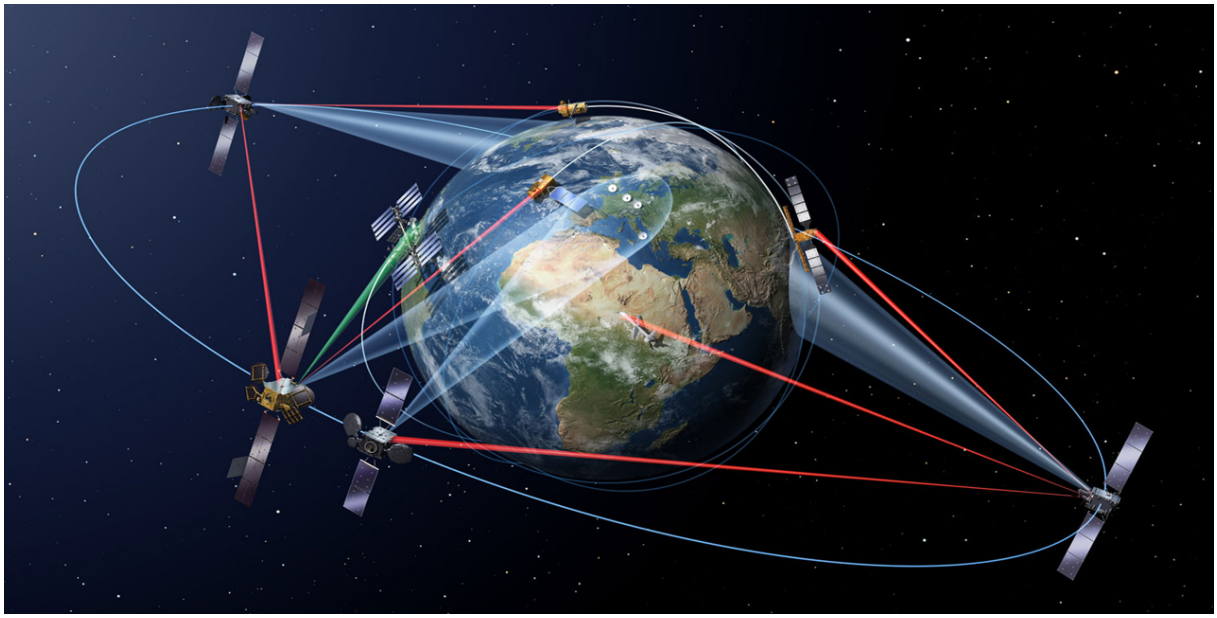
Ph.D. Program in Civil, Chemical and Environmental Engineering,
Curriculum in Fluid Dynamics and Environmental Engineering

Department of Civil, Chemical and Environmental Engineering
Polytechnic School, University of Genoa, Italy.



Enhancing the usability of Satellite Earth Observations through Data Driven Models. An application to Sea Water Quality.

Candidate: Stefania Magrì

Cover image from Copernicus website

Enhancing the usability of Satellite Earth Observations through Data Driven Models.

An application to Sea Water Quality.

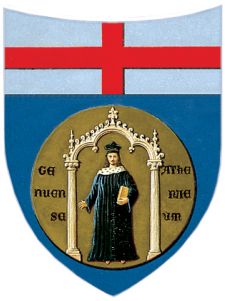
BY

Stefania Magrì

Dissertation discussed in partial fulfillment of
the requirements for the Degree of

DOCTOR OF PHILOSOPHY

Civil, Chemical and Environmental Engineering
curriculum in Fluid Dynamics and Environmental Engineering,
Department of Civil, Chemical and Environmental Engineering, University of Genoa,
Italy



30 October, 2023

Supervisors:
Prof. Giovanni Besio
Prof. Bruno Fabiano
Prof. Bianca Federici

Coordinator of the PhD program:
Prof. Rodolfo Repetto

External Reviewers:
Prof. Sonia Silvestri - Department of Biological, Geological, and Environmental Sciences, University of Bologna
Prof. Claudio Parente - Departiment of Science and Technology, University Parthenope of Naples

Examination Committee:
Prof. Francesco Pirotti - Department of Land, Environment, Agriculture and Forestry, Università di Padova
Prof. Giuseppe Ciraolo - Department of Enigneering, Università degli Studi di Palermo
Prof. Nicoletta Tambroni - Department of Civil, Chemical and Environmental Engineering, Università di Genova

Ph.D. program in Civil, Chemical and Environmental Engineering
Curriculum in Fluid Dynamics and Environmental Engineering
Cycle XXXV

Acknowledgements

I would like to thank my thesis advisors, who made the realisation of my doctoral studies possible: Prof. Bruno Fabiano, who encouraged me to take this path, Prof. Giovanni Besio who once again accepted to supervise my research work, and Prof. Bianca Federici who guided me with her expertise and enthusiasm.
I would like to extend my gratitude to all the people at DICCA that I have had the opportunity to meet over these last three years, who in one way or another have influenced and helped me during this period, with their passion, hard work and positive vibes. I hope that life will make our professional paths cross again.
I would like to express my sincere gratitude to my office managers in ARPAL, Rosella Bertolotto and Fabrizia Colonna, who have always taken an interest in my research and supported my decisions.
I am also grateful to Prof. Ennio Ottaviani and Enrico Prampolini, for their valuable feedback and discussions, which have significantly improved the quality of this thesis.
My successes are largely due to the quiet support of my family. A special thank you to Tomaso, for being there for me throughout my Ph.D journey, and for all the insightful discussions, your dedication and enthusiasm are a constant source of inspiration.

ABSTRACT

Earth Observation from satellites has the potential to provide comprehensive, rapid and inexpensive information about land and water bodies. Marine monitoring could gain in effectiveness if integrated with approaches that are able to collect data from wide geographic areas, such as satellite observation.
Integrated with in situ measurements, satellite observations enable to extend the punctual information of sampling campaigns to a synoptic view, increase the spatial and temporal coverage, and thus increase the representativeness of the natural diversity of the monitored water bodies, their inter-annual variability and water quality trends, providing information to support EU Member States' action plans.

Turbidity is one of the optically active water quality parameters that can be derived from satellite data, and is one of the environmental indicator considered by EU directives monitoring programmes. Turbidity is a visual property of water, related to the amount of light scattered by particles in water, and it can act as simple and convenient indirect measure of the concentration of suspended solids and other particulate material.
A review of the state-of-the-art shows that most traditional methods to estimate turbidity from optical satellite images are based on semi-empirical models relying on few spectral bands. The choice of the most suitable bands to be used is often site and season specific, as it is related to the type and concentration of suspended particles.
When investigating wide areas or long time series that include different optical water types, the application of machine learning algorithms seems to be promising due to their flexibility, responding to the need of a model that can adapt to varying water conditions with smooth transition, and their ability to exploit the wealth of spectral information. Moreover, machine learning models have shown to be less affected by atmospheric and other background factors. Atmospheric correction for water leaving reflectance, in fact, still remains one of the major challenges in aquatic remote sensing.
The use of machine learning for remotely sensed water quality estimation has spread in recent years thanks to the advances in algorithm development, computing power, and availability of higher spatial resolution data. Among all existing algorithms, the choice of the complexity of the model derives from the nature and number of available data.
The present study explores the use of Sentinel-2 MultiSpectral Instrument (MSI) Level-1C Top of Atmosphere spectral radiance to derive water turbidity, through application of a Polynomial Kernel Regularized Least Squares regression. This algorithms is characterized by a simple model structure, good generalization, global optimal solution, especially suitable for non-linear and high dimension problems.
The study area is located in the North Tyrrhenian Sea (Italy), covering a coastline of about 100 km, characterized by a varied shoreline, embracing environments worthy of protection and valuable biodiversity, but also relevant ports, and three main river flow and sediment discharge.
The coastal environment in this area has been monitored since 2001, according to the 2000/60/EC Water Framework Directive, and in 2008 EU Marine Strategy Framework Directive 2008/56/EC further strengthened the investigation in the area.
A dataset of combination of turbidity measurements, expressed in nephelometric turbidity units (NTU), and values of the 13 spectral bands in the pixel corresponding to the sample

location was used to calibrate and validate the model.
The developed turbidity model shows good agreement of the estimated satellite-derived surface turbidity with the measured one, confirming that the use of ML techniques allows to reach a good accuracy in turbidity estimation from satellite Top of Atmosphere reflectance.
Comparison between turbidity estimates obtained from the model with turbidity data from Copernicus CMEMS dataset named 'Mediterranean Sea, Bio-Geo-Chemical, L3, daily observation', which was used as benchmark, produced consistent results.
A band importance analysis revealed the contribution of the different spectral bands and the main role of the red-edge range.
Finally, turbidity maps from satellite imagery were produced for the study area, showing the ability of the model to catch extreme events and, overall, how it represents an important tool to improve our understanding of the complex factors that influence water quality in our oceans.

INDEX

List of Abbreviations

The following table describes the meaning of the various abbreviations and acronyms used throughout the document. The page on which each one is defined or first used is also given.

Abbreviation	**Meaning**	Page
AC	Atmospheric Correction	40
AERONET	Aerosol Robotic Network	43
AOD	Aerosol Optical Depth	40
AOP	Apparent Optical Properties	9
BD	Birds Directive	2
BOA	Bottom Of Atmosphere	38
CDOM	Color Dissolved Organic Matter	10
CFP	Common Fisheries Policy	2
CMEMS	Copernicus Marine Environment Monitoring Service	27
EMR	Electromagnetic radiation	4
EMS	Electromagnetic Spectrum	4
EOS	Earth Observing System	11
EPA	Environmental Protection Agency	31
EQS	Environmental Quality Standards	2
ESA	European Space Agency	4
ETM+	Enhanced Thematic Mapper Plus	25
EUMETSAT	European Organisation for the Exploitation of Meteorological Satellites	12
FNU	Formazin Nephelometric Units	31
GEO	Geostationary Orbit	5
GES	Good Environmental Status	2
GSO	Geosynchronous Orbit	5
HD	Habitats Directive	2
HEO	High-Earth Orbit	5
HH	Hanna and Henold index	56
IOP	Inherent Optical Properties	9
ISO	International Organization for Standardization	31
KRLS	Kernel Kernel Regularized Least Squares	53
LEO	Low-Earth Orbit	4
LOOCV	Leave-One-Out Cross Validation	50

Abbreviation	**Meaning**	Page
MAE	Mean Absolute Error	56
MEO	Medium-Earth Orbit	5
ML	Machine Learning	2
MODIS	Moderate Resolution Imaging Spectroradiometer	11
MSFD	Marine Strategy Framework Directive	2
MSI	MultiSpectral Instrument	8
NRMSE	Normalized Root Mean Square Error	61
NAP	Non-Algal Pigments	10
NASA	National Aeronautics and Space Administration	4
NTU	Nephelometric Turbidity Units	31
OLCI	Ocean and Land Color Instrument	13
OLI	Operational Land Imager	25
OSC	Optically Significant Constituents	10
RMSE	Root mean squared error	56
RS	Remote Sensing	4
RSC	Regional Seas Conventions	2
SAR	Synthetic Aperture Radar	3
SLSTR	Sea and Land Surface Temperature Radiometer	13
SRAL	Synthetic Aperture Radar Altimeter	13
SSO	Sun-synchronous orbit	5
SST	Sea Surface Temperature	3
TCI	True Color Image	8
TM	Thematic Mapper	25
TOA	Top Of Atmosphere	38
TSM	Total Suspended Matter	10
USGS	U.S. Geological Survey	4
WFD	Water Framework Directive	2

INTRODUCTION

Water quality is an essential climate variable and knowledge about its state and dynamics is vital. Remote sensing instruments mounted on satellites are currently our only way to obtain frequent observations of sea water surface. Satellite observations from the European Union's Copernicus programme are freely and openly available and, once appropriately processed, provide regular, comprehensive coverage of wide water bodies.
Satellite optical sensors can be used to monitor the quality of sea water, but interpreting the vast amounts of data generated by these sensors can be challenging. Machine learning techniques can be extremely useful in processing satellite images by automating the processing, creating visualizations, through thematic maps, that make it easier to interpret the data and identify patterns and trends, e.g. time series visualizations can show changes over time. Overall, machine learning can help to improve our understanding of the complex factors that influence water quality in our oceans.
The main goal of the present work is to answer the following open research question:
Can optical satellite data support sea water quality monitoring in a sufficiently accurate way, and how can Machine Learning contribute to the scope?

The first Chapter, or Introduction, is meant as a reference chapter that essentially presents general information about the three main blocks, or topics, that constitute the work, in order to provide a common background knowledge that will be used later in the document: (i) a description of the Marine Environment Monitoring Directives, (ii) an overview of Remote Sensing technology, and (iii) an account on Machine Learning techniques.
The second Chapter will introduce and describe the objective of the research, together with an outlook on the state-of-the art.
Chapter 3 will present the proposed methodology and its application to a case study: Chapter 3.1 defines sea water turbidity and describes its role in water quality assessment, Chapter 3.2 presents the study area in the North Thyrrenian Sea, and relative in situ turbidity data acquisition; Chapter 3.3 describes satellite data used, with an insight on the issue of atmospheric correction and cloud detection. The proposed model is presented in detail in Chapter 3.4, from the choice of the model to its implementation and application.
Finally in Chapter 4 results and their implications will be presented and discussed. In Chapter 5 some conclusions are withdrawn and recommendations on further activities presented.

1.1. Monitoring of the Marine Environment

Oceans cover over seventy per cent of the Earth's surface, but they are probably the least understood, most biologically diverse, and most undervalued of all ecosystems. The ocean and marine environment plays a vital role in regulating the planet's climate and provides a significant source of goods and services, such as food, fossil fuels, construction materials, transport and recreation for millions of people worldwide. It also hosts an invaluable biodiversity that forms complex ecosystems which are threatened by human activities such as agriculture, fisheries, aquaculture, shipping, urbanisation and tourism. Pollutants from anthropogenic activities include plastic, oil, toxic chemicals, radioactive waste, domestic and industrial sewage, excessive nutrients from agricultural fertilizers and pesticides (Islam and Tanaka, 2004).
The quality of seawater is essential for humanity, and monitoring the marine environment is crucial to ensure that it remains safe and healthy. By doing so, we can protect the health of marine life, support the economy, and better understand the impacts of climate change on the ocean. Knowledge of the pollution sources and impacts is important not only for a better understanding on the ecosystem responses but also to formulate prevention measures.

Increasing concerns about pollution levels in the oceans and coastal regions have led to establishment and implementation of multiple monitoring programmes efforts for assessing the environmental status of marine waters, in order to implement mitigation strategies and achieve sustainable water quality. To this aim, the Water Framework Directive (WFD, 2000/60/EC and amendments) and European Union Marine Strategy Framework Directive (MSFD, 2008/56/EC) require Member States to monitor the quality of the marine environment, based on the status of several elements, and to take measures to achieve and maintain a Good Environmental Status (GES) of all marine waters by 2020.
The monitoring programmes shall integrate and complement the monitoring requirements imposed by other EU legislation, such as the Habitats Directive (HD, 92/43/EEC) and Birds Directive (BD, 2009/147/EC), the Environmental Quality Standards Directive (EQS, 2008/105/EC), the Data Collection Framework Regulation for the Common Fisheries Policy (CFP) and other international agreements, such as the Regional Seas Conventions (RSCs). Consistency, coherence and comparability within marine regions and sub regions should be ensured by coordination of monitoring methods, taking also into account trans-boundary features and impacts (Zampoukas et al., 2013).
A monitoring programme can be considered integrated when it provides data relevant to different indicators, to different pieces of legislation, for more than one Member State and collected in comparable way. However, according to OSPAR (OSPAR Commission, 2008), monitoring activity efforts would benefit in efficiency and cost effectiveness from better coordination and data sharing among the different programmes and the involved institutes.
Considering that the marine territory of the EU is larger than its land territory, a considerable effort is needed to fulfill this legislative requirement in a meaningful and pragmatic manner. Marine monitoring could gain in effectiveness if integrated with approaches that are able to collect data from wide geographic areas, where only few observations can be

conducted by traditional methods using oceanographic vessels (Zampoukas et al., 2012). Remote sensing from satellites offers observations of the sea surface over large areas in a cost-effective way, as an always increasing amount of data can be freely accessed.
Confidence in ecological status assessment would improve, with increased representativeness of the natural diversity of waterbodies that are monitored, their inter-annual variability and water quality trends within larger waterbodies. Moreover, using standardised approaches, it would allow better comparison and standardisation of water quality assessment across Member States, facilitating the management of transboundary waters in particular.
Finally, by increasing spatial and temporal coverage, satellite observation is expected to enhance the effectiveness of the EU Member States' action plans through early detection of deterioration, improving knowledge of the potential extent of management impacts, improving monitoring of the effectiveness of action plans and providing information to support more strategic, in situ sampling.

Recent developments in sensor technologies have turned remote sensing into an effective means of monitoring marine areas. Different remote sensing sensors have their own capabilities for mapping and monitoring water pollution of different types, characteristics, and concentrations, such as oil pollutants, suspended solid concentrations, algal blooms, and floating plastic waste in marine waters (Hafeez et al., 2018).
Scanning radiometers and microwave sounders measure Sea Surface Temperature (SST) data, altimeters collect wave height data, and synthetic aperture radar (SAR) can measure the sea surface roughness patterns, that can be related to the presence of oil pollutants or plastic debris. Optical sensors can determine optical water quality variables such as chlorophyll-a, diffuse attenuation coefficient, ocean color. The retrieval of optical water quality information from satellites, such as chlorophyll, total suspended matter, turbidity, pigmented fraction of dissolved organic matter and phytoplankton functional groups, has gained recognition as a valid integrative tool for monitoring water bodies with the availability of new imagery from multispectral sensors satellites with a higher spatial, spectral, and temporal resolution, like those of Landsat and Sentinel.
The following chapter will present and discuss characteristics of satellite remote sensing techniques, with a particular focus on optical sensors.

1.2. Satellite Remote Sensing

1.2.1. Satellite Remote Sensing - Basic Principles

Satellite Remote Sensing (RS) can be defined as the technique to acquire information about the Earth's surface and atmosphere from out of space using sensors onboard satellites (Lillesand and Kiefer, 1993).
Satellite RS records the electromagnetic energy reflected or emitted by the atmosphere and the Earth's surface. The amount of radiation from an object (called radiance) is influenced by both the properties of the object (physical, chemical) and the radiation hitting the object (irradiance). Depending on the source of energy used for sensing, we can distinguish between passive and active sensors (Weng, 2013). The first ones depend upon an external energy source, generally the sun, while for the second ones the energy is provided by the platform itself. Since passive sensors exploit sun energy, they can detect data only during daylight, while active sensors can acquire information both at day and night. Sensors characteristics determine the wavelength regions of the electromagnetic spectrum (EMS) that can be detected (Figure 1.1).
In the microwave region of the spectrum, the sensor is radar, which is an active sensor. The electromagnetic radiation (EMR) produced by the radar is transmitted to the earth's surface and the EMR reflected (back scattered) from the surface is recorded and analyzed. The microwave region can also be monitored with passive sensors, called microwave radiometers, which record the radiation emitted by the terrain in the microwave region.
Optical sensors record reflected solar radiation from the visible (380 to 700 nanometers) to the mid-infrared regions of the EMS. Optical sensors cannot acquire information of the Earth's surface in presence of clouds, while the majority of active sensors operate in the microwave band of the EMS, which gives them the ability to penetrate the atmosphere under most conditions, including clouds and rain.
Thermal infrared systems are passive sensors that measure the energy emitted by the earth surface itself in the thermal infrared wavelength region.
Hereby, the main characteristics of Satellite Remote Sensing are described, with particular reference to satellites carrying optical sensors. Most of information reported are based on the official NASA (National Aeronautics and Space Administration) [1], USGS (U.S. Geological Survey)[2], European Space Agency (ESA)[3] and Copernicus Programme[4] web sites.

Satellite Orbits

Satellites can be placed in several types of orbits around Earth, depending on what the satellite is designed to achieve (Figure 1.2). Three common classes of orbits are:

- Low-Earth Orbit (LEO), as the name suggests, this orbit is relatively close to Earth's surface (approximately 160 to 2'000 km above Earth). Due to the vicinity to the

[1] https://www.earthdata.nasa.gov; https://science.nasa.gov
[2] https://www.usgs.gov/landsat-missions
[3] https://sentinel.esa.int/
[4] https://sentinels.copernicus.eu

Earth surface, they offer images at higher resolution. LEO satellites do not always have to follow a particular path around Earth in the same way and their plane can be tilted. This means there are more available routes for satellites in LEO, which is one of the reasons why LEO is a very commonly used orbit.

- Medium-Earth Orbit (MEO) comprises a wide range of orbits from 2'000 to 35'500 km above Earth. It does not need to take specific paths around Earth. It takes approximately 12 hours to complete an orbit, thus in 24-hours, the satellite crosses over the same two spots on the equator every day. This orbit is consistent and highly predictable; as a result, this is an orbit used by many telecommunications and GPS satellites.
- High-Earth Orbit (HEO), above 35'500 km above Earth. Satellites orbiting at 35'786 km are at an altitude at which their orbital speed matches the planet's rotation, and are in what is called geosynchronous orbit (GSO). A satellite in GSO directly over the equator, circling the Earth from west to east travelling at the exact same rate as Earth, will have a geostationary orbit (GEO), that is, it will maintain its position directly above one particular place over Earth, which makes it appear to be 'stationary' over a fixed position. GEO is used by satellites that need to stay constantly over the same position, such as telecommunication satellites.
  While both geosynchronous and geostationary satellites orbit at 35'786 km above Earth, geosynchronous satellites have orbits that can be tilted above or below the equator. Geostationary satellites, on the other hand, orbit Earth on the same plane as the equator.

A particular type of LEO that travels past Earth from North to South rather than from West to East, passing roughly over Earth's poles, are the Polar orbits.
Sun-synchronous orbit (SSO) is a particular kind of polar orbit, synchronous with the Sun.

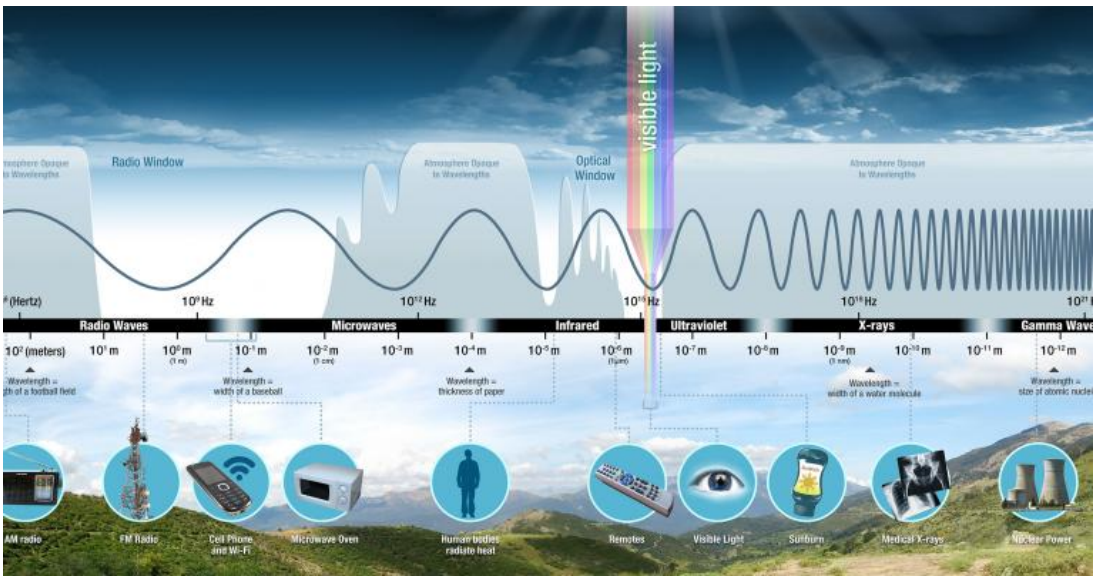


Figure 1.1: Diagram of the Electromagnetic Spectrum (Source: NASA website).

This means they are synchronised to always be in the same 'fixed' position relative to the Sun. This means that the satellite always visits the same spot at the same local time. This type of orbit serves a number of applications, for example, to capture satellite images used to monitor changes over time of a specific area. A satellite in a Sun-synchronous orbit would usually be at an altitude of between 600 to 800 km.

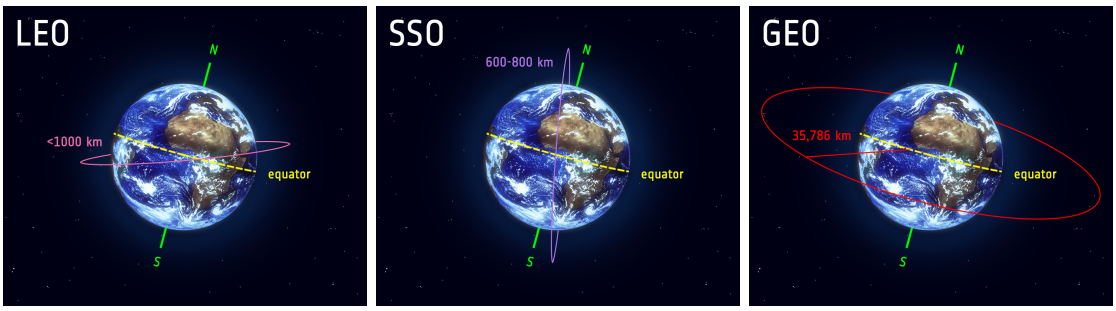


Figure 1.2: Different types of orbites (from left to right): Low-Earth Orbit (LEO), LEO Sun-synchronous orbit (SSO), High-Earth Goestationary Orbit (GEO) (Source: ESA website).

Resolution

Resolution plays a key role in how data from a sensor can be used. There are four types of resolution to consider, that can vary depending on the satellite orbit and sensor design: spatial, spectral, radiometric, and temporal:

- spatial resolution is defined by the size of each pixel within a digital image and the area on Earth's surface represented by that pixel. It is related to the discernible detail you can see on the image.
- spectral resolution refers to the number of sensed bands (i.e. intervals of the EMS for which the average reflected energy is measured) and the width of the wavelength range (Figure 1.3). Sensors can be classified as multispectral (from 3 to 16 bands) or hyperspectral (more that 16 bands).
- radiometric resolution refers to the ability of the sensor to discriminate between slight differences in sensed energy. It is described by the number of bits representing the energy recorded in each pixel. For example, an 8 bit resolution indicates that the sensor has 2^8 potential digital values (0-255) to store information.
- temporal resolution refers to the time necessary for the sensor to revisit the same observation area. It depends on the satellite orbit and the swath width.

