DELIVERABLE D2:
INPUT DATA INVENTORY

BY
ANDREY KUREKIN, EDSON SILVA, VICTOR MARTINEZ VICENTE,
LASSE H. PETTERSSON, ANTON SALGADO, CATHERINE DOWNY AND ANTONIO BONADUCE

AN ESA FUNDED PROJECT IMPLEMENTED BY
NANSEN ENVIRONMENTAL AND REMOTE SENSING CENTER, NORWAY AND
PLYMOUTH MARINE LABORATORY, UK

NERSC
STIFTELSEN NANSEN SENTER FOR MILJØ OG FJERNMÅLING
NANSENSENTERET · BERGEN · NORGE

PML | Plymouth Marine Laboratory

EUROPEAN SPACE AGENCY
CONTRACT REPORT

The work described in this report was done under ESA contract.
Responsibility for the contents resides in the author or organisation that prepared it.

Customer	European Space Agency (ESA)
Contract Name	Earth Observations for Sustainable Aquaculture (EO4SA)
Contract Number	4000146872/24/I-EF
Project Contractor & Manager	Nansen Environmental and Remote Sensing Center (NERSC), Bergen, Norway. Lasse H. Pettersson
Contributing institutions	NERSC and PML
Authors	Andrey Kurekin, Edson Silva, Victor Martinez Vicente, Lasse H. Pettersson, Anton Salgado, Catherine Downy, and Antonio Bonaduce
Document Reference	D2: Input Data Inventory
SoW Deliverable Reference	Deliverable D2: Input Data Inventory
Version/Revision	1.0
NERSC report no.	N/A
Date of issue	5 December 2025
Distribution	ESA and Consortium

Approved by NERSC	Lasse H. Pettersson NERSC Project Manager	
Approved by ESA	Marie-Helene Rio ESA Technical Officer	

Revision Change log

Issue	Date	Type	Who	Change description
0.1		Structure	AK, LHP	Report ToC and structure
1.0	05.12.2025	To ESA PO	AK + all	Submitted to ESA

Acronyms

ASF	Alaska Satellite Facility
CDSE	Copernicus Data Space Ecosystem
EO	Earth Observation
PAR	Photosynthetically Available Radiation
PP	Primary Production
SAR	Synthetic Aperture Radar
VHR	Very-high-resolution

Table of Contents

List of Figures

List of Tables

Preface

The **Earth Observations for Sustainable Aquaculture (EO4SA)** project is funded by the European Space Agency (ESA) in their EO for Sustainable Blue Economy project under the ESA PEOPLE-program. The project is implemented by the Nansen Environmental and Remote Sensing Centre (project contractor and manager) in Bergen, Norway and Plymouth Marine Laboratory, UK.

The main objective of EO4SA is to consolidate the requirements and demonstrate the information opportunities, needed by the aquaculture industry and monitoring agencies using EO and other data sources. For selected Early Adopters (EA) the project implements four Use Cases, at several locations in Norway, Spain and the Philippines, related to:

1. Predicting risks of salmon lice infestation in Norway.
2. Forecasting toxic algal blooms (HABs) impacting shellfish farming in Norway.
3. Optimising location of sustainable shellfish farming and tourism in Galicia, Spain.
4. Mapping aquaculture structures and use of marine resources in Palawan, Philippines.

Through these four use cases, EO4SA will contribute to enhancing the future operations of aquaculture management authorities, stakeholders and industry, by taking up new information based on satellite Earth observation data. This report is the project deliverable D2: Input Data Inventory, which describes the input data needed for each use case for the algorithm development and testing.

1. Input Data Inventory Database

The project input database contains all the input data that we require for the algorithm development, testing and validation activities. It has been discussed and decided at the early stage of the project that the database will be distributed between NERSC and PML and stored separately between these two locations. As the input dataset contains **terabytes** of data, it would be inefficient to keep all these data in one location. This is also not required since NERSC and PML independently work on their own use cases that have different geographical locations and the data they use do not overlap. This ensures that NERSC and PML organisation cannot work on different versions of the same data.

2. Use Case 1: Predicting risks of salmon lice infestation in Norway

2.1. Input data

This section presents the input datasets used in the Use case 1 application - Predicting risks of salmon lice infestation in Norway (Figure 1). The input data are historical lice density across different life stages and delousing employments in trout and salmon farms in Norway and the Norkyst-800 oceanographic model sea surface temperature (SST) and sea surface salinity (SSS) outputs. Those inputs are used to forecast the next probability of adult female lice (*Lepeophtheirus salmonis*) crossing the 0.2 and 0.5 lice per fish threshold, which are used for regulatory and environmental monitoring.

