087
P90 threshold:	0.0088
Max error:	1.198E-4

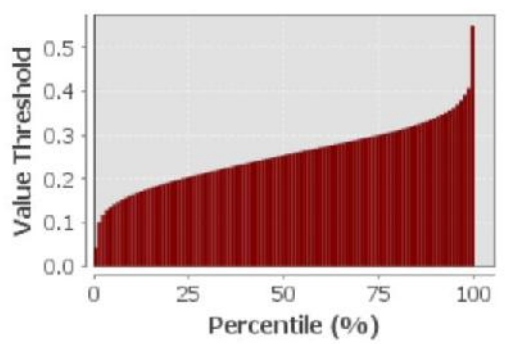




EV_250_Aggr1km_RefSB_2 with ROI-mask **Land with Clouds**

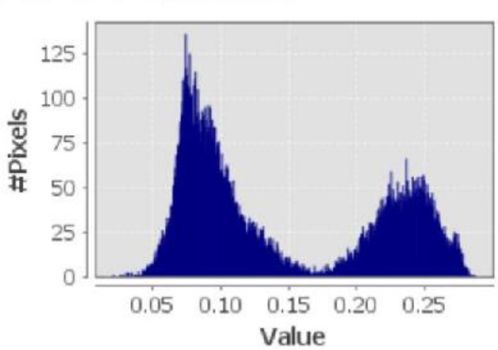
#Pixels total:	19295
Minimum:	0.0416
Maximum:	0.5503
Mean:	0.2554
Sigma:	0.0702
Median:	0.2561
Coef Variation:	0.5295
ENL:	3.5663
P75 threshold:	0.3016
P80 threshold:	0.3128
P85 threshold:	0.3260
P90 threshold:	0.3433
Max error:	5.087E-4





EV_250_Aggr1km_RefSB_2 with ROI-mask **Land without Clouds**

#Pixels total:	26782
Minimum:	0.0207
Maximum:	0.2862
Mean:	0.1435
Sigma:	0.0726
Median:	0.1070
Coef Variation:	0.9115
ENL:	1.2037
P75 threshold:	0.2238
P80 threshold:	0.2326
P85 threshold:	0.2411
P90 threshold:	0.2488
Max error:	2.6556E-4



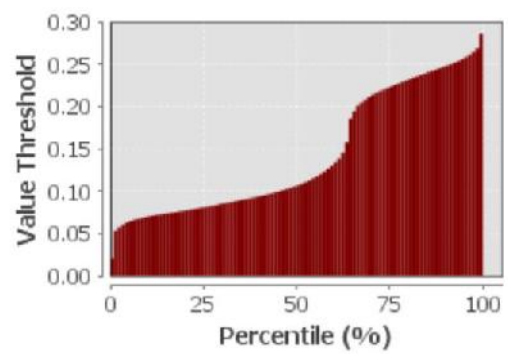


Figure 9.3 - Statistics

Result:

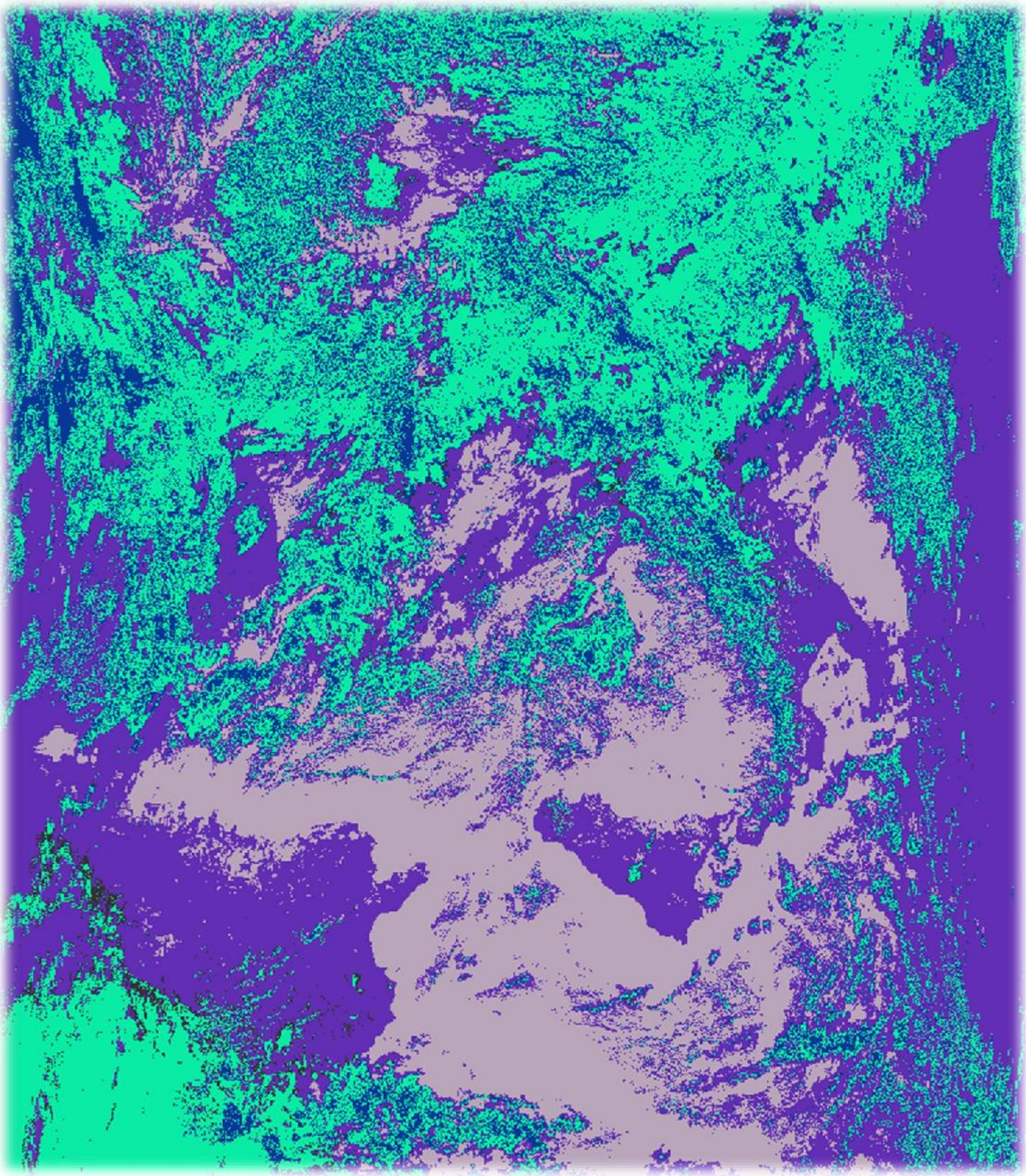


Figure 9.4 - Supervised Classification – Random Forest Classifier Result

RF_Classifier_Winter - Notepad

File Edit Format View Help

```
RandomForest classifier RF_Classifier_Winter

Cross Validation
Number of classes = 4
   class 0.0: Land
    accuracy = 1.0000 precision = 1.0000 correlation = 1.0000 errorRate = 0.0000
    TruePositives = 1250.0000 FalsePositives = 0.0000 TrueNegatives = 3335.0000 FalseNegatives = 0.0000
   class 1.0: Sea
    accuracy = 1.0000 precision = 1.0000 correlation = 1.0000 errorRate = 0.0000
    TruePositives = 1250.0000 FalsePositives = 0.0000 TrueNegatives = 3335.0000 FalseNegatives = 0.0000
   class 2.0: sea_clouds
    accuracy = 0.8674 precision = 0.6291 correlation = 0.6216 errorRate = 0.1326
    TruePositives = 553.0000 FalsePositives = 326.0000 TrueNegatives = 3424.0000 FalseNegatives = 282.0000
   class 3.0: Land_clouds
    accuracy = 0.8674 precision = 0.7662 correlation = 0.7067 errorRate = 0.1326
    TruePositives = 924.0000 FalsePositives = 282.0000 TrueNegatives = 3053.0000 FalseNegatives = 326.0000

Using Testing dataset, % correct predictions = 86.7394
Total samples = 9171
RMSE = 0.36415151375867877
Bias = -0.00959651035986897

Distribution:
   class 0.0: Land                      2500    (27.2598%)
   class 1.0: Sea                       2500    (27.2598%)
   class 2.0: sea_clouds                1671    (18.2205%)
   class 3.0: Land_clouds               2500    (27.2598%)

Testing feature importance score:
Each feature is perturbed 3 times and the % correct predictions are averaged
The importance score is the original % correct prediction - average
   rank 1   feature 1  : virtual_red             score: tp=0.4073 accuracy=0.2036 precision=0.2976 correlation=0.4397 errorRate=-0.2036 cost=-0.6576 GainRatio = 0.5476
   rank 2   feature 3  : virtual_blue            score: tp=0.1038 accuracy=0.0519 precision=0.1190 correlation=0.1303 errorRate=-0.0519 cost=-0.1921 GainRatio = 0.5524
   rank 3   feature 2  : virtual_green           score: tp=0.0939 accuracy=0.0470 precision=0.1075 correlation=0.1162 errorRate=-0.0470 cost=-0.1698 GainRatio = 0.5480


Power set evaluation:
RF_Classifier_Winter.1: cv 86.24 % virtual_red, virtual_green,
RF_Classifier_Winter.2: cv 86.26 % virtual_red, virtual_blue,
RF_Classifier_Winter.3: cv 86.24 % virtual_green, virtual_blue,
RF_Classifier_Winter.4: cv 86.37 % virtual_red, virtual_green, virtual_blue,

TOP Classifier = RF_Classifier_Winter.4 at 86.37 %
```

