

Framework Service Contract EEA/IDM/15/026/LOT 1 for Services supporting the European Environment Agency's (EEA) implementation of cross-cutting activities for coordination of the in-situ component of the Copernicus Programme Services

---

# HYDROLOGICAL IN SITU DATA REQUIREMENTS AND AVAILABILITY

---

Issue: 1.0  
Date: 05/12/2019

## [DOCUMENT RELEASE]

	Name(s)	Affiliation
Coordinated by:	Matt Fry	UK CEH
Contributions:	Matt Fry Vicente Fernandez Jana Pöldnurk	UK CEH EuroGOOS Estonian Environment Agency
Approval:	Henrik Steen Andersen	European Environment Agency

Prepared for:	European Environment Agency (EEA)
Represented by: (Project Manager)	Henrik Steen Andersen
Contract No.	EEA/IDM/15/026/LOT1

## [Change Record]

Version	Date	Changes
0.1	18/06/2019	Initial draft release
0.2	16/11/2019	Second draft release
1.0	05/12/2019	Updated release after initial comments by Copernicus Services

## Table of Contents

Executive Summary .....	5
1. Introduction .....	7
2. Approach .....	8
3. River flows and levels .....	9
3.1. Requirements for hydrological in situ data .....	9
3.2. Current use of in situ data .....	12
3.3. Other potential sources of in situ data .....	18
3.4. Gaps .....	20
3.5. Avenues for future coordination work .....	21
4. Lake and river water level .....	22
4.1. Requirements for in situ data .....	22
4.2. Current use of in situ data .....	23
4.3. Other potential sources of in situ data .....	24
4.4. Gaps .....	24
4.5. Avenues for future coordination work .....	25
5. River water quality .....	25
5.1. Requirements for in situ data .....	25
5.2. Current use of in situ data .....	26
5.3. Other potential sources of in situ data .....	26
5.4. Gaps .....	27
5.5. Avenues for future coordination work .....	28
6. Lake water quality .....	28
6.1. Requirements for in situ data .....	28
6.2. Current use of in situ data .....	30
6.3. Other potential sources of in situ data .....	30
6.4. Gaps .....	31
6.5. Avenues for future coordination work .....	32
7. Lake Surface Water Temperature .....	32
7.1. Requirements for in situ data .....	32
7.2. Current use of in situ data .....	33
7.3. Other potential sources of in situ data .....	33
7.4. Gaps .....	34
7.5. Avenues for future coordination work .....	34
8. Soil moisture .....	35

## HYDROLOGY IN SITU DATA REQUIREMENTS AND AVAILABILITY

8.1.	Requirements for in situ data .....	35
8.2.	Current use of in situ data .....	37
8.3.	Other potential sources of in situ data .....	39
8.4.	Gaps.....	39
8.5.	Avenues for future coordination work.....	40
9.	Lake extents .....	40
10.	Summary and conclusions.....	41
	References.....	45
	Appendix A. CMEMS MFC requirements .....	48

## EXECUTIVE SUMMARY

The requirements for in situ data for validation of satellite-based data products relating to water / hydrology currently produced across the Copernicus services were identified, largely in discussion with data developers. The availability and limitations of the available in situ data are summarised in this report, with a view to identifying potential avenues for improving access to such data in future. Results are summarised below across the different data types addressed. Across all of these data types there are recommended actions on two common themes.

For short to medium term improvement there are a range of activities to help coordinate data sharing between the Copernicus services and to match current requirements for in situ data with data sources currently available. To improve the situation in the long-term, there is a need to define in detail a vision for a functioning system of in situ hydrological monitoring that will support future satellite product delivery, and to promote this via global and regional organisations coordinating monitoring and data collation activities.

### River flows:

- Measurements are available globally, though real-time access is often not possible
- The CEMS EFAS Hydrological Data Collection Centre (HDCC) is very effective at collating river levels and flows for Europe, and is a potential source of river flow data for other services (e.g. in development of new coastal products)
- Adaptation of the EUMETNET data licence for other organisations would help EFAS with consistency of licensing across data providers
- Encouraging the implementation of technical measures within EFAS HDCC that would enable data to be shared across services, and potentially, for data available under an open licence, more widely.
- Promotion of web-based data sharing approaches is needed to improve access to data from countries outside Europe, and the WMO WHOS initiative offers an existing pathway for this
- Support for National Hydrological Services in developing countries is needed to enable better data sharing, including for real time applications.
- A review of data available from National Hydrological Services would help provide an understanding of the availability of river flows (and levels) globally and would help to focus future activities and support the WHOS initiative.

### Water level (lakes and rivers)

- River and lake level in situ measurements exist, but significant effort is required to collate data from sources currently.
- Coordination efforts may be able to identify and improve access to existing sources of river and lake levels (e.g. EFAS HDCC and others).
- The requirements identified should be used to help shape a longer term vision for global databases that better support satellite product development and delivery.

### River water quality:

- The requirements for access to in situ river water quality are developing, in particular for proposed coastal zone modelling, and coordination may be beneficial at this moment.
- Summary information on European and global databases, including contents and operational model, should be produced to aid understanding of whether data will meet these requirements.
- Requirements for river water quality data for improved satellite products should be developed and promoted

**Lake water quality:**

- The Globalakes / LIMNADES database, used within CLMS-GL, is a comprehensive database of existing in situ data for validation of satellite derived bio-optical measures of lake water quality. It needs to be sustained and enhanced to support further product development.
- More bio-optical measurements are needed as the products develop to include more lakes, and UAVs have potential to deliver these.
- Development and promotion of the requirements and protocols for more suitable monitoring offers potential to improve the availability of appropriate in situ data in future, including from statutory monitoring.

**Lake surface water temperature:**

- Lake surface water temperature measurements are relatively simple and should be widespread and accessible
- Current global databases do not offer suitable data for validation of products
- Better coordination of lake surface water temperature measurements is needed to meet the needs of satellite validation.

**Soil moisture:**

- Available in situ data is too sparse for much more than minimal validation
- The International Soil Moisture Network is an effective resource for access to the soil moisture data that is collected globally
- Development and promotion of requirements and protocols for more suitable in situ monitoring may help to ensure that future soil moisture monitoring produces more usable data.

## 1. INTRODUCTION

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 (CIS<sup>2</sup>). 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 this exercise was initially identified to include river flows, river water quality, lake extents and depths, lake water quality, and soil moisture. A watching brief was maintained across related marine, ice and snow activities within the in situ component. Groundwater data was not considered as there are currently no Copernicus service products directly focussed on groundwater, though this may need consideration in future. Precipitation is dealt with elsewhere within the Copernicus In Situ component, along with other meteorological variables.

There are a number of Copernicus data products, both current and planned, which need, or would benefit from, use of hydrological in situ data for validation / ground-truthing. The importance of in situ data varies significantly between services, depending on the focus of activities.

The requirements for in situ data were identified through discussions with key representatives of each service, or of organisations undertaking the development of data products under contract.

Currently available sources of in situ data were identified, and some summary information is provided here about the content of these data sources, but no detailed matching of the Copernicus service requirements for in situ data with the specific availability of data types at locations was undertaken, though in some cases such an exercise is suggested as an appropriate next step.

This report summarises these requirements and the in situ data currently available to meet them, and proposes some potential activities that could help to fill gaps in data availability in future.

The services focussed on in this report are:

- The Copernicus Emergency Management Service (CEMS)
- The Copernicus Climate Change Service (C3S)
- The Copernicus Land Management Service (CLMS) which itself has three components:
  - CLMS Global Land Component (CLMS-GL)
  - CLMS pan-European Component (CLMS-PE)
  - CLMS Local Component (CLMS-L)
- The Copernicus Marine Environment Monitoring Service (CMEMS)

The Copernicus Atmospheric Monitoring and Security Services were not believed to require the use of in situ hydrological data, and are not discussed in this report.

## 2. APPROACH

This reports aims to summarise the requirements for in situ hydrological data across the Copernicus services, to identify gaps in the data currently available, and consider potential coordination activities that may help to improve access to suitable in situ data.

Information has been gathered largely through discussion with those leading product developments, and through reviewing additional available data sources.

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 hydrological data across the services. The information presented for each of these types of data is broken down under the following headings:

### **Requirements for in situ data**

The existing and upcoming products are discussed and general requirements for validation are identified. The approach to validation is often determined by the currently available in situ data, though the broader requirements beyond this are addressed where possible.

### **Current use of in situ data**

All of the Copernicus data products discussed in this report have undergone some level of validation using in situ data. In all cases some effort, often considerable, has been focussed on acquiring suitable in situ data for this purpose. The groups working on these data products have, in general, good knowledge of the availability of data to meet their requirements. Some opportunities exist to immediately improve this availability through sharing of knowledge of in situ data, but in many cases the datasets currently used represent the best readily available data for the purpose. For this reason we focus initially on the availability and limitations of the data currently used, and discuss other datasets that may potentially yield useful additional in situ data.

### **Other potential sources of in situ data**

Hydrological in situ measurements originate from numerous organisations, often operating at national scales but also at sub-national region and basin / catchment scales. Measurements can be made for a range of purposes but are generally for statutory and regulatory environmental monitoring or for research. Citizen science initiatives can also provide useful data in some circumstances. Many important hydrological measurements are made by organisations within the private sector (e.g. reservoir levels and outflows), though access can often be restricted. There is still little in the way of consistent and standardised mechanisms for sharing even information on the existence of this data directly from the source, less still standardised sharing of the data itself.

Aggregations of these datasets are made to create databases of in situ measurements, again for purposes ranging from regulatory / statutory monitoring, to research, and sometimes specifically to underpin the development of particular data products. Usually these aggregated datasets are super-sets of smaller databases, comprising copies of data, sometimes with added data validation, and are focussed at national, regional or global scales.

The Global Terrestrial Network – Hydrology aims to provide a linking mechanism to global datasets relating to hydrology, including the WMO Global Runoff Data Centre, the FAO Aquastat system, and the UNEP GEMS/Water Programme. These datasets can provide a lot of the available in situ measurements, but in many cases they were set up before the satellite era, and before the era of

real-time access via the web, and are not specifically designed to address the needs of satellite validation, which can be seen to limit their utility in some cases.

### Gaps

The descriptions of datasets used and otherwise available include some of the limitations for validation and production of Copernicus data products. For each type of data we also consider a broader set of factors relating to in situ data that have the potential to limit the ability to undertake this validation, now or in future:

- Observations do not exist
- Observations exist but data is not freely available
- Observations exist but have not been collated into a usable database
- Data collections exist but data quality is limited, or insufficient for satellite data validation
- Technology gaps, both in measurements or in data access
- Sustainability gaps, i.e. risks to the future availability and continuation of existing datasets.

### Avenues for future coordination work

We suggest some potential areas where coordination work may help to improve access to in situ data, addressing some of the gaps.

## 3. RIVER FLOWS AND LEVELS

In situ measurements of river flows (discharge) and levels are considered together here as they are often measured by the same instruments or at the same locations (gauging stations). However it is not always the case that the resulting data is made available through the same networks or data centres. There are also, in many countries, considerably more water level monitoring stations than flow due to the cost of maintaining the latter and utility of the former in flood warning. In situ river level data for altimetry products is also discussed separately in section 4 below.

### 3.1. Requirements for hydrological in situ data

**CEMS** is the main user of river flow data of all the services. It operates the **European Flood Awareness System (EFAS)**. EFAS “produces European overviews of ongoing and forecasted floods up to 10 days in advance and contributes to better protection of the European citizens, the environment, properties and cultural heritage”. It has been developed at the Joint Research Centre (JRC), in close collaboration with national hydrological and meteorological services and policy DG's of the European Commission. EFAS provides direct support to the EU's Emergency Response Coordination Centre (ERCC) of DG ECHO and the hydrological services in the Member States. A Spanish consortium (REDIAM and SOOLOGIC) has been awarded the contract for the EFAS Hydrological Data Collection Centre and is responsible for collecting discharge and water level data across Europe, and has been doing so since the service became operational in 2012.

EFAS operates the LISFLOOD hydrological model, taking meteorological inputs (managed via another data collection centre within a separate consortium) and producing modelled estimates of runoff, river flow, and river level. This model is applied over a spatial domain that covers greater Europe and surrounding countries, including river basins across Turkey and North Africa (but not extending to the entire Nile basin) [Arnal et al, 2019].

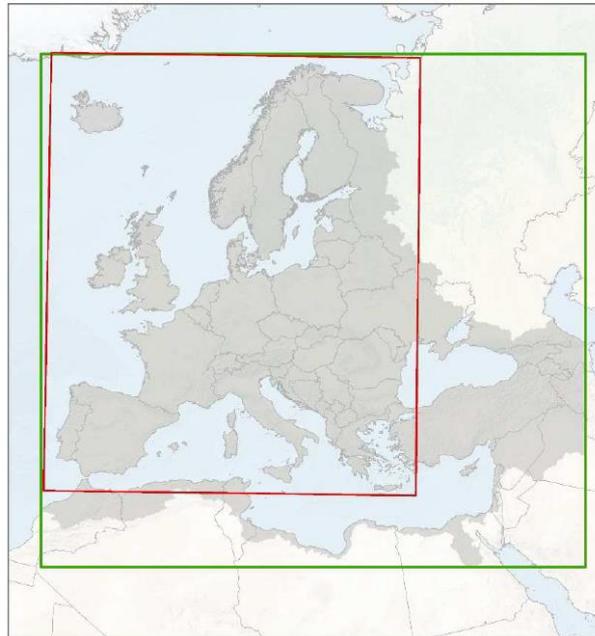


Figure 1. Current domain limits (green box) for the EFAS model. The shaded grey areas show the hydrological catchments modelled. [ Arnal et al, 2019].

River flows and levels are required over historic periods and in real-time for different purposes. Historic data is required for model development, model calibration, and understanding of the historic skill of the model. Real-time river flows are needed for bias correction of forecasts (for some sites), real-time assessment of modelled flows, including event-based verification which identifies if predicted flood events have occurred or if flood events have occurred that were not predicted (the results of which are included in regular reporting such as EFAS Bulletins). Sub-daily (e.g. 15-minute) data are required for accurate estimation of flood flows and levels on all but the largest rivers. Daily flows can be useful for model validation and calibration in the absence of sub-daily data. EFAS does not currently undertake data assimilation using river flow data, a process by which the various model states are updated in real-time based on observations.

Since April 2018 CEMS has also been operating the **Global Flood Awareness System (GloFAS)**, “a global hydrological forecast and monitoring system independent of administrative and political boundaries” ([www.globalfloods.eu/](http://www.globalfloods.eu/)). GloFAS also uses the LISflood model, though configured differently to that used by EFAS, and only used to simulate the groundwater and river routing processes (surface runoff is generated by the HTESSEL model). The development of GloFAS creates a requirement for global river flow and river level data. The GloFAS model has been calibrated historically, but it is understood that it does not use real-time data at present, though it does state on its website that in situ data are important for calibration, validation and post-processing.

**C3S** do not currently make direct use of in situ hydrological measurements, but they are important to users of C3S services. Should the Climate Data Store evolve to include more in situ measurement data, river flows would likely be an important component, representing a terrestrial ECV with a wide range of uses in relation to climate data analysis and applications, though this could require a more open licensing model.

C3S produces information about the past, present and future climate of Europe and the world, in particular medium-term (seasonal) weather forecasts and long-term climate projections for the coming century and beyond. The outputs are taken up by end users directly (i.e. through

## HYDROLOGY IN SITU DATA REQUIREMENTS AND AVAILABILITY

applications that are developed independently of C3S), and C3S also commissions the development of Sectoral Information Systems to deliver to key sectors. Examples include snow forecasting for the tourism industry, and forecasts and long-term climate-change projections of wind for the energy and insurance industries. The water sector is an important end user, and C3S has developed two prototype (SWICCA, <http://swicca.eu/>; EDgE, <http://edge.climate.copernicus.eu/>) and one operational Sectoral Information System in this area.

