

Organisation NIVA  
Department Oceanography

## D9.4

Report on remote sensing matchups with *in situ* observations during  
Norwegian and Aegean Sea demonstration

Date: 01/07/2025  
Doc. Version: V1  
doi: [10.5281/zenodo.18245045](https://doi.org/10.5281/zenodo.18245045)

### Document Control Information

Settings	Value
<b>Deliverable Title</b>	Report on remote sensing matchups with <i>in situ</i> observations during Norwegian and Aegean Sea demonstration
<b>Work Package Title</b>	Data modelling and data sharing
<b>Deliverable number</b>	D9.4
<b>Description</b>	Report on matchups between NAUTILOS' instruments and Satellites and development of marine litters detection algorithm.
<b>Lead Beneficiary</b>	
<b>Lead Authors</b>	Sabine Marty (NIVA), Manolis Ntoumas (HCMR), Natalia Stamataki (HCMR), Elmar Berghöfer (DFKI), Daniel Lukats (DFKI), Andrew King (NIVA)
<b>Contributors</b>	Felix Becker (DFKI), Eike Rodenbäck (DFKI), Daniel Lukats (DFKI), Zacharias Kapelonis (HCMR), Manolis Pettas (HCMR)
<b>Submitted by</b>	Sabine Marty
<b>Doc. Version (Revision number)</b>	V1
<b>Sensitivity (Security):</b>	Public, Basic, High
<b>Date:</b>	01/07/2025

### Document Approver(s) and Reviewer(s):

NOTE: All Approvers are required. Records of each approver must be maintained. All Reviewers in the list are considered required unless explicitly listed as Optional.

Name	Role	Action	Date
<b>Gabriele Pietri, Caterina Lemos</b>	Review #1	<i>skeleton approved</i>	04/06/2025
<b>Kostas Tsiaras</b>	Review #2	approved	30/06/2025

### Document history:

The Document Author is authorized to make the following types of changes to the document without requiring that the document be re-approved:

- Editorial, formatting, and spelling
- Clarification

To request a change to this document, contact the Document Author or Owner.

Changes to this document are summarized in the following table in reverse chronological order (latest version first).

Revision	Date	Created by	Short Description of Changes
<b>V0</b>	01/06/2025	Sabine Marty	Skeleton created
<b>V1</b>	27/06/2025	Sabine Marty	contribution from all partners added

### Configuration Management: Document Location

The latest version of this controlled document is stored in <location>.

Nature of the deliverable		
<b>R</b>	Report	X
<b>DEC</b>	Websites, patents, filing, etc.	
<b>DEM</b>	Demonstrator	
<b>O</b>	Other	

Dissemination level		
<b>PU</b>	Public	X
<b>CO</b>	Confidential, only for members of the consortium (including the Commission Services)	

### ACKNOWLEDGEMENT

This report forms part of the deliverables from the NAUTILOS project which has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101000825. The Community is not responsible for any use that might be made of the content of this publication.

NAUTILOS - New Approach to Underwater Technologies for Innovative, Low-cost Ocean observation is an H2020 project funded under the Future of Seas and Oceans Flagship Initiative, coordinated by the National Research Council of Italy (CNR, Consiglio Nazionale delle Ricerche). It brings together a group of 21 entities from 11 European countries with multidisciplinary expertise ranging from ocean instrumentation development and integration, ocean sensing and sampling instrumentation, data processing, modelling and control, operational oceanography and biology and ecosystems and biogeochemistry such, water and climate change science, technological marine applications and research infrastructures.

NAUTILOS will fill-in marine observation and modelling gaps for chemical, biological and deep ocean physics variables through the development of a new generation of cost-effective sensors and samplers, the integration of the aforementioned technologies within observing platforms and their deployment in large-scale demonstrations in European seas. The fundamental aim of the project will be to complement and expand current European observation tools and services, to obtain a collection of data at a much higher spatial resolution, temporal regularity and length than currently available at the European scale, and to further enable and democratise the monitoring of the marine environment to both traditional and non-traditional data users.

NAUTILOS is one of two projects included in the EU's efforts to support the European Strategy for Plastics in a Circular Economy by supporting the demonstration of new and innovative technologies to measure the Essential Ocean Variables (EOV).

More information on the project can be found at: <http://www.nautilus-h2020.eu>.

**COPYRIGHT**

© NAUTILOS Consortium. Copies of this publication – also of extracts thereof – may only be made with reference to the publisher.

## TABLE OF CONTENTS

<b>ACKNOWLEDGEMENT</b> .....	<b>3</b>
<b>COPYRIGHT</b> .....	<b>4</b>
<b>TABLE OF CONTENTS</b> .....	<b>5</b>
<b>EXECUTIVE SUMMARY</b> .....	<b>6</b>
<b>LIST OF FIGURES</b> .....	<b>7</b>
<b>LIST OF TABLES</b> .....	<b>7</b>
<b>LIST OF ACRONYMS AND ABBREVIATIONS</b> .....	<b>8</b>
<b>I. INTRODUCTION</b> .....	<b>9</b>
<b>II. METHODS</b> .....	<b>9</b>
1. Matchups .....	9
1.1. Field data.....	9
a. IR sensor.....	9
b. Multispectral and Hyperspectral Cameras and LIF-LIDAR.....	10
1.2. Satellite data .....	10
a. SST.....	10
b. Ocean color .....	11
2. Marine litter detection.....	12
2.1. Field data.....	12
2.2. AI image detection .....	14
<b>III. RESULTS</b> .....	<b>14</b>
1. Matchups .....	14
1.1. SST (Aegean Sea).....	14
1.2. Ocean colour (Norwegian Sea) .....	16
2. Marine litter detection.....	20
<b>IV. SUMMARY</b> .....	<b>23</b>
<b>V. APPENDIX 1: REFERENCES AND RELATED DOCUMENTS</b> .....	<b>25</b>

## EXECUTIVE SUMMARY

Deliverable D9.4 describes the activities carried out in WP9 Task 9.4, which is concerned with improvements on remote sensing algorithms and observations. This deliverable explores how NAUTILOS-developed downward-looking sensors, including multi- and hyperspectral cameras, an infrared sea surface temperature sensor, and a Light-induced Fluorescence Light Detection And Ranging instrument on remote sensing can enhance validation and fill gaps in observations of ocean variables like chlorophyll-a (Chl-a), turbidity, sea surface temperature (SST) and litter detection. Data from demonstrations in the Aegean Sea for SST and the North Sea (Oslofjord) for Chl-a, turbidity, and plastic litter detection are analysed and presented in the context of matchups with remote sensing products. Recommendations are provided for how cost-effective sensors, observing platforms, and algorithms can be used to fill crucial observational gaps.

## LIST OF FIGURES

- Figure 1 : Screenshot from the Copernicus data portal ([data.marine.copernicus.eu](https://data.marine.copernicus.eu))
- Figure 2: RGB composites of L8 OLI, S3B OLCI and S2A MSI scenes from passes over Oslofjord on 05/09/25 for L8 OLI and S3B OLCI and the 06/09/25 for S2A MSI, after atmospheric corrections.
- Figure 3: The difference in temperature between raw (blue) and corrected (orange) IR data and satellite data
- Figure 4: Comparison statistics for IR and SAT temperature data
- Figure 5: Turbidity measured by L8 OLI, S3 OLCI and S2 MSI (from left to right) from the same scenes shown in Fig. 2.
- Figure 6: Chl-a estimation from OC2, OC3, Mishra (from left to right) from the same S2 scene shown in Fig. 2.
- Figure 7: Top panel: Matchups between S2A/B MSI and UAV-multispectral camera Rrs. Middle panel: Matchups between L8/9 OLI and UAV-multispectral camera Rrs. Bottom panel: Matchups between S3A/B OLCI and UAV-multispectral camera Rrs. Matchup times are offset by ~13-1314 minutes depending on satellite matchup data used.
- Figure 8: Object detection with the MaskRCNN model on drone imagery from a beach in Norway.
- Figure 9: Example of marine litter detection based on citizen science beach images.
- Figure 10: F1 scores of the shape-based litter detection on RGB images with different GSD.
- Figure 11: Visualization of plastic detection using pixel-wise multispectral image analysis. Left: Composite RGB image generated from the multispectral camera. Center: Result of pixel-wise plastic detection based on multispectral data. Right: Overlay of the RGB image with the plastic detection map, highlighting identified plastic regions.
- Figure 12: Accuracy and F1 scores of the pixel-based litter detection on multispectral images and RGB images with different GSD
- Figure 13: Pixels displaying litter at the original resolution (left), GSD=100 mm (center) and GSD=500 mm (right). Yellow pixels are litter, whereas purple pixels indicate environment.

## LIST OF TABLES

Table 1: Chl-a ( $\mu\text{g Chl-a L}^{-1}$ ) as measured by UAV-multispectral camera, FerryBox fluorometer (FB), Landsat8 (L8) OC3 product, Sentinel 2 (S2) OC2 product, Sentinel 3 (S3) OC4 product, LIF-LIDAR, CT -fluorometer deployed from a research vessel (CTD flu), and water samples collected from an autonomous surface vehicle and rigid-inflatable boat and manually analysed via extraction/spectrophotometry.