Optimisation of one type of resolution requires a trade-off with the other types, for example, to acquire observations with high spatial resolution, a narrower swath is required, which requires more time between observations of a given area, resulting in a lower temporal resolution. It is thus very important to understand the type of data needed for the specific application of interest.

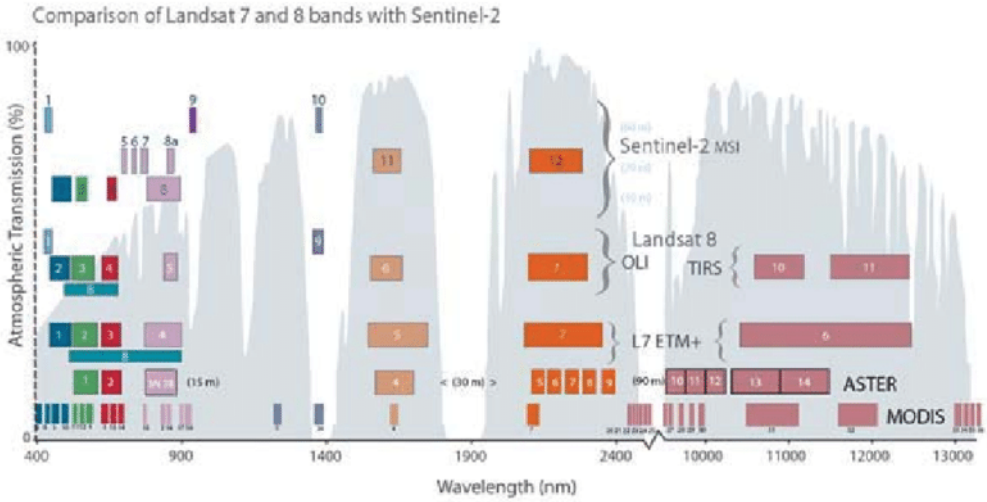


Figure 1.3: Comparison of AQUA MODIS, Landsat7 ETM+, Landsat 8 OLI and Sentinel-2 MSI spectral resolution (Source: U.S.G.S.

Satellite Images

The output of a RS system is a set of images representing the observed scene (Weng, 2013). A satellite digital image is a raster file, a regular grid of columns and rows whose cells are known as pixels.
The digital number associated to each pixel represents the intensity of the energy received from the corresponding portion of the surface, in a given wavelength interval. A satellite product is thus composed of a number of raster images equal to the number of the sensed bands, where each pixel can assume a range of values according to the sensor radiometric resolution (Figure 1.4).

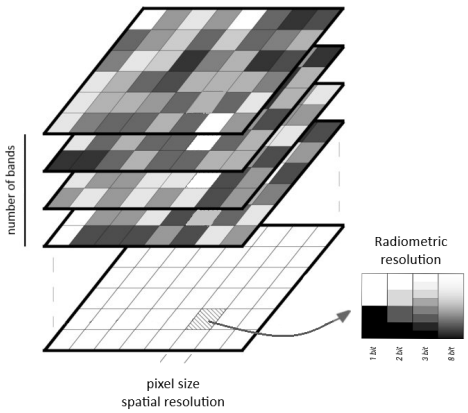


Figure 1.4: Characteristics of a satellite raster image.

Remote sensing data acquired from instruments aboard satellites require processing before the data are usable. Pre-processing applied is related to geometric and radiometric effects. Geometric correction is the process of removing geometric distortions from an

image caused by the satellite position, sensor orientation, and the terrain below. These distortions can cause errors in the measurement of distances, areas, and angles in the image. Geometric correction involves matching control points in the satellite image with corresponding points on a reference map or image, and then using a transformation model to rectify the image. Most geometric errors are corrected by the ground segment of the satellite mission before data distribution, often including the registration of the image to a map projection and coordinate system. The radiometric correction, instead, consists in removing the effects of sensor-related distortions, and to convert the digital number values to physically meaningful measurements of radiance values.
Further processing can be done, for example removing the effects of the atmosphere (for more details see Chapter 3.3.2.) and data of different bands can be combined for imagery interpretation. An example is given in Figure 1.5 where a True Color Image (TCI) is obtained from Sentinel-2 MultiSpectral Instrument (MSI) Level-1 Top of Atmosphere and Level-2 Bottom-of-Atmosphere products combining bands in the visible range (Blue-Green-Red).

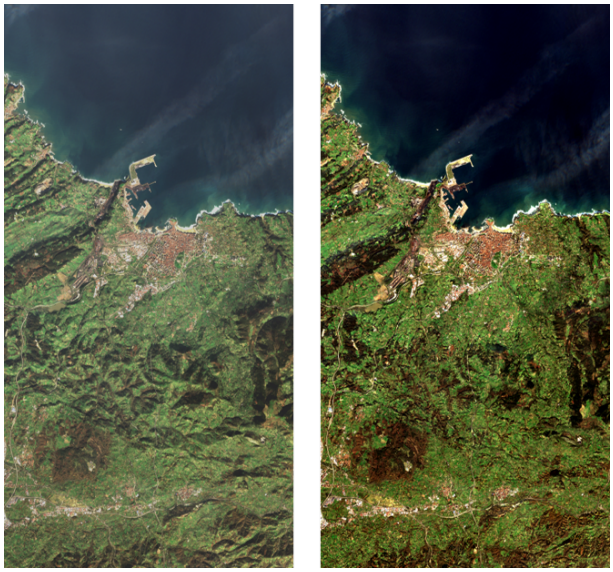


Figure 1.5: A True Color Image obtained from Sentinel-2 MSI Level-1C (left) and Level-2 (right) products.

Spectral Signature

Everything in nature has its own unique distribution of reflected, emitted and absorbed radiation (Figure 1.6). In fact, different objects return different amount of energy in different bands of the electromagnetic spectrum, incident upon it, depending on the property of material (structural, chemical, and physical), surface roughness, angle of incidence, intensity, and wavelength of radiant energy (Moore, 1980).

These spectral characteristics are detected by the remote sensor and if ingeniously exploited, can be used to distinguish one thing from another or to obtain useful information about the object of interest, such as shape, size and other physical and chemical properties (Aggarwal, 2004).

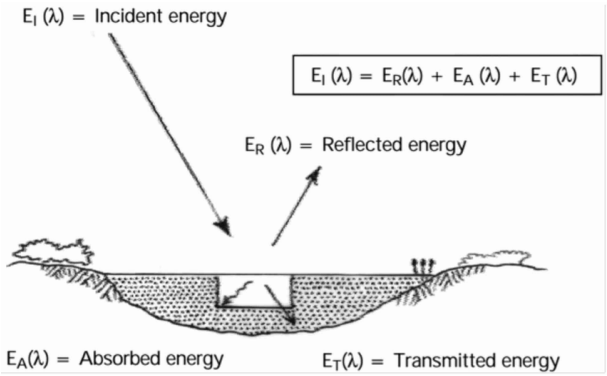


Figure 1.6: Interaction of Energy with the earth's surface (Source: Lillesand and Kiefer, 1993).

The reflectance characteristics of the earth's surface features are expressed by spectral reflectance, $\rho(\lambda)$ or R_{rs}, which is the ratio of reflected energy to incident energy as a function of wavelength:

$$\rho(\lambda) = E_R(\lambda)/E_I(\lambda) \ * 100$$

where:

$\rho(\lambda)$ = spectral reflectance (reflectivity) at a particular wavelength

$E_R(\lambda)$ = Energy of wavelength reflected from object

$E_I(\lambda)$ = Energy of wavelength incident upon the object.

Spectral reflectance is responsible for the color of objects: trees appear green because they reflect more of the green wavelength. The values of the spectral reflectance of objects averaged over different, well-defined wavelength intervals comprise the spectral signature of the objects or features by which they can be distinguished.
The plot between $\rho(\lambda)$ and λ is called a spectral reflectance curve. The spectral response patterns are averaged to get a generalized form, which is called generalized spectral response pattern for the object concerned. Spectral signature is a term used for unique spectral response pattern, which is characteristic of a surface feature.
The number of spectral bands detected by a given sensor, its spectral resolution, determines how much differentiation a researcher can identify between materials.

1.2.2. Remote Sensing of Water

Radiative transfer theory divides the optical properties of water into inherent (IOP) and apparent (AOP). The inherent optical properties depend on the characteristics of the water mass, and are characterised by absorption and scattering coefficients of the water mass for electromagnetic radiation at a specific wavelength. Total absorption is the sum of absorption by phytoplankton (microalgae), non-algal pigments (NAP), Color Dissolved Organic Matter (CDOM), and absorption by water, whereas light scattering is mainly controlled by Total Suspended Matter (TSM).

Variations in IOP change the reflectance of water which is captured by a RS sensor, and this is known as the apparent optical properties (AOP) of water. Indeed, the presence, the concentration and the type of materials and substances in water highly influence reflection, absorption, and transmittance of electromagnetic radiation (Hafeez et al., 2018). The term Ocean Colour refers to the spectral composition of light that is reflected by water bodies. Ocean color data can be used to estimate the concentrations of these optically significant constituents (OSC).

In general, clear waters have low reflectance in the visible region and have no reflectance in the near infrared (NIR) spectrum ($\sim$ 850 nm), as it is absorbed by water. High concentrations of suspended sediments in water block the transmittance from and to lower depths and therefore increases reflectance from the water surface, determining high reflectance measurements in red ($\sim$ 650 nm) and NIR regions. Similarly, chlorophyll (Chl-a) concentrations show a correlation with high reflectance measurements in green band and high absorption in blue and red bands due to photosynthetic activity (Figure 1.7).
Algae absorb visible light for the photosynthesis process and emit excessive energy in the form of fluorescence signal (681 nm, the fluorescence band) when chlorophyll molecule comes to the non-excitation state during the photosynthesis process. A portion of absorbed incident energy by the earth's feature is also re-emitted in the thermal infrared region of the electromagnetic spectrum.
Ocean colour is primarily governed by chlorophyll and any of its accessory pigments in open sea (also known as Case1 waters), while in inland and coastal waters (Case2), the colour is further modulated by the presence of organic and inorganic particles and dissolved matter (Pahlevan et al., 2021).
Most satellite-borne optical sensors provide a spectral resolution covering the visible and NIR spectral regions required for the estimation of biogeochemical material.

1.2.3. Satellite Earth Observation Programmes

On July 1972 the NASA/USGS Landsat 1, formerly known as the Earth Resources Technology Satellite (ERTS), was the first Earth-observing satellite of the United States' Landsat program launched with the express intent to study and monitor our planet's landmasses. To perform the monitoring, Landsat 1 carried two instruments: a camera system built by the Radio Corporation of America (RCA), and the Multispectral Scanner (MSS), built by the Hughes Aircraft Company (El Segundo, CA; NASA contract NAS 5-11255). The MSS recorded data in four spectral bands: a green, red, and two infrared bands. Landsat 1 was designed to provide high-resolution images of the Earth's surface, which could be used for a wide range of applications, including mapping, land-use planning, resource management, and environmental monitoring.
Since then, the Landsat program has achieved to launch other 8 satellite missions, with Landsat 9 being launched on September 2021 (Figure 1.8) and a new mission, Landsat Next, being currently in its early phase, with the aim of continuing the existing decades-long data record. Landsat is the longest-running Earth observation mission to date.
Landsat missions provide images at a high spatial resolution, of 30 meters for most of

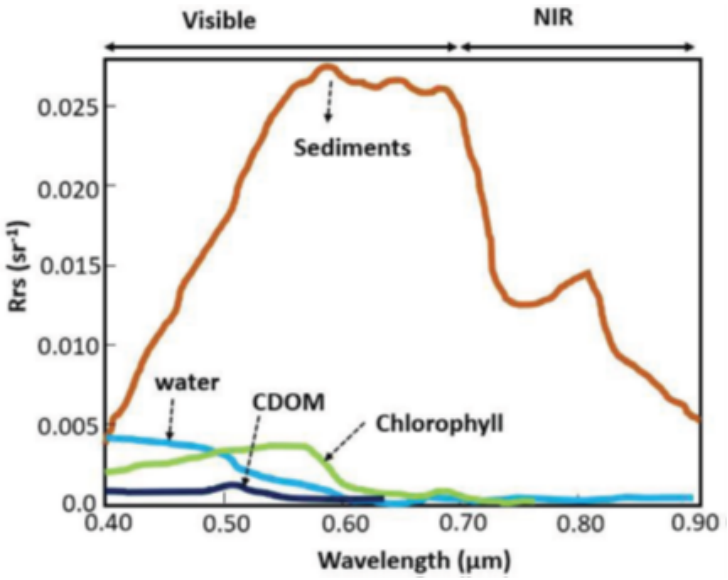


Figure 1.7: Reflectance (R_{rs}) of clear water (light blue line), water with chlorophyll content (green line), with sediments (orange line), and with Colored Dissolved Organic Matter CDOM (dark blue line) (Hafeez et al., 2018).

the spectral bands, and have a nominal repeat cycle of 16 days. Landsat imagery can be visualized and download from NASA's Earthdata Search platform[5].

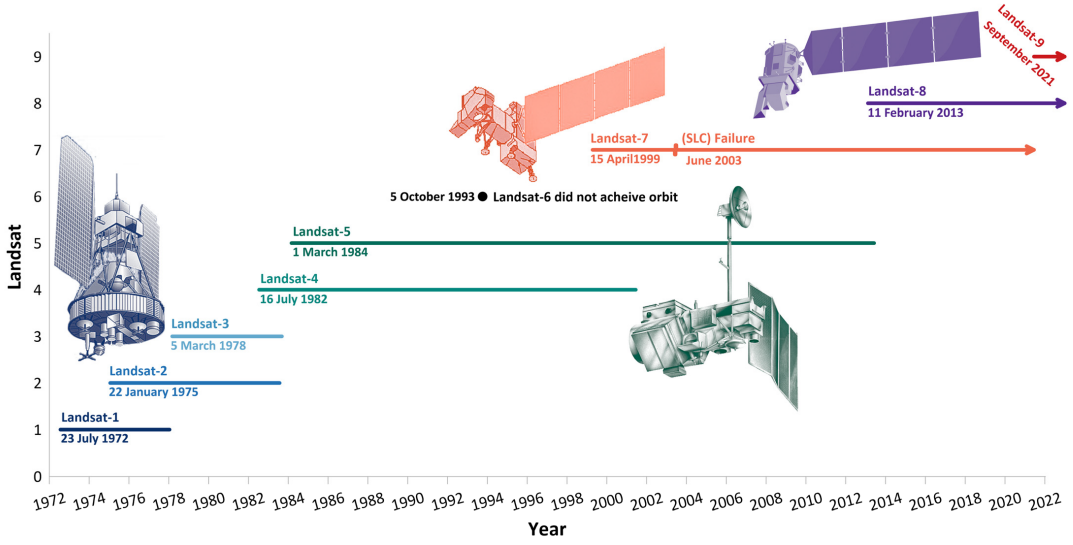


Figure 1.8: Landsat program timeline. (Source: NASA's Scientific Visualization Studio).

MODIS (Moderate Resolution Imaging Spectroradiometer) is a key instrument on board two NASA Earth Observing System (EOS) satellites Terra and Aqua. Terra was

[5]https://search.earthdata.nasa.gov

launched on December 1999, and Aqua was launched on May 2002, with the aim of observing the Earth's land, oceans, and atmosphere, monitoring global climate and environmental changes. The MODIS instrument has 36 spectral bands, with resolutions ranging from 250 meters to 1 kilometer, depending on the band. The instrument provides global coverage of the Earth's surface every one to two days, with a swath width of 2'330 kilometers. MODIS data is freely available to the public through the NASA Earthdata website, allowing researchers and the general public to access and analyze the satellite's observations. Aqua and Terra focus on broader observations of the Earth, while the Landsat program is specifically focused on high-resolution land observation.

At European level, the Copernicus Programme is the European Union's Earth Observation (EO) Programme, which delivers data, information and services based on satellite EO and in situ (non-space) data. The Programme is funded, coordinated and managed by the European Commission, with the aim of establishing a European capacity for the provision and use of monitoring information for environment and security applications (Drusch et al., 2012). The development of the observation infrastructure is performed under the aegis of the ESA for the space component and of the European Environment Agency (EEA) and Member States for the in situ component. Since the launch of Sentinel-1A in 2014, the European Union has initiated a process to place a complete constellation of almost 20 satellites in orbit before 2030. The EO satellites which provide the data exploited by the Copernicus services are split into two groups of missions:

- The Sentinels, which are developed for the specific needs of the Copernicus programme;
- The Contributing Missions, which are operated by National, European or International organisations and already provide a wealth of data for Copernicus services.

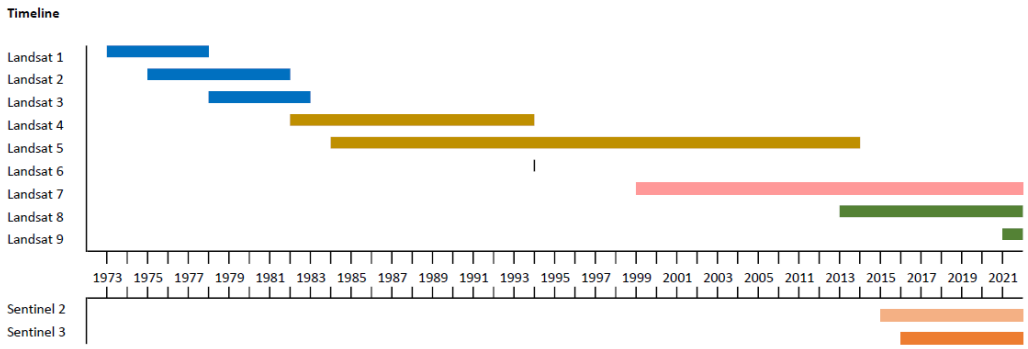


Figure 1.9: Timeline of Landsat and Sentinel satellites carrying optical sensors.

ESA Sentinel missions are designed to provide data for environmental monitoring, climate change studies, and other scientific research, to better understand the planet and sustainably manage the environment we live in, streamlined through six thematic Copernicus services:

- Atmosphere

- Marine
- Land
- Climate Change
- Security
- Emergency

Currently, the Copernicus Space Component includes the following missions and satellites:

- Sentinel-1 is a polar-orbiting, all-weather, day-and-night SAR imaging mission (Davidson et al., 2010) for land and ocean services. Sentinel-1A was launched on April 3, 2014 and Sentinel-1B on April 25, 2016.
- Sentinel-2 is a polar-orbiting multi-spectral optical imaging mission designed to monitor land use and cover changes, vegetation health, and water quality. Sentinel-2A and Sentinel-2B were launched respectively on on June 23, 2015 and on March 7, 2017.
- Sentinel-3 is a polar-orbiting and multi-instrument mission for marine and land services like ocean and land surface temperature, and ocean and land color. Sentinel-3A was launched on on February 16, 2016 and Sentinel-3B on April 25, 2018. Sentinel-3 carries an Ocean and Land Color Instrument (OLCI) for ocean color measurements, a Sea and Land Surface Temperature Radiometer (SLSTR) for temperature measurements, and a Synthetic Aperture Radar Altimeter (SRAL) for altimetry measurements.
- Sentinel-5P: Launched on October 13, 2017, this mission is a single-satellite atmospheric monitoring mission designed to measure air quality and trace gases that affect the Earth's climate.
- Sentinel-6 : Launched on November 21, 2020, this mission is a collaboration between ESA, NASA, NOAA, and the European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT) to continue a long-term sea level monitoring record.

In addition to these missions, more Sentinel missions are planned in the coming years, including Sentinel-4, Sentinel-5, and Sentinel-6B.
All Sentinel data are freely accessible from several sources among which the official Copernicus Open Access Hubsites[6] managed by ESA.
A comparison of the spectral resolution of the optical sensors Aqua and Terra MODIS, Landsat ETM+, Landsat OLI, and Sentinel 2 MSI is shown in Figure 1.3.

Datasets stretching back for years allow the examination and monitoring of changes in patterns. Maps created from imagery allow to identify features and anomalies and extract statistical information. Data fusion, combining data from multiple sensors or platforms, such as Landsat and Sentinel, allow to generate more complete information and allow

[6]https://scihub.copernicus.eu

wider temporal coverage. However, transforming this wealth of satellite data into value-added information can be a challenging task and requires processing and analysing the data, taking in consideration various factors that can affect the image quality and interpretation.
Also, computing infrastructure is required to store, process and analyze the amount of data collected, which can be expensive and time-consuming to develop and maintain.
Expert knowledge and computing software are required to ensure data quality, as satellite data is prone to errors and inaccuracies, including signal distortions, atmospheric interference, and sensor noise, which can impact the quality and accuracy of the derived information. Finally, to extract useful information from satellite data, it often needs to be integrated with other data sources, such as ground-based sensors, weather models, and historical data. This integration can be complex and require advanced analytical techniques.
The aim of this work is to investigate how the use of Machine Learning Techniques can contribute to enhance the usability of satellite data and allow its use to complement and support marine water quality monitoring, as required by EU monitoring programmes.

1.3. Machine Learning and Data Driven Models

Machine Learning (ML) is a field of computer science that aims to teach computers how to learn and act without being explicitly programmed. Arthur Samuel is credited for coining the term "Machine Learning" with his research around the game of checkers in 1959 (Samuel, 1959). He described machine learning as the "field of study that gives computers the ability to learn without being explicitly programmed". He concluded that programming computers to learn from experience should eventually eliminate the need for detailed programming.
According to Tom Mitchell (Mitchell, 1997), professor of Computer Science and Machine Learning at Carnegie Mellon, "a computer program is said to learn from experience E with respect to some task T and some performance measure P, if its performance on T, as measured by P, improves with experience E." A mathematical way of saying that a program uses machine learning if it improves at problem solving with experience.
Alan Turing's seminal paper (Turing, 1950) introduced a benchmark standard for demonstrating machine intelligence, such that a machine has to be intelligent and responsive in a manner that cannot be differentiated from that of a human being.

Machine learning involves the construction of algorithms that adapt their models to improve their performance at some task through experience (Awad and Khanna, 2015). Experience exists in the form of training datasets, which encompass an existing problem domain that the learner uses to build a general model about that domain. This enables the model to generate predictions in new cases. Models that are designed to learn patterns and relationships in the data through algorithms and statistical techniques, rather than relying on pre-existing knowledge or assumptions, are also known as *Data Driven Models* (DDM) .
ML and DDM have gained in popularity in the last decades and will further find widespread uses due to the empowerment of computation power, and the availability of big data.

This chapter discusses the essentials of ML, with the aim to present the key algorithms and theory that form the core of machine learning. Most of definitions and concept described here are taken from Prof. Mitchell textbook "Machine Learning" (Mitchell, 1997) and "Efficient Learning Machines: Theories, Concepts, and Applications for Engineers and System Designers" by Awad and Khanna (Awad and Khanna, 2015). A focus will be specifically made on statistical learning theory, the branch of ML algorithms from which the applied model belongs.

1.3.1. Algorithms classification

To have a well-defined learning problem, we must identity these three features: the class of tasks, the source of experience and the measure of performance to be improved. The first design choice we face is to choose the type of training experience from which our system will learn. In general, learning is most reliable when the training examples follow a distribution similar to that of future test examples. Secondly, we reduce the learning task to the problem of discovering an operational description of the ideal target function f. We

expect learning algorithms to acquire only some approximation to the target function, and for this reason the process of learning the target function is often called function approximation. We must choose a representation that the learning program will use to describe the function. In general, this choice of representation involves a crucial trade-off. On one hand, we wish to pick a very expressive representation to allow representing as close an approximation as possible to the ideal target function f. On the other hand, the more expressive the representation, the more training data the program will require in order to choose among the alternative hypotheses it can represent. Finally, we reduce the problem of improving performance P at task T to the problem of learning a particular target function. Machine learning algorithms all aim to learn and improve their accuracy as they process more datasets.
We can thus think of breaking out the learning system of a machine learning algorithm into three main parts:

1. A Decision Process: the type of task (e.g. make a prediction or classification). Based on some input data, which can be labeled or unlabeled, the algorithm will produce an estimate about a pattern in the data.

2. An Error Function: it evaluates the prediction of the model. If there are known examples, an error function can make a comparison to assess the accuracy of the model.

3. A Model Optimization Process: If the model can fit better to the data points in the training set, then weights are adjusted to reduce the discrepancy between the known example and the model estimate. The algorithm will repeat this "evaluate and optimize" process, updating weights autonomously until a threshold of accuracy has been met.

One way we can classify ML algorithms is based on the training experience and by how much feedback they receive. We can identify four primary categories:

- Supervised Learning: it's defined by the use of labeled datasets to train algorithms. It involves creating a mathematical function that relates input variables to the preferred output variables. A large amount of labeled training datasets are provided which provide examples of the data that the computer will be processing. Classification or regression problems are supervised learning tasks.

- Semi-supervised Learning: some labelled training data is provided, but most of the training data is unlabelled. Image recognition is a good example of semi-supervised learning. In this example, we might provide the system with several labelled images containing objects we wish to identify, then process many more unlabelled images in the training process.

- Unsupervised Learning: all input is unlabelled and the algorithm must create structure out of the inputs on its own. Clustering problems (or cluster analysis problems) are unsupervised learning tasks that seek to discover groupings within the input datasets.

- Reinforcement Learning: similar to supervised learning, but the algorithm is not trained using sample data. This model learns as it goes by using trial and error. A sequence of successful outcomes will be reinforced to develop the best recommendation or policy for a given problem.

Another way to distinguish algorithm is according to the class of task they aim to solve:

- Clustering;
- Dimensionality reduction;
- Classification;
- Regression.

Finally, we can try group algorithms by similarity in terms of their function or representation for the target function.

Clustering Algorithm

Clustering describes the class of problem and the class of methods. A cluster analysis attempts to group objects into "clusters" of items that are more similar to each other than items in other clusters. Using unsupervised learning, clustering algorithms can identify patterns in data so that it can be organized into groups of maximum commonalities. Clustering methods are typically organized by the modeling approaches such as centroid-based and hierarchal. The most popular clustering algorithms are:

- k-Means
- k-Medians
- Expectation Maximization (EM)
- Hierarchical Clustering

Clustering is a common task in statistical analysis and data mining.

