Earth Observation Data Analysis - Homework 02

Gioia Mancini, Christian Brignone

August 31, 2020

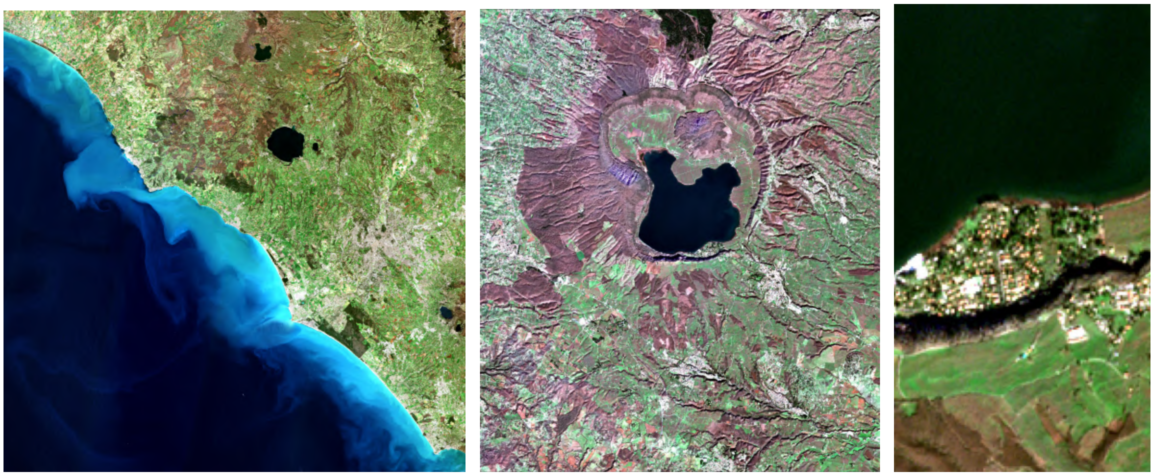
In this homework, we will explore Sentinel-2 MSI data in order to estimate vegetation cover, inland water, chlorophyll-a sea concentration and supervised classification in a particular ROI.

Data exploration Let us consider the first product: the Tyrrhenian coast of Lazio (in figure 1 (a) we can see the RGB view obtained by combining Sentinel-2 MSI Red, Green and Blue bands), in July 2017. In particular, in figure 1 (b), we have considered the ROI around Ronciglione lake. This was done by using the SNAP GraphBuilder tool, through which we first resampled the image at $10m$ pixel resolution for all bands, except for the bands at 20 and 60 m resolution, for which an upsampling *nearest* method is needed; then we computed the subset for all bands around the desired region, specified by geographic coordinates with the predefined geometry in 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)).`

(a) (b) (c)

Figure 1: RGB image of the Tyrrhenian coast of Lazio, Italy taken in July 2017 (a), ROI around Ronciglione lake (b) and zoom in to observe visible details (c).

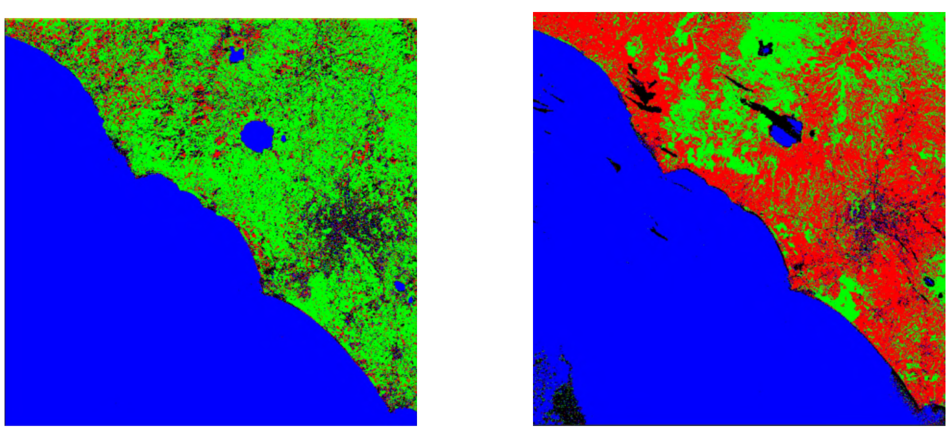
Then, by using the batch processing tool, it is possible to apply the created graph to the other considered S2 products, for example in figure 2 (a) we can see the same region as before but in a winter season, and then cut around the ROI and resampled (b): all the bands have now the same resolution.



(a) (b) (c)

Figure 2: RGB image of the Tyrrhenian coast of Lazio, Italy taken in February 2018 (a), ROI around Ronciglione lake (b) and zoom in to observe visible details (c).

It is possible to use the scene classification masks provided by the products to generate an RGB image, for both winter and summer seasons, as we can see in figure 3. In fact, we can use *not vegetated* as the red band, *vegetation* as the green band and *water* as the blue band.