## LIST OF ACRONYMS AND ABBREVIATIONS

Abbreviation	Definition
AI	Artificial Intelligence
ASV	Autonomous Surface Vehicle
cDOM	Coloured Dissolved Organic Matter
Chl-a	Chlorophyll-a
CMEMS	Copernicus Marine Environment Monitoring Service
CTD	Conductivity Temperature Depth
GSD	Ground Sampling Distance
IR	InfraRed
LIF LIDAR	Laser Induced Fluorescence Light Detection And Ranging
MED	Mediterranean Sea
MFC	Monitoring and Forecasting Center
MSI	Multi Spectral Imager
OLCI	Ocean and Land Colour Instrument
OLI	Operational Land Imager
QC	Quality Control
R/V	Research Vessel
Rrs	Remote Sensing Reflectance
SST	Sea Surface Temperature
UAV	Unmanned Aerial Vehicles

## I. INTRODUCTION

This Deliverable reports on Task 9.4, which focused on assessing how instruments developed in NAUTILOS WP3 and WP4 can contribute to remote sensing matchups between *in situ* observations and remotely sensed observations, obtained from downward-looking sensors (integrated on research vessels, autonomous surface vehicles (ASVs), and unmanned aerial vehicles (UAVs)) and satellite remotely sensed data (e.g., Sentinel 2/3). The overall objective of Task 9.4 was to assess how new observations supported by instruments developed in NAUTILOS can improve remote sensing products by providing better and novel ground-truth data for algorithms used for SST, ocean colour (Chl-a, turbidity, cDOM), and marine litter. These instruments include an infrared SST sensor, hyper- and multi-spectral cameras for SST, ocean colour, and marine litter, and a Laser-Induced Fluorescence Light Detection And Ranging (LIF-LIDAR) for Chl-a and cDOM detection. Additionally, an artificial intelligence model was evaluated for marine litter detection on a multispectral image taken with a UAV-based camera system and on images taken from citizen science contributions. Furthermore, the impact of different Ground Sampling Distances was investigated using augmented versions of the multispectral image. The regional coverage of the matchups includes the Norwegian Sea and Aegean Sea demonstrations, carried out in WP7, and introductory/background aspects of the demonstrations can be found in Deliverable 7.3.

## II. METHODS

### 1. MATCHUPS

#### 1.1. Field data

The following NAUTILOS instruments have collected *in situ* data related to ocean parameters that are potentially observable by remote sensing from space.

##### *a. IR sensor*

As part of NAUTILOS Sub-Task 5.3.2, which focuses on the integration of sensors and samplers onto FerryBox ships of opportunity, the intended demonstration platform for the infrared sea surface temperature (SST) sensor was the POSEIDON-HCMR FerryBox system. However, due to the unavailability of the FerryBox system during the demonstration period, the SST sensor was instead installed on the HCMR research vessel R/V *Philia*. Integration between the sensor and the hosting platform was successfully tested in May 2024. Additionally, calibration and validation activities (T6.1.2) were conducted at NIVA's field station, with results and methodology documented in Deliverable D6.1: Report on the results and methodology of the calibration/validation experiments. The infrared SST sensor was deployed on the R/V *Philia* from June 2024 to November 2025, during which it operated for 30 days across the North Aegean and Ionian Seas, collecting over 250,000 measurements along 2,000+ nautical miles. Data were gathered during two main periods: 10 June–24 July 2024 (North Aegean) and 3 September–13 October 2024 (Ionian Sea). For the validation of the data, a SeaBird CTD 19 plus was used during the ship surveys. The quality control (QC) and the processing of the data are reported in D.7.3.

*b. Multispectral and Hyperspectral Cameras and LIF-LIDAR*

Downward looking sensors developed in T3.2 were demonstrated in T7.2 at NIVA’s field station in Solbergstrand (approximately 40 km south of Oslo, Norway; 59.62 °N, 10.65 °E) in the Oslofjord, where the MS Color Fantasy FerryBox operates. Two commercial off-the-shelf (COTS) sensors from T.3.2 were used: a Specim AFX10 hyperspectral camera, a Micasense Altum PT multispectral camera. The multispectral and hyperspectral cameras were integrated with two different multirotor UAVs, a DJI Matrice 300 RTK for multispectral camera deployments, and a DJI Matrice 600 Pro for the deployments of the hyperspectral camera. The processing of the camera’s data included calibration, georeferencing and application of ocean colour algorithms to retrieve Turbidity/TSM, Chl-a and cDOM concentrations. Additionally, a LIF-LIDAR developed in T3.2 was demonstrated in the same area deployed on a Martine Robotics Otter Pro ASV (autonomous surface vehicle). A detailed description of the instruments is included in D3.3, D5.3, and D6.1, and a description of the camera-UAV and LIF-LIDAR-ASV setups and full data analysis process from the demonstrations is described in D7.3. The data collection occurred in synchronisation with FerryBox and satellite passes in May 2023, July 2025, and September 2024. Data collected from the middle of the fjord, along the MS Color Fantasy FerryBox route (described in D7.2), were selected to match the satellite overpasses from the drone transects across the fjord. The pixel values of a window of 13\*13 pixels were averaged to be compared to satellite data. Coefficients from matchups between Ferrybox and UAV found in D7.3 were applied to UAV data to retrieve Chl-a concentration from Reflectances.

1.2. Satellite data

*a. SST*

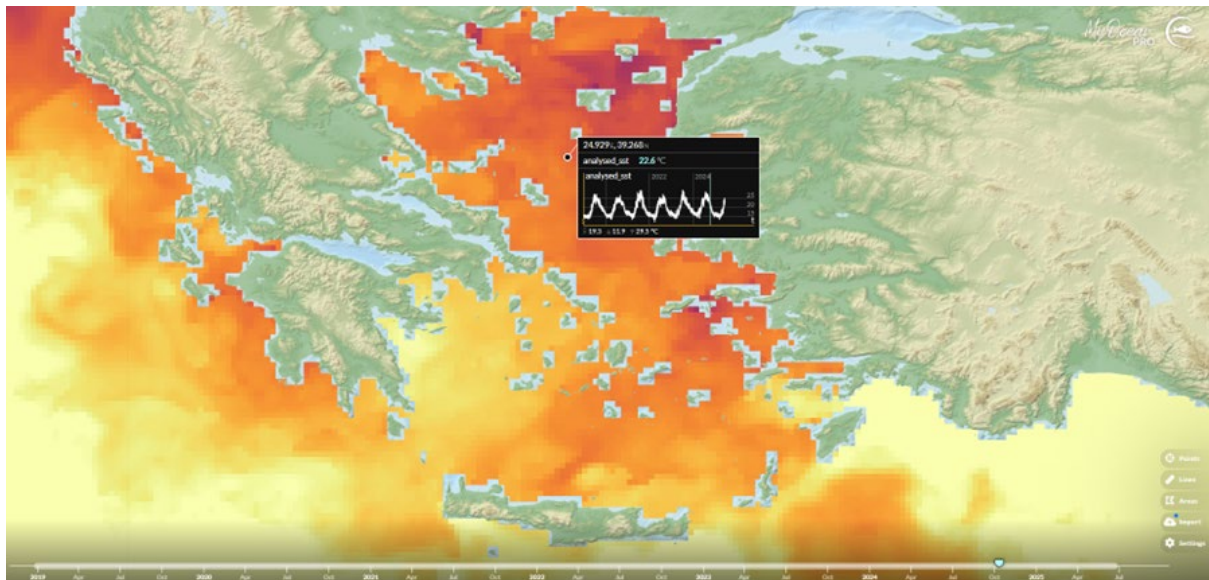


Figure 1: Screenshot from the Copernicus data portal ([data.marine.copernicus.eu](https://data.marine.copernicus.eu))

The dataset used is the CNR diurnal sub-skin Sea Surface Temperature (SST) product, which provides daily gap-free Level 4 (L4) maps of hourly mean sub-skin SST at a horizontal resolution of 1/16° (0.0625°) over the Copernicus Marine Environment Monitoring Service

(CMEMS) Mediterranean Sea (MED) domain. This product combines infrared satellite data with model outputs, as described in Marullo et al. (2014), and builds on operational SST processing chains that produce daily mean SST fields for the same region (Buongiorno Nardelli et al., 2013).

Sub-skin temperature refers to the temperature at the base of the thermal skin layer (Fig. 1). At night, it is equivalent to the foundation SST, but during the day it may differ significantly due to diurnal warming under clear skies and low wind conditions.

The L4 sub-skin SST product is generated by blending geostationary satellite observations from SEVIRI with model data from the CMEMS MED Monitoring and Forecasting Center (MFC), used as a first-guess. This methodology leverages the high temporal resolution of geostationary satellites to produce hourly, gap-free SST fields. The difference between satellite observations and model outputs (i.e., the SST anomaly field) is largely devoid of diurnal cycles, allowing for reliable interpolation of SST anomalies across various times of the day (Marullo et al., 2014).

