Project	MONOCLE H2020 (grant 776480)	Start / Duration	1 February 2018/ 48 Months
Dissemination	PUBLIC	Nature	REPORT
Date	20/11/2018	Version	1.1

MONOCLE

Multiscale Observation Networks for Optical monitoring of Coastal waters, Lakes and Estuaries

Deliverable 7.1

Report on existing EO services, existing in situ data resources and protocols

Project Description

Funded by EU H2020 MONOCLE creates sustainable *in situ* observation solutions for Earth Observation (EO) of optical water quality in inland and transitional waters. MONOCLE develops essential research and technology to lower the cost of acquisition, maintenance, and regular deployment of *in situ* sensors related to optical water quality. The MONOCLE sensor system includes handheld devices, smartphone applications, and piloted and autonomous drones, as well as automated observation systems for e.g. buoys and shipborne operation. The sensors are networked to establish interactive links between operational Earth Observation (EO) and essential environmental monitoring in inland and transitional water bodies, which are particularly vulnerable to environmental change.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 776480

Project	MONOCLE H2020 (grant 776480)	Start / Duration	1 February 2018/ 48 Months
Dissemination	PUBLIC	Nature	REPORT
Date	20/11/2018	Version	1.1

Deliverable Contributors:	Name	Organisation	Role / Title
Deliverable Leader	Liesbeth De Keukelaere	VITO	WP7 Lead
Contributing Author(s)	Jaume Piera	CSIC	Contributor
	Andrew Tyler	USTIR	Contributor
	Caitlin Riddick	USTIR	Contributor
	Evangelos Spyrakos	USTIR	Contributor
	Thomas Jackson	PML	Contributor
	Sindy Sterckx	VITO	Contributor
Reviewer(s)			
Final review and approval	Stefan Simis	PML	Project Coordinator

Document History:

Release	Date	Reason for Change	Status	Distribution
1.0	31/07/2018	Initial document, reviewed	Completed	EC
1.1	20/11/2018	Digital Object Identifier added	Released	Public

To cite this document:

De Keukelaere, L., Piera, J., Tyler, A., Riddick, C., Spyrakos, E., Jackons, T., Sterckx, S. (2018). D7.1 Report on existing EO services, existing in situ data resources and protocols. *Deliverable report of project H2020 MONOCLE (grant 776480)*, doi: 10.5281/zenodo.1492283

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 776480

TABLE OF CONTENTS

1. Executive Summary 4
2. Scope 4
3. Introduction 4
4. Existing EO services 5
   4.1 Copernicus Marine Environment Monitoring Service (CMEMS) 5
   4.2 Copernicus Global Land Service (CGLS) 6
5. Existing in situ resources 6
   5.1 Fixed platforms 7
   5.2 Buoy data 7
   5.3 Moving platforms 9
6. In situ open observation databases 11
7. Integration techniques 13
   7.1 In situ data for algorithm validation and calibration 13
   7.2 In situ data for EO product validation 15
   7.3 Protocols for in situ data integration in EO services 17
   7.4 Existing gaps and MONOCLE solutions 18
8. Conclusions and recommendations 19
9. Exploitation and Dissemination 19
10. References 19

1. Executive Summary

This report summarizes existing EO services, validation resources, best practices/protocols and the methods employed to integrate multiple sources of in situ reference data with EO data processing. A wide variety of in situ data sampling techniques are currently available from fixed platforms (e.g. AERONET-OC or CEFAS buoys) or moving platforms (e.g. boats, ships and increasingly remotely operated aircraft), collected in various databases. Despite these efforts, many waterbodies and most inland waters remain under-sampled hampering comprehensive validation of satellite remote sensing of water quality variables. Lowering the cost of sensors, deployment methods and including citizen science have a high potential to fill these observation gaps.

To integrate existing and new in situ data into EO services, they have to be accessible in a clearly structured way. A pre-defined data format with metadata, containing information on source, location, calibration and applied algorithms, allows easy interpretation and traceability of the data. Quality control checks and flagging suboptimal observations are required to ensure their correct use.

2. Scope

This report summarizes existing EO services, validation resources, best practices/protocols and current methods employed to integrate in situ reference data that are relevant to MONOCLE. This report is publicly available for benchmarking of newly developed methods. Within the MONOCLE project, this report provides guidelines for development and testing in WP4 (Field Testing), WP5 (System interoperability and data integration), WP6 (Signal and image analysis) and will lead to the selection of a use case (T7.2) to highlight the added value of MONOCLE for EO validation or improvement of EO products in WP7 (In situ data integration in EO services).

3. Introduction

This document provides an overview of relevant EO services and in situ datasets suitable for EO product validation. A description is given on current practices to combine both data sources and any shortcomings are identified.

For the EO services, two large public services running under the Copernicus framework are considered. These are the Copernicus Marine Environment Monitoring Service and the Copernicus Global Land Service. For both services we consider only components delivering optical water quality as these relate to the challenge addressed in MONOCLE for better in situ validation and calibration. These services are intended to provide access to the state-of-the-art in free and documented satellite products and are therefore expected to have a wide global impact on management of water resources. For the in situ platforms this document focusses on methods that support a high degree of automation, and with various degrees of automated quality control. Such methods have increasingly been used to complement manually collected observation data from research cruises, speeding up the process of satellite validation. These systems are more common today in the marine environment than in inland waterbodies.

In terms of data integration methods, present methods focus on techniques to match in situ observations with (near-) coinciding satellite observations.

The current state-of-the-art for each of the above elements is described in the following sections.

4. Existing EO services

Two examples of EO services are described and investigated in terms of water quality validation: the Copernicus Marine Environment Monitoring Service (CMEMS) and the Copernicus Global Land Operations Lot-2: "Water & Cryosphere" (CGLOPS-2). Together, these services provide large global coverage of both inland and marine optical water quality.

