Norwegian institute for
water research

7920-2023

# Satellite water quality monitoring

## Status report 2023

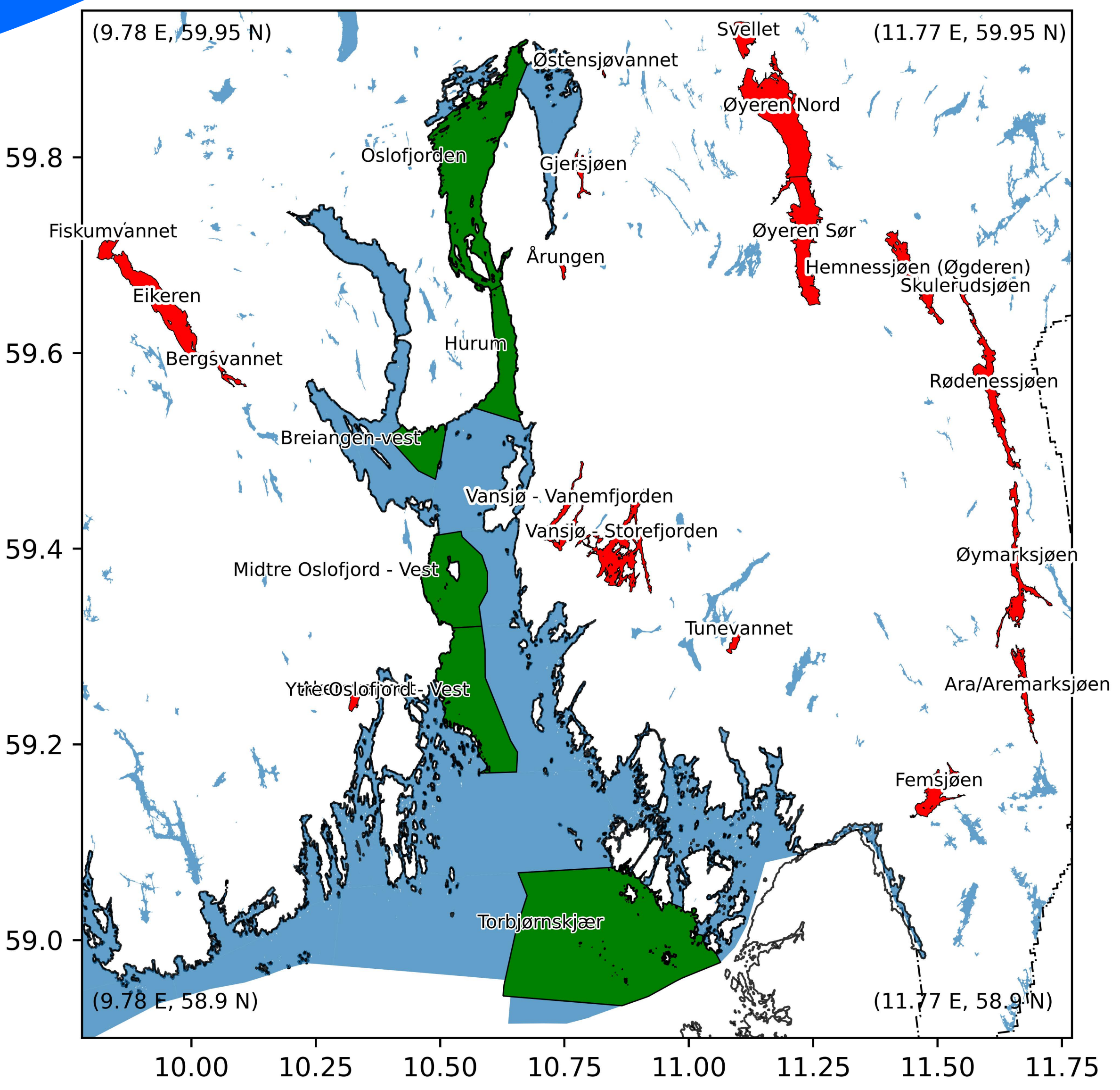


Foto: Pipatthra Saesin, NIVA

# Report

## Norwegian Institute for Water Research

Serial no: 7920-2023

ISBN 978-82-577-7656-5
NIVA report
ISSN 1894-7948

This report has been quality assured according to NIVA's quality system and has been approved by:

Therese Harvey
Project Manager/Lead Author

Andrew Luke King
Quality Assurer

Åse Åtland
Research Manager



www.niva.no

| Title | Pages | Date |
| --- | --- | --- |
| Satellite water quality monitoring: Status report 2023 / Vannovervåking med satellitt: Status rapport 2023 | 24 | 20.12.2023 |



| Authors | Topic group | Distribution |
| --- | --- | --- |
| Therese Harvey, Sivert Bakken, Carsten Brockmann, Martin Böttcher, Anette Engesmo, Sigrid Haande, Carole Lebreton, Petra Philipson, Pipatthra Saesin, Kerstin Stelzer, Kai Sørensen, Jan-Erik Thrane | Monitoring | Open |



| Client(s) | Client's contact person |
| --- | --- |
| Miljødirektoratet | Steinar Sandøy |

**Published by NIVA**
220242

**Abstract**
This report covers the 1-year status report of the project Vannovervåking med satellitt 2023 – 2024 (2025)/ Satellite water quality monitoring 2023 – 2024 (2025) funded by Miljødirektoratet/the Norwegian Environmental Agency (NEA). It describes the work done in Work package 1, 3, 5 and partly 6 as well as plans for 2024. In situ data have been gathered for the most monitored Norwegian lakes (for ØKOSTOR and ØKOFERSK) and the coastal stations included in ØKOKYST. Both Sentinel-2 and Sentinel-3 data have been processed between 2016-2023. A demo viewer for exploring the data for some dedicated lakes and coastal areas has been set up with granted access to project members. The infrastructure have been set up for the initial validation of 10+10 water bodies (WB) as well as tests with MET and EUMETSAT for the future host of the service have been tested. The validation and evaluation of different algorithms of chl-a and Secchi depth data have started for the 10+10 WBs. The selection of WB for the initial validation was based on water quality and optical parameters to cover a wide range of water types and conditions.

**Keywords:** Water Quality, Remote sensing, Water Framework Directive, Classification assessment
**Emneord:** Vannkvalitet, Fjernmåling, Vann Direktivet, Klassifikasjon

# Table of contents

# 1 Introduction

This report covers the 1-year status report of the project Vannovervåking med satellitt 2023 – 2024 (2025)/ Satellite water quality monitoring 2023 – 2024 (2025) funded by Miljødirektoratet/the Norwegian Environmental Agency (NEA) (Avtalenummer:23087474).

## 1.1 Background

There is a need for a complementary dataset of water quality, in addition to the limited number of samples collected in situ, that can be used for deriving water quality status classification for the EU Water Framework Directive (WFD) according to Vannforskriften for Norwegian lakes (Direktoratsgruppen, 2018) for water bodies in both lakes and coastal waters.

Phytoplankton biomass, or the use of chlorophyll a (chl-a) as a proxy, is one of four biological quality elements used for classification of ecological status in lakes and coastal areas. In lakes or coastal water bodies that are monitored, the ecological status of chl-a is currently classified on the basis of monthly, biweekly or weekly measurements during the growing season. This is done either at a single station on the deepest part of the lake or at a few stations within the WB (WB). However, the concentration of chl-a varies widely over time, and there can also be great variation within the lake or the WB. Thus, the current sampling practice as described above is comparably sparse in time and space. The use of remote sensing (RS) to determine chl-a can potentially contribute to a more comprehensive determination of the ecological status as the temporal and geographical coverage is higher. In addition, only a few Norwegian lakes are currently part of a monitoring program where chl-a concentration is measured. In the non-monitored water bodies, the classification is done with expert assessment based on similar ones and where the assessment of impact type (påvirkningstype) and the degree of impact is important. The use of satellite RS data with a tested and validated method, will greatly increase the availability of chl-a monitoring data and increase the reliability of the classification of ecological status in lakes.

With the EU´s Copernicus program and the two twin-satellite sets of Sentinel-2 A and B (S2), and Sentinel-3 A and B (S3), the possibilities for satellite-based monitoring have never been better. Earth Observation from the S3 sensor OLCI (Ocean and Land Colour Instrument, 300m spatial resolution) occurs twice per day in Norway and every 2-3 days from the S2 sensor MSI (Multi-Spectral Instrument (MSI), 60m spatial resolution). However, the data retrieval is limited by cloud coverage. The spectral properties and resolution of OLCI are adapted to water applications with 21 spectral bands between 400-1025 nm. MSI has 10 broader bands between 492-1376 nm. Satellite-based water quality products provide a valuable supplementary source for water quality information with high spatial coverage and temporal resolution.

HYPSO-1 is a Norwegian experimental CubeSat with a primary payload capable of acquiring optical hyperspectral images. These hyperspectral images will be able to resolve spectral features in the range from 440 to 800 nm in 120 distinct spectral bands with a bandwidth ranging from 3 to 8 nm. This wavelength range is used to infer biological and chemical parameters related to water quality monitoring.