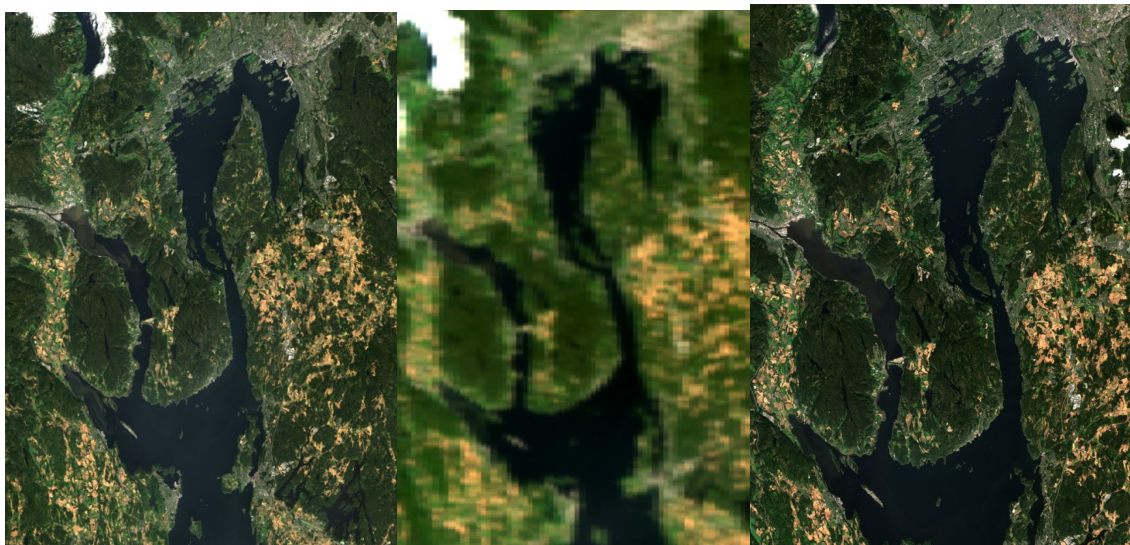
#### *b. Ocean color*

Ocean colour data from several relevant instruments onboard Sentinel-2 (A and B), Sentinel-3 (A and B), and Landsat-8/9 satellite platforms were used based on matchups in space and time. On Sentinel-2A/B, data were retrieved from the Multispectral Instrument (MSI) that is used for land and coastal monitoring with 13 bands (443-2190 nm) with a spatial resolution of 10 m (4 bands in the visible and near infrared wavelengths), 20 m (6 bands in the red-edge and shortwave infrared wavelengths), and 60 m (3 bands for atmospheric correction wavelengths). On Sentinel-3A/B, Ocean and Land Colour Instrument (OLCI) data, which include 20 bands (400-1020 nm) offering enhanced suitability for ocean color studies due to its higher spectral resolution and band configuration tailored for this purpose., and with a spatial resolution of 300 m which are were retrieved. On Landsat 8/9, data were retrieved from the Operational Land Imager (OLI) instrument that is used for water quality, amongst other, Earth Observations with 9 spectral bands (435-2300 nm) and one panchromatic band with a spatial resolution of 30 m for the multispectral bands and 15 m for the panchromatic band. The revisit times for the Sentinel-2, Sentinel-3, and Landsat-8/9 satellites are approximately 5, 2, and 8 days, respectively. Figure 2 shows the study area as observed by Landsat-8 (L8), Sentinel-3B (S3B) on September 5, and Sentinel-2A (S2A) on September 6, 2025—one day after the LIF lidar demonstration, which took place under overcast conditions for L8 and S3B, and two days after the lidar demonstration for S2A, highlighting the limitations of remote sensing due to cloud coverage preventing data acquisition on the day of the deployment, and temporal constraints of revisiting period, as no Sentinel-2 overpass occurred on the same day as the other two platforms. The figure 2 also displays the coarser spatial resolution of OLCI (on S3B).

Atmospheric correction for OLI and MSI data was made using the software package ACOLITE developed by the Royal Belgian Institute of Natural Sciences version 20250402.0. ACOLITE can be used for atmospheric correction of data and generation of water quality products from coastal and inland waters and supports MSI, OLCI, and OLI data that were retrieved as part of this work. The atmospheric correction converts the top-of-atmosphere reflectance into water-leaving reflectance, which is the signal that originates from the sea surface (Fig.2). The remote sensing Reflectance (Rrs) data product was generated for use in matchups with observations made during the T7.2 demonstrations in Oslofjord.

Chlorophyll-a retrieval algorithms were used to estimate Chl-a concentrations using Rrs data products. These include OC2 (Ocean Color 2-band algorithm), OC3 (Ocean Color 3-band algorithm) (Franz et al., 2015), OC4 (Ocean Color 4-band algorithm) (O'Reilly and Werdell 2019) and the Mishra algorithm (Mishra and Mishra, 2012). The OC2 algorithm is based on the blue-green reflectance ratio (typically uses two bands: 443 and 555 nm) and is generally well-suited for open ocean and oligotrophic waters. The OC3 algorithm is also based on the blue-green reflectance ratio, but instead can use two blue bands and one green band (e.g., 443, 490, and 555 nm). OC3 is better suited for mesotrophic regimes and handles spectral variability better. The OC4 algorithm is a four-band blue-green reflectance ratio-based algorithm (412, 443, 490, and 555 nm) that is usually used with open ocean OLCI data. The Mishra algorithm is based on the green-red reflectance ratio (typically 560 and 665 nm) and is applicable for turbid and eutrophic waters. For turbidity, the Nechad algorithm was used to estimate turbidity from Rrs data. The Nechad algorithm is a single-band, semi-analytical algorithm derived from radiative transfer simulations and validation with in situ turbidity data (Nechad et al., 2016). It is based on Rrs for a red/near-infrared wavelength (usually 665, 681, or 708 nm, depending on whether data is from MSI, OLCI, or OLI) and usually works well in water masses that are turbid and optically complex like the Oslofjord.

The satellite data selected for the matchups were acquired on the same day for S3A/B, and within a one-day tolerance for L8/9 and S2A/B. For L8/9 and S2A/B, values were averaged within a 3×3 pixel window around the matching coordinates, while for S3A/B, only the closest pixel value was used, considering the coarser spatial resolution of OLCI (Fig. 2).



*Figure 2: RGB composites of L8 OLI, S3B OLCI and S2A MSI scenes from passes over Oslofjord on 05/09/25 for L8 OLI and S3B OLCI and the 06/09/25 for S2A MSI, after atmospheric corrections.*

## 2. MARINE LITTER DETECTION

### 2.1. Field data

AI models for marine litter detection rely on field data for training of the model and its evaluation. This Deliverable uses a number of data sets: Existing data sets from another

project were used by DFKI for training of the models. Evaluation was conducted on data recorded by NIVA or with the NAUTILOS citizen science app.

RGB, Multispectral and hyperspectral observations with a Sony A7R V 61MP, Micasense altum PT and Specim AFX10 camera systems (presented in detail in D3.3) were collected by NIVA in May 2023 on a beach in Oslofjord, where plastic litter was placed over a beach nearby NIVA research infrastructure. The dataset was presented in D6.5 as it was also used to validate the capabilities of the integrated camera on their UAVs. This data is scaled to different GSD of 9 mm, 28 mm, 33 mm and 47 mm corresponding to resolutions DFKI's plastic detection models were trained and evaluated on in other projects. Furthermore, GSD of 100 mm, 250 mm and 500 mm are simulated corresponding to tentative multi-spectral satellite image resolutions. Data from the NAUTILOS Citizen Science campaign app (see D12.3) contained 12 images recorded by users with different smartphone camera systems. This data set contained too few images to conduct robust training of an AI model for marine litter detection. Therefore, this data set was only used to evaluate the capabilities of the model.

Training data for a segmentation-based marine litter detection was obtained in the project PlasticObs\_plus. Data was collected in the air using a DJI Matrice 210 V2 UAV for high-resolution, site-specific mapping at different altitudes of 4 meters (Ground Sampling Distance (GSD) 2 mm) and 20 meters (GSD 9 mm). It was equipped with a multispectral MicaSense Altum V04 camera system, which is comparable to NIVA's MicaSense Altum PT; it merely lacks a panchromatic band. The dataset (Rettig et al. 2025) comprises 432 training images of litter. As the data set was collected a few days after the litter was brought into the environment, it contained mostly intact objects. This is a downside to this approach, because beach litter can be degraded.

Training data for a pixel-based litter detection was collected during three field tests in northern Germany, encompassing a variety of terrains and configurations. The data was collected through the utilization of a DJI Mavic M210 drone, which was equipped with a Micasense Altum multispectral camera. The drone operated at an altitude of 75 meters, and the collected images were subsequently utilized to create an orthomosaic. In the initial two field tests, five artificial plastic targets were utilized, including LDPE blue (low-density polyethylene), LDPE transparent, PS white (polystyrene), PS cream, and PP black (polypropylene). These targets were placed on three different surfaces: grass, sand, and water. The third field test was conducted on a German landfill and therefore contains plastic waste of unknown types. A dataset was created from the labelled images, with a vector for each pixel containing the pixel values and the corresponding class. Due to the considerable imbalance present within the dataset, the random undersampling method is employed (van den Broek et al. 1998). During the labelling process, various classes were utilized, including different types of plastic, meadow, water, and beach. For each of these classes, a maximum of 1,000 pixels were randomly selected from each image for the dataset. All data sets were annotated by DFKI using the Computer Vision Annotation Tool (Ref ID: DFKI2).

Instead of using hyperspectral images for the detection of marine litter, as originally planned in the proposal, DFKI altered its approach, leveraging the geometric characteristics of objects based on RGB or multi-spectral images. Hyperspectral data has a few drawbacks, as it is more complicated to get correct reflectance images (reflectance values normalised to the light intensity), and usually creates line scans that have to be combined to a total image which can

lead to very blurry images (see D6.5). Additionally, hyperspectral cameras are quite expensive compared to RGB or multispectral cameras.

