

<b>Project Title:</b>	Baltic+ SeaLaBio
<b>Document Title:</b>	Dataset User Manual
<b>Version:</b>	1.2 Final version
<b>Author(s) and affiliation(s):</b>	Thomas Neumann, IOW Sampsa Koponen, Jenni Attila, Kari Kallio, Sakari Väkevä, Mikko Kervinen, SYKE Carsten Brockmann, Dagmar Müller, BC Constant Mazeran, SOLVO Petra Philipson, BG
<b>Version history:</b>	1.0 June 25, 2020. Version for ESA review 1.1 Sep 30, 2020. Version for ESA review 1.2 Nov 20, 2020. Final version
<b>Distribution:</b>	Public

## Contents

Abstract .....	3
Glossary.....	3
List of Symbols.....	3
1 Introduction .....	4
2 Generated datasets .....	5
2.1 EO based input data to ERGOM .....	5
2.1.1 Generation of data .....	5
2.1.2 Access to data .....	8
2.2 ERGOM simulations .....	9
2.2.1 Generation of data .....	9
2.2.2 Access to data .....	9
2.3 Satellite products .....	11
2.3.1 Generation of data .....	11
2.3.2 Access to data .....	11
3 References .....	13
Appendix 1: Instructions for accessing data in THREDDS .....	13
Appendix 2: Instructions for accessing data in WMS .....	16

## Abstract

This document contains detailed descriptions of the datasets generated during WP4 of the SeaLaBio project and explains how to access the data.

## Glossary

AC	Atmospheric correction
aCDOM	Absorption coefficient of Coloured Dissolved Organic Matter
ATBD	Algorithm Theoretical Basis Document
B6	Band 6 of a satellite sensor
B8	Band 8 of a satellite sensor
C2RCC	Care 2 Regional Coast Color
CDOM	Coloured Dissolved Organic Matter
CSV	Comma-separated values
EO	Earth Observation
ERGOM	Ecological Regional Ocean Model
GIS	Geographic Information System
IOW	Leibniz Institute for Baltic Sea Research
L1b	Level 1b (satellite data processed to top-of atmosphere radiances)
L2	Level 2 (satellite data processed to reflectances and/or geophysical parameters)
MSI	MultiSpectral Instrument
OGC	Open Geospatial Consortium
OLCI	Ocean and Land Color Imager
Pt	Platinum-Cobalt Scale (used in the measurement of water color)
S2	Sentinel-2
S3	Sentinel-3
WAF	Web Accessible Folder
WMS	Web Map Service

## List of Symbols

Symbol	Definition	Dimension/Unit
ay400	Absorption coefficient of yellow substances (CDOM) at 400 nm	$m^{-1}$
ay440	Absorption coefficient of yellow substances (CDOM) at 440 nm	$m^{-1}$
N_valid_cdom	The number of valid pixels extracted from an area	
N_max	The maximum number of pixels that can be extracted from the area	

## 1 Introduction

The overall goal of the WP4 of the Baltic+ SeaLaBio project was to generate EO and model datasets using the methods developed in WP3 and then assess the impact that these new and/or improved products have on the understanding of biogeochemical processes and carbon cycle. One specific goal formulated during the project was to assess if the performance of the ERGOM model can be improved by using EO based data as input values. This document describes the datasets generated during WP4. These currently include:

- aCDOM (absorption by Coloured Dissolved Organic Matter) values estimated from S2 MSI data with the C2RCC V1 processor from areas representing the ERGOM input locations
- aCDOM maps simulated with ERGOM
- aCDOM maps estimated from S3 OLCI data with the Baltic+ processor developed in WP3

In addition to the technical descriptions about the characteristics and generation of the products this document shows how the products can be accessed. For this the following have been utilized:

- **TARKKA** ([www.syke.fi/tarkka/en](http://www.syke.fi/tarkka/en)) is a public service provided by the Finnish Environment Institute for browsing and viewing SYKE's open satellite data. The map application is designed to display satellite imagery at both high resolution (10 m – 60 m) and moderate resolution (300 m – 1 km) and allows the user to overlay and combine various other datasets with EO products. With the application the user can e.g. zoom in and out on the map, pan to different areas and click the image to obtain parameter values. More information about TARKKA and its use is available in [http://www.i4.ymparisto.fi/i4/eng/tarkka/tarkka\\_guide.pdf](http://www.i4.ymparisto.fi/i4/eng/tarkka/tarkka_guide.pdf).
- **Data interfaces** (machine to machine) are standardized ways to share data so that a user can access the information without downloading the data as a file in the local environment. The main advantage of this is that if the information in the server is updated, the client has immediate access to the new version. For this SYKE utilizes an OGC (Open Geospatial Consortium) compliant GeoServer while IOW distributes ERGOM data through a THREDDS system.
- **SYKEs EO Web Accessible Folder** (EO-WAF) located at [eo.ymparisto.fi/data/water/Baltic\\_SeaLaBio/](http://eo.ymparisto.fi/data/water/Baltic_SeaLaBio/) is simply a folder visible to the Internet. It is convenient for sharing files in different format (images, text, etc.).

Comparisons of EO, in situ and model data are presented in the Impact Assessment Report.

