
| | |
|---|---|
| **ESA Contract No.** | 4000135758/21/I-EF |
| **Project Name** | BIOMONDO |
| Towards Earth Observation supported monitoring of freshwater biodiversity | |

| | |
|---|---|
| **Deliverable** | D2.2 Algorithm Theoretical Baseline Documents |
| **Short description** | A detailed specification of the final versions of the algorithms/models. |
| **Work package No.** | WP2 |
| **Lead Partner** | BC |
| **Distribution** | ESA, Project team, AB, BEs, EAs |
| **Version** | V2.1 |
| **Submission date** | 2023-06-30 |
| **Contributors** | Jorrit Scholze (Brockmann Consult)<br>Carsten Brockmann (Brockmann Consult)<br>Tamara Keijzer (PBL)<br>Miguel Dionisio Pires (Deltares)<br>Tineke Troost (Deltares)<br>Petra Philipson (Brockmann Geomatics)<br>Susanne Thulin (Brockmann Geomatics)<br>Jelle Lever (EAWAG)<br>Daniel Odermatt (EAWAG) |

# Table of Contents

# List of Figures

# List of Tables

# 1 Introduction

Freshwater ecosystems are some of the most important ecosystems on the planet, providing a range of ecosystem services to humans, including clean water, food, and recreation. They are also home to a vast array of biodiversity, including many species that are found nowhere else on earth. Unfortunately, freshwater ecosystems are under threat from a range of human activities, including pollution, habitat destruction, and climate change. Current biodiversity policies and strategies as well as assessments of progress towards set targets, point out that there has been a general failure to halt the negative trend of biodiversity loss and that different approaches are needed to reverse the situation. This includes revision of targets and the indicators that inform the targets and a greater emphasis on the links between biodiversity, ecosystems and their services and people.

The European Space Agency (ESA) activity called Biodiversity+ Precursors is a contribution to the joint EC-ESA Earth System Science Initiative launched in February 2020 to jointly advance Earth System Science and its response to the global challenges that society is facing. The ESA Biodiversity+ Precursors include four projects on different themes and BIOMONDO is the freshwater project, and has a focus on biodiversity in lakes, wetlands, rivers, and streams.

BIOMONDO aims to improve our understanding of freshwater biodiversity around the world and to support freshwater biodiversity monitoring through development of solutions that integrate EO data and state-of-the-art biodiversity modelling using advanced data science and information and communications technology. Three BIOMONDO Pilots have been developed and will demonstrate how novel Earth Observation and Biodiversity modelling products can be integrated to enhance scientific understanding and support decision systems for biodiversity monitoring addressing policy priorities such as the EU Biodiversity Strategy for 2030.

To develop a broad outlook on ongoing changes in freshwater biodiversity and how these changes can be monitored using EO data, our three BIOMONDO pilots each address pilot objectives and knowledge gaps corresponding to one of the following three drivers of global environmental change in freshwater ecosystems: ‘pollution and nutrient enrichment’ (Pilot 1 - Eutrophication), ‘climate change’ (Pilot 2 – Heat Tolerance of Fish), and ‘habitat change’ (Pilot 3 – River Connectivity). The resulting combination of data from Earth Observation, in-situ measurements and model outputs is made available to the scientific and policy community through the BIOMONDO Freshwater Biodiversity Laboratory.

# 2 Scope of this document

Within the BIOMONDO project, data from various sources has been collected and own data products have been produced to develop three pilot studies for freshwater system biodiversity observation and monitoring.

This document is a description of the novel methods and products developed as well as the specification of the models and methods used for the BIOMONDO Freshwater Laboratory. The novel methods include new algorithmic approaches, new models and interactions and combinations of different datasets and results. Furthermore, the three BIOMONDO pilots are descripted in detail and the results are discussed.

The structure of the document includes the description of the BIOMONDO approach, followed by the description of the novel methods and their implementation. The last chapter includes the application and use of the methods within the three pilots and showcase the applicability. For publicly available operational services, a description and literature references are given. For own processed data sources and models an evaluation was performed during the data collection phase, and validation results to justify the final selection are provided in D2.3. For a more detailed overview look into related documents which are described in Table 1.

