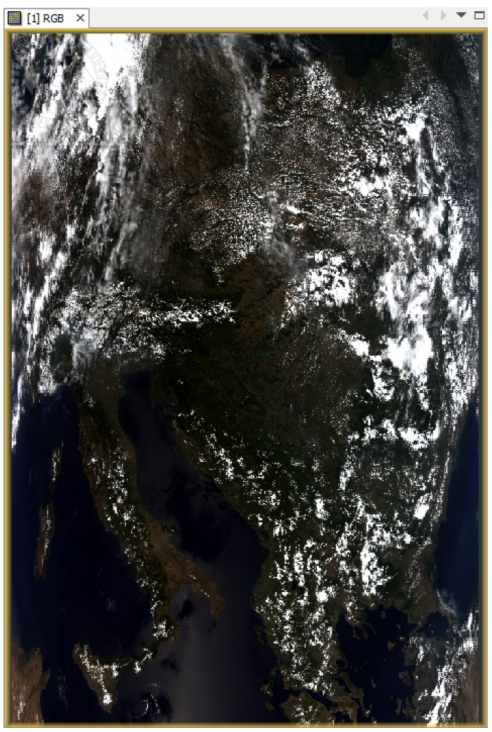
Earth Observation Data Analysis - Homework 01

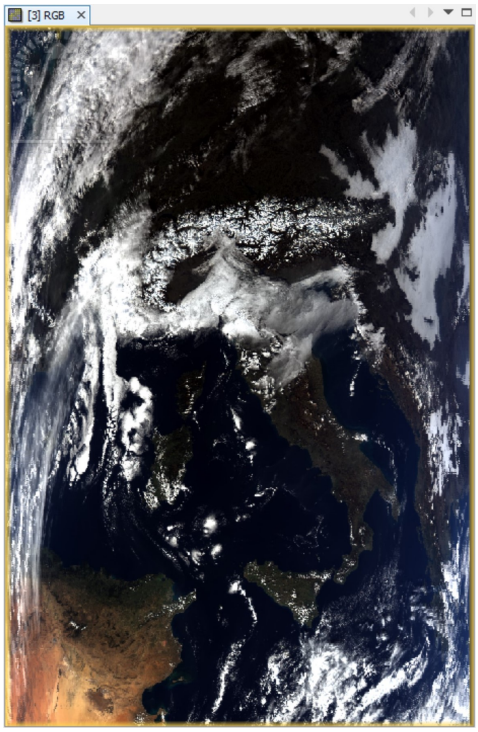
Gioia Mancini, Christian Brignone

August 31, 2020

The topic of this homework is related to the analysis of MODIS (Moderate-resolution Imaging Spectroradiometer) images coming from the Aqua satellite. In particular two images at 1Km resolution over Italy, taken in different seasons, have been selected in order to be compared (figure 1).



(a) *MYD021KM.A2019201.1150.061.2019202151857*



(b) *MYD021KM.A2020015.1220.061.2020016154721*

Figure 1: RGB images of Italy taken in July 2019 (a) and January 2020 (b).

Data analysis by spectrum, histogram, profile tools and data correlation. Starting from the first image (July 2019), before going deeper into the analysis, a quality check of the product is needed. Looking at the channels, besides reflective channel number 6 and emissive channels number 24 and 36 that seem to contain failures and/or stripes, all is ok.
Now it is possible to perform data analysis by spectrum, histogram and profile tools. In order to do that, we can first take into account two different channels: EV_250_Aggr1km_RefSB_1 and EV_1KM_Emissive_31 (figure 2).

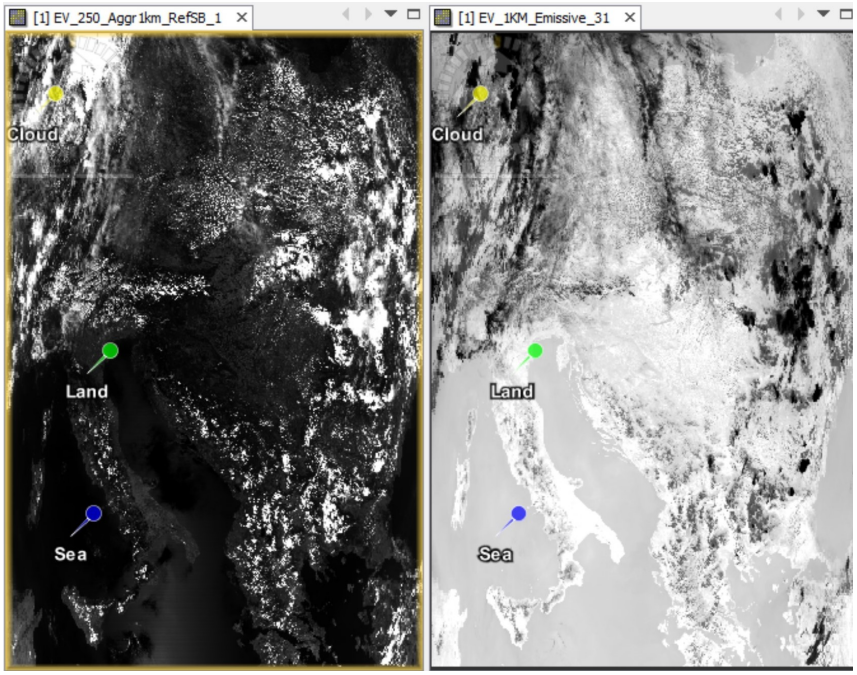


Figure 2: EV_250_Aggr1km_RefSB_1 channel, on the left, and EV_1KM_Emissive_31 channel, on the right: the pins are placed in correspondence of land, sea and clouds.

First, we can place three different pins in place of interesting features, such as land, sea and clouds; then, we can inspect the spectrum view of the image by considering the pins and their positions. By looking at figure 3, it's possible to see that clouds reflect higher in the visible band (VIS) while land and especially sea reflect in the near infrared band (NIR); for what concern emissivity, land and sea have a similar behaviour in the middle infrared (MIR) and in the far infrared (FIR), where they assume the highest values; also clouds emissivity assume values growing from the middle infrared till the first part of the far infrared (but with much lower values than land and sea) up to $14\mu m$ wavelength, where all of them assume similar values.