4.1 Copernicus Marine Environment Monitoring Service (CMEMS)

The Copernicus Marine Environment Monitoring Service (CMEMS) has been designed to respond to issues emerging in the environmental, business and scientific sectors. It provides state-of-the-art analyses and forecasts daily, which offer an unprecedented capability to observe, understand and anticipate marine environment events. CMEMS provides regular and systematic reference information on the physical state, variability and dynamics of the ocean and marine ecosystems for the global ocean and the European regional seas. The CMEMS delivers a core information service to any user related to 4 areas of benefit: Maritime Safety, Coastal and Marine Environment, Marine Resources, and Weather, Seasonal Forecasting and Climate activities. The CMEMS has been working in operational mode since May 2015, providing a reliable, free service, providing a single access point to a catalogue of products. The service is meant for any user requesting generic information on the ocean, and especially downstream service providers who use this information as an input to their own value-added services to end-users. The observations and forecasts produced by the service support a range of marine applications.

Details of the CMEMS catalogue can be found at http://marine.copernicus.eu/wp-content/uploads/2016/06/r2421_9_catalogue_services.pdf

In the context of MONOCLE the most relevant elements of the catalogue are the global and regional products for reflectance, chlorophyll-a concentration, turbidity, suspended particulate matter (SPM), and Sea Surface Temperature (SST). The global products are available at resolutions of up to 4km and the regional products are available at resolutions down to 1km.

The CMEMS EO service comprises a component called Copernicus In Situ Thematic Assemble Centre (INS-TAC) which ensures consistent and reliable access to a range of in situ data for the purpose of service production and validation. The in situ observations are on-site local measurements of sea water properties using sensors on board of a wide range of platforms from autonomous observations at sea (e.g. floats, drifting buoys, gliders, moorings, ferrybox, marine mammals) and ships of opportunity or research ships, and allow to capture ocean variability at different scales. INS TAC provides two types of products:

- The Near Real Time (NRT) products are automatically quality controlled within 24 hours from acquisition.
- The reprocessed products are built for reanalysis activities or Climate research and provide integrated products over the past 25 to 50 years. They are assessed by scientific team and all suspicious measurements are checked to discriminate sensor anomaly from real ocean signal.

4.2 Copernicus Global Land Service (CGLS)

The Copernicus Global Land Service covers Vegetation, Energy, Cryosphere, Hot Spots, and Water, providing public access to state-of-the-art products derived from EO sensors. Under the Water category it currently provides Lakes Surface Water Temperature (LSWT), Lake Water Quality (LWQ), Water Bodies and Water Level. LSWT and LWQ are of primary relevance to MONOCLE. LSWT is produced by the University of Reading, and LWQ by PML. Both products are managed by Brockmann Consult who also subcontract additional partners for specific product evolution. The LWQ products include Lake Surface Reflectance (LSR) at all available satellite bands that were available on MERIS (2002-2012) and OLCI (2016 onward). Trophic State Index (TSI) and Turbidity products are derived from the LSR and individually calibrated. Documentation and algorithms are available through https://land.copernicus.eu/global/products/lwq.

In summary, the processing chain for LWQ includes daily segmentation of new OLCI imagery to the areas around >1000 lake sites, distributed globally. Every 10 days, the results are aggregated and mapped to a global grid. LSR is provided as the spectrum that is most representative for the period, based on the lowest mean absolute percentage difference between a given observation and the median spectrum for the 10-day period. The Trophic State Index (derived from Chlorophyll-a observations) and Turbidity are provided as 10-day averages.

The Service currently provides data derived from Sentinel-3A at 300m and 1km resolutions. Evolutions to include Sentinel-2 at 100m have started and it is expected that Sentinel-3B will be included when the sensor has passed in-flight calibration.

The processing chain supporting LWQ is *Calimnos* (v1.2) which was developed for the UK GloboLakes project and since then adapted for operational processing for the CGLS. *Calimnos* is based on the principle of detecting optical water types, and using a mapping of suitable algorithms to each water type. Thus, algorithms that perform best for a given type (e.g. humic, turbid, clear, or productive waters) are automatically selected. The algorithms are validated and tuned against the LIMNADES database held at the University of Stirling. The algorithm calibration/validation is therefore based on the global collection of in situ data for a given water type, and therefore considered suitable to address the remote sensing of poorly characterized (or seasonally under-sampled) waterbodies.

5. Existing in situ resources

Anno 2018, a wide variety of in situ data exists which can be addressed for validation or calibration of Earth Observation products. This section describes the most relevant sources and databases of in-situ data currently available.

5.1 Fixed platforms

Sensors mounted on fixed platforms provide continuous data, measured in a consistent way over time. These time series of in situ data yield important information on temporal variation and dynamics taking place in the water. Once positioned, they are not intended to be relocated except for sensor calibration. In this category, we discuss the Aerosol Robotic Network (AERONET) and the Ocean Colour extension (AERONET-OC). AERONET data is used regularly in the validation of atmospheric correction algorithms applied on EO data. Next to AERONET, buoy data will be addressed, which are very used in validating EO water quality products. The AERONET(-OC) data are relevant in MONOCLE in comparison against new sensor solutions for Reflectance and Atmospheric Optical Depth.

AERONET (– OC)

The Aerosol Robotic Network (AERONET), developed by Aeronautics and Space Administration (NASA), is a worldwide network containing autonomous sun-photometer measurements to sustain and support atmospheric studies (Zibordi et al., 2009). AERONET measures, amongst other parameters, the aerosol optical thickness (AOT) at different wavelengths. This is a major asset in optical remote sensing since light detected by a satellite sensor has to pass the atmosphere twice (sun – target and target – satellite). AOT is rarely known, hard to retrieve from optical satellite imagery and has an important contribution in the atmospheric correction, particularly over water.

The AERONET network has been extended in AERONET-OC to support marine applications, by providing additional capability of measuring the radiance emerging from the sea (i.e. water –leaving radiance) with modified sun-photometers installed on offshore platforms. AERONET-OC is instrumental in satellite ocean colour validation activities through standardized measurements (i) performed at different sites with a single measuring system and protocol, (ii) calibrated with an identical reference source and method, and (iii) processed with the same code. Version 3 AERONET data are computed at three data quality levels: Level 1.0 (unscreened), Level 1.5 (cloud-screened and quality controlled), and Level 2.0 (quality-assured). Currently the radiance data are measured in a few discrete narrow bands (corresponding to the SeaWifs satellite sensor), which do not cover the broader spectral range of current satellite missions.