C3S sees the requirement for in situ hydrological data to be driven by its potentially numerous and diverse end users, and therefore hard to specify in detail. However, the Sectoral Information Systems provide insight into this requirement. As with EFAS, these systems are based on hydrological modelling, and the developers of the prototype services describe a need for extensive in situ river flows and water levels at daily or sub-daily resolutions (depending on the applications) for similar purposes. Historical data is needed for calibration and validation, and real-time or near real-time data for initialisation of states prior to forecasting, and verification and skill assessment. Spatial coverage will depend on the application, but basin level analysis of modelled results against in situ data is often desirable. Many C3S outputs, and therefore their potential users, are global as well as European. Users of such services often request data at very high resolutions (e.g. catchments of <10km<sup>2</sup>) effectively requiring the use of any in situ data available from national monitoring schemes, therefore defining a potential requirement for more detailed data than is currently required by EFAS.

**CMEMS** operates 8 separate regional Marine Forecasting Centres (MFC). The MFCs have a requirement for information on river discharges to the marine environment, as a land forcing input to ocean circulation models, often relating to ocean forecasting. Currently access to in situ data is variable across the different MFCs, with some making use of modelled discharge (e.g. E-Hype model data from SMHI, Sweden) or simple seasonal climatologies, alongside limited directly sourced in situ river discharge data. Even where discharge data is not directly used as model input there is a need for access to in situ data for validation and understanding of modelled river inputs, though this is a requirement of the model developers rather than of the service itself. MFCs provided detailed information about requirements in relation to the spatial coverage for discharge data (annex A). River discharges at coastal outlets are required, and so measurements at the furthest downstream locations are preferable. Both historic data (reference periods started between 1980 and 1993) and near-real time data are required. A daily temporal resolution for flow / discharge is sufficient.

**CLMS** does not currently make use of in situ river flow data, but the CLMS-GL Operational Service has a global water level product which covers lakes and rivers, and makes use of in situ river level measurements for validation. This combined river and lake product and its requirements for in situ water quality data are discussed below in section 5.

**CLMS local/pan-European** do not currently make use of in situ river flow or level data, but have a requirement in relation to the future development, together with CMEMS, of planned products for coastal zones. Rivers provide important inputs to these zones in terms of water quality (e.g. nutrients) and water quantity (e.g. flood volumes). Coastal products may require modelled inputs of river flows in some areas but in situ measurements are likely to be important in their calibration, validation and development. Flows are also required to convert measurements of water quality (generally taking the form of observed concentrations) into total pollutant / nutrient loads. The availability of in situ measurements at or near locations of in situ water quality measurements is therefore important. The requirements for in situ data reflect those required by CMEMS MFCs for

ocean modelling, though there may be a heightened requirement for in situ river flows as the impact in coastal zones (both in dynamics and in water quality) will be far more significant.

Currently the **CMEMS service evolution project LAMBDA** (Land-Marine Boundary Development and Analysis, <http://www.cmems-lambda.eu/>) is collating ocean inflows from European rivers to underpin research to improve CMEMS MFCs' thermohaline circulation modelling, coupling watershed modelling and regional ocean models. They have identified additional river flow sites with available data, in particular across the Iberian peninsular and Atlantic coasts, and systems to harvest and provide data in near real time.

## 3.2. Current use of in situ data

### EFAS Hydrological Data Collection Centre

One of the components of EFAS is the Hydrological Data Collection Centre (HDCC) which is operated by a consortium based in Spain comprising the Agencia de Medio Ambiente y Agua de Andalucía in collaboration with Soologic SL (a technology company). HDCC acquires and manages river level and flow data for EFAS, providing both historic and real-time data. At the present time data from the HDCC are not accessed by other Copernicus services.

Data are acquired from national level hydrological services or hydro-meteorological services in most cases, though for some countries (e.g. UK, Germany, Spain) data is provided by regional bodies. It is largely the case that in western and northern Europe these agencies are specifically hydrological or environmental agencies, and in Eastern Europe they tend to be combined national meteorological and hydrological agencies (i.e. EUMETNET members).

The HDCC access data from providers over the web using a range of technologies. There are very few standards-based services (e.g. using the WMO-endorsed OGC WaterML2 standard for data transfer), but the variety of technologies does not create a barrier to data access. Data are acquired at the highest temporal resolution available (i.e. instantaneous measurements, or 5 / 15 / 60 minute averages) and aggregated by HDCC to common intervals (hourly, 6-hourly, daily). Historic data tend to be daily but there are ongoing efforts to source hourly or 6 hourly data directly from data providers. Near real-time data are more likely to be hourly, but can also be at 6-hourly and daily intervals. Input data are processed providing data quality control as well as temporal averaging. A series of processed data products at 1 hour, 6 hour and 24 hour resolution is then available to EFAS for use within its modelling and validation processes via consistent and standards-based services (REST, WaterML, with WMS' showing maps of operational stations and stations exceeding thresholds). This is a complex process, both in terms of data management, and the technical software and server infrastructure required to support it. HDCC has a suite of software visualisation tools for understanding both the data availability and data quality, and the status of the levels and flows in the rivers providing data.

The availability of data is regularly reviewed and the status reported. In 2018 HDCC received real-time data from 35 data providers for 1301 stations across the greater European area, comprising 289 stations delivering flow only, 273 providing level only, and 739 providing level and flow (CEMS, 2019). Data completeness is an important factor in the utility of records, and 958 stations met the HDCC target of being fully operational throughout the year and with at least 75% data completeness over the year. Of these 809 stations met the completeness criteria over a historic period of 1991 to 2016. Many stations provide far higher levels of data completeness (98% of gaps are shorter than 1 day, 74% less than 1 hour). A series of processes are in place to identify, classify and process gaps

in the data series, where appropriate infilling through interpolation and updating retrospectively with data from the data providers.

The spatial distribution of stations is shown in Figure 1. Whilst there is very good density of the data in most areas, many countries and regions can be seen to be missing, notably France, Poland, regions of Germany, and most regions of Italy. In some cases it is believed that data is now available for some of these countries. HDCC undertakes ongoing efforts to source additional data to fill gaps to meet EFAS requirements. The reasons for limited data availability in some areas are varied. In many cases it is the lack of an existing technical solution within the data providing agency, i.e. they are not currently sharing data over the web with other agencies in real time using systems that can be opened up for HDCC access. In others there is a lack of resources to make any changes to existing systems that are required for EFAS access, and it is often not the case that provision of data to EFAS is able to be prioritised by staff within data providing agencies, however willing. In some cases there is no agreement to share data with EFAS.

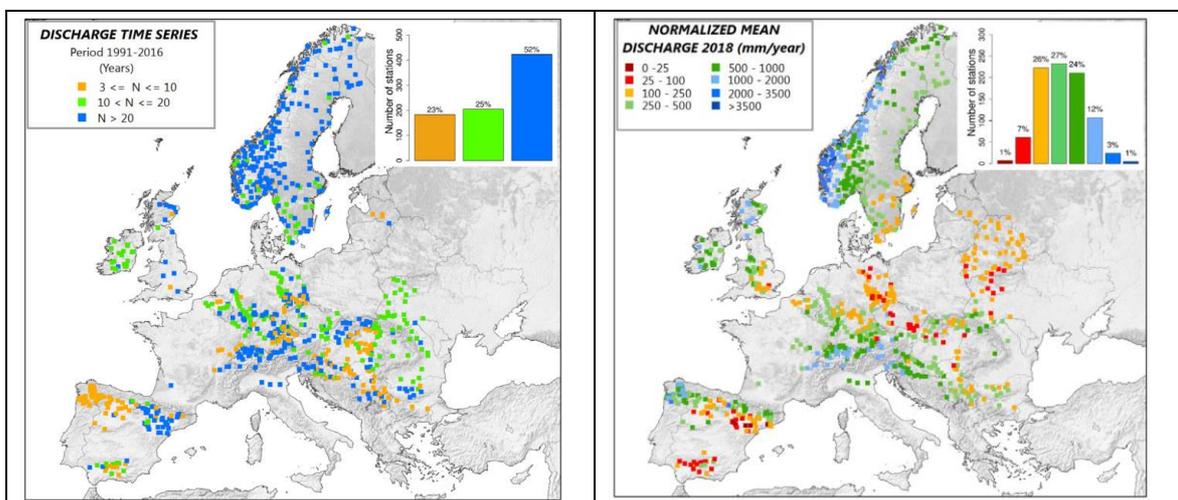


Figure 2 HDCC stations for a) historic period, symbolised by length of record, b) real-time data, symbolised by 2018 mean flow [from CEMS, 2019].

Whilst the maps of the data may show significant gaps in the coverage, it should be noted that most major and transboundary rivers have stations that are providing data. In addition the most recent EFAS report on model upgrades shows greater coverage of stations used for historical calibration and validation (Figure 2) which may mean that additional data is accessed from other sources (e.g. the GRDC, described below) for this purpose though, as noted below, this may be at daily resolution and therefore less ideal than sub-daily data for flood model assessment.

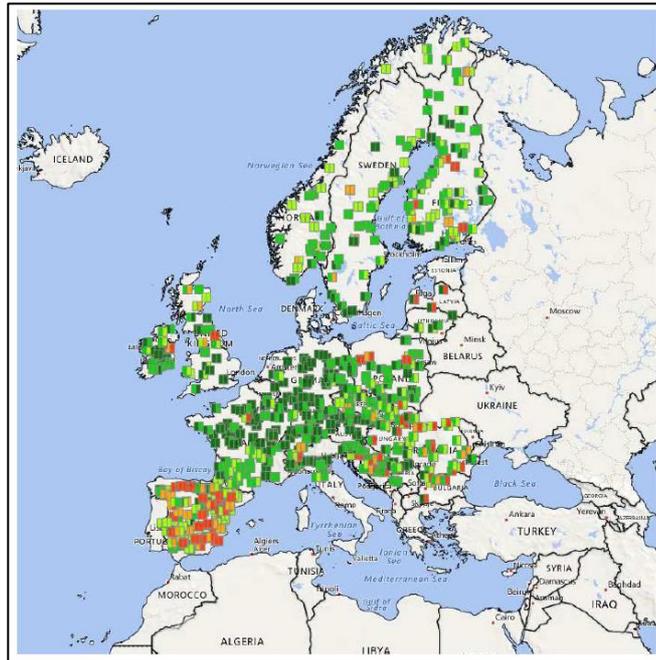


Figure 3 Map of stations used for EFAS validation [from Arnal et al, 2019]

Overall it can be seen that HDCC provides a very comprehensive and robust service for real-time and historical river level and flow data for Europe. Whilst it is not an open dataset (and this is precluded by the license conditions by which the HDCC acquires the data), there may be potential for data to be shared across services to reduce unnecessary effort and improve data quality and provenance.

The HDCC does not currently acquire data for countries outside Europe, e.g. to support the work of GloFAS. The calibration of GloFAS is described in detail by Hirpa et al (2018). Data were largely acquired from the Global Runoff Data Centre (GRDC).

### Global Runoff Data Centre

The principal source of accessible global scale data on river flows is the WMO mandated Global Runoff Data Centre operated by the German Federal Institute of Hydrology. The GRDC maintains a database of historic daily mean river flows, and some associated metadata, covering 6000 stations in over 120 countries. Data are provided by countries (generally the National Hydrological Services) under a standard licence, enabling GRDC to make data available on request for non-commercial usage. Data provision is currently via manual transfer of files, with no web service option, limiting its operational utility for some purposes.

The geographical representation of countries globally is good, with only a few notable exceptions (e.g. India, Egypt, Sudan, Namibia). Figure 4 shows the number of years of data per unit area for each country. This number is a conflation of the density and length of record, and can be misleading, particularly in the largest and smallest countries, as it is not directly representative of the ability of the GRDC to reflect the status of water resources, but it provides a useful measure of the overall data availability by country.

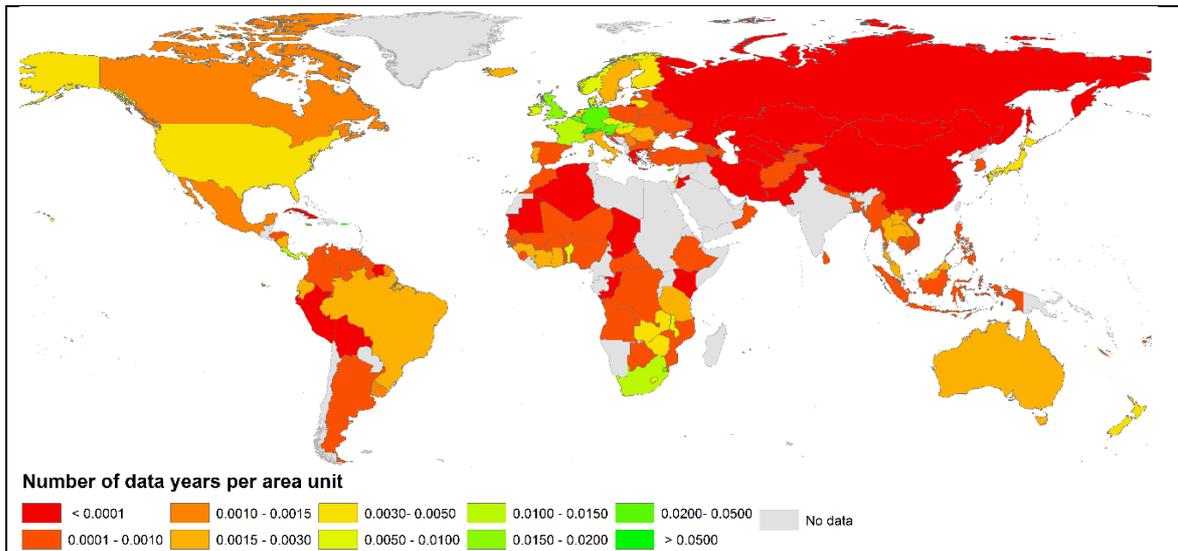
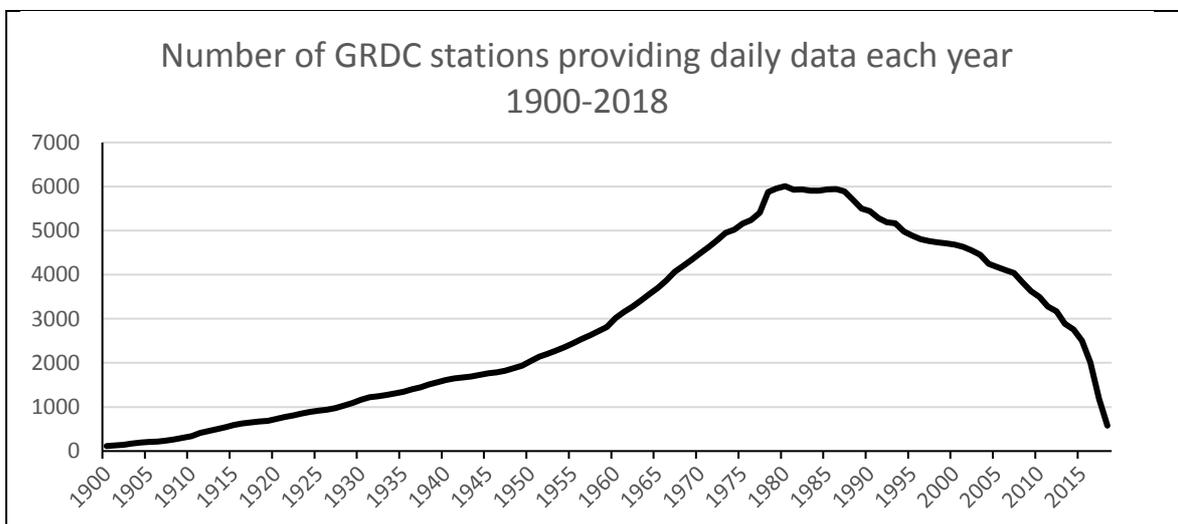


Figure 4 Map of geographical representation of countries within the Global Runoff Data Centre as expressed by the number of years of data per unit area (as of December 2018)

One of the key constraints of the GRDC is the process by which data is updated. Data are actively updated by GRDC through manual requests made to national data providers but this is dependent on the willingness and ability of the data provider to make new data available, and can lead to a long delay before data is accessible. There is no real time data available. Figure 5 shows graphs of the number of stations providing data for each year (since 1900) and the number of countries providing data for the same period. Inclusion of historical data in the earlier period of the record is generally quite comprehensive based on start dates of river flow time series (prior to the 1950s/1960s river flow data measurements were very sparse). However the drop-off in measurements from the mid-1980s is not solely indicative of the decline in hydrometric monitoring in many countries around this period, but also seems to reflect a decrease in the number of countries represented within the GRDC database. This may limit the ability of the GRDC to meet the global requirements for hydrological modelling, particularly as some of the key globally available meteorological datasets, used as inputs to hydrological models, such as the C3S produced ERA5 dataset, start in 1979, meaning there is a limited period for calibration of hydrological models using in situ data.