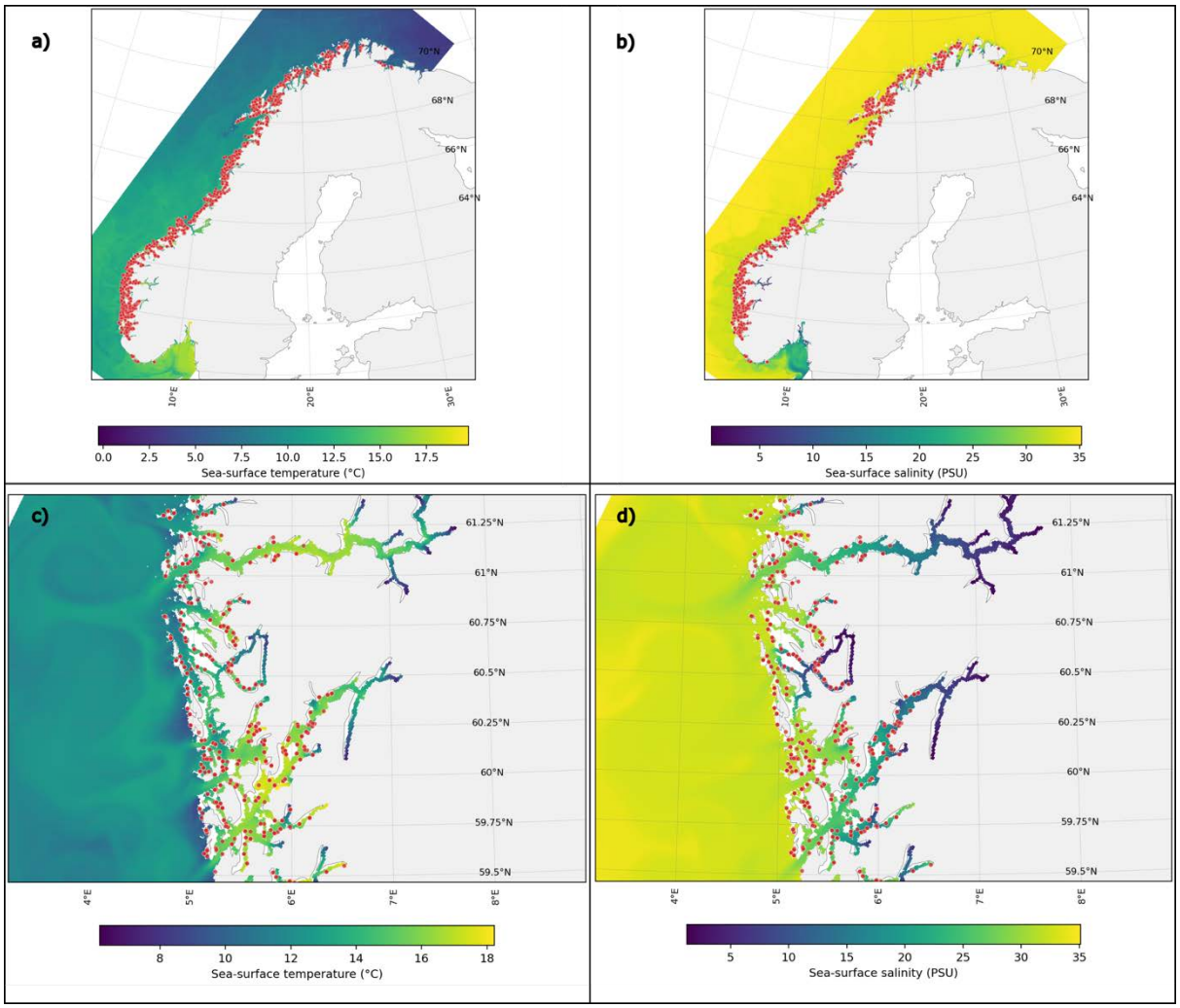

Figure 1. Norkyst-800 and Fish farms with reported lice counts and delousing actions. The subplots show the overall coverage of Norkyst-800 for 2024-06-01 for SST (a) and SSS (b) and a zoomed regional area for SST (c) and SSS (d). Fish farms with reported lice counts and delousing employments from 2012 to 2015 are shown as red dots.

Salmon lice input data

The lakselus_per_fisk.csv file contains weekly salmon lice density (counts per fish) reports from Norwegian aquaculture facilities (Table 1), which are reported to the Norwegian Food Safety Authority (NFSA). The counting consists in taking the farmed fish from the cages to a separated area (basin), where for each fish, lice attached in their skin are counted in different life stages: sessile lice, adult male lice, and adult female lice. When counting is finished, the density of each life stage per fish is estimated. The number of counted fish is variable and not reported or made available in public repositories. Each record is identified by a specific farm (region_id) and week of the year. Most observations are weekly although farmers can report lice bi-weekly when SST is relatively low. The data is made public available in near-real time (NRT) on the BarentsWatch website.

Table 1. Salmon lice input data summary obtained through BarentsWatch.

Dataset Property	Description
Dataset Name:	lakselus_per_fisk.csv
Source:	BarentsWatch (https://www.barentswatch.no/data/fishhealth)
Format:	CSV (Comma-Separated Values)
Size:	752,377 records
Temporal Coverage:	2012 - 2025
Temporal Resolution:	Weekly measurements
Spatial Coverage:	Norwegian coastal salmon aquaculture facilities (Latitude: 58.02N – 71.02N)
Coordinate System:	WGS84 (EPSG:4326) - Latitude/Longitude

Delousing treatment input

The tiltak_mot_lakselus.csv dataset tracks individual delousing treatments at aquaculture sites (Table 2), covering several methods that are grouped as medicinal (e.g., Emamectin benzoat) and non-medicinal (e.g., thermal bath) methods. Each record also contains the application scope, which separates delousing methods employed in the whole farm or specific cages. Similar to the lice data, the delousing records are identified by a specific farm (region_id) and week of the year.

Table 2. Delousing treatment input summary.

Dataset Property	Description
Dataset Name:	tiltak_mot_lakselus.csv
Source	BarentsWatch
Format	CSV (Comma-Separated Values)
Size	59,656 records
Temporal Coverage	2012 - 2025
Temporal Resolution	Event-based (individual treatment occurrences)
Spatial Coverage	Norwegian coastal salmon aquaculture facilities (Latitude: 58.02N – 71.02N)
Coordinate System	WGS84 (EPSG:4326)

NorKyst-800m input

NorKyst-800m is an operational ocean forecasting system for the Norwegian coastal waters. It provides several ocean variables – including SST and SSS that are relevant for and used by us – at 800 m spatial resolution (Table 3).

Table 3. Norkyst-800 input summary.

Dataset Property	Description
Dataset Name:	Multiple NetCDF files.
Source:	Norwegian Meteorological Institute (MET Norway)
Format:	NetCDF-4 (Network Common Data Form)
Size:	10 TB
Temporal Coverage:	Version 1 (v1): 2012-2017, Version 2 (v2): 2017-2024.
Temporal Resolution:	Daily averages
Spatial coverage:	Norwegian coastal waters (Norkyst domain)
Coordinate System:	EPSG: 5939. v1: ROMS grid with X/Y in fractional grid units (0.8 km resolution), v2: ROMS grid with X/Y in meters (800 m resolution)

Although Norkyst-800 is an ocean model rather than an Earth Observation (EO), through a variational data assimilation scheme it aggregates several EO data sources (Figure 2) during the ocean analysis to produce high quality ocean variables at the resolution and frequency needed for developing our applications. The integrated and EO derived products include the CMEMS Baltic Sea model, TOPAZ, OSTIA, as well as satellite radiance and altimetry systems. Despite some EO being able to provide SST and SSS, their coarse spatial resolution combined

with the several islands and narrows fjords produce many observations gaps where many fish farms operate. This not only reduce the coverage of the use case application but also de data available for further training the lice probabilistic models. Therefore, Norkyst-800 allows us to maximise the use of several EO systems integrated in an ocean model while improving the coverage of our application in all Norwegian coastal waters.

