Framework Service Contract EEA/ DIS/R0/20/001 Lot 1 for Services supporting the European Environment Agency's (EEA) cross-cutting coordination of the Copernicus programme's in-situ data activities – Observational data

Lake Water Quality In-Situ Data Requirements and Availability

Date: 21/12/2021

[Document Release]

	Name	Affiliation
Coordinated by:	Laurence Carvalho	UK Centre for Ecology & Hydrology (UKCEH)
Contributions:	Ana Ruescas[1], Kerstin Stelzer[1], Carsten Brockmann[1], Philip Taylor[2], Anne Dobel[2], Gemma Nash[2], Laurence Carvalho[2], Matt Fry[2]	[1]Brockmann Consult [2]UKCEH
Approval:		European Environment Agency
Prepared for:	European Environment Agency (EEA)	
Represented by: (Project Manager)	Henrik Steen Andersen	
Contract No.	EEA/DIS/R0/20/001 Lot 1	

[Change Record]

Version	Date	Changes
0.1	30th April 2021	Initial draft
0.2	1st Oct 2021	Second draft
1.0		Final draft
1.1	21st Dec 2012	Published version incl minor edits

Executive Summary

This report summarises the requirements, availability and limitations of in-situ data for development of satellite-EO products relating to lake water quality across the Copernicus services. It identifies gaps in available data and provides recommendations for coordination activities that may help improve access and usefulness of in-situ data.

The review has been structured in relation to four aspects:

1. Key data centres for in-situ data
2. Inherent Optical Properties required for development of satellite EO water quality products
3. Water quality data required for calibration/validation (cal/val) of satellite EO water quality products
4. Citizen science data that has the potential to support cal/val of satellite EO water quality products

Results

The Copernicus Global Land Service (CGLS) Lake Water Quality product is identified as the key current service requiring high quality in-situ data for calibration and validation of their products. Currently it monitors water quality in over 4000 permanent and seasonal water bodies, natural lakes and artificial reservoirs, with an area larger than 50 ha (0.5 km^2). The selection of lakes is based on the size and shape of the water bodies for being suitable for EO data retrieval. A few smaller waterbodies are included in the service, via demonstration products, with a spatial resolution of 100 m. The water quality products include:

1. Turbidity (water clarity)
2. Trophic state index based on the amount of the phytoplankton pigment chlorophyll-a.
3. Lake surface reflectances measuring the apparent colour of the water body.

The next version of products is proposed to include:

4. Total suspended matter concentrations.
5. Chlorophyll-a concentration as a direct measure of phytoplankton abundance.
6. Harmful algal blooms of cyanobacteria

In addition to these, future water quality requirements could potentially include measures of coloured Dissolved Organic Matter (cDOM) due to its potential relevance to IPCC greenhouse gas emission inventories and carbon cycling in inland waters.

Data needs cover two aspects:

1. Optical data such as lake water reflectances and Inherent Optical Properties (IOPs) needed for algorithm development and algorithm validation.
2. Water quality data for calibrating algorithms and validating the water quality products listed above

There are three main criteria to consider in relation to in-situ water quality data:

- Sampling frequency (temporal match-up)
- Sampling locations (spatial match-up, surface vs depth-integrated)

- Comparability of data (e.g. Phycocyanin reflectance, fluorescence, cyanobacteria cell counts)
- Coverage of a wide range of water types across different levels of water quality, seasons and atmospheric conditions
- Documentation of methods is key, as is QA/QC of data and estimations of uncertainties.

LIMNADES is the only dedicated global data centre collecting relevant optical properties and associated water quality data specifically for the purpose of supporting development of satellite EO-based water quality products for inland waters. In addition to LIMNADES, there are some key global and regional in-situ water quality data centres, such as GEMS/Water and EIONET (Europe) that have a suitable structure, follow FAIR principles and have relatively secure on-going funding to be a repository for in-situ water quality data specifically to support the development of the Copernicus Inland Water Service.

IOP data is relatively scarce and mainly available for wealthier countries or regions, such as the USA, Europe and Australia. This is largely because of the expensive specialist kit required for data collection and the expertise required to use it. Relatively cheaper portable sensors are coming onto the market to improve this situation. The review highlighted regional data gaps and, therefore, the need for regional campaigns and for funding and capacity building to support better coverage in less developed countries.

Relevant in-situ data are not always readily available. For example, global networks, such as GLEON, could be enhanced to allow for easier data discovery. More meta-data are needed to understand sampling locations, methodologies, detection limits, etc. In addition to this, the comparability and uncertainty of in-situ measurements between different sensors, laboratories and citizen science campaigns is largely unknown. How representative in-situ data is in relation to EO measurements needs further evaluation, e.g. shoreline vs open water measurements, surface vs integrated water column data.

There are very few global citizen science campaigns delivering relevant water quality data, with the exception of good schemes for turbidity (Secchi Dip-in, Freshwater Watch). There are also some potentially relevant regional schemes for monitoring harmful algal blooms (Bloomin' Algae CyanoWatch) and water colour (Eye on Water) which could be scaled-up and tailored more to deliver useful data for validation of Copernicus water quality products.

Recommendations

There are a number of areas for future coordination between in-situ and EO data communities on data gaps, comparability, accessibility and licensing and the need to document information using meta data standards. We recommend the two communities work more closely together to provide more consistent and reliable access to a range of in-situ data for the purpose of EO service production and validation. Ideally this would be developed through a global data portal providing links to regional in-situ data centres. COINS should particularly work to support initiatives such as GEO AquaWatch and the European Union Water-ForCE Project to support further enhancement of in-situ data collection and availability for the development of the Copernicus Inland Water Service.

Specific activities that could be undertaken (in part through COINS) include:

- Compiling detailed guidance (user manuals and standard operating procedures) on the design of in-situ sampling programmes and sampling protocols to enhance match-up with satellite overpasses.
- Reviewing and cataloguing (affordable) portable sensors that can deliver high quality hyperspectral IOP data to increase in-situ data availability.
- Reviewing and cataloguing (affordable) portable sensors that can deliver high quality water quality data (especially chlorophyll-a, cyanobacteria and cDOM) for increasing in-situ data availability for these parameters.
- Document case-study examples on the use of sensors on fixed monitoring buoys (e.g. GLEON) or sensors combined with targeted citizen science to demonstrate innovation in delivering matched satellite and in-situ datasets
- Working with communities of practice, such as GEO AquaWatch, to organise capacity building / training workshops or online courses on in-situ data collection, quality checking and archiving, particularly in Africa, Asia and South America.
- Work alongside GLEON network to provide a more transparent catalogue of in-situ data available on key water quality parameters (turbidity, chlorophyll-a, etc.)
- Collaborate with UN GEMS/Water and LIMNADES to further promote and support the data integration and connection to additional in-situ data bases.
- Organise workshop(s) with specific providers of monitoring data in order to align measurements to be better suited for purpose for EO validation (at least in terms of metadata). This includes bringing together EO and citizen science (CS) communities with shared interests in water quality data to identify priorities for cooperative working and shared campaigns.
- Further support or elaborate case-studies where CS data could be used to support calibration or validation of satellite EO-data products. Specifically for turbidity (Secchi dip-in, Freshwater Watch), harmful algal blooms (Bloomin' Algae) and water colour/cDOM (Eye on Water)

Contents

1. Introduction

Background to review

The Copernicus In Situ component has previously identified requirements for hydrological data, as referenced within its 2017 and 2018 State of Play reports [Copernicus In Situ, 2018], and the Copernicus In Situ Information System (CIS2). In September 2018 the Hydrology project of the Copernicus In Situ component was established comprising members with expertise in European and global in-situ hydrological data, and a mandate to develop an improved understanding of these requirements and a series of coordination activities to improve access to in-situ hydrological data across the services.

The scope of the work was initially identified to include river flows, river water quality, lake extents and depths, lake water quality and soil moisture. This was reported in the Copernicus In Situ project report on “Hydrology in-situ data requirements and availability” (Fry et al.,2019).

This work is being continued within the COINS consortium project. A second phase follow-on project has been developed to elaborate aspects of the first review. One of the areas considered important to elaborate in a further review was in-situ data to support development of Copernicus services in lake water quality. In this report we particularly focus on the needs of the Copernicus Global Land Service (CGLS) inland water quality product. Fry et al. (2019) reported that the quality of the underpinning in-situ data, and its representativeness of types of water and conditions, is a fundamental aspect of the quality of resulting EO products. Similarly the comparability between in-situ and EO data was highlighted as well as the need for more detailed metadata on the exact location of in-situ data within the lake, as well as the measurement / analysis techniques used. These issues are all explored in much more detail in this report.

There is also growing interest in satellite EO water quality products for rivers, estuaries and coastal waters too but these are being considered elsewhere within the COINS consortium.

The report aims to summarise the requirements for in-situ lake water quality data, to identify gaps in currently available data, and consider potential coordination activities that may help improve the design, comparability and access of in-situ water quality data for use in the development of Copernicus services.

Approach

Information has been gathered largely through expert review of available data sources online and discussion with those leading Copernicus service product developments.

The requirements for in-situ data are considered here by the type of data, rather than by service or product, in order to identify commonalities in uses of different water quality data types across the services.

The review has been structured in relation to four aspects:

1. Key data centres for in-situ water quality data
2. Inherent Optical Properties (IOPs) required for development of satellite EO water quality products

3. Water quality data required for cal/val of satellite EO water quality products
4. Citizen science data that has the potential to support cal/val of satellite EO water quality products

The information presented for each of them is considered under the following headings:

- Current data availability
- Data accessibility/licencing
- Gaps in data
- Recommendations for increasing data availability/quality/access

2. In-situ water quality data needs for Copernicus products

Requirements for in-situ water quality data

What current Copernicus services have an interest in in-situ water quality data?

Two large public services running under the Copernicus framework provide products concerning water quality. These are the Copernicus Marine Environment Monitoring Service (CMEMS) and the Copernicus Global Land Service (CGLS). For both services we consider only components delivering optical water quality products as relevant for reflecting the requirements for in-situ data. And here, the focus is laid on the Copernicus Global Land Service in relation to water quality products for inland waters.

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 Lake Surface Water Temperature (LSWT), Lake Water Quality (LWQ), Water Bodies (extent) and Water Level. The processing chain supporting LWQ is Calimnos (v1.4) which was developed during the UK GloboLakes project and since then has been adapted for operational processing for the Lake Water products within 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 for calibration/validation is based on a global collection of in-situ data. There is a need for LIMNADES to hold sufficient data for a number of lakes across optical water types, to be considered suitable for global applicability. There is also a need to address data gaps of poorly characterized (or seasonally under-sampled) waterbodies.

The CGLS Lake Water Quality product has been applied to a large number (nominally 4,264) of permanent and seasonal water bodies, natural lakes and artificial reservoirs, with an area larger than 50 ha (0.5 km^2). The selection of lakes was based on the size and shape of the water bodies for being suitable for EO data retrieval. A few smaller waterbodies are included in the service, via demonstration products, with a spatial resolution of 100 m. The products consist of three main parameters:

1. The **turbidity** of a lake describes water clarity. Turbidity often varies seasonally, both through the discharge of rivers, through growth of phytoplankton (algae and cyanobacteria) and through wind resuspension of lake sediments.
2. The **trophic state index** is an indicator of the productivity of a lake in terms of phytoplankton and is a biological response to the nutrient status of a water body. It is (in this case) based on the amount of algal pigment, chlorophyll-a.
3. **Lake surface reflectances** describe the apparent colour of the water body, intended for scientific users interested in further development of algorithms. The reflectance bands are also used to produce true-colour images of the water bodies by combining the visual wavebands.

The next version of products is proposed to include:

- **Total suspended matter concentrations,** providing information about sediment transport mainly as inflow from rivers into lakes or reservoirs or resuspension.
- The direct measure **chlorophyll-a concentration** as a proxy for phytoplankton abundance.

- Harmful algal blooms of **cyanobacteria** due to their relevance to human and animal health and their impact on water supply, fisheries and recreational water use.

Apart from these, future water quality requirements could potentially include measures of coloured Dissolved Organic Matter (cDOM) due to the impacts that this water quality parameter has on water treatment and its potential relevance to IPCC greenhouse gas emission inventories and carbon biogeochemistry of inland waters.

Data needs cover two aspects:

3. Optical data such as lake water reflectances and Inherent Optical Properties (IOPs) needed for calibrating and validating the atmospheric correction algorithm(s) and optical processes in the water (absorption and scattering). This data is needed for both – algorithm development and algorithm validation.
4. Water constituents – data relevant for calibrating the in-water algorithms as well as validating the derived in-water parameters. For this review we will focus on requirements for the following satellite EO products: turbidity, trophic state (chlorophyll-a), suspended sediment concentrations, cyanobacteria and cDOM.

The Copernicus Climate Change Service (C3S) Global Land and Marine Observations Database service provides access to integrated historical surface meteorological holdings in collaboration with the US National Oceanic and Atmospheric Administration's National Centers for Environmental Information. The service is building upon existing data holdings and capabilities for both land and marine domains and could be developed for in-situ WQ data needs.

The EO data used for deriving water quality parameters are mainly the specifically designed ocean (water) colour sensors such as Sentinel-3 OLCI (ESA). Other ocean colour sensors, but less suitable for inland waters due to their spatial and spectral resolution, are MODIS onboard AQUA (NASA) and VIIRS onboard Suomi NPP and NOAA-20 (NOAA). MERIS, the ocean colour sensor onboard ENVISAT (ESA) was delivering data from 2002 - 2012 and was used for archive production building a longer time series. MERIS and OLCI provide data at 300m resolution. The spectral band setting is designed for performing a good atmospheric correction and retrieving in-water constituents. Water colour sensors need to have high requirements for the signal to noise ratios and absolute calibration of the spectral bands. In order to retrieve information at higher resolution, sensors originally designed for land applications are used as well, albeit with knowledge on their limitations concerning the sensor specifications. For example, Sentinel-2 has shown to deliver useful spectral data for monitoring inland water quality, especially for "greener" eutrophic and hypertrophic lakes.

What future services/products may have an interest?

A future coastal service might also cover coastal lagoons. Being predominantly land-locked, this service may, therefore, have needs for in-situ data on inland water quality. The in-situ data needs will be very similar to current needs for inland waters.

Another service with potential interest could include services on air quality and greenhouse gas emissions. New IPCC guidance to include artificial waters (reservoirs) in national GHG inventories (IPCC, 2019), highlights a potential need to measure inland water quality to measure and model emissions from these sources. Water quality data may be important for validation of measurements or atmospheric models.

Current use of in-situ water quality data

In situ data are used for two aspects in the process of water quality estimation using EO data: for algorithm calibration and validation; and for WQ product validation.

There are different categories of in-situ data, which are collected with different purposes. The highest level of quality and in terms of protocols are the so-called fiducial reference measurements (FRMs). They are a suite of independent, fully characterized, and traceable ground measurements that follow the guidelines outlined by the GEO/CEOS Quality Assurance framework for Earth Observation (QA4EO). These FRM provide the required confidence in data products, in the form of independent validation results and satellite measurement uncertainty estimation, over the entire end-to-end duration of a satellite mission (FRM4OC). Secondly, scientific measurement campaigns which follow sampling and analysis protocols are valuable for algorithm calibration and validation. Especially when collected specifically for EO algorithm development. Finally, some in-situ data collected in the framework of monitoring programmes may not have a specific purpose for EO product development. This data source may still be useful for validation work but it is important to consider that sampling strategy, frequency and location might not be ideally suited for EO validation. The documentation of metadata for in-situ data sets is key for searching for the most suitable data sets.

For algorithm calibration and validation: If in-situ data are used for calibration of algorithms, they need to fulfill certain quality criteria. Ideally, they should be very consistent in terms of measurement methods and follow specific sampling and analysis protocols. Documentation is key. The number of suitable match-ups between measurements (satellite and in-situ) are needed and a wide range of water types of interest across different levels of water quality, seasons and atmospheric conditions needs to be covered in order to train or calibrate algorithms. For calibration, simulated data can also be used.

All types of developed algorithms need to be validated. For algorithm validation, in-situ measurements can be used, together with simulations. The wider the range of concentrations and IOPs, the better. A good coverage of seasonal variability is also needed. For validation, it is not necessary to have such a close match-up with EO data and a check of data quality and estimation of uncertainties helps identify which in-situ data are best to use. However, not many in-situ data sets provide this kind of information. An exception is CMEMS, which assigns flags to measurements based on their quality (Table 1).

Table 1 CMEMS quality control flags

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

EO product validation. Finally, the validation of derived parameters, i.e. the water leaving reflectances, optical properties such as absorption and scattering and production-water constituents, require match-ups between in-situ measurements and remote sensing algorithm estimations. There are established protocols to perform the match-up analysis with in-situ data (Durant et al., 2006). The protocols depend on the spatial and temporal resolution of the sensor to select a proper macro-pixel size (multiple pixels, n x n) from the EO data to be comparable with in-situ measurements. Macro-pixels are preferred compared to single pixels in order to detect and filter for outliers (within the macro-pixel) and to provide a larger spatial representativeness. The size of macro-pixels depend on the spatial resolution of the input EO data. For very small-scale and patchy waters, the use of macro-pixels might not always be suitable. Simillarly, for particular circumstances, e.g. monitoring cyanobacterial blooms, the targeted water quality variable (e.g. Chl-a) may vary significantly across pixels, and use of macro-pixels may not be appropriate. For the validation of water leaving reflectances, the comparability between measured (in-situ) and derived (satellite sensor) data will differ depending on the differences in the band settings and spectral resolution of the instruments. The units of the variables should be the same or equivalent and should represent equivalent surfaces (e.g. water leaving reflectances or remote sensing reflectances; water leaving reflectances or surface water reflectances). Another issue to be taken into account are the uncertainties of the in-situ measurements, the possible mis-calibration of ground instruments, the time difference between the satellite overpass and related in-situ measurement, and the fact that we are comparing the radiometric signal (and derived parameters) received by a satellite sensor determined by signals from millions of cubic meters of water with in-situ and in vitro measurements typically made on sample volumes typically only of several cubic centimeters.

What parameters are needed?

There are two types of data needed for development of satellite EO services of inland water quality. The first is optical data, such as water leaving reflectances, needed for validating atmospheric correction (the most critical step within EO water quality retrieval). The second type is the in-water parameters that characterize the quality of the water. The EU MONOCLE Project (https://monocle-h2020.eu/Home) reviewed the water quality data needs of a range of practitioners and stakeholders (Heard et al., 2018). The survey respondees were not specific to stakeholders developing EO products, but covered a wide range of experts in water quality monitoring, sensor development, EO research, and related stakeholder roles. The results of the most relevant water quality variables are shown in Figure 1. The results highlight that data on nutrients was selected as the most important WQ parameter needed, followed by other chemical (dissolved oxygen) and biological (chlorophyll-a) water quality parameters. Parameters that can be derived from EO data include chlorophyll-a, turbidity and total suspended solids. Temperature was also selected as an important parameter, but is not strictly a water quality parameter and will not be reviewed in this report as it was considered in a previous report for Copernicus Programmes In Situ Data activities (Fry et al., 2019).