## 2 Generated datasets

### 2.1 EO based input data to ERGOM

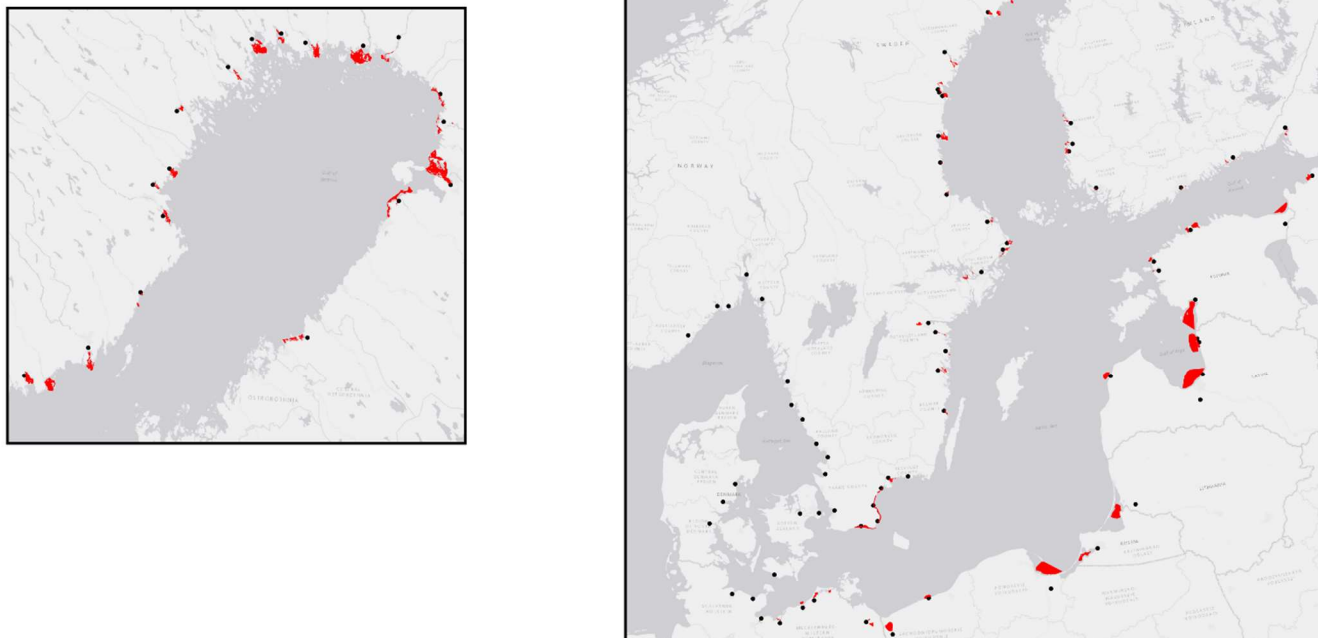
#### 2.1.1 Generation of data

Regional, marine ecosystem models, like ERGOM for the Baltic Sea, need boundary conditions providing information and forcing from outside the model domain. At the sea surface, meteorological data from other models and/or observations are used. The data include *inter alia* wind, temperature, radiation, precipitation, and humidity. In addition, ecosystem models require atmospheric gas concentrations (partial pressure) for e.g. oxygen and carbon dioxide. Especially nitrogen enters the coastal ocean also due to deposition (wet and dry) and therefore is needed as an external forcing. Another important boundary for the model is the land sea flux. It is mainly represented by riverine input carrying fresh water, nutrients, and other substances. For coastal waters with fresh water inputs and especially for the Baltic Sea, terrestrial CDOM is an important part of the marine ecosystem. aCDOM is used in the model to estimate the light climate controlling primary production. The previous version of the model estimated aCDOM from salinity and thus did not represent well the amount of CDOM entering the Baltic Sea from terrestrial sources. The updated version of the model can utilize river CDOM values as input data.

For this study we utilized high resolution images of the Sentinel-2 satellite (S2) to provide the aCDOM values. The spectral characteristics (number of bands, widths and wavelengths of bands) of this instrument are not optimized for water applications. However, it has provided good results in previous studies and having a high resolution is essential for the estimation of CDOM values in river estuaries where the ERGOM input locations are (Figure 1). As the rivers themselves are usually too narrow even for S2 data the extraction areas were defined on the coast in front of the river.

The processing steps are shown in Table 1. The C2RCC V1 processor (Brockmann et al., 2016) was utilized to process satellite data from L1b (radiances) to L2 (concentrations) while IDEPIX was used to remove effects of clouds and shadows. Pixel values were extracted from 69 areas defined to represent the ERGOM input locations (typically a river estuary but sometimes representing a larger area with multiple rivers) and calibrated to absorption by CDOM at 400 nm values (aCDOM(400)) using in situ data from Finnish coastal areas (example shown in Figure 2). All available data from years 2017 - 2019 were processed and monthly means were derived. Optical satellites cannot observe water in winter time due to lack of light and the presence of ice cover. Thus, we used interpolation to generate data for the missing months. Figure 3 shows the final aggregated monthly values at four input locations. Figure 4 show two other examples together with river runoff values as visualized in TARKKA. The behavior of the values is logical: High values in the spring time when melting snow water brings terrestrial matter to coastal areas, low values during summer and again higher values in autumn when the rainfall increases.

The dataset was generated using the 75<sup>th</sup> and 95<sup>th</sup> percentiles of the pixel values extracted from each area so that the impact of different input values could be tested. The results of these tests are presented in the Impact Assessment Report.

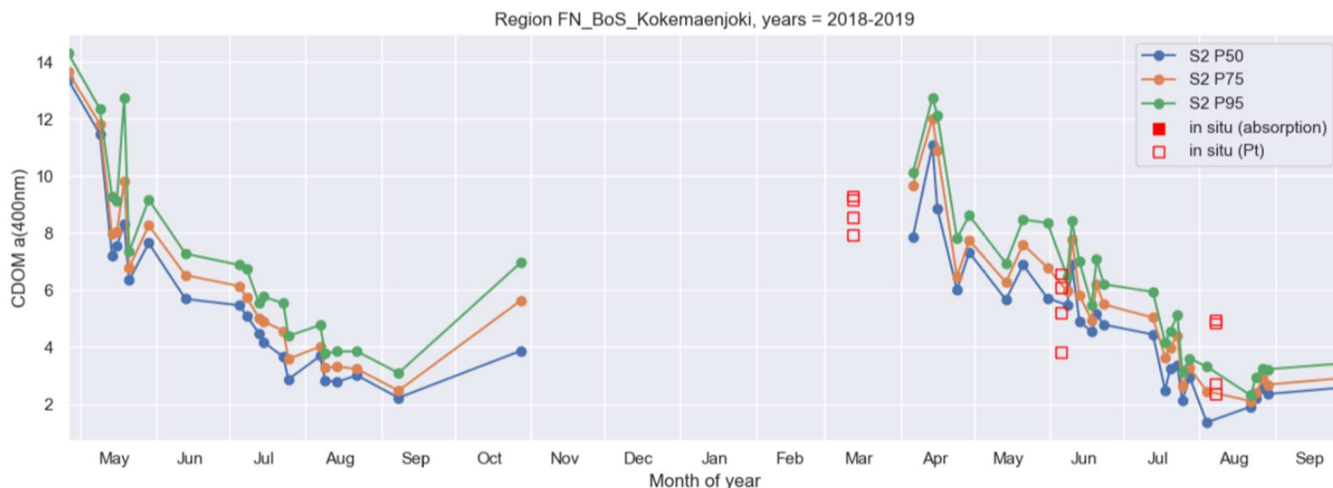


**Figure 1. Pixel extraction areas defined for the data flow test. The extraction areas were drawn manually at river mouths near the input locations used in ERGOM. Note that the areas were not defined for locations closest to the North Sea (e.g. west coast of Sweden, Denmark and Norway).**