Relevant products for the WFD classifications, which can be derived from S2 or S3, are e.g. seasonal or monthly averages of chl-a and Secchi depth. Data is available from 2015 (S2) and 2016 (S3) and thereby covers both the previous and the upcoming WFD assessment periods. RS also has the potential for increased monitoring on other parameters that are important for aquatic ecology and can be used to assess the effects of climate change and the impacts of, for example, agricultural inputs, the spring flood, storm events and run-off. This applies to the concentration of total suspended matter (TSM) in water and coloured dissolved organic matter (CDOM) as well as changes in light conditions within the water column.

## 1.2 Objectives

The main objective of this project is to develop an operational system for the use of RS for monitoring and classification of lakes and coastal waters. Satellite RS data can then be used as a complementary method for monitoring and assessing the ecological status in Norwegian lakes and coastal water bodies. The satellite data aims to support the conventional monitoring programmes for lakes, i.e. ØKOSTOR and to some extent ØKOFERSK, and for coastal waters via ØKOKYST and the Ocean acidification programme (Havforsuring). An additional objective of this project is to develop and improve the methods for using satellite data to calculate the concentration of chl-a and other water quality parameters, such as Secchi depth, particles (i.e. TSM), light attenuation (Kd) and turbidity.

The goal of the project is to demonstrate a supplementary method that can complement the in situ data with satellite RS data in water bodies that are currently not monitored so that as many lakes and coastal water bodies as possible can be classified.

Furthermore, a system for operational and quality assured data processing, operated on an IT-platform that can be hosted on a private or public cloud, and that can be used for ecological status classification will be developed with the goal of implementation during the assessment period 2025-2027. The overall system will be composed of the software tools that allow operational processing of multi-sensor satellite data, hosted on a cloud platform, and running the best possible algorithms to derive above water quality parameters. It shall be in place to be used for ESA (EFTA Surveillance Authority) reporting in March 2028. The project consists of 6 Work packages (Figure 2.1) for which WP2 and WP4 have been postponed and WP6 pushed forward to 2023.

# 2 Work package 1 - Algorithms and performance

Work package 1 has tested the applicability of the S-3 and S-2 processing schemes and algorithms to Norwegian water bodies. The work for creating a framework for selecting and adapting algorithms in different water types has started and will continue in 2024. WP1 provides the basis for WP3.

## 2.1 In situ data

In situ data has been compiled between 2016-2023 for 76 coastal stations and 27 lake stations. Parameters included are chl-a, salinity, Secchi depth, turbidity, CDOM, TSM, colour number for lakes and WB size. The amount of in situ data within a WB was important for selection of the water bodies to start the evaluation of RS products, so that there would be enough in situ samples for comparisons.

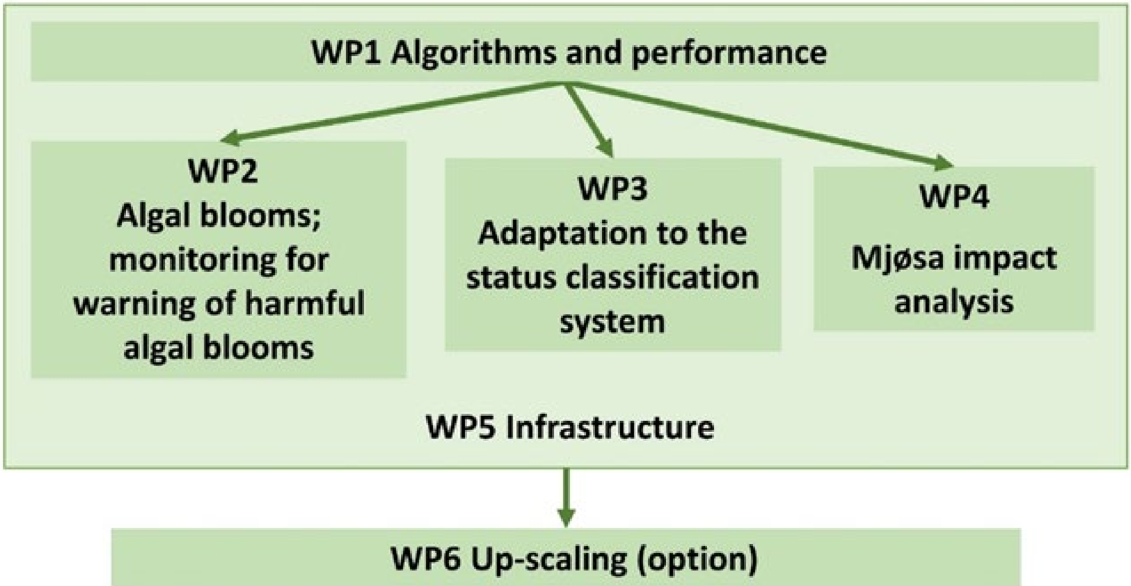


*Figure 2.1 Work packages and the structure of the project.*

### 2.1.1. Coastal water bodies

Biological and physical data, for the water bodies that are included in the ØKOKYST monitoring program, were used to select 10 water bodies for the first evaluation round. The size of the water bodies was taken into consideration as large water bodies require more than 1 scene of Sentinel-2 data to be covered, which means more data to be processed and stored. Therefore, the largest water bodies were not selected in 2023. It is also important that the average depth is deep enough so that bottom reflectance doesn't disturb the data retrieval. The water bodies' characteristics were plotted, and an initial selection was done to cover a wide range of water types and fora representative selection of different water types (Figure 2.2 and Figure 2.3).

The in situ data for 76 monitoring stations was gathered from AquaMonitor (NIVAs data base and from Vannmiljø (NEAs database). When compiling the dataset for comparison with RS data, many issues were discovered with the data as we were missing certain depth, parameters or years where there should be data etc. Therefore, a coastal data set was re-compiled in collaboration with other current NIVA/NEA projects focusing on the same issue (which will be ready in mid Dec). The naming convention of the parameters and stations are not coherent, and a massive data cleaning effort was needed. This work is tedious and takes a lot of time, but it is crucial to have a clean data set that can be used for validating the RS algorithms, presented in Ch. 2.4.

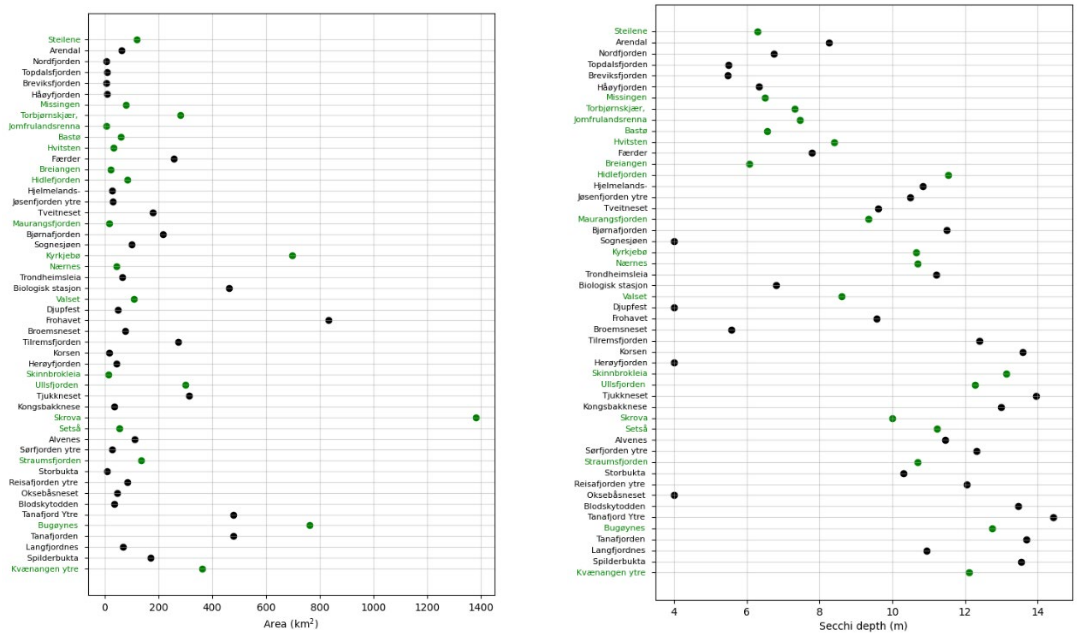


*Figure 2.2 Selection of coastal water bodies based on size (left) and Secchi depth (right). The green marks include the first selection. The list on the y-axis represents the different names of the coastal water bodies, where the green was the first selection of which 10 WBs were further selected.*