Dimensionality Reduction Algorithm

Like clustering methods, dimensionality reduction seeks and exploit the inherent structure in the data, but in this case in order to summarize or describe data using less information. This can be useful to visualize dimensional data or to simplify data which can then be used in a supervised learning method. Examples of algorithms are:

- Principal Component Analysis (PCA)
- Principal Component Regression (PCR)
- Partial Least Squares Regression (PLSR)
- Sammon Mapping
- Multidimensional Scaling (MDS)

- Projection Pursuit
- Linear Discriminant Analysis (LDA)
- Mixture Discriminant Analysis (MDA)
- Quadratic Discriminant Analysis (QDA)
- Flexible Discriminant Analysis (FDA)

Regression Algorithm

Regression is concerned with modeling a relationship between features of the training data, iteratively refined using a measure of error in the predictions made by the model, with the aim of using the model to predict the value for new data. The most popular regression algorithms are:

- Linear Regression
- Logistic Regression
- Least Squares Regression (LSR)
- Stepwise Regression
- Multivariate Adaptive Regression Splines (MARS)

Instance-based Algorithms

Instance-based learning methods typically build up a database of example data and compare new data to the database using a similarity measure in order to find the best match and make a prediction. For this reason, instance-based methods are also called memory-based learning. Focus is put on the representation of the stored instances and similarity measures used between instances. The most popular instance-based algorithms are:

- k-Nearest Neighbor (kNN)
- Learning Vector Quantization (LVQ)
- Self-Organizing Map (SOM)
- Locally Weighted Learning (LWL)
- Support Vector Machines (SVM)

Regularization Algorithms

Regularization is a technique used to prevent overfitting in statistical models. It does this by adding a penalty term to the objective function that the model is trying to optimize. The penalty term is designed to discourage the model from assigning too much importance to any particular feature, thereby reducing the model's sensitivity to noise in the data

and improving its ability to generalize to new data. Regularization indeed encourages the model to produce simpler solutions that generalize better to new data.
The most popular regularization algorithms are:

- Ridge Regression
- Least Absolute Shrinkage and Selection Operator (LASSO)
- Regularized Least Squares (RLS)
- Elastic Net
- Least-Angle Regression (LARS)

Bayesian Algorithms

Bayesian methods are those that explicitly apply Bayes' Theorem for problems such as classification and regression. The most popular Bayesian algorithms are:

- Naive Bayes
- Gaussian Naive Bayes
- Multinomial Naive Bayes
- Averaged One-Dependence Estimators (AODE)
- Bayesian Network (BN)
- Bayesian Belief Network (BBN)

Decision Tree Algorithms

Decision trees use a branching sequence of linked decisions that can be represented with a tree diagram. They can be used for both regression and classification problems. They are often fast and easy to validate. The most popular decision tree algorithms are:

- Classification and Regression Tree (CART)
- Iterative Dichotomiser 3 (ID3)
- C4.5 and C5.0 (different versions of a powerful approach)
- Chi-squared Automatic Interaction Detection (CHAID)
- Decision Stump
- Conditional Decision Trees

Ensemble Algorithms

Ensemble methods are models composed of multiple weaker models that are independently trained and whose predictions are combined in some way to make the overall prediction. Much effort is put into what types of weak learners to combine and the ways in which to combine them. This is a very powerful class of techniques and as such is very popular. Examples are:

- Random Forest
- Boosting
- Bootstrapped Aggregation (Bagging)
- AdaBoost
- Weighted Average (Blending)
- Stacked Generalization (Stacking)
- Gradient Boosting Machines (GBM)
- Gradient Boosted Regression Trees (GBRT)

Artificial Neural Networks

Artificial Neural Networks are models that are inspired by the structure and/or function of biological neural networks, as they simulate the way the human brain works, with a huge number of linked processing nodes. They use a series of functions to process an input signal or file and translate it over several stages into the expected output. They are a class of pattern matching that are commonly used for regression and classification problems but are really an enormous subfield comprised of hundreds of algorithms and variations for all manner of problem types. The most popular classical artificial neural network algorithms are:

- Perceptron
- Multilayer Perceptron (MLP)
- Back-Propagation
- Stochastic Gradient Descent
- Hopfield Network
- Radial Basis Function Network (RBFN)

Deep learning Algorithms

Deep learning refers to algorithms that make heavy use of artificial neural networks, where "deep" refers to the number of layers, or iterations between input and output. The most popular deep learning algorithms are:

- Convolutional Neural Network (CNN)
- Recurrent Neural Networks (RNNs)
- Long Short-Term Memory Networks (LSTMs)
- Stacked Auto-Encoders
- Deep Boltzmann Machine (DBM)
- Deep Belief Networks (DBN)

Finally, the field of machine learning is concerned with answering questions such as the following:

- Which algorithms perform best for which types of problems and representations?
- How much training data is sufficient? What general bounds can be found to relate the confidence in learned hypotheses to the amount of training experience and the character of the learner's hypothesis space?
- When and how can prior knowledge held by the learner guide the process of generalizing from examples?
- What is the best way to reduce the learning task to one or more function approximation problems?
- Can the learner automatically alter its representation to improve its ability to represent and learn the target function?

It is also important to judge whether ML is the suitable approach for solving a given problem. By its nature, ML cannot deliver perfect accuracy. For solutions requiring highly accurate results in a bounded time period, ML may not be the preferred approach. In general, the following conditions are favorable to the application of ML: (a) very high accuracy is not desired; (b) large volumes of data contain undiscovered patterns or information to be synthesized; (c) the problem itself is not very well understood owing to lack of knowledge or historical information as a basis for developing suitable algorithms; and (d) the problem needs to adapt to changing environmental conditions.

1.3.2. Learning from data - Statistical learning theory

We will discuss here statistical learning theory, a branch of ML that focuses on studying the properties and behavior of learning algorithms in a statistical framework. At the heart of statistical learning theory is the concept of a hypothesis space, which is a set of possible models or functions that the learning algorithm can choose from to fit the training data. The goal of the learning algorithm is to select the best model from this hypothesis space that minimizes some notion of error on the training data, while still generalizing well to new, unseen data.

We consider the approach to machine learning of supervised learning, based on the so called *learning from examples* paradigm. Given the so called training set S, that is a set of n input-output pairs:

$$S = \{(x_1, y_1), ..., (x_n, y_n)\}$$

where the inputs x belong to the input space X and the outputs y belong to an output space Y, the goal of supervised learning is to find (or *learn*) an underlying input-output relation:

$$f(x_{new}) \sim y,$$

The training examples come from some generally unknown probability distribution (considered representative of the space of occurrences) and the learner has to build a general model about this space that enables it to produce sufficiently accurate predictions in new cases. It is necessary to postulate the existence of a model for the data, that should take into account the possible uncertainty in the task and in the data.
One of the most fundamental aspects of machine learning is how to optimize the performance of the prediction and, thus, how to minimize the expected error. A loss function l needs to be fixed, which is a measure of the error *l(y, f(x))* of predicting *f(x)* in place of y. Given a loss function the target function $f^* : \mathrm{X} \longrightarrow \mathrm{Y}$ is the one that minimize the expected loss, which can be seen as a measure of the error on past as well as future data:

$$\epsilon(f) = \mathbb{E}[l(y, f(x))]$$

A good learning algorithm should be able to fit the training data, and at the same time be stable with respect to noise and sampling, and avoid overfitting. Generalization in this context is the ability of a learning machine to perform accurately on new, unseen data.
The performance of the algorithm to predict is given by the expected error which can be analyzed decomposing it as the sum of bias, variance and the irreducible error σ^2, resulting from the noise in the data:

$$Err(x) = bias^2 + variance + \sigma^2$$

The square of the bias of the learning method can be thought of as the error caused by the simplifying assumptions built into the method. Underfitting, which happens when a model is unable to capture the underlying pattern of the data, result in high bias. The variance represents how much the learning method will move around its mean; it's related to sensitivity to small fluctuations in the training set. High variance may result from an algorithm modeling the random noise in the training data (overfitting).
For any given training set, the best choice for the complexity of the model would be the one striking the optimal trade-off between bias and variance, that is the value minimizing their sum (Figure 1.10).
Most algorithms depend on one or more regularization parameters that control the trade-off between data fitting and stability. Statistical learning theory employs the principle of Occam's razor, which states that the simplest model that explains the data is usually the best one. This principle is captured by the concept of regularization, which adds a penalty term to the loss function to discourage complex models and to prevent overfitting.

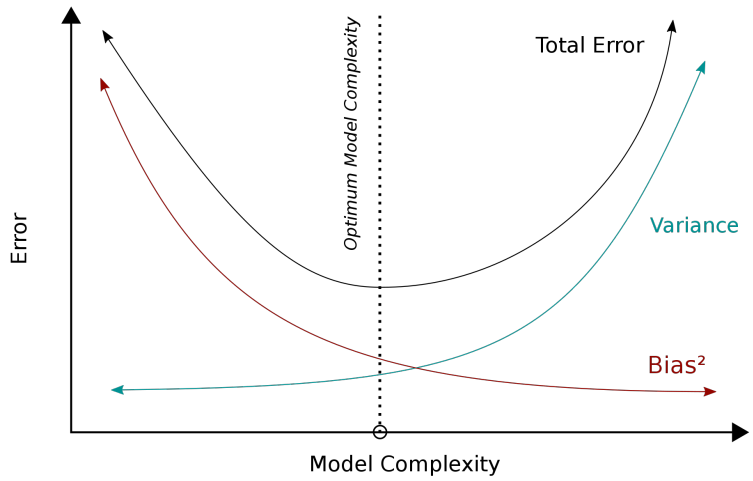


Figure 1.10: The bias-variance trade-off.

Since the data distribution, hence the expected loss, is not accessible, data driven procedures are used to find a proxy for the expected loss. Cross validation procedure is used to compute a hold-out error, obtained from part of the training set, to be used as a proxy of the expected error and thus to achieve an empirical bias variance trade-off choosing the model parameters that achieve minimum hold-out error.

The process of developing supervised machine learning algorithm may be decomposed into the following steps:

1. Collect the data. Gather the data necessary to the problem.

2. Pre-process the data, which consists of the following three steps:

   - Formatting. The data needs to be in a usable format for ML algorithm.
   - Cleaning. The data needs to be cleaned by removing, substituting, or fixing corrupt or missing data. In some cases, data needs to be normalized, discretized, averaged, smoothed, or differentiated for efficient usage. In other cases, data may need to be transmitted as integers, double precision, or strings.
   - Transform the data. Transform the data specific to the algorithm and the knowledge of the problem. Transformation can be in the form of feature scaling, decomposition, or aggregation. Features can be decomposed to extract the useful components embedded in the data or aggregated to combine multiple instances into a single feature.

3. Select the model. Choose a suitable ML algorithm based on the problem you are trying to solve and the characteristics of the data.

4. Train the algorithm. Select the training and testing datasets from the transformed data. An algorithm is trained on the training dataset and evaluated against the

test set. The transformed training dataset is fed to the algorithm for extraction of knowledge or information. Apply cross validation to chose the parameters of the model.

5. Test the algorithm. Evaluate the algorithm to test its effectiveness and performance. A trained model exposed to test dataset is measured against predictions made on that test dataset which are indicative of the generalization performance and potential overfitting. If the performance of the model needs improvement, repeat the previous steps by changing the data streams, sampling rates, transformations, outliers' removal methodology, and biasing schemes.

6. Model deployment. Apply the validated model to perform an actual task of prediction. If new data are encountered, the model is retrained by applying the previous steps. The process of training may coexist with the real task of predicting future behavior.

The end-to-end construct that orchestrates the flow to produce a machine learning model, from data extraction and pre-processing, to model training and deployment, is referred to as machine learning pipeline. A simplified scheme of ML pipeline is reported in Figure 1.11.

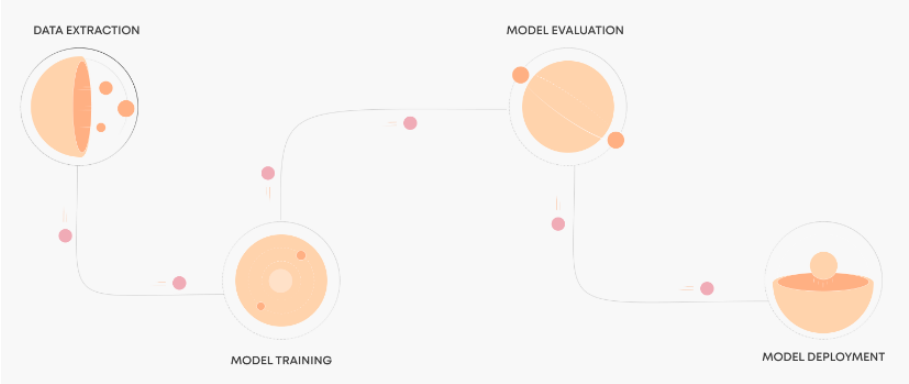


Figure 1.11: Example of a generic ML pipeline (Source: *https://valohai.com*).

RESEARCH GOAL

The spectral signature of the seawater surface is a function of its optically significant constituents: phytoplankton pigments, CDOM, TSM, etc.. which may vary independently of one another. It is possible to attempt to discriminate the presence, types and concentration of materials and substances in sea water by measurements of the water reflectance. Ocean color remote sensing onboard satellite platforms can be very useful to complement field measurements and monitor surface turbidity and other optically active parameters in natural waters.
Remote sensing from satellites offers observations of the sea surface over large areas in a cost-effective way, as an always increasing amount of data can freely be accessed through specific websites.
Water quality retrieval from satellites has gained recognition as a valid integrative tool for monitoring water bodies with the availability of freely available new imagery from optical multispectral sensors with a higher spatial, spectral, and temporal resolution, like those of Landsat and Sentinel.
Many studies have been conducted to estimate turbidity from satellite images in ocean, coastal or inland waters. Commonly used methods are based on semi-empirical algorithms relying on reflectance in bands that have shown the highest physical correlation to the parameter (see Chapter 1.2.2.).
For low concentration of suspended sediments, turbidity is mostly related to water reflectance in the green and red parts of the spectrum. Red reflectance is more sensitive than green reflectance in medium-high turbid waters, while for sediment dominated highly turbid waters, a saturation of the water reflectance in the green and red bands is usually observed, so a NIR band is usually more appropriate (Doxaran et al., 2002b).
A Normalized Difference Turbidity Index (NDTI), defined as (Red - Green) / (Red + Green), is often used to estimate the turbidity in inland water bodies (Baughman et al., 2015).
Terra and Aqua MODIS spectral data have been successfully applied to retrieve turbidity and TSM in coastal and estuarine environments (Doxaran et al., 2002a; Chen et al., 2007; Miller and McKee, 2004; Petus et al., 2010; Hudson et al., 2017).
Numerous studies reported effective ways to utilize medium-resolution Landsat Thematic Mapper (TM), Enhanced Thematic Mapper Plus (ETM+), Operational Land Imager (OLI) and Sentinel-2 MultiSpectral Instrument (MSI) sensors images through semi-empirical algorithms for inland waters, lagoon, marshes and estuarine coastal waters (Akbar et al., 2014; Nas et al., 2010; Liu and Wang, 2019; Pereira et al., 2018; Quang et al., 2017; Joshi et al., 2017; Baughman et al., 2015; Bustamante et al., 2009; Abirhire et al., 2020; Katlane et al., 2020; Caballero et al., 2018).
Despite their good general performance, these types of models are dependent on reflectance ranges and turbidity levels, therefore their validity and accuracy often reveal to be limited to the domain (Dogliotti et al., 2015) and season (Joshi et al., 2017) over which they have been developed. This often leads in practice to site-specific models whose

coefficients need to be locally adapted for a defined coastal domain (Han et al., 2016). Hence, it is difficult to select one model that will provide accurate turbidity retrieval from low to high-turbidity waters, limiting the application of studies over large coastal areas. For this reason, geographically weighted regression (GWR) and geographically and temporally weighted regression (GTWR) models have been proposed in literature (Chu et al., 2018) to take into consideration spatio-temporal variation in model coefficients.

Nechad et al. (2010) proposed a multi conditional single band algorithm for TSM in estuarine coastal environments, based on red band for medium-low turbidity waters and NIR band in highly turbid waters. The difficulty resides in the selection of the limiting bounds for each model. The algorithm scheme has been applied with robust results to many studies in coastal and estuarine environments, where the switching thresholds is based on ranges of turbidity/TSM concentration (Feng et al., 2014) or reflectance values (Dogliotti et al., 2015; Han et al., 2016; Novoa et al., 2017; Caballero et al., 2018). The automated switching algorithm described in Novoa et al., 2017 constitutes the basis for the turbidity data, expressed in FNU, obtained from the Sentinel-2 MSI data and introduced into Copernicus Marine Service (CMEMS) website[1] in May 2021, named "Mediterranean Sea, Bio-Geo-Chemical, L3, daily observation" (Figure 2.1).

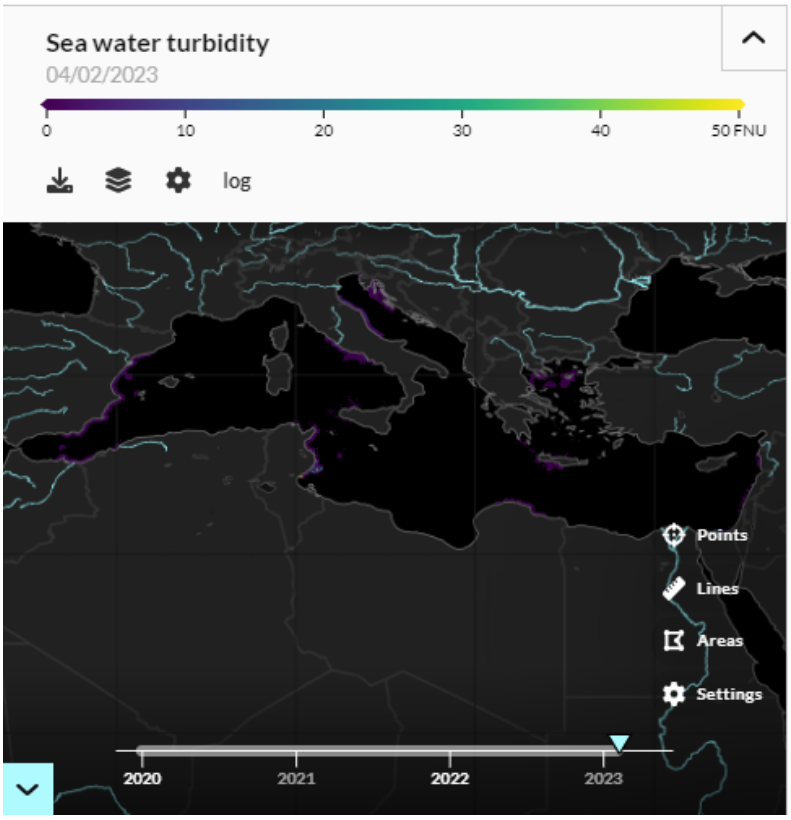


Figure 2.1: 'Mediterranean Sea, Bio-Geo-Chemical, L3, daily observation' CMEMS turbidity product.

[1]https://data.marine.copernicus.eu/

An alternative approach to semi-empirical models is provided by non-parametric regression, where no explicit relationship between reflectance and the parameter of interest is assumed, but the functional form is inferred from the data. Very flexible relations can be accommodated, responding to the need of a model that can adapt to varying water conditions with smooth transition, and the wealth of spectral information can be exploited.
Use of ML for remotely sensed optically active water quality parameters estimation has spread in recent years thanks to the advances in algorithm development, computing power, sensor systems, and higher spatial resolution data availability (Peterson et al., 2018).
Ruescas et al. (2018) applied and compared five ML regression methods to Sentinel-2 MSI and Sentinel-3 OLCI simulated reflectance data for the retrieval of CDOM, showing that best results were obtained when all bands from visible to NIR, together with band ratios, were used as input of non-linear ML models. Similarly, a Neural Network (NN) model was developed by Chebud et al. (2012) to quantify Chl-a, turbidity and phosphorus from reflectance values from the seven Landsat TM bands, including SWIR and thermal bands. El Din (2019) showed, through application of Principal Components Analysis (PCA), that also Coastal Aerosol band provides important information in high turbid waters. Pahlevan et al. (2020) introduced a Mixture Density Network (MDN) approach for estimations of Chl-a from seven MSI and twelve OLCI bands, that largely outperforms existing algorithms when applied across different bio-optical regimes in inland and coastal waters. Similarly, Peterson et al. (2020) applied NN Deep learning model fed by seven raw harmonized Landsat-8 and Sentinel-2 (HLS) spectral bands, from Coastal Aerosol to SWIR-2, supplemented with a series of additional spectral features to add key information based on standard band ratios method for the estimation of blue-green algae (BGA), Chl-a, fluorescent dissolved organic matter (fDOM), dissolved oxygen (DO), specific conductance (SC), and turbidity in inland waters.
Neural Network-based approaches have gained popularity due to their robust ability to capture complex statistical trends typical of water quality remote sensing data, however, the use deep learning is still limited by the high computational costs and the amount of training samples required to adequately calibrate the numerous parameters (Sagan et al., 2020).
By analyzing the state-of-the art, it appears clear that the application of ML algorithms seems to be promising for the retrieval of optically active parameters when investigating wide areas or long time series that include different optical water types.
When it comes to a complex parameter, such as turbidity, in coastal regions, where the effects of organic and inorganic material become more intricate, the use of ML algorithms can be more appropriate. In fact, it allows evaluating nonlinear relationships, and to exploit the whole spectrum of available bands, evaluating the contribution of the different bands in the model prediction.
These considerations guided the research objective, which can be synthesized in the following research question: Can satellite optical multispectral data support seawater quality monitoring with sufficient accuracy, and how can machine learning contribute to this?

MATERIALS & METHODS for Model Development

To achieve this goal, an area in the North Thyrrenian Sea has been chosen as study area for its environmental and economic relevance, and the intense monitoring activity carried on from the competent authorities (Chapter 3.2.).
Turbidity was chosen as study parameter for its wide use in monitoring the marine environment related to its versatility and practicality (Chapter 3.1.).
Sentinel-2 MSI images were identified as the best satellite product for the scope, due to the sensor spectral, spatial and temporal resolution (Chapter 3.3.). Based on the problem to solve and the characteristics of the data, a suitable machine learning algorithm was identified, calibrated and validated (Chapter 3.4.) .

The model implementation workflow is shown in Figure 3.1, and is the topic of the present Section.

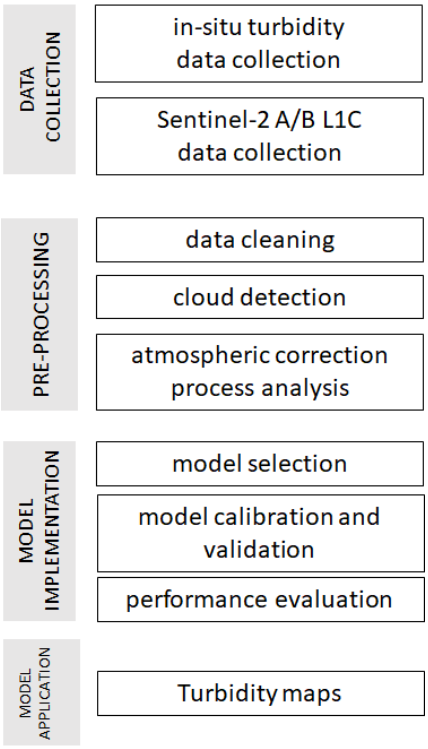


Figure 3.1: Model development workflow.

The Model Development follows the steps presented in the different chapters of the present Section:

- Data selection and collection (Chapters 3.1., 3.2. and Chapter 3.3.);

- Data pre-processing, including some specific processing of optical images, such as atmospheric correction analysis (Chapter 3.3.2.) and cloud detection (Chapter 3.3.3.);
- Model implementation: algorithm selection, model calibration and validation (Chapter 3.4.). The model performance was evaluate through validation, considering different statistical coefficients;
- Model deployment: application of the validated model to produce turbidity maps. The identified model was finally applied to different satellite images to produce dense maps (Chapter 4.3.) that can support marine monitoring extending the punctual information of sampling campaigns to a synoptic view.

3.1. Sea Water Turbidity

Turbidity is a visual property of water; it is a measure of the amount of light scattered by particles in water, due to the presence of suspended particles, i.e., suspensoids, which operationally can be defined as the fraction in water that can be removed by a pore size filter of 0.22 μm (Kirk, 1985). Suspensoids include plankton and inanimate, organic or inorganic, particles. Turbidity is thus particularly useful for studying phenomena such as eutrophication and sediment transport in water bodies.
Even though algal blooms can seasonally be observed, in the coastal marine environment high turbidity events are dominated by suspended sediments: sediment load from rivers runoff, sea bottom resuspension due to wave action, shoreline erosion, anthropogenic activities that cause the resuspension or load of fine sediments, such as beach nourishment, coastal infrastructures, dredging activities.
It is widely recognised that the increase of suspended solid concentration can determine negative environmental effects on the marine ecosystem (Erftemeijer et al., 2012; Wilber and Clarke, 2001). In coastal marine systems at a high level of anthropization, sediments may contain toxic substances: heavy metals, polycyclic aromatic hydrocarbons (PAHs), tributyltin (TBT), polychlorinated (Garbolino et al., 2014). Sediments thus represent a potential source of contaminants whose resuspension and dispersion under different hydrodynamics conditions can contribute to propagate pollution (Lisi et al., 2019; Magrì et al., 2020).
Rivers serve as the main channel for the delivery of significant amounts of dissolved and particulate materials from terrestrial environments to the ocean. Along with freshwater, they discharge suspended particulate matter, that not only modifies the color and transparency of the water, but is often associated with contaminants and bacteria that affect water quality.

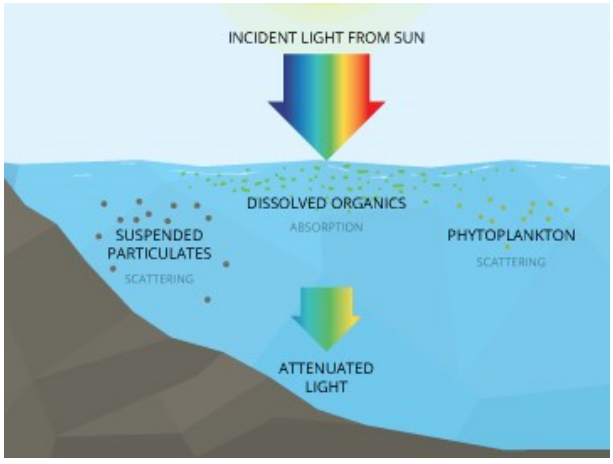


Figure 3.2: Effects of attenuation of the incident light due to dissolved and suspended particles in water.