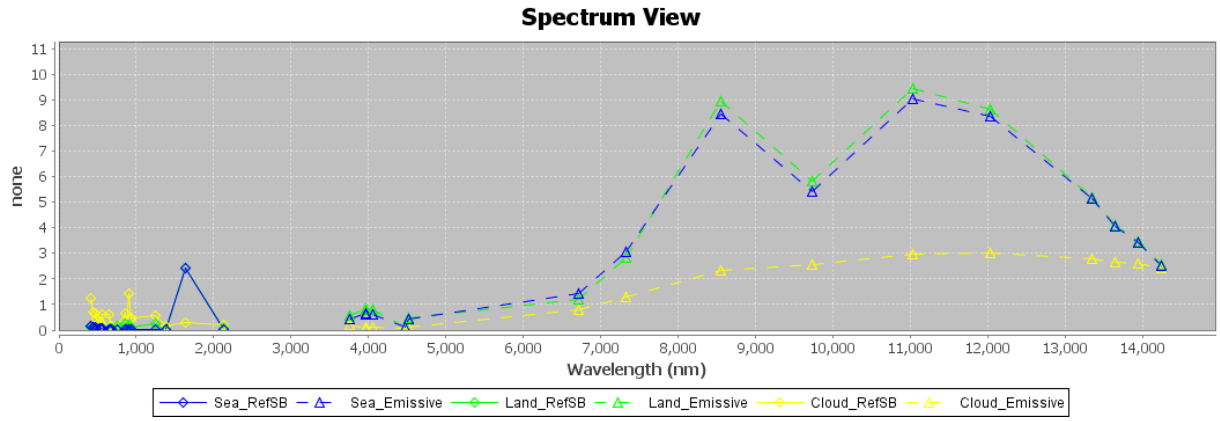
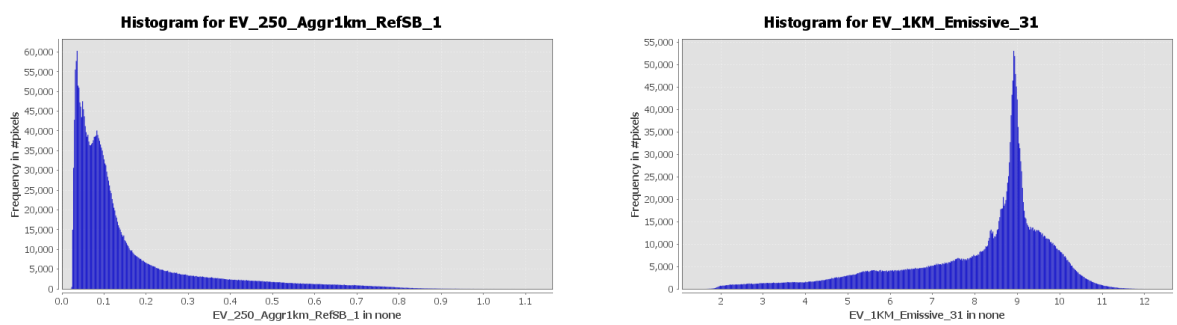


Figure 3: Spectrum.

Then, we can see the histograms of reflectance and emissivity bands (respectively in figure 4 (a) and (b)). On the y axis of the histogram plots, there are the frequencies expressed in terms of number of pixels, so the histogram tells us how many pixels of the image fall into each bin. It is possible to see that we have a higher concentration for reflectance around very small values, while for emissivity the highest value is reached around 9.



(a) (b)

Figure 4: Histogram: reflectance (a) and emissivity (b).

In addition, it is possible to perform channel data correlation by using the scatter plot, which in general allows to plot one raster data node against another. In particular, for each band, the value range is divided into 512 bins and this provides $512 * 512$ cells. For each cell we can see the amount of pixels within the according bands' value ranges, so we can compare the results given by the scatter plot with respect to the ones we had in doing the histograms. In fact, as we expected (figure 5), the highest amount of pixels (which appears in yellow color) is around the coordinate $(0, 9)$.

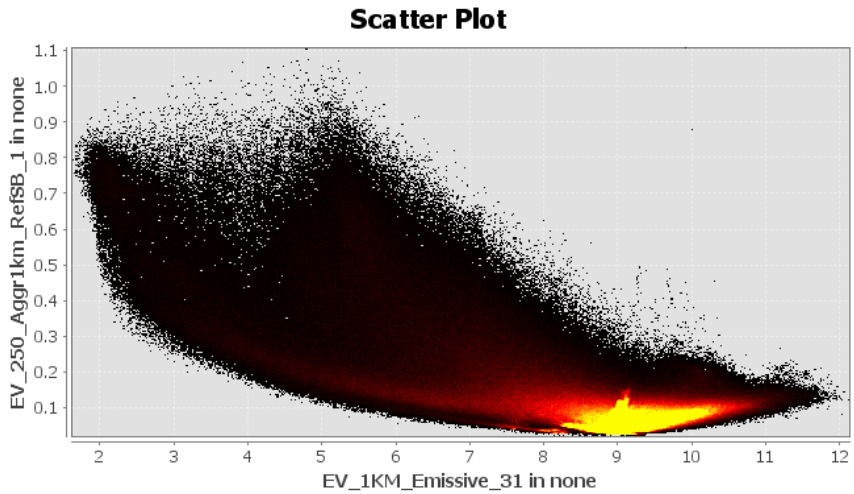


Figure 5: Scatter Plot data Correlation of the whole image.

Consider now a particular region of interest (ROI), obtained by selecting a rectangular geometry shape over central Italy (figure 6). Computing the profile plots for reflectance and emissivity of the ROI, the figures 7 and 8 are obtained.

