POLITECNICO
MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE E DELL'INFORMAZIONE E
SCUOLA DI INGEGNERIA CIVILE, AMBIENTALE E TERRITORIALE

Regression of Chlorophyll-a values in Insubric Lakes with PRISMA hyperspectral imagery

Tesi di Laurea Magistrale in Geoinformatics Engineering - Ingegneria Geoinformatica

Author: **Juan Francisco Amieva**

Student ID: 974722
Advisor: Prof. Maria Antonia Brovelli
Academic Year: 2022-23

Contents

1 | Acronyms

Here are mentioned the acronyms used in the current document.

Acronym	*Explanation*
AaSI	Aalto-1 Spectral Imager
AHSI	Advanced Hyperspectral Imager
AIS	Airborne Imaging Spectrometer
AIS-2	Airborne Imaging Spectrometer-2
AIUS	Atmospheric Infrared Ultraspectral
ARTEMIS	Advanced Responsive Tactically Effective Military Imaging Spectrometer
ASI	Italian Space Agency (Agenzia Spaziale Italiana)
AVIRIS	Airborne Visible-Infrared Imaging Spectrometer
C2RCC	Case 2 Regional Coast Colour
CASI	Compact Airborne Spectrographic Imager
CHIME	Copernicus Hyperspectral Imaging Mission for the Environment
Chl-a	Chlorophyll-a
CHRIS	Compact High-Resolution Imaging Spectrometer
CNN	Convolutional Neural Network
CRF	Conditional Random Fields
DAIS	Digital Airborne Imaging Spectrometer
DESIS	DLR Earth Sensing Imaging Spectrometer
DLR	German Aerospace Center
DN	Digital Number
DPC	Directional Polarization Camera
EMI	Environment Monitoring Instrument
ENVISAT	Environmental Satellite
ESA	European Space Agency
FLEX	Fluorescence Explorer

FLI	Fluorescence Line Imager
FLORIS	Fluorescence Imaging Spectrometer
FOV	Field of View
FTHSI	Fourier Transform Hyperspectral Imager
GF-5	GaoFen-5
GHGSat-D	Greenhouse Gas Satellite - Demonstrator
GISAT-1	Geo Imaging Satellite
GMI	Greenhouse-gases Monitoring Instrument
GomX-4B	GomSpace Express-4B
GRU	Gated Recurrent Unit
GSD	Ground Sampling Distance
HICO	Hyperspectral Imager for the Coastal Ocean
HIRIS	High-Resolution Imaging Spectrometer
HJ-1	Huan Jing - 1
HSI	Hyperspectral Imager
HYDICE	Hyperspectral Data and Information Collection Experiment
HySI	Hyperspectral Imager (Indian)
HyspIRI	Hyperspectral Infrared Imager
IMS-1	Indian Microsatellite -1
ISA	Israeli Space Agency
ISRO	Indian Space Research Organization
ISS	International Space Station
JPL - NASA	NASA's Jet Propulsion Laboratory
LEO	Low Earth orbit
LSTM	Long Short-Term Memory
MAE	Mean Absolute Error
MERIS	Medium Resolution Imaging Spectrometer
MHRIS	Miniature Hyperspectral SWIR Imaging Spectrometer
MIVIS	Multispectral Infrared and Visible Imaging Spectrometer
MRF	Markov Random Fields
MSI	Multispectral Imager
MSX	Midcourse Space Experiments
MWIR	Mid-wave infrared
NASA	National Aeronautics and Space Administration
NDVI	Normalized difference vegetation index

NDWI	Normalized difference water index
NSO	Netherlands Space Office
OCI	Ocean Color Instrument
OLCI	Ocean and Land Color Imager
PACE	Plankton, Aerosol, Cloud, Ocean Ecosystem
PCA	Principal Component Analysis
PRISMA	PRecursore IperSpettrale della Missione Applicativa
PROBA-1	Project for On-Board Autonomy-1
RBF	Radial Basis Function
RF	Random Forest
RMSE	Root Mean Square Error
RNN	Recurrent Neural Network
S3	Sentinel 3
SFSI	SWIR Full Spectrographic Imager
SHALOM	Spaceborne Hyperspectral Applicative Land and Ocean Mission
SIMILE	Informative System for the Integrated Monitoring of Insubric Lakes and their Ecosystems
SISEX	shuttle imaging spectrometer experiment
SMIRR	Shuttle Multispectral Infrared Radiometer
SNR	Signal to Noise Ratio
SSI	Spectral Sampling Interval
SVM	Support Vector Machine
SVR	Support Vector Regressor
SWIR	Shortwave Infrared
TIR	Thermal Infrared
UV	Ultraviolet
VIMS	Visual and Infrared Multispectral Sensor
VNIR	Visible and Near Infrared
VSWIR	Visible to Short Wave Infrared
WFD	Water Framework Directive

2 | Introduction and main objectives

The main objective of this work was to determine the concentration of chlorophyll-a in a selection of Insubric Lakes, namely Lake Maggiore, Lake Como, and Lake Lugano, located in the cross-basin between Italy and Switzerland. These lakes were previously analyzed in the "Informative System for the Integrated Monitoring of Insubric Lakes and their Ecosystems" (SIMILE) project [14], which provided the baseline for the production of ground truth information used in our study. Further details about SIMILE will be provided in Section 4.3.

Considering the importance of freshwater for communities and the role of lakes in ecosystems, it becomes evident that monitoring and managing water quality is a critical requirement for society. Specifically, this study aimed to develop a model to predict chlorophyll-a concentrations using hyperspectral imagery, with a set of previously computed chlorophyll-a (Chl-a) concentration maps from the SIMILE project serving as the ground truth data.

Chlorophyll-a is a relevant parameter in the monitoring of the lakes because it can be used as an estimator of their biomass concentration. This parameter is particularly important because it allows understanding the trophic state of the lakes, which is associated with their biodiversity: the presence of nutrients in a lake can produce an overgrowth of algae and other aquatic plants that when they die are decomposed causing consumption of the oxygen in the water and affecting the biodiversity, process known as eutrophication [72].

The decision of using hyperspectral imagery was mainly based on the advantage provided by this kind of information which offers a higher spectral resolution with respect to multispectral imagery. Given that the hyperspectral images provide a high spectral resolution, they form a datacube in which the number of channels adds a third dimension and for each pixel, it is provided the spectral response of each of the measurement bands [60]. This

high spectral resolution can be exploited to understand the physicochemical composition of the Earth [60]. In our specific case, we used this information to train machine and deep learning models in order to predict chlorophyll-a concentrations on the mentioned insubric lakes.

To understand the complete development of the hyperspectral imagers throughout history, our project described in a detailed manner, the different missions that were established making use of aircraft and spacecraft vehicles, and provided information about the main technical characteristics of the analyzed imagers. Also, it specified the future evolution of this field, introducing the main aspects of ongoing missions.

Additionally, our study aimed to present the importance of chlorophyll-a parameter for measuring the lake water quality and it introduced the SIMILE project which signifies a relevant initiative to understand and manage the lake water quality of the previously mentioned lakes under study.

From the technical perspective, there was also introduced the background related to the machine and deep learning methods which could be used for classifying hyperspectral imagery.

Our study selected the Italian hyperspectral imager PRecursore IperSpettrale della Missione Applicativa (PRISMA) as the instrument to get the input data used for the proposed objective. The PRISMA mission was led by the Italian Space Agency (ASI), launched in 2019, and it is composed of VNIR and SWIR spectrometers and a panchromatic camera [60][18]. For this reason, a relevant aspect that was also covered in the project was related to the presentation of the PRISMA acquisitions used in the study and their associated chlorophyll-a maps. A first analysis of the available hyperspectral PRISMA images was carried out to determine which were the most suitable images to be considered according to a set of quality controls: cloud coverage, the extension of chlorophyll-a associated maps and the presence of glint. After the selection of the acceptable use cases to be taken into consideration for the study, an statistical analysis was performed for each of the finally selected acquisitions to understand the main statistics of the related chlorophyll-a maps.

Then, in the experimental development section 6, were explained the complete set of manipulations carried out on the input data: coregistration of input images, the intersection between hyperspectral PRISMA images and the associated chlorophyll-a maps,

pre-processing steps related to the removal of anomalous pixels, among others. Also, was mentioned the approach considered for determining which images corresponded to the training-validation set and which to the test set, and the normalization and dimensionality reduction techniques applied. Apart from this, there were specified the results obtained with each of the four model typologies taken into account in this project: Random Forest Regressor, Support Vector Regressor, Long-Short Term Memory networks and Gated Recurrent Unit networks. Finally, the document details the results achieved while inferring with 30-meter GSD inputs.

All the code developed in this project is publicly available in the following repository: `https://github.com/juanfranciscoamieva54/Regression-Chl-a-PRISMA.git`

3 | Structure of the document

The current document is structured in different chapters and the content of the main ones is detailed in the following:

Chapter 4 presents a review of the state of the art related to the main theoretical aspects that form the background of the entire project. The main motivation here was to provide a comprehensive theoretical foundation to the readers, introducing the various aspects covered in the project. These aspects include:

1. An explanation of the different hyperspectral missions that have been previously developed, as well as the ongoing ones, along with their main characteristics.
2. The importance of the chlorophyll-a parameter in assessing the water quality of lake bodies.
3. The SIMILE project, which was considered to obtain ground truth chlorophyll-a maps for this study.
4. An overview of the various classification approaches available for hyperspectral imagery.

Then, in Chapter 5, the input data that was considered and ultimately used for this work is presented. The main aim here was to provide a comprehensive introduction to:

1. The identification and pre-selection process of the available PRISMA hyperspectral acquisitions and their associated chlorophyll-a maps.
2. The final selection process after conducting a series of quality controls.
3. A detailed statistical analysis of each of the use cases, presenting the key statistics of the corresponding chlorophyll-a maps.

This approach has the objective of fully presenting the different aspects necessary to understand the data considered in the study.

Chapter 6 explains the entire experimental development that was carried out in our work

with the intention of communicating the complete process and convey the main concepts learned. The chapter covers the following aspects:

1. The pre-processing applied to the input hyperspectral images and the related chlorophyll-a maps.
2. The approach followed for defining the training-validation and test sets, as well as the normalization and dimensionality reduction techniques applied.
3. Detailed descriptions of the different experiments performed for each selected model typology (Random Forest Regressor, Support Vector Regressor, Long-short Term networks, Gated Recurrent Unit networks), specifying the main concepts learned.
4. The results obtained from performing inference on 30-meter GSD inputs.
5. A concluding discussion that highlights the main aspects of the chapter.

The objective of this chapter is to provide a comprehensive understanding of the experimental process, communicate the methodologies employed, and convey the key insights gained from the conducted experiments.

Finally, Chapter 7 was prepared with the intention of providing a summary of the main conclusions of the work. Among other aspects this chapter discusses:

1. The main insights of the experimental development carried out.
2. Limitations of the amount of available data.
3. Highlighting the significance of the developed models as tools for monitoring Chlorophyll-a levels in the analyzed lakes.
4. Exploring the possibility of expanding the models to different regions.

The objective of this chapter is to consolidate the main findings and outcomes of the study, address any limitations, and suggest avenues for further investigation.

4 | State of the Art

This chapter explores the state of the art providing an analysis of the main theoretical background required for the development of the present work:

- specifies the current and future evolution of hyperspectral imagery,
- describes the importance of the chlorophyll-a parameter for the monitoring and assessment of the lake water,
- introduces the SIMILE project
- describes some of the alternative approaches for classifying hyperspectral imagery.

4.1. Hyperspectral imagery evolution: from airborne to satellite missions

This section analyzes the evolution of hyperspectral imagery by describing in chronological order the initial developments in the field, by means of aircraft vehicles, then explores some of the main spaceborne missions and finally provides a summary of some of the main upcoming hyperspectral spaceborne missions.

4.1.1. Introduction to Imaging Spectrometry

In the remote sensing field, data is acquired by using different types of sensors. There are active sensors such as Synthetic Aperture Radar (SAR) and passive sensors, specifically optical sensors that rely on the acquisition of reflections from the Earth's surface. In this latter group, they are commonly mounted on top of various platforms like aircraft or satellites, and the characteristics of the instruments can also vary significantly.

Optical imagers have been well-established since their initial development in the early 1970s. Nowadays, many satellite missions have been launched that incorporate this type of instrument. Initially, they consisted of multispectral imagers, which had a limited

number of bands. However, a more recent development has been the introduction of Hyperspectral Imagers or Imaging Spectrometers. These instruments have a substantial characteristic: high spectral resolution, offering continuous coverage of the spectrum with a higher number of spectral bands.

Regarding hyperspectral imagers, in the early 1980s, they were mounted on top of aircraft vehicles. However, the first well-known spaceborne hyperspectral imager was released at the beginning of this millennium [60]. Hyperspectral imagers have the capability to measure the spectrum of each pixel in a certain acquisition, which is a powerful tool for studying and analyzing the Earth [60]. The spectral measurements are useful for determining the physicochemical composition of the Earth, leading to scientific research and applications on a regional scale [60].

Hyperspectral imagers order their information, producing a stack with different spectral bands. This structure is typically called a datacube, and it allows for the exploitation of the spatial and spectral information [60]. The sampled information for each pixel is a spectral vector that, in each of its components, has the value associated with each spectral band of the corresponding acquisition. Given the characteristics of this data, Hyperspectral Imagers are suitable for analyzing different challenging problems in various fields of application such as ecology, atmosphere, ocean, agriculture, and forestry, among others [71].

Figure 4.1 provides a representation of the acquisition of data by a generic spaceborne hyperspectral imager. It also presents an example of the mentioned datacube structure formed by the different spectral bands and some representations of spectral signatures associated with various components.

4.1.2. Initial development of airborne hyperspectral imagers

The first airborne hyperspectral imager was the Airborne Imaging Spectrometer (AIS) which was built by the National Aeronautics and Space Administration (NASA) at the beginning of the 1980s and covers the Shortwave Infrared (SWIR) spectrum from 1200 to 2400 nm with a Spectral Sampling Interval (SSI) of 9.3 nm [60].
Among the main drawbacks of this imager were found the following ones[60]:

- It did not cover the Visible and Near-infrared (VNIR) region, just the SWIR one.
- also had a narrow Field Of View (FOV) of around 3.7°.
- A third drawback was that it had a small detector array which meant that only a wavelength range of 32 x 9.6 nm = 307 nm could be covered.

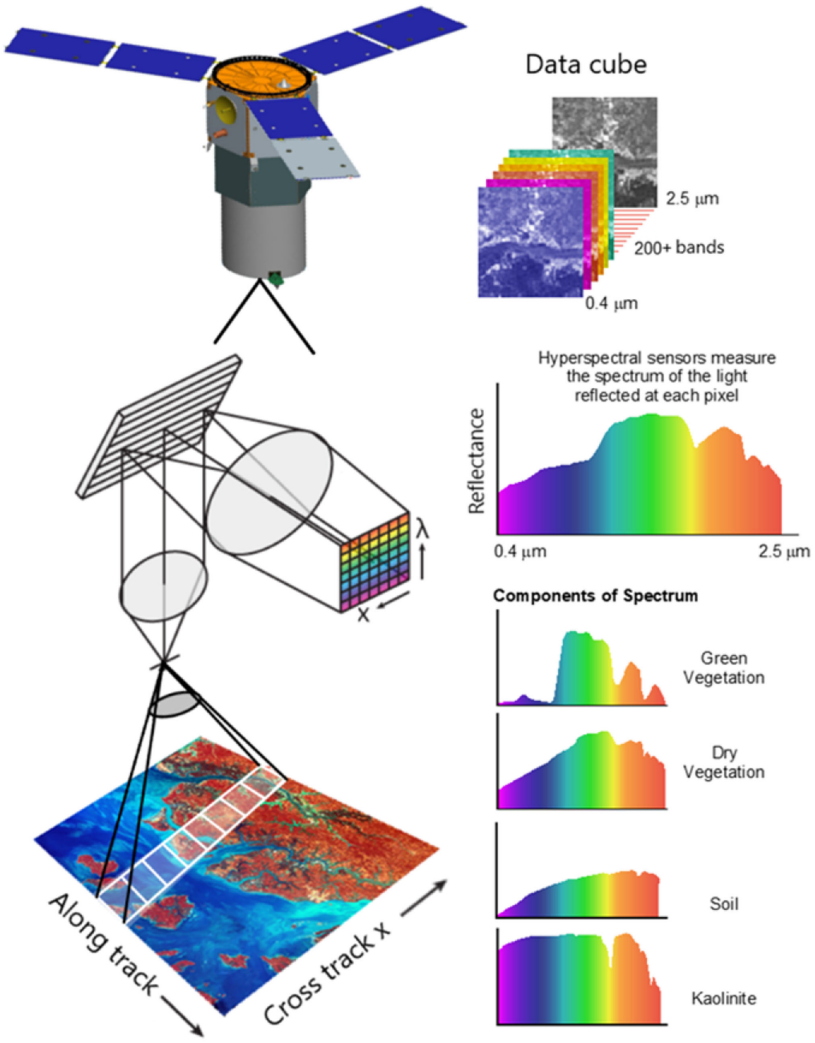

Figure 4.1: Representation of data acquisition by a generic hyperspectral satellite and examples of signatures [60] (Source: Qian, Shen-En, 2021).

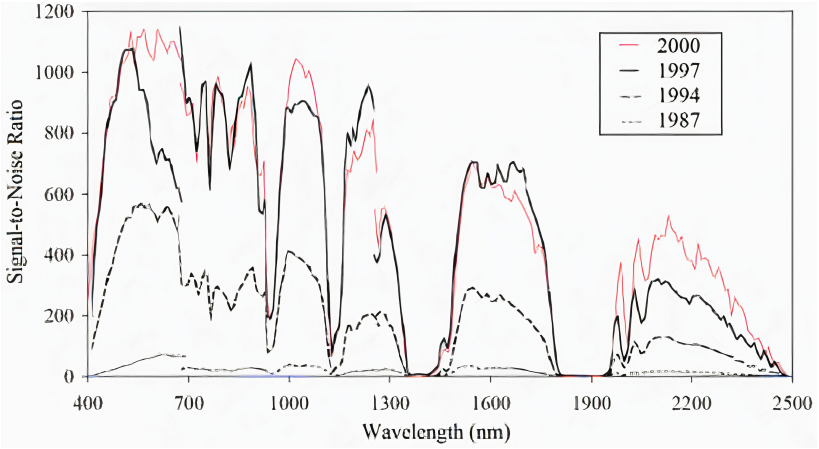

Figure 4.2: Evolution of SNR - AVIRIS [60] (Source: Qian, Shen-En, 2021).

In order to cover the complete range from 1200 to 2400 nm it had to rotate its grating into four positions within a short integration time.
Also, it is mentioned in [60] that AIS flew aboard NASA C-130 aircraft to be tested over a set of geological, geobotanical, natural and agricultural targets in the United States. These flights together with a later one over the Cuprite Mining District were valuable experiences in order to understand the potential of hyperspectral imaging and helped as a starting point for the later development.
Some years later, an improved version was developed: AIS-2. This version consisted of an imager covering a spectral range from 800 to 2400 nm and it improved some of the drawbacks of the previous version [60].

In 1983, JPL-NASA proposed the development of the *Airborne Visible-Infrared Imaging Spectrometer* (AVIRIS) imager which is considered the **first operational airborne hyperspectral imager**. Its development started in 1984 and the first flight was performed in 1986 aboard the NASA-ER2 aircraft [60].
Regarding its characteristics, it covers the spectrum that ranges from 400 to 2500 nm at 10 nm SSI, offering the best calibrated hyperspectral data due to its high Signal-to-noise ratio (SNR= and its calibration system and procedure. Also, it is worth mentioning that in order to cover its spectral range it uses four spectrometers instead of a rotating grating [60]. The instrument also suffered a set of upgrades which implied a relevant improvement over the initial expectations [60].

Due to the previous aspects mentioned, AVIRIS is the main hyperspectral data provider to the scientific community since its first flight [60].

In 1981 Canada's federal government Department of Fisheries and Oceans (DFO) also started the development of the first Canadian airborne hyperspectral imager: the *Fluorescence Line Imager* (FLI). The imager was built by two Canadian companies called Moniteq Ltd. of Toronto and ITRES Research Ltd. of Calgary and it was possible to be flown in 1983 [60].
This imager had two modes for its operation, one called spatial mode (digitizes and records a limited number of bands while maintaining full spatial resolution) in which it was possible to obtain a Ground Sampling Distance (GSD) of 2.5 m at 2 Km altitude and the second one called spectral mode in which it was offering a total of 288 bands covering the range from 430 to 800 nm. The imager had a distinct characteristic that consisted of its high SNR of about 1900:1 for a band of 16 elements [60].
FLI airborne flew over many different target locations, including phytoplankton mapping, benthic vegetation, and measurement of water depths in lakes and coastal areas. This imager was successful in determining the chlorophyll fluorescence which was one of its main objectives [60].

The Canadian company ITRES Research Limited developed another hyperspectral airborne imager but in this case for commercial purposes: in 1989 it was marketed the *Compact Airborne Spectrographic Imager* (CASI) [60]. This imager had the same modes of operation as FLI and also among its main aspects it had a total of 288 bands covering the range from 400 to 926 nm with an SSI of 1.8 nm, a GSD around 2-5 m and a swath width of 1-5 Km [60].
Some years after the release of CASI, ITRES added some modifications to CASI models in order to improve its performance [60].

In the same period, ITRES was involved in the development of another line of hyperspectral imagers which were focused on covering other regions of the spectrum: SWIR, mid-wave infrared (MWIR), and thermal infrared (TIR) [60].

Also, the Canada Center for Remote Sensing developed the first Canadian SWIR hyperspectral imager called *SWIR Full Spectrographic Imager* (SFSI) in 1992. Its main characteristics were its high spatial resolution of 0.5 m (from an airborne platform) and high spectral resolution of 10 nm covered by 122 bands from 1200 to 2400 nm [60].

Table 4.1 compares the main airborne imagers developed by Canada and the United

States.

Parameters	***FLI***	***CASI***	***SFSI***	***AIS***	***AVIRIS***
Period of Operation	1984-1990	Since 1989	Since 1992	1983-1985	Since 1987
Country	Canada	Canada	Canada	USA	USA
Spectral Range (nm)	430-805	418-926	1200-2400	1200-2400	approx. 410-2450
Number of spectral elements	288	288	122	128 after rotating to the 4 positions	224
SSI (nm)	1.3	1.8	10	9.3	10
Airplane altitude (km)	2	2	2	6	20
Swath width (km)	4.8	1-5	0.26	0.37	11
GSD (m)	2.5	2-5	0.5	11.4	20

Table 4.1: Comparison of the main airborne of Canada and United States. Elaboration based on [60]

During the 1980s-1990s, there were developed also in other countries a series of hyperspectral airborne imagers. Among them can be distinguished [60]:

- *Multispectral Infrared and Visible Imaging Spectrometer* (MIVIS): Developed in 1993 by a company called Daedalus Inc. it covered the visible, near-IR, mid-IR, and thermal-IR regions with a total of 102 spectral bands.
- The *Hyperspectral Data and Information Collection Experiment* (HYDICE) was operated in 1994. It covered the range from 400 to 2500 nm with a GSD of 3 m at a 6 Km altitude. Its SNR exceeded considerably the one of AVIRIS. It was acquired under the control of the Department of Defense (DoD) of the United States.
- The *Digital Airborne Imaging Spectrometer* (DAIS) was a development of Geophysical Environmental Research of Millbrook, NY which entered into operation in 1994, covering visible, SWIR, and TIR with a total of 79 bands.
- The *Australian Hyperspectral Imager* (HyMap) was built by the Hy Vista Corporation. It had a swath of 2.3 km with 5 m GSD at 1.3 km altitude or 4.6 km with 10 m GSD at 2.6 km altitude, it covered a total of 32 bands in each of its 3 spectrographic modules: visible module, and the near-infrared and short-wave infrared-1 (SWIR1) and 2 (SWIR2) modules.

4.1.3. Evolution of hyperspectral spaceborne imagers

The beginning of the development of the spaceborne hyperspectral imagers could be associated in some way with the implementation of the *Shuttle Multispectral Infrared Radiometer* (SMIRR) led by JPL- NASA which was initially planned to fly in 1979 but was finally launched in November 1981 [60].

The instrument comprised a set of ten bands spanning the spectral regions of the VNIR and SWIR [60]. The main aim of this mission was to determine if it was feasible to identify minerals from space using hyperspectral imaging [60].

Given the success of the AIS and SMIRR missions, NASA started developing a hyperspectral program that initially considered three new projects [60]: i) *AVIRIS* which was the next-generation aircraft hyperspectral imager, ii) the *shuttle imaging spectrometer experiment* (SISEX), and iii) the free-flyer *High-Resolution Imaging Spectrometer* (HIRIS). Although the development and operation of AVIRIS were successfully reached, in the case of SISEX it was cancelled because of the space shuttle Challenger disaster in 1986, and HIRIS was not launched because the complete program was reduced to one-fourth of its original size [60].

The first spaceborne hyperspectral imager was for military purposes and it was called *Ultraviolet and Visible Imagers and Spectrographic Imagers* (UVISI). It was mounted on the Midcourse Space Experiments (MSX) satellite mission of the United States Department of Defense, which was launched in 1996 [60]. It had five spectrographic imagers in UV to VNIR and four UV and visible multispectral imagers. The main drawback of this imager was that it had a GSD of 770 m, which is another reason why it is not so well-known in the scientific community. The coverage of this instrument specifically comprised the region of wavelengths between 110 to 900 nm with a total of 1360 spectral bands and a swath width of 15 Km [60].

During the beginning of the 90's NASA associated with TRW Inc in order to develop a spaceborne imaging spectrometry system *HyperSpectral Imager* (HSI) for the LEWIS mission. It comprised 128 bands in the VNIR region of 400 - 1000 nm (5 nm SSI) and 256 bands in the SWIR region of 900–2500 nm (6.5 nm SSI) [60]. The instrument also

considered a swath width of 7.7 Km and a GSD of 30 m. Although HSI was launched, the control of the LEWIS satellite was lost on the third day after its launch in September 1997 [60].
The next hyperspectral imager developed by NASA was ***Hyperion*** which was mounted on top of the EO-1 satellite; it represented a really successful mission which had a long period of operation considering that it was launched in the year 2000 and decommissioned in March 2017 after 17 years of operation. Given its success, it is well known in the scientific community and it is **usually considered the first spaceborne hyperspectral imager** [60].
It is reported in [60] that the main characteristics of Hyperion were that it comprised a telescope and two spectrometers, one on the VNIR region comprising 400–1000 nm range of the spectrum and the other in the SWIR around the 900–2500nm. The overlapping region between 900-1000 nm was used for the cross-calibration of the spectrometers. Hyperion had an SSI of 10 nm, a GSD of 30 m and the swath width was around 7,65 Km. According to [60], the SNR was between 140:1 and 190:1 in the VNIR region from 550 to 700 nm, 96:1 at 1225 nm, and 38:1 at 2125 nm.

The table 4.2 details the relevant information regarding access to Hyperion's data.

Data portal's URI	`https://earthexplorer.usgs.gov/`
Conditions of access and use	Registration is required. Data are freely available for registered users.
Product levels	***L1R:** This level corresponds to radiance and radiometrically corrected data but is not georeferenced. ***L1T:** In this case the data of L1R level is orthorectified and geocoded (UTM projection). ***L1Gst:** It corresponds with terrain-corrected information and it is delivered in 16-bit radiance values.
Data formats	L1R Product in HDF format, L1Gst Product in GeoTiff format, and L1T Product in GeoTiff format.

Table 4.2: Hyperion's data main characteristics [61] [51]

The next hyperspectral spaceborne imager was the *Compact High-Resolution Imaging Spectrometer* (CHRIS), led by the European Space Agency (ESA) and it was launched in

PROBA-1 satellite in the year 2001. In [28] are reported the main parameters of CHRIS: its spatial resolution is 18 m at the nadir, the swath width is 14 Km, the spectral range goes from 400 to 1050 nm in VNIR, a spectral resolution from 1.25 to 11 nm and an image size of 13 Km x 13 Km. Regarding the available modes it is possible to use the following ones [28]:

- MODE 1: Full swath width, 62 spectral bands, 773 nm / 1036 nm, nadir ground sampling distance 34 m at 556 km.
- MODE 2 WATER BANDS: Full swath width, 18 spectral bands, nadir ground sampling distance 17 m at 556 km.
- MODE 3 LAND CHANNELS: Full swath width, 18 spectral bands, nadir ground sampling distance 17 m at 556 km.
- MODE 4 CHLOROPHYLL BAND SET: Full swath width, 18 spectral bands, nadir ground sampling distance 17 m at 556 km.
- MODE 5 LAND CHANNELS: Half swath width, 37 spectral bands, nadir ground sampling distance 17 m at 556 km.

In table 4.3 are described the main characteristics of CHRIS' data.

Data portal's URI	`https://eocat.esa.int/`
Conditions of access and use	Registration is required. Data are freely available for registered users.
Product levels	The 5 mentioned modes can be requested.
Data formats	HDF format.

Table 4.3: CHRIS' data main characteristics

In the year 2002 was launched the Environmental Satellite (ENVISAT) which had on-boarded the *Medium Resolution Imaging Spectrometer* (MERIS). The first reception of data from the instrument was in May 2002 and it was operational until 2012 [32].

MERIS was mainly dedicated to ocean colour observations but also comprised in its scope objectives related to atmospheric and land surface studies [32]. According to [60], MERIS could record 520 bands in the range from 390 to 1040 nm with a high spectral resolution of 1,25 nm. The wide swath was of 1150 Km and the GSD varied from 260 m

at the nadir to 390 m at the extremities.
Another fortress of MERIS was its SNR which was around 1700 in the ocean colour bands. Although the high spectral resolution, MERIS had a restriction in the downlink and only transmitted 15 channels which was an average of the native spectral elements.

Details about MERIS' data characteristics can be observed in table 4.4.

Data portal's URI	`https://meris-ds.eo.esa.int/oads/access/collection`
Conditions of access and use	Registration is required. Data are freely available for registered users.
Product levels	There are four available product levels: ***L1 Full Swath Full ResolutionInfo, *L2 Full Swath Full ResolutionInfo, *L1 Reduced ResolutionInfo, and *L2 Reduced Resolution**: Level 1 products are composed of 22 measurements data files (15 radiance bands + 7 annotation data files) [1] while the Level 2 products are composed by composed of 64 measurement files containing: Water-leaving reflectance, Land surface reflectance, and TOA reflectance, and additional measurement on Ocean, Land and Atmospheric parameters and annotation [2]. The full-resolution products cover the complete instrument swath and the Reduced Resolution data was acquired over 80 % of the descending track.
Data formats	NetCDF 4.

Table 4.4: MERIS' data main characteristics [33]

The National Committee for Disaster Reduction and State Environmental Protection Administration of China developed a minisatellite constellation called HJ-1 (Chinese abbreviation of "Huan Jing" which means "environment"). It comprises three satellites HJ-1A, HJ-1B, and HJ-1C.
In [56] it is explained that the objective of the mission is mainly to establish an earth-observing system for disaster monitoring and mitigation.

In the year 2008 was launch of the HJ-1A and HJ-1B satellites of the constellation. The HJ-1A satellite comprises two optical cameras and the *Fourier Transform Hyperspectral Imager* (FTHSI). The FTHSI imager is characterized by 115 spectral bands, a spectral range that goes from 450 to 950 nm (VNIR region), 100 m GSD a 50 Km of swath width, and an SNR around 50-100. The data of FTHSI is not publicly available.

Also, in 2008 the Indian Space Research Organization (ISRO) launched the Indian Microsatellite – 1 (IMS-1) which supported a multispectral camera and the *Hyperspectral Imager* (HySI). It is mentioned in [57] that the main objectives were to obtain experience in this type of payload and in the handling of hyperspectral data to generate application models.
Regarding the main technical aspects of HySI, it is also explained in [57] that it covered a spectral range from 400 to 950 nm (VNIR) with a total of 64 bands, characterized by an SSI of 8 nm a GSD of 550 m at nadir and a swath width of 128 Km. It just had coverage over India.
The IMS-1 mission had its end of life on 31 July 2013 [57].

Table 4.5 describes the main characteristics of HySI data.

Data portal's URI	`https://bhuvan-app3.nrsc.gov.in/data/download/index.php`
Conditions of access and use	Registration is required. Data are freely available for registered users.
Product levels	Radiance and TOA Reflectance
Data formats	GeoTiff.

Table 4.5: HySI's data main characteristics

In May 2009 the TacSat-3 satellite was successfully launched, supporting the *Advanced Responsive Tactically Effective Military Imaging Spectrometer* (ARTEMIS). It is explained in [60] that it was the third in a series of satellites of the United States Department of Defense and here the objective was to demonstrate the ability to provide real-time data collected from space to combatant commanders in the field.
Although TacSat-3 was initially designed for six months of operation, with a goal of one year, [55] explains that it outlives its originally designed life and also surpassed the initial goals and in the year 2010 it transitioned from experimental to operational status,

transferring the control from the Air Force Research Laboratory and Space and Missile Systems Center to the Air Force Space Command. Finally, in the year 2012, the satellite reentered Earth´s atmosphere almost three years after its launch.
Regarding the technical characteristics of ARTEMIS, in [55] it is detailed that this instrument comprised a spectral range from 400 nm to 2500 nm (coverage in VNIR and SWIR regions), provided an SSI of 5 nm implying a total of more than 400 bands, also [60] mentions that ARTEMIS had a GSD of 4 m and a swath width of 4 km.

Information about how to access ARTEMIS data is not available.

On the 23rd of September 2009 was deployed in the International Space Station (ISS) the *Hyperspectral Imager for the Coastal Ocean* (HICO). [19] explains that HICO was the first spaceborne coastal Maritime Hyperspectral Imager designed to have a high SNR for dark coastal scenes, capturing large scenes at moderate spatial resolution and offering high sensitivity in the blue and full coverage of water-penetrating and near IR wavelengths.
Also, it is detailed in [19], that given the mentioned requirements, the main applications of HICO are to derive bathymetry, water optical properties, bottom type, and terrain and vegetation maps for coastal areas of interest around the world.
HICO was developed by The Naval Research Laboratory for the Office of Naval Research as an Innovative Naval Prototype [46]. After its success as a prototype, it continued the operation for five years ending in 2014.
About the main parameters in [60] is described that HICO operated in the VNIR region from 380 to 960 nm with an SSI of 5.7 nm. It had a swath width of 51 km when the ISS altitude was 420 km. Its GSD was 100 m and it offered a reasonably high SNR: peak SNR 470:1 at 480 nm, SNR $>$ 200:1 in the spectral range 400–600 nm.