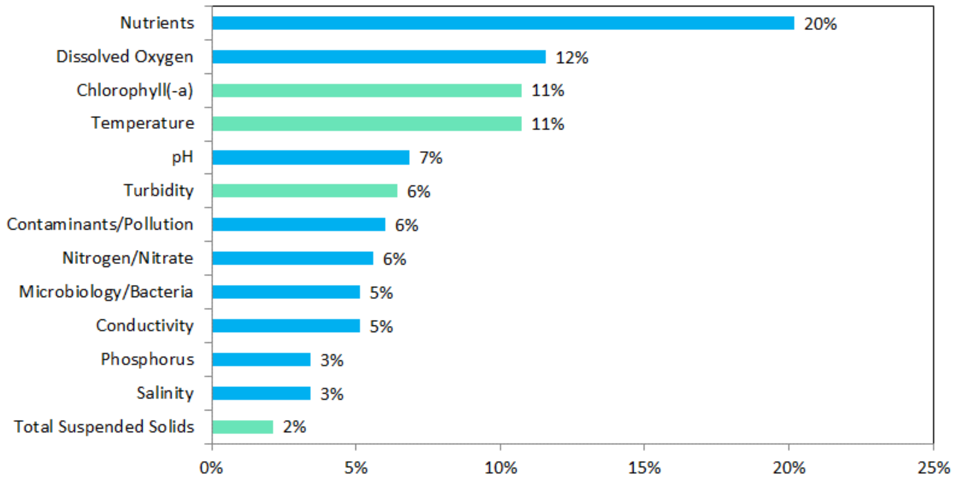


Figure 1. MONOCLE survey results on sampling variables of water quality, in green those that can be derived by remote sensing

What criteria are important for in-situ data (spatial/temporal frequency)?

There are three main criteria to consider in relation to in-situ water quality data needs:

- Sampling frequency (temporal match-up)
- Sampling locations (spatial match-up)
- Comparability of data (e.g. surface vs depth integrated samples, Phycocyanin vs Cyanobacteria)

In addition to this, we would recommend that the EO and in-situ communities also consider complementarity of data products. For example, satellite EO can potentially provide a comprehensive picture of the distribution of harmful algal blooms across large water bodies, but this could be complemented by citizen data along shorelines, where harmful algal blooms can accumulate in high densities but where adjacency effects with land preclude a clear EO signal. Similarly, there may be in-situ data needs on nutrient data, not explicitly for cal/val of EO data products but to provide complementary understanding on the water quality drivers that are causing spatial variability and trends in trophic state observed in EO data.

Sampling frequency

EO data on optical properties of inland waters are typically available every few days (e.g. Sentinel 2 and 3). For algorithm calibration, the temporal match with the satellite overpass should be short, ranging from 1 hour to ~3 days. The ideal time difference depends on the temporal variability of the water quality parameter of interest, so defining a specific frequency is context dependent. A

recommendation is to derive match-ups within 3 hours. Measurements taken during late morning (local time) are closest to the overpasses of optical sensors across the globe.
For providing match-up with EO data, sampling frequency of in-situ data should by high, ideally, every day or every few days. Permanent measurement installations help to provide a large number of match-ups, though not all parameters can be measured with automated systems. Sampling frequencies of monthly or less are likely to have fewer close match-ups with satellite data, unless monthly in-situ sampling programmes are explicitly designed to ensure match-up with satellite overpasses.

Sampling frequency varies a lot across the different variables and depends on the measurement technique. The frequency ranges from every few minutes for parameters that can be measured automatically using sensors (e. g. temperature, dissolved oxygen, chlorophyll-a fluorescence, turbidity) to parameters that need to be processed in a laboratory, which are more typically sampled at weekly to monthly frequency (e.g. microbiology, cyanobacteria, chlorophyll-a concentration, suspended matter concentration). The possibility of making campaigns though the year or years, also determines the sampling frequency, from months to years. Monthly sampling is most common and corresponds to the sampling frequency typically required for monitoring inland water quality by regulations (e.g. the European Water Framework Directive). However, hourly to weekly sampling can be required for some water quality parameters for research or investigative monitoring purposes to adequately capture natural variability (e.g. diurnal changes in dissolved oxygen or turbidity changes associated with storm events). Permanent measurement stations such as AERONET-OC stations and the GLEON network provide very valuable data sets as they enable match-ups with each overflight and provide key data on reflectance and/or water quality in very many different conditions (light, sun/viewing angles, atmospheric conditions).

Sampling location

EO data provide very comprehensive spatial coverage, with relevant multi-spectral optical data available from 10-60 x 10-60 m resolution (e.g. Sentinel-2 MSI) to 300m x 300m resolution (Sentinel-3 OLCI). The shorelines of inland waters are typically discarded in EO water quality data products in order to remove errors associated with mixed pixels of land and water, adjacency effects from land and shallow littoral areas of lakes where sediment colour and submerged vegetation can affect reflectance signal. For example, the Calimnos processing chain, used in CGLS Inland water quality product, removes a 600 m boundary around a lake shoreline to reduce errors arising from these effects. This may limit the minimum size of lake that can be processed in this way, and may still not address some issues of vegetation growth (including benthic vegetation in shallow lakes and floating vegetation). Another methodology is using algorithms for pixel identification and flagging. Since the flagging can fail occasionally, it is recommended to mask a single pixel buffer from shorelines (e.g. CGLOPS2_PUM). This still typically provides thousands to millions of "sampling points" (pixels) per lake. EO data also provide comprehensive spatial and synoptic coverage across a landscape, potentially increasing representativeness of lakes across a region. If the sampling position of In situ measurements is too close to the shore, the location for the extraction of EO data can be moved away from the shoreline to ensure an open water pixel is considered. This assumes that the water type does not change significantly within this distance. For high resolution sensors, it is even mandatory to move the position of the sampling coordinates for some pixels in order to avoid influence by any in-situ sampling infrastructure. This is relevant for larger measurement infrastructures such as AERONET-OC stations.

The comprehensive spatial coverage of EO compares with a much-reduced spatial coverage of in-situ datasets. The number of in-situ sampling locations for most regulatory or research monitoring is one location, either open water around the centre of the lake basin or a shoreline location, often near the lake outflow. Some lakes have more sampling locations, for example the EU WFD requires additional locations if there is significant spatial heterogeneity in water quality between sub-basins of a lake. Shore sampling is more common in regulatory monitoring for the WFD, although CEN standards typically recommend open water sampling as providing a more representative picture of the majority of the lake basin. Bias towards easily accessible areas (near shore) is one of the problems of in-situ data sets that could be addressed through better communication and guidance on in–situ data needs and through wider adoption of ISO and CEN protocols for water quality measurement procedures. The spatial range should capture the heterogeneity of each lake. This analysis could be supported by including optical water type classifications in the design of in-situ sampling campaigns.

Comparability of data

It is important to recognise that water quality measurement parameters are not always the same between EO and in-situ campaigns:

- Turbidity – frequently measured in-situ using a Secchi disc or a turbidity meter (absorbance), compared with reflectance in EO data
- Trophic state (Chlorophyll-a) – measured in relation to reflectance in EO campaigns, but through fluorescence (in sensors) or absorption (in laboratory analyses).
- HABs – typically measured using reflectance associated with phycocyanin pigment in cyanobacteria, but phycocyanin is infrequently measured in-situ. Data for cal/val of EO data products is likely to be more widely available for cyanobacteria cell counts and biovolume. The toxicity of HABs is not visible in optical sensors, although abundance of cyanobacteria, measured by Phycocyanin has been shown to be a good proxy (Hunter et al., 2010).
- cDOM is derived from the absorption characteristics of the blue bands from EO data, or the green to red ratio where the blue band water-leaving signal is minimal. In situ data analyses absorption with spectrophotometers, or directly using a fluorometer sonde, based on a correlation between fluorescence and CDOM absorption at excitation wavelengths, or the relative fluorescence intensity produced by excitation over a spectrum of wavelengths.

There is a major gap in understanding the comparability of these measurements and in particular how this comparability may be affected by seasonality, weather (affecting in-situ data collection and cloud cover of EO data) and global location (affecting reflectance). The need for specific campaigns to evaluate this comparability is clear.

Data Gaps / usage restrictions

General gaps concerning in-situ data can be summarized as follows (source: Heard et al., 2018). More detail will be provided in chapters 4 and 5:

- **Coverage**: there is a lack of in-situ data from countries that have less resources (Kirsche et al., 2020) and/or poor accessibility to water bodies of interest (e.g. lakes in remote or mountainous areas). Even in high income countries with extensive monitoring, the coverage of the area of interest (AOI) can be very irregular in terms of temporal and spatial sampling. The main focus of many in-situ lake sampling campaigns are waters affected by eutrophication (both in coastal and inland waters), and less impacted waters are often sampled less frequently, which leads to an underrepresentation of some water types.
- **Measurement techniques**: the most important issue to consider are the techniques used for in-situ water sampling (manually and with buoys) and how they compare with remote sensing surveys (airborne and space). There is also limited knowledge in the in-situ community on available hand-held measurement solutions (sensors) for spectral measurements. It would be useful for drones to be used more in in-situ sampling campaigns for collecting reflectance measurements, with the possibility of using hyper-spectral sensors (though the current cost of instruments and risk of loss when flying over water are likely to remain limiting factors). AERONET-OC stations have the advantage of providing permanent measurements for measuring key parameters needed, such as surface reflectance, but have the disadvantage of fixed spectral settings. Hyperspectral data is needed for flexibility in band selection and validation of atmospheric correction from different sensors.
- **Funding**: most surface waterbodies remain unsampled or under sampled due to the high costs of field campaigns on the ground. Funding comes mainly from public entities. Involvement of the private sector is low and mainly focuses on specific sectors, such as the water industry. Given the importance of water quality for the public and several industries (e.g. water supply, aquaculture, tourism, agriculture) there is potential for public-private partnerships to increase useful in-situ monitoring to support development of satellite-based water quality moniytoring .
- **Standards & Metadata**: A key issue when working with data from different sources are the metadata. They need to fulfil certain criteria that enables the user of the data a good quality assessment of their suitability and allow filtering for certain criteria (e.g. sampling depth, extraction method, sampling method). More effort must be undertaken to promote open data standards, such as Open Geospatial Consortium standards, for data exchange and harmonisation. Accessibility and readability are key, with clear agreements on licencing needed. Ownership and terms of use should be clearly specified. See more detailed information on this in the following section.
- **Availability**: Data which are collected for scientific purposes are often not available for validation purposes by third parties (due to IPR). They are often provided by many different data originators making data collation time consuming and expensive and a need for rigorous quality control. The existence of many useful datasets may remain unknown to all, apart from the data collectors.

Potential future coordination work

There are a number of areas for future coordination between in-situ and EO data communities on data gaps, comparability, accessibility and licensing and the need to document information using meta data standards. Specifically, we recommend the two communities work more closely together on the following:

- Providing detailed guidance (user manuals) on the design of in-situ sampling programmes and sampling protocols to enhance match-up with satellite overpasses.

- Document case-study examples on the use of sensors on fixed monitoring buoys (e.g. GLEON) or sensors combined with targeted citizen science to demonstrate innovation in delivering matched datasets

- Providing more consistent and reliable access to a range of in-situ data for the purpose of EO service production and validation through development of a global data portal linked to regional data centres

In addition to these, approaches adopted in the marine community include the following points (Heard et al., 2018):

1. Access to Real-Time and historical in-situ data collected and validated for a specific region.
2. Products stored using the NetCDF format
3. Enhanced meta data for in-situ products to guide those involved in the collection, processing, QC and exchange of data. The metadata file requires the following information as a minimum (Jaccard et al. 2015):

    - Position of the measurement (latitude, longitude, depth/height, coordinate system)
    - Date and Time of the measurement (date and time in UTC or clearly specified local time zone)
    - Method of the measurement (instrument type)
    - Specification of the measurement (platform code, in addition to e.g. station numbers, cast numbers, name of the data distribution centre).
    - PI of the measurement (name and institution of the data originator for traceability reasons).
    - Processing of the measurement (date of last sensor calibration, details of processing and calibration already applied, algorithms used to compute derived parameters).
    - Calibration method used
    - Comments on measurements (e.g. problems encountered, comments on data quality, references to applied protocols).

A number of initiatives are being developed to take forward closer cooperation between EO and in-situ data communities, most notably GEO AquaWatch: https://www.geoaquawatch.org/
AquaWatch is an Initiative within the Group on Earth Observations (GEO) that aims to develop and build the global capacity and utility of Earth Observation-derived water quality data, products and information to support water resources management and decision making. It is a global Community of Practice that coordinates activities to achieve this goal. In relation to this, there are specific

shorter-term funded projects to help deliver the goals set out by GEO AquaWatch. This includes the UK-funded GloboLakes Project (http://www.globolakes.ac.uk/) and the EU H2020 funded Water-ForCE project (https://cordis.europa.eu/project/id/101004186).

Water-ForCE is co-creating a Roadmap for the development of the next phase of Copernicus Inland Water Services. A key aim is to align in-situ and remote observation as this is considered essential to further the exploitation of operational earth observation platforms. A strategy to integrate in-situ networks will be defined, integrating approaches to product validation and filling observation gaps. Technical requirements for the future Copernicus sensors will also be specified for optimal inland water monitoring needs and future service development.

3. Data centres

This section summarizes existing national regional or global data bases and data centres for freshwater quality data, both satellite and in-situ. Table 2 provides an overview of data sources; a few key freshwater data centres highlighted in bold black text are reviewed in detail in this chapter.

Table 2 Summary of data acquisition platforms and databases

Instruments	Fixed Platforms	Buoy data	Moving Platforms
	AERONET-OC	CEFAS smart buoys	Boats
		MOBY	Ferrybox
		CoASTS	Drones
		BOUSSOLE	
		GLEON	
Databases	Inland waters	Coastal	
	LIMNADES	MERMAID	
	GLEON	CCRR	
	GEMS/Water		
	EIONET (Europe)		
Services	SeaDataNet	EMODNet	

LIMNADES

What is LIMNADES?

LIMNADES (Lake Bio-optical Measurements and Matchup Data for Remote Sensing), established and maintained by the University of Stirling, provides a database of in-situ bio-optical measurements and satellite match-up data from lakes and coastal waters worldwide. The purpose of LIMNADES was to collect lake water quality data from, and for, scientific purposes.

At time of writing, the LIMNADES database securely holds datasets from over 25 verified research groups from over 200 lakes across the globe, representing arguably the most diverse centralised bio-optical data repository in the world in terms of temporal, spatial and coverage of inland lakes.

LIMNADES was developed in the framework of the GloboLakes project (http://www.globolakes.ac.uk/), and its objective was to serve as a scientific database.

LIMNADES has been further developed in the framework of H2020 projects (MONOCLE, CERTO, Coastops, EOMORES) and a new version will be released in autumn 2021. That version will contain improved data up- and download and the metadata definition has been harmonized. Additional data sets will be available, also via the call by GeoAquaWatch “be a LIMNADES Betatester”. Data within LIMNADES is partly public, partly rejected for a certain embargo period before also going public (e.g. project data).

Overview of current data and accessibility

Currently LIMNADES provides a repository for inherent and apparent optical properties and water constituents. There are around 40,000 data measurements from 3547 stations which extend back nearly 30 years.

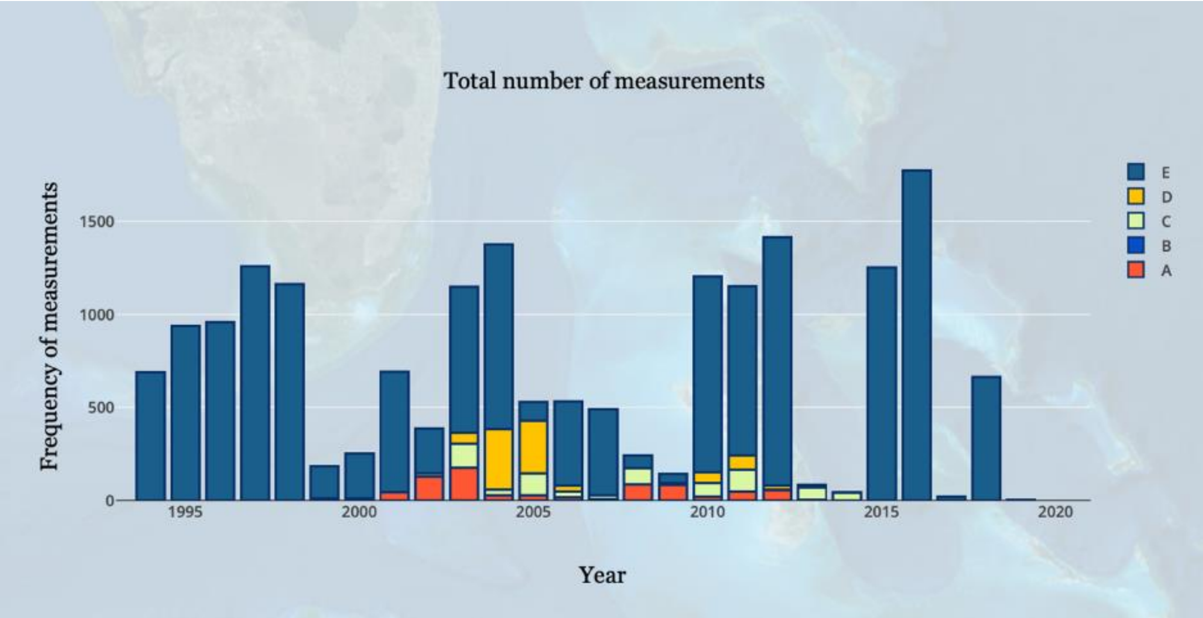


Figure 2. Total number of measurements of the LIMNADES database: (A) Fully open licence, (b) Fully open-non-commercial licence, (c) Open, attribution licence, (D) Open, attribution licence non-commercial, (E) Open, attribution with co-authorship, (F) Temporary storage and use in internal projects

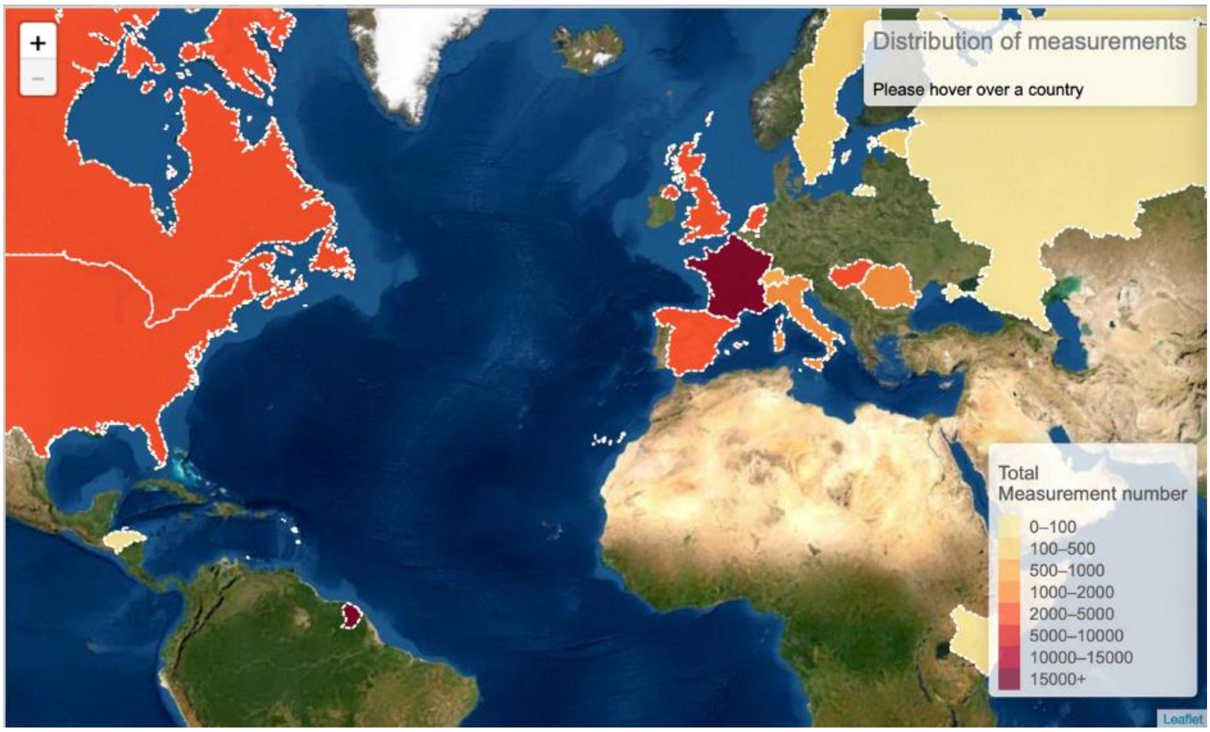


Figure 3. Global distribution of LIMNADES data collection

Gaps in data

The current version of LIMNADES database stopped collating data in 2019, but will contain additional data sets after the release of a new version in autumn 2021.