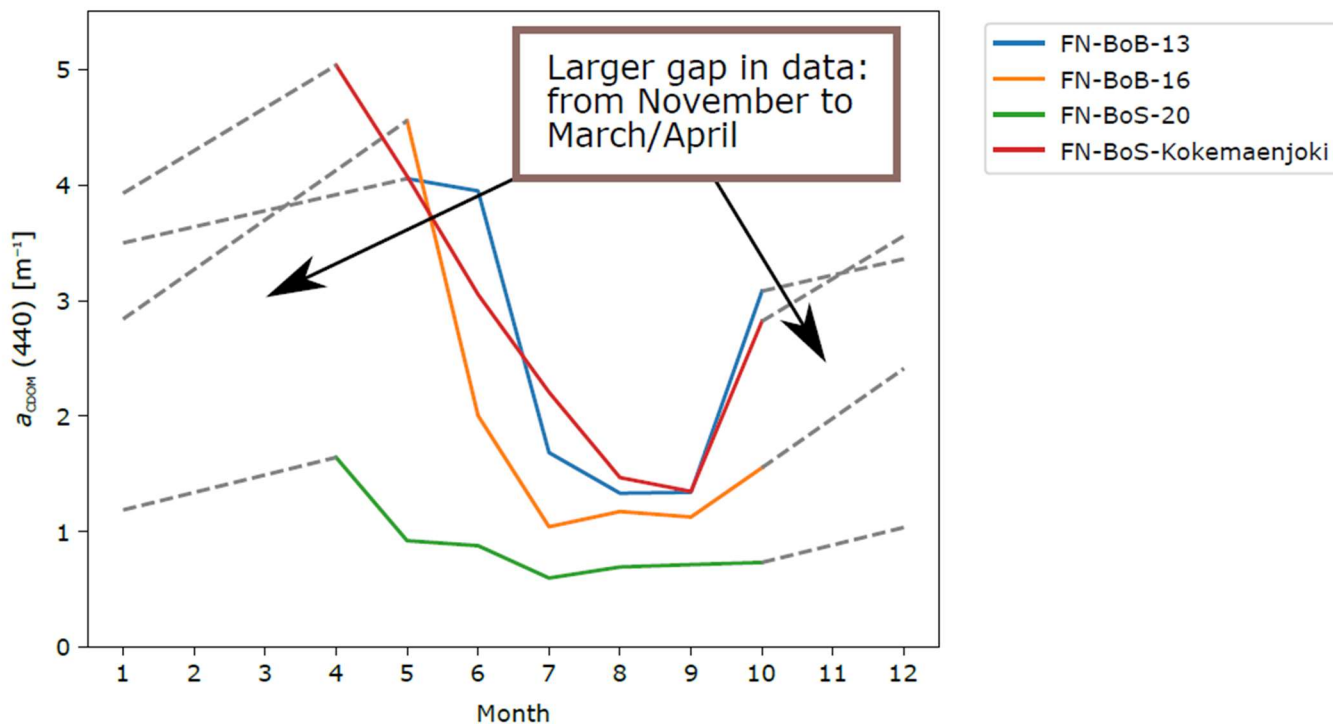
**Table 1. Details of the ERGOM input data processing (with S2 data).**

	Step	Method
<b>Level 2 processing</b>	1. AC	C2RCC V1
	2. In water	C2RCC V1
	3. Data extraction	PixEx
	4. Cloud masking	Idepix
<b>Post processing</b>	5. Local calibration	Empirical equation based on data from Finland (for C2RCC V1)
	6. Filtering (rules used to remove poor quality data)	Remove dates with low number of pixels ( $N_{\text{valid\_cdom}} / N_{\text{max}} < 50\%$ )*
	7. Derive monthly aCDOM means (from all years available) for each area	Arithmetic mean of the 75 <sup>th</sup> and 95 <sup>th</sup> percentile pixel values of all valid days within each month
	8. Derive values for winter months	Linear interpolation from fall to spring
	9. Transform aCDOM values to ERGOM units [ $\text{m}^{-1}$ to Mol/kg]	$\text{aCDOM} [\text{mg}/\text{m}^3] = 21 \cdot \text{ay} + 5$ [ $\text{m}^{-1}$ at 440 nm]. Conversion factor from 400→440: $\text{ay}(440) = 0.5 \cdot \text{ay}(400)$ . Conversion to [Mol/kg] in ERGOM.
	10. Data export	Write a CSV file based on a template

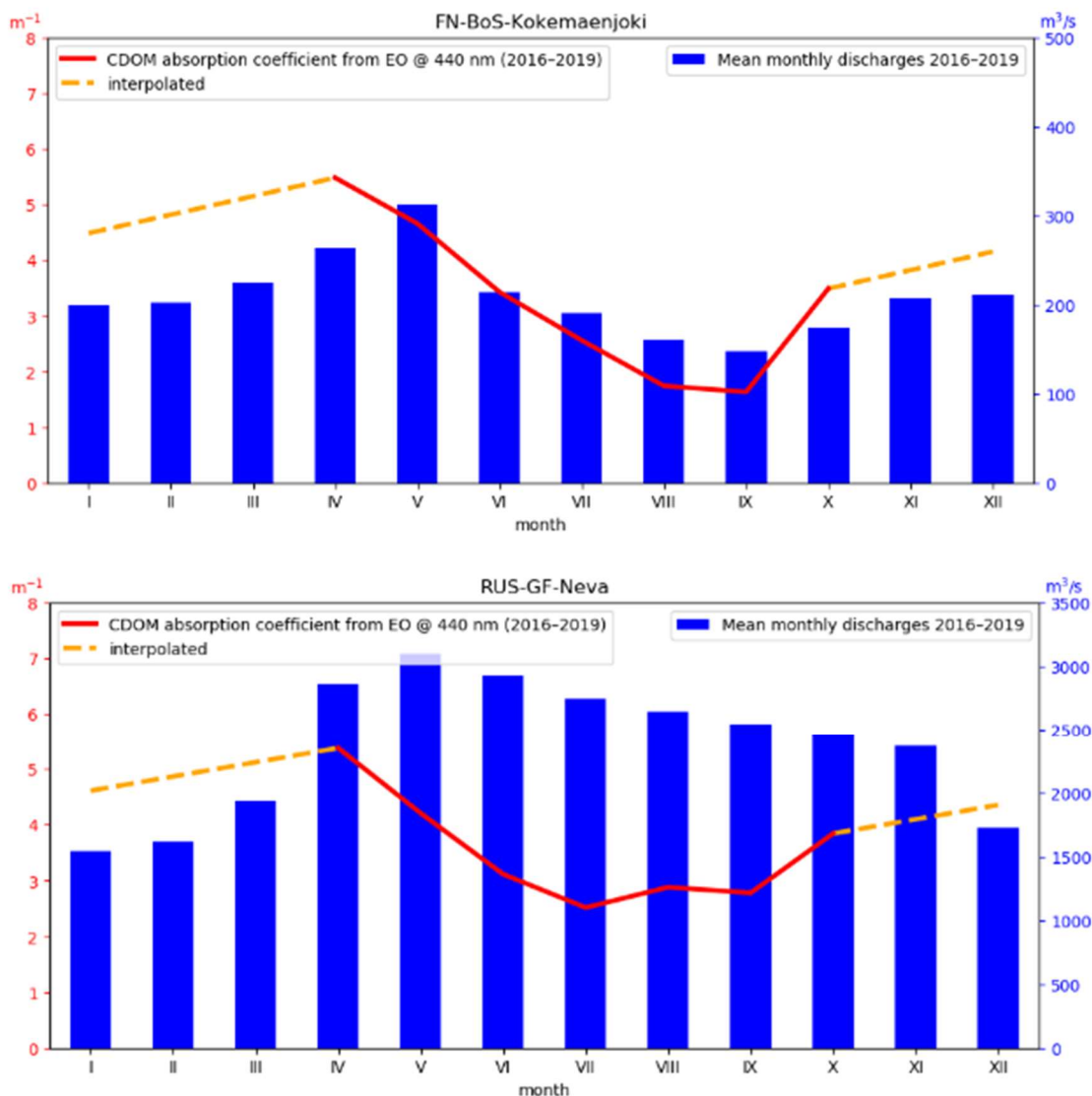
\*  $N_{\text{valid\_cdom}}$  is the number of valid pixels extracted from an area while  $N_{\text{max}}$  is the maximum number of pixels that can be extracted from the area.