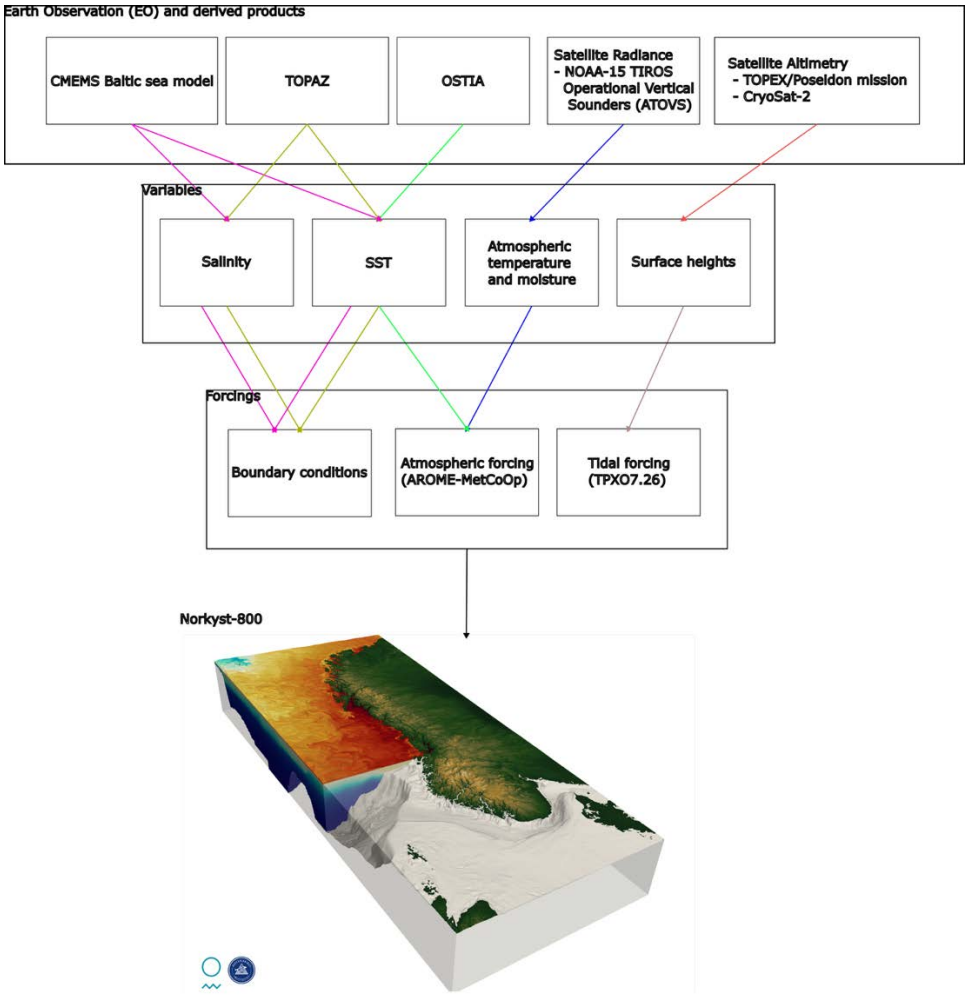

Figure 2. Simplified scheme of Norkyst-800 and EO dependency.

2.2. Data quality and other issues

Salmon lice input

There are many records with missing lice density, and that is explained by BarentsWatch as not reported data, probably because the farm is temporarily not operating. Consequently, this creates some gaps in the local farm time series used. Although it is not a problem for training the models, it might create gaps for demonstrating the application of the model predictions. Furthermore, the methods of counting lice are not completely standardized for all farms and time series, to the extent of our knowledge, including how many fishes should be samples and where in the cages they should be picked up. This might create bias in the dataset which we cannot estimate.

Delousing treatment

Although not considered a delousing treatment, the use other similar mitigation strategies are common in Norway, such as the use of skirts, sinking of cages in the sub-surface, and the use of cleaner fish. Those mitigation strategies are not reported in the public repositories, and their effect cannot be considered by the calibration of lice models. Cleaner fish was mandatory only until 2018, after which it ceased due to regulatory changes with regard to fish welfare. Furthermore, treatments are recorded in the week following their application (e.g., treatments applied during week t are documented in week t+1). Although not explicitly mentioned on the source of data, we notice this aspect by analyzing the time series. This means that for calibration of our prediction models, we should shift the delousing record from t+1 to t+0.

NorKyst-800 input

The main issue is the discontinuity of model versions, where version 1 has changed to version 2. Furthermore, a new online version (v.2) was updated in the second semester of 2025. The version changes might introduce biases that come from slight differences in the grid cells or improvement of the SST and SSS estimations in the newest versions.

2.3. Access and Restrictions on data use

BarentsWatch lice and treatment input

BarentsWatch data is publicly accessible (https://www.barentswatch.no/fiskehelse/2025/45), and no restrictions are required to the extent of our knowledge, although proper source acknowledgement is recommended.

NorKyst-800 ocean model data

The data are publicly available on thredds (https://thredds.met.no/thredds/fou-hi/norkyst800v2.html) and the use of data should acknowledge the responsible studies (Asplin et al., 2020; Idžanović et al., 2017; Myksvoll et al., 2018; Röhrs et al., 2014). Besides, MET recommends that for the use of Norkyst-800 in operational applications we should contact them as the public repository might be unstable on some occasions. Furthermore, MET also recommends not downloading several files at once in parallel. However, for the EO4SA Use case we will have sufficient and relevant model data available.

3. Use Case 2: Forecasting Harmful Algal Blooms (HABs) Impacting Shellfish Farming in Norway

In consultation with the EAs the product of the use case has been changed from originally *Forecasting Harmful Algal Blooms (HABs) Impacting Shellfish Farming in Norway* to **Forecasting toxin-safe areas to harvest blue mussels.** This is reflected in the further deliverables, while the original use case title remains the same.

3.1. Input data

This section presents the input dataset necessary for forecasting toxin-safe areas to harvest blue mussels (*Mytilus edulis*) in Norway. This Use Case was previously named Forecasting toxic HABs but we changed the name as we changed the target and information given, although we are still addressing the same problem for the shellfish farmers (and recreational harvesting). Note that this change is due to the feedback and recommendations from early adopters. The main input data for developing the use cases is the Norkyst-800 model, assimilating EO data sources, and the in-situ algae toxin monitoring data (Figure 3) of Diathermic Shellfish Toxins (DST) and Paralytic Shellfish Toxins (PST).