Ln 1, Col 1 | 100% | Windows (CRLF) | UTF-8

Figure 9.5 - Results

10. Implement at least 3 formulas of 2-band and 3-band vegetation index (VI) using SNAP processing tools

We have chosen the following 3 formulae: **NDVI**, **EVI** and **SAVI**.

NDVI

A Vegetation Index (VI) is a spectral transformation of two or more bands designed to enhance the contribution of vegetation properties and allow reliable spatial and temporal inter-comparisons of terrestrial photosynthetic activity and canopy structural variations. Natural surfaces reflect solar radiation almost similarly in the Near Infrared (NIR) and Red wavelengths. We used:

- NIR = EV_250_Aggr1km_RefSB_1 (645nm)
- RED = EV_250_Aggr1km_RefSB_2 (859nm)

$$NDVI = \frac{NIR - Red}{NIR + Red}$$

NDVI - Normalized Difference Vegetation Index (performed with the Raster Band Math Calculator)

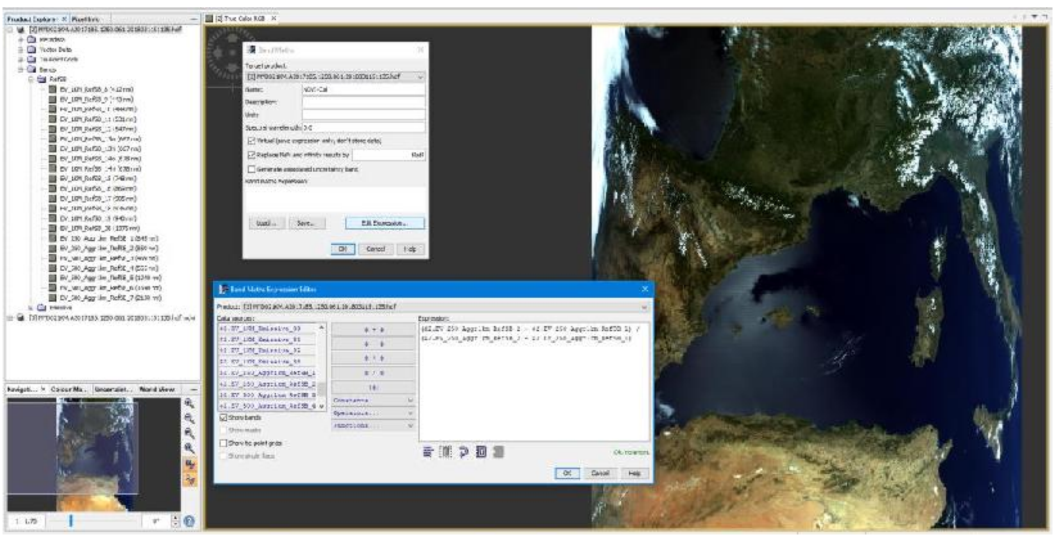


Figure 10.1

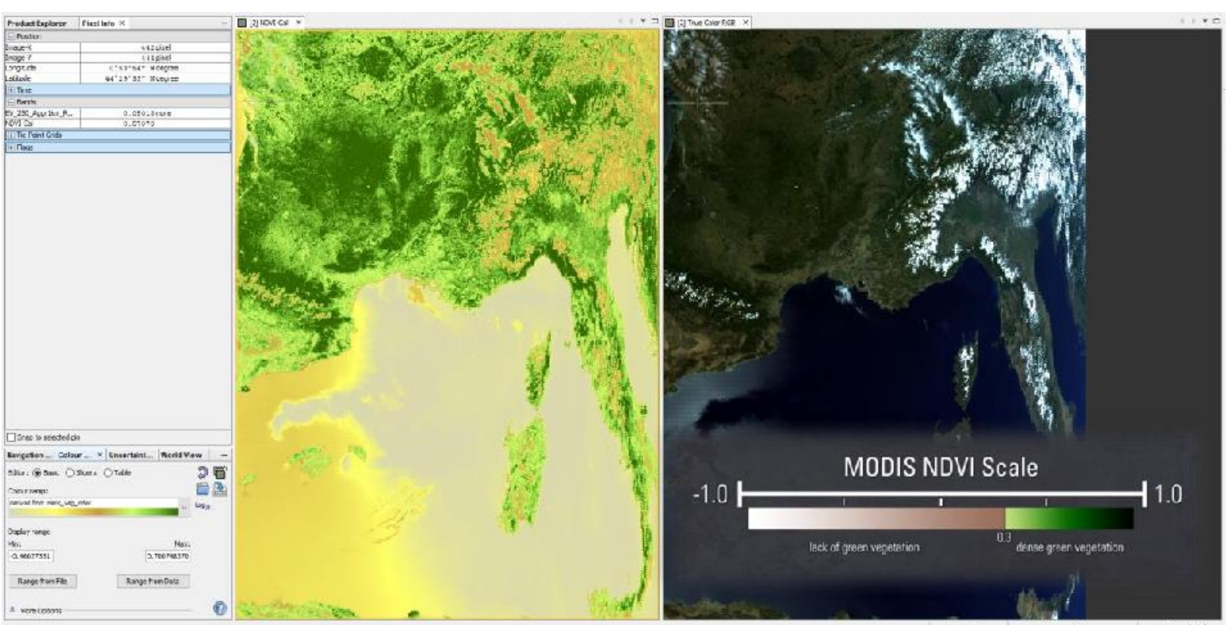


Figure 10.2 – NDVI as indicator of vegetation greenness

Once the process is finished, we apply a color ramp to obtain a better visualization (Figure 10.2), we see how the pixel values are higher where there is vegetation (Values close to 1.0), instead, we obtain negative values for the regions covered with water

NDVI - Normalized Difference Vegetation Index (performed with the Optical Land thematic Processing tool)

