Earth Observation Data Analysis – Homework 01

Andrea Pisani

1) Install SNAP

To download SNAP go to the link http://step.esa.int/main/download/

2) Download the MODIS/AQUA image

The image used in the Homework 01 is the one provided as “image example”.

3) Data quality check

For the emissive Channels the “good” images and the ruined ones are classified:

- Good images: 20, 22, 23, 25, 31, 32
- Ruined images (failures/strips): 21, 24, 27, 28, 29, 30, 33, 34, 35, 36

The images below show the comparison within the bands 20, 27 and 36: the first has no failures, while the 27 and 36 have evident horizontal strips. Those images cannot be used in the analysis

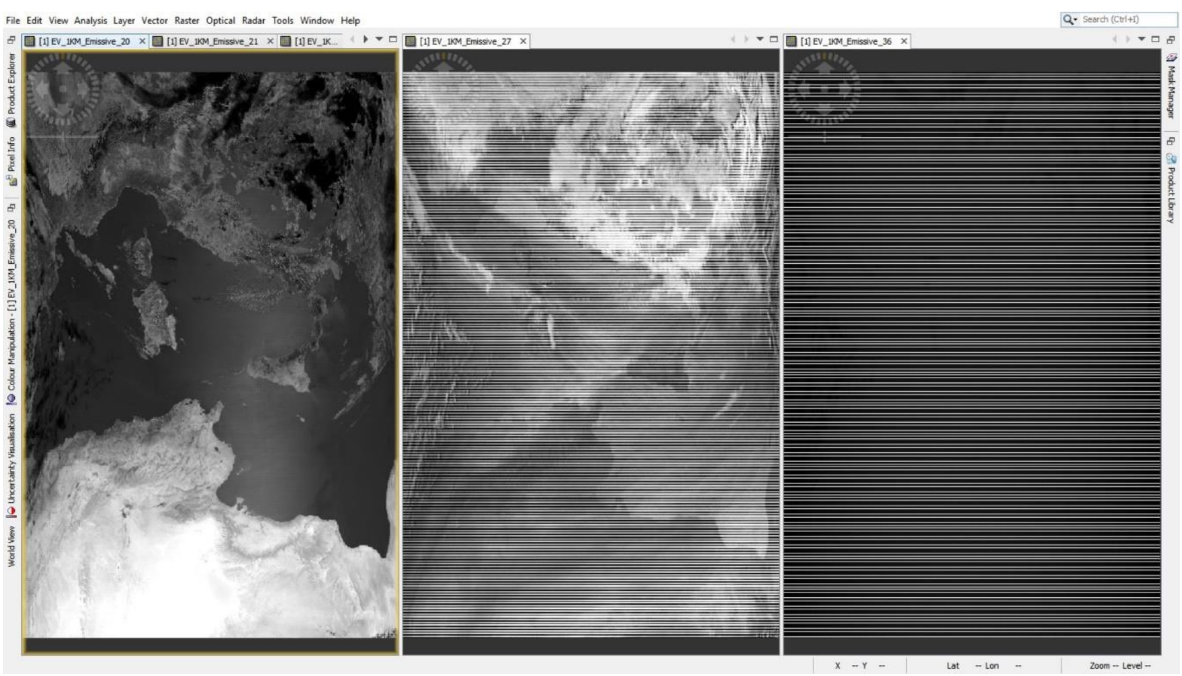

Figure 1 - band 20, band 27 and band 36

For the reflective Channels those are the “good” images:

- Good Images: 1, 2, 3, 4, 5, 6, 7

Other images in the reflective channels will not be considered.

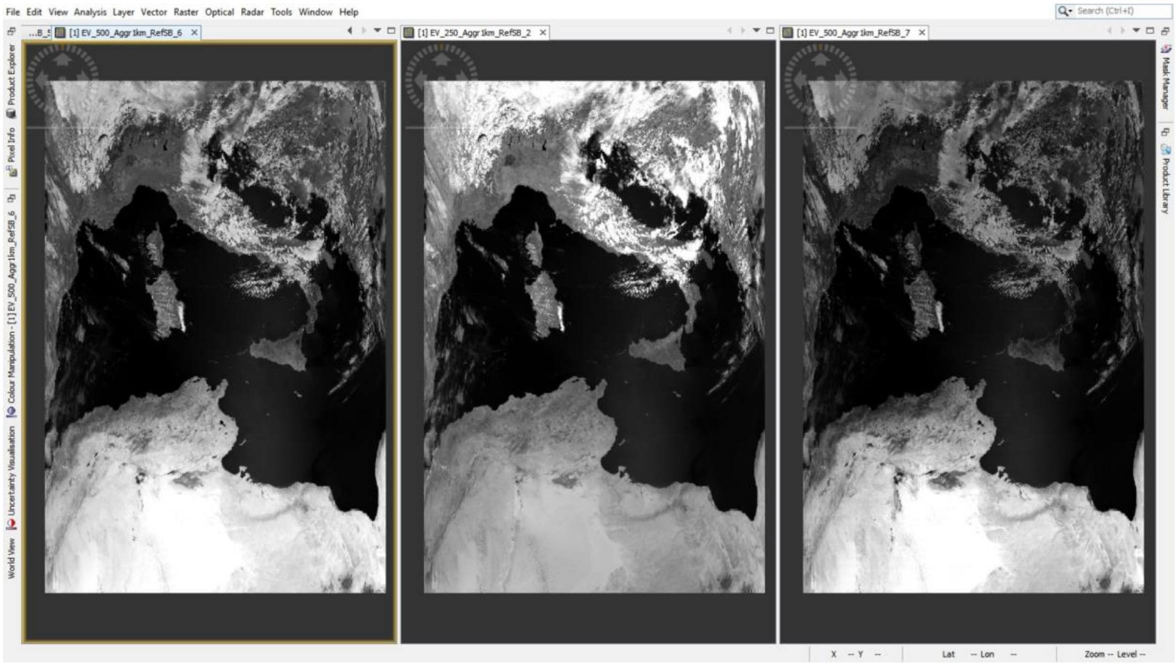

Figure 2 - band 6, band 2 and band 7

4) Data analysis by spectrum

To analyze the spectrum **5 pins** are placed in different places: vegetation (green), clouds over the terrain (white), clouds over the sea (light blue), sea (blue) and desert (red) as shown in the next image.