**Figure 2.** Time series aCDOM (at 400 nm) derived from satellite data (S2 & C2RCC V1) and in situ measurement at the Kokemäenjoki (in Finland) input location of ERGOM. Pt refers to the measurement of water color in the Platinum-Cobalt Scale.



**Figure 3.** The aggregated monthly values of aCDOM based on EO data (S2 & C2RCC V1) for four ERGOM input locations in the western coast of Finland. During winter months EO data is not available and the values are obtained through interpolation.



**Figure 4. Seasonal behavior of aCDOM (at 440 nm) estimated from satellite data at 2 ERGOM input locations in red and the runoff used in the ERGOM in blue (note the different scales).**

### 2.1.2 Access to data

The data are available as follows:

- Plots in the TARKKA service (click the red circles which represent the ERGOM input locations). See the examples in Figure 4.
- Time series plots (\*.PNG) in the EO-WAF (see Chapter 1, folder CDOM\_input\_to\_ERGOM)
- Time series text files (\*.CSV) in the EO-WAF:
  - summary\_statistics\_ay440\_P75\_with\_interpolation\_round.csv
  - summary\_statistics\_ay440\_P95\_with\_interpolation\_round.csv

The format of these files is comma separated text and its contents looks like this:

```
# COLUMNS: areaid, month (1-12), ay440 (1/m)
FN_BoB_14,1,4.05
FN_BoB_14,2,4.21
FN_BoB_14,3,4.38
FN_BoB_14,4,4.54
```



FN\_BoB\_14,5,4.71  
FN\_BoB\_14,6,4.95  
FN\_BoB\_14,7,2.50  
FN\_BoB\_14,8,1.99  
FN\_BoB\_14,9,3.17  
FN\_BoB\_14,10,3.56  
FN\_BoB\_14,11,3.72  
FN\_BoB\_14,12,3.89

## 2.2 ERGOM simulations

### 2.2.1 Generation of data

The modifications made to the model during this project are described in the ATBD. The main change was to allow the model to take CDOM concentrations as input data along with runoff (see Chapter 2.1). The model was first run with 3 n.m. resolution from year 1948 to 2019. After this the model was run with improved resolution (1 n.m.) for years 2016 to 2019 using the 3 n.m. result as a starting value. ERGOM provides results with a time step of 12 hours. However, the values do not change quickly and in order to facilitate the comparison with EO datasets the averages for each calendar month were derived.

### 2.2.2 Access to data

These data are available as follows:

- Full dataset with all parameters and 12 h time step through the THREDDS interface of IOW:
  - o Parameter 1
  - o Parameter 2
  - o etc
- Monthly CDOM averages as:
  - o Images in the TARKKA service
  - o GeoTiffs in the EO-WAF (folder ERGOM\_results)
  - o WMS interface

#### How to access data in THREDDS:

The THREDDS data server is a web server that provides metadata and data access using remote data access protocols (<http://www.unidata.ucar.edu/software/tds/current/TDS.html>). IOW is running a public server providing data from model simulations. Data generated in the project SeaLaBio can be accessed via the link [thredds-iow.io-warnemuende.de/thredds/catalogs/projects/SeaLaBio/catalog\\_sealabio.html](http://thredds-iow.io-warnemuende.de/thredds/catalogs/projects/SeaLaBio/catalog_sealabio.html).

Appendix 1 shows the instructions for accessing the data in IOW THREDDS

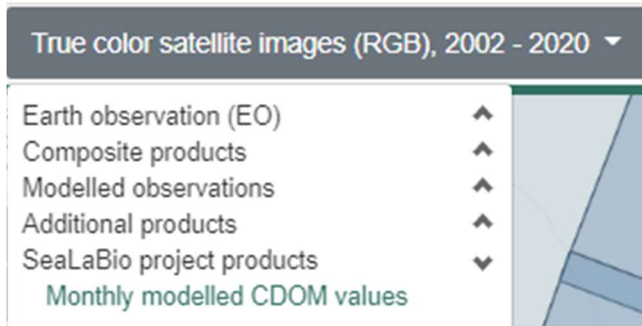
#### How to view the data in TARKKA:

In TARKKA ([www.syke.fi/tarkka/en](http://www.syke.fi/tarkka/en)) the monthly CDOM images can be browsed by selecting SeaLaBio project products → Monthly modeled CDOM values from the menu in the left upper part of the window (Figure 5). The selection of the month can be done with the list of images in the right side of the window.

The ERGOM input locations and the corresponding extraction areas used in the generation of the input data are shown over the CDOM values. These datasets can be turned on and off in the panel below the product selection menu (Figure 6). An example product is shown in Figure 7.