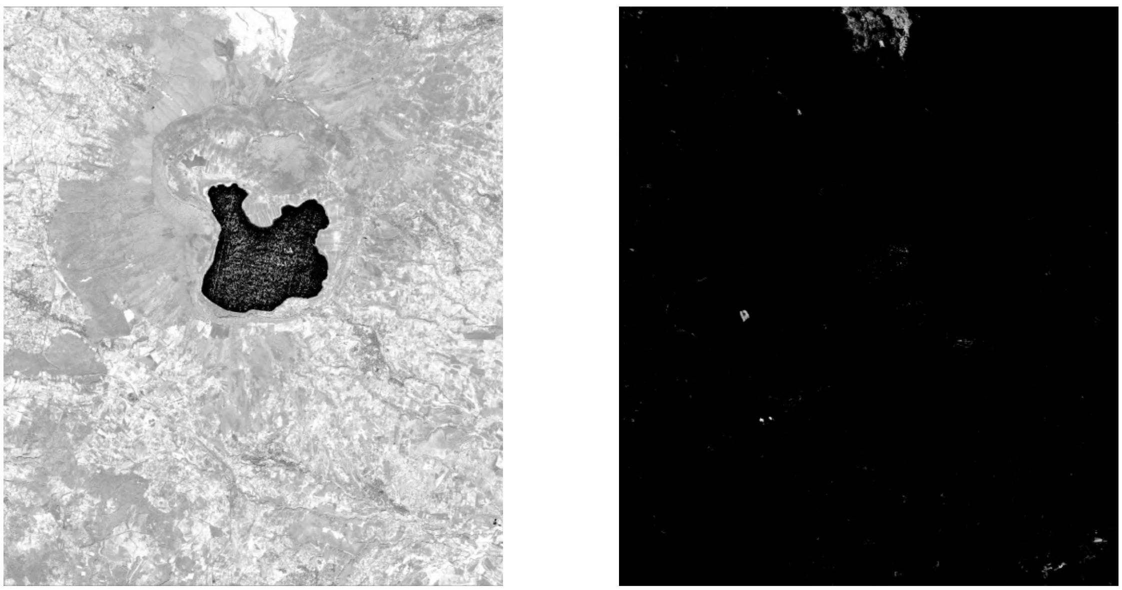
(a) (b)

Figure 3: RGB images generated by the scene classification masks: *no vegetation*, *vegetation* and *water* in February (a) and July (b).

Water and vegetation normalized indexes Now, starting from the resampled subset products, we will calculate the Normalized Difference Water Index (NDWI) and the second NDWI (NDWI2): the first one to put in evidence the water content in the leaves at canopy level; the second one to put in evidence in-land water surface.

We first perform the NDWI over Ronciglione Lake in February 2018: in figure 4 (a) it is possible to see the result of the application of the NDWI index, which is described by the following formula. We will define, respectively, B12 for MIR and B8 for NIR source bands.

$$NDWI = \frac{\rho_{NIR} - \rho_{SWIR}}{\rho_{NIR} + \rho_{SWIR}} \tag{1}$$



(a) (b)

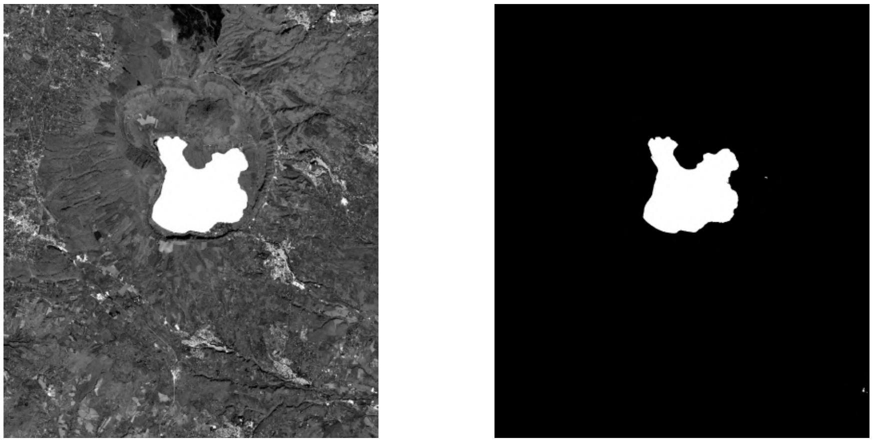
Figure 4: NDWI index (a) water content in vegetation (b) in February.

If we modify the thresholds in the colour histogram to put in evidence only very high values, what we obtain is shown in figure 4 (b): the image is quite dark and only some white spots are left in the whole region. The bigger white spot in the upper part of figure 4 (b) corresponds to a wild vegetation area in the RGB image. This index can be used in this way to monitor the status of vegetation, in particular the water content in the leaves at the canopy level.

The NDWI2, instead, can be used to put in evidence the water content in the lake. B3 and B8 bands will be chosen for Green and NIR respectively.

$$NDWI2 = \frac{\rho_{Green} - \rho_{NIR}}{\rho_{Green} + \rho_{NIR}} \tag{2}$$

The Second NDWI (figure 5 (a)) seems to be very high in the lake, while it is very low elsewhere. So, we can manipulate the colour histogram to isolate the water in the lake from the rest of the image (figure 5 (b)).

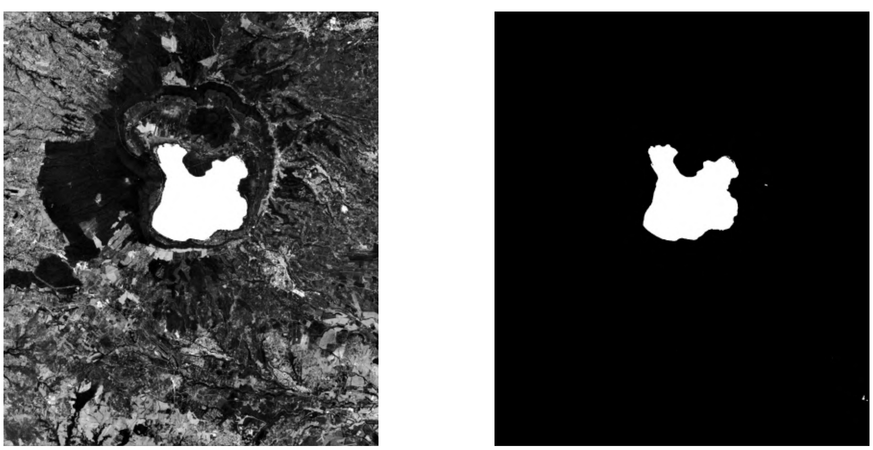


(a) (b)

Figure 5: NDWI2 index in February (a) and water content in the lake (b).