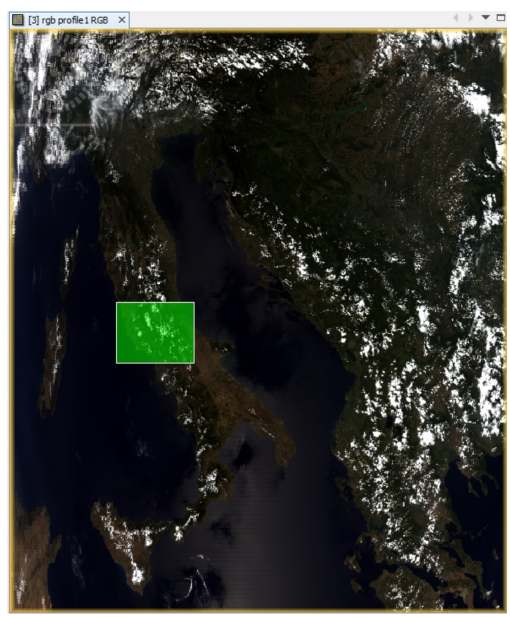


Figure 6: RGB ROI.

Finally, we have the scatter plot for the considered ROI (figure 9).

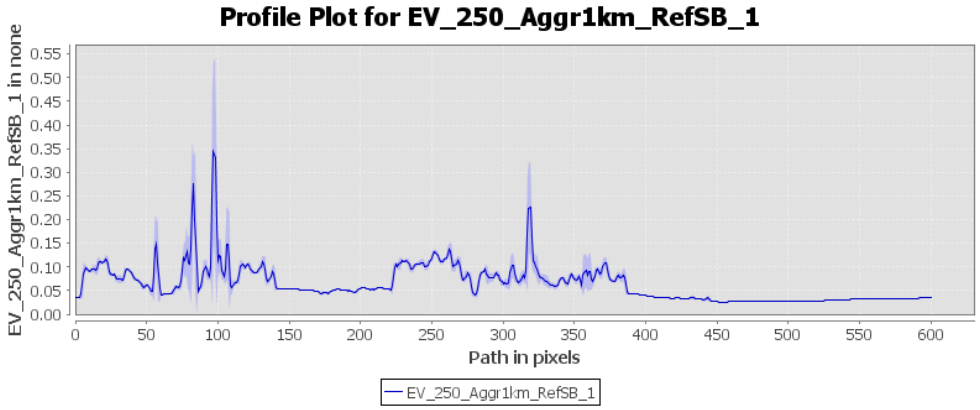


Figure 7: RGB ROI.

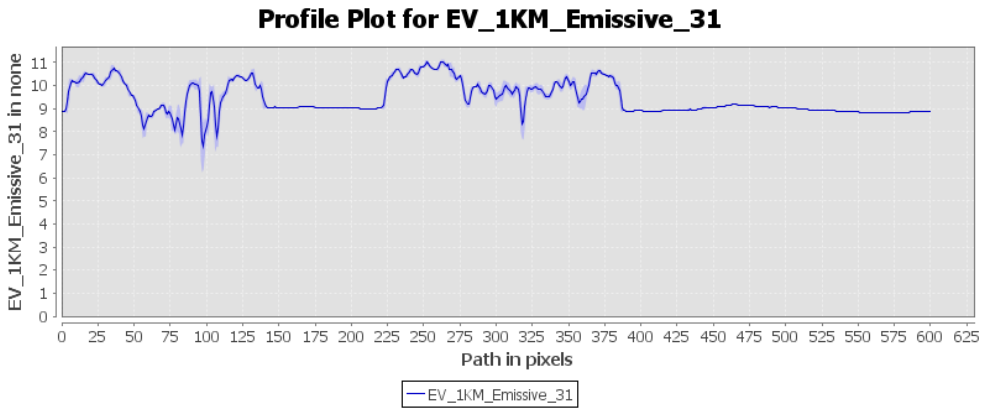


Figure 8: RGB ROI.

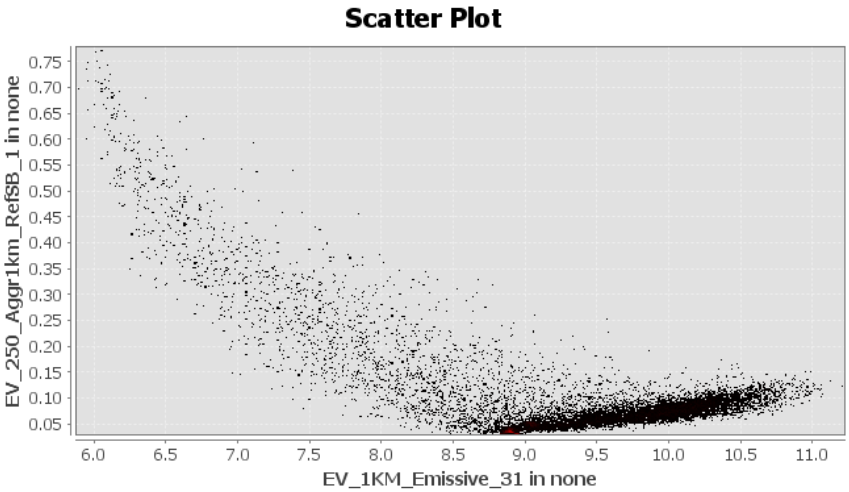


Figure 9: Scatter Plot data Correlation ROI.

Principal Component Analysis. A Principal Component Analysis (PCA) has been performed. In fact, additional features derived from the original ones can replace or complete the different spectral bands in order to support image interpretation. We considered the first seven reflective bands, the resulting components can be seen in figure 10.