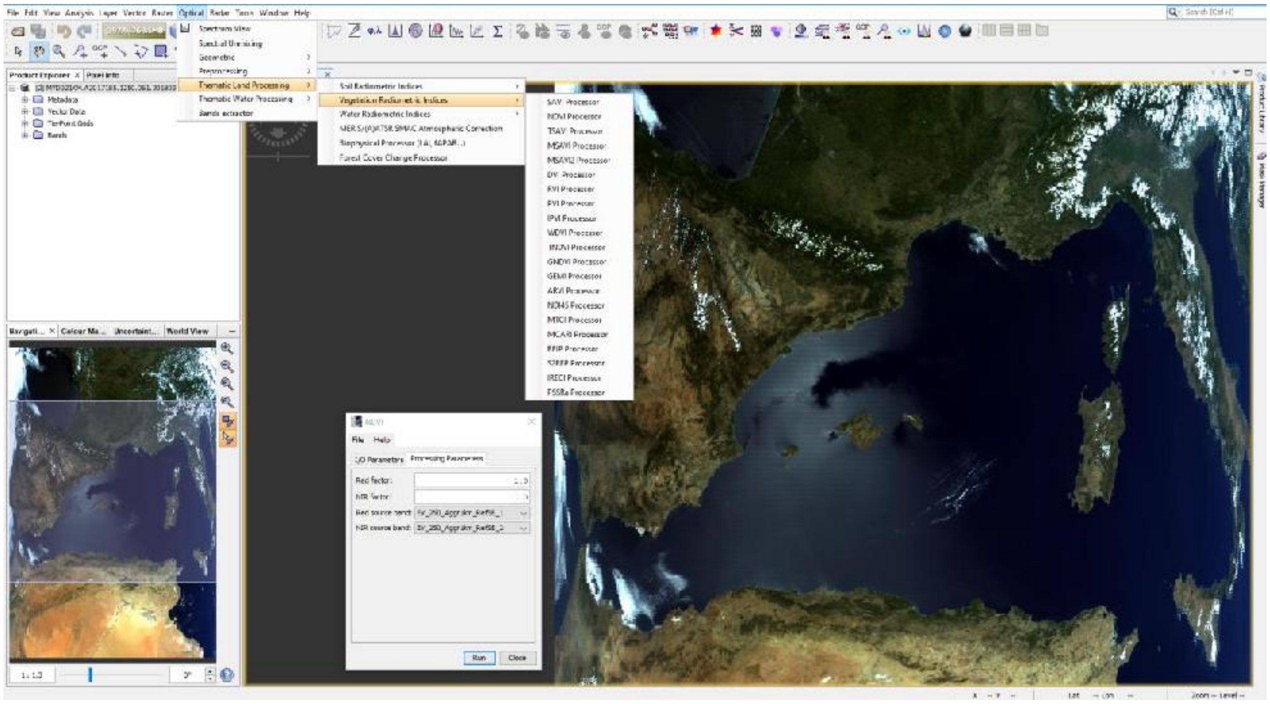


Figure 10.3

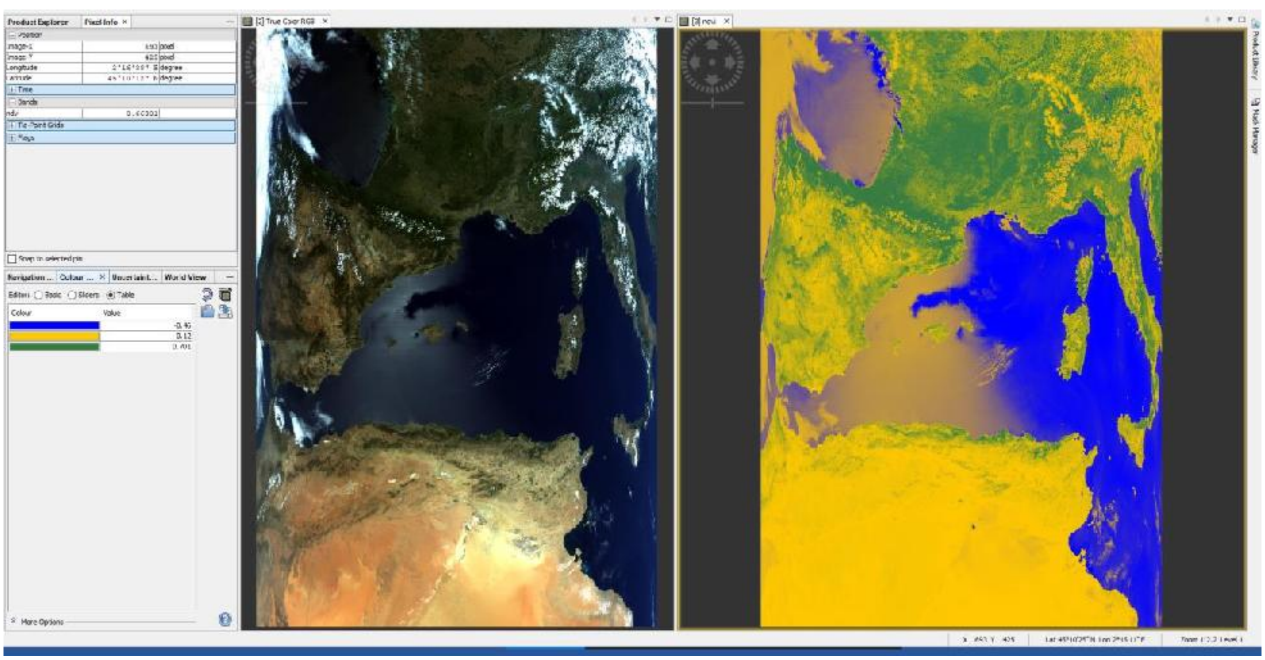


Figure 10.4

Through this tool, the results obtained were practically identical, this time the visualization was done by manually coloring the table

EVI

The Enhanced Vegetation Index was invented by Liu and Huete to simultaneously correct NDVI results for atmospheric influences and soil background signals, especially in areas of dense canopy. The value range for EVI is -1 to 1, and for healthy vegetation it varies between 0.2 and 0.8. Is designed to enhance the vegetation signal with improved sensitivity in high biomass regions and improved vegetation monitoring through a de-coupling of the canopy background signal and a reduction in atmosphere.

- Reduces saturation on dense biomass
- Decreases influence of bare soil
- Uses a ratio of Red, NIR and Blue

$$EVI = G * \frac{NIR - Red}{NIR + C1 * Red - C2 * Blue + L}$$

EVI - Enhanced Vegetation Index (performed with the Raster Band Math Calculator)