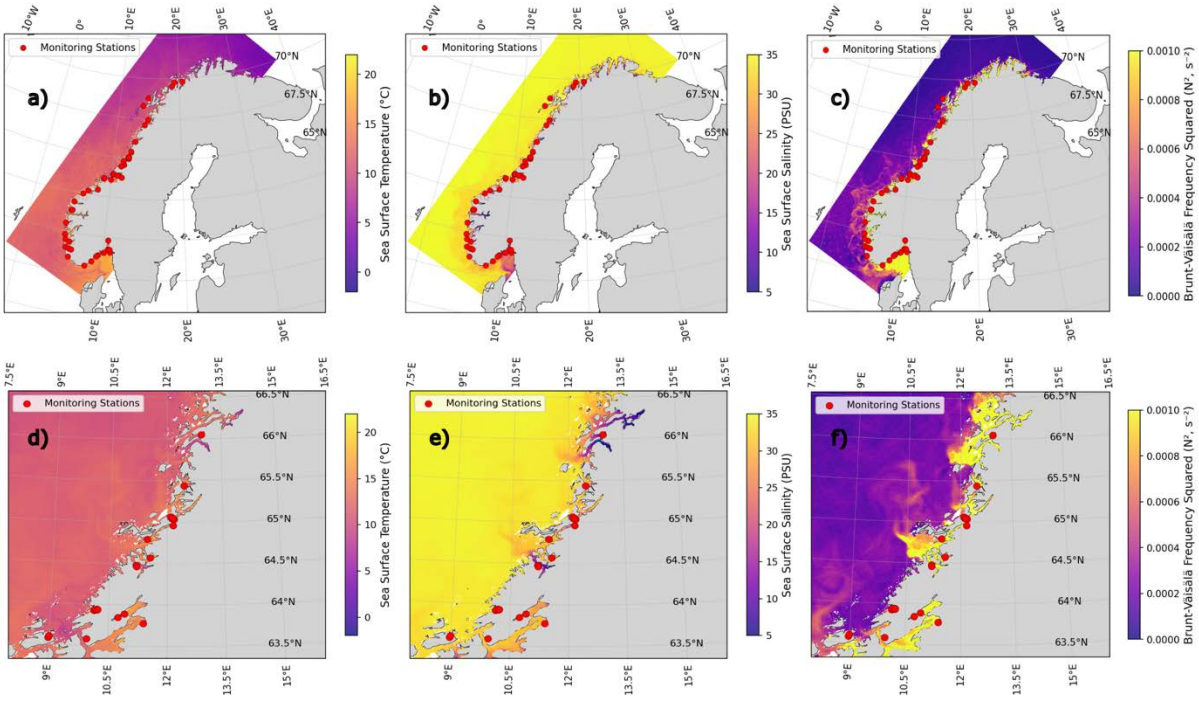

Figure 3. Norkyst-800 and monitoring stations with DST and PST concentration from mussel flesh. The subplots show the overall coverage of Norkyst-800 example for 2024-06-01 for SST (a), SSS (b), and N^2(C), and a zoomed area for SST (d), SSS (e), and N^2 (f). The algae toxin monitoring stations are shown as red dots.

Norkyst-800m input

The Norkyst-800 used for Use Case 2 is the same used for Use Case 1, and the same details can be found in the Section 2, Table 3, and Figure 2. The only difference of the Use case 2 is that we retrieve the Brunt–Väisälä frequency (N^2) using the density profiles and 25 m depth as a reference.

Toxin data

Paralytic Shellfish Toxins (PST) and Diarrhetic Shellfish Toxins (DST) are retrieved from monitoring locations in controlled areas, as required by the NFSA. Field observation data encompasses systematic toxin monitoring that is on average 32 stations every year since 2012 (Table 4). Considering the total temporal range of 2012-2024, a total of 54 distinct coastal locations is retrieved. The frequency of observations is mostly monthly and exclude the winter period, although some exceptions occur in the dataset. The dataset consists of sampling blue mussels on the surface area, which are send to a laboratory for assessing the toxin levels per kg mussel flesh. This dataset is used as an input only for training the models as the target variables but not as a predictor.

Table 4. Toxins input data summary.

Dataset Property	Description
Dataset Name:	Toxins.csv
Source	Norwegian Food Safe Authority
Format	CSV (Comma-Separated Values)
Size	3,800
Temporal Coverage	2012 – 2024, February to November
Temporal Resolution	Event-based (but mostly monthly)
Spatial Coverage	Norwegian coastal salmon aquaculture facilities
Coordinate System	WGS84 (EPSG:4326)

3.2. Data quality and other issues

Norkst-800 presents the same issues as discussed on page 12, and one should consult that section for more details. Considering the toxins.csv, the raw files on which toxins.csv was built present limited geographical information. Toxin observations are organised by year (2012-2024), location ID, and week of the year. Other files for the coordinates are organised in the same structure but are only available for 2010, 2015, 2020, and 2021. Hence, for the years with missing location files, we assumed the location is the same as the closest recent year. For example, observations of 2012 are registered with the location of 2010. We verified that IDs found in all files have some variability in the locations, but quite small and probably reflect slight changes in the location of the farm. Only one farm we found substantial change (Egersund), and we removed it from model training. Furthermore, some IDs present in years

such as 2012 are not present in the available years with coordinate information, and therefore they are removed from the final file (Toxins.csv). Furthermore, the depth of sampling the mussels are not clearly described. Although NFSA and EAs referring to surface samples, we are aware from previous collaborators in other projects some samples might be in deeper waters. The impact should be minimal but might slightly contribute with some uncertainties in the models.

3.3. Access and restrictions of data use

The Norkyst-800 model data are publicly available on thredds (https://thredds.met.no/thredds/fou-hi/norkyst800v2.html) and the use of data should acknowledge the responsible studies (Asplin et al., 2020; Idžanović et al., 2017; Myksvoll et al., 2018; Röhrs et al., 2014). Besides, MET recommends that for the use of Norkyst-800 in operational applications we should contact them as the public repository might be unstable on some occasions. Furthermore, MET also recommends not downloading several files at once in parallel. However, for the EO4SA Use case we will have sufficient and relevant model data available.

The Toxins.csv is made available by the NFSA to NERSC and only allowed for internal project research use, with further external sharing not allowed. However, NFSA may make the data available for research applications on justified requests.

4. Use Case 3: Optimising the location of sustainable shellfish farming and tourism in Galicia, Spain.

4.1. Input datasets

Case study area

The EO datasets for Case Study 3 were generated for two sites. The bounding boxes of these sites are shown in red in Figure 4. The first site is located in Rías Baixas, and the second site covers Ría de Ares–Betanzos.

Figure 4. Areas covered by the EO dataset in Case Study 3: Rías Baixas and Ría de Ares–Betanzos.

Sentinel-3 OLCI 300m resolution dataset

Sentinel-3A and Sentinel-3B OLCI 300 m datasets were processed using PML's in-house Polymer V16.1 processing chain. The product includes remote sensing reflectance (Rrs), chlorophyll-a concentration (Chl-a), and optical water types (OWT), as detailed in Table 5. Data from Sentinel-3A and Sentinel-3B were combined to create daily files. Chlorophyll-a concentration was estimated using the POLYMER algorithm (Steinmetz et al., 2011), the OC5 algorithm (Gohin et al., 2002), and the OCI algorithm (Hu et al., 2012). The OWTs were derived from the reflectance data using a fuzzy c-means clustering approach (Jackson et al., 2017).

Table 5. Description of the Sentinel-3 OLCI 300m dataset.