The geometry of the lake is perfectly represented by this index, which can be used to build a binary water mask (figure 7 (a)). This was done by using the BandMaths operator and defining a new band, described by the following expression: `if NDWI2 > 0.3 then 1 else 0`.

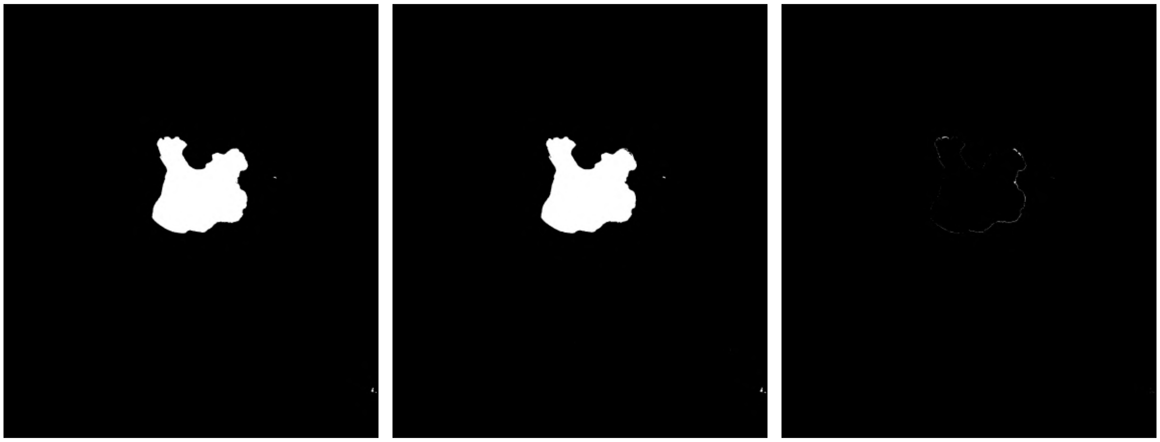
Consider now the same process but in a summer season, for example August, in order to compare the water content in Ronciglione lake in two different seasons. In figure 7 (b) we can see the resulting water mask obtained by using the NDWI2 in august and the same expression as before. Then, in figure 7 (c) it is described the difference between the water mask in February and the water mask in august.



(a) (b)

Figure 6: NDWI2 index in august (a) water content in lake (b).

Through the computed difference between the two masks, we obtained the contour of the lake, which is the water content difference between the two seasons. As we expected, the water content

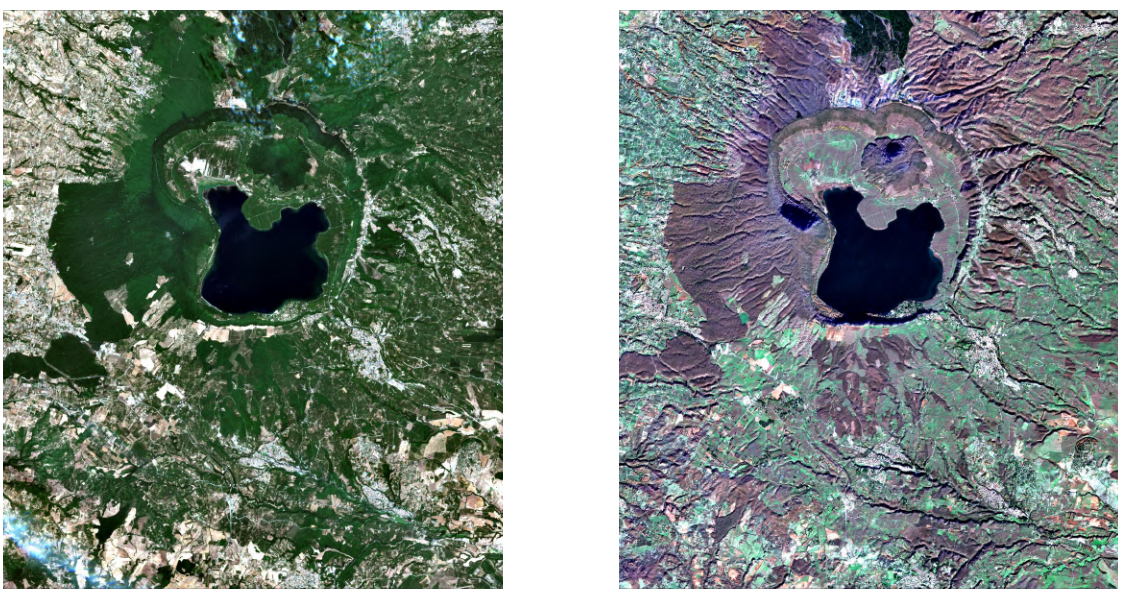


(a) (b) (c)

Figure 7: Water mask in February (a), august (b) and water mask difference between the two (c).

was higher in the winter season.