Additionally, as a result of climate change, extreme weather events are expected to occur more frequently, increasing coastal and soil erosion susceptibility, generating high levels

of sediments and turbidity in rivers and coastal waters.
The European Union MSFD and WFD Directives require member states to monitor the quality of the marine environment and to achieve and maintain a good environmental status (GES) of all marine waters. The directives require member states to assess the ecological quality status of water bodies, based on the status of several elements, including water transparency and turbidity. Monitoring the spatio-temporal distribution of turbidity in estuarine and coastal waters is of particular importance, not only to assess water clearness, but also to study sediment transport dynamics and to evaluate the impacts of human activities.
Turbidity values can act as simple and convenient indirect measures of the concentration of suspended solids and other particulate material. However, a direct relationship between turbidity data and suspended solids concentrations is weakened by the complex interactions of light energy with suspended particles. This interplay is heavily dependent on many factors, including: concentration of scattering particles suspended in the medium, size distribution of the scattering particles; shape, orientation, and surface condition of the scattering particles; refractive index of the scattering particles; refractive index of the suspension medium; wavelength of the light source employed (Lawler, 2005). Natural waters contain an ensemble of variably sized, irregularly-shaped, and randomly oriented particles, for which theory is still being developed. Furthermore, processes become highly complex when concentrations are so great that multiple scattering occurs (i.e., particles receive light previously scattered from other particles: this normally increases opportunities for light absorption).

3.1.1. Turbidity measurements

The field of turbidity measurements is hampered by a lack of standardization in units, measurement devices and calibration techniques (Li and Liu, 2018). Modern measurement devices (also known as turbidimeters) use photosensitive cells to quantify scattered and/or transmitted light. The instruments usually have five basic components: a light source of known, constant intensity and given wavelength characteristics; a lens to collimate the light beam; a sample cell; photosensor(s); and a meter or logger to record the output signals from the photosensor(s).
In general, turbidity measurement tools fit into two categories measurement devices and calibration techniques, nephelometers and spectrophotometers (Figure 3.3):

- Nephelometers emit and measure the amount of light scattered by particles in the sample. The units depend on the wavelength of the light and the angle at which the detector is positioned. Typical units include Formazin Nephelometric Units (FNU) and Nephelometric Turbidity Units (NTU). Both methods measure scattered light at 90 degrees from the incident light beam, but the FNU is measured with an infrared light source (wavelength of 860 ± 60 nm) according to the ISO-7027 method, whereas the NTU is measured with a white light (wavelength from 400 to 680 nm), according to EPA method 180.1.

- Spectrophotometers (or transmissometer) direct a beam of light at a specific wavelength and known initial intensity through the sample, while a detector determines the amount of light transmitted. Turbidity is measured by quantifying the degree of 'attenuation' of the beam of light. Depending on the light source, the units used are Attenuation Units (AU) or Formazin Attenuation Units (FAU). Attenuation turbidity instruments are not approved by most regulatory agencies.

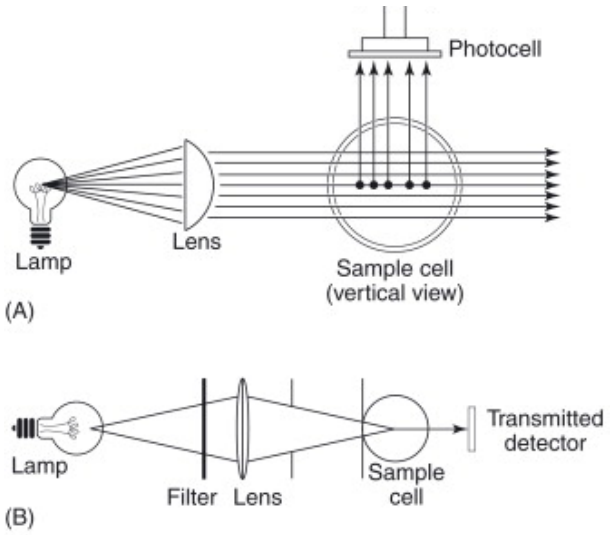


Figure 3.3: Basic designs of: (A) the nephelometer, which directly measures light scattered (usually at 90° to the beam direction) by suspended particles; (B) the transmissometer, where the transmitted light is detected, in relation to initial beam intensity. Reproduced from Lawler, 2005

.

A B

Figure 3.4: Example of a submersible probe (A), and detail of different sensors (B).

As FNU or NTU turbidity measurement are usually required in field applications, in situ sea water turbidity is commonly measured by a nephelometer, mounted on a submersible

probe (see Figure 3.4).
Field measurements are time-consuming, expensive and specific to a time and/or geographical location, and therefore do not always accurately represent the temporal and spatial dynamics of coastal systems.
The accurate, but punctual and sparse information provided by traditional sampling methods could gain in effectiveness if integrated with satellite observations, that offer high-spatial-resolution images over large areas in a cost-effective way.

3.2. Study Area

The research area is located in the North Tyrrhenian Sea – Italy (Figure 3.5), covering a coastline of about 100 km, characterized by a varied shoreline, embracing environments worthy of protection and valuable biodiversity, including Cinque Terre Marine Protected Area (North Western part of the area), aquaculture and shellfish farms, the Versilia extensive sandy shoreline with a strong tourist vocation, but also relevant ports such as Livorno and La Spezia. Three main rivers flow in the area with a significant sediment discharge: Magra River, whose basin covers an area of about 1700 km^2 and separates Liguria and Tuscany regions, with an average flow of about 40 m^3/s; Serchio river, whose basin covers approximately 1455 km^2 and an estimated average annual discharge of approximately 40 m^3/s. Arno River, which represents the main watercourse in Tuscany and the second largest Apennine river in Italy, with a basin of about 8200 km^2 and an estimated average annual discharge of approximately 110 m^3/s.
The coastal environment in this area has been monitored since 2001, according to the 2000/60/EC WFD, and in 2008 EU MSFD 2008/56/EC further strengthened the investigation in the area, resulting in 36 monitoring stations located along the coast. Water quality monitoring includes physical and chemical parameters along the water column and in sediments.
Figure 3.5 shows the study area, location of Marine Protected Areas (MPA) and Zones of Special Conservation (ZPC), according to official cartography[1] and the position of the monitoring stations according to WFD and MSFD.

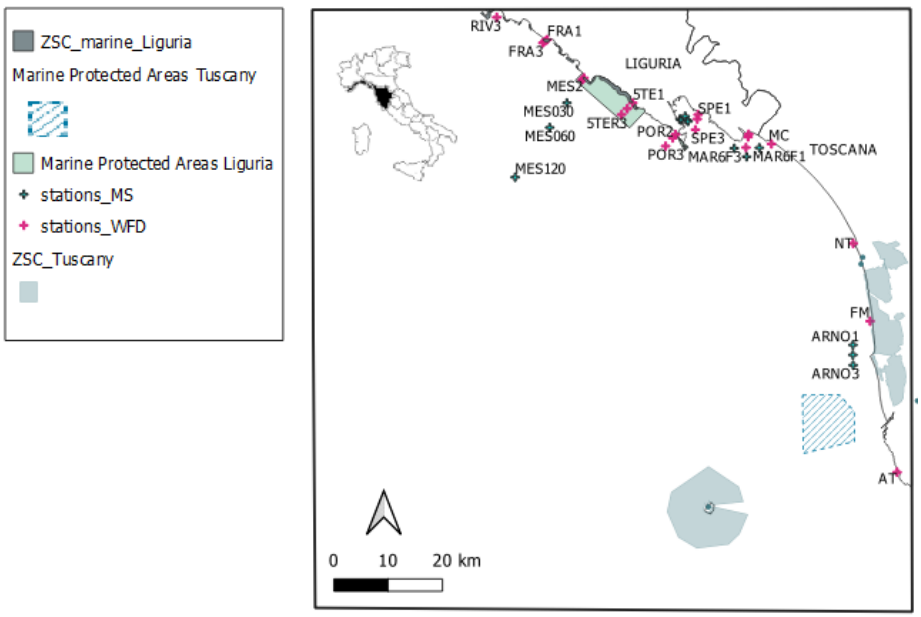


Figure 3.5: Location of the study area, location of MPA and ZPC, and distribution of the sampling stations according to WFD and MSFD.

[1] http://www502.regione.toscana.it/; https://geoportal.regione.liguria.it/

3.2.1. Turbidity data acquisition

On-field data collected according to the WFD and MSFD by the Regional Environmental Agencies ARPAL (Liguria Region) and ARPAT (Toscana Region) in the 36 stations located in the area were used. The monitoring stations are located at a minimum depth of about 4 m (station MAR1), up to water depths of about 500 m (station MES120).
The considered area presents a wide variety of turbidity contributions and conditions, ranging from clear water, especially within the Marine Protected areas, to high turbidity, particularly related to sediment loads carried from rivers.
Turbidity was measured using a multi-parameter submersible probe, together with electrical conductivity, temperature, salinity, and pressure of seawater (Figure 3.6). Turbidity was measured in nephelometric turbidity units (NTU) (see Chapter 3.1. for more details). At each station, turbidity profiles were conducted from the surface to the sea bottom, by lowering the probe at a rate of 1 m per minute with values recorded every second. In order to obtain a good representation of surface turbidity that is consistent for each station, the average turbidity at 0.5 m was used.

Figure 3.6: Turbidity measurements by a turbidimeter mounted on a submersible probe.

The different monitoring stations have a specific sampling frequency according to the reference Directive, therefore on-site measurements are not available for all stations on the same days, but mostly for groups of stations. In general, the sampling campaigns are conducted from about 9 a.m to about 2 p.m.; the time difference between measurements at different points, on the same day, depends on the distance between stations and is at most a couple of hours.
Figure 3.7 shows for each monitoring station the amount of available data (represented by the dimension of the circle), and the maximum observed value (represented by the color of the circle). We can notice that the highest values are found at the Serchio river mouth (station “NT”) and Magra river (stations “MC”).
All data are available on the official “Regional Environmental Information System”

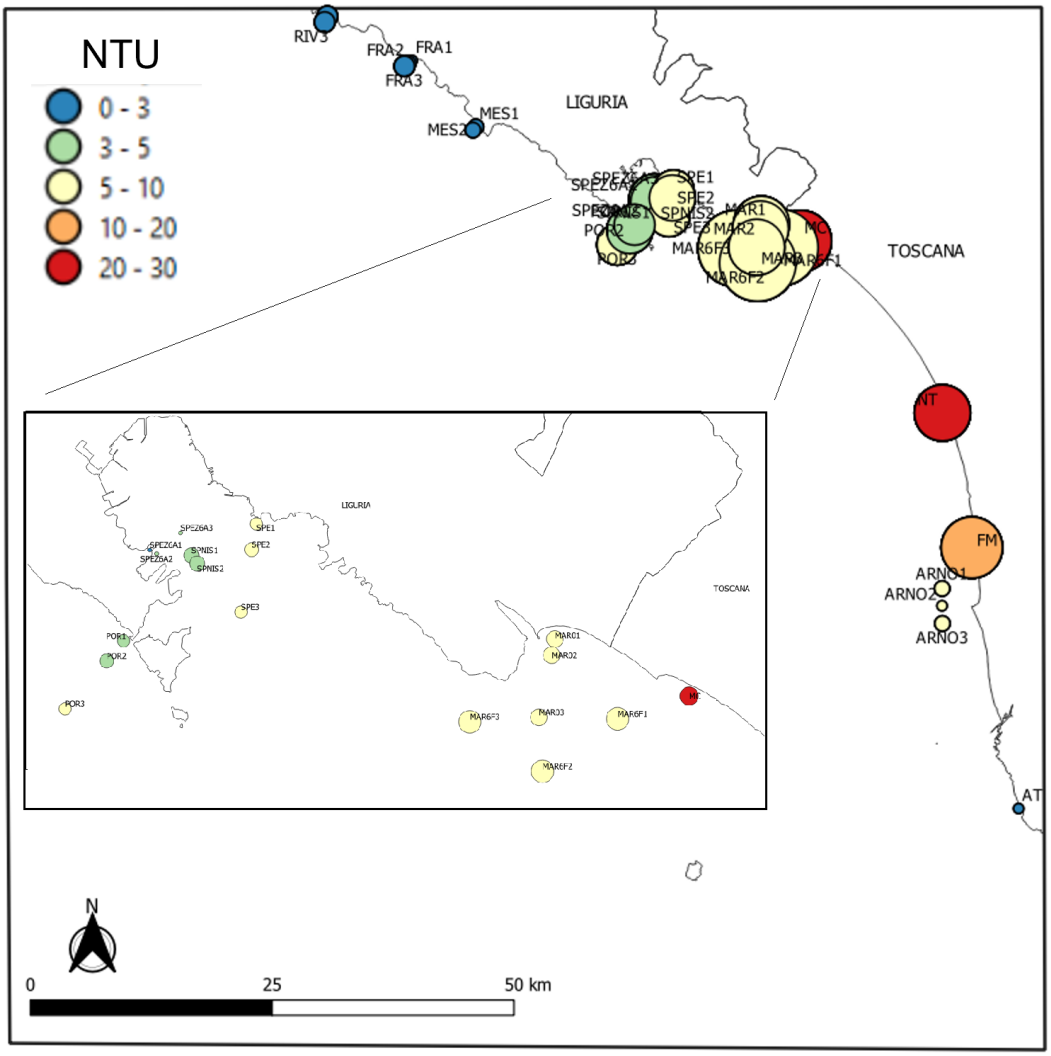


Figure 3.7: Measured turbidity (NTU) for the different monitoring stations. The size of the circle refers to the number of observations, while the color refers to the maximum measured turbidity value, according to the legend.

(SIRA) websites[2]. SIRA is a digital platform used by the Italian regions to collect, manage, and disseminate environmental data. The system is in fact designed to provide public access to environmental information.
The used database covered the period from April 2015 (when first Sentinel-2 satellite was launched) to December 2021.
Out of 234 days of field campaigns in the area, in the defined period, only 84 had a Sentinel-2 satellite overpass on the same day, corresponding to 394 co-located pairs of turbidity records and water-leaving reflectances in the different bands, that were further analysed. Only the daily correspondence was considered, with a maximum time lag be-

[2]https://http://www.banchedati.ambienteinliguria.it/; https://sira.arpat.toscana.it

tween the two type of observations (on site and remote) of few hours.
Available measures include a variety of turbidity conditions (clean water, algal blooms, river discharge events), with turbidity values ranging from 0.1 NTU to 28.7 NTU. The complete histogram of measured turbidity data distribution is shown in Figure 3.16, highlighting the predominance of low turbidity values.
It should be remarked that higher turbidity scenarios often correspond to extreme or at least intense meteorological events, not suitable for in situ sampling and indeed associated to high cloudiness, when optical satellite images do not provide useful information.

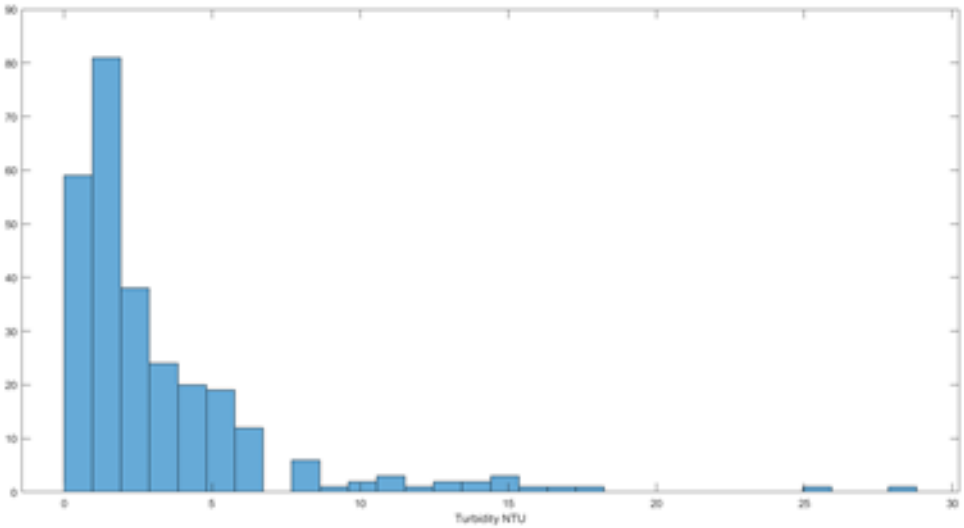


Figure 3.8: Histogram of measured turbidity dataset in NTU used for model development.

3.3. Satellite Data Processing

For the present work, the Sentinel-2 optical images have been used, freely downloaded from Copernicus Open Access Hub website[3]. The choice of Sentinel-2 data was due to its properties: spectral capabilities, wide coverage and short revisit time, but most of all its medium-high spatial resolution. In fact, even though Sentinel-3 mission is designed for ocean and land monitoring, and provides measurements of sea surface temperature, sea surface topography and ocean color through a set of spectral bands with radiometric range and radiometric resolution aimed at an accurate estimation of the inherent optical properties of water, it provides global coverage of oceans at a spatial resolution ranging from 300 to 1000 meters, making it more appropriate for oceanography applications. On the other hand, Sentinel-2, which was designed for land monitoring, has a lower signal-to-noise ratio for water reflected radiance values, but provides global coverage of Earth areas at a spatial resolution up to 10 meters, allowing to study coastal dynamics, for which higher spatial resolution is necessary.

3.3.1. Sentinel-2 products

Sentinel-2 is composed of two identical satellites 2A and 2B, launched in June 2015 and March 2017, respectively, by the European Space Agency. Its sun-synchronous orbit has a mean altitude of 786 Km and an inclination angle of 98.62°.

The two satellites, placed on the same orbit but phased at 180°, allow to acquire images

Band	Spatial resolution [m]	Central wavelength [μm]	Band width [m]	Description
1	60	0.443	20	Coastal aerosol
2	10	0.490	65	Blue
3	10	0.560	35	Green
4	10	0.665	30	Red
5	20	0.705	15	vegetation red-edge
6	20	0.740	15	vegetation red-edge
7	20	0.783	20	vegetation red-edge
8	10	0.842	115	NIR
8a	20	0.865	20	vegetation red-edge
9	60	0.945	20	Water vapour
10	60	1.380	30	Cirrus
11	20	1.610	90	SWIR
12	20	2.190	180	SWIR

Table 3.1: Sentinel-2 spatial resolution for each spectral band.

[3]https://scihub.copernicus.eu.

about every 2 to 3 days at the latitudes of the area of interest, and, in particular, the satellite visits the study area at about 10.30 UTC. The swath width of Sentinel-2 is 290 km (ESA website[4]).
The MultiSpectral Instrument mounted on Sentinel-2 covers 13 spectral bands from visible to short-wave infrared wavelengths, from 10 to 60 m spatial resolution, dependently on the particular spectral band (Table 3.1). The radiometric resolution of Sentinel-2 is 12-bit, this gives a potential range of brightness levels from 0 - 4095.

Products are a compilation of elementary granules, also called tiles, of fixed size, within a single orbit. A granule is the minimum indivisible partition of a product, containing all 13 spectral bands. Image stacks of 13 bands are generated in spatial tiles, ortho-images in WGS84/UTM projection, each one covering a fixed area of approximately 100x100 Km. For the study, tile T32TNP was used, which covers the entire study area with a single image stack (Figure 3.9).

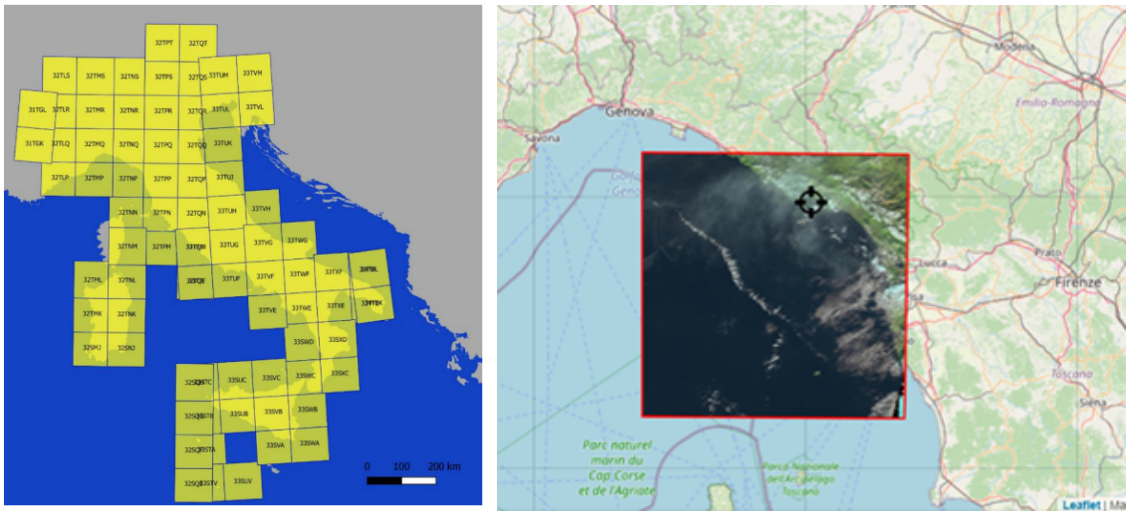


Figure 3.9: Tiles covering Italy (left image) and tile T32TNP covering the study area (right image).

Data are distributed in the Sentinel-SAFE format, that consists in an archive folder (.SAFE) in which all data and added information are stored. The SAFE folder contains the image files in JPEG2000 format for each band, metadata files, quality indicators and auxiliary data.
Since March 2018, the Sentinel-2 processing pipeline produces both Level-1C (L1C) Top Of Atmosphere (TOA) and Level-2A (L2A) Bottom Of Atmosphere (BOA) products:

- L1C products are the result of radiometric and geometrical corrections. Images are otho-rectified and georeferenced using a Digital Elevation Model (DTM) to project them in a WGS84/UTM cartographic projection. Per-pixel measurements are provided as TOA reflectance;

[4]https://sentinel.esa.int/web/sentinel/missions/sentinel-2

Figure 3.10: Comparison of a True Color Image in the study area obtained from Level-1C (left image) and Level-2A (right image) Sentinel-2 MSI products.

- L2A products are the result of atmospheric correction performed by the specific image processor *Sen2Cor* to the LC1 images. Other image corrections like cirrus and topographic corrections are applied.

An example of TOA and BOA products for tile T32TNP is presented in Figure 3.10, using a TCI built from the bands B02 (Blue), B03 (Green) and B04 (Red).

The effect of atmosphere is to reduce the apparent resolution of satellite imagery and to cause errors in the classification of surface fields.
The possibility of using the L2A BOA product was considered: different methods to apply the atmospheric correction were tested and compared with the aim of identifying the most suitable algorithm. The analysis is described in the following section.

3.3.2. Atmospheric Correction

Atmospheric Effects

The electromagnetic radiation from the sun that is reflected by the Earth and detected by the satellite sensor must pass through the atmosphere twice, once on its journey from the sun to the Earth and second after being reflected by the surface of the Earth back to the sensor.
Interactions of the direct solar radiation and reflected radiation from the target with the atmospheric constituents are called as atmospheric effects. This interaction with the atmosphere is important to remote sensing for two main reasons: first, it can be used to obtain useful information about the atmosphere itself; second, when studying the Earth's surface we need to take into account that the signal measured by the sensor, the TOA reflectance, is composed by the reflectance of the target surface, or BOA, perturbed by contributions from the atmosphere:

$$\rho_{TOA}(\lambda) = \rho_{BOA}(\lambda) + \rho_{atm}(\lambda)$$

The effect of atmosphere is mainly related to scattering and absorption of sunlight by atmospheric constituents, e.g. molecules and particles of haze, aerosol, water vapor, and other atmospheric components.
Both scattering and absorption vary in their effect from one part of the spectrum to the other. The gas molecules present in the atmosphere strongly absorb the EMR passing

through the atmosphere in certain spectral bands. Ozone, carbon dioxide and water vapour are responsible for most of absorption of solar radiation.
Also, the effects vary in time and space due to the composition of the atmosphere: the amount of scattering depends upon the size of the particles, their abundance, the wavelength of radiation, depth of the atmosphere through which the energy is traveling and the concentration of the particles, which all vary both in time and over season. Thus the effects of scattering will be uneven spatially and will vary from time to time.

A useful parameter to estimate to which extent the electromagnetic wave is prevented along its propagation path by the presence of aerosol is the Aerosol Optical Depth - AOD. It is a dimensionless parameter related to the amount of aerosol in the vertical column of atmosphere over the observation station. AOD is defined as the natural logarithm of the ratio of incident solar radiation to the radiation that reaches the Earth's surface after passing through a column of atmosphere containing aerosols. It ranges from 0 to 1, where values less than 0.1 corresponds to an extremely clean atmosphere with high visibility, whereas a value greater than 0.4 would correspond to a very hazy condition.
More details on the interaction of the electromagnetic radiance with the atmosphere can be found in Aggarwal, 2004.
Deriving reliable information on the ocean color from remote sensing data is very challenging because of atmospheric influences. When studying water-leaving reflectance, the signal measured by the satellite can be decomposed into several terms (Gordon, 1997):

$$\rho_{TOA}(\lambda) = \rho_r(\lambda) + \rho_a(\lambda) + \rho_{ra}(\lambda) + T(\lambda)\rho_g + t_d(\lambda)\rho_{wc}(\lambda) + t_d(\lambda)\rho_w(\lambda)$$

where $\rho_{TOA}(\lambda)$ is the top of atmosphere reflectance measured by satellite; $\rho_r(\lambda)$ is the Rayleigh reflectance, $\rho_a(\lambda)$ is the reflectance from multiple scattering by aerosols, $\rho_{ra}(\lambda)$ is the aerosol-molecular scattering, $\rho_{wc}(\lambda)$ is the reflectance from whitecaps, ρ_g is the sun-glint reflectance and ρ_w is the water-leaving reflectance; t_d and T are diffuse and direct transmittance, respectively.
In addition to the mentioned parameters, there is another factor that can influence the retrievals, which comes from the proximity of land and scattering of surface-reflected radiance, called the adjacency effect. The process to remove the atmospheric effects and obtain the Earth surface reflectance is known as *Atmospheric Correction* (AC).