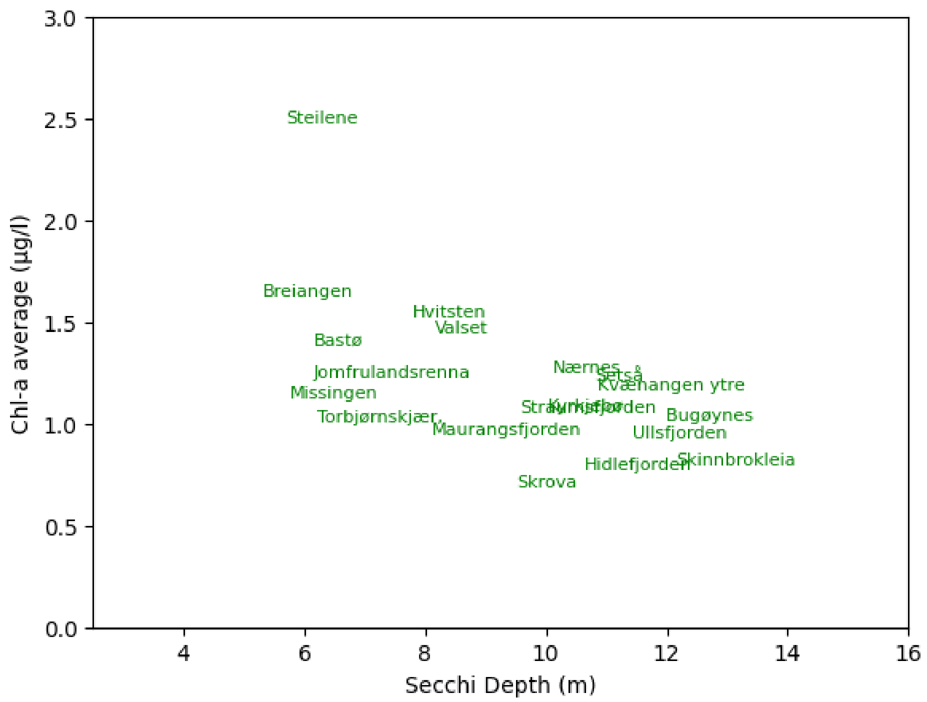


*Figure 2.3 Selection of coastal water bodies on the relationship between chl-a conc. and Secchi depth.*

### 2.1.2. Lake water bodies

Biological and physical data and placement of stations, for the water bodies that are included in the ØKOSTOR, ØKOFERSK and regional monitoring program, were used to select 10 water bodies for the first evaluation round during the reporting period. As many of the Norwegian lakes are clear with low chl-a concentrations, which can be problematic for the satellites to retrieve data from, it was decided to focus on more eutrophic lakes as a starting point. The relationship between the different parameters were used as a basis for collection (Figure 2.4 and Figure 2.5). Further, lakes in Southern Norway were selected to ensure a synergy between in situ sampling and RS data collection within this project and the NTNU lead project for HYPSO-1, financed by the Norwegian Space Center. Water quality data for 27 lake stations between 2016-2023 were extracted and compiled from AquaMonitor and Vannmiljø, the data set was cleaned, and quality assured and used for the validation in Ch. 2.4.

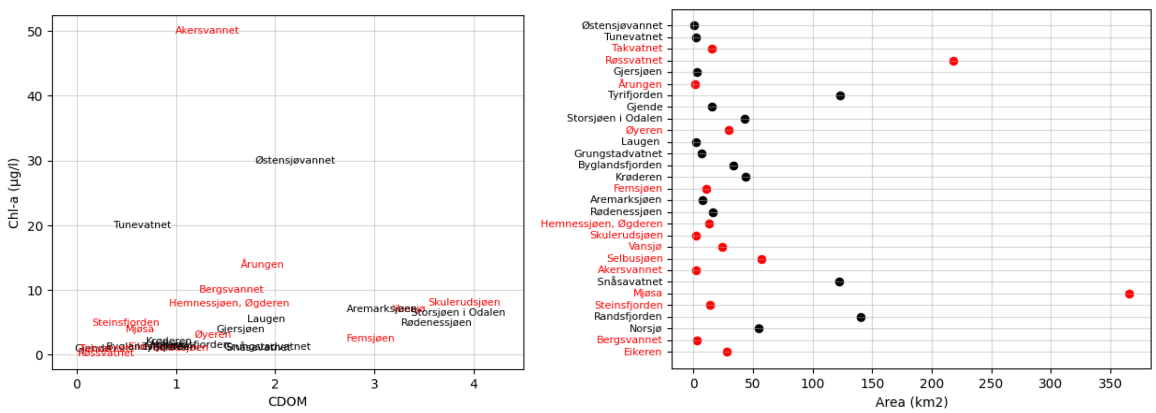


*Figure 2.4 Example of selection of lake water bodies based on the relationship between average chl-a and average CDOM (re-calculated from the colour number) (left) and area (right).*

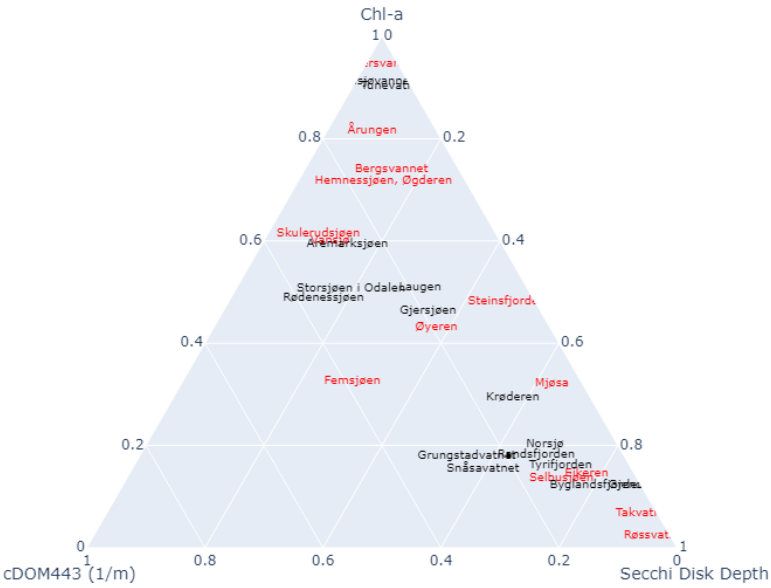


*Figure 2.5 Ternary plot showing the relationship between key parameters, chl-a concentration, Secchi depth and CDOM, of the lake water bodies, used as a basis for a representative selection. Note that the values are normalised between 0-1 for easier comparison.*

## 2.2 Included water bodies

Tables 1 and 2 show the coastal and lake water bodies that were used for the initial evaluation during the project period. Some of the coastal water bodies were exchanged due to the availability of the in situ data. The map in Figure 2.6 show examples of the coastal and lake water bodies included.

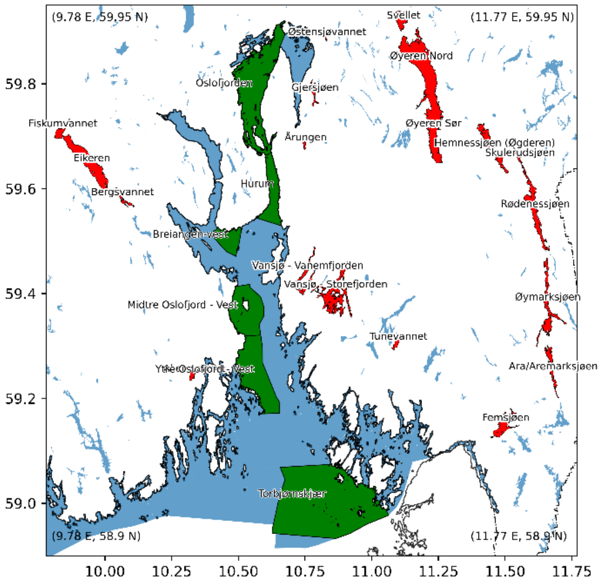


*Figure 2.6 Map showing some of the selected coastal (green) and lake (red) water bodies in Southern Norway.*

**Table 1. Selected coastal water bodies for the project period.**

| | Waterbody (WB) | WB ID | Watertype | Area | Station | Station code | Class chl-a |
|---|---|---|---|---|---|---|---|
| 1 | **Midtre Oslofjord - Vest** | 0101020200-2-C | Moderat eksponert kyst (S2) | Horten - Larvik | Basto | VT2/OF2 | **High** |
| 2 | **Ytre Oslofjord – Øst** | 0101020101-2-C | Moderat eksponert kyst (S2) | Horten - Larvik | Missingen | VT65 | **High** |
| 3 | **Breiangen-vest** | 0101021000-1-C | Beskyttet kyst/ fjord (S3) | Indre Oslofjord Vest | Breiangen | VT10/OF5 | **Good** |
| 4 | **Hurum** | 0101020500-C | Beskyttet kyst/ fjord (S3) | Indre Oslofjord Vest | Hvitsten, Elle | VT4/OF4 | **High** |
| 5 | **Torbjørnskjær** | 0101000030-1-C | Åpen eksponert kyst (S1) | Glomma sør for Øyeren | Torbjornskjear | VT3/OF1 | **High** |
| 6 | **Skrurenna** | 0110000035-C | Åpen eksponert kyst (S1) | Kragerøvass-draget | Jomfrulandsrenna | VT68 | **High** |
| 7 | **Mossesundet-ytre** | 0101020400-3-C | Beskyttet kyst/ fjord (S3) | Innlandet og Viken | Kippenes | MO-2 | **Good** |
| 8 | **Langesunds-fjorden** | 0110010801-C | Beskyttet kyst/ fjord (S3) | Vestfold og Telemark | Spilderbukta | VT67 | **High** |
| 9 | **Færder** | 0101000030-2-C | Åpen eksponert kyst (S1) | Horten - Larvik | Færder | VT84 | **High** |