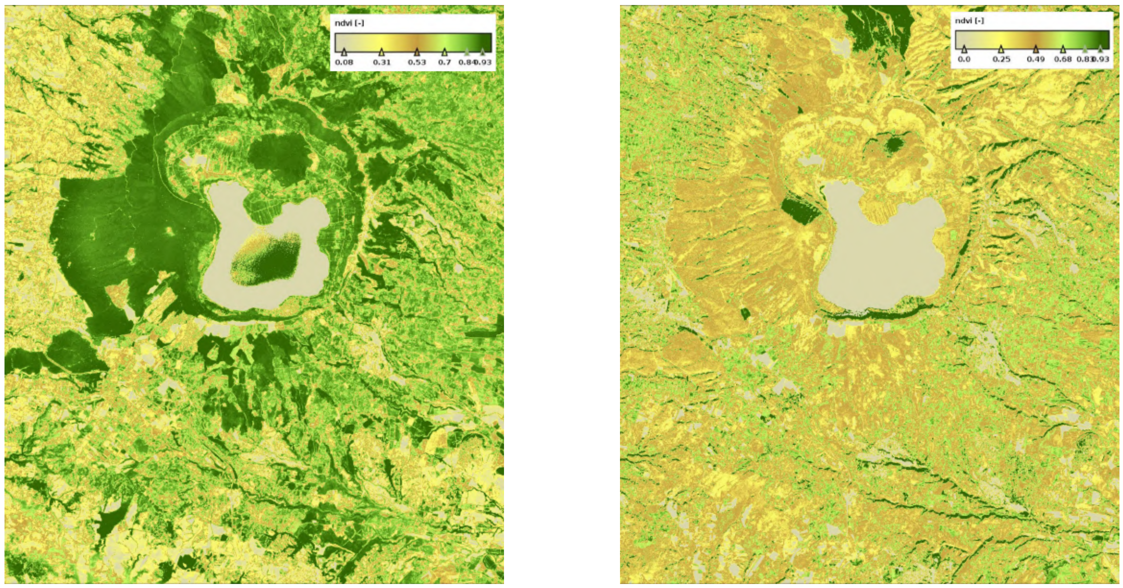
Now, it is interesting to consider the NDVI index, in order to see the amount and state of vegetation on the ground. This was done over two resampled and subset products, in July and December 2017 (figure 8).



(a) (b)

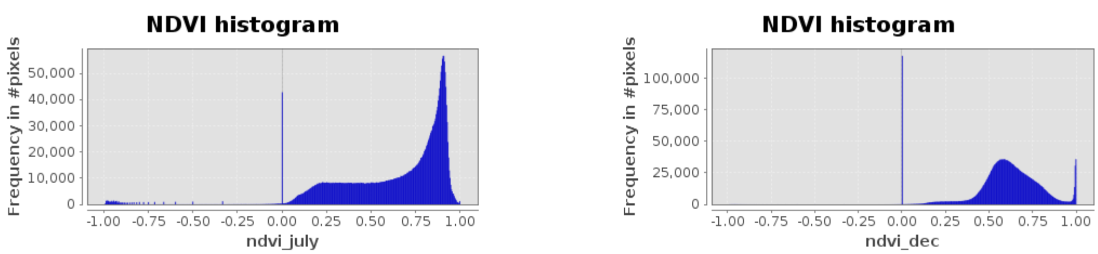
Figure 8: Ronciglione lake in July (a) and December (b).

So, the NDVI band for July and December is shown in figure 9. B4 and B8 were chosen for red and NIR source bands; green is associated with large values and brown with small values: it is easily possible to appreciate the difference between the two seasons in small agricultural fields. In fact, this is very useful to monitor the status of an agricultural field, in order to decide the most suitable time for harvesting, for example.



(a) (b)

Figure 9: NDVI result for July (a) and December (b) over Ronciglione lake.



(a) (b)

Figure 10: NDVI histograms for July (a) and December (b) over Ronciglione lake.

From the histogram we can deduce that in average value there are more pixels associated with green in the summer image; we can see a high concentration of pixels around zero in the winter image, but there is also a peak in 1 where about 40000 pixels are associated with green: they correspond to the small but very vigorous vegetation areas in the upper part of the image.

Sea chlorophyll-a and sediment estimation Now, we will consider the Maximum Chlorophyll Index, which is used to study the presence of chlorophyll in the water; it is also an indicator of the presence of algae. The MCI exploits the height of the measurement in a spectral band above a baseline; the band combination for Sentinel-2 MSI data is B4, B5 and B6. The formula for calculating the MCI is the following;

$$MCI = L_2 - k\left(L_2 + (L_3 - L_1)\frac{\lambda_2 - \lambda_1}{\lambda_3 - \lambda_1}\right) \tag{3}$$

For this purpose, we will consider the Tiber river in Lazio, in both august 2017 and February 2018. It is possible to resample and subset the images as done previously, this time along the coastline and near the Tiber estuary.

(a)

(b)

Figure 11: Tiber river RGB view in February (a) and august (b).

Note that, if we consider the spectrum view, by moving over the image we see that in the clean water area, there is a peak in the spectrum shape in blue; if we move near the estuary the spectrum is peaked in green: this denotes the presence of suspended matter and higher concentration of phytoplankton.

