

2014 - 2020 Interreg V-A  
Italy - Croatia CBC Programme  
Call for proposal 2019 Strategic

## MARLESS (MARine Litter cross-border awareNess and innovation actions)

Priority Axis: Environment and cultural heritage; Specific objective: 3.3 - Improve the environmental quality conditions of the sea and coastal area by use of sustainable and innovative technologies and approaches

### 3.3.3 Maps of probability of accumulation or removal of waste

Activity 3.3

WP 3

Version: FINAL  
Distribution: PUBLIC  
Date: 30/06/2023

P

P  
A

## PROJECT MARLESS

<b>Work Package:</b>	<b>WP3 Monitoring optimization</b>
<b>Activity:</b>	<b>Activity 3.3 Marine litter hot spots identification</b>
<b>WP Leader:</b>	<b>PP2</b>
<b>Deliverable:</b>	<b>D3.3.3 Maps of probability of accumulation or removal of waste</b>

<b>Version:</b>	<b>Final</b>	<b>Date:</b>	<b>30/06/2023</b>
<b>Type:</b>	<b>Report</b>		
<b>Availability:</b>	<b>Public</b>		
<b>Responsible Partner:</b>	<b>PP2</b>		
<b>Involved Partner</b>	<b>ARPA FVG</b>		
<b>Editor:</b>	<b>ARPA FVG</b>		
<b>Contributors:</b>	<b>/</b>		

DISCLAIMER: PP2 reflects the project MARLESS views; the IT-HR Programme authorities are not liable for any use that may be made of the information contained therein.

# Maps of probability of accumulation or removal of waste

## Activity Deliverable 3.3.3

### WP3 ACTIVITY 3.3

Deliverable D3.3.3

Version: final

Confidential level: partnership

Data of release: 30/06/2023

1

1

# TABLE OF CONTENTS

INTRODUCTION .....	4
PART 1 SATELLITE IMAGES ANALYSIS .....	5
1. Marine litter detection with space-borne instruments .....	5
1.1 Properties of remote sensing instruments .....	6
1.2 How the data from satellite cameras are stored .....	9
1.3 How to process raster images .....	9
2. The ESA Copernicus Sentinel-2 Mission .....	10
2.1 Sentinel-2 spacecraft technical details .....	11
2.2 Contents of Sentinel-2 products .....	15
3. Investigating marine litter with Sentinel-2: literature review .....	20
4. Analysis of Sentinel-2 products .....	23
4.1 Pre-processing steps for Level-1C products. ....	24
4.2 Post-processing .....	33
4.2.1 Pixel spectra .....	33
4.2.2 Spectral signature .....	35
4.2.3 Band’s indexes .....	37
4.2.4 Classification of pixels by means of machine learning algorithms .....	40
4.2.5 Selection of an effective unsupervised classification methods. ....	43
5.1 Analysis of training areas .....	47
5.2 Search for marine litter hotspots on selected beaches .....	50
5.1 How we determine the fraction of potential marine litter pixels in a stretch of coast near the area of “Grado Pineta” in Grado (Italy) .....	51
5.2 Results of the search of marine litter with Sentinel-2 in two MARLESS beaches .....	55

PART 2: MAPS OF MARINE LITTER HOT SPOTS PROBABILITY .....	57
1. Numerical models results:.....	58
2. Satellite analysis results .....	61
3. Integration of the numerical model and the satellite analysis results . .....	62
4. Integration of these results with the in situ monitoring campaigns .....	65
CONCLUSIONS.....	68
BIBLIOGRAPHY .....	69

# INTRODUCTION

Deliverable 3.3.3 (D3.3.3) “*Maps of probability of accumulation or removal of waste*” concerns the realization of tools to identify areas of highest accumulation and removal with high probability to be a hotspot of marine litter. These maps should provide predictions about the presence of marine litter in a coastal environment of the Adriatic Sea, potentially being a support tool to guide future on field monitoring campaigns.

The MARLESS project schedule the realization of these maps for the *Work-Package 3 (WP3): “Monitoring optimization”* inside the *Act 3.3: Marine litter hotspots identification*. Project partner Friuli Venezia Giulia Autonomous Region is the main responsible for the realization of this directive. The production of these maps is based on an innovative approach which aim to combine together the outputs of the dispersion model developed in the directive *D3.3.2: “Model of dispersion of floating waste”* with the results of satellite images analysis of the ESA satellite mission Sentinel-2 which is developed as a preliminary step in D3.3.3.

# PART 1

## SATELLITE IMAGES ANALYSIS

In this section we provide a general overview of properties and key points about the monitoring of Earth's surface with remote sensing instruments, in particular describing the technical properties of the ESA Copernicus space mission Sentinel-2.

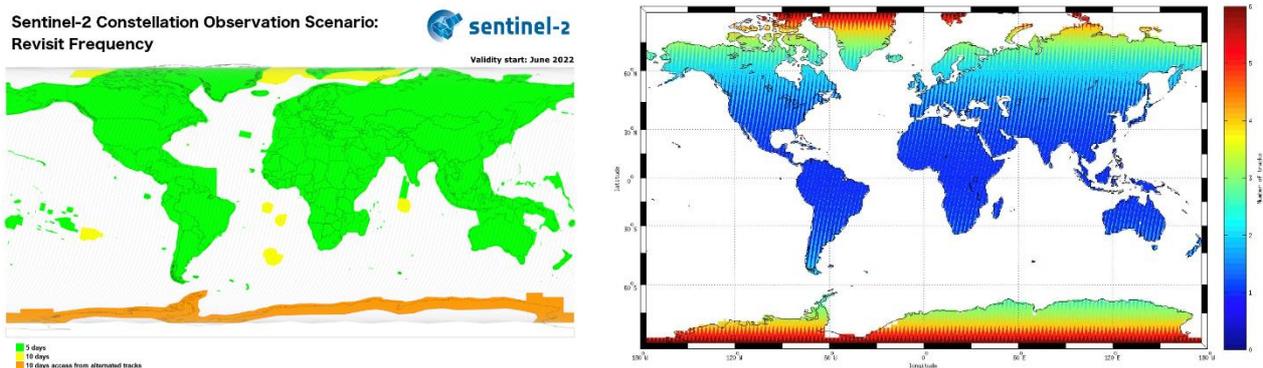
### 1. Marine litter detection with space-borne instruments

Marine litter detection has become a topic theme nowadays in the agenda of several Nations and Institutions worldwide due to a widely awareness about the consequences of human activities on the maritime environment. A wide range of sources produces the waste that pollute our seas, varying from urban/industrial to maritime activities. The debris they produce are transported by riverine currents into seawaters and then dispersed by sea currents and wind streams, letting the garbage to pollute environments at large distances. Marine litter is not simply scattered over the sea or along coastlines, it can form filaments or patches in seawaters or it could be aggregated when deposited on beaches, in both cases it is usually found mixed with organic materials (seaweeds, algae, woods etc.). The composition and the spatial distribution of the marine litter changes over time: marine litter is a dynamical problem. Visual ship surveys or land-based monitoring campaigns are useful to record in detail the presence of marine litter and the kind of the debris it is composed by, but these activities take a static picture of the reality only once a time and are too expensive to be conducted on a large scale with higher frequency. As appointed by several studies in the recent years, space-born sensing

instruments can carry relevant improvements to improve marine litter monitoring, thanks to a coverage of Earth’s surface on large spatial areas and a more frequency passage of the instruments over time [1][2][3].

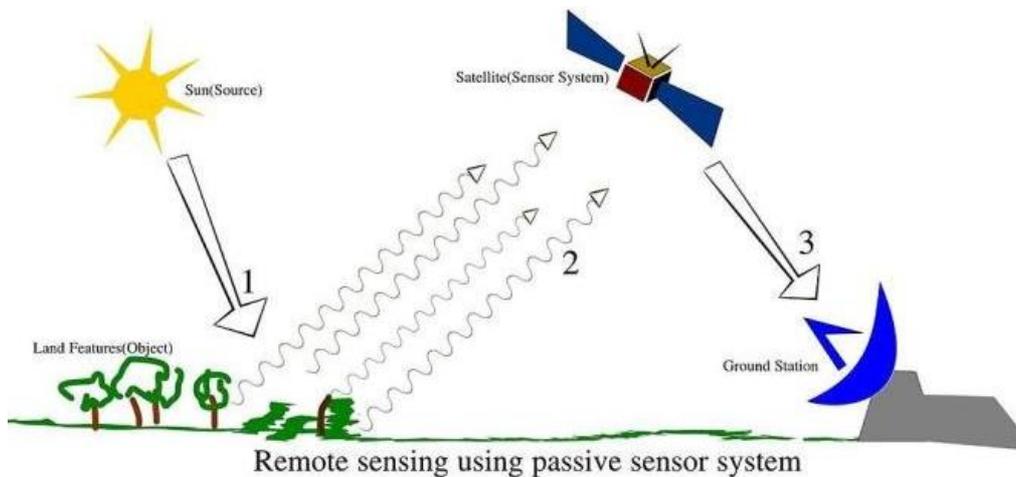
### 1.1 Properties of remote sensing instruments

Satellites dedicated to monitor Earth’s environment move in polar orbits that let them to achieve a *revisit time*, i.e., the time interval from two sequential passages over the same point above a location on Earth, of about a week. This property is usually referred as the *temporal resolution* of a satellite. For example, the Sentinel-2 mission achieve of about 3-5 days (see **Fig. 1.1**). Moreover, the orbit altitudes, about 700 km, allows the instruments equipped on a spacecraft to achieve a *swath width*, i.e., the wide of the imaged area on the ground, of hundreds kilometres.



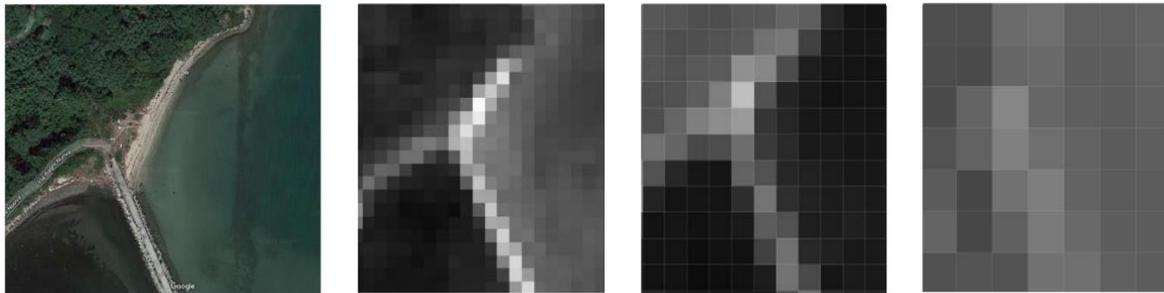
**Figure 1.1.** Left: revisit frequency achieved by the twin satellite Sentinel-2A and Sentinel-2B. Right: due to the large swath width, several orbits areas overlap, increasing the coverage frequency of the satellite especially at higher latitudes [4].

We can divide remote sensing instruments into two classes: *active and passive instruments*. The first send to Earth’s surface a signal (typically radio) and measure the echo, the latter simply collect reflected sunlight over objects on the ground. Active sensors pose more difficult challenges to reconstruct the signal ground-truth, thus in this work we adopt and describe only passive instruments properties and imagine analysis techniques (see **Figure 1.2**).

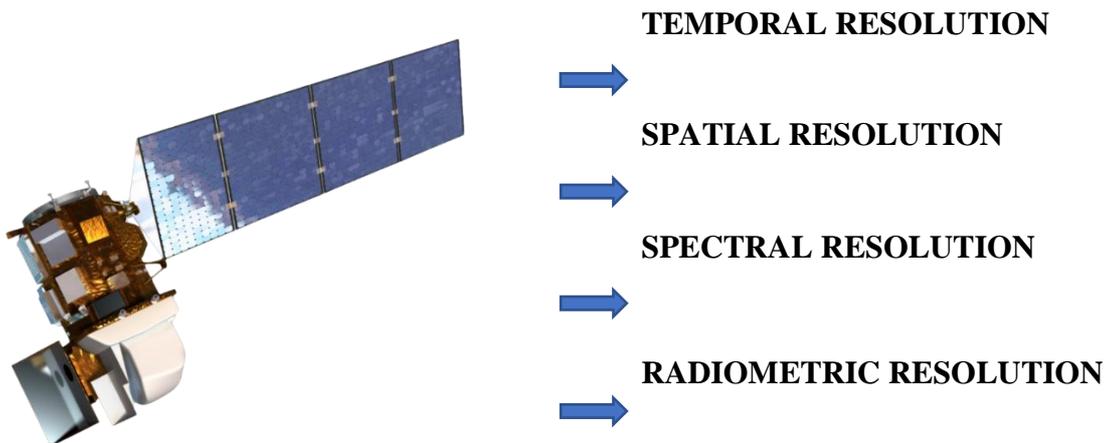


**Figure 1.2.** Schematic view of the working principle of passive remote sensing instruments. The data collected by the instruments aboard the satellite are periodically downloaded to the *ground stations*, i.e., a series of surface-based facilities that communicate with all the satellite of a constellation and can perform the first processing steps on the raw data. Image credits: DOI:[10.13140/RG.2.2.22741.12007](https://doi.org/10.13140/RG.2.2.22741.12007)

Remote sensing instruments cannot detect objects on the ground with high detail. Optical properties of satellite camera typically let to achieve a *spatial resolution* that vary from about 100 m down to 10 m. Spatial resolution specifies how much ground corresponds to the side of a pixel, as the next image shows (**figure 1.3**). Satellite detectors supply the spatial resolution limits by having the ability to see in a wide spectral range that typically goes from the visible to the infrared part of the electromagnetic spectrum. The collected light is not recorded as a continuous signal but is spitted into a discrete set of channels that could range from a few dozen to a hundred, defining the *spectral resolution* of the satellite. There are two classes of instruments: if the bands are about a decade, we have a *multispectral instrument*, if the bands are about hundreds we have a *hyperspectral instrument*. The last parameter that characterizes a remote sensing instrument is its *radiometric resolution* that indicates the ability of the sensors to distinguish two different intensity of the measured light. Usually expressed as a bit number, radiometric resolution can range from 8 to 16 bits. **Figure 1.4** resumes the main properties of a remote sensing mission.



**Figure 1.3.** The leftmost picture shows a part of the Island of Bagni in Monfalcone (Italy) as seen with Google Maps. The next three pictures display the same area as seen by three different monochromatic bands of Sentinel-2, having pixels with an increasing spatial resolution of 10, 20 and 60 m. The increasing of the pixel dimensions implies a loss of the details of the object on the ground smaller than the spatial resolution.



**Figure 1.4.** The main four properties of a remote sensing mission. The picture above is the rendering view of the NASA Landsat-8 mission (credits: <https://landsat.gsfc.nasa.gov/article/landsat-data-continuity-mission/>).

## 1.2 How the data from satellite cameras are stored

Remote sensing missions store data they collect into one or more raster images with associated auxiliary data and metadata. A raster is a matrix of pixels which contains the values of the measured signals in a suitable format to save memory and preserve the loss of information, i.e., the values are stored as *Digital Numbers (DN)*. Space missions don't provide the User the raw data collected by spacecraft's instruments, but several post-processing elaborations are performed over them before they are ready for Users analysis. Metadata and auxiliary data are associated to rasters to provide information about how values have been calculated. Geo-referencing, i.e., associating a Geo Information System (GIS), and orthorectification, i.e., parsing data with a Digital Elevation Model (DEM), are two fundamental steps to elaborate raw data. The metadata attached to a raster provides information about these processes, along other properties such as the date-time of the acquisition of the image, the spacecraft orbit status, recording instrument and electronics technical properties, data calibrations, etc. Thus, when we open a raster image, the raw pixel values have been already (slightly) altered by aforementioned operations and dedicated software are needed to handle properly the metadata. In our work we use mainly QGIS (<https://www.qgis.org/en/site/>) and specific Python libraries to read such geospatial data. The collection of raster, metadata and other auxiliary data are usually referred as *the products*. Each raster contains data from a single instrument band.

## 1.3 How to process raster images

When we open a raster, we can visualize and manipulate data in several ways, according to our purposes. Similar to the most common steps for the manipulation of images, we can define a *pre-processing phase* and a *post-processing phase*. We provide now a brief survey of the main actions to take in the two phases.

### - Pre-processing phase

This phase includes steps that vary depending on the level of processing of the products. Missions tend to provide to Users products already georeferenced and ortho-rectified, but they

make available also lower level products. The orthorectification step is quite complicated, thus typically the User don't perform it. The main pre-processing steps are (listed not in order):

- i) Land/terrain/cloud Masking
- ii) Atmospheric correction
- iii) Clipping
- iv) Resampling

#### - **Post-processing phase**

Once the pre-processing phase is complete, the User has setup the data to analyse them. The reflectance value of each pixel of every band can be read and put into a reflectance Vs wavelength plot to make *spectral plots*, or two or more (clipped and masked) raster can be algebraically combined making *environmental indexes*, which are displayed as heatmaps. Classification technique and other filters can also be applied to interpret data to recover their ground-truth. The User can use all these techniques together or independently. Here is the resume of such techniques:

- i) Making of spectral plots
- ii) Making of heatmaps from environmental indexes
- iii) Classification of pixels via machine learning techniques
- iv) Applying filters or masks

We will go back to details of how we have processed Sentinel-2 rasters, but first we have to describe the Sentinel-2 mission itself.

## 2. The ESA Copernicus Sentinel-2 Mission

**Table 2.1** reports a list of the main spatial missions with their properties. The most suitable space mission providing open data access to fulfil our objectives is the ESA Copernicus Sentinel-2

mission. Spectral data are freely downloadable from the mission website, with each new image ready to be downloaded in less than 24 hours after its acquisition. Revisit time and spatial resolution are a valid compromise between other open data missions and the number of bands are enough to have a good set of spectral data points to distinguish some prominent signatures in the spectral plots. Sentinel-2 has been already tested with some success for its detection capability of marine litter, as reported in literature by several studies (see the section “*Searching marine litter with Sentinel-2: literature review*”). Before discussing forward about marine litter investigation, we are going now to briefly summarize the technical details of the Sentinel-2 spacecraft and its products.