The table 4.6 explains the data access details of this instrument.

In the launch of ESA's Sentinel 3A satellite on February 2016, it was put in orbit the *Ocean and Land Color Imager* (OLCI). According to [60], OLCI is the successor of MERIS (which was out of service in 2012) and it was designed to *"provide global and regional measurements of ocean and land surface with high radiative accuracy based on the heritage design from MERIS"* [60].
OLCI transmits 21 spectral bands covering the spectral range from 400 to 1020 nm, it has a swath width of 1270 Km, a GSD of 300 m [35] and its revisit time is 3 days [60]. In order to minimize the impact of the solar reflection from sea surfaces ("sun-glint"), an asymmetric swath width with respect to the satellite ground track was adopted [60]. The

Data portal's URI	`https://oceancolor.gsfc.nasa.gov/cgi/browse.pl?sen=amod`
Conditions of access and use	Registration is required. Data are freely available for registered users.
Product levels	***L1A:** applies instrument/radiometric calibrations to the raw data. ***L2:** derived geophysical variables at the same resolution as the source Level 1 data.
Data formats	NetCDF.

Table 4.6: HICO's data main characteristics [45]

design lifetime is 7.5 years [35].

In table 4.7 are described the main aspects related to data accessing for this instrument.

Data portal's URI	`https://scihub.copernicus.eu/dhus/#/home`
Conditions of access and use	Registration is required. Data are freely available for registered users.
Product levels	***Level-1 TOAR:** This level corresponds with the top of atmosphere radiance (mW/m2/sr/nm) maps. ***Level-2 OTCI:** The second level OTCI corresponds with OLCI Terrestrial Chlorophyll Index maps. ***Level-2 GIFAPAR:** In this case the OLCI Green Instantaneous Fraction of Absorbed Photosynthetically Available Radiation information is provided. ***Level 2 IWV:** The last available level refers to the OLCI Integrated Water Vapour it is expressed in kg/m2.
Data formats	NetCDF.

Table 4.7: OLCI's data main characteristics [26] [9]

In June 2016 was launched Greenhouse Gas Satellite - Demonstrator (GHGSat-D) has the *Miniature Hyperspectral SWIR Imaging Spectrometer* (MHRIS) as a secondary payload. GHGSat is a commercial venture of GHGSat Inc. of Montreal, Canada, a subsidiary of Xiphos Systems Corporation [58].
The mission of GHGSat is to be a global reference for the measurement of greenhouse gases and air quality gas emissions from industrial locations using remote sensing [58].

Among the technical requirements of the GHGSat-D mission is the quantification of carbon dioxide total atmospheric columns above target sites and quantifying greenhouse gas columns at a ground sample distance of 50 m or less.

During its operation, the satellite performed observations of Oil gas facilities (wells, compressor stations, refineries, LNG, etc), Power stations (thermal and hydroelectric), Coal mines (open-pit and underground), Landfills, Animal feedlots and Natural sources (e.g. mud volcanos, localized seeps, etc) [58].

The MHRIS hyperspectral imager of the GHGSat-D satellite covers from 1600 to 1700 nm wavelengths with a total of more than 300 bands, implying an SSI of around 0.1 nm. Regarding the swath width, it is 15 Km and the GSD is under 50 m [60].
Table 4.8 describes the main characteristics of the instrument data accessing.

Data portal's URI	`https://earth.esa.int/eogateway/catalog/ghgsat-archive-and-tasking`
Conditions of access and use	ESA offers for scientific research and application development, access to the archive, and new tasking data from the GHGSat mission upon submission and acceptance of a project proposal.
Product levels	***Abundance dataset (Level 2):** This level offers per-pixel abundances (ppb or mol/m2) for a single species, and per-pixel measurement error. ***Concentration Maps (Level 2):** It corresponds with a high readability pseudocolour map which combines surface reflectance, and column density expressed in ppb or mol/m2.***Emission Rates (Level 4):** It is the instantaneous emission rates from certain targeted sources, estimated using abundance datasets.
Data formats	GeoTiff.

Table 4.8: MHRIS' data main characteristics [34]

Aalto-1 Spectral Imager (AaSI) is the main payload of the Aalto-1 satellite developed by VTT Technical Research Centre of Finland and launched on June 2017. AaSI has a spectral range from 500 to 900 nm with a spectral resolution that can be adjusted elec-

tronically from 10 to 30 nm [60].

Information about how to access AaSI data is not available

The *DLR Earth Sensing Imaging Spectrometer* (DESIS) is a hyperspectral imager developed by the company Teledyne Brown Engineering and the German Aerospace Center (DLR) [15]. DESIS is predominantly a commercial mission that was launched on June 2018 and it was installed on the exterior of the ISS on August 2018 where it is hosted in the multiuser system for earth sensing [60].
A main difference of DESIS with respect to the other hyperspectral imagers is that it has the capability to point forward and back directions up to ±15°. It operates in static mode and also in a dynamic mode which allows the acquisition of continuous observations of the same targets [60]. Among other relevant technical parameters, it is worth mentioning that it operates in the VNIR region from 400 to 1000 nm with a minimum SSI of 2.55 nm. It has a ground swath width of 30 km with a GSD of 30 m [60], the revisit time is around 3-5 days (mainly depending on the manoeuvres) and the expected lifetime is from 2018 to 2023 [15].

Table 4.9 mentions the main aspects related to data access for the DESIS mission.

Data portal's URI	`https://eoweb.dlr.de/egp/main`
Conditions of access and use	Can be ordered for free for scientific purposes. It requires a short description of the use of the data in form of a proposal and a subsequent review process.
Product levels	***L1B:** Level 1B applies corrections on certain systematic effects: odd-even, non-linearity, and non-uniformity. Also converts digital numbers (DN) to physical at-sensor radiance values. ***L1C:** Level 1C applies geometric correction of the L1B Earth Products. This level is a geometrically corrected image, resampled and transformed into a map projection system.***L2A:** It is derived from the L1C product, but it is applied atmospheric correction and the data is converted to ground surface reflectance values.
Data formats	GeoTiff.

Table 4.9: DESIS' data main characteristics [16] [8]

HyperScout is a hyperspectral satellite on board the CubeSat GomSpace Express-4B (GomX-4B) developed by the ESA and the Netherlands Space Office (NSO) which was launched on February 2018 and It is planned to be operative until December 2022 [30].
The GomX-4B is part of a constellation of the two CubeSats GomX-4A and GomX-4B which has as one of the principals aims the demonstration of the inter-satellite linking radio communication subsystem and integration of partner contributions [59].
The main objective of HyperScout is the demonstration of hyperspectral applications on a CubeSat and the demonstration of early warning capabilities. Among its main applications are the detection of floods, change detection, crop water factor and vegetation monitoring [30].
The coverage of HyperScout is from 400 to 1000 nm in the VNIR range with an SSI of 15 nm a swath width of 200 Km and a GSD of 70 m [60].

For HyperScout was determined in [31] that the data format is HDF5 but its access is constrained.

In May 2018 was launched the Gao Fen – 5 (GF-5) satellite, developed by the Chinese National Space Administration with a planned lifetime of 8 years [42].
The main payload GF-5 is the *Advanced Hyperspectral Imager* (AHSI) and its additional payloads are a Visual and Infrared Multispectral Sensor (VIMS), a Greenhouse-gases Monitoring Instrument (GMI), an Atmospheric Infrared Ultraspectral (AIUS), an Environment Monitoring Instrument (EMI) and a Directional Polarization Camera (DPC) [42].
The AHSI imager has 330 spectral bands which cover the spectrum from 400 to 2500 nm in the VNIR and SWIR regions. In the VNIR region (400 – 1000 nm) the SSI is 5 nm and, in the SWIR, (1000 – 2500 nm) it is 10 nm. The swath width is 60 Km, the GSD is 30 m and the peak SNR is 654:1 which is remarkably high if it is compared with the one of Hyperion with a peak SNR of 190:1 [60]. Information about AHSI imager is not publicly available.

In March 2019 was launched the *PRecursore IperSpettrale della Missione Applicativa* (PRISMA) satellite with the objective of qualifying the technology, contributing to the development of applications, and providing products to institutional and scientific users for environmental observation and risk management [60].

The PRISMA satellite was built by the Italian Space Agency (ASI) with OHB Italia Spa as the prime contractor and Leonardo Space Airborne Systems was responsible for the payloads. About the payloads, they are VNIR and SWIR spectrometers and a panchromatic camera with a high spatial resolution of 5 m [18].

Regarding the main technical attributes, it is characterized by a lifetime of 5 years, a revisit time is 29 days, a GSD of 30 m in the hyperspectral imagers (VNIR-SWIR), a swath width of 31 Km, coverage of the VNIR range from 400 to 1010 nm with 66 spectral bands and 174 bands cover the SWIR range from 920 to 2500 nm implying an SSI of 12 nm [18].

Table 4.10 describes the main characteristics related to data access for PRISMA.

Data portal's URI	`http://prisma.asi.it/`
Conditions of access and use	First, It is required to submit a user request specifying the further use of the information. Registered users can access the data freely.
Product levels	***L1:** It corresponds with radiometrically corrected and calibrated radiance data in physical units. ***L2B:** Here is provided geolocated at Ground (Bottom-of-Atmosphere) Spectral Radiance Product. ***L2C:** This level offers geolocated At-surface (Bottom-of-Atmosphere) Reflectance Products. ***L2D:** This is the higher processing level and corresponds with the geocoded version of the level 2C products.
Data formats	HDF5.

Table 4.10: PRISMA's data main characteristics [4]

The *Hyperspectral Imager Suite* (HISUI) is a development of the Japanese Ministry of Economy, Trade and Industry which was launched on December 2019 and is currently onboard the ISS. The main objective of the project is to acquire data to start a full-scale practical application development for hyperspectral remote sensing. In particular, it is intended to determine if the instrument is useful in applications such as oil resource exploration and evaluate the potential of the imager to acquire know-how for later developments [69].
A relevant aspect to mention is that HISUI has an HSI and an MSI which can operate

simultaneously or not. Among other main technical aspects of the HSI, it consists of a telescope and two spectrometers covering VNIR and SWIR. In the VNIR it has an SSI of 10 nm with a total of 57 bands and in the SWIR it possesses 128 bands with an SSI of 12.5 nm, the swath width is 20 Km and the GSD is 30 m [60].
Also, it is worth mentioning that given the downlink constraints in the ISS the data is partially transmitted to the ground stations, and the remaining data are recorded, stored, and shipped to earth three or four times a year [60].
The key aspects of HISUI's data access are detailed in table 4.11.

Data portal's URI	Not found.
Conditions of access and use	Data Policy under consideration. For research purposes, HISUI will give priority observation, downlink, and distribution for free. For other science users, HISUI will distribute archive data.
Product levels	**L0:** Raw data. **L1A:** It refers to the raw DN product with all radiometric calibration coefficients. **L1R:** Top-of-atmosphere spectral radiance product. **L1G:** This level offers a geometrically corrected / orthorectified top-of-atmosphere spectral radiance product. **L2:** Atmospherically corrected surface spectral reflectance product corresponds with this level.
Data formats	Tif, GeoTiff, TBD.

Table 4.11: HISUI's data main characteristics [68] [25]

One of the latest missions was the Indian *Geo Imaging Satellite* (GISAT-1) class **which failed in its launch on August 2021** [43]. This instrument was designed with the objective of offering a high revisit capability orbiting in a geostationary orbit.
There were two identical GISAT satellites, each of them with an MSI and an HSI. The HSI was a VNIR spectrometer, covering the range from 380 to 1000 nm with a GSD of 318 m and another spectrometer capturing the SWIR from 900 to 2500 nm with a GSD of 191 m [60].
According to [43] it is planned a second satellite GISAT-2 which consist in a repeat of the first one.

On March 2022 was launched the *Environmental Mapping and Analysis Program* (EnMAP) german hyperspectral satellite[23]. This satellite has the mission of monitoring

and characterizing the Earth's environment on a global scale, measuring and modelling the Earth's ecosystems processes, and extracting geochemical, biochemical, and biophysical information critical to further understand various ecosystems [24].
The main parameters of the imager are its spectral range which goes from 420 to 1000 nm in the VNIR region with an SSI of 6.5 nm and from 900 to 2450 nm in the SWIR with an SSI of 10 nm, the swath width of 30 Km and a GSD of 30 m. It is also characterized by a better SNR than the other available spaceborne hyperspectral imagers: SNR 500:1 at 495 nm (VNIR) and 150:1 in the SWIR region can be achieved [60].

There are detailed in table 4.12 the main aspects of the access to EnMAP data.

Data portal's URI	`https://planning.enmap.org/`
Conditions of access and use	Registration is required. Data are freely available for registered users. Access levels are approved after a required submission.
Product levels	***L1B:** Raw values converted into at-sensor radiance physical values. ***L1C:** Orthorectified images. ***L2A:** The atmospherically corrected imagery reflectance values for land and water areas. Also provided several quality masks.
Data formats	Tif.

Table 4.12: ENMAP's data main characteristics [22]

Table 4.13 summarizes the main aspects already mentioned about the different hyperspectral spaceborne imagers explored in the document.

Hyperspectral Imager	*Satellites - Platform*	*Launch Year*	*Spectral range (nm)*	*Number of Spectral bands*	*SSI (nm)*	*GSD (m)*	*Swath width (Km)*	*Data Platform's URI*
HSI	LEWIS	1997	400 - 1000 and 900 - 2500	128 and 256	5 and 6.5	30	7.7	Not found
Hyperion	EO-1	2000	400 - 2500	220	10	30	7.7	https://earthexplorer.usgs.gov/
CHRIS	PROBA	2001	400 - 1000	19-62	1.25-11	25-50	13	https://eocat.esa.int/
MERIS	ENVISAT	2002	390 - 1040	520 (transmit 15)	1.25	300	1150	https://meris-ds.eo.esa.int/oads/access/collection
FTHSI	HJ-1A	2008	450 - 950	115	4	100	50	Not publicly available
HySI	IMS-1	2008	450 - 950	64	8	500	130	https://bhuvan-app3.nrsc.gov.in/data/download/index.php
ARTEMIS	TacSat-3	2009	400 - 2500	400	5	4	4	Not found
HICO	ISS	2009	350 - 1080	128	5.7	90	51	https://oceancolor.gsfc.nasa.gov/cgi/browse.pl?sen=amod
OLCI	Sentinel 3A	2016	390 - 1040	520 (transmit 21)	1.25	300	1270	https://scihub.copernicus.eu/dhus/#/home
MHRIS	GHGSat-D	2016	1600 - 1700	512	0.2	50	15	https://eoiam-idp.eo.esa.int/
AaSI	Aalto-1	2017	500 - 900	6 - 20	7 - 10	192	97	Not found
DESIS	ISS	2018	400 - 1000	235	2.55	30	30	https://eoweb.dlr.de/egp/main
HyperScout	GomX-4B	2018	400 - 1000	45	15	50	200	Constrained access
AHSI	GaoFen-5	2018	400 - 2500	330	5 (VNIR) - 10 (SWIR)	30	60	Not publicly available
PRISMA	PRISMA	2019	400 - 2510	237	12	30	30	http://prisma.asi.it/
HISUI	ISS	2019	400 - 2500	185	10 (VNIR) - 12.5 (SWIR)	30	20	Not found
HSI - GISAT 1	GISAT	Failed launch	380 - 1000 and 900 - 2500	158 and 256	3.92 (VNIR) - 6.25 (SWIR)	318 and 191	160 and 190	Not available
EnMAP	German HS	2022	420 - 2500	244	5 (VNIR) - 10 (SWIR)	30	30	https://planning.enmap.org/

Table 4.13: Summary of the main information of the principal spaceborne hyperspectral imagers.

In figure 4.3 are summarized the different spaceborne hyperspectral missions discussed in the current sub-section.

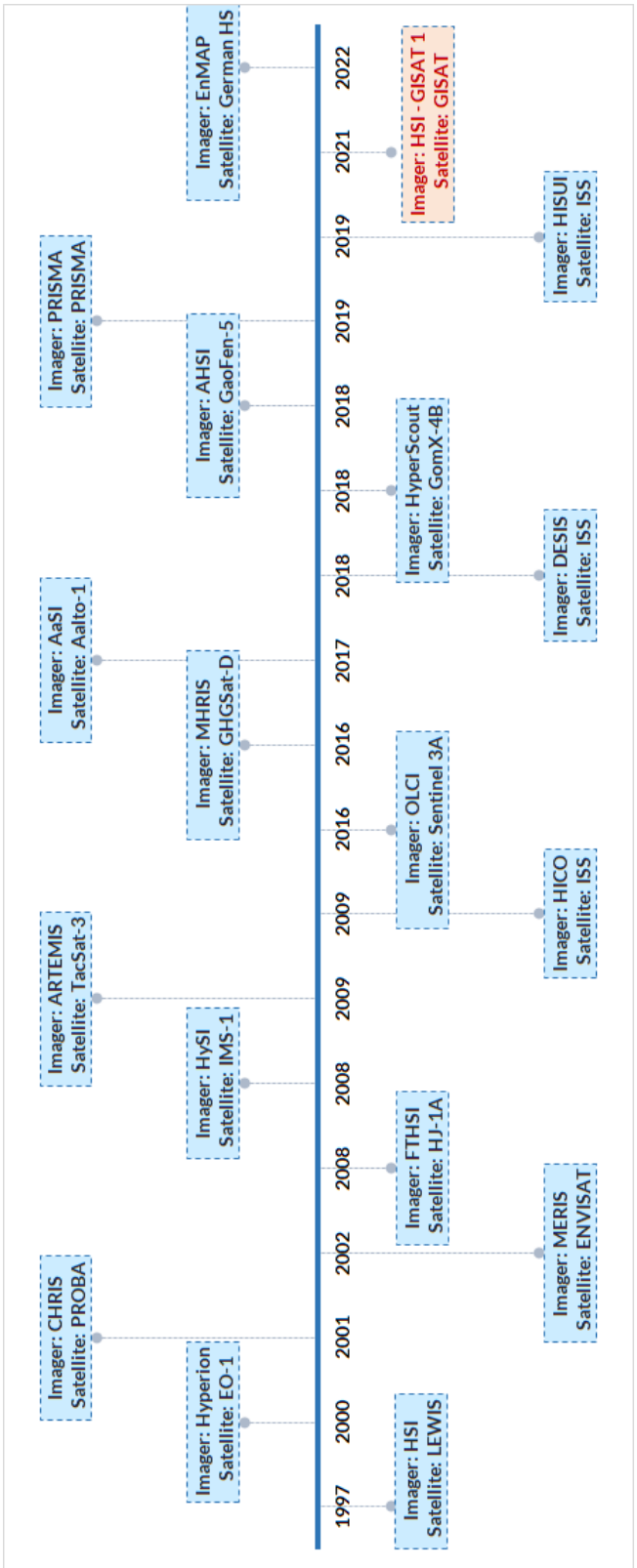

Figure 4.3: Summary of main hyperspectral spaceborne missions

4.1.4. Planned Spaceborne Hyperspectral Imagers

According to [52], the company Planet plans to launch a set of two hyperspectral satellites during the year 2023 to complement its constellation of optical imagers. The name of this planned mission is *Tanager* and their main characteristics are its spatial resolution of 30 meters, an SSI of 5 nm covering a total of 400 bands that range from 400 nm to 2500 nm [53]. The main applications considered by this mission are related to the detection of social, environmental and climate risks by the understanding of human activities and their economic and environmental impacts [53].

The *Ocean Color Instrument* (OCI) is a hyperspectral imaging radiometer onboard NASA's Plankton, Aerosol, Cloud, Ocean Ecosystem (PACE) [60] which is planned to be launched in 2024 [49].
Its coverage extends from 340 to 890 nm in the ultraviolet (UV) to near-infrared spectrum at 5 nm resolution but it is a goal to increase this SSI to 2,5 nm. It also includes 7 discrete bands from 940 nm to 2260 nm in the shortwave infrared (SWIR) spectrum [48]. It also will be characterized by a GSD of 1000 m and a swath width of 2663 Km [60].

Another upcoming mission is NASA's *Hyperspectral Infrared Imager* (HyspIRI) that was recommended in the 2007 National Research Council Decadal Survey requested by NASA, NOAA, and USGS [47]. The main objectives of this instrument are to *"study the world's ecosystems and provide critical information on natural disasters such as volcanoes, wildfires, and drought. HyspIRI will be able to identify the type of vegetation that is present and whether the vegetation is healthy. The mission will provide a benchmark on the state of the world's ecosystems against which future changes can be assessed. The mission will also assess the pre-eruptive behaviour of volcanoes and the likelihood of future eruptions, as well as the carbon and other gases released from wildfires"* [47].
The main characteristics of HyspIRI are that it will consist of two instruments: a multispectral imager and a hyperspectral imager. Regarding the HSI, it will cover the Visible to Short Wave Infrared (VSWIR) from 380 to 2510 nm with an SSI of 10 nm, the GSD will be 60 m at nadir, and the revisit time 19 days. The satellite will be launched in LEO but at the moment it is not specified the date [60].

The European Space Agency (ESA) is planning to launch Mid-2024 on the Vega-C rocket from Kourou, French Guiana the Fluorescence Explorer (FLEX) satellite that will support the *Fluorescence Imaging Spectrometer* (FLORIS) which is a high-resolution imaging spectrometer [29].

The main goals of the mission are to provide globally a measurement of the vegetation fluorescence, give a quantification on the photosynthetic activity and plant stress, allow the capability to understand the functioning of the photosynthetic machinery and the health and performance of the terrestrial vegetation, and to produce global maps with a monthly frequency[29].
The prime contractor of FLORIS is Leonardo Space and Airborne Systems and it will consist of two spectrometers covering the spectral range between 500 and 780 nm. One of the spectrometers will measure the fluorescence spectrum within two oxygen absorption bands called O2A and O2B with an SSI that varies from 0.1 nm to 0.5 in the range of 677 to 780 nm. The other spectrometer is focused on the measure of other parameters of the atmosphere and the vegetation, and it covers the range of 500 to 740 nm with an SSI that goes from 1 to 2 nm. The GSD is 300 m and the swath width is 150 Km [60]. The estimated lifetime of the mission is 3.5 years [29].

Another ongoing mission of ESA is the *Copernicus Hyperspectral Imaging Mission for the Environment* (CHIME) that will contribute to the generation of information to help the design of policies to manage natural resources in the European Union and also it will support the actions to ensure food security at a global scale [27]. At the moment it is not reported a specific date for its launch.
CHIME will cover the VSWIR from 400 to 2500 nm with an SSI of 10 nm, a GSD of 20-30 m, and a 128 Km swath width. The satellite will operate in a sun-synchronous LEO orbit and offer a revisit time of around 10 to 12.5 days [60].

Finally, it is worth mentioning the Spaceborne Hyperspectral Applicative Land and Ocean Mission (SHALOM) which is a co-funded and managed mission by the Italian Space Agency (ASI) and the Israeli Space Agency (ISA) [50]. SHALOM is introduced in [50] as a next-generation mission because it expects to provide a commercial service with high spatial resolution, high daily area coverage, short revisit times and precise geo-location [50]. Specifically, [50] describes the following mission requirements:

- **GSD:** Less than 10 meters.
- **Swath width:** More than 10 Km.
- **Revisit time:** Below 4 days.
- **Spectral coverage:** 400-2500 nm.
- **Spectral sampling interval:** 10 nm.

- **Daily area coverage:** More than 200.000 Km^2.
- **SNR:** > 200 in the range between 400 nm and 1750 nm, and higher than 100 in the range from 1750 nm until 2350 nm.
- **Geolocation accuracy:** Higher than 30 meters.
- **Panchromatic Camera:** 2,5 - 5 meters GSD, 10 Km Swath width, VNIR band SNR > 240.

According to [41], the launch of this satellite mission is planned for the year 2025 and it is expected an operation time of 5 years.

In table 4.14 there is a summary of the previously introduced planned spaceborne hyperspectral imagers.

Mission	*GSD [m]*	*Spectral Range [nm]*	*SSI [nm]*	*Agency or Company*	*Planned year*
Tanager	30	400 - 2500	5	Planet	2023
OCI	1000	340 - 890 and 7 bands in SWIR	2,5 to 5	NASA	2024
HyspIRI	60	380 - 2510	10	NASA, USGS and NOAA	Not specified
FLORIS	300	500 - 780	0,1 to 0,5	ESA	2024
CHIME	20 - 30	400 - 2500	10	ESA	Not specified
SHALOM	<10	400 - 2500	10	ASI - ISA	2025

Table 4.14: Summary of Planned Spaceborne Hyperspectral Imagers.

4.2. Chlorophyll-a parameter for Lake Water quality assessment

When nutrients like nitrogen and phosphorus are introduced into a lake, they can cause an overgrowth of algae and other aquatic plants. This overgrowth is known as an algal bloom, and it can have a significant impact on water quality. As these algae and plants die and decompose, they consume oxygen in the water, leading to a process known as eutrophication which can produce changes in the species composition of the lakes (biodiversity) [72].

According to [73] the plant pigments of algae and cyanobacteria consist of chlorophylls and carotenoids. Specifically, chlorophyll-a is the most dominant pigment and considering it alone is possible to estimate the algal biomass.

In conclusion, a high concentration of chlorophyll-a can indicate an overgrowth of algae and other plants, which may be a sign of poor water quality. For this reason, chlorophyll-a is a valuable tool for assessing the health of aquatic ecosystems. By monitoring their levels, it is feasible to detect changes in water quality and take decisions to prevent or mitigate the harmful effects of algal blooms and eutrophication.

4.3. SIMILE project: monitoring of lake water quality parameters in Insubric Lakes

The "Informative System for the Integrated Monitoring of Insubric Lakes and their Ecosystems" (SIMILE) [1] is a project which objective is to monitor the water quality of the insubric lakes which are in the cross-basin between Italy and Switzerland improving their coordinated management and promoting the stakeholder participation [14]. The Prealp lakes which are part of the SIMILE project are an important source of fresh water for their region: Lake Maggiore, Lake Como and Lake Lugano (During the last time also included in the SIMILE project Lake Varese but it is not part of this study). Specifically, in the case of Italy, these lakes together with other Prealp lakes represent around 80% of the freshwater of the whole country [62].

For complying with their objectives the project employs in-situ measurements but also remote sensing techniques. Among the remote sensing monitoring, the project computed three main indicators for the water quality assessment of the lakes: Total Suspended Matter, Lake Water Surface Temperature and Chlorophyll-a [70]. In particular, the latter is the one which is also of relevance to our study.

In order to generate Chl-a maps, in [70] were employed images from the ESA's Sentinel-3 A/B OLCI instrument, which have a daily revisiting time and a spatial resolution of 300 m. The authors used the Case 2 Regional Coast Colour (C2RCC) processing technique, described in [13], to perform radiometric and atmospheric corrections and calculate Chl-a concentrations using a Neural Network model. Furthermore, they utilized the C2RCC neural net flags to identify water spectra anomalies and potential cloud cover, with flagged pixels being excluded. Lastly, the authors applied the 3 σ rule to remove any outliers. The result of the previously described process is a set of 389 chlorophyll-a maps whose initial acquisition dates from 15/01/2019 and the last one is from 05/11/2022 [7].

[1] https://progetti.interreg-italiasvizzera.eu/it/b/78/sistemainformativoperilmonitoraggiointegratodeilaghiinsubriciedeiloroe

4.4. Classification of Hyperspectral imagery

Although the aim of the current work is not focused on the classification of chlorophyll-a ranges from hyperspectral data but instead it is desired to perform regression, by predicting their continuous values, this section introduces some of the state-of-the-art techniques used to perform the classification of Hyperspectral data. Then, in most cases, their adaptation to regression problems is quite simple because it just requires the output to be a continuous variable instead of a categorical one.

4.4.1. Traditional Approaches

In this sub-section are described some of the traditional approaches that could be applied for classifying hyperspectral imagery. They are categorized as traditional because they rely in some statistical approaches or in Machine Learning methods not based on data driven techniques.

Spectral Classification

According to [6], the more straightforward way to perform HSI classification is by considering each pixel as a spectral signature that then is used to fit a statistical model. Among the approaches described in the previously mentioned reference are:

- **Unmixing:** Considering that usually, a pixel corresponds to a surface made of several materials, it will produce a spectral mixture. For this reason, the method involves finding the individual materials present in a mixed spectrum by computing their abundance maps. These abundance maps correspond to the proportional contributions of each material to each pixel. [6] also explains that to obtain the abundance maps it is necessary to solve a linear system by means of linear algebra and numerical methods. Additionally, [6] details that it is possible to solve this kind of problem by using for example clustering methods and finding unknown end-members.
- **Dimensionality Reduction:** [6] describes that the neighbour intensities are highly correlated and so the spectral signature is quite redundant. For this reason, it is quite useful many times to reduce the spectral dimension of the inputs and then make use of any classification techniques to perform the desired task, among which the paper mentions: Decision Trees, Random Forests and Support Vector Machines. In order to reduce the spectral dimension, [6] proposes the traditional PCA tech-

nique, band selection or the computation of indices that summarize the physical priors of the band response (i.e.: Normalized difference vegetation index (NDVI), Normalized difference water index (NDWI), etc.).

Spatial-Spectral Classification

In [6] it is also highlighted the limitations of using a spectral-only approach for hyperspectral image classification as it ignores the spatial structure of the image because neighbouring pixels often share structural relationships, and accounting for these dependencies can improve the model's robustness and efficiency. The paper identifies three main approaches to incorporating spatial information in the classification process, which differ based on when the spatial aspect is considered.

- **Spatial Regularization:** It consists in first classifying the individual spectra and then regularising the result of the classification with a spatially-structured model. For the regularization, [6] introduces that it is possible to use the Markov Random Fields (MRF) or Conditional Random Fields (CRF) models.
- **Pre-segmentation:** This second approach performs an unsupervised spatial regularization. This regularization involves segmenting the hyperspectral image first, and then aggregating in the spectrum dimension the features of each segmented region in order to ensure local consistency.
- **Joint-learning:** This last approach explained in [6] consists in learn the spatial and spectral dimensions at the same time by means of kernels. The before-mentioned reference emphasizes that a possible way to implement this approach is by designing a spatial-spectral kernel for SVMs.

4.4.2. Deep Learning Approaches

In this sub-section are introduced some of the data driven approaches used for the classification of hyperspectral data.

Spectral classification

In [6] are also introduced the Deep Learning approaches based on Spectral Classification. In this case, the difference with respect to the traditional cases relies on the fact that instead of using a standard classifier, a deep fully-connected network is utilized. In the cited reference, it is explained that these approaches could alternatively make use of Recurrent Neural Networks (RNN) because of the intrinsic characteristics of the Hyperspectral Data:

it is sequential data.
RNNs are well suited for the modelling of sequential data because they are helpful to model both long and short-range dependencies in the spectral dimension.
Finally, also [6] describes that it is feasible to follow unsupervised approaches by performing dimensionality reduction by means of autoencoders and then followed by classification with a simple perceptron.

Spatial-spectral classification

- **1D or 2D Convolutional Neural Networks (CNNs):** Typically, in computer vision, the CNNs are designed in a way that the first part is a set of convolutional filters which extract the features and the final one is a fully connected layer that performs the classification task [6]. The issue with HSI data is their high dimensionality implies the need for several filters because they depend on the number of channels of the inputs. For the previous reason, according to [6] different works proposed to reduce the dimensionality of the inputs in order to allow their treatment in the CNN networks. The issue with these methods is that they force the inputs to be reduced and in some cases, they rely on unsupervised approaches to perform that reduction which could imply an uncontrolled loss of the Hyperspectral imaging properties [6].
- **2D + 1D CNNs:** Considering the importance of the spectral dimension, the document [6] details that there were developed approaches in which the CNN architectures alternate spatial and spectral convolutions, processing the complete hypercube of data but reducing in an alternate manner the size of the feature maps. Alternatively, an unsupervised approach also was proposed to handle the hypercubes with a 2D CNN with a residual learning paradigm that is able to obtain an efficient low-dimensional representation of the hyperspectral pixels and their neighbours [6].
- **3D CNNs:** The document [6] finally introduces as a promising approach the use of 3D CNNs, which work on all three dimensions simultaneously producing 3D feature cubes that are better suited for pattern recognition in a volume. Compared to spectral or 2D+1D CNNs, these architectures combine both pattern recognition strategies into one filter, requiring fewer parameters and layers. Studies have shown that 3D CNNs perform better than their 2D counterparts for hyperspectral image classification [6].

5 | Dataset description

The data available for achieving the objectives of the study consist of the following two main components:

- **Hyperspectral imagery:** A set of PRISMA acquisitions were selected and extracted from the official platform of this satellite mission. The initial set consisted of a total of 27 images and after refinement there were filtered a set of 12 images.
- **Chlorophyll-a:** The second dataset available was the introduced in the section 4.3. It consists of a total of 389 chlorophyll-a maps from where there were chosen the ones that correspond to the selected PRISMA images.

5.1. Initial selection

A first search was performed on the official platform of PRISMA mission to obtain images in the area that covers the lakes of the SIMILE project: Maggiore, Lugano and Como. Then, there were also considered the dates on which were available Chlorophyll-a maps. According to these criteria, a total of 27 PRISMA images were pre-selected: for each pre-selected PRISMA acquisition, a corresponding chlorophyll-a map was assigned according to the dates.