Table 1 Reference documents

| **Document** | **Version** | **Short description** |
|---|---|---|
| Development Database | D2.1 v.2.1 | A detailed structure of all input datasets and how to access them. Can made be available on request. |
| Product Validation Report | D2.3 v2.0 | A detailed specification of the validation methods, metrics and description of the reference data. |
| Experimental Datasets | D2.4 v1.0 | A detailed description of the Experimental Datasets and how to explore them. |

# 3 BIOMONDO Approach

The BIOMONDO approach includes the development and implementation of the technical and practical basis in the BIOMONDO Freshwater Laboratory. An overview of the BIOMONDO Lab is given in Figure 1. On the left side we have the EO based inputs. In green are satellite data which require own thematic processing, e.g. to detect dams and other human disturbances, or S2 based high resolution water quality products. In brown are EO based products which have already been processed and which are available from operational services. On the bottom right side in purple are non-raster data including information from biodiversity databases, in-situ measurements, and citizen science activities. On the top right, red colour, are ecosystem and biodiversity models at different spatial scales. In orange on the center right side are the user interfaces and data analysis APIs. In the center are the BIOMONDO datacubes which combine all these inputs and are made accessible and processable. We almost completely rely on existing building blocks (EO operational services, xcube, xcube viewerand Delft3D-ECO models, Jupyter Lab), but have adapted the required methods to extended data sources, established the interactions between these data sources and developed a combination of data sources into thematic ecosystem change indices (TECIs). These novel methods are described in this ATBD and are show in Figure 1 within the grey boxes. The novel methods include novel EO products and applications, a heat tolerance model for fish species, the implemented use of EO data within models and the resulting data cube-model loop, which feeds model outputs back into a data cube, and the methodology and implementation of the **T**hematic **E**cosystem **C**hange **I**ndices. Already existing algorithms like the Delf3D model, the Connectivity Model or existing EO operational services will not be described in detail, but their use and reference will be outlined in the description of the pilots in Chapter 5.

The BIOMONDO Freshwater Lab is set-up with access to all required inputs and interactions for the biodiversity pilots and will be kept in operation for the potential use by the biodiversity experts and early adopters within WP3 and WP4.