#### How to access data through SYKE's WMS:

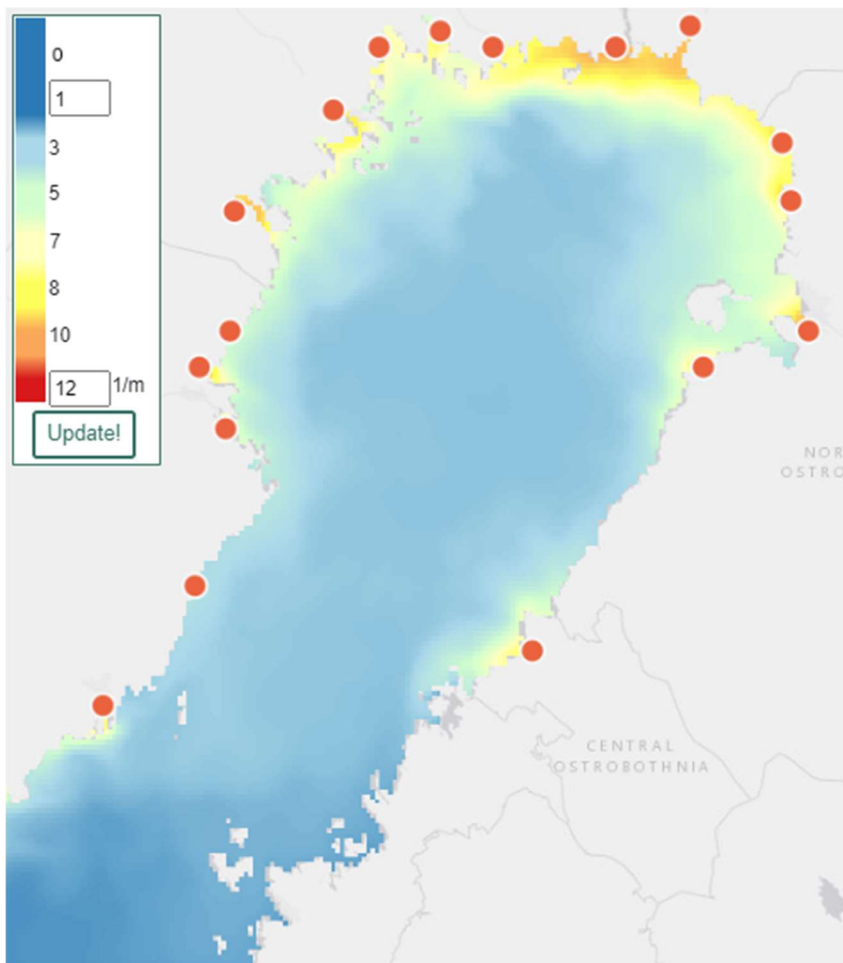
See Appendix 2.



**Figure 5.** An example of a monthly ERGOM CDOM map visualized in TARKKA. The list of months from which the CDOM values are available is shown in the right side of the window.



**Figure 6.** Panel showing the GIS data related to the generation of the data.



**Figure 7. aCDOM (at 400 nm) simulated with ERGOM for May 2019. Red dots are the ERGOM input locations.**

## 2.3 Satellite products

### 2.3.1 Generation of data

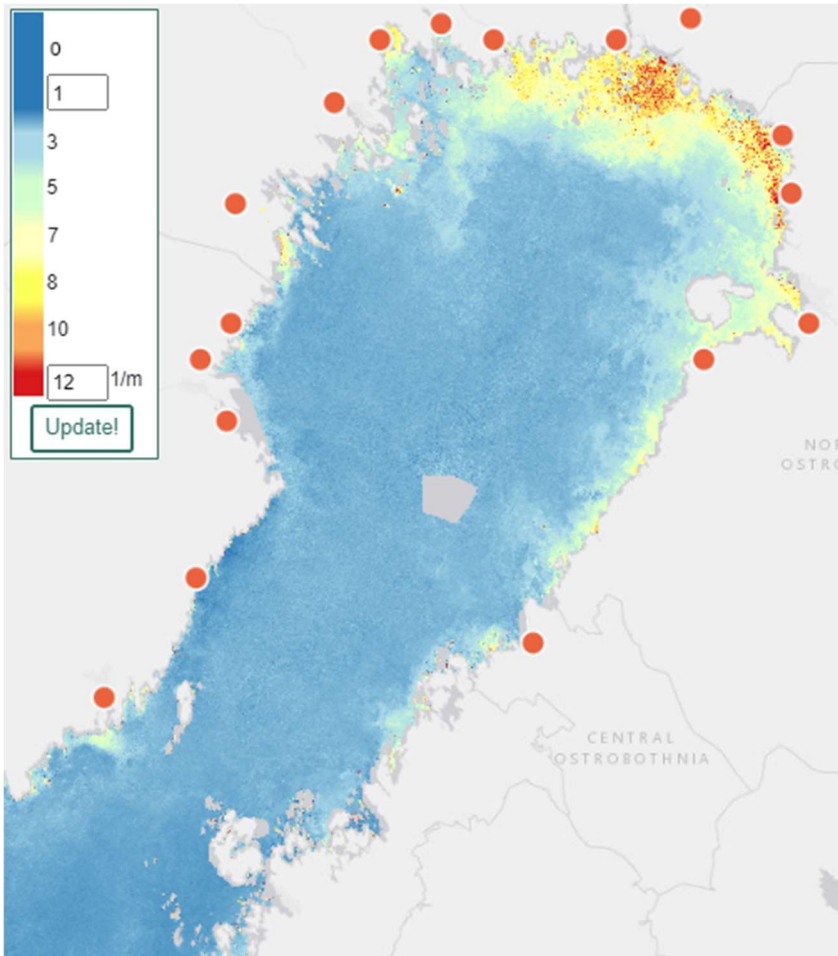
The monthly aCDOM composites were derived from S3 OLCI data with the Baltic+ atmospheric correction processor and a band ratio calibrated with in situ data (see the Validation Report of this project for details). Due to processing capacity restrictions the monthly products were calculated for three subareas (Gulf of Finland, Bothnian Bay, Archipelago Sea) of the Baltic Sea separately and then mosaicked together. The no-data strips between the tiles were filled and the land and cloud areas of the composite were buffered to remove the high product values near islands and land where water is very shallow. The remaining product still had some faulty areas related to clouds and sea ice, especially in the Bothnian Bay in April. These were manually masked before publishing. An example is shown in Figure 8.

### 2.3.2 Access to data

Similarly to the monthly aCDOM values of ERGOM (Chapter 2.2) these dataset are available as follows:

- Images in the TARKKA service
- GeoTiff (\*.tiff) files in the EO-WAF (folder EO\_results)
- WMS interface

See above for details.



**Figure 8. Sentinel-3 OLCI one-month composite of aCDOM (at 400 nm) for May 2019 derived with a band ratio B8/B6 after atmospheric correction with the Baltic+ AC. Red dots are the ERGOM input locations.**

### 3 References

Brockmann, C., Doerffer, R., Peters, M., Stelzer, K., Embacher, S., Ruescas, A. (2016). Evolution of the C2RCC neural network for Sentinel 2 and 3 for the retrieval of ocean colour products in normal and extreme optically complex waters. European Space Agency, (Special Publication) ESA SP, SP-740