Mission Name	Spacecraft	Best Spatial Resolution	Spectral Class	Number of Bands	Revisit time
<b>Landsat</b> <sup>[29]</sup>	Landsat 8	30 m	Multispectral	8	16 days
<b>PlanetScope</b> <sup>[30]</sup>	130+ Cubesat	5 m	Multispectral	8	Daily
<b>Sentinel</b> <sup>[31]</sup>	Sentinel 2A-2B	10 m	Multispectral	13	3 – 5 days
<b>WorldView3</b> <sup>[32]</sup> (*)	/	< 10 m	Multispectral	17	1 – 4.5 days
<b>PRISMA</b> <sup>[33]</sup> (*)	/	30 m + 5 m pan camera	Hyperspectral	~ 250	7 – 14 days

**Table 2.1** List of main remote sensing missions dedicated to monitor Earth environment to give the reader a quick comparison to spacecraft and cameras properties discussed in the text. For each mission, the best spatial resolution available is reported, therefore the **number of bands** column indicate how many bands there are with that spatial resolution. Symbol (\*) indicates mission with no open data access policy.

## 2.1 Sentinel-2 spacecraft technical details

ESA Sentinel missions are a family of satellite that are part of the Copernicus programme, the European Union project of monitoring Earth’s environments through satellite and in-situ data [5][6]. Sentinel-2 is a mission dedicated to land, marine and atmospheric monitoring and consists of a planned constellation of four identical satellites. Up to now, only half of the Sentinel-2 constellation is operative, with the other members scheduled to launch in the further years.

The two satellites working nowadays are Sentinel-2A and Sentinel-2B. They orbit Earth at 700 km of altitude in the same polar orbit shifted by 180°. They are passive instruments that collect reflected sunlight via their optical camera named MultiSpectral Instrument (MSI) [7]. **Figure 2.1** shows the external appearance of a Sentinel-2 spacecraft and the layout of the MSI.



**Figure 2.1.** *Left panel:* schematic view of a Sentinel 2 satellite. The MSI payload is attached on the left of the spacecraft body. *Right panel:* inside view of the MSI instrument. Further details are available through the mission website.

MSI is a multispectral camera with wide swath covering wavelengths from infrared to visible and is designed specifically to satisfy the land-monitoring applications established for the Sentinel-2 mission. The MSI camera sensors collect sun radiation reflected by Earth atmosphere or land/sea surfaces in 13 channels with different resolutions:

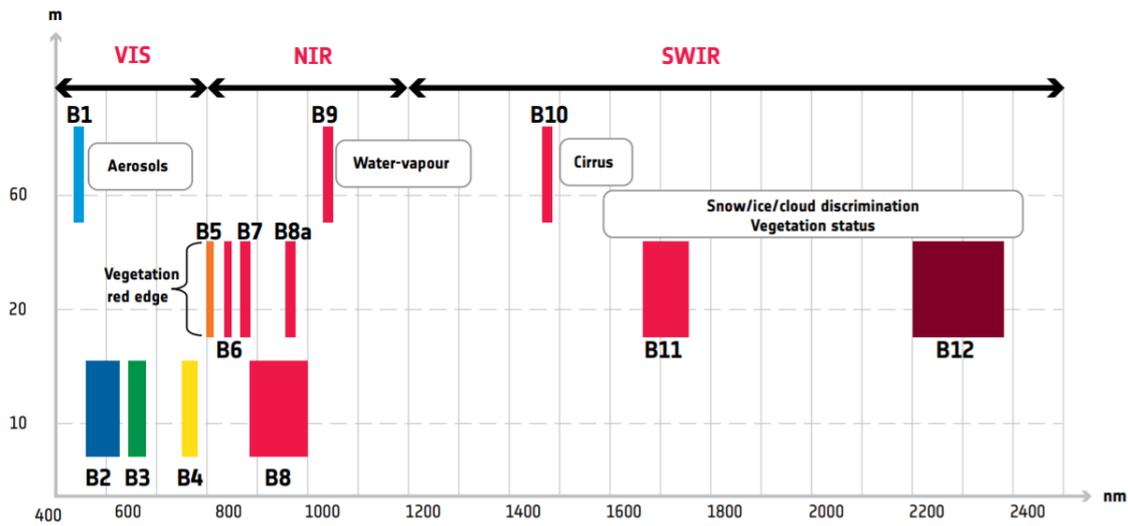
- Three channels centred at wavelengths in the visible (Red, Green, Blue) and one in the near-infrared collect light at 10 meter resolution.
- Six channels operate at 20-meter resolution, of which one works in the Red-Edge band, three in the NIR band and two in the SWIR band.
- The last three channels have a 60-meter resolution where the channel at a shorter wavelength operates in the Ultra Blue band, the other two in the SWIR bands.

**Table 2.2 and figure 2.2** reports full specifications of the wavelength of the bands and their main usage for the Sentinel-2 mission.

The pair of Sentinel-2 satellites operate continuously, taking images in a push-broom scanning mode and observing Earth surface with a 290 km swath width and achieving a revisit time that is about 5 days and can be decreased thanks swath overlapping at middle latitudes. The revisit time over Italy has been tested by counting all the tiles available in the two years 2020-2021 depicting the coastal area of the Friuli Venezia Giulia Autonomous Region. Averaging over all the images taken monthly for the two years, the calculated revisit time is about 2.5 days. **Table 2.3** shows the frequency of the images taken by Sentinel-2A/B for each month.

Spatial Resolution (m)	Band Number	S2A		S2B	
		Central Wavelength (nm)	Bandwidth (nm)	Central Wavelength (nm)	Bandwidth (nm)
10	2	492.4	66	492.1	66
	3	559.8	36	559.0	36
	4	664.6	31	664.9	31
	8	832.8	106	832.9	106
20	5	704.1	15	703.8	16
	6	740.5	15	739.1	15
	7	782.8	20	779.7	20
	8a	864.7	21	864.0	22
	11	1613.7	91	1610.4	94
	12	2202.4	175	2185.7	185
60	1	442.7	21	442.2	21
	9	945.1	20	943.2	21
	10	1373.5	31	1376.9	30

**Table 2.2.** MSI list of spectral bands grouped for their spatial resolution for each of the satellite of the Sentinel-2 mission. There are 13 bands numbered 1 to 12 due to the special name assigned to “Band 8a”, which cover with lower spatial resolution part of the same spectra seen by Band 8 [7].



**Figure 2.2.** The chart shows the position and the amplitude at FWHM of all Sentinel 2 spectral bands over the electromagnetic spectrum, grouped for their spatial resolutions. The picture show information of the bands main monitoring land purposes. Bands B1, B9, B10, regarding atmospheric analysis, are not used in the proceeding of our study [8].

2020					2021				
Month	S2A Images	S2B Images	Total Monthly Images	Av. Img.	Month	S2A Images	S2B Images	Total Monthly Images	Av. Img.
JAN	6	7	13	2,4	JAN	6	6	12	2,6
FEB	6	5	11	2,5	FEB	5	6	11	2,5
MAR	6	7	13	2,4	MAR	7	6	13	2,4
APR	6	6	12	2,5	APR	6	6	12	2,5
MAY	6	6	12	2,6	MAY	5	6	11	2,8
JUN	6	6	12	2,5	JUN	6	6	12	2,5
JUL	6	6	12	2,6	JUL	6	6	12	2,6
AUG	7	6	13	2,4	AUG	7	6	13	2,4
SEP	6	6	12	2,5	SEP	6	6	12	2,5
OCT	5	6	11	2,8	OCT	6	6	12	2,6
NOV	6	5	11	2,7	NOV	5	6	11	2,7
DEC	6	6	12	2,6	DEC	6	7	13	2,4

**Table 2.3.** Monthly images available for each spacecraft. The column “Av. Img.” reports the average number of images available each month. The calculation is done by counting how many images depict the city of Monfalcone (Italy) during a month and dividing the result by the total amount of images showing the same area in a year. It can be notice that the number of passages of a spacecraft is not constant over the time and also changes between the two spacecraft, but the total number of passages is constant ( $N_{pass} = 144$ ).

## 2.2 Contents of Sentinel-2 products

### ➤ Where to find and how to download Sentinel-2 products

MSI images are transmitted to the Ground Segment of the Sentinel-2 mission, which assigns a geometric model and radiometric corrects the raw data. The outputs of this process are then compressed, radiometric corrected, orthorectified and made disposal as *Level 1C* products, which are downloadable at the mission webpage: <https://scihub.copernicus.eu/dhus/> (hereafter referred as *SciHub portal*) as zipped files (**figure 2.3** shows the main page of the website).

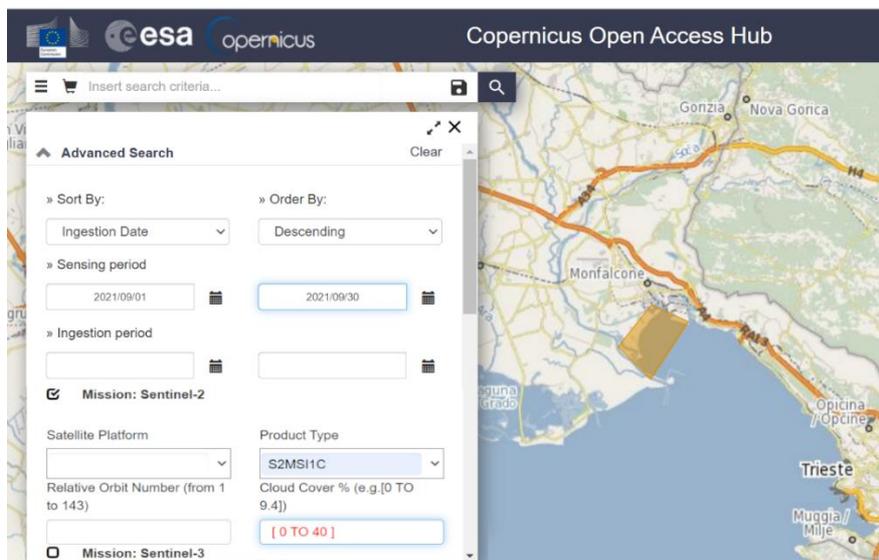


**Figure 2.3.** The SciHub portal main page shows an interactive map of the whole Earth surface. On the top left there is the search bar with a dropdown menu. The access to the products is subject to a free sign up to the website.

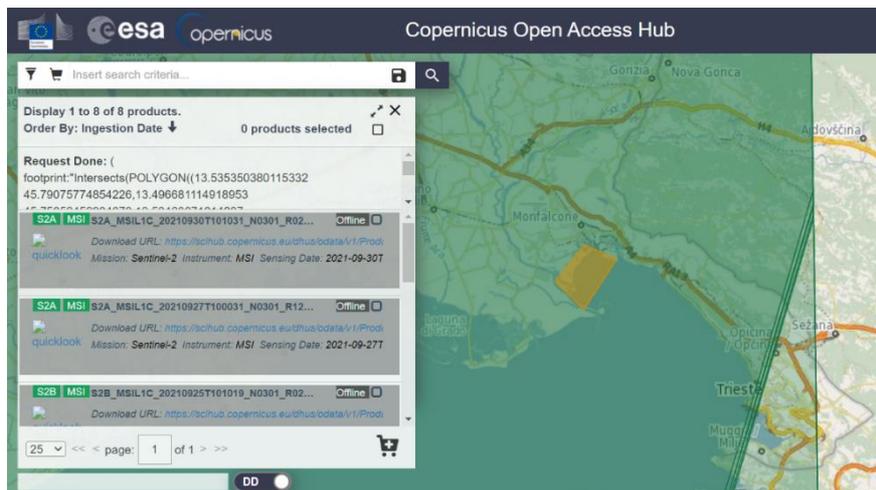
Searching and downloading a Sentinel-2 product is quite straightforward: first, we have to draw on the map displayed in **figure 2.4** a polygon containing the area we want to study. Then we have to select from a cascade menu the *sensing period*, i.e., a time interval that have to comprise at least one

orbital passage of a Sentinel-2 spacecraft. In the end we have to select from the same menu the product type and the cloud filter intensity and start the search. These two steps are shown in **figure 2.5**.

If there are one or more products available, they will be displayed in a list where we can choose to immediate download them, look a quick RGB preview of the whole area imaged in the product or save in a virtual cart for late downloading (see **figure 2.5** for the query results).



**Figure 2.4**



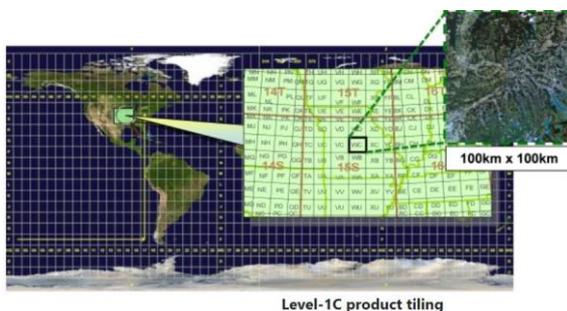
**Figure 2.5**

➤ **Sentinel-2 products content**

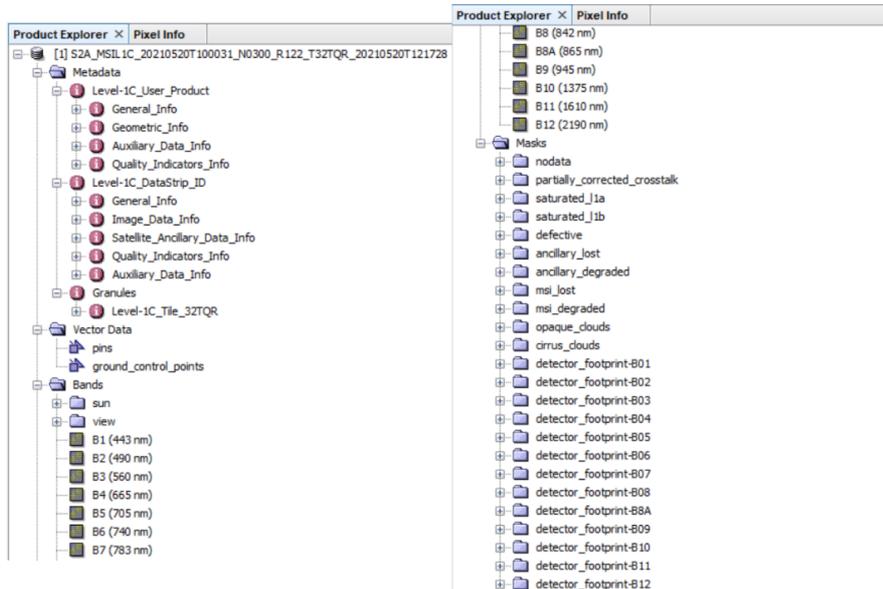
*Level 1C* products most important content is the set of 13 raster images, one for each Sentinel-2 band, and the *true color image (TCI)*, that is a RGB preview of the product imaged area. The extension of such area depends on how Earth’s surface was imaged during the spacecraft orbital passage and how it fits into the tiling grid of the Sentinel-2 mission (**figure 2.6**). Each tile is a squared area of about 110 km<sup>2</sup> (**figure 2.7**). Other Level 1C product contents are metadata and vector data. The former contains important information that helps to contextualize raster data (**figure 2.8**).



**Figure 2.6.** *Left picture:* jpeg preview obtained by conversion of a TCI raster of a Sentinel-2C tile over the Gulf of Trieste. In the picture is visible the entire coastline of Friuli Venezia Giulia Region up to the eastern part of the Venice Lagoon (left) and of the Istrian Peninsula (right). The black slice on the left top corner contains no data. *Right picture:* Sentinel-2B TCI of the same area, the image cuts are slightly different due to not exactly overlapped trajectories of the orbits of the two spacecraft.



**Figure 2.** How the Earth’s surface is sampled by the tiling grid. Due to the overlapping of adjacent orbits, there could be two consecutive tiles depicting the same area.



**Figure 2.8** Tree structure of a generic Level 1C Sentinel-2 product when opened with the ESA SNAP software. More information about each files is available on the online mission documentation [9].

### ➤ Rasters

A raster is simply a collection of values on a two-dimensional grid. The physical meaning of raster data depends on the context. In remote sensing field, the values we found in a raster are the radiance measured by the MSI pixel camera and distributed over a pixelated grid. These values are stored as DN converted as integers to memory save reasons. The radiance values are converted to the reflectance ones thanks to the formula:

$$\text{radiance} = \text{reflectance} * \cos(\text{radians}(\text{SunZAng})) * \text{solarIr} * U/\pi$$

Where the value of SunZAng (the angle of the Sun at the zenith), the values of U and solarIr (the solar irradiance) for the considered band can be found in the metadata. Moreover, the metadata contains the value of the scaling factor to restore reflectance values as floating point numbers.

While the latitude and longitude edges of a raster are fixed by the tile extension, the total number of pixels in a raster changes according to the spatial resolution of the given band. For example, a single

20 m pixel band B6 can contain four 10 m pixels of band B2. In table 2.4 we report the size of a raster in function of its spatial resolution.

Bands	Spatial resolution	Tile size (km)	Tile size (pixels)
B1, B9, B10	60 m	110 km x 110 km	1830 x 1830
B5, B6, B7, B8, B8A, B11, B12	20 m	110 km x 110 km	5490 x 5490
B2, B3, B4, B8	10 m	110 km x 110 km	10980 x 10980

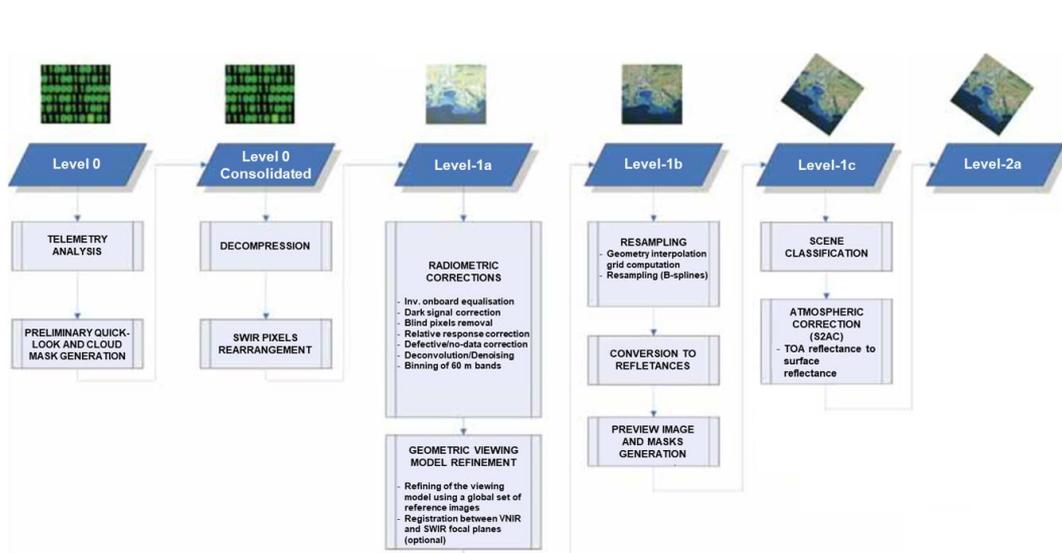
**Table 2.4**

To fully understand satellite raster contents we need to know if the data are correctly georeferenced and what is the world geodetic system (WGS) associated to the pixel grid (for more information about WGS, see [https://en.wikipedia.org/wiki/World\\_Geodetic\\_System](https://en.wikipedia.org/wiki/World_Geodetic_System)). Metadata contains such information and every time we act on a raster we have to care about if the WGS is correct.