Acq. ID	*PRS Date*	*L. Como*	*L. Maggiore*	*L. Lugano*	*Pix. >90 pct Chl-a*	*Chl-a date*	*Spatial ext.*	*Clouds*	*Accept. Glint*
1	24/04/20	**YES**	**NO**	**YES**	0	23/04/20	**OK**	**OK**	**OK**
2	24/04/20	**YES**	**NO**	**NO**	0	23/04/20	**KO**	**OK**	**OK**
3	25/04/20	**NO**	**YES**	**NO**	0	23/04/20	**OK**	**KO**	**OK**
4	25/04/20	**NO**	**YES**	**NO**	0	23/04/20	**OK**	**OK**	**OK**
5	24/05/20	**NO**	**YES**	**NO**	0	24/05/20	**OK**	**OK**	**KO**
6	03/07/20	**NO**	**YES**	**NO**	0	05/07/20	**OK**	**OK**	**OK**
7	19/08/20	**NO**	**YES**	**NO**	0	20/08/20	**OK**	**KO**	**OK**
8	17/09/20	**NO**	**YES**	**NO**	3	17/09/20	**OK**	**KO**	**OK**
9	28/06/21	**YES**	**NO**	**NO**	0	25/06/21	**KO**	**KO**	**OK**
10	09/07/21	**YES**	**NO**	**YES**	0	09/07/21	**KO**	**OK**	**OK**
11	25/08/21	**YES**	**NO**	**NO**	0	26/08/21	**KO**	**KO**	**OK**
12	25/08/21	**YES**	**NO**	**NO**	0	26/08/21	**KO**	**KO**	**OK**
13	31/08/21	**YES**	**NO**	**NO**	0	31/08/21	**OK**	**KO**	**OK**
14	31/08/21	**YES**	**NO**	**NO**	0	31/08/21	**OK**	**KO**	**OK**

15	29/09/21	YES	NO	NO	0	30/09/21	KO	KO	OK
16	29/09/21	YES	NO	YES	0	30/09/21	KO	KO	KO
17	16/10/21	NO	YES	NO	336	16/10/21	KO	OK	OK
18	22/10/21	YES	NO	NO	112	22/10/21	OK	OK	OK
19	22/10/21	YES	NO	YES	112	22/10/21	OK	OK	OK
20	08/11/21	YES	NO	NO	964	08/11/21	KO	KO	OK
21	26/11/21	NO	NO	YES	404	24/11/21	OK	KO	OK
22	03/02/22	NO	YES	NO	13	03/02/22	KO	OK	OK
23	09/02/22	NO	YES	NO	426	09/02/22	OK	OK	OK
24	27/03/22	YES	NO	NO	56	25/03/22	OK	OK	OK
25	27/07/22	YES	NO	YES	0	27/07/22	KO	KO	OK
26	31/08/22	YES	NO	YES	0	29/08/22	KO	KO	OK
27	06/09/22	YES	NO	YES	0	08/09/22	OK	KO	OK

Table 5.1: Initial Selection of PRISMA Acquisitions.

5.2. Final dataset

Starting from the initially selected set of images, a second filtering was performed and it was possible to define a set of 12 images. These images were chosen as the final dataset because they passed a set of controls related to the cloud coverage, the spatial extension of the Chlorophyll-a map associated with each image and the acceptability of the glint disturbance. Although ideally, the area of coverage of each chlorophyll-a map should comprise all the lakes of the SIMILE project, on some dates, due to the removal of the anomalous pixels explained in section 4.3 and the further pre-processing related to the removal of anomalous pixels but by means of spectral ratios that will be introduced later on sub-section 6.2.3, the extension of the PRISMA acquisitions and chlorophyll-a maps considered was reduced.

The final selection is detailed in the table 5.2 and then are also analyzed the chlorophyll-a maps associated with each acquisition.

Acq. ID	***PRS Date***	***L. Como***	***L. Maggiore***	***L. Lugano***	***Pix. >90 pct Chl-a***	***Chl-a date***	***Spatial ext.***	***Clouds***	***Accept. Glint***
1	24/04/20	YES	NO	YES	0	23/04/20	OK	OK	OK
2	24/04/20	YES	NO	NO	0	23/04/20	KO	OK	OK
4	25/04/20	NO	YES	NO	0	23/04/20	OK	OK	OK
6	03/07/20	NO	YES	NO	0	05/07/20	OK	OK	OK
10	09/07/21	YES	NO	YES	0	09/07/21	KO	OK	OK
13	31/08/21	YES	NO	NO	0	31/08/21	OK	KO	OK
17	16/10/21	NO	YES	NO	336	16/10/21	KO	OK	OK
18	22/10/21	YES	NO	NO	112	22/10/21	OK	OK	OK
19	22/10/21	YES	NO	YES	112	22/10/21	OK	OK	OK
21	26/11/21	NO	NO	YES	404	24/11/21	OK	KO	OK
23	09/02/22	NO	YES	NO	426	09/02/22	OK	OK	OK

24	27/03/22	YES	NO	NO	56	25/03/22	OK	OK	OK

Table 5.2: Final Selection of PRISMA Acquisitions.

In order to provide a standardized reference of the season associated with each acquisition, the Water Framework Directive (WFD) seasons were taken into account, in the following sub-sections, while analyzing each case, the corresponding season from this framework is detailed. The seasons of the WFD are specified in the table 5.3.

Case ID 1

The date of the PRISMA acquisition is 24/04/2020 and its associated Chl-a map is from 23/04/2020, therefore it corresponds to the spring season according to the WFD seasons described in table 5.3. From figure 5.1 it is possible to appreciate that the Chl-a map covers almost the whole extension of Lake Como and the right side of Lake Lugano. Additionally, figure 5.2 presents the distribution of their values.
Regarding the main statistics it is possible to highlight the following ones:

- Min: 1,42 µg/L
- Max: 5,48 µg/L
- Mean: 3,40 µg/L
- Std. dev.: 0,57 µg/L

Case ID 2

Also, the PRISMA acquisition related to this case took place on 24/04/2020 and for that reason, it was associated with the Chl-a map from 23/04/2020, which, according to the table 5.3 also corresponds to the Spring Season of the WFD. Figure 5.3 shows that this

WFD Season	Initial date	Final date
Winter	01-Jan	20-March
Spring	01-Apr	15-May
Transit. spring-summer	15-May	15-Jun
Summer	01-Jul	31-Aug
Transit. summer-autumn	01-Sep	01-Oct
Autumn	01-Oct	30-Nov

Table 5.3: Water Framework Directive seasons' initial and final dates. [11]

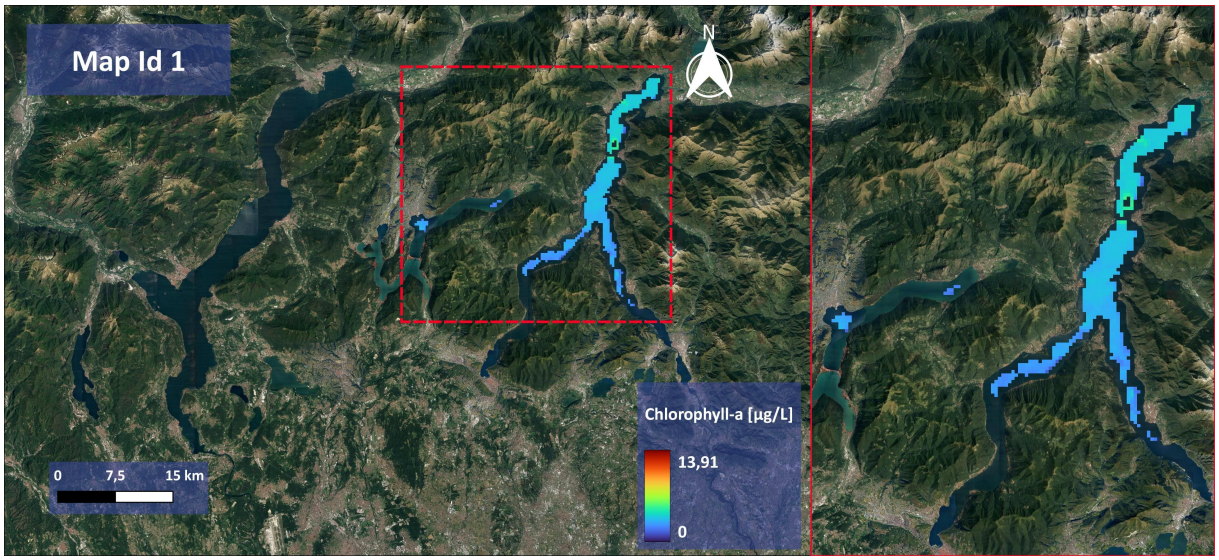

Figure 5.1: Chlorophyll-a map ID 1.

acquisition has poor spatial coverage because it is just present in some bottom areas of Lake Como and a small region of Lake Lugano, however, it was kept for the final dataset because the PRISMA acquisition does not have problems of cloud coverage. Figure 5.4 describes the distribution of the Chl-a values and their main statistics are the next:

- Min: 1,71 µg/L
- Max: 3,81 µg/L
- Mean: 2,79 µg/L
- Std. dev.: 0,67 µg/L

Case ID 4

In this case, the PRISMA acquisition was from 25/04/2020 and its associated Chl-a map is from 23/04/2020, therefore it corresponds with the spring season of the WFD detailed in table 5.3. Figure 5.5 displays the chlorophyll-a map corresponding to this acquisition and from there, it is observed that it covers most of the bottom part of Lake Maggiore. Figure 5.6 presents the histogram for this acquisition and their main statistics are the following ones:

- Min: 0,61 µg/L
- Max: 4,49 µg/L
- Mean: 3,07 µg/L

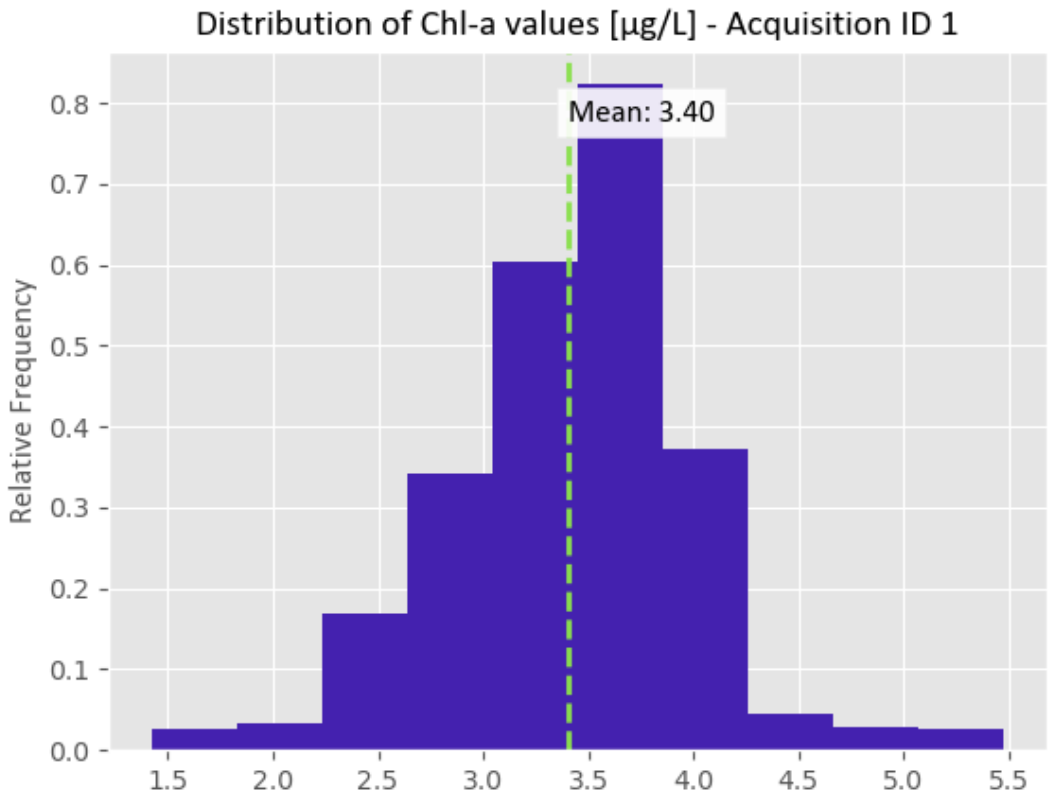

Figure 5.2: Distribution of Chl-a map ID 1.

- Std. dev.: 0,37 µg/L

Case ID 6

This case corresponds to a PRISMA acquisition from 03/07/2020 which has been associated with a Chl-a map from 05/07/2020. In this case, the period of acquisition corresponds to the Summer season from the WFD seasons described in table 5.3. By observing figure 5.7 it is noticeable that the chlorophyll-a map for this case covers most of Lake Maggiore and also from figure 5.8 it is possible to observe its data distribution where it is evident that in comparison with the previous cases, the data distribution is centred in higher values.
The main statistics for this case are the following ones:

- Min: 0,61 µg/L
- Max: 5,51 µg/L
- Mean: 4,18 µg/L

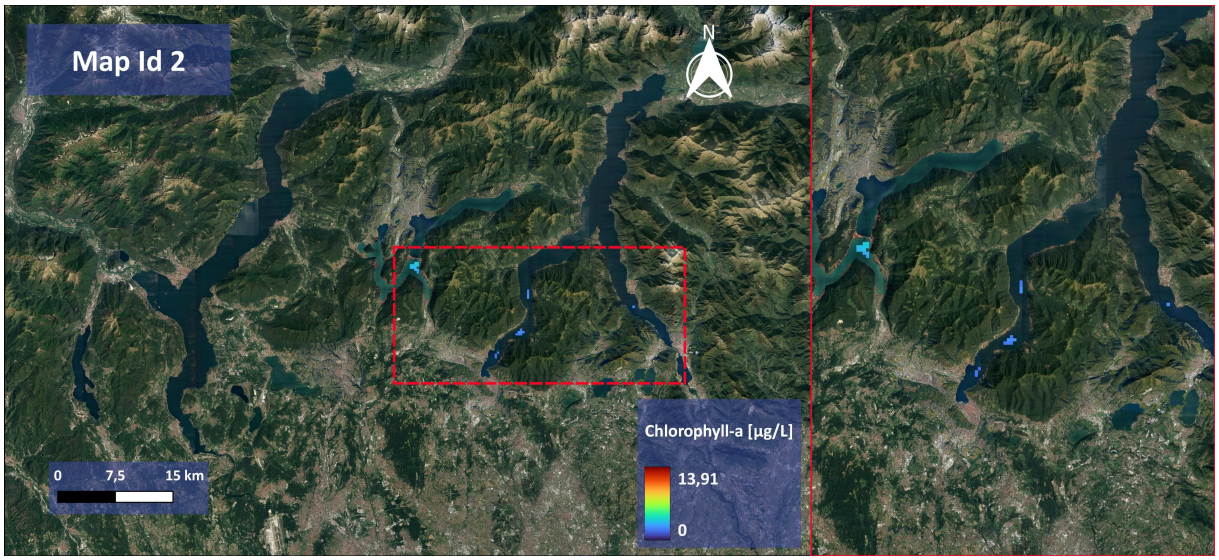

Figure 5.3: Chlorophyll-a map ID 2.

- Std. dev.: 0,46 µg/L

Case ID 10

In this case, both the PRISMA image and the Chl-a map date from 09/07/2021, associated with the Summer season of WFD seasons, detailed in table 5.3. The Chl-a map is displayed in figure 5.9 and its data distribution is presented in figure 5.10. Here are the main statistics for the Chl-a map distribution:

- Min: 1,38 µg/L
- Max: 3,62 µg/L
- Mean: 2,44 µg/L
- Std. dev.: 0,53 µg/L

Case ID 13

The PRISMA image and the Chl-a map date from 31/08/2021 for this case. Observing table 5.3 is possible to confirm that the acquisition date lies exactly at the end date of the Summer season from the WFD seasons. Figure 5.11 presents the Chl-a map and its associated histogram is in figure 5.12. The main statistics for the Chl-a map distribution are the following:

- Min: 1,44 µg/L

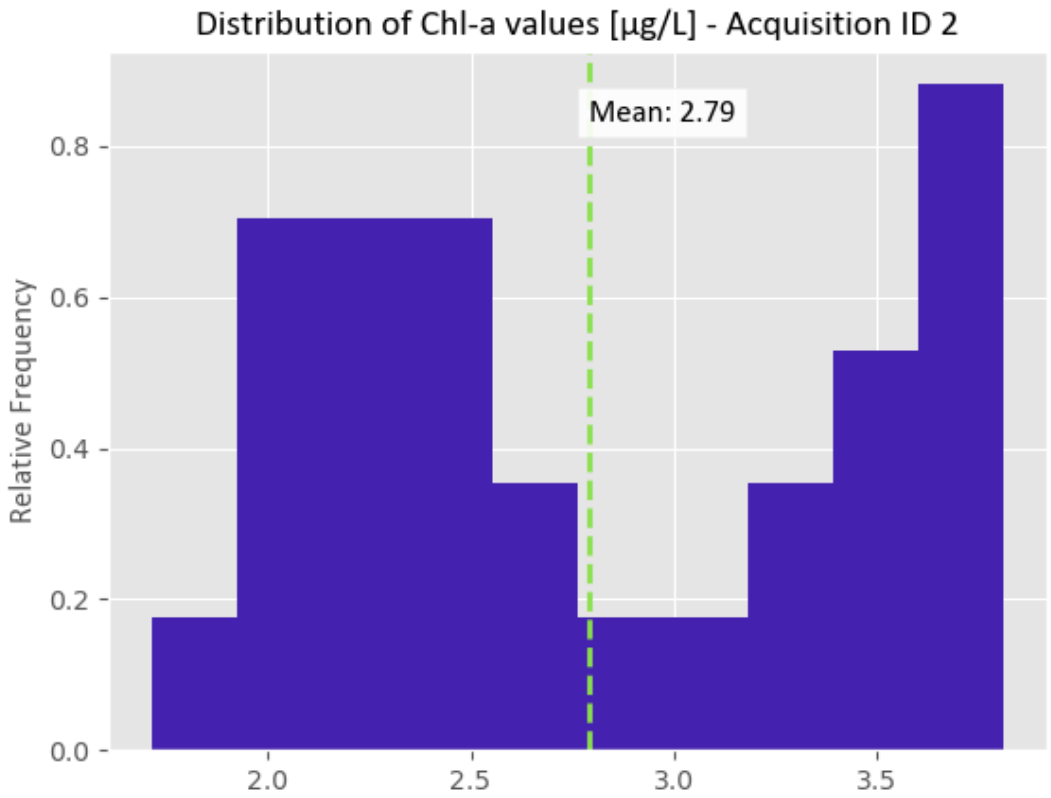

Figure 5.4: Distribution of Chl-a map ID 2.

- Max: 5,71 µg/L
- Mean: 3,64 µg/L
- Std. dev.: 0,43 µg/L

Case ID 17

In this case, the corresponding PRISMA acquisition and the Chl-a map are from 16/10/2021 which, according to the table 5.3 is associated with the Autumn season from the WFD seasons. Figure 5.13 presents the Chl-a map from where it is noticeable that the spatial coverage is quite sparse, however, it is important to observe that there are certain areas in which the pixels take high values. Its associated histogram is in figure 5.14 and from the principal statistics which are presented in the following, can be appreciated that the maximum value and the standard deviation, are higher with respect to the previously presented cases:

- Min: 0,21 µg/L

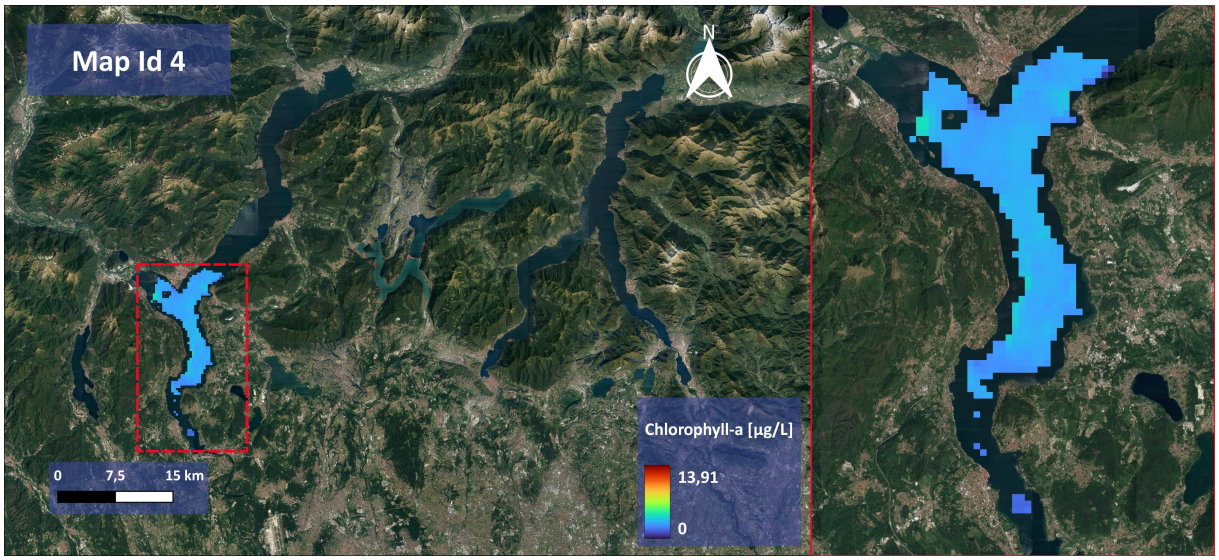

Figure 5.5: Chlorophyll-a map ID 4.

- Max: **9,56 µg/L**
- Mean: 5,16 µg/L
- Std. dev.: **1,41 µg/L**

Case ID 18

The PRISMA image and the Chl-a map associated with this case are from 22/10/2021. Also in this case the acquisition date is related to the Autumn season from the WFD seasons described in table 5.3. As it can be observed in figure 5.15 the Chl-a map covers most of the top region of Lake Como. Its corresponding histogram is in figure 5.16 and the main statistics for the Chl-a map are:

- Min: 0,49 µg/L
- Max: 7,33 µg/L
- Mean: 4,88 µg/L
- Std. dev.: **1,20 µg/L**

Case ID 19

Also in this case both the PRISMA image and the Chl-a map associated are from 22/10/2021, corresponding to the Autumn season of the WFD seasons. Figure 5.17 presents the Chl-a map which covers the two bottom branches of Lake Como and the

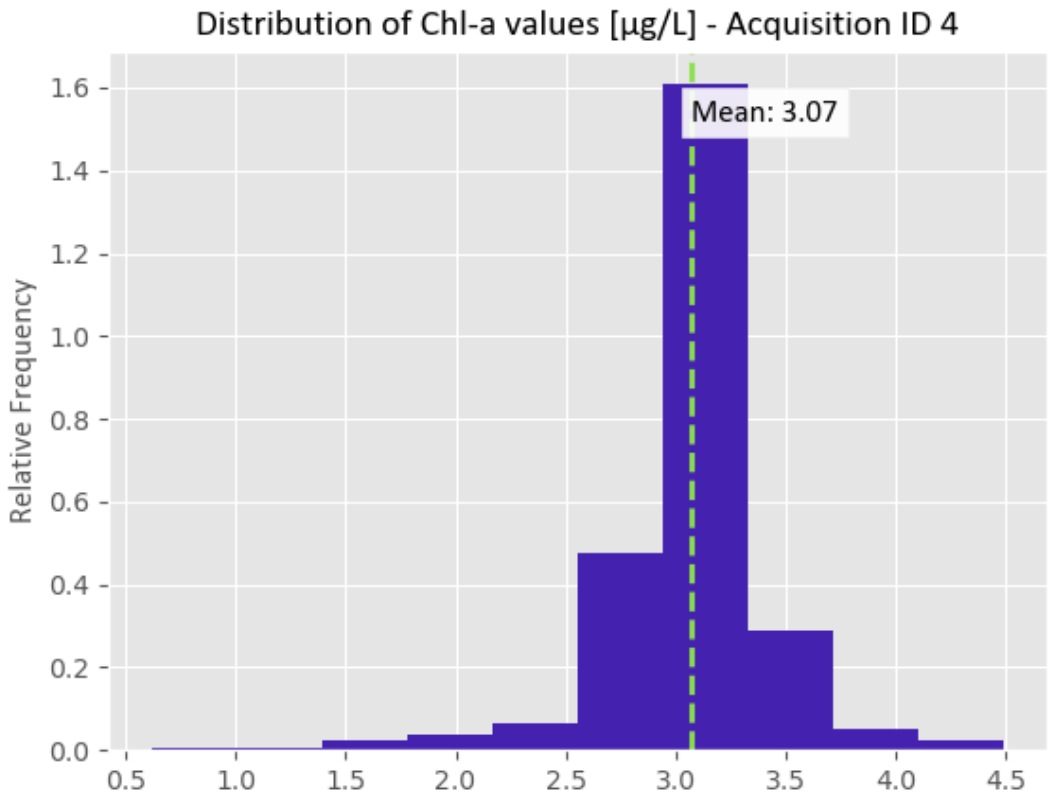

Figure 5.6: Distribution of Chl-a map ID 4.

right side of Lake Lugano. The data distribution of the Chl-a map is also presented in figure 5.18 and the statistics for the Chl-a map of this case are:

- Min: 0,32 µg/L
- Max: 7,60 µg/L
- Mean: 4,49 µg/L
- Std. dev.: **1,25 µg/L**

Case ID 21

The PRISMA acquisition associated with this case is from 26/11/2021 but the Chlorophyll-a map is from 24/11/2021. Again, in this case, the acquisition date lies in the autumn season according to table 5.3. The spatial distribution of the Chl-a map is quite sparse, it covers most of the top edge of Lake Lugano and some small regions of lakes Como and Maggiore, this can be noticed in the figure 5.19. Also figure 5.20 presents the data

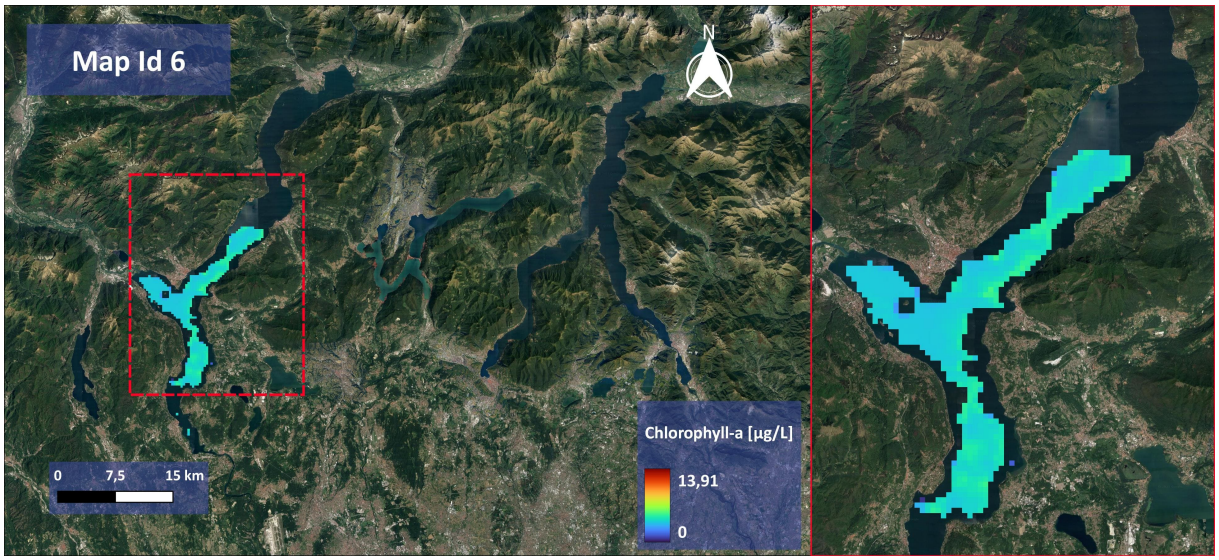

Figure 5.7: Chlorophyll-a map ID 6.

distribution of the Chl-a map and from there can be understood that this case has higher values of chlorophyll-a in comparison with the previously mentioned cases, also this can be confirmed from the main statistics of this case which are introduced next. This case also presents the highest standard deviation among all the cases considered in the study:

- Min: 0,01 µg/L
- Max: **9,86 µg/L**
- Mean: 4,85 µg/L
- Std. dev.: **2,71 µg/L**

Case ID 23

In this case, both the PRISMA acquisition and the Chl-a map are from 09/02/2022. Observing table 5.3 is possible to determine that the acquisition date for this case corresponds to the winter season. From figure 5.21 can be appreciated that the Chl-a map covers most of the bottom part of Lake Maggiore. Also figure 5.22 displays the data distribution of the Chl-a map it allows us to comprehend that this case also presents high values of the parameter. The principal statistics that are useful to summarize the values of the Chl-a map for this case are presented in the following:

- Min: 0,88 µg/L
- Max: 8,92 µg/L

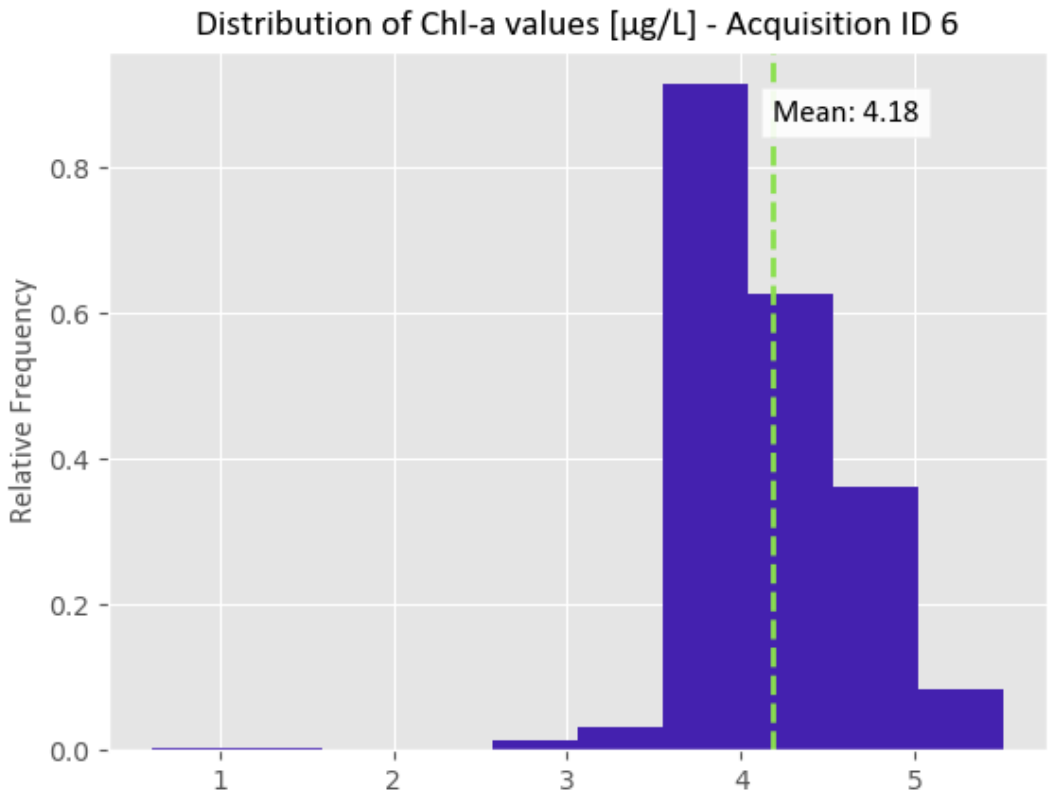

Figure 5.8: Distribution of Chl-a map ID 6.

- Mean: 5,21 µg/L
- Std. dev.: 0,85 µg/L

Case ID 24

The PRISMA image associated with this case is from 27/03/2022 and the Chl-a map is from 25/03/2022, which, accordingly to the table 5.3 is closest to the Spring season. In this case, as it can be verified from figure 5.23, the spatial coverage is on the bottom part of the Lake Como. Additionally, figure 5.24 presents the distribution of the Chl-a values and from there is possible to notice that this case also presents high values of the Chl-a indicator, considering the following main statistics:

- Min: 1,93 µg/L
- Max: 7,06 µg/L
- Mean: 5,21 µg/L

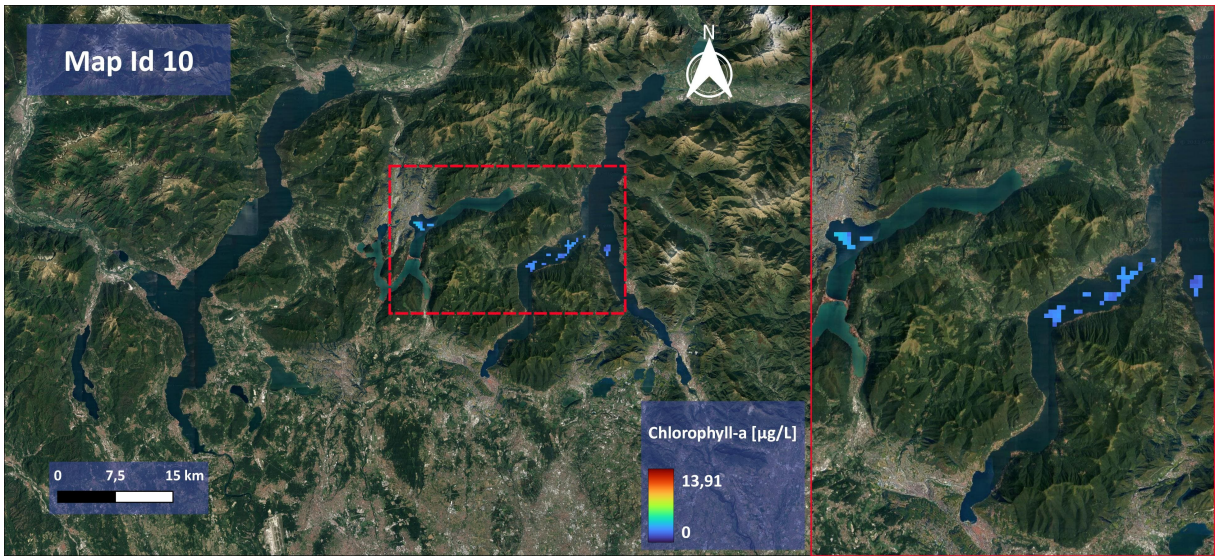

Figure 5.9: Chlorophyll-a map ID 10.