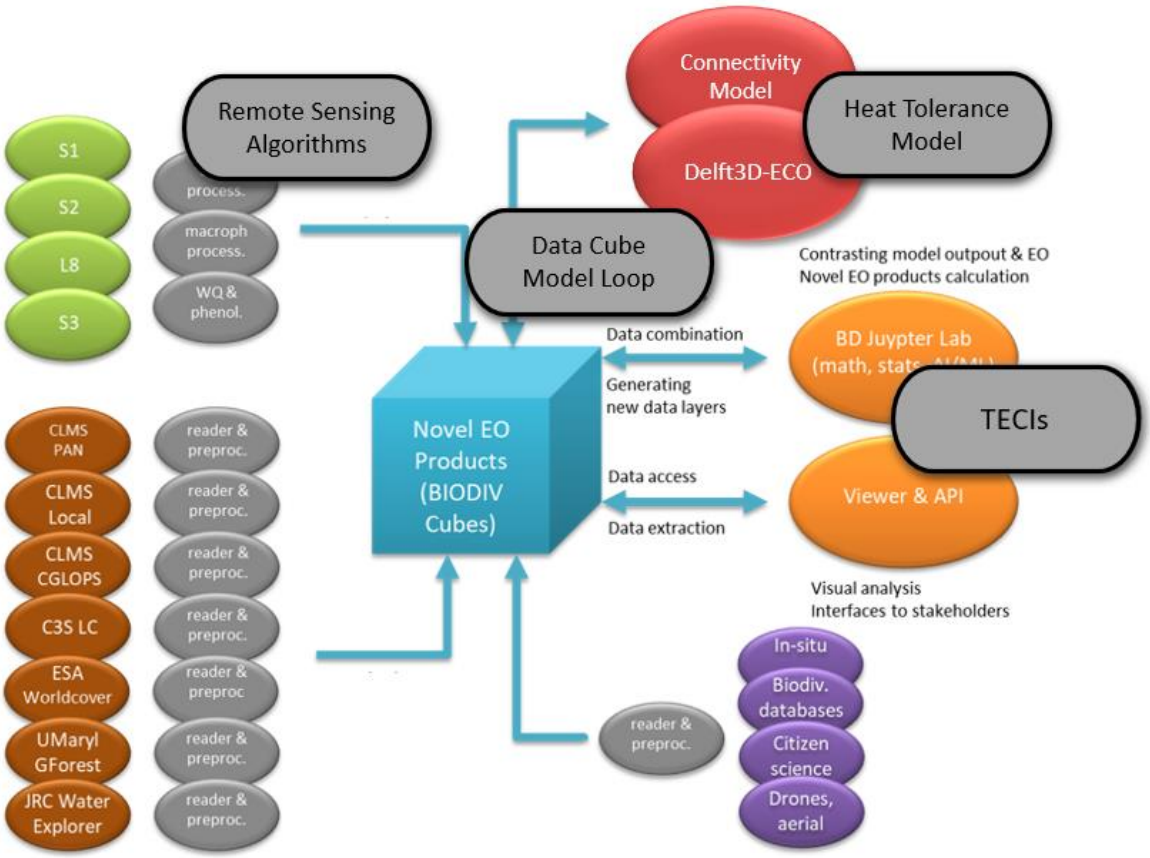



Figure 1 Concept of the BIOMONDO Laboratory with the novel methods described in this ATBD.

The BIOMONDO Freshwater Laboratory is a federation of all the data and is made accessible via state-of-the-art methods (e.g. JupyterNBs, xcube Viewer). Visualisation interfaces and export function complements the BIOMONDO Lab. Providing access to the Lab is an important task within this project and we believe that this way of enabling scientists and other users to get access to novel EO products and models is the best way to proceed. It is not the scope of this precursor project to implement a fully-fledged Biodiversity Lab but to test critical concepts and provide recommendations for the roadmap. We will use existing software, namely the xcube ecosystem of python tools and services, to implement basic functionalities (more information about the xcube software can be found here: https://xcube.readthedocs.io). This will already be suitable to provide the BIOMONDO Lab to biodiversity experts of our pilot sites and the Early Adopters, to work with the EO data and model results in an agile manner. More information about the BIMONDO Freshwater Lab can be found in the D2.4 document (Table 1).

# 4 Algorithms and Methods

Remote sensing and modelling are powerful tools for monitoring and understanding freshwater biodiversity. By providing researchers with the ability to gather data on a large scale and simulate different scenarios, these methodologies are helping to improve our understanding of the complex and interconnected nature of freshwater ecosystems and how best to protect and conserve them.

## 4.1 Remote Sensing Algorithms

Earth observation data is essential for studying and monitoring freshwater biodiversity because it provides information on the spatial and temporal dynamics of aquatic ecosystems. EO data provides information on various environmental factors such as water quality, temperature, and land use. For the BIOMONDO project the focus is set on operational EO services and often used and generally accepted EO algorithms to feed the BIOMONDO Freshwater Lab with EO data. Nevertheless, a few novel methods and applications were developed to support the Pilots in this project.

### 4.1.1 Water Quality Retrieval

Water quality retrieval using remote sensing data is important for freshwater biodiversity monitoring because it provides valuable information about the physical and chemical properties of water that can impact the survival and health of freshwater organisms. Remote sensing data can be used to estimate several water quality parameters, such as chlorophyll-a, suspended solids, turbidity, and dissolved organic matter, which can help researchers to understand the state of the ecosystem and the potential impacts of environmental stressors.