5.2 Buoy data

CEFAS SmartBuoys

CEFAS SmartBuoys are autonomous water quality monitoring buoys located at key locations around the UK. All SmartBuoys use the same basic suite of sensors based on the Cefas ESM2 data logger. Two sets of instruments are used at each SmartBuoy site, with maintenance cruises deployment a "fresh" set and recovering the previous buoy. Prior to 2014 SmartBuoys operated on a 10-minute measurement burst every 30 minutes. Since 2014 this was reduced to a 5-minute burst to enable longer endurance.

Marine Optical Buoy

The Marine Optical Buoy (MOBY) is an autonomous optical buoy moored off the island of Lanai in Hawaii, and it is supported by the National Oceanic and Atmospheric Administration (NOAA, USA) to provide vicarious calibration of ocean colour satellites. The MOBY buoy is 15 meters tall floating vertically in the water with approximately 3 meters above the surface and 12 meters below. It has 3 underwater arms (top, middle and bottom arms at approximately 1, 5 and 9 m depth respectively). Each arm has an upwelling radiance (Lu) sensor and a downwelling irradiance (Ed) sensor. Water-leaving radiance (Lw) is calculated by propagating Lu measurements to just below and then across the surface. Additionally, above-water downwelling irradiance (Es) is measured from sensors mounted on top of the buoy. MOBY radiometry data are regularly used by the NASA Ocean Biology Processing Group (OBPG) as part of ocean colour validation and vicarious calibration activities.

Quality Control: Several tests, shown in Table 1, are used to flag erroneous data. If any of the following tests fail the listed limits then all output is flagged and not used for either validation or vicarious calibration.

Table 1 Quality error flags (source https://seabass.gsfc.nasa.gov/wiki/MOBY)

Flag	Description
Lw_RMS <= 5	The RMS (root mean square) of the percent error between Lw computed using all 3 arms and Lw computed using just the top 2 arms
Es_RMS <= 10	The RMS of the percent error between Es (surface irradiance measurements) and underwater Ed measurements extrapolated to Es
ES_stability <= 10	The percent error between the min and max measured Es
TheoryES_diff <= 15	The RMS difference between modeled clear sky Es and the measured Es between 425-575 nm.
Sensor Tilt and Roll <= 5 degrees	

MOBY has generated calibrated measurements of ocean colour at the sea surface since 1996. MOBY served as the primary sea surface calibration for satellite borne sensors such as the sea-viewing wide field-of-view sensor (SeaWiFS) and the moderate-resolution imaging spectroradiometer (MODIS). MOBY has contributed to the calibration of the Ocean Color and Temperature Sensor (OCTS), the polarization detection environmental radiometer (POLDER), and the Modular Optoelectronic Scanner (IRS1-MOS).

CoASTS

The Coastal Atmosphere and Sea Time Series (CoASTS) project (Zibordi et al. 2002) have provided a time series of bio-optical marine and atmospheric measurements performed from an oceanographic tower in the northern Adriatic Sea, near Venice. The data set collected (Berthon et al. 2002) includes spectral measurements of the in-water AOP (diffuse attenuation coefficient, reflectance, Q-factor, etc.) and IOP (absorption and scattering coefficients), as well as the concentrations of the main optical components (pigment and suspended matter concentrations).

BOUSSOLE

The buoy named “Bouée pour l’acquisition de Séries Optiques à Long Terme” (BOUSSOLE) was deployed in the Mediterranean Sea between the French Riviera coast and Corsica. One general objective is to perform match-up analyses and vicarious calibration experiments for ocean colour satellite sensors. In parallel to these operational objectives, the assembled data set was used for more fundamental studies in marine optics and bio-optics. The BOUSSOLE buoy was specifically designed to perform radiometric measurements at sea with the following objectives (Antoine et al., 2008): (1) to measure the upward and downward plane irradiances (Eu and Ed, respectively), and the upwelling radiance at nadir (Lu) at two depths in the water column, plus the above-surface downward irradiance (Es), (2) to minimize the shading of the radiometers while maximizing their stability (i.e., keeping them as horizontally level as possible), and finally (3) to allow deployment at a deep-water site with swells up to 8m (but low currents).

5.3 Moving platforms

Moving platforms cover a wide range of in situ data collection, including measurements from boats, Ferryboxes, gliders, drifting profilers, drones and even sensors mounted on surfboards (http://www.euronews.com/2018/04/19/sentinel-ocean). These datasets do not have a fixed temporal frequency and cannot always meet the high accuracy standards of fixed platforms, but they allow a larger flexibility in sampling location, and often at much lower cost due to better accessibility for maintenance.

Boats

Measurement by boat is a widely used approach for inland, coastal and marine surveys, and provides flexibility in terms of survey timing and sampling locations (Figure 1). A planned field campaign by boat allows for data collection during optimal environmental conditions and for dedicated experiments and testing. Sampling locations or transects can be chosen as desired, especially across gradients of change in bio-geo-optical properties or to assess the impact of adjacency effects on remote sensing reflectance. Depth profiles can also be taken for the characterisation of the underwater light climate (e.g. light attenuation, absorption and scattering properties). *In situ* water samples for validation of remote sensing products can also be collected coincident with any optical measurements. Instruments can be either handheld (e.g. WISP-3 by Water Insight) or fixed to the boat (e.g. TriOS Ramses radiometers) in order to collect measurements by the most suitable means and with varying levels of control over observation conditions. Another advantage with boat surveys is that power sources and laptops can remain on the boat, therefore the instruments and components do not require the same rigour in weather- and waterproofing as for permanently exposed installations. However, radiometric measurements from (small) boats are subject to errors from tilt and roll (greater than 5°); therefore, affected data must be screened during data processing (Tilstone et al., 2012). Additionally, underway measurements are subject to possible contamination of lenses from spray, therefore careful attention to the cleaning of sensors prior to measurement is required.

Standard field protocols have been developed for ocean (NASA 2002, 2003 and 2004) and coastal waters (Tilstone et al., 2012) for the measurement of inherent and apparent optical properties by boat, and these have also been widely implemented for inland waters (e.g. UK GloboLakes and FP7-INFORM projects). For example, the measurement of absorption properties can be undertaken using a winched package (which is especially useful for depth profiles), a slow decent rate optical package (slowDROP), an in-line flow through system or as a towed system on the front of the vessel

(IOCCG, 2017). Above-water reflectance (ρ_w or R_{rs}) can also be collected by boat using spectroradiometers installed on a panel or steel frame to collect downwelling irradiance, water surface radiance and sky radiance (Tilstone et al., 2012).