First, we will implement the *EmpReg* (empirical regressive) algorithms [1] to estimate Chlorophyll-a concentration and Total Suspended Matter using the SNAP formula processing tool, then we will compare the result with the MCI application from Snap plugged-in algorithm. The empirical regressive algorithm for Chl-a is a blue-to-green reflectance maximum band ratio (MBR) model:

$$r_{MBR} = \frac{max(R_{wlB1}, R_{wlB2})}{R_{wlB3}} \tag{4}$$

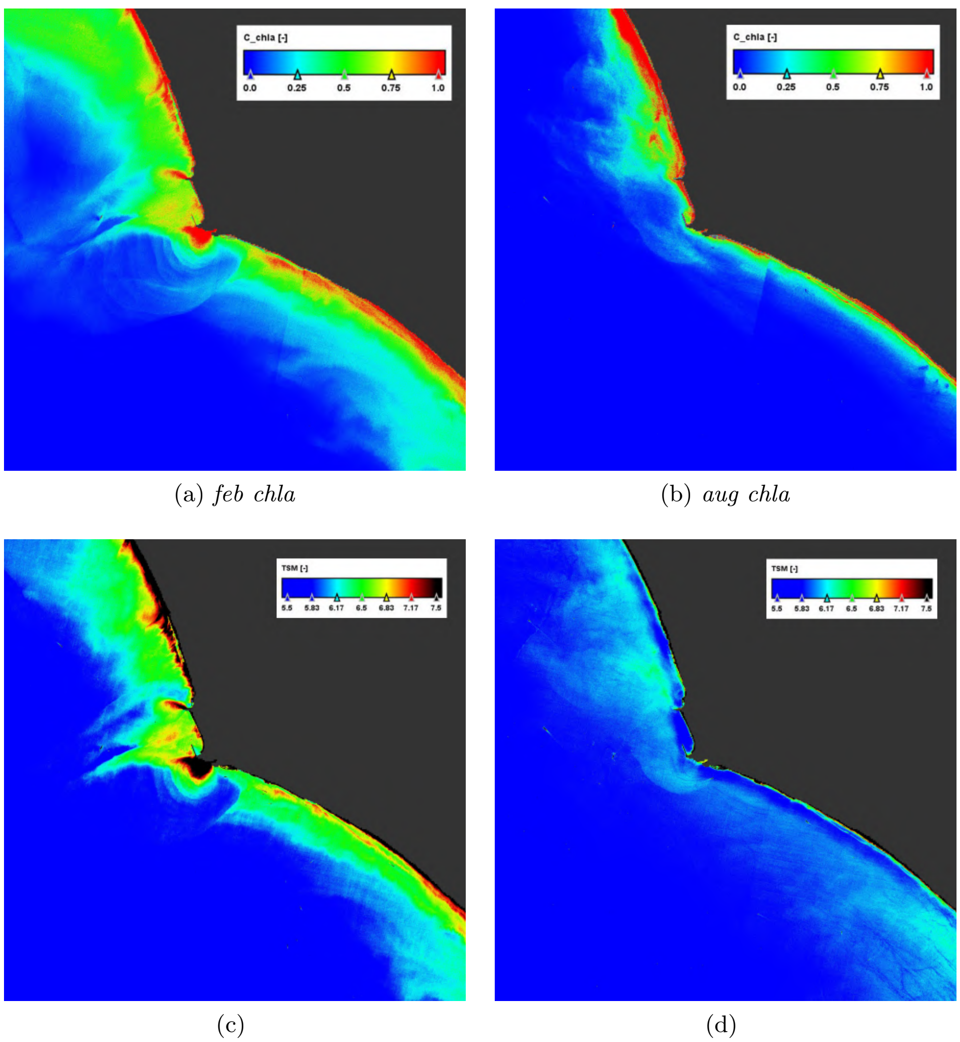
This model makes use of the higher reflectance at the green wavelength caused by phytoplankton, so that a high value of MBR corresponds to low Chl-a concentrations and vice versa. Red-edge wavelengths are effective in waters with very high Chl-a concentrations but more susceptible to reflection effects near the coast. The expression of the optimal EmpReg retrieval algorithm is the following:

$$\hat{C}_{Chla} = a_1 \exp(-a_2 r_{MBR}) \tag{5}$$

where $a_1 = 59.795 mg/m^3$ and $a_2 = 4.559$. Another exponential model, making use of the water-leaving reflectance at B4 is obtained for TSM retrieval, as described in the following expression:

$$\hat{C}_{TSM} = b_1 \exp(b_2 R_{wlB4}) \tag{6}$$

In the following images (figure 12) Chla-a and TSM are shown for February and august products,

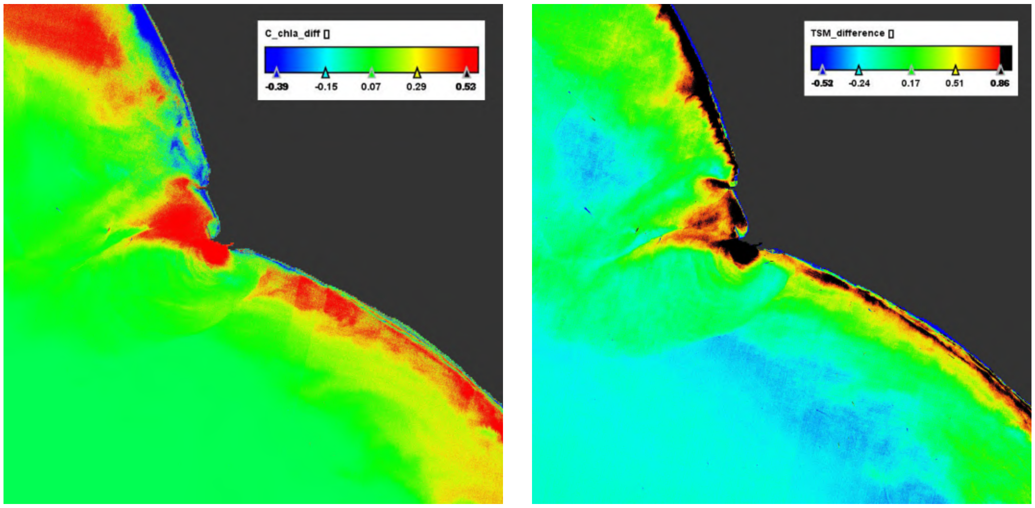


(a) *feb chla* (b) *aug chla*

(c) (d)

Figure 12: Chlorophyll-a concentrations (a), (b) and TSM (c), (d) over the winter and summer products.

around the Tiber estuary in Lazio. Also, in figure 13 we can see the Chl-a and TSM differences respectively from winter to summer season.