- Std. dev.: 1,02 µg/L

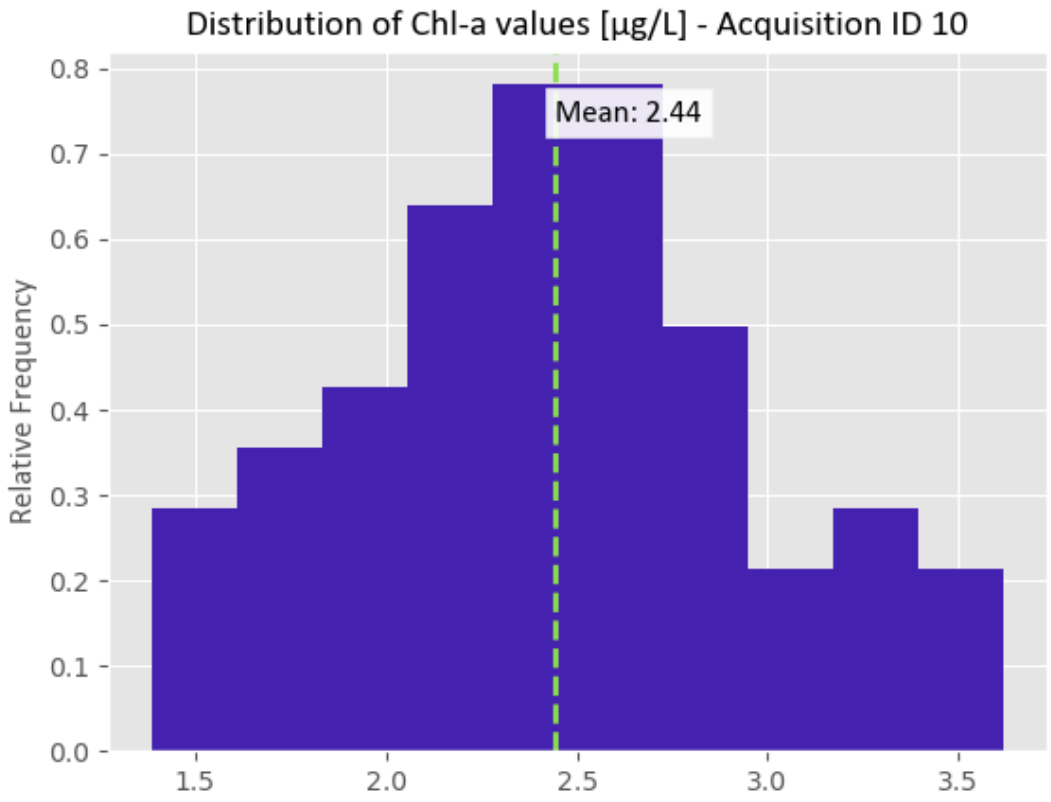

Figure 5.10: Distribution of Chl-a map ID 10.

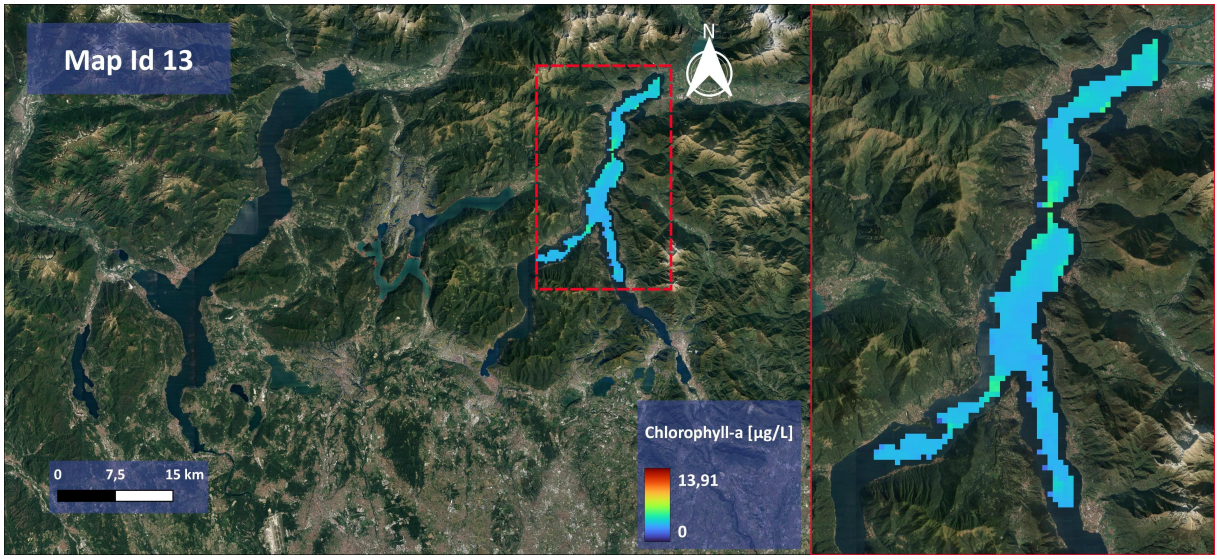

Figure 5.11: Chlorophyll-a map ID 13.

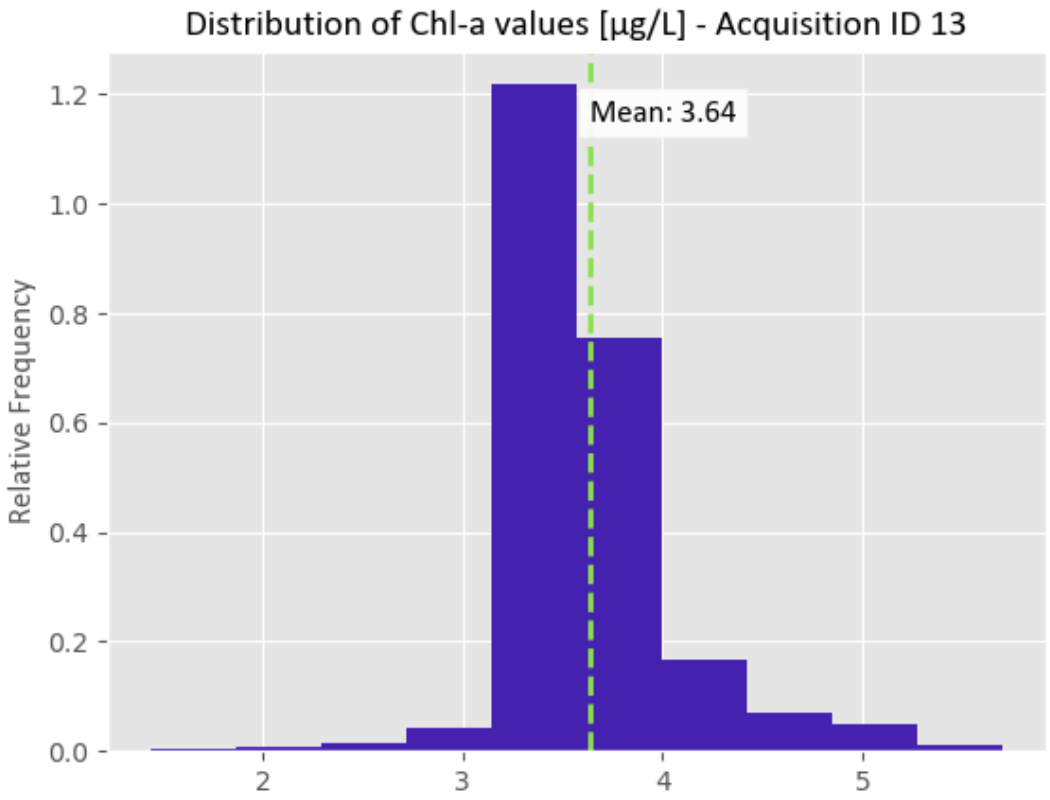

Figure 5.12: Distribution of Chl-a map ID 13.

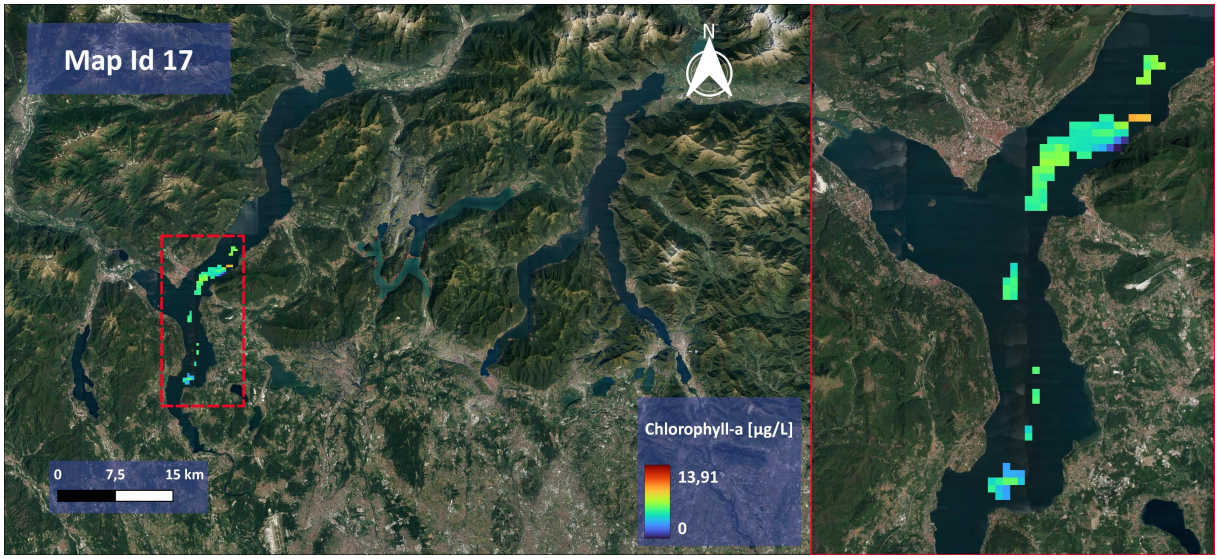

Figure 5.13: Chlorophyll-a map ID 17.

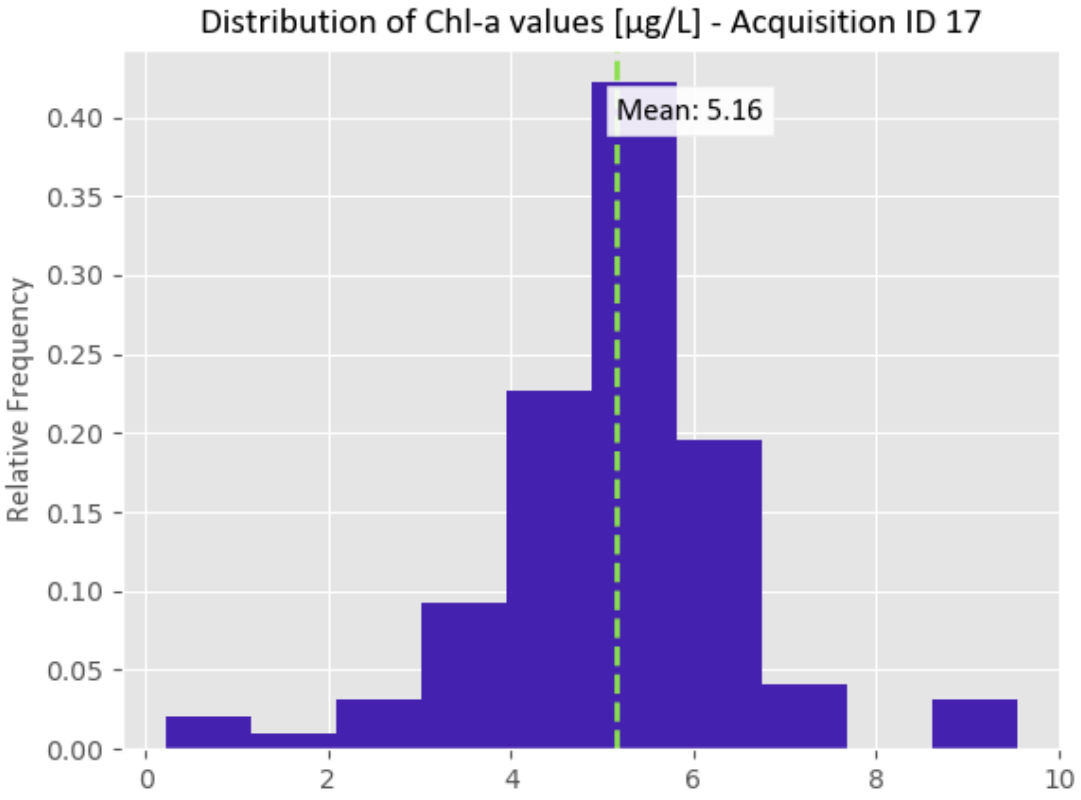

Figure 5.14: Distribution of Chl-a map ID 17.

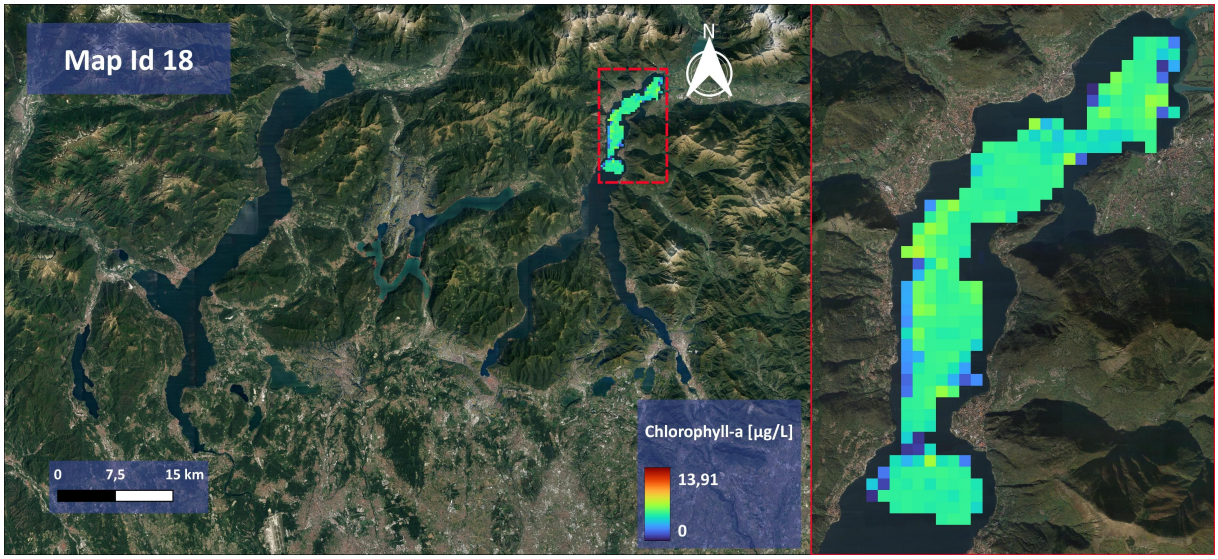

Figure 5.15: Chlorophyll-a map ID 18.

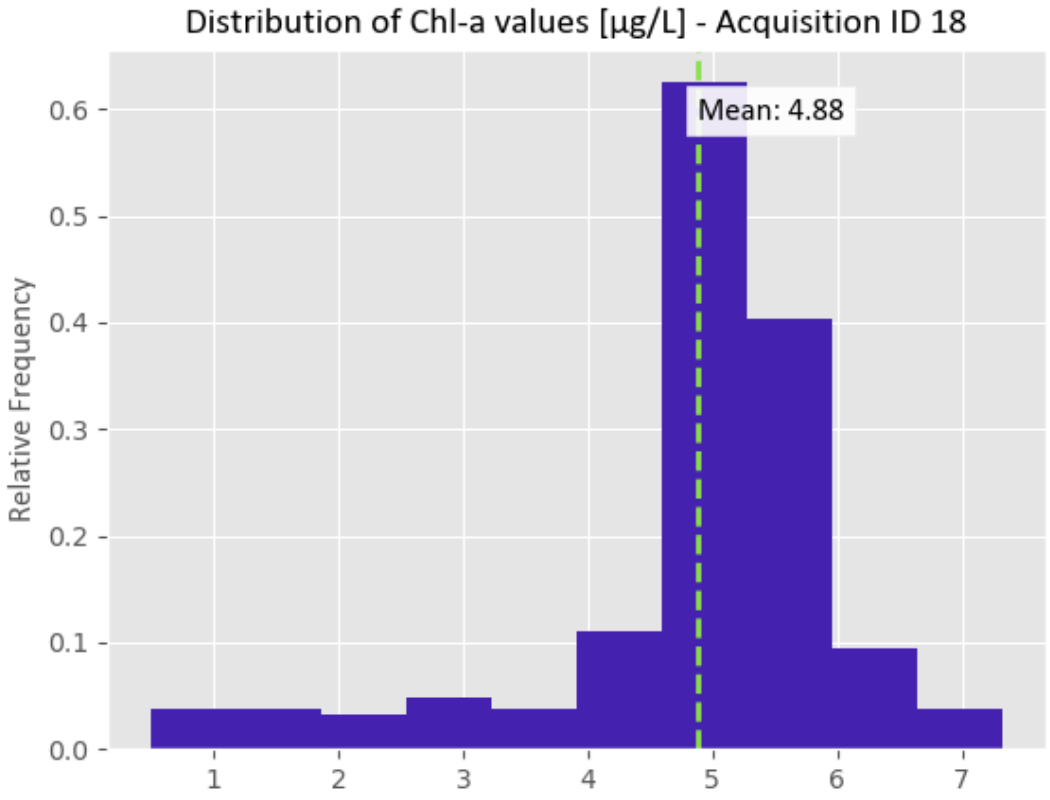

Figure 5.16: Distribution of Chl-a map ID 18.

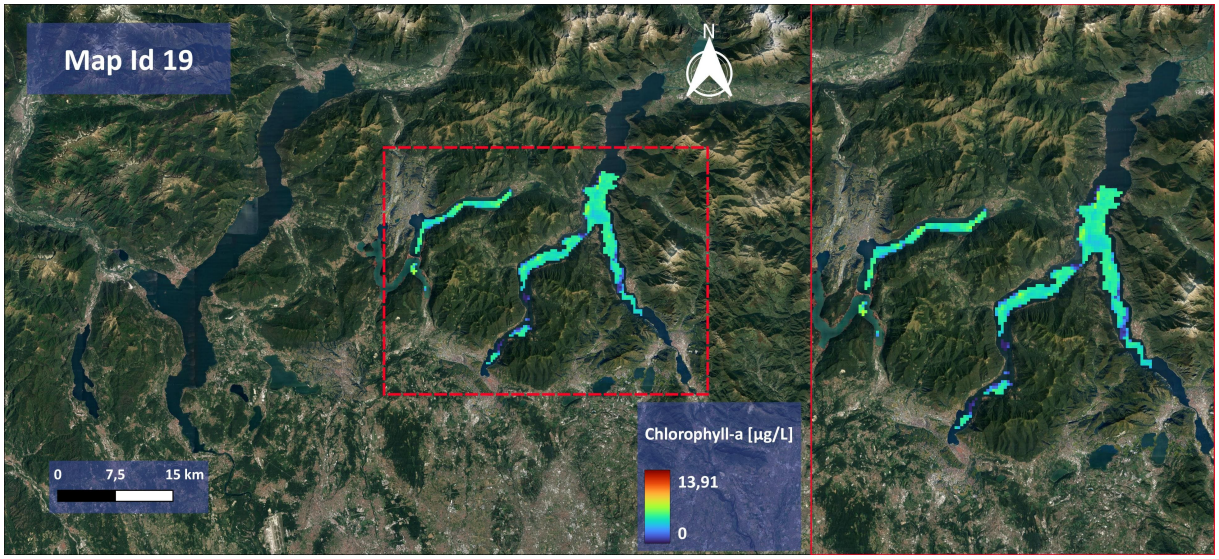

Figure 5.17: Chlorophyll-a map ID 19.

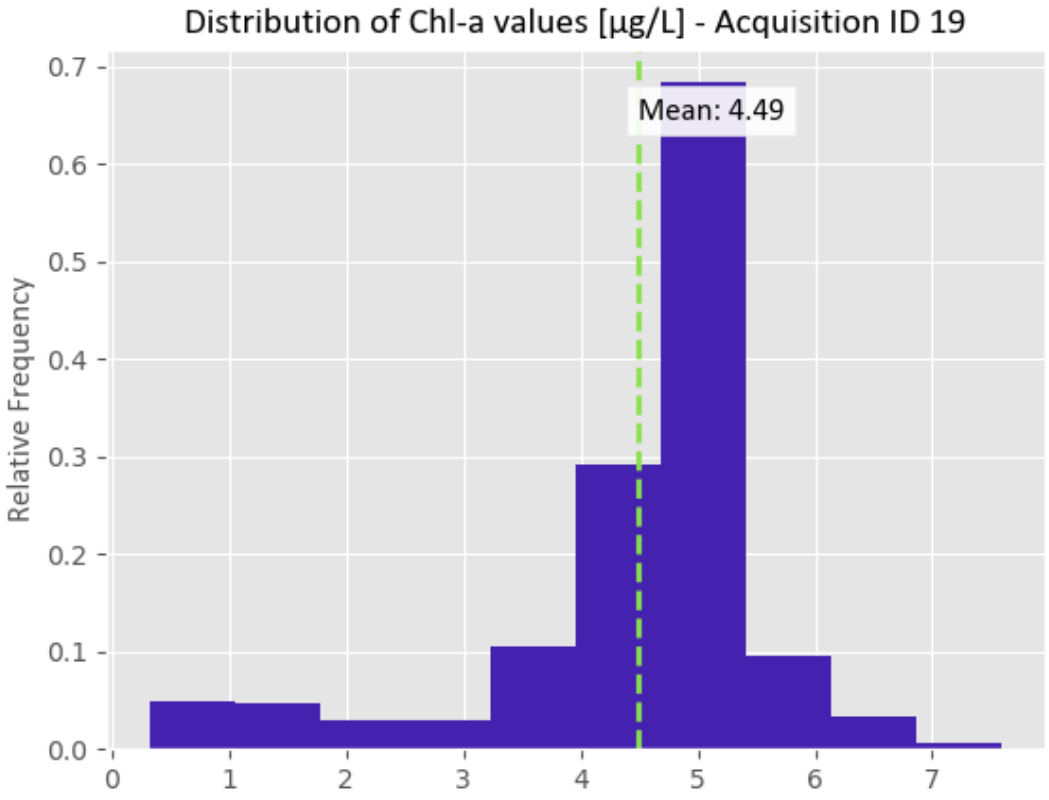

Figure 5.18: Distribution of Chl-a map ID 19.

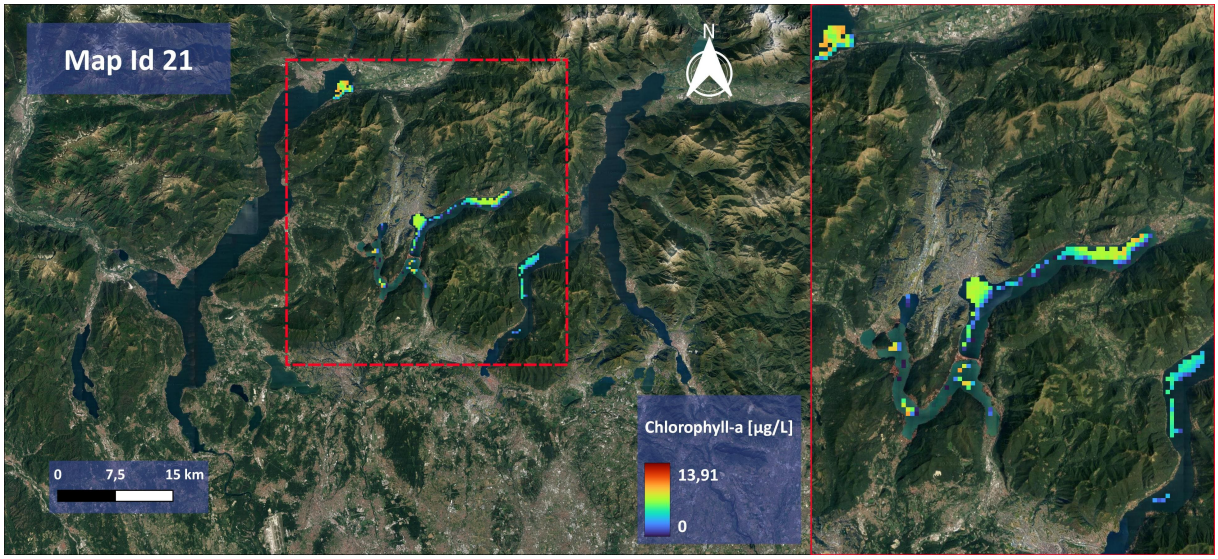

Figure 5.19: Chlorophyll-a map ID 21.

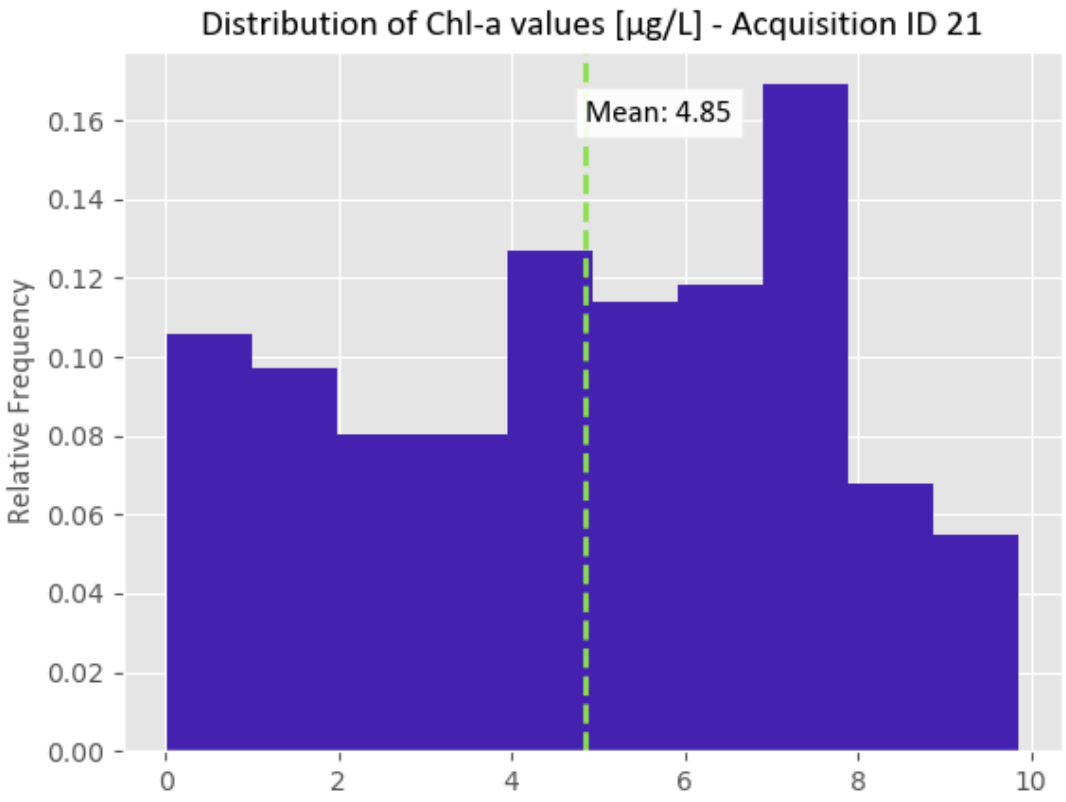

Figure 5.20: Distribution of Chl-a map ID 21.

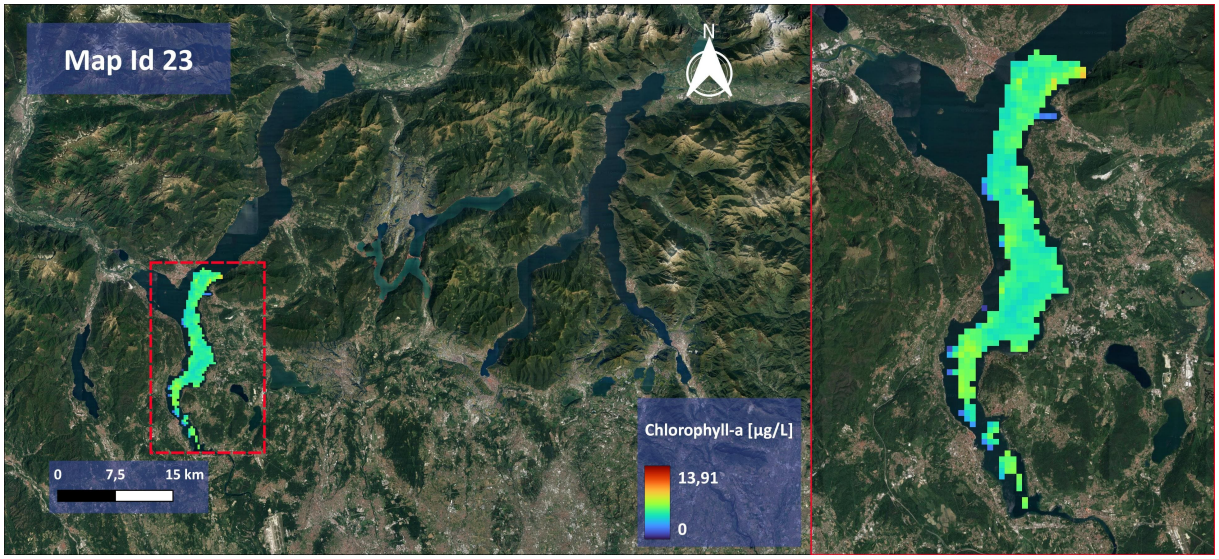

Figure 5.21: Chlorophyll-a map ID 23.

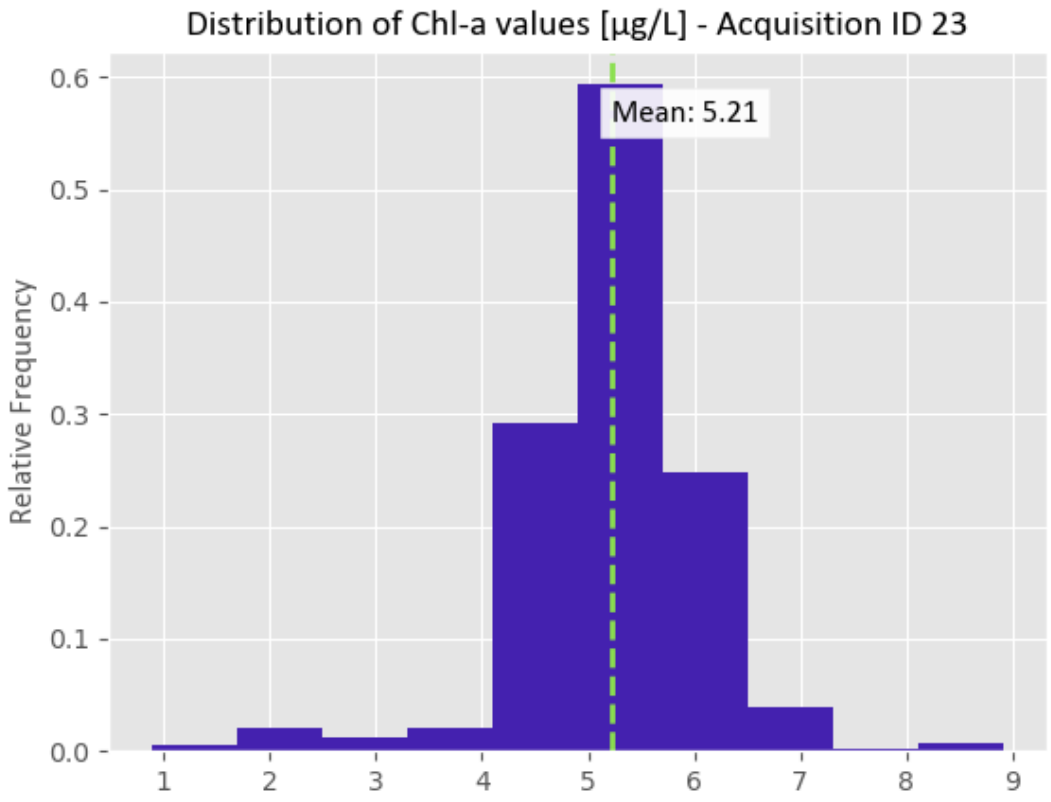

Figure 5.22: Distribution of Chl-a map ID 23.

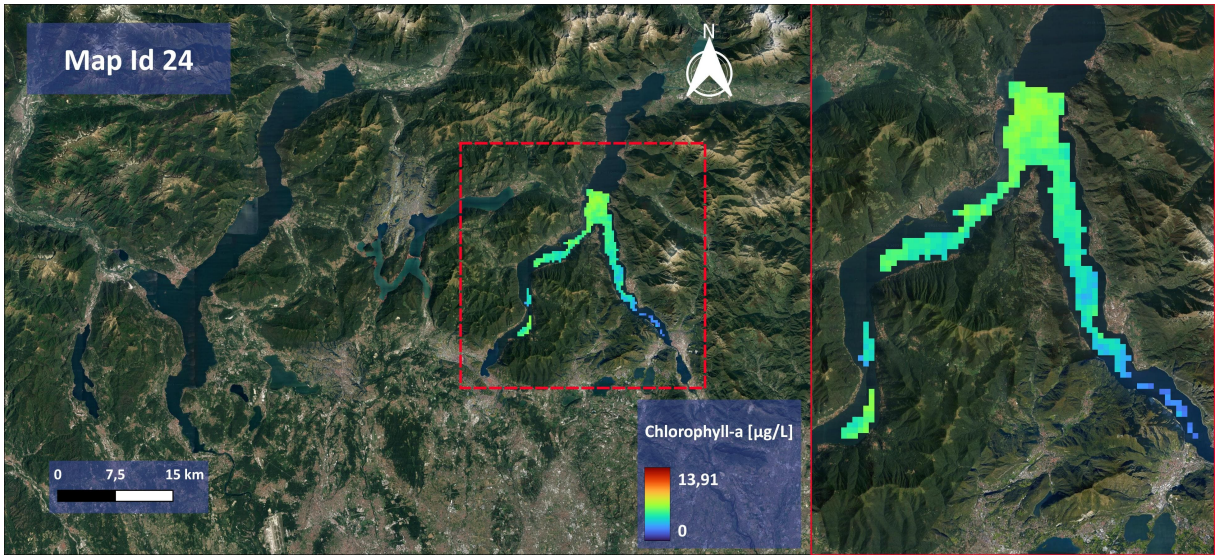

Figure 5.23: Chlorophyll-a map ID 24.