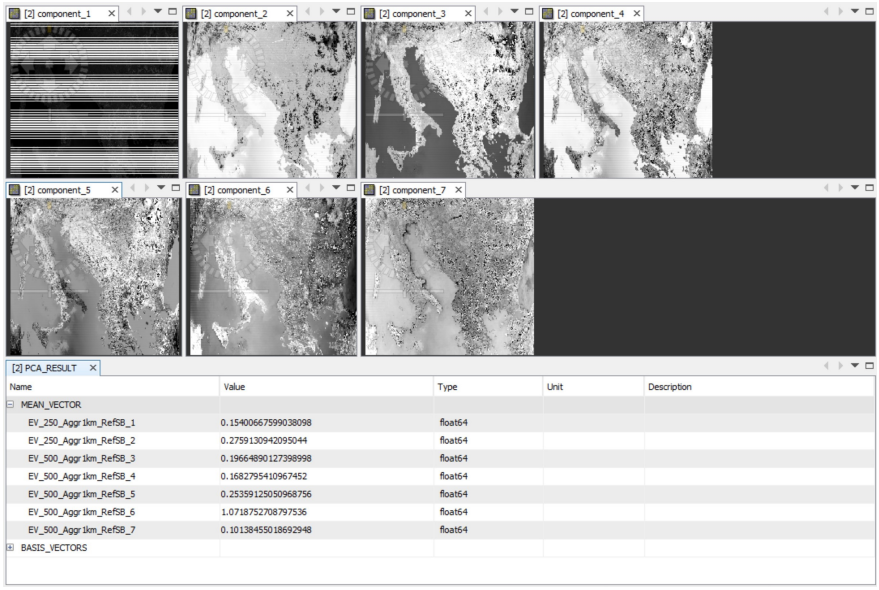


Figure 10: PCA result.

By looking at the MEAN_VECTOR file from the output provided by the PCA, it is possible to order the new bands (the components returned by the PCA) from the one which carries the highest information to the one with less information and noise (figure 11).

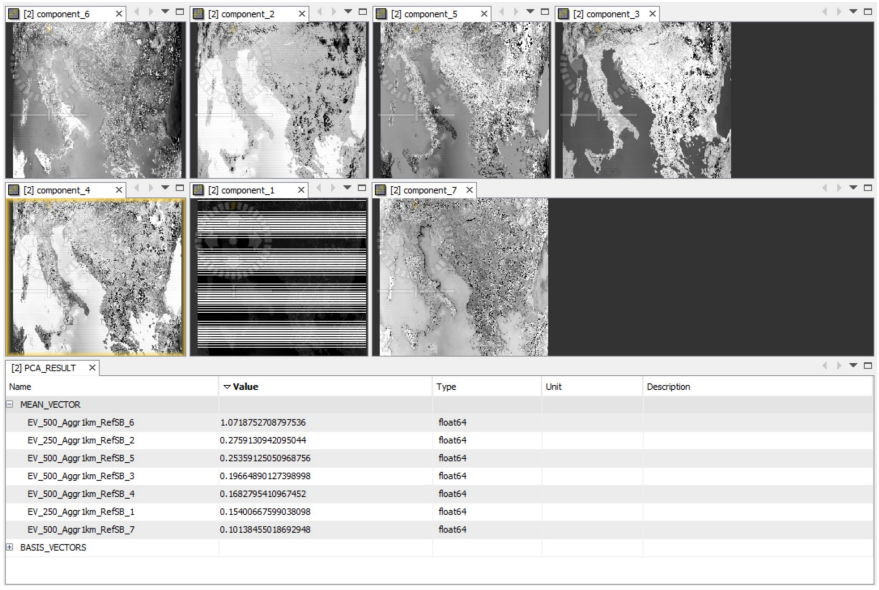


Figure 11: PCA result with ordered components.

Unsupervised and supervised classification. It is also interesting to perform an unsupervised classification. In order to do that, we need to use the mask manager and create coverage masks, masking out the land. Hence, it is possible to perform a k-means clustering, by using three clusters and ten iterations. In figure 12 we can see the result of the algorithm: land, sea and clouds are represented by, respectively, light-yellow, violet and lilac.

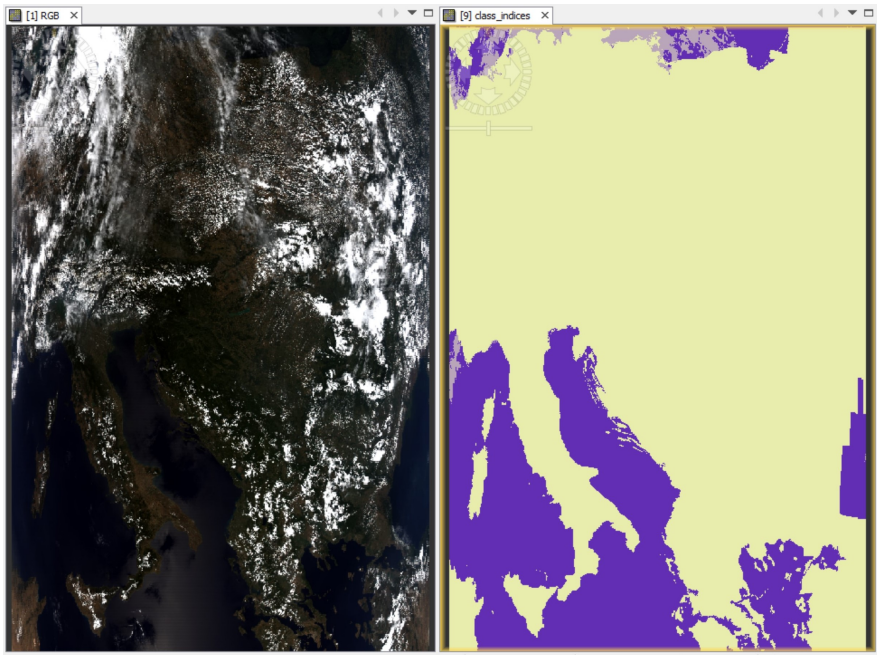


Figure 12: Unsupervised classification performed by K-means, with 3 classes: land (light-yellow), sea (violet) and clouds (lilac).

A comparison can be done by performing, now, a supervised classification. The objective is to distinguish land, sea, clouds and desert. First, we need to create labels for those classes; the SNAP tool allows us to do this by creating new vector data, one for each class, and by selecting, with the polygon drawing tool, small regions of the image, associating them to the corresponding vector data (figure 13). We tried different classification algorithms, such as Random Forest and Maximum Likelihood, with differently labelled data: the best results we obtained can be seen in figure 14 and in the following table.

Table 1: Random Forest results.

Class	Accuracy	Precision	Correlation	Error Rate
CLOUDS	1.0000	1.0000	1.0000	0.0000
DESERT	0.9982	0.9944	0.9952	0.0018
LAND	0.9982	0.9944	0.9952	0.0018
SEA	1.0000	1.0000	1.0000	0.0000

The difference between the supervised and unsupervised classification, in this case, is that the first seems to be highly more precise in identifying clouds over land: this is due to the fact that for unsupervised classification we masked out the land, so the clouds over the land were not considered.