Table 2. Selected lake water bodies for the project period.

| | Waterbody (WB) | WB ID | Watertype | Code | Colour (mgPt-l) | Class Chl-a |
|---|---|---|---|---|---|---|
| 1 | **Eikeren** | 012-542-2-L | Stor, kalkfattig, klar (TOC2-5) | L105b LEL32113 | 13 | High |
| 2 | **Bergsvannet** | 012-519-L | Middels, moderat kalkrik, klar (TOC2-5) | L107 LEL23112 | 28 | Good |
| 3 | **Steinsfjorden** | 012-522-1-L | Stor, kalkrik, klar (TOC2-5) | L109 LEL34112 | 9 | Good |
| 4 | **Mjøsa** | 002-118-1-L | Svært stor, moderat kalkrik, klar (TOC2-5) | L107 LEL43113 | 12 | High |
| 5 | **Akersvannet** | 014-314-L | Middels, moderat kalkrik, klar (TOC2-5) | L107 LEL23112 | 24 | Very Bad |
| 6a | **Vansjø - Storefjorden** | 003-291-2-L | Stor, kalkfattig, humøs | L106 LEL32212 | 52 | Good |
| 6b | **Vansjø - Vanemfjorden** | 003-291-1-L | Stor, kalkfattig, humøs | L108 LEL33212 | na | Moderate |
| 7 | **Hemnessjøen, Øgderen** | 001-327-L | Stor, kalkfattig, humøs | L1068 LEL33212 | 28 | Good |
| 8 | **Femsjøen** | 001-316-L | Stor, kalkfattig, humøs | L106 LEL32212 | 58 | High |
| 9 | **Skulerudsjøen** | 001-324-L | Middels, moderat kalkrik, humøs | L108 LEL23212 | 71 | Good |
| 10 | **Øyeren Sør** | 002-113-2-L | Svært stor, moderat kalkrik, klar (TOC2-5) | L107 LEL43113 | 25 | High |

## 2.3 Remote sensing data

### 2.3.1. Sentinel data

The water quality information assessed during the reporting period is based on satellite data from Sentinel-3 OLCI (Ocean and Land Color Instrument) and Sentinel-2 MSI (Multispectral Imager). Sentinel-3 and Sentinel-2 are two of the satellites in the European space program Copernicus. For both types of Sentinel, two identical satellites (3a/3b and 2a/2b) are in orbit, both collecting information.

Sentinel-3a data is available from 2016 and Sentinel-3b from 2018. OLCI has a ground resolution of 300 meters and spectral characteristics adapted for water applications. The satellite data cover Norway on a daily basis. In theory, this means that one can get two estimates of, for example, the chl concentration for a body of water or for a position every day. In Norway, however, clouds are a limiting factor. The satellite covers large areas at the same time and data for the whole of Norway is collected during a short time period.

Sentinel-2a is available from 2015 and Sentinel-2b from 2017. The MSI sensor has a ground resolution of 10-60 meters depending on the colour band. The satellites go in a polar orbit around the Earth and pass over the same area 2-3 times per week. As with Sentinel-3, however, cloud is a limiting factor. Sentinel-2 was developed primarily for land applications, but good results have also been demonstrated for aquatic environments.

### 2.3.2. HYPSO-1

The HYPSO-1 satellite, a 6U CubeSat, carries a hyperspectral imager aimed at observing ocean colour and detecting harmful algal blooms. The imager covers the spectral range of 450 and up to 800 nm, with a spectral bandpass near 5 nm on average, and has a swath width of 70 km for the nominal captures in nadir. It can perform a slew maneuverer during a scan to induce more overlapping frames during data acquisition. To aid the work in ØKSAT data collection has been focused on relevant water bodies as well as data processing.

## 2.4 Initial validation

For the first validation effort, Sentinel-3 data was extracted from the selected coastal water bodies and their respective monitoring stations. With respect to Sentinel-2 data, it was extracted from the selected coastal and lake water bodies and their respective monitoring stations. For both satellites the extracted data corresponds to the time period between 2016-2023. Time series were generated and plotted for all stations and compared to available in situ data (chl-a and Secchi Disk Depth) as presented in Chapter **Error! Reference source not found.**. The S2 and S3 processing chains implemented on Brockmann's Calvalus system generates several alternatives for e.g. chl-a information. The purpose of this initial validation was to define if appropriate algorithms for each WB / station and parameter could be identified within the available alternatives. The validation is performed in parallel by several members of the team and the results are compared. Example plots are provided below in Chapter 2.4.1 and 2.4.2. The time series are used to evaluate if the satellite data is in the same range as the in situ data and if the peaks are consistent and captured correctly. Differences are expected as there are much fewer in situ samples, and the satellite data and the methods per se are very different (i.e. comparing about 1 l of water to a 300*300 m area of the top layer in of the WB). The satellite data are expected to capture more of the phenology and might identify other peaks as the number of observations are higher.

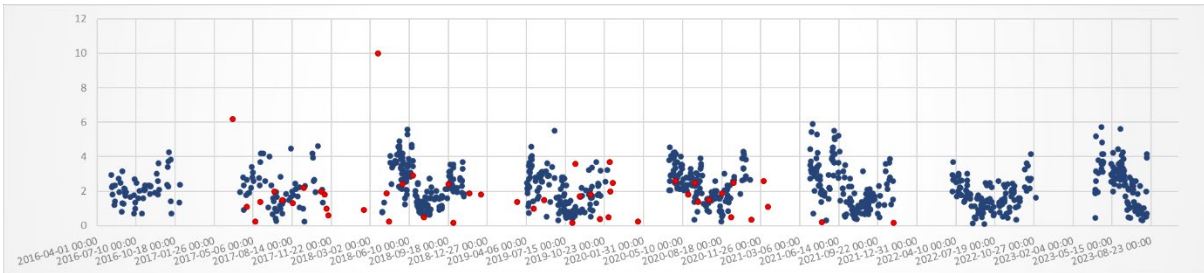

*Figure 2.7 Sentinel-3 chlchl-a time series for station Bastø in the outer Oslofjord. The blue dots indicate the Sentinel-3 data and the red dots the surface in situ samples.*

### 2.4.1. Coastal water bodies

The available Sentinel-3 OLCI processing scheme and algorithms are described in detail in the project description and a full reference list is provided in that document. An example of a time series based on Sentinel-3 data, corresponding to the monitoring station Bastø, is provided in Figure 2.7. The in situ data corresponds to a surface sample. The time series for Bastø provides good results with S3 OLCI data for one of the tested algorithms with chl-a values in the same range. The S3 range was between 0.11- 5.9 µg/l compared to the in situ data with a range of 0.16-10 µg/l. The two highest in situ points of 6.2 and 10 µg/l are both in March during the spring bloom, which can have a large effect on individual in situ samples. These measurements could also be on days without any S3 overpasses.

### 2.4.2. Lakes water bodies

The available Sentinel-3 OLCI and Sentinel-2 MSI processing scheme and algorithms are described in detail in the project description and a full reference list is also provided in that document. An example of a time series based on Sentinel-3 data, corresponding to the monitoring station Akersvannet, is provided in Figure 2.8and Figure 2.9 and for Sentinel-2 data in Figure 2.9 and Figure 2.11. The in situ data corresponds to integrated samples from 0 (surface) to varying max depth between 0-4 meters. In the time series, the chl-a and Secchi depth values are in the same range for both methods (RS and in situ) for both S3 and S2. There are more RS data from S3 as the satellites passes every day, whereas they only pass every 3-4 days with S2, which is reflected in the number of observations in the graphs.