#### 4.1.1.1 C2RCC with C2X-Complex Nets

The C2RCC (Case 2 Regional Coast Colour) processor is used for the atmospheric correction and retrieval of water quality parameters from Sentinel-2 satellite imagery. It is specifically designed for processing imagery of coastal and inland water bodies, where the water is not perfectly clear and contains a variety of constituents, such as dissolved organic matter, suspended sediments, and phytoplankton. The C2RCC algorithm (Doerffer and Schiller, 2007, Brockmann et al., 2016) has evolved from the "Case 2" algorithm originally developed for MERIS and is now based on a suite of neural networks. A water-inherent optical properties model and a coupled water-atmosphere radiative transfer model are used to generate a very large number of water reflectance and TOA reflectance spectra from different combinations of water constituent concentrations and different atmospheric conditions. These simulated TOA reflectance spectra are used to train a neural network that can then quickly invert each input TOA reflectance spectrum to present optimal estimates of water reflectance and water constituent concentrations. The C2RCC algorithm shows robust performance under a range of water and atmospheric conditions,

including extremely absorbing and extremely scattering water. A limitation of C2RCC is that the method assumes that the water reflectivity matches the spectra that were present in the neural network training set. C2RCC is a coupled algorithm because it calculates both atmospherically corrected water reflections and the water constituents. In the context of BIOMONDO, C2RCC is used as a coupled algorithm for Sentinel-2 applying the C2X-COMPLEX nets. These neural networks were trained with a dataset representing optical models of complex water bodies, covering the properties of inland waters and their concentration ranges for chlorophyll-a concentration and turbidity. The novelty of this method within the BIOMONDO project is the application of the C2RCC with C2X-Complex nets to the unknown Mekong River waters. The C2X-Complex net was trained with sampled data from inland waters of Germany and never tested for the high turbid und unknown waters within the monsoon season in Asia and the Mekong. The application of the C2RCC with C2X-Complex nets proved to be very reliable for the volatile river water constituents. Visualizations and validation of the datasets can be found in the BIOMONDO D2.3 document (Table 1).

#### 4.1.1.2 Forel Ule Index

Increasing brownification of inland waters and rivers have ecological and social consequences (Kitzberg et al., 2020). The brownification has profound physical, chemical and biological repercussions for aquatic ecosystems affecting water quality, biological community structures and aquatic productivity. Potential drivers of brownification trends are complex and include reductions in atmospheric acid deposition, changes in land use/cover, increased nitrogen deposition and climate change (Meyer-Jacob et al., 2019). The Forel-Ule (FU) color index has been calculated to assess the watercolor within the Mekong and its side arms. The FU scale was initially developed as a color index that would allow the visual classification of the watercolor (Wernand and Van Der Woerd, 2010) and divides water reflectance spectra into 21 color classes from dark blue to yellowish-brown. The classification into the 21 colors is based on a calculated hue angle. For Sentinel-2 the calculation of the FU value uses the water-leaving reflectance ($R_{rs}$) data calculated with the C2RCC. We used the FU approach to investigate the hypotheses that the watercolor changes due to the increasing number of dams in the Mekong basin, resulting in reduced sediment transportation (Schmitt et al., 2018). The results of the FU calculation can be found in the BIOMONDO D2.3 document (Table 1).

### 4.1.2 LSWT Gap filling with DINEOF

Remote sensing data can be used to estimate lake surface temperature by measuring the amount of thermal radiation emitted by the water surface. This information is used to create a continuous map of lake surface temperature and to monitor changes over time and across space. However, remote sensing data may have limitations, such as cloud cover or poor spatial resolution, which can limit its accuracy.

Daily LSWT gap filled L4 products have been generated using the DINEOF (Data Interpolating Empirical Orthogonal Functions) approach, which reconstructs missing data in geophysical datasets by using a truncated Empirical Orthogonal Functions (EOF) basis in an iterative approach. DINEOF reconstructs missing data in a geophysical dataset by extracting the main patterns of temporal and spatial variability from the data. DINEOF has been successfully applied to sea surface temperature (e.g. Alvera-Azcárate et al., 2005),

chlorophyll-a and winds (e.g. Alvera-Azcárate et al., 2007), and total suspended matter (e.g. Alvera-Azcárate et al., 2015). While originally designed for low resolution data products, recent research has resulted in the optimization of DINEOF to handle high resolution data provided by Sentinel-2 MSI, including cloud shadow detection (Alvera-Azcárate et al., 2020). As data source the ESA Lakes v2.0.1 LSWT was used and interpolated for Lake Marken and Lake Mälaren. The L4 LSWT products are used for the Heat Tolerance Model described in Chapter 4.2. The validation of the L4 LSWT products are available in the BIOMONDO D2.3 document (Table 1).