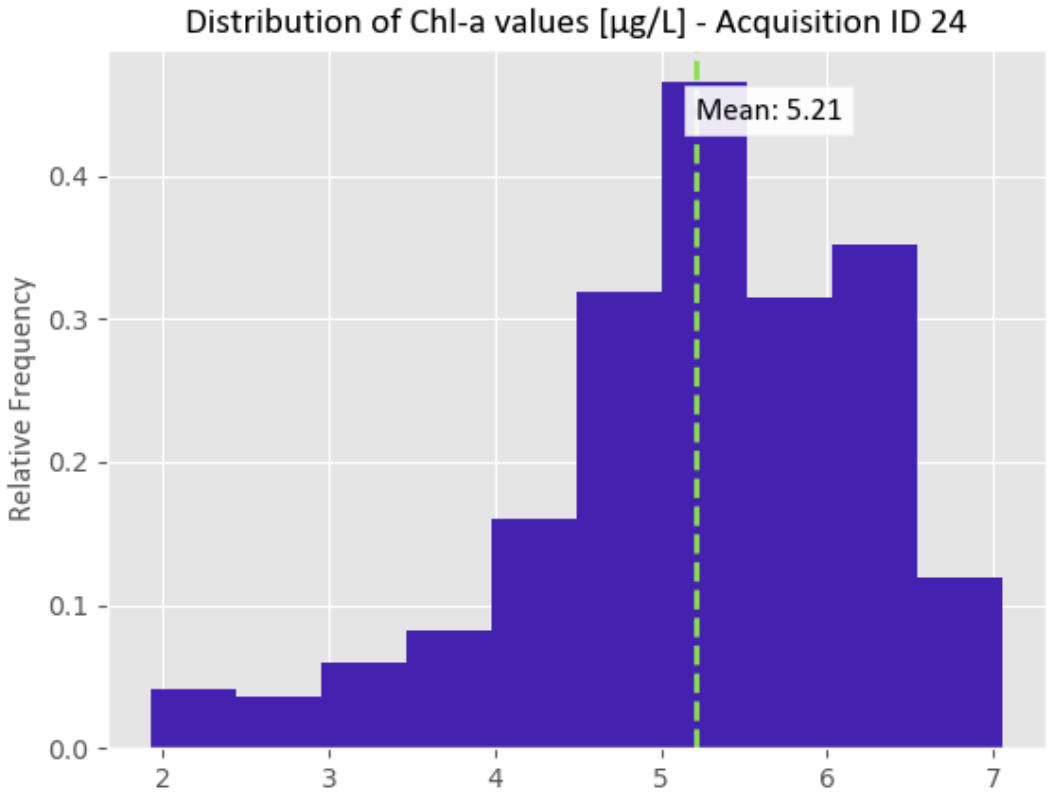

Figure 5.24: Distribution of Chl-a map ID 24.

6 | Experimental development

In order to predict the concentration of chlorophyll-a starting from hyperspectral imagery, different experiments were carried out in this study. This chapter will describe all the relevant aspects to introduce the reader to the experimental development that was done. Additionally, the results of each experiment are provided and analyzed to arrive at the definition of the best model.

6.1. Acquisition of L2D PRISMA imagery from ASI official portal

In order to acquire the input PRISMA imagery, a submission request on the official web platform of ASI[1] was performed for each of the required input images. Among the different processing levels available for downloading the PRISMA images, it was selected the L2D level [3]. The decision of this selection was based on the fact that this level provides geocoded bottom-of-atmosphere information [3].
Also, it is worth mentioning that the document [44] carried out an analysis of different atmospheric correction approaches against the standard atmospheric correction procedure considered by default in the PRISMA processing chain and the final conclusion was that the atmospheric correction correspondent to the PRISMA L2D products was outperforming the other explored options.

6.2. Pre-processing of the inputs

The PRISMA images and the corresponding chlorophyll-a maps had to be manipulated to be properly fed into the different models. This section explains each of the steps taken into consideration for the proper use of the data. A graphical summary is provided in figure 6.1.

[1] http://prisma.asi.it/js-cat-client-prisma-src/

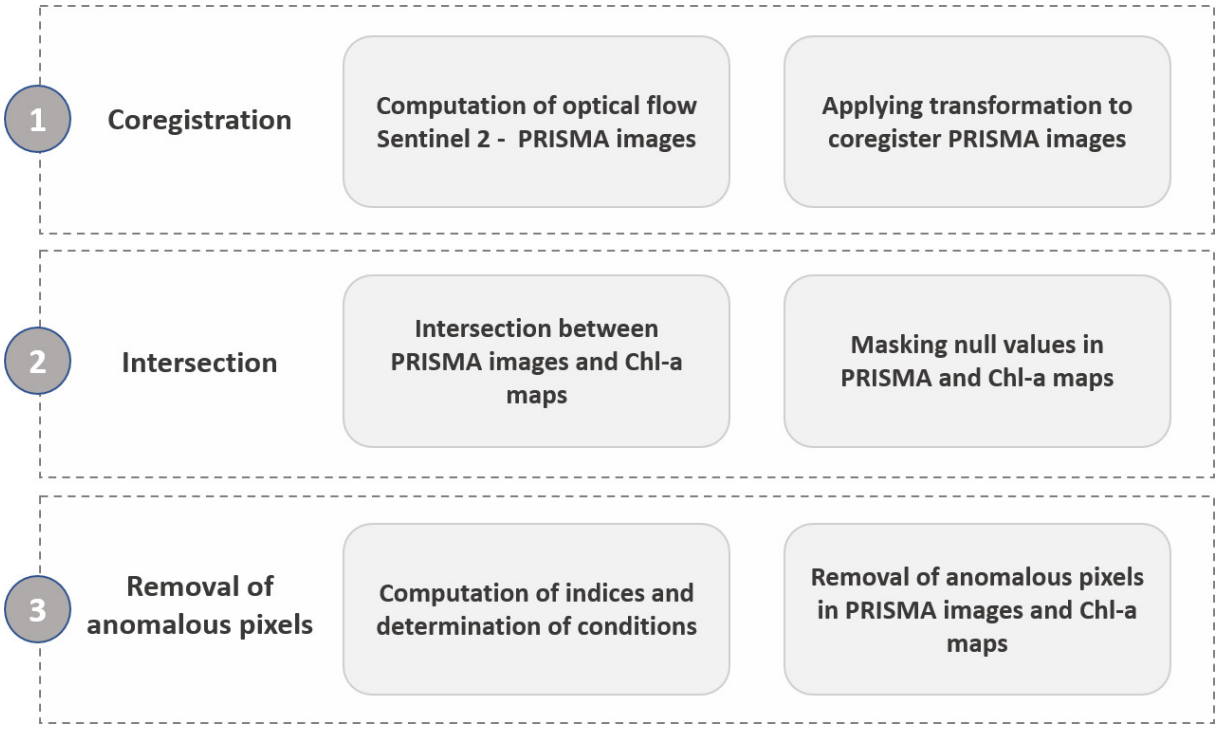

Figure 6.1: Pre-processing steps for the input data.

6.2.1. Coregistration of PRISMA images

When working with geospatial information, one of the critical steps is that the information to be processed is properly coregistered in order to correctly represent the reality of each pixel. In the case of our study, the coregistration of the chlorophyll-a maps was previously performed in the SIMILE project which was introduced in section 4.3 but the PRISMA images had to be coregistered.
For the coregistration of the PRISMA acquisitions, the approach considered in this work was to use the Gefolki python library [2] which considers two main steps:

- **Optical flow computation:** The algorithm takes an image as a reference and another image as a follower in order to compute the optical flow, i.e. the vertical and horizontal displacements for each pixel between the follower and the reference image.
  As reference image was considered the green band of the Sentinel 2 mosaic built with acquisitions from the period from 11/09/2022 until 18/09/2022 and, in the case of the follower image, it was considered the equivalent wavelength band of each PRISMA acquisition individually.

- **Apply coregistration:** Once the optical flows for each PRISMA acquisition were

[2]Its Github repository is available in: `https://github.com/aplyer/gefolki`

determined, those matrices of vertical and horizontal displacements were used to coregister all the bands of each PRISMA acquisition.

6.2.2. Intersection of PRISMA images with Chl-a maps

After the coregistration of the PRISMA images, there were performed two additional tasks but this time related to the Chlorophyll-a maps:

- **Determination of intersected areas:** Taking for each case the PRISMA image and its corresponding Chlorophyll-a map, it was determined the intersection area among them. Once the intersected areas were determined, the areas outside them were removed.
- **Mask null values:** Additionally, it was relevant to mask the pixels that were not null in the PRISMA image of a specific case but null in its chlorophyll-a map or vice versa.

6.2.3. Removal of anomalous pixels

The final pre-processing step for the input data involved removing anomalous pixels. To accomplish this, three indices were computed from the reflectances of Sentinel 3 - OLCI (S3-OLCI) corresponding to each chlorophyll-a map. Based on the values of these indices, a decision was made regarding whether or not to remove the corresponding pixels from the PRISMA image and Chl-a map.

The following lines provide a summary of the computed indices and their associated conditions:

- **Index 1:** If for a certain pixel of one of the Chlorophyll-a maps its value is above 10 µg/L but the following index [38] is smaller than 1, the pixel is considered anomalous and should be removed:

$$\textbf{Index 1} = \frac{\text{Band 11 (708nm)}}{\text{Band 8 (665nm)}} \tag{6.1}$$

  Specifically, band 11 of S3-OLCI is associated with the Chlorophyll fluorescence baseline red edge transition and band 8 is related to the 2nd Chlorophyll absorption maximum [36].

- **Index 2:** It consists of the ratio between the bands 12 and 11 of the OLCI instrument [12]. The pixels of the S3-OLCI acquisition, where this ratio is higher than

one, refer to an anomaly so it implies the removal of its corresponding PRISMA and Chl-a pixels:

$$\textbf{Index 2} = \frac{\text{Band 12 (753nm)}}{\text{Band 11 (708nm)}} \tag{6.2}$$

Band 12 of the OLCI instrument is related to Oxygen absorption, clouds and vegetation [36].

- **Index 3:** Finally, the third condition is formed by two components. The first one measures the ratio between bands 6 and 3:

$$\textbf{Index 3a} = \frac{\text{Band 6 (560nm)}}{\text{Band 3 (442.5nm)}} \tag{6.3}$$

And its second component refers to the ratio between bands 6 and 4:

$$\textbf{Index 3b} = \frac{\text{Band 6 (560nm)}}{\text{Band 4 (490nm)}} \tag{6.4}$$

If one of the two components is higher than 1 but the other is not, then it refers to an anomalous pixel and it should be removed. Instead, if both components are below 1 it means that it is related to an area with very blue water and if is the case that both components are above 1 it means that it is expected to find phytoplankton in that area.

Band 3 is related to the Chlorophyll maximum absorption, band 4 is associated with high chlorophyll and band 6 is considered as a reference for the chlorophyll minimum [36].

Figure 6.2 summarizes the procedure previously described.

6.3. Determination of input sets: Train, validation, test

In order to determine the acquisitions that were considered in each of the sets the next steps were followed.

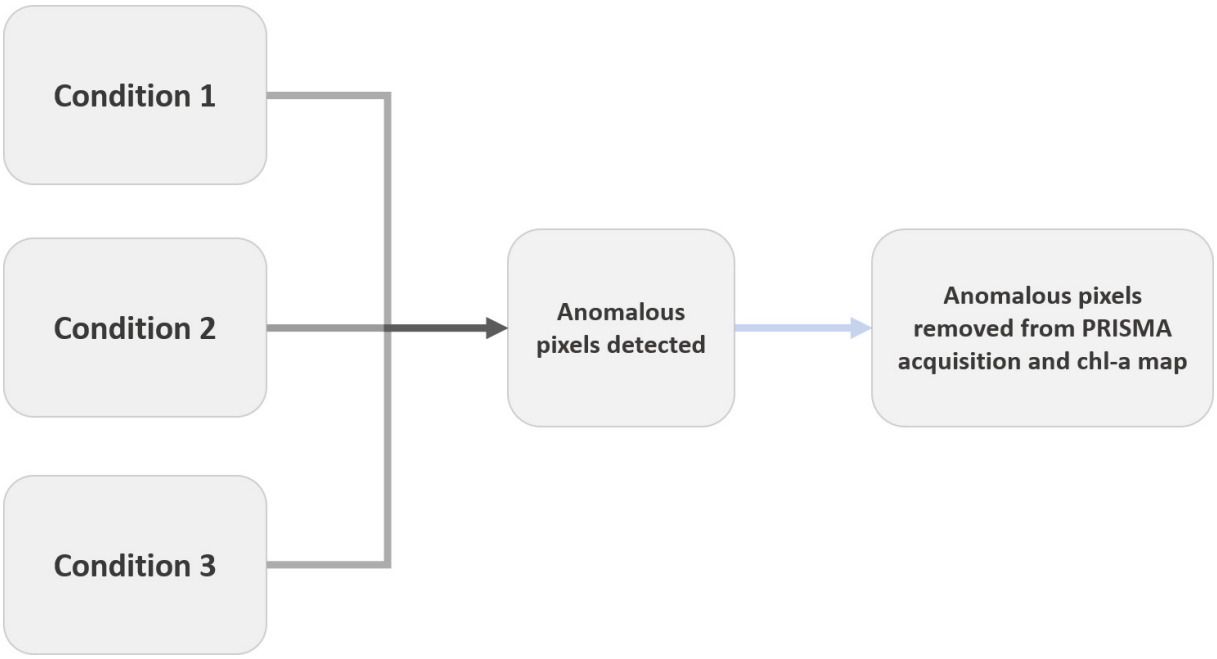

Figure 6.2: Summary of procedure to remove anomalous pixels from each case.

1. Initially, it was intended to perform a homogeneous distribution between the training and validation set with respect to the testing set. For this purpose, it was performed an iterative process in which different sets of images were assigned to one of the two alternative groups (testing vs. train-validation) until achieving the best fitting of the Quantile-Quantile-plot (QQ-plot) that relates both distributions. The final QQ plot of both data distributions is detailed in figure 6.3 and from that figure it is possible to appreciate that the data points of the two distributions fit the identity function with a R^2 of 0.876.

2. In the second place, it was performed a split of the training-validation set. The split assigned 80% of the total input data from the training-validation set to the training subset and the remaining 20% was assigned to the validation set. To perform this split, it was taking into account a stratification procedure to keep the same data distribution in both subsets.

6.4. Dimensionality reduction, normalization techniques and data imbalance manipulations

To properly model the inputs, different alternatives of normalization were considered: min-max scaling, standard scaling and normalization to reflectance units. Also, was manipulated the amount of null data fed to the different models in order to avoid class

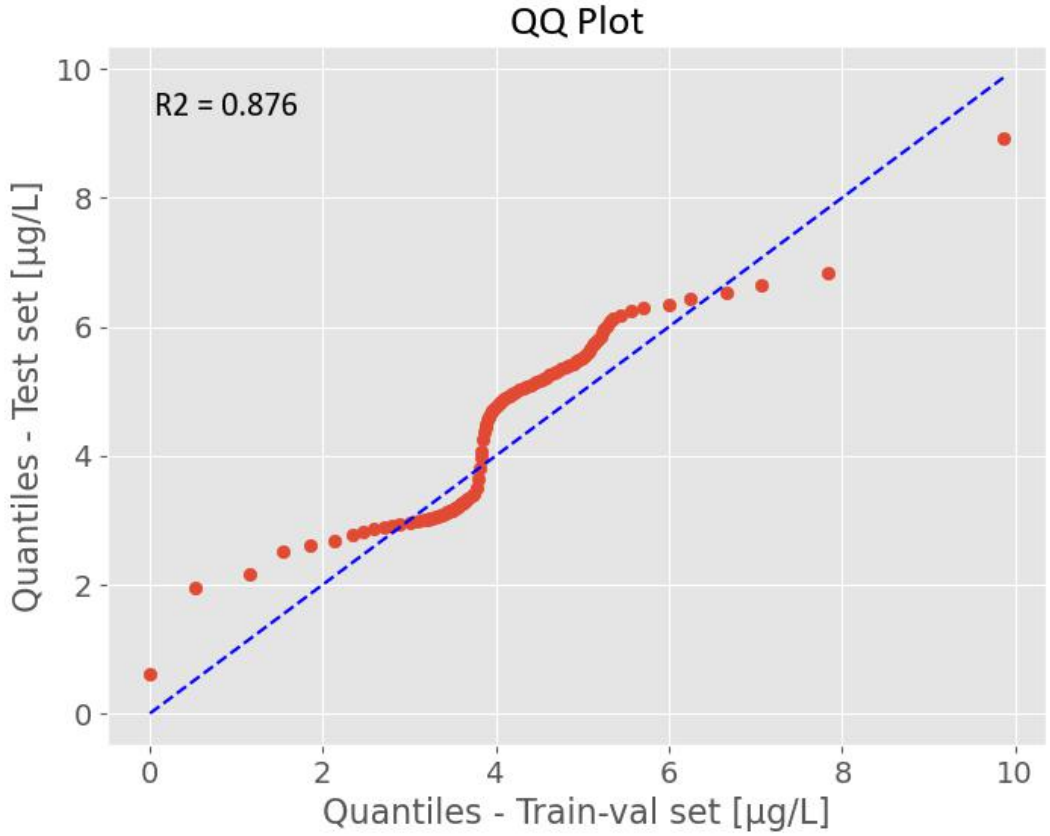

Figure 6.3: QQ-plot of Training-validation set vs. Testing set.

imbalance.
Additionally, the traditional dimensionality reduction technique of Principal Component Analysis (PCA) was explored for reducing the spectral dimension of the inputs. Furthermore, an alternative approach of downsampling the spatial dimension instead of the spectral dimension was also investigated.

In the following lines, there is a summary of each of these processes and how they were implemented.

- **Min-max scaling:** This technique considers the maximum and minimum values of the inputs to normalize them to the range [0-1]. It means that for the case of this study, the Digital Number values of the PRISMA images were scaled to the range previously specified by taking into account the following formula.

$$x_{scaled} = \frac{x - x_{min}}{x_{max} - x_{min}}$$

- **Standard scaling:** This is another common data preprocessing step used to scale the values in a dataset. This technique scales the values so that they have a mean of 0 and a standard deviation of 1. For this study also were scaled the Digital Numbers of the PRISMA images. The formula of their implementation is described in the following.

$$x_{scaled} = \frac{x - \mu}{\sigma}$$

- **Normalization to reflectance units:** Also it was explored the possibility to scale the input digital numbers to reflectance units by the normalization formula proposed in the PRISMA user manual [5] which is detailed next:

$$x_{scaled} = L2scaleXXX_{min} + x_{DN}\frac{.(L2scaleXXX_{max} - L2scaleXXX_{min})}{65.535} \quad (6.5)$$

  Where XXX refers to the specific area of the input spectrum: "VNIR" or "SWIR". L2scale min and max are the minimum and maximum scaling factors specified in the metadata related to each PRISMA image. The number 65.535 used for this normalization, came from the computation $2^{16} - 1$ because the digital numbers used to store the information of each pixel are 16 bits coded.

- **Manipulation on null values:** To mitigate issues related to data imbalance caused by areas of the acquisitions where values are not present, all the null values which are associated with the areas outside the lakes were excluded from the training and validation set.

- **Principal Component Analysis:** Principal Component Analysis (PCA) is a widely used technique for dimensionality reduction in machine learning and data analysis, originally proposed in 1901 by Pearson [54]. This technique is used to identify patterns in data and to reduce the dimensions of a large dataset by transforming it into a new coordinate system that aligns with the principal components of the data. The steps for performing PCA are as follows:

  – Calculation of the mean of the input data:

$$\mu = \frac{1}{n}\sum_{i=1}^{n} x_i$$

where x_i is the i^{th} observation in the dataset and n is the number of observations. In our case, it refers specifically to the digital numbers of the PRISMA images.

- Subtract the mean from each observation to center the data:

$$x_{centered} = x - \mu$$

where x is an observation in the dataset.

- Compute the covariance matrix of the centered data:

$$C = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \mu)(x_i - \mu)^T$$

where ^{T} denotes the transpose of a matrix.

- Compute the eigenvectors and eigenvalues of the covariance matrix:

$$C\mathbf{v} = \lambda\mathbf{v}$$

where $\mathbf{v}$ is an eigenvector and λ is the corresponding eigenvalue.

- Sort the eigenvectors in descending order based on their eigenvalues, and choose the first k eigenvectors as the principal components. The number of principal components, k, is chosen based on the amount of variance in the data that needs to be retained.

- Project the centered data onto the principal components to obtain the reduced-dimension representation of the data:

$$x_{reduced} = \mathbf{v}^T x_{centered}$$

where $\mathbf{v}^T$ is the transpose of the matrix of eigenvectors.

The eigenvectors and eigenvalues of the covariance matrix represent the principal components of the data. The principal components are orthogonal to each other, and they capture the directions of maximum variance in the data. By projecting the data onto these directions, we can obtain a lower-dimensional representation of the data that retains most of the information.

By applying PCA the spectral dimension of the input data was reduced from the initial 230 bands of the PRISMA images to 30 principal components.

- **Spatial Downsampling:** Considering that the input images have a spatial resolution of 30 meters but the ground truth one of 300 meters, it was explored also if spatial downsampling of the PRISMA input images to 300 meters was beneficial for the objective of this study. To this aim, there was considered the nearest neighbour interpolation approach.

  Additionally, during some of the experiments that will be later detailed, this approach was complemented with the extraction of additional features by means of some traditional filters such as the following ones:

  - **Mean filter:** Also known as Box Filter, this filter replaces the center pixel of a 3×3 window with the average of its surrounding pixels [39]. The filter kernel is given by the following matrix [39]:

$$\begin{bmatrix} 1/9 & 1/9 & 1/9 \\ 1/9 & 1/9 & 1/9 \\ 1/9 & 1/9 & 1/9 \end{bmatrix}$$

  - **Sobel-X filter:** This filter emphasizes vertical edges in an image by convolving the input with the Sobel-X kernel [66]. The Sobel-X kernel is given by the next matrix [66]:

$$\begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}$$

  - **Sobel-Y filter:** This filter emphasizes horizontal edges in an image by convolving the input with the Sobel-Y kernel [66]. The Sobel-Y kernel is the following [66]:

$$\begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix}$$

6.5. Regression of Chl-a values

One of the primary goals of the study was to estimate chlorophyll-a values in the selected lakes. To accomplish this objective, several models were employed, and various data processing techniques were utilized. This sub-section summarizes the main models implemented, the different configurations of parameters taken into account, and finally the results are described for each case.

6.5.1. Models

Among the possible models to be implemented for this task, there were considered the following ones:

1. **Random Forest Regressor:** This type of model corresponds to the category of Traditional Approaches - Spectral Classification, previously described in subsection 4.4.1. In a Random Forest Regressor [10, 20], the algorithm builds multiple decision trees on randomly selected subsets of the training data, and the final prediction is the average of the predictions of all trees. This algorithm was implemented by means of the Python library "Scikit-learn" [3]. The parameters that were sensitized were the next:
    - **Number of estimators:** Number of decision trees chosen [63]. Higher numbers could imply better performance but also higher computational costs.
    - **Minimum number of samples-leaf:** This parameter determines the minimum number of samples to be a leaf node of the tree (to take a decision) [63]. This parameter also works in the regularization task because by increasing its value can reduce overfitting.
    - **Maximum depth of each decision tree:** By increasing its value is also possible to achieve more complex models and then conduce to overfitting [63]. For this reason, it is another parameter to be set to regulate the tradeoff between variance and bias.
2. **Support Vector Regressor (SVR):** Also this model corresponds to the category of Traditional Approaches - Spectral Classification introduced in subsection 4.4.1. SVR is used for both linear and nonlinear regression tasks and is known for its ability to handle high-dimensional data. In a Support Vector Regressor [21, 65],

[3] `https://scikit-learn.org/stable/`

the algorithm finds a hyperplane that maximizes the margin between the training data points and the regression line. The hyperplane is selected based on the support vectors, which are the data points closest to the hyperplane. Support Vector Regressor is suitable for hyperspectral data because it can handle a large number of features. In our particular case, considering the kernels offered by the Python library "Scikit-learn" it was decided to use the radial basis function (RBF). With the selected kernel, several trials were performed sensitizing the following key parameters:

- **Gamma**: This is the kernel coefficient for the RBF kernel and it determines the shape of the decision boundary [64]. A larger gamma value means a more complex decision boundary and may lead to overfitting.
- **C**: Controls the width of the margin and how many data points are allowed to be misclassified [64]. For this reason, it is a regularization parameter and defines the tradeoff between achieving a low training error and a low testing error.

3. **Long Short-Term Memory (LSTM):** This model lies in the category of Deep Learning Approaches - Spectral classification, introduced in subsection 4.4.2. LSTM [37, 40] is a type of recurrent neural network (RNN) that is widely used for time-series data. In the case of hyperspectral data, the spectral bands can be considered as a time series. LSTMs are designed to handle long-term dependencies and can remember information from previous time steps. In an LSTM neural network, the algorithm uses multiple memory cells and gates to control the flow of information. The gates regulate the information flow by deciding what information to keep or forget. LSTM neural network is suitable for chlorophyll-a estimation from hyperspectral data because it can capture the sequential dependencies between the spectral bands. For its implementation, it was used the Python package called "Tsai"[4]. Among the main parameters that were sensitized in this case are the following ones.
   - **Hidden state:** It refers to the internal memory of the LSTM cell that stores information about the previous steps in a sequence.
   - **Number of layers:** The number of LSTM cells stacked on top of each other.
   - **Dropout in the recurrent neural network cells:** Dropout is a regularization technique that prevents overfitting by randomly dropping out a fraction

[4] `https://timeseriesai.github.io/tsai/`

of the neurons [67]. In the case of recurrent neurons, the dropout technique is applied in the connection between consecutive recurrent hidden cells.

- **Dropout in the fully-connected layer:** here the technique is applied to the outputs of the fully-connected layer.
- **Directionality:** There were explored unidirectional and bidirectional LSTM networks. The difference between them is that the bidirectional networks compute the hidden state at each time step taking into account both the past and future inputs but the unidirectional networks instead just consider the past inputs.

4. **Gated Recurrent Unit (GRU):** Also this model corresponds to the classification of Deep Learning Approaches - Spectral classification, described in subsection 4.4.2. GRU was proposed in [17] and it is another type of RNN architecture. GRU is similar to LSTM, but it has fewer parameters and is faster to train. In this case, the input sequence is passed through a series of hidden layers, and each layer uses a gated mechanism to selectively remember or forget the information from previous sequence steps. Also, this architecture was implemented by relying on the Tsai package. Regarding the parameters that were sensitized for this architecture, they coincide with the ones described for the LSTM architecture.

6.5.2. Results of each model typology and discussion

In this subsection, the results of the different experiments conducted during this study are presented. Alternative models were trained, and various parameters were sensitized to achieve the best possible configuration for the project's objective. To measure and evaluate the performance of all the experiments, two well-known metrics commonly used in the scientific community were adopted, with the unit of measure being µg/L.:

- **Mean Absolute Error (MAE):** This metric measures the average absolute error between the predictions and the actual values [74], in the following equation n refers to the number of samples, y_i to the actual value and $\hat{y}_i$ is the prediction.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$$

- **Root Mean Square Error (RMSE):** This second metric, is another way to evaluate the predictions where the residuals between the actual values and the predictions are computed and squared, then their summation is performed and finally the result

is divided by the total number of samples and the squared root is computed [74].

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

Random Forest Regressor

Initially, experiments were conducted with the three normalization techniques introduced in section 6.4 (RF-1 to RF-3) to determine the most appropriate alternative. The result achieved by the three options was almost identical in terms of the evaluated metrics as it can be appreciated from table 6.1.

Then, the possibility of applying PCA to reduce the dimensionality was evaluated in order to extract the 30 principal components (PCs) from the original 230 bands of PRISMA images. This operation has shown a worse result in comparison with the previous attempts, by analyzing table 6.1 can be observed that the MAE and the RMSE of experiment RF-4 are higher with respect to the first three experiments.

Next, it was analyzed during experiment RF-5, if including additionally extracted features was improving the performance of the model. In this experiment was kept the same normalization approach as in experiment RF-2 (Standard Scaling) and no dimensionality reduction was applied. For each pixel were applied the Mean, and the Sobel x and Sobel y filters. The result of experiment RF-5 suggested that the addition of the features was not helpful for the model because experiment RF-2 which had the same parameters and normalization on the input data was achieving a better performance without the need of including additional features.

Then, experiments were conducted from RF-6 to RF-11 to evaluate the various possible configurations for the hyperparameters of the model and select the best alternative (i.e. solving the typical model selection problem). Among these experiments, RF-10 stood out with a better performance with respect to the other alternatives.

Finally, it was evaluated if the use of PRISMA images with a spatial resolution of 30 meters (original resolution) instead of 300 meters (downsampled images) allowed us to achieve a better performance in the predictions. In order to perform this evaluation, experiment RF-12 used almost the same configuration as the one used for experiment RF-10 -except the number of trees parameter due to computational constraints- but the PRISMA images were considered in their original resolution and the chlorophyll-a maps were upsampled by the nearest neighbour method to 30 meters (original resolution of PRISMA images). Observing the table 6.1 is possible to appreciate that the result of RF-12 exhibits a higher error (MAE and RMSE) with respect to RF-10.

Id	*Res [m]*	*PCA*	*Norm.*	*Data augm.*	*N° trees*	*Min. leaf*	*Max. depth*	*MAE*	*RMSE*
RF-1	300	No	Minmax	No	1000	3	10	0,931	1,112
RF-2	300	No	Std.	No	1000	3	10	0,931	1,112
RF-3	300	No	Reflect.	No	1000	3	10	0,931	1,112
RF-4	300	30 PC's	Std.	No	1000	3	10	1,020	1,245
RF-5	300	No	Std.	Yes	1000	3	10	1,106	1,296
RF-6	300	No	Std.	No	1000	3	5	1,032	1,192
RF-7	300	No	Std.	No	1000	3	20	0,930	1,113
RF-8	300	No	Std.	No	100	3	20	0,947	1,128
RF-9	300	No	Std.	No	10000	3	20	0,924	1,107
RF-10	300	No	Std.	No	10000	2	20	0,915	1,099
RF-11	300	No	Std.	No	10000	10	20	0,934	1,114
RF-12	30	No	Std.	No	1000	2	20	0,986	1,181

Table 6.1: Results of Random Forest experiments.

Among the different attempts, the best performance was obtained in the experiment RF-10. The results associated with each test case are described in table 6.2, the corresponding predictions, ground truths and their residuals are in figure 6.4 and the figure 6.5 presents the histogram with the distribution of the residuals for each of the cases of the testing set.

Id	*MAE-4*	*RMSE-4*	*MAE-23*	*RMSE-23*	*MAE-24*	*RMSE-24*
RF-10	0,464	0,622	0,903	1,106	1,378	1,570

Table 6.2: Results of experiment RF-10 for each of the three test cases.

Support Vector Regressor

In the analysis of the normalization techniques carried out with experiments SVR-1 to SVR-3, it was deduced that the best alternative for normalising the inputs is the Standard Scaling technique. This can be appreciated by observing in table 6.3 the result of experiment SVR-2 with respect to the results of experiments SVR-1 and SVR-3 which refer to the other two normalization techniques.

Regarding the utility of applying PCA on the input PRISMA images, it was determined from experiment SVR-4 that the performance increases. The previous affirmation can be seen in table 6.3 by comparing the result achieved in experiment SVR-2 with respect to the one obtained by experiment SVR-4. For this reason, this technique was also considered for the following experiments.

Experiment SVR-5 evaluated the benefit of using data augmentation on the input data. It was determined that it was not helpful to improve the performance of the model because it produced a worse result with respect to SVR-4.

From experiments SVR-6 to SVR-9 were evaluated different possible setups for the hyper-parameters C and gamma of the Support Vector Regressor model but it was determined that the best results were still achieved by experiment SVR-4 and for that reason this configuration was preferred.
Finally, experiment SVR-10 analyzed the performance of using 30 meters of spatial resolution for the input data, which corresponds to the original PRISMA images and Chl-a maps upsampled by the nearest neighbour method. The results of this experiment were not promising because the MAE and RMSE metrics were higher than the previous best model (SVR-4).

Id	*Res [m]*	*PCA*	*Norm.*	*Data augm.*	*Gamma*	*C*	*MAE*	*RMSE*
SVR-1	300	No	Minmax	No	0,001	15	1,285	1,431
SVR-2	300	No	Std.	No	0,001	15	0,699	0,898
SVR-3	300	No	Reflect.	No	0,001	15	1,253	1,394
SVR-4	300	30 PC's	Std.	No	0,001	15	0,687	0,895
SVR-5	300	30 PC's	Std.	Yes	0,001	15	0,909	1,126
SVR-6	300	30 PC's	Std.	No	0,0001	15	0,752	0,993
SVR-7	300	30 PC's	Std.	No	0,01	15	0,956	1,152
SVR-8	300	30 PC's	Std.	No	0,001	1,5	0,756	0,955
SVR-9	300	30 PC's	Std.	No	0,001	150	1,106	1,307
SVR-10	30	30 PC's	Std.	No	0,001	15	1,260	1,555

Table 6.3: Results of Support Vector Regressor experiments.

According to the experimental development of this set, the best configuration corresponds to the one of experiment SVR-4. Table 6.4 presents the results of this model for each of the three test cases. In figure 6.6 are exposed the predictions, ground truths and the corresponding residuals for each case and figure 6.7 presents the distribution of the errors in each of the test cases.

Id	*MAE-4*	*RMSE-4*	*MAE-23*	*RMSE-23*	*MAE-24*	*RMSE-24*
SVR-4	0,544	0,688	0,712	0,961	0,806	1,036

Table 6.4: Results of experiment SVR-4 for each of the three test cases.