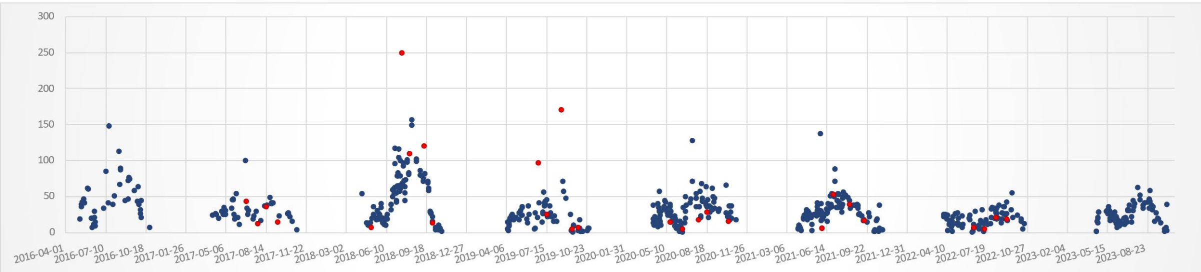


*Figure 2.8 Sentinel-3 chl-a time series for station Akersvannet. The blue dots indicate the Sentinel- 3 data and the red the surface in situ samples.*

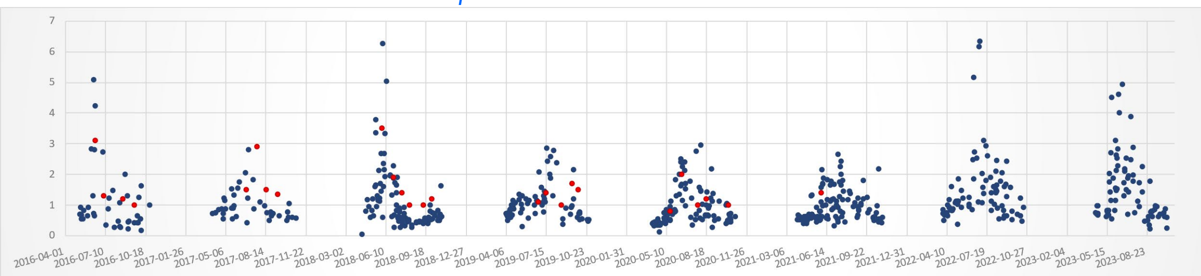


*Figure 2.9 Sentinel-3 Secchi Disc Depth time series for station Akersvannet. The blue dots indicate the Sentinel- 3 data and the red dots situ samples.*

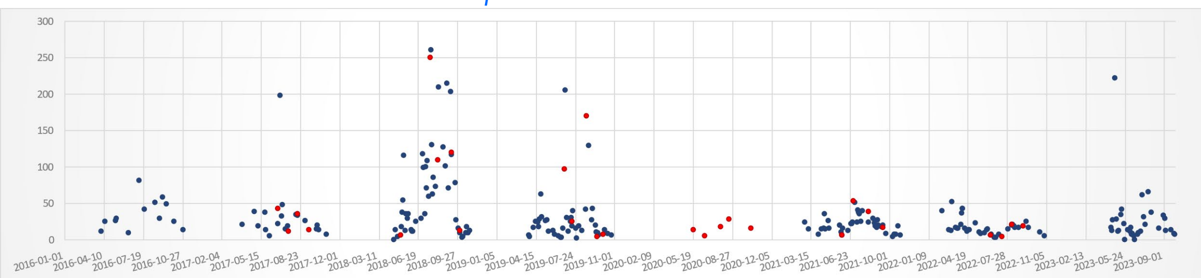


*Figure 2.10 Sentinel-2 chl-a time series for station Akersvannet. The blue dots indicate the Sentinel-2 data and the red dots the surface in situ samples.*

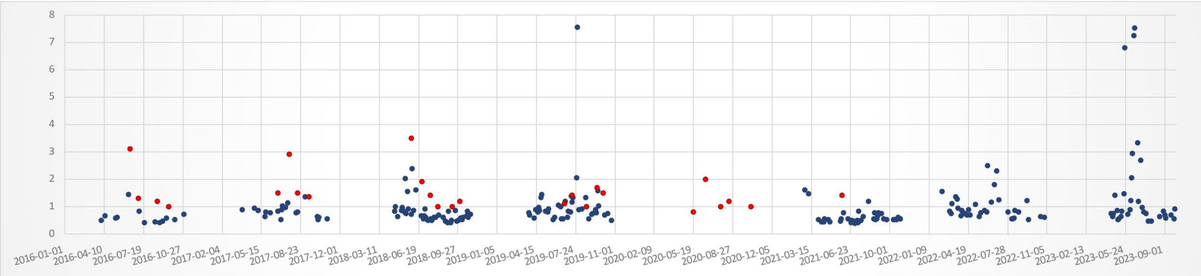

*Figure 2.11 Sentinel-2 Secchi Disc Depth time series for station Akersvannet. The blue dots indicate the Sentinel-2 data and the red dots the in situ samples.*

# 3 Work package 3 - Adaptation to the status classification system

In WP3, "Adaptation to the status classification system", discussion on potential products based on RS data has started. Two workshops between Miljødirektoratet representatives and classification/ satellite experts from the team have been performed. The main philosophy of products must be to use the RS data's strength to cover large areas and to cover close to the whole water bodies. A selection of lakes and marine locations has been tested based on the developed ØKOSAT DEMO VIEWER (WP5) that has been made available also for Miljødirektoratet. In the viewer one can easily select either the whole WB a circular point of different size or polygons. These examples illustrate the benefits and challenges of using the RS data. One needs to find optimal size of the WB to avoid interference from land and bottom reflectance but maximise the strengths of RS spatial coverage. The RS data will also increase the number of "station based" observations to be used in the classification and cover other years where in situ programs are not running or where classification is based on expert judgement. Here we must consider if RS data should "trump" expert judgement – that is, an assessment must be made as to whether RS data is more reliable than expert judgement and should be used in lieu of expert judgement. There is also further work to be done to incorporate new products based on RS data into the "Veileder" and to inform end-users e.g. "Statsforvalteren", what RS data is and how to use it. The work in WP3 will be intensified in 2024.

In Figure 3.1 the WB Storfjorden in Vansjø isare illustrated with available satellite data for 2018 and 2019. The next figure (Figure 3.2 shows two graphs for the situation on 22.7.18 with the monitoring station (VAN1) as a red dot and a polygon drawn around the station, which are illustrating the different amount of chl-a one can extract using the full lake information or at least the part of the lake that have data. In Figure 3.3 the results for 4 dates in July 2018 illustrate the difference of using 1 point station information, a large area around the station or the "full" WB data.

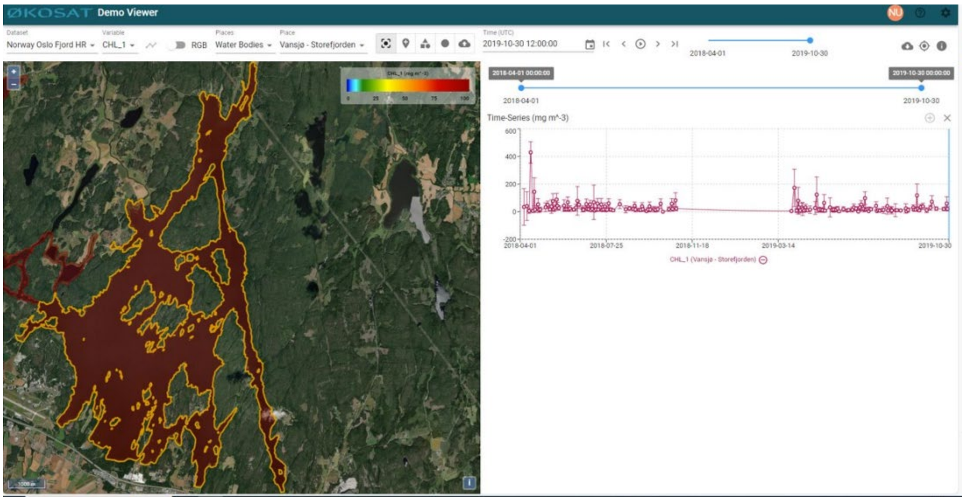


*Figure 3.1 chl-a concentrations extracted from RS data for a full WB with the plot displayed on the side of the image.*

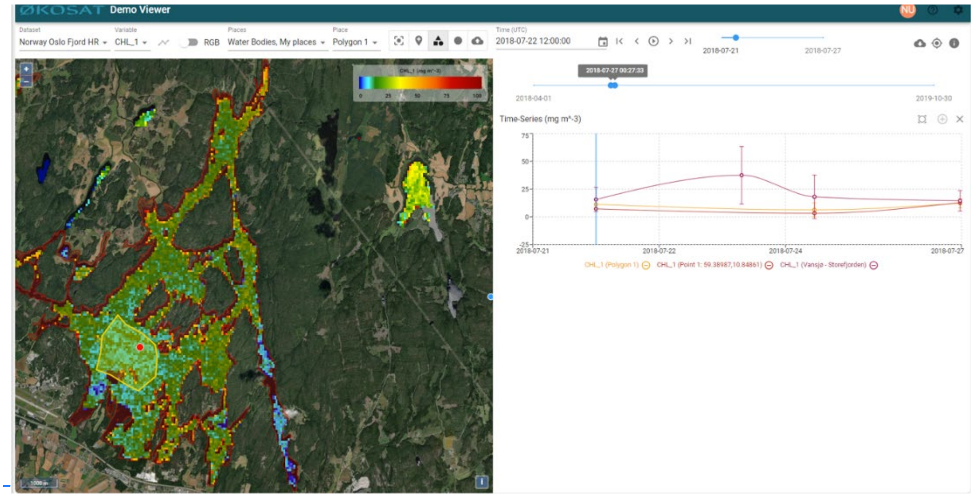


*Figure 3.2 chl-a concentrations extracted from RS data for the full WB (purple line), from a polygon around the monitoring station (yellow line), the full lake (puple line) and at the monitoring station coordinate (red line).*