Comparison of Atmospheric Correction methods

Atmospheric correction is very challenging and a considerable number of algorithms have been developed over time. An interesting review of available atmospheric correction literature and applications over water surfaces can be found in Hadjimitsis et al., 2004. Atmospheric correction methods can be divided into two main groups:

- Image-based approach;
- Physically-based approach.

Image-based approaches rely only on the spectral and radiometric characteristics of the processed image, from which the atmospheric effects are derived and then subtracted

to the TOA reflectance to obtain BOA reflectance. No atmospheric measurements are required.
Physically-based approaches, instead, use a radiative transfer model which accurately computes the atmospheric effects of scattering and absorption of light and remove them from the sensor observations to obtain the surface reflectance. This approach requires many atmospheric measurements and parameters (e.g. AOD, water vapour, atmospheric and aerosol models, etc.) to parameterise the radiative transfer model. All the required atmospheric information must be coherent in space and time with the processed image.
In general, the physically-based approach is more accurate than the image-based one, but, since the required atmospheric parameters can be unavailable, its use can be challenging (Chavez, 1996).
A comparison of three AC models was made, applying three different open source software:

- The Dark Object Subtraction method, through the *Semi-Automatic Classification Plugin* (Congedo, 2021) in QGIS software (QGIS Development Team, 2022);
- The physically-based Second Simulation of Satellite Signal in the Solar Spectrum (6S) algorithm (Kotchenova et al., 2006), applied using *i.atcorr* in GRASS GIS software (GRASS Development Team, 2022a,b);
- *L2A_AtmCorr* model of *Sen2Cor* processor in the Sentinel Application Platform (SNAP) ver. 8.0.0 software (Lantzanakis et al., 2017), provided by ESA/ESRIN (European Space Agency, 2020), also based on a physically-based algorithm.

The complete work on AC algorithms comparison was presented at FOSS4G2022 (Free and Open Source Software for Geospatial) conference and published in the proceedings with a paper titled "Seawater turbidity analysis from Sentinel-2 images: atmospheric correction and bands correlation" (Pisanti et al., 2022).

Dark Object Subtraction (DOS) model

The Dark Object Subtraction (DOS) model is an example of image-based approach. It is based on the assumption that the lower value of reflectance of the image, that is, the reflectance of the darkest object, is almost only due to the atmospheric effect of scattering (Chavez, 1988). Hence, this method consists in the subtraction of a constant value from each pixel of the processed image, which changes depending on the considered image. The greatest advantage of this method is that it doesn't require atmospheric measurements.
The QGIS Semi-Automatic Classification Plugin (SCP) (Congedo, 2021) allows to apply the DOS method for AC. The only required inputs are the directory containing Level-1C TOA images and the corresponding Metadata file (MTD-MSI file).
The dark pixel assumption for deriving the aerosol contribution can be used in open ocean waters (i.e., the ocean is totally absorbent in the near-infrared bands). However, in optically-complex waters such as turbid coastal waters this hypothesis is not valid anymore (Siegel et al., 2000) and this leads to large errors when retrieving the output products. To deal with this issue, many different algorithms have been developed for the ocean color remote sensors (Pahlevan et al., 2021). These algorithms show strong dis-

crepancies in terms of performance, they do not cover the entire variability of the coastal waters and still need more assessing. Moreover, not all algorithms are open-source, easy to use and well-supported for Sentinel-2 products (Bui et al., 2022).

Second Simulation of Satellite Signal in the Solar Spectrum (6S) algorithm

In GRASS GIS the atmospheric correction is made possible applying the 6S algorithm through the *i.atcorr* module[5], defining the following input data, atmospheric parameters and models:

- Sensor and solar geometries (provided in the metadata);
- Date and time, longitude and latitude (provided in the metadata);
- mean target elevation above sea level (estimated from the input digital elevation model DEM);
- atmospheric model (e.g. Tropical, Mid Latitude Summer /Winter, Subartic Summer/Winter);
- aerosol model (e.g. Continental, Maritime, …)
- Visibility or Aerosol Optical Depth (AOD) at 550 nm.

The most difficult parameters to set is the AOD, as the value is time and site specific and is not provided in the metadata file. This parameter can be estimated from the Aerosol Robotic Network (AERONET), a federation of ground-based globally distributed remote sensing aerosol networks with more than 25 years of data (Holben et al., 1998).

From the data visualisation interface of the AERONET website[6], it is possible to search for the nearest measurement site or one with atmospheric characteristics potentially closest to the site under study and select the AOD value at 550 nm. The value should be measured at the same time as the satellite image was taken.

As no AERONET stations are available close to the study area, the following AERONET stations were selected taking into account proximity to the site, altitude and distance from the coastline (see Figure 3.11):

1. Palma de Mallorca, Spain (altitude of 10 m)
2. CeSMA, Naples, Italy (altitude of 50 m)
3. La Sapienza, Rome, Italy (altitude of 75 m)
4. Modena, Italy (altitude of 125 m)
5. Ispra, Varese, Italy (altitude of 235 m)

[5]https://grass.osgeo.org/grass78/manuals/i.atcorr.html
[6]https://aeronet.gsfc.nasa.gov/new_web/aerosols.html

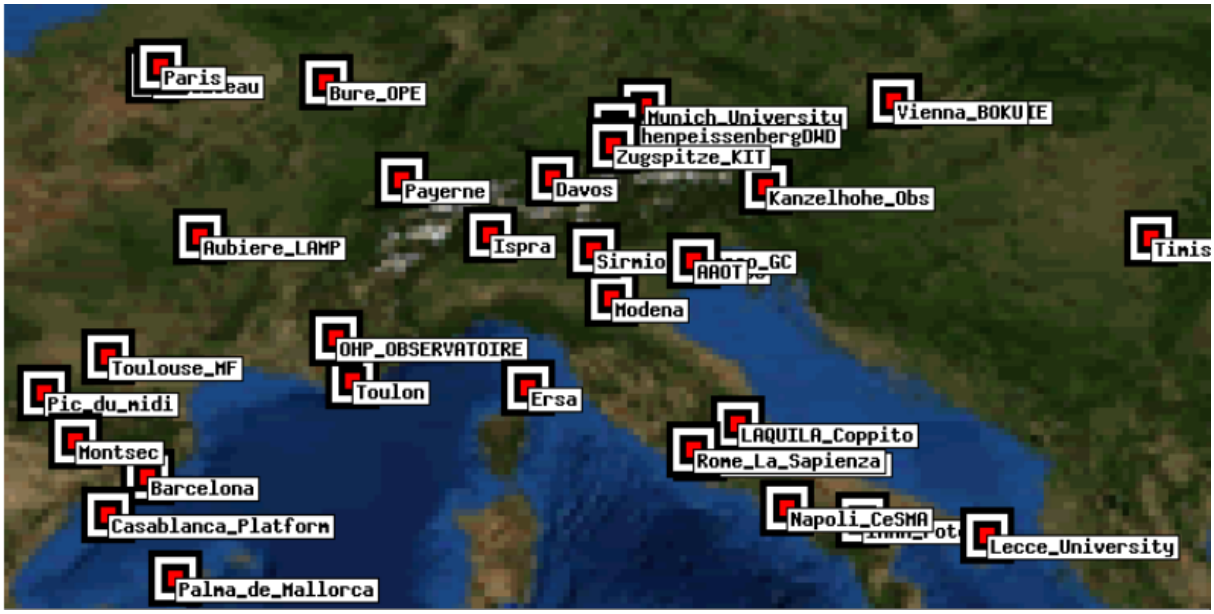


Figure 3.11: AeroNet stations. Source: *https://aeronet.gsfc.nasa.gov*

The first three stations are located near the coast, but far from the study area, while the last two are closer to the study area even though they have higher altitude.
The AOD variability during the day in the different stations can be high and measurements in each station are not continuous nor available everyday or at required hour of the day. For instance, Figure 3.12 shows the daily variability of the AOD parameter on May 06th, 2020 for the different stations. It can be noticed that for the specific day,

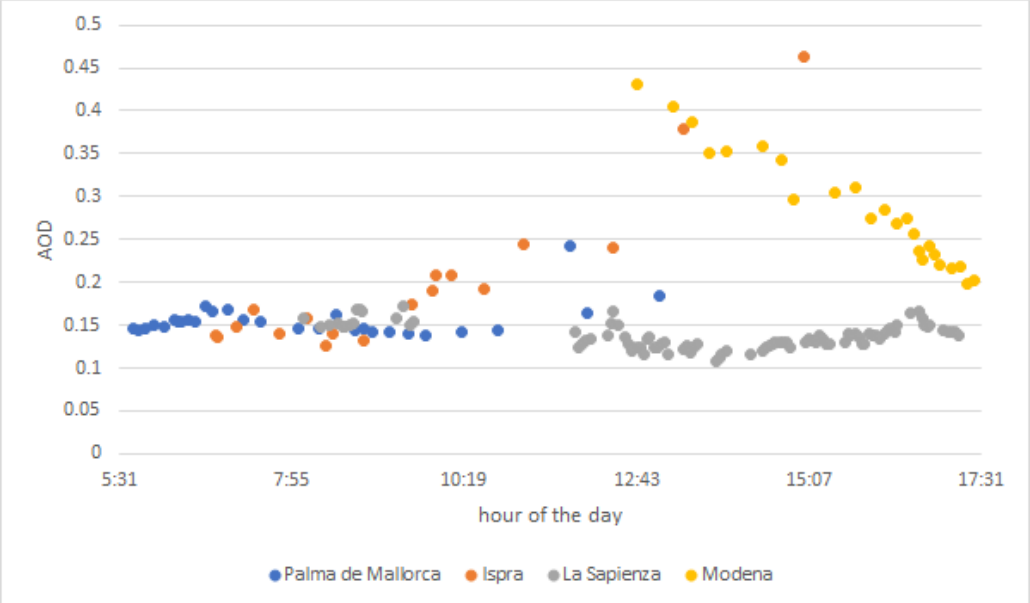


Figure 3.12: Example of AOD daily and spatial variability (e.g for the 06th may 2020).

CeSMA-Naples data were not available.
The different values registered in the different stations were used to perform a sensitivity analysis of *i.atcorr* to the AOD value. Figure 3.13 shows an example on how small variations of the parameter lead to significant differences in the resultant BOA reflectance. It can be noticed that the value of 0.15, which could be assumed by analysing the observations from the selected AERONET stations (Figure 3.12) lead to a BOA value far from the Sentinel-2 L2A product, considered as ground truth.
The difficulty of identifying the proper AOD parameter for the model, which might not have a temporal correspondence to a satellite overpass time, makes its use questionable.

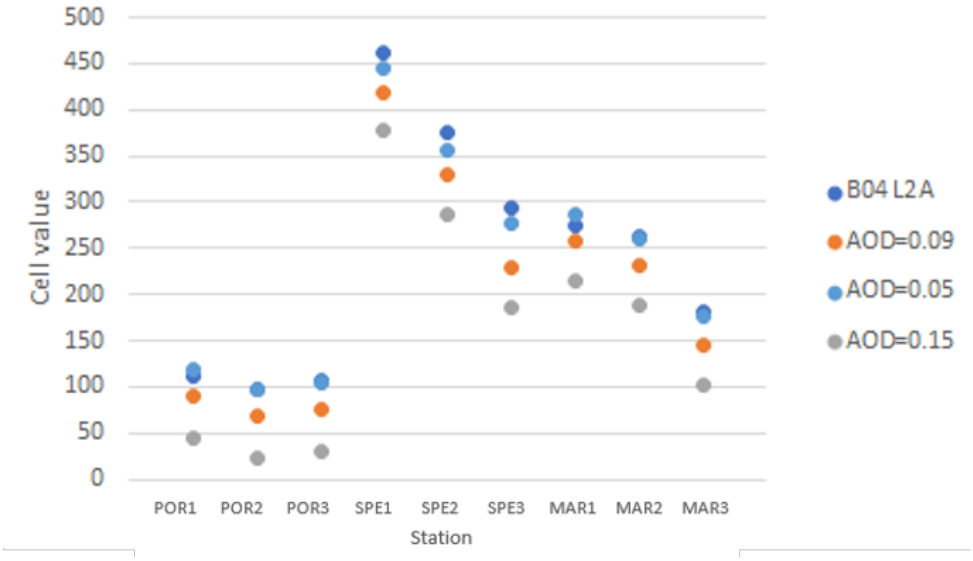


Figure 3.13: Comparison between Sentinel-2 L2A BOA and L1C TOA reflectance for e.g. the red band (B04) obtained using *i.atcorr* with different AOD values. Results are shown for different pixels, corresponding to different in situ stations.

L2A_AtmCorr

The Sentinel-2 Level-2A BOA reflectance products are obtained from Level-1C TOA products through the atmospheric correction processor *Sen2Cor*, developed by Telespazio VEGA Deutschland GmbH on behalf of ESA (Louis et al., 2016).
The L2A_AtmCorr algorithm runs on the Sentinel Application Platform (SNAP), an architecture for all Sentinel processing and analysis, published under the GPL license and with its sources code available on GitHub.
It is a physically-based approach based on an algorithm proposed in Atmospheric/Topographic Correction for Satellite Imagery (Richter and Schläpfer, 2011); it is dependent on the computation of radiative transfer functions for different sensor and solar geometries, ground elevations and atmospheric parameters, through the libRadtran radiative transfer model (Mayer and Kylling, 2005).

The model is run once to generate a large LUT of sensor-specific functions required

for the AC (path radiance, direct and diffuse transmittances, direct and diffuse solar fluxes, and spherical albedo) that accounts for a wide variety of atmospheric conditions, solar geometries and ground elevations. This database is generated with a high spectral resolution (0.6 nm) and then resampled with Sentinel-2 spectral responses. This LUT is used as a simplified model (running faster than the full model) to invert the radiative transfer equation and to calculate BOA reflectance.

All gaseous and aerosol properties of the atmosphere are either derived by the algorithm itself or fixed to an a priori value. The AOD parameter can be derived from the images themselves above reference areas of known reflectance behaviour, preferably Dark Dense Vegetation (DDV) targets and water bodies. Water vapour retrieval over land is performed using the Atmospheric Pre-corrected Differential Absorption (ADPA) algorithm which is applied to band 8a and band 9. The processor reads the parameter in form of Look Up Tables (LUTs) pertaining to this parameter space and interpolates, if required. The LUTs have been generated via *libRadtran*, a library for the calculation of solar and thermal radiation in the Earth atmosphere.
Sen2Cor needs the input of the directory containing the metadata file (MTD-MSI) of the satellite image and the LUTs can be automatically determined by the processor, or they can be configurable by the user in the configuration file L2A*GIPP.xml* located in the *cfg* folder of the subdirectory of the processor.
The atmospheric correction can be obtained changing few parameters of the L2A_*GIPP.xml* with respect to the default ones. *Sen2Cor* additional outputs are an Aerosol Optical Thickness (AOT) map, a Water Vapour (WV) map and a Scene Classification (SCL) map with Quality Indicators for cloud and snow probabilities (Figure 3.14).

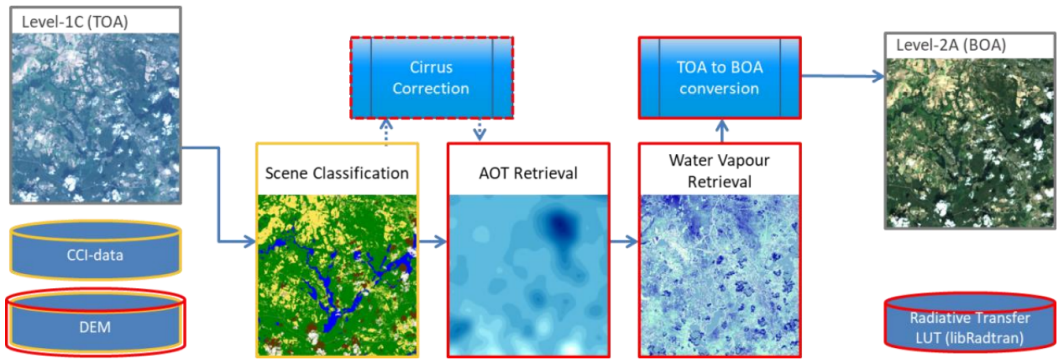


Figure 3.14: Flow of L2A_AtmCorr *Sen2Cor* AC processing chain.

The use of L2A_AtmCorr *Sen2Cor* is therefore quite simple. However, it must be highlighted that it is not designed for AC above water, as it does not estimate the aerosols over water and does not correct for the sun-glint effect. Therefore results obtained can not be considered accurate without any validation.
A critical appraisal of the results, and of problems encountered in applying the various methods, leads to an evaluation of their value in practice. The application of AC methods requires the availability of local parameters, otherwise the inherent error introduced

can be significant still not quantifiable. The accuracy of the Sentinel-2 BOA products obtained through different AC processors should be, in fact, validated with field spectrometry data.
Inaccurate AC leads to large uncertainties in satellite data products, with effects on quantitative retrieval of optically active water quality parameters (Pahlevan et al., 2020). The Atmospheric Correction Intercomparison Exercise, ACIX-Aqua, (Pahlevan et al., 2021) examined the performance of eight state-of-the-art AC methods available for Landsat-8 and Sentinel-2 data processing over lakes, rivers, and coastal waters. The work highlights that performance of the different methods depends on optical water type, and that best-performer's uncertainties across the different bands are not uniform. Finally, the work attempted to evaluate how uncertainties in AC impact on Chl-a and SPM retrieval. The work reveals that despite high improvement in processors available, AC over inland and coastal waters remains one of the major challenges in aquatic remote sensing.
As no atmospheric parameters, coherent in space and time with the processed images, are available for the present study and no in-situ optical data is available to support the choice of an AC algorithm, to avoid uncertainties in BOA products, Level-1C TOA product was considered to train the predictive model.
This choice is one of the mot innovative aspect of the proposed approach and is supported by literature review.
In fact, while physically-based relationships require robust atmospheric correction to obtain BOA reflectance, ML algorithms are less affected by the atmospheric and other background factors (Chebud et al., 2012). An examination of relevant published work, reports some satellite-based methods for the detection, for instance, of Sea Surface Salinity (Medina-Lopez, 2020) or Harmful Algal Blooms (Binding et al., 2021) that rely on Level-1 TOA quantities in order to avoid large uncertainties in Level-2 BOA products introduced by poor AC.

3.3.3. Cloud detection

A significant obstacle in the use of satellite optical data is related to the presence of clouds as they interfere with this process by scattering and absorbing the EMR.
Temporal resolution of images is indeed related to the weather conditions during the image acquisition, and cloud coverage masking must be applied to filter data and avoid wrong interpretation of the reflectance. Omission of clouds can lead to errors that propagate to images processing, whereas over detection of clouds can lead to a reduced number of valid observations.
Clouds are generally divided into two types: a thick and almost completely opaque (dense cloud) and a light, semi-transparent cloud (cirrus cloud). The most difficult to manage are cirrus, because they can alter the data without being clearly confirmed by visual inspection; also, their signals are strongly affected by the reflectance of surfaces underneath.
The detection of cloud shadows is even more difficult (Zhu and Woodcock, 2012) as they have very similar spectral signature to other dark covers.
Cloud and shadow detection methods can be divided into two main approaches: single date and multi-date. The former utilize the intrinsic characteristics of a single satellite

image and rely mainly on physical properties of cloud reflectance. The latter exploits time series or computes differences of the same scene with respect to a cloud-free reference (Zhu and Woodcock, 2014). The multi-date approaches provide more accurate results but need several images and require cloud free reference images.
Development of cloud masking algorithms remains an area of active research in the remote sensing community.
The Sentinel-2 L1C product contains rough pixel classification maps for both phenomena, using spectral bands thresholds based on cloud physical properties to identify potential cloud pixels and clear-sky pixels. However, there are several reliable tools available in literature which are the result of deep statistical analysis of large amounts of data, capable of classifying pixels in a multispectral image according to their type (water, soil, clouds, snow, etc.). A comprehensive review of many cloud masking algorithms for Sentinel-2 images can be found in Skakun et al., 2022.
For the present work, the so-called *s2cloudless* algorithm (Zupanc, 2017) was chosen, which allows to get a precise cloud coverage masking for MSI images, obtained on the basis of supervised learning conducted by the authors on thousands of tiles in different operating conditions. The *s2cloudless* algorithm is single date, pixel-based classification, where we assign each image pixel a probability being covered with a cloud solely based on satellite's spectral response for that pixel. It does not consider any spatial context, and therefore can be executed at any resolution. The *s2cloudless* algorithm can, unlike many others, be executed also on averaged Sentinel-2 reflectance values over arbitrary user-defined geometries and still provide meaningful results.
The algorithm was developed by the EO Research team at Sinergise and is published under the MIT License on *https://github.com/sentinel-hub/sentinel2-cloud-detector*.
The statistical model is based on the Gradient Boosting Decision Tree (GBDT) technique and relies on a widely used Python package such as *LightGBM* (Ke et al., 2017).
Gradient Boosting works on the principle that many weak learners can together make an accurate predictor. It involves three elements to be defined:

1. a loss function to be optimized, that must be differentiable;

2. weak learners to make predictions, usually shallow decision trees;

3. an additive model to add weak learners to minimize the loss function.

Specifically, regression trees are used that output real values for splits and whose output can be added together, allowing subsequent models outputs to be added and correct the residuals in the predictions. A gradient descent procedure is used to minimize a set of parameters; after calculating error, the parameters are updated to minimize that error, moving in the right direction by reducing the residual loss. *LightGBM* has the benefit of being faster that conventional GBDT while achieving almost the same accuracy.

The technique has many hyper-parameters to be tuned, in particular the hyper-parameters that were not set to their default values are:

1. min_data_in_leaf =1500;

2. n_estimators =170;

3. num_leaves = 770.

The input features of *s2cloudless* algorithm are Sentinel-2 Level-1C TOA reflectance values of the following ten bands: B01, B02, B04, B05, B08, B8A, B09, B10, B11, B12 and output of the algorithm is a cloud probability map. Users of the algorithm can convert the cloud probability map to a cloud mask by defining a threshold in the cloud probability map.
An example of the obtained contour map of cloud masked pixels, with a threshold of 50% probability, is shown in Figure 3.15, overlaid with an original TCI image.
The depicted contours also include the coastline, derived from openly available cartography, and the delimitation of the valid part of the image, due to the possible cut of the orbit inside the tile itself.
The threshold of 50% probability of cloud detection was used to selection the training data, corresponding to 222 observations.

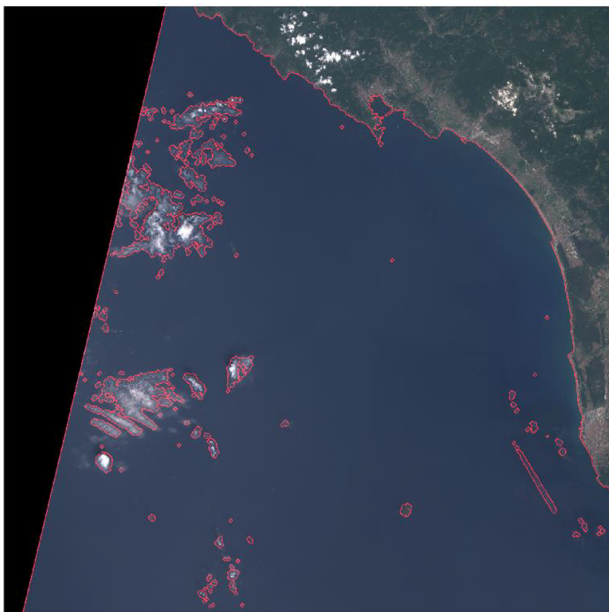


Figure 3.15: Cloud masking by *s2cloudless* algorithm.

3.4. Turbidity Model implementation

As discussed, after data selection and preprocessing, the first step for the model implementation is represented by the selection of a suitable algorithm, based both on the problem and the characteristics of the data. The possible complexity of the model derives, in fact, from the nature and number of training data.

3.4.1. Dataset description

We hereby briefly recall the available dataset composition and characteristics, while all detailed information can be found in the previous Chapters.

As previously described (see Chapter 3.1.), turbidity values used to build the model range from 0.1 NTU to 28.7 NTU, but they are highly unbalanced towards low values, with a median of 2.0 NTU and a mean of 3.4 NTU (Figure 3.16).

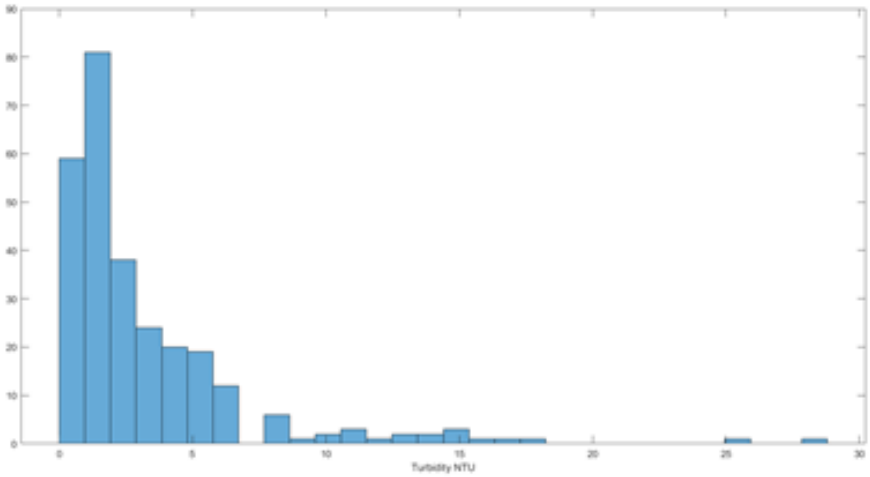


Figure 3.16: Histogram of measured turbidity data in NTU.

Unbalanced datasets can pose several problems for ML models, including:

- Biased predictions: Since the model has seen more examples of one class than the other, it may become biased towards the majority class and tend to predict that class more often. This can result in lower accuracy for the minority class, which may be of greater interest.
- Lack of representative examples: In some cases, the minority class may be underrepresented in the dataset, making it difficult for the model to learn useful patterns. This can lead to poor performance on the minority class, even if the model is not biased towards the majority class.
- Overfitting: If the model is trained on an unbalanced dataset, it may overfit to the majority class and fail to generalize well to new examples. This can lead to poor

performance on new data, especially if the distribution of classes in the test set is different from the training set.