(a) (b)

Figure 13: Chl-a (a) and TSM (b) differences from winter to summer.

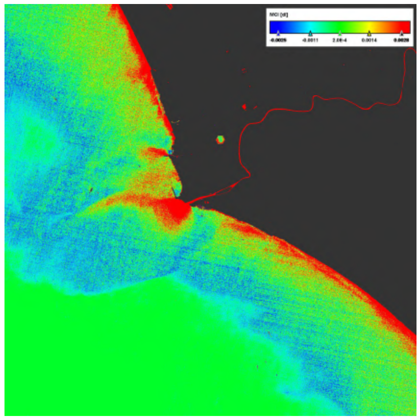
What is evident is that chlorophyll-a and TSM concentrations are more intense over the coastline, as we expected. In fact, the blue part means an area of clean water, the green part corresponds to a higher concentration of phytoplankton, while dead organic and inorganic matter plus suspended materials presence is shown in the red-yellow part of the image.

We will now show the Maximum Chlorophyll Index (MCI) computed from Snap plugged-in algorithms over this area, in order to compare it with the one previously computed with the EmpReg algorithms in the BandMaths tool. So, we use the S2 MCI Processor, selecting B6 and B4 as lower and upper baseline bands; then, a mask expression is defined: the *scl_water*, which means that the computation is performed only on water and the land is masked out. What we obtained is shown in figure 14.

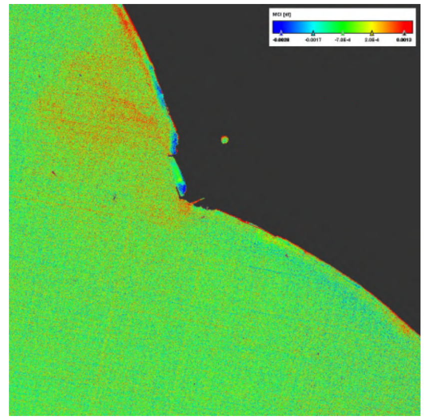
It is clear that the higher values, in red, are along the coastline and the river, which is more evident in particular in the winter season. We can also see the images in Google Earth in figure 15.

In the end, we can make the comparison between the Chlorophyll-a concentration computed with the EmpReg retrieval algorithm (figure 12 (a) and (b)) and the MCI computed through Snap (figure 14): it is clear that the computation given by the EmpReg algorithm seems more precise and realistic w.r.t. the MCI computed by Snap, which in some cases happens to be very noisy.

A final consideration can be done by comparing our results with a Chl-a product downloaded from the International Copernicus Marine Service [4]. The Global Ocean Satellite monitoring and marine ecosystem study group (GOS) of the Italian CNR, in Rome, distributes surface chlorophyll concentration (mg m-3, 1 km resolution, while our products are at 10 m resolution) derived from multi-sensor (MODIS-AQUA, NPP-VIIRS, and Sentinel3-OLCI) and Sentinel3-OLCI Rrs spectra. Chlorophyll datasets are obtained by means of the Mediterranean Ocean Colour regional algorithms.

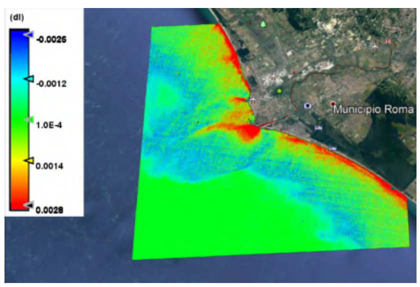


(a)

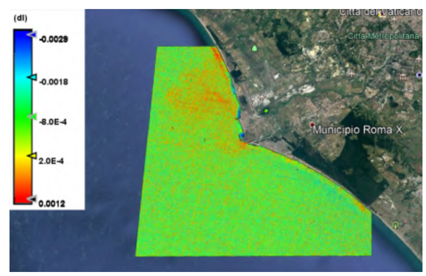


(b)

Figure 14: Maximum Chlorophyll Index computed over the winter (a) and summer (b) products.



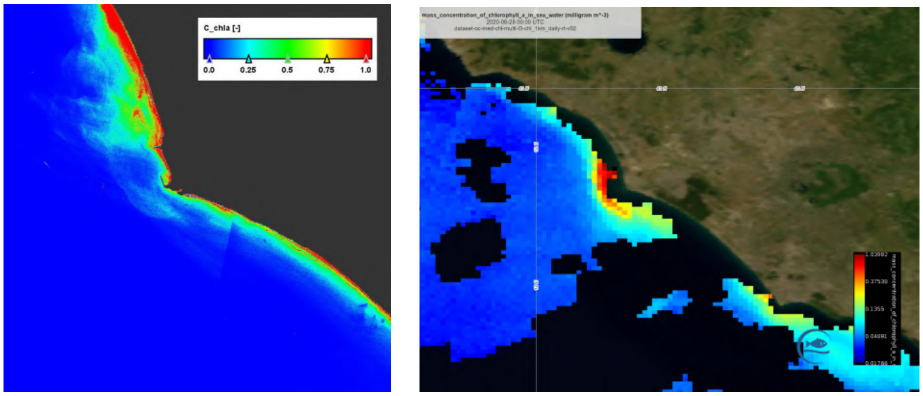
(a)



(b)

Figure 15: MCI results visualized in Google Earth.