An important and frequently overlooked aspect of sampling from boats using multiple isolated instrument packages, is to synchronize the time-stamps of each package with a common source, either a GPS record that is updated every second (and preferably shared between data loggers), or an internet time service used to synchronize the data logger at regular intervals.

Existing (accessible) data obtained from small boats (for lakes) and research vessels (coastal areas) relevant to MONOCLE are normally found in the databases described in the next chapter.

Figure 1. Example of a boat survey for measurement of optical properties on Loch Lomond (from globolakes.ac.uk)

Ferrybox

Ferryboxes gained popularity in the 1990s as a first approach to collect seasonally varying observations in whole coastal marine systems. Ferryboxes today typically have a dedicated water inlet 3-5 m below the water line, are driven by a pump to force water through the system, and measured variables include hull temperature, line temperature, salinity, turbidity and fluorescence of one or more substances but usually including chlorophyll-a. Validation relies on regular calibration of sensor data against water samples (which can be taken with autonomous refrigerated samplers) which is taken care of by research institutes hosting managing a particular route, and occasional inter-calibration of sensors between routes.

Advantages of using merchant vessels and particularly ferries are access to electricity, data transfer, technical resources onboard, support from ship crews, and set routes which help in regular maintenance.

Fluorometer readings for Chl-a, phycocyanin, and to an extent coloured dissolved organic matter are of limited value in satellite product validation, due to unknown diurnal variation in the fluorescence signal, particularly for biological components. However, these readings can be used for qualitative or semi-quantitative product validation, e.g. following detrending analysis, calibration against pigment extracted from water samples along a measured transect, or signal coherence analysis (Groetsch et al. 2014).

In 2013, a shipborne spectroradiometer set to measure remote-sensing reflectance with continuous azimuth control was installed on a Baltic Sea ferry by the Finnish Environment Institute, using a set of TriOS RAMSES spectroradiometers (Figure 2, left) supported by free (but not fully open) software to control both viewing azimuth angle and radiometers. A commercial solution is now also available for the HyperSas sensors from Sea-bird Scientific (Figure 2, right). A versatile, low-cost open source solution (supporting more than one type of sensor) for azimuth control is still lacking and under development at PML for MONOCLE. Data collected during 2013-2015 may be obtained through the BONUS FerryScope website (http://ferryscope.org/). It is not clear whether data continue to be collected.

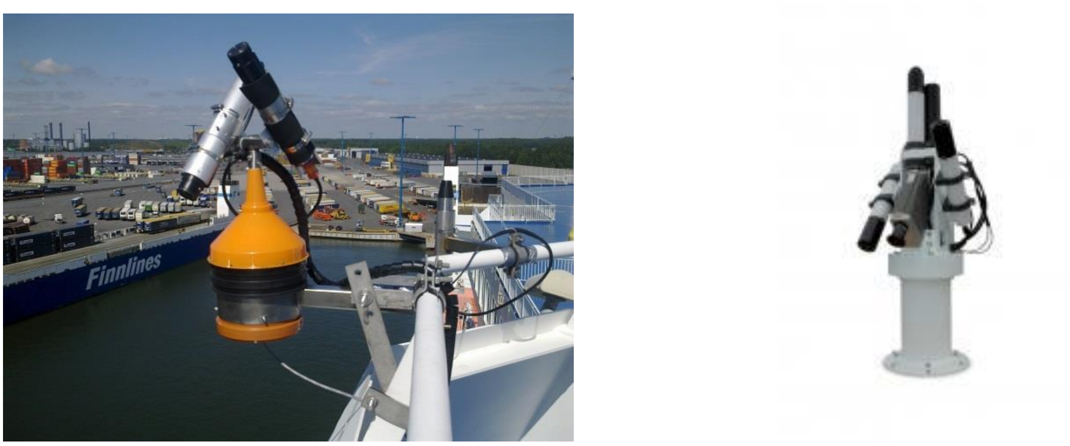


Figure 2 (left) Spectroradiometers on a low-cost azimuth controller on a Baltic Sea ferry by the Finnish Environment Institute. (right) The Sea-bird scientific SAS solar tracker.

Drones

Drone technology is fast evolving and spans a wide range of remotely controlled systems, including underwater drones, Autonomous and Remotely Operated Surface Vessels (ASV/ROSV) and Remotely Piloted Aircraft Systems (RPAS). Each of these systems has a lot of potential in collecting data in shallow or otherwise hard-to-reach waters. While ASV/ROSVs are a special case of a boat platform, on which standard sensors may be mounted, RPAS can carry optical sensors but these require dedicated engineering effort. Despite the potential of RPAS systems for water quality monitoring, well-defined protocols, needed to obtain quality-assured data, are still lacking. This is one of the focus points in the MONOCLE project. There are no current data sources identified for Drone imagery of aquatic environments.

6. In situ open observation databases

After the collection of in situ data, the main aim is to make the data accessible to others so that the data can be used where it is intended, and without undue delays. Dedicated processing chains are often needed, so it is no surprise that each of the existing sensor solutions have their own databases and websites (e.g. https://aeronet.gsfc.nasa.gov/ for AERONET data), which may or may not be accessible through automated methods (APIs and geospatial web services). Data collected ad-hoc generally lack such an infrastructure. This section describes three well-known databases: LIMNADES, MERMAID and the Coast Colour Round Robin dataset.

LIMNADES

LIMNADES is a database which comprises global lake and inland water bio-optical measurements and matchup data for remote sensing (www.limnades.org). It was developed by and is held in trust and maintained by the UK GloboLakes project (www.globolakes.ac.uk). This database presently comprises *in situ* data from over 250 lakes, including coincident Inherent Optical Properties (IOPs), radiometric quantities and biogeochemical data. The database only contains archived (no real-time) observations.

Account registration for the LIMNADES database can be done on the website (https://www.limnades.org/register.psp). Once registered, data must be first submitted to the repository in order to view and download any publicly available data. A request must be made in order to access restricted datasets.