## 2.2. AI image detection

Due to the limited availability of data in NAUTILOS (too few to build new models just on them), a model was used, which was originally developed in the project PlasticObs\_plus (Ref ID: DFKI3). The objective of the models was to evaluate high-resolution multispectral images to achieve more precise quantification of plastic waste. A Mask R-CNN model (PyTorch with ResNet50 backbone) was deliberately selected for the instance segmentation to circumvent the commercial license restrictions of YOLO models and to preserve the project's open-source character. The development of models was conducted for both images of the multispectral (five channels) and RGB color spectrum (Ref ID: DFKI4). To efficiently evaluate image data from different flight altitudes and the resulting spatial resolution, a scaling process was implemented, which was supplemented by the adaptation of annotation classes (Ref ID: DFKI5). This approach was found to be conducive to the maximum adaptability of the ground truth data basis, extending its utility to alternative use cases involving different sensors.

While instance segmentation models, such as Mask R-CNN, have proven to be highly effective for detecting and delineating distinct objects in high-resolution imagery, they are not always optimal in scenarios involving lower-resolution data or smaller debris (Jia et al. 2023). In such cases, the analysis of individual pixels could also be beneficial, particularly when dealing with low-resolution images and smaller plastic particles, since spatial information such as shape can no longer be obtained. An artificial neural network with three hidden layers and 100 nodes per layer was trained using this dataset. The input is represented by a pixel vector, and the output is binary, with 1 indicating "plastic" and 0 indicating "no plastic" (Tholen et al. 2024)

## III. RESULTS

### 1. MATCHUPS

#### 1.1. SST (Aegean Sea)

For each hourly IR data group, the corresponding satellite overpass hour was first identified. Then, satellite grid points located within a 30 km radius of the IR measurement location were selected using the Haversine distance formula to account for Earth's curvature. The sea surface temperature (SST) values from these satellite grid points were averaged to obtain a representative satellite SST for that hour and location. Simultaneously, the mean IR temperature was computed from all IR measurements taken within that hour to facilitate direct comparison.

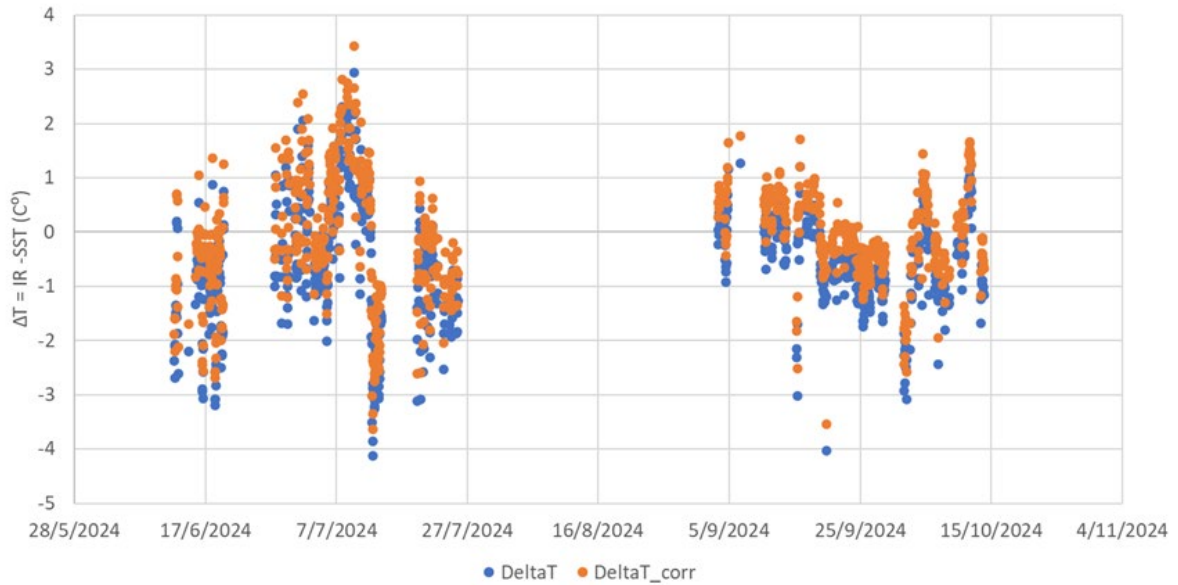


Figure 3: The difference in temperature between raw (blue) and corrected (orange) IR data and satellite data

Figure 3 shows the difference between the IR sensor temperature data and the satellite product before and after the IR data correction. The mean differences before and after correction were 0.61°C and 0.11°C, respectively. The comparison statistics between the IR corrected dataset and the satellite SST data are presented in figure 4.

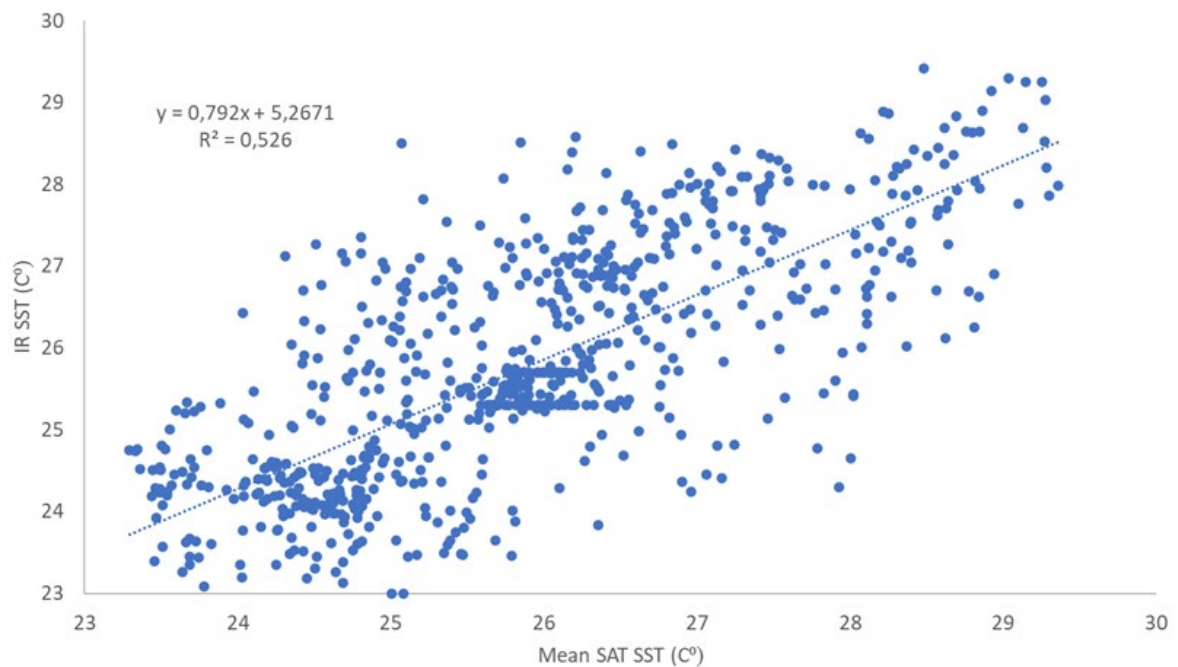


Figure 4: Comparison statistics for IR and SAT temperature data

## 1.2. Ocean colour (Norwegian Sea)

Final products such as chlorophyll-a (Chl-a) and turbidity extracted in the study area were particularly challenging due to their proximity to land, complex atmospheric conditions, and the low reflectance of Oslofjord waters caused by high absorption from elevated CDOM concentrations. For example, Figure 5 illustrates turbidity derived from the same reflectance data as in Figure 2 using the Nechad turbidity algorithm. Turbidity could not be retrieved from S3 OLCI data (Fig. 2, middle panel), and results from L8 OLI and S2 MSI (Fig. 2, left and right panels, respectively) show noticeable spatial discrepancies in turbidity values. Additionally, different algorithms applied to the same reflectance values can yield highly variable estimates, which is demonstrated for Chl-a derived using OC2, OC3, and Mishra algorithms from the same S2A scene presented in Figures 2. The Chl-a products vary by one order of magnitude with maximum values in Drammensfjord (fjord that extends in the NW direction in the center-right part of the scenes) approaching  $\sim 10 \mu\text{g Chl-a L}^{-1}$  in OC2 and OC3 products and  $\sim 5 \mu\text{g Chl-a L}^{-1}$  in the Mishra product. Therefore, the algorithm should be carefully selected based on the specific conditions of the study areas.