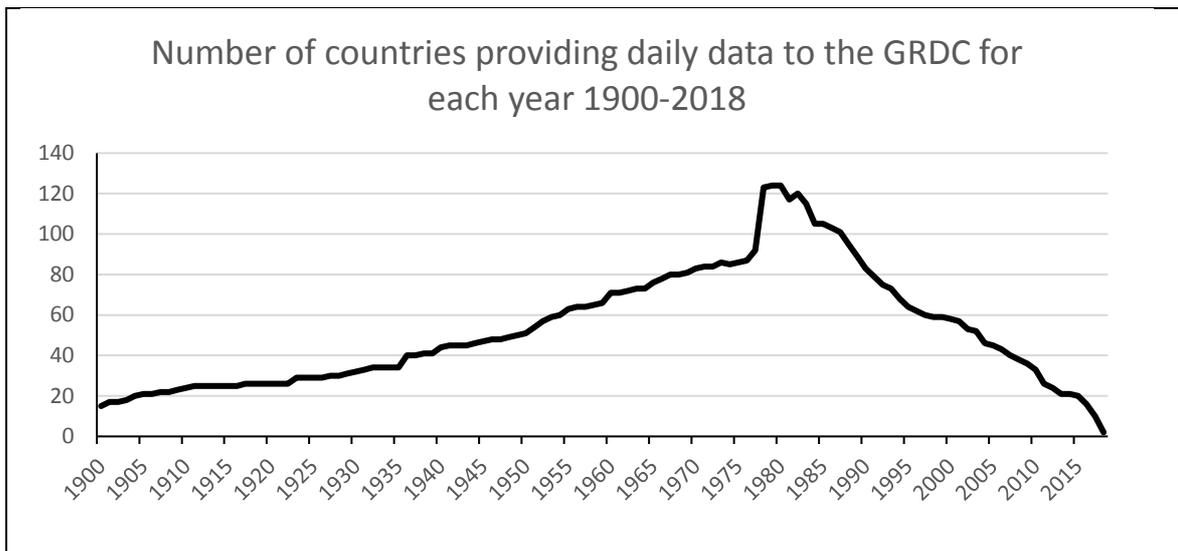


Figure 5 Graphs showing the number of stations in the GRDC with data for each year, and the number of countries providing data for each year (as of December 2018)

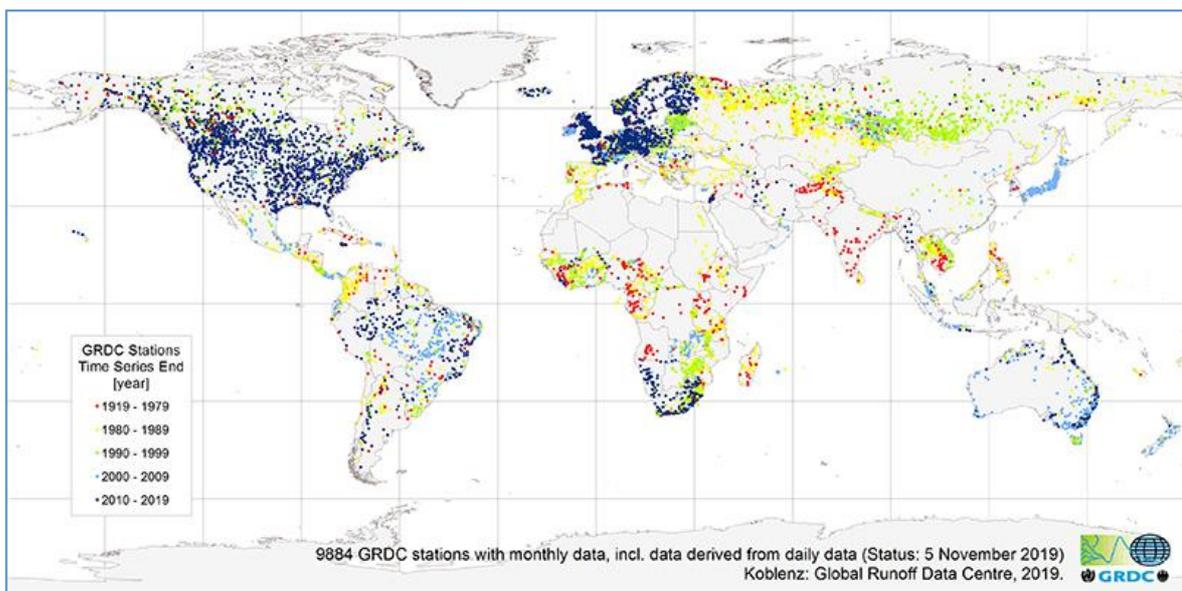


Figure 6 Map showing end dates of monthly flow data in the GRDC

Whilst the spatial and temporal coverage is good, there is no sub-daily data, limiting its utility for flood-related applications, particularly for smaller basins, though it should be noted that in many countries it is very difficult, even when requesting data directly from national monitoring agencies, to obtain long historic sub-daily river flow records.

Similarly, historic river level data are not held within the GRDC, and there is no global database of historic river levels. The changing relationship between river level and flow at any one location can mean that a long-term record of level is less useful than a record of the flow, which is the primary driving variable. River levels are more useful in real-time, or for the recent past, to establish contemporary relationships between flows and resulting flood extents.

Hirpa et al (2018) describe the calibration and validation of the GloFAS model. This used data for 1287 stations around the world, largely sourced from GRDC. Stations with more than four years of data for the study period of 1995-2015 were used, and split into a calibration and a validation period. Whilst the study period was defined by the availability of a specific meteorological driving dataset (in this case a dataset of retrospective forecasts), it underlines how important up to date historic records are for these purposes. The areas with data available for the GloFAS calibration (Figure 7) can be contrasted to those with data available for earlier periods (Figure 4). The study assesses model skill across different hydro-climatic regions and notes the need for representative in situ measurements in this respect. It also notes the limited availability of in situ data in some regions, particularly for flood-prone rivers in developing regions (Indus, Ganges, Mekong, etc.), and issues with the quality of some of the flow data used and the time required to resolve quality issues.

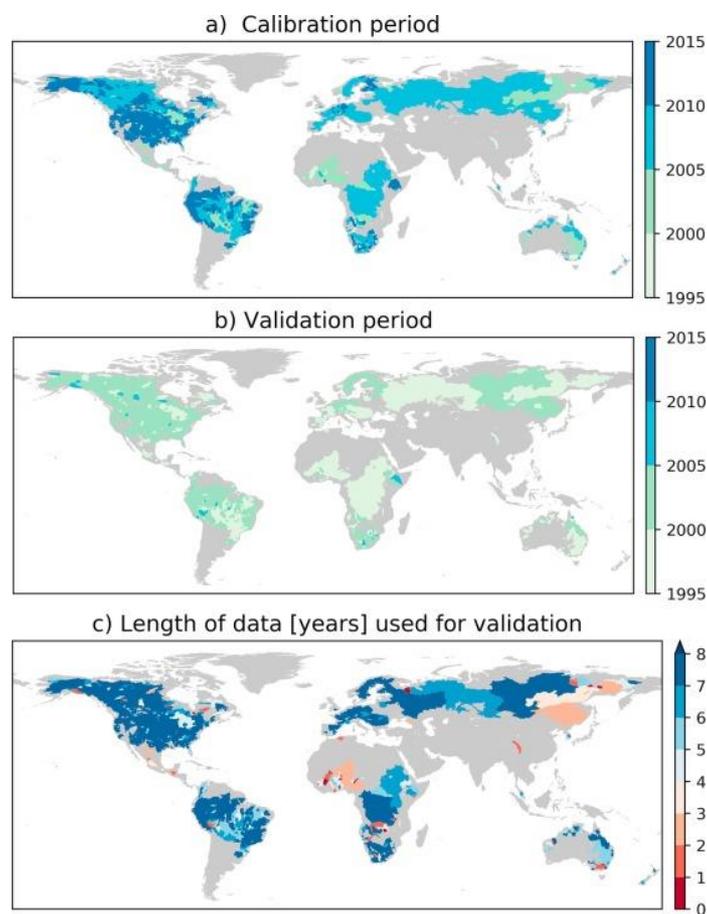


Figure 7 Spatial distribution of data used for calibration of the GLOFAS model, showing median year (a and b) and length of record (c). From [Hirpa et al \(2016\)](#) reproduced [under licence](#).

As an existing recognised data centre with a sustained source of funding, the GRDC still represents the most comprehensive source available of global river flow data, though its potential to fully meet Copernicus service requirements is limited by the speed of updates and its ease of access. The limitations of the data are largely due to the lack of provision of data by data providers, which can be for a variety of reasons, including prioritisation of data provision against sometimes stark resource limitations, and occasionally lack of cooperation with international goals. The WMO is currently considering the evolving roles of GRDC and other GTN-H data centres (WMO, 2016 – pages 12 and 36), and there may be an opportunity for Copernicus In Situ to express Copernicus requirements for global river flow (and level) in support of this activity.

### 3.3. Other potential sources of in situ data

#### Real-time data

There is no single open source of real-time river flow and level data globally, or at a European level. Many countries are now able to provide real-time data over the web as part of national flood forecasting services, but these tend to be restricted access (which could be a due to their status as components of an emergency service, needing to ensure services are resilient) and there are few completely open services.

#### National Hydrological Services

River levels and flows are recorded in every major river in the world. Almost all countries maintain a database of historical and contemporary hydrometric data either at a national level or at some regional administrative level. Real time data are very prevalent, in most countries from at least a proportion of the most significant river stations, though there are many countries where the transmission of data from field to central database is achieved with considerable lag time. In some increasingly rare cases data are created in analogue form and transcribed to digital form at a later date, though this situation rarely occurs at sites on important (larger) rivers.

Access to data from national and sub-national databases is very frequently difficult, even, as the coverage of stations within the HDCC attests, for Europe. In other global regions it can be yet more difficult, as confirmed by the variations in GRDC data over time, in which case there is a historical precedent to provide data, a strong international agreement around data sharing, and considerable active effort focussed on data acquisition.

The USGS provides fully accessible data for its gauging stations, with a queryable API (Application Programming Interface) for downloading data. Data is available to a varying extent in a limited number of other countries, generally via user-interfaces rather than via programmatically accessible APIs (e.g. France, <https://www.vigicrues.gouv.fr/>; South Africa, <http://www.dwa.gov.za/Hydrology/Default.aspx>; Brazil, <http://www.snirh.gov.br/hidroweb/>). In England live river levels and some flow data are available directly from gauging stations, both via a user interface (<https://flood-warning-information.service.gov.uk/river-and-sea-levels>), and an Application Programming Interface (API) (<https://environment.data.gov.uk/flood-monitoring/doc/reference>). But as the data from stations that require flow to be derived from ratings has not passed through the authoritative source of these conversions, it is often not accurate, and only the levels are provided via the user interface. For most sites an additional set of information on the most up-to-date conversion equations is required to derive accurate real-time flows from the level data. For similar reasons it is likely that real time river levels, which can be provided directly from the monitoring station, are more widely available on a global basis than accurate real-time river flows.

For the UK overall the National River Flow Archive (<https://nrfa.ceh.ac.uk>) undertakes the role of a national coordinating body, providing quality controlled historic flows from the agencies within England, Scotland, Northern Ireland and Wales, together with services to link records to API end points for real time data from those agencies. Whilst the specific structure of national and regional services varies greatly around the world, there is a role that national level coordinating bodies can play in resolving the aforementioned issue whereby regionally devolved responsibilities for hydrometric monitoring lead to inconsistent or inaccessible data.

A comprehensive survey of national hydrological agencies and their data access has not been undertaken for this report, but has potential to identify useful publicly accessible repositories of river levels and flows, and should be undertaken as a next step in understanding the availability, globally, of usable river level and flow data.

### Global services

Recognising the issues with international data sharing of hydrological data, WMO has a number of initiatives aiming to improve this situation. Global exchange of meteorological data is generally well advanced under the WMO Integrated Global Observing System (WIGOS). To advance hydrological data sharing in a similar way the WMO have established WHOS, the WMO Hydrological Observing System. WHOS provides information on accessible hydrological services, and is undertaking a number of initiatives to demonstrate the potential for harmonisation and web-based delivery of data between countries. Projects within the River Plata, the Sava, and others, are aiming to demonstrate how new technologies and standardised data exchange could provide benefits within basin-wide, national, and regional approaches. Currently these pilot projects are not providing publicly accessible data. Figure 8 shows the WHOS map indicating (in red) hydrological services providing data online. The spatial distribution can be seen to be non-homogeneous. Furthermore, the inconsistency of formats, data access methods, and metadata on data quality, observation methods, etc., means that use of data from multiple services would require substantial effort in downloading, transformation and manual checking. It is this issue that WHOS wishes to address through its ongoing efforts. In the meantime it is likely that global access to data will require significant dedicated effort, targetting countries where it is particularly important to address gaps in in situ data and where the data access is not excessively complex.

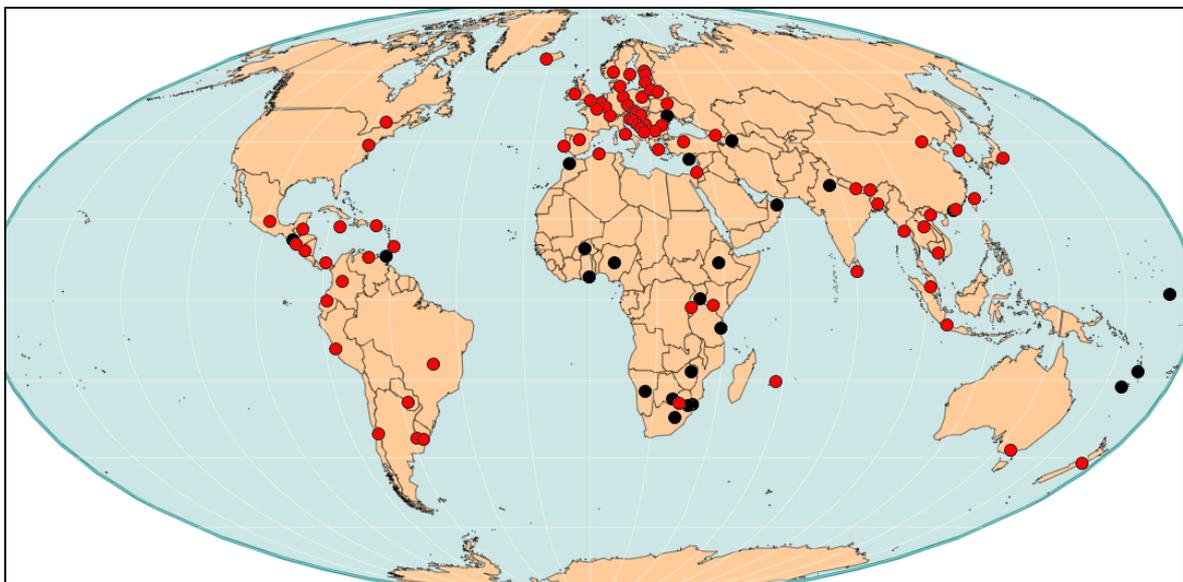


Figure 8 Map of National Hydrological Services accessible through the WHOS website (red: website and data access; black: website only).