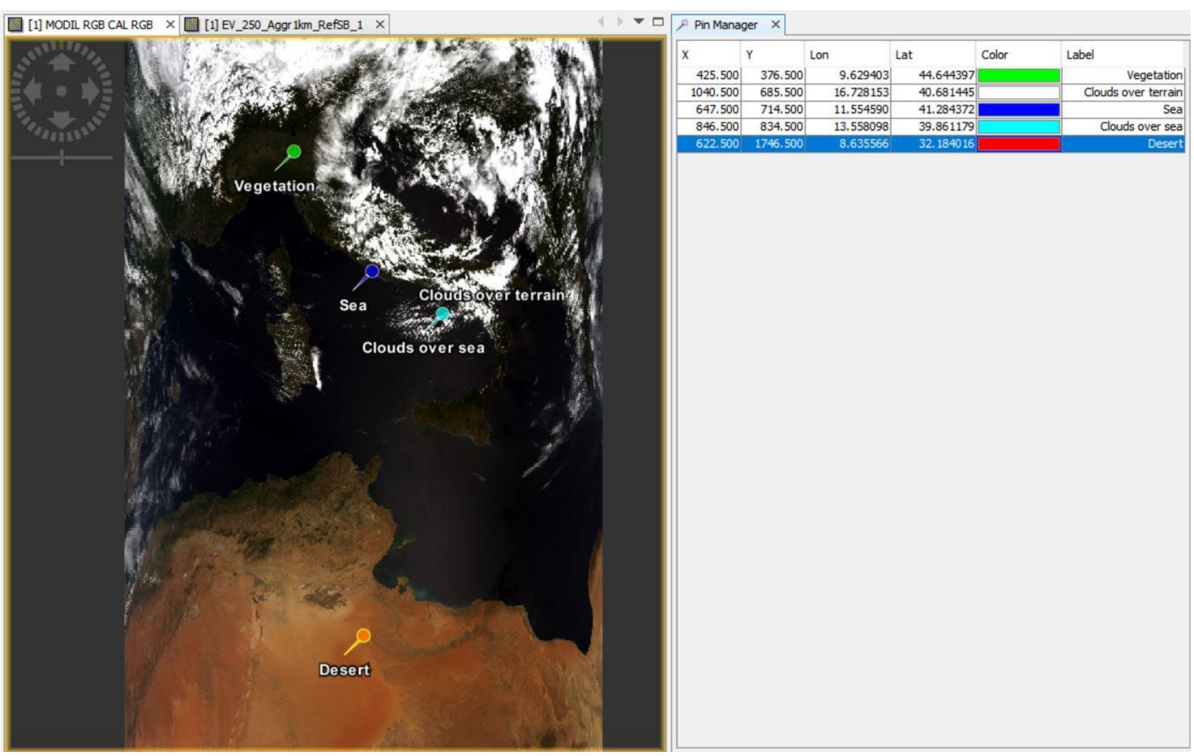

Figure 3 - 5 pins placed on the image

In the next images there's the plot of the **spectrum view** in these five pins.

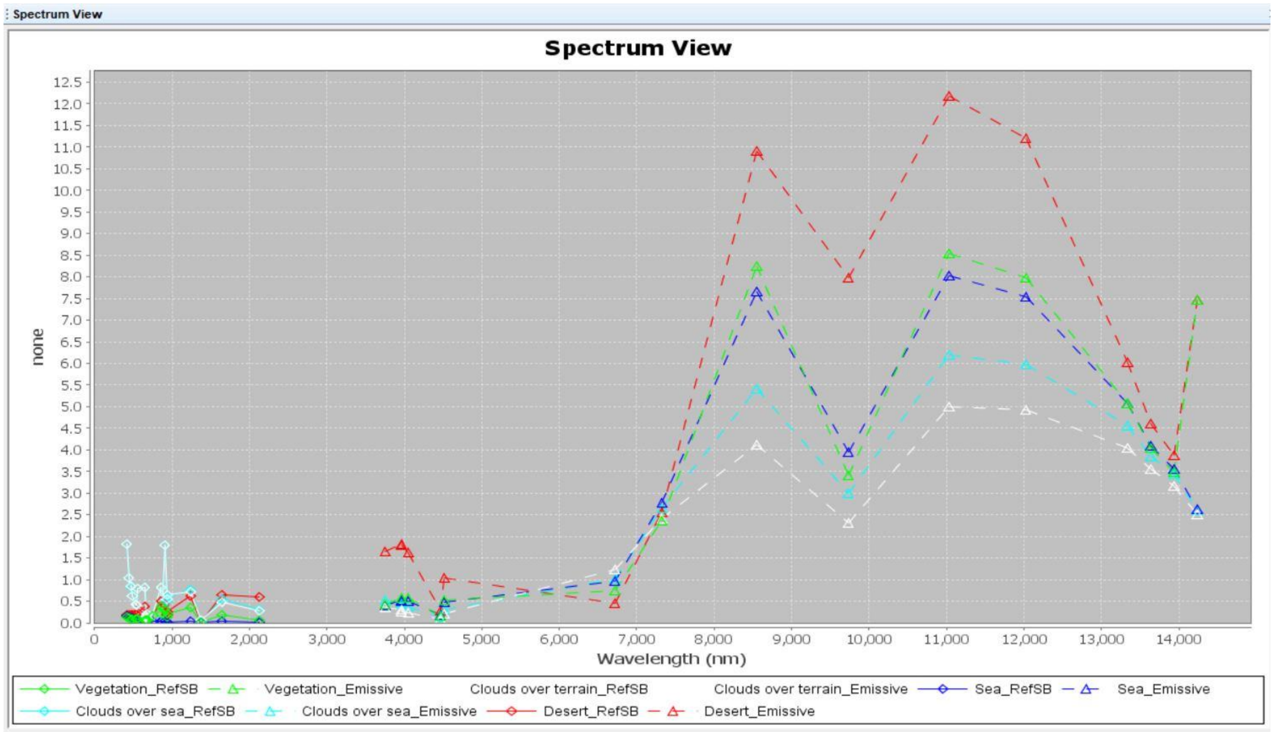

Figure 4 - spectrum view using the pins placed earlier

In the left part of the previous image the reflective part of the spectrum can be seen, while the emissive part is in the middle-right side from **4000 nm to 14000 nm**. In the reflective part of the spectrum we can easily see that the clouds (white in the image) are the most reflective portion of the image and the less emissive ones. In the emissive part of the spectrum the plot is dominated by the desert (red). The interpretation can be trivial knowing that it is hotter than the sea and the clouds.

In the next 2 images the histograms of band 1 (reflective) and band 31 (emissive) are shown. On the x-axis there's the intensity of the pixels, while in the y-axis there's the frequency of pixels at a certain intensity. The histogram of the reflective band 1 has darker pixels with respect to the one of the emissive band 31.