Dataset Property	Description
Dataset identity	Remote Sensing Reflectance, chlorophyll-a concentration in sea water, optical water types
Data format	NetCDF-4
Processing level	Level 3
Spatial Resolution	300m
Temporal Resolution	Daily
Timespan	September 2018 to August 2025
Spatial coverage	Rias Baixas, Ria de Ares-Betanzos
Data location	Plymouth Marine Laboratory
Produced by	Plymouth Marine Laboratory
Further Documentation	https://pml.ac.uk/projects/certo/ Atwood, E.C., Jackson, T., Laurenson, A., et al. 2024. Framework for Regional to Global Extension of Optical Water Types for Remote Sensing of Optically Complex Transitional Water Bodies. Remote Sensing 16, 17, 3267.
Size of dataset	34Gb

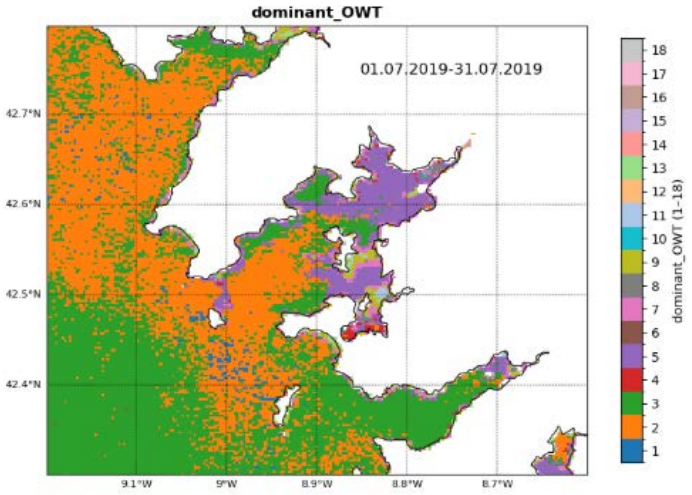

Figure 5. Example of a Sentinel-3A OLCI optical water type product (monthly composite) for Rias Baixas , acquired in July 2019.

Sentinel-2 MSI 60m resolution dataset

The Sentinel-2 MSI dataset at 60 m resolution was produced using a processing chain developed in the CERTO project (Copernicus Evolution: Research for harmonised and Transitional water Observation), coordinated by PML. CERTO MSI v1 uses the POLYMER algorithm for atmospheric correction to generate Rrs. Total suspended matter (TSM) and chlorophyll-a concentration (Chl-a) are calculated using the TOP2 and TOP3 algorithms, blended according to optical water types (OWT). Details of the Sentinel-2 MSI dataset are provided in Table 6.

Table 6. Description of a Sentinel-2 MSI 60m dataset.

Dataset Property	Description
Dataset identity	Remote sensing reflectance, chlorophyll-a concentration, total suspended matter, optical water types
Data format	NetCDF-4
Processing level	Level 3
Spatial Resolution	60m
Temporal Resolution	Daily
Timespan	January 2016 – December 2024
Spatial coverage	Rias Baixas, Ria de Ares-Betanzos
Data location	Plymouth Marine Laboratory
Produced by	Plymouth Marine Laboratory
Further Documentation	https://pml.ac.uk/projects/certo/ Atwood, E.C., Jackson, T., Laurenson, A., et al. 2024. Framework for Regional to Global Extension of Optical Water Types for Remote Sensing of Optically Complex Transitional Water Bodies. Remote Sensing 16, 17, 3267.
Size of dataset	21.8 Gb

Sea Surface Temperature (SST)

The Operational Sea Surface Temperature and Sea Ice Analysis (OSTIA) product (Good et al., 2020) provides high-resolution (1/20° – approximately 5 km) daily maps of average sea surface temperature (SST) at 20 cm depth for the global ocean. To determine surface temperature, OSTIA uses satellite data supplied by the GHRSST project together with in situ observations. The product is used as input to the primary production (PP) model. Details of the SST dataset are provided in Table 7.

Table 7. Description of the sea surface temperature dataset.

Dataset Property	Description
Dataset identity	Global Ocean OSTIA Sea Surface Temperature
Data format	NetCDF-4
Processing level	Level 4
Spatial Resolution	0.05 degree (≈5.5km at equator)
Temporal Resolution	Daily
Timespan	November 1997 to September 2023
Spatial coverage	Global Ocean Lat -89.97° to 89.97°Lon -179.98° to 179.98°
Data location	https://data.marine.copernicus.eu/product/SST_GLO_SST_L4_NRT_OBSERVATIONS_010_001/description
Produced by	SST-METOFFICE-EXETER-UK
Further Documentation	Good, S., Fiedler, E., Mao, C., Martin, M.J., Maycock, A., Reid, R., Roberts-Jones, J., Searle, T., Waters, J., While, J., Worsfold, M. The Current Configuration of the OSTIA System for Operational Production of Foundation Sea Surface Temperature and Ice Concentration Analyses. Remote Sens. 2020, 12, 720, doi:10.3390/rs12040720.
Size of dataset	251 Gb

Satellite derived Photosynthetically Active Radiation (PAR)

An innovative product has been developed based on a robust algorithmic framework for daily PAR estimation from ENVISAT/MERIS and Sentinel-3/OLCI. Several original methods have been implemented in the PAR dataset to correct known issues, such as the diurnal variation of cloudiness, and the product is complemented with secondary, experimental variables, including UV irradiance above and below the sea surface and spectral PAR. This product is used as input to the primary production (PP) model. Details of the PAR dataset are provided in Table 8.

Table 8. Description of PAR dataset.

Dataset Property	Description
Dataset identity	OLCI-A daily PAR
Data format	NetCDF-4
Processing level	Level 3
Spatial Resolution	0.041 degree (≈4.6km at equator)
Temporal Resolution	Daily
Timespan	May 2016 to December 2024
Spatial coverage	Global Ocean
Data location	Hygeos
Produced by	Hygeos
Further Documentation	https://seom.esa.int/page_project019.php
Size of dataset	456 Gb

4.2. Validation datasets

Hyperboost dataset

This dataset comes from a 2 year sampling cruise around the European coastal waters (Figure 6, www.hyperboost.info). The relevance of this dataset is that it is representative of the optical conditions of the area of study and contains all the required variables. So it will be used to validate the Stramski et al. (2022).

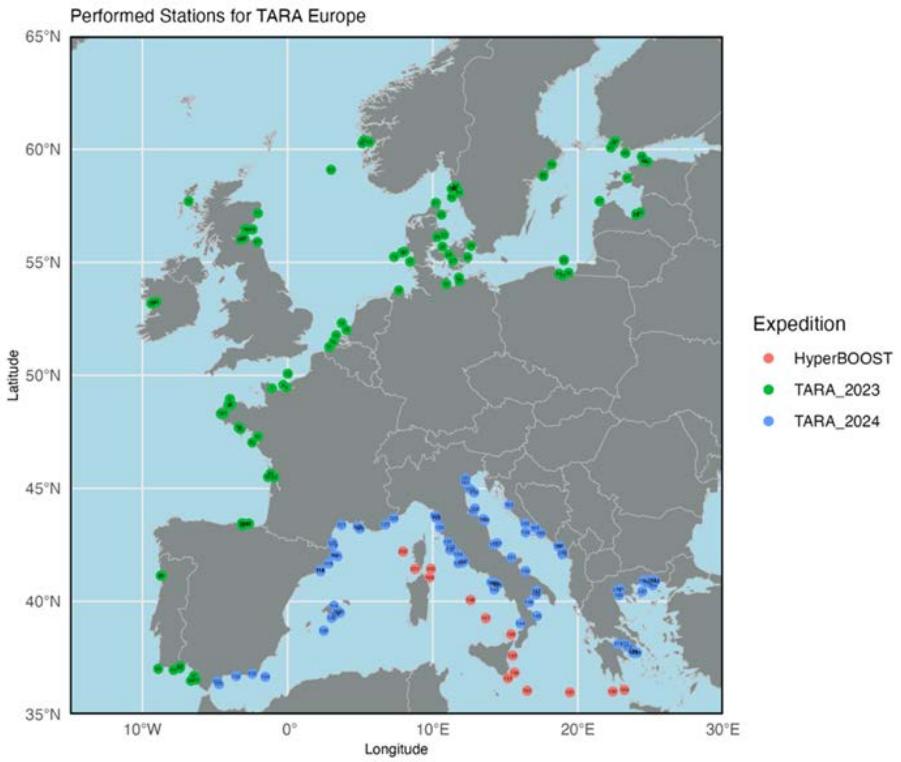

Figure 6. Location of samples from the HyperBoost cruise. Despite not having exact locations around the area of the Galician rias, coastal bio-optical relationships will be explored between in situ data and reflectance to compare with the products developed in EO4SA.