### 4.1.3 Phenology

Phenology refers to the seasonal timing of recurring biological events, such as the emergence of leaves, the flowering of plants, or the migration of animals. In freshwater ecosystems, phenology is an important driver of biodiversity and ecosystem function, as it influences the timing and availability of resources for aquatic species. For example, the timing of the spring thaw and the onset of snowmelt can influence the timing of breeding and reproduction for many fish and amphibians, while changes in the timing and duration of algal blooms can affect food availability for zooplankton and other aquatic consumers and are one of the most visible effects of lake eutrophication. Phenology is also sensitive to changes in climate and other environmental factors and can therefore be used as an indicator of ecosystem health and change over time.

Remote sensing data can provide valuable information on the phenology of freshwater ecosystems, allowing for the monitoring and analysis of changes in the timing and intensity of biological events over time. To study these changes, metrics of phenology were extracted from daily chlorophyll-a concentrations (on a pixel level) with a method that roughly follows Grey *et al.* (2019), i.e. the method used for the extraction of global land-surface phenology for MODIS Land Cover Dynamics product (MCD12Q2.v6.1). More specifically, periods of sustained in- and decreases in chlorophyll-a concentration were identified that occurred over a minimum period of 16 days and of which the amplitude corresponded to at least 20% of the maximum amplitude of change in chlorophyll concentration observed in each year (the minimum period and amplitude were varied to see whether this affects results). For each of these periods the day at which a "peak" in chlorophyll concentration was then determined (i.e. when the first derivative changes sign), as well as a variety of related metrics of phenology such as the day of "peak onset" and "mid-greenup", and "peak termination" and "mid-greendown" defined as the first day at which chlorophyll-a concentrations cross 15 and 50% of the in- or decrease preceding or following each peak (see Figure 2.a-b). In addition to this, we also extracted metrics of phenology in ways that are more commonly used in oceanography in which case periods of sustained in- and decrease in chlorophyll concentration were defined as periods in which the chlorophyll concentrations were above 1.05 times the median concentration in each year. Once these metrics of phenology were determined for each pixel, we then proceeded by determining the timing of "recurring algal blooms" which are defined as a cluster of peaks of which the period of in- and decrease in chlorophyll-a concentration overlaps strongly between years and between pixels. Trends of change in the timing of these blooms were then studied by studying changes in the median day of the year at which a peak occurred across all peaks belonging to a bloom (see Figure 2.c-d).

Natural variability between years can make it difficult to determine trends of change in phenology when time-series are short. As a data source, we therefore used mean chlorophyll-a concentrations from ESA Lakes v2.0.2 which combines data from different sensors. We checked for consistency across sensors and determined trends of change in the timing of recurring algae blooms on a lake level with a focus on changes occurring in Lake Marken and 38 other lakes for which MERIS, MODIS-Aqua, and OLCI data are available.