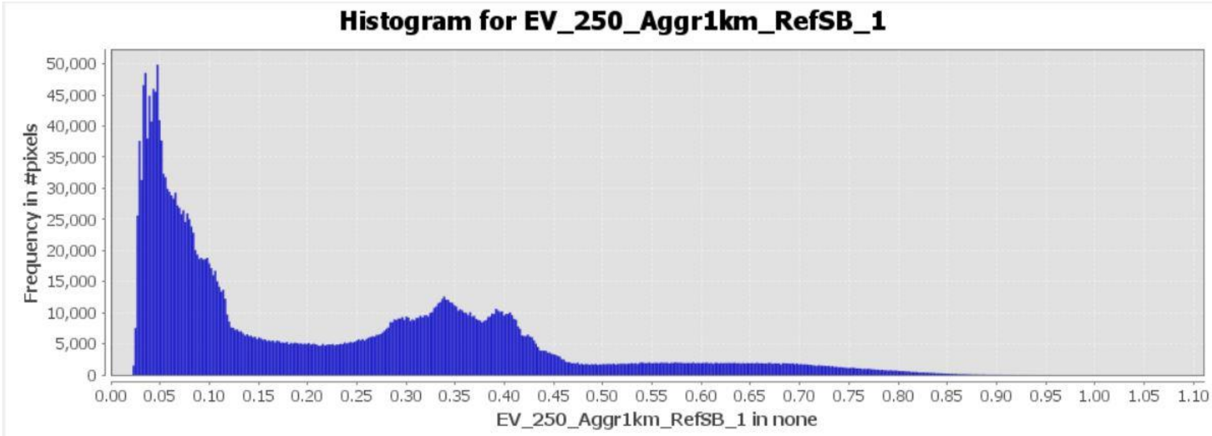

Figure 5 - histogram of the reflective band 1

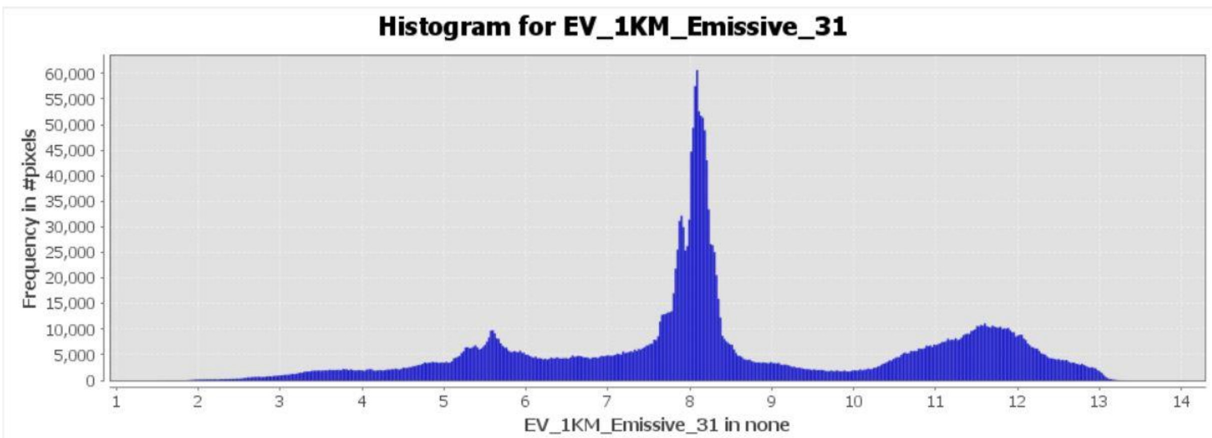

Figure 6 - histogram of the emissive band 31

The next 2 images are associated with the reflective band 1 and the emissive band 31: we can see by inspecting them that the first one is darker than the second one. This is the information shown also by the 2 histograms. We can also point out that the whiter color of the desert in the emissive band 31 is associated to an higher emission, meaning that the desert is hotter as we already stated by inspection of the spectrum view.

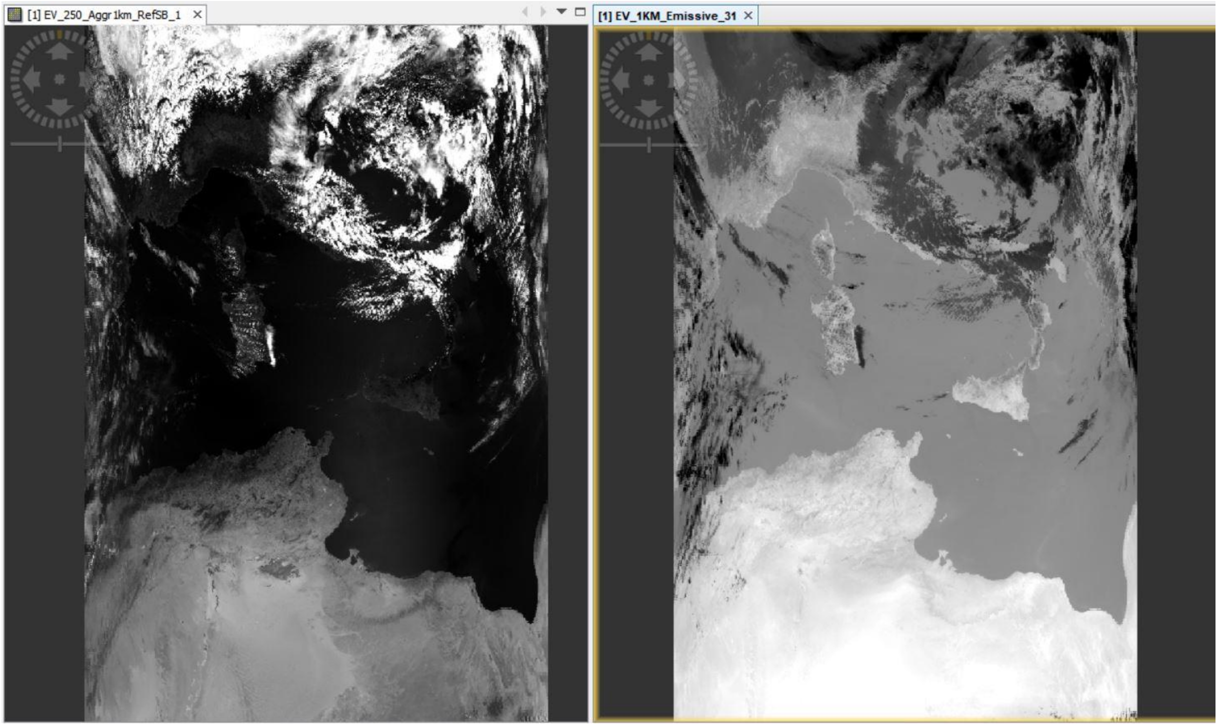

Figure 7 - inspection of the reflective band 1 and the emissive band 31

5) Channel data correlation of the whole image

The scatter plot showed in the next image is a correlation between the histogram of the reflective band 1 at 0.645 µm on the x-axis and the histogram of the emissive band 31 at 11 µm on the y-axis.

The pixels corresponding to a yellower color are the ones where the two histograms both have a high frequency in terms of number of pixels. For example, the two graph have approximately the same amount of pixels at intensity 0.05 for the 0.645 µm band and intensity 8 for the 11 µm band.

On the other hand the darker pixels correspond to the areas where the histograms both have the same low frequency in terms of number of pixels.

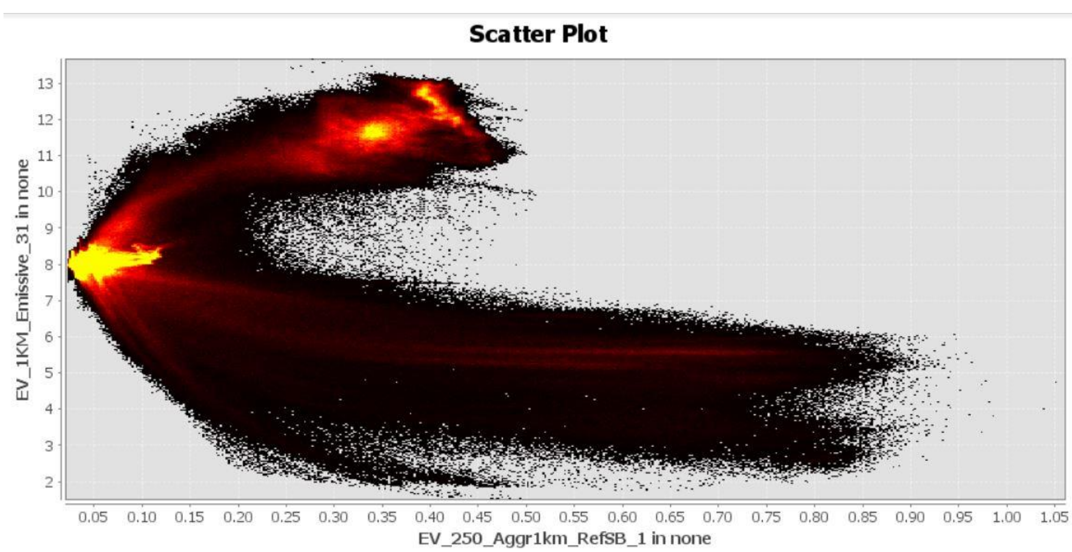

Figure 8 - scatter plot between the reflective band 1 and the emissive band 31

Using the scatter plot the information provided in the SNAP short guide can be interpreted:

- Land: low reflective at 0.6 µm and highly emitting at 11 µm.
- Sea: low reflective at 0.6 µm and medium-to-high emitting at 11 µm.
- Clouds: highly reflective at 0.6 µm and low-to-medium emitting at 11 µm.