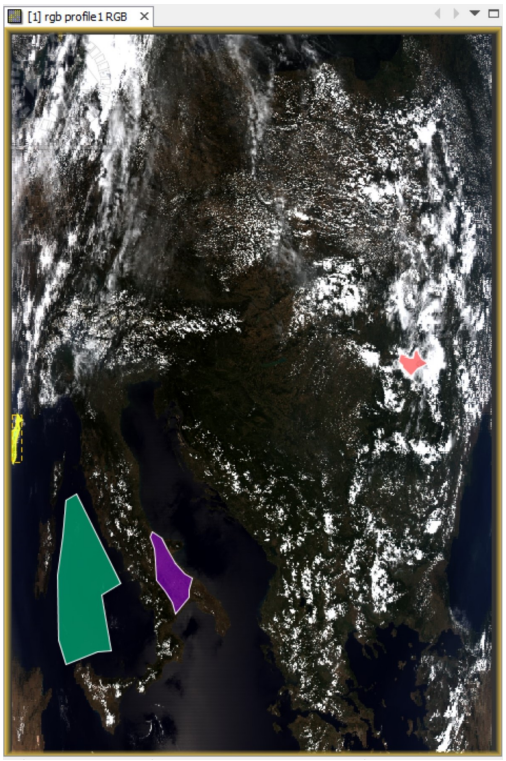


Figure 13: Data vectors 4 classes for supervised classification.

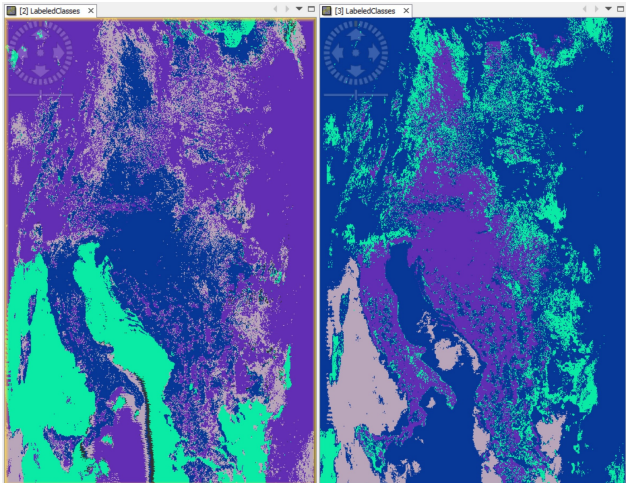


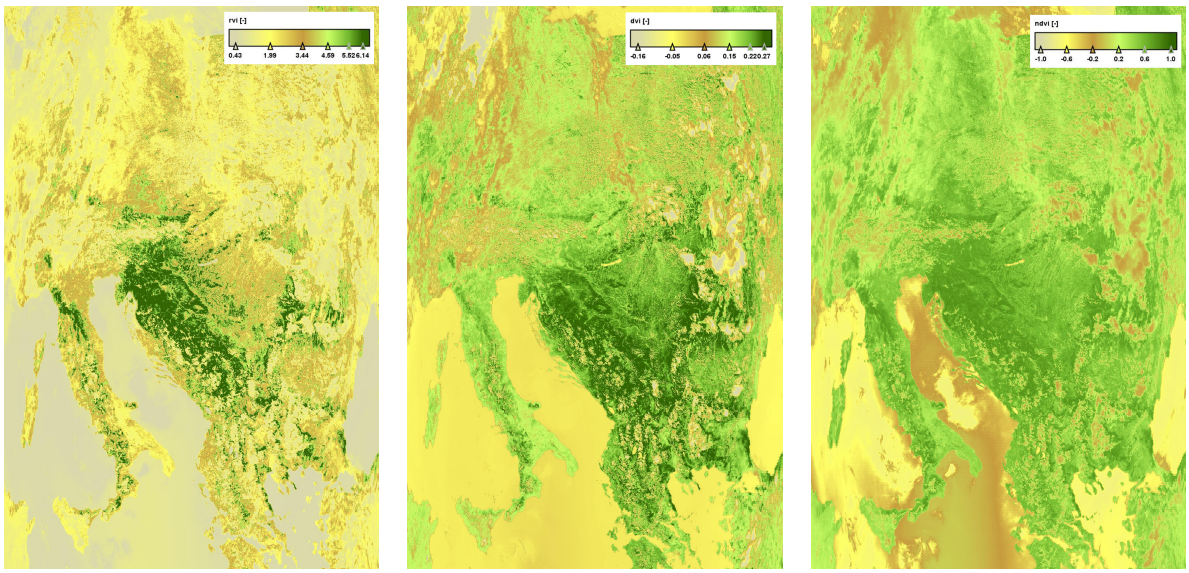
Figure 14: Supervised classification with four classes: Random Forest and Maximum Likelihood results.

Vegetation indices The Vegetation Index (VI) is a quantitative estimate of biomass and vegetation vigour/turgor from space. It is based on the fact that natural surfaces reflect solar radiation similarly in the red and NIR bands, except from vegetation; moreover, vegetation absorbs red wavelengths but at the same time it reflects the NIR ones.

Consider the following 2-band vegetation indices:

- $RVI = NIR/RED$
- $DVI = NIR - RED$
- $NDVI = \frac{NIR-RED}{NIR+RED}$

In figure 15 we can see the results of the application of the three indices over the image, after applying the Meris Vegetation Index colour palette.