Table 9. Description of A Coruña dataset.

Dataset Property	Description
Dataset identity	Daily samples around the European Coastal Waters
Measured parameters	POC, SPM, Chla , Remote Sensing Reflectance
Data format	.csv
Frequency of observations	daily
Temporal coverage	2023-2024
Data location	https://doi.org/10.5281/zenodo.15849296
Produced by	Various , hyperboost project
Restrictions of use	Only for internal use
Further Documentation	HyperBOOST_D2.1_Final in-situ dataset and Dataset description_v3.doc

CSIC dataset in Galician Rias

Information about the in-situ concentration of particulate organic matter (POM) or particulate organic carbon (POC) in the Galician rías (NW Iberian upwelling system) since the Sentinel satellites family was launched (in 2016) is scarce and concentrates in the Rias Centrais, mainly A Coruña and Ares-Betanzos. In A Coruña, the CSIC Instituto Español de Oceanografía (Spanish Institute of Oceanography), under the leadership of Prof. Antonio Bode, has maintained two time series stations since 1989 (see Figure 7). The sites, one within the Ría the A Coruña (E4CO) and the other on the inner shelf (E2CO) are visited monthly, collecting water samples from 5 depths through the photic layer. POC is measured at these sites, together with PON, Chlorophyl a, primary production, inorganic nutrients, salinity and temperature. The time series from 1989 to 2016 is publicly available, and it is incorporated in the ESA project PRIMUS open-access dataset. Data from 2016 to 2020 has been kindly shared by Prof. Antonio Bode for internal use only. Note that mussels are not cultured in the Ría de A Coruña. On the contrary, the 3 time-series stations in the Ría de Ares-Betanzos (Figure 4), are located within the mussel farming polygons of Arnela and Lorbé. In fact, samples are taken next to the mussel rafts. The mussel farming company PROINSA is the owner of this dataset, which can only be used for the purpose of validating satellite images and cannot be publicly disseminated. The four points have been sampled (at 3 m) with weekly to bi-weekly frequency, from 2007 to 2019. POM, particulate inorganic matter (PIM), suspended particulate matter (SPM), Chlorophyl a, salinity and temperature were measured. Details of two in-situ datasets can be found in Table 10 and Table 11.

Figure 7. Positions of the Ría de Ares-Betanzos and A Coruña observation stations.

Table 10. Description of A Coruña dataset.

Dataset Property	Description
Dataset identity	Monthly POC and PP profiles in two stations into Ría de A A Coruña (E4CO) and adjacent inner shelf (E2CO).
Measured parameters	Salinity, temperature, Chl-a, POC, PON, PP, inorganic nutrients
Data format	.csv
Frequency of observations	monthly
Temporal coverage	1989-present
Data location	Ría de A Coruña (E4CO), inner shelf off Ria de A Coruña (E2CO)
Produced by	Prof. Antonio Bode, IEO-CSIC A Coruña, RADIALES program
Restrictions of use	Data 1989-2016 (PRIMUS open-access dataset, 2016-present, only for internal use
Further Documentation	

Table 11. Description of the Ría de Ares-Betanzos dataset.

Dataset Property	Description
Dataset identity	1-2 wk surface (3 m) POM, PIM and SPM values in the mussel farming areas of Arenla and Lorbé (Ria de Ares-Betanzos).
Measured parameters	Salinity, temperature, Chlorophyl a, POM, PIM and SPM
Data format	.csv
Frequency of observations	Weekly to biweekly
Temporal coverage	2007 to 20219
Data location	Mussel farming areas of Arnela and Lorbé (Ría de Ares-Betanzos)
Produced by	PROINSA
Restrictions of use	Only internal use
Further Documentation	

4.3. Data quality and other issues

Hyperboost dataset is a high quality data, with harmonised data collection and analysis, it is complete with in situ radiometry so allows the characterisation of the algorithm uncertainty without taking into account satellite data processing. The A Coruna dataset was also collected by a scientific expert, however it is not complete (no optics), so the comparisons will have to be made with satellite products. Finally, the PROINSA dataset has been collected within the area of the mussel rafts, and it may have the largest uncertainties when compared to satellite, in particular the measurements are of POM (particulate organic matter) which is not a chemical measurement like POC (particulate organic Carbon), which may lead to difficulties with interpretation of the results. However, these data are still valuable to link with Early adopter's real life operational observations.

4.4. Restrictions on data use

The EO data described in Section 4.1 Input Datasets are publicly available and freely distributed by PML. The in situ validation dataset for the Galician Rías (1989–2016) is part of the ESA PRIMUS open-access dataset. The time series from 2007 to 2019 for the three stations in the Ría de Ares–Betanzos are owned by the company PROINSA and are not publicly accessible. These data are used solely for validation purposes.

5. Use Case 4: Mapping aquaculture structures and use of marine resources in Palawan, Philippines

5.1. Input datasets

Case study area

The developed input dataset covers three case study areas: Puerto Princesa City, Honda Bay and Ulugan Bay,in the Philippines, as outlined in Figure 8.