Types of data use licences are varied, but in many cases, data can only be distributed and used with some limitations (D and E) that required direct permission from data owners, co-authoring if documents or papers are written, and have a strong requirement on non-commercial use. For future datasets, a more open data policy is envisaged.

Chapter 5 provides a more detailed overview of the water quality data in LIMNADES. Only chlorophyll-a and TSM have a reasonably good representation in LIMNADES; there is a lot less data available on absorption of CDOM and phycocyanin. Remote sensing reflectance (Rrs) is only available from 2000 to 2015 and most of it has licence E. The release of the new version would need a new assessment of the available data sets.

Restrictions in data

There are several types of data licences (see Figure 1 and Table 1). Many of the licences held within the LIMNADES database are request only data licences (C-E), meaning that if you are interested in that data you will have to make a request to that particular data owner. A smaller percentage is open access and free to download (A-B).

The data base is currently only accessible to users who also provide data to the database. This currently poses a major limitation on access to the data for researchers that cannot contribute to measurements. This is planned to change in the future, when an embargo time (approximately two years) is put in place for new project data contributed, before these data go public.

Table 3. Types of licences and measurements included in current version 1 of LIMNADES

Type of measurement	Table name	Licence A	Licence B	Licence C	Licence D	Licence E
Apparent Optical Property (AOP)	Remote Sensing Refectance (λ) (sr^{-1})	45		557	377	1157
Apparent Optical Property (AOP)	Secchi disk depth (m)	36		626	63	983
Biogeochemical constituent	Chlorophyll a (mg m^{-3})	785		242	490	10100
Biogeochemical constituent	Inorganic suspended matter (mg l^{-1})	39		508	72	7754
Biogeochemical constituent	Phycocyanin (mg l^{-1})				237	389
Biogeochemical constituent	Total suspended matter (mg l^{-1})	39		508	72	7754
Inherent Optical Property (IOP)	a_{ph} (λ) (m^{-1})				367	718
Inherent Optical Property (IOP)	aCDOM (λ) (m^{-1})	39			367	719
Inherent Optical Property (IOP)	aNAP (λ) (m^{-1})				367	500
Physical parameter	Water depth (m)			380		0
Physical parameter	Water temperature (°c)			352		175
Physical parameter	Wind speed (m s^{-1})				63	0

AERONET OC lake stations

What is AERONET-OC?

The National Aeronautics and Space Administration, NASA, has established a global network of stations to measure atmospheric properties, the so-called Aerosol Robotic Network (AERONET). Since 2006, the Joint Research Centre, JRC, and NASA started establishing the Aerosol Robotic Network-Ocean Color (AERONET-OC), which besides aerosol optical properties also measures water-leaving radiance. The Ocean Color component of the Aerosol Robotic Network (AERONET-OC) was established to support satellite ocean colour validation activities in coastal waters through standardized measurements of atmospheric and marine optical quantities. Specifically, AERONET-OC can provide in-situ values of the normalized water-leaving radiance and aerosol optical thickness through autonomous radiometers operated on fixed platforms in coastal waters.

AERONET_OC stations are equipped with a Sun photometer called the Sea-viewing Wide Field-of-view Sensor (SeaWiFS) Photometer Revision for Incident Surface Measurements (SeaPRISM), which is a modified CE-318 Sun photometer (CIMEL, Paris). The instrument is calibrated yearly. The instrument scheme and data processing are detailed in Zibordi et al. (2009). Briefly, the SeaPRISM performs measurements of direct solar irradiance which is used to derive the aerosol optical thickness at various wavelengths.

There are a few AERONET-OC stations located in lakes, one in Sweden and four in the USA:

- Lake Vänern, Sweden. In 2008, Stockholm University established the AERONET-OC station Pålgrunden in Lake Vänern. The station is situated on the light house 'Pålgrunden' in lake Vänern (58° 45.32′ N, 13° 09.09′ E) and has been operational since April 2008. Pålgrunden was the first AERONET-OC station to be located in a lake, and is one of only a few high-latitude stations.
- Lake Okeechobee, Florida, USA: there are installed two towers, one called Lake Okeechobee located near the middle of the lake The instrument is installed at a height of approximately 20 feet above the water level. And Lake Okeechobee North, located one degree north of the previous tower.
- Lake Erie, USA, has a Fixed Coast Guard structure in western Lake Erie at a water depth of 8.5 m.
- Lake Michigan, USA, has a station in the south of Green Bay, a small arm off the main basin of Lake Michigan.

Overview of current data and accessibility

The measured water-leaving reflectance and aerosol optical properties can be used to validate atmospheric models used in satellite data processing. These models estimate the aerosol properties such as AOT (aerosol optical depth) and the Ångström exponent and convert the top-of-atmosphere radiance measured by the satellite into water-leaving reflectance at sea surface level.

The AERONET-OC data has been quality assured, which implies that the data is checked for cloud contamination, high variance of multiple sea- and sky-radiance measurements utilized for computing Lwn, elevated differences between pre- and post-calibrations of SeaPRISM Sun photometers and spectral inconsistency of Lwn data (D'alimonte and Zibordi 2006; Smirnov et al. 2000).

Gaps in data

Very few lakes have an AERONET-OC station and most were installed recently. Data from high latitudes also only occur from May to October.

- Palgrunden has a long time series of L1 Lwn data, from 5/7/2008 until present (6912). L2 data from 5/7/2008 to 23/9/2020 (1642). Data are only available for summer months.
- Lake Okeechobee (Florida): Lwn from 30/5/2018-25/10/2020 in L1 (3669), and Lwn from 9/8/2918 to 24/10/2020 in L2 (997). The new tower located further north in the lake has L1 data from 19/1/2021 to 27/4/2021 (1060). Data are available for all months,
- Lake Erie: L1 Lwn data from 19/7/2016 to 23/9/2019 (1334), only for summer months (June to September, depending on the year). Same dates for L2 data (255).
- Lake Michigan (South Green Bay): L1 Lwn data from July 4/6/2018 to 26/9/2020 (923); and L2 Lwn data from 14/7/2018 to 16/9/2019 (90) only for summer months.

One complaint from data users is the lack of flexibility of the spectro-radiometers to change their configuration, especially to improve the results of the atmospheric correction of remote sensing data. This affects the quality of reflectance estimates which propagates to the measures of the water quality parameters. A major factor bounding the uncertainties of AERONET-OC data products is the CE-318 measurement technology.

Restrictions in data

No restrictions in data access, but raw data are not easy to handle.

UN GEMS/Water

What is GEMS/Water?

The Global Environment Monitoring System for freshwater (GEMS/Water: https://www.unep.org/explore-topics/water/what-we-do/monitoring-water-quality) provides the world community with data on fresh water quality to support scientific assessments and decision-making. The programme was established in 1978 to collect world-wide water quality data for assessments of status and trends in global inland water quality. The GEMS/Water monitoring network provides surface and ground water quality monitoring data.

Overview of current data and accessibility

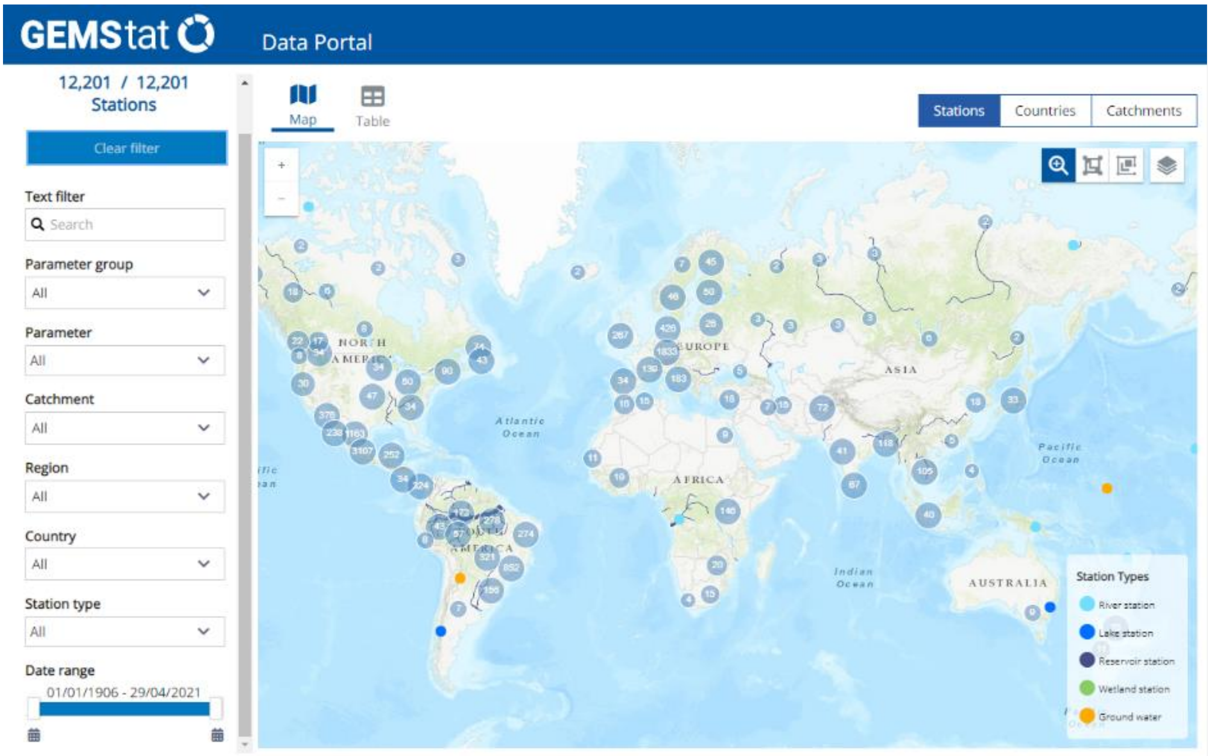
The water quality of water bodies worldwide is assessed by means of water quality indicators for the SDG 6.3.2 core parameters: dissolved oxygen, nitrogen, phosphorus and pH. The indicators have been calculated at station, basin and country level and can be accessed in an interactive dashboard or as separate time-animated maps. Numerous other indicators are measured depending on the site, location and time period. These include data relevant to cal/val of EO water quality data products, such as optical, organic, phytoplankton and pigment data.

The GEMS/Water Data Centre (GWDC: https://gemstat.org/data/data-portal/) provides access to the data and information on the state and trend of global inland water quality. As an operational part of the GEMS/Water Programme of the United Nations Environment Programme (UNEP), GEMStat is hosted by the GEMS/Water Data Centre (GWDC) within the International Centre for Water Resources and Global Change (ICWRGC) in Koblenz, Germany.

The GEMStat metadata catalogue (https://gemstat.org/data/metadata-catalogue/) provides ISO 19115-compliant metadata for all monitoring programmes and monitoring stations of the global water quality database as well as several open web services including an OGC CSW service (http://catalog.gemstat.org/geonetwork/srv/eng/csw?SERVICE=CSW&VERSION=2.0.2&REQUEST=GetCapabilities).

The GEMStat data portal also offers a map viewer (https://gemstat.bafg.de/applications/public.html?publicuser=PublicUser#gemstat/Stations) allowing data to be viewed across space and time. The parameter groups offer a high-level overview of what is available, and include the following parameter groups and record counts:

- Flux (2572 records)
- Indicator Organism (6857)
- Inorganic (10713)
- Nutrient (10808)
- Optical (5413)
- Organic (6437)
- Oxygen Demand (5755)
- Phytoplankton (24)
- Pigment (1631)
- Temperature (9320)
- Water (8232)



GEMStat Data Portal – Map viewer

The data portal, map viewer and metadata catalogue can all be used to explore site specific data, which shows the parameters recorded in more detail. For example, this dataset for Lake Victoria (https://gemstat.bafg.de/geonetwork/srv/eng/catalog.search#/metadata/9ac25ae0-f32d-4c26-a762-a2696d05f48a) has the following parameters recorded:

- Alkalinity
- Ammonia
- Calcium
- Carbon Dioxide
- Chloride
- Chlorophyll A
- Coliforms
- Dissolved Solids
- Electrical Conductance
- Fecal Streptococci
- Fluoride
- Hardness
- Iron
- Kjeldahl Nitrogen
- Magnesium
- Manganese
- Orthophosphate
- Oxidized Nitrogen
- Oxygen
- Oxygen Demand
- Phosphorus
- Potassium
- Silicon dioxide
- Sodium
- Sulfate
- Suspended Solids
- Temperature
- Total Nitrogen
- Turbidity
- pH

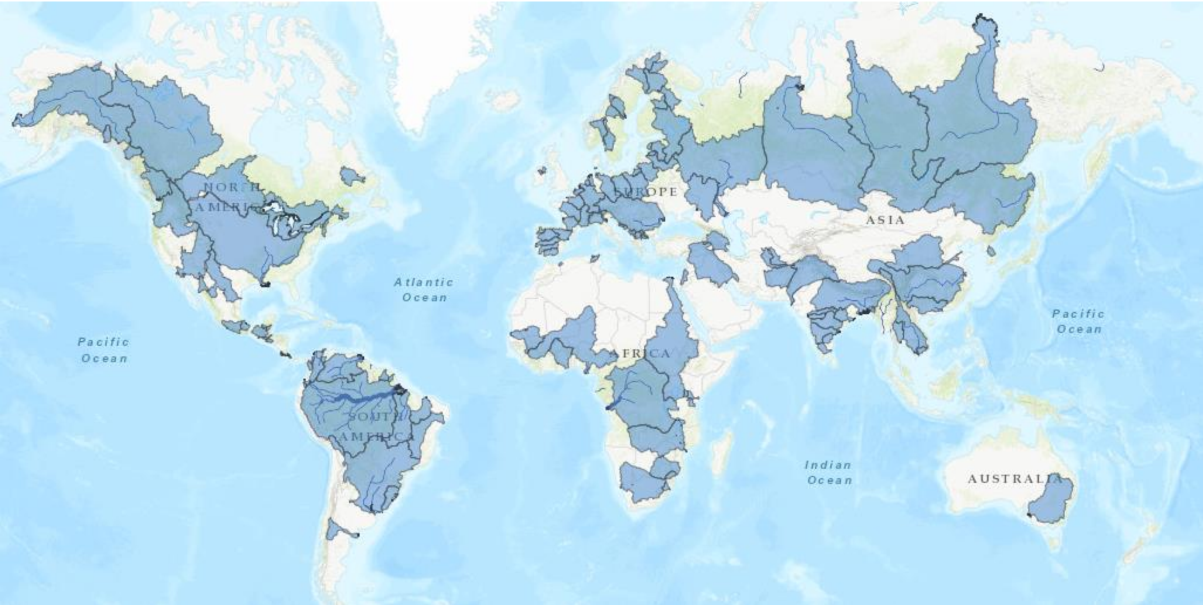
Lake Victoria dataset – parameters available

Specific parameters can therefore be explored for cal/val of EO water quality data products, providing a global overview of data availability.

After data is selected for download, a link is provided to start the download process. To download the data, a contact form needs to be filled in with contact information and details for the requested download. Help can be gained during the download process via the following email address: gwdc@bafg.de

Gaps in data

Parameters available vary across sites and through time, but availability can be explored using the data portal, map viewer and metadata catalogue. The catchments of all available data are shown below, providing global coverage but with noticeable gaps in some regions:



GEMS/Water data – catchment coverage

Restrictions in data

The download of water quality data from the Portal is currently restricted to a maximum of 500 stations. If larger datasets, such as global data, are required a data request form needs to be submitted.

GLEON

What is GLEON?

The Global Lake Ecological Observatory Network (GLEON: https://gleon.org/) conducts innovative science by sharing and interpreting high resolution sensor data on lake monitoring buoys, recording water quality and water temperature typically sub-hourly, to understand, predict and communicate the response of lakes in a changing global environment.

Overview of current data and accessibility

GLEON provides a variety of sensor data for lakes across the globe, with the majority of sites offering very high resolution data (sub-hourly):

The lakes that form GLEON, with yellow dots showing where high resolution data is available

The data is input into one of two archives, which are then made discoverable and re-usable for others. These two services are CUAHSI and EDI:

CUAHSI (http://hiscentral.cuahsi.org/) is the Consortium of Universities for the Advancement of Hydrologic Science, Inc.

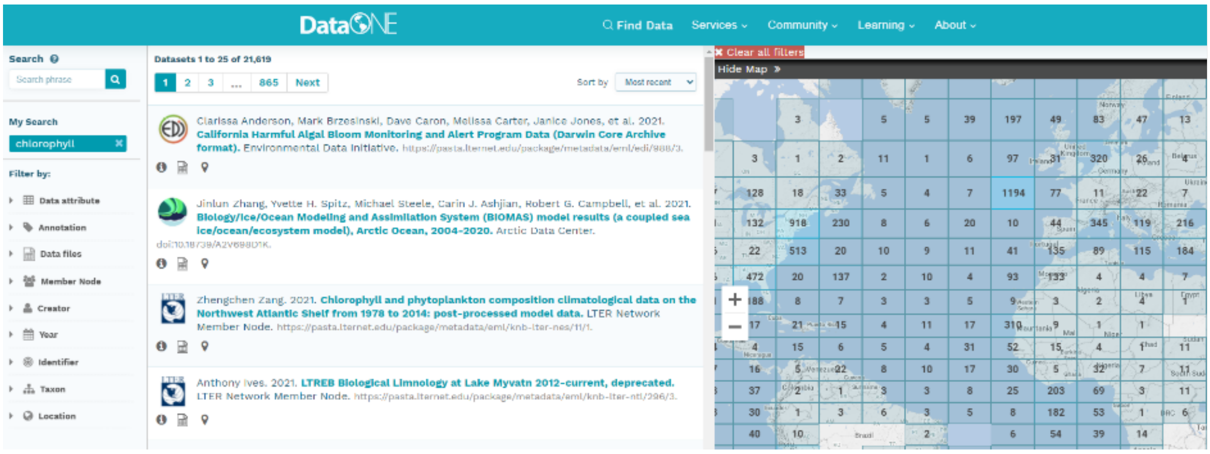
EDI (https://environmentaldatainitiative.org/) is the Environmental Data Initiative. It's a data repository providing open, persistent, robust, and secure access to well-described and easily discovered Earth observational data.