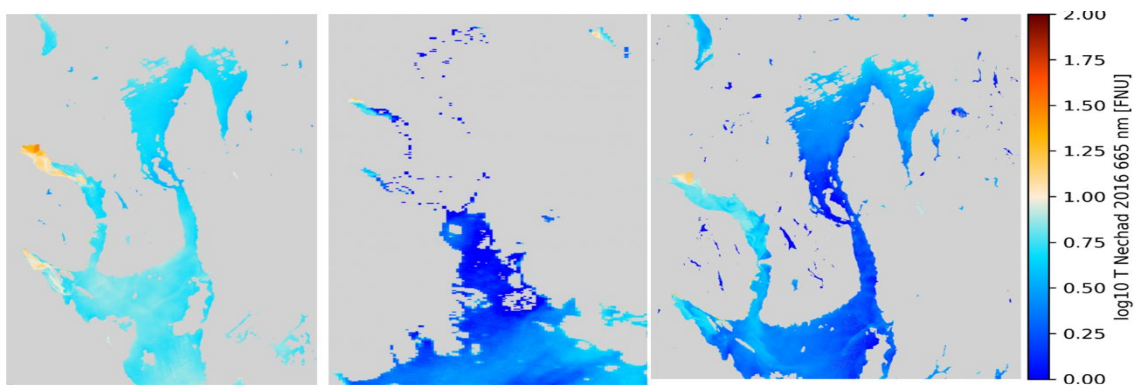


Figure 5: Turbidity measured by L8 OLI, S3 OLCI and S2 MSI (from left to right) from the same scenes shown in Fig. 2.

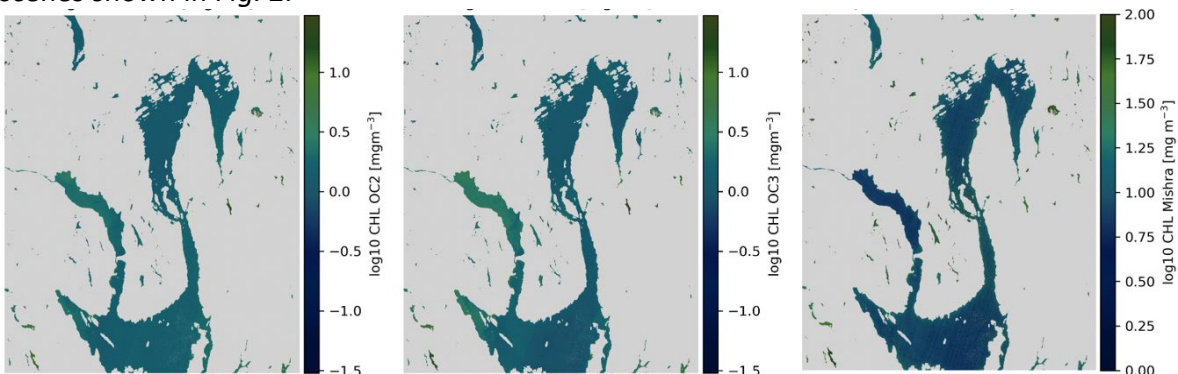


Figure 6: Chl-a estimation from OC2, OC3, Mishra (from left to right) from the same S2 scene shown in Fig. 2.

Matchups between turbidity estimated from S2 MSI observation with Nechad 2016, based on 665nm band, and reflectance measured by the UAV based multispectral imager at 668nm showed a good correlation (result not shown here).

The best matchups between UAV-camera and satellite passes were made with the UAV-multispectral camera observations, so the remaining section will be focused on multispectral camera data. Rrs measured in the field by the UAV-multispectral camera and by satellite-borne instruments, MSI (S2A/B), OLI (L8-9) and OLCI (S3A/B) are shown in Fig. 7. Because the wavelengths at which the satellite-borne instruments and the multispectral camera did not exactly coincide, the closest wavelength pairs were matched up for Fig. 7. For example, both MSI and the multispectral camera had measurements at 560 nm, so both have Rrs data at 560 nm; however, the closest wavelengths in the blue-cyan band was 492 nm for MSI and 475 nm for the multispectral camera, and therefore 475/492 nm data are matched up in Fig. 7 (top panel). While there was generally a linear relationship between Rrs from UAV-multispectral camera and S2 MSI, L8/9 OLI, and S3 OLCI, Rrs from the UAV-multispectral camera were higher by a factor of ~5-8-fold. Also, there was some variability around the linear relationship, for example, a positive residual for the infrared data (842 v. 833-865 nm) and a relatively smaller positive and negative residual for the visible wavelengths in the range of 560-717 nm. Several factors may contribute to the variability observed between UAV-based multispectral reflectance (Rrs) and Sentinel-2 MSI Rrs. These include low signal-to-noise ratios caused by strong atmospheric scattering relative to the water-leaving signal—especially in highly absorptive waters—resulting in reduced data quality; spatial integration challenges in satellite data near land-water boundaries; sun glint effects; differences in spectral band configurations between sensors; and calibration uncertainties associated with the UAV-mounted multispectral camera.

Matchups of Chl-a measurements are presented in Table 1. These include Chl-a estimated over a 4-day period (3-6 September 2024) by the UAV-multispectral camera, MS Color Fantasy FerryBox at ~4 m depth, Landsat-8 OLI OC3, Sentinel-2 OLI OC2, Sentinel-3 OLCI OC4, ASV-LIF-LIDAR integrated between 0-5 m, CTD Chl-a sensor at 0 and 5 m, and manually-collected seawater samples. The CTD Chl-a sensor was a Sea-Bird SBE 9 with a Chl-a fluorometer deployed by R/V Trygve Braarud. Chl-a from seawater samples collected with an autonomous surface vehicle and rigid-inflatable boat were measured spectrophotometrically from filtrations with  $\varnothing$  47 mm GF/F filters. The pigments were then extracted in a 100 % methanol solution and analysed on a Perkin Elmer UV/VIS Lambda 365 spectrophotometer, following Norwegian NS-4767 protocol. In situ measurements of Chl-a ranged from 1.07 to 5.16  $\mu\text{g Chl-a L}^{-1}$  in the 0-5 m depths (FerryBox, LIF-LIDAR, CTD fluorometer, manually-processed seawater samples). The UAV-multispectral camera measurement on 5 September was 3.86  $\mu\text{g Chl-a L}^{-1}$  from a depth interval likely representing the euphotic zone which was ~5 m deep, while in situ data from the same day ranged from 2.86-4.31  $\mu\text{g Chl-a L}^{-1}$ . In this study, the limited number of coincident satellite and in situ measurements (only three per spaceborne instrument) was insufficient to determine which chlorophyll-a estimation algorithm performed best in the study area. To ensure greater consistency across sensors, an OC (ocean color) algorithm was selected for each instrument, as Landsat does not have spectral bands in the red-edge region required by certain chlorophyll-a algorithms. The satellite Chl-a products ranged from 1.32-3.41  $\mu\text{g Chl-a L}^{-1}$  over the whole time period, however the most reliable Chl-a product for comparison was from Landsat-8 OLI OC3 of 1.32  $\mu\text{g Chl-a L}^{-1}$  since Sentinel-2 OLI OC2 did not pass on the same day, as the UAV-multispectral camera demonstration, and Sentinel-3 OLCI OC4 Chl-a products were found to overestimate concentrations.

The Rrs and Chl-a results indicate that satellite-derived Rrs values are generally underestimated relative to UAV-multispectral camera values (Fig. 7), which in turn leads to an underestimation of Chl-a by satellite Chl-a retrieval algorithms (Table 1). This issue is commonly observed in CDOM-rich, low-particle waters such as those found in Scandinavia, where high CDOM absorption strongly attenuates blue wavelengths and reduces satellite Rrs (Tilstone et al., 2022). The accuracy of satellite-derived products in such environments can be improved through region-specific vicarious calibration using field-measured Rrs, along with in situ measurements of Chl-a, CDOM, and turbidity—highlighting the importance of new technologies for collecting these critical data. Furthermore, other issues can affect satellite products such as the above-mentioned issue of land interferences, as well as the high level of heterogeneity in coastal systems - in water masses, Chl-a concentrations, or even phytoplankton community structure showing a strong variability in both space and time. UAV-multispectral camera observations can provide better links between point-measurements in heterogeneous and complex water/EO conditions, while satellite observations provide a broader spatial and temporal view with some additional uncertainties.

*Table 1: Chl-a ( $\mu\text{g Chl-a L}^{-1}$ ) as measured by UAV-multispectral camera, FerryBox fluorometer (FB), Landsat8 (L8) OC3 product, Sentinel 2 (S2) OC2 product, Sentinel 3 (S3) OC4 product, LIF-LIDAR, CT-fluorometer deployed from a research vessel (CTD flu), and water samples collected from an autonomous surface vehicle and rigid-inflatable boat and manually analysed via extraction/spectrophotometry.*

date	UAV	FB (~4m)	L8 OC3	S2 OC2	S3 OC4	LIF- LIDAR (0-5m)	CTD flu 0m	CTD flu 5m	Water sample 0m	Water sample 3m
03.09.24	n.d.	n.d.	n.d.	n.d.	n.d.	3.68	1.4	n.d.	2.81	3.64
04.09.24	n.d.	3.08	n.d.	n.d.	n.d.	4.28	2.75	3.22	5.16	3.95
05.09.24	3.86	n.d.	1.41	n.d.	3.41	n.d.	n.d.	n.d.	2.86	n.d.
06.09.24	n.d.	1.07	n.d.	1.32	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.