Following the descriptions above the land is the in the middle-left/top part of the image, the sea is in the right/middle part and the clouds are in the middle/low part as the next image shows. The land and the sea are yellower because there are more pixel associated to those parts of the image. This seems reasonable if we inspect the original pictures: the land and the sea are larger than the clouds.

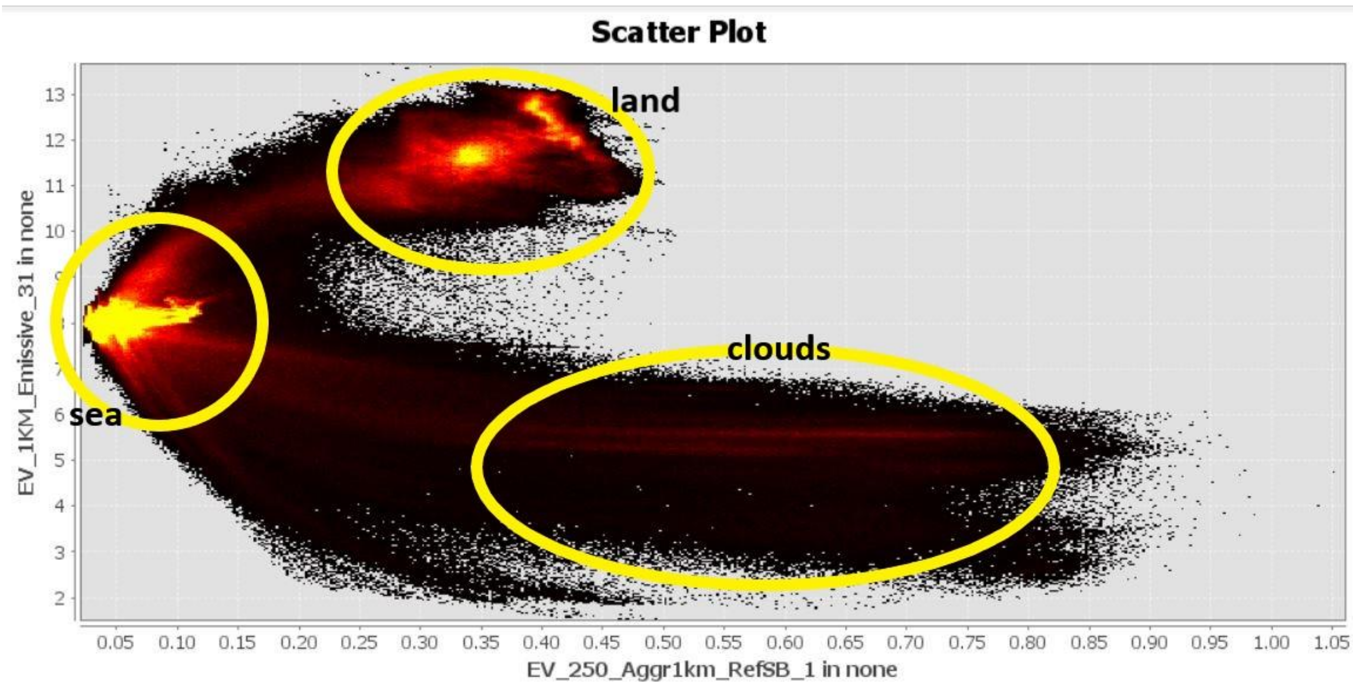

Figure 9 - scatter plot with interpretation

6) Channel data correlation of the selected ROI

Now a ROI is selected in the reflective band 1 channel to obtain a profile plot of the terrain. A line is traced starting in the clouds, continuing in the terrain and ending again in the clouds. As it can be seen in the profile plot the first and the last part of the path are more reflective (clouds) and the middle part has low reflectivity (terrain).

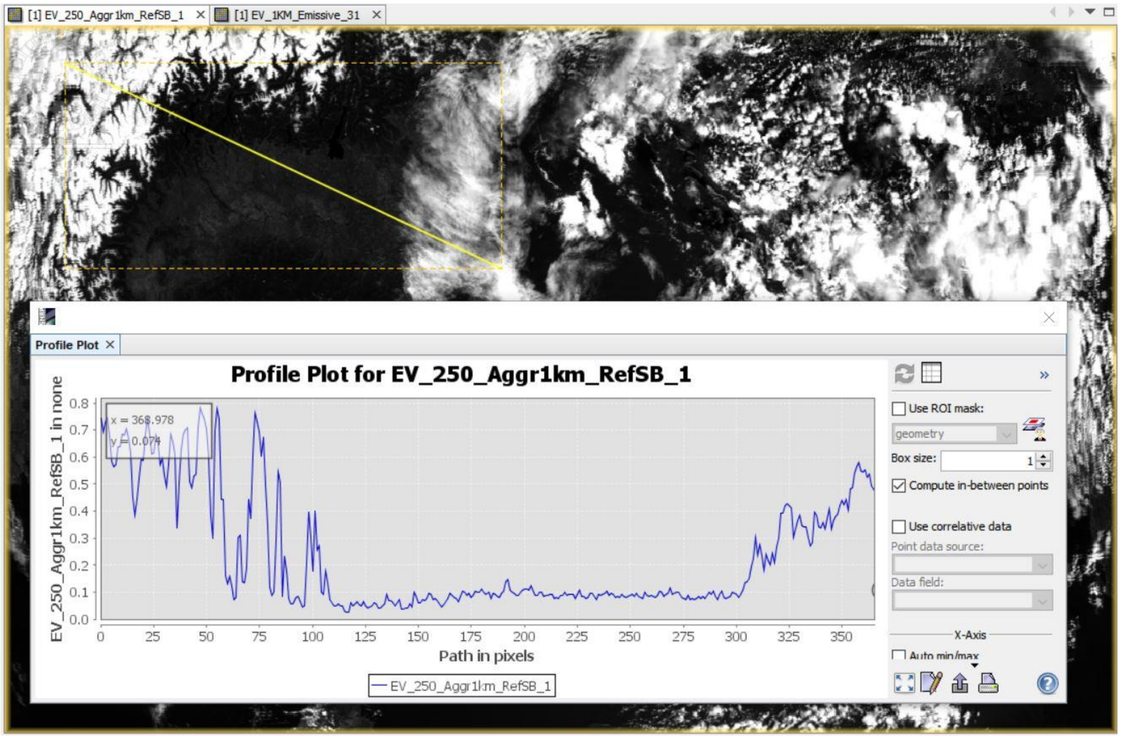

Figure 10 - profile plot of a line

Another ROI is selected in the reflective band 1 channel to plot the histogram in this region. This area, as shown in the histogram below, has a high number of darker pixels. As a matter of fact the square chosen in the reflective band 1 image has more terrain than clouds.