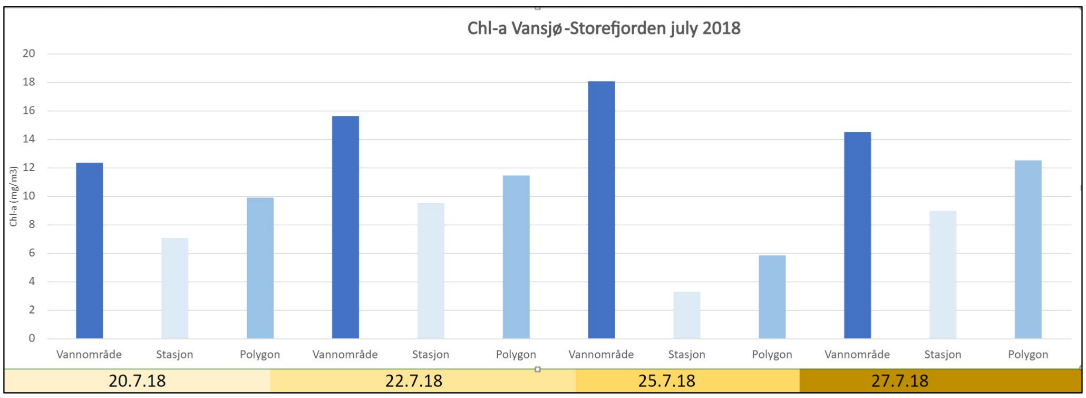


*Figure 3.3 Comparison of the average chl-a concentrations extracted from RS data for the WB (dark blue bars), from a polygon around the monitoring station (blue bars) and at the monitoring station coordinate (light blue bars) from different dates in July 2018.*

# 4 Work package 5 - Infrastructure

## 4.1 Achievements in system setup and data production

The processing and data access system for the project has been configured and set up using three infrastructures: The Brockmann Consult (BC) cluster, a cluster operated by BC and hosted on Copernicus Data Space Ecosystem (CDSE) Creodias, and the data cube services of BC hosted on Amazon Web Services (AWS).They have been used to generate the first data cubes and to extract time series for different coastal locations and lake stations (Figure 4.1). In the first year, data has been ingested, processed and three data cubes have been generated with provided access.

- The BC cluster hosts a mirror of the S3 OLCI L1B data and has been used for processing and extracting time series of 300m patches for the complete mission (2016-2023). The time series have been further analysed in Ch. 2.4 Initial validation (see above).
- It has also been used to generate a S3 data cube for the Oslo region. This cube has 10 variables, 1024 x 1024 pixels and 1921 time steps.
- The Sentinel-2 data is available online in CDSE. Therefore, the time series for the higher resolution has been processed and extracted on the BC cluster on Creodias. This was a larger processing task which has used 15 VMs for more than a week (11000 core hours) to process 34000 MSI L1C input data products.

An excerpt of the result of time series extraction is shown in Figure 4.2.

- The result of time series extraction is a set of CSV files with 3*3 lines for each observation (3*3=9 pixels extracted) and as many 3*3 blocks of lines as there are locations and observations in the respective year.
- There is one CSV file for S3 OLCI and one CSV file for S2 MSI per year.
- The columns of the CSV file contain all variables that the processing chain produces (CHL, ...)

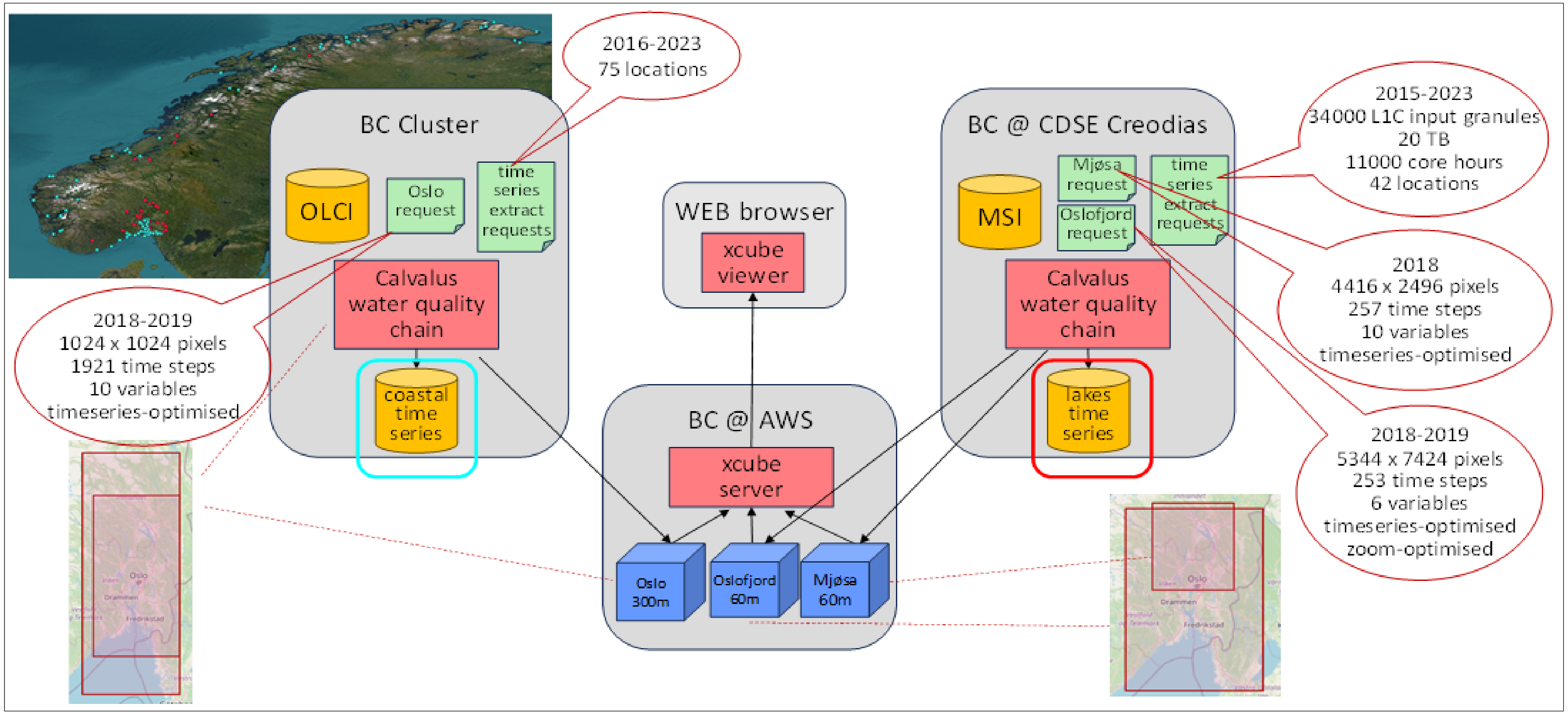


*Figure 4.1 Time series extraction and data cube generation for ØKOSAT from OLCI and MSI data*


*Figure 4.2 Coastal and lakes time series extraction.*

## 4.2 Data Cubes

The three demo data cubes for the project have been generated into gridded data products. One S3 OLCI data cube for the Oslofjord region has been developed with a spatial resolution of 300m and temporal coverage between Jan 2018 - Dec 2019. The variables included are absorption_443, chl concentration (2 versions), turbidity (2 versions), Secchi depth (2 versions) and suspended sediment (TSM) concentration. Two data cubes from Sentinel-2 have been produced: the Mjøsa cube for 2018, and Oslofjord cube for the spring and summer seasons of 2018 and 2019. They are larger due to the higher resolution of the sensor.

The data cubes are generated and directly stored in AWS cloud storage. From there, they are served by BC's xcube server and can be investigated using the xcube viewer in web browsers. Access to the viewers has been granted. Figure 4.3 shows a schematic view of the xcube production chain and Figure 4.4 the result of data cube generation.