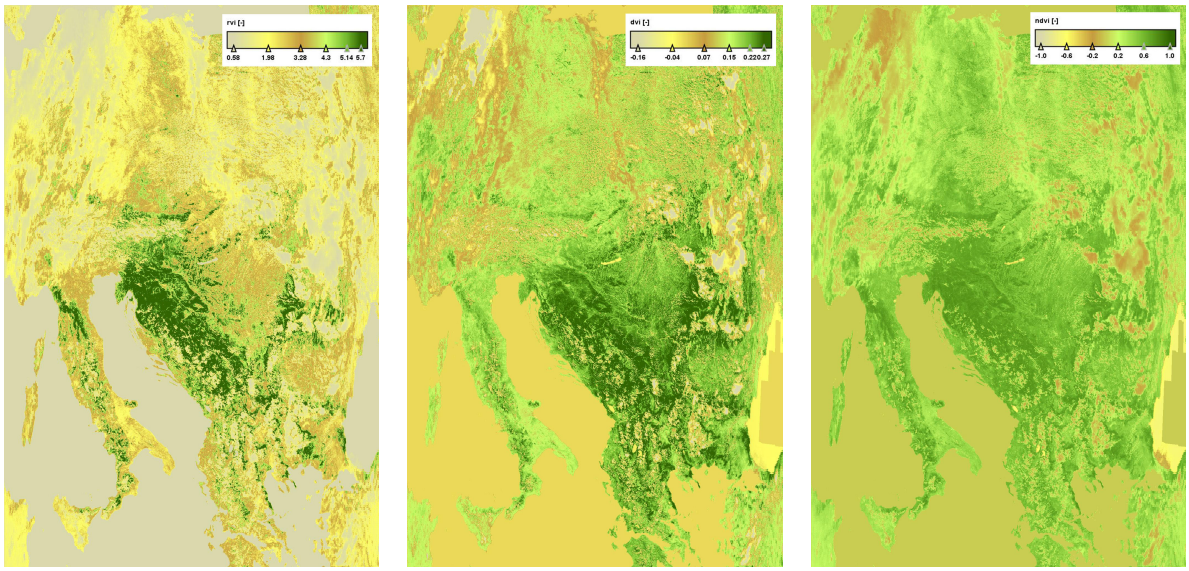


(a) *RVI* (b) *DVI* (c) *NDVI*

Figure 15: Vegetation indices.

RVI (Ratio Vegetation Index) is based on the principle that leaves absorb relatively more red than infrared; it is widely used for green biomass estimations and monitoring, in particular, at high density vegetation coverage since this index is very sensitive to vegetation. But, when the vegetation cover is sparse, RVI is very sensitive to atmospheric effects. The Difference Vegetation Index (DVI) is instead very sensitive to changes in soil background. In the end, the NDVI (Normalized Difference Vegetation Index) is the most used one; this is due to the fact that it is computed through a normalization procedure so it has a very sensitive response to green vegetation even for low vegetation covered areas.

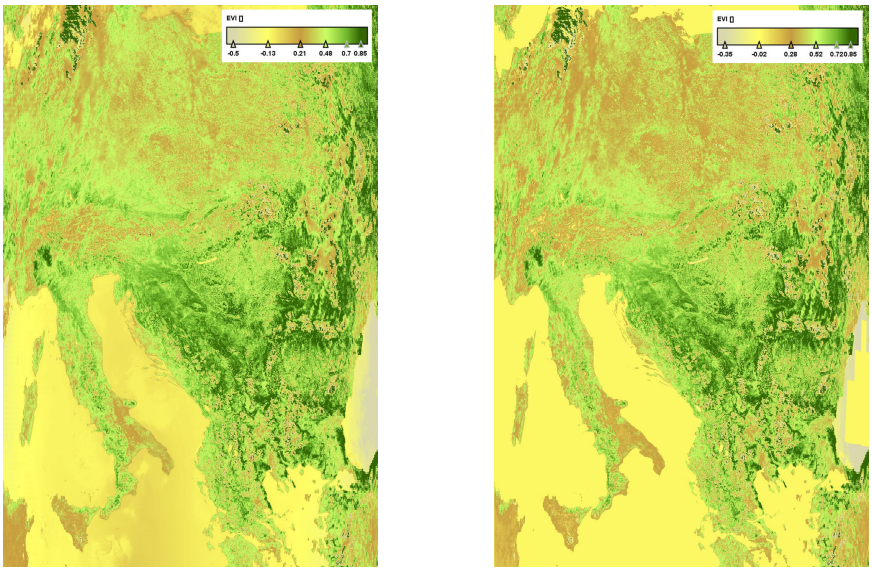
Better results can be obtained by considering a 3-band vegetation index, such as the EVI (Enhanced Vegetation index). It is an optimized VI designed to enhance the vegetation signal; some advantages wrt NDVI are the correction of some distortions in the reflected light caused by the particles in the



(a) *RVI* (b) *DVI* (c) *NDVI*

Figure 16: Vegetation indices with masks.

air (the ground cover below the vegetation) and the avoidance of saturation as easily as the NDVI when viewing areas of Earth with a large amount of chlorophyll.



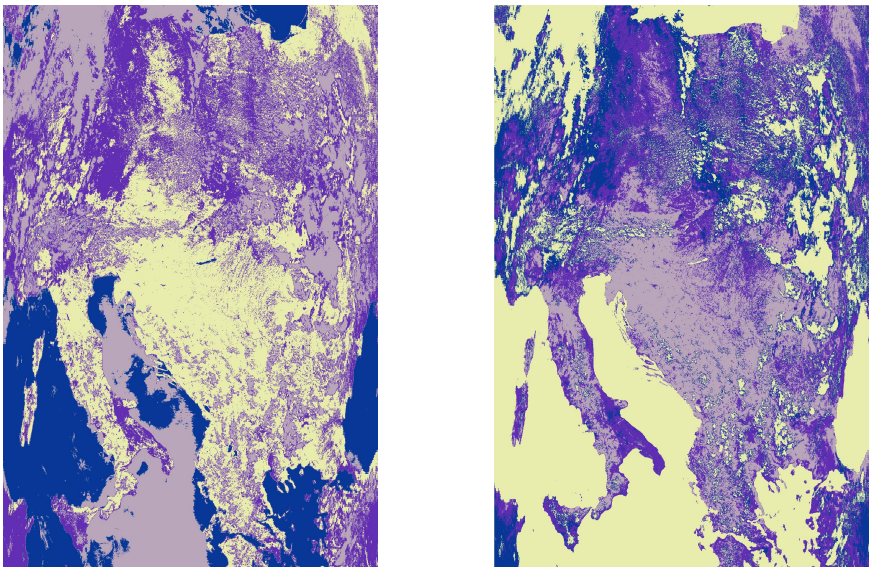
(a) *EVI* (b) *EVI mask*

Figure 17: Comparison between the application of the EVI index (3-band vegetation index) over the image without (a) and with mask (b).