➤ **Sentinel 1C vs Sentinel 2A**

A final word Level-1C: these products are also labelled *Top of Atmosphere (TOA)* files, meaning that the reflectance values have not been corrected due to the atmospheric disturbance. In our work, we use Level 1C products and evaluate by ourselves the best way to apply an atmospheric correction to our data. Level-2A are not used because the automatic atmospheric correction provided by the Sen2Cor ESA software is known to not work well with coastal environments.

**Figure 2.9** shows the chain of data processing involved to make a Level-1C product (more details are provided on Sentinel-2 mission website).



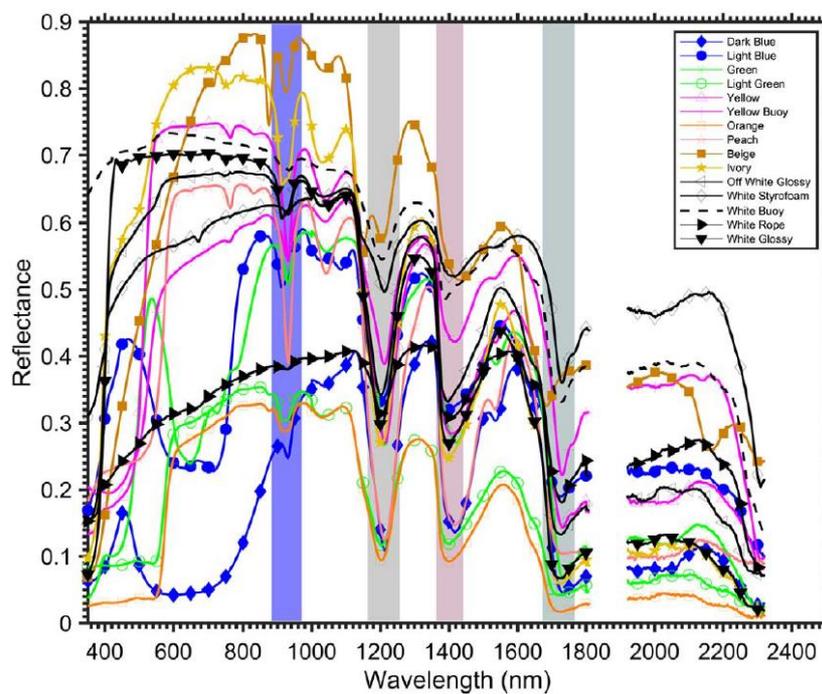
**Figure 2.9.** Level-0, Level-1A and Level-1B products are internal products not available to Users.

### 3. Investigating marine litter with Sentinel-2: literature review

Marine litter investigation using aerial or space-born instruments is a new field of research, hence the literature is still debating the most reliable and replicable process to analyse Sentinel-2 images finding pixels containing marine litter.

Difficulties arises not only by the technical limits of the instrumentation, but also by the lack of validated a-priori knowledge about the nature of the objects that produce the measured signals. We already know that the “marine litter” word indicates object of different nature. Only looking at their main component, i.e., plastics, we have to distinguish between a plethora of types, colours and dimension of plastics. Moreover, marine litter components form heterogeneous patches within them or mixed with other biological material. These patches could be distributed randomly in the environment and have various extension. Therefore, there are no prefixed spectral signatures or well-defined bounding areas to search. The identification of marine litter hotspots in coastal or marine environments is nevertheless made possible by adopting some strategy and approximations.

Due to the remarkable problem of plastic pollution of marine environments, the search of marine litter with remote sensing instruments is mostly dedicated to find plastic spectral signatures. This field of research is new and still open to debate. The studies focus on testing limits in the detectability of plastic signals and in establish replicable methodologies. One important conclusion that shares such studies is that Sentinel-2 is a space mission that can help finding plastic debris thanks its infrared spectra bands that matches the spectral range where plastic spectra measured in laboratory exhibits important absorption lines [10] (**figure 3.1**).



**Figure 3.1.** The measured spectra of plastic harvested in USA beaches. Shaded areas indicate the absorption tips of the plastics.

To solve the problem of the ground-truth of the pixels measured reflectance, some studies have focused their efforts in measuring signals emitted by handmade plastic rafts afloat near coastal areas (these studies have been conducted by the University of the Aegean in two campaigns named ‘Plastic Litter Project’ for the years 2018-2019 [11][12] and by a team of researchers in Cyprus [13]) (**figure 3.2**).



**Figure 3.2.** Aerial view of the three plastic rafts deployed at Tsamakia beach. [11]

Other studies have instead focused their efforts searching plastic debris under different environmental conditions around the globe, and only a study try to extend the search of marine litter over the time [14][15][16][17].

All the works listed before show that plastic marine litter is detectable using Sentinel-2, but the environmental conditions and the in-situ data of the pixels are always well known or there is some knowledge on the ground-truth. No open scenario are studied, and are preferred the cases of marine litter floating over seawater. These researches have nevertheless appointed some important facts that we have to take into account into the proceedings of our work:

- i) The importance of the NIR band in highlighting plastic is proved, but the signal in this band is often difficult to distinguish from the spectral response from other objects and in particular from seawater when dealing with transparent or blue plastics.
- ii) Atmospheric correction is needed to compare images at different date times, but due to the action of the correction algorithms in the blue band, it may happen that the signal of plastic can be significantly altered.
- iii) Patches that occupy less than 50% of a raster pixel may not be detectable. [18].

- iv) Machine learning algorithms are useful to catalogue automatically large area, but the results of unsupervised or supervised method need to be fine-tuned by more in-situ or synthetic data spectra.

## 4. Analysis of Sentinel-2 products

As appointed before, up to nowadays there is no validate and trustable procedure already developed in literature to find marine litter. Several efforts focuses on floating marine litter with little attention to search beached marine litter of any kind or hotspots. We are not capable to study of the composition of such debris, but we are concerned in establishing a procedure to distinguish between pixels ‘clean’ of marine litter form pixels ‘contaminated’. This requires the identification two main class of pixels: those containing only the spectral signatures of natural features of the ground (water, sand, woods, scrublands ...) or artificial ones (backwaters, wharves ...) and those covered by natural or artificial elements that are beached there. The latter objects could occupy a surface heterogeneously, thus a pixel could exhibit a spectra that is not comparable with the spectra of a natural/artificial feature, making this a potential marine litter pixels. Moreover, even if a pixels is fully occupy by marine litter, the exact shape of its spectral signature is unknown The analysis we propose in this work is therefore driven by the following methodology: first, we focus on the creation of a database of spectral signatures of features of the ground, then we will start the classification of the pixels excluding those with known spectral signatures, distinguishing between two main categories, ‘*reference pixels spectra (RPS)*’ and ‘*potential litter pixels spectra (PLPS)*’. To build these classes, we have to pass through several step, letting us to properly process Sentinel-2 Level-1C and get pixels spectra plots. This methodology will require the application some pre-processing techniques and in the post-processing phase the application of machine learning algorithms.

#### 4.1 Pre-processing steps for Level-1C products.

The pre-processing phase of a satellite image usually includes the assignation of a georeferenced coordinate system. As appointed before, Level-1C are already provided in the EPSG4326 projection system, thus we do not need to assign an own. The other steps we have perform to setup a Sentinel-2 product to further analysis are:

- 1) cloud filtering;
- 2) product resampling;
- 3) bands atmospheric correction;
- 4) clipping.

##### **1) Cloud filtering:**

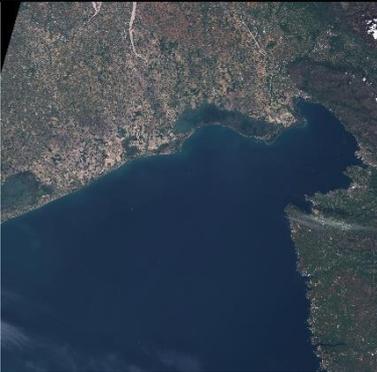
Cloud filtering consists in discarding products with a cloud coverage occluding the area interested for our analysis. We perform cloud filtering both automatically and manually.

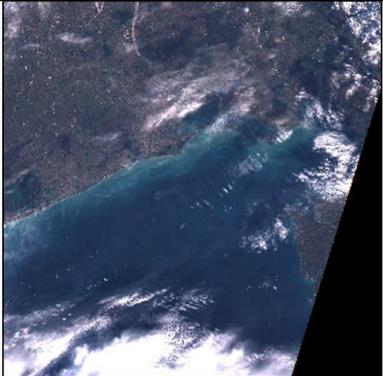
When done automatically, the filtering is done simply by defining the cloudiness threshold in the product searching menu in the SciHub portal. Since a large extension of clouds over a tile very likely implies a coverage of our area, we have established that a value of cloud coverage equal or above the threshold of 60% indicates the partial or total darkness of our area of study. The discussion of how we have selected this value is reported in the next page. This parameter is passed to the SciHub portal every time we search a new product.

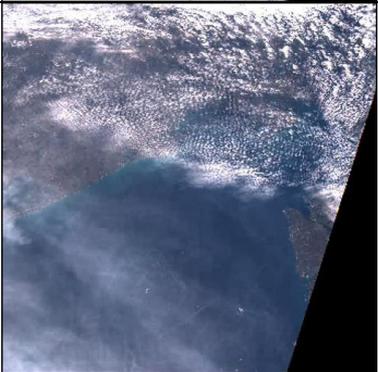
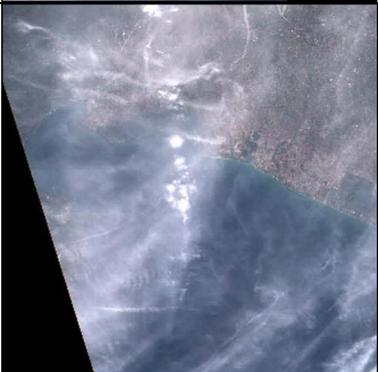
Nevertheless, clouds could occlude an area of study also if they occupy a small fraction of the whole tile. To check if our area of study is covered by such clouds, we have to visual look at the TCI image, that is a RGB preview of the product tile. Completing this second “cloudiness” control ends this pre-processing step.

Hereafter in **figure 4.1** we report a summary of the visual analysis of all the products available during the year 2021 that allow us to identify the 60% automatic threshold parameter.

**Figure 4.1:** Sample of TCI images of products affected by different cloudiness listed in increasing order. The images are referred to all the tiles containing the city of Monfalcone. We lists two images around the 60% value to show that also those that visually seems to be cloud-free can have a high cloud coverage. This is surely due to the fact that the bands B1, B9, B10, used to evaluate cloudiness, are not all in the visible light range.

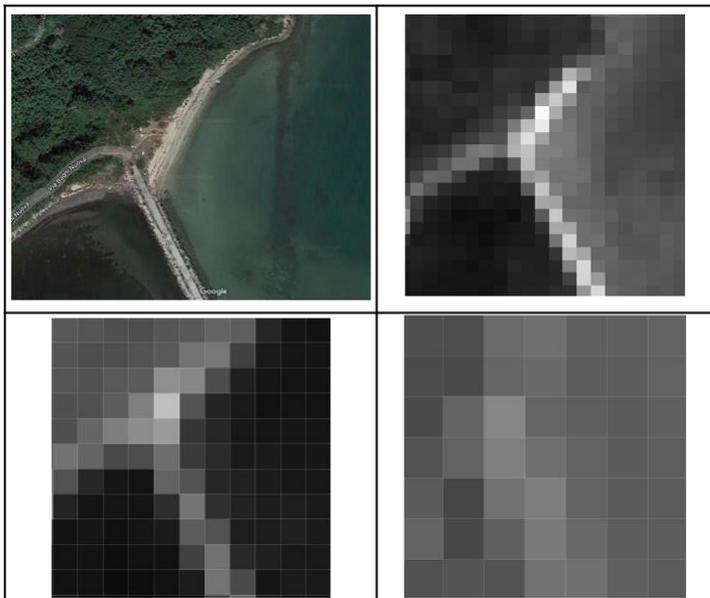
DATA	IMAGE PREVIEW	CLOUD COV. %
12 - 09 - 2021		0 %
2021-08-26		12 %
2021-11-20		20 %

29 - 11 - 2021		31 %
31 - 12 - 2021		46 %
2021-09-07		59 %
2021-12-14		60 %

2021-12-19				73 %
2021-12-04				87 %
2021-10-25				99 %

## 2) Product resampling:

We have already discussed about the different spatial resolutions of the Sentinel-2 bands (Figure PS1), in this section we explain how we manipulate this parameter to obtain a common spatial resolution. We do not use the 60 m bands since they are dedicated to atmospheric measurements. Hence, we decide to downscale all the 20 m bands to 10 m. To avoid altering significantly the original signal content of each 20 m pixel, we use the simple *nearest neighbourhood* resampling algorithm to perform this task. This procedure operates as follows: the coordinate of each pixel in the output map is determined. Then, for each output pixel, the pixel value of the input pixel closest to this coordinate is used as the output value. The resampling process produces as output a raster having all the 9 bands at 10 m per pixel spatial resolution.



**Figure 4.2.**

Aerial views of a portion of the Panzano Bagni beach in the city of Monfalcone (IT).

From top left clockwise: Google Maps screenshot, band 4 at 10 m spatial resolution, band 6 at 120 m spatial resolution, band 1 at 10 m spatial resolution.

### 3) Atmospheric correction:

The atmospheric correction of a product is needed to have comparable results between different days, since this operation corrects disturbances affecting light collected by the MSI camera sensors. Typical disturbance of reflected light moving from the ground back to the spacecraft are scattering or the absorption of light rays by atmospheric particles, whose intensity vary day by day due to the normal change environmental conditions of the atmosphere. Another important effect that such algorithms can correct is the *sun glint*, i.e., the phenomenon for which under certain angle of view the camera lens sees the sun mirrored into water bodies, measuring a reflectance value higher than the real one (**figure 4.3**).



**Figure 4.3.** Comparison between two Sentinel-2 Level 1C tiles displaying the same maritime area with and without sun glint effect disturbance.

Several software are available to perform atmospheric correction and each one has its own characteristics and field of application, moreover literature doesn't point toward a unique algorithm capable to handle marine litter detection in the proper way. From all the software we have at disposal, we have selected the software ACOLITE developed at the Royal Belgian Institute of Natural Sciences (<https://odnature.naturalsciences.be/remsem/software-and-data/acolite>) because this software has

been developed specifically to correct Sentinel-2 products when the studying coastal environments, that is indeed the environment of our work.

Moreover, ACOLITE manual and technical details are open access and an online forum webpage helps the User with responsive and clear answer to any problems encountered during the installation, configuration or application steps [19][20][21].

ACOLITE takes as input any Sentinel-2 product and then processes producing as outputs resampled rasters for each band plus a RGB preview of the corrected area. There are several input parameters to handle, but a few set of them is sufficient to make a good correction of Sentinel-2 products, an example of a basic setup is provided by ACOLITE authors in the documentation. As general warning, the intensity of the rescaling of the reflectance values applied to the raster bands depends on the size of the area we select for the correction. It's better to choice a large squared area, i.e., with an extension of at least 10 km side and comprising both inland than water surface, and check if the output spectral band pixels exhibits negative reflectance values. If this happens to a large set of points then the software is not working well and we have to change the input parameters or consult the user's forum.

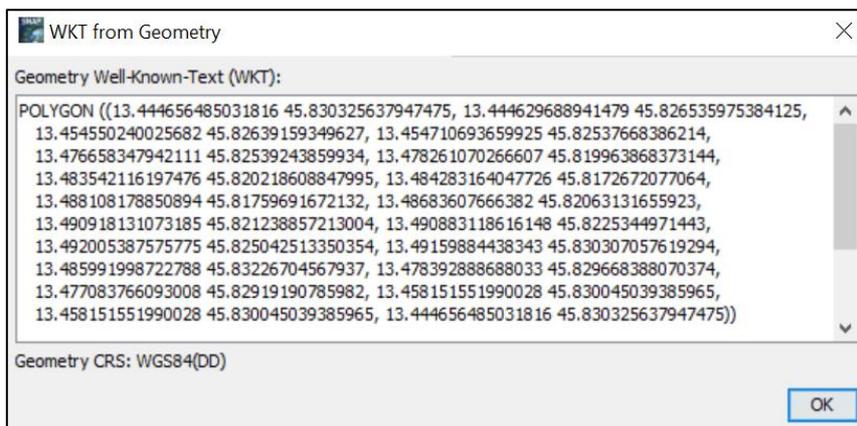
#### 4) Clipping

Clipping an image means the operation of sub setting the raster using as input a text that describes a polygon on Earth surface trough the geographical coordinates of its vertices. *Well Known Text (WKT)* and *GeoJson* are the common file standards we have used in our work. Such files can contain the description of a single or multiple polygons. Typically, the area to cut is drawn manually by the User using a suitable coordinate reference system. We used the ESA SNAP Toolbox and QGIS to open Sentinel-2 products or satellite maps and draw with the desired accuracy the polygons containing the area of study. In the following images (**figure 4.4 and 4.5**), we shows an example extraction of a WKT string. The first picture shows a portion of a SNAP RGB preview of a Sentinel-2 products over the low part of the Friuli Venezia Giulia autonomous region. The polygonal line follow the borders of the Trieste Airport perimeter. SNAP allow us to extract the polygon vertices

coordinates simply by right-clicking over the violet area and that selecting “WKT from Geometry”. At this point appear a window showing the text defining the polygon geometry. It’s very important to check the reference coordinate system (in this example, the WGS84) to check if the latitude/longitude pairs are compatible with the reference coordinate system of the raster bands. If the two system are different, we have to convert one of them done using online tools or the QGIS utilities.