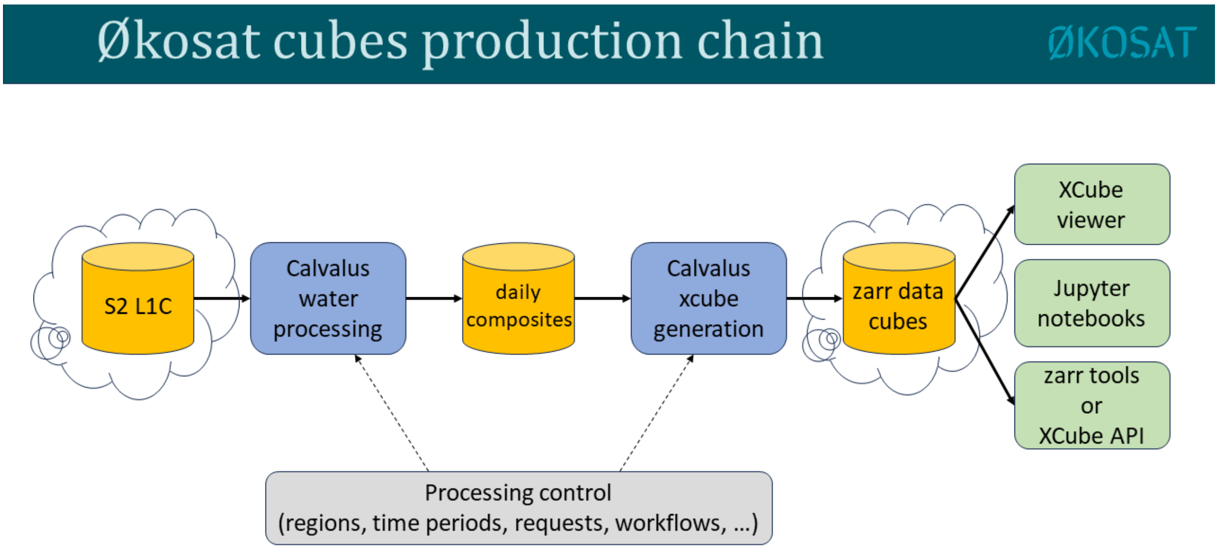


*Figure 4.3 Schematic view of cube generation.*

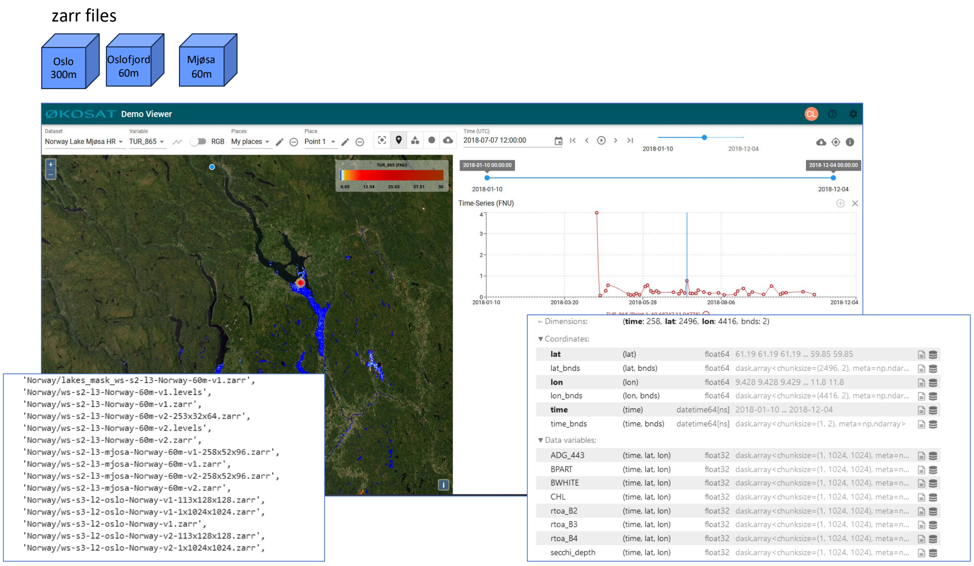


*Figure 4.4 Example of gridded data products (data cubes).*

- There are three data cubes represented as zarr files (directory trees).
- The list in the lower left shows that there are in fact additional zarr files for a time-chunked version of the same data (time series-optiomised) and for a leveled version (zoom-optimised).
- The lower left shows the inspection of one zarr file with its variables.

The three cubes differ in size where the S2 cube demands much higer storing capacaty (Figure 4.5), although they are covering smaller areas.

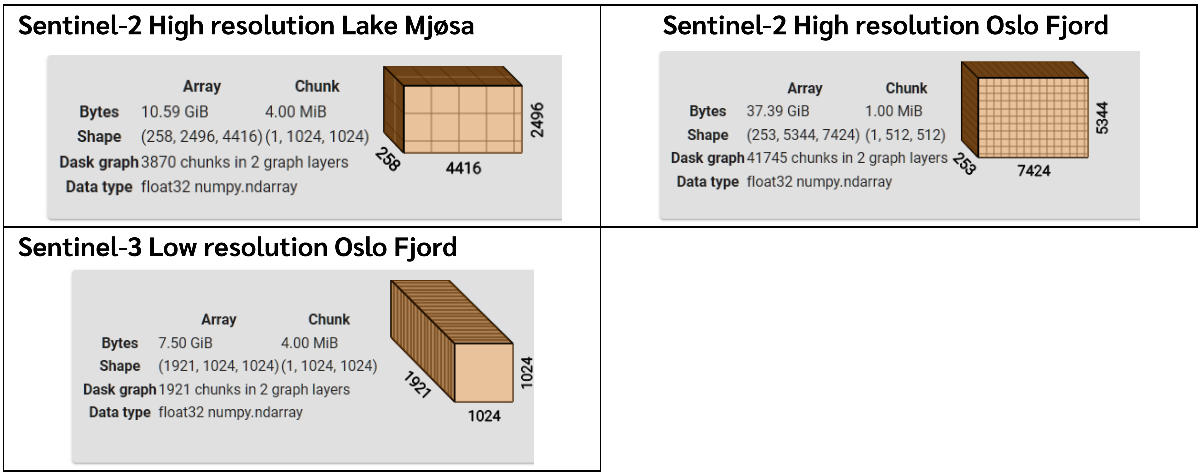


*Figure 4.5 Storing for the three cubes in the data viewers.*

# 5 Work package 6 - Up-scaling

We have started investigating the use of MET.no's cloud infrastructure ahead of schedule of the original plan. MET.no's cloud infrastructure is part of the European Weather Cloud (EWC). The sites involved in the test are BC as client, EWC, MET.no, and Eumetsat and ECMWF as backend of EWC. The elements and their interactions are depicted in Figure 5.1.

We had an onboarding session with Eumetsat. We got an account in a larger tenant Eumetsat-no-met-federation, which is fine for testing. With this account we can login into the Morpheus Dashboard.

- We got access to MET.no's S3 cloud storage which is independent of EWC. We have copied the Mjøsa cube to this cloud storage as a real-world test. This was successful.
- We have used the Morpheus Dashboard of EWC to provide the first VM. But there is an issue with accessing the VM with ssh from BC and the internet. We have raised a ticket but could not do so with the provided account. We have used a different project's account at ECMWF that luckily is on the same shared Jira system. Reaction on the ticket was quick, and the issue has been temporarily resolved for this particular VM. The ticket has been escalated to Morpheus.
- With the temporary solution we can login to the test VM in the MET.no cloud.
- We have tested the Morpheus command line client as well. This will be needed to automatically set up VMs in the cloud environment. Listing our test VM worked fine.
- During the onboarding session we also discussed input data access. Only OData is offered as a protocol, not S3 or some other shared file system. This is not suitable for processing because all input data would have to be ordered and downloaded before processing instead of streaming it directly by the data processors. (It would be possible to access CDSE's S3 storage of Sentinel data inputs. But this is not the preferred solution network-wise.)

### Summary of test results

- Access to VMs provisioning provided
- Network security issue open
- Shared tenant eumetsat-no-met-federation
- Morpheus interface (not openstack)
- Access to input data on met.no only via OData protocol (ordering)
- Cloud storage for result data provided on met.no infrastructure

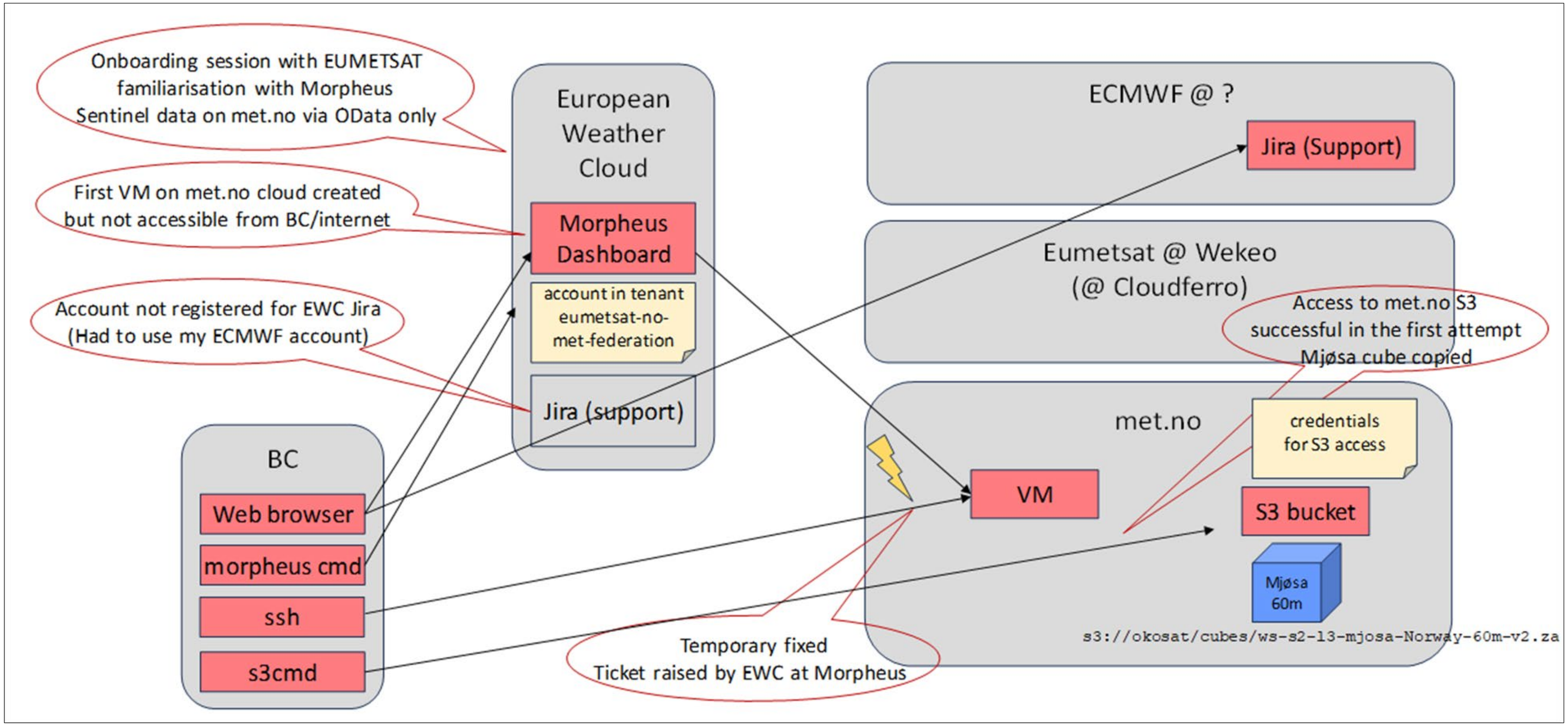


*Figure 5.1 Testing access to MET.no infrastructure.*

# 6 Plan for 2024

The overall goal with the project is:

- To develop and demonstrate a supplementary method that can complement the in situ data with satellite remote sensing data in the water bodies without monitoring so that as many lakes and coastal ware bodies as possible can be classified.

And specific goals are:
1) to develop and improve the methods for using satellite data to calculate the concentration of chl-a and other water quality parameters, such as turbidity and Secchi depth

2) to develop a system for operational and quality assured data processing, operated on an IT-platform that can be hosted on a private or public cloud, and that can be used for ecological status classification with the goal of implementation during the assessment period 2025-2027.

The work during the current project period was to prepare for this step by step, but there are many things still remaining. In 2024 the work will be focused on:

- Set up a working plan for the evaluation of RS products
- Detailed evaluation of lake and coastal WB data
  - Evaluate the 10+10 WB and if needed adjust any algorithms/processing and re-evaluate
  - Expand the evaluation to the other WB that we have in situ data on (i.e. Northern and western coastal areas + northern lakes)
- Focus on WP3, WB classification products, several workshops will be held during the year
- Upscaling discussion and test on WP5 and WP6
- Prepare the work on the upscaling to all Norwegian WBs in 2025 in dialogue with NEA via the “upscaling plan” and “roll-out plan” (described below)

### Table 3. Milestones and Timeplan

| 2024 Milestones | Topic | Date |
|---|---|---|
| **3rd project meeting/Half-time meeting** | • Presentation of 2nd evaluation of data products for lakes & coast<br>• Presentation of 1st evaluation of products in the supplementary classification system<br>• Presentation of initial results from Mjøsa impact analysis<br>• Challenges raised<br>• Inputs and way forward | Jun. 2024 |
| **4th project meeting** | • Brief status update<br>• Challenges raised<br>• Inputs and way forward | Oct. 2024 |
| **End-meeting** | • Presentation of final results for lakes, coastal areas<br>• Presentation of a supplementary classification system<br>• Presentation of results and recommendations from Mjøsa impact analysis<br>• Challenges raised<br>• Inputs and plan for up-scaling | Dec. 2024 |
| **Report** | • End of first phase report | Jan. 2025 |

## Upscaling Plan

-**Decide on the infrastructure for the production service, its deployment and operations** (Calvalus production system configured with the Norwegian WQ graph).

-As of today, options include to continue using the service offered by Brockmann Consult, or alternatively migrate the software and its operations to another place, e.g. on a suitable national Norwegian processing center, institute or a commercial cloud as AWS or the EC DAS (successor of the DIAS'ses).

-**Decide on a priority list for water bodies to be included**; technically, all water bodies of interest can be included in the production system. However, validation will be preliminary and based on general water types. Validation of all water bodies will take time and a priority list should be developed by NEA in which sequence validation shall be executed. It should also be decided if only validated water bodies shall be made public or if all water bodies shall be included after upscaling and a quality label shall be assigned according to validation status.

## Roll-out Plan

-This list is not exhaustive and will be completed during the course of this project in collaboration with NEA.

-Depending on the decision for the production infrastructure, include a **migration plan for Calvalus** (with the Norwegian Processing graph included).

-**Configure a cloud system for the datacubes** which is large enough to (a) cover all water bodies and (b) serve all users that can be expected concurrently. The latter will be supported by autoscaling as offered by cloud systems like AWS.

-**Testing** of the technical readiness for production of whole country.

-**Validation of water bodies** as defined (see above). Additional validation in 2025 means that the results of the new waterbodies that are upscaled will be checked by comparing with similar WB types that is already validated in 2024 and see if there are any that are problematic or limited for different reasons that should be excluded, changed etc. A possibility can be to adapt or optimise the algorithms for those. Also new updates of the L1 data from EUMETSAT can be considered (e.g. improved cloud flagging etc.).

-**Adaptation/optimisation of algorithms** to water bodies or water types, following the validation.

-Identification of the minimum size and /or optical complexity of lakes/coastal water bodies that can be classified by satellite data.

**- A plan for the classification** towards the users from e.g. Stadsforvaltaren needs to be set up in combination with info material/training sessions etc.

Below is a suggestion of an adapted time schedule for the whole project based on the project plan, but now including WP4 and WP6 in 2024, and extension of WP1 and WP3 into 2025 and beyond (Figure 6.1).

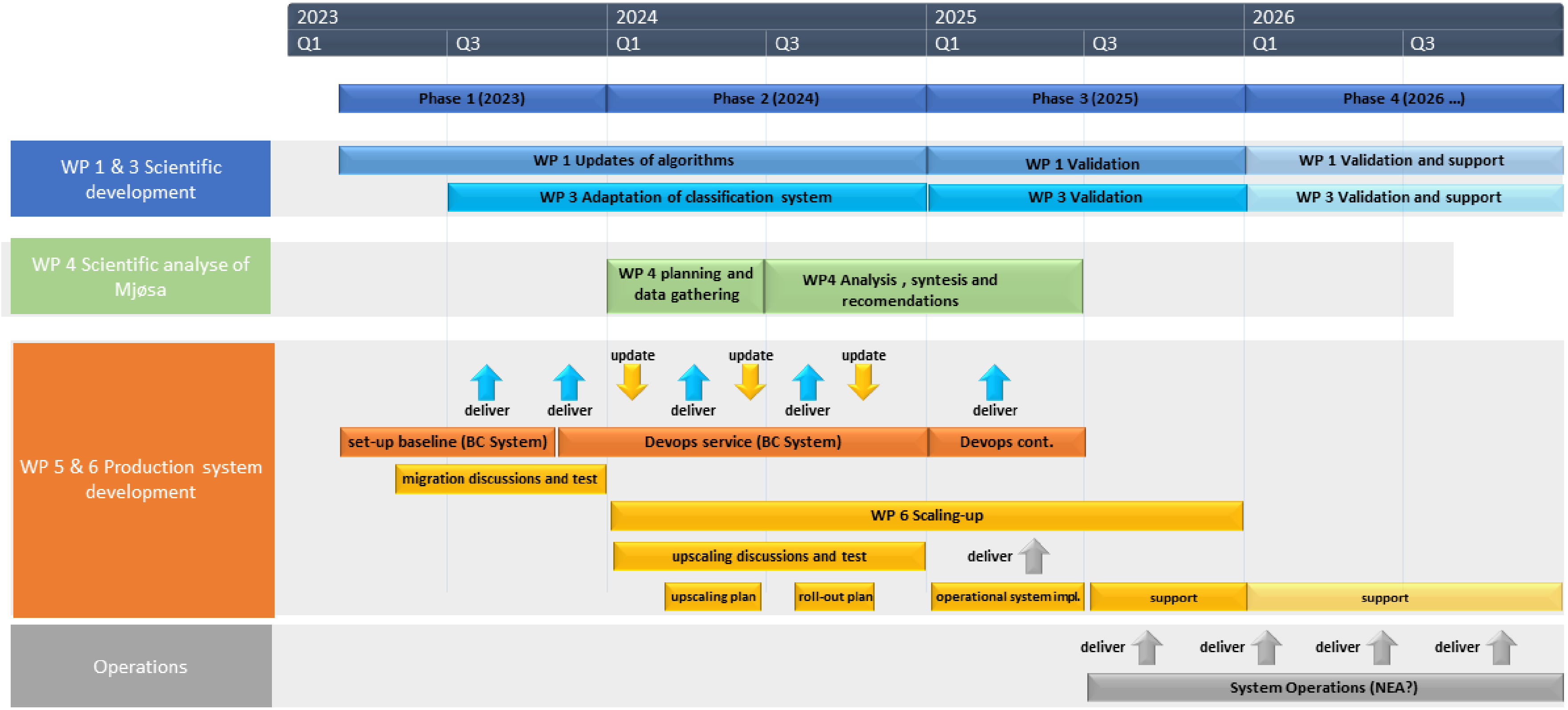


*Figure 6.1 Schedule for 2024-2026.*