To address these issues, simpler algorithms less prone to overfitting should be preferred, and techniques specifically useful for small or imbalanced datasets can be used, such as Leave One Out Cross Validation.
The 13 MSI bands coming from L1C (TOA) imagery were selected (Chapter 3.3.2.) as predictive variable for turbidity, directly downloaded from Copernicus Open Access Hub website[7]. Cloud masking was applied on a pixel basis to eliminate those pixels affected by the effects of clouds (Chapter 3.3.3.).
The final database for training and validation was composed of 222 combinations of turbidity measurements and TOA spectral data from Sentinel-2A/B MSI.

Figure 3.17 shows the boxplot of the 222 set of spectral bands values resulting after cleaning operations and representing the final predictors.
Box plots divide the data into four quartiles, with the middle 50% of the data falling between the first and third quartiles (the box). The value dispersion is larger in the visible bands (B02, B03, B04), and in band B01, affected by Coastal Aerosol. Some outliers can be observed, indicating the possible presence, even after cloud detection, of data corrupted by some other effects.

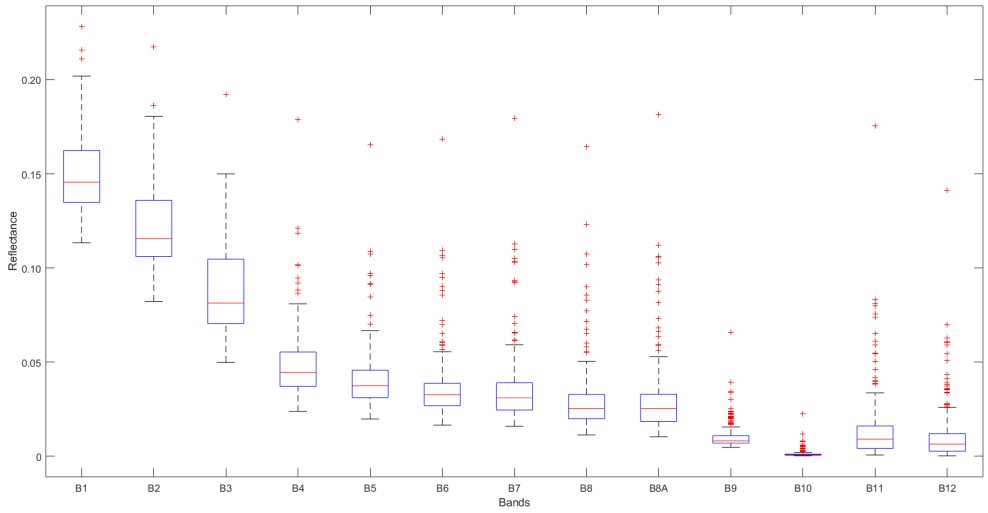


Figure 3.17: Boxplot of the values of the 13 TOA spectral bands for the 222 pixels used as predictors for turbidity estimates.

A min-max normalization was applied to predicting variables in order to scale the values of each feature to a range between 0 and 1. Data normalization is a common pre-processing step to rescale the values of numerical features to a common scale. This is important because features with large values can dominate the algorithm's objective function.

[7]https://scihub.copernicus.eu

3.4.2. Model selection: the Kernel Regularized Least Squares

As previously discussed (see Chapter 1.3.), for any given training set, the best choice for the complexity of the model would be the one striking the optimal trade-off between bias and variance.
Due to the limit of the available dataset, in terms of quantity and imbalance, for the present work a Kernel Regularized Least Squares model was selected, characterized by a simple model structure, good generalization, global optimal solution, especially suitable for non-linear and high dimension problems. The theoretical basis of the model are hereby described.

Feature Map and Kernels

The use of feature maps and kernel allows to go beyond linear models, still applying a simple model structure. Given the vector x of our input (or predictive) variables, we consider the feature map Φ(x) mapping the input vector x from the input space X (see Chapter 1.3. for more details on statistical learning theory) to a new space F, also called feature space, that can potentially be infinite dimensional:

$$\Phi(x) : X \to F \tag{3.1}$$

If we choose a feature map $\Phi(x)$ obtained as a non-linear function of our input vector x, we can classify our data by a linear model in the feature space, while the model is no longer linear in the original input space:

$$f_{\omega}(x) = \omega^T \Phi(x)$$

In this way, we can consider nonlinear models applying computations that are essentially the same as in the linear case, as long as we change $x \in R^D$ into $\Phi(x) \in R^p$ (with p > D).
The Kernel method is based on the so-called "kernel trick", that is, given a kernel function K (x, x'), symmetric and positive definite, we will implicitly have a nonlinear function $\Phi(x)$ only through the inner product of the input points x. The advantage of using a kernel function as a similarity measure is that it allows us to construct algorithms in dot product spaces. Popular examples of positive definite kernels include linear kernel, polynomial kernel and Gaussian kernel.

Polynomial kernel was used (eq. 3.2), where the degree of the polynomial d ($d \in N$) is a hyper-parameter to be tuned, by minimising the loss over the available training samples.

$$K(x, x') = (x^T x' + 1)^d \tag{3.2}$$

The polynomial kernel of degree d thus computes a dot product in the space spanned by all monomials up to degree d in the input coordinates. More details on Kernel methods can be found in Hofmann et al., 2008.

Kernel Regularized Least Squares

As introduced, the use of a Feature map allows to consider nonlinear models applying computations for linear models. Let's consider for example Regularized Least Squares

(RLS), an algorithm defined by the square loss:

$$\min_{\omega \in R^D} \frac{1}{n} \sum_{i=1}^{n} \left(y_i - \omega^T x_i\right)^2 + \lambda \omega^T \omega, \qquad \lambda \geq 0 \tag{3.3}$$

where the term $\omega^T\omega$ is a regularizer and helps preventing overfitting by controlling the stability of the solution; the parameter λ balances the error term and the regularizer. Algorithm (3.3) is an instance of Tikhonov regularization, also called penalized empirical risk minimization. The space of possible solution, or hypothesis space, is therefore the space of linear functions:

$$H = \left\{ f : \mathbb{R}^D \to \mathbb{R} : \exists \omega \in \mathbb{R}^D \mid f(x) = x^T \omega, \ \ \forall x \in \mathbb{R}^D \right\},$$

so that finding a function f_ω(x) reduces to finding a vector ω.
When introducing the feature map $\Phi(x)$, we will have:

$$f_\omega(x) = \omega^T \Phi(x) = \sum_{j=1}^{p} \omega_j \Phi(x)_j$$

which, as result of the representer theorem, further becomes:

$$\hat{f}(x) = \sum_{j=1}^{p} K(x_j, x) c_j$$

where $x_1, ..., x_n$ are the inputs in the training set and c = $(c_1, ..., c_n)$ a set of coefficients. For the square loss function, the vector c of coefficients can be computed solving the following linear system:

$$(K_n + \lambda n I)c = Y_n.$$

where Y_n is the vector of observed output variables, and I is the identity matrix.
The running time of the algorithms varies from model to model and it also depends on the inputs and amount of fine-tuned hyper-parameters. The computation cost and memory cost of KRLS can be expressed as functions of the dimensionality of the training set, d, and the number of training samples, n.
The computation cost of KRLS is dominated by the computation of the kernel matrix, which has a computational complexity of O(n^2d) for a standard kernel. The memory cost of KRLS is proportional to the size of the kernel matrix, which is O(n^2). Since the kernel matrix is symmetric, it is typically stored as a lower triangular matrix or a vector, reducing the memory requirements to O(n). Main characteristics of the KRLS is given in Table 3.2.
As anticipated, KRLS algorithm is particularly well-suited for non-linear and high-dimensional problems. We can summarize the characteristics that make it a powerful algorithm:

- KRLS has a simple model structure, consisting of a linear combination of kernel functions applied to the training examples. This simple structure makes it easy to interpret and understand the model.

Train Cost	Test Cost	Train Memory	Test Memory
O(n^2d)	*O(nd)*	*O(n^2)*	*O(nd)*

Table 3.2: Main characteristics of the regularized kernel regression (KRLS): the order O of the computational and memory cost for training and testing as a function of dimension d (spectral bands), number of samples n for training.

- KRLS is able to model non-linear relationships between the input features and the output variable. This is due to the use of kernel functions, which can map the input features into a higher-dimensional space, where non-linear relationships can be more easily modeled.
- KRLS has a closed-form solution that guarantees a global optimum. This means that it is able to find the best possible solution to the problem, rather than getting stuck in a local optimum.
- KRLS has good generalization properties, meaning that it is able to make accurate predictions on new, unseen data. This is due to the use of regularization, which helps to prevent overfitting and improve the model's ability to generalize. This is particularly useful in problems with few or unbalanced training data.
- KRLS is able to handle high-dimensional problems, where the number of input features is much larger than the number of training examples. This is due to the use of kernel functions, which can project the input features into a higher-dimensional space, where the problem can be more easily solved.

3.4.3. Model calibration and validation

The choice of the model parameters (hyperparameter tuning) aims to achieve an empirical bias variance trade-off. This is obtained through a cross validation procedure, used to compute a hold-out error, obtained from part of the training set, to be used as a proxy of the expected error and to be minimized. Cross-validation is a verification technique that evaluates the generalization ability of a model for an independent dataset, and is used for testing the model during the training phase for overfitting. Cross-validation can also be used to evaluate the performance of the prediction function. It is a process for creating a distribution of pairs of training and test sets out of a single data set.
In k-fold cross-validation, the training dataset is arbitrarily partitioned into k mutually exclusive non-overlapping subsets, or folds. The model is trained k times, where each iteration uses one of the k subsamples for testing (cross-validating), and the remaining k-1 subsamples are applied toward training the model. The k results of cross-validation are averaged to estimate the accuracy as a single estimation (Figure 3.18).
A Leave-One-Out Cross Validation (LOOCV) procedure, a special case of k-fold cross-validation, has been used throughout the study to assess the prediction performance in an objective way (Efron, 1982). For the LOOCV, the learning algorithm is applied once for

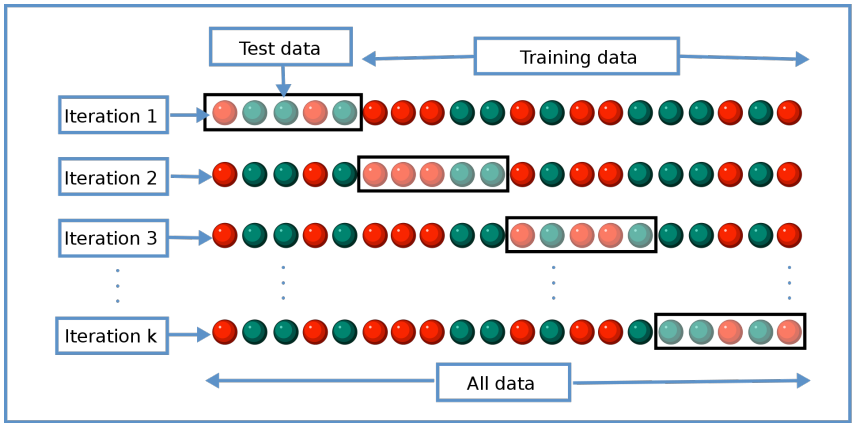


Figure 3.18: Sketch of k-Fold cross-validation (Source: https://commons.wikimedia.org).

each instance, using all other instances as a training set and using the selected instance as a single-item test set. The method can be computationally expensive for large data sets, however, it has the advantage of using all available data for training, and its estimates of model performance tend to have low bias, but high variance. This method is particularly useful when the available dataset is small. In these cases, LOOCV can provide a more accurate estimate of the model's performance than k-fold CV, because it uses all but one data point for training, providing a more representative estimate of the true performance of the model.

3.4.4. Model performance evaluation

To determine the strength of the relationships, the coefficient of determination (R^2) was used, while root mean squared error (RMSE) and the mean absolute error (MAE) were used to determine model accuracy. The cited coefficients are calculated as:

$$RMSE = \sqrt{\sum_{i=1}^{N}(y_i - x_i)^2}$$

$$MAE = \frac{\sum_{i=1}^{N} |(y_i - x_i)|}{n}$$

where y_i is the value predicted by the model and x_i the values observed. The Normalized Root Mean Square Error (NRMSE), which relates the RMSE to the observed range of the variable, and the symmetrically normalized root square error (HH) proposed by Hanna and Heinold, 1985 were also calculated, applied per class of values, as these indexes have shown to provide a more reliable information than RMSE about the accuracy of the results of numerical models (Mentaschi, 2013). The HH index is defined as the RMSE divided by

the absolute value of the mean of the product of the observations and modelled values:

$$NRMSE = \sqrt{\frac{\sum_{i=1}^{N}(y_i - x_i)^2}{\sum_{i=1}^{N}(x_i)^2}}$$

$$HH = \sqrt{\frac{\sum_{i=1}^{N}(y_i - x_i)^2}{\sum_{i=1}^{N}(y_i x_i)}}$$

Being normalized, the error indicators are useful metric for comparing datasets with different scales.
Finally, a comparison among turbidity estimates obtained from the developed models, data from Copernicus CMEMS dataset named 'Mediterranean Sea, Bio-Geo-Chemical, L3, daily observation', extracted for the days of the surveys in the pixel corresponding to the sampling location, and in-situ measurements was conducted.

3.4.5. Feature Importance

Input data of a ML algorithms consist of a set of variables, that can be combined as input features. Features often become one of the most important components in ML modeling. Basically, model predictions directly depend on the quality of features; as a result, a special emphasis is placed on feature engineering and feature selection. To know which features contribute more to the model and select or create new features, it's necessary to measure their importance somehow. Feature (or variable) importance indicates how much each feature (or variable) contributes to the model prediction. Basically, it determines the degree of usefulness of a specific variable for a current model and prediction.
Feature importance scores play an important role in a predictive modeling project, including providing insight into the data, insight into the model, and the basis for dimensionality reduction and feature selection that can improve the efficiency and effectiveness of a predictive model on the problem.
There are many ways of calculating feature importance, but generally, we can divide them into two groups:

- Model dependent methods, specific to one particular ML model (e.g. impurity-based importance in decision trees);
- Model agnostic methods, they are not specifically related to one particular ML method, but can be applied to most of the ML models that solve one problem (e.g. regression).

Examples of model agnostic feature importance methods are:

- **Correlation criteria**. The most simple way to measure the feature importance is to correlate them with target variables. This is possible when the target has

continuous value. Basically, correlation is a statistical measurement that quantifies the relationship between two variables. The most popular correlation methods are Pearson and Spearman correlation. Pearson correlation measures linear dependence between two variables while Spearman measures monotonicity, and they both might assume values between -1 and 1. However, they only relate to linear correlation.

- **Permutation Feature Importance**. It is a model inspection technique that can be used for any fitted estimator when the data is tabular. This is especially useful for non-linear or opaque estimators. The permutation feature importance is defined to be the decrease in a model score when a single feature value is randomly shuffled (Breiman, 2001).This procedure breaks the relationship between the feature and the target, thus the drop in the model score is indicative of how much the model depends on the feature.

The permutation importance algorithm can be schematized as follow. Given:

- the fitted predictive model M,
- a tabular dataset D_{dj} used for training and/or validation,
- the calculated reference score s of the trained model on data (for instance the R^2, MAE or RMSE for a regressor);

For each feature f_j , column of D:

- randomly shuffle the column f_j to generate a corrupted version of the data named $\tilde{D_{dj}}$. Compute the score s_{nj} of model on corrupted data $\tilde{D_{dj}}$ for each of the N repetitions;
- Compute importance i_j for feature j defined as:

$$i_j = s - \frac{1}{N}\sum_{n=1}^{N}(s_{nj})$$

Permutation importance reflects how important this feature is for a particular model rather than the intrinsic predictive value of a feature by itself. The scores can be used in a range of situations in a predictive modeling problem, such as:

- Provide insight into the dataset. The relative scores can highlight which features may be most relevant to the target, and the converse, which features are the least relevant. This may be interpreted by a domain expert and could be used as the basis for gathering more or different data.
- Better understanding a model. Most importance scores are calculated by a predictive model that has been fit on the dataset. Inspecting the importance score provides insight into that specific model and which features are the most important and least important to the model when making a prediction. This is a type of model interpretation that can be performed for those models that support it.
- Reducing the number of input features. This is a type of feature selection and can simplify the problem that is being modeled, speed up the modeling process (deleting features is called dimensionality reduction), and in some cases, improve the performance of the model.

For the proposed model, the predictive features are simply the 13 TOA spectral bands, thus a feature importance analysis provides an estimate of the relative importance of each band for the predictive model.
For the purpose, a permutation analysis was conducted: after random re-orderings (shuffling) of the predictive variable the test statistics is recalculated (Anderson and Robinson, 2001).
The statistics used are the MAE and RMSE, in accordance to the statistic used for the model evaluation (see Chapter 3.5.4) and the number of runs used to achieve stable feature rankings was N = 30.

The above reported methodology, as well as the following results, are published in *Remote Sensing Applications: Society and Environment* with the title "Application of machine learning techniques to derive sea water turbidity from Sentinel-2 imagery" (Magrì et al., 2023).

RESULTS AND DISCUSSION

4.1. Turbidity model evaluation

As previously discussed in Chapter 3.4., the degree d of the polynomial KRLS is the hyper-parameter to be tuned, by minimising the square loss function. To ensure that the minimization problem is well defined, we must add a regularisation term.
The hyper-parameter tuning was obtained through a k-fold cross validation procedure, with k = 10, while a LOOCV was used to evaluate the performance of the calibrated prediction function.
A 3-degree polynomial Kernel showed the best cross-validated fitting with measured data. The tested KRLS shows good agreement of the estimated satellite-derived surface turbidity with measured one, with $R^2 = 0.725$, RMSE = 2.07 NTU and MAE =1.40 NTU. Figure 4.1 shows the scatter plot depicting cross-validated turbidity (NTU) estimations vs. in situ measured values.

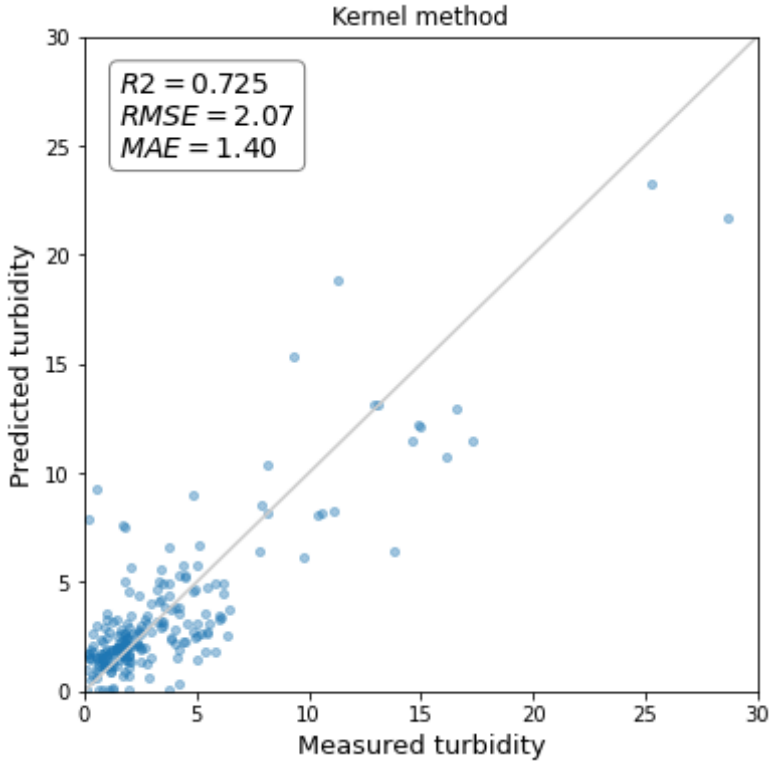


Figure 4.1: Scatter plot of the Measured vs Predicted turbidity values (NTU) with the KRLS model.

From the plot, we can notice a higher dispersion occurs in particular for medium values of turbidity (> 10 NTU).
The Normalized Root Mean Square Error (NRMSE) and the symmetrically normalized root square error (HH) were calculated for the different ranges of the observed values, as shown in Table 4.1.
It is possible to notice that the two indexes give similar results: the KRLS model provides

Turbidity (NTU)	# data	MAE (NTU)	RMSE (NTU)	NRMSE	HH
0-2	114	1.0	1.7	1.26	1.07
2-5	70	1.2	1.6	0.44	0.46
5-10	23	2.4	2.7	0.41	0.45
> 10	15	3.7	4.4	0.27	0.30

Table 4.1: Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Normalized Root Mean Square Error (NRMSE) and the symmetrically normalized root square error (HH) for different ranges of turbidity values.

an average NRMSE value of 0.40 and an HH index of 0.43, with higher relative error for the lowest values of turbidity. In fact, while the RMSE increases for higher values of turbidity, the normalised errors NRMSE and HH decrease.
The table also reports the number of record available for each range of turbidity values: as previously marked, low values represents the most numerous class. Values higher than 20 NTU, which represent extreme events, were included in the class of values higher than 10 NTU, as their quantity is too small to infer statistical consideration.
Residuals, obtained as the difference between the predicted value and the actual value for a given observation, were calculated and used to build the histogram of residuals (Figure 4.2).

By examining the shape and distribution of the histogram, we can gain insights into how well the model is making predictions and identify some potential issues that need to be addressed.
A normal distribution centered around zero value can be observed, this means that the model's predictions are, on average, unbiased. The histogram is symmetrical, this indicates that the model is making predictions that are equally likely to be too high or too low. Most residuals have values in +/-2 NTU as expected, while only a few exceed this range.

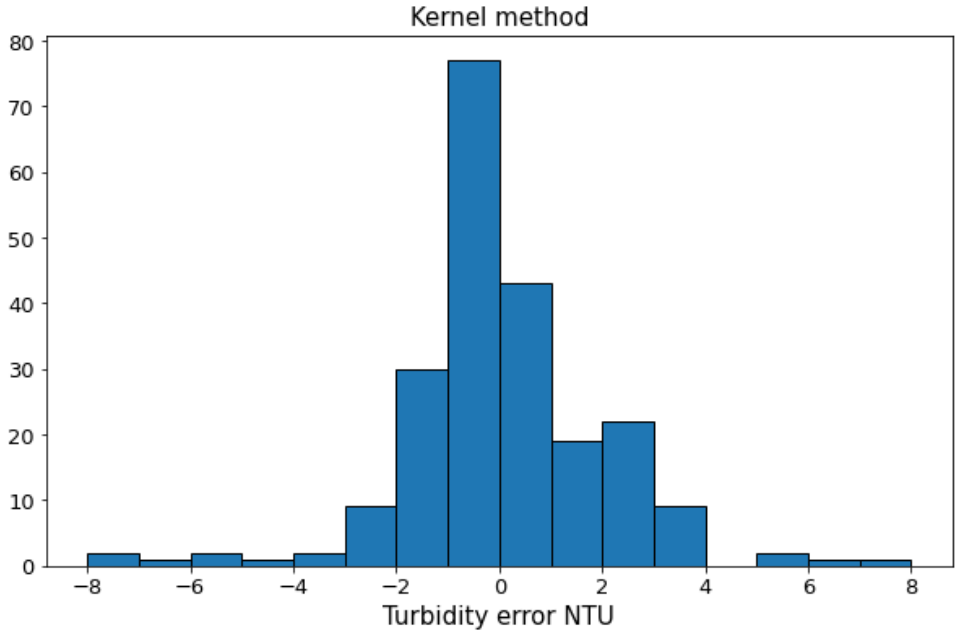


Figure 4.2: Histogram of residuals in NTU.

As Data Driven Models are highly influenced by the training data, the availability of more data, especially in the less represented class (medium-high turbidity), is expected to improve the model accuracy.

4.2. Spectral bands importance

A permutation analysis was conducted to estimate the relative importance of each band for the proposed predictive models: after random re-orderings of the predictive variable the test statistics is recalculated (Anderson and Robinson, 2001). The statistic used are the MAE and the RMSE; the number of runs used to achieve stable feature rankings was N = 30. Chapter 3.4.4. describes in detail the procedure applied.
Figures 4.3 and 4.4 show the increase in MAE and RMSE after shuffling each of the bands: the higher the increase in the error value when a particular band is shuffled, the more weight it has in the model.
The ranking of the features is approximately the same for the two different metrics: bands B05, B06, B07 in the red-edge part of the spectrum (705-783 nm) caused a higher MAE and RMSE when shuffled compared with the other bands, followed by the B8A (865 nm) and the B04 (665 nm - Red) bands. This result shows importance of bands in the Red-NIR range (665 nm - 865 nm), which agrees with the physical knowledge about the bands contributing more to turbidity determination (Chapter 2); moreover it revels the importance that bands in the red-edge can be played in turbidity retrieval.

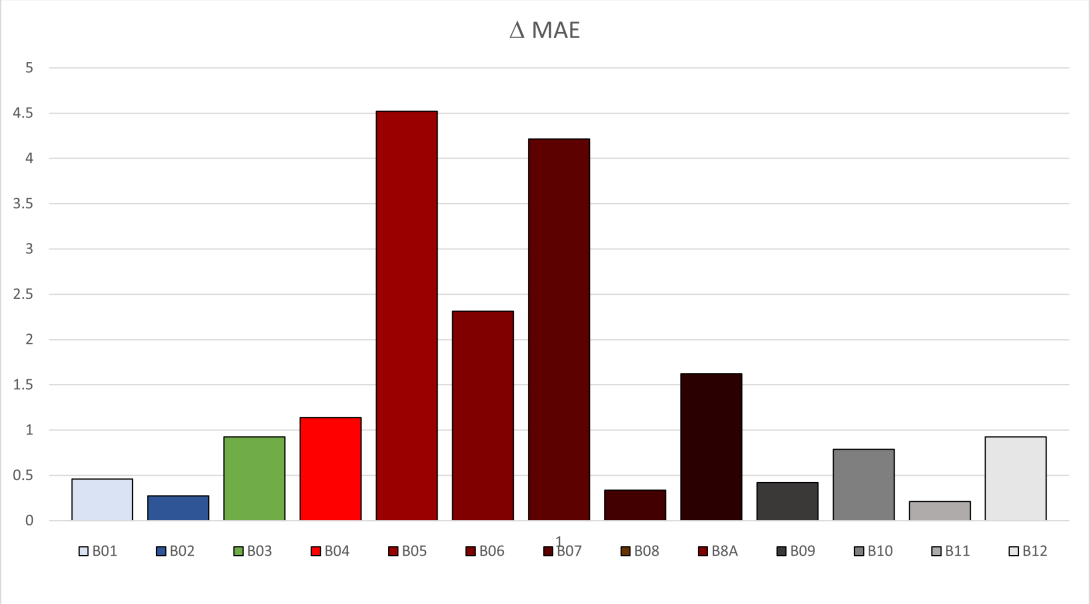


Figure 4.3: Spectral bands importance obtained as increase of MAE.