**Figure 4.4**

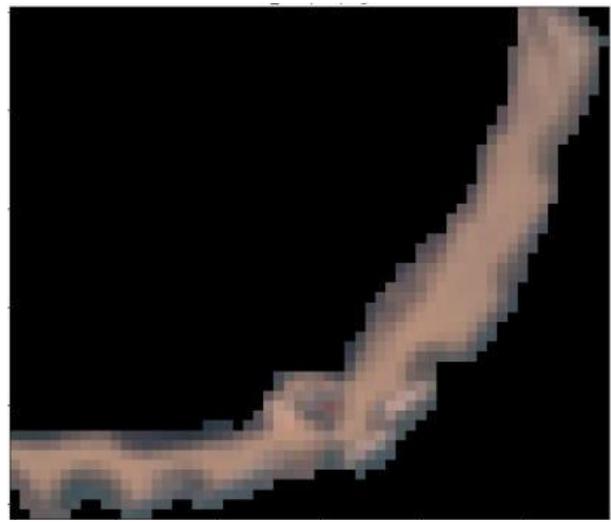


**Figure 4.5**

The final output of all these processes is then a new raster having a size smaller than the original tile and with 10 bands at a spatial resolution of 10 m. Here below we show an example of a clipped area resulting from the clipping of a Sentinel-2 Level-1C tile on the 12<sup>th</sup> of September, 2021. The clipped area contours Punta Tagliamento beach in Veneto Region, near the mouth of the river Tagliamento. A Google Maps image is provided to better understand the features on the ground.



Sentinel-2 Level 1C tile



Clipped view of the tile showing the Punta Tagliamento beach.



Google Maps satellite image showing with a red contour the clipped area

## 4.2 Post-processing

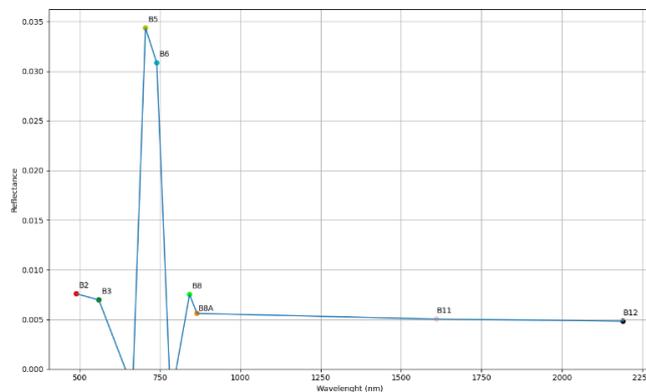
In our work we pre-process each Sentinel-2 product applying a cloud filter, the atmospheric correction (which also included the resampling) and the clipping of the tile. Post-processing analysis steps of satellite image includes all the analysis techniques adopted to elaborate raster data and the choice of a particular one depends on the outputs we want to produce. We report the principal methods adopted in our work to perform such analysis:

- 0) plotting pixel spectra and spectral signatures;
- 1) producing band indexes maps;
- 2) the classification of pixels by means of machine learning algorithms.

### ➤ Plotting pixel spectra and spectral signatures

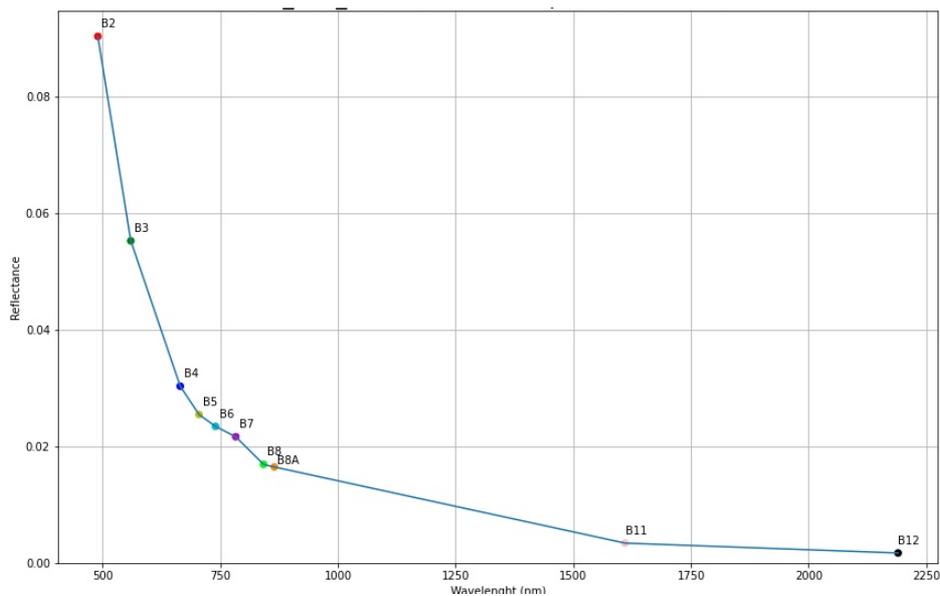
#### Pixel spectra

**Pixel spectra** are reflectance vs wavelengths charts where we display reflectance values for all the bands of a single pixel. Since reflectance is a non-negative physical quantity, we don't expect any of the band values to be lower than zero. If this case happens, we have to discard the pixel, or check if the atmospheric correction has push down too much the reflectance values (for example, the figure below show a spectra with unrealistic negative values in band B4 and B7 as consequence of having activated the glint correction parameter over an area not affected by such problem).



Plot's points can be join together with a broken line to better visualize a “trend”. It is important to remind that a multispectral instruments measure reflected light in few bands, thus peaks and valleys that appears in spectra plots have not necessarily the true shape of the real-world spectra due to the lack of several intermediate points. However, the set of ten spectral points allow us to see the general trend of the spectral data and then make some assumptions of what is the environmental features that is producing such spectra.

For example, one of the most recognizable spectra is that of the seawater, which has higher reflectance values in the visible band (typically, it peaks in the blue band B2), and then decrease towards zero-values in the SWIR part of the electromagnetic spectrum. This is a behaviour expected from the nature of seawater, which is black in the IR band. We show in **figure 4.6** an example of a water spectra.

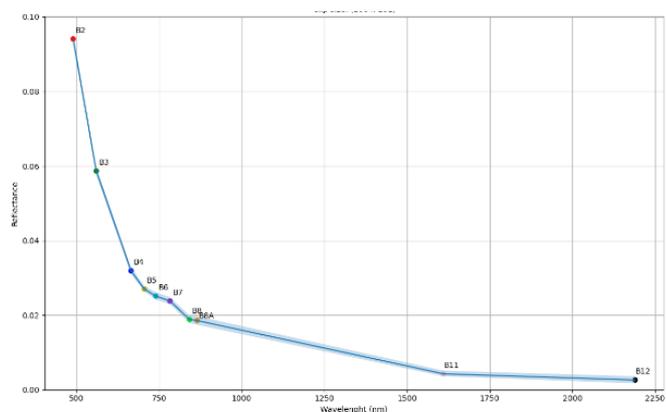
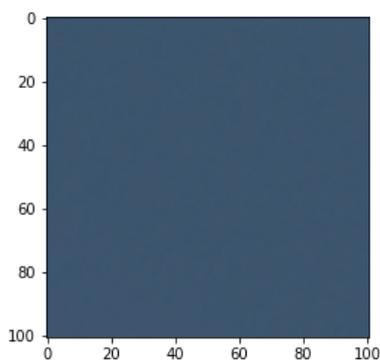


**Figure 4.6.** Example of a typical seawater spectra signature. Data plots are averaged band values over a seawater area in the Gulf of Trieste of size 100 x 100 meters, on September 12, 2020. Intensity of the measured signal can vary due to a number of conditions but the trend remains the same.

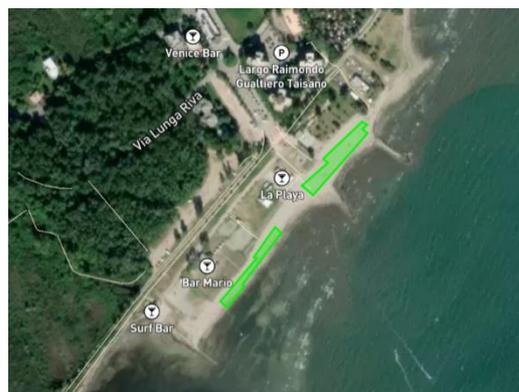
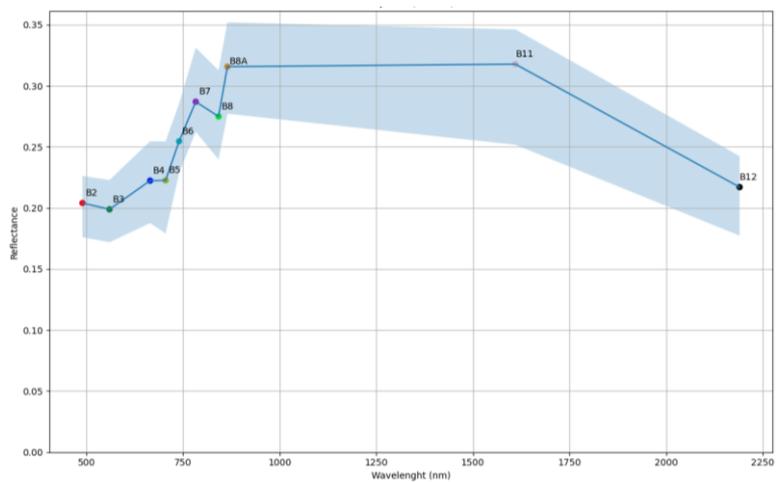
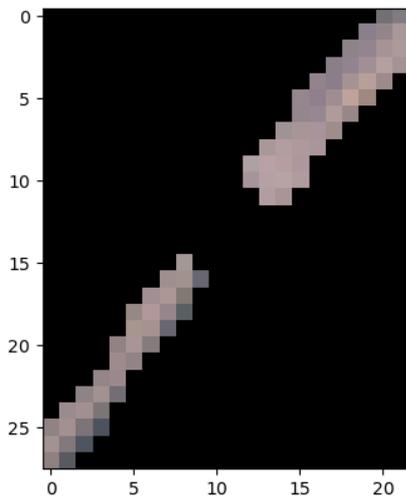
## Spectral signature

If we identify the nature of the object producing a spectra, we can say we have found its **spectral signature**. A single pixel of seawater, as it shown above, exhibits the expected trend for a seawater body. However, a single pixel could exhibit some random fluctuations of band points due to the fact we are studying the natural environment. Thus if we want to obtain a more representative and statistically significant spectral signature we have to take into account several pixels containing the same object. We can achieve this goal by clipping a tile over single or multiple areas where we are very confident that only a kind of material lies on the ground and then collect look at some statistical indicators such as: the mean of all the reflectance values, the value of the first venitle (or the quantile at 5% of the distribution) and last ventile (or the quantile at the 95% of the distribution). Hereafter are some examples to clarify our statements.

First, we go back to water and we clip the tile of the Level 1C product of the orbital passage of Sentinel-2B over the Gulf of Trieste on 12<sup>th</sup> September, 2021. The clipping polygon is a square of 1 km<sup>2</sup> size and a visual inspection of the area in the RGB bands seems to not show nothing else than seawater in that area. Having an area of 1 km<sup>2</sup> size means that we have in total 10 000 pixels. Averaging over all these values gives a series of ten points that let us to draw the thick blue line in the middle of the blue shadowed area. The borders of lower and upper side of the shadowed area are instead build by the reflectance values at the quantile 5% and quantile 95%. Thus with this simple plot we can say that almost the majority of the pixels of the distribution are very similar to the seawater spectral signature.



We report another example of a spectral signature, this time from a typical coastal environment like a sandy beach. It's a bit more difficult to know exactly if the surface of a beach is completely free of other objects like beached seaweed, woods or strand with sun umbrellas, but at least we can exclude the latter if we not study the area during the summer time and the other by averaging over a large number of pixels, eventually manually discarding some strange and unexpected pixel spectra signature. The following example is made by clipping the same Sentinel-2 tile of the seawater example over the pebbles beach of Marina Julia in Monfalcone (Italy). This time we used QGIS to define the borders of the clipping polygons to be sure to avoid to select buildings or the seawater. The bottom image is a screenshot from OpenStreetMaps showing a satellite view of the area and the two polygons.



## Band's indexes

Because pixel of different bands refers to the same surface area, raster bands can be combined to produce new images, therefore enhancing some ground features that could not be seen using information of one band alone. The simplest way to combine raster is to select a triplet of bands and make RGB images, i.e., assigning to each pixel the right amount of red, green and blue to produce coloured images, or False-Colour images, i.e., substituting the red band with an infrared one. **Figure 4.7** shows examples of these images over the island of Barbana, in the city of Grado lagoon.

Raster bands values can also be algebraically matched producing band indexes. Several studies have developed a huge amount of indexes for Sentinel-2 mission [22]. Since marine litter research is a newly and unforeseen field of study for this mission, researches have investigated the effectiveness of adopting old and new indexes. No research has found yet a set of indexes to highlight with high confidentiality the presence of waste in seas or coasts; indeed most of the studies use those indexes to discard unwanted pixels and guide to selected areas neural network algorithms.

In the following, we list the set of indexes we have used for our study and that has been selected from a comparison review of literature about marine litter detection. Terms of the indexes formula refer to Sentinel-2 bands as reported in **Table 2.2**. Numbers right to indexes names refer to bibliography entries.

- *Normalized Difference Vegetation index* <sup>[23]</sup>  $NDVI = \frac{B08 - B04}{B08 + B04}$
- *Normalized Difference Water index* <sup>[24]</sup>  $NDWI = \frac{B03 - B08}{B03 + B08}$
- *Plastic Index* <sup>[13]</sup>  $PI = \frac{B08}{B08 + B04}$
- *Floating Debrids Index* <sup>[2]</sup>  $FDI = B08 - B06 - (B11 - B6) * k$  ,  $k = \frac{\lambda_{B08} - \lambda_{B04}}{\lambda_{B11} - \lambda_{B04}} * 10$
- *Simple Ratio* <sup>[25]</sup>  $SR84 = \frac{B08}{B04}$

All the indexes listed below highlights different features of Earth surface providing both a qualitative and quantitative way to analyse ground features (see **figures 4.8 and 4.9**). The use and the meaning of first two indexes is well established in literature: these indexes are adopted to investigate, respectively, areas occupied by vegetation or water. The Plastic Index was first developed to find signal of plastic in the experiment conducted at Limassol's Old Port, Cyprus, by a group of researchers working at marine litter detection [13]. In their study, a raft made of blue empty plastic bottles was placed in the sea near the coast and measurements of reflected light were collected by Sentinel-2 and by an unmanned aerial vehicle. The plastic index uses Band 8 over contribute of Band 4 plus Band 8 because the study founded that plastic bottles emitted well detectable signals in the NIR band. The Floating Debrids Index (FDI) has been developed specifically to try to improve the set of indexes at disposal able to finds material objects floating on seawaters [2]. SR84 is a simple index to evaluate how much a pixel is brighter in NIR than Red. It has been used in our work to simply obtain a quick view of the most prominent infrared features of an area.



**Figure 4.7.** Eastern part of the Grado Lagoon (Italy) from a Google Maps screenshot. *Left Panel:* RGB image. Each pixel is a combination of the blue band (B02), green band (B03) and red band (B04). Image resolution is 10 m per pixel. *Right Panel:* False colour image. Respect to the RGB image the red band has been replaced with the B08 NIR band. This representation enhance the pixels with high coverage of grass or other plants.



**Figure 4.8.** Indexes maps of Eastern part of the Grado Lagoon. *Left:* NDVI map. Darker areas identify urban or water areas, whiter areas identifies green lands. *Right:* NDWI map. The colour meaning is complementary to the NDVI map, i.e., water bodies are whiter than land.



**Figure 4.9.** Indexes maps of Eastern part of the Grado Lagoon. From top left clockwise: i) PI map. ii) FDI map. iii) SR84 map.

➤ **Ineffectiveness of indexes analysis for ML search**

As appointed before, literature largely discusses the use of indexes to help to search marine litter. Unfortunately, no clear and unique methodology is provided to solve this problem. We have

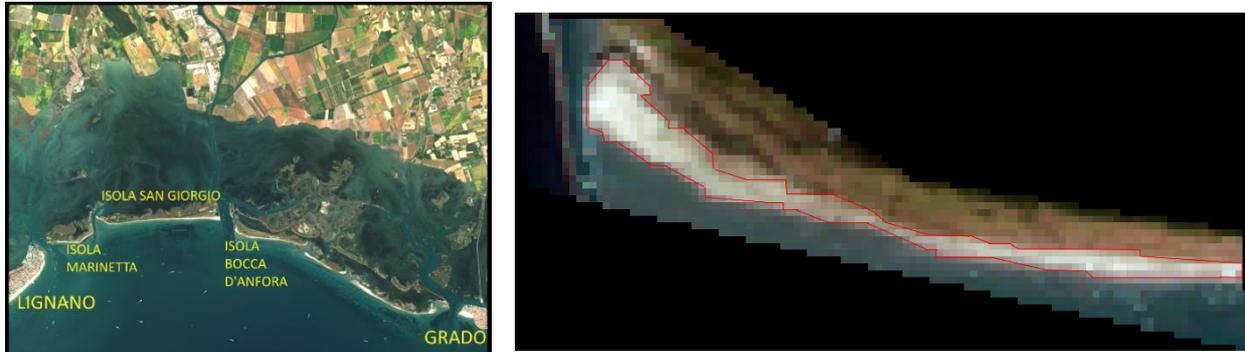
tested such indexes during the first part of our work but we ended up with the evidence that those indexes are ineffective to highlight marine litter features on the ground for our areas of investigation. The main problem is that indexes are thought to highlight ground features by comparing only a restricted set of reflectance values. But the problem is that we don't know a priori what the relative distance between band points when we look at a marine litter pixel. Indexes are good to quickly highlight features on the ground but doesn't exist a reference catalogue or database to assign particular values to a precise class of objects. For example, the NDVI index use the normalized distance between the bands B8 and B4. If we take a coastal area and we get a value of  $NDVI = 0.5$  for a pixel, we don't know if this indicates the presence of marine litter in that pixel. Maybe the object is reflecting more infrared light than red because is part of a scrubland, or maybe we are looking at beached seaweed (in this case, it could be the case that marine litter is hiding under the seaweed, but surely we don't see it signal). Thus the use of indexes, from our point of view, implies the loss of fundamental information from other band values that let to classify more properly a pixel. For any pixel, we should use the combined information of the whole set of indexes, but this statement points towards the use of the spectral signatures. Finally we recall that marine litter is accumulated in shores in several ways and we cannot compare our literature results with our cases of study.