Each service has different availability and restrictions, as shown in the table below:

EDI	CUAHSI
Datasets are immutable; updates will create new versions of the dataset	Datasets are mutable by data provider without notification to data users
Providers can submit records consisting only of metadata with provider contact information so that users can contact data provider directly for access	
Datasets can be set as public, or access can be restricted to selected users	All data are public
Each dataset receives a unique DOI	
Current cost: FREE	Current cost: FREE
Major data wrangling needed to combine data from different users	Little data wrangling needed by the user
Lower effort to submit metadata and data	Significant effort to set up metadata the first time
Dataset-by-dataset QA/QC, visualization, and analysis	Potential for streamlined QA/QC, visualization, and analysis for multiple datasets
Web search and access to datasets or access data directly through scripting languages	Websearch and download
Any data types	Designed for high frequency time series sensor data
Data downloads are tracked	
Metadata structure is defined, but content is free form	Highly defined metadata structure and content
	Defined semantics and controlled vocabulary
Not queryable within datasets	Can search within and across datasets and download the combined results
Data are long-term archived and backed up centrally by EDI	Datasets are stored in a cloud location and backed up centrally by CUAHSI
No requirements for data structure and format (no harmonization required by data provider). Any schema may be used.	Provider required to submit harmonized (consistent) data structure. No harmonization required by data user to combine data sets.
Web services available: REST API	SOAP but not REST services
Uses XML such as EML, FGDC, etc. Can hold any data type	Only WaterML data types

Accessibility information of GLEON data for the two main data repositories

GLEON data may be accessed via the EDI search (https://portal.edirepository.org/nis/home.jsp), the DataONE search (https://search.dataone.org/data) or the Google data set search (https://toolbox.google.com/datasetsearch). Of these, the DataONE search provides the simplest interface, showing presenting search results in a detailed list and dynamic gridded global map (shown below). It’s important to note that although these search facilities link to datasets held in EDI and CUAHSI, they have a wider remit and thus present results from other repositories as well.



DataONE – details from a spatial search for the keyword 'Chlorophyll'

Gaps in data

As shown in the sites map above, there are lake sites across the globe, but no sites monitored in Africa and very limited sites in Asia. As there is no clear overview of all sensor parameters measured across all GLEON lakes, it is hard to assess data gaps at this stage.

Restrictions in data

The use and restrictions of GLEON data is shown in the accessibility table above, where the two main archives have different restrictions. As GLEON data can be downloaded from one of three data portals, different restrictions may apply. GLEON is committed to making data findable, accessible, interoperable and reusable (FAIR), although its accessibility is confused somewhat by the two different archives and the multiple routes to query the portal, none of which present results specific to GLEON, rather providing global searches of all their collated data. These discovery portals are themselves useful though for finding data not represented in the portals and services listed in this chapter.

Other Global Datasets

The Global Open Data Index

Shows water quality data from 2013 - 2016, where available, for each country. The table below shows a variety of statistics per country.

Data, measured on water quality outlines minimum requirements for this category for the following chemicals and these should be updated at least weekly:

- faecal coliform
- arsenic
- fluoride levels
- nitrates
- TDS (Total dissolved solids)

Of these TDS is the only parameter that is relevant for EO-derived water quality measures. Data on nitrates and faecal coliform may be complementary.

KNB Repository

The KNB repository is potentially really useful being based on a big survey that had a lot of publicity at the time. Data is from 1956 to 2020 for various lakes, with lots of other datasets available on the KNB for other WQ determinands. Filazzola et al. (2020) provides a detailed description of the repository.

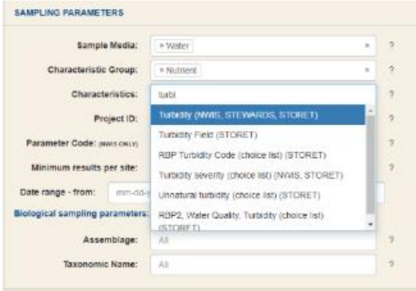
Gaps and restrictions in data

These other global portals vary in their adherence to FAIR principles. By the nature of their inclusion in this report, they are findable, but not all accessible, which in turn makes their interoperability and reusability hard to determine. Some offer detailed information on the data coverage and availability, as well as visualisation tools and data graphic services, but most don't compare in global spatial coverage to GEMS/Water and GLEON. Recent data is also harder to come by, possibly due to update schedules – meaning current and future COPERNICUS products can't easily be compared to in-situ data from the same timeframe. In combination, these data portals could provide useful data but this will bring in potential issues of unit conversion and overcoming differences in sampling methodologies.

Major Regional Datasets

EIONET (Europe)

EIONET data from the European Environment Agency (EEA) is available for 2019-2020. Waterbase is the generic name given to the EEA's databases on the status and quality of Europe's rivers, lakes, groundwater bodies and transitional, coastal and marine waters, on the quantity of Europe's water resources, and on the emissions to surface waters from point and diffuse sources of pollution.



The Waterbase Water Quality ICM (part 1) csv data can be filtered and analysed for any determinand relevant to EO-derived water quality (shown below)

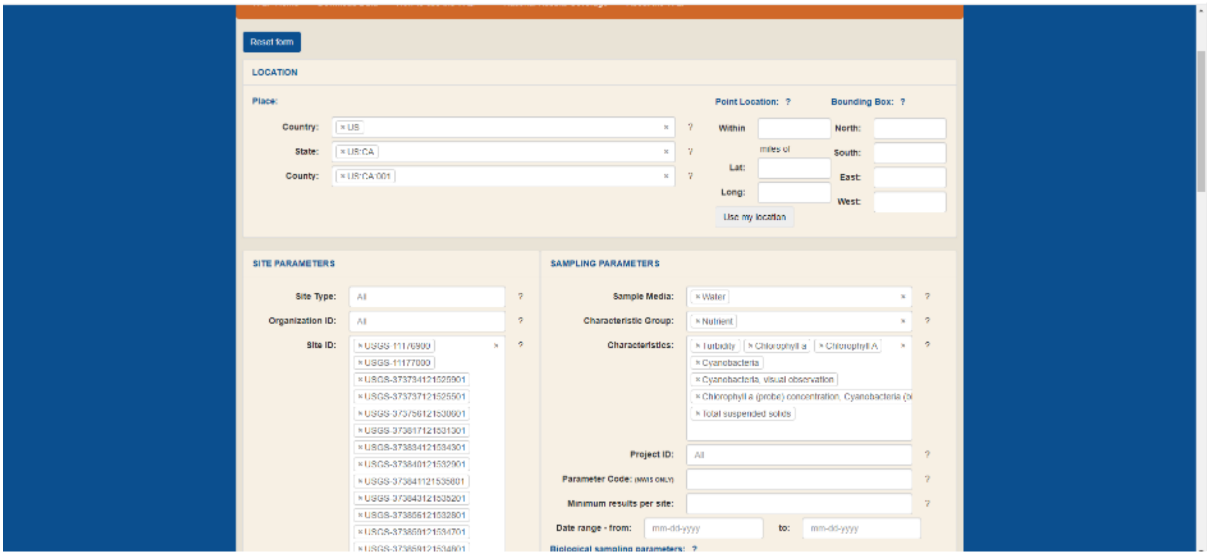
Determinand	ID	Total no. of samples	Mean	Min	Max	Number of sites	Avg. sample count / site
Turbidity	EEA_3112-01-4	1178	6.25	0.2	849	140	17
Total suspended solids	EEA_31-02-7	10009	8.33	0	1000	946	21
Cyanobacteria biomass	EEA_11-06-3	1482	2.65	0	459.00	256	11
Chlorophyll a	EEA_3164-01-0	31827	118.16	0.00083	532000	2440	26

National Water Quality Monitoring Council (USA)

The Water Quality Portal (WQP) is a cooperative service sponsored by the United States Geological Survey (USGS), the Environmental Protection Agency (EPA), and the National Water Quality Monitoring Council (NWQMC). It serves data collected by over 400 state, federal, tribal, and local agencies.

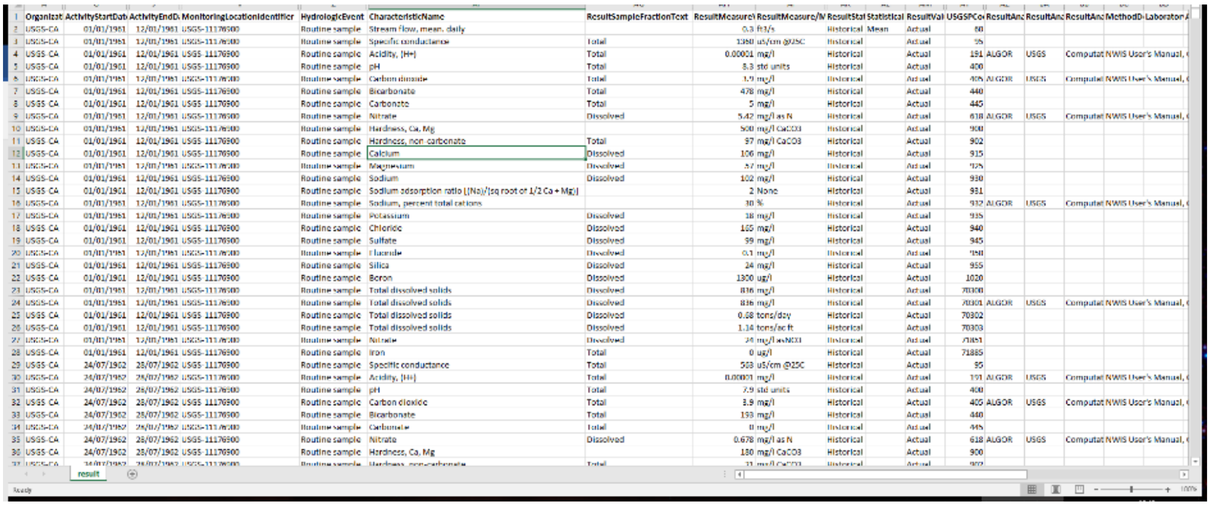
Given that there are thousands of monitoring sites in the USA, filtering has to be applied before viewing data. There are a multitude of options available, and specific 'characteristics' can be chosen from the sampling parameters section. Finding the specific determinand can be difficult as there are several results from a search. The data within the portal is very diverse as it is gained from different monitoring programmes and agencies. Filtering is needed to retrieve comparable results in terms of measurement technique, analysis method, depths, coordinate systems etc. Metadata is not always complete making them less reliable for use in cal/val of satellite water quality data.

WQP – sampling parameters search



WQP – detailed search form

Data are available for download as kml, csv and tsv. No online graphing options seem to be available. Often, due to the scale of data being queried, the website times out before returning data.



Example download of all determinands available for selected sites in California from the WQP.

USGS Lake monitoring and research (USA)

Related science, publications, data, tools and maps are available for selected lakes. USGS water quality data for the nation has 2311 sites of current and historical data plus daily data and statistics (probably the predecessor to the National Water Dashboard).

National Water Dashboard (USA)

Real-time water quality data of 690 lakes for temperature, specific conductance, pH, turbidity, dissolved oxygen.

CDC Healthy Water (USA)

Lots of water quality information including CDC HAB-Associated Illness and links to state WQ datasets. Also Global WASH data and information.

Hypernets / Waterhypernet

The Hypernets and Waterhypernet activities are both aiming to produce hyperspectral data relevant for satellite validation, including inland water sites. Hypernets is developing a new automated hyperspectral radiometer integrated with other equipment across a range of different sites. Waterhypernet aims to develop a federated network of automated hyperspectral radiometers deployed on fixed structures. No data is yet available at the time of writing.

LAGOS (USA)

LAGOS is a multi-scaled database system and a set of tools to study lake water quality at macroscales.

Irish Lakes Waters

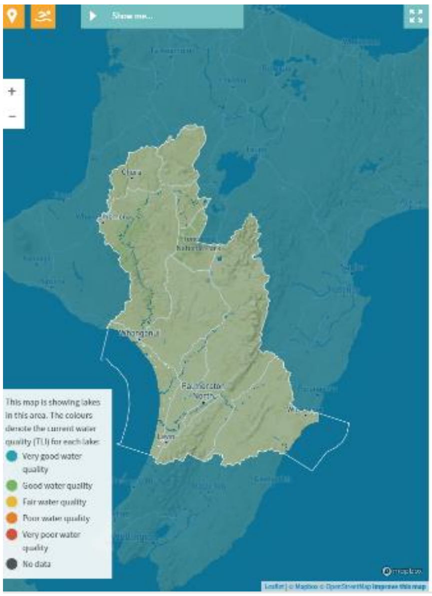
Reporting period 2007-2009.

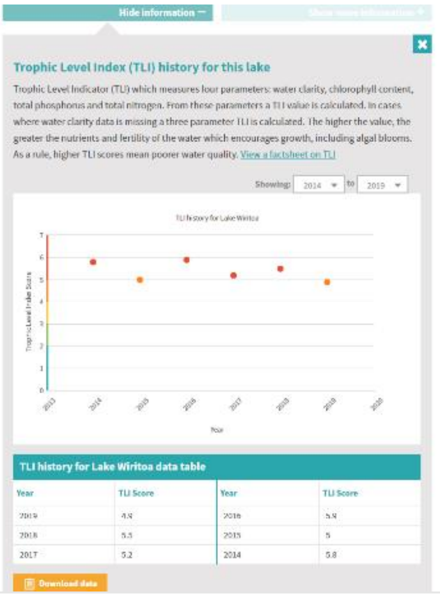
Slovenian water quality data

Comprehensive Slovenian water quality data for 2007 - 2020 available. There is also Surface water quality monitoring data available, which has evaluations of surface water quality reports Their map portal links to data for their numerous sites. There are specific downloads for lakes for each year that has a variety of data including temperature, chlorophyll, turbidity, and many others.

LAWA (New Zealand)

Interactive maps and graphs of Trophic Level Index data for rivers and lakes in New Zealand. Other environmental data also available.





Gaps and restrictions in data

The majority of these data portals are at a national scale, which limits their use for assessing COPERNICUS products across the globe. Even in combination, the global coverage would be limited and focused in particular areas – mostly North America and Europe. Some of this data will also be fed in to global portals, meaning potential duplication of data and effort in analysis. However, some of the portals are extensive with good search and query functions, and countries like the USA and Canada are large and diverse, meaning satellite products could still be compared against a number of lake typologies, temperature and altitude gradients.

Other Regional Datasets

A number of other portals were found with either limitations on access, restricted sets of determinands and / or spatio-temporal coverage, but they could still potentially be useful for targeted cal/val analysis, especially in areas where the global data portals have little coverage. They are listed here with brief notes, where applicable:

Data portals and descriptions

- Lake Titicaca water quality data for 2003-2011
- Nova Scotia - surface water quality monitoring network (doesn't seem to be working properly)
- Great Lakes Water Quality Monitoring and Aquatic Ecosystem Health Data - 2021 data and archive data from 1960 available.
- Water quality monitoring Queensland, Australia
- Water quality Inland waters 2016 , Australia
- EOLakeWatch (cyanobacteria), Open Maps Data Viewer for WQ – both from the Government of Canada
- Note about 2020 Lake Victoria data in GEMstat

- Swedish environmental data - data via download and API since 1960
- Data from WorldBank for Lake Victoria – Data for 2014-16. NB link is currently broken.
- ICPR (International Commission for the Protection of the Rhine) - Water quality data up to 2019. Limited recent data available.
- AQUASTAT – has several different datasets available in the catalogue.
- Stats New Zealand – modelled lake water quality
- JRC portal - generalised data for Europe.

Recommendations for data centres

LIMNADES is the only dedicated global data centre collecting relevant optical properties and water quality data specifically for the purpose to support development of satellite-EO based water quality products for inland waters. Currently there are many limitations to this including on-going maintenance and collation of data and data access only to contributors, although this is proposed to change in the future. In addition to LIMNADES, there are some key global and regional in-situ water quality data centres, such as GEMS/Water and EIONET (Europe) that have a suitable structure, follow FAIR principles and relatively secure on-going funding to be developed further to support in-situ water quality data specifically to support EO services develop water quality products, such as CGLS. We recommend COINS provides specific help to these data centres in terms of providing guidance on in-situ data requirements for EO and design of ideal sampling protocols and metadata.

There may be potential to support development of regional data centres to help overcome some of the gaps in in-situ data availability. We recommend COINS supports the EO community, such as through GEO AquaWatch, to work closely with GEMS/Water or existing regional data centres, to more explicitly outline regional needs and design practical and affordable monitoring programmes to support filling gaps. Geo AquaWatch already supports LIMNADES evolution. Further activity for interchangeable data among different data bases would be useful. Plans that take a long-term perspective for sustainable data availability are essential.

4. Optical properties of lake waters

The optical properties of lakes are highly sensitive to global changes, such as climate warming and stratospheric ozone depletion, as well as to local changes in land use and associated pollution (deforestation, farming, urbanisation). The loss of water clarity or changing colour are often a clear indicator of declining water quality (Belzile et al. 2004). Water clarity, underwater light quality, the blueness of the water (spectral reflectance) and other optical properties of lakes are controlled by their inherent optical properties (IOPs), specifically absorption and scattering. The IOPs depend on the substances within the aquatic medium and not on the illumination conditions. By contrast, the apparent optical properties (AOPs) depend both on the properties of the water and on the ambient light field. The main IOPs are:

- the absorption coefficient (a) (m^{-1})
- the scattering coefficient (b) (m^{-1})
- the attenuation coefficient (c), with c = a + b (m^{-1})
- the scattering phase function (β) (sr^{-1})

The spectral absorption coefficient of natural waters, can be subdivided into four additive components: water molecules, coloured dissolved organic matter (CDOM), phytoplankton and non-algal particles (NAP) to give an overall absorption at each wavelength. Particle absorption is composed of heterotrophs such as bacteria, detritus, mineral particulates and bleached algal cells.

Similarly, scattering can be partitioned into the additive contributions from water molecules and suspended particles. Dissolved substances are generally assumed to have a negligible contribution to scattering.

Case 1 waters (ocean waters) are defined as those where all components, except the absorption and scattering of water medium, are assumed to covary with chlorophyll a concentrations (Chlorophyll-a). In these systems the IOPs can be estimated as a function of Chlorophyll-a (Morel, 1988). In coastal and inland ecosystems (case 2 waters), variations in absorption and scattering are more complex, as are the relationships between Chlorophyll-a and other components (S. Sathyendranath (Ed.) (2000). The particle backscattering coefficient in particular strongly affects the spectral reflectance (and thus colour) of lakes.

Summary of current data availability

Most of the IOPs and reflectance spectra taken from lakes belong to scattered research groups around the world, and there is not a centralized database that incorporates all. LIMNADES and AERONET-OC are the most populated datasets (see Chapter 3 for overview). Table 3 show the reflectances and IOP data that can be found in LIMNADES. Only Remote Sensing Reflectances (2136) and absorption (3038) are available.

AERONET-OC data contain several parameters of interest, many of them related to atmosphere characterization (NO2 optical depth (OD), Rayleigh OD, ozone OD, aerosol OD), meteorological data (water vapour, wind speed, pressure); information about sun and sensor angles; and the light-in water related variables like the sun irradiance the sky radiance and the radiance emerging from the water (Lw, Lwn) at various centre wavelengths in the visible and near-infrared spectral regions (412-1020 nm).

In addition to these data centres, individual research groups hold data collected by a range of specialist submersible instruments available from several specialist manufacturers, generally small companies established by researchers, who have themselves made significant contributions to the field. Examples include WET Labs, HobiLabs and Bioshperical Instruments in the US, Satlantic in Canada and Trios in Germany.

Other specialist instruments are needed to collect IOPs, including: Optical Laser diffraction instruments (LISST), Optical backscatter point sensor (OBS) and Use of Lidar for coastal habitat mapping. The sensor can be configured for autonomous operations, mounted in research vessels on a protective frame for profiling, or deployed for long periods in moored buoys. Sensors are available for planar and scalar irradiance measurements, using flat and spherical diffusers, and for radiance measurements with typical acceptance angles of 5-10°. The sensitivity of a radiometric sensor is largely determined by size of the detecting element, typically square millimetres for a multi-waveband sensor and square microns for a high resolution sensor.

Data accessibility/licencing

Most of the data in LIMANDES is licensed as D and E, that is:

- Licence D: the owner of the dataset can dictate how they want the data used and how they would like to be acknowledged. Co-authorship is not strictly required, but should be offered particularly where a publication draws heavily on data from one or more providers. Specific for non-commercial use.
- License E: users should contact the original data provider for approval at an early stage where there is an intent to use their data in publications. Co-authorship is not strictly required, but should be offered particularly where a publication draws heavily on data from one or more providers.