### Other data collections

**R-ArcticNET** (A Regional, Electronic, Hydrographic Data Network for the Arctic Region), run from the University of New Hampshire (<http://www.r-arcticnet.sr.unh.edu/v4.0/index.html>) is an archive of river flows for stations flowing into the arctic region. Some of the data was used in addition to the GRDC data in calibration and validation of the GloFAS model. Most of the flow time

series end in the early 2000s, and it is not clear how open the dataset is or whether the initiative is sustained. However there may be useful data available for the Arctic MFC. It is possible that existing data collation activities that are specifically focused on the Arctic, across a number of environmental variables, could prove more successful in coordinating a coherent and up-to-date database of hydrological information for this region. The Copernicus In Situ activities focused on Arctic data should help to improve awareness of opportunities in this area.

**EMODnet-Physics portal** (<http://www.emodnet-physics.eu/portal/>) aims to provide harmonized access to in situ observational data pertinent to ocean physics, as part of the overall EMODnet (European Marine Observation and Data Network) programme of the EU Directorate General of Maritime Affairs and Fisheries (DG-MARE). It has river flow data for many inflow sites around Europe, though the coverage is quite selective, and many sites only provide historic data from the GRDC which is already otherwise available. However there are some additional river flow sites providing real-time data that is not readily accessible elsewhere. EMODnet-Physics' mission is to unlock all possible river data sources which are of interest to ocean and coastal communities.

The CMEMS LAMBDA project portal (<http://www.cmems-lambda.eu/Map/>) also provides river flow data in real or near-real-time for European coastal inflow sites some of which are accessed via EMODNET-Physics. Additional sites useful for coastal and marine modelling are being added to the LAMBDA database and portal and will be accessible from the EMODnet-Physics portal..

### 3.4. Gaps

There are no specific technological gaps in relation to river level and flow monitoring. In situ measurements, where available, are generally considered adequate for the purposes of hydrological modelling, and as accurate as is practicably possible.

In terms of the coverage of in situ measurements, it is generally the case that higher income countries in most cases have far higher density of monitoring than those with lower incomes. However, it is the limited access to the data from this monitoring that is currently imposing the most significant limitations on data usage.

At the European level, the EFAS HDCC is providing a functioning service for hydrological data access for EFAS for both real-time and historic data. It can be seen that there are some limitations to data access for some countries. Whilst this is continually improving as data providers implement new technologies and overcome barriers to data sharing, there may be avenues for coordination activities to help improve this. Currently no other Copernicus services make use of data from the EFAS HDCC, though discussions have already started within the Copernicus In Situ component.

Licensing between EFAS and its data providers is currently inconsistent, with bespoke licences for non-EUMETNET members. This has implications for the consistent management and use of in situ data held by the HDCC. There are opportunities to help to introduce a licence that is comparable between EUMETNET and non-EUMETNET members.

The CMEMS MFCs' modelling and CLMS pan-European future coastal inputs do not have sources of collated quality controlled historic and real-time data and could benefit significantly from access to data from EFAS, though it should be noted that in some cases modelled outflows are adequate, and easier to access and use. Furthermore EFAS HDCC is not directly focussed on coastal inflows and so the representation of sites many not be adequate. The EMODnet-Physics and LAMBDA projects

have undertaken significant work on identifying useful sources of coastal inflows, and there may be benefit in liaison between EFAS HDCC and EMODnet to identify future avenues to improve data access for all.

The GRDC is a good source of historic river flow data at a global level, providing data that would either not be possible to access otherwise, or which would take a huge amount of effort to collate, as well as preserving this data against the risk of loss within national archives. There are limitations in the spatial coverage, particularly for most recent periods which are represented in historic meteorological driving datasets, and for developing countries, which limit their use for applications for GloFAS.

There is no consistent access point for real-time river flow and level data on a global level. While there are many national and sub-national data portals, they are inconsistent in their approaches, and there is no catalogue or registry of these portals and the data they make available. WHOS is aiming to coordinate access to hydrological data from national services, and an exercise to identify, review and document the availability to data would produce a very useful resource to data users within the services and elsewhere. In the longer term WHOS has the potential to provide the global coordination required to create a federated system of national services providing consistent hydrological data, which would be of huge value to Copernicus. There may be potential for Copernicus to improve the situation through supporting the expansion of the WHOS programme.

In the long-term it is hoped that global data will be accessed directly from data providers, using standards-based approaches for requesting and delivering data over the internet. While data standards such as WaterML2 have been adopted by WMO, their uptake is minimal, in part due to a lack of readily available cost-effective software solutions that are appropriate for low income countries, and potentially a lack of requirement for data sharing in this way. WHOS aims to coordinate activities in this area. In some regions pilot projects have been undertaken to demonstrate this; these pilot projects could be more widely promoted. To improve access to river flow and level information globally, more support is needed for National Hydrological Services, and WHOS may provide an avenue for Copernicus to do this. In the longer term the future of hydrological in situ data access is likely to be via federated, and coordinated, national services providing consistent content within consistent formats, readily collated by global data services, and Copernicus In Situ should highlight and promote its requirements to help shape the development and evolution of this federated global data system.

### **3.5. Avenues for future coordination work**

The following activities are identified to help address some of these gaps:

- Assist EFAS with licensing for non-EUMETNET members
- Further discuss access to EFAS HDCC for other services, and provide any input required to a formalised agreement, ensuring it meets data provider licence conditions.
- Encourage EFAS HDCC to implement technical solutions to enable the selective opening of data services to end users based on licensing permissions.
- Liaise between the CMEMS MFCs and CLMS pan-European (and EMODnet) to describe availability of in situ river flow data and finalise requirements for in situ hydrological data for a future coastal service. Liaise between these services and EFAS HDCC to ensure monitoring data includes the sites required for coastal discharge (and matching to river water quality sites).

- Further develop and document the requirements for the Copernicus services for hydrological data access. This should clarify what a functioning hydrological data access system should look like, and be appropriate to highlight these requirements to other national and international organisations.
- Identify, with the appropriate components of the services (i.e. GloFAS) where GRDC updates would most benefit historic model calibration (if an updated calibration is planned) and communicate this requirement to GRDC.
- Support developments to GRDC and contribute to defining a vision for its future evolution.
- Work with WHOS to promote the needs of the services for access to in situ river flow data, and undertake a review of national hydrological data portals to help Copernicus services to identify readily available data resources.
- Liaise with WMO to explore where efforts to support data exchange could be aligned to reduce repetition, e.g. promotion of best practice for hydrological data exchange.

## 4. LAKE AND RIVER WATER LEVEL

### 4.1. Requirements for in situ data

There are two global water level products produced for the services, one for CLMS-GL (CGLOPS-2, 2018 e) and one for C3S (currently unpublished). The CLMS-GL water level product provides global near real-time water level time series at inland water locations on rivers and lakes at daily, 10-day mean and monthly mean intervals. The C3S water level product provides historic and up-to-date lake water level time series at daily to monthly intervals with a focus on long-term consistency.

These two requirements, for lake and river levels, are considered here together as they are strongly linked through the validation of a single global water level product. However, the requirements for river levels should be considered alongside those described above in section 3.

The spatial resolution and locations of the ground points for which data is provided is determined by the path of the satellite. Altimetry measurements are not (yet) undertaken across swaths and therefore only the intersections of the narrow ground track with rivers / lakes of appropriate dimensions are available. Rivers and lakes are treated fundamentally differently. Data from 2 satellites is currently used (though a third is in the process of being integrated): JASON-3 with a ~100km track spacing and a 10-day repeated flyover, and Sentinel 3A with a smaller track spacing but with a 28 day repeated flyover. The temporal resolution of data will therefore vary between stations. The period of record of the rivers product is from 2002 / 2008 to present, and the lakes product is from 1992 to present.

For rivers these intersections, called “virtual stations”, are single points. Whilst there could potentially be in the order of 1M of these globally, the number within the final product is limited for a variety of reasons, and is currently around 2000. The width of rivers is a major factor: rivers over 100m in width are mostly included, and rivers down to 30m can be included. Below this the resolution of the data is not sufficient to accurately distinguish the river surface. Obstacles such as riparian vegetation, sandbanks or bridges can also be a problem. However, developments in on-board technology (including the ability to improve accuracy where approximate height is known, even making use of in situ data directly) are improving this all the time, and automated methods for resolving issues at virtual stations are being developed, meaning the number of virtual stations

is likely to increase over time. In addition future satellites (such as SWOT, Sentinel-3 NG Topography) are likely to have altimetry data from swaths, meaning a move to a continuous measurement surface along rivers.

It is unlikely that a virtual station will coincide with an in situ measurement (in some experimental networks new in situ sensors are being located at virtual stations for validation, but this is a rare case). In most cases data from some distance up or downstream is used for validation. The appropriateness of an in situ station for validation of a virtual station will depend upon the distance and the comparability of that station to the virtual station, and so knowledge of what might affect this comparability are important. This could include factors creating differences between the amount of water flowing in the river between locations, such as lakes and dams, or significant incoming tributaries (and potentially diversions or bifurcations in rivers), or factors causing a different relationship between flow and water level, such as changes in channel geometry or slope. For some larger rivers (e.g. the Congo) the adjustments required to estimate absolute levels at virtual stations from nearby in situ stations are minimal. For many stations it is not possible to convert absolute measurements at in situ stations to levels at the virtual stations. However, the in situ data is nonetheless very useful, as strength of correlation with data at virtual stations can still be assessed and is one of the principal validation approaches. Nonetheless, good quality metadata on local reference datums that can enable accurate calculation of absolute in situ river levels is beneficial, as is information on the measurement approach or technology used, and the accuracy or uncertainty of the in situ data.

Lakes, being larger water bodies, can effectively have multiple altimetry measurements along a track across the surface of the lake, and larger lakes can be covered by multiple tracks. This leads to a different approach for calculating mean level, and also often a higher resolution in the temporal domain where a lake is crossed more frequently along different tracks (the Caspian Sea has a daily series as there is always at least one satellite above it). The product currently contains 60-70 lakes, though this is increasing continually, and new automated procedures for adding new lakes are being prototyped, which should increase this number significantly.

The use of in situ data is also different for lakes, where measurements at any point on the water body should be able to be used to validate the satellite altimetry measurement, though variations in the surface level across a lake must be accounted for. There are similar metadata requirements for in situ data for lake level validation: accurate location, information on methods used and changes in methods, clarity of the levels of uncertainty however basic (e.g. whether measurements are GPS calibrated).

Good quality real time data on levels for rivers and lakes could potentially be used in the production process of the product itself (e.g. on-the-fly calibration, or derivation of uncertainty metrics) but the timeliness of access to available data is not currently adequate, and there is not sufficient confidence in the quality of the data that is available.

## **4.2. Current use of in situ data**

Limited validation is currently undertaken for the CLMS-GL lake level product. The LEGOS database has been used for a study of Lake Issyk-Kul, and more validation work is planned. The GTN-H HYDROLARE (<http://hydrolare.net/database.php>) database is expected to provide useful measurements of lake levels historically, though data is not real time and has a significant lag time. It contains basic metadata for 1081 lakes globally, of which 420 have mean monthly water level

measurements, and 203 of which have instantaneous water level data for the first day of the month. Data for some lakes start pre-1900, and the latest data is for 2012 for stations from countries outside Russia, where data is available to 2016. Other databases from Argentina and the US Great Lakes are also being considered. No other global databases of in situ lake water level have been identified.

In situ data for rivers has been accessed from a number of sources as no global database exists. These include data for the Amazon (SO HYBAM), Niger basin (ABN), Congo (CICOS), France (SHAPI), and flows from the GRDC. River levels are preferable though flows can be used to some extent. Data access is a limiting factor, with manual requests and transfer of files required, and no real time access, meaning products can only be validated post hoc. Data licensing can be an issue in some cases.

For both rivers and lakes the validation and use of in situ data is still evolving and so there may be potential to provide significant benefits through timely coordination of access to in situ data.

### **4.3. Other potential sources of in situ data**

As described in section 3 above, river level measurements are widespread and may be available via services provided by national hydrological agencies. Many national and sub-national hydrological services also include water levels from lakes within their hydrometric databases and services, and it is likely there is publicly accessible data available, but this would require collating data from diverse sources in a range of formats and using different approaches. WMO's WHOS initiative has the aim of helping to develop capabilities within national services to deliver consistent data that could be used for this purpose.

As discussed above, GRDC provides widespread historic coverage of river flow data, which may have temporal signatures that are useful for validation of river levels. The EFAS HDCC has real-time and historic data (levels and flows) for rivers in Europe which may be able to be used by other Copernicus services.

An ESA project (S3TART), funded by Copernicus, on in situ Fiducial Reference Measurements for satellite altimetry has been proposed. Availability of in situ measurements co-located with virtual stations would be highly beneficial to the development and validation of Copernicus river and lake level products. There would be benefits in providing to this project further information from the developers of Copernicus river and lake level products on their requirements for in situ data.

### **4.4. Gaps**

There are no specific technological limitations to acquisition of in situ lake level measurements that would be suitable for validation of satellite products.

There is a lack of a global dataset of up-to-date in situ lake level measurements. Many national and sub-national hydrological services include water levels from lakes within their hydrometric databases and services, and it is likely there is publicly accessible data available, but this would require collation of data from diverse sources, likely to be using different approaches. In some cases the wider use of this data is hampered by licensing restrictions.

For river levels, there are a number of limitations in the available in situ data, including data quality, lack of information on the technologies used and their uncertainties (approaches can range from

manual reading from a fixed stage board, to GPS readings to optical and acoustic technologies), lack of datum information, and undocumented shifts in datums due to flooding or subsidence. The time required to manually resolve the various data quality issues is prohibitive in many cases.

The overall coverage is currently not ideal for validation, being limited in many areas and with a lack of extensive in situ measurements in any area. A large number of good quality measurements within a limited geographical area could be sufficient for validation of the global product, as the quality should not vary on a spatial basis. Temporally the in situ measurements available should be sufficient to cover the relatively limited historic period.

#### **4.5. Avenues for future coordination work**

- Liaise between CLMS-GL and CEMS to explore the potential for use of HDCC river levels for validation
- Summarise and promote the requirements for in situ river and lake level data for satellite product validation
- Highlight the need for a database of lake levels, and the detailed requirements, to GTN-H
- Highlight the potential benefits of the proposed Copernicus / ESA S3TART project on in situ data for altimetry, and feed in requirements from Copernicus product developers.
- Identify National Hydrometric services providing real time and / or historic lake and river level data

### **5. RIVER WATER QUALITY**

River water quality can encompass a huge range of measurements of pollutants and other factors which are important at a global scale, and relevant to Copernicus services.

#### **5.1. Requirements for in situ data**

The principal requirement for in situ measurement of water quality data from across the Copernicus services comes from **CMEMS and CLMS local/pan-European**, via the MFCs, for coastal modelling. Currently, it is believed, there is no in situ water quality measurement data used for ocean modelling. Within the current delegation agreement period there is preparatory work being undertaken, jointly between CMEMS and CLMS pan-European, to investigate the development of coastal zone products, including identification of in situ data required for these products. The requirements for inputs to the coastal areas are being identified by CMEMS, via the different forecasting centres. As data products develop from the coarser scales appropriate for open oceans to more detailed information required to understand the implications of water quality at more local levels within coastal zones (including sediment levels, bathing waters, etc.), there will be a greater requirement for in situ water quality data. Another main use of river quality data are the EU Directives regarding coastal marine and land data quality: The Marine Strategy Framework Directive and the Water Framework Directive. The principal categories of water quality information that were stated as required are “nutrients”, temperature and salinity. The nutrient load data requirements were described in different levels of detail by each MFC, but overall dissolved inorganic nitrogen and phosphorous were considered most important, followed by silicate; particulate and organic nitrogen and phosphorous would be useful if available, as would dissolved oxygen; dissolved inorganic carbon and bicarbonate data was identified specifically by the Arctic and Atlantic European North West Shelf MFCs. These requirements are summarised in Appendix A.