MERMAID

The MERis Matchup In-situ Database (MERMAID) is a European Space Agency (ESA) initiative initially developed for validation of MERIS satellite products (http://mermaid.acri.fr). This database incorporates *in situ* data from over 30 global sites, including:

- Optical data – radiometry (water and sky), IOPs and Apparent Optical Properties (AOPs)
- Biogeochemical data – chlorophyll-*a*, sediment, coloured dissolved organic matter (CDOM), phytoplankton accessory pigments
- Atmospheric data

Data is provided as text files of *in situ* data, matched with concurrent and comparable MERIS L2 products (including flags, auxiliary information and intermediate processing outputs). Users can select matchups according to site, parameters of interest, flags and statistical screening. The extraction interface additionally produces validation statistics and plots. Optical measurement protocols for all datasets providing *in situ* data are also available.

MERMAID was originally developed for support of the MERIS Maintenance and Evolution project. Access for external investigators can be requested by sending a short description of the project and intended use of MERMAID data to mermaid@acri.fr, with access granted through a service level agreement (SLA).

Coast colour round robin dataset

The coast colour dataset is a comprehensive database compiling a total of 1,293,396 measurement records. This database provides information about metadata, radiometry, inherent optical properties (IOPs) of the water body and biochemical optical data (Table 2).

A short summary of the most relevant parameters and their availability per test site is given in the table below. Besides acquisition time and geo-location, which are preliminary information for the usability of the data, the most frequently measured parameters are water depth, temperature, salinity, chlorophyll a (Chl_a), the dissolved material absorption coefficient at different wavelengths ag(λ), total suspended matter TSM, spectral downwelling attenuation coefficient Kd(λ), water leaving radiance Lw(λ) and remote sensing reflectance Rrs(λ) at different wavelengths λ.

Table 2 Availability of selected in situ measured parameters per test site in the Coast Colour data set.

		North Sea	Baltic Sea	Mediterranean & Black Sea	Morocco	Acadia	Chesapeake Bay	Oregon and Washington	Plumes and Blooms	Puerto Rico	Benguela	China, Korea, Japan	Great Barrier Reef	Red Sea	Indonesian Waters	Cape Verde	Central California	Antares-Ubatuba	Tasmania	Gulf of Mexico
METADATA																				
start Date	Date of begining of data aquisition [yyyy-mm-dd]	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
start Time	Time of begining of data aquisition [hh:mm:ss]	x	x	x	x	x	x	x	x	x	x	x	x		x	x	x	x	x	x
LAT	Minimum Latitude [degree]	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
LONG	Maximum Longitude [degree]	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Wind_Speed	Wind_Speed [m/s]	x	x												x					
Cloud_Cover	Cloud Cover [%]	x	x												x					
Secchi_Depth	Secchi_Depth [m]	x	x									x		x	x					
Water_Depth	Water_Depth [m]	x	x	x	x	x	x	x	x	x	x	x			x	x	x	x		x
Conductivity	Conductivity [mS/cm]		x	x	x															
Temperature	Temperature [°C]	x	x	x	x	x	x	x	x	x		x					x			
EUz_m	Depth of euphotic layer [m]	x																		
BIOGEOCHEMICAL_OPTICAL_DATA																				
Salinity	Salinity [psu]	x		x	x		x	x		x		x	x				x		x	
Density	Density [sigma]	x										x								
turbFNU	Turbidity: Formazine Nephelometric Units	x																		
chl_a	Chlorofill a by Fluorometric Method [mgm-3]	x	x	x	x	x	x	x	x	x	x	x		x	x	x	x	x		x
agXXX	Dissolved Material Absorption Coefficient at wavelength XXX [1/m]	x	x	x	x	x	x		x			x	x		x	x		x	x	
SPM/ ss/ STM	Suspended Particulate Matter / Suspended solids / Total Susp. Matter [mg/l]	x	x	x	x							x	x	x	x				x	x
kpar	Diffuse downwelling coefficient for PAR (Photosynthetically Available Rad.)	x			x	x	x		x	x					x	x				
z_YY	Depth of YY% light level of PAR [m]	x			x	x	x		x	x					x	x				
RADIOMETRIC_DATA																				
kdXXX	Spectral downwelling irradiance Attenuation Coeff. at wavelength XXX [1/m]	x			x	x	x		x	x	x	x			x	x				
LwXXX	Water Leaving Radiance at wavelength XXX [uW/cm²/nm/sr]	x			x	x	x	x	x	x	x				x	x	x	x		x
RrsXXX	Remote Sensing Reflectance at wavelength XXX [1/sr]	x						x			x	x	x		x					x
IOPs																				
aXXX	Absorption Coefficient at wavelength XXX a(z,l)			x					x											x
bXXX	Scattering Coefficient at wavelength XXX [1/m]																			x
bbXXX	Backscattering Coefficient at wavelength XXX (fit) [1/m]	x			x	x			x							x				

The number of available data per test site varies strongly. For the North Sea thousands of Ferry-box records are included in the database, whereas for other sides only very few measurement records have been taken for a specific parameter.

7. Integration techniques

7.1 In situ data for algorithm validation and calibration

Algorithms transform observed variable A into a variable B based on physical principles, empirical relationships or a combination of both (Brockman, 2015). These algorithms are subject to assumptions, limitations and scope (e.g. min and max of the concentrations), and should be validated to test if the expected behaviour is confirmed. The test dataset for validation can comprise simulated data, based on radiative transfer simulations, and/or in situ measurements. To set-up, define and validate algorithms, it is important to have a large dataset of measurements with a wide range in concentrations and Inherent Optical Properties (IOPs), avoiding spatio-temporal bias. There is no need of a satellite overpass in this stage of the process.

An illustrative example for in situ data integration is how the *Calimnos* processor for MERIS, OLCI and MSI sensors groups in situ data by their apparent Optical Water Types (OWT). Data from multiple sites that belong to a given OWT are grouped and jointly used to validate satellite algorithms intended for that OWT. Data points observed at a given station can belong to different OWTs at different times.