A way to improve the results can be applying vegetation indices only on land pixels. In order to

do this, we have to mask out the sea and apply again the indices to the masked image. In figure 16 and 17 b) it is possible to see the new images obtained with this procedure.

Classification with NDVI Now, consider the NDVI result in both cases (without and with the mask over the sea): it is possible to use it in order to perform an unsupervised classification. As we did previously, we perform the K-means algorithm with four classes and 30 iterations. Hereafter we can see the results obtained (figure 18) and in the following table, it is possible to investigate the NDVI values obtained for each class.



(a) *K-means NDVI* (b) *K-means NDVI mask*

Figure 18: K-means results from NDVI without (a) and with the mask (b).

Table 2: kmeans results.

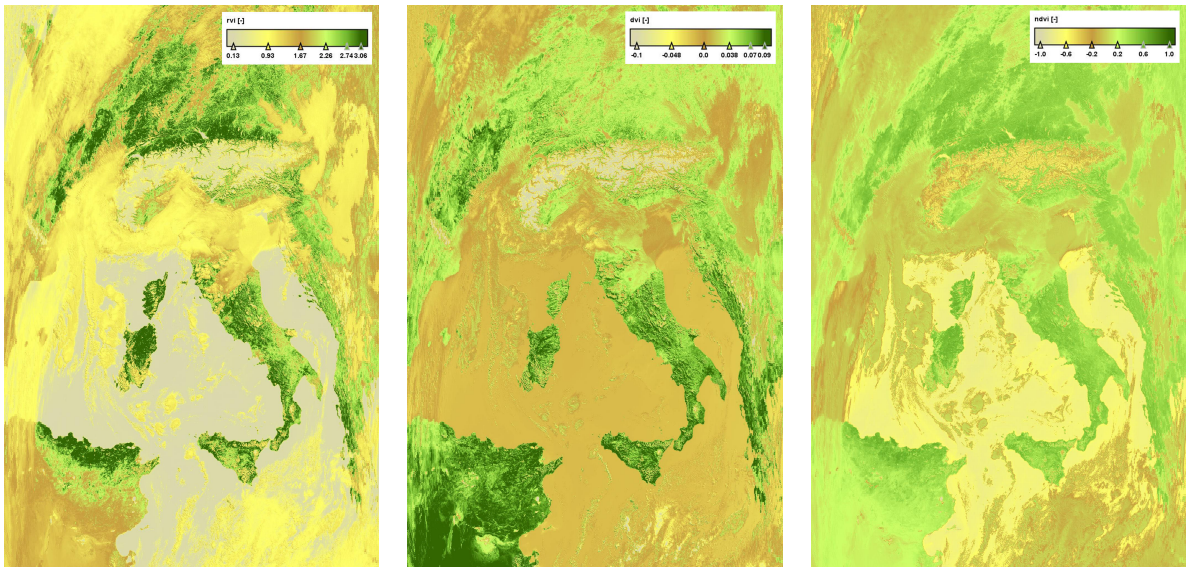
(a) *K-means, NDVI*

Cluster	Colour	Centroid
0	Light Yellow	NDVI = 0.581
1	Purple	NDVI = 0.306
2	Lilac	NDVI = −0.045
3	Blue	NDVI = −0.584

(b) *K-means, NDVI with sea masked out*

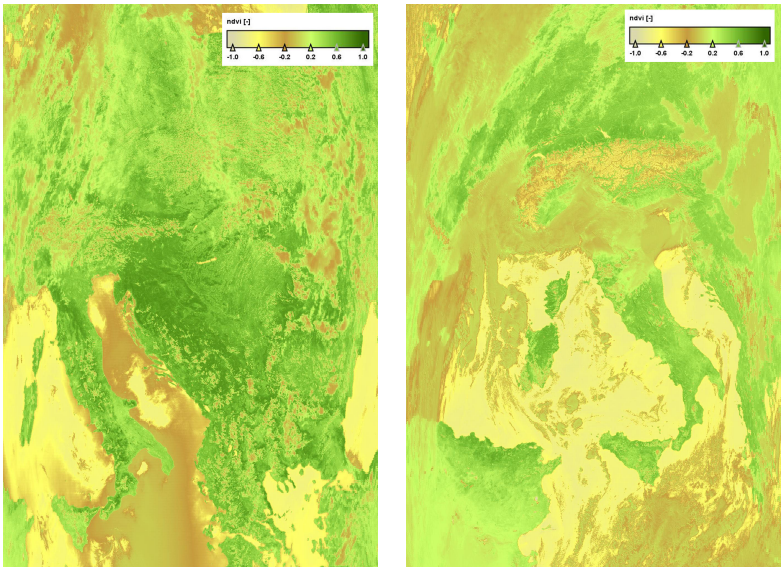
Cluster	Colour	Centroid
0	Light Yellow	NDVI = −0.018
1	Purple	NDVI = 0.443
2	Lilac	NDVI = 0.638
3	Blue	NDVI = 0.227

Winter and summer seasons comparison and quantitative change detection. Now, if we consider the same region but in a winter season, we can compare the vegetation status w.r.t. the summer; the following images show the three vegetation indexes over Italy in January 2020.



(a) *RVI* (b) *DVI* (c) *NDVI*

Figure 19: Vegetation indices January 2020.