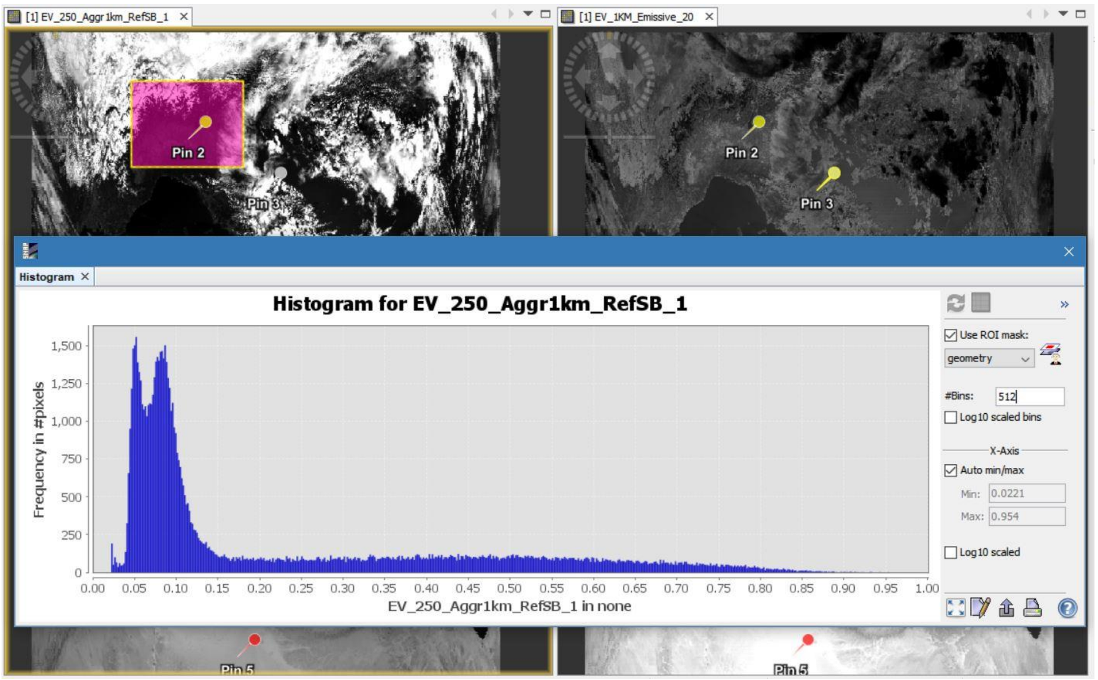

Figure 11 - histogram of a ROI

7) Principal Component Analysis

To perform the principal component analysis 7 images only the “good” images listed in the 3rd part of the homework are used.

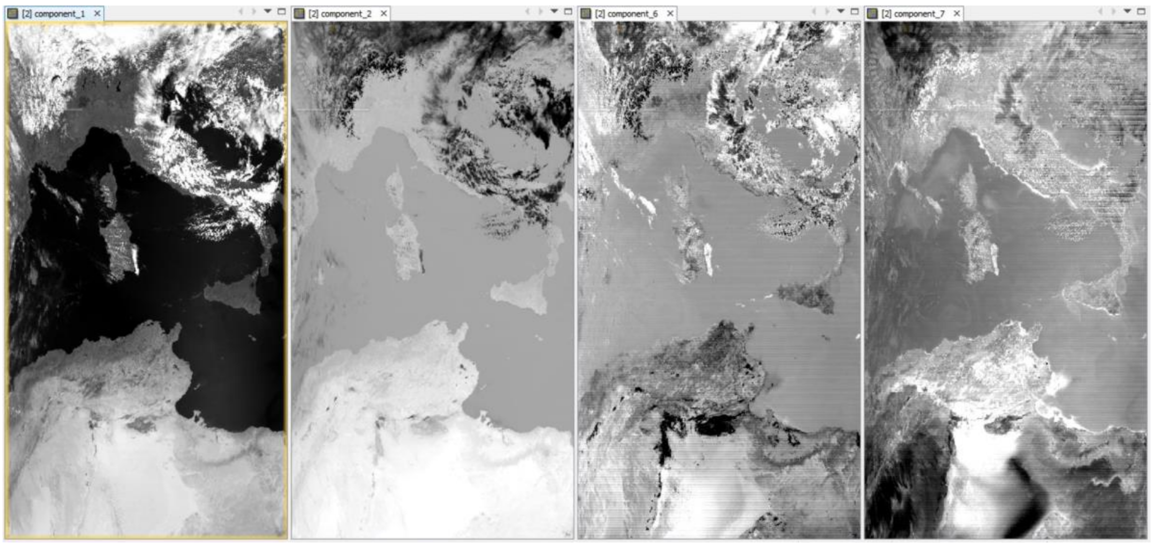

Figure 12 - PCA analysis

The first 2 component shown above are the ones associated to a low variation of the images and carry most of the information, while the 6th and the 7th component display a higher contrast and are associated to an high variation of the image.

8) Unsupervised classification

For the unsupervised classification the k-means method is used. The best results are obtained using the components of the PCA of the previous point of the homework.

All the 7 components of the PCA analysis are provided to the k-means algorithm to find the clusters in the image. The method needs the parameter k to be chosen, meaning that the number of clusters to find must be provided to obtain result.

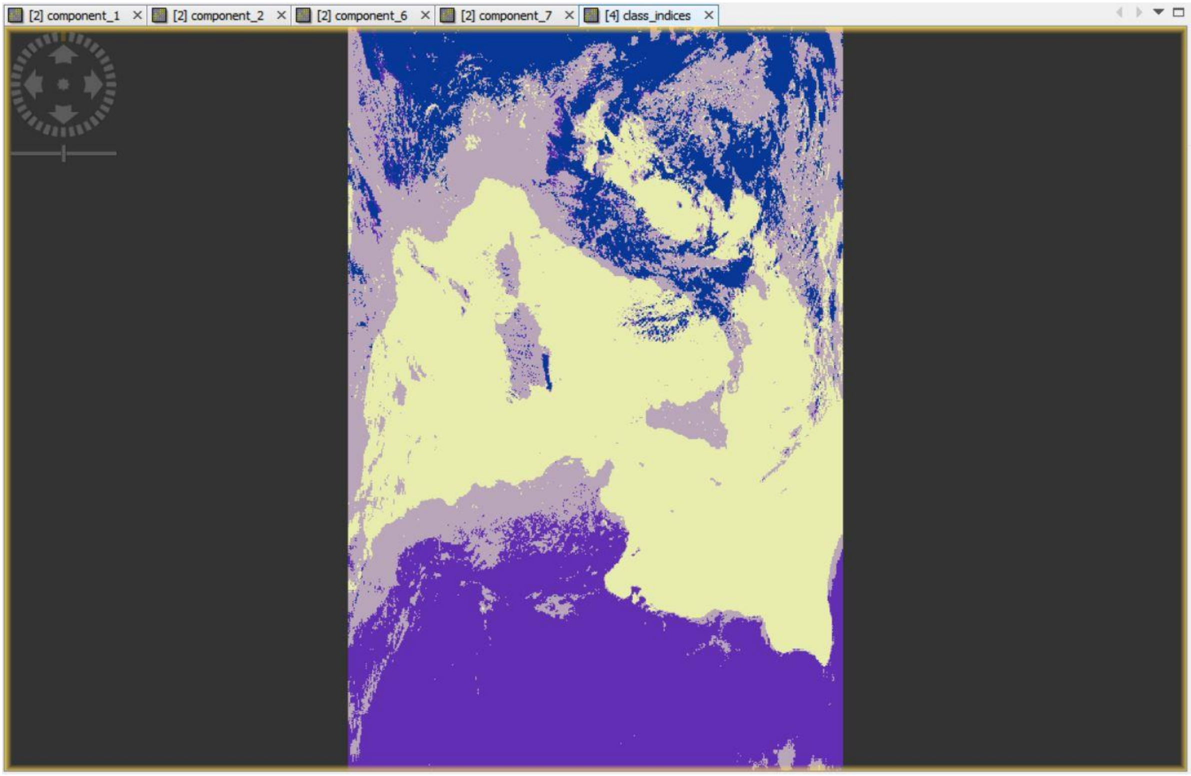

Figure 13 - k-means results

For the k-means unsupervised clustering algorithm the number of clusters (k) must be chosen before the computation starts. The best results are obtained with 4 clusters as we can see in the image above: the 4 classes identified by the algorithm are the desert, the clouds, the sea and the terrain with vegetation.