## Classification of pixels by means of machine learning algorithms

The study of coastal areas implies the analysis of hundreds of pixels. For example, if we look at the west portion of the Isola Bocca d'Anfora in the Grado Lagoon (Italy) (**figure 4.10**), we count approximatively 400 pixels inside the sandy beach!

This example shows clearly that in general we are dealing with a very large amount of pixels and that visual inspection and classification of coastal areas by the only human action is not an affordable effort. We have to use the aid of machine learning techniques to overcome this problem. The open source Python library `scikit-learn` (<https://scikit-learn.org/stable/index.html>) (**figure 4.11**) is a useful tools to develop simple codes to analyse raster and classifying their pixels, i.e., grouping together pixels that shares a common feature. As already explained, we study pixels by

analysing their spectra. Thus, by classifying pixels we mean collecting together all the pixels with similar spectral shape.



**Figure 4.10.** *Left:* atmospheric corrected Level-1C product on the 12<sup>th</sup> September, 2021, displaying a view of the Grado Lagoon. *Right:* the same area clipped with a polygon encircling the west part of the Isola di Bocca d'Anfora. The red contour highlights the sandy beach where we have counted the total amount of pixels inside lying inside it.



**Figure 4.11.** *Left:* screenshot of the home page header of the scikit-learn library. *Right:* the main menu of the User Guide [26].

From the set of scikit-learn proposed machine learning algorithms, we adopted the unsupervised learning ones. Unsupervised techniques are able to classify objects but cannot establish their nature, i.e., if we are looking at a picture made by yellow and red dots, unsupervised methods

will find out that there are two different groups of objects but will be unable to say that they are colours and especially *what* colours.

We use unsupervised algorithms to classify spectra we don't know a priori what we are looking on the ground and in particular we don't know the shape of marine litter spectra.

However, we don't move totally blindly when analysing a coastal environment and some prominent features, such as seawater, sand, Mediterranean scrublands ... are easily recognizable. So, we can adopt a "semi-supervised approach", as it is defined in literature, that is the technique of apply unsupervised methods on a "training area", then label the extracted classes and use such labels to classify a completely new area. Due to the lack of the working time needed to develop, test and apply a fully operative Python algorithm reproducing this technique, we have to perform the labelling phase and the classification of the new areas by hand, thus we have built a "semi-automatized" semi-supervised approach. Here we lists the steps to perform this operation:

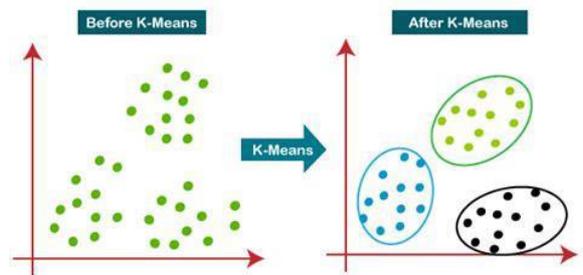
- i) Unsupervised classification, with a Python code, of a training area;
- ii) manual labelling of the spectral signatures of the founded classes;
- iii) enrichment of a spectral database with the founded classes;
- iv) unsupervised classification, with a Python code, of a new area;
- v) manual labelling of the spectral signatures of the new classes by looking at the classes in the spectral database.

In the next section we recall briefly the unsupervised techniques adopted for our work and describe more in detail how we can build the *database of spectral signatures* from selected training areas.

Selection of an effective unsupervised classification methods.

Several tutorial and research articles indicates the unsupervised **K-means** technique as the more simple and suitable technique to classify satellite images.

K-means is a clustering algorithm, i.e., it uses as input the expected number of classes  $N$  we expect to have in a sample and then it try to group sample members to match this requirement. In statistical terms, the algorithm try to minimize inner variance of points when grouped in a set of  $N$  groups [27].



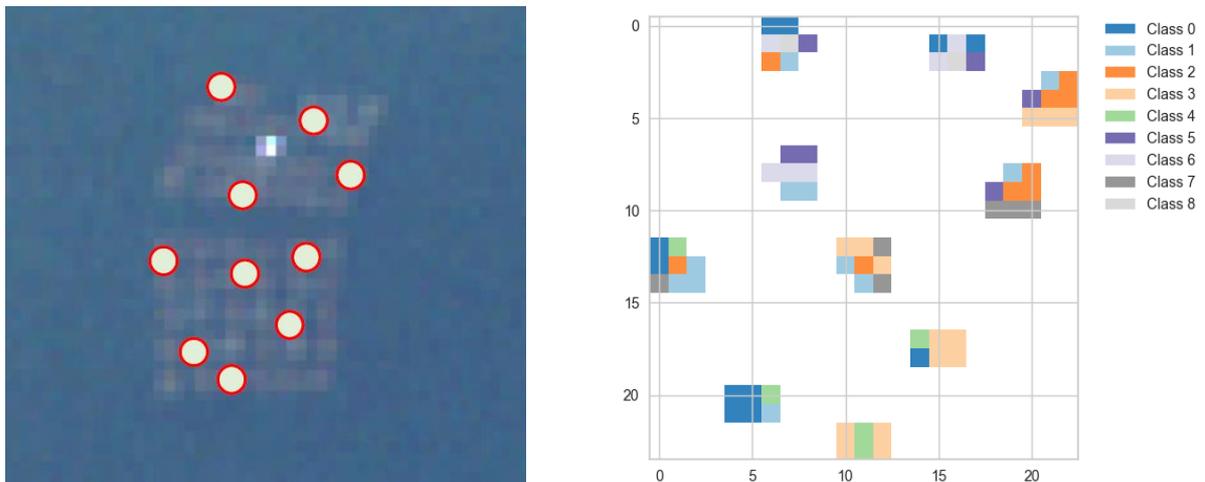
However, we have found that K-Means technique don't perform well with our set of spectral data and we have abandoned it. The main problem of K-Means is that it cannot handle well outliers. Outliers are those points of a sample that are statistically outside the point distribution. Since marine litter spectra are expected to be outliers, of course with the use of this algorithm we are going to miss them. We have also tested the performances of the algorithm in doing a good classification of pixels, finding bad results. Here we show as example the results of the classification of a maritime area near the town of Duino (Italy) where there is a fish farming made of 58 floating farms (**figure 4.12**).



**Figure 4.12.** Left: Google Maps Satellite view of the fish farming (bottom left) in front of the town of Duino (top right). Right: photo of a fish farming depicting the floating circles of the farms (credits: Jebulon-Own work, CC0, <https://commons.wikimedia.org/w/index.php?curid=53116580>)

We selected the usual tile of Sentinel-2 Level-1C product on 12<sup>th</sup> September 2021, clipping it with 10 circular polygons to select 10 random circles.

Applying the K-means algorithm over this selection (to predict the best number N of classes as input, we have used the silhouette score metrics, which is described in details at the scikit-learn webpage: [https://scikit-learn.org/stable/auto\\_examples/cluster/plot\\_kmeans\\_silhouette\\_analysis.html](https://scikit-learn.org/stable/auto_examples/cluster/plot_kmeans_silhouette_analysis.html)) we obtain a value of N = 9 and the clipped circles appears classified as shown in **figure 4.13**:

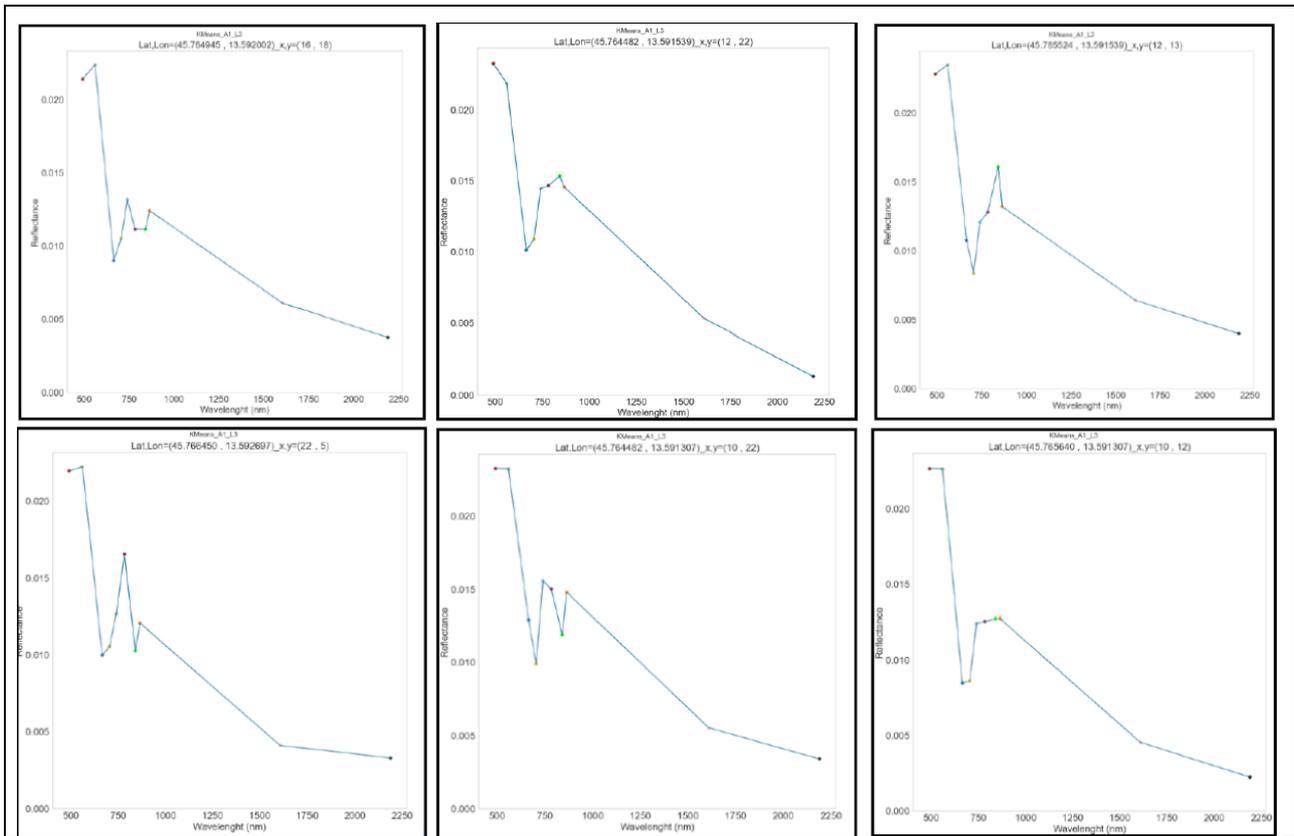


**Figure 4.13.** Left: position of the 10 random circles. Right: classification results for the clipped polygons.

Alongside the classification, we save in separate folders the plots of all the pixel spectra of each class. This allow us to visually check the *internal coherence* of the class, i.e., verify that all the spectral shapes are similar. The following table shows how the internal coherence is around 50% - 60% for several classes.

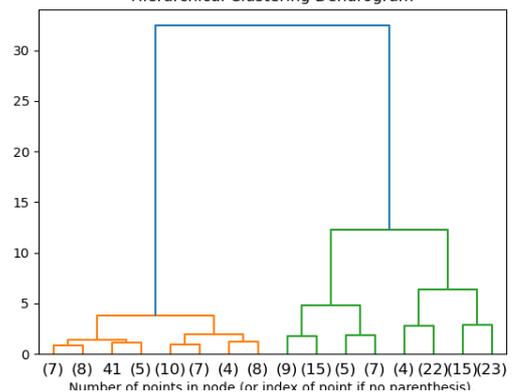
Classes	0	1	2	3	4	5	6	7	8
# pixels	11	11	9	14	5	6	6	6	2
Internal choerence	65%	80%	66%	50%	60%	60%	100%	60%	50%

These are not acceptable results. The figure below shows that in class 3 the algorithm has grouped together 6 different spectral shapes (**figure 4.14**)! We have also tested the increasing the number N of cluster, but it have not shown an improvement of the internal coherence.



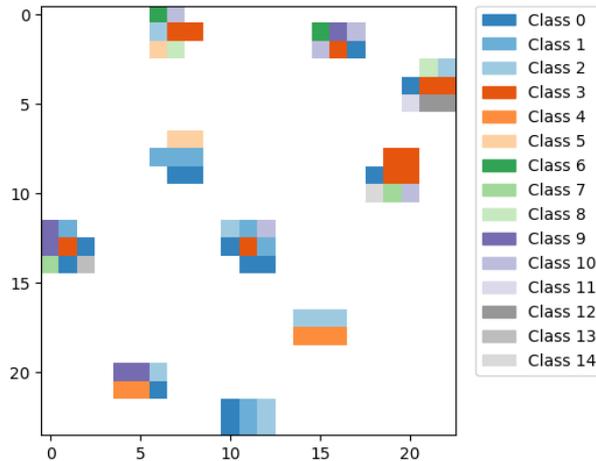
Luckily, consulting the scikit-learn User Guide, we have found an alternative unsupervised technique that work better than K-Means and reach very good results in the internal coherence parameter. This new technique is the **Agglomerative Hierarchical Clustering (AHC)** [28]. The principle of work of this technique is the following: the algorithm tries to cluster points following ‘decisional trees’, i.e., making dendograms. At start, each point is treated as a one-component cluster, then the algorithm recursively

Hierarchical Clustering Dendrogram



Example of dendnogram. The first two branches are on the top and then they split into more branches going downward.

cluster points according to a distance threshold. This method allow us to extract outliers and we can also make classes composed by a single member. As for K-Means, the AHC method requires to give as input the number N of expected clusters. If we look at the classified ten polygons we obtain far better results than K-Means.

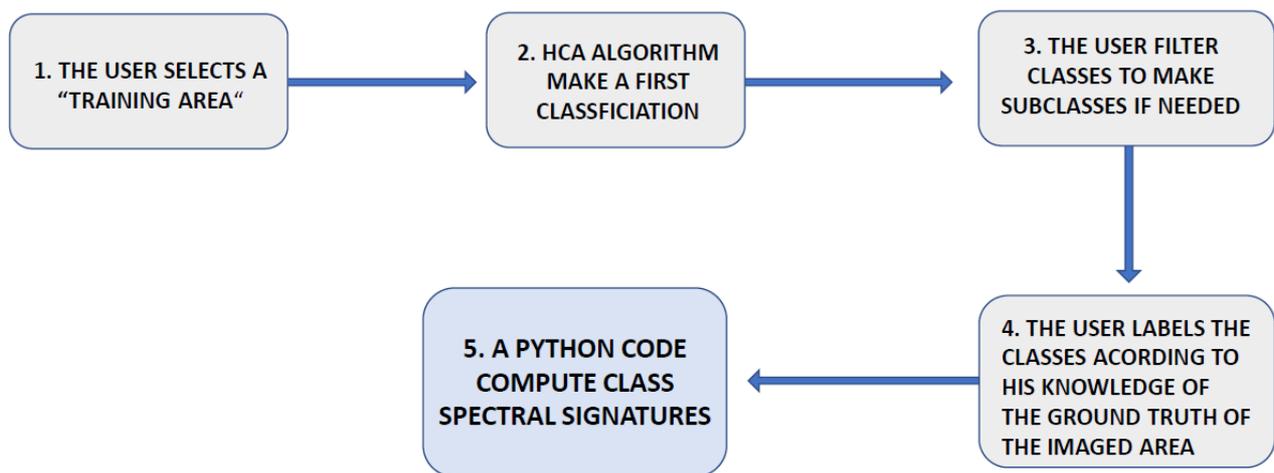


Classes	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
# pixels	3	16	3	6	2	2	7	2	5	2	5	1	1	6	9
Internal coherence	100%	65%	100%	100%	100%	100%	70%	100%	36%	50%	80%	100%	100%	80%	70%

Only three classes shows a medium internal coherence and only one class has bad results. This hint us of the necessity of a second manual screening of the classes, i.e., we look at medium-low internal coherence classes and manually divide into more new subclasses the pixel spectra. With this operation we obtain a good set of classes and for each one we can compute their spectral signature. We are ready now to process *training areas* and extract spectral signatures of the main ground features of a coastal environment, i.e., making a spectral database to use as a collection of *training classes* to be used to manually check if in new areas there are potential marine litter pixels.