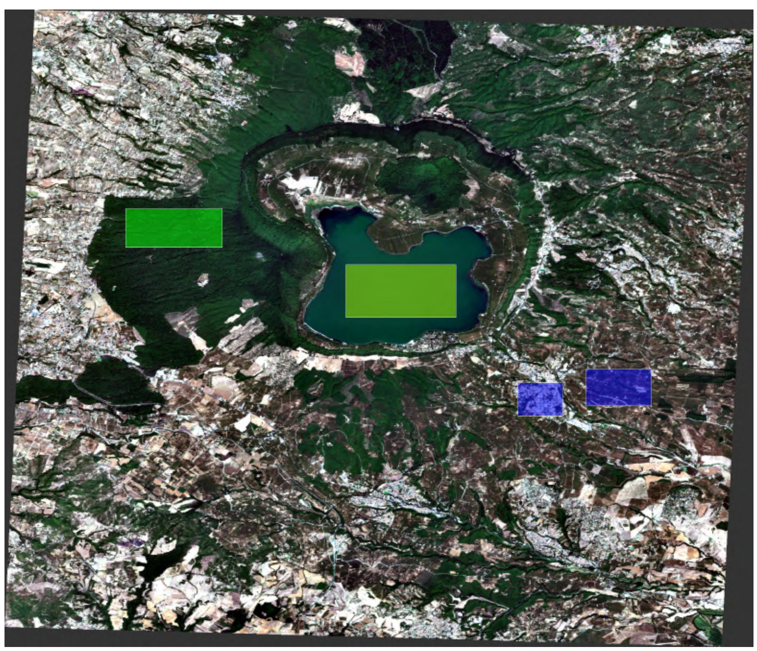
This product identifies the average chlorophyll content of the surface layer as defined by the first optical depth (roughly one-fifth of the euphotic depth). For multi-sensor observations, single sensor Rrs fields are band-shifted, over the SeaWiFS native bands (using the QAAv6 model, Lee et al., 2002) and merged with a technique aimed at smoothing the differences among different sensors.



(a) (b)

Figure 16: Chl-a concentration in august: comparison between the chl-a computed through the EmpReg algorithm over Tiber estuary in august 2017 (a) and the chl-a Copernicus product over the same area in august 2018 (b).

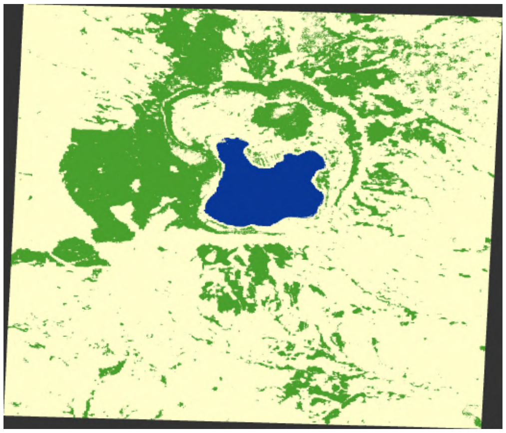
Scene classification The last exercise is about scene classification: in particular, we will consider a training dataset from the Ronciglione Lake resampled product. We define four vector data containers: *vegetation*, *water*, and *not vegetated* area; they will be used as training areas for the Maximum Likelihood classifier. Before running the algorithm, we reproject the image (figure 17).



(a)

Figure 17: Reprojected image of august over Ronciglione lake, with labels.

Then, after the application of the Maximum Likelihood classifier, what we obtain is shown in figure 18. We have vegetation in green, water in blue and not vegetated area in light yellow.



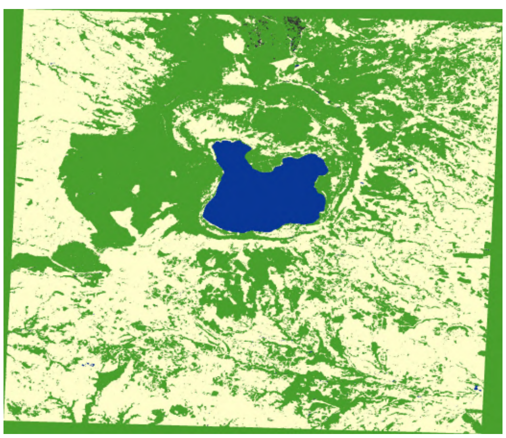
(a)

Label	Colour	Value	Frequency	Description
no data		-1	0.000%	no data
vegetation		0	21.622%	
not_vegetated		1	74.833%	
water		2	3.546%	

(b)

Figure 18: Result of Maximum Likelihood classification over Ronciglione lake in august, with three classes.

If we want to compare the result obtained with another classifier, we can for example consider the Random Forest one and look at the results in figure 19. The water area is the same, but the vegetation area seems to be much larger than the one recognized by the Maximum Likelihood classifier.