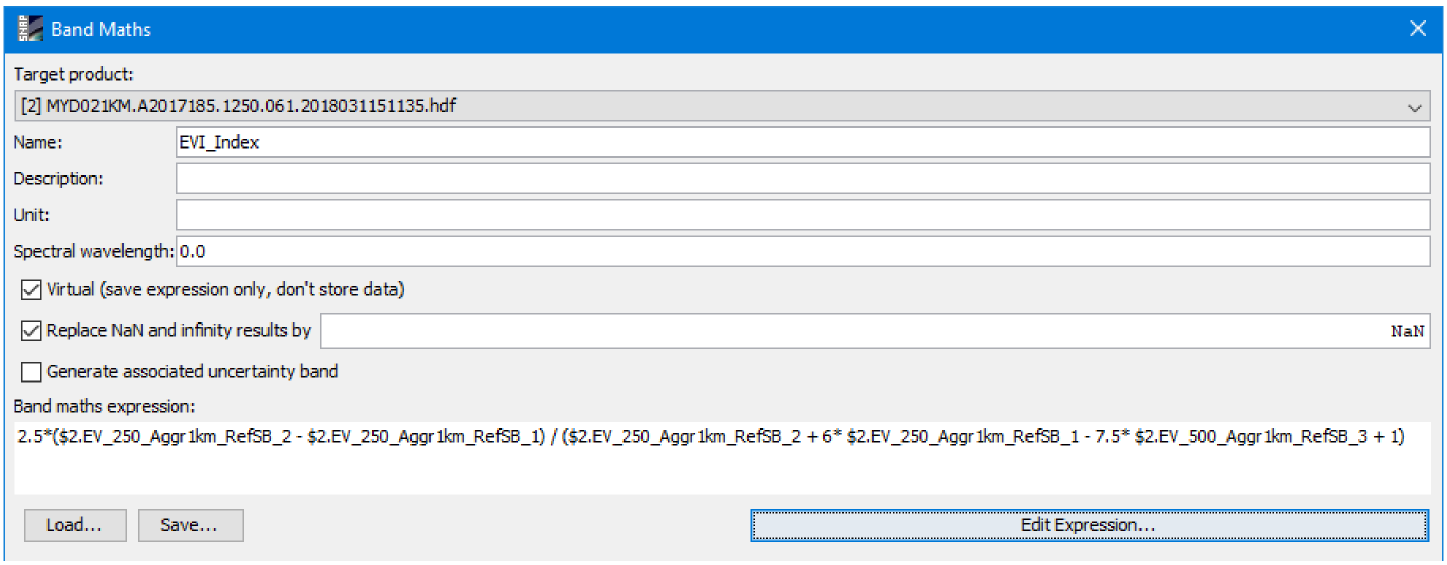


Figure 10.5 – Band Math Expression for EVI

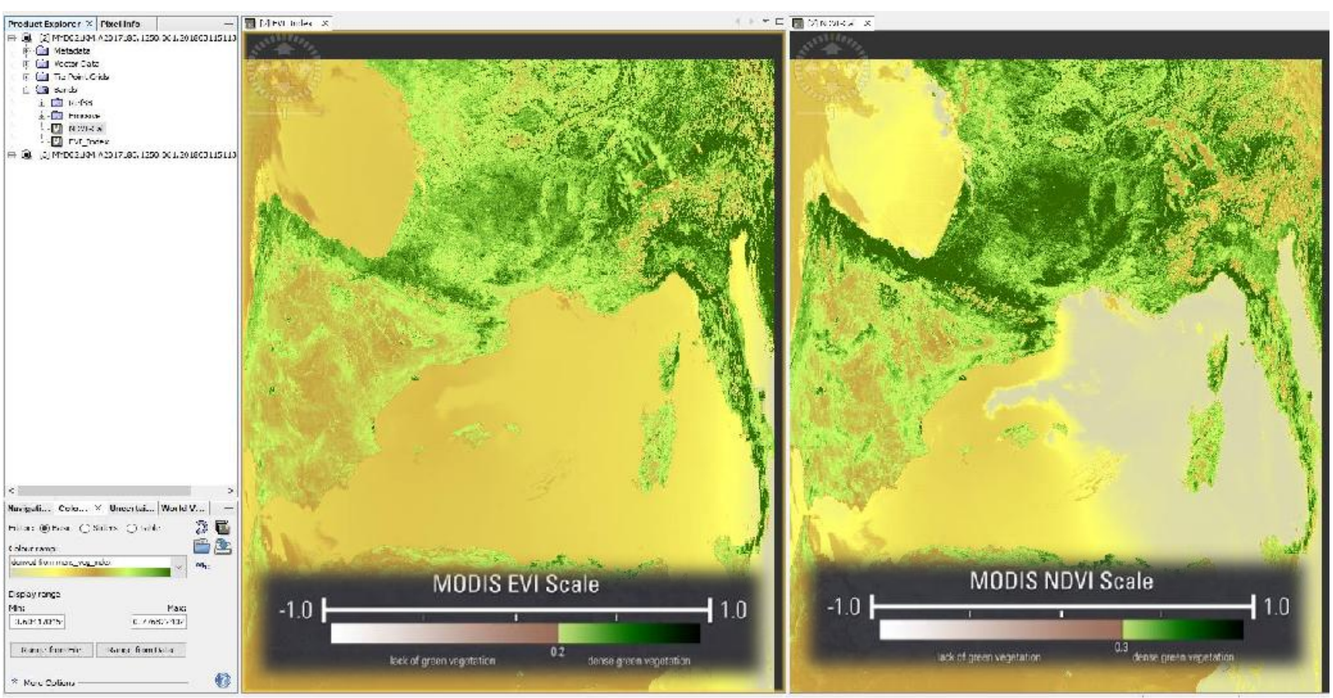


Figure 10.6 Comparison between EVI and NDVI Index

SAVI

$$SAVI = \frac{NIR - Red}{NIR + Red + L} * (1 + L)$$

The Soil Adjacent Vegetation Index was developed as a modification of the Normalize d Difference Vegetation Index to correct for the influence of soil brightness when vegetative cover is low. The SAVI is structured similar to the NDVI but with the addition of a soil brightness correction factor.
In areas where vegetative cover is low (i.e., < 40 percent) and the soil surface is exposed, the reflectance of light in the red and near -infrared spectra can influence vegetation index values. This is especially problematic when comparisons are being made across different soil types that may reflect different amounts of light in the red and near infrared wavelengths. The value of L varies by the amount or cover of green vegetation: in very high vegetation regions, L=0; and in areas with no green vegetation, L=1.Generally, an L=0.6 works well in most situations and is the default value used. When L=0, then SAVI = NDVI

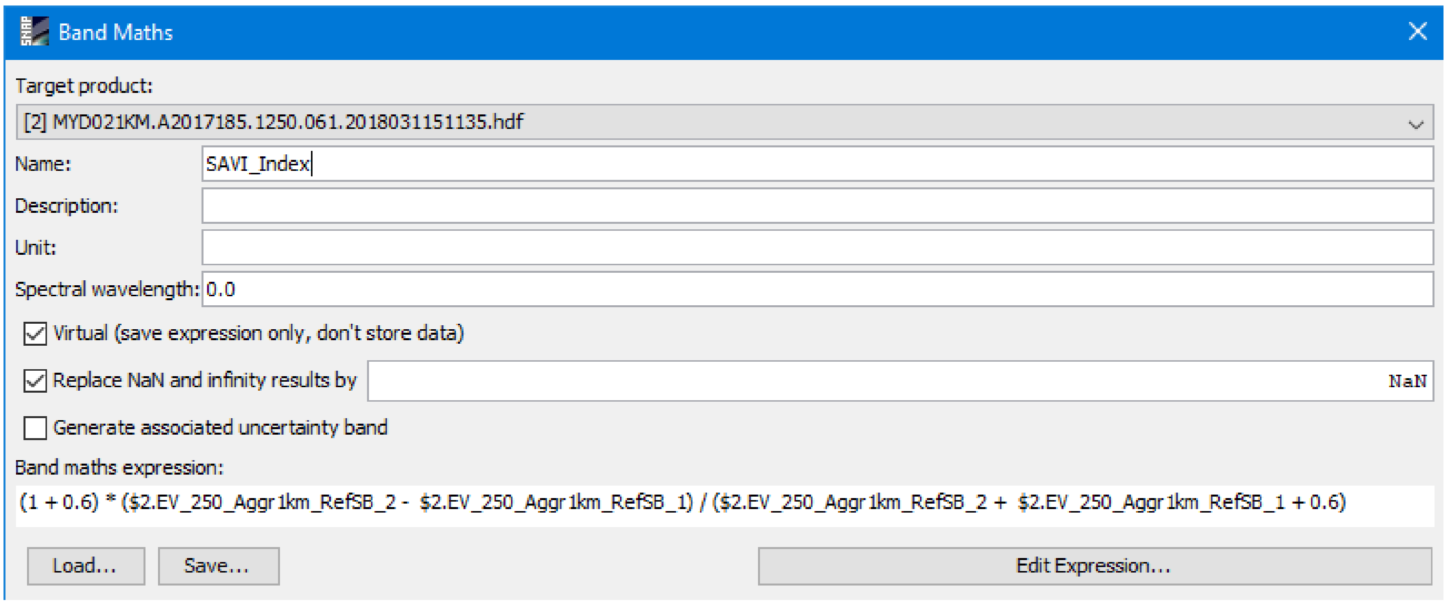


Figure 10.7 – Band Math Expression for SAVI

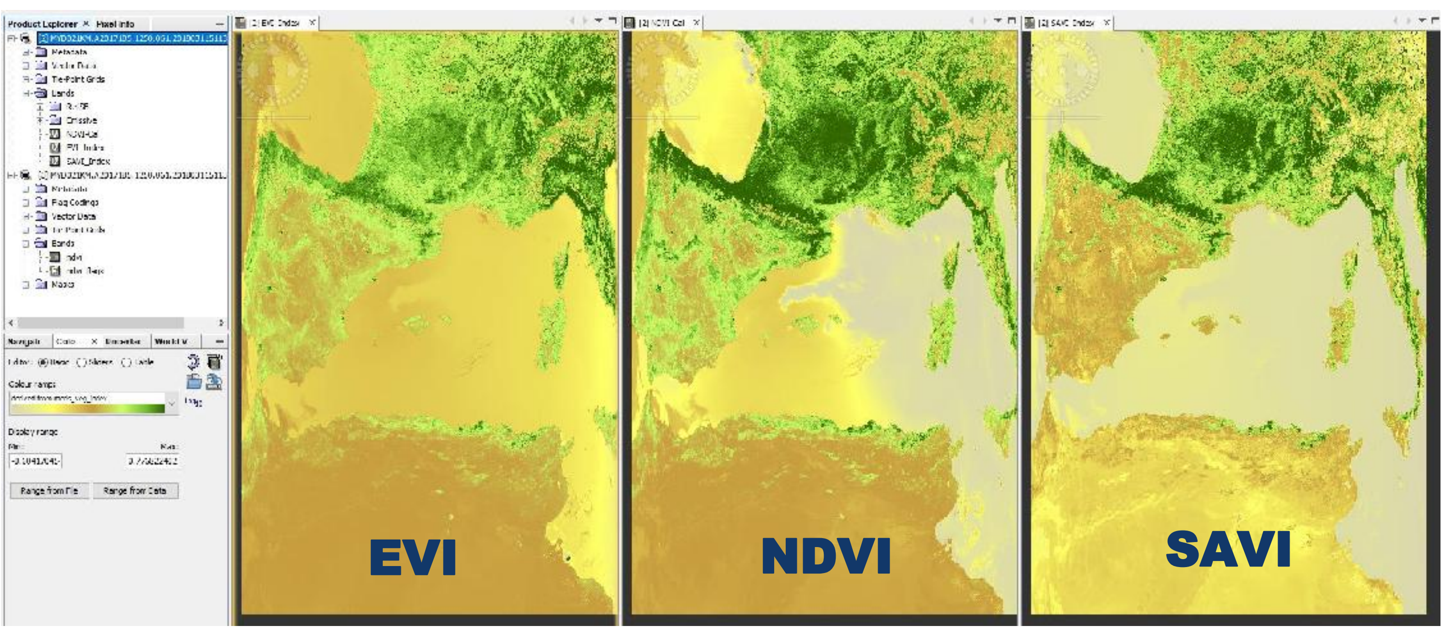


Figure 10.8 Comparison between EVI NDVI and SAVI Index

11. Apply at least 3 VI formulas to land pixels and interpret their output results and differences.

To apply VI formulas to land pixels, we first have to mask the sea, using Raster, Masks, Land / Sea Mask, and then selecting the "mask the sea" option. Then we apply the same NDVI, EVI and SAVI in this image, and then we compare it with the results obtained previously. (Figure 11.1)