In order to maintain standards and ensure consistency and reliability of the collected in situ data, the data should be delivered with a quality flag or uncertainty estimate. This gives an indication of the quality of the values in a file or a given station. The quality control flags CMEMS use are presented in Table 3. The data should also be subjected to a series of quality control checks to exclude inconsistencies. A list of CMEMS quality action control points is given in Table 4. These action points were mainly defined for water temperature, salinity, chlorophyll-a and oxygen, but can also be addressed for other parameters. Detailed protocols can be found in:

- http://archimer.ifremer.fr/doc/00251/36230/34790.pdf
- http://eurogoos.eu/download/RTQC_BGC_recommendations_v2.5.pdf

Table 3 CMEMS quality control flags assigned to measurements after quality control.

Code	Meaning	Comment
0	No QC was performed	-
1	Good data	All real-time QC tests passed.
2	Probably good data	-
3	Bad data that are potentially correctable	These data are not to be used without scientific correction.
4	Bad data	Data have failed one or more of the tests.
5	Value changed	Data may be recovered after transmission error.
6	Not used	-
7	Nominal value	Data were not observed but reported. Example: an instrument target depth.
8	Interpolated value	Missing data may be interpolated from neighbouring data in space or time.
9	Missing value	An observation was performed, but it is not available

Table 4 CMEMS quality control action code (Jaccard et al. 2015)

Code	Meaning
Spike	Flag changes because of a spike detection
GlobalRange	Flag changes after applying a gross filter on the observed values.
RegionalRange	Flag changes if measured values fall outside the expected range for the relevant region. This range can be time-dependent (e.g. expected increase of Chl concentrations during due to algae blooms).
BadDate	Flag changed for bad date (e.g. number of month > 12-.
BadLocation	Flag changed for a bad location such as position on land or impossible speed
Analysis	Flag changed after a statistical analysis
VisualInspection	Flag changed after visual inspection
Climatology	Flag changed after a climatology test
Model	Flag changed after a model diagnostic
StuckValue	
IncreasingPressure	

Other	

7.2 In situ data for EO product validation

Satellites capture radiance emitted from the water surface over a certain spatial extent at a certain time. This signal is converted into physically meaningful values (e.g. water-leaving reflectance, chlorophyll-a concentrations) using defined algorithms, and ultimately we need to know the accuracy of the whole EO processing chain. This can be done by conducting a match-up analysis for water-leaving reflectance and water quality parameters. Typical considerations for collecting data and performing match-up analysis are listed in the ESA GlobCOLOUR Validation protocol (Durand et al., 2006):

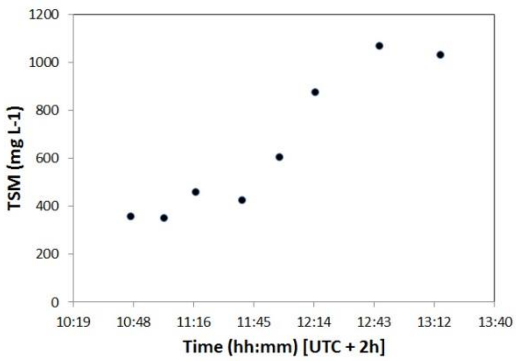
- http://www.globcolour.info/validation/report/GlobCOLOUR_VP_v2.1.pdf

The considerations refer to

- Satellite-specific issues: spatial and temporal resolution, 'box' used for pixel averaging (spatial representativeness) and comparability between the quantity measured by in situ instrument and that derived by the satellite sensor (e.g. difference in band settings and spectral resolution between in situ radiometer and satellite sensor, or fluorescence versus absorption proxies).
- Measurement-specific issues: although in situ measurements are referred to as ground truth, they are rarely the absolute truth. The errors associated with an in situ measurement should be adequately characterized and considered when evaluating validation results. Secondly, the applicability of an in situ measurement towards validation of a satellite product strongly depends on the time the measurement was collected relative to the time the satellite imaged the in situ location. The acceptable time difference is dependent on the stability of the geophysical parameter being compared and can range from minutes to days.
  Environment-specific issues: while the radiometric signal received by the satellite sensor is determined by an order of millions cubic metres of water, in situ and in vitro measurements are typically made on sample volumes ranging from several cubic metres down to millilitres. Also be aware of 'out of bounds' conditions and understand the limitations of the algorithms which are tested/validated, so that validation results can be interpreted correctly.

Two case studies illustrate the challenges related to match-up data collection in dynamic environments: (i) an APEX airborne data acquisition of the Blaye pontoon in the Gironde (FR) for sediment concentration mapping (Knaeps et al., 2015) and (ii) CASI-2 acquisition at Bartom broad for chlorophyll-a and phycocyanin (C-PC) (Hunter et al., 2008).

- At the Blaye pontoon, water samples were collected at the most southern end of the pontoon, Figure 3 (r). The graph shows how the measured in situ TSM values tripled in less than 3 hours, while the map shows the spatial sensitivity of sampling: in the southernmost part of the pontoon, where in situ samples were taken, TSM values dropped from 600 mg/L in the centre of a sediment plume to 200 mg/L a few hundred meters further.



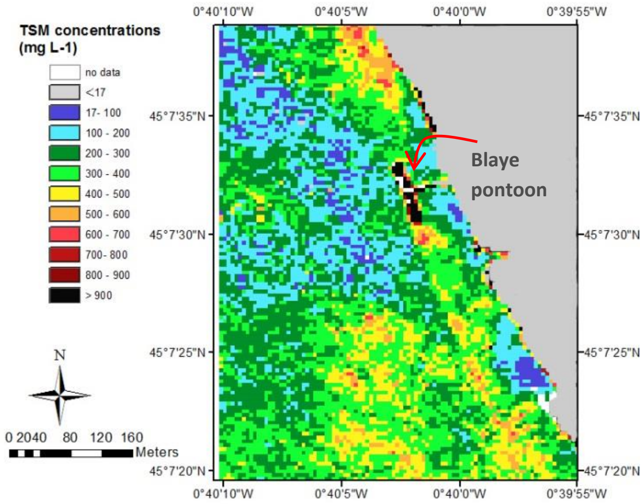


Figure 3. Sediment concentrations at Blaye pontoon: (l) sediment dynamics over time using in situ measurements, (r) sediment dynamics in space using APEX airborne data (Knaeps et al., 2015).