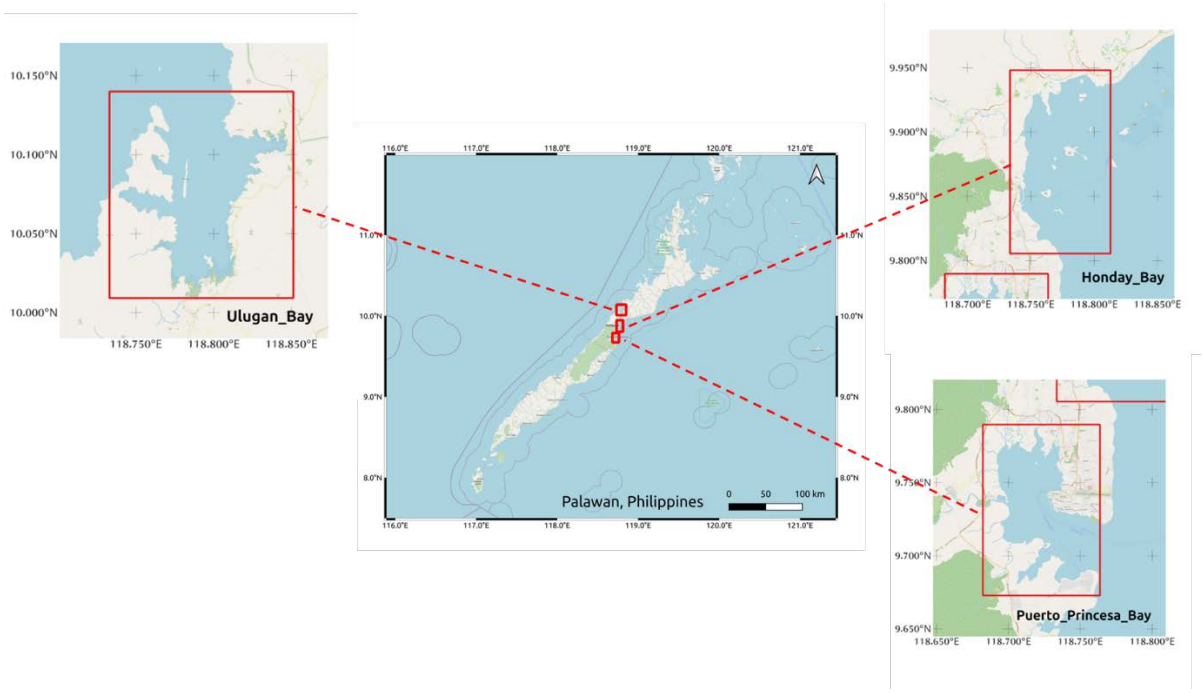

Figure 8. Geographical extent of the Case Study 4 areas covered by the input EO dataset.

Sentinel-1 SAR dataset

A Sentinel-1 SAR Single Look Complex (SLC) product contains focused SAR images provided in slant-range geometry. Each image pixel is represented by complex (I and Q) values, containing both amplitude and phase information. The images are georeferenced using satellite orbit and attitude data. The data are provided as Sentinel-1 SLC bursts, which are segments of radar echoes acquired by cyclically switching the antenna beam. Details of the Sentinel-1 SAR SLC dataset are provided in Table 12.

Table 12. Description of the Sentinel-1 SAR dataset.

Dataset Property	Description
Dataset identity	Focused SAR image, Interferometric Wide Swath (IWS) mode with radar signal transmitted at vertical (V) polarisation and received at both vertical and horizontal (H) polarisations. Single Look Complex (SLC) data format.
Data format	SAFE
Processing level	Level 1
Spatial Resolution	2.3 m (range) × 14.1 m (azimuth) (varies slightly per sub-swath)
Temporal Resolution	around 12 days (for the case study area)
Timespan	2021 to present
Spatial coverage	Puerto Princesa City, Honda Bay, Ulugan Bay
Data location	CDSE: https://catalogue.dataspace.copernicus.eu/odata/v1/ ASF: https://api.daac.asf.alaska.edu/services/search/
Produced by	the European Space Agency (ESA) as part of the Copernicus Programme
Further Documentation	https://sentiwiki.copernicus.eu/web/s1-products https://documentation.dataspace.copernicus.eu/APIs/Sentinel-1%20SLC%20Burst.html
Size of dataset	574 Gb

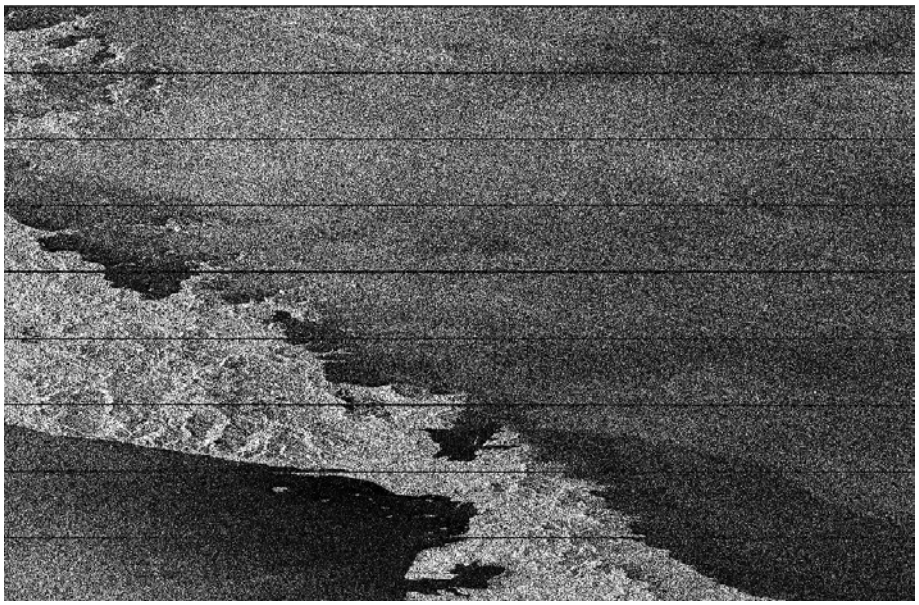

Figure 9. Example of a Sentinel-1A SAR SLC product (intensity at VV polarisation, IW mode, sub-swath 1), showing intensity derived from complex data, acquired on 11 October 2021.

Sentinel-2 MSI dataset

The Sentinel-2A MSI Level-2A dataset provides atmospherically corrected surface reflectance images derived from Level-1C data. It also includes additional products such as scene classification, aerosol optical thickness, and water vapor. The images are orthorectified to ensure accurate geolocation. Details of the Sentinel-2A MSI Level-2A dataset are provided in Table 13.

Table 13. Description of the Sentinel-2 MSI dataset.

Dataset Property	Description
Dataset identity	Surface reflectance data (bottom of atmosphere), scene classification, aerosol optical thickness and water vapour.
Data format	SAFE
Processing level	Level 2
Spatial Resolution	10x10m, 20x20m, 60x60m
Temporal Resolution	around 10 days (for the case study area)
Timespan	2021 to present
Spatial coverage	Puerto Princesa City, Honda Bay, Ulugan Bay
Data location	CDSE: https://catalogue.dataspace.copernicus.eu/odata/v1/
Produced by	the European Space Agency (ESA) as part of the Copernicus Programme
Further Documentation	https://sentinels.copernicus.eu/sentinel-data-access/sentinel-products/sentinel-2-data-products/collection-1-level-2a
Size of dataset	212 Gb

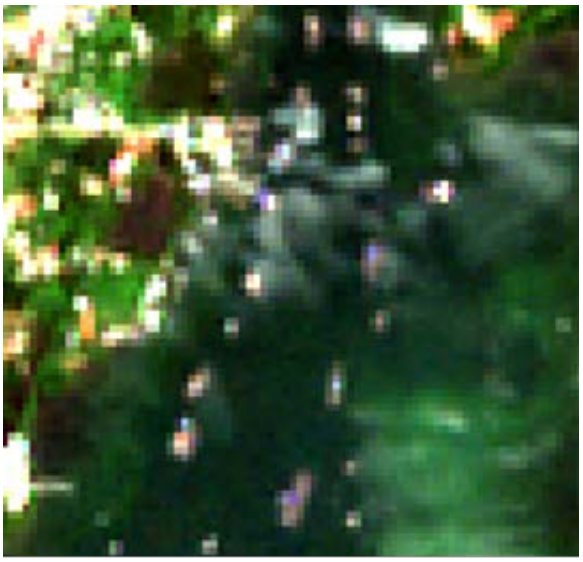

Figure 10. Example of a Sentinel-2 MSI surface reflectance product (10m resolution, RGB representation) showing the Puerto Princesa City site on 17 October 2023.