## 5.1 Analysis of training areas

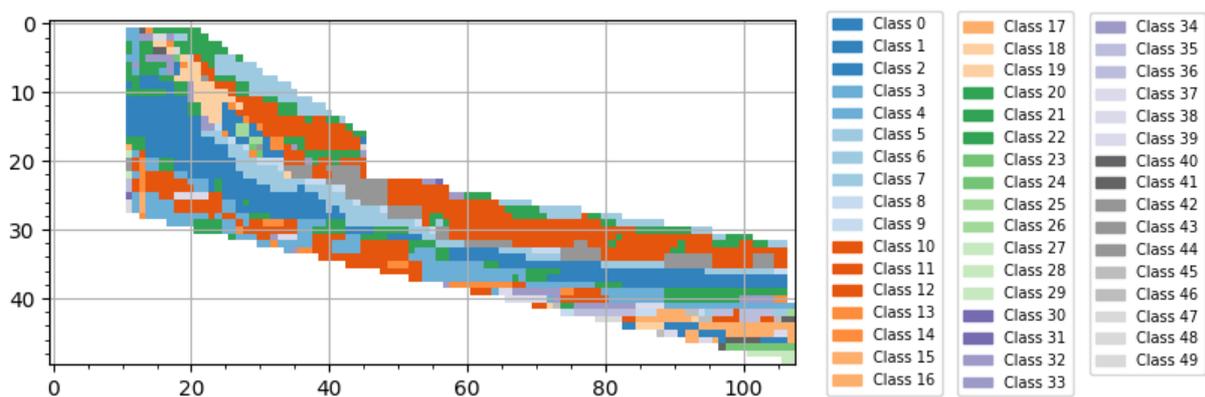
The steps described in the previous chapter have shown how we get spectral signatures from the classification of pixels. Now we describe the extraction of spectral signatures from main environmental features on a coastal area that has been selected to have a good control on the ground-truth of the imaged area. Here we provide a scheme of the steps we have to perform to analyse a training area using both the HCA algorithm and the manual filtering (when it's needed) and get as output a set of spectral signatures.



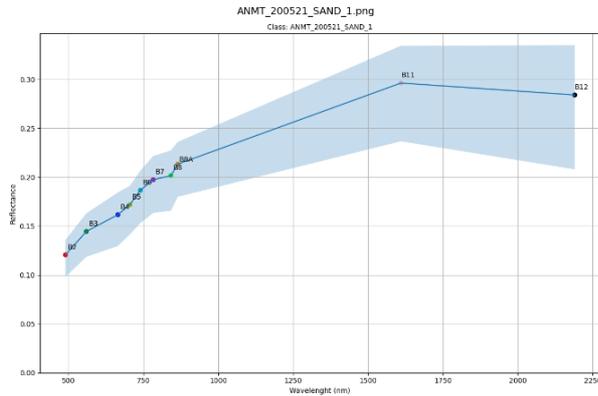
As example, we presents our results for the training area of Isola Bocca d’Anfora, already cited in **figure 4.10**. This place is a good training area since it’s located in the middle of the Grado Lagoon, far away from urban settlements and inhabited. We have studied a clipped area from the Sentinel-2 Level 1C tile on 20<sup>th</sup> of May, 2021. The area of study presents a sandy beach surrounded by Mediterranean scrubland on the back and seawater on the front and on the left side. Respect to **figure 4.10** it can be seen that there is a change of the seawater position, which is disappeared from the middle sections of the picture: this is due to low tide on the 20<sup>th</sup> of May.



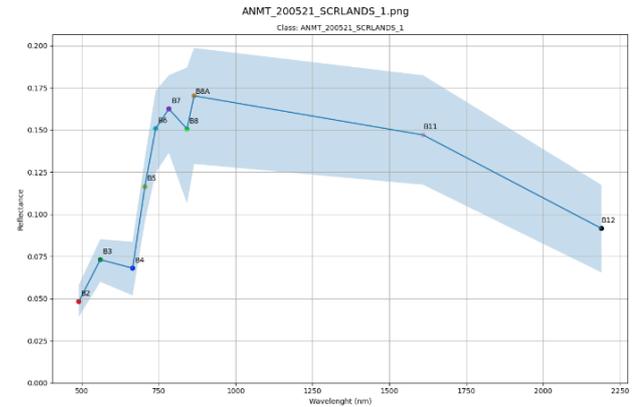
When processing this area with the AHC algorithm we find a reasonable number of classes with a good internal coherence if we input  $N = 50$ . Once the algorithm has made the classification, we obtain a colourful map showing the classified pixels and their position:



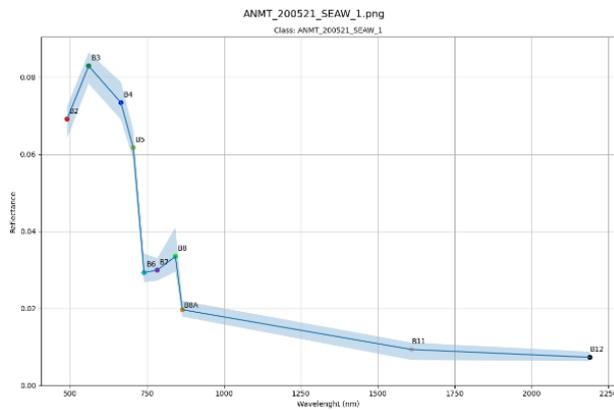
Here starts the most time consuming part of the work: by comparing the grouped pixels place and the RGB preview of the area, we try to deduce what was the environmental feature on the ground producing the observed spectral signature. At the end of this process, for this scenario, we have identified 4 main features: seawater (SEAW), sand (SAND), Mediterranean scrublands (SCRLANDS), intertidal zone (TIDAL). We report here as example the first spectral signature of each category and the position of the group of pixels producing it. Other spectral signatures are attached to the deliverable data folders.



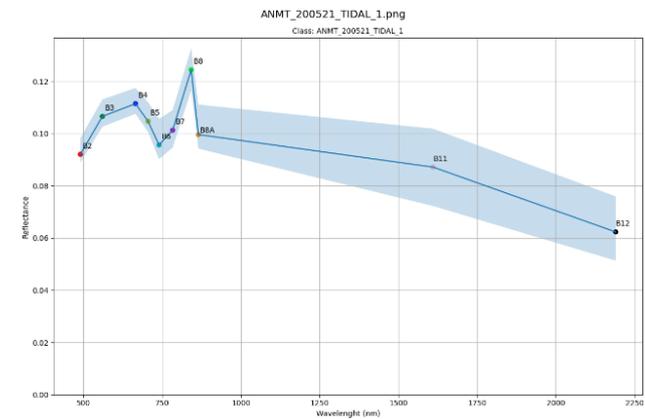
Example of the sand spectral signature.



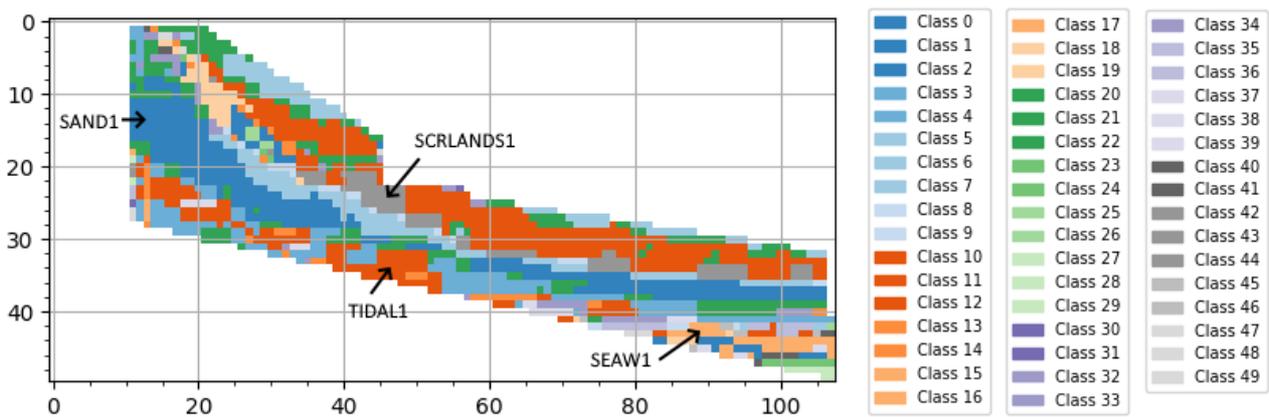
Example of the Mediterranean scrubland spectral signature.



Example of seawater spectral signature.



Example of the intertidal zone spectral signature.



The position of the pixels producing the four spectral signatures shown above.

## 5.2 Search for marine litter hotspots on selected beaches

Based on our knowledge of the marine litter phenomenon and the technological limits in using the Sentinel-2 mission, when we look at a coastal area we can label a pixel as a potential host of marine litter if one or more of the following conditions are satisfied:

- 1) The pixel shows a spectral shape that is atypical respect to the “reference classes” of our spectral database;
- 2) The pixel is placed in an area near the coastline, including the intertidal zone (if we are analysing a day with the seawater in low tide condition, we could look at submerged litter);
- 3) Pixels are sparse or form a little cluster localized in a small part of the whole area of study;

Conditions from 1 to 3 help us to define a methodology to afford the analysis of a beach and select the pixels having the highest probability of having marine litter inside: in a shoreline scenario, we select a group of pixels, then we look at our spectral database (condition 1) and/or at the spatial position of the pixels (condition 2) to discard such pixels, the remaining ones, if also satisfies condition 3), are identified as potential marine litter pixels.

Once we found all the potential marine litter pixels, we can compute the ratio between the total amount of them ( $N_{pm}$ , “number of potential marine litter”) and the total amount of pixels of the analysed coastal area ( $N_{tp}$ , “number of total pixels”), according to the formula:

$$FP_{ml} = \frac{N_{pm}}{N_{tp}}$$

This will give us the fraction of pixels affected by marine litter in such area ( $FP_{ml}$ ). In the next section, we describe in detail how we compute this index by analysing a beach zone near the maritime city of Grado, in the Friuli Venezia Giulia Autonomous Region (Italy), on the 12<sup>th</sup> September, 2021.

## 5.1 How we determine the fraction of potential marine litter pixels in a stretch of coast near the area of “Grado Pineta” in Grado (Italy)

In this section we describe step-by-step the procedure to compute the fraction of pixels that have spectral signatures of potential marine litter.

### 1. Selection of the area of study

The area of study is a coastal zone near the residential suburb named ‘Grado Pineta’ in the proximity of the city of Grado (see the Google Maps satellite view of the



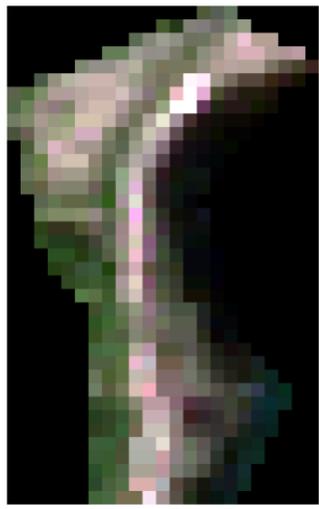
Grado Island). We selected this beach because it is one of the beaches studied by the on field monitoring campaigns for the MARLESS project.

The next pictures show in detail the extension of the area of study: on the left is a screenshot from Google Earth, the red line contours approximatively the area of study; on the right we show the same area as seen by Sentinel-2. The area of study is limited northeast by the beach concessionary “Spiaggia Villaggio Ca’ Laguna”, on the south by the beach concessionary “Stabilimento Costa del Sol”. The depth of the polygon is determined to satisfy two conditions:

- a. due to the spatial resolution uncertainty, we want to be sure to include all pixels affected by shore pixels;
- b. we want to include large portions of ‘macro areas’ such as scrublands, sand, seawater, etc., to be sure to not confuse hotspots of such features with marine litter.

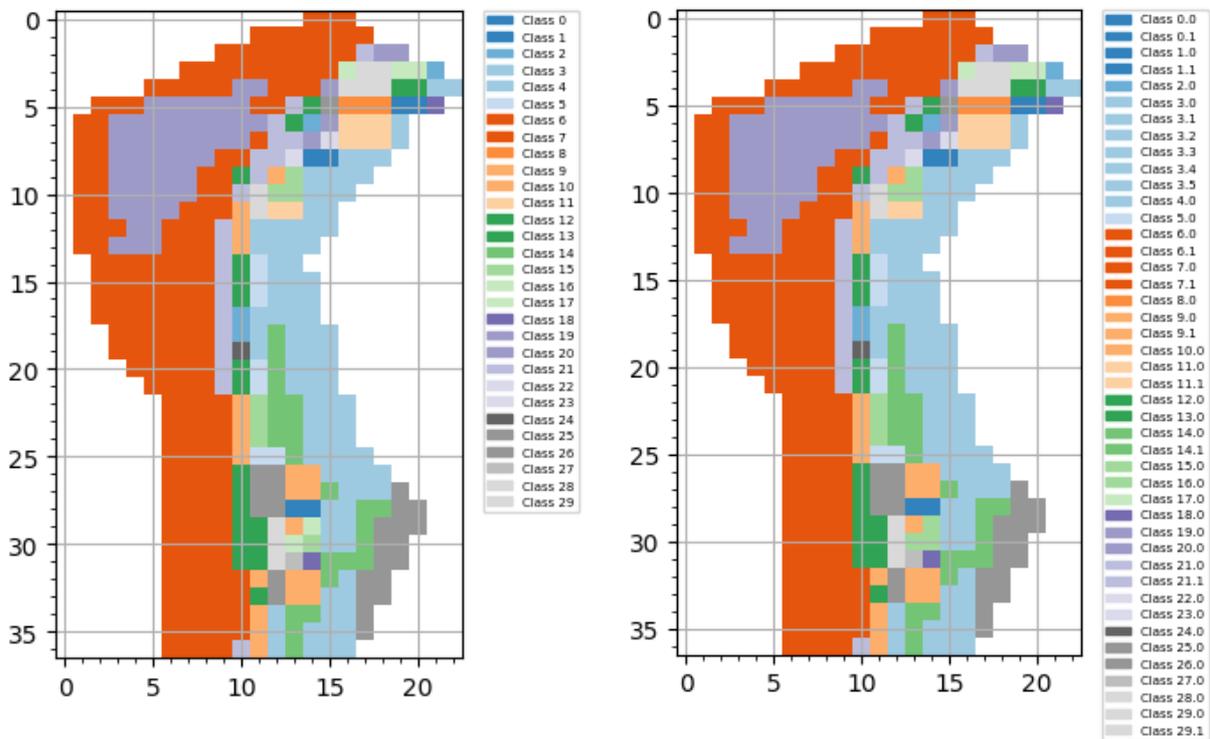


The analysis we present is performed on the level-1C Sentinel-2B product of the 12 September 2021. The area of study was cloud free and according to the image processing chain explained in the previous sections we pre-processed the product. The RGB preview of the output tiff, containing 502 pixels, is shown below.



## 2. Application of the semi-supervised classification algorithm

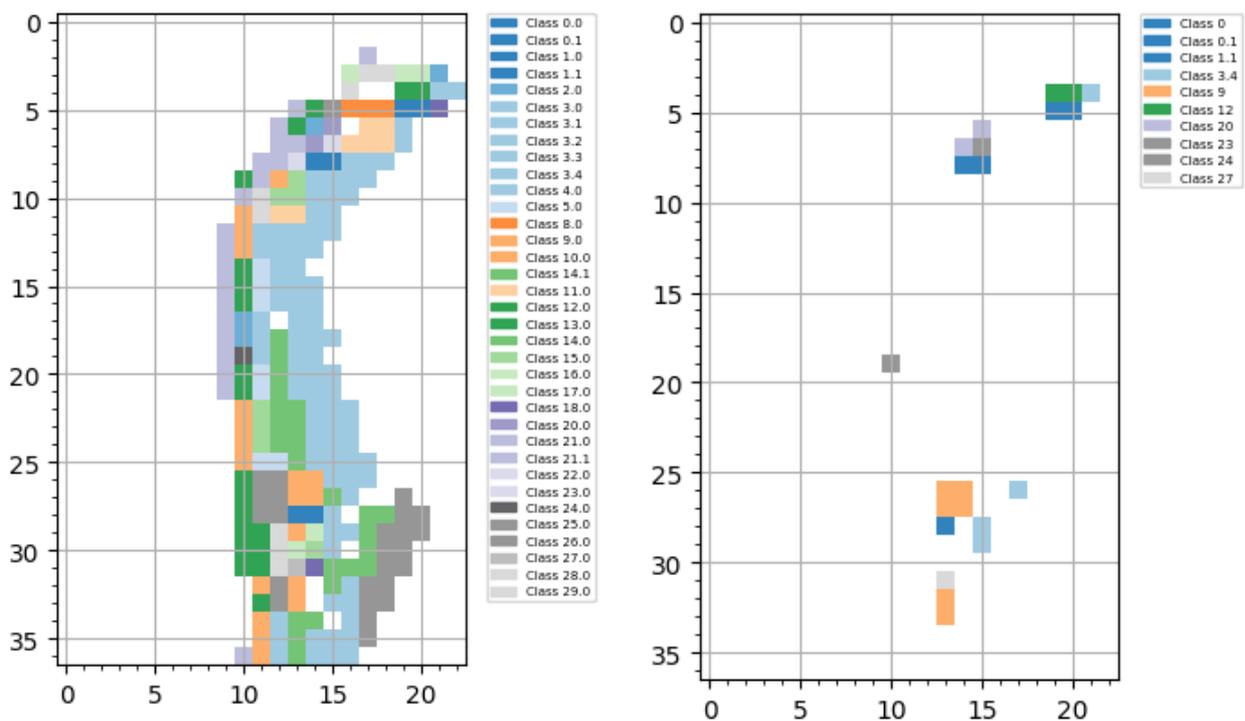
The application of the semi-supervised algorithm suggests the presence of 30 classes (see the left image in the box below). Manually filtering the spectral pixels let us to identify several new sub-classes (see the right image in the box below). At this point, we plot the mean spectral signature of each class along with the shape of the pixel spectra at the first and third quartile of the statistical distribution. This help us to quick check the goodness of the filtering.



## 3. Finding potential marine litter spectra classes

We are now ready to investigate each class to find out potential marine litter pixels. We follow the receipt listed at the start of this chapter to discard classes that are unlikely to be associated to pixels hosting marine litter. First, we have to establish the extension, in terms of pixels, of the area that is interested by the phenomena of accumulation of debris by sea currents and winds. Visually

comparing the Google Earth screenshot of the area of study and the Sentinel-2B RGB preview we identify a group of pixel classes matching our requirements. Hereafter we refer to this area as the ‘Grado Pineta Beach’. Counting the number of pixels inside Grado Pineta Beach we get  $N_{tp} = 246$ . Now we use the reference classes to compare them with the mean spectral signatures of all the Grado Pineta Beach classes. We remind that the reference classes are those we have determined from the analysis of the Anfora and Fossalon beaches. The comparison process allow us to discard the majority of the classes (see figure for the remaining classes after this step). Following the working principles stated before, we should now search for pixels that are unlikely covered by marine litter originated from the sea, but we cannot find a suitable class satisfying this condition. Thus, we can compute the total amount of pixels in the remaining classes,  $N_{pm} = 22$ .