### Appendix 1: Instructions for accessing data in THREDDS

The following scrips accesses the monthly aCDOM composites in THREDDS and writes the result into GeoTiff files.

```
from pydap.client import open_url # NOTE: does not support proxy connections
import numpy as np
from rasterio.profiles import DefaultGTiffProfile
import rasterio
import datetime as dt

print('Logging on server...')

## INPUT AND OUTPUT PARAMETERS - MODIFY HERE

# URL to the Thredds endpoint
dataset_url = (r'https://thredds-iow.io-warnemuende.de/thredds/dodsC/Baltic'
+
              r'/IOW-THREDDS-Baltic_pocNP_V03R04_1nm_agg_time_monthly_2020-
03-18-12.nc')

# the time is indicated as 'whole days after this timestamp'
start_time = dt.datetime(2016, 1, 1, 0, 0)

# prefix for output files
file_prefix = 'IOW-THREDDS-Baltic_1nm_'

# postfix for output files, including suffix
file_postfix = '.tif'

# variable to use from the Thredds endpoint
variable_name = 't_cdom'

# scalar factor to multiply with, used e.g. for unit conversion
scalar_factor = 888000.0

# compress type that rasterio supports
compress = 'LZW'

# nodata value to be assigned in the output files (convert default values)
new_nodata = 0.0

# tiled or nope?
tiled = True

## DO NOT MODIFY AFTER THIS LINE
```

```
# open handle to the remote dataset and the variable of interest
dataset = open_url(dataset_url)
variable = dataset[variable_name][variable_name]

# declare this as the original nodata value of the gridded dataset
original_nodata = dataset[variable_name]._FillValue

# generate array of timestamps based on data and the starting date
time = [start_time + dt.timedelta(days=i) for i in dataset.time[:].data]

# define bounding boxes of georeferenced data
bbox_definitions = {'dx': (dataset.xt_ocean_sub01[1].data[0] -
                          dataset.xt_ocean_sub01[0].data[0]),
                   'x_min': dataset.xt_ocean_sub01[0].data[0],
                   'x_max': dataset.xt_ocean_sub01[-1].data[0],
                   'dy': (dataset.yt_ocean_sub01[1].data[0] -
                          dataset.yt_ocean_sub01[0].data[0]),
                   'y_min': dataset.yt_ocean_sub01[0].data[0],
                   'y_max': dataset.yt_ocean_sub01[-1].data[0]
                  }

# update GTiff creation profile
print('Creating GTiff profiles...')
prof = DefaultGTiffProfile()
prof.update(dtype=rasterio.float32, count=1)

transform = rasterio.Affine(a=bbox_definitions['dx'],
                            b=0.0,
                            c=(bbox_definitions['x_min'] -
                                0.5*bbox_definitions['dx']),
                            d=0.0,
                            e=-bbox_definitions['dy'],
                            f=(bbox_definitions['y_max'] +
                                0.5*bbox_definitions['dy']))

prof['transform'] = transform
prof['nodata'] = new_nodata
prof['tiled'] = tiled
prof['compress'] = compress

# go through all timesteps in dataset
print('STARTING LOOP over timesteps')

for i in range(variable.shape[0]):
    print('Reading timestep No. {0}'.format(i))

    try:
        # collect grid data at the particular timestep
        grid = variable[i, 0, :, :]
        grid = grid.data[0, 0, :, :]
    except:
        # If reading the data fails, the server has probably closed the
        # connection. Reinitialize.
```

```
dataset = open_url(dataset_url)
variable = dataset[variable_name][variable_name]
grid = variable[i, 0, :, :]
grid = grid.data[0, 0, :, :]

# replace original nodata values with new nodata values
grid[grid == grid.dtype.type(original_nodata)] = new_nodata

# apply scalar factor for unit conversion etc.
grid[grid != new_nodata] *= scalar_factor

# by convention, the data array is upside down: change the orientation
grid = np.flipud(grid)

# collect height and width for writing the data
height, width = grid.shape

# generate timestamp for output file
time_str = time[i].strftime("%Y%m%d%H%M%S")

# write output file
with rasterio.open((file_prefix + time_str + '_' +
                    variable_name + file_postfix),
                  'w', width=width, height=height,
                  crs='EPSG:4326', **prof) as dst:
    dst.write(grid.astype(rasterio.float32), 1)
```

## Appendix 2: Accessing monthly composites over WMS

The ERGOM monthly  $a_{CDOM}$  products are available through Finnish Environment Institute's (SYKE) GeoServer, which offers WCS, WFS, WMS and WMTS capabilities. The WMS endpoint is at <https://geoserver2.ymparisto.fi/geoserver/eo/wms?service=WMS&version=1.1.1> .

The GeoServer hosts the following ERGOM datasets:

- eo:MODEL\_ERGOM\_REGIONS (polygon) – Stations showing the input locations of the Ecological Regional Ocean Model (ERGOM).
- eo:MODEL\_ERGOM\_SOURCEPOINTS (point) – Areas used in the extraction of CDOM values derived from satellite data.
- eo:MODEL\_ERGOM\_WQ\_CDOM (raster) – Baltic Sea ecosystem model CDOM-results (monthly average of ERGOM acdom400, Ver 3, Run 3). Time dimension indicates the first day of the month of the average.

A list of the available months, along with other dataset-specific parameters, can be obtained in machine-readable form using the Capabilities document which can be requested with the link <https://geoserver2.ymparisto.fi/geoserver/eo/wms?SERVICE=WMS&VERSION=1.1.1&REQUEST=GetCapabilities> .

The WMS endpoint supports standard calls such as GetMap, GetCapabilities and GetFeatureInfo. The GetMap call is used for generating a map with the specified layers, image format, coordinate system, dimensions and geographical extent.

A typical GetMap call might have the following parameters supplied as a URL encoded query string:

- FORMAT – format for the map output (e.g., image/png)
- TRANSPARENT – whether the map background should be transparent (e.g., true)
- LAYERS – layers to display (e.g., MODEL\_ERGOM\_WQ\_CDOM)
- TIME – time value of data (e.g., 2019-01-01)
- ENV – product values to be passed to the band rendering method (e.g.,

S01:1;S02:3.5;S03:6;S04:8.5;S05:11;S06:13.5;S07:16

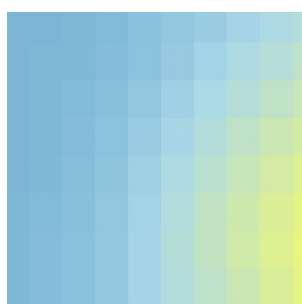
extends the fixed colormap from 1 to 16  $m^{-1}$  with the seven specified breakpoints)

- WIDTH – width of output in pixels (e.g., 256)
- HEIGHT – height of output in pixels (e.g., 256)
- SRS – the EPSG number of the requested spatial reference system (e.g., EPSG:3857)
- BBOX – map extent (minx,miny,maxx,maxy in units of the SRS)

The example request

[https://geoserver2.ymparisto.fi/geoserver/eo/wms?SERVICE=WMS&VERSION=1.1.1&REQUEST=GetMap&FORMAT=image%2Fpng&TRANSPARENT=true&LAYERS=MODEL\\_ERGOM\\_WQ\\_CDOM&TIME=2019-01-01&ENV=S01%3A1%3BS02%3A3.5%3BS03%3A6%3BS04%3A8.5%3BS05%3A11%3BS06%3A13.5%3BS07%3A16&WIDTH=256&HEIGHT=256&SRS=EPSG%3A3857&BBOX=2560609.1933018393%2C9329934.57040346%2C2593152.750688435%2C9362478.127790056](https://geoserver2.ymparisto.fi/geoserver/eo/wms?SERVICE=WMS&VERSION=1.1.1&REQUEST=GetMap&FORMAT=image%2Fpng&TRANSPARENT=true&LAYERS=MODEL_ERGOM_WQ_CDOM&TIME=2019-01-01&ENV=S01%3A1%3BS02%3A3.5%3BS03%3A6%3BS04%3A8.5%3BS05%3A11%3BS06%3A13.5%3BS07%3A16&WIDTH=256&HEIGHT=256&SRS=EPSG%3A3857&BBOX=2560609.1933018393%2C9329934.57040346%2C2593152.750688435%2C9362478.127790056)