Even if the results are quite accurate there are some errors when the clouds have a small area, as we can see in the left and right part of the image. As a matter of fact the small and distorted clouds are interpreted as vegetation. In the next image this error is highlighted comparing the RGB image and the result of the k-means clustering using 6 different classes. As we can see below choosing 6 clusters doesn't improve the overall quality of the clustering.

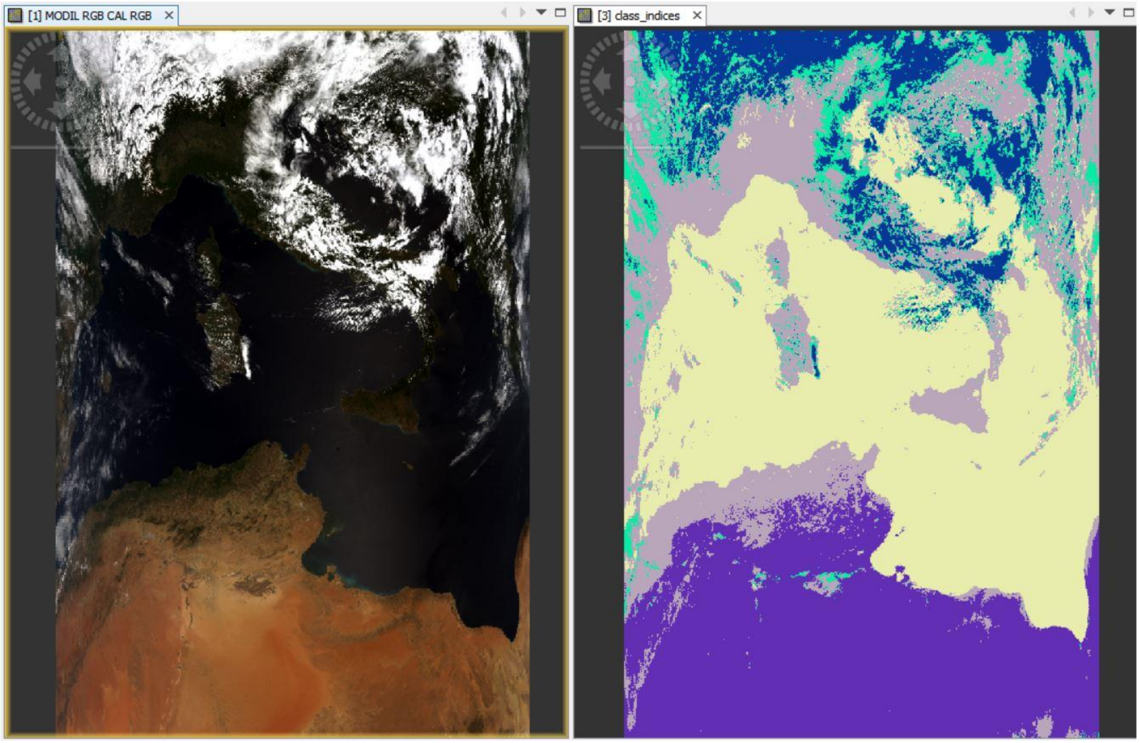

Figure 14 - RGB image vs k-means with 5 classes

9) Supervised classification

To perform the supervised classification and identify the 4 main classes (clouds, sea, vegetation/terrain and desert) the RGB image is used to construct the vector data. As shown in the image below there are different areas of the picture selected to train our model and identify the 4 classes listed before.

For the sea 2 portion are selected in yellow, for the desert a light blue area is highlighted, for the vegetation/terrain an orange area is selected and finally 2 pink areas are displayed in pink. Those vector data are used in the supervised classification as train vectors.

All methods have been used but the best results are obtained with the maximum likelihood classifier: this algorithm maximize the likelihood function given a certain parameter representing the hidden information. In our case the hidden information are the vector data.

As we can see in the image the results are quite accurate, even if there are some (large) errors in the identification of the sea on the left part of the image. This is normal because in this portion of the map the clouds are all over the sea, and the method fails to recognize the color of the water.

On the left and right side of the picture the vegetation/terrain (blue) is overestimated in the area where the sea become darker.

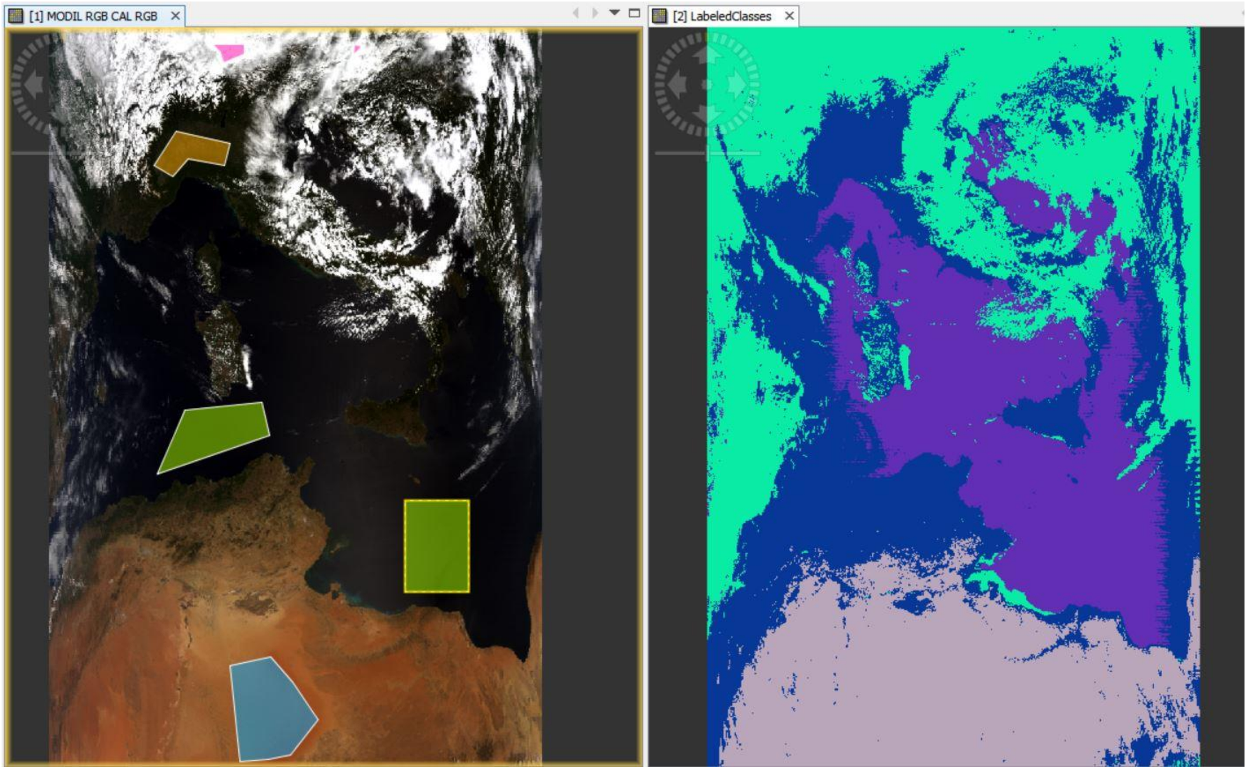

Figure 15 - results of the maximum likelihood classification

10) Vegetation indexes

The idea behind the vegetation index is to highlight the vegetation in a map using the different bands. In particular we will use the spectral reflectance acquired in the near infrared band (NIR) and the reflectance acquired in the red band (RED): the vegetation is characterized by an high reflectance in the NIR band with respect to the low reflectance in the RED band. If we subtract the RED to the NIR we find the areas where the vegetation is present, removing all the information about other elements (cloud, sea, …).