#### 4. Computing the probability of the area to be a marine litter hotspot

Following the definition given in beginning of this section, we can compute the Grado Pineta Beach coefficient  $FP_{hsp}$  :

$$FP_{ml} = \frac{22}{246} = 0,089$$

$$FP_{ml,\%} = 8,9 \%$$

#### 5.2 Results of the search of marine litter with Sentinel-2 in two MARLESS beaches

We have done the analysis presented so far on two beaches, one is the already presented Grado Pineta beach, and the other is a stretch of coast near the Brenta mouth called “Isola Verde” in the Veneto Region. These two beaches are interested by the MARLESS on field monitoring campaign and have been analysed in D3.3.2 with the numerical simulation too. These are sandy beaches that are ‘large’ in terms of ‘pixel extension’, this property let us to collect statistically robust spectral signatures from the environment and to have a more clear distinction of the boundaries between the various parts of the beach. Moreover, these areas are not occupied by beaches concessionaries buildings neither are too close to urban areas or roads. Thus we can assume that changes on the beaches surfaces are caused only by the action of the natural elements or by sporadic cleaning campaigns.

Under these assumptions, we have analysed the two areas using the methodology described in the previous section performing the research of marine litter in different days of the year, excluding the summertime for the obvious reason of having a high probability of area occupancy by sun bathers. Hereafter we show an example of clipped area of the Isola Verde beach and then we report, in **table 5.1** and **table 5.2**, the results of the search of potential marine litter pixels on these two beaches.



Date	$N_{pm}$	$N_{tp}$	$FP_{ml}$	$FP_{ml,\%}$
24-02-2021	22	246	0.089	8.9
20-05-2021	16	231	0.069	6.9
12-09-2021	19	239	0.079	7.9
16-12-2021	39	234	0.167	16.7

**Table 5.1.** Results for the Grado Pineta beach.

Date	$N_{pm}$	$N_{tp}$	$FP_{ml}$	$FP_{ml,\%}$
24-02-2021	135	443	0.305	30.5
20-05-2021	111	415	0.267	26.7
01-12-2021	122	466	0.261	26.1

**Table 5.2.** Results for the Isola Verde beach.

## PART 2

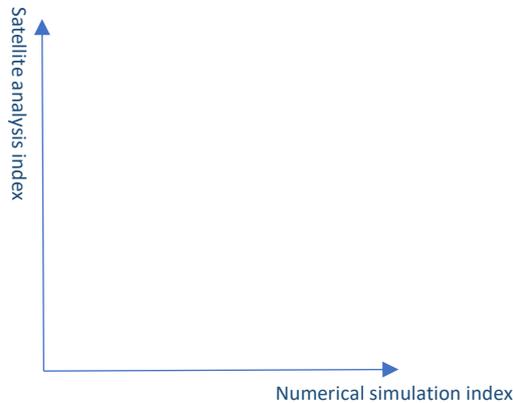
# MAPS OF MARINE LITTER HOTSPOTS PROBABILITY

The identification of beaches, which are potential marine litter hotspots, is based on the integration of the two main sources of information considered in the MARLESS project, namely the satellite data and the numerical simulations of marine litter dispersion. Since those sources are coming from quite different approaches, we have considered worth to avoid to merge their contribution in one scalar index, which would represent the sum of the information of the hotspot potential.

According to this choice, the identification of a marine litter hotspot is conducted in plane where the two dimensions are related to the two sources of information. Two indexes are identified for each beach in study. The x-axis reports the index summarizing the results of the numerical simulations, the y-axis reports the index computed thanks to the satellite image analysis. As a consequence, the analysed beach will be represented by a point on the plane.

It is expected a correlation between the indexes, since a hotspot should be identified by means of both the sources of information. Thus, beaches with high probability to be a hotspot will lie far from the plane origin in the upper right corner, while beaches expected to be marine litter free will cluster very close to the origin of each axes that is in the lower left corner. The agreement on the hotspot potential for both indexes, that is a high correlation, results in points populating a row in the first quadrant of the plane. On the other hand, points belonging to beaches that are far from the linear path connecting the area of potential hotspots with the plane region of clean beaches are considered of quite difficult

interpretation, because of the disagreement of indexes. Those points represents not-classifiable beaches.



Hereafter we present the results from the numerical simulations and the satellite analysis and then their integration in a map of probability.

### 1. Numerical models results:

As far as the numerical simulation index is concerned, this index is calculated considering the results of the D3.3.2 deliverable, i.e., the deliverable concerning the forward dispersion of marine litter. Two statistical parameters are considered: `daily_max(median)` and the `daily_median(median)`. For each day of simulation, we compute the median from the hourly count of the beached particles; then, from this indicator, we calculate the two statistical parameters considered from all the days of simulation: the maximum and the median. These parameters are computed for both the beached particles on the beaches in study and for the entire Adriatic basin, i.e. considering the beached particles on the overall Adriatic coasts. The parameters calculated for the Adriatic basin are necessary to find the constraints from which a beach can be considered with a low/intermediate/ high probability of accumulation of marine litter. Since the satellite image pixel has a spatial resolution of  $10*10\text{ m}^2$ , the simulation parameters too are expressed in  $10*10\text{ m}^2$  both for the beach in study and both for the overall Adriatic. For the former we considered the length and the width of the coast on which the

beached particles are counted, for the latter we considered the shoreline length of the Adriatic Sea which is of about 10 000 km.

In conclusion, the *Numerical simulation index* is going to represent the maximum number of elements found on 10\*10 m<sup>2</sup> on a day and the median number of elements found on 10\*10 m<sup>2</sup> on a day.

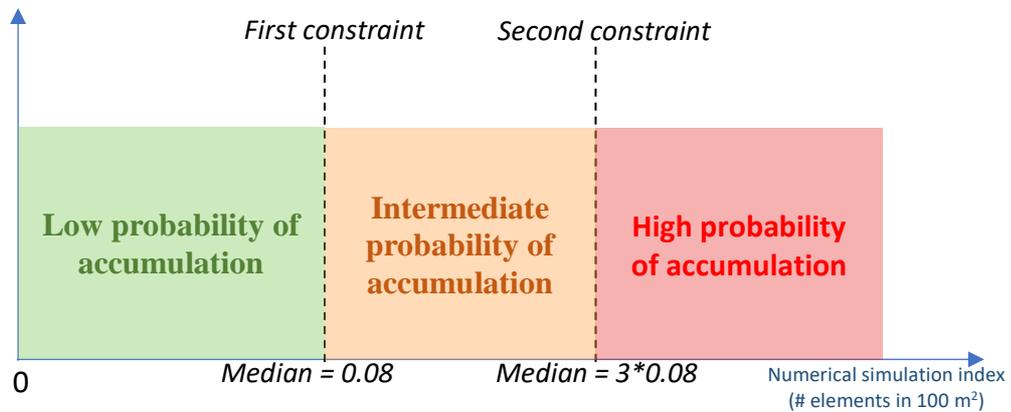
The constraints values found for the overall Adriatic basin are reported in the following table:

Maximum constraint (elements on 100 m <sup>2</sup> )	Median constraint (elements on 100 m <sup>2</sup> )
<b>0.16</b>	<b>0.08</b>

From these values, we consider a beach with different probability of accumulation:

- a **low probability** of accumulation if the calculated parameters are going to be lower than the median constraint;
- an **intermediate probability** of accumulation if the calculated parameter are going to be between the median constraint and 3 times this constraint;
- a **higher probability** of accumulation if the calculated parameters are going to be higher than 3 times the median constraint.

This parametrization is shown in the graph below to better visualize the concept.



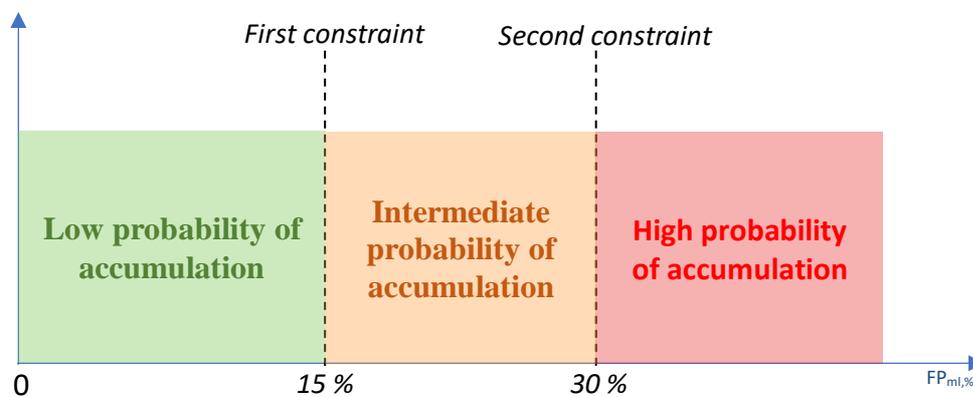
On the table below, the results obtained from the beaches in study are shown together with the two constraints found analysing the overall Adriatic Sea beached particles. The studied beaches are coloured in green if they have a low probability of accumulation, in orange if they have an intermediate probability of accumulation and in red if they have a high probability of accumulation.

Adriatic Sea shorelines: first constraint	0.08 # elements 100 m <sup>2</sup>
Adriatic Sea shorelines: second constraint	0.08*3 = 0.24 # elements 100 m <sup>2</sup>

	Median # elements 100 m <sup>2</sup>	Maximum # elements 100 m <sup>2</sup>
<b>Grado Pineta beach</b>	<b>0.76</b>	<b>1.30</b>
<b>Isola Verde beach</b>	<b>1.45</b>	<b>3.31</b>
<b>Riccione beach</b>	<b>0.19</b>	<b>0.40</b>
<b>Sole Luna beach</b>	<b>0.14</b>	<b>0.33</b>
<b>Opcina Umag beach</b>	<b>0.09</b>	<b>0.11</b>
<b>Trstenik beach</b>	<b>0.01</b>	<b>0.01</b>

## 2. Satellite analysis results

As discussed in the **chapter 5** of this document, we have analysed two MARLESS beaches in different days of the year. For each case we have calculated the parameter  $FP_{ml}$  that describes how much (potential) marine litter pixels we have found on a beach. We can go further in the interpretation of this parameter by noticing that it indicates also the level of probability of an area to be occupied by marine litter. We can establish the following interval of probability of having low, intermediate or high probability of accumulation of marine litter on a beach:

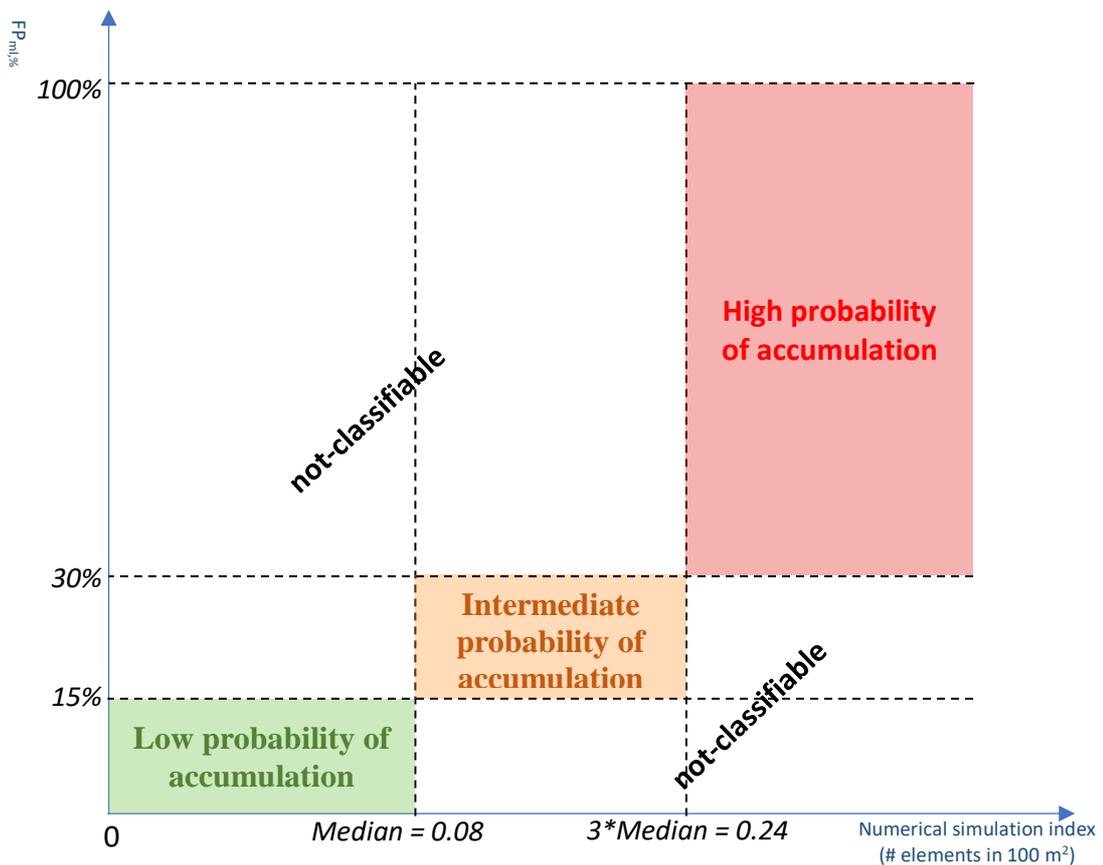


For each beach we compute two  $FP_{ml}$  values: the max and the median, and we will use these two values to the integration with the numerical simulations results. Here we report the data for the two MARLESS beaches we have analysed.

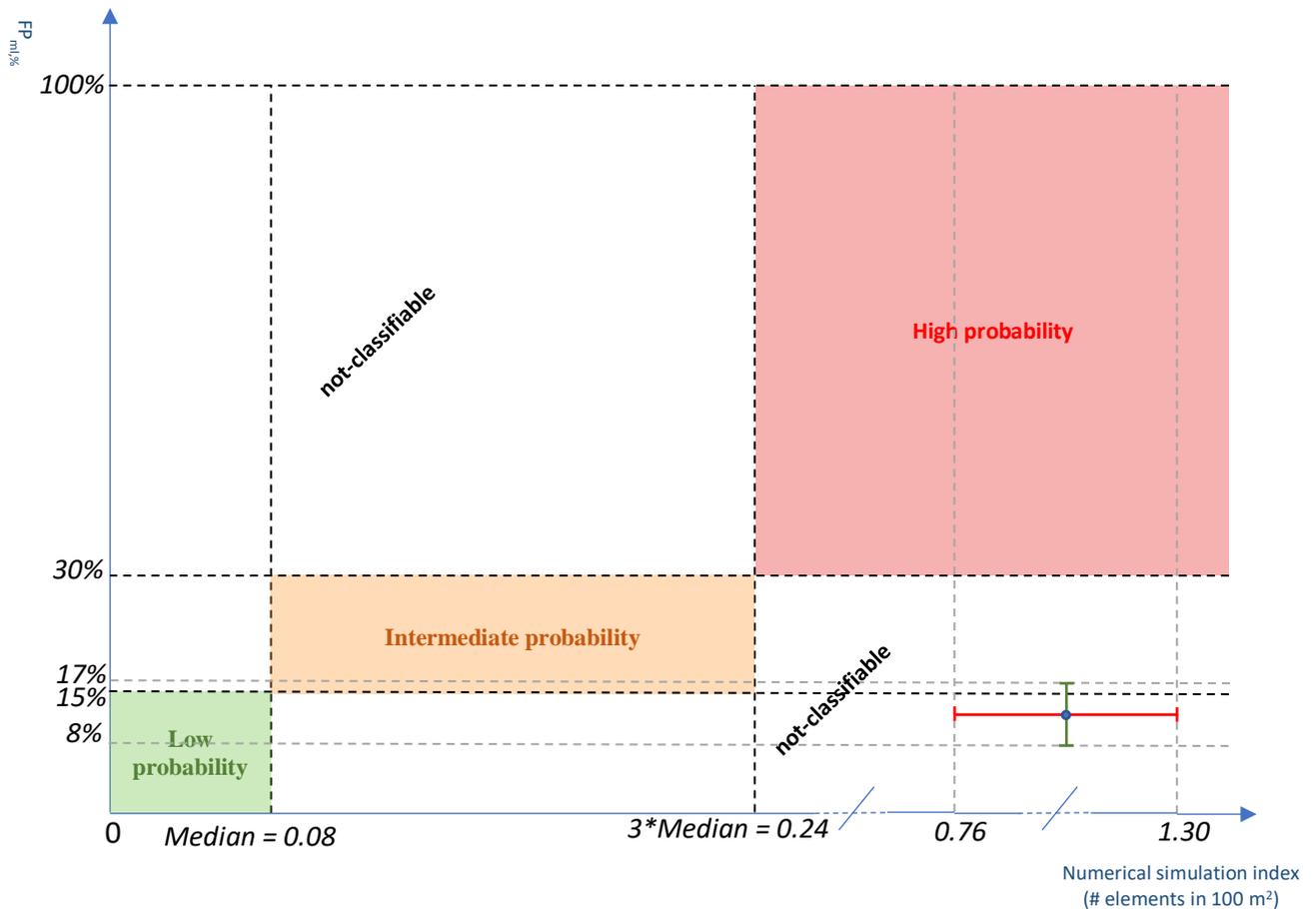
	Median $FP_{ml,\%}$	Maximum $FP_{ml,\%}$
<b>Grado Pineta beach</b>	<b>8.4</b>	<b>16.7</b>
<b>Isola Verde beach</b>	<b>26.7</b>	<b>30.5</b>

### 3. Integration of the numerical model and the satellite analysis results in the maps of probability

By combining the two indexes, a probability map will result: the intersection of the respective parametrization will show the combined areas in which a specific beach can fall. The concept is shown in the figure below. If the beach will fall on the green area (i.e., the low probability one), it will have a low probability of accumulation, and so on with the orange area (i.e., the intermediate probability one) and the red area (i.e., the high probability one). Otherwise, if the beach will fall in a not combined area the result is going to be classified as not-classifiable.