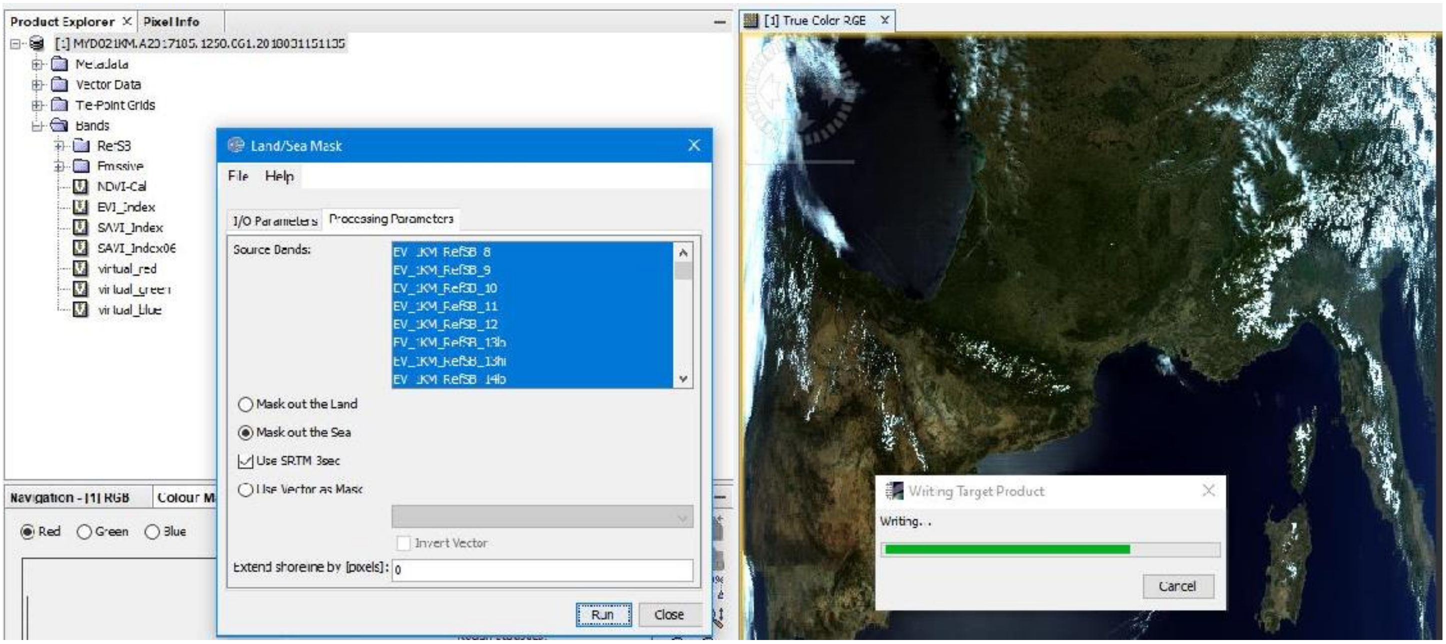


Figure 11.1 – Mask out the sea.

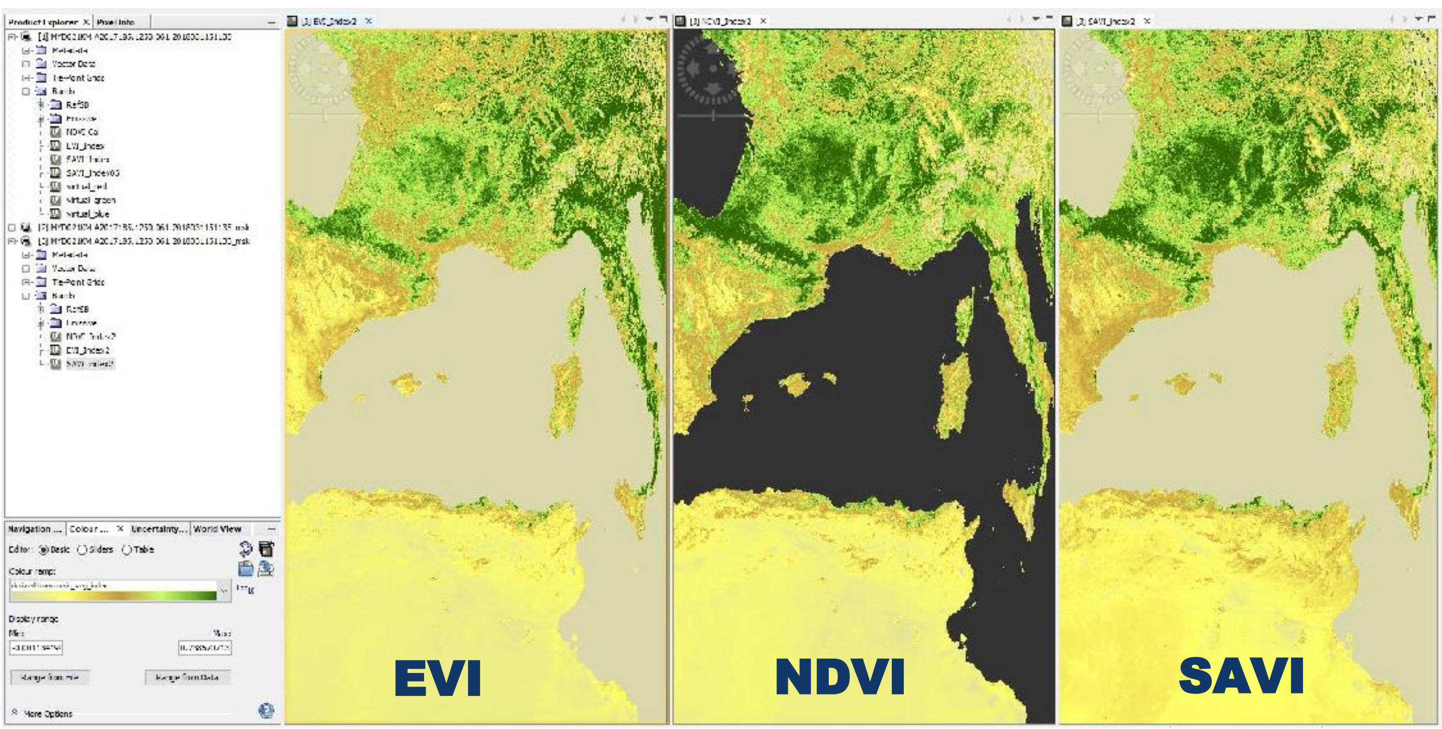


Figure 11.2 – Mask out the sea.

When comparing the different VI indexes applied to the original pixels of the image and to the terrestrial, we found that a better result is obtained when the formulas are applied to the terrestrial pixels only.

It could also be said that the resulting image is easier to interpret for those who analyze it.

Comparisons:

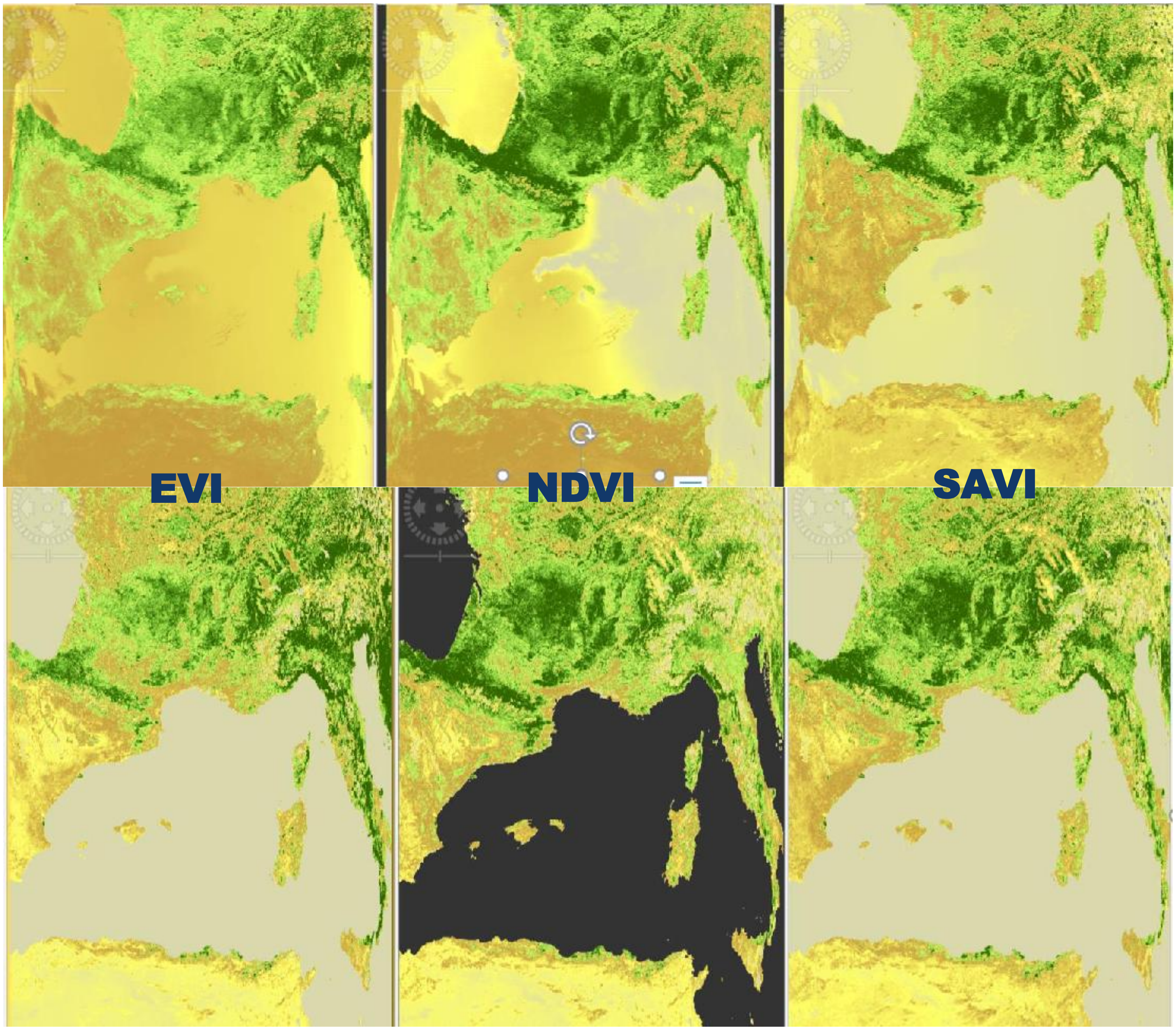


Figure 11.3 – Comparison: Top = No Mask, Bottom: Mask out

- There is a consistency in the defined vegetation areas
- When applying the mask, the ground pixels are better defined
- When the formula is applied to the entire image in some places, land and sea are confused
- From our personal point of view, EVI presents a better definition regarding the classification of the vegetal layer.

12. Perform an unsupervised classification by using an arbitrary VI index instead of MODIS channel

In this case, we used the EVI index to perform an unsupervised clustering, one for K-Means and one for EM Cluster. We compare both of them with the previous results with 6 and 14 classes

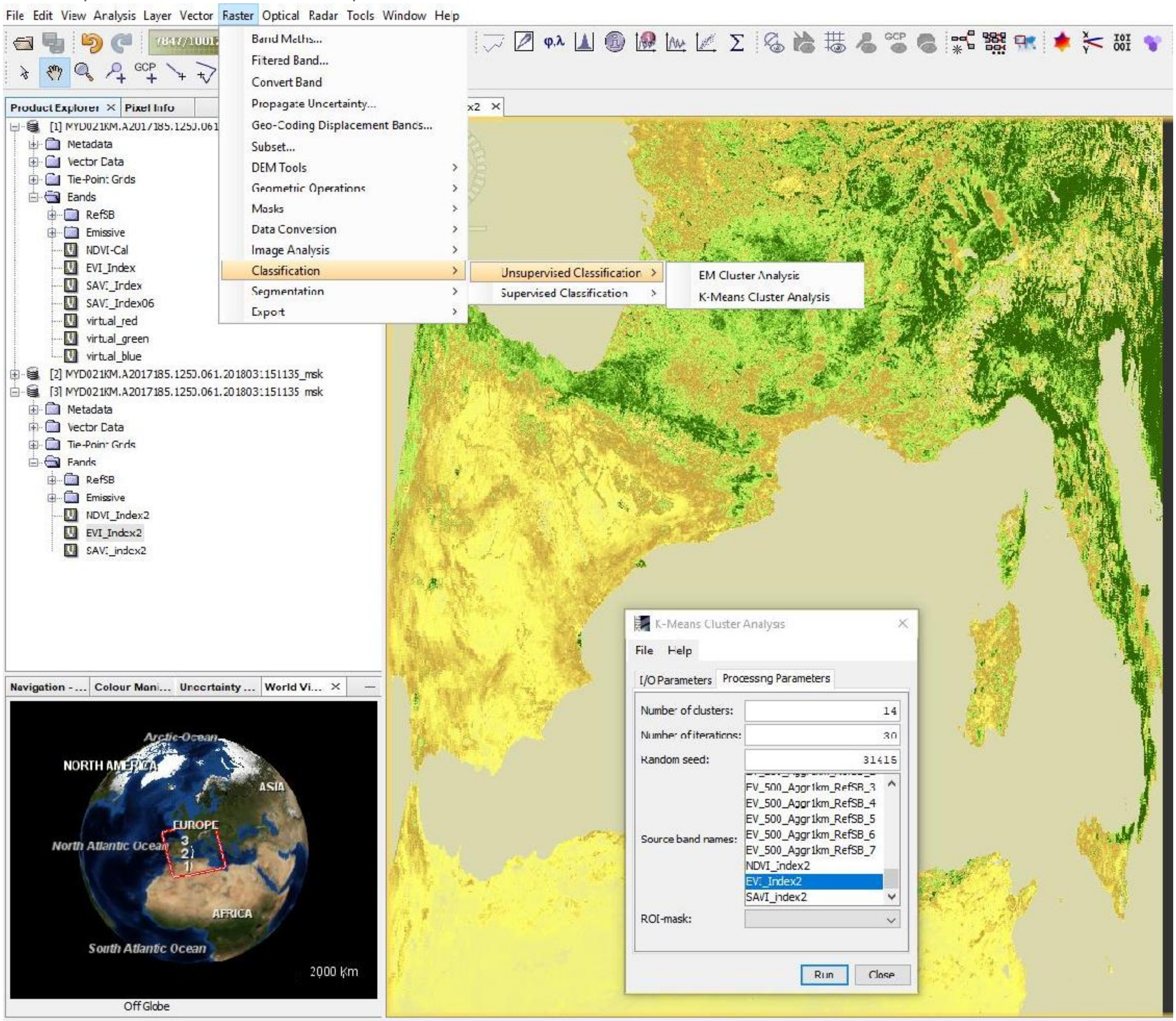


Figure 12.1 – Unsupervised Classification - KMeans

Result:

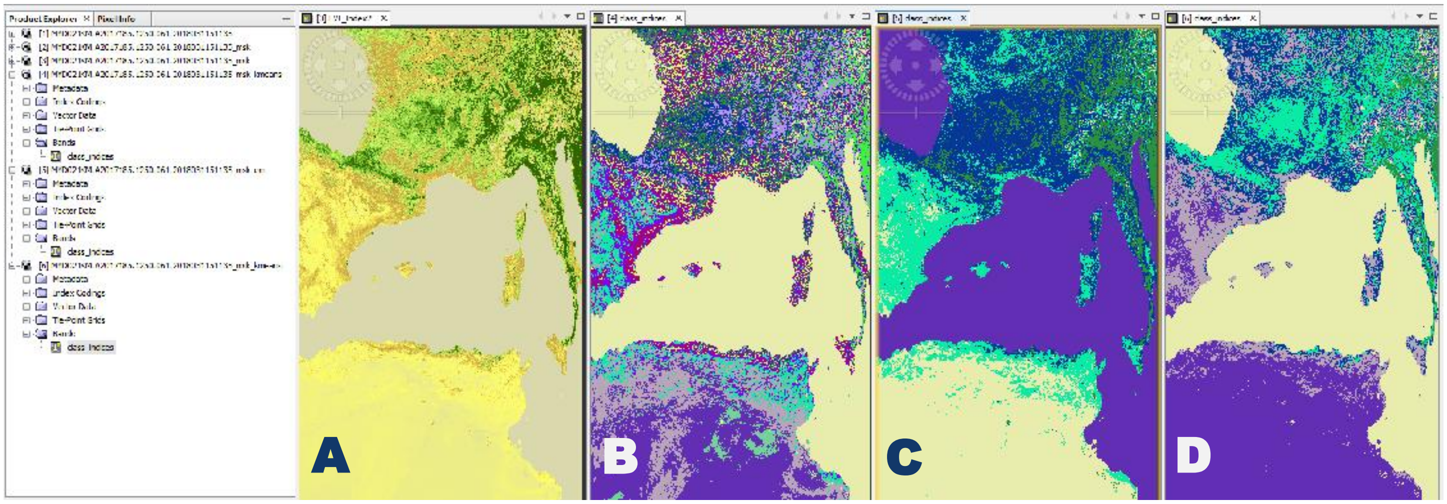


Figure 12.2 – Unsupervised Classification Results

A: EVI image (Mask out over water)
B: EM Cluster Analysis with 14 classes
C: K-Means with 6 classes
D: EM Cluster Analysis with 6 clases

13. Download a second MODIS image at 1-km resolution over the same geographical area of interest during a winter season. Apply the same selected VI index and qualitatively compare the differences.