The first index used to highlight the vegetation is a 2-band vegetation index called the **normalized difference vegetation index (NDVI):**

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

This index is already implemented in SNAP as “NDVI processor”. The result of this procedure will be displayed in the next image with other results obtained using different VI.

The **difference vegetation index (DVI)** is a 2-band vegetation index that works in a similar way, but it has no normalization factor. As it will be shown in the pictures below, the results are worst and even the desert terrain is recognized as vegetation in the image.

$$DVI = NIR - RED$$

The last index implemented is another 2-band VI called **transformed vegetation index (TVI)** which is:

$$TVI = \sqrt{|NDVI + 0.5|}$$

This one is not implemented in SNAP, but it is possible to write a customized formula via band math to compute it. This index gives improved results (for example in the desert area, where the colors are less green with respect to the NDVI) but still small. We'll use the TVI to compare the results of the 2-band VI vs the 3-band VI.

In all the images the color ramp used is the "meris_veg_index" which substitute the darker colors with yellower colors and whiter colors whit greener colors. The sliders in the color manipulation window have been all moved to the right to increase the contrast and highlight the vegetation.

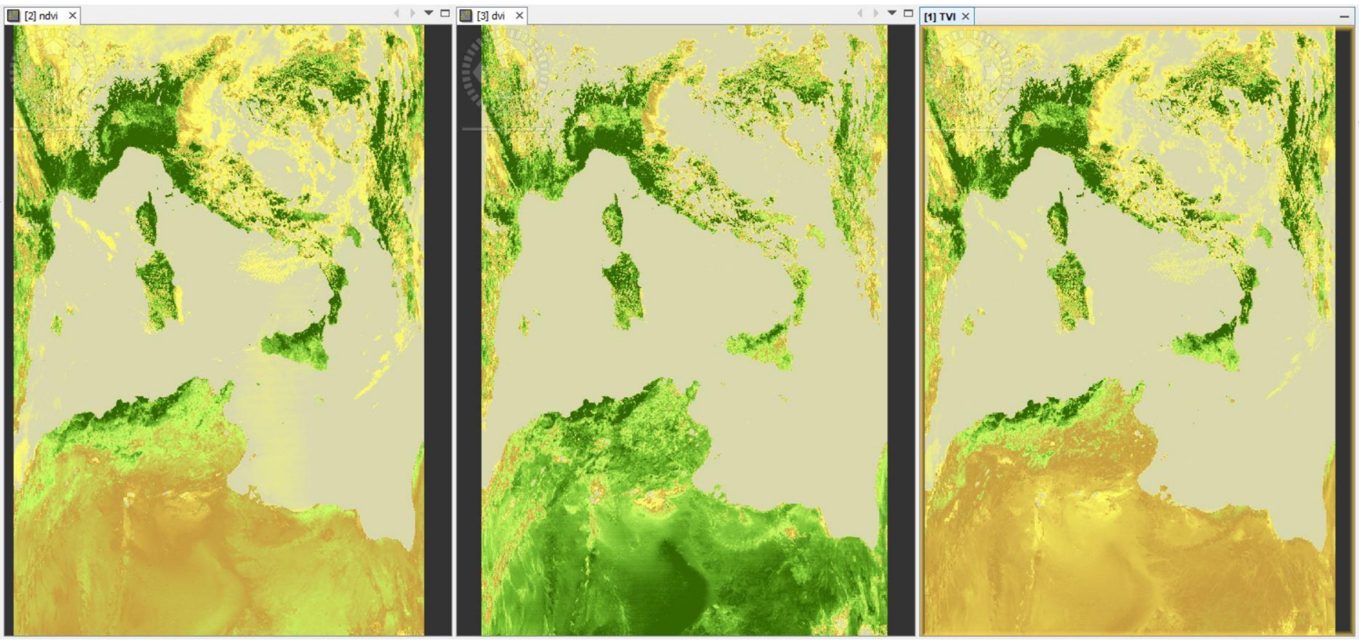

Figure 16 - NDVI vs DVI vs TVI

A 3-band is used in the next images, the **triangular greenness index (TGI)** useful to estimate leaf chlorophyll:

$$TGI = GREEN - 0.39 * RED - 0.61 + BLUE$$

The index is used in Nebraska and it's easy to see that there is an overestimation of vegetation in the desert area.

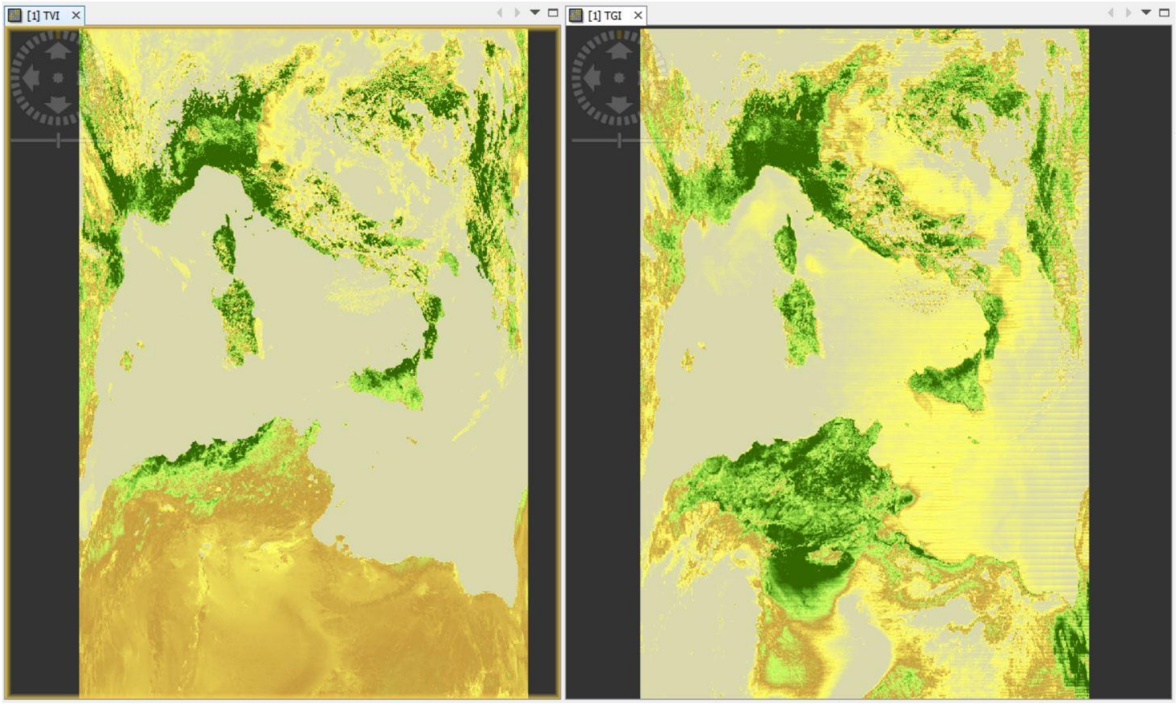

Figure 17 - TVI vs TGI

11) Vegetation index on land pixels

Using the NDVI we inspect the land pixels to determine their value in different regions. If we place the cursor over the clouds the values of the pixels will be negative as shown in the next picture.