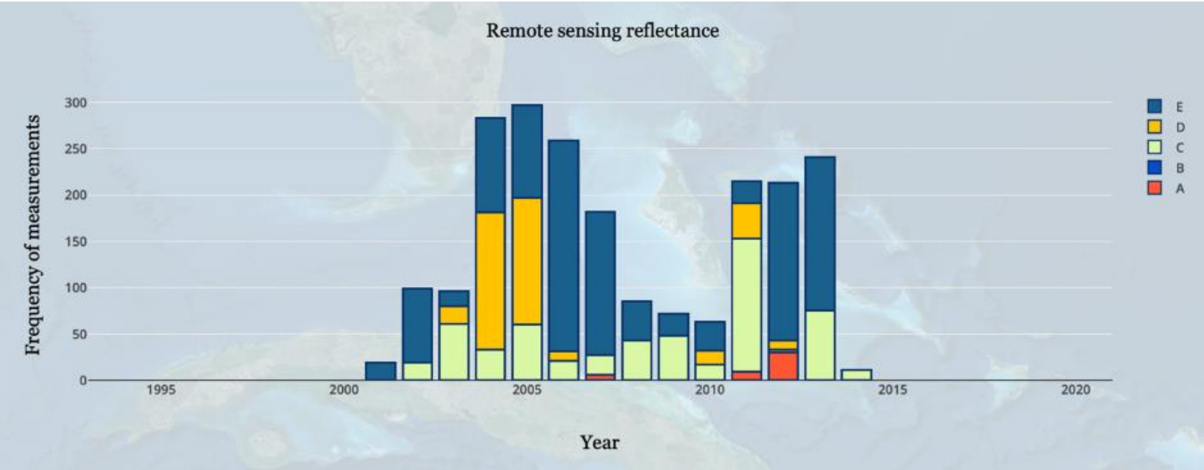


Figure 4. availability of reflectance data in LIMNADES

AERONET-OC data are free and downloadable from the webpage: https://aeronet.gsfc.nasa.gov

Currently data delivered are the results of AERONET processing version 3. The data are automatically cloud cleared and quality assured with pre-field and post-field calibration applied. Use of these data requires offering co-authorship to the Principal Investigator (PI). The public domain data are contributed by the International AERONET Federation. Each site has a Principal Investigator(s) (PI) responsible for deployment; maintenance; and data collection. The PI has priority use of the data collected at the site. The PI is entitled to be informed of any other use of that site data.

Gaps in Data

- Data from more regions and optical water types are required - some parts of the world are not covered by open or public datasets.
- Smaller lakes are poorly represented (e.g. AERONET-OC only covers six very large lakes.
- Hyperspectral data is especially needed, both for improving AC models, and to be able to adjust data to more satellite sensors.

Recommendations for increasing data availability/quality/access

The specialist kit required for data collection and the expertise required to use it are two key reasons why IOP data is mainly available for wealthier countries or regions, such as the USA, Europe and Australia. It highlights the need for regional campaigns where data gaps exist and for funding and capacity building to support better coverage in less developed countries.

A review and catalogue of (affordable) portable sensors that can deliver high quality hyperspectral IOP data could help increase data availability. Programmes to provide these alongside capacity building activities are needed, particularly in Africa, Asia and South America.

5. In-situ lake water quality data

The major factors which can influence the quality of inland water bodies include suspended sediments (turbidity), phytoplankton and harmful algal blooms of cyanobacteria, dissolved organic matter (DOM), nutrients, pesticides, metals and pathogens (Giardino et al., 2014). Only some of these factors affect the optical and/or thermal properties of waters, changing the signal acquired by optical sensors over water bodies.

The molecular scattering of pure water follows an approximately parabolic trend with higher values at short (ultraviolet) wavelengths, while the absorption is highest in the red-infrared region.

The particulate constituents that attenuate incoming light include suspended sediments (both organic and inorganic) and phytoplankton, particularly the main light-absorbing photosynthetic pigment of algal cells, chlorophyll-a, but also accessory pigments, such as phycocyanin in cyanobacteria (PC) (Simis et al., 2007). Light scattering by suspended sediments strongly depends on the particles size, shape, and composition while absorption by mineral particulates is usually low. The organic fraction of suspended sediments (suspended particulate organic matter, SPOM) and phytoplankton both absorb and scatter light appreciably. The spectral absorption of SPOM is similar to that of CDOM and contrasts with absorption by phytoplankton, where chlorophyll-a has two distinct absorption peaks at approximately 440 nm and 675 nm. The inorganic fraction of suspended sediments (suspended particulate inorganic matter, SPIM) scatters light significantly while its absorption is usually negligible. Overall, the absorption and back-scattering of light by these components of the water influences the shape and magnitude of the water-leaving reflectance, which is the information that can be retrieved by remote sensing sensors. The absorption of CDOM decreases exponentially with increasing wavelengths while it has negligible backscattering. CDOM dynamics are a subject of active research, particularly the importance of autochthonous (lake-derived) CDOM versus that originating from soil and vegetation decomposition in the catchment (allochthonous), and the variations in CDOM optical properties as a function of sources, composition and previous light exposure.

Summary of current data availability

LIMNADES contains water quality data that concerns four variables: phytoplankton chlorophyll-a, absorption by coloured dissolved organic matter, total suspended matter and phycocyanin (cyanobacterial pigment). Data availability and accessibility are shown in Figures 5 to 8.

Chlorophyll-a measurements can be made with a fluorometer or with a spectrophotometer. Data are available from 1994 until 2019.

There is much less LIMNADES data on absorption by coloured dissolved organic matter with datasets starting in 2001 and intermittent inputs until 2019.

The TSM dataset in LIMNADES spans a similar time period as the chlorophyll-a dataset, which suggests that in many campaigns both data are collected (Figure 7).

Phycocyanin, a pigment present in cyanobacterial cells, is much less abundant and only available for relatively few lakes for a scatter of years (Figure 8).

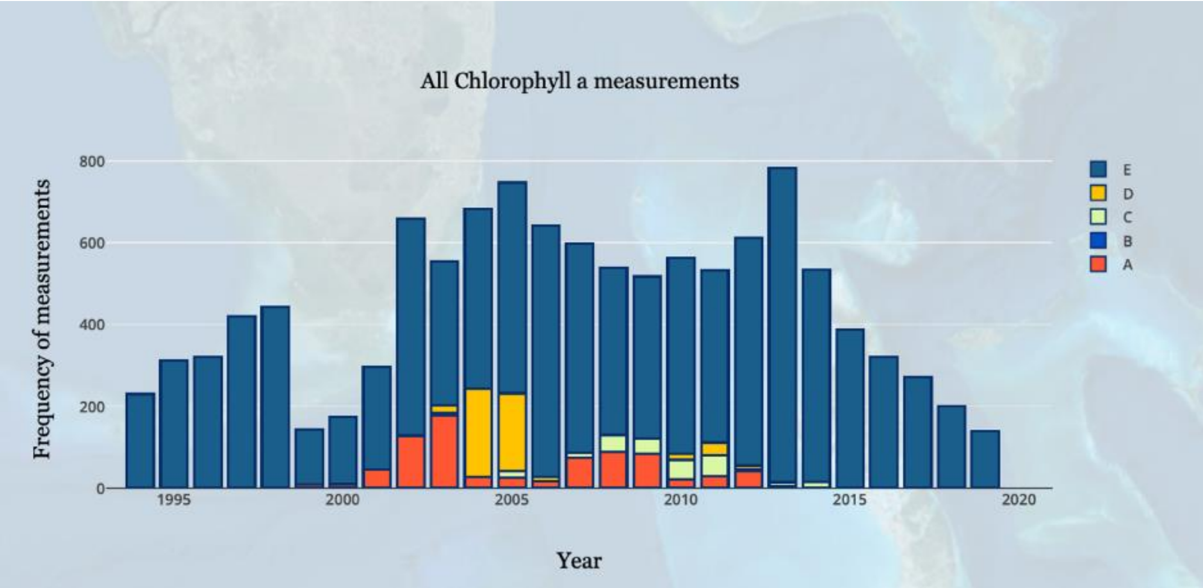


Figure 5. All chlorophyll-a measurements in the LIMNADES dataset

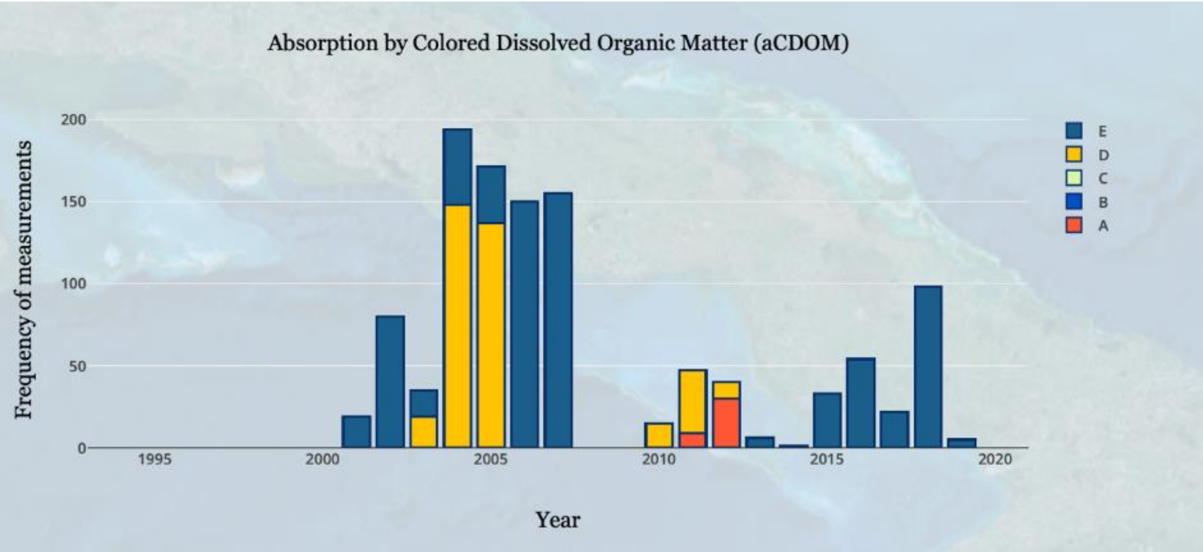


Figure 6, CDOM measurements in the LIMNADES dataset

Figure 7. TSM measurements in the LIMNADES dataset

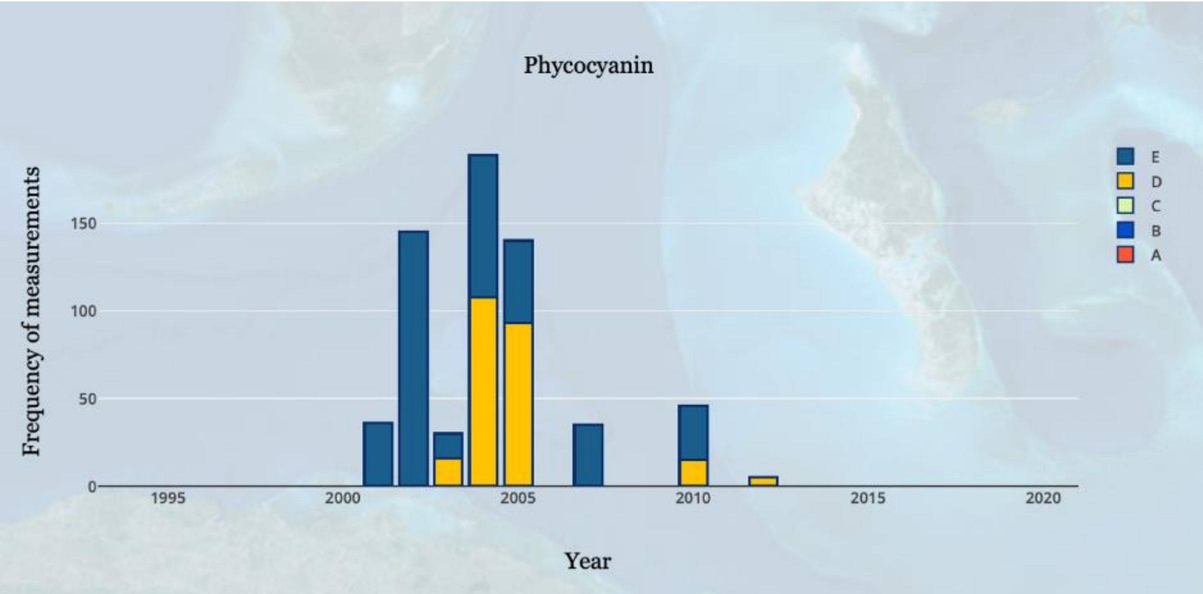
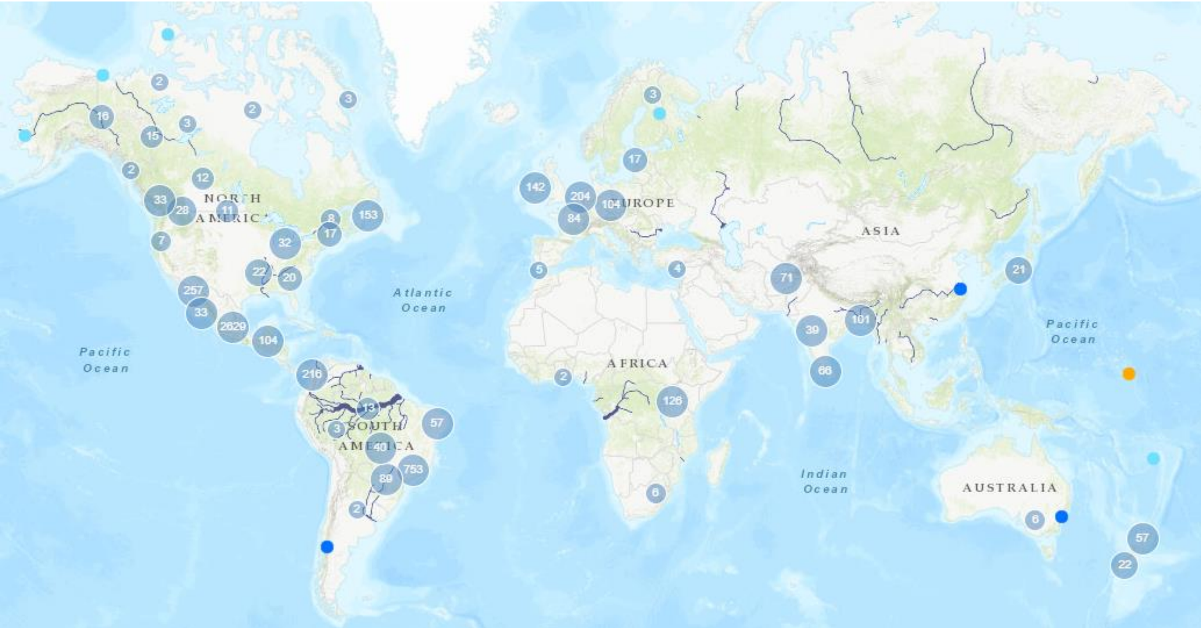


Figure 8. Phycocyanin measurements in the LIMNADES dataset

GEMS/Water

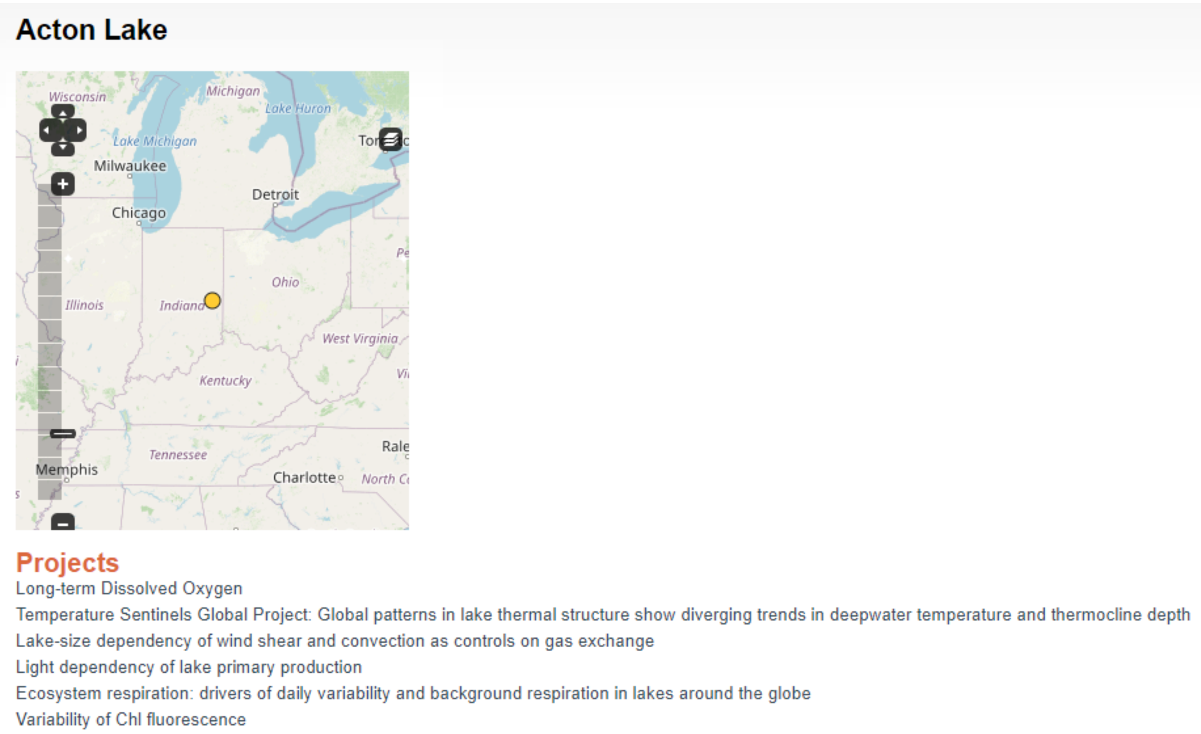
All GEMS/Water sites with chlorophyll-a and turbidity records are shown below, illustrating reasonable global coverage:



GEMS/Water chlorophyll-a / turbidity records

GLEON

Details on parameters recorded can be found using the data searches outlined in Chapter 3, or by selecting a specific GLEON site for more details on the projects being run there, e.g. for Acton Lake (https://gleon.org/lakes/acton-lake), below:



GLEON – details of Acton Lake and its associated projects

Each project can then be further explored for information on the parameters available, e.g. for the Long-term Dissolved Oxygen project above: https://gleon.org/research/projects/long-term-dissolved-oxygen.

Because of this need for site-specific data interrogation it is difficult to provide an overview of the data available for specific parameters in GLEON. This also highlights a complication for accessing data from many sites in GLEON, making it a time-consuming task.

AERONET-OC data contain chlorophyll-a data and IOPs in several bands, all derived from the water leaving radiance through bio-optical algorithms. Data availability per station are presented in section 3.2.3.

EIONET (Europe) - Waterbase

EIONET data from the European Environment Agency (EEA) Waterbase is available for 2019-2020. The available Waterbase Water Quality data relevant to EO-derived water quality is shown below

Determinand	ID	Total no. of samples	Mean	Min	Max	Number of sites	Avg. sample count / site
Turbidity	EEA_3112-01-4	1178	6.25	0.2	849	140	17
Total suspended solids	EEA_31-02-7	10009	8.33	0	1000	946	21
Cyanobacteria biomass	EEA_11-06-3	1482	2.65	0	459.00	256	11
Chlorophyll a	EEA_3164-01-0	31827	118.16	0.00083	532000	2440	26

Like LIMNADES there is particularly good coverage of chlorophyll-a and total suspended matter and lesser coverage of cyanobacterial biomass and turbidity. As LIMNADES was developed in Europe, much of these data may be common to both datasets.