5.2. Validation dataset

WorldView VHR dataset

The dataset contains Maxar Standard Satellite Imagery products from WorldView satellites and covers three case study areas. The images are orthorectified and map-ready. They include a combination of spectral bands: four multispectral bands (BLUE, GREEN, RED, NIR1) and a panchromatic band (PAN). Details of the dataset are provided in Table 1414.

Table 14. Description of a WorldView VHR dataset.

Dataset Property	Description
Dataset identity	High Res Optical: Panchromatic and 4-bands, map ready (Ortho) 1:12.000 orthorectified.
Data format	GeoTIFF
Processing level	Standard(2A)/View Ready Standard (OR2A)
Spatial Resolution	1.2m multispectral, 0.4x0.4m panchromatic
Temporal Resolution	N/A
Timespan	08.10.2023, 07.10.2023, 14.02.2025
Spatial coverage	Puerto Princesa City, Honda Bay, Ulugan Bay
Data location	Maxar Technologies
Produced by	Maxar Technologies, ordered through ESA's Third Party Missions programme
Further Documentation	https://earth.esa.int/eogateway/catalog/worldview-3-full-archive-and-tasking https://earth.esa.int/eogateway/documents/20142/37627/Map-Ready-imagery-data-sheet.pdf/83b9a034-5548-1fc7-4744-8d0fdb922f52
Size of dataset	11 Gb

Figure 11. Example of a WorldView multiband product (1.2m resolution, RGB representation) showing the Puerto Princesa City site on 8 October 2023.

5.3. Data quality and other issues

Sentinel-1 SAR dataset

SAR images are affected by speckle, a granular, noise-like texture that can hinder the detection of aquaculture objects and cause false alarms. Speckle arises from the coherent interaction of microwaves with scatterers, which combine with different phases. The intensity of speckle is higher in Sentinel-1 SLC products, and special measures have been implemented to suppress speckle for the CS4 application (see D3, Section 5.3, Methodology Description).

Sentinel-2 MSI dataset

The application of Sentinel-2 MSI data for aquaculture mapping is affected by several quality issues, which can be summarized as:

- Clouds and cloud shadows creating gaps in Sentinel-2 data.
- Classification errors in cloud masking leading to artifacts in masked images.
- Partially detected cloud shadows impacting reflectance data.
- Sun glint errors.

These adverse factors have been successfully mitigated by compositing time series of Sentinel-2 images over a long time interval before applying them for land–water masking and cage detection. This approach minimized the effects of sun glint and other artifacts in the composite images (for details, see D3, Section 5.3, Methodology Description). Additionally, to reduce the influence of clouds, only Sentinel-2 scenes with cloud cover below 15% were used.

Worldview VHR dataset

The quality of WorldView VHR data is affected by atmospheric effects, including haze, aerosols, and clouds. To minimize these adverse factors, we carefully inspected the ordered WorldView scenes to ensure they were free of clouds and other artifacts.

5.4. Restrictions on data use

Sentinel-1 SAR and Sentinel-2 MSI data

Sentinel-1 and Sentinel-2 data are provided under the Copernicus Open Access Licence. The data are publicly accessible to both registered and non-registered users without restrictions. They can be used for a wide range of applications, and it is permitted to adapt, modify, and combine them with other data sources.

WorldView VHR data

The data are distributed by ESA on behalf of the data owner, after an application from the EO4SA project. They are owned and licensed by Maxar Technologies, and their use is governed by strict commercial licenses. Sharing these data outside the project is not permitted. These restrictions do not affect the output aquaculture maps, as the WorldView data are used solely for validation purposes (see D3, Section 5.5, Accuracy) and are not publicly distributed further by the project.

6. Conclusions

The Input Data Inventory Database, collected in the EO4SA project, includes EO data acquired by different sensors and in-situ measurements from various sources. It is geographically divided into two parts: Use Cases 1 and 2 in one part and Use Cases 3 and 4 in the other. These parts are stored at NERSC and PML facilities, respectively. The database is used for the development, validation, and analysis of EO4SA products and algorithms, as described in EO4SA *Deliverable D3: Algorithm Theoretical Basis Documents (ATBD)*.

7. References

Atwood, E.C., Jackson, T., Laurenson, A., et al. 2024. Framework for Regional to Global Extension of Optical Water Types for Remote Sensing of Optically Complex Transitional Water Bodies. Remote Sensing 16, 17, 3267

Gohin, F., Druon, J.N., and Lampert, L. 2002. A five channel chlorophyll concentration algorithm applied to SeaWiFS data processed by SeaDAS in coastal waters. International Journal of Remote Sensing 23, 8, 1639–1661.

Good, S., Fiedler, E., Mao, C., Martin, M.J., Maycock, A., Reid, R., Roberts-Jones, J., Searle, T., Waters, J., While, J., Worsfold, M. The Current Configuration of the OSTIA System for Operational Production of Foundation Sea Surface Temperature and Ice Concentration Analyses. Remote Sens. 2020, 12, 720, doi:10.3390/rs12040720.

Hu, C., Lee, Z., and Franz, B. 2012. Chlorophyll algorithms for oligotrophic oceans: A novel approach based on three-band reflectance difference. Journal of Geophysical Research: Oceans 117, C1.

Jackson, T., Sathyendranath, S., and Mélin, F. 2017. An improved optical classification scheme for the Ocean Colour Essential Climate Variable and its applications. Remote Sensing of Environment 203, 152–161.

Steinmetz, F., Deschamps, P.-Y., and Ramon, D. 2011. Atmospheric correction in presence of sun glint: application to MERIS. Optics Express 19, 10, 9783.

Stramski, D., Joshi, I., & Reynolds, R. A. (2022). Ocean color algorithms to estimate the concentration of particulate organic carbon in surface waters of the global ocean in support of a long-term data record from multiple satellite missions. Remote Sensing of Environment, 269, 112776. https://doi.org/10.1016/j.rse.2021.112776.

------- END of Document -------