Long-Short Term Memory network

From experiments LSTM-1, LSTM-2 and LSTM-3 was possible to determine that the best normalization approach was the Standard scaling technique which corresponds to experiment LSTM-2. These results are detailed in table 6.5.
Later, was evaluated if applying PCA to the input data was useful or not for the performance of the model. From the results of experiment LSTM-4 is possible to determine

that it was helpful for improving the performance because it exhibits lower values of MAE and RMSE with respect to experiment LSTM-2.
Starting from experiment LSTM-5 until experiment LSTM-12 were sensitized all the hyper-parameters of this model architecture which were previously introduced in the subsection 6.5.1. Among these experiments, the best alternative referred to the experiment LSTM-10 because, as it can be observed from table 6.5 it obtained the lowest values for MAE and RMSE.

Additionally, experiment LSTM-13 kept all the parameters with the same values as in experiment LSTM-10 but it considered bidirectional flow. In this case, the result obtained was superlative with respect to the previous best model (LSTM-10).

Finally, was analyzed the use of inputs with a spatial resolution of 30 meters in experiment LSTM-14, keeping the same configuration for the model hyper-parameters and normalization of the inputs as in experiment LSTM-13 but the result was not improving against the previous best model (LSTM-13).

Id	***Res [m]***	***PCA***	***Norm.***	***Hidden state***	***Layers***	***Drop. RNN***	***Drop. FCN***	***Bidir.***	***MAE***	***RMSE***
LSTM-1	300	No	Minmax	10	2	0,6	0,4	No	1,443	1,584
LSTM-2	300	No	Std.	10	2	0,6	0,4	No	1,303	1,431
LSTM-3	300	No	Reflect.	10	2	0,6	0,4	No	1,897	2,012
LSTM-4	300	30 PC's	Std.	10	2	0,6	0,4	No	1,298	1,428
LSTM-5	300	30 PC's	Std.	5	2	0,6	0,4	No	1,386	1,522
LSTM-6	300	30 PC's	Std.	15	2	0,6	0,4	No	1,323	1,452
LSTM-7	300	30 PC's	Std.	10	4	0,6	0,4	No	1,494	1,635
LSTM-8	300	30 PC's	Std.	10	1	0,6	0,4	No	1,334	1,490
LSTM-9	300	30 PC's	Std.	10	2	0,2	0,4	No	1,342	1,475
LSTM-10	300	30 PC's	Std.	10	2	0,8	0,4	No	1,278	1,407
LSTM-11	300	30 PC's	Std.	10	2	0,8	0,6	No	1,366	1,498
LSTM-12	300	30 PC's	Std.	10	2	0,8	0,2	No	1,305	1,434
LSTM-13	300	30 PC's	Std.	10	2	0,8	0,4	Yes	1,211	1,345
LSTM-14	30	30 PC's	Std.	10	2	0,8	0,4	Yes	1,278	1,455

Table 6.5: Results of LSTM model experiments.

Based on the previous information, the best-performing model for this set of experiments was LSTM-13. Table 6.6 describes the error achieved by this model in each of the three cases of the test set. The predictions of this experiment, the corresponding ground truths for each test case and the differences between predictions and ground truths are described in figure 6.8. Additionally, figure 6.9 exposes the histogram of the differences between the predictions and the ground truths for each case.

Id	*MAE-4*	*RMSE-4*	*MAE-23*	*RMSE-23*	*MAE-24*	*RMSE-24*
LSTM-13	0,992	1,053	1,288	1,442	1,355	1,538

Table 6.6: Results of experiment LSTM-13 for each of the three test cases.

Gated Recurrent Unit network

The first set of experiments (GRU-1 to GRU-3) were analyzing the best normalization approach, obtaining as a result the standard scaling method (GRU-2). The results can be appreciated in table 6.7.
Experiment GRU-4 allows us to determine that the use of PCA method for this model was not useful because the experiment provided a lower performance with respect to GRU-2. Then, from experiment GRU-5 until experiment GRU-15 were sensitized all the hyperparameters previously introduced in the sub-section 6.5.1, achieving the best performance with the configuration adopted in experiment GRU-8.
Later, in experiment GRU-16 was analyzed if taking the configuration of GRU-8 and setting the network with bidirectional flow improved the performance or not: the result of this experiment exhibited a lower performance according to what is detailed in table 6.7.

Finally, experiment GRU-17 was conducted to analyze the model's performance when using input data with a spatial resolution of 30 meters, keeping the same configuration as in experiment GRU-8. The obtained result of this new experiment was worse with respect to GRU-8.

Id	*Res [m]*	*PCA*	*Norm.*	*Hidden state*	*Layers*	*Drop. RNN*	*Drop. FCN*	*Bidir.*	*MAE*	*RMSE*
GRU-1	300	No	Minmax	10	2	0,6	0,4	No	1,367	1,499
GRU-2	300	No	Std.	10	2	0,6	0,4	No	1,287	1,416
GRU-3	300	No	Reflect.	10	2	0,6	0,4	No	1,559	1,698
GRU-4	300	30 PC's	Std.	10	2	0,6	0,4	No	1,305	1,433
GRU-5	300	No	Std.	5	2	0,6	0,4	No	1,435	1,575
GRU-6	300	No	Std.	20	2	0,6	0,4	No	1,235	1,366
GRU-7	300	No	Std.	40	2	0,6	0,4	No	1,221	1,352
GRU-8	300	No	Std.	60	2	0,6	0,4	No	1,186	1,321
GRU-9	300	No	Std.	100	2	0,6	0,4	No	1,271	1,408
GRU-10	300	No	Std.	60	1	0,6	0,4	No	1,236	1,373
GRU-11	300	No	Std.	60	10	0,6	0,4	No	1,231	1,362
GRU-12	300	No	Std.	60	2	0,2	0,4	No	1,272	1,419
GRU-13	300	No	Std.	60	2	0,8	0,4	No	1,194	1,340

GRU-14	300	No	Std.	60	2	0,6	0,2	No	1,202	1,355
GRU-15	300	No	Std.	60	2	0,6	0,8	No	1,260	1,399
GRU-16	300	No	Std.	60	2	0,6	0,4	Yes	1,213	1,363
GRU-17	30	No	Std.	60	2	0,6	0,4	No	1,203	1,382

Table 6.7: Results of GRU model experiments.

According to the previous explanation, the best model was corresponding to experiment GRU-8, which metrics are described in table 6.8. In figure 6.10 can be observed the predictions of this experiment for the three cases of our test set, their associated ground truths of chlorophyll-a maps and the difference between the predictions and the ground truths. Additionally, figure 6.11 describes the distribution of the residuals for each of the test cases.

Id	*MAE-4*	*RMSE-4*	*MAE-23*	*RMSE-23*	*MAE-24*	*RMSE-24*
GRU-8	0,929	0,997	1,262	1,420	1,365	1,544

Table 6.8: Results of experiment GRU-8 for each of the three test cases.

6.5.3. Summary of best models

In this subsection are briefly summarized the results of the best models of each type. Among the different alternatives, the SVR achieved with its best model (experiment SVR-4) the higher performance. The results of the best models for each type of model are detailed in the table 6.9.

Id	*Model*	*MAE general*	*RMSE general*	*MAE-4*	*RMSE-4*	*MAE-23*	*RMSE-23*	*MAE-24*	*RMSE-24*
SVR-4	SVR	0,687	0,895	0,544	0,688	0,712	0,961	0,806	1,036
RF-10	RF	0,915	1,099	0,464	0,622	0,903	1,106	1,378	1,570
GRU-8	GRU	1,186	1,321	0,929	0,997	1,262	1,420	1,365	1,544
LSTM-13	LSTM	1,211	1,345	0,992	1,053	1,288	1,442	1,355	1,538

Table 6.9: Comparison of metrics for the best models.

6.5.4. Inference on 30 meters GSD inputs

After the determination of the best models from the experimental development that was carried out, it was possible to observe that the experiment which achieved the higher

performance used input data with 300 m GSD. This spatial resolution is the same as the Sentinel 3 imagery and it refers to a value of low spatial resolution. However, considering that the PRISMA images were originally of 30 m GSD, it was evaluated the performance of the best models summarized in the previous subsection 6.5.3, when predicting chlorophyll-a maps of 30 meters instead of 300 meters. For this reason, both Chl-a ground truth and the associated PRISMA images of each of the three test cases (id 4, 23 and 24) were upsampled to 30 meters. The upsampled PRISMA images were input into the trained models to obtain the corresponding predictions and finally, the performances were determined for each of the model typologies by comparing the obtained predictions against the upsampled Chl-a ground truths. Table 6.10 details the performance achieved for each of the three test cases. From table 6.10 is possible to appreciate that using the best models previously determined, which were trained with 300 meters GSD input data, is feasible to achieve acceptable results while predicting cases which are with 30 meters spatial resolution. This is quite relevant because it implies that when having a new PRISMA image over the area of interest, instead of obtaining a chlorophyll-a map at 300 meters, the resulting map will have 30 meters which is 10 times better than the Sentinel 3 imagery.

Additionally, it is worth mentioning that in this new evaluation with 30 meters of input data, the best model among the four preselected ones is the Random Forest Regressor (exp. RF-10) instead of the Support Vector Regressor (exp. SVR-4). Also, observing the figure 6.14 is possible to notice that the output produced an average value for each of the three test cases which is something undesired, for this reason, it is also preferred the RF-10 when evaluated with 30-meters input data.

Then, were compared the results of the models trained with 300-meter inputs evaluated with 30-meter inputs with respect to the models both trained and evaluated with 30-meter inputs. In table 6.10 is presented a summary of the results. It can be observed that in the case of Random Forest and Gated Recurrent Unit models, the performances are better. In contrast, training and inferring on 30-meter inputs but in the case of Support Vector Regressor and LSTM network models it was better to train with 300-meter inputs and then evaluate on 30-meter inputs.

The conclusion is that, among all the possible alternatives, if the final objective is to produce chlorophyll-a maps with 30-meter GSD, it is recommended to use the Random Forest model (exp. RF-12) trained with 30-meter input data.

The following figures provide further details about the results of each of the models trained

with 300-meter inputs but evaluated with 30-meter data:

- Figures 6.12, 6.13: predictions, ground truths and differences for experiment RF-10
- Figures 6.14, 6.15: idem for experiment SVR-4.
- Figures 6.16, 6.17: idem for experiment LSTM-13.
- Figures 6.18, 6.19: idem for experiment GRU-8.

And the figures that present the visual results of the models both trained and evaluated on 30-meter inputs are:

- Figures 6.20, 6.21: predictions, ground truths and differences for experiment RF-12
- Figures 6.22, 6.23: idem for experiment SVR-10.
- Figures 6.24, 6.25: idem for experiment LSTM-14.
- Figures 6.26, 6.27: idem for experiment GRU-17.

Id	*Model*	*Train GSD [m]*	*Eval. GSD [m]*	*MAE general*	*RMSE general*	*MAE-4*	*RMSE-4*	*MAE-23*	*RMSE-23*	*MAE-24*	*RMSE-24*
RF-10	RF	300	30	1,076	1,241	0,988	1,071	0,836	1,068	1,405	1,585
RF-12	RF	30	30	0,986	1,181	0,815	0,921	0,707	0,987	1,435	1,635
SVR-4	SVR	300	30	1,107	1,266	1,578	1,620	0,778	1,017	0,964	1,161
SVR-10	SVR	30	30	1,260	1,555	1,052	1,571	1,043	1,235	1,686	1,859
LSTM-13	LSTM	300	30	1,234	1,369	0,826	0,905	1,413	1,556	1,462	1,648
LSTM-14	LSTM	30	30	1,278	1,455	1,004	1,112	1,004	1,214	1,826	2,039
GRU-8	GRU	300	30	1,248	1,393	0,643	0,746	1,294	1,448	1,808	1,986
GRU-17	GRU	30	30	1,203	1,382	0,598	0,727	1,518	1,732	1,493	1,686

Table 6.10: Evaluation of the best models using 30 m GSD inputs.

Finally, taking into consideration the model that achieved the highest performance when inferring 30-meter input data (RF-12), it was analyzed if the use of the bilinear interpolation method for upsampling the ground truth chlorophyll-a maps to 30 meters GSD was better than the previously used method of Nearest Neighbour.

To make this comparison, the edges of the lakes had to be removed from all the acquisitions because the bilinear interpolation method produced a border effect that was not desired. An example of the mentioned border effect can be appreciated in Figure 6.28 where the original chlorophyll-a map is presented together with the two evaluated upsampling methods.

The removal of the lakes' edges was done by applying a buffer of 300 meters which corresponds to the size of a pixel from the original chlorophyll-a maps to be upsampled. This operation was considered in the two experiments that were set for the comparative analysis (RF-13 and RF-14).

The conclusion of this comparison of upsampling methods was that the bilinear interpolation achieved slightly better performance with respect to the nearest neighbour technique but in the third test case (acquisition Id 24) the error was in a higher order of magnitude for the bilinear interpolation method. For this reason and because of the previously introduced border effect, it was preferred the nearest neighbour method. This conclusion can be confirmed by observing Table 6.11 where the results of experiments RF-13 and RF-14 are presented.

Exp. Id	*Ups. method*	*MAE general*	*RMSE general*	*MAE-4*	*RMSE-4*	*MAE-23*	*RMSE-23*	*MAE-24*	*RMSE-24*
RF-13	Bilinear	1,04	1,20	0,54	0,65	0,59	0,79	2,00	2,14
RF-14	Nearest	1,05	1,20	0,79	0,86	0,69	0,88	1,67	1,86
Diff. 13 - 14	-	-0,01	0,00	-0,25	-0,21	-0,10	-0,08	0,33	0,28

Table 6.11: Results of comparison between upsampling methods.

6.6. General discussion of the chapter

This chapter described the complete process developed to produce a model able to perform predictions of chlorophyll-a on lakes: it started detailing the aspects related to the acquisition and preparation of the data, then specified all the different experiments and the criteria taken into consideration to decide how to select the best possible models among the existing alternatives and, once was determined the most performing option for each of the model typologies, a second evaluation for each of the best models were performed in order to determine the capabilities to be used for predicting chlorophyll-a maps with 30 meters spatial resolution.

According to the achieved results, while tuning each of the model typologies, was ob-

served that the use of 30 meters of input data for training the models was not improving the performances, for this reason, the experiments with the higher performances of each typology used 300 meters of input data. Specifically, among the four selected experiments (RF-10, SVR-4, LSTM-13 and GRU-8), the best performance referred to SVR-4.

Then, to understand if those preselected models that were trained with 300-meter GSD input data were useful for predicting 30 meters chlorophyll-a maps, a second evaluation was done. The result of this second evaluation allowed us to understand that the Random Forest Regressor (experiment RF-10) was the model which outperform the others. However, it was also relevant to determine if these models that were trained with 300-meter inputs and evaluated on 30-meter inputs were achieving a better performance than the experiments in which 30-meter inputs were used also for training. The results showed that in two of the four model typologies, it was better to use 30-meter inputs for training and evaluating and among all the possible alternatives, the best option corresponds with experiment RF-12. The conclusion of this second analysis is that if the objective is to achieve a chlorophyll-a map with a GSD of 300 meters, the best alternative is to follow the settings considered in experiment SVR-4, but if the objective is to predict a chlorophyll-a map with a 30-meter GSD, is better to follow the configuration of experiment RF-12 which used 30-meter inputs for training and testing.

Also, taking into account the configuration used for experiment RF-12 (which achieved the highest performance when predicting 30-meter chlorophyll-a maps), it was evaluated if the bilinear interpolation method could be better than the used nearest neighbour method when upsampling the original chlorophyll-a maps to 30 meters of spatial resolution. For this purpose, there were set experiments RF-13 and RF-14 and from their results was possible to observe that although the performance of the bilinear interpolation method was slightly better, for one of the test cases, the result with this upsampling method was strongly affected and also this technique produced undesired border effects. For these two reasons, the preferred method is the nearest neighbour for upsampling the ground truth chlorophyll-a maps to 30 meters GSD.

Finally, another important insight was the fact that generally, the model typologies which achieved higher performances were part of the category of Traditional approaches (Random Forest and Support Vector Regressor) instead of Deep Learning ones (Long-short Term Memory and Gated Recurrent Unit networks).

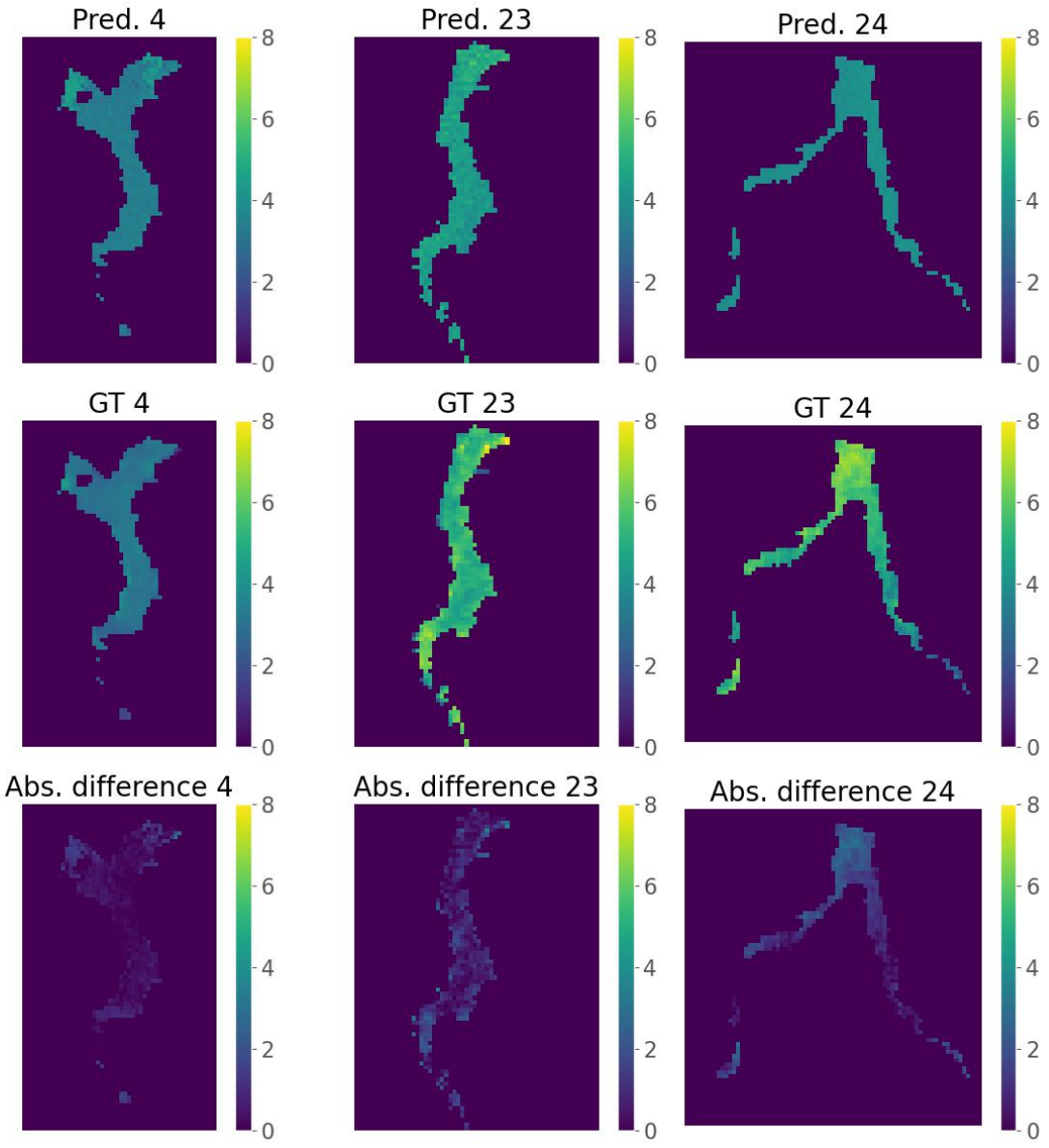

Figure 6.4: Predictions, Ground Truths and difference GT-Preds. of experiment RF-10

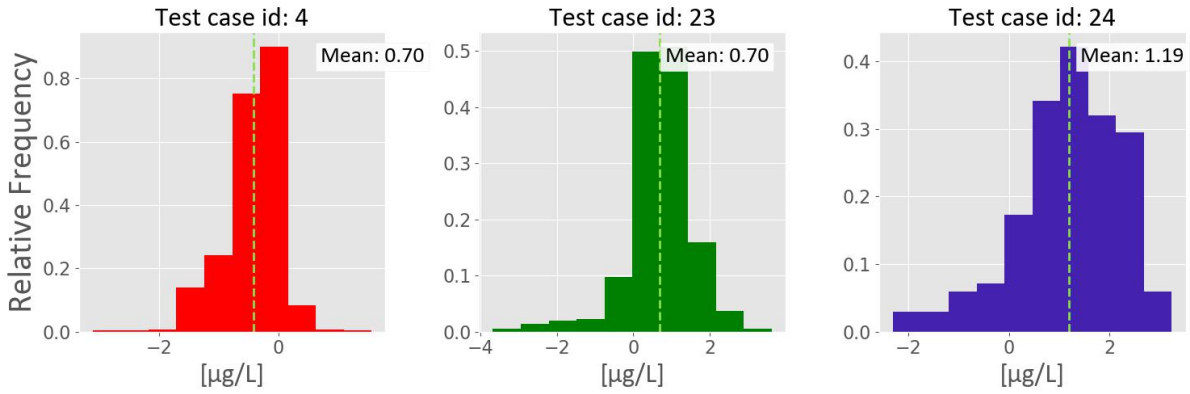

Figure 6.5: Distribution of the errors of experiment RF-10

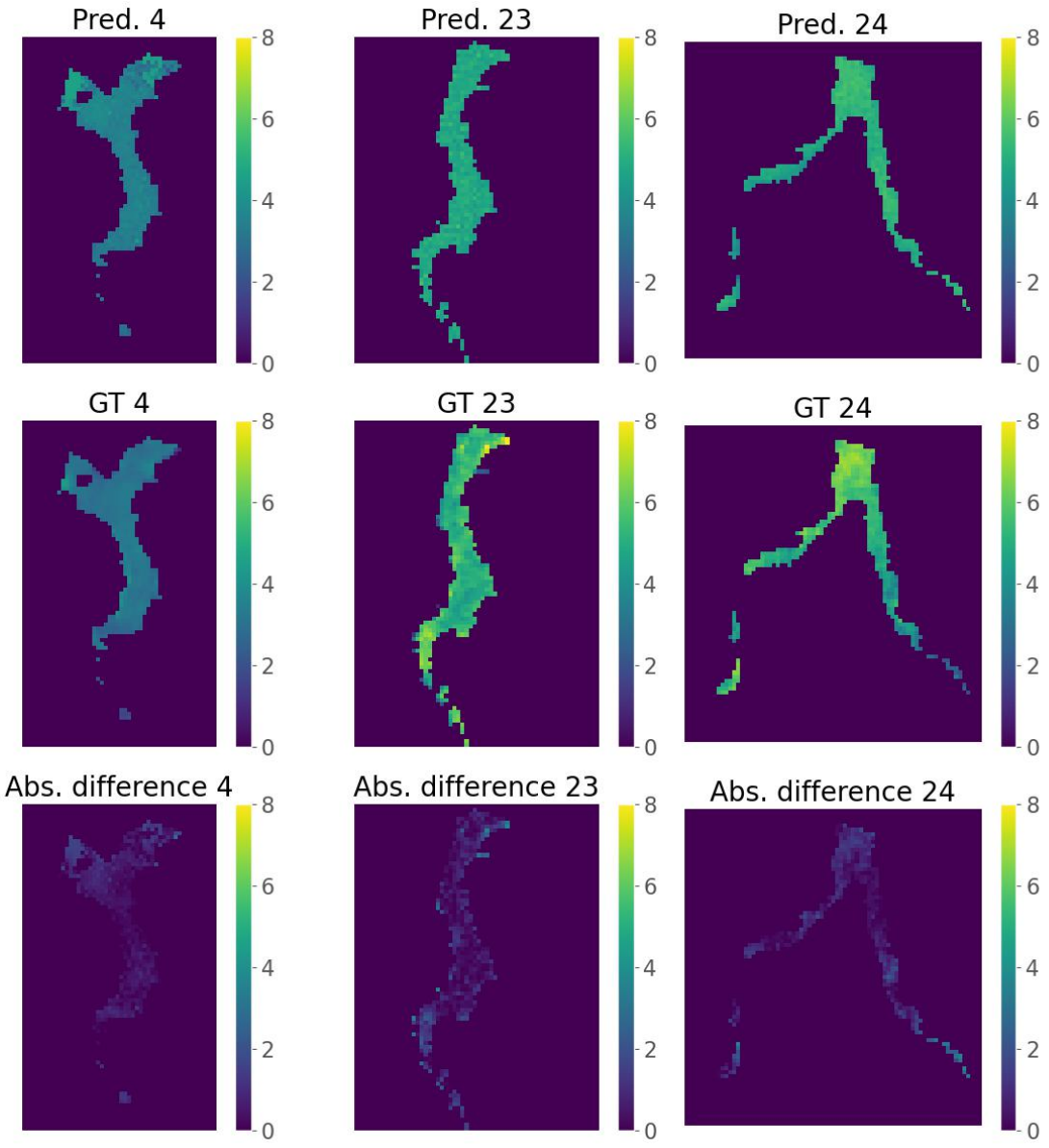

Figure 6.6: Predictions, Ground Truths and difference GT-Preds. of experiment SVR-4

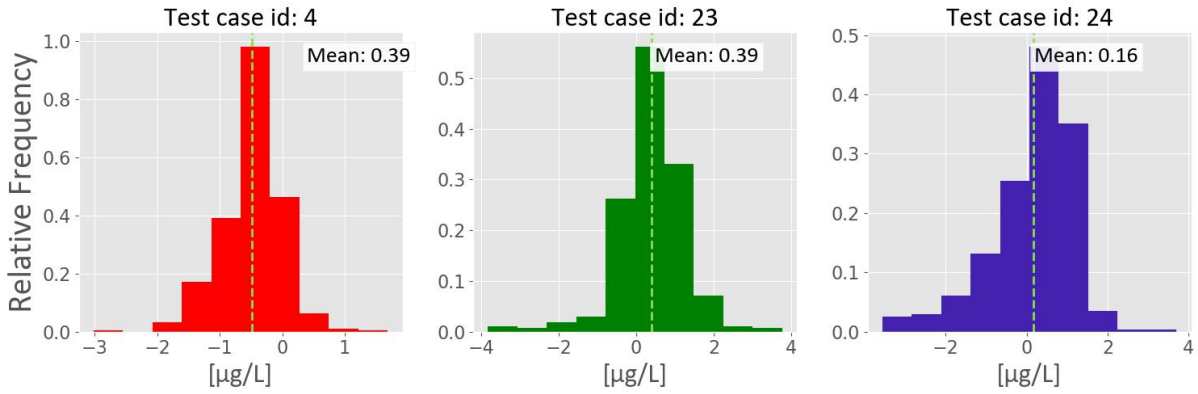

Figure 6.7: Distribution of the errors of experiment SVR-4

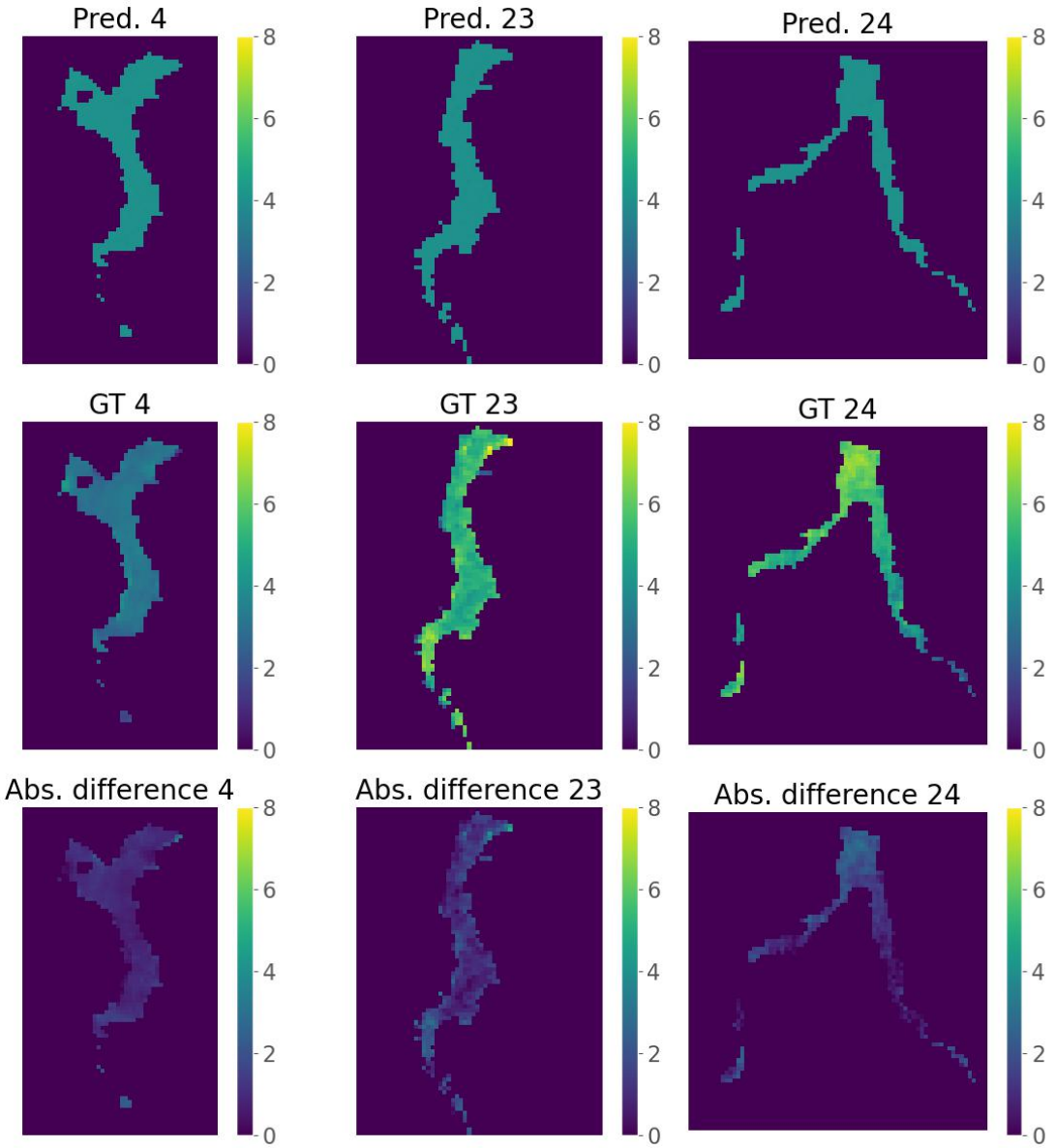

Figure 6.8: Predictions, Ground Truths and difference GT-Preds. of experiment LSTM-13

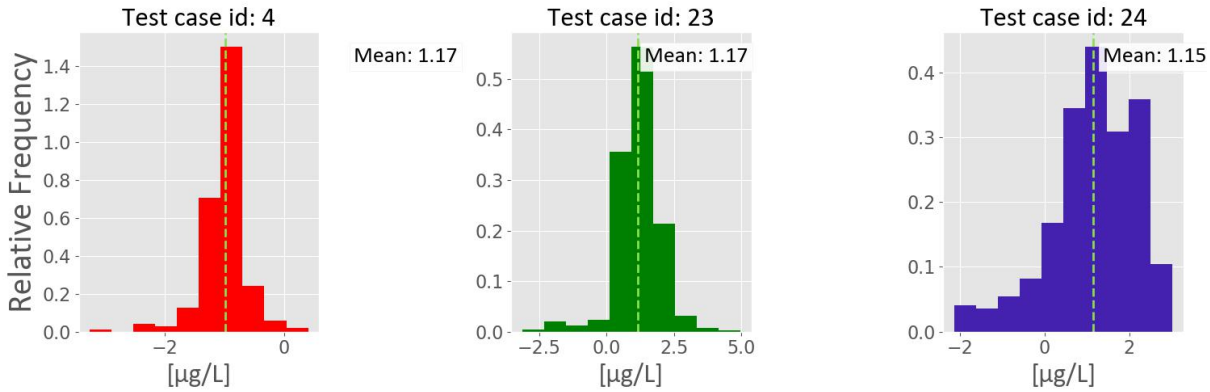

Figure 6.9: Distribution of the errors of experiment LSTM-13

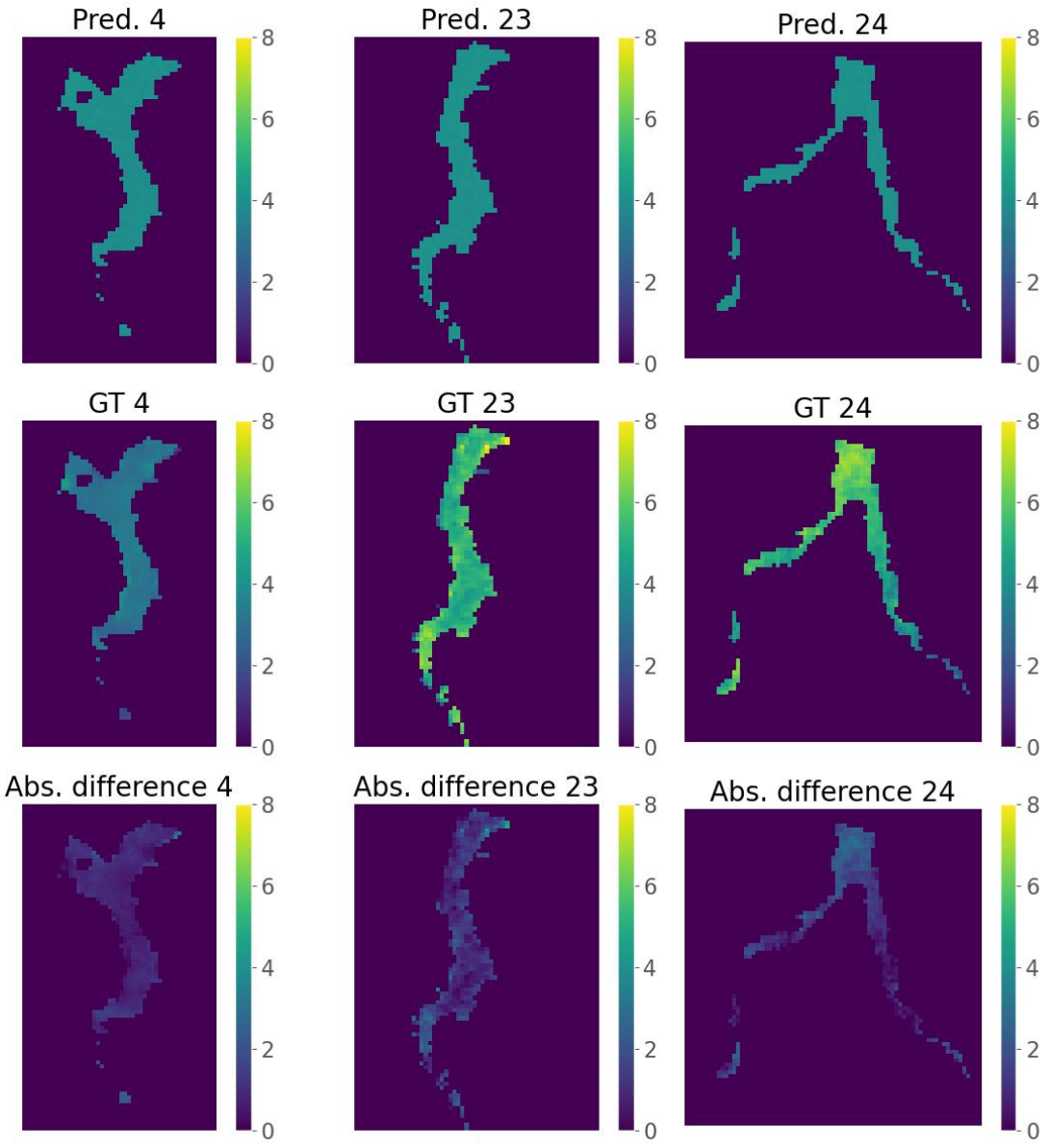

Figure 6.10: Predictions, Ground Truths and difference GT-Preds. of experiment GRU-8.