- In Barton Broad (Norfolk, UK), marked changes were observed in the surface spatial distribution of chlorophyll-a (Chl-a) and C-PC over as little as 2.5 hours, linked to vertical migrations of a *Microcystis aeruginosa* bloom (Hunter et al., 2008).

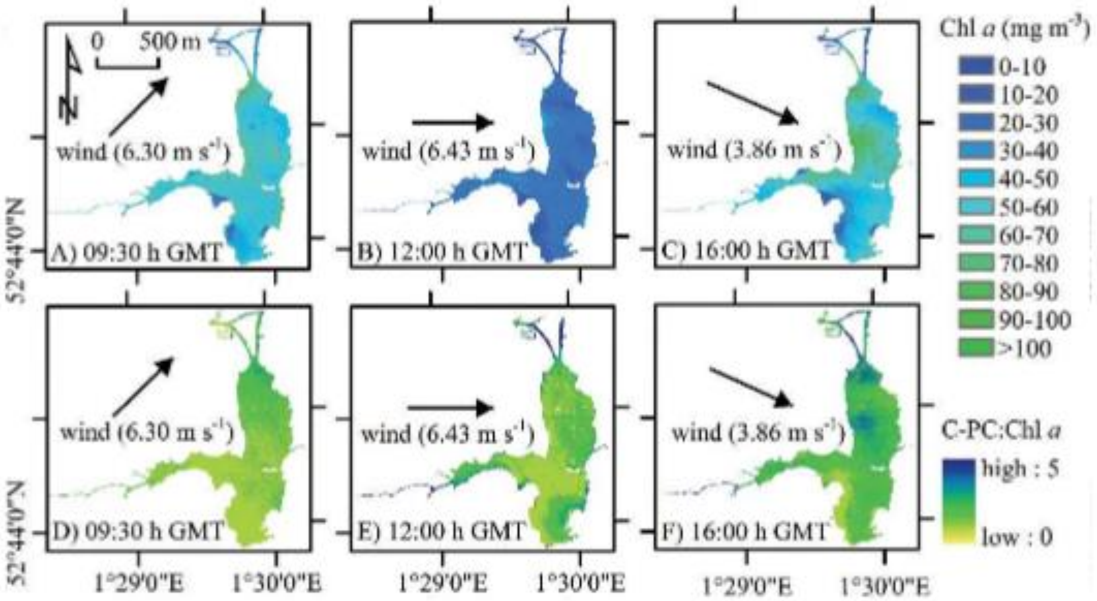


Figure 4. Time series of CASI-2 images showing changes in the distribution of (A-C) Chl-*a* and (D-F) C-PC:Chl-*a* in Barton Broad on 29 Aug 2005 (from Hunter et al., 2008)

Matchup data for MONOCLE will be collected over a wide spatial range to sufficiently capture the bio-geo-optical heterogeneity of each lake and *in situ* data will be ideally collected within a short temporal window (hours) of the satellite acquisition. In MONOCLE, we will also look at how to increase the number of match-up data by sending triggers to the in situ sensors when a satellite is passing by, giving a sign to start measuring.

7.3 Protocols for in situ data integration in EO services

In situ data can only be included in EO services when they are easily accessible, in a generic data format, accompanied by metadata and quality controlled.

The Product User Manual for In situ Products (INS TAC partners, 2017) describes the protocols CMEMS uses for their in situ data and can be used as a starting point for new sensors and data sources.

The **distribution** of in situ data is based on regional portals that allow access to the Real-Time and historical data collected and validated for a specific region. These portals provide access to the best available version of a data at that time. The different portals follow the same structure:

- A single format: the OceanSITES NetCDF format has been chosen because it is CF compliant, it relies on SeaDataNet vocabularies, it is able to handle profiles and time series data coming from floats, drifters, moorings, gliders and vessels.
- Data organization on FTP resources: these have three main directories:
    - Latest: providing access to a sliding window on the latest 30 days of observations for real-time applications
    - Monthly: accumulating the best copy of a dataset, organized by platform and by month
    - History: providing access to the best quality copy of an observation organised by platform
    - Index files to enable automatic download
- Standardized web services (WMS, CSW, WFS) to allow viewing and downloading applications.
- Extendable to other parameters to fulfil the EUROGOOS region needs.

On a regular basis (i.e.: every hour), each regional Distribution Unit (DU) distributes all its new data on its regional portal. The regional portal is a FTP site where data files are regularly distributed. On a regular basis (i.e.: every hour), the global Distribution Unit (DU) collects all new files from all Regional centres and integrate them on the Global centre server. The Global DU does not apply any further quality control (quality control is under Regional data centres responsibility).

The products are stored using the NetCDF **format** following the CMEMS and EuroGOOS recommendations and handles 3 types of data: vertical profiles, time-series and trajectory data. The NetCDF format description is available at:

- http://www.coriolis.eu.org/Data-Products/Data-Delivery/Copernicus-In-Situ-TAC.
- http://dx.doi.org/10.13155/40846.

In-situ products require **metadata** (Jaccard et al. 2015) to guide those involved in the collection, processing, QC and exchange of data. The metadata file requires at least following information:

1. Position of the measurement (latitude, longitude, depth/height, coordinate system)
2. Date of the measurement (date and time in UTC or clearly specified local time zone)

3. Method of the measurement (instrument type)
4. Specification of the measurement (platform code, in addition to e.g. station numbers, cast numbers, name of the data distribution centre).
5. PI of the measurement (name and institution of the data originator for traceability reasons).
6. Processing of the measurement (date of last sensor calibration, details of processing and calibration already applied, algorithms used to compute derived parameters).
7. Calibration method used
8. Comments on measurements (e.g. problems encountered, comments on data quality, references to applied protocols).

7.4 Existing gaps and MONOCLE solutions

Lakes, estuaries and rivers are optically complex waters with variable composition of water constituents, possibly subjected to adjacency effects from neighbouring land and proximity of terrestrial sources of atmospheric pollution. This section highlights the current gaps left to further improve EO products and EO services.

Despite all the efforts in in situ data collection, most inland waters remain unsampled, due to their large number and the costs of site visits, equipment and maintenance. In addition, field sampling has in recent decades focussed on water bodies that are affected by eutrophication, which has left large gaps in the data availability from other sites. Consequently, satellite retrieval algorithms cannot be fine-tuned or adapted to certain optical water types. To overcome this gap, MONOCLE will look at improving automation, adding low-cost alternatives and involvement of citizen science.