These requirements are of course strongly dictated by the potential availability for in situ or modelled data (itself dependent on appropriate in situ measurements being made). There may be value in clarifying these requirements to identify what would be needed in an ideal situation and also what would be prioritised in relation to the actual availability of currently accessible data.

The comments in section 3.1, describing the needs of **C3S** Sectoral Information System developers and users for in situ monitoring data also apply to water quality. Whilst current water quality products from these systems are limited, there is certainly potential for the development and application of medium and far future forecasting of river water quality, and in situ data would be needed in the development and exploitation of such products. One of the water sector demonstrators (SWICCA: Service for Water Indicators in Climate Change Adaptation) includes modelled water quality indicators under different climate change scenarios of: water temperature; inorganic, organic, and total nitrogen concentrations and loads, and soluble, particulate and total phosphorous concentrations and loads. Whilst the outputs are modelled, there are requirements for in situ monitoring data for the calibration and validation of water quality models, generally requiring data at temporal and spatial resolutions that are far higher than those available from current monitoring activities. Such outputs are also planned for the operational product which, like the demonstrator, will be focussed on Europe.

## 5.2. Current use of in situ data

In situ water quality measurement data is not believed to be used currently within the production of satellite products by the Copernicus services, though, as discussed, there are plans to start developing coastal zone products that will need a greater level of detail about the quality of water discharging into oceans.

## 5.3. Other potential sources of in situ data

The **GEMS/Water Programme** (<https://www.unenvironment.org/explore-topics/water/what-we-do/monitoring-water-quality>), established in 1978, is the primary source of global water quality data. GEMS/Water is considered as an inter-agency body of the UN and is functionally part of the Science Division of the United Nations Environment Programme (UN Environment). It is described as a “multi-faceted water science centre oriented towards building knowledge on inland water quality issues worldwide”. This includes research into water quality data collection, and capacity building in respect of water quality monitoring with the aim of improving the ability of participating countries to assess the state of, and trends in, national, regional and global water quality. As such it takes an active interest in the requirements for in situ data for the validation and improvement of satellite water quality products and ensuring such data is available at the global level, as well as the development of new in situ sources such as could be provided by citizen science approaches.

One function within GEMS/Water is the **GEMStat** database, which contains data for rivers, lakes, and groundwater systems. The GEMStat data portal (as of June 2019) contains data for 3730 rivers stations globally (309 in Europe), 2762 of which have data for “nutrients” (203 in Europe), and 309 have water temperature data (275 in Europe). The number of stations across Europe varies from country to country. There is limited non-European data available for the required inflows to the Mediterranean and Black Sea. Some non-European Arctic stations have good coverage (e.g. several hundred nutrient measurements for the Lena and Pechora rivers), but others have limited data (e.g. 70 or so nutrient measurements for the Mackenzie river through the 1980s). The GEMStat portal

provides a very good interface for identifying and accessing data, as well as simple downloadable station catalogues for offline use.

The European Environment Information and Observation Network (**EIONET**) hosts river and lake water quality data from EU member states. Data are reported on an annual basis, often with several years lag time. Data are openly accessible via the **EEA Waterbase** dataset (<https://www.eea.europa.eu/data-and-maps/data/waterbase-water-quality-2>). Data are available as annual or seasonal aggregations or as instantaneous measurements (~monthly). Waterbase contains 33,848,578 records of water quality across ~21,600 river, lakes and groundwater sites. The whole dataset is openly accessible over the internet, but is very large, and has no summarised catalogue (e.g. showing start and end dates, types of site, numbers of measurements and determinands across sites). The structure of data is tightly controlled, including use of a number of vocabularies of site types, determinand codes, and analytical methods, and it should therefore be relatively straightforward to extract data for a selected set of determinands across sites to produce a uniform dataset, though this assumes data providers have declared this metadata. It is likely to be the most accessible source of water quality data for Europe.

The **International Commission for the Protection of the Danube River** (ICPDR/IKSD) makes its water quality database available publicly (<https://www.icpdr.org/wq-db/>), though this is based on national monitoring networks and may contain the same data as the EIONET system. There are 79 monitoring locations across the basin, with up to three sampling points across the river at each location, with a minimum sampling frequency of 12 times per year. Additional water quality data is collected via occasional Joint Danube Surveys (2001, 2007, 2013).

Similarly the **International Commission for the Protection of the Rhine** (IKSR) provides water quality data publicly (<http://iksr.bafg.de/iksr/>). There are 18 sites on the main Rhine river, and numerous determinands available at approximately monthly resolution with a lag time of several years.

The **EMODNet-Chemistry** portal (<http://www.emodnet-chemistry.eu>) provides chemical information in ocean and coastal zones, and appears to contain some in situ river water quality data, collated from other source such as SeaDataNet, although it is difficult to interrogate the portal for river-specific information. SeaDataNet itself also contains data for some European rivers, though does not seem to provide coverage across all countries, and it is not clear whether they offer data that is not held elsewhere.

As discussed above, data on, or estimates of, river flows are needed for locations at which water quality measurements are made (usually as concentrations) in order to calculate total pollutant loads. It is difficult to use the query interfaces of the various portals described here to identify co-location in situ flow and water quality data. Any in situ data scoping exercise should jointly consider flow and water quality data. It is likely that modelled river flows, e.g. from EFAS for Europe, may be available to Copernicus services and be more readily accessible and consistent, and therefore more useful for this purpose than in situ data. Such wider use of modelled data may have implications for the coverage of in situ data required for the modelling activity.

## 5.4. Gaps

The availability of water quality data at a regional or global level is generally driven by statutory monitoring requirements, which leads to relatively coarse measurements both temporally (e.g. monthly) and spatially. This level of monitoring often fails to capture the quite significant variability

in pollutants, particularly where this is driven to some extent by rainfall (e.g. agricultural runoff and sewage system overflows). However it is likely that, in the first instance, access to some data on river water quality will be useful for validation or improvement of modelling of coastal and ocean waters. It may be the case that modelling water quality outflows to take account of temporally varying factors could provide better overall estimates of discharges to coastal waters, though in situ data would in turn be important in the validation of these models.

In situ data are likely to be available for major rivers, which will account for a large proportion of the areas contributing runoff to coastal zones, but for the areas between these rivers it is unlikely that suitably high resolution data will be available, and an approach based on estimation or modelling would be required.

In terms of the water quality parameters measured, most datasets contain data on nutrients and water temperature, as well as other parameters required such as silicates and alkalinity.

The discussion here is on data for Europe, as this is where the current requirements are focussed. Data for the rest of the world is far more limited. For areas of interest to the CMEMS MFCs that receive water from non-European countries (North Africa, Black Sea, Arctic Canada and Russia), there may be benefit in looking at existing collations of data specific to these regions.

## **5.5. Avenues for future coordination work**

The following activities are identified as potential avenues for future coordination:

- Liaise with the Arctic in situ project over availability of data for water quality for inflows to the Arctic within other datasets.
- Provide coordination as required by CMEMS and CLMS pan-European in relation to a more detailed review of available datasets for in situ river water quality, starting with the identification of EIONET / Waterbase stations on the major rivers identified by the MFCs in Appendix A.
- Undertake the review of EFAS HDCC sites discharging to coastal waters alongside the review of river water quality sites, in order to identify coincident sites.
- This activity should be coordinated so as to avoid repeated effort by separate MFCs.
- Develop longer term requirements for global in situ water quality data for satellite product delivery and promote these to global and regional bodies (e.g. GCOS, GEMStat, GEO Aquawatch).

## **6. LAKE WATER QUALITY**

Lake water quality is an area where new satellite data has huge potential to increase our ability to monitor and understand the status and change of the environment at global scales. Lakes are relatively isolated systems with locally varying pressures. Practical and financial constraints limit the extent of in situ monitoring but remote sensing provides an opportunity to produce the lake-specific measurements of water quality needed to understand the status of water bodies more extensively.

### **6.1. Requirements for in situ data**

**CLMS-GL** produces lake water quality products. The water quality product is based on the work of the NERC (UK) Globolakes project and initially covers 1000 lakes of medium and large size or of specific environmental interest. The data product consists of three parameters:

- Turbidity (describing water clarity and the ability for light to penetrate the lake water, and derived from estimates of suspended solids concentrations)
- Trophic State Index [TSI] (an indicator of the productivity of a lake in terms of phytoplankton, which over longer time scales can provide an indirect indication of the eutrophication status, and is derived from estimates of phytoplankton by proxy of chlorophyll-a)
- Lake surface reflectance (describes the apparent colour of the water body, and is intended for scientific users for further product development, including checking of the sometimes challenging atmospheric corrections, and also for producing “true colour” images). Lake Water Reflectance is an Essential Climate Variable.

Products are provided for these 1000 lakes at a nominal 300m and 1km resolution on a common grid, and at 10-day temporal resolution (1<sup>st</sup>, 11<sup>th</sup>, 21<sup>st</sup> day of each month). The current version of the products is described as a demonstration product.

The process by which the TSI and turbidity products are produced involves classification of the observed (after spectral adjustment for atmospheric corrections) water reflectance spectra into “optical water types” developed within the Globolakes project using measurement data (CGLOPS-2, 2018 a). No in situ data is used in the production processing chain (as real-time data are rarely available).

The product report states that the most significant of the assumptions within the processing chain is “that optical water types which have been defined from a large set of in situ data from optically complex waters (lakes, reservoirs, lagoons, estuaries, and coastal areas) can be assigned successfully to each satellite observation (pixel) containing open water”. Whilst issues with this assignment are often due to misclassification of water pixels or failure of the atmospheric correction algorithms, 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 the resulting product.

No uncertainty information is currently provided with the data product, partly due to the difficulty of deriving this through a complex processing chain, but also because of the small number of in situ optical water type measurements.

Validation of the products is described in detail in the Quality Assessment Report (CGLOPS-2, 2018 b). Validation was undertaken manually for a small subset of lakes against in situ data from the LIMNADES database and is summarised below. The report stresses that validation cannot be undertaken in an automated way for all 1000 water bodies, but does not set out what would be required from in situ networks in order to enable this automated analysis.

The ideal measurements required are optical surveys with associated concentration measurements of optically active water constituents, but these are generally only acquired in research activities. Standard regulatory measurements, such as those made for WFD monitoring, would be of use (e.g. Secchi depth, water colour, chlorophyll-a) but the observations require detailed metadata on the exact location within the lake, as well as the measurement / analysis technique used. Furthermore, for measurements to be representative of satellite data pixels, they should be located clearly within a pixel, ideally one which is located entirely over water. Hence lake-shore measurements are far

less likely to be appropriate for ground-truthing, and ideally in situ measurements should be made from a jetty or pier, or boat. There are significant opportunities to obtain useful optical measurements from UAVs which can capture the same measure (e.g. water reflectance across the optical spectrum) as satellite systems, and can do so away from lake shores and even at multiple locations across lakes.

The next framework contract will require a heightened focus on in situ data for lake water quality to characterise product uncertainty, particularly as the service is likely to expand to higher resolution sensors and more (smaller) lakes.

## 6.2. Current use of in situ data

A precursor activity within the UK Globolakes project undertook to assemble available lake bio-optical in situ data for use in remote sensing. This dataset is now managed as the LIMNADES database, held at the University of Stirling (UK). The Globolakes project was very successful in galvanising the community to provide available in situ data. Data from around 1500 lakes was collated, though this is very variable in quality and quantity. For in situ measurements of remote sensing reflectance only around 250 lakes have measurements. Similarly for chlorophyll and organic matter. Other measurements of optical properties required for remote sensing analysis are even more limited in their availability, though some measurements were made within the project.

The LIMNADES database is not an openly available dataset, but it is currently accessible to CLMS-GL contracted organisations working on lake water quality product validation. The LIMNADES data policy states that the data will be accessible to organisations contributing data.

There are many issues with the usability of the data that is available. For example, measurements of chlorophyll are not undertaken to any specific standard, with different methods of preservation, analysis and description used in different projects, countries and catchments. In addition, the LIMNADES in-situ data are biased towards measurements of chlorophyll-a whereas turbidity and particularly reflectance data are scarce. Furthermore, eutrophic waterbodies are represented more frequently than oligotrophic and mesotrophic waterbodies.

Other projects are also taking on this work (with University of Sterling involved), such as the EU MONOCLE project (<http://www.monocle-h2020.eu>), and an instrumentation node within the Danubius Research Infrastructure Preparatory Phase (<https://danubius-pp.eu/>) looking at use of satellites for looking at deltas / estuaries and large rivers. While these projects are likely to deliver data to the LIMNADES database, it is likely that other research activities may be producing data that is useful but not accessible.

## 6.3. Other potential sources of in situ data

Standard regulatory measurements, such as those made for WFD monitoring, and made available via the **Waterbase** dataset, would be of use (e.g. Secchi depth, water colour, chlorophyll-a). However, the observations available often do not provide sufficient metadata to identify where they are made (e.g. measurement for a lake given with only lake name or centroid rather than exact location) and many will have been taken from the lake shore. Measurements should ideally be taken from a boat or long pier/jetty.

The **GEMStat** database, described above, also contains data on lake water quality measurements. However these offer limited utility for the validation of satellite products, with a limited number of the type of measurements required (e.g. 53 lake sites with data on Total Suspended Solids, 64 with transparency data). In addition, data are likely to be taken near to shore, or using simple measures not directly comparable with the satellite data.

The World Lake Database (<http://wldb.ilec.or.jp/>) contains scientific and socio-economic data on the environments of 217 lakes and reservoirs in 73 countries compiled as part of a UNEP (United Nations Environment Programme) project. Data are registered within a catalogue. Currently about six hundred lakes are registered, though the available data seems to be relatively limited.

Some countries have suitable in situ measurements available directly, e.g. US EPA, Sweden, but globally there is no adequate data available.

## 6.4. Gaps

It is likely that the Globolakes / LIMNADES work has collated most of the appropriate data available globally. Future research projects obtaining bio-optical data should be encouraged to make it available for satellite calibration.

Current use of in situ data is based on traditional water quality monitoring that was not designed for the validation of satellite data. One challenge is the atmospheric correction of satellite imagery, where uncertainties exist even for the comparison with in situ data. There are now numerous remote sensing techniques that can be used to produce comparable optical measurements that remove this uncertainty, providing true optical measures for direct comparison, but there is little data of this type being recorded, and it is largely undertaken within research projects.

Overall there is a genuine awareness amongst national agencies of the potential benefits of remote sensing of lake water quality, but there is an inertia to the types of changes in practice that would be required to enable more useful data to be routinely collected.

In terms of data access, the CLMS-GL team working on lake water quality product development have access to the LIMNADES database, and there is nothing to suggest that this will be a problem for the Copernicus services in future. There may be potential for opening the LIMNADES database up for less restricted access.

There are opportunities with new UAV-based technologies to undertake measurements that are directly comparable to satellite readings, whilst avoiding atmospheric effects. Such measurements have the potential to provide better in situ data for validation. Research projects are already investigating these approaches, but wider uptake may require on the development of better protocols and guidance. There is a need to promote the benefits of such measurements within regulatory monitoring agencies, and their ability to improve the quality and usability of satellite products.

In relation to sustainability issues, the LIMNADES database is clearly the most useful resource for data for satellite lake water quality validation. It is currently hosted by the University of Stirling, which sets out, on the LIMNADES website, that there is a "long-term vision to maintain this database beyond the end of the GloboLakes project (ends 2017)".

Supporting the development of satellite water quality products is within the remit of GEMS/Water. There is an intention to evolve the contents and focus of GEMStat to include use as a calibration and validation database for satellite data. A short ESA funded project SPONGE (<http://www.odermatt-brockmann.ch/sponge>) worked with GEMStat national nodes to collect water quality monitoring for satellite validation, and identified some of the issues in doing so on a widespread basis.

There may be potential for citizen science to address the lack of in situ data, particularly in some countries. There are efforts underway in respect of this, e.g. through GEM/Water Capacity Development Centre.