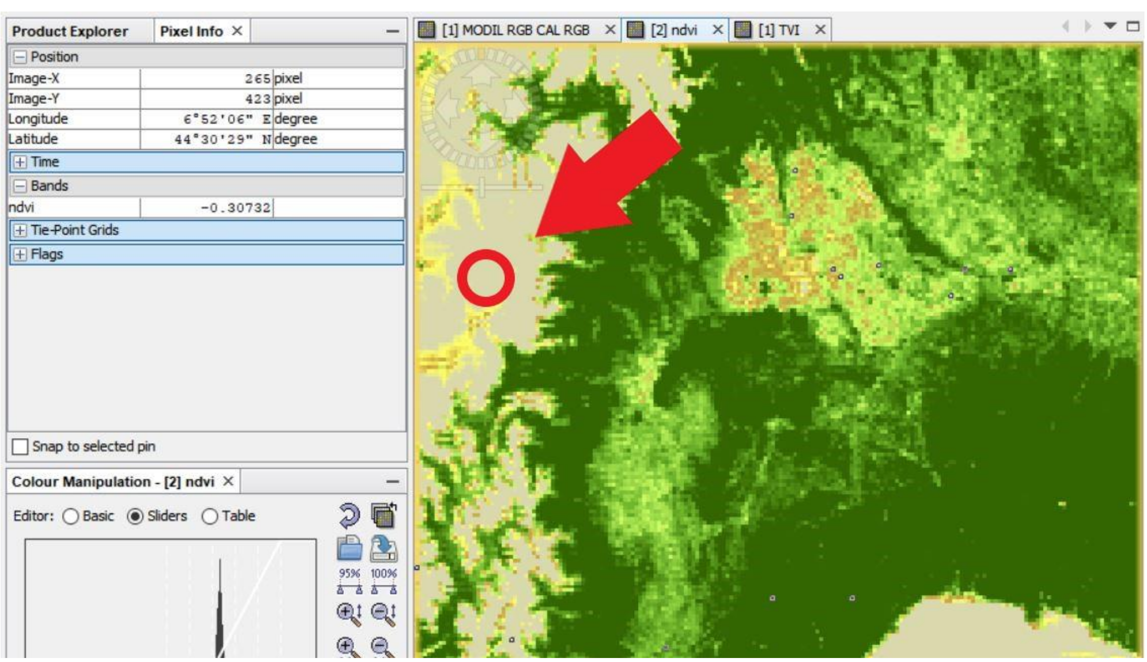

Figure 18 - pixel value of the clouds

If we move the cursor on the greener area the values of the pixels will be positive and closer to one.

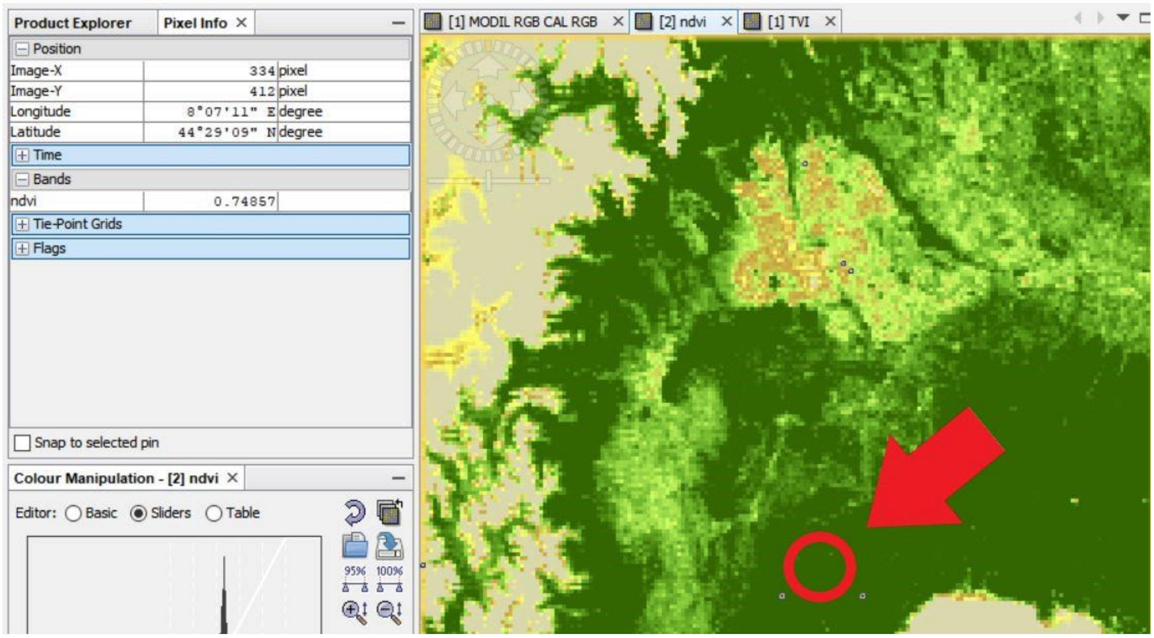

Figure 19 - pixel value of the vegetation

We can use a formula on band math to retrieve only the pixels greater than a certain value to remove everything that is too different from the vegetation.

As shown in the next images only the pixels with value greater than 0.4 are displayed, while the others now have a value equal to zero.

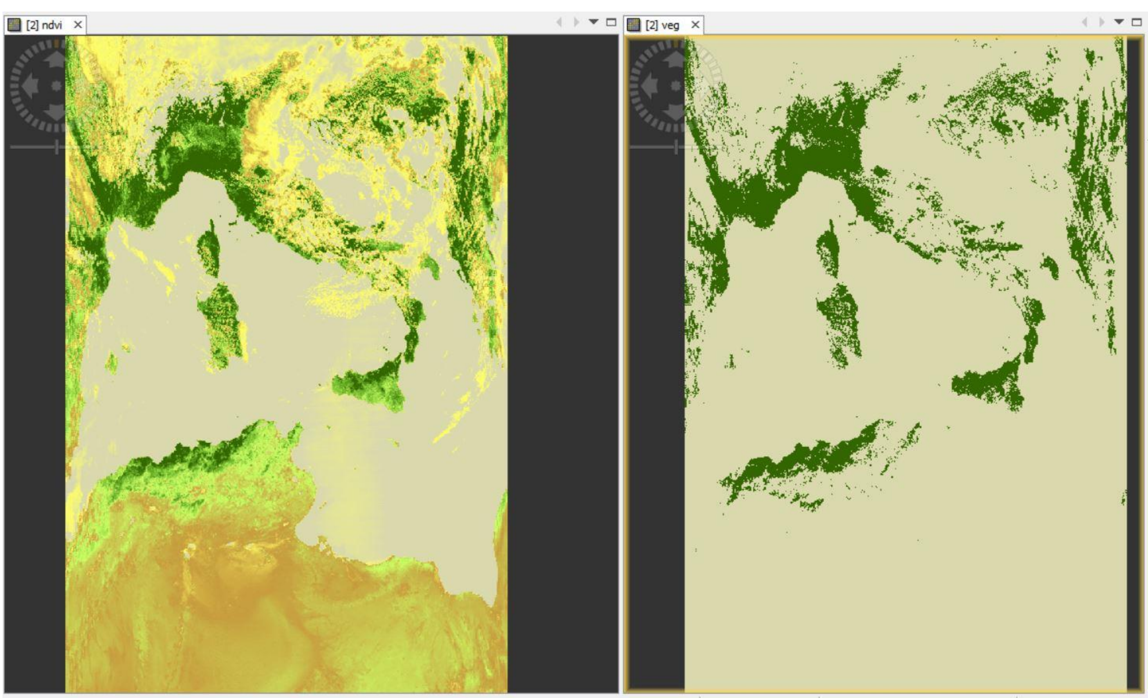

Figure 20 - NDVI vs pixels with value >= 0.4