A major source of uncertainty in EO products is the atmospheric correction. AERONET-OC stations are a good initiative to collect atmospheric parameters and water-leaving reflectance in situ, using standard protocols at fixed locations. Currently 26 AERONET-OC stations are available, most of them located in coastal areas where adjacency effects are less crucial. The radiance data do not cover the spectral range of the broader bands in satellite missions. These instruments do not allow a large flexibility and don't always represent the optical complexity of inland waters. Hand-held spectroradiometers, used in sampling campaigns from vessels or boats, can collect information in these environments while automation is relatively unexplored. MONOCLE is looking at optical cameras mounted under drones to measure water-leaving reflectance and the HSP-1 Hyperspectral Pyranometer to measure direct and diffuse irradiance for water and atmosphere, to complement further automation of hyperspectral radiometers on buoys and ships.

In terms of data exchange, there is much progress to be made in adopting the Open Geospatial Consortium standards, which can publicly advertise the capabilities and content of data hosted on web services, allowing a high degree of automation. Combined with NetCDF standards for data and metadata organisation, these services can offer fast extraction for data requests. This functionality should be implemented at the lowest possible level, which may be the sensor itself, the data logger or measurement platform.

8. Conclusions and recommendations

A wide variety of in-situ data sampling techniques are currently present taken from fixed platforms (e.g. AERONET-OC or CEFAS buoys) or moving platforms (e.g. boats) and collected in databases (e.g. LIMNADES). Despite these efforts there exist large spatiotemporal gaps in in situ data availability.

To integrate new in-situ data into EO services, they have to be accessible in a clearly structured way, preferably developed as data services with quality control.

This report is a starting point and resource for development in WP7 of MONOCLE, to ensure these developments take full opportunity of existing data sets and procedures. In the next step (D7.2) MONOCLE case studies are disseminated where in situ data (streams) are used to (semi) continuously validate EO products.

9. Exploitation and Dissemination

This report will be advertised on the MONOCLE website and disseminated with a communication package towards relevant (H2020) projects, agencies and service operators. The inventory of existing relevant validation data and methods will be communicated internally and externally for benchmarking of newly developed methods.

10. References

Brockman, C. (2015). Validation of ocean Colour Data. 4th ESA Advanced training on Ocean Remote Sensing. Ifremer, Brest, France.

Copernicus Marine In Situ Tac Data Management Team (2017). Copernicus in situ TAC - CMEMS System Requirements Document. CMEMS-INS-SRD. http://doi.org/10.13155/40846

Data-MEQ working group. (2010). Recommendations for in-situ data Near Real Time Quality Control. Available at: http://archimer.ifremer.fr/doc/00251/36230/34790.pdf

Durand, D., Mangin A., Pinnock, S. (2006). GlobCOLOUR: An EO based service supporting global ocean carbon cycle research – Validation Protocol. GC-PL-NIVA-VP-02. Issue 2. Available at: http://www.globcolour.info/validation/report/GlobCOLOUR_VP_v2.1.pdf

Groetsch, P.M.M., Simis, S.G.H., Eleveld, M.A. and Peters, S.W.M. (2014) Cyanobacterial bloom detection based on coherence between ferrybox observations. Journal of Marine Systems 140, Part A(0), 50-58.

Hunter, P.D., Tyler, A.N., Willby, N.J. and Gilvear, D.J. (2008). The spatial dynamics of vertical migration by Microcystis aeruginosa in a eutrophic shallow lake: A case study using high spatial resolution time-series airborne remote sensing. *Limnology and Oceanography*, 53(6): 2391-2406.

Hunter, P.D., Tyler, A.N., Carvalho, L., Codd, G.A. and Maberly, S.C. (2010). Hyperspectral remote sensing of cyanobacterial pigments as indicators for cell populations and toxins in eutrophic lakes. *Remote Sensing of Environment*, 114(11): 2705-2718.

INS TAC partners (2017), Product User Manual For In Situ Products. CMEMS-INS-PUM-013. Issue 1.7; Version 3.0 http://dx.doi.org/10.13155/43494.

IOCCG (2017). Inherent Optical Property Measurements and Protocols: Absorption Coefficient, available at http://ioccg.org/wp-content/uploads/2015/10/absorption_protocol_forcommunitydistribution-sept2017-2.pdf

Jaccard, P., Norli, M., Birgita Ledang, A., Oystein Hjermann, D., Roberto Reggiani, E., Sorensen, K., Wehde, H., Kaitala, S., Folkestad, A., (2015). MyOcean Project: WP15 – Real Time Quality Control of biogeochemical measurements, Version 2.5. Available at http://eurogoos.eu/download/RTQC_BGC_recommendations_v2.5.pdf

Knaeps, E., Ruddick, K.G., Doxaran, D., Dogliotti, a. I., Nechad, B., Raymaekers, D., Sterckx, S., 2015.A SWIR based algorithm to retrieve total suspended matter in extremely turbid waters. Remote Sens. Environ. 168, 66–79.

NASA /TM – 2004-, Ocean Optics Protocols For Satellite Ocean Color Sensor Validation, Revision 5 Volume VI, Part 2: Special Topics in Ocean Optics Protocols.

NASA/TM, 2003. Ocean Optics Protocols For Satellite Ocean Color Sensor Validation, Revision 4, Volume III: Radiometric Measurements and Data Analysis Protocols. Mueller, J.L., Fargion, G.S., and McClain, C.R. (Eds). NASA/TM-2003-21621/Rev-Vol III

NASA/TM, 2002. Ocean Optics Protocols For Satellite Ocean Color Sensor Validation, Revision 4, Volume IV: Inherent Optical Properties: Instruments, Characterizations, Field Measurements and Data Analysis Protocols, NASA/TM-2003-211621/Rev4-Vol.IV

Tilstone, G. and Martinez-Vicente V., 2012. Protocols for the Validation of Ocean Colour Satellite data in Case 2 European Waters. *IESCA satellite validation protocols* 07-027-FR-ISECA. Plymouth Marine Laboratory, UK.