There are a number of other initiatives in the area of water quality. The UNEP World Water Quality Assessment is forming a World Water Quality Alliance with the aim of improving the data and methods available to assess the current state and future projections of water quality. GEO AquaWatch 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”, and has a specific working group on observations and data. A number of these initiatives are developing requirements and protocols for in situ water quality measurements. There may be benefits in ensuring the specific requirements of the Copernicus services are properly described and that they feed into these various initiatives.

## **6.5. Avenues for future coordination work**

The following activities are identified as potential avenues for future coordination:

- Work with CLMS-GL teams to fully capture the technical specifications for in situ measurements of lake water quality, including UAV-derived measurements, and coordinate the development of protocols to feed in to other global initiatives such as GEO Aquawatch.
- Identify, with the developers of data products, where citizen science might be able to provide useful data on a widespread basis.
- Liaise with the hosts of the LIMNADES database to identify the potential to open it up for less restricted access to Copernicus, and understand future sustainability.

## **7. LAKE SURFACE WATER TEMPERATURE**

Lake Surface Water Temperature is an Essential Climate Variable with environmental and meteorological applications including water quality management and weather forecasting, as well as long-term monitoring.

### **7.1. Requirements for in situ data**

There are 2 Copernicus lake surface water temperature products. CLMS-GL has a 10-day mean ~1km product, to go alongside its water quality parameter product, produced in near real-time (CGLOPS-2, 2018 c). C3S has a daily resolution ~1km product (daily time slices with data where available) produced on an annual basis for the previous year (currently unpublished). Both products apply to the same ~1000 lakes as the CLMS-GL water quality product, run from 1996 (currently) and are each based on a related methodology, using remotely sensed infra-red data, and are validated using the same in situ data. No in situ data is used in the actual creation of the products, only for validation.

The ~1000 lakes are at least 2km across, and due to the pixel size, measurements should ideally be 1km or more from the lake shore. However, due to the limited number of such measurements, validation is undertaken with any available data. To be usable for validation, in situ data cannot be averaged over periods longer than one day, and ideally are instantaneous values (with metadata on the time of measurement). Monthly mean data is not usable. Continuous measurements at a location are preferable, covering as much as the past 20 years as possible. However sparser measurements or individual measurements at different times and locations would still be useful.

The data product being validated is the lake surface temperature, and so surface temperature measurements are preferable, though the validation process can handle measurements at depths, as long as these depths are known. Uncertainty in the in situ measurements would be useful, though for standard thermistors, etc., these can be estimated. Consistently available metadata, particularly on the precise location (preferably to within 0.001 degrees of latitude / longitude) and depth (specified to within 10 cm), would be beneficial.

## **7.2. Current use of in situ data**

Various sources of data have been investigated in the process of validating the CLMS-GL lake surface water temperature product. A new dataset of lake water temperature was assembled on the basis that there are currently no suitable open global databases of lake surface water temperature. Over time this database of in situ data for validation has been built up with significant effort, obtaining data directly from individual institutions where suitable contacts could be found. These can be understood to be roughly evenly split between observations taken by environmental monitoring agencies, by researchers and research projects, and by meteorological agencies.

The dataset currently comprises data from 56 locations (on 48 lakes), a limited proportion of the 1000 lakes represented within the data product. The lakes represented are listed within the CLMS-GL product validation report (CGLOPS-2, 2018 d), which also describes the validation approach and results. The spatial coverage is acceptable for Europe, but is lacking data from Africa and China (specifically the Tibetan lakes) and becomes sparser back in time (particularly before 2000). As with the HYDROLARE database, it is understood that much of the available digitised data for Russian lakes is aggregated to monthly average temperatures, and therefore of limited use for satellite data validation.

## **7.3. Other potential sources of in situ data**

The GTN-H HYDROLARE initiative (<http://hydrolare.net/database.php>) is the international data centre for hydrology of lakes and reservoirs, supported by WMO. The HYDROLARE has not been investigated in detail as it is believed to contain only monthly average temperature measurements. It contains basic metadata for 1081 lakes globally, of which 21 have water temperature data. A data catalogue is provided electronically on the HYDROLARE website, but the data is otherwise not possible to interrogate online.

The US Great Lakes Monitoring data is freely available from NOAA and the US EPA, with measurements taken from a research boat within the lakes system. This has already contributed useful data for validation.

Some of the validation data used has been accessed from the Global Lake Ecological Observatory Network (GLEON) database, though many lakes in the network are smaller water bodies.

There may be open sources of data that could provide information for additional locations. For example the GEMStat database may contain data for a limited number of additional sites. It contains 134 sites with lake water temperature data, though many of these are likely to be from smaller lakes. The EIONET / Waterbase repository is also likely to contain useful data for European lakes in a consistent format that might be made use of relatively easily. It is difficult to query in order to identify the number of potential sites that hold lake temperature data, but for example there are ~800 measurements of water temperature across 159 lakes in Finland in 2016. The level of detail of the metadata or the accuracy of the spatial locations is not known, and this may limit its utility for validation purposes.

An open dataset of global lake water temperature measurements from 1985-2009 was published in 2015 (Sharma, 2015; dataset available at <https://portal.lternet.edu/nis/mapbrowse?packageid=knb-lter-ntl.10001.3>), within which 186 appear to have data be based on in situ measurements. This database is related to the Global Lake Temperature initiative in the US (<http://www.laketemperature.org/>). However the data in the open database is monthly means only, and therefore of limited suitability. It may be worthwhile contacting this initiative to identify whether instantaneous data is available from other sources.

## 7.4. Gaps

Lake water temperature is a simple measurement, regularly observed and reported. There are therefore no sustainability or technology gaps in relation to the observations.

The CLMS-GL validation report explicitly notes an “important limitation related to the validation is the lack of in situ measurements and the existence of a solid quality controlled in situ measurements database”. The principal issue is the lack of daily or instantaneous values within global level collections of lake water temperature data.

However, there are also limitations in the data that has been collated by CLMS-GL. There is a lack of off-shore temperature measurements, as well as a lack of metadata in relation to specific monitoring locations which can limit the usability of data. Some mobile platforms such as those used in the US Great Lakes provide off-shore measurements but are withdrawn during the winter to avoid the ice season, and so lack data for the coldest part of the year.

Licensing of data can also be an issue, as with many data sources, limiting the ability to pass on data to 3<sup>rd</sup> parties for independent verification of the validation, and limiting the ability to publish outputs in scientific journals (though not explicitly due to limitations on use within commercial products).

## 7.5. Avenues for future coordination work

The following activities are identified as potential avenues for future coordination:

- Work with CLMS-GL product developers to summarise and promote the requirements for in situ validation data for lake surface water temperature data, specifically for instantaneous or daily values.
- Explore the contents of the HYDROLARE database to understand if underlying instantaneous measurements may be available.

- Liaise with WMO and the hosts of the HYDROLARE database to highlight these requirements and discuss potential future avenues for improving in situ lake temperature data for satellite product validation.
- Review the EIONET / Waterbase dataset to identify whether lake surface water temperature measurements are appropriate for validation.

## 8. SOIL MOISTURE

Soil moisture is a difficult hydrological variable to measure, but is an important factor in the hydrological cycle, influencing flooding and drought and driving feedbacks with the atmosphere, and also in agriculture and irrigation.

### 8.1. Requirements for in situ data

CLMS-GL produces two soil moisture related data products, a Surface Soil Moisture (SSM), and a Soil Water Index (SWI).

The SSM product is a 1km resolution dataset, based on Sentinel-1 radar backscatter data, of relative soil moisture as a percentage of saturation levels, updated every 1.5-4 days for swaths over the European continent (though a global product is foreseen for future release) (CGLOPS-1, 2018 b). Spatial coverage varies by location due to the repeat pattern of the swath paths and width of swath. **Error! Reference source not found.** (extracted from the product manual) shows that in some areas there is very limited coverage, and so there are implications for the suitability of in situ measurements.

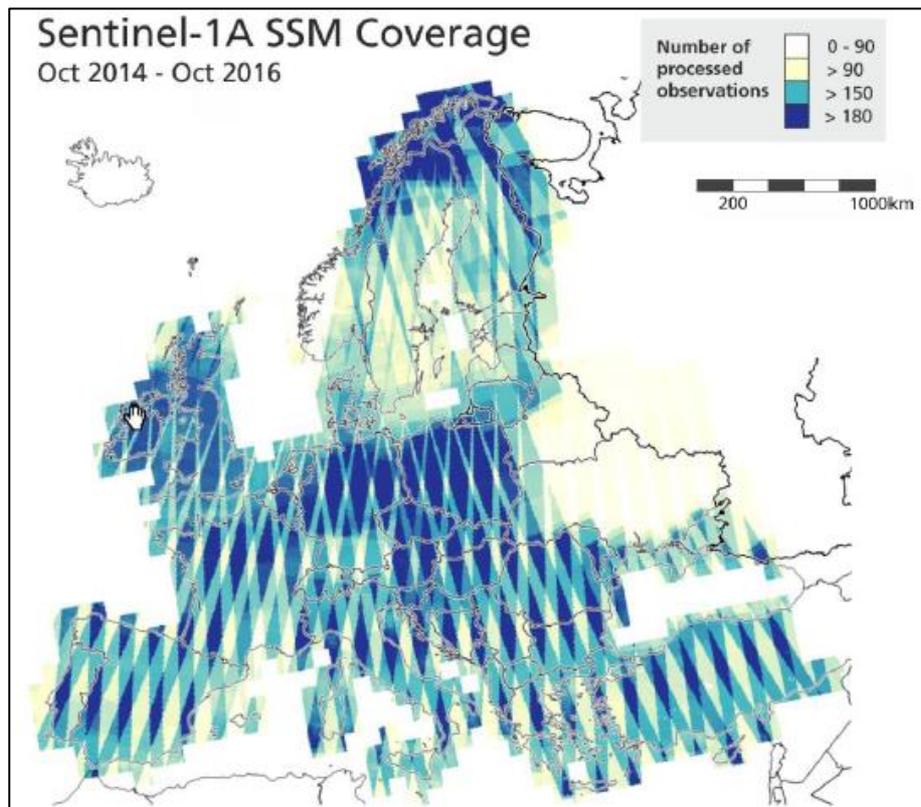


Figure 9 Map showing the number of processed observations of Sentinel-1A cglfor SSM retrieval over the period October 2014 – October 2016 (excerpt from CGLOPS-1, 2018 b)

The SWI dataset is based on a combination of high resolution Sentinel-1 SAR with ASCAT data and provides an estimate, also over Europe, of soil water saturation at 8 nominal depths of 2, 5, 10, 15, 20, 40 and 100 mm, at 1km on a daily timestep, updated daily. These depths are nominal as their relation to the equivalent absolute depth will vary with soil texture.

Validation of the data products is currently undertaken on an annual basis in a number of ways, both with in situ measurements and with data from other products (e.g. modelled soil moisture from the GLDAS Noah model). For the SSM product, in situ data is used to validate both the time series for individual sites, and the spatial variability across an area.

C3S produces one soil moisture product representing surface soil moisture (C3S, 2018 a) comprising three datasets: active, passive and combined. All C3S products feature daily, 10-day and monthly mean intervals, spanning a time range from 1978 to present (passive only prior to 1991), on a 0.25 degree grid globally. Two versions of this product exist: the Intermediate / Interim Climate Data Record, updated every 10 days; and the Thematic Climate Data Record, updated on a 6-monthly basis. The products are based directly on the ESA Climate Change Initiative Soil Moisture ECV.

The C3S products are validated to understand the variations in accuracy of the satellite data relative to available in situ measurements with a number of factors such as soil depth (i.e. how well it compares to in situ data at different depths), soil texture, climatic zone (as defined by Koeppen-Geiger class), land cover (grassland and tree), and season (C3S, 2018 b). The quality assurance report highlights an issue with the representativeness of the available in situ measurements in relation to these factors. A detailed gap analysis has been undertaken and documented (C3S, 2018

c) but this focusses on gaps in the production process and instrumentation, and does not mention any specific limitations of the in situ data used for calibration.

Soil moisture is highly variable spatially, and dependent on very local factors. For all products spatially dense networks are required that capture the local variability across the satellite footprint. This is made more important by the fact that many point soil moisture sensors are heavily influenced by conditions directly around the sensor and so a number of sensors are required to ensure a good representation.

For validation of the time series at a point it is important to have continuous observations, though most instruments provide data at higher temporal resolution than the satellite products. In future the in situ data will be complemented by 1km modelled data, to overcome inadequacies in the spatial density. In many cases the at-site variability is used for validation rather than absolute values, e.g. for the SSM product validation the satellite data was fitted to the distribution of the in situ data.

The C3S and CLMS-GL SWI products require surface and / or near surface measurements as well as a range of other depths for proper calibration.

Many electrical soil moisture sensors produce soil temperature outputs as well. For the SWI data product, soil temperature was also used as a validation dataset, and so the inclusion of this additional variable within soil moisture databases is important.

Data of consistent quality, with sufficient metadata about locations, technologies, uncertainties of measurements, local assessments of soil type and land cover, etc., are all important.

## 8.2. Current use of in situ data

All validation of the soil moisture products against in situ data makes use of data from the International Soil Moisture Network (ISMN, <https://ismn.geo.tuwien.ac.at/en/>). This network, operated from the Technical University of Wien, is a “combined effort of the Global Energy and Water Exchanges Project (GEWEX), the Committee on Earth Observation Satellites (CEOS), the Global Climate Observing System - Terrestrial Observation Panel for Climate (GCOS-TOPC), the Group of Earth Observation (GEO), and the Global Terrestrial Network on Hydrology (GTN-H).” It is currently supported financially by ESA’s Earth Observation Programme, and whilst there have been some uncertainties over its long term funding it has recently (November 2019) obtained funding from ESA for a further year to December 2020. Copies of data from individual sites and regional and national networks are provided to ISMN under an open license, and data is freely downloadable. Many networks are focused specifically on soil moisture measurement while others record soil moisture as part of a range of research-focused observations (e.g. for meteorology, or catchment hydrological studies). Soil temperature is also available for some sites and, as stated above, this can also be useful for in situ validation of some satellite soil moisture products.

ISMN is likely to contain the majority of longer-term soils moisture measurements from distributed networks, and the harmonization of data structures hugely improves the ability for researchers to access global soil moisture data. Despite this it can be seen from **Error! Reference source not found.** that measurements are very spatially heterogeneous. There is a fundamental lack of soil moisture measurement being undertaken globally. The USA is the only country with a resemblance of a representative coverage (the UK has a new network of ~50 spatially extensive sites but these are

not yet represented within the ISMN). Whilst a heterogeneous distribution is not necessarily problematic in itself (in fact some of the current validation approaches require clustered sites) it is likely to mean that some soil types, land cover types, topographic conditions and climatological zones are under-represented, and lead to considerable bias. This is noted in some of the validation reports, but there is no specific information on where this representation needs to be improved.