Data accessibility/licencing

As described in Chapter 3 and section 4.2 for LIMNADES and AERONET-OC

Gaps in Data

As described in Chapter 3 and section 4.3 for LIMNADES and AERONET-OC

Recommendations for increasing data availability/quality/access

Recommendations include establishing stronger communication lines between in-situ and EO communities. Other recommendations cover the following limitations highlighted earlier:

- Comparability with EO data (frequency, locations and measurements)
- Availability
- Accessibility
- QA/QC

Comparability

Most in-situ data is collected for regulatory or research purposes and not with the purpose of cal/val of EO data. There is, therefore, a clear need of more explicit design or fit of in-situ water quality sampling campaigns to ensure temporal and spatial match-up with satellite overpass.

For algorithm development and calibration purposes, collection of more comparable parameters is needed (e.g. phycocyanin concentrations) Figure 9 highlights how awareness could help increase data availability/quality/access of water colour and transparency. If in-situ data providers were more aware of instrumentation options for water sampling beyond a Secchi disk, more measures of higher quality data could be made. This involves more investment in instrumentation, and a better coordination for calibration and measurement protocols, but it opens up new possibilities, largely unexplored.

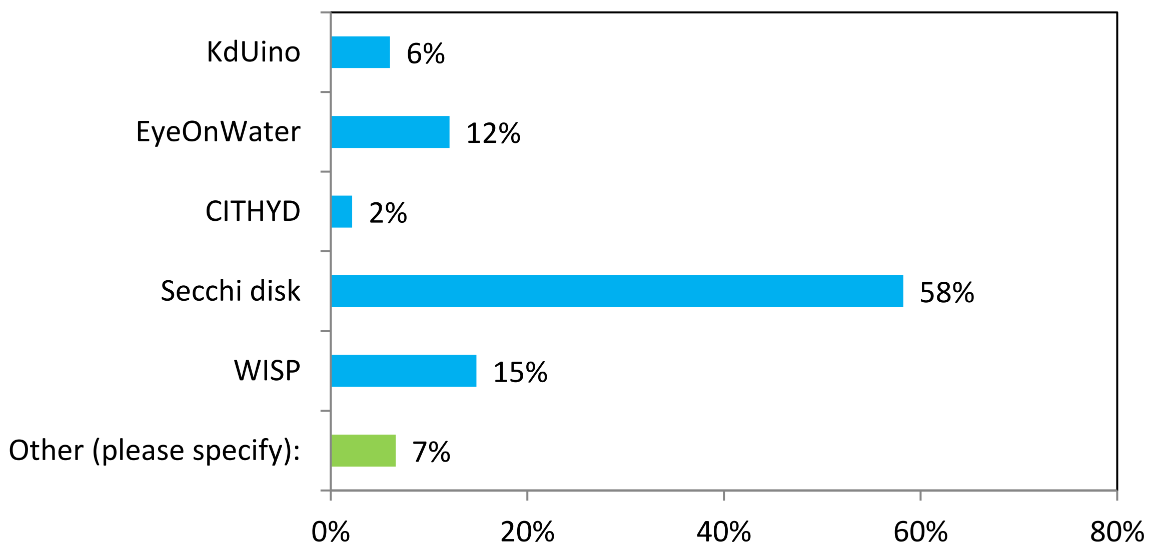


Figure 9: Familiarity with currently available hand-held solutions to measure water colour and transparency (MONOCLE survey)

Accessibility

Availability of metadata information also needs to improve, with the need for a more centralised collection of meta data to generate a database useful for many sectors. Metadata information should include data about the location, the protocols followed, the calibration method, uncertainties, the license type and ownership and some kind of identification code, to understand if the measure is part of a campaign, deployment, sensor, etc.

Table 4 summarizes the minimum requirements established by the MONOCLE project concerning metadata information (Heard et al., 2018).

Table 4.Minimum metadata requirements to enter new observations into the MONOCLE system (courtesy of Stefan Simis)

Category	Element	Description	Possible values, data type, conventions
Location/time	Latitude	Geographic location	Float decimal degrees, north positive
	Longitude	Geographic location	Float decimal degrees, east positive
	Elevation	Height above reference ellipsoid	Float in meters
	Reference Coordinate System		Default WGS84
	Time	Time in Coordinated Universal Time (UTC)	Character string formatted according to ISO8601
	Location_source	Source of the Geodetic information	e.g. GNSS
	Time_source	Source of the Time information	e.g. GNSS, internet time pool
Processing	Processing_level	Sensor-specific	0, 1, 2 … n including sublevels such as 1A, 1B, 1C. Defined by manufacturer and described in the reference documentation. Level 0 is uncalibrated sensor output and not distributed; Level 1 is calibrated data prior to any corrections or interpretation; Level 2 is interpreted data; Level 3 is aggregated or regridded data.
	Processing_procedure	Reference to protocols and algorithms describing the steps involved in data processing	URL
	Processing_version	Version of the data processing software	Free form, recommended: major.minor.build
	Processing_revision	Incremental version of the processed data	Free form, likely an integer
	Calibration_procedure	For calibrated data: documentation describing the calibration procedure. Can be the same as Processing procedure reference	URL
	Calibration_reference	Identifier of calibration information	Flexible, system-specific
	Calibration_time	Date/time stamp of applicable (uncalibrated data, if available) or applied (calibrated data) sensor calibration.	Character string formatted according to ISO8601
	Calibration_version	Version of the calibration processing software	Free form
Identifier	Sensor_id	Unique identifiers used to prevent data duplication with data consumers	Sensor serial number
	Platform_id		Platform serial number or randomly assigned identifier (UUID) used with all connected sensors. May be left empty if not applicable.
	Deployment_id		Randomly assigned identifier (UUID) specific to deployment sequence (e.g. cruise, campaign, vertical profile) of **this** sensor. Not shared with other sensors.
	Sample_id		Randomly assigned identifier (UUID) generated with each distinct data record from any **set** of sensors belonging to a single observation.

	Observer_id		Randomly assigned identifier (UUID) repeated with each data record from this and/or other sensors when operated by a specific observer.
License	Owner_contact	An email address where the owner of the data can be contacted now and in future	Sustained email address (e.g. data@organisation.org rather than individual@organisation.org)
	Operator_contact	An email address where the current operator can be contacted	Email address to operator or group of operators
	License	A licence string or coding that is either self-explanatory or detailed in the License_reference field.	Free form
	License_reference	A reference describing the data license in detail.	URL
	Embargo_date	A date following which the data may be used according to the specified license. Used, for example, to hide the data record in NRT visualization until quality control is completed.	Character string formatted according to ISO8601

6. Citizen Science

Citizen Science (CS) has long been recognised as an inexpensive way of gathering large amount of environmental data in many locations, potentially also at high frequency. Studies have evaluated how this data can be used complementarily in environmental research and agency monitoring (Hadj-Hammou et al. 2017; Loiselle et al. 2017) as well as for achieving specific monitoring and assessment goals e.g. the United Nations sustainable development goal SDG 6.3.2 (Quinlivan et al. 2020a; Quinlivan et al. 2020b).

In this chapter we review the challenges and benefits of using CS data for monitoring of water quality in general, and specifically related to those water quality parameters considered the greatest relevance to the calibration and validation of EO products i.e. chlorophyll-a, algal blooms, water colour, dissolved organic carbon and water clarity / turbidity. We also review CS data on nutrients which are highly complementary to satellite EO data on water quality, particularly as a driver of chlorophyll-a and cyanobacterial blooms. Where available, we sought any CS data explicitly gathered in relation to water quality monitoring by satellite Earth Observation (EO).

The chapter includes a data review of existing CS schemes and projects used in monitoring of the key EO-sensed water quality parameters listed above. This review of existing schemes is used to identify potential sources of data for calibration and validation of EO water quality monitoring services, or potentially schemes that could be encouraged to collect data to support development of EO water quality products.

This overview of challenges, benefits and data availability concludes with final recommendations related to the use of CS data to validate, calibrate and complement EO services for water quality monitoring.

Summary of current data availability

An initial literature review was carried out using Web of Science (WoS Advanced Search) with the following scope: English language, all documents, Dates: 1970-2021, all collections. The search string was for titles (TS) with the following terms:

TS = (“citizen Science” or “community science” or “volunteer monitoring” or “crowd-source” or “public engagement”)

AND

TS = (nutrients or turbidity or “water clarity” or DOC or Chlorophyll-a or chlorophyll-a or cyanobacteria or “trophic state index” or “surface reflectance” or DOC or CDOM or “optical properties” or “suspended sediments”)

The number of relevant papers identified in the WoS search (dated 21/07/2021) are summarised in **Error! Reference source not found.**. The title and abstracts of the 104 records where both search terms were present were reviewed to produce a shortlist of publications for more detailed review (advantages, disadvantages, etc.) of CS methods for collecting water quality data.

Table 5: Results from Web of Science search of relevant citizen science literature

Set	Results	Search string
#1	9,795	TS=("citizen Science" or "community science" or "volunteer monitoring" or "crowd-source" or "public engagement")
#2	744,866	TS=(nutrients or turbidity or "water clarity" or DOC or chlorophyll-a or Chlorophyll-a or cyanobacteria or "trophic state index" or "surface reflectance" or DOC or CDOM or "optical properties" or "suspended sediments")
#3	104	#1 AND #2

A structured google search was also carried out (e.g. "citizen science" + "water quality" + "chlorophyll") to produce a catalogue of potentially relevant existing schemes/data sets. In addition to this, we were provided with an inventory of schemes identified in a similar review carried out for GEO AquaWatch (Justine Spore, pers. comm). Based on the Web of Science and Google searches, the following information was gathered on data available from the catalogue of relevant citizen science schemes:

- Campaign-based or continuous?
- Period available
- Region of interest: regional to global
- No. of records/samplings
- No. of sites – easily available or not?
- Frequency of observations – easily available or not?
- Accessibility of data – especially sources with APIs

The review of existing CS schemes identified 25 different global, regional or national schemes relating to lake data from one, or more, of the EO-related water quality parameters (nutrients, chlorophyll, algal blooms, water clarity / turbidity and water colour / DOC) (Table 7). Most additional sub-national schemes identified in the USA, or schemes for rivers, streams, or coastlines were not reviewed further in this study.

The monitoring parameters most associated with citizen science data collection were algal blooms (10) and water clarity (9). CS schemes designed to collect nutrients (6), water colour (5) and chlorophyll-a (2) were less common and there were no schemes for DOC. Several projects monitored more than one parameter (Table 7).

Table 6: Summary of citizen science schemes collecting water quality data of key EO-sensed parameters. NA = not available

Water quality parameter	CS schemes	Sources of data availability	Spatial scale	Temporal scale (earliest to latest)
Nutrients	6	Available by login, membership or request (2) Freely available (2) NA (2)	Global (1) Germany (1) USA (4)	1993 Ongoing
Chlorophyll	2	Available by login, membership or request (1) NA (1)	Local USA (2)	1993 Ongoing
Algal blooms	10	Available by login, membership or request (3) Freely available (2) NA (5)	UK, Belgien, Netherlands (1) USA (5) Local USA (3) Local Ivory Coast (1) Local NI, Ireland border (1)	2001 Ongoing
Water clarity / turbidity	9	Available by login, membership or request (2) Freely available (3) NA (4)	Global (5) Local USA (3) Local India (1)	1988 Ongoing
Water colour / DOC	5	Available by login, membership or request (1) Freely available (2) NA (2)	Global (4) Local India (1)	2019 Ongoing

Figure 10 shows that 11 out of 25 projects are still ongoing and 4 of these have a long-term dataset that started in 2001 or earlier; Bloomwatch and Phytoplankton Monitoring Network (PMN) specialise in the monitoring of harmful algal blooms in the USA and North America respectively, the Secchi Dip-in for water clarity (USA and globally) and Waquoit BayWatchers (local to Waquoit Bay, USA) (). For 12 of the schemes we were unable to identify the start date of the project (Figure 10), however all these were still ongoing and the start date is likely to be available upon request.

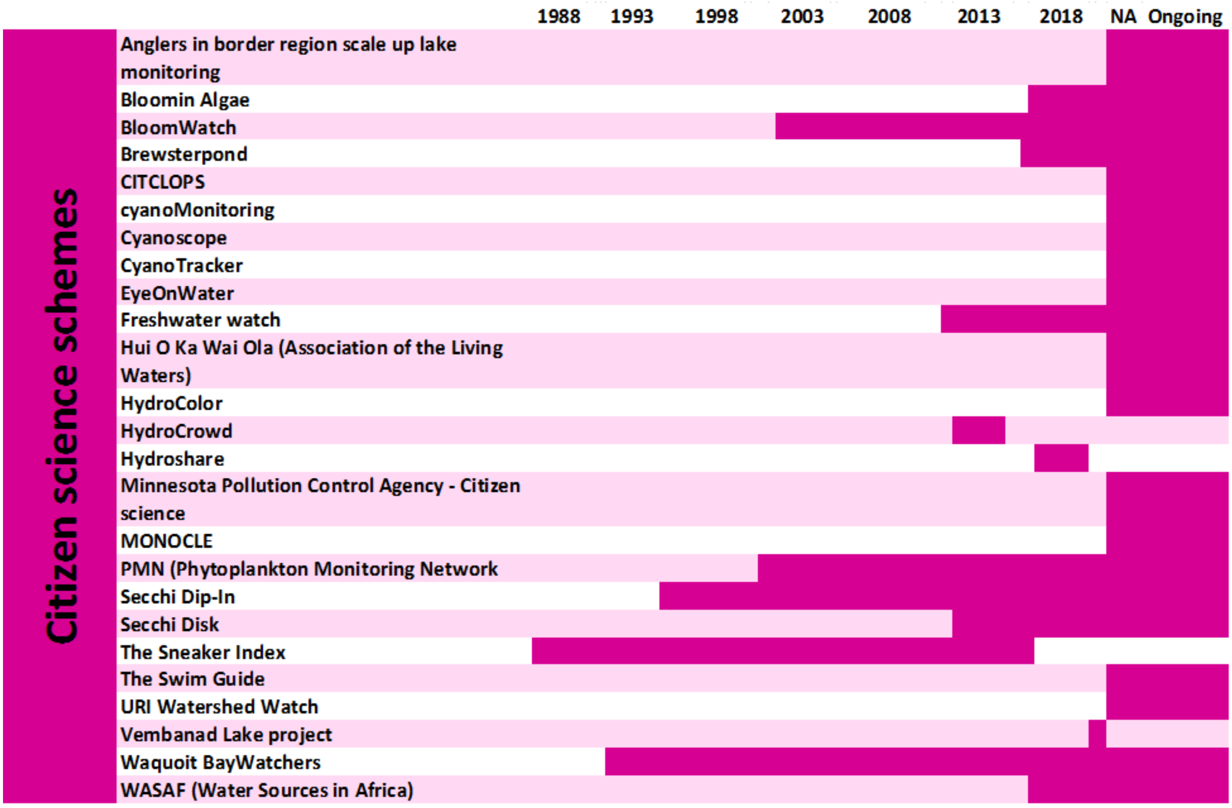


Figure 10: Overview of the temporal scale of data availability of 25 relevant CS schemes. Where there was no available start date, they are labelled as NA.

The CS schemes found were unequally distributed spatially (Table 7). Eight schemes had a more or less global scope (some of which included specific locations for campaigns or spatial biases from one country i.e. USA or Australia; ten schemes were completely local (USA: 7, Ireland / Northern Ireland: 1, India: 1 and Ivory Coast: 1); five were specific to the USA; one to UK, Belgium and the Netherlands; and one specific to Germany. The spatial bias highlights a severe under-representation of CS data from Africa, Asia and South America.

Table 7: Summary of citizen science schemes collecting water quality data of key EO-sensed parameters. NA = not available

Water quality parameter	CS schemes	Sources of data availability	Spatial scale	Temporal scale (earliest to latest)
Nutrients	6	Available by login, membership or request (2) Freely available (2) NA (2)	Global (1) Germany (1) USA (4)	1993 Ongoing
Chlorophyll	2	Available by login, membership or request (1) NA (1)	Local USA (2)	1993 Ongoing
Algal blooms	10	Available by login, membership or request (3) Freely available (2) NA (5)	UK, Belgien, Netherlands (1) USA (5) Local USA (3) Local Ivory Coast (1) Local NI, Ireland border (1)	2001 Ongoing
Water clarity / turbidity	9	Available by login, membership or request (2) Freely available (3) NA (4)	Global (5) Local USA (3) Local India (1)	1988 Ongoing
Water colour / DOC	5	Available by login, membership or request (1) Freely available (2) NA (2)	Global (4) Local India (1)	2019 Ongoing

A number of schemes were considered particularly relevant:

Citizen Science Project	Website	Water Quality Parameter	Area
Bloomin'Algae	https://www.ceh.ac.uk/algal-blooms/bloomin-algae	Harmful Algal Blooms (Cyanobacteria)	UK
CyanoTracker	http://www.cyanotracker.uga.edu/	Harmful Algal Blooms (Cyanobacteria)	USA
EyeOnWater	https://www.eyeonwater.org/	Water Colour and Clarity	Global
FreshWaterWatch	https://freshwaterwatch.thewaterhub.org/	Clarity, Nitrate, Phosphate	Global
HydroColor	http://misclab.umeoce.maine.edu/research/HydroColor.php); Leeuw and Boss (2018)	Remote-Sensing Reflectance	Global
Secchi Dip In	http://www.secchidipin.org/	Water Clarity	Global
Secchi Disk	http:// http://www.secchidisk.org/	Water Clarity	Global

Benefits and challenges for monitoring specific water quality parameters using Citizen Science

This section describes the benefits and challenges associated with retrieving data using CS on the parameters of main interest.

The benefits can be summarised as:

- Cost-effectiveness of sampling and analysis
- Potential for high spatial and temporal coverage (relative to agency monitoring)
- Complementarity with EO data
- Public engagement in water quality (environmental stewardship / ownership)
- Accessibility – normally public

The challenges include:

- Data quality and credibility (user or analytical error difficult to measure or verify)
- Collection bias (spatially to more populated regions, seasonally to warmer weather and bias towards weekends)
- Comparability of data
- Organisational challenges – maintaining volunteer interest or low participation rates generally, low participant diversity leading to sampling bias

CS challenges also include some of the challenges described earlier for in-situ monitoring data more generally, with regards to sampling water that is representative for the entire waterbody and the types of water sample taken (surface sample vs. integrated water sample). Citizen science is likely to be more likely collected from a shore than in a boat, whereas EO services often mask out shoreline sections of lakes in order to avoid interference from reflections of shoreline, or growth of benthic macrophytes and sediment resuspension in the shallows. A shoreline collected sample (whether calculated using a smartphone adaptor or analysed in a lab) may, therefore, not be directly linked to EO images of the same lake even if the satellite data and volunteer collection happened on the same day. For coarser validation purposes, CS data from shoreline samples may still be useful. The types and quality of data available differ for each parameter and are, therefore, discussed separately below:

Chlorophyll

Very few studies identified in the literature focus on using CS for measuring or estimating chlorophyll-a in lakes for water quality (2 out of 108 reviewed) and only 2 out of 25 CS projects were identified which aimed to measure chlorophyll-a, both at sub-national scale: Waquoit BayWatchers (WBB) and URI Watershed Watch (URIWW)). URIWW volunteers sample waterbodies, samples are then sent for chlorophyll-a analysis at a certified lab. No information was available on how WBB collect their data, but both CS projects are closely linked to local official monitoring of specific water. The two studies indicate that the process of collecting water samples for chlorophyll-a analysis is something volunteers could seemingly easily undertake. The model of using certified labs should ensure high quality data, highlighting an approach that could be used for cal/val of chlorophyll-a algorithms.