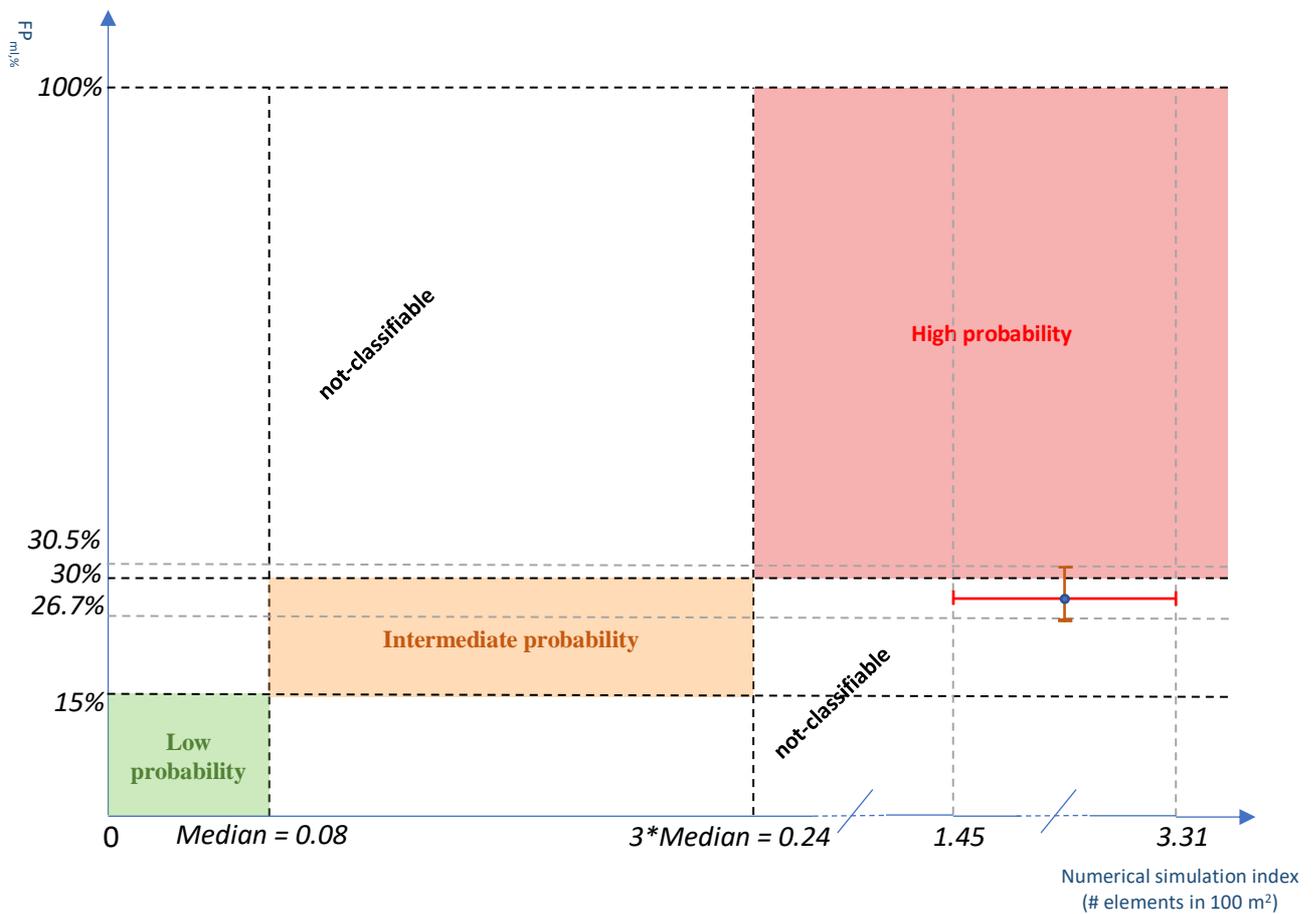


Considering the results obtained for the Grado Pineta beach, the probability map obtained is the following:



As it can be noticed, from the numerical model point of view the beach fall in the high probability area, while from the satellite point of view the beach fall in the low probability one: in conclusion, the Grado Pineta beach is considered not-classifiable.

Considering now, the results obtained for the Isola Verde beach, the probability map obtained is the following:



As it can be noticed, from the numerical model point of view the beach fall in the high probability area, on contrary from the satellite point of view the beach fall in the intermediate probability one. In conclusion, the Isola Verde beach is considered not-classifiable but considering the error bars extension of the satellite index the result doesn't exclude the possibility of this beach to be a high probability accumulation area.

#### 4. Integration of these results with the in situ monitoring campaigns

A normalization of the two indexes is going to improve the power of the two dimensional approach in the hotspots identification. In fact, if both indexes range over the same domain, for example [0,1], the correlation between the indexes is described by points lying on the bisector of the first quadrant of the plane. To this end, we linked to a beach a relative value with respect to the most polluted beach (which is going to have a maximum value of 1) and the less polluted one (which is going to have a minimum value of 0). The information of the most and less polluted beaches will come from the in situ monitoring campaigns measurements.

Moreover, we aimed to verify if the results obtained with the numerical simulation are linearly related to the one of the in situ monitoring campaigns. To prove this, we compute the ratio between the values obtained for two different beaches for both the data; for the numerical simulations we use both the median and both the maximum number of elements in 100 m<sup>2</sup> while for the in situ monitoring measurements we use the total amount of litter found in all the campaigns divided in categories. The categories used for the in situ data are shown below: the first category (from now on called bottles) includes all the item shown in the first image; the second category (from now on called nets) includes the marine litter shown in the second image; the third category (from now on called plastic pieces) includes the items shown in the third image.

Rifiuti: Cod. - Cat. - Descrizione	Spiaggia: Cod. - Nome
(Multiple values) ▼	FVGGP10 - Grado Pineta
bottles ×	
<input checked="" type="checkbox"/> MA7 - Drink bottles <=0.5l	
<input checked="" type="checkbox"/> MA8 - Drink bottles >0.5l	
<input checked="" type="checkbox"/> MA9 - Cleaner/ cleanser bottles & containers	
<input checked="" type="checkbox"/> MA11 - Beach use related cosmetic bottles and containers, e.g. Sunblocks	
<input checked="" type="checkbox"/> MA12 - cosmetics bottles & containers	
<input checked="" type="checkbox"/> MA13 - other bottles & containers (drums)	
<input type="checkbox"/> MA14 - Engine oil bottles & containers <50 cm	
<input type="checkbox"/> MA15 - Engine oil bottles & containers > 50 cm	
<input checked="" type="checkbox"/> MA200 - Bottles incl. Pieces	

Rifiuti: Cod. - Cat. - Descrizione	Spiaggia: Cod. - Nome
(None) ▼	SBA_VEN_03 - Verde
nets ×	
<input checked="" type="checkbox"/> MA45 - Mussels nets, Oyster nets	
<input checked="" type="checkbox"/> MA53 - Nets and pieces of net < 50 cm	
<input checked="" type="checkbox"/> MA54 - Nets and pieces of net > 50 cm	
<input checked="" type="checkbox"/> MA56 - Tangled nets/cord	
<input checked="" type="checkbox"/> MA62 - Floats for fishing nets	
<input checked="" type="checkbox"/> MA142 - Rope, string and nets	

<input checked="" type="checkbox"/> MA75 - Plastic/polystyrene pieces 0-2.5 cm
<input checked="" type="checkbox"/> MA76 - Plastic/polystyrene pieces 2.5> <50 cm
<input checked="" type="checkbox"/> MA77 - Plastic/polystyrene pieces > 50 cm
<input checked="" type="checkbox"/> MA78 - Plastic pieces 0-2.5 cm
<input checked="" type="checkbox"/> MA79 - Plastic pieces 2.5 cm > < 50cm
<input checked="" type="checkbox"/> MA80 - Plastic pieces > 50 cm
<input type="checkbox"/> MA89 - Plastic construction waste
<input type="checkbox"/> MA90 - Plastic flower pots
<input checked="" type="checkbox"/> MA124 - 12 plastic/polystyrene items (identifiable)
<input type="checkbox"/> MA359 - Plastic, latex or vinyl disposable medical gloves

The data is shown in the table below:

Beach	Numerical simulation: Median # elements 100 m <sup>2</sup>	Numerical simulation: Maximum # elements 100 m <sup>2</sup>	Total amount of bottles found:	Total amount of nets found:	Total amount of plastic pieces found:
Grado Pineta	0.76	1.30	48	46	183
Isola Verde	1.45	3.31	52	353	1169
Opcina Umag	0.09	0.11	126	160	869
Trstenik	0.01	0.01	665	35	765

The following ratios have been computed and they are shown in the table below:

Ratios	<i>Grado Pineta</i> <i>Isola Verde</i>	<i>Trstenik</i> <i>Opcina Umag</i>
$\frac{\text{Median \# elements } 100 \text{ m}^2_{\text{beach1}}}{\text{Median \# elements } 100 \text{ m}^2_{\text{beach2}}}$	0.52	0.11
$\frac{\text{Maximum \# elements } 100 \text{ m}^2_{\text{beach1}}}{\text{Maximum \# elements } 100 \text{ m}^2_{\text{beach2}}}$	0.39	0.09
$\frac{\text{Bottles found}_{\text{beach1}}}{\text{Bottles found}_{\text{beach2}}}$	0.92	5.28
$\frac{\text{Nets found}_{\text{beach1}}}{\text{Nets found}_{\text{beach2}}}$	0.13	0.22
$\frac{\text{Plastic pieces found}_{\text{beach1}}}{\text{Plastic pieces found}_{\text{beach2}}}$	0.16	0.88

As it can be noticed from the results shown in the last table, the values obtained from the simulations are not linearly related to the ones obtained in the in situ campaigns measurements. Besides approximations adopted in the simulation approach and the uncertainty characterizing the remote sensing imagery, the in situ measurements too are affected by too many variables: for example, the beaches of Trstenik and Opcina Umag are not inside the Marine Strategy so they had never been cleaned-up by the colleagues before the MARLESS project, on contrary Grado Pineta and Isola Verde beaches are inside the Marine Strategy (so they have been cleaned more often and the marine litter are removed more often). In conclusion, we do not have enough information to compare the in situ data measurements and the simulation results.

# CONCLUSIONS

The method defined and developed thanks to the MARLESS project to generate information to create maps of probability shows a great potential, in spite of the uncertainties that are implicit in both the numerical simulation and the satellite data approaches.

The method to integrate information of potential hotspots, that are coming from very different sources of information, has the capability to reveal also the level of correlation between the indexes which are summarizing the results coming from each of the approach.

First, its extensive application to all the beaches along the Adriatic coasts is just a matter of computation, since satellite information covers the whole cooperation area and the computational time, which was spent to simulate trajectories and back trajectories of marine litter pieces, can be tackled with medium size computational infrastructures. All of these warranty the transferability of the MARLESS method to further explorations, also in other marine basins.

Second, the improvement of the identification power of the indexes used in the hotspot identification, can proceed independently on both of them, resulting in a reduction of the uncertainty in the hotspot identification. Furthermore, the normalization of the indexes, thanks to the in situ marine litter monitoring, adds the value to the interpretation of the congruence of the hotspot identification by means of the two independently approaches, because of the straightforward interpretation of the points representing the beaches on the hotspot identification plane.

The reduction of uncertainty, in the identification of the hotspots, is foreseen coming from the new generation of hyperspectral satellite, that are going to bring the required details in the spectral information associated to each image pixel, together with the increase of spatial resolution. From the numerical model simulations, the reduction of hotspot index is expected to come from a better representation of beaching and refloating parametrization in addition to a refinement identification of the potential sources, which are responsible of marine litter input in the sea.

In the end, we can say that the application of this method to rank the beaches according to the hotspot indicator leads naturally to the creation of probability maps of marine litter hotspots.

# BIBLIOGRAPHY

- [1] Martínez-Vicente, V.; Clark, J.R.; Corradi, P.; Aliani, S.; Arias, M.; Bochow, M.; Bonnery, G.; Cole, M.; Cózar, A.; Donnelly, R.; et al. Measuring Marine Plastic Debris from Space: Initial Assessment of Observation Requirements. *Remote Sens.* 2019, *11*, 2443.
- [2] Biermann, L., Clewley, D., Martinez-Vicente, V. *et al.* Finding Plastic Patches in Coastal Waters using Optical Satellite Data. *Sci Rep* 10, 5364 (2020).
- [3] Maximenko, N.; Corradi, P.; Law, K.L.; Van Sebille, E.; Garaba, S.P.; Lampitt, R.S.; Galgani, F.; Martinez-Vicente, V.; Goddijn-Murphy, L.; Veiga, J.M. Toward the integrated marine debris observing system. *Front. Mar. Sci.* 2019, *6*, 1–25
- [4] <https://sentinel.esa.int/web/sentinel/user-guides/sentinel-2-msi/revisit-coverage>
- [5] <https://www.copernicus.eu/en>
- [6] [https://www.esa.int/Applications/Observing\\_the\\_Earth/Copernicus/The\\_Sentinel\\_missions](https://www.esa.int/Applications/Observing_the_Earth/Copernicus/The_Sentinel_missions)
- [7] <https://sentinel.esa.int/web/sentinel/missions/sentinel-2/instrument-payload/resolution-and-swath>
- [8] [https://esamultimedia.esa.int/docs/EarthObservation/Sentinel-2\\_ESA\\_Bulletin161.pdf](https://esamultimedia.esa.int/docs/EarthObservation/Sentinel-2_ESA_Bulletin161.pdf)
- [9] <https://sentinel.esa.int/web/sentinel/technical-guides/sentinel-2-msi/level-2a/product-formatting>
- [10] GARABA, S. P. & DIERSSEN, H. M. 2018. An airborne remote sensing case study of synthetic hydrocarbon detection using short wave infrared absorption features identified from marine-harvested macro-and microplastics, *Remote sensing of environment* 205: 224 – 235.
- [11] Konstantinos Topouzelis, Apostolos Papakonstantinou, Shungudzemwoyo P. Garaba, Detection of floating plastics from satellite and unmanned aerial systems (Plastic Litter Project 2018), *International Journal of Applied Earth Observation and Geoinformation*, Volume 79, 2019, Pages 175-183, ISSN 1569-8432.
- [12] Topouzelis, K.; Papageorgiou, D.; Karagaitanakis, A.; Papakonstantinou, A.; Arias Ballesteros, M. Remote Sensing of Sea Surface Artificial Floating Plastic Targets with Sentinel-2 and Unmanned Aerial Systems (Plastic Litter Project 2019). *Remote Sens.* **2020**, *12*, 2013.

- [13] Themistocleous, K.; Papoutsas, C.; Michaelides, S.; Hadjimitsis, D. Investigating Detection of Floating Plastic Litter from Space Using Sentinel-2 Imagery. *Remote Sens.* 2020, *12*, 2648.
- [14][http://www.cleanatlantic.eu/wp-content/uploads/2022/02/CA\\_WP5\\_5.2\\_Analysis-\\_of-\\_strategies-\\_for-\\_the-\\_monitoring-\\_and-\\_evaluation-\\_of-\\_accumulations-\\_of-marine-\\_litter-\\_on-\\_the-\\_coast\\_EN.pdf](http://www.cleanatlantic.eu/wp-content/uploads/2022/02/CA_WP5_5.2_Analysis-_of-_strategies-_for-_the-_monitoring-_and-_evaluation-_of-_accumulations-_of-marine-_litter-_on-_the-_coast_EN.pdf)
- [15] Kikaki, A., Karantzalos, K., Power, C.A., & Raitsos, D.E. (2020). Remotely Sensing the Source and Transport of Marine Plastic Debris in Bay Islands of Honduras (Caribbean Sea). *Remote. Sens.*, *12*, 1727.
- [16] Ciappa AC. Marine plastic litter detection offshore Hawai'i by Sentinel-2. *Mar Pollut Bull.* 2021 Jul;168:112457.
- [17] Biermann, L., Clewley, D., Martinez-Vicente, V. *et al.* Finding Plastic Patches in Coastal Waters using Optical Satellite Data. *Sci Rep* 10, 5364 (2020).
- [18] Konstantinos Topouzelis, Dimitris Papageorgiou, Giuseppe Suaria, Stefano Aliani. Floating marine litter detection algorithms and techniques using optical remote sensing data: A review. *Marine Pollution Bulletin*, Volume 170, 2021, 112675, ISSN 0025-326X.
- [19] <https://odnature.naturalsciences.be/remsem/acolite-forum/viewforum.php?f=4&sid=7f4c913a51c187b4bd60da6a0b2a53b6>
- [20] Quinten Vanhellemont, Kevin Ruddick. Atmospheric correction of metre-scale optical satellite data for inland and coastal water applications. *Remote Sensing of Environment*, Volume 216, 2018, Pages 586-597, ISSN 0034-4257.
- [21] [https://odnature.naturalsciences.be/downloads/remsem/acolite/acolite\\_manual\\_20210802.0.pdf](https://odnature.naturalsciences.be/downloads/remsem/acolite/acolite_manual_20210802.0.pdf)
- [22] [https://www.indexdatabase.de/db/is.php?sensor\\_id=96](https://www.indexdatabase.de/db/is.php?sensor_id=96)
- [23] N. Carlson, David A. Ripley, On the relation between NDVI, fractional vegetation cover, and leaf area index, *Remote Sensing of Environment*, 1997.
- [24] McFeeters, S.K. (1996). The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. *International Journal of Remote Sensing*, *17*, 1425-1432.
- [25] <https://www.indexdatabase.de/db/i-single.php?id=345>
- [26] [https://scikit-learn.org/stable/user\\_guide.html](https://scikit-learn.org/stable/user_guide.html)

- [27] <https://scikit-learn.org/stable/modules/generated/sklearn.cluster.KMeans.html>
- [28] <https://scikit-learn.org/stable/modules/clustering.html#hierarchical-clustering>
- [29] <https://landsat.gsfc.nasa.gov/satellites/landsat-8/>
- [30] <https://earth.esa.int/eogateway/missions/planetscope#instruments-section>
- [31] <https://sentinels.copernicus.eu/web/sentinel/technical-guides/sentinel-2-msi/msi-instrument>
- [32] <https://earth.esa.int/eogateway/missions/worldview-3>
- [33] <https://directory.eoportal.org/web/eoportal/satellite-missions/p/prisma-hyperspectral>