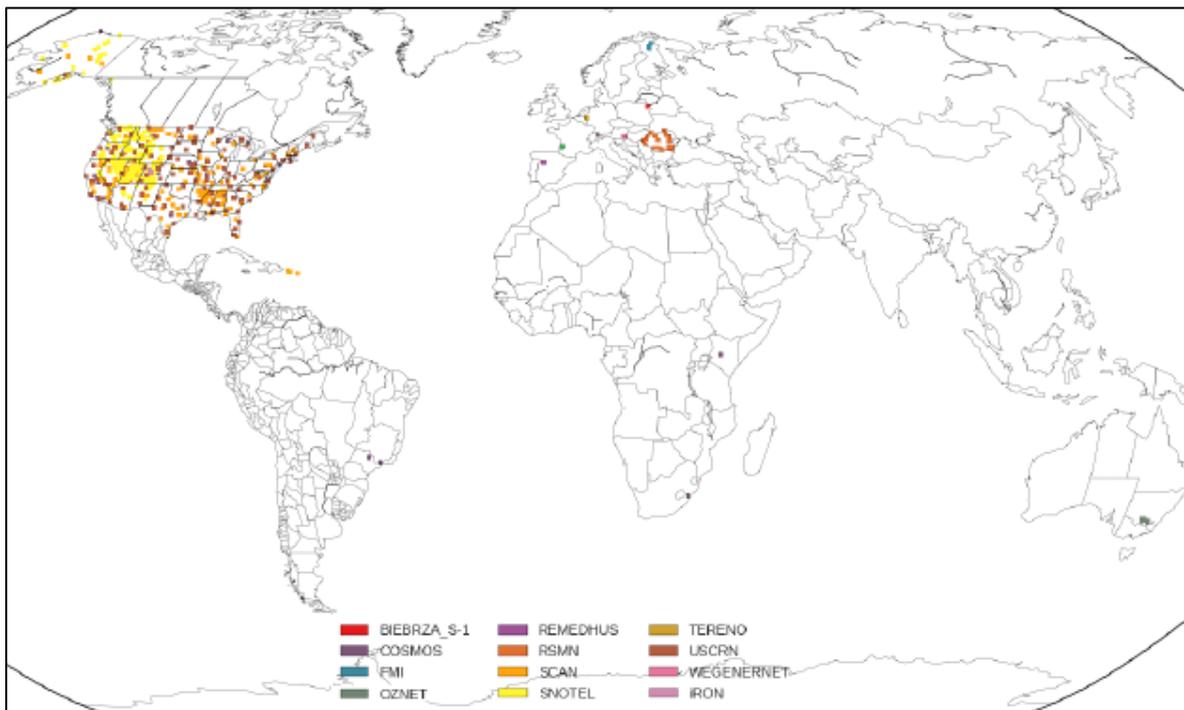


Figure 10. Distribution of sites within the International Soil Moisture Network (ISMN)

**Error! Reference source not found.** shows the ISMN sites in Europe together with the coverage of SSM1km product for a single day (image extracted from the product validation report). Whilst other days have better coincidence between sites and swaths, this map demonstrates the impact of the limited in situ coverage on validation of the product.

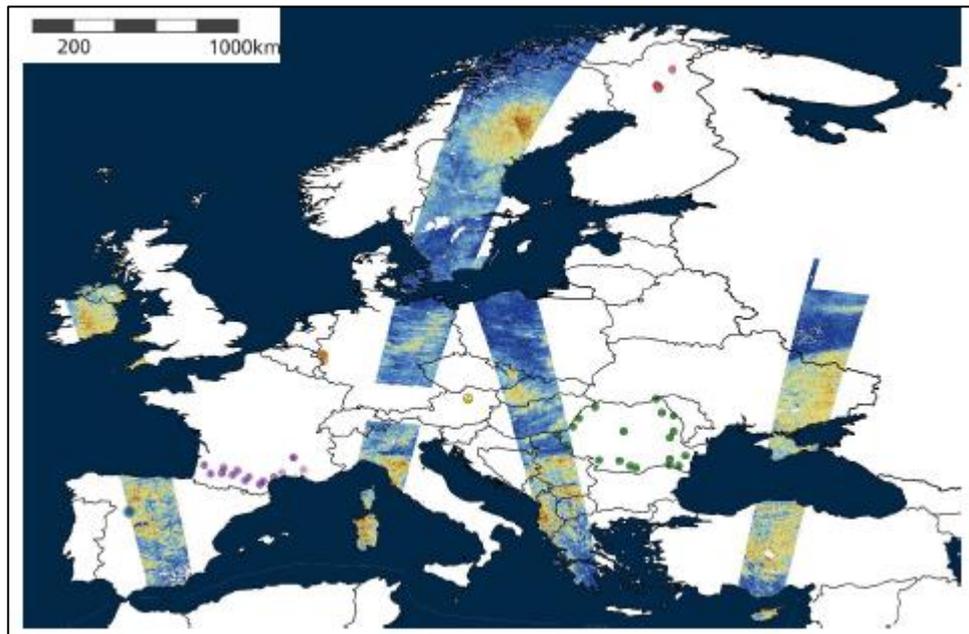


Figure 11 Map showing the coverage of ISMN sites in Europe and the SSM1km product for 6/9/2015 (excerpt from CGLOPS1, 2018 b)

The validation reports describe how in some cases the spatial representativeness across the satellite footprint also limited the ability to use the data. More dense networks, with far greater numbers of sites, are required for a proper validation of the satellite data.

ISMN undertakes its own standard quality control checks and data cleaning, but data quality is an issue within the network, with one validation report describing how “many soil temperature time series are contaminated with inconsistent and erroneous data, especially in the winter season, so that the ISMN dataset on soil temperature suffers from irregular coverage and quality”. Point soil moisture sensors are prone to shifts in measurements over time as instruments change the extent to which they have contact with the soil, sometimes due to disturbance. This can mean that even at a single location, multiple sensors, and quality control processes across these sensors, are required to provide a consistent long-term record.

### 8.3. Other potential sources of in situ data

As described above, it is likely that the ISMN contains the majority of useful coordinated soil moisture measurements globally. It is possible there are a few more that could be accessed in addition. The US National Soil Moisture Network (<http://nationalsoilmoisture.com/>) lists additional US networks that are not currently part of ISMN (e.g. West Texas Mesonet, Oklahoma Mesonet), though their absence from ISMN could be due to data licensing limitations. In the UK there is a new national network of sites with both COSMOS (wide area) and point sensors (<https://cosmos.ceh.ac.uk/>) and so there may be opportunities to increase the coverage slightly.

### 8.4. Gaps

The principal gap is in the very limited number of soil moisture samples, particularly outside of the USA. The currently available data is inadequate for the proper validation of satellite soil moisture products over Europe, and most of the world. A coordinated network representing a range of soil types, land cover types, and climatic zones is required, with continuous measurements from sites

with multiple sensor across satellite footprints measuring at a range of depths, and multiple nearby sites to help understand spatial variability.

Gruber et al (2012) describe how, even within the ISMN dataset of 1400 sensors globally, there were significant variations in the error of sensor data both between networks, and at sites within networks, with 36% of sites failing to meet the ECV accuracy requirement of  $0.04 \text{ m}^3\text{m}^{-3}$  as assessed by the described method. Certain networks and sensor types were particularly prone to errors. The authors recommend “development of procedures to reliably select representative existing or future sites for the in situ–satellite inter-comparison”. The developers of the Copernicus soil moisture products have been involved in the development and assessment of in situ datasets.

## 8.5. Avenues for future coordination work

The following activities are identified as potential avenues for future coordination:

- Liaise with the ISMN to help to describe and widely promote the requirement for an adequate in situ soil moisture monitoring network.
- Liaise between ISMN and operators of networks not currently represented (e.g. COSMOS-UK)

## 9. LAKE EXTENTS

Lake extents (Water Bodies) data products are produced by CLMS-GL (CGLOPS-2, 2018, f). It is effectively a 300m/1km resolution dataset of the existence of water bodies in any given pixel, plus quality information about the certainty of water body occurrence. It has global coverage and is calculated on a 10-daily basis and made available in near real-time. As a time series it provides information on changing surface areas of lakes. An existing satellite-based product (the JRC Global Scale Water History Record, Pekel 2016) was used both to identify initial lake outlines, and also to validate the product over the period of record (CGLOPS-2, 2018 g). No in situ data was used in the production or validation of the dataset. It is unlikely that in situ measurements of lake shore extents would provide enhanced validation at the scales at which this is currently undertaken. For this reason this data product is not discussed further within this document.

## 10. SUMMARY AND CONCLUSIONS

We have identified Copernicus products that relate to hydrology and have summarised the level of current and potential use of in situ hydrological data in their production. In situ data requirements are described for key variables of river level and flow, lake level, river and lake water quality including water temperature, and soil moisture.

River flows and levels are particularly important, underpinning operational flooding services from CEMS, and wider use is currently being considered in the development of future coastal zone products, as well as for more thorough validation of water level / altimetry products. In other areas there are specific current demands but a clear lack of available data, as summarised below.

In all cases the development of products is ongoing, and the use of in situ data varies significantly between products. In situ hydrological data is currently mainly used for assessment of data product quality, focussed on historic data. Very few products make direct use of in situ hydrological data to improve products, and only the CEMS Flood Awareness Service makes use of real time data.

In most cases the extent to which in situ data is used is limited by the availability of data, and, if more readily available, in situ data would be more extensively used and would lead to improvements in the products, or at least a better understanding of uncertainty in the products, an important consideration in relation to uptake with end users.

The findings of this work can be summarised within the following categories:

### Monitoring

In most areas hydrological monitoring is not designed for the purposes of validating or improving satellite data products, and so the monitoring undertaken is often not fit for this purpose. Monitoring schemes have not yet been designed to help yield benefits from complementary satellite monitoring. In the long term combined satellite and in situ products will not be able to be underpinned by research-level monitoring alone, and so an evolution of statutory monitoring regimes will be required to provide the level and quality of in situ data needed. Whilst agencies responsible for statutory monitoring may understand that satellite products will provide enhanced monitoring in future – helping to achieve levels of spatial and temporal coverage that are not possible with in situ monitoring – this is not reflected by an evolution of monitoring regimes towards those that would support the development of products. There may be many reasons for this, and the evolution of statutory monitoring regimes will take time, but Copernicus In Situ should be aiming to assist in providing the evidence to support the case for change, and in particular clarifying requirements for in situ monitoring.

The current level of monitoring **of river levels and flows** undertaken globally is broadly sufficient to meet Copernicus product requirements, though access to data is inadequate in many areas.

For **river water quality**, the developing requirement for in situ data is for coastal zone monitoring / modelling for European waters, where European monitoring should be able to provide adequate data, although access to real-time or near real-time data will always be challenging for measurements requiring ex-situ laboratory analysis. Monitoring in other countries bordering on these waters may not provide the in situ data required (e.g North Africa, Arctic). Further

## HYDROLOGY IN SITU DATA REQUIREMENTS AND AVAILABILITY

development of global products would be likely to be significantly limited by the availability of in situ water quality monitoring data.

Globally, little of the **lake water quality** monitoring activity currently undertaken produces measurements which are suitable for validation of satellite products (which are currently focussed on turbidity, trophic state index / phytoplankton / chlorophyll, and reflectance / colour). Validation activities at present focus on a small set of lakes and do not adequately represent the range of lake types or conditions. New measurements are required that focus on delivering consistent in situ data for satellite validation and calibration, and that are more widespread. UAVs have potential to produce very useful datasets, in particular for improving processes for the removal of atmospheric effects. With some potentially minor changes (though not without practical implications), such as sampling at distance from the lake shore, standard and widespread water quality measurements, e.g. of turbidity and chlorophyll concentration, would be usable for satellite data comparability. There is an opportunity to use the expertise within the Copernicus services to define detailed requirements in this area, and to coordinate the development of monitoring protocols. The requirements for in situ data should be summarised and strongly promoted through appropriate avenues (e.g. international initiatives such as GEO Aquawatch) to highlight this need and to drive change in monitoring practices by national statutory agencies.

**Lake water temperature** product development / analysis is undertaken separately from other aspects of lake water quality. Standard monitoring approaches, ongoing in most countries, should be able to provide the data required. An increase in measurements away from lake shores would be beneficial. Activities collating global datasets should be supported to improve access to this in situ data and avoid ad hoc and repeated data collation activities.

The number of available **soil moisture** measurements is well below the level required for adequate validation or calibration of satellite-based soil moisture products. Point soil moisture measurements do not adequately represent the spatial variability of soil moisture over the scales of current satellite footprints; more field scale measurements, or arrays of point measurements, are required. Spatially extensive monitoring of soil moisture, i.e. across multiple satellite pixels, is required to provide data suitable for comparison with satellite products. Measurements that represent a wider range of land cover and soil types are needed. Longer term soil moisture monitoring programmes are required. Where possible soil moisture measurements should be undertaken with the specific aim of comparability with satellite-derived data. The Copernicus In Situ programme should summarise these requirements, as well as the potential for satellite soil moisture measurements, and identify avenues through which to promote the need for better in situ soil moisture measurements, supporting existing activities collating global soil moisture measurements.

### Data standards and exchange technologies

Where in situ measurements exist, many of the limitations to its use are around data exchange technologies and standards. National services do not have the capacity, or the incentive, to deliver datasets that are consistent in either content or format. Technology developments are rapid and approaches to data sharing evolve constantly, but it is not the inconsistent use of technology that is the current limiting factor in many areas. Many monitoring agencies simply do not make data openly available, and in many cases (e.g. water quality monitoring) the metadata provided with in situ measurements limits their utility. Copernicus in situ should help to clarify data and metadata

requirements for in situ hydrological data and promote these requirements to influence the evolution of environmental data sharing initiatives, both at the European and Global scales.

In some areas data sharing via the internet has increased hugely in recent years, e.g. of river levels and flows from National Hydrological Services. However the range of techniques for publishing data means the collation of data is not feasible. In these cases there would be benefits in promoting technologies and approaches for more consistent data sharing, moving towards standards-based web services, and more importantly the standardised content and metadata, that could underpin a global federated hydrological data system. Copernicus In Situ should support global initiatives such as the WMO WHOS programme which are already working towards this same aim. In the short term an assessment should be made of the level of hydrological in situ data, in particular levels and flows, available from National Hydrological Services globally.

### **Data collation and coordination**

Global databases exist for many of the types of measurements reviewed. However these were generally set up before the satellite era and for different purposes than satellite product development, and the contents and data provision are not ideally suited to this purpose. In particular real time data are generally not covered by these databases, and for many the historic data periods are not focussed on the satellite era. There is a strong case for these data centres and databases to further evolve and meet these modern requirements.

For flow, and to some extent, level, there is a good database of European data, including real time data, that is well curated and supported by Copernicus funding: the CEMS Hydrological Data Collection Centre (HDCC). Other Copernicus initiatives are starting to develop their own databases which may ultimately lead to unnecessary or duplicated effort. CEMS are happy for HDCC data to be used across other services, though there is likely to be a need for agreement about mechanisms for this, and how a shared data a service might operate in the longer term. The wider use of this data by services should be investigated. Specific recommended activities include working with CEMS to standardise licence arrangements with data providers, disseminating information on the data available within the HDCC to other Copernicus services, and liaison between CEMS and HDCC and potential users to identify how sharing arrangements might function in future.

Outside of this there is little adequate data provision and in order to make use of in situ data Copernicus services will need to either spend effort on collating data sources, or to make use of existing global initiatives. In most cases it is likely that the latter could provide a more effective solution. Capabilities and approaches within national services are often the major limiting factor to better data availability. As described above, for hydrometric data (rivers levels and flows, lake levels) the WMO WHOS initiative offers an avenue to develop national capacity and consistent approaches. Copernicus should actively support the work of WHOS in achieving this. Many national services offer some sort of data provision. This should be documented in the short term and the information used to provide an understanding of future potential.

There is a need to engage with global data centres (e.g. GTN-H components focused on river flows, lake water quality and level, river water quality, and soil moisture) both to support their work, and to understand how the data availability could be improved for the purposes of satellite data product development. This includes promotion of the requirements for real time data. As more agencies provide data online, the focus of these global databases may change (e.g. to discoverability, curation, QA, dissemination, and archive of data from federated services) and Copernicus In Situ

## HYDROLOGY IN SITU DATA REQUIREMENTS AND AVAILABILITY

should help to develop a vision for their future operation and use, specifically how such a model could support Copernicus satellite products.

Other databases could provide useful data with some initial effort. For example the EEA Waterbase may contain the in situ water quality data required by services but access to this data is not currently sufficiently simple. Copernicus In Situ could undertake a more detailed survey of such data sources, matching data available to the requirements of services, to assess whether the spatial and temporal resolution of sampling is adequate for satellite product development, and promote the need for more timely and efficient access to datasets to support its work.

### Concluding remarks

A future of hydrological monitoring using high quality satellite-derived data will continue to require extensive and high quality in situ monitoring to support not just the development, but the ongoing production and validation of satellite outputs. Current in situ monitoring must evolve to underpin satellite based products in order to realise the potential benefits of widespread and real-time monitoring. A number of short term coordination activities can improve the immediate access to in situ data across the Copernicus services. But to achieve the long term benefits there is a need to properly define the long-term vision for in situ hydrological monitoring for Copernicus products, based on detailed requirements, and to promote this vision at the European and global level to help the evolution of monitoring approaches, national level data exchange, and sustainable global data collections.