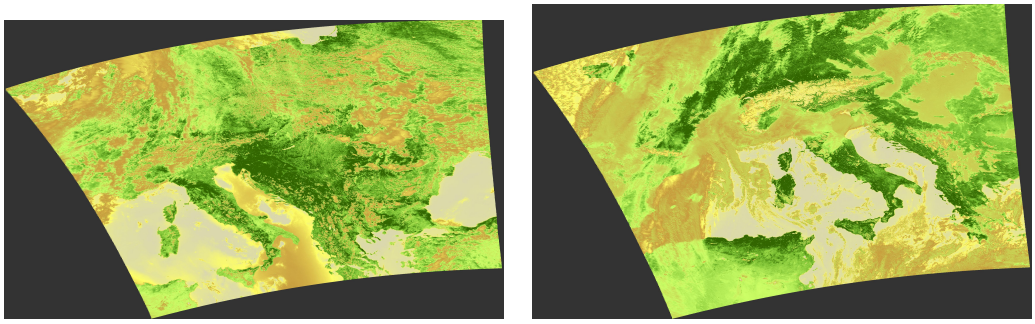
(a) *NDVI in July* (b) *NDVI in January*

Figure 20: Comparison between NDVI indices in July 2019 and January 2020 over Italy.

At the first glance, it seems that there is more vigorous vegetation in summer rather than in winter (figure 20); but we can see that the desert area is considered as a vegetation area, which should

not be the case, and also the clouds in the winter season affect the analysis. These problems and misunderstandings may be avoided with a more precise quantitative comparison.

The quantitative change detection between the two seasons can be done by first reprojecting the two images over the same grid (figure 21) and then by performing the difference between the two NDVI bands, over a selected ROI (figure 22).



(a) (b)

Figure 21: Reprojected summer and winter images.

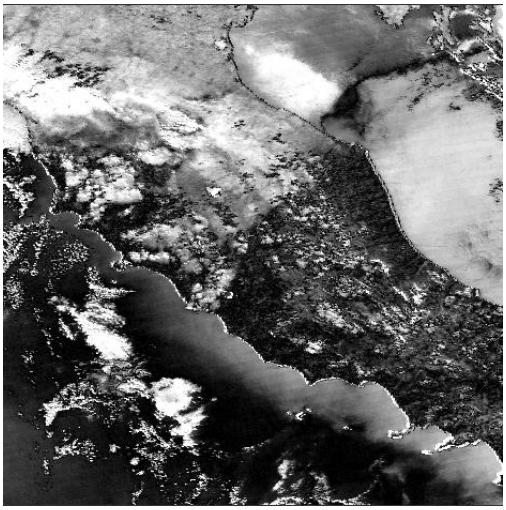


Figure 22: Quantitative change detection of the vegetation coverage between winter and summer images.

Better results can be achieved by first masking out the clouds through the Band Maths tool in Snap, with the expression: `if ndvi >= 0.3 then ndvi else 0` and then by performing again the difference (figure 23).

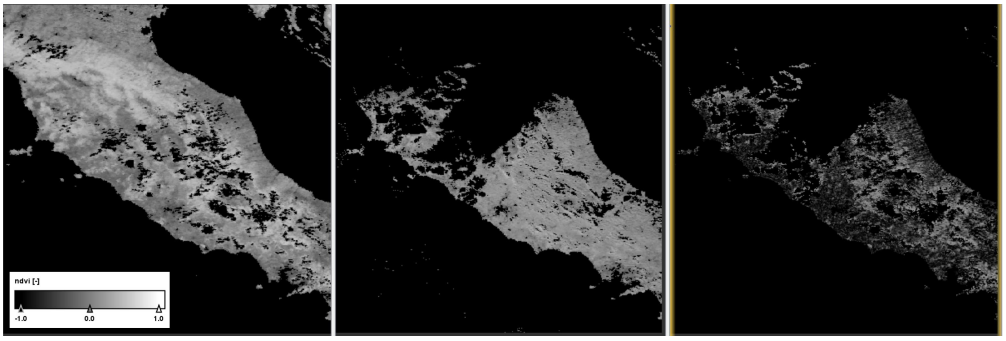


Figure 23: Quantitative change detection (difference in absolute value) with masks.

Moreover, we can highlight the difference between the vegetation coverage through the two seasons through figure 24; in fact, the pixels in dark green represent a higher NDVI value in July w.r.t. January, while the pixels in yellow represent the opposite. So, the prevalence of green pixels in this image shows that in July there is an healthier and more vigorous vegetation than in January.

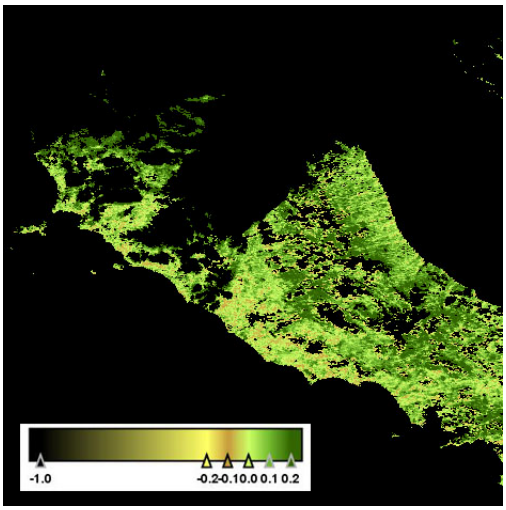


Figure 24: Change detection.

Conclusions In this homework, MODIS-Aqua images have been analyzed; we have seen how spectrum, histogram and profile tools and data correlation can be very useful in the interpretation of different features of the images, such as land, sea or clouds, in particular in the summer season. We performed PCA, unsupervised and supervised classification over the data by means of different classifiers with three and four classes, showing the main benefits and differences of the choices made. Then, we implemented 2-band and 3-band formulas for the vegetation indexes, which were applied to the land pixels in the winter and summer images showing the differences in terms of health status of vegetation and coverage. The NDVI result was used to perform an unsupervised classification through K-means algorithm. Finally, we performed a quantitative change detection of the vegetation coverage class by reprojecting the two images in winter and summer seasons over the same grid and in the same ROI, proving that there is a more vigorous and healthier vegetation

in summer rather than in winter.

References

[1] Andrés Viña, Anatoly A. Gitelson, Anthony L. Nguy-Robertson, Yi Peng, "Comparison of different vegetation indices for the remote assessment of green leaf area index of crops", Elsevier 2011

[2] Ranga B. Myneni, Forrest G. Hall, Piers J. Sellers and Alexander L. Marshak, "The interpretation of spectral vegetation indexes", IEEE Transactions on Geoscience and Remote Sensing, vol. 33, No.2, March 1995