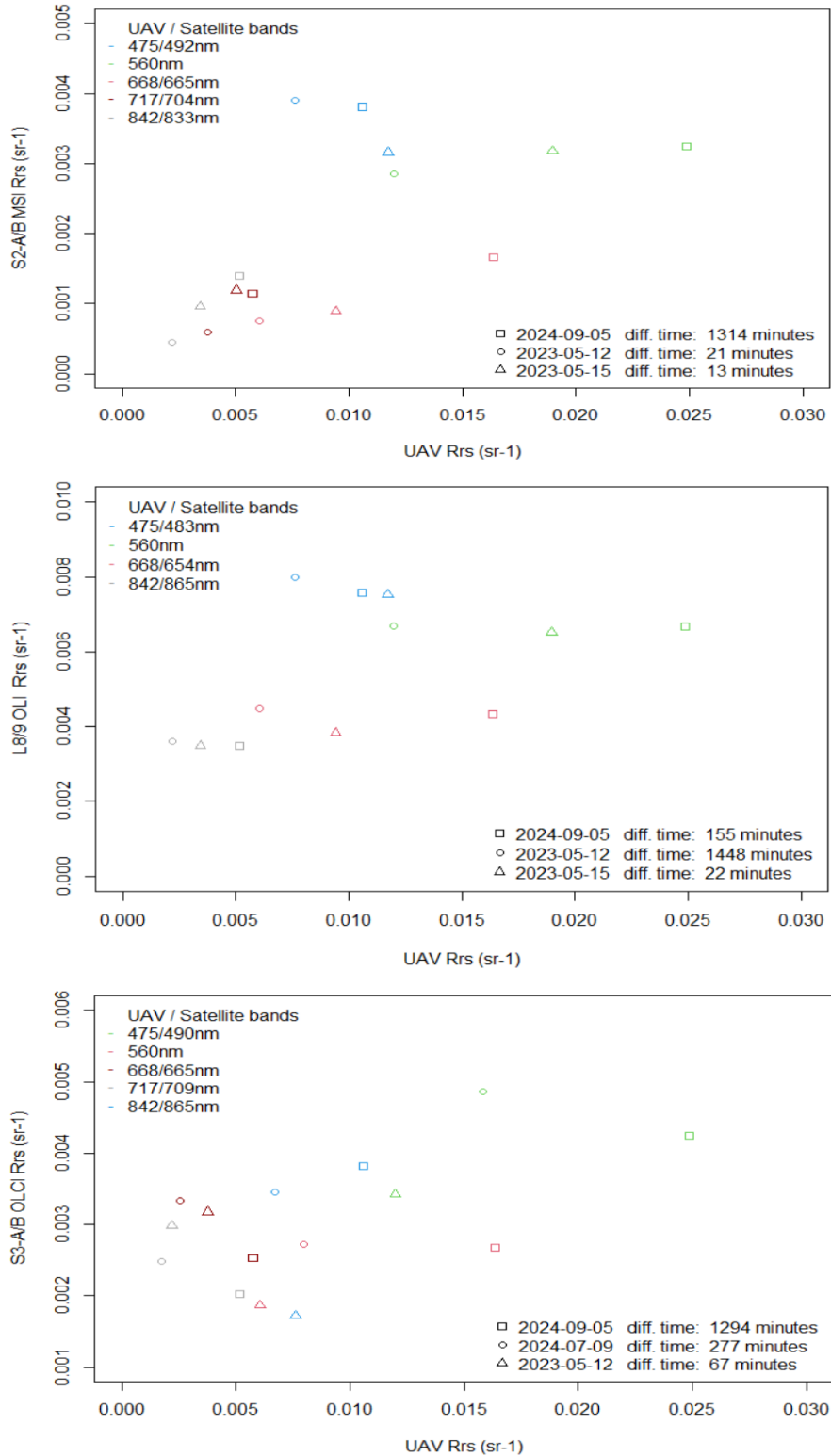


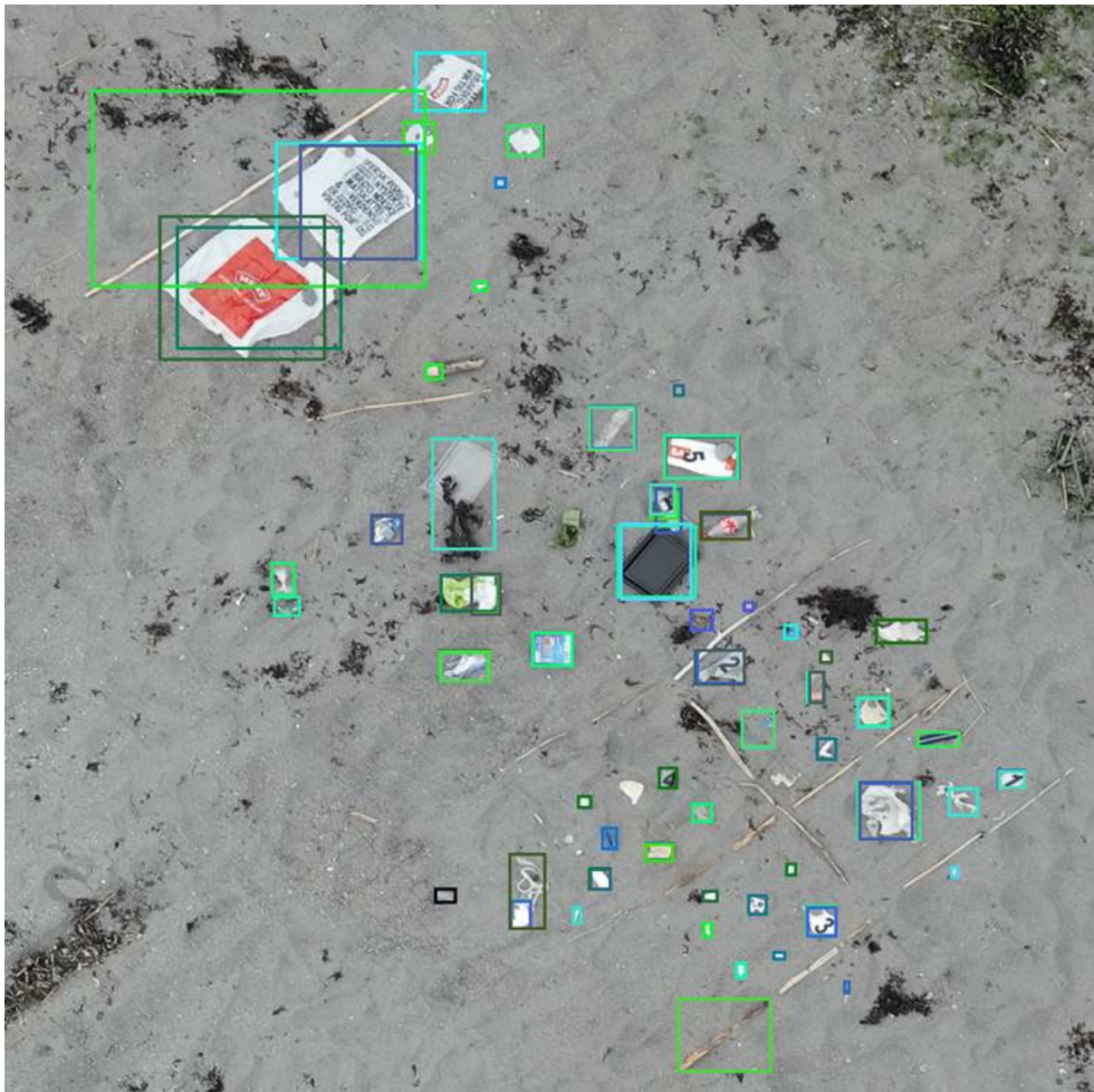
Figure 7: Top panel: Matchups between S2A/B MSI and UAV-multispectral camera Rrs. Middle panel: Matchups between L8/9 OLI and UAV-multispectral camera Rrs. Bottom panel: Matchups between S3A/B OLCI and UAV-multispectral camera Rrs. Matchup times are offset by ~13-1314 minutes depending on satellite matchup data used.

## 2. MARINE LITTER DETECTION

---

In the following section, the results obtained from the AI models are presented. The analysis of the F1 scores indicates a clear influence of the spatial resolution.

The model obtained an F1 score of 0.68 for the RGB image (Figure 8), provided by NIVA (see D6.5). This relatively strong score can be attributed to the model's ability to correctly identify and localize a significant portion of the litter present in the scene. Specifically, the score indicates that while the model effectively captures the majority of true litter instances, a few objects may have been missed (false negatives) or detected more than once (duplicate bounding boxes), which slightly lowers the overall precision. However, results are worse than on the training data set, because the training data set predominantly comprises mostly intact objects instead of the more fragmented and degraded forms characteristic of marine litter.



*Figure 8: Object detection with the MaskRCNN model on drone imagery from a beach in Norway.*

As illustrated in the following images (Figure 9), while the model has been trained on drone data, it is relatively feasible to utilize it with smartphone images for citizen science applications, as those collected in D12.3. It is evident that the modifications made to the images, including the alteration in perspective and the introduction of novel objects not present in the training dataset, have led to certain drawbacks. These novel objects bear no resemblance to the environment observed in the DFKI training image set. Consequently, they are identified as objects of interest and are classified within one of the waste categories. In particular, stones have a high tendency to be misidentified as waste in the marine environment when employing a model that has not been trained on smartphone images of beaches.



*Figure 9: Example of marine litter detection based on citizens science beach images*

The impact of scaling the image in Fig. 8 to different GSD is shown in Fig. 10. As the GSD increases, the performance drops severely until it ultimately reaches an F1 score of 0 at a GSD of 500 mm. At this GSD the image has a size of 7 x 8 pixels and contains no visible traces of plastic anymore, as each individual object is too small to be visible. Even at a GSD of 47 mm, which corresponds to an altitude of 100 meters with the camera systems used by NIVA, the F1 score is so low that operational use of the plastic detection models is no longer advisable.