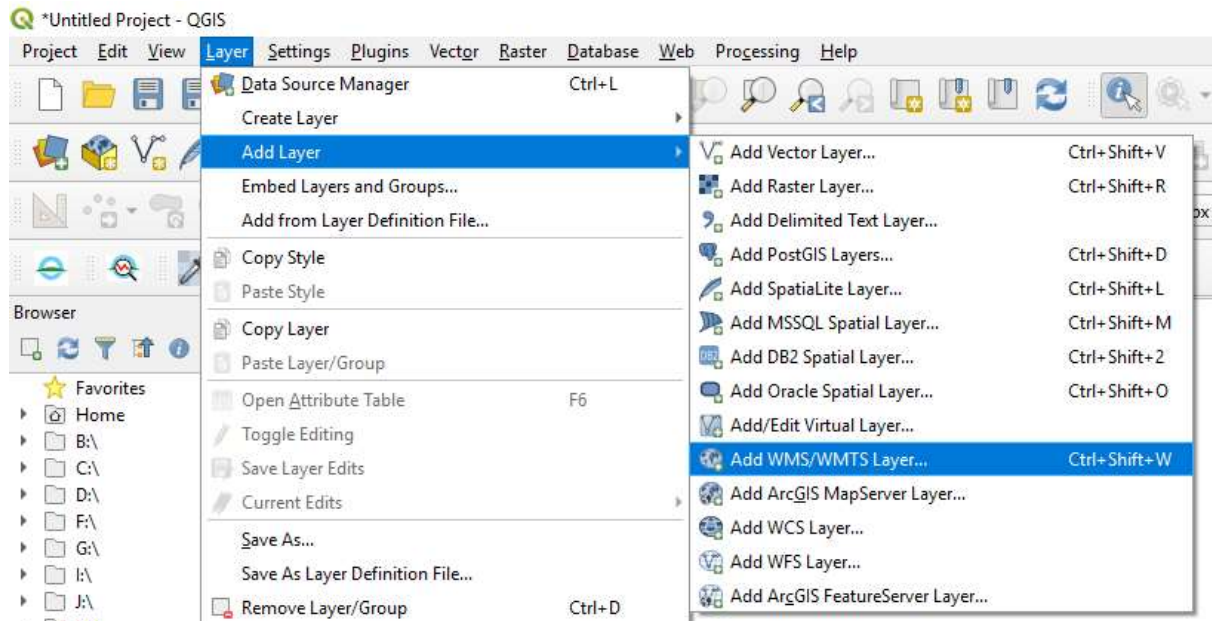
generates the following map:



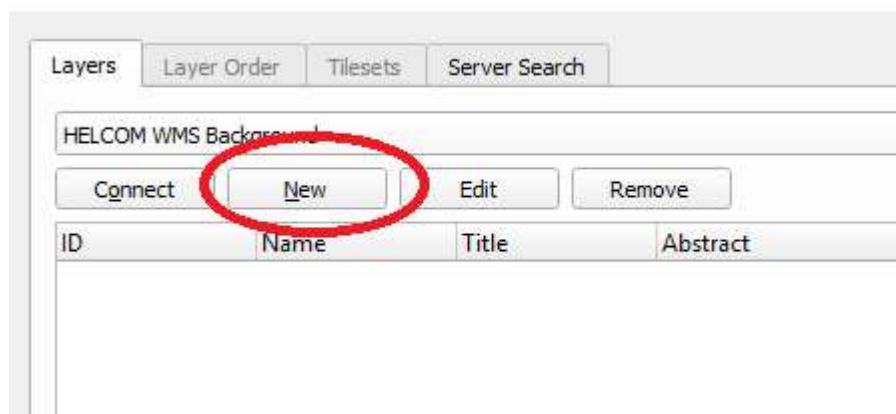


The monthly rasters can be accessed from QGIS 3.4 with the following instructions. The example describes the insertion of the January 2019 layer to the QGIS project. Modify the red strings to change the month.

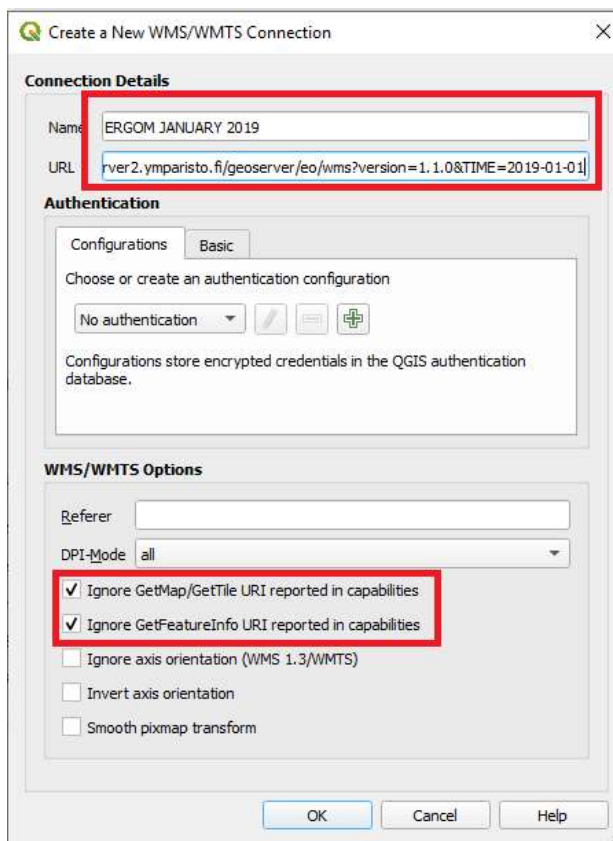
1. From **Layer** menu, choose **Add Layer** → **Add WMS/WMTS layer** or use the keyboard shortcut **Ctrl+Shift+W**