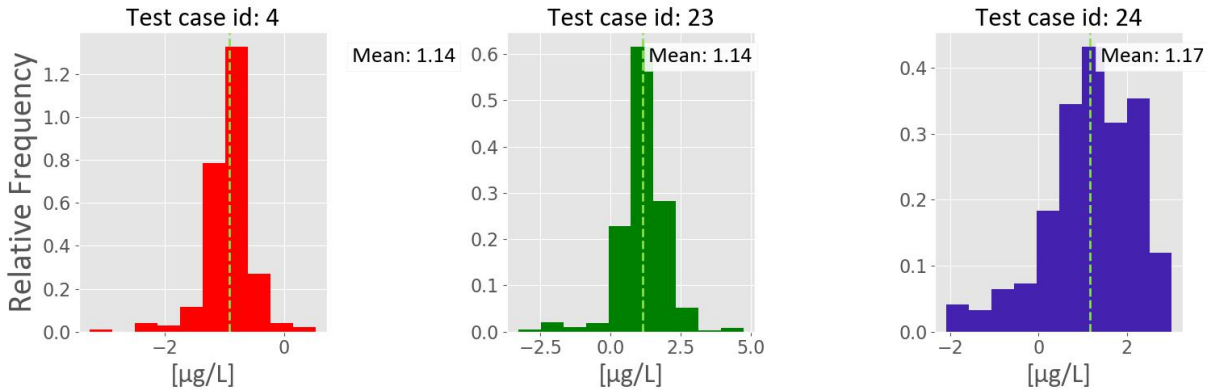

Figure 6.11: Distribution of the errors of experiment GRU-8.

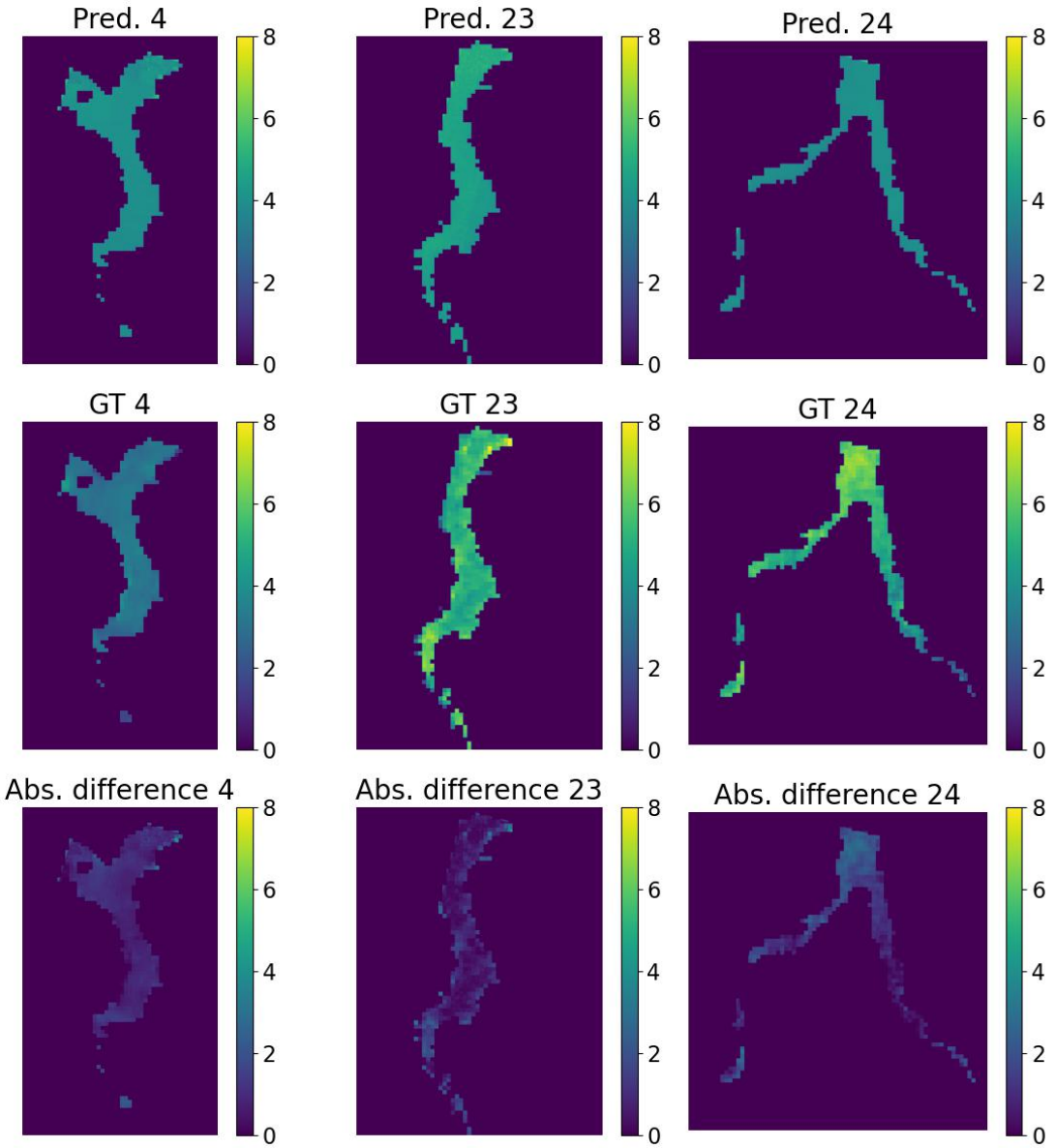

Figure 6.12: Predictions, ground truths and differences when exp. RF-10 was evaluated with 30-meter inputs.

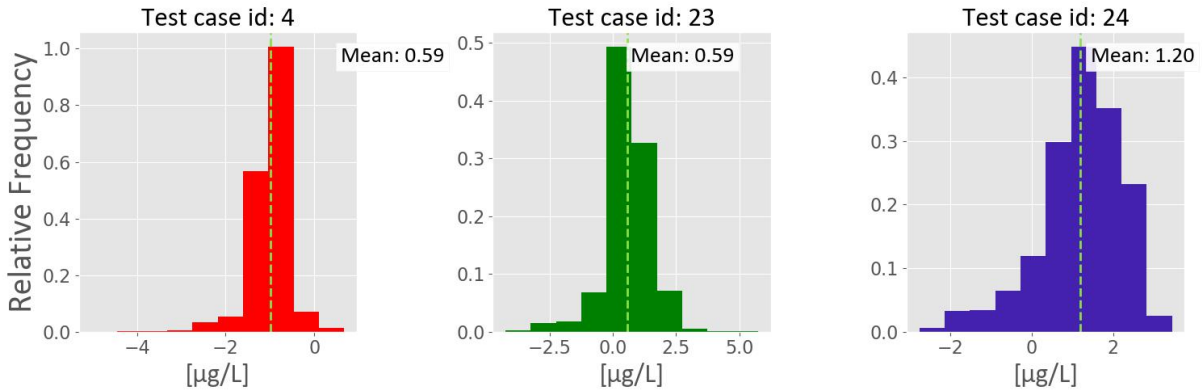

Figure 6.13: Distribution of the errors of experiment RF-10 when evaluated with 30 m inputs.

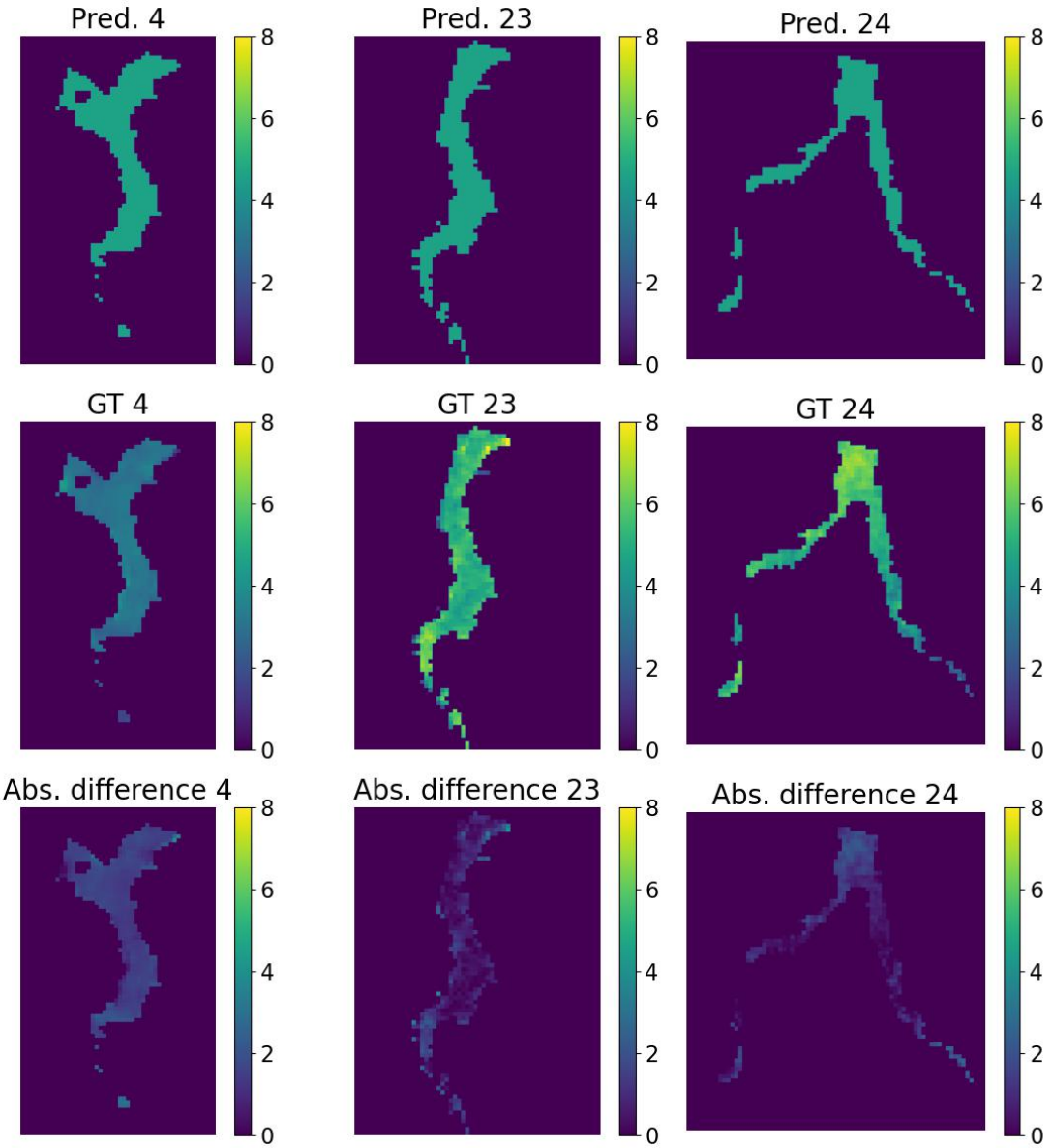

Figure 6.14: Predictions, ground truths and differences when exp. SVR-4 was evaluated with 30-meter inputs.

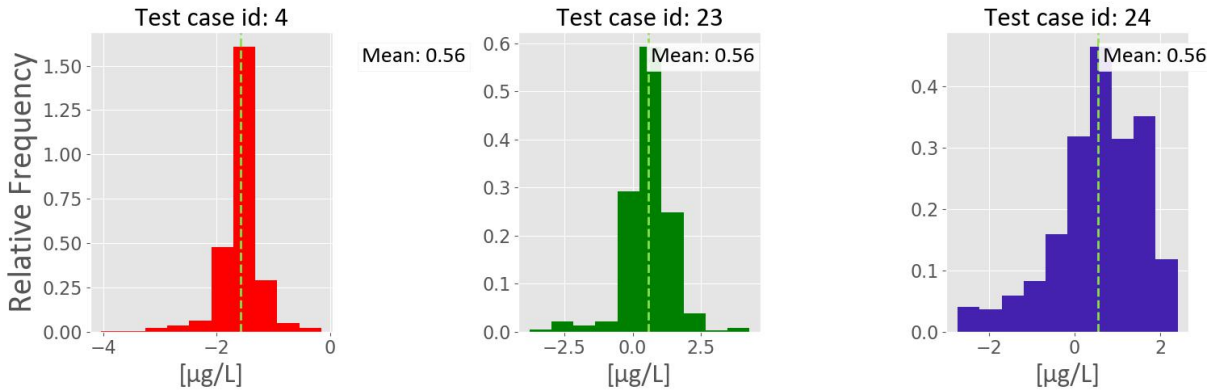

Figure 6.15: Distribution of the errors of experiment SVR-4 when evaluated with 30 m inputs.

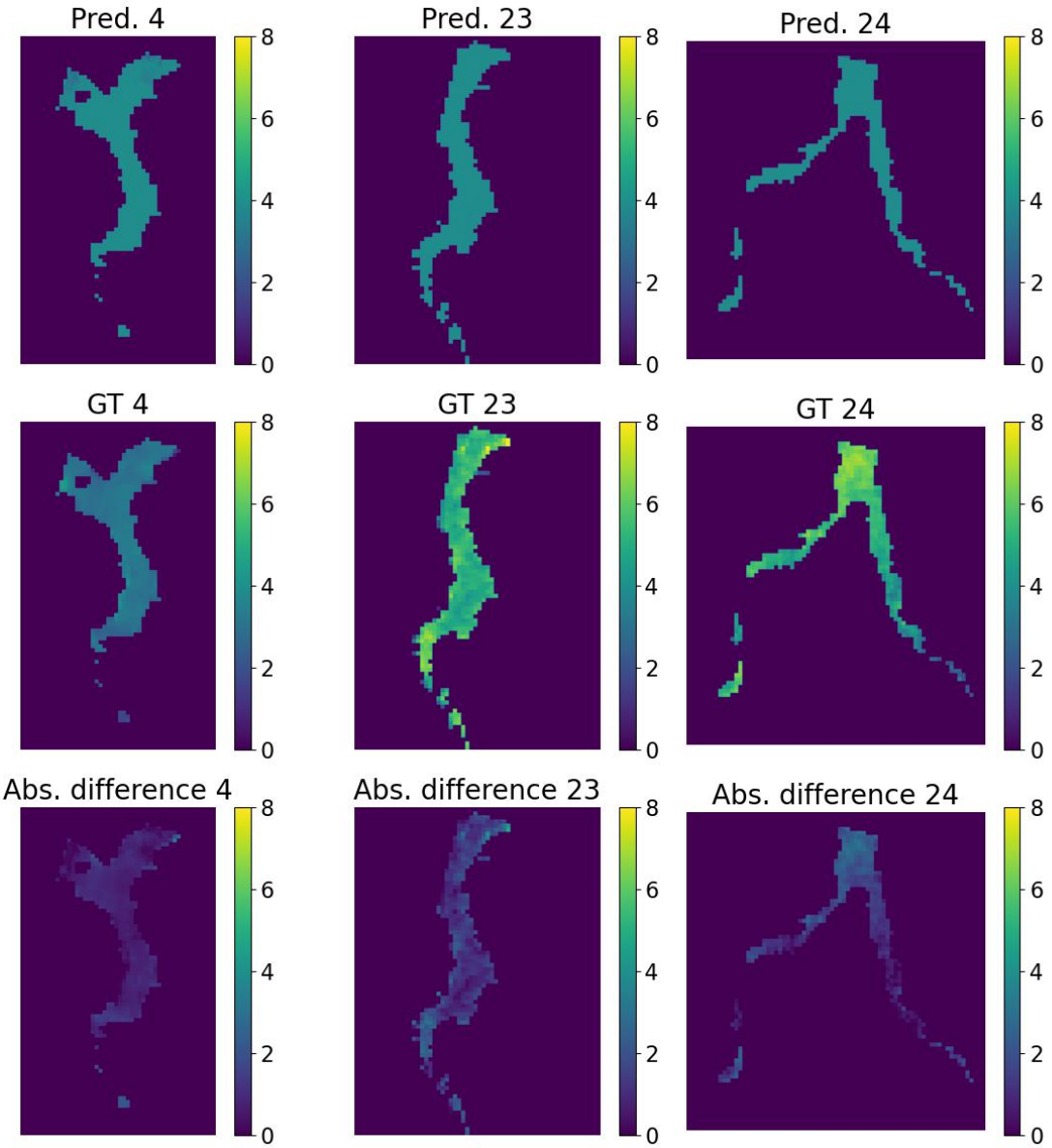

Figure 6.16: Predictions, ground truths and differences when exp. LSTM-13 was evaluated with 30-meter inputs.

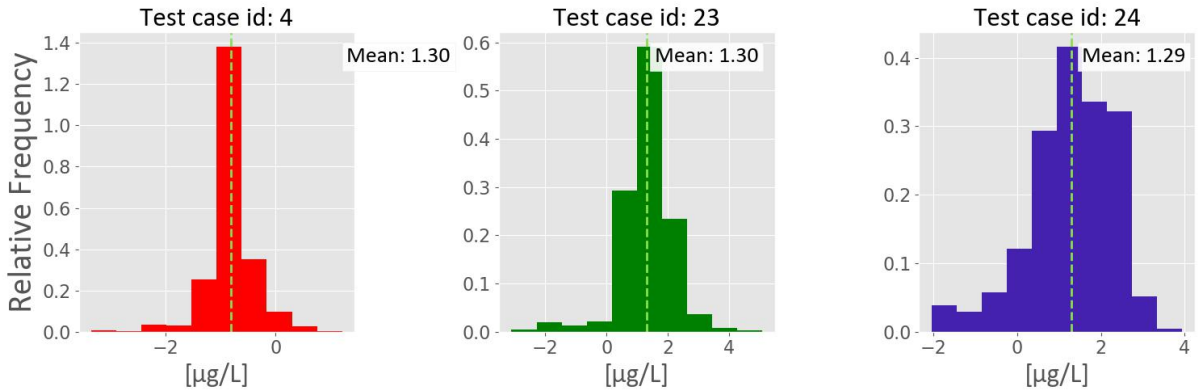

Figure 6.17: Distribution of the errors of experiment LSTM-13 when evaluated with 30 m inputs.

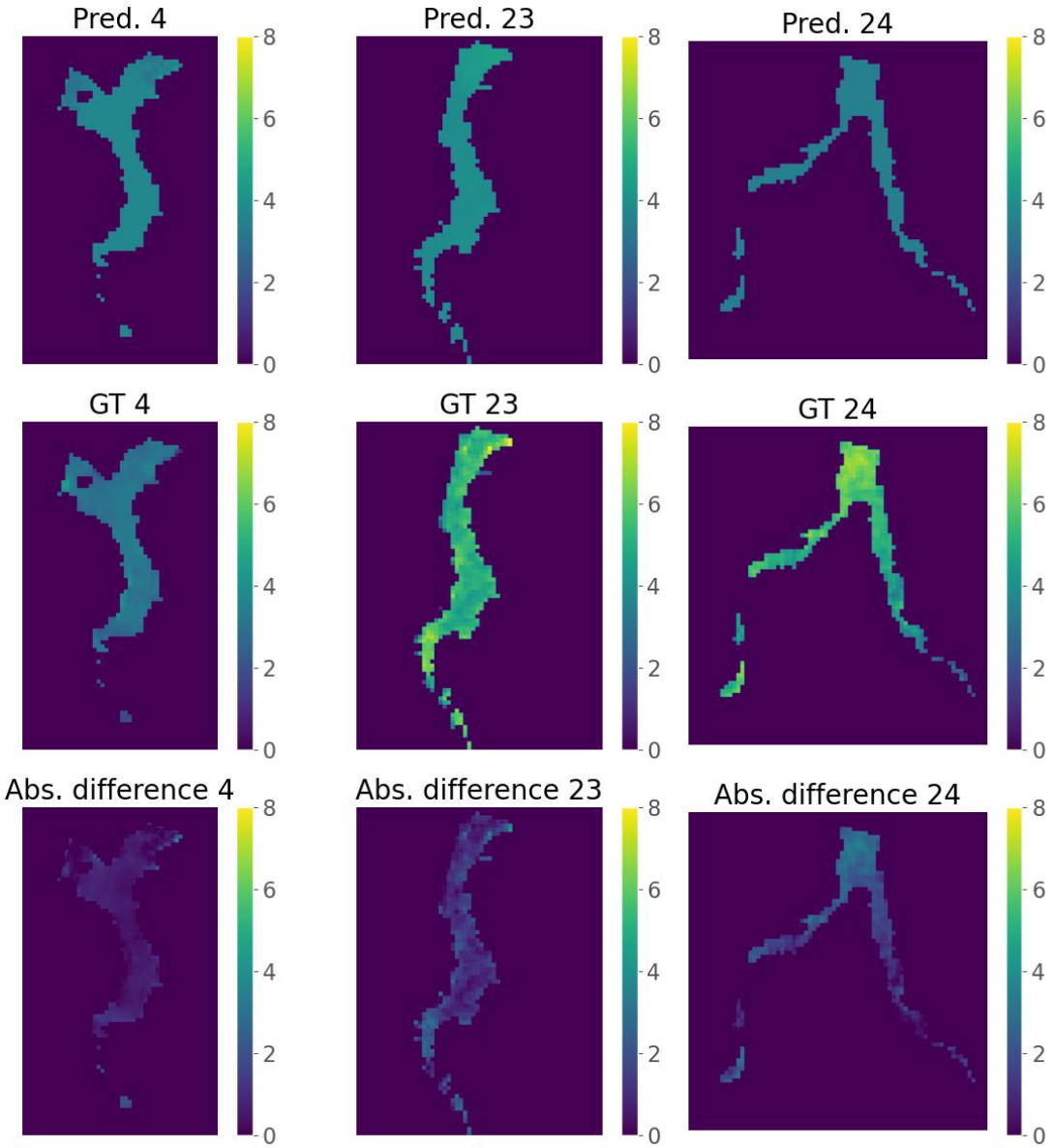

Figure 6.18: Predictions, ground truths and differences exp. when GRU-8 was evaluated with 30-meter inputs.

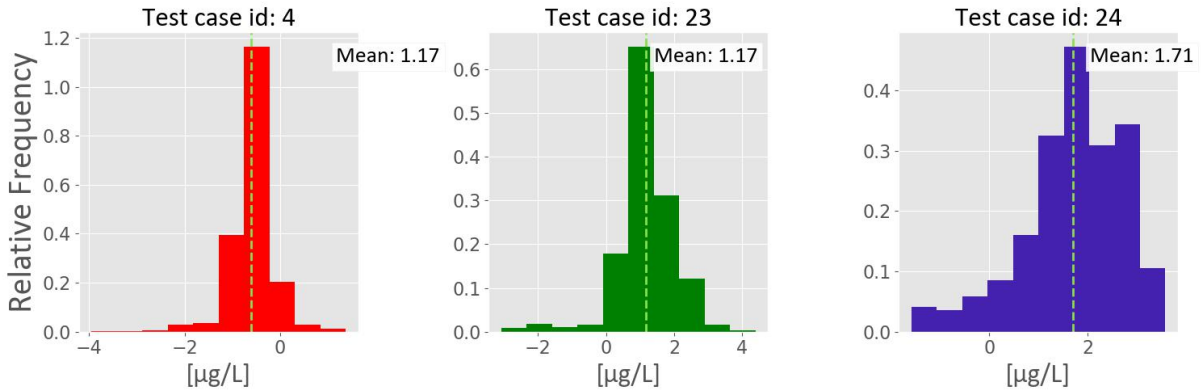

Figure 6.19: Distribution of the errors of experiment GRU-8 when evaluated with 30 m inputs.

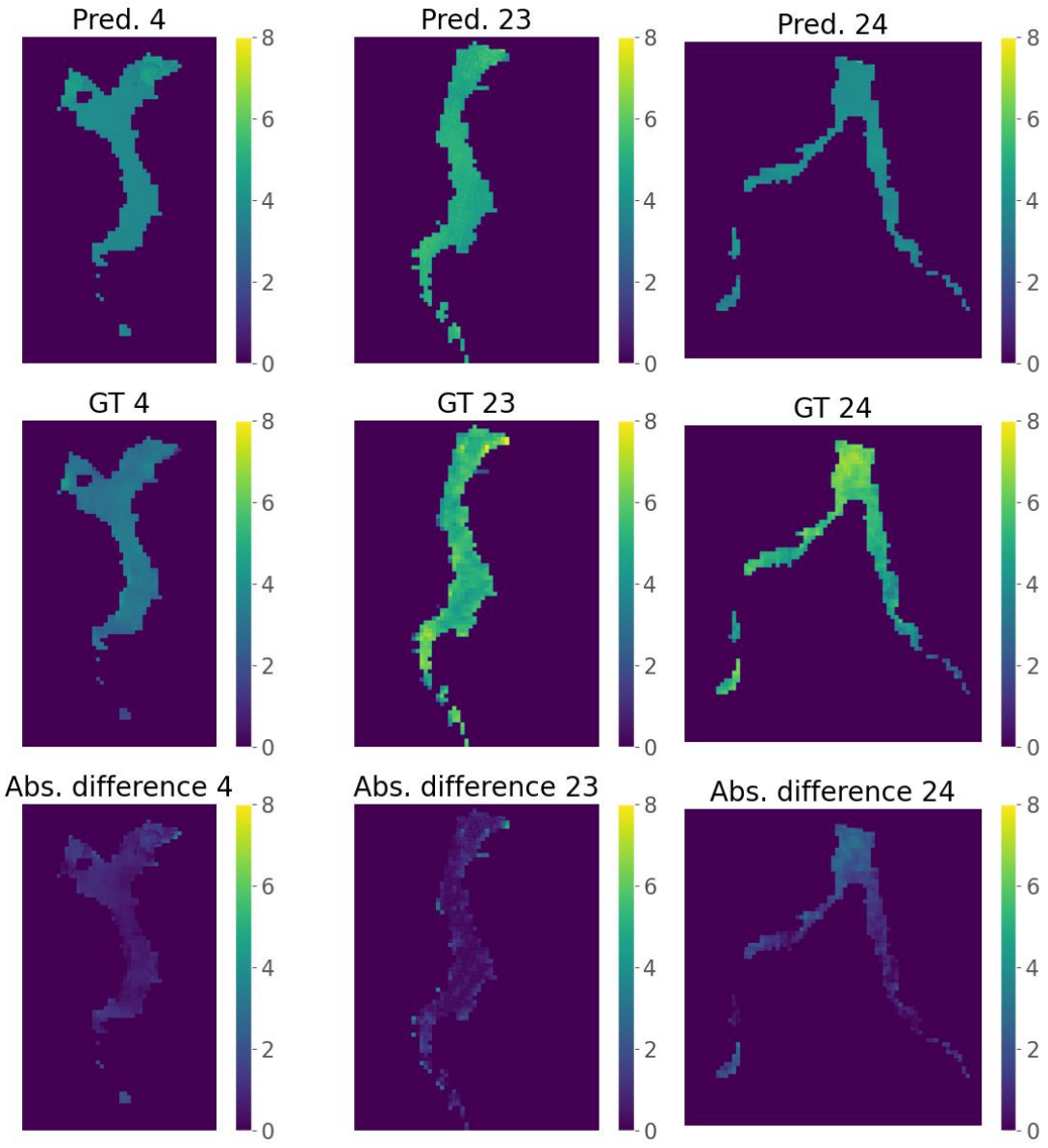

Figure 6.20: Predictions, ground truths and differences when exp. RF-12.

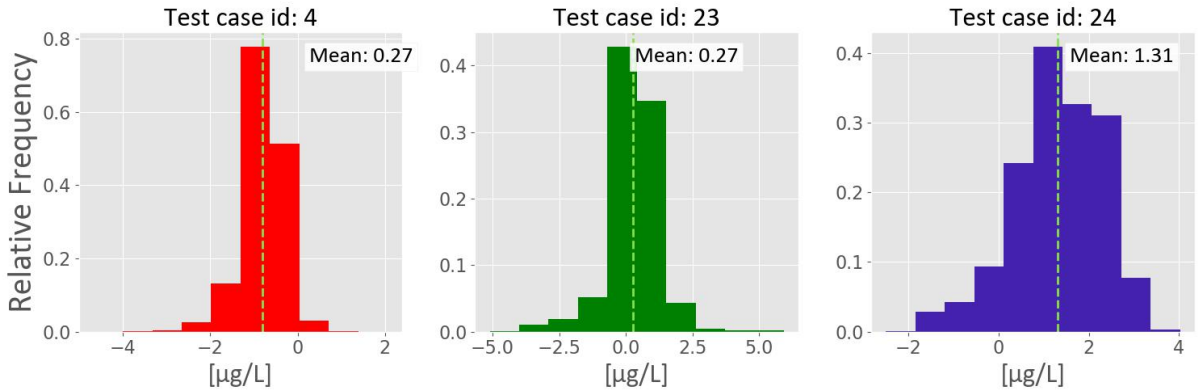

Figure 6.21: Distribution of the errors of experiment RF-12.

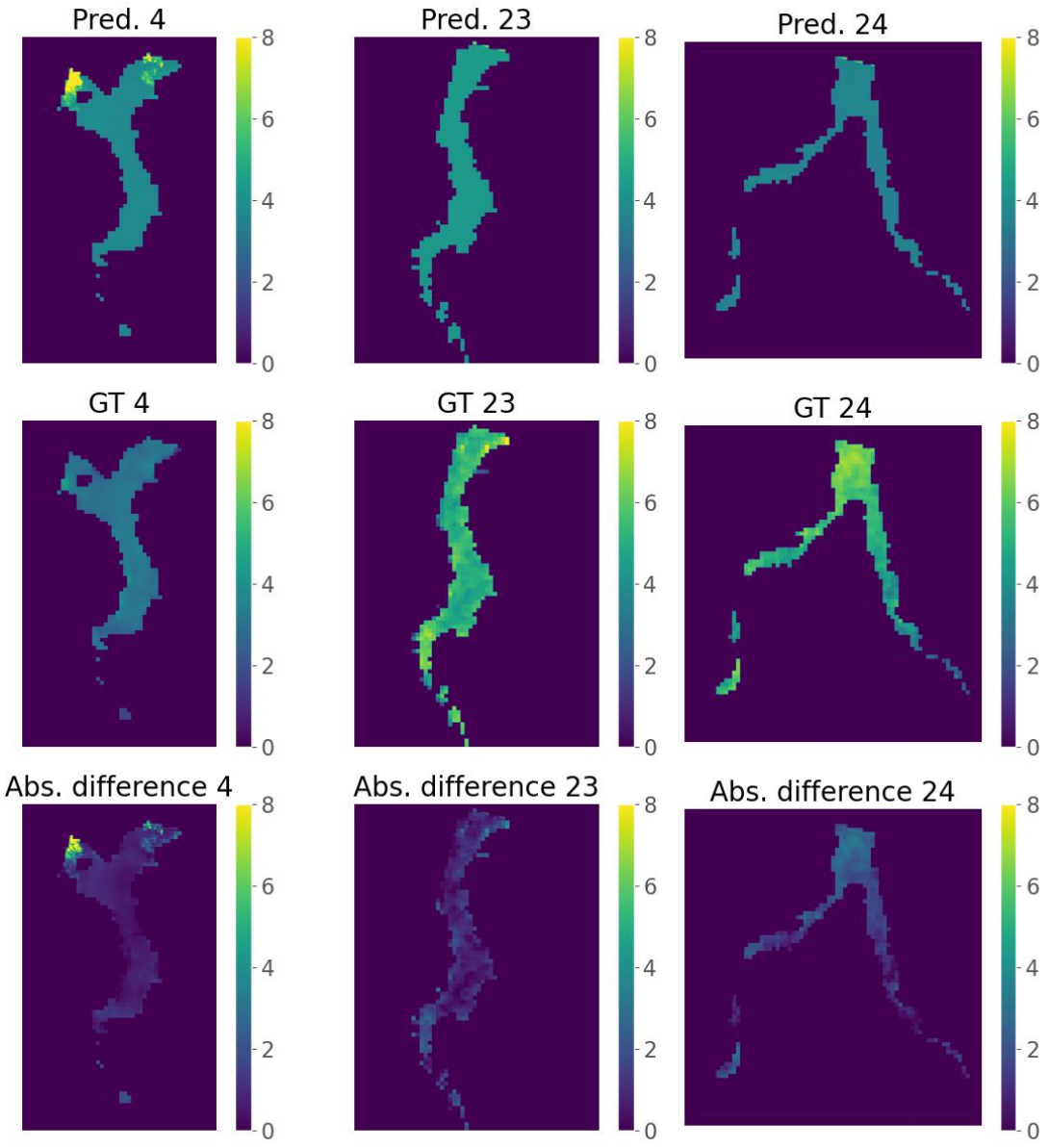

Figure 6.22: Predictions, ground truths and differences when exp. SVR-10.