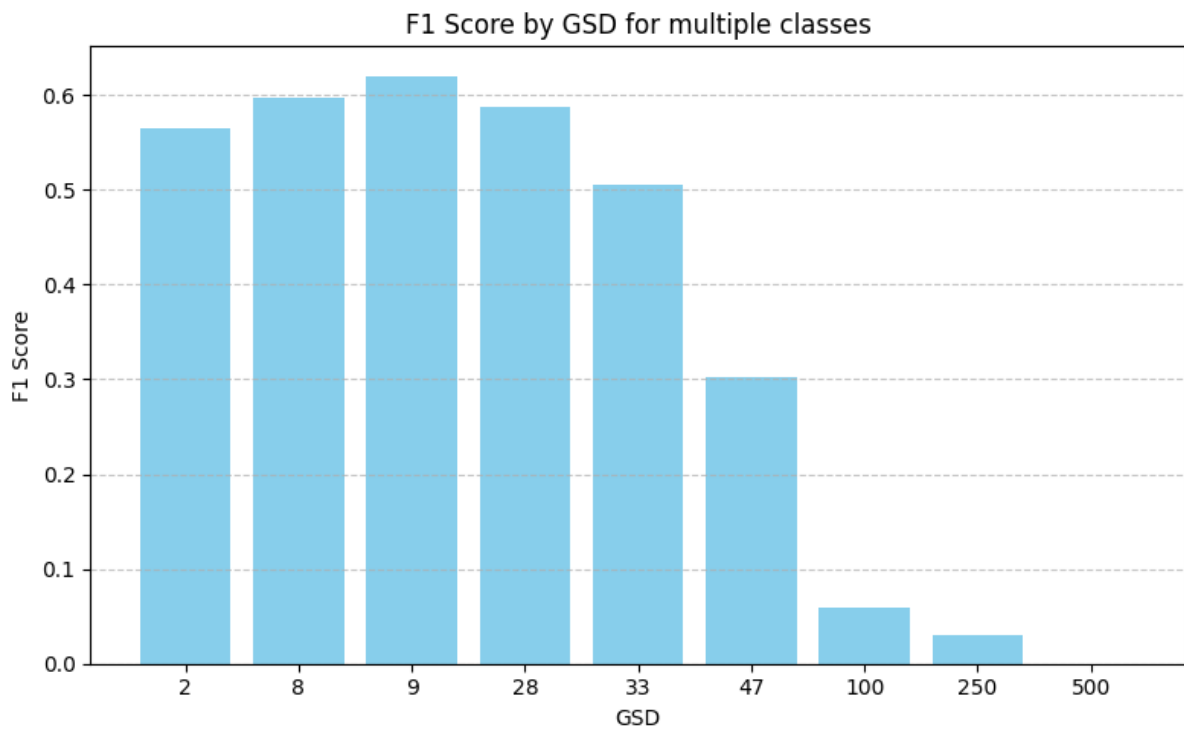


Figure 10: F1 scores of the shape-based litter detection on RGB images with different GSD

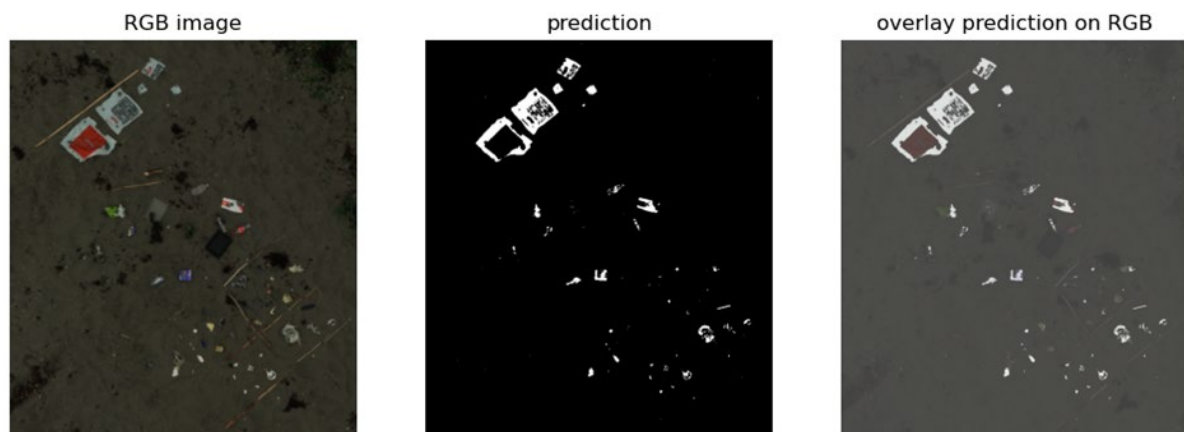


Figure 11: Visualization of plastic detection using pixel-wise multispectral image analysis. Left: Composite RGB image generated from the multispectral camera. Center: Result of pixel-wise plastic detection based on multispectral data. Right: Overlay of the RGB image with the plastic detection map, highlighting identified plastic regions.

Although Mask R-CNN demonstrated encouraging outcomes, we also conducted experiments using the pixel-based approach. As illustrated in Figure 11, the model's predictions can be observed, indicating a significant number of plastic parts that are not recognized. This phenomenon may be attributed to the training data set, which does not encompass the full spectrum of colors. However, the proportion of plastic parts identified as false positives is minimal. This approach yielded an accuracy of 96% and an F1 value of 0.47 for the unbalanced, pixel-by-pixel multispectral dataset derived from the orthomosaic (see Figure 10). In contrast, the model achieved an accuracy of 95% and an F1 value of 0.27 when using only RGB channels.

As expected, the pixel-based detection of marine litter declines with an increased GSD, although the multispectral approach outperforms detection of marine litter based solely on RGB (see Fig. 12). The performance increase at GSD=500 mm is an anomaly, which might be attributed to the number of pixels available for evaluation. Only 4 pixels of litter remain (see Fig. 13, right). Moreover, Fig. 13 shows that smaller particles of litter disappear from the images at a larger simulated GSD. As a consequence, only larger objects and congregations of several particles can be detected from larger altitudes. On this limited data set, multispectral data provides an improvement, although larger data sets are needed for the training of better models and more informative evaluation.

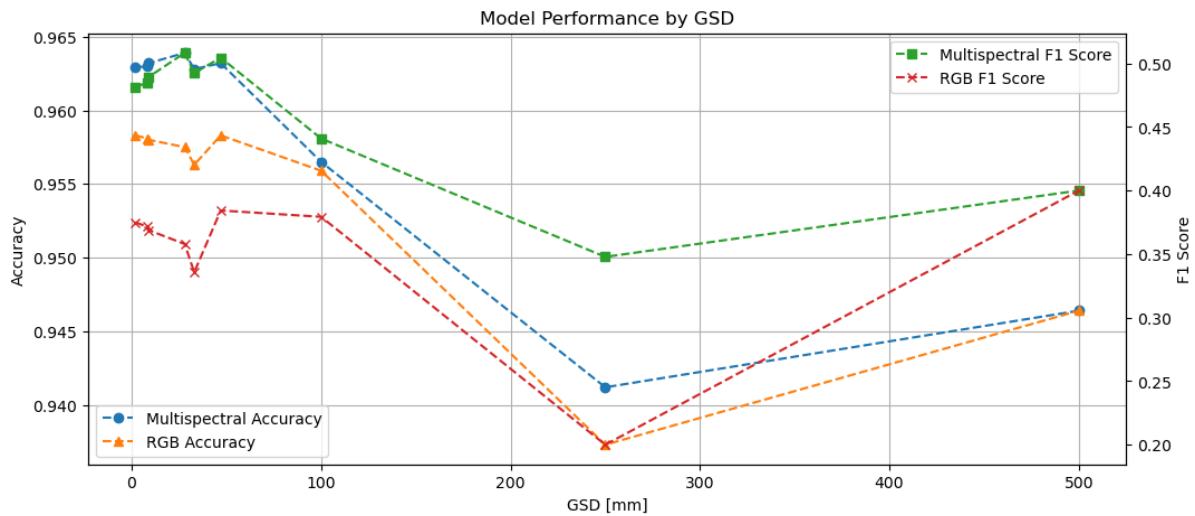


Figure 12: Accuracy and F1 scores of the pixel-based litter detection on multispectral images and RGB images with different GSD

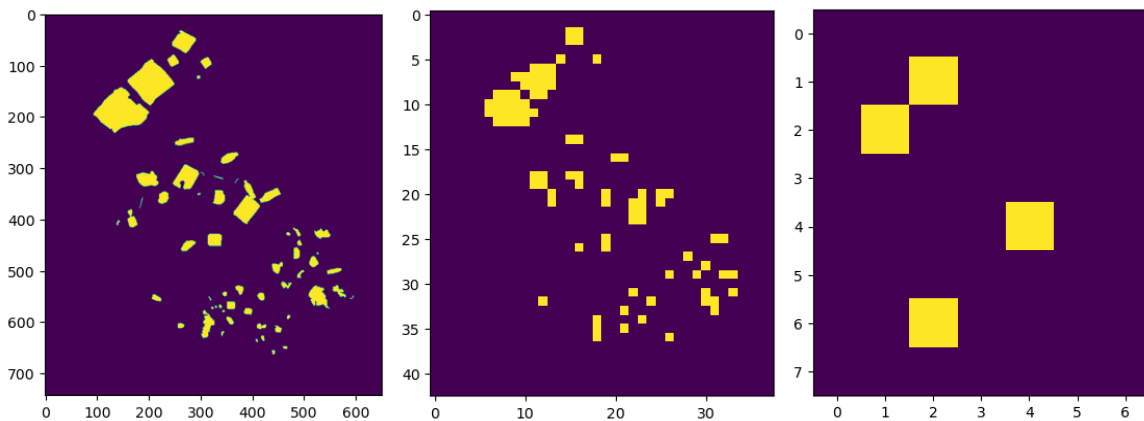


Figure 13: Pixels displaying litter at the original resolution (left), GSD=100 mm (center) and GSD=500 mm (right). Yellow pixels are litter, whereas purple pixels indicate environment.