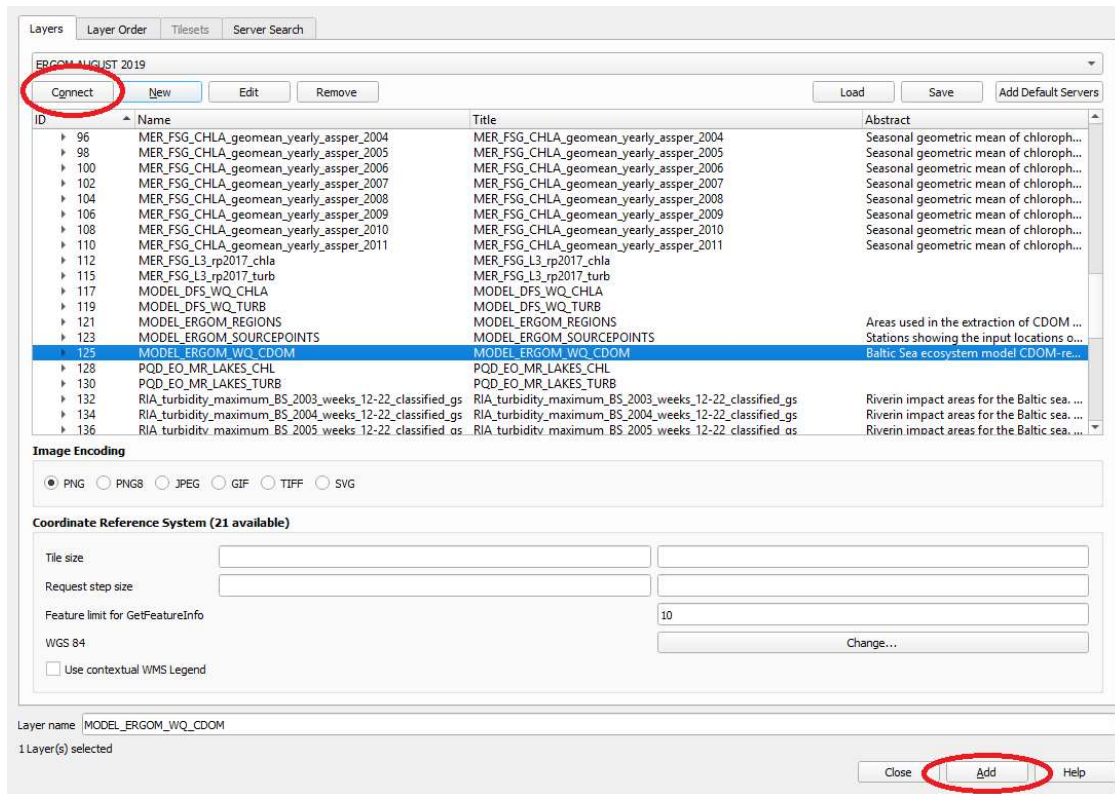
2. From the Layers tab, click **New**. A popup window opens asking for the connection details.



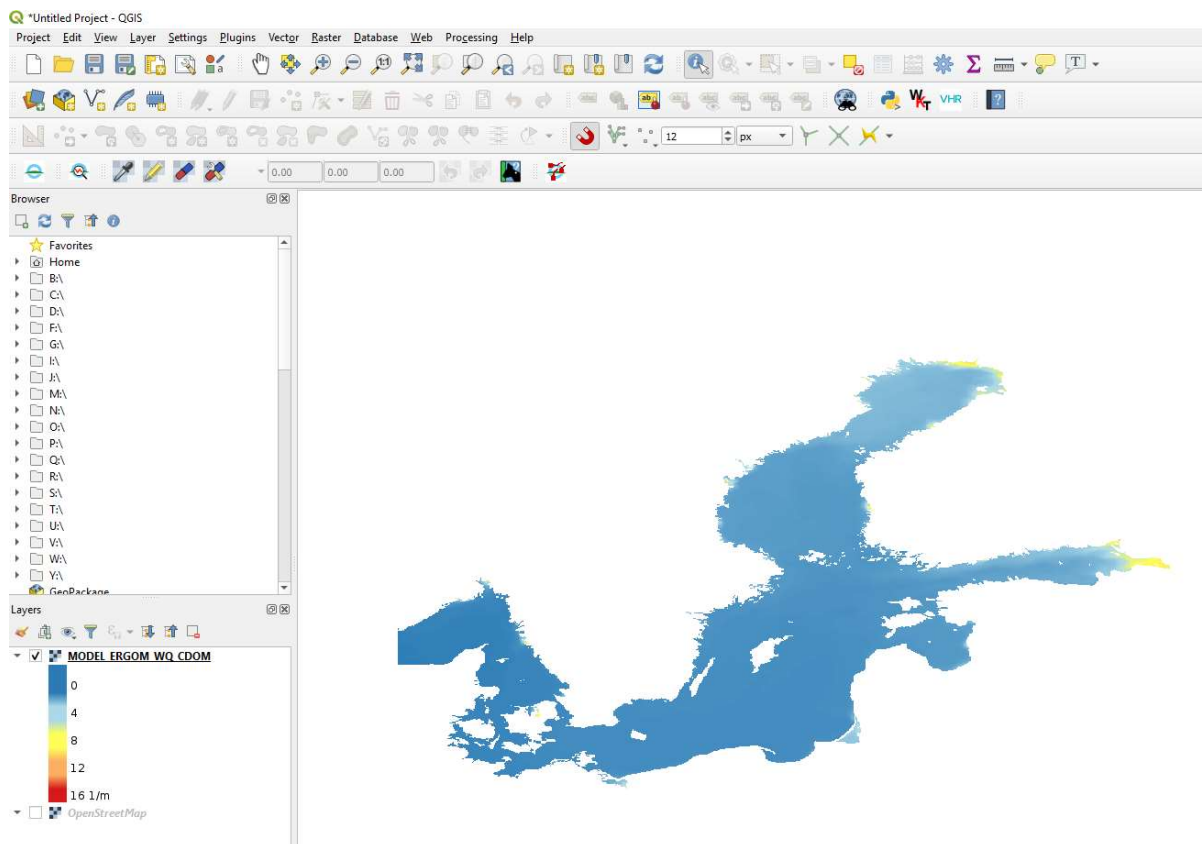
3. Use the following connection details:
  - a. Name = ERGOM **JANUARY 2019**
  - b. URL = <https://geoserver2.ymparisto.fi/geoserver/eo/wms?service=WMS&version=1.1.0&request=GetMap&TIME=2019-01-01>
  - c. Check “Ignore GetMAP/GetTile URI reported in capabilities” and “Ignore GetFeatureInfo URI reported in capabilities”.



4. Click **OK** and **Connect**. Choose the layer eo:MODEL\_ERGOM\_WQ\_CDOM in the list of available layers and click **Add**.



5. The January layer should now successfully be inserted into the QGIS project.



This method only adds one month at a time to the project, as QGIS does not natively support time-stepping WMS layers.

There are some third-party addons that can be used to provide the capability of stepping through WMS-T layers, such as [TimeManager](#) (versions above 2.1.5). Karolina Alexiou has described its use in the tutorial “[Using TimeManager for WMS-T layers](#)” (2015-08-10).