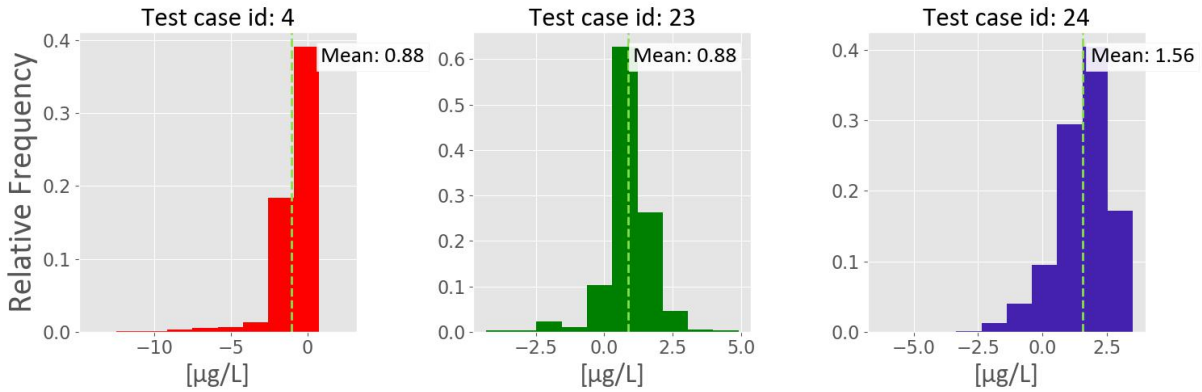

Figure 6.23: Distribution of the errors of experiment SVR-10.

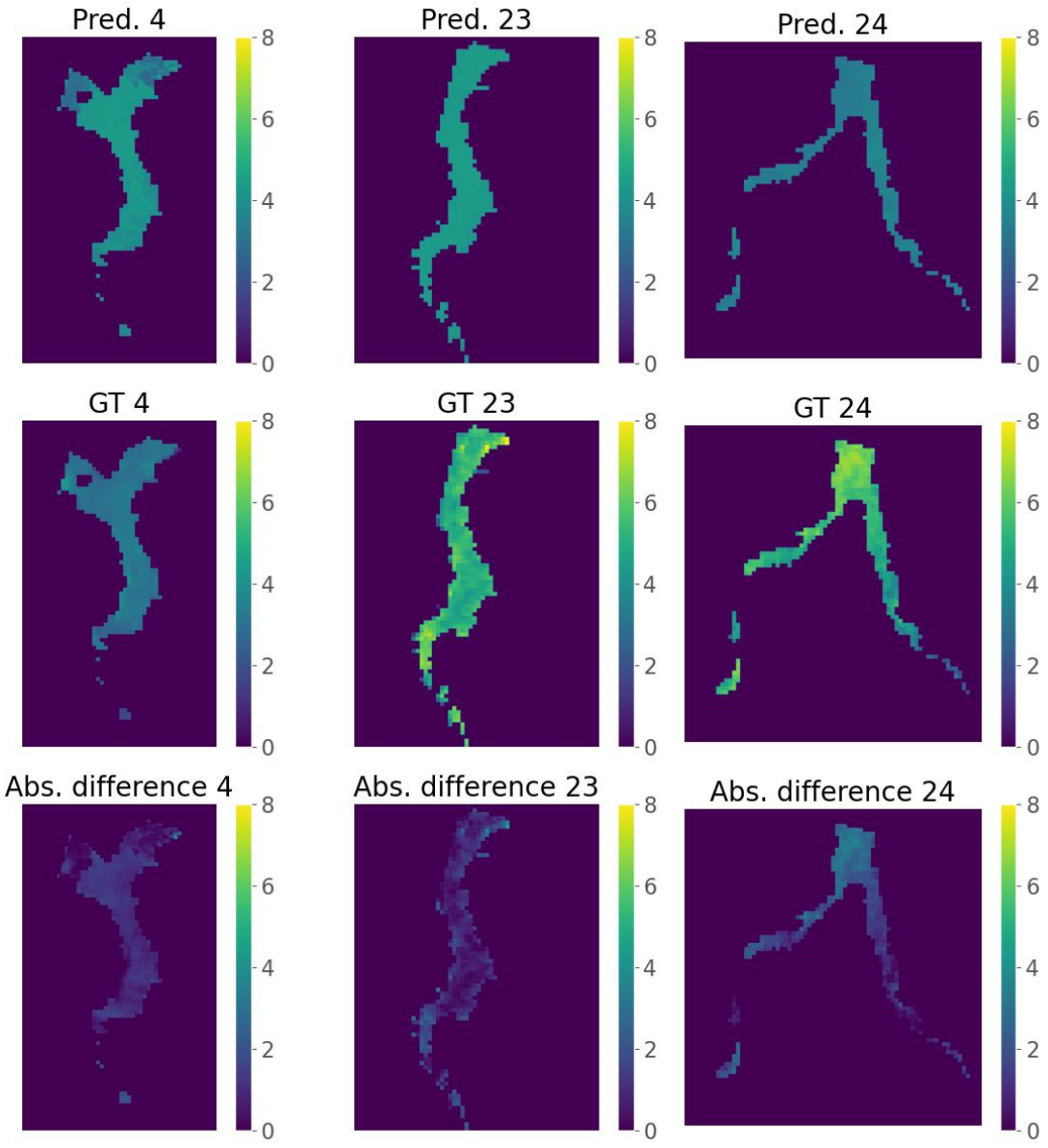

Figure 6.24: Predictions, ground truths and differences when exp. LSTM-14.

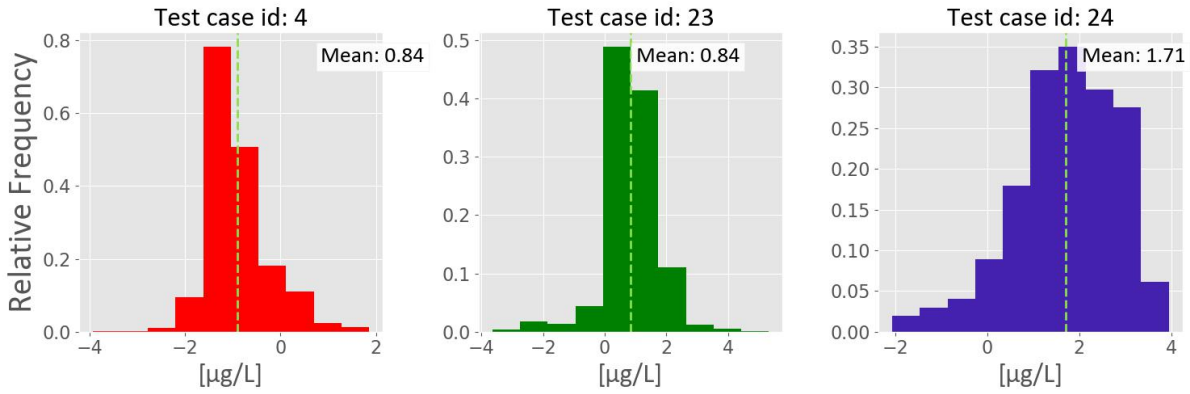

Figure 6.25: Distribution of the errors of experiment LSTM-14.

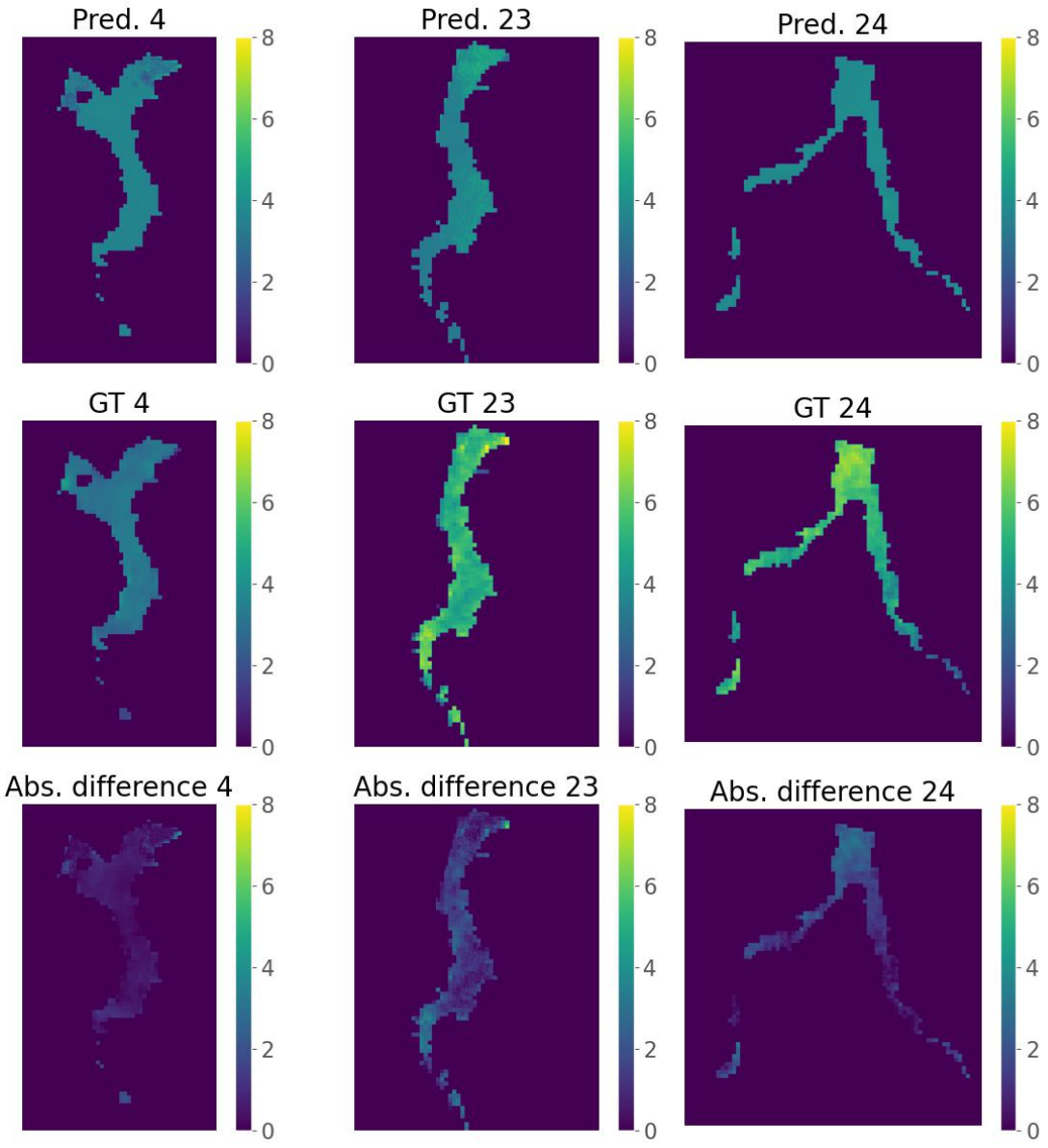

Figure 6.26: Predictions, ground truths and differences exp. when GRU-17.

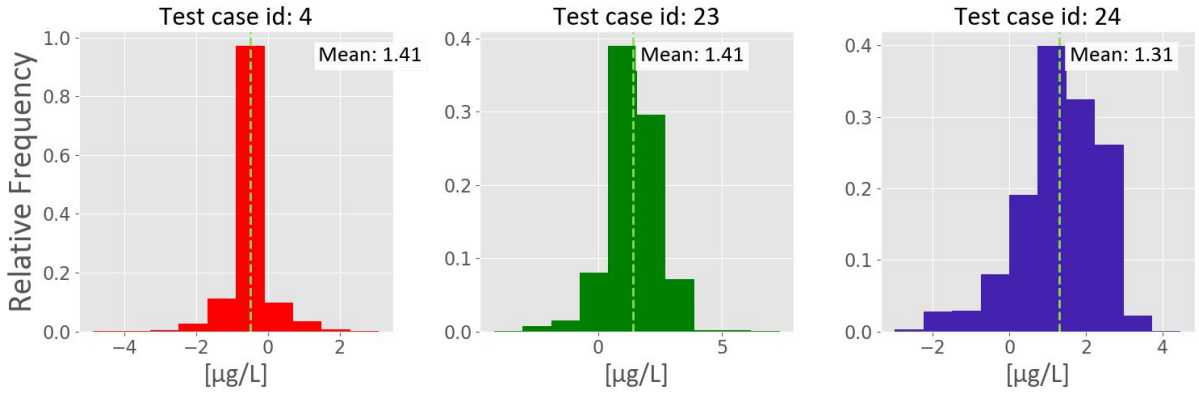

Figure 6.27: Distribution of the errors of experiment GRU-17.

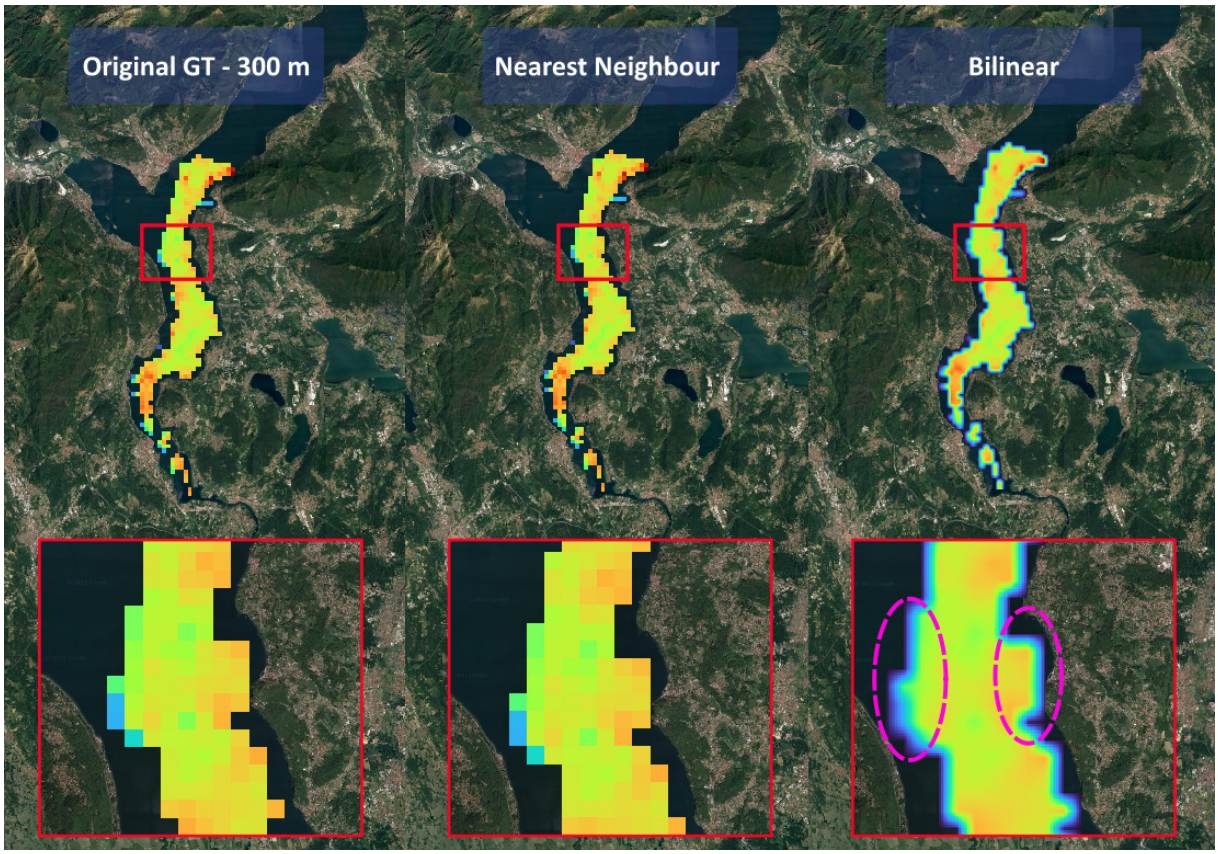

Figure 6.28: Visual comparison of nearest neighbour and bilinear upsampling methods.

7 | Conclusions

The project covered theoretical and practical aspects related to the modelling of Chlorophyll-a from hyperspectral imagery, developing a valuable tool for the monitoring of the Chlorophyll-a parameter in the lakes under the objective of this study.

In this sense, initially, the theoretical background was described to allow any reader to understand properly the theme of hyperspectral imagery, knowing deeply the relevance of the Chlorophyll-a parameter, the situation of the SIMILE project and the methods that allow the modelling of the hyperspectral information.

Then were analyzed the different cases that compound the initial and final datasets, explaining the criteria used for the selection of the final set.

Finally, the experimental development was completely described: All the manipulations on the inputs were specified, the different models employed were detailed, and the decisions that were taken to determine which was the most suitable model.

After choosing the best options for each model typology, a second evaluation was done but taking into account models trained with 300 meters GSD inputs and tested on 30-meter inputs. In this case, the model performances of the best models were decreased but the main insight was that the models trained with 300-meter inputs were able to at least provide an acceptable prediction when tested with 30-meter inputs.

Additionally, a comparative analysis between nearest neighbour and bilinear interpolation methods was performed to understand which of them was the most suitable approach to upsample the ground truth chlorophyll-a maps to 30-meter GSD. The results achieved allowed us to conclude that the nearest neighbour method was better than the bilinear interpolation one because the bilinear method produced undesired border effects and did

not improve the performance with respect to the nearest neighbour method.

Also, it is worth mentioning that among the four evaluated model typologies, it was determined that the traditional approaches (RF and SVR) always achieved higher performance than the deep learning methods (LSTM and GRU). Specifically, it was determined that the best alternative when testing with 300-meter inputs is the Support Vector Regressor but when testing with 30-meter inputs the best alternative considered was Random Forest Regressor.

However, in order to complement the previously mentioned analysis was performed a comparison by evaluating if the results achieved with those models trained with 300-meter inputs and evaluated on 30-meter inputs were achieving higher performance than the corresponding ones but trained-evaluated with 30-meter inputs. The conclusion of this comparison allowed us to determine that in two of the four model typologies was better to consider also 30-meter input data for training if the objective is to predict 30-meter chlorophyll-a maps because the final performances are better in this way. Among all the alternatives used for inferring 30-meter GSD chlorophyll-a maps, the best option was set in experiment RF-12 which used both for training and testing 30-meter input data.

Also, the following additional aspects could be remarked:

- By increasing the number of images available and improving their quality in terms of the considered parameters (mainly: clouds, and glint), a higher performance would be achieved with any of the four model typologies already analyzed in the document.
- The trained models achieved an acceptable performance and they could represent a valuable tool to monitor and mitigate any distortion in the water quality of the lakes of this study.
- Although the current study just analyzed a group of three lakes located in the shared basin between Switzerland and Italy (Lake Como, Lugano and Maggiore), it could be interesting to expand the study to different locations and evaluate if the models prepared are able to generalize and produce acceptable predictions for different lake conditions.

Bibliography

[1] E. S. Agency. Envisat meris full resolution - level 1 [mer_frs_1p/me_1_frg], 2023. URL https://earth.esa.int/eogateway/catalog/envisat-meris-full-resolution-full-swath-level-1-mer_frs_1p-.

[2] E. S. Agency. Envisat meris full resolution - level 2 [mer_frs_2p/me_2_frg], 2023. URL https://earth.esa.int/eogateway/catalog/envisat-meris-full-resolution-full-swath-level-2-mer_frs_2p-.

[3] I. S. Agency. *PRISMA Products Specification Document - Issue 2.3*. Italian Space Agency, 2020.

[4] I. S. Agency. *PRISMA Algorithm Theoretical Basis Document (ATBD)*. Italian Space Agency, 2021.

[5] ASI. *PRISMA product specification*. ASI, Italy, 2020.

[6] N. Audebert, B. L. Saux, and S. Lefèvre. Deep learning for classification of hyperspectral data: A comparative review. *CoRR*, abs/1904.10674, 2019. URL http://arxiv.org/abs/1904.10674.

[7] A. Austoni, J. F. Amieva, M. Bresciani, and M. A. Brovelli. Analysis of the remotely sensed water quality parameters of the insubric lakes: Methods and results of the interreg simile project. *International Symposium on Remote Sensing of Environment (ISRSE-39)*, 2023.

[8] J. Avbelj, H. Krawczyk, R. Müller, A. Makarau, M. Bachmann, V. Ziel, R. de los Reyes, K. A. G. Birgit Gerasch, and E. Carmona. *DESIS-PAV ATBD L1A, L1B, L1C, L2A Processors*. German Aerospace Center, 2015.

[9] L. Bourg, J. B. (ACRI-ST), H. Morris, and S. D. E. J. Dash (University of Southampton), R. Preusker (FUB). *Copernicus Sentinel-3 OLCI Land User Handbook*. European Space Agency (ESA), 2021.

[10] L. Breiman. Random forests. *Machine Learning*, 45(1):5–32, 2001.

[11] M. Bresciani, D. Stroppiana, D. Odermatt, G. Morabito, and C. Giardino. Assessing remotely sensed chlorophyll-a for the implementation of the water framework directive in european perialpine lakes. *Science of The Total Environment*, 409(17):3083–3091, 2011. ISSN 0048-9697. doi: https://doi.org/10.1016/j.scitotenv.2011.05.001.

[12] M. Bresciani, M. Adamo, G. De Carolis, E. Matta, G. Pasquariello, D. Vaičiūtė, and C. Giardino. Monitoring blooms and surface accumulation of cyanobacteria in the curonian lagoon by combining meris and asar data. *Remote Sensing of Environment*, 146:124–135, 2014. doi: https://doi.org/10.1016/j.rse.2013.07.040.

[13] C. Brockmann, R. Doerffer, M. Peters, S. Kerstin, S. Embacher, and A. Ruescas. Evolution of the C2RCC Neural Network for Sentinel 2 and 3 for the Retrieval of Ocean Colour Products in Normal and Extreme Optically Complex Waters. In *Living Planet Symposium*, volume 740 of *ESA Special Publication*, page 54, Aug. 2016.

[14] M. A. Brovelli, M. Cannata, and M. Rogora. Simile, a geospatial enabler of the monitoring of sustainable development goal 6 (ensure availability and sustainability of water for all). *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-4/W20:3–10, 2019. doi: 10.5194/isprs-archives-XLII-4-W20-3-2019.

[15] G. A. Center. Desis mission, 2022. URL `https://www.dlr.de/eoc/en/desktopdefault.aspx/tabid-13618/23664_read-54267/`.

[16] G. A. Center. Desis data access, 2022. URL `https://www.dlr.de/eoc/en/desktopdefault.aspx/tabid-13629/23675_read-54295/`.

[17] K. Cho, B. van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio. Learning phrase representations using RNN encoder-decoder for statistical machine translation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1724–1734. Association for Computational Linguistics, 2014.

[18] S. Cogliati, F. Sarti, L. Chiarantini, M. Cosi, R. Lorusso, E. Lopinto, F. Miglietta, L. Genesio, L. Guanter, A. Damm, S. Pérez-López, D. Scheffler, G. Tagliabue, C. Panigada, U. Rascher, T. Dowling, C. Giardino, and R. Colombo. The prisma imaging spectroscopy mission: overview and first performance analysis. *Remote Sensing of Environment*, 262:112499, 2021. ISSN 0034-4257. doi: https://doi.org/10.1016/j.rse.2021.112499. URL `https://www.sciencedirect.com/science/article/pii/S0034425721002170`.

[19] M. R. Corson, D. R. Korwan, R. L. Lucke, W. A. Snyder, and C. O. Davis. The

hyperspectral imager for the coastal ocean (hico) on the international space station. In *IGARSS 2008 - 2008 IEEE International Geoscience and Remote Sensing Symposium*, volume 4, pages IV – 101–IV – 104, 2008. doi: 10.1109/IGARSS.2008.4779666.

[20] D. R. Cutler, T. C. Edwards Jr, K. H. Beard, A. Cutler, K. T. Hess, J. Gibson, and J. J. Lawler. Random forests for classification in ecology. *Ecology*, 88(11):2783–2792, 2007.

[21] H. Drucker, D. Wu, and V. N. Vapnik. Support vector regression machines. *Advances in neural information processing systems*, 9:155–161, 1997.

[22] EnMAP. Data access - enmap, 2022. URL `https://www.enmap.org/data_access/`.

[23] EnMAP. Enmap successfully launched, 2022. URL `https://www.enmap.org/news/2022-04-04/`.

[24] EnMAP. Mission, 2022. URL `https://www.enmap.org/mission/`.

[25] E. EO-Portal. Hisui gds (ground data system), 2017. URL `https://www.eoportal.org/satellite-missions/iss-hisui#hisui-gds-ground-data-system`.

[26] ESA. Sentinel 3 - data access and products, 2015. URL `https://sentinels.copernicus.eu/documents/247904/1848151/Sentinel-3_OLCI_Data_Access_and_Products.pdf`.

[27] ESA. Going hyperspectral for chime, 2021. URL `https://www.esa.int/Applications/Observing_the_Earth/Copernicus/Going_hyperspectral_for_CHIME`.

[28] ESA. Chris, 2022. URL `https://earth.esa.int/eogateway/instruments/chris`.

[29] ESA. Facts and figures, 2022. URL `https://www.esa.int/Applications/Observing_the_Earth/FutureEO/FLEX/Facts_and_figures`.

[30] ESA. Ceos eo handbook – mission summary - gomx4, 2022. URL `http://database.eohandbook.com/database/missionsummary.aspx?missionID=917`.

[31] ESA. Ceos eo handbook – instrument summary - hyperscout, 2022. URL `http://database.eohandbook.com/database/instrumentsummary.aspx?instrumentID=1890`.

[32] ESA. Meris, 2022. URL `https://earth.esa.int/eogateway/instruments/meris`.

[33] ESA. Discover and download the earth observation data you need from the broad catalogue of missions the european space agency operate and support.,

2022. URL https://earth.esa.int/eogateway/search?text=&&category=Data&&filter=MERIS&&subFilter=Data.

[34] ESA. Ghgsat archive and tasking, 2022. URL https://earth.esa.int/eogateway/catalog/ghgsat-archive-and-tasking.

[35] ESA. Olci instrument payload, 2022. URL https://sentinel.esa.int/web/sentinel/missions/sentinel-3/instrument-payload/olci.

[36] ESA. Radiometric resolution - 21 bands in vis/swir, 2023. URL https://sentinels.copernicus.eu/web/sentinel/user-guides/sentinel-3-olci/resolutions/radiometric.

[37] F. A. Gers, J. Schmidhuber, and F. Cummins. Learning to forget: Continual prediction with lstm. *Neural computation*, 12(10):2451–2471, 2000.

[38] A. Gilerson, A. Gitelson, G. D. Zhou, J., W. Moses, I. Ioannou, and S. Ahmed. Algorithms for remote estimation of chlorophyll-a in coastal and inland waters using red and near infrared bands. *Optics Express*, 8(23):24109–24125, 2010. doi: https://doi.org/10.1364/OE.18.024109.

[39] R. C. Gonzalez and R. E. Woods. *Digital Image Processing*. Pearson, 4th edition, 2018.

[40] S. Hochreiter and J. Schmidhuber. Long short-term memory. *Neural computation*, 9 (8):1735–1780, 1997.

[41] A. S. Italiana. *Annual report 2020*. Agenzia Spaziale Italiana, 2020.

[42] G. D. Krebs. Gaofen 5, 5-02 (gf 5, 5-02), 2022. URL https://space.skyrocket.de/doc_sdat/gf-5.htm.

[43] G. D. Krebs. Gisat 1, 2 (eos 03, 05), 2022. URL https://space.skyrocket.de/doc_sdat/gisat-1.htm.

[44] T. M. A. d. Lima, C. Giardino, M. Bresciani, C. C. F. Barbosa, A. Fabbretto, A. Pellegrino, and F. N. Begliomini. Assessment of estimated phycocyanin and chlorophyll-a concentration from prisma and olci in brazilian inland waters: A comparison between semi-analytical and machine learning algorithms. *Remote Sensing*, 15(5), 2023. ISSN 2072-4292. doi: 10.3390/rs15051299. URL https://www.mdpi.com/2072-4292/15/5/1299.

[45] NASA. Product definitions, 2022. URL https://oceancolor.gsfc.nasa.gov/products/.

[46] NASA. Mission overview, 2022. URL https://oceancolor.gsfc.nasa.gov/hico/.

[47] NASA. Welcome to hyspiri mission study website, 2022. URL https://hyspiri.jpl.nasa.gov/.

[48] NASA. Ocean color instrument, 2022. URL https://pace.oceansciences.org/oci.htm.

[49] NASA. Timeline, 2022. URL https://pace.oceansciences.org/timeline.htm.

[50] V. G. Natale, A. Kafri, G. A. Tidhar, M. Chen, T. Feingersh, E. Sagi, A. Cisbani, M. Baroni, D. Labate, R. Nadler, A. Leizer, S. Signorile, P. Tempesta, M. L. Magliozzi, C. Catallo, and A. Pietropaolo. Shalom — space-borne hyperspectral applicative land and ocean mission. In *2013 5th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, pages 1–4, 2013. doi: 10.1109/WHISPERS.2013.8080667.

[51] E. R. Observation and S. E. Center. Usgs eros archive - earth observing one (eo-1) - ali, 2019. URL https://www.usgs.gov/centers/eros/science/usgs-eros-archive-earth-observing-one-eo-1-ali.

[52] P. L. PBC. Planet announces new details of hyperspectral offering, 2022. URL https://www.planet.com/pulse/planet-announces-new-details-of-hyperspectral-offering/.

[53] P. L. PBC. The hyperspectral constellation - reach new scales of sight, 2023. URL https://www.planet.com/products/hyperspectral/.

[54] K. Pearson. On lines and planes of closest fit to systems of points in space. *Philosophical Magazine*, 2(11):559–572, 1901. doi: 10.1080/14786440109462720.

[55] E. PORTAL-ESA. Tacsat-3, 2012. URL https://www.eoportal.org/satellite-missions/tacsat-3#eop-quick-facts-section.

[56] E. PORTAL-ESA. Hj-1, 2012. URL https://www.eoportal.org/satellite-missions/hj-1#wvc-wide-view-ccd-cameras.

[57] E. PORTAL-ESA. Ims-1, 2012. URL https://www.eoportal.org/satellite-missions/ims-1#spacecraft.

[58] E. PORTAL-ESA. Ghgsat-d, 2016. URL https://www.eoportal.org/satellite-missions/ghgsat-d#eop-quick-facts-section.

[59] E. PORTAL-ESA. Gomx-4, 2017. URL https://www.eoportal.org/satellite-missions/gomx-4#gomx-4-gomspace-express-4-mission.

[60] S.-E. Qian. Hyperspectral satellites, evolution, and development history. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14:7032–7056, 2021. doi: https://doi.org/10.1109/JSTARS.2021.3090256.

[61] C. Rogass, L. Guanter, C. Mielke, D. Scheffler, N. K. Boesche, C. Lubitz, M. Brell, D. Spengler, and K. Segl. An automated processing chain for the retrieval of georeferenced reflectance data from hyperspectral eo-1 hyperion acquisitions. 2014.

[62] M. Rogora, F. Buzzi, C. Dresti, B. Leoni, F. Lepori, R. Mosello, M. Patelli, and N. Salmaso. Climatic effects on vertical mixing and deep-water oxygen content in the subalpine lakes in italy. *Hydrobiologia*, 824, Apr. 2018. doi: https://doi.org/10.1007/s10750-018-3623-y.

[63] Scikit-Learn. Randomforestregressor, 2023. URL `https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestRegressor.html`.

[64] Scikit-Learn. Support vector machines, 2023. URL `https://scikit-learn.org/stable/modules/svm.html`.

[65] A. J. Smola and B. Schölkopf. A tutorial on support vector regression. *Statistics and computing*, 14(3):199–222, 2004.

[66] I. Sobel. An isotropic 3x3 image gradient operator. In *Proceedings of the 5th annual symposium on theory of computing*, pages 271–272. ACM, 1973.

[67] N. Srivastava, G. E. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov. Dropout: a simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15(1):1929–1958, 2014. URL `http://www.cs.toronto.edu/~rsalakhu/papers/srivastava14a.pdf`.

[68] J. S. Systems. Hisui - hyper-spectral imager suite, 2021. URL `https://www.jspacesystems.or.jp/en/project/observation/hisui/`.

[69] J. S. Systems. Hisui hyperspectral imager suite, 2021. URL `https://www.jspacesystems.or.jp/en/project/observation/hisui`.

[70] J. F. Toro Herrera, D. Carrion, M. Bresciani, and G. Bratić. Semi-automated production and filtering of satellite derived water quality parameters. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLIII-B3-2022:1019–1026, 2022. doi: 10.5194/isprs-archives-XLIII-B3-2022-1019-2022.

[71] E. Vangi, G. D'Amico, S. Francini, F. Giannetti, B. Lasserre, M. Marchetti, and

G. Chirici. The new hyperspectral satellite prisma: Imagery for forest types discrimination. *Sensors*, 21(4), 2021. ISSN 1424-8220. doi: 10.3390/s21041182. URL https://www.mdpi.com/1424-8220/21/4/1182.

[72] R. Vollenweider and J. Kerekes. *Eutrophication of waters: monitoring, assessment and control*. OECD, 1982.

[73] R. Wetzel and G. Likens. *Limnological analyses*. Springer Science Business Media, 2000.

[74] C. J. Willmott and K. Matsuura. Advantages of the mean absolute error (mae) over the root mean square error (rmse) in assessing average model performance. *Climate research*, 30(1):79–82, 2005.