We can see in the images that there is an increase in cloudiness for both areas (land, sea) in the winter period, compared to what we had in the same region in the summer.

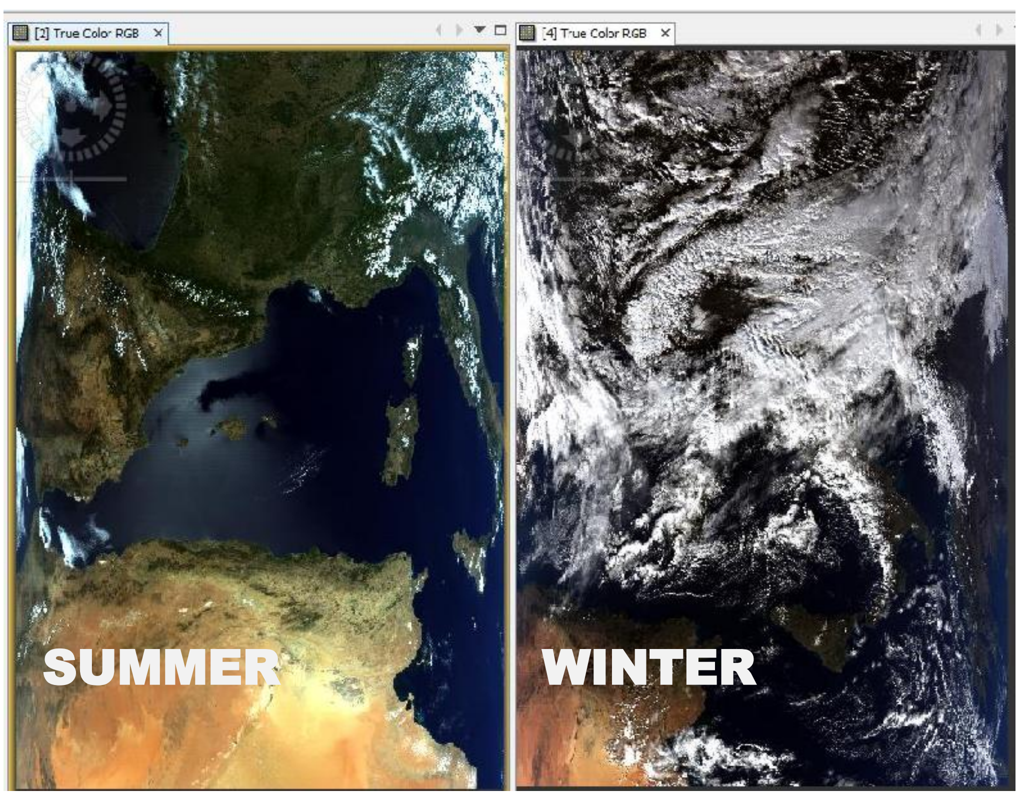


Figure 13.1 – Italy Summer and Winter

Winter Image in False Color band Composite with NDVI and NDVI with Mask Out over Sea