There are some developing technologies, which may make more direct data collection of chlorophyll-a data by citizens possible in the near future. Friedrichs et al. (2017a) developed the SmartFluo, an extension for smartphones, which enables the attachment of a cuvette for measuring chlorophyll fluorescence using a smartphone camera. This makes in-field measurements possible, eliminates costs for lab analysis and shipping as well as decreases any errors generated by any delay in lab analysis and shipping the samples.

More sophisticated (and well financed) CS schemes could incorporate portable sensors, such as the BBE AlgaeTorch, which measures total phytoplankton chlorophyll-a, cyanobacterial chlorophyll-a and turbidity in real time along with GPS coordinates. These are generally easy to use devices, which take little training to use. For example, as part of the Loch Leven long-term monitoring programme, a volunteer from the local fishery uses an algal torch to record data during spring, summer and autumn seasons from a boat at several locations in the lake at least weekly. The volunteer is also provided with a calendar of the dates and times that the ESA Sentinel satellites are passing over, to sample on any cloud-free day. This is a time demanding job for a volunteer, which is only possible because of the well-established relationship between the volunteer and a specialist water quality research group (UKCEH). However, coordination of this sort of volunteer arrangement at several sites, particularly for regions where cal/val data is missing, may be a solution to deliver high quality chlorophyll-a data for cal/val of EO products.

Algal blooms

Harmful algal blooms of cyanobacteria (cyanoHABs) are a widespread problem globally which impact on public and animal health and because they form surface blooms, often collecting along shorelines, they can often be readily identified by members of the public, with potentially minimal training of CS volunteers (Thornhill et al. 2018). For this reason, concerned citizens, such as swimmers or dog owners, can be highly motivated to take part in citizen science to validate the presence (or absence) of blooms. Delivering quantitative data on cyanobacteria requires collaboration between citizens and professional laboratories or the use of sophisticated sensors (such as the BBE Algal Torch) as described for chlorophyll-a. However, more widely available photographs of blooms sent in by citizens alongside date and location data using a smartphone's built-in GPS, still have value in providing qualitative data for validating the presence or absence of cyanoHABs identified in satellite imagery. Due to other algal pigments having absorption spectra that can potentially interfere with the phycocyanin pigment signal from cyanobacteria, the spectral data provided by satellites is less clearly attributed to cyanobacteria alone. This technical difficulty is one reason why qualitative data from citizens provide a quality check to cyanobacterial blooms identified by satellites, and vica versa, satellites can provide confirmatory evidence of cyanoHABs when citizens photographs are not sufficiently clear.

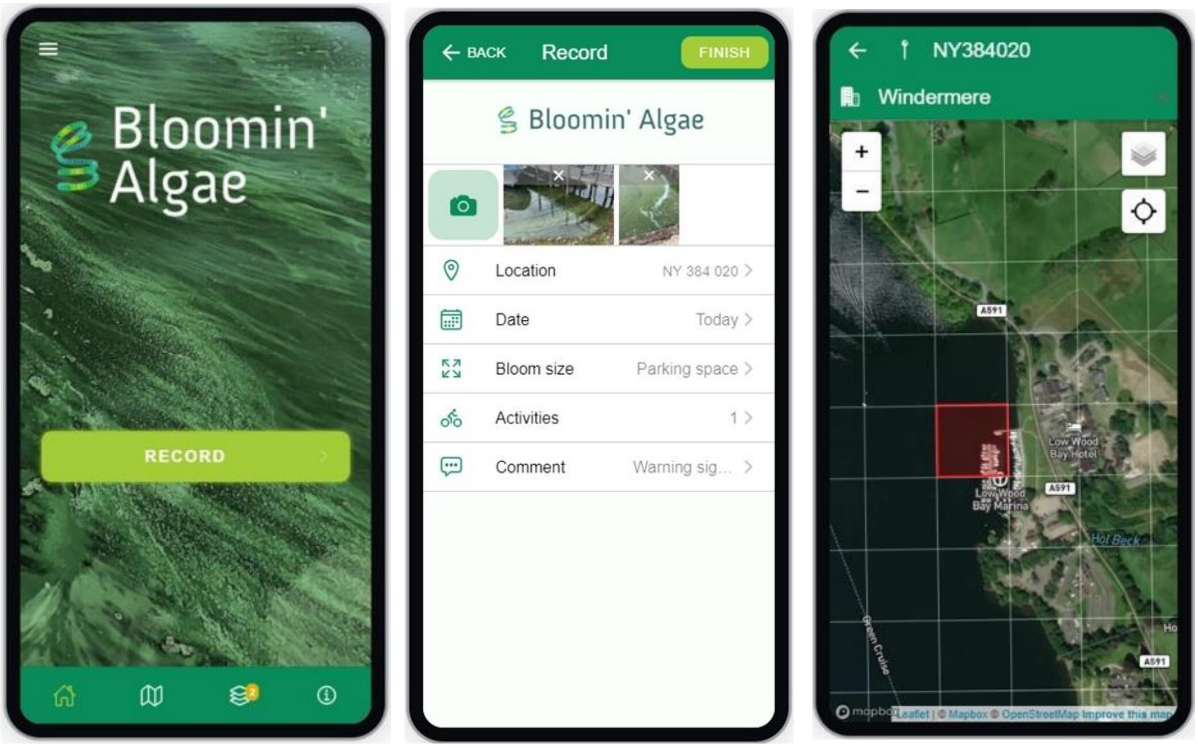
Although cyanoHABs are more common in nutrient enriched lakes, they can occur under suitable weather conditions in more localised accumulations even in large, nutrient-poor lakes at any time of the year. The unpredictable nature of blooms, and their widespread occurrence, makes regular in-situ monitoring by agencies practically impossible. This highlights the benefits of using wide-scale, high frequency monitoring by satellites and citizens for public and animal health surveillance. Apart from the (cross-validation purpose highlighted above, there are many other reasons why these two forms of monitoring are highly complementary for cyanoHABs:

- EO can provide early warning of developing cyanoHABs before they are obvious to the public

- EO provides consistent high-frequency monitoring across whole water bodies and across large landscapes, particularly where citizen scientists may not be active
- Citizens provides data on cyanoHABs during cloudy periods
- Citizens can provide evidence of cyanoHABs accumulating along shorelines and in very small waterbodies, not easily monitored by satellites

Two of the CS projects that use mobile apps for recording location and images of algal blooms are Bloomin' Algae and CyanoTracker (Carvalho 2021; Scott et al. 2016).

Bloomin' Algae case-study



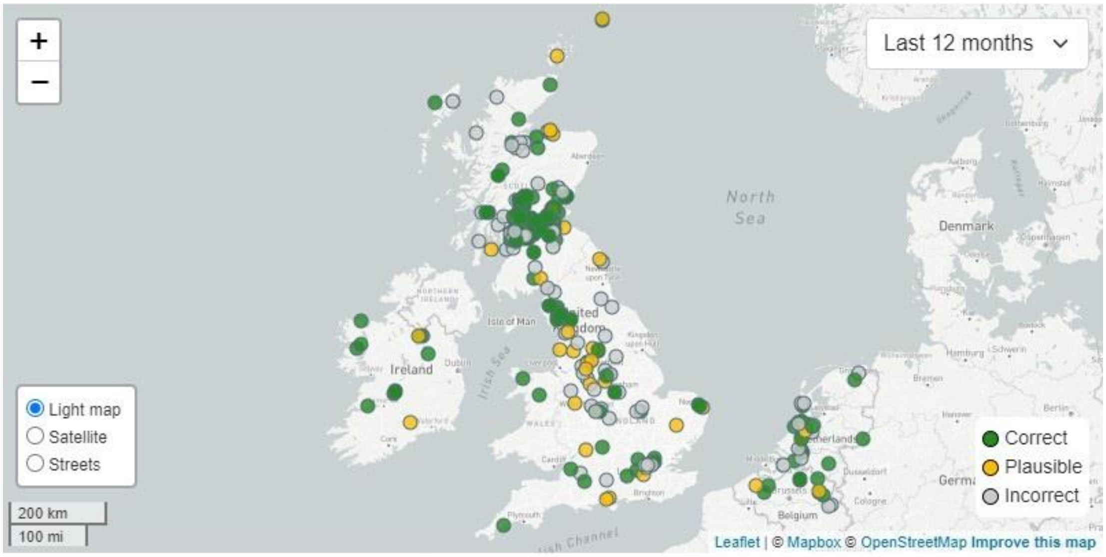
Bloomin' Algae is a Citizen Science (smartphone) app for reporting the presence of harmful algal blooms of blue-green algae (https://www.ceh.ac.uk/algal-blooms/bloomin-algae). The app helps speed up public health warnings and can help teach citizens how to recognise risks to public and animal health. As well as recording an accurate location using a phone GPS and the date and time of a suspected bloom, app users submit a photo of the bloom for verification by experts. For this reason the data on the presence/absence of cyanoHAB is usually high quality.

Experts aim to check the photos and provide rapid feedback on the record, within one or two days of recording. CyanoHAB experts provide a quality checked dataset of confirmed cyanobacterial blooms vs other confounding water quality issues (e.g. surface blooms of blanketweed or duckweed) which may also interfere with the satellite signal

Submitted record data and photos are open access and can be viewed on maps in the app or on the CS scheme website.

The app does not produce quantified data on cyanobacterial abundance but produces qualitative data on the presence or absence of cyanobacterial blooms, which could provide important validation data for development of cyanobacterial bloom algorithms. There are possibilities to develop the app to incorporate satellite images of blooms from the CyanoLakes service with prompts to app users of potential blooms in their area for validation checks (Carvalho, pers. comm.). The addition of satellite-EO data is likely to further engage citizens and provide a greater weight of evidence approach for public and animal health risk management.

Although the app can be used anywhere in the world, until recently, it has only been promoted in the UK, and within the UK there is a spatial bias to Scotland where it has been most promoted (Dobel, A., pers. comm. 2021). In 2021, the spatial coverage has expanded to three other European countries: Belgium, Netherlands and Ireland (Figure 11). The number of sites monitored is >150 lakes and reservoirs by >100 recorders with blooms recorded in most months if the year, although there is, as expected, more records in summer when blooms are most likely to occur in temperate latitudes (Figure 12).



Correct	Plausible	Incorrect
202	56	112

Figure 11. Overview of records from 2021 submitted via the Bloomin' Algae app from UK, Belgium, Netherlands and Ireland.

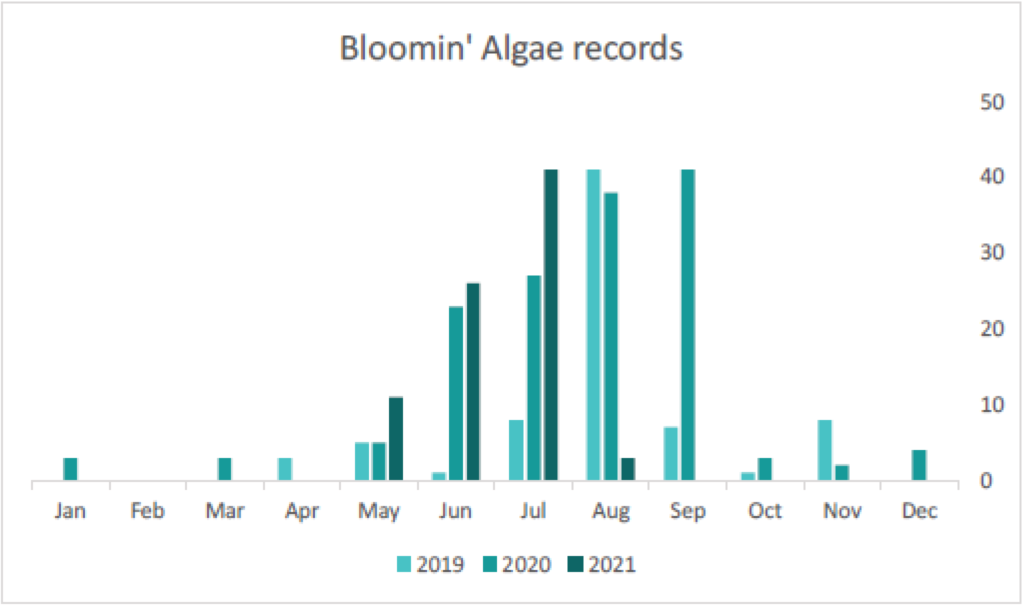


Figure 12. 2019, 2020 and 2021 records submitted to the Bloomin' algae app with a seasonal pattern in bloom records, with majority of records submitted from June to September.

Water colour and Dissolved Organic Carbon (DOC)

Water colour and DOC are often correlated, although water colour can also be affected by other water constituents such as phytoplankton biomass. Here we discuss them together as the publishd literature on both parameters were very limited (1 out of 108 captured references for both cases). The Google search was more successful and showed water colour being used in 5 out of 25 CS projects. This is likely a reflection of the simplicity in recording water colour, which makes it ideal for untrained volunteers. DOC was not, however, explicitly covered in any of the five citizen science schemes.

Two smartphone apps, i.e. Hydrocolour (HC) and EyeOnWater (EoW), were evaluated for assisting water quality monitoring in inland waterbodies in Australia based on water colour records (Malthus et al. 2020). They show that the apps are near, but not quite at, the same standard as scientific spectrometers, but that this improves with replicated records. The EoW app represents water colour with high reliability. HC experienced issues with recording of images (glare, sunlight, screen light settings on the phone).

According to the reviewed literature, water colour, and potentially CDOM, is a good parameter for monitoring by CS projects, as there seem to be affordable and simple measures to use for recording these parameters. Some studies have, however, found a low correlation between CS and researcher made measurements of colour (George et al. 2021), suggest that there could be bias in water colour data when obtained from CS.

Several smartphone adapters have been developed for measuring water colour.

Friedrichs et al. (2017b) report an extension to the SmartFluo adapter (described under chlorophyll section) for measuring FDOM (a proxy for Chromophoric or coloured Dissolved Organic Matter, cDOM). The device was of DIY type and the measurement was based on fluorescence with quality assurance checked by a lab luminescence spectrometer, with a high level of agreement (R^2 = 0.99)

(Friedrichs et al. 2017b). In 2020 Burgraff et al. also introduced an adapter (iSpex 2) that allows a smartphone camera to measure aerosols and ocean water colour. Production, calibration and validation of the adaptor was meant to commence in 2020 and a demonstration of iSpex 2 performance must be close at hand.

Water clarity / turbidity

Water clarity relates to turbidity or transparency of the water, which is dependent on the amount of suspended solids in the water (e.g. sediments or phytoplankton). According to our data review both parameters are often collected using CS (9 out of 25 CS schemes collected water clarity or turbidity data), with one record starting in 1988 and continuing until 2018 (the Sneaker Index), while another one started in 1994 and is still ongoing (The Secchi Dip-in).

The two parameters have been collected in different ways:

Using a turbidity tube Scott and Frost (2017) found that turbidity measures collected by CS had a good relationships to lab based results ($r^2 = 0.68$), however one limitations shown was that the majority of data (52%) were at, or below, the lowest possible value that could be recorded with the equipment used (Scott and Frost 2017).

Water clarity assessed using a Secchi disc, an inexpensive simple wooden disc painted black and white, lowered into the water until it is no longer visible. George et al. (2021) tested a 3D printed mini Secchi disc for CS measurements of water clarity at Vembanad Lake in Kerala, India. The study found good coherence between the CS measured Secchi depth and verification results and following a cost-effectiveness analysis, they also found that the project data had been delivered at practically no cost (George et al. 2021). This suggests that water clarity is an ideal and easy to assess parameter in CS projects for water quality. George et al. (2021) also derived turbidity data from the Secchi disc measurements and found good similarity with scientist measurements using CTD. Deutsch et al. (2021) found that a global algorithm could accurately depict the relationship between in-situ records of Secchi disc depth and satellite observations (Landsat 8 Blue/Red band ratio) in different lake types in Canada. The fit improved further using a filter to remove outliers (Deutsch et al. 2021).

Poisson et al. (2020) found that CS contributed 82 % of lake water clarity data in a US database containing data from seven different states during the last 31 years.

In our CS data review there are 5 fully or partially global CS schemes, which collect water clarity data: Freshwater Watch (specific locations, from 2012 and still ongoing, data available under creative commons), the Secchi Dip-in (mostly USA, from 1994 and still ongoing, data available from two web portals), Secchi Disk (from 2013 and still ongoing, data-explorer only), Eye-On-Water (EoW) (Start date of project unavailable, started as an Australian initiative, however, now described as a global scheme with data available via web portal), Citclops (Start date of project unavailable, but the data collection is ongoing, data available via EoW web portal).

It was not possible in this brief review to readily evaluate the spatial and temporal scale of the available data (Freshwater Watch, Secchi Dip-in, EoW and Citclops). There is clear possibility that the data could be compiled and be complimented by further local or regional scale CS campaigns. This highlights the real potential for CS to expand and be the primary source of data for cal/val of EO-derived water clarity/turbidity products across large geographical scales.

Nutrients

Nutrients cannot be directly measured using EO optical or thermal data. In-situ data may, however, be useful for validating measures of algal and cyanobacterial biomass as they are highly correlated. The measurement of nutrients in water is a quantitative measure that typically depends on highly developed and sensitive field equipment. It is, therefore, one of the more difficult measurements to deliver through CS sampling campaigns without relatively expensive equipment. The largest CS campaign to produce nutrient data is Freshwater Watch with data from 2,500+ sites globally, with some datasets more than a decade long. The test kits used by FWW produces a categorical classification for a sample's nitrate or phosphate concentration using colorimetric methods. For example, for nitrate citizens identify the nitrate concentration within seven specific classes ranging from 0.2 to 0.5, 1, 2, 5 and 10.0 mg/L. Scott and Frost (2017) argued that even though FWW data on nutrient concentrations could be useful to see patterns on a global scale, there were limitations of class-based classification for finer scale analysis. This semi-quantitative data could, however, potentially be used for validation of EO-derived measures of algal biomass, if chlorophyll-a or cyanobacteria data did not exist.

Data accessibility/licencing

Data was directly available from portals or websites in seven schemes; five schemes had data available upon request and acknowledgement or by membership (potentially unpaid); 13 had either an online data explorer or no clear indication of data availability (Table 7).

Gaps in Data

- Parameters missing from CS schemes: DOC and limited for Chlorophyll-a and nutrients
- Poorly monitored regions – bias to USA and Australia
- Temporal coverage – few long-term datasets

Recommendations for increasing data availability/quality/access

Our literature and data review has outlined a number of opportunities to be further examined:

- Better communication lines between CS and EO communities to improve data collection and use of both CS and EO data by both communities
- Developing tailored CS schemes specifically to support development of EO data products
- Better (cross-)validation of EO and CS data products
- New smartphone sensors to support high quality data collection e.g. iSpex-2

There are also some specific partnerships which merit further consideration where better communication could lead to benefits for both EO and CS communities:

- o Turbidity products: engaging with Secchi dip-in
- o Harmful algal blooms of cyanobacteria: engaging with Bloomin' Algae
- o Water colour and cDOM: engaging with Eye on Water

7. Summary and Conclusions

Developments in communication and coordination

The GEO AquaWatch Initiative has as its over-arching goal to develop and build global capacity and utility of Earth Observation-derived water quality data, products, and information to support water resources management and decision making. GEO AquaWatch focus its actions on improving the coordination, delivery and utilisation of water quality information for the benefit of society. The objetives of AquaWatch are:

- Objective 1: Facilitate effective partnerships between the producers, providers and users of water quality data, products and information.
- Objective 2: Improve analysis and integration of in-situ and remote sensing water quality data.
- Objective 3: Develop and deliver fit-for-purpose water quality products and information services.
- Objective 4: Support technology transfer and access to water quality data products and information.
- Objective 5: Advocate for increased education and capacity for the use of water quality information for decision making.