## REFERENCES

Note: CLMS-GL reports may not be available at the links below where reports have been updated and version numbers have changed. All documents can be found from the “documents” tab for each product listed here: <https://land.copernicus.eu/global/products>

C3S. 2018 a. Product User Guide and Specification. Soil Moisture. V2.3. C3S\_D312a\_Lot7.3.3.1\_201801\_PUG\_v2.3.docx. <http://datastore.copernicus-climate.eu/c3s/published-forms/c3sprod/satellite-soil-moisture/product-user-guide-v2.3.pdf>

C3S. 2018 b. Product Quality Assessment Report. Soil Moisture. V2.3. C3S\_D312a\_Lot7.3.1.2\_201711\_PQAR\_v1.1.docx. <http://datastore.copernicus-climate.eu/c3s/published-forms/c3sprod/satellite-soil-moisture/product-quality-assessment-report-v1.1.pdf>

C3S. 2018 c. Gap Analysis Document. Soil Moisture. V2.3. D312a\_Lot7.1.1.4\_2017\_GAD\_V1.0.docx. <http://datastore.copernicus-climate.eu/c3s/published-forms/c3sprod/satellite-soil-moisture/gap-analysis-v1.0.pdf>

CEMS. 2019. Analysis of the CEMS hydrological data collection, Annual Report 2018. CEMS Hydrological Data Collection Centre. [http://www.efas.eu/sites/default/files/2019-06/AnnualReport2018\\_compressed.pdf](http://www.efas.eu/sites/default/files/2019-06/AnnualReport2018_compressed.pdf)

CGLOPS-1. 2018 a. Copernicus Global Land Operations, “Vegetation and Energy”, “CGLOPS-1”, Validation Report, Surface Soil Moisture, Collection 1km, Version 1. [https://land.copernicus.eu/global/sites/cgls.vito.be/files/products/CGLOPS1\\_VR\\_SSM1km-V1\\_I1.20.pdf](https://land.copernicus.eu/global/sites/cgls.vito.be/files/products/CGLOPS1_VR_SSM1km-V1_I1.20.pdf)

CGLOPS-1. 2018 b. Copernicus Global Land Operations, “Vegetation and Energy”, “CGLOPS-1”, Product User Manual, Surface Soil Moisture, Collection 1km, Version 1. [https://land.copernicus.eu/global/sites/cgls.vito.be/files/products/CGLOPS1\\_PUM\\_SSM1km-V1\\_I1.30.pdf](https://land.copernicus.eu/global/sites/cgls.vito.be/files/products/CGLOPS1_PUM_SSM1km-V1_I1.30.pdf)

CGLOPS-2. 2018 a. Copernicus Global Land Operations, “Cryosphere and Water”, “CGLOPS-2”, Product User Manual, Lake Waters 300m and 1km products, Version 1.2.0. [https://land.copernicus.eu/global/sites/cgls.vito.be/files/products/CGLOPS2\\_PUM\\_LWQ300\\_1km\\_v1.2.0\\_I1.06.pdf](https://land.copernicus.eu/global/sites/cgls.vito.be/files/products/CGLOPS2_PUM_LWQ300_1km_v1.2.0_I1.06.pdf)

CGLOPS-2. 2017. Copernicus Global Land Operations, “Cryosphere and Water”, “CGLOPS-2”, Quality Assessment Report, Water Bodies, Collection 300m, Version 1. [https://land.copernicus.eu/global/sites/cgls.vito.be/files/products/CGLOPS2\\_QAR\\_WB300m-V1\\_I1.00.pdf](https://land.copernicus.eu/global/sites/cgls.vito.be/files/products/CGLOPS2_QAR_WB300m-V1_I1.00.pdf)

CGLOPS-2. 2018 b. Copernicus Global Land Operations, “Cryosphere and Water”, “CGLOPS-2”, Quality Assessment Report, Lake Waters 300m and 1km products, Version 1.2.0. [https://land.copernicus.eu/global/sites/cgls.vito.be/files/products/CGLOPS2\\_QAR\\_LWQ300\\_1km\\_v1.2.0\\_I1.07.pdf](https://land.copernicus.eu/global/sites/cgls.vito.be/files/products/CGLOPS2_QAR_LWQ300_1km_v1.2.0_I1.07.pdf)

## HYDROLOGY IN SITU DATA REQUIREMENTS AND AVAILABILITY

CGLOPS-2. 2018 c. Copernicus Global Land Operations, "Cryosphere and Water", "CGLOPS-2", Product User Manual, Lake Surface Water Temperature, 1km products, Version 0.1.0. [https://land.copernicus.eu/global/sites/cgls.vito.be/files/products/CGLOPS\\_PUM\\_LSWT\\_1km\\_v0.1.0\\_I1.02.pdf](https://land.copernicus.eu/global/sites/cgls.vito.be/files/products/CGLOPS_PUM_LSWT_1km_v0.1.0_I1.02.pdf)

CGLOPS-2. 2018 d. Copernicus Global Land Operations, "Cryosphere and Water", "CGLOPS-2", Quality Assessment Report, Lake Surface Water Temperature, 1km products, Version 0.1.0. [https://land.copernicus.eu/global/sites/cgls.vito.be/files/products/CGLOPS\\_QAR\\_LSWT\\_1km\\_v0.1.0\\_I1.02.pdf](https://land.copernicus.eu/global/sites/cgls.vito.be/files/products/CGLOPS_QAR_LSWT_1km_v0.1.0_I1.02.pdf)

CGLOPS-2. 2018 e. Copernicus Global Land Operations, "Cryosphere and Water", "CGLOPS-2", Quality Assessment Report, Lake and River Water Level, 300m, Version 2. [https://land.copernicus.eu/global/sites/cgls.vito.be/files/products/CGLOPS2\\_PUM\\_LakeAndRiverWaterLevel-V2\\_I2.03.pdf](https://land.copernicus.eu/global/sites/cgls.vito.be/files/products/CGLOPS2_PUM_LakeAndRiverWaterLevel-V2_I2.03.pdf)

CGLOPS-2. 2018 f. Copernicus Global Land Operations, "Cryosphere and Water", "CGLOPS-2", Product User Manual, Water Bodies, PROBA-V 300m, Version 1. [https://land.copernicus.eu/global/sites/cgls.vito.be/files/products/CGLOPS2\\_PUM\\_WB300m\\_V1\\_I1.20.pdf](https://land.copernicus.eu/global/sites/cgls.vito.be/files/products/CGLOPS2_PUM_WB300m_V1_I1.20.pdf)

Copernicus In Situ. 2018. State of Play reports. <https://insitu.copernicus.eu/state-of-play>

L. Arnal, S.-S. Asp, C. Baugh, A. de Roo, J. Disperati, F. Dottori, R. Garcia, M. Garcia-Padilla, E. Gelati, G. Gomes, M. Kalas, B. Krzeminski, M. Latini, V. Lorini, C. Mazzetti, M. Mikulickova, D. Muraro, C. Prudhomme, A. Rauthe-Schoch, K. Rehfeldt, P. Salamon, C. Schweim, J.O. Skoien, P. Smith, E. Sprokkereef, V. Thiemig, F. Wetterhall, M. Ziese, *EFAS upgrade for the extended model domain – technical documentation*, EUR 29323 EN, Publications Office of the European Union, Luxembourg, 2019, ISBN 978-92-79-92881-9, doi:10.2760/806324, JRC111610.

A. Gruber, W.A. Dorigo, S. Zwieback, A. Xaver, and W. Wagner. 2013. Characterizing coarse-scale representativeness of in situ soil moisture measurements from the International Soil Moisture Network. *Vadose Zone J.* 12. (2)

Hirpa, F. A., Salamon, P., Beck, H. E., Lorini, V., Alfieri, A., Zsoter, E., Dadson, S. J. 2018. Calibration of the Global Flood Awareness System (GloFAS) using daily streamflow data. *Journal of Hydrology.* 566. Pp 595-606. <https://doi.org/10.1016/j.jhydrol.2018.09.052>

Pekel, J.-F., Cottam, A., Gorelick, N, Belward, A., High resolution mapping of the global surface water and its long-term changes, *Nature*, 540, 418-422, 2016

Sharma S., D. Gray, J. Read, C. Oreilly, P. Schneider, A. Qudrat, C. Gries, S. Stefanoff, S. Hampton, S. Hook, J. Lenters, D. Livingstone, P. McIntyre, R. Adrian, M. Allan, O. Anneville, L. Arvola, J. Austin, J. Bailey, J. Baron, J. Brookes, Y. Chen, R. Daly, M. Dokulil, B. Dong, K. Ewing, E. de Eyto, D. Hamilton, K. Havens, S. Haydon, H. Hetzenauer, J. Heneberry, A. Hetherington, S. Higgins, E. Hixcon, L. Izmet'eva, B. Jones, K. Kangur, P. Kasprzak, O. Koster, B. Kraemer, M. Kumagai, E. Kuusisto, G. Leshkevich, L. May, S. MacIntyre, D. Mueller-Navarra, M. Naumenko, P. Noges, T. Noges, P. Niederhauser, R. North, A. Paterson, P. Plisnier, A. Rigosi, A. Rimmer, M. Rogora, L. Rudstram, J. Rusak, N. Salmaso, N. Samal, D. Schindler, G. Schladow, S. Schmidt, T. Schultz, E. Silow, D. Straile, K. Teubner, P. Verburg, A. Voutilainen, A. Watkinson, G. Weyhenmeyer, C.

Williamson, K. Woo. 2014. Globally distributed lake surface water temperatures collected in situ and by satellites; 1985-2009. Environmental Data Initiative. <https://doi.org/10.6073/pasta/379a6cebee50119df2575c469aba19c5>.

WMO. 2016. World Meteorological Organisation Commission for Hydrology, Fifteenth Session. Abridged final report with resolutions and recommendations. WMO-No. 1184. [https://library.wmo.int/doc\\_num.php?explnum\\_id=3404](https://library.wmo.int/doc_num.php?explnum_id=3404)

## APPENDIX A. CMEMS MFC REQUIREMENTS

River discharge locations required by CMEMS Marine Forecasting Centres:

**Mediterranean Sea MFC:** Po, Ebro, Nile, Rhone, Vjosë, Seman, Buna/Bojana, Piave, Tagliamento, Soca/Isonzo, Livenza, Brenta-Bacchiglione, Adige, Lika, Reno, Krka, Arno, Nerveta, Aude, Trebisjnica, Tevere/Tiber, Mati, Volturno, Shkumbini, Struma/Strymonas, Meric/Evros/Maritsa, Axios/Vadar, Arachtos, Pinios, Acheloos, Gediz, Buyuk Menderes, Kopru, Manavgat, Seyhan, Ceyhan, Gosku, Medjerda, Asi/Orontes

**Black Sea MFC:** Danube, Batova, Kamchiya, Khandzhiliska, Repotamo, Rezovo, Riva, Sakarya, Melet, Yenice, Koca-Irmak, Kokacal, Aidos, Kizil-Irmak, Samsun, Yesil-Irmak, Balami, Aksu, Yagilidere, Harsit, Degermen, Karadere, Kolopotamos, Chorokhi, Rioni, Kodori, Apasta, Bzyb, Sochi, Demerdzhi, Dereolka, Kacha, Alma, Eupatoria, Dnieper, South Bug, Dneister, Lake Alibey, Lake Sasyk, Danube, Dyursi, Pshada, Shapsukho, Shakhe, Red River, Don.

**Baltic Sea MFC:** Neva, Vistula and Orde rivers are the largest of the several hundred inflowing rivers.

**Arctic MFC:** Ob, Lena, Yenissei, Mackenzie, Yukon, Kolyma

**Atlantic, Iberian, Biscay, Irish Seas MFC:** Rhine, Rhone, Garonne, Loire, Elbe, Glama, Seine, Duero, Vaenern-Goeta, Ebro, Tejo, Minho, Meuse, Weser, Adour, Shannon, Guadalquivir, Tay, Guadiana, Otra, Sebou, Barrow, Mondego, Corrib, Lower Bann, Erne, Trent, Thames, Ems, Severn, Vilaine, Spey.

**Atlantic European North West shelves MFC:** All the UK, Norway, South North Sea, French-Spanish Portuguese Atlantic coast rivers.

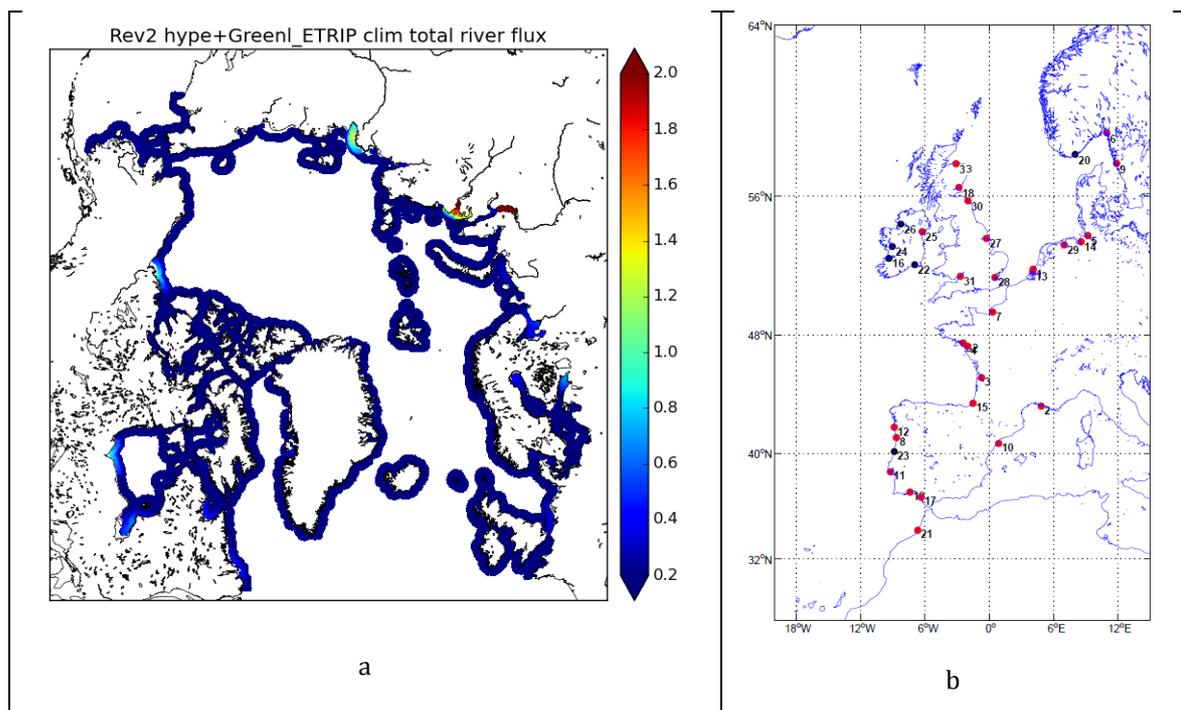


Figure 12 Spatial areas requiring river discharge information for a) Arctic MFC, b) Atlantic, Iberian, Biscay, Irish Seas MFC

River water quality inputs required by CMEMS Marine Forecasting Centres

**Arctic MFC:** Nutrients (nitrate, silicate, phosphate, dissolved inorganic carbon, alkalinity).

**Atlantic, Iberian, Biscay, Irish Seas MFC:** Nutrients.

**Mediterranean Sea MFC:** transport, salinity, temperature, nutrient load.

**Atlantic European North West shelves MFC:** Nutrients (dissolved inorganic nitrogen and phosphorus are the most important, and we would also like silicate if possible. We can also use particulate and organic N, P if available; also dissolved oxygen), Dissolved inorganic carbon and bicarbonate.

**Black Sea MFC:** Nutrient discharge (nitrate, phosphate, ammonium, silicate)

**Baltic Sea MFC:** Nutrient load