The red edge is a key wavelength range in remote sensing that is sensitive to vegetation conditions and can be used to support different vegetation parameter retrieval products (Cui and Kerekes, 2018) and to extract water body information, as it was seen that the reflectivity of the water body to the red edge band was significantly different from that of other ground objects to red edge band (Lu et al., 2021).
As these spectral bands are provided by Sentinel-2 MSI, but not incorporated in other spectral sensor, such as Landsat and MODIS (see Figure 4.5), these results drive the attention on potential use of red-edge spectral bands in water quality applications.

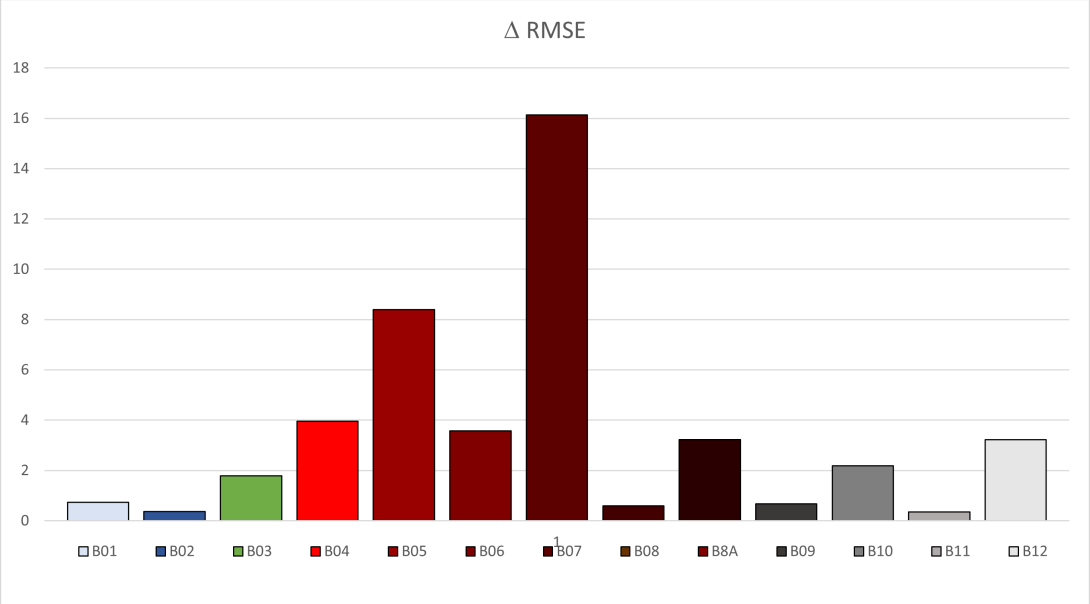


Figure 4.4: Spectral bands importance obtained as increase of RMSE.

For instance, Landsat Next (sometimes referred to as Landsat 10) is a planned constellation of three superspectral satellites in continuation of the Landsat series. The mission is a collaboration between NASA and the USGS with the objective to extend the ability to detect and quantitatively characterise changes on the global land surface at a scale where natural and human-induced causes of land surface change can be detected and differentiated. New spectral bands were added to Landsat Next mission requirements in response to emerging science and user needs, such as red edge and shortwave bands.

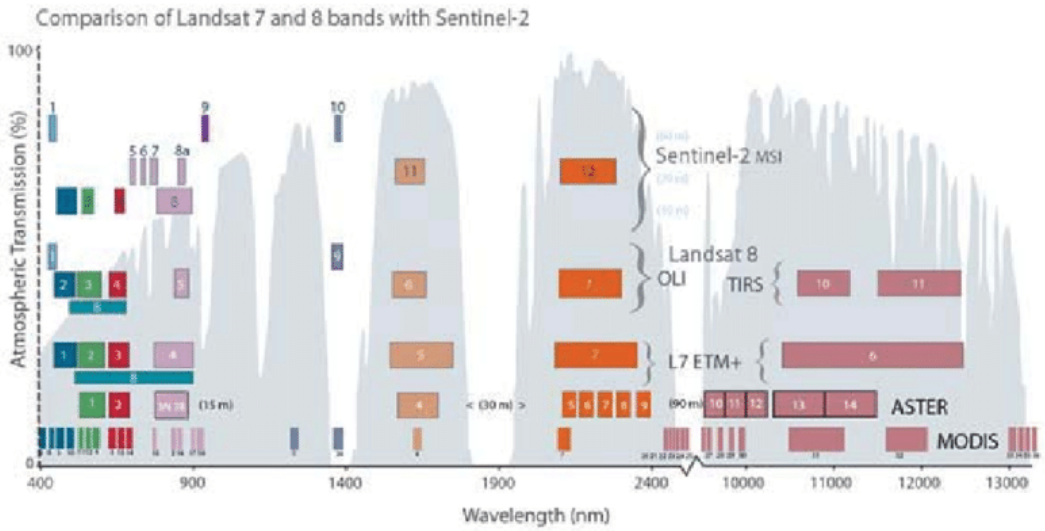


Figure 4.5: Comparison of Landsat 7 ETM+, 8 OLI, Sentinel-2 MSI, TERRA ASTER, and AQUA MODIS bands (El Atillah et al., 2019).

Bands mostly affected by atmospheric effects (Coastal-Aerosol B01, Water Vapour B09, and Cirrus B10) provide a smaller but not neglectable contribution to turbidity prediction.
Bands ranking obtained through permutation test confirmed that all bands contribute to the model prediction and that ML techniques enhance the exploitation of all spectral information. In general, the plots stress the importance of widening the range of bands used for turbidity estimates when applying ML techniques.
The KRLS approach can be considered a good balance between the ability of a ML algorithm to use all available bands as inputs, to capture complex statistical trends, and, on the other hand, prefer a simplest solution over the more sophisticated and computational time consuming Deep Learning ML techniques, that require adequate training samples due to the corresponding increment of parameters to be tuned.

4.3. Turbidity Maps

The calibrated model has been applied on a pixel-by-pixel basis to produce full turbidity maps of the area. An example is shown in Figure 4.7, where the turbidity map corresponding of the image of October 8th 2020 (Figure 4.6) is reported.

Figure 4.6: A True Color Image of the study area (tile T32TNP) obtained from Level-1C Sentinel-2 MSI products (October 8th, 2020).

The shown map is produced at 60 m resolution, that is 1830x1830 pixels. Higher spatial resolution could be provided re-sampling MSI data to a given level of detail, up to the maximum available spatial resolution of 10 m.
Such dense maps allow a direct inspection of possible high turbidity areas. In the map, a smooth transition from clear water to the turbidity peak at the mouth of river Arno is clearly visible. It can be observed from querying the map, that the Kernel method tends to over-estimate to unity low values in open sea.
Also, it is possible to see how clouds are detected and identified by the cloud detection tool (see Chapter 3.4.).
The considered area presents a wide variety of turbidity contributions and conditions, ranging from clear water to high turbidity, particularly related to sediment loads carried from rivers.
The model has been applied to produce quantitative turbidity maps from satellite imagery with different turbidity conditions, showing promising results on the flexibility of

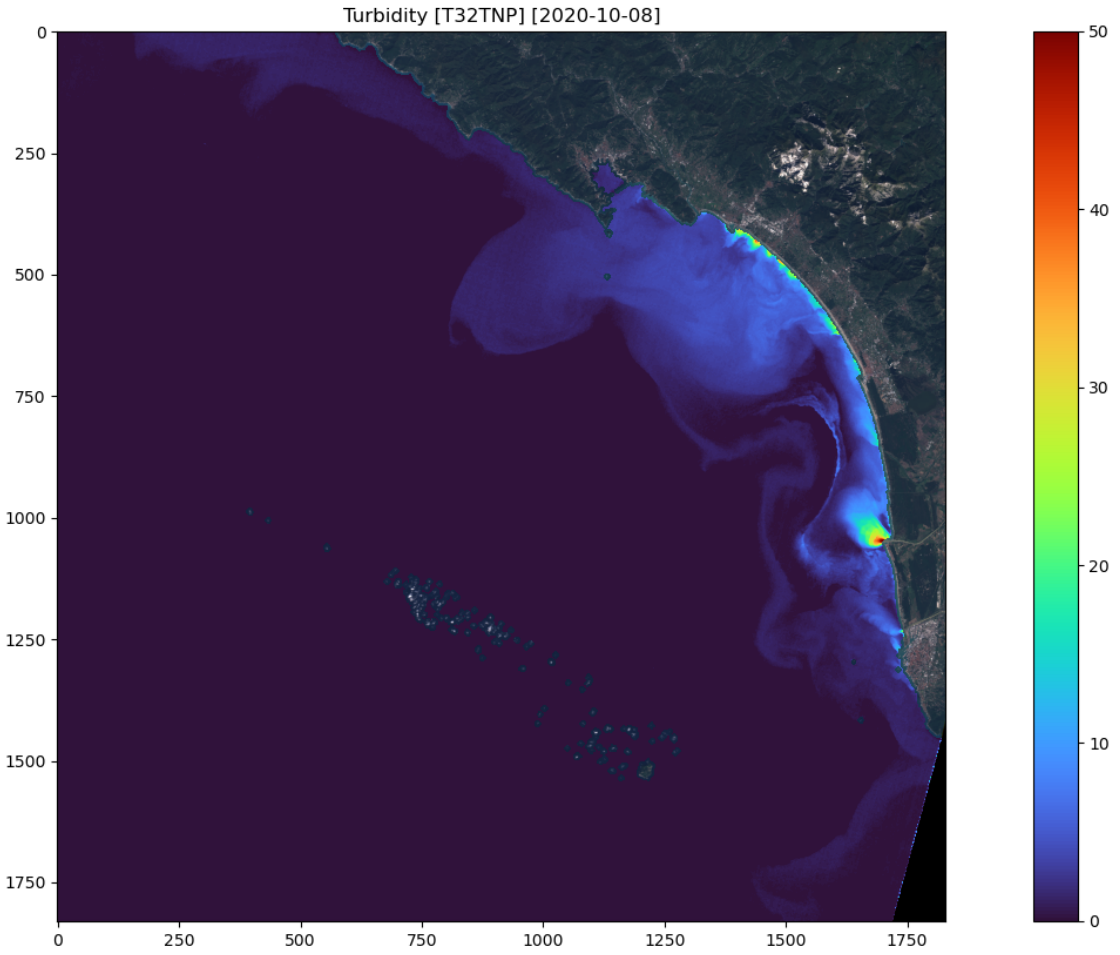


Figure 4.7: Dense turbidity map in NTU obtained by the KRLS model from Level-1C Sentinel-2 MSI products (October 8th, 2020).

the models to operate in the entire tile area, and in turbidity conditions tat differ from the specific training ones. The maximum observed value used for training is, in fact, 28.7 NTU, while in the shown map (Fig. 4.7) the plume reaches up to 62 NTU.
Although the accuracy of results outside the range of the training dataset can be evaluated and quantified only with availability of more in situ data, these results are confirmed by the estimates from CMEMS product, used as benchmark, that show similar turbidity values for the same image (see Figure 4.10 in Chapter 4.4.).
The creation of turbidity maps represents valuable tool to support the traditional on field monitoring, as it allows to extend the punctual measured information to a synoptic view.

The applicability of the calibrated algorithm in other areas of the North Tyrrhenian Sea will be tested in the future, comparing results of the models with in-situ turbidity data, to evaluate whether these algorithms can be applied in close geographical sites, with supposedly similar hydrography and environments, without the need for region-specific calibration.

At the moment, in fact, the algorithm can be applied to any of the tiles covering the Liguria Region, however accuracy of results needs to be evaluated by comparison with in-situ data. An example of application of the developed model to tile T32TLP (covering the Western part of Liguria Region and a small part of French coast) is shown in Figure 4.8, representing the effect of the meteoric event known as "Storm Alex" that hit from Tuscany to France on October 2020 and brought torrential rainfall and flooding. The image shows the plume from rives Roya and Var (France).

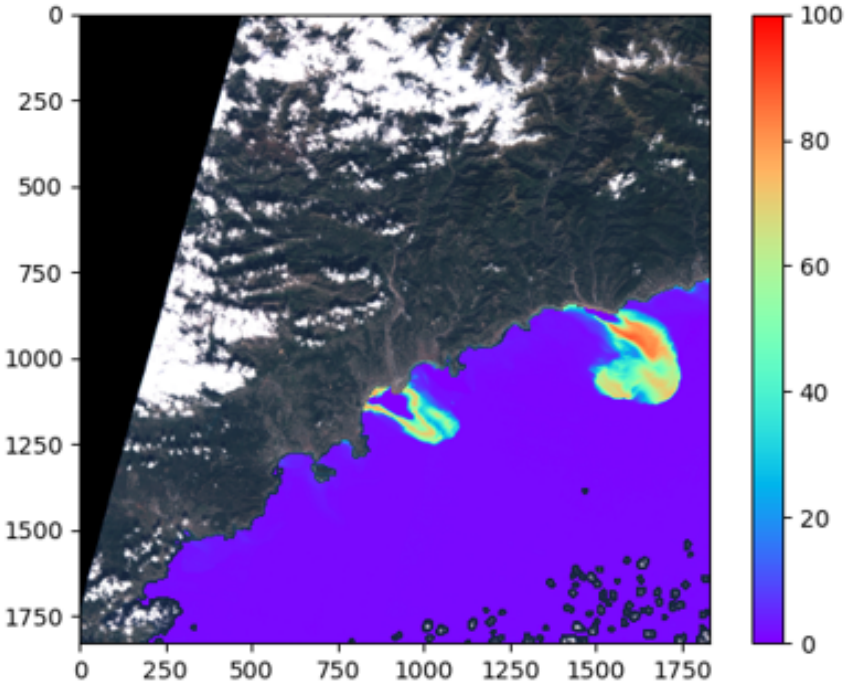


Figure 4.8: Dense turbidity map in NTU obtained in tile T32TLP by the KRLS model from Level-1C Sentinel-2 MSI products (October 3th, 2020).

4.4. Comparison with state-of-the-art

As previously introduced (Chapter 2), the state of the art in turbidity retrieval from Sentinel-2 MSI images is represented by the Copernicus Marine Environment Monitoring Service "Mediterranean Sea, Bio-Geo-Chemical, L3, daily observation" product, which was thus used as benchmark to show the performance of the developed turbidity model.

CMEMS provides a wide range of oceanographic data and products for both scientific and operational users. The "Mediterranean Sea, Bio-Geo-Chemical, L3, daily observation" dataset contains daily observations of bio-geo-chemical variables in the Mediterranean Sea derived from Sentinel-2 MSI images. Turbidity estimates are based on the semi-analytical algorithm proposed by Nechad et al., 2010 that uses remote sensing reflectance at two wavelengths: red band (665 nm) for medium-low turbidity waters and NIR (704 nm) band in highly turbid waters. TOA reflectance is obtained applying different atmospheric correction algorithms depending on turbidity levels. Turbidity maps have a spatial resolution of 100 m.
More details on the AC algorithm and the applied multi conditional automated switching single band algorithm can be found in Novoa et al., 2017 and Pahlevan et al., 2021.
The product was released into Copernicus Marine Service website in May 2021, when the resent research work was ongoing.
A comparison among in-situ measurements, turbidity estimates obtained from the developed model, and CMEMS turbidity product, extracted for the days of the surveys in the pixel corresponding to the sampling location, was conducted.
Since the Copernicus product is available since January 2020, the comparison covered the period 2020-2021, including 75 turbidity values, ranging from 0.1 NTU to 28.7 NTU.

The Pearson correlation coefficient, which measures the linear dependence between two variables, was computed, and is shown in Table 4.2.
The Pearson correlation coefficient is the covariance of the two variables (x and y) divided by the product of their standard deviations:

$$r = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2(y_i - \overline{y})^2}}$$

	CMEMS (FNU)	KERNEL (NTU)	Measure (NTU)
CMEMS (FNU)	1	0.973	0.923
KERNEL (NTU)		1	0.932
Measure (NTU)			1

Table 4.2: Linear correlation coefficient or Pearson coefficient (R) between in-situ measured turbidity in NTU, turbidity estimated from the Kernel model in NTU and turbidity in FNU from the CMEMS "Mediterranean Sea, Bio-Geo-Chemical" dataset.

A value of 1 validates a perfect linear relationship between x and y, in which the data variability of x tracks that of y. A value of 0 suggests lack of linear correlation between the variables.

It is important to notice that the minimum and maximum measured turbidity values are included in the dataset, and this is particularly important as the Pearson coefficient is highly sensitive to extreme values and may be misleading if x and y have not been analyzed in terms of their full ranges.

The consistency of results of the comparison between estimates obtained from the two proposed models with turbidity data from Copernicus CMEMS dataset named "Mediterranean Sea, Bio-Geo-Chemical, L3, daily observation" confirms robustness of obtained turbidity estimates.

The graph in Figure 4.9 plots how data are distributed and compares estimates from the two methods with observations.

Samples with high residuals are nearly the same for both methods (Figure 4.9), in par-

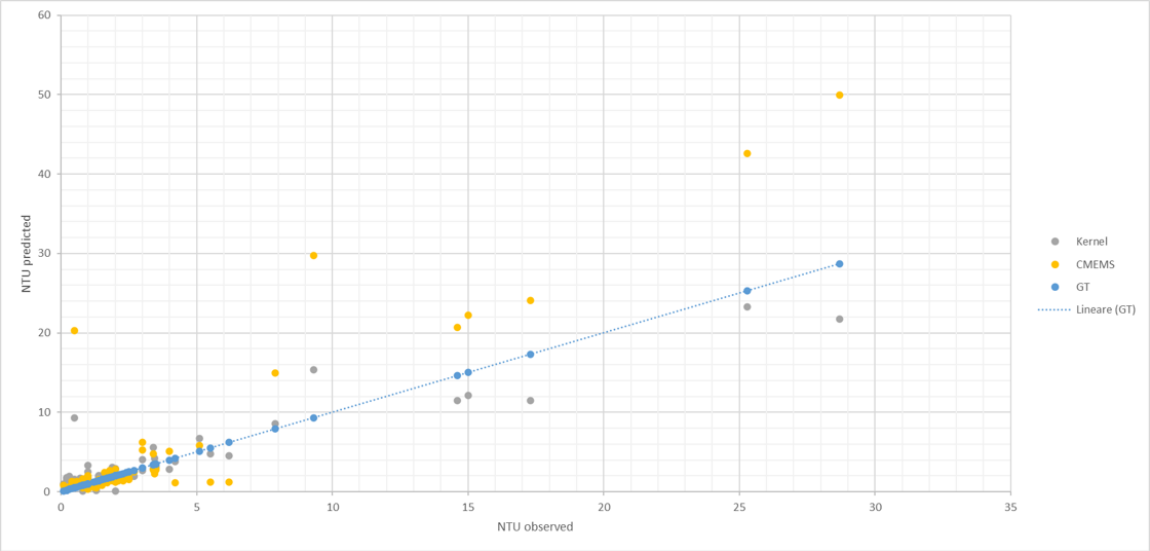


Figure 4.9: Comparison among turbidity estimates from the proposed model (Kernel), CMEMS Mediterranean Sea, Bio-Geo-Chemical product (CMEMS) and observations (ground truth - GT).

ticular for low-medium values of turbidity (8-15 NTU), thus indicating a possible problem in other stages of the data processing chain, starting from the dataset construction itself. The daily, rather than hourly, synchronism of the in situ and satellite measurements can lead to significant differences between the turbidity values observed by the two instruments, especially during phenomena associated with an increase in suspended solids with high hydrodynamics, such as river runoff.

The use of samples collected by different operators (Liguria and Tuscany Environmental Agencies) helped achieve a more representative set for training, but, on the other hand, the lack of previous inter-calibration increases the error related to the accuracy of the reference in situ data. Despite this, the model showed a good capability to generalise with respect to the used instruments, thus it is expected that new available data could

be easily integrated for further calibration.
Finally, the total accuracy of the model is affected by other factors, such as probe calibration and maintenance.
Figure 4.10 shows turbidity maps obtained from Sentinel-2 MSI data for the CMEMS product and the proposed KRLS model on October 8th, 2020.
We can notice that both methods reproduced turbidity plumes at Arno river mouth that reach up to 88 FNU and 62 NTU respectively.

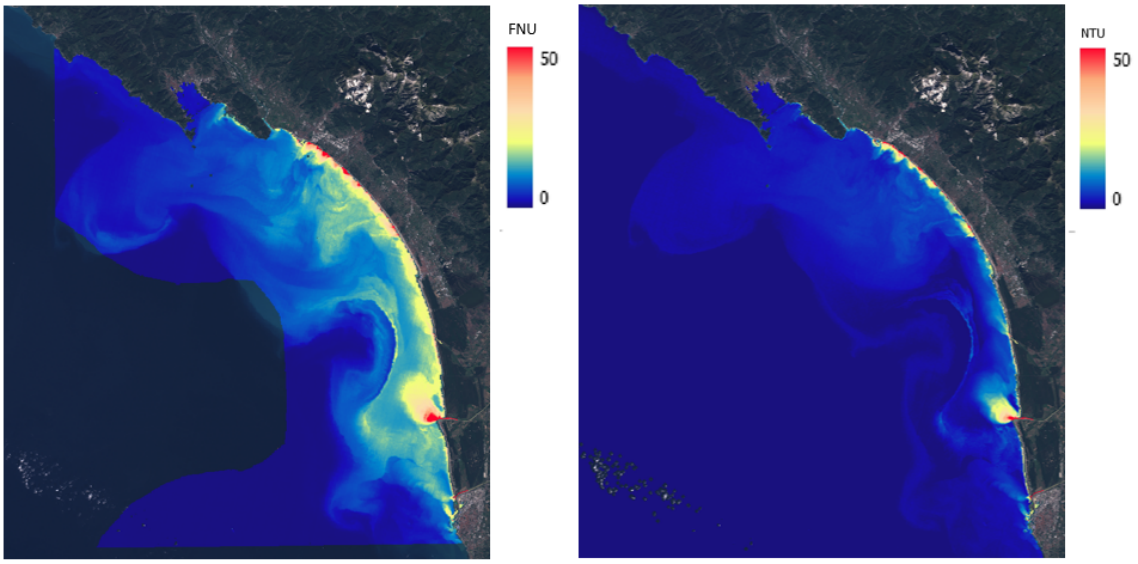


Figure 4.10: Turbidity maps obtained from CMEMS in FNU (left image) and from the proposed model in NTU (right image).

As previously discussed, this corroborates the remarkable ability of the model to generalize out of the training set, thanks to an optimized trade-off between bias and variance. Compared to CMEMS product, the developed model work as an input-output relationship; it can be considered valid for all turbidity ranges and across different bio-optical regimes and it doesn't require the intermediate step of atmospheric correction, which can even lead to large uncertainties in satellite data products if not supported by validation. The obtained results confirm that it is possible to reach a good accuracy in turbidity estimation from satellite TOA reflectance using ML techniques, and in particular, that it is possible to predict reliable nephelometric turbidity values from Sentinel-2 MSI images, using the proposed methodology.
The potential of the model will be exploited by reconstructing historical series of turbidity in the study area to derive statistical analysis, such as seasonal or annual climatology, etc. Such results can be of crucial interest for authorities in charge of water quality monitoring, such as the Regional and National Agencies for the Environmental Protection, who will benefit from EO to integrate in situ measurements and to gather knowledge of the sea environment, and, in cascade, decision makers, such as the authorities in charge of spatial planning (e.g. for identifying areas suitable for aquaculture, for which turbidity levels represent a critical indicator).

The possibility to create historical series of turbidity, complementing existing measurements if available, can support the analysis of the relation between increase and persistence in turbidity and hydro-meteorological variables, adding important knowledge of natural seasonal turbidity fluctuations in the area. Reconstructed turbidity levels from satellite images can further lead to the identification of natural turbidity levels to be used within the Environmental Impact Assessment (EIA, Directive 2014/52/UE) for projects that require it.
Our understanding of turbidity events can be valuable in the management of the sea water quality, allowing to investigate and early detect water quality deterioration.

CONCLUSIONS

To ensure a sustainable marine ecosystem, it is essential to monitor marine pollutants and reduce pollutants flow into the sea. Conventional approach for marine monitoring rely on in-situ measurements that are accurate but time consuming and geographically constrained. However, measurements across a range of spatial scales are necessary for monitoring and managing the environmental impacts and understanding the processes governing their spatial distribution.

Marine monitoring could gain in effectiveness if integrated with approaches that are able to collect data from wide geographic areas, where only few observations can be conducted by traditional methods using oceanographic vessels. Remote Sensing technology has the potential to provide spatially synoptic information of optically active water quality parameters, that can complement conventional water sampling.

Many studies have demonstrated the successful application of the use of remote sensing optical data to determine optical water quality variables, including turbidity. However, the use of available satellite data still has some limitations, such as the required preprocessing, related for instance to the atmospheric correction and cloud detection, or the availability of ground truth data to compare retrieved results with.

Atmospheric correction over coastal waters still remains one of the major challenges in aquatic remote sensing. Inaccurate AC leads to large uncertainties in satellite data products, with effects on quantitative retrieval of optically active water quality parameters. A state-of-the-art analysis was reported in the document and limitation and open problems were described.

The use of Machine Learning techniques is hypothesised to enhance the usability of optical remote sensed data to retrieve turbidity in optically complex areas, as they provide flexible relationships that can better cope with the complexity of the phenomena. While physically-based relationships require robust atmospheric correction to obtain Bottom Of Atmosphere (BOA) reflectance, ML algorithms are less affected by the atmospheric and other background factors. An added advantage of the use of ML technique is that they allow to exploit the information of the whole available spectrum of sensed bands.

The application of a Kernel Regularized Least Squares model was tested to retrieve turbidity from Sentinel-2 MSI Top Of Atmosphere (TOA) optical data in a study area in the North Thyrrenian Sea.

The developed turbidity model shows good agreement of the estimated satellite-derived surface turbidity with measured one, confirming that the use of ML techniques allows to reach a good accuracy in turbidity estimation from satellite TOA reflectance.

Moreover, promising results emerged on the flexibility of the model to operate in turbidity conditions that differ from the specific training ones, as observed from the comparison of turbidity maps obtained from the proposed model and CMEMS turbidity product, where both methods provide, in images captured after intense meteorological event, values that exceed the maximum turbidity value available for training.