Objectives 1 and 2 are particularly focused on improving communication between communities of practice working in satellite EO of water quality and in-situ data providers: including buoy networks (GLEON), sensors (e.g. AERONET), global water quality data centres (GEMS/Water) and citizen scientists (GEO CitSci). Their ambition, therefore, is in close alignment with the goals of COINS and they provide the key network to develop the linkages.

In addition to this, there are some shorter-term, project-based initiatives that align with the objectives of COINS. This includes projects, such as the completed GloboLakes project which developed the LIMNADES database and more recently the EU-funded MONOCLE Project.

Currently, the most relevant is the EU-H2020 project: Water-ForCE (Water scenarios For Copernicus Exploitation). This project aims to develop a Roadmap for Copernicus Inland Water Services by:

- Addressing the disconnection between in-situ and remote sensing observations and the user community.
- Highlighting the needs and expectations of public and private sectors in delivering research and innovation opportunities
- Advising on a strategy to ensure effective uptake of water related services

Water-ForCE is led by University of Tartu, and is a consortium of 20 organisations from Europe, connecting experts in water quality and quantity. Their specific objectives are to:

- Analyse EU policies to identify where the Copernicus Services can improve monitoring programs and how Copernicus data can be more effectively used in developing and delivering future versions of EU directives.
- Specify the requirements for future Copernicus missions (e.g. optical configuration of Sentinel-2E and onward, hyperspectral sensors).

- Optimize future exploitation for inland water monitoring & research and, consequently, (a) enlarge the service portfolio and (b) improve the performance of current Services.

In order to achieve this, Water-ForCE developed a work concept that consist of four overarching WPs and four technical WPs. The first overarching WP1 analysed current and future policies, end-users need, innovation needs, need for supporting water related SDG's, etc. WP1 provides direction to the technical WP's (WP2-5) (see Figure 11)

Water-ForCE has held three workshops:

1) March 15, 2021: "On the use of remote sensing for monitoring and modelling the water cycle"
2) April 20, 2021: "Stakeholder Input on the Evolution of Copernicus Water Services"
3) May 17-18 and 20: "In situ calibration and validation of satellite products of water quality and hydrology"

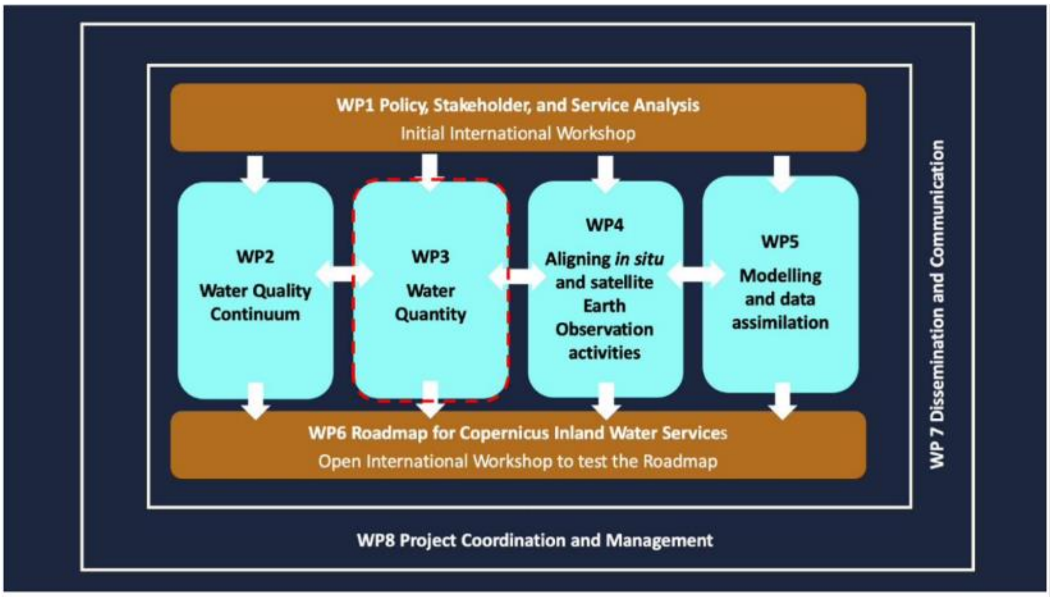


Figure 11. Organizational structures of the different work packages in the Water-ForCE project (Water-ForCE, 2021)

Only one document is currently public - concerning WP3 (water quantity) (Water-ForCE, 2021). The general objective of WP3 is to provide insights into products that are relevant to inland water services, thereby supporting integrated water resource management and improving coverage of EU policies regarding water quantity.

Our COINS consortium contributed to the third workshop in May, where we presented the findings from the first draft of this document and the COINS reports on water quantity.

GEO AquaWatch and Water-ForCE originated from the EO communities and are led by EO scientists and, therefore, we recommend COINS continues to engage with these key networks to provide understanding and expertise in in-situ data collection, including citizen monitoring, of water quality.

Key gaps

DATA COORDINATION

- Our review highlights that there is no on-going dedicated global data repository providing in-situ data for the EO community (i.e. like SeaDataNet). There is such a diversity of in-situ data that a single data repository is unlikely to be practical, but there is a clear need for a connected system of data bases that support identification of relevant validation data via one entry and a good selection / filtering functionality by different criteria. Here, the specification and harmonisation of metadata is very relevant and needs to be a requirement for submission. There are on-going initiatives that are currently pushing this agenda (e.g. Water-FORCE, GEO AquaWatch).
- Further discussions with data centres like LIMNADES, EOINET and GEMS/Water should be held to outline activities that could support their development for this purpose
- A lack of in-situ data: aquatic vegetation (floating weed and submerged cover, algal pigment data, particularly Phycocyanin and Phycoerythrin, primary productivity, nutrients, faecal coliforms) that could help evaluate water quality products and understand potential sources of error.

DATA & METHODS

- There is a need to produce standard operating procedures (SOPs) for sampling and analysis of water quality data that optimises their match with EO data protocols. This includes spatial and temporal aspects of sampling, parameters measured and estimates of uncertainty of in-situ water quality data
- Need to document algorithms used in the different types/brands of in-situ water quality sensors (e.g. chlorophyll-a fluorescence sensors) to evaluate their comparability
- An up-to-date review drone technologies for RS of water quality products to support satellite EO cal/val
- Gaps in-situ measurements at different lake optical water types, locations (especially tropical and sub-tropical regions) and seasons
- Limited availability of hyperspectral data available to stimulate research
- In terms of CS data, there are no existing CS schemes specifically elaborated to support in-situ data suitable for ca/val of EO water quality products

FUNDING

- Document potential funding opportunities to support in-situ collection campaigns, particularly in less developed countries
- Supporting innovation in sensors: underwater spectrophotometers, CDOM/fDOM sensors and cheaper sensors (hyperspectral still very expensive)
- Review sustainable funding models for long-term support of in-situ data centres and their interoperability

Key actions to take forward

COORDINATION

- Consider role of EEA (or ESA) as centralised body for Europe to store (or signpost to) high quality in-situ data or support initiatives that support projects to transfer data into sustainable long-term repositories.
- Work alongside GEO AquaWatch to support international capacity building programmes, such as online, accessible training courses for training people in in-situ data collection, data handling and data archiving to help stimulate data collection and sharing following FAIR principles.
- Support activities to enhance cooperation of in-situ sampling programmes with EO communities to deliver better match-up with satellite overpasses
- Create networks of users with sensors collecting IOP data (hyperSAS, PANTHYR, Iecplore, Wisp Sttaions)
- Develop tools and standards to help data sharing and building links between different data bases
- Engaging with specific ongoing citizen science schemes i.e. Secchi Dip-in (turbidity), Bloomin' Algae (cyanoHABs) and Eye on Water (water colour and cDOM) to consider how they could be developed or elaborated to provide in-situ data suitable for ca/val of EO water quality products This could be encouraged through developing tailored CS schemes specifically to support development of EO data products. Developing such tailored schemes would also advance the (cross-)validation of EO and citizen science data products
- Consider groups (e.g. fishing sector) or highly motivated citizens (anglers, watersports) to support targeted collection of in-situ data, particularly where water quality affects them (e.g. harmful algal blooms)
- Encourage CS schemes to collect in-situ water quality data in specific parts of the world (I.e. Asia, South America and Africa) where suitable data for cal/val are poorly represented

DATA & METHODS

- Differentiate data requirement for calibration (high quality, well-matched) and for validation (can be qualitative and less rigorously matched)
- Better metadata definitions
- Merge of vertical and horizontal data with ML approaches, data assimilation methods
- Intercalibration of in-situ data providers and sensors
- Much more research on uncertainties in in-situ data are needed related to spatial and temporal variability in sampling and analyst/method errors
- Develop guidance for wider use of drones for measurements of in-situ data across long transects
- Enhance autonomous fixed platforms with hyperspectral radiometers and increase coverage across different lake types and regions.
- Investigate potential innovations for CS recording DOC, chlorophyll-a and nutrients as these key parameters were found to never or rarely be collected using CS

FUNDING

- Identify opportunities, or provide support, to businesses developing innovations in instrumentation that could be used to expand in-situ water quality data collection e.g. spectroradiometers, smartphone sensors to support high quality citizen science data collection e.g. iSpex 2
- Support relevant in-situ projects in less developed countries in delivering discoverable QA/QC datasets

8. References

Ali, J. M., B. C. Noble, I. Nandi, A. S. Kolok & S. L. Bartelt-Hunt, 2019. Assessing the Accuracy of Citizen Scientist Reported Measurements for Agrichemical Contaminants. Environmental Science & Technology 53(10):5633-5640 doi:10.1021/acs.est.8b06707.

Belzile, C., W. F. Vincent, C. Howard-Williams, I. Hawes, M. R. James, M. Kumagai, and C. S. Roesler (2004), Relationships between spectral optical properties and optically active substances in a clear oligotrophic lake, Water Resour. Res., 40, W12512, doi:10.1029/2004WR003090.

Burggraaff, O., Perduijn, A.B., van Hek, R.F., Schmidt, N., Keller, C.U. and Snik, F., 2020, May. A universal smartphone add-on for portable spectroscopy and polarimetry: iSPEX 2. In *Micro-and Nanotechnology Sensors, Systems, and Applications XII* (Vol. 11389, p. 113892K). International Society for Optics and Photonics.

Carvalho L., Poikane S., Lyche Solheim A., Phillips G., Borics G., Catalan J., De Hoyos C., Drakare S., Dudley B., Jarvinen M., Laplace-Treyture C., Maileht K., McDonald C., Mischke U., Moe J., Morabito G., Nõges P., Nõges T., Ott I., Pasztaleniec, A., Skjelbred B. & Thackeray S. 2013. Strength and uncertainty of lake phytoplankton metrics for assessing eutrophication impacts in lakes. Hydrobiologia, 704, 127-140.

Carvalho, L., 2021. Bloomin' Algae. In: UK Centre for Ecology & Hydrology. https://www.ceh.ac.uk/our-science/projects/bloomin-algae Accessed 15/09/2021 2021.

Castilla, E. P., D. G. F. Cunha, F. W. F. Lee, S. Loiselle, K. C. Ho & C. Hall, 2015. Quantification of phytoplankton bloom dynamics by citizen scientists in urban and peri-urban environments. Environmental Monitoring and Assessment 187(11) doi:10.1007/s10661-015-4912-9.

Cunha, D. G. F., S. P. Casali, P. B. de Falco, I. Thornhill & S. A. Loiselle, 2017. The contribution of volunteer-based monitoring data to the assessment of harmful phytoplankton blooms in Brazilian urban streams. Science of the Total Environment 584:586-594 doi:10.1016/j.scitotenv.2017.01.080.

Deutsch, E. S., J. A. Cardille, T. Koll-Egyed & M. J. Fortin, 2021. Landsat 8 Lake Water Clarity Empirical Algorithms: Large-Scale Calibration and Validation Using Government and Citizen Science Data from across Canada. Remote Sensing 13(7) doi:10.3390/rs13071257.

Friedrichs, A., J. A. Busch, H. J. van der Woerd & O. Zielinski, 2017a. SmartFluo: A Method and Affordable Adapter to Measure Chlorophyll a Fluorescence with Smartphones. Sensors 17(4) doi:10.3390/s17040678.

Friedrichs, A., O. Ferdinand, M. L. M. Montenegro & O. Zielinski, 2017b. SmartFluo goes FDOM: Advancement of the DIY fluorometer approach towards UV excitation Oceans 2017 - Aberdeen. Oceans-Ieee.

Fry, M., Fernandez V. and Põldnurk J., 2019. Hydrology in-situ data requirements and availability. Report to European Environment Agency on Contract No. EEA/IDM/15/026/LOT1, December 2019. Accessed on 22nd April 2021 at https://insitu.copernicus.eu/library/reports/copy2_of_Hydrologyinsitudataavailabilityreport1.0.pdf

Garaba, S. P., A. Friedrichs, D. Voss & O. Zielinski, 2015. Classifying Natural Waters with the Forel-Ule Colour Index System: Results, Applications, Correlations and Crowdsourcing. International Journal of Environmental Research and Public Health 12(12):16096-16109 doi:10.3390/ijerph121215044.

George, G., N. N. Menon, A. Abdulaziz, R. J. W. Brewin, P. Pranav, A. Gopalakrishnan, K. G. Mini, S. Kuriakose, S. Sathyendranath & T. Platt, 2021. Citizen Scientists Contribute to Real-Time Monitoring of Lake Water Quality Using 3D Printed Mini Secchi Disks. Frontiers in Water 3 doi:10.3389/frwa.2021.662142.

Giardino, C., Bresciani, M., Stroppiana, D., Oggioni, A., Morabito, G. 2014. Optical remote sensing of lakes: an overview on Lake Maggiore, J. Limnol., 73(s1):201-214, DOI:14.4081/jlimnol.2014.817

Hadj-Hammou, J., S. Loiselle, D. Ophof & I. Thornhill, 2017. Getting the full picture: Assessing the complementarity of citizen science and agency monitoring data. PLoS One 12(12):e0188507.

Heard, J., Simis, S.G.H., Ceccaroni, L., Clymans, W. 2018. Water Quality Survey of the Multiscale Observation Networks for Optical monitoring of Coastal waters, Lakes and Estuaries (MONOCLE) project. For original content see www.monocle-h2020.eu. doi: 10.5281/zenodo.1625594

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: 2705-2718.

Kirschke, S., Avellán, T., Bärlund, I., Bogardi J., Carvalho L., Chapman D., Dickens C.W.S., Irvine K., Lee S., Mehner T. and Warner S., 2020. Capacity challenges in water quality monitoring: understanding the role of human development. Environmental Monitoring and Assessment, 192, 298. Published 19th April 2020. https://doi.org/10.1007/s10661-020-8224-3.

Loiselle, S. A., P. C. Frost, E. Turak & I. Thornhill, 2017. Citizen scientists supporting environmental research priorities. Science of the Total Environment 598:937.

Malthus, T. J., R. Ohmsen & H. J. van der Woerd, 2020. An Evaluation of Citizen Science Smartphone Apps for Inland Water Quality Assessment. Remote Sensing 12(10) doi:10.3390/rs12101578.

McGoff, E., F. Dunn, L. M. Cachazo, P. Williams, J. Biggs, P. Nicolet & N. C. Ewald, 2017. Finding clean water habitats in urban landscapes: professional researcher vs citizen science approaches. Science of the Total Environment 581:105-116 doi:10.1016/j.scitotenv.2016.11.215.

Mitroi, V., K. C. Ahi, P. Y. Bulot, F. Tra, J. F. Deroubaix, M. K. Ahoutou, C. Quiblier, M. Kone, J. C. Kalpy & J. F. Humbert, 2020. Can participatory approaches strengthen the monitoring of cyanobacterial blooms in developing countries? Results from a pilot study conducted in the Lagoon Aghien (Ivory Coast). Plos One 15(9) doi:10.1371/journal.pone.0238832.

Poisson, A. C., I. M. McCullough, K. S. Cheruvelil, K. C. Elliott, J. A. Latimore & P. A. Soranno, 2020. Quantifying the contribution of citizen science to broad-scale ecological databases. Frontiers in Ecology and the Environment 18(1):19-26 doi:10.1002/fee.2128.

Quinlivan, L., D. V. Chapman & T. Sullivan, 2020a. Applying citizen science to monitor for the Sustainable Development Goal Indicator 6.3. 2: a review. Environmental monitoring and assessment 192(4):1-11.

Quinlivan, L., D. V. Chapman & T. Sullivan, 2020b. Validating citizen science monitoring of ambient water quality for the United Nations sustainable development goals. Science of the Total Environment 699:134255.

Rose, N. L., S. D. Turner, B. Goldsmith, L. Gosling & T. A. Davidson, 2016. Quality control in public participation assessments of water quality: the OPAL Water Survey. Bmc Ecology 16 doi:10.1186/s12898-016-0063-2.

Scott, A. B. & P. C. Frost, 2017. Monitoring water quality in Toronto's urban stormwater ponds: Assessing participation rates and data quality of water sampling by citizen scientists in the FreshWater Watch. Science of the Total Environment 592:738-744 doi:10.1016/j.scitotenv.2017.01.201.

Scott, M. D., L. Ramaswamy & V. Lawson, 2016. CyanoTRACKER: A Citizen Science Project for Reporting Harmful Algal Blooms 2016 Ieee 2nd International Conference on Collaboration and Internet Computing. 391-397.

Shupe, S. M., 2017. High resolution stream water quality assessment in the Vancouver, British Columbia region: a citizen science study. Science of the Total Environment 603:745-759 doi:10.1016/j.scitotenv.2017.02.195.

Stelzer, K., Simis, S., Selmes, N., Müller, D. 2020. Copernicus Global Land Operations "Cryosphere and Water" (CGLOPS-2) Product User Manual I1.02, pp.26

Thackeray, S.J., P. Nõges, M.J. Dunbar, B.J. Dudley, B. Skjelbred, G. Morabito, L. Carvalho, G. Phillips, U. Mischke, J. Catalan, C. de Hoyos, C. Laplace, M. Austoni, B.M. Padedda, K. Maileht, A. Pasztaleniec, M. Järvinen, A. Lyche Solheim & R.T. Clarke. 2013. Quantifying uncertainties in biologically-based water quality assessment: a pan-European analysis of lake phytoplankton community metrics. Ecological Indicators, 29, 34–47. 10.1016/j.ecolind.2012.12.010

Thornhill, I., A. Chautard & S. Loiselle, 2018. Monitoring Biological and Chemical Trends in Temperate Still Waters Using Citizen Science. Water 10(7) doi:10.3390/w10070839.

Water-ForCE (2021), Review document on Copernicus products related to the hydilogical water balance, D3.2, pp.43

Zhang, Y. C., R. H. Ma, M. Q. Hu, J. H. Luo, J. Li & Q. C. Liang, 2017. Combining citizen science and land use data to identify drivers of eutrophication in the Huangpu River system. Science of the Total Environment 584:651-664 doi:10.1016/j.scitotenv.2017.01.093.

Zibordi, G., Holben, B. N., Talone, M., D'Alimonte, D., Slutsker, I., Giles, D. M., & Sorokin, M. G. (2020). Advances in the Ocean Color component of the Aerosol Robotic Network (AERONET-OC), Journal of Atmospheric and Oceanic Technology, https://journals.ametsoc.org/view/journals/atot/aop/JTECH-D-20-0085.1/JTECH-D-20-0085.1.xml