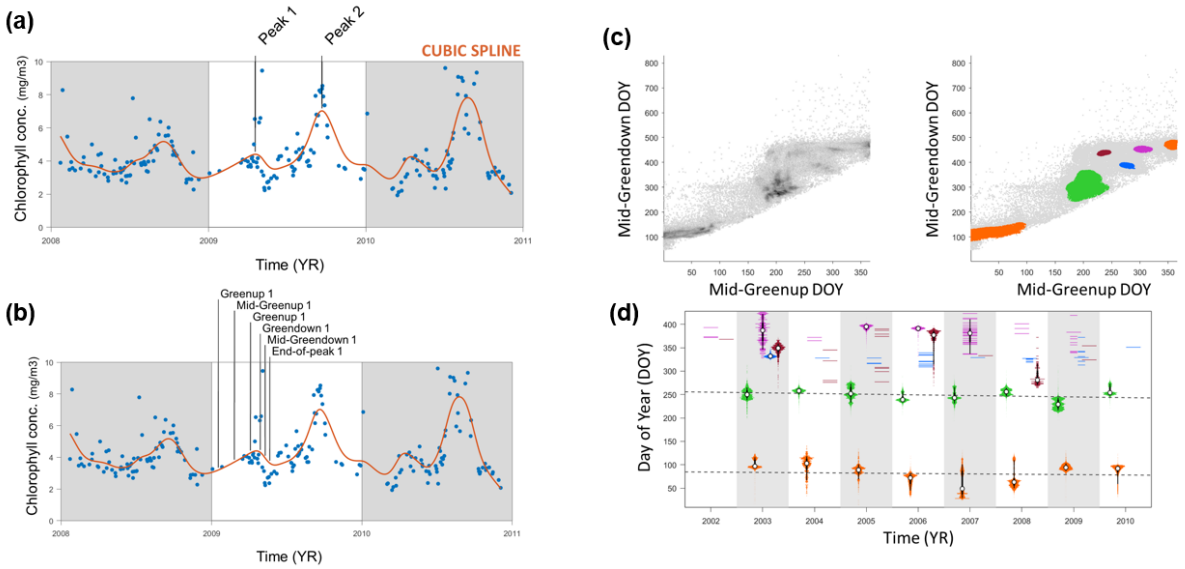



Figure 2 A novel method to determine changes in the timing of algae blooms. (a) A cubic spline (orange line) is fitted through daily estimates of Chlorophyll-a concentration (blue dots) for each lake pixel, and the timing (day of year) of peaks in chlorophyll-a concentration determined. **(b)** Related metrics of phenology are determined for each peak. **(c)** Peaks of which the period of in- and decrease in chlorophyll-a concentration shows a strong overlap with peaks in other pixels and/or in other years are clustered. We call these clusters of peaks recurring algae blooms. **(d)** Trends of change in the timing of these algal blooms over multiple years are determined.

# 4.2 Models

Models are recognized as crucial for biodiversity monitoring and assessment because they can help to synthesize and integrate large amounts of complex data, making it easier to identify patterns and trends in biodiversity across different spatial and temporal scales. They can be used to identify the drivers of biodiversity change, predict future trends, and assess the effectiveness of different management strategies. Combining models with remote sensing data is often necessary to fully understand biodiversity patterns and trends. In the following chapter the models used within BIOMONDO are described. Whereas the Delft3D and the River Connectivity model are well known models and do not count into the novel methods of the BIOMONDO project the Heat Tolerance model is stated as a novel method and was implemented and tested for the first time within the BIOMONDO project.

## 4.2.1 Delft3D

Delft3D is a 3D modelling suite to investigate hydrodynamics, sediment transport and morphology, and water quality for fluvial, estuarine and coastal environments. These models require a large set of input data, and their performance heavily relies on the

accuracy and resolution of these input data. The model performance may thus be improved by improving the input data, for example by using EO based products, which has been explored and validated within BIOMONDO. These input data include the meteorological forcings over space and time, as well as the initial values of the state variables. The selected model application for the Markermeer is based on the open-source Delft3D software of Deltares. Delft3D models are process-based models that are often used in a water management context. They combine different modules such as hydrodynamics, waves, and biogeochemistry that can be interconnected. More details are given in the paragraphs below.

**Hydrodynamics (Delft3D-FLOW)**

Delft3D-FLOW is a multi-dimensional (2D or 3D) hydrodynamic (and transport) simulation program which calculates non-steady flow and transport phenomena that result from tidal and meteorological forcing on a rectilinear or a curvilinear, boundary fitted grid.

Typically, the output of the Delft3D-FLOW module provides not only information on water transport, but also on salinity and temperature, but for this pilot only output on transport was generated.

Boundary conditions/forcing functions required for the hydrodynamic model:

- bathymetry
- air temperature
- irradiation
- wind speed and direction
- water discharges into the system (based on local measurements)

**Water Quality (Delft3D-WAQ)**

DELWAQ is the computational engine of the D-Water Quality and D-Ecology programmes of the Delft3D suite. It is based on a rich library from which users and developers can pick relevant substances and processes to quickly put water and sediment quality models together.