Although the accuracy of results outside the range of the training dataset cannot be ver-

ified and quantified without in situ data, the coherence of the two estimates support the obtained results.
Good generalization is, in fact, one of the typical properties of KRLS algorithms, due to the use of regularization, which helps to prevent overfitting and improve the ability of the model to generalize.

The applicability of the calibrated algorithm in other areas of the North Tyrrhenian Sea will be tested in the future, comparing results of the models with in-situ turbidity data, to evaluate whether these algorithms can be applied in close geographical sites, with supposedly similar hydrography and environments, without the need for region-specific calibration.
At the moment, in fact, the algorithm can be applied to any of the tiles covering the Liguria Region, however accuracy of results needs to be evaluated by comparison with in-situ data. model to generalize includes the ability of use data from different sources and to predict in other areas
The selected model allows a good level of explainability, that is the degree to which we can understand the reasoning and decision-making processes of the algorithm. This was achieved through feature importance analysis, that allows identifying the most important variables that the model is using to make its prediction.
The feature importance analysis, conducted through permutation analysis, allowed to estimate the relative importance of each band for the predicting model, reveling the importance that bands in the red-edge can played in turbidity retrieval. The results also confirmed the potential of the use of all available bands, as each band provide a positive contribution to the model prediction.
The three main research goals were then achieved:

1. Define and calibrate a model based on Machine learning algorithms to retrieve turbidity from Sentinel-2 MSI in a study area.

2. Explore the use of the whole spectrum of MSI bands.

3. Create turbidity maps from Sentinel-2 images in the study are to support in-situ observations.

We can conclude that the developed model is a step toward the use of Satellite Remote Sensing to provide observations over large areas, representing a valuable tool to support and integrate the traditional monitoring campaigns, enabling to extend the punctual measured information to a synoptic view. The integration of traditional measurements and satellite observation is believed to represent a win-win approach to improvement of marine water quality monitoring: high-resolution data from field observations con contribute to validate and improve the accuracy of satellite based data, while satellite observations allow to extend the scale of field observations. This approach can help to fill gaps in knowledge and provide a more comprehensive picture of a particular system or process.
Finally, by increasing spatial and temporal coverage, satellite observation is expected to enhance the effectiveness of the EU Member States' action plans through early detection of deterioration, improving knowledge of the potential extent of management impacts,

improving monitoring of the effectiveness of action plans and providing information to support more strategic, in situ sampling.

List of Figures

List of Tables

REFERENCES

Abirhire, O., Davies, J. M., G., X., Hudson, J., 2020. Understanding the factors associated with long term reconstructed turbidity in Lake Diefenbaker from Landsat imagery. *Science of The Total Environment*, 724, 138222. https://doi.org/10.3390/rs10050786.

Aggarwal, S., 2004. Introduction to Remote Sensing Systems, Data, and Applications. In Remote Sensing of Natural Resources. *Satellite Remote Sensing and GIS Applications in Agricultural Meteorology*, World Meteorological Organisation, 23–38.

Akbar, T. A., Hassan, Q. K., Achari, G., 2014. Development of Remote Sensing Based Models for Surface Water Quality. *CLEAN – Soil, Air, Water*, 42(8), 1044–1051. https://doi.org/10.1002/clen.201300001.

Anderson, M., Robinson, J., 2001. Permutation Tests for Linear Models. *Australian and New Zealand J. Statistics*, 43, 75-88. https://doi.org/10.1111/1467-842X.00156.

Awad, M., Khanna, R., 2015. *Machine Learning*. Apress, Berkeley, CA, 1–18.

Baughman, C. A., Jones, B. M., Bartz, K. K., Young, D. B., Zimmerman, C. E., 2015. Reconstructing Turbidity in a Glacially Influenced Lake Using the Landsat TM and ETM+ Surface Reflectance Climate Data Record Archive, Lake Clark, Alaska. *Remote Sensing*, 7(10), 13692–13710. https://www.mdpi.com/2072-4292/7/10/13692.

Binding, C., Pizzolato, L., Zeng, C., 2021. EOLakeWatch; delivering a comprehensive suite of remote sensing algal bloom indices for enhanced monitoring of Canadian eutrophic lakes. *Ecological Indicators*, 121, 106999. https://doi.org/10.1016/j.ecolind.2020.106999.

Breiman, L., 2001. Random Forests. *Machine Learning*, 45, 5-32. https://doi.org/10.1023/A:1010933404324.

Bui, Q.-T.and Jamet, C., Vantrepotte, V., Mériaux, X., Cauvin, A., Mograne, M., 2022. Evaluation of Sentinel-2/MSI Atmospheric Correction Algorithms over Two Contrasted French CoastalWaters. *Remote Sens*, 14(1099). https://doi.org/10.3390/rs14051099.

Bustamante, J., Pacios, F., Díaz-Delgado, R., Aragonés, D., 2009. Predictive models of turbidity and water depth in the Doñana marshes using Landsat TM and ETM+ images. *Journal of Environmental Management*, 90(7), 2219-2225. https://doi.org/10.1016/j.jenvman.2007.08.021.

Caballero, I., Steinmetz, F., Navarro, G., 2018. Evaluation of the First Year of Operational Sentinel-2A Data for Retrieval of Suspended Solids in Medium- to High-Turbidity Waters. *Remote Sensing*, 10(7). https://doi.org/10.3390/rs10070982.

Chavez, P. S., 1988. An improved dark-object subtraction technique for atmospheric scattering correction of multispectral data. *Remote Sensing of Environment*, 24(3), 459–479. https://doi.org/10.1016/0034-4257(88)90019-3.

Chavez, P. S., 1996. Image-Based Atmospheric Corrections - Revisited and Improved. *Photogrammetric Engineering and Remote Sensing*, 62, 1025–1036.

Chebud, Y.and Naja, G., Rivero, R., Melesse, A. M., 2012. Water Quality Monitoring Using Remote Sensing and an Artificial Neural Network. *Water Air Soil Pollut*, 223, 4875–4887. https://doi.org/10.1007/s11270-012-1243-0.

Chen, Z., Hu, C., Muller-Karger, F., 2007. Monitoring turbidity in Tampa Bay using MODIS/Aqua 250-m imagery. *Remote Sensing of Environment*, 109(2), 207-220. https://doi.org/10.1016/j.rse.2006.12.019.

Chu, H.-J., Kong, S.-J., Chang, C.-H., 2018. Spatio-temporal water quality mapping from satellite images using geographically and temporally weighted regression. *International Journal of Applied Earth Observation and Geoinformation*, 65, 1-11. https://doi.org/10.1016/j.jag.2017.10.001.

Congedo, L., 2021. Semi-Automatic Classification Plugin: A Python tool for the download and processing of remote sensing images in QGIS. *Journal of Open Source Software*, 6(64). https://doi.org/10.21105/joss.03172.

Cui, Z., Kerekes, J., 2018. Potential of Red Edge Spectral Bands in Future Landsat Satellites on Agroecosystem Canopy Green Leaf Area Index Retrieval. *Remote Sensing*, 10, 1458. https://doi.org/10.3390/rs10091458.

Dogliotti, A., Ruddick, K., Nechad, B., Doxaran, D., Knaeps, E., 2015. A single algorithm to retrieve turbidity from remotely-sensed data in all coastal and estuarine waters. *Remote Sensing of Environment*, 156, 157-168. https://doi.org/10.1016/j.rse.2014.09.020.

Doxaran, D., Froidefond, J., Castaing, P., 2002a. A reflectance band ratio used to estimate suspended matter concentrations in sediment-dominated coastal waters. *International Journal of Remote Sensing*, 23(23), 5079-5085. https://doi.org/10.1080/0143116021000009912.

Doxaran, D., Froidefond, J.-M., Lavender, S., Castaing, P., 2002b. Spectral signature of highly turbid waters: Application with SPOT data to quantify suspended particulate matter concentrations. *Remote Sensing of Environment*, 81(1), 149-161. https://doi.org/10.1016/S0034-4257(01)00341-8.

Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., Hoersch, B., Isola, C., Laberinti, P., Martimort, P., Meygret, A., Spoto, F., Sy, O., Marchese, F., Bargellini, P., 2012. Sentinel-2: ESA's Optical High-Resolution Mission for GMES Operational Services. *Remote Sensing of Environment*, 120, 25-36. https://doi.org/10.1016/j.rse.2011.11.026.

Efron, B., 1982. The jackknife, the bootstrap and other resampling plans. *In CBMS-NSF Regional Conference series in applied mathematics 1982*, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA.

El Atillah, A., El Morjani, Z. E. A., Mustapha, S., 2019. Use of the Sentinel-2A Multispectral Image for Litho-Structural and Alteration Mapping in Al Glo'a Map Sheet (1/50,000) (Bou Azzer–El Graara Inlier, Central Anti-Atlas, Morocco). *Artificial Satellites*, 54, 73-96. https://doi.org/10.2478/arsa-2019-0007.

El Din, E. S., 2020. A novel approach for surface water quality modelling based on Landsat-8 tasselled cap transformation. *International Journal of Remote Sensing*, 41(18), 7186-7201. https://doi.org/10.1080/14498596.2019.1674197.

Erftemeijer, P. L., Riegl, B., Hoeksema, B. W., Todd, P. A., 2012. Environmental impacts of dredging and other sediment disturbances on corals: A review. *Marine Pollution Bulletin*, 64(9), 1737-1765. https://doi.org/10.1016/j.marpolbul.2012.05.008.

European Space Agency, 2020. SeNtinel Application Platform (SNAP) ver 8.0.0. http://step.esa.int, last accessed on 25 May 2022.

Feng, L., Hu, C., Chen, X., Song, Q., 2014. Influence of the Three Gorges Dam on total suspended matters in the Yangtze Estuary and its adjacent coastal waters: Observations from MODIS. *Remote Sensing of Environment*, 140, 779-788. https://doi.org/10.1016/j.rse.2013.10.002.

Garbolino, E., Aqua, J., Abriak, N., 2014. Applicability of H14 Protocol for Sediments in Order to Consider Their Valorization: Limits and Benefits. *Chem. Eng. Trans.*, 36, 631-636. https://doi.org/10.3303/CET1436106.

Gordon, H., 1997. Atmospheric Correction of Ocean Color Imagery in the Earth Observing System Era. *Journal of Geophysical Research: Atmospheres*, 102, 17081- 17106. https://doi.org/10.1029/96JD02443.

GRASS Development Team, 2022a. Geographic Resources Analysis Support System (GRASS) Software. Open Source Geospatial Foundation. grass.osgeo.org last accessed on 25 May 2022.

GRASS Development Team, 2022b. i.atcorr - GRASS GIS MANUAL. Open Source Geospatial Foundation. https://grass.osgeo.org/grass80/manuals/i.atcorr.html last accessed of 25 May 2022.

Hadjimitsis, D. G., Clayton, C. R. I., Hope, V. S., 2004. An assessment of the effectiveness of atmospheric correction algorithms through the remote sensing of some reservoirs. *International Journal of Remote Sensing*, 25(18), 3651–3674. https://doi.org/10.1080/01431160310001647993.

Hafeez, S., Wong, M. S., Abbas, S., Kwok, C. Y., Nichol, J., Lee, K. H., Tang, D., Pun, L., 2018. Detection and monitoring of marine pollution using remote sensing technologies. *Monitoring of Marine Pollution*. https://doi.org/10.5772/intechopen.81657.

Han, B., Loisel, H., Vantrepotte, V., Mériaux, X., Bryère, P., Ouillon, S., Dessailly, D., Xing, Q., Zhu, J., 2016. Development of a Semi-Analytical Algorithm for the Retrieval of

Suspended Particulate Matter from Remote Sensing over Clear to Very Turbid Waters. *Remote Sensing*, 8(3). https://doi.org/10.3390/rs8030211.

Hanna, S. R., Heinold, D. W., 1985. *Development and application of a simple method for evaluating air quality models*. American Petroleum Institute.

Hofmann, T., Schölkopf, B., Smola, A. J., 2008. Kernel methods in machine learning. *The Annals of Statistics*, 36(3), 1171 – 1220. 10.1214/009053607000000677.

Holben, B. N., Eck, T. F., Slutsker, I., Tanré, D., Buis, J. P., Setzer, A., Vermote, E., Reagan, J. A., Kaufman, Y. J., Nakajima, T., Lavenu, F., Jankowiak, I., , Smirnov, A., 1998. AERONET – A Federated Instrument Network and Data Archive for Aerosol Characterization. *Remote Sens. Environ.*, 66, 1–16. https://doi.org/10.1016/S0034-4257(98)00031-5.

Hudson, A., Talke, S., Jay, D., 2017. Using Satellite Observations to Characterize the Response of Estuarine Turbidity Maxima to External Forcing. *Estuaries and Coasts*, 40, 343–358. https://doi.org/10.1007/s12237-016-0164-3.

Islam, M. S., Tanaka, M., 2004. Impacts of pollution on coastal and marine ecosystems including coastal and marine fisheries and approach for management: a review and synthesis. *Marine Pollution Bulletin*, 48(7), 624-649. https://doi.org/10.1016/j.marpolbul.2003.12.004.

Joshi, I. D., D'Sa, E. J., Osburn, C. L., Bianchi, T. S., 2017. Turbidity in Apalachicola Bay, Florida from Landsat 5 TM and Field Data: Seasonal Patterns and Response to Extreme Events. *Remote Sensing*, 9(4). https://doi.org/10.3390/rs9040367.

Katlane, R., Dupouy, C.and El Kilani, B., J.C., B., 2020. Estimation of Chlorophyll and Turbidity Using Sentinel 2A and EO1 Data in Kneiss Archipelago Gulf of Gabes, Tunisia. *Int. J. Geosci.*, 11, 708-728. https://doi.org/10.4236/ijg.2020.1110035.

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W.and Ye, Q., Liu, T., 2017. LightGBM: A Highly Efficient Gradient Boosting Decision Tree.

Kirk, J. T. O., 1985. Effects of suspensoids (turbidity) on penetration of solar radiation in aquatic ecosystems. *Hydrobiologia*, 125(1), 195–208. https://doi.org/10.1007/BF00045935.

Kotchenova, S. Y., Vermote, E. F., Matarrese, R., Klemm, F., 2006. Validation of a vector version of the 6S radiative transfer code for atmospheric correction of satellite data. Part I: Path radiance. *Applied Optics*, 45(26), 6762–6774. https://doi.org/10.1364/AO.45.006762.

Lantzanakis, G., Mitraka, Z., Chrysoulakis, N., 2017. Comparison of physically and image based atmospheric correction methods for Sentinel-2 satellite imagery. *Perspectives on Atmospheric Sciences*, 255–261.

Lawler, D., 2005. Spectrophotometry | turbidimetry and nephelometry. P. Worsfold, A. Townshend, C. Poole (eds), *Encyclopedia of Analytical Science (Second Edition)*, second edition edn, Elsevier, Oxford, 343–351.

Li, D., Liu, S., 2018. Sensors in water quality monitoring. *Water Quality Monitoring and Management: Basis, Technology and Case Studies*, Academic Press, 1–54.

Lillesand, T. M., Kiefer, R. W., 1993. *Remote Sensing and Image Interpretation*. John Wiley & Sons, New York.

Lisi, I., Feola, A., Bruschi, A., Pedroncini, A., Pasquali, D., Risio, M. D., 2019. Mathematical modeling framework of physical effects induced by sediments handling operations in marine and coastal areas. *Journal of Marine Science and Engineering*, 7(149). https://doi.org/10.3390/jmse7050149.

Liu, L.-W., Wang, Y.-M., 2019. Modelling Reservoir Turbidity Using Landsat 8 Satellite Imagery by Gene Expression Programming. *Water*, 11(7). https://doi.org/10.3390/w11071479.

Louis, J., Debaecker, V., Pflug, B., Main-Knorn, M., Bieniarz, J., Mueller-Wilm, U., Cadau, E., Gascon, F., 2016. Sentinel-2 Sen2Cor: L2A Processor for users, in: Proceedings living planet symposium 2016. *Spacebooks Online*, 1–8.

Lu, Z., Wang, D., Deng, Z., Shi, Y., Ding, Z., Ning, H., Zhao, H., Zhao, J., Xu, H., Zhao, X., 2021. Application of red edge band in remote sensing extraction of surface water body: a case study based on GF-6 WFV data in arid area. *Hydrology Research*, 52(6), 1526–1541. https://doi.org/10.2166/nh.2021.050.

Magrì, S., De Gaetano, P., Feola, A., Lisi, I., Salmeri, A., Venti, F., Pedroncini, A., 2020. Numerical Modelling for Environmental Impact Assessment of Sediment Dispersion in Port Areas. *Comput. Aided Chem. Eng.*, 48, 337-342. https://doi.org/10.1016/B978-0-12-823377-1.50057-4.

Magrì, S., Ottaviani, E., Prampolini, E., Federici, B., Besio, G., Fabiano, B., 2023. Application of machine learning techniques to derive sea water turbidity from Sentinel-2 imagery. *Remote Sensing Applications: Society and Environment*, 30, 100951. https://doi.org/10.1016/j.rsase.2023.100951.

Mayer, B., Kylling, A., 2005. Technical note: The libRadtran software package for radiative transfer calculations - description and examples of use. *Atmos. Chem. Phys.*, 5, 1855–1877. https://doi.org/10.5194/acp-5-1855-2005.

Medina-Lopez, E., 2020. Machine Learning and the End of Atmospheric Corrections: A Comparison between High-Resolution Sea Surface Salinity in Coastal Areas from Top and Bottom of Atmosphere Sentinel-2 Imagery. *Remote Sensing*, 12(18). https://doi.org/10.3390/rs12182924.

Miller, R. L., McKee, B. A., 2004. Using MODIS Terra 250 m imagery to map concentrations of total suspended matter in coastal waters. *Remote Sensing of Environment*, 93(1), 259-266. https://doi.org/10.1016/j.rse.2004.07.012.

Mitchell, T. M., 1997. *Machine Learning*. McGraw-Hill Science/Engineering/Math, chapter Concept Learning and the General-to-Specific Ordering.

Moore, G. K., 1980. Satellite remote sensing of water turbidity / Sonde de télémesure par satellite de la turbidité de l'eau. *Hydrological Sciences Bulletin*, 25(4), 407–421. https://doi.org/10.1080/02626668009491950.

Nas, B., Ekercin, S., Karabörk, H., Berktay, A., Mulla, D. J., 2010. An Application of Landsat-5TM Image Data for Water Quality Mapping in Lake Beysehir, Turkey. *Water Air Soil Pollut*, 212, 183–197. https://doi.org/10.1007/s11270-010-0331-2.

Nechad, B., Ruddick, K., Park, Y., 2010. Calibration and validation of a generic multisensor algorithm for mapping of total suspended matter in turbid waters. *Remote Sensing of Environment*, 114(4), 854-866. https://doi.org/10.1016/j.rse.2009.11.022.

Novoa, S., Doxaran, D., Ody, A., Vanhellemont, Q., Lafon, V., Lubac, B., Gernez, P., 2017. Atmospheric Corrections and Multi-Conditional Algorithm for Multi-Sensor Remote Sensing of Suspended Particulate Matter in Low-to-High Turbidity Levels Coastal Waters. *Remote Sensing*, 9(1). https://doi.org/10.3390/rs9010061.

OSPAR Commission, ., 2008. Marine biodiversity monitoring and assessment: Activities to improve synergies between eu directives and international conventions.

Pahlevan, N., Mangin, A., Balasubramanian, S. V., Smith, B., Alikas, K., Arai, K., Barbosa, C., Bélanger, S., Binding, C., Bresciani, M., Giardino, C., Gurlin, D., Fan, Y., Harmel, T., Hunter, P., Ishikaza, J., Kratzer, S., Lehmann, M. K., Ligi, M., Ma, R., Martin-Lauzer, F.-R., Olmanson, L., Oppelt, N., Pan, Y., Peters, S., Reynaud, N., Sander de Carvalho, L. A., Simis, S., Spyrakos, E., Steinmetz, F., Stelzer, K., Sterckx, S., Tormos, T., Tyler, A., Vanhellemont, Q., Warren, M., 2021. ACIX-Aqua: A global assessment of atmospheric correction methods for Landsat-8 and Sentinel-2 over lakes, rivers, and coastal waters. *Remote Sensing of Environment*, 258, 112366. https://doi.org/10.1016/j.rse.2021.112366.

Pahlevan, N., Smith, B., Schalles, J., Binding, C., Cao, Z., Ma, R., Alikas, K., Kangro, K., Gurlin, D., Hà, N., Matsushita, B., Moses, W., Greb, S., Lehmann, M. K., Ondrusek, M., Oppelt, N., Stumpf, R., 2020. Seamless retrievals of chlorophyll-a from Sentinel-2 (MSI) and Sentinel-3 (OLCI) in inland and coastal waters: A machine-learning approach. *Remote Sensing of Environment*, 240, 111604. https://doi.org/10.1016/j.rse.2019.111604.

Pereira, L. S. F., Andes, L. C., Cox, A. L., Ghulam, A., 2018. Measuring Suspended-Sediment Concentration and Turbidity in the Middle Mississippi and Lower Missouri Rivers Using Landsat Data. *JAWRA*, 54, 440– 450. https://doi.org/10.1111/1752-1688.12616.

Peterson, K. T., Sagan, V., Sidike, P., Cox, A. L., Martinez, M., 2018. Suspended Sediment Concentration Estimation from Landsat Imagery along the Lower Missouri and Middle Mississippi Rivers Using an Extreme Learning Machine. *Remote Sensing*, 10(10). https://doi.org/10.3390/rs10101503.

Petus, C., Chust, G., Gohin, F., Doxaran, D., Froidefond, J.-M., Sagarminaga, Y., 2010. Estimating turbidity and total suspended matter in the Adour River plume (South Bay of Biscay) using MODIS 250-m imagery. *Continental Shelf Research*, 30(5), 379-392. https://doi.org/10.1016/j.csr.2009.12.007.

Pisanti, A., Magrì, S., Ferrando, I., Federici, B., 2022. Seawater turbidity analysis from Sentinel-2 images: atmospheric correction and bands correlation. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, 258, 371–378. https://doi.org/10.5194/isprs-archives-XLVIII-4-W1-2022-371-2022.

QGIS Development Team, 2022. QGIS Geographic Information System. https://www.qgis.org, last accessed on 25 May 2022.

Quang, N. H., Sasaki, J., Higa, H., Huan, N. H., 2017. Spatiotemporal Variation of Turbidity Based on Landsat 8 OLI in Cam Ranh Bay and Thuy Trieu Lagoon, Vietnam. *Water*, 9(8). https://doi.org/10.3390/w9080570.

Richter, R., Schläpfer, D., 2011. Atmospheric/topographic correction for satellite imagery: Atcor-2/3 userguide.

Ruescas, A., Hieronymi, M.and Mateo-Garcia G.and Koponen, S., Kallio, K., Camps-Valls, G., 2018. Machine Learning Regression Approaches for Colored Dissolved Organic Matter (CDOM) Retrieval with S2-MSI and S3-OLCI Simulated Data. *Remote Sensing*, 10, 786. https://doi.org/10.3390/rs10050786.

Sagan, V., Peterson, K. T., Maimaitijiang, M., Sidike, P., Sloan, J., Greeling, B. A., Maalouf, S., Adams, C., 2020. Monitoring inland water quality using remote sensing: potential and limitations of spectral indices, bio-optical simulations, machine learning, and cloud computing. *Earth-Science Reviews*, 205, 103187. https://doi.org/10.1016/j.earscirev.2020.103187.

Samuel, A. L., 1959. Some Studies in Machine Learning Using the Game of Checkers. *IBM Journal of Research and Development*, 3(3), 210–229. http://people.csail.mit.edu/brooks/idocs/Samuel.pdf.

Siegel, D., Wang, M.and Maritorena, S., Robinson, W., 2000. Atmospheric Correction of Satellite Ocean Color Imagery: The Black Pixel Assumption. *Appl. Opt.*, 39, 3582–3591. https://opg.optica.org/ao/abstract.cfm?URI=ao-39-21-3582.

Skakun, S., Wevers, J., Brockmann, C., Doxani, G., Aleksandrov, M., Batič, M., Frantz, D., Gascon, F., Gómez-Chova, L., Hagolle, O., López-Puigdollers, D., Louis, J., Lubej, M., Mateo-García, G., Osman, J., Peressutti, D., Pflug, B., Puc, J., Richter, R., Roger, J., Scaramuzza, P., Vermote, E., Vesel, N., Zupanc, A., Žust, L., 2022.

Cloud Mask Intercomparison eXercise (CMIX): An evaluation of cloud masking algorithms for Landsat 8 and Sentinel-2. *Remote Sens. Environ.*, 274(10), 891–921. https://doi.org/10.1016/j.rse.2022.112990.

Turing, A. M., 1950. Computing machinery and intelligence. *Mind*, 433–460.

Weng, Q., 2013. *Introduction to Remote Sensing Systems, Data, and Applications. In Remote Sensing of Natural Resources*. Taylor & Francis Group.

Wilber, D. H., Clarke, D. G., 2001. Biological effects of suspended sediments: A review of suspended sediment impacts on fish and shellfish with relation to dredging activities in estuaries. *North American Journal of Fisheries Management*, 21(4), 855 – 875. https://doi.org/10.1577/1548-8675(2001)021<0855:BEOSSA>2.0.CO;2.

Zampoukas, N., Hanke, G., Piha, H., Bigagli, E., Cardoso, A., Hoepffner, N., 2012. doi/10.2788/77640. *Monitoring for the Marine Strategy Framework Directive : requirements and options*. Joint Research Centre and Institute for Environment and Sustainability Publications Office.

Zampoukas, N., Piha, H., Bigagli, E., Hoepffner, N., Hanke, G., Cardoso, A. C., 2013. Marine monitoring in the European Union: How to fulfill the requirements for the marine strategy framework directive in an efficient and integrated way. *Marine Policy*, 39, 349-351. https://doi.org/10.1016/j.marpol.2012.12.004.

Zhu, Z., Woodcock, C. E., 2012. Object-based cloud and cloud shadow detection in Landsat imagery. *Remote Sensing of Environment*, 118, 83-94. https://doi.org/10.1016/j.rse.2011.10.028.

Zhu, Z., Woodcock, C. E., 2014. Automated cloud, cloud shadow, and snow detection based on multitemporal Landsat data: an algorithm designed specifically for land cover change. *Remote Sensing of Environment*, 152, 217-234. https://doi.org/10.1016/j.rse.2014.06.012.

Zupanc, A., 2017. Improving cloud detection with machine learning.