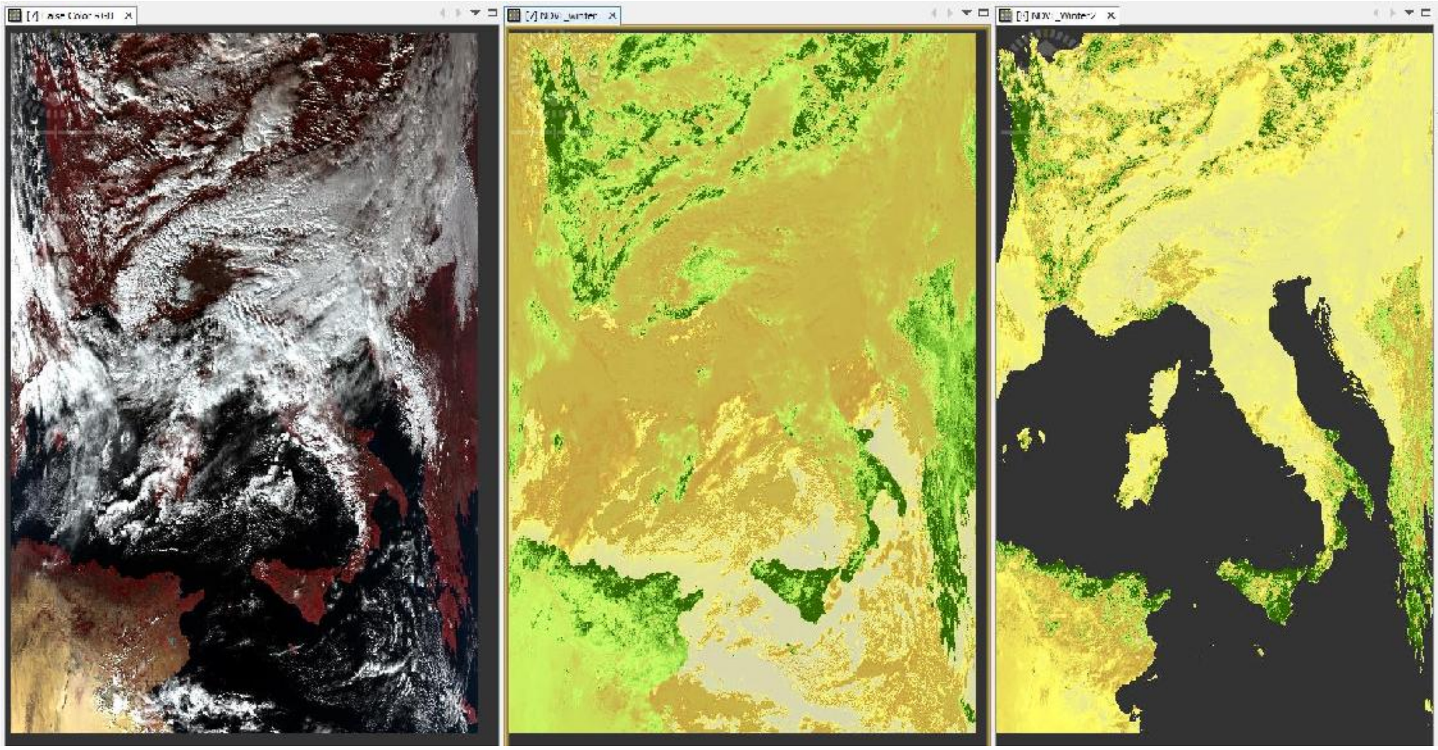


Figure 13.2 – Winter False Color Composite and NDVI

After perform the three different VI

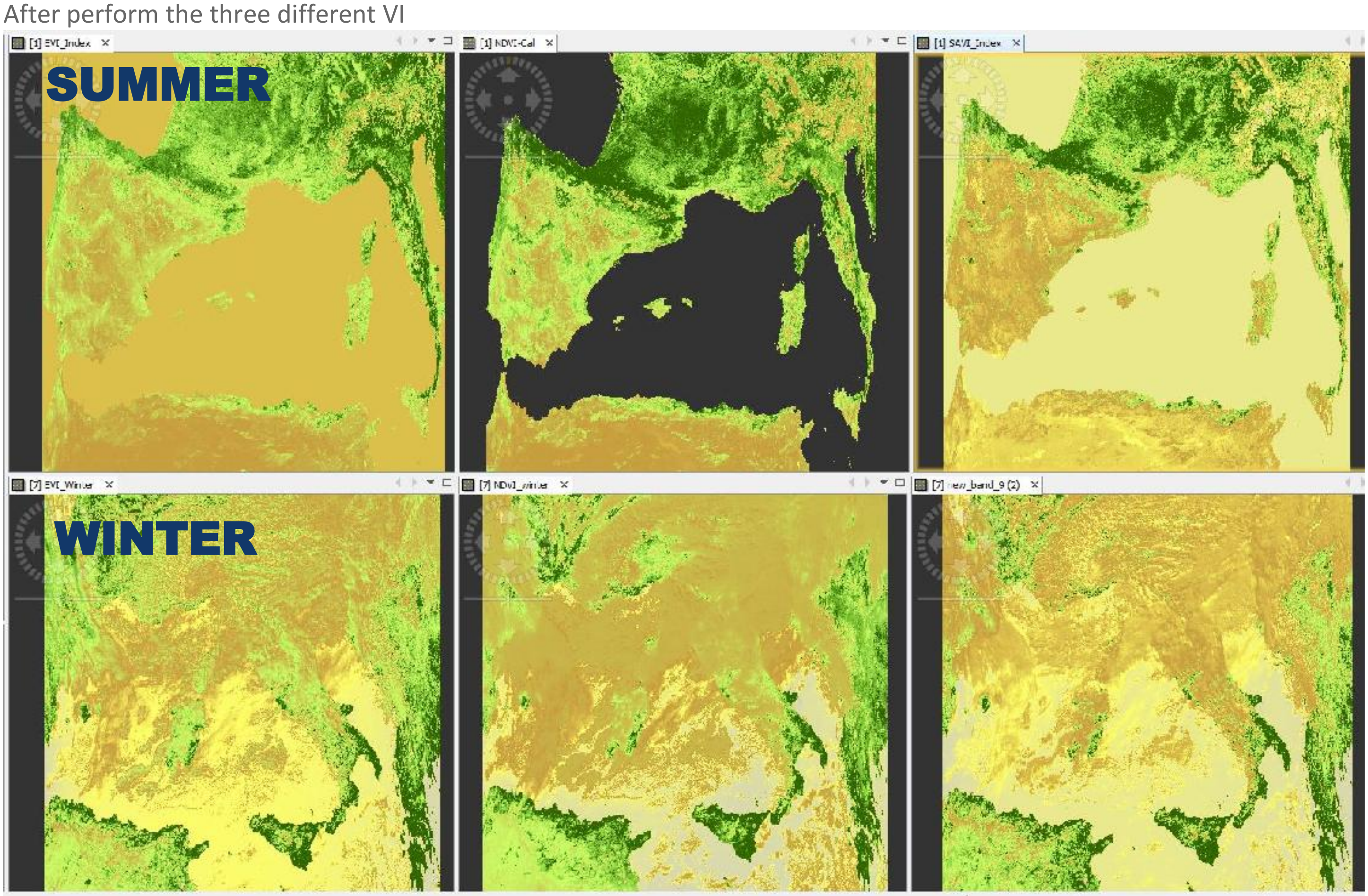


Figure 13.3 – Winter / Summer NDVI comparison

Conclusions:

- In summer cloudiness is much less, so we can appreciate the variations of the biomass layer over the ground
- Winter cloudiness influences the result of the vegetation index
- Despite the cloudiness, it can be seen that in winter the land presents an increase in biomass, which translates into healthier vegetation

14. To perform a quantitative change detection (difference) of the vegetation coverage class by reprojecting the 2 winter and summer MODIS images over the same grid in a selected region of interest (ROI).

We perform collocation using the Nearest Neighbor Resampling method, Using Summer Image as a Master and Winter Image as Slave.

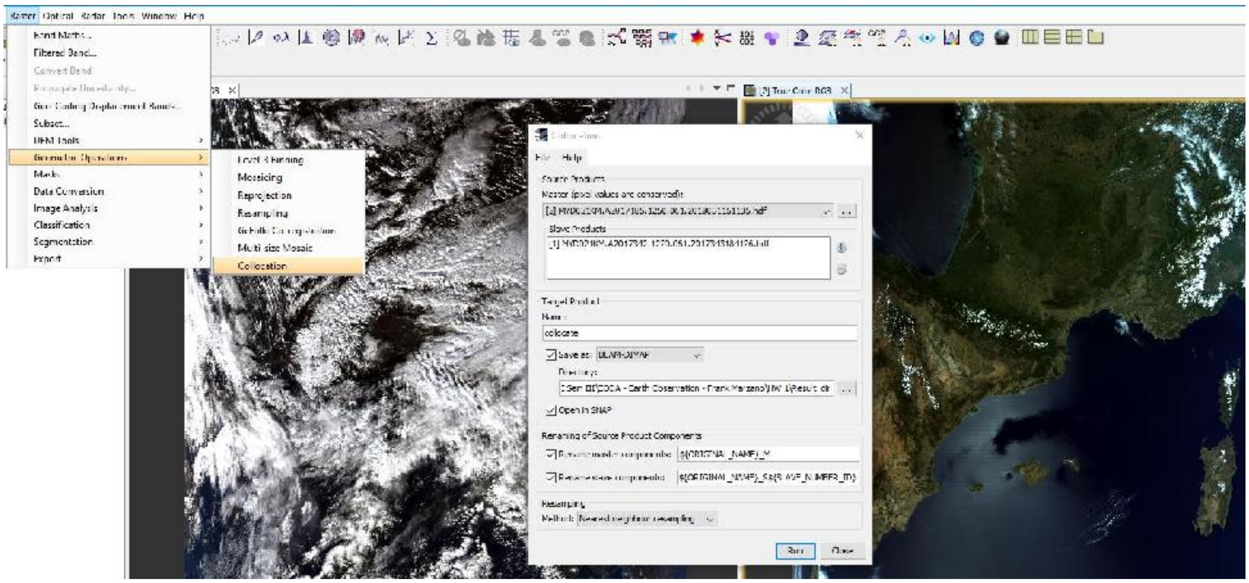


Figure 14.1 – Performing Nearest Neighborr Resampling

Result:

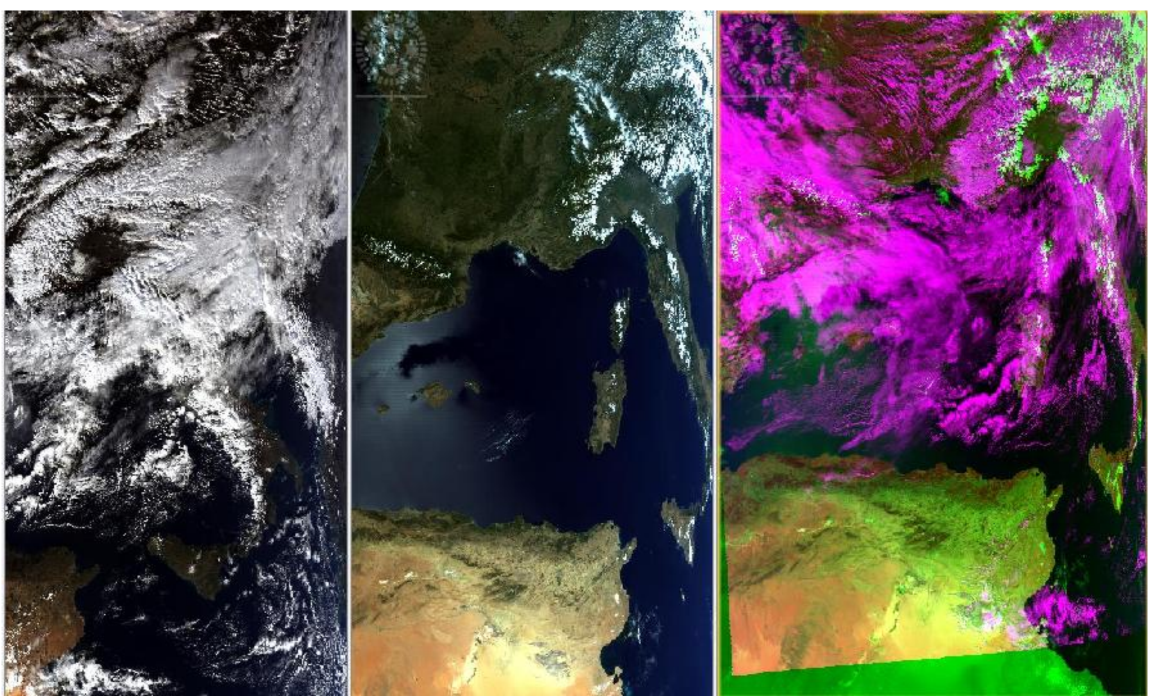


Figure 14.2 – Collocation Result

When two products are placed on top of each other, a new product containing the information of all the components of the originals is created.