#### IV. SUMMARY

The spatial and temporal coverage of future observations for validation of remotely-sensed EOVs and emerging EOVs can be supported by other instruments developed within NAUTILOS. A cost-effective network of compact infrared SST sensors can be installed on small

commercial vessels—such as fishing and leisure boats—to collect sea surface temperature data in targeted areas of the Mediterranean. Based on D7.3 recommendations, these low-cost, easy-to-install modules are ideal for widespread deployment. Using ship density maps to guide sensor placement ensures frequent and broad data coverage along high-traffic maritime routes. A new generation of small, robust, and user-friendly Chl-a sensors developed in Subtask 3.1.2 will enable additional observations in unconventional ocean observing applications such as fisheries- and aquaculture-observing systems, and eventually citizen science campaigns. The Sampler for phytoplankton and other suspended matter can be deployed on autonomous observation lines such as FerryBoxes (which T7.2 carried out a demonstration for, presented in D7.3), or Autonomous Surface Vehicles that can accommodate the necessary payload requirements, and collect samples for variables like Chl-a, total suspended matter, phytoplankton carotenoid pigments, etc. that are useful for remote sensing validation but are traditionally only sampled/filtered manually on research cruises that are generally low in frequency and spatial extent.

Discrepancies were observed between satellite and UAV data. These differences can be driven by complex coastal conditions that affect atmospheric correction and reflectance retrievals. These findings underscore the critical need to expand in situ observations of Chl-a, CDOM, and turbidity, to support regional algorithm tuning and validation, as well as Rrs, intermediate product, critical for atmospheric correction calibration and validation. Increasing the density and frequency of such measurements is essential for improving the reliability of satellite-derived water quality products in heterogeneous and optically complex coastal waters.

The detection of plastic litter on beaches is possible, although it is affected by the size of the plastic waste. The smaller the particles, the worse the predictive performance. Transfer from mostly intact waste to more degraded is difficult, but models still predict the majority of annotated particles. Especially the low rate of false positives means the algorithms can still be used for support in cleanup activities, even though they cannot be depended upon to identify all litter. The use of additional channels from a multispectral camera system led to a minor improvement of 1 percentage point in accuracy and a notable improvement of the F1 score by 0.2. The lower impact on accuracy is easily explained by the fact that images mostly contain non-litter pixels, which the models already classified very reliably.

At larger GSD, model performance decreases. However, leveraging multispectral data yields better results than using only RGB channels. Finally, smaller plastic particles become harder to see, until only large objects and congregations of several objects remain visible.

## V. APPENDIX 1: REFERENCES AND RELATED DOCUMENTS

ID	Reference or Related Document	Source or Link/Location
1	D3.2	Report on the development of Dissolved Oxygen and Chlorophyll-a sensors for fishery vessels
2	D3.3	Report on laboratory tests of downward looking sensors
3	D5.3	Report on integration of payloads/sensors on ASV
4	D5.4	Report on integration of payloads/sensors on UAV
5	D5.6	Validation and integration report on ships of opportunity
6	D6.1	Report on the results and methodology of the calibration/validation experiments
7	D6.5	Report on the testing results of the joint operations of sensors and UAV in ST6.3.3
8	D7.2	Fisheries and Aquaculture Observing Systems demonstration final report
9	D7.3	Platforms of Opportunity and Ferryboxes demonstration final report
10	D12.3	Report on citizen science campaigns
11	Mishra, S. and Mishra, D.R., 2012. Normalized difference chlorophyll index: A novel model for remote estimation of chlorophyll-a concentration in turbid productive waters. Remote Sensing of Environment, 117, pp.394–406.	<a href="https://doi.org/10.1016/j.rse.2011.10.016">https://doi.org/10.1016/j.rse.2011.10.016</a>
12	Acolite Manual	<a href="https://github.com/acolite/acolite/releases/tag/20210802.0">https://github.com/acolite/acolite/releases/tag/20210802.0</a>
13	O'Reilly, J.E.; Werdell, P.J. Chlorophyll Algorithms for Ocean Color Sensors—OC4, OC5 & OC6. Remote Sens. Environ. 2019, 229, 32–47	<a href="https://doi.org/10.1016/j.rse.2019.04.021">https://doi.org/10.1016/j.rse.2019.04.021</a>

14	Nechad, B., Ruddick, K.G. and Park, Y., 2010. Calibration and validation of a generic multisensor algorithm for turbidity mapping in coastal waters. <i>Remote Sensing of Environment</i> , 114(4), pp.854–866.	<a href="https://doi.org/10.1016/j.rse.2009.11.022">https://doi.org/10.1016/j.rse.2009.11.022</a>
15	Franz, B.A.; Bailey, S.W.; Kuring, N.; Werdell, P.J. Ocean color measurements with the Operational Land Imager on Landsat-8: Implementation and evaluation in SeaDAS. <i>J. Appl. Remote Sens.</i> 2015, 9, 096070	<a href="https://doi.org/10.1117/1.JRS.9.096070">https://doi.org/10.1117/1.JRS.9.096070</a>
16	Tilstone, G. H., Pardo, S., Simis, S. G. H., Qin, P., Selmes, N., Dessailly, D., & Kwiatkowska, E. (2022). Consistency between Satellite Ocean Colour Products under High Coloured Dissolved Organic Matter Absorption in the Baltic Sea. <i>Remote Sensing</i> , 14(1), 89.	<a href="https://doi.org/10.3390/rs14010089">https://doi.org/10.3390/rs14010089</a>
HCMR1	CNR diurnal sub-skin Sea Surface Temperature (SST) product	<a href="https://doi.org/10.48670/moi-00170">https://doi.org/10.48670/moi-00170</a>
HCMR2	Buongiorno Nardelli B., C.Tronconi, A. Pisano, R.Santoleri, 2013: High and Ultra-High resolution processing of satellite Sea Surface Temperature data over Southern European Seas in the framework of MyOcean project, <i>Rem. Sens. Env.</i> , 129, 1-16,	<a href="https://doi.org/10.1016/j.rse.2012.10.012">doi:10.1016/j.rse.2012.10.012.</a>
HCMR3	Marullo, S., Santoleri, R., Ciani, D., Le Borgne, P., Péré, S., Pinardi, N., ... & Nardone, G. (2014). Combining model and geostationary satellite data to reconstruct hourly SST field over the Mediterranean Sea.	<a href="https://doi.org/10.1016/j.rse.2013.11.001">https://doi.org/10.1016/j.rse.2013.11.001</a>

	Remote sensing of environment, 146, 11-23.	
DFKI1	van den Broek, Wienke, Melssen, & Buydens (1998): Plastic material identification with spectroscopic near infrared imaging and artificial neural networks	<a href="https://doi.org/10.1016/S0003-2670(98)00012-9">https://doi.org/10.1016/S0003-2670(98)00012-9</a>
DFKI2	Computer Vision Annotation Tool	<a href="https://www.cvat.ai/">https://www.cvat.ai/</a>
DFKI3	PlasticObs_plus	<a href="https://www.dfki.de/en/web/research/projects-and-publications/project/plasticobs-plus">https://www.dfki.de/en/web/research/projects-and-publications/project/plasticobs-plus</a>
DFKI4	PlasticObs_plus AI library	<a href="https://github.com/DFKI-NI/PlasticObs-_AI_Library/tree/main/eoir-ai">https://github.com/DFKI-NI/PlasticObs-_AI_Library/tree/main/eoir-ai</a>
DFKI5	Adapting Annotation Datasets	<a href="https://github.com/DFKI-NI/Adapting_Annotation_Datasets">https://github.com/DFKI-NI/Adapting_Annotation_Datasets</a>
DFKI6	Jia et al. (2023): Deep learning for detecting macroplastic litter in water bodies: A review	<a href="https://doi.org/10.1016/j.watres.2023.119632">https://doi.org/10.1016/j.watres.2023.119632</a>
DFKI7	Tholen, Rodenbäck, Nolle, Rettig, & Stahl (2024): On the Development of a Pixel-Wise Plastic Waste Identification System for Multispectral Remote Sensing Applications	<a href="https://doi.org/10.1007/978-3-031-77915-2_4">https://doi.org/10.1007/978-3-031-77915-2_4</a>
DFKI8	Rettig et al. (2025): Multiscale Waste PlasticObs_plus	<a href="https://doi.org/10.5281/ZENODO.14844219">https://doi.org/10.5281/ZENODO.14844219</a>