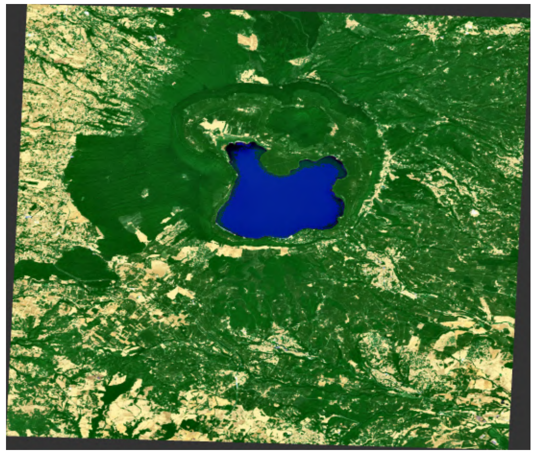
(a)

Label	Colour	Value	Frequency	Description
no data		-1	0.000%	no data
vegetation		0	44.872%	
not_vegetated		1	51.447%	
water		2	3.681%	

(b)

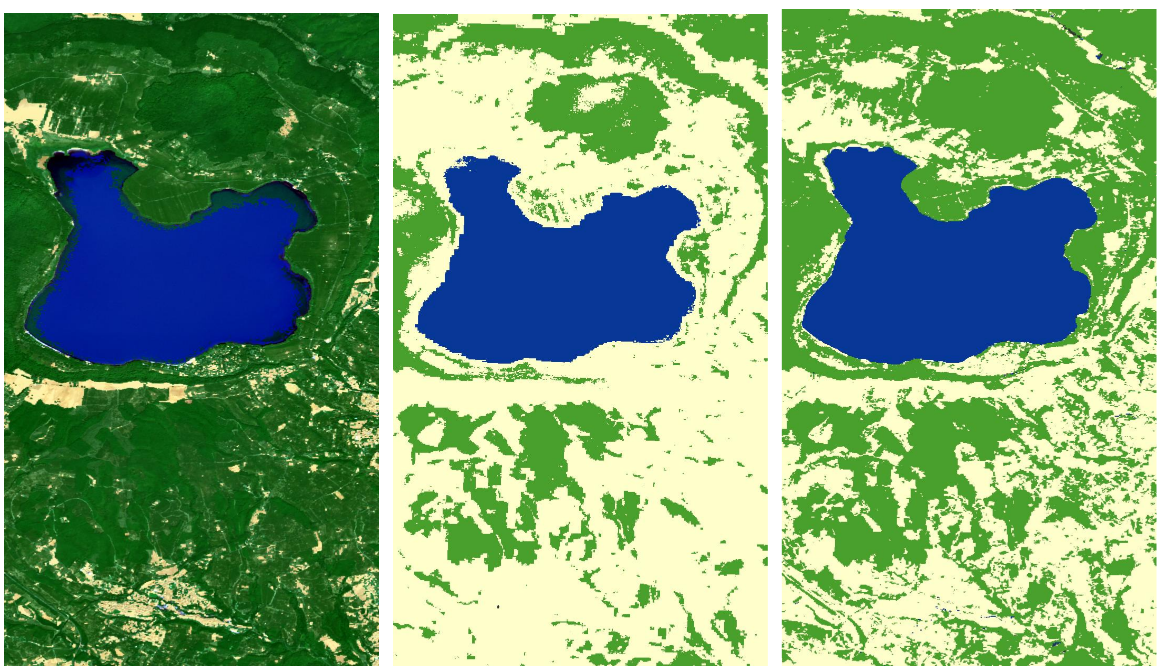
Figure 19: Result of Random Forest classification over Ronciglione lake in august, with three classes.

So, we will compare the results of the scene classification with the image obtained by applying to the image the scene classification masks of the product itself, as shown in figure 20. Moreover, we can also show some differences between the two classifiers by looking closer at a small area around the lake: figure 21. It is evident that the best result is given by the Random Forest classifier, which areas are in fact the most similar to the masks provided by the product.



(a)

Figure 20: Image built with the scene classification masks *vegetation*, *water* and *not vegetated*



(a) (b) (c)

Figure 21: Comparison between the image built with the scene classification masks *vegetation*, *water* and *not vegetated* (a) and the results from Maximum Likelihood (b) and Random Forest Classification (c)

Conclusions We can conclude that the S2 MSI data can be properly analyzed and processed in order to use them to provide very useful information about both vegetation coverage and health status variations between different seasons or water content in the leaves and in a lake, with the help of the water and vegetation normalized indexes (NDVI, NDWI and NDWI2); another very important application is the sea chlorophyll and sediments estimation which, by means of tools and techniques such as EmpReg algorithms, can provide a valid understanding of the water quality near a river estuary or, in general, along a coastline. Finally, we have shown some results about scene classification and it turned out that, through a comparison with the scene classification masks provided by the products, the best performance was given by the Random Forest classifier.

References

[1] Frank S. Marzano, Fellow, IEEE, Michele Iacobelli, Massimo Orlandi, and Domenico Cimini, "Coastal water remote sensing from Sentinel-2 satellite data using physical-statistical and neural-network retrieval approach", IEEE TGRS 2020

[2] N. Sravanthi, I. V. Ramana, P. Yunus Ali, M. Ashraf, M. M. Ali, and A. C.Narayana, "An Algorithm for Estimating Suspended Sediment Concentrations in the Coastal Waters of India using Remotely Sensed Reflectance and its Application to Coastal Environments", 2013

[3] Anatoly A. Gitelson, Giorgio Dall'Olmo, Wesley Moses, Donald C. Rundquist, Tadd Barrow, Thomas R. Fisher, Daniela Gurlin and John Holz, "A simple semi-analytical model for remote estimation of chlorophyll-a in turbid waters: Validation", Elsevier 2008

[4] `https://resources.marine.copernicus.eu/?option=com_csw&view=details&product_id=OCEANCOLOUR_MED_CHL_L3_NRT_OBSERVATIONS_009_040`