State variables:

- inorganic nutrients in the water column (NO3, NH4, PO4, Si)
- dead organic material (C,N,P,Si) in both the water and the sediment
- algae in the water column (various species)
- oxygen

Not explicitly included are floc formation, sulfur bacteria, aquatic plants, shellfish, zooplankton (algae grazing by zooplankton is implicitly taken into account in the parameter setting of algal mortality).

Considered processes:

- algae processes (growth and mortality, respiration, competition, etc)
- remineralization (in water column and soil)
- (de)nitrification (in water column and soil)

- light extinction (due to silt, and other living & dead material in the water phase)
- settling (of algae and detritus)
- reaeration

Required boundary conditions/forcing functions:

- water transport (from hydrodynamic model)
- water temperature (based on in-situ measured air temperature)
- irradiation
- wind speed
- nutrient concentration of incoming water (based on local measurements for the year 2016)
- suspended silt concentration (based on dynamically modelled silt field for the year 2016)

The model output is provided in the form of a netCDF-file containing various variables related to primary production e.g. chlorophyll a, light extinction, algae concentrations for different species (groups), and the total net primary production flux itself, as shown in Table 2.

Table 2 Parameters of the Delft3D model output

| Variable Name | Dimen-sions | Description |
|---|---|---|
| mesh2d_Chlfa | time, lat, lon | Chlorophyll-a concentration (mg/m3) |
| mesh2d_ExtVl | time, lat, lon | total extinction coefficient visible light (1/m) |
| mesh2d_fPPtot | time, lat, lon | total net primary production (gC/m2/d) |
| mesh2d_FDIATOMS | time, lat, lon | Freshwater Diatoms** (gC/m3) |
| mesh2d_FFLAGELA | time, lat, lon | Freshwater flagellates** (gC/m3) |
| mesh2d_GREENS | time, lat, lon | Green algae** (gC/m3) |
| mesh2d_ANABAENA | time, lat, lon | Anabaena*(gC/m3) |
| mesh2d_APHANFIX | time, lat, lon | N-fixing Aphanizomenon* (gC/m3) |
| mesh2d_OSCILAT | time, lat, lon | Oscillatoria* (gC/m3) |
| mesh2d_MICROCYS | time, lat, lon | Microcystis* (gC/m3) |

* Cyanobacteria, **Other algae

### 4.2.2 Heat Tolerance Model

Fish, like all organisms, have a range of temperatures within which they can survive and thrive. Heat tolerance refers to the upper limit of this range, beyond which fish may experience stress or mortality.

To model freshwater fish species' physiological tolerance to water temperature, we retrieved species-specific data on heat tolerance (critical thermal maximum) from existing databases (Comte and Olden, 2017; GlobTherm by Bennett et al., 2018; Leiva et al., 2019) and a systematic literature review. Experimental heat tolerance values were standardized to a duration of 1 hour to correct for differences in experimental methodology.

Next, we established a novel phylogenetic regression model that estimates the heat tolerance of a given species as a function of the species' morphological and ecological properties, their phylogeny, and their ability to acclimate to heat. This approach accounts for the non-independence of data points, i.e., multiple observations per species, and the phylogenetic similarity between species (Figure 3A). We used the R package brms. Realm and length traits were collected from Fishbase (Froese & Pauly, 2022). Maximum habitat temperature, habitat temperature variability and range size were retrieved from Barbarossa et al. (2021). We obtained the phylogenetic tree from the Open Tree of Life (OTL; https://tree.opentreeoflife.org/, tree version: opentree13.4). To incorporate phylogenetic relatedness in the predictions of species not present in the training data, we imputed the phylogenetic group-level effect for these species using the group-level effect of the species in the training data (using Rphylopars; Goolsby et al., 2017).

We tested the predictive ability of different model configurations (i.e. variable included or excluded) based on phylogenetic blocking cross-validation (see Roberts et al., 2016) to evaluate how it performs for species that were not included in the model training. The phylogenetic blocking approach accounts for the introduction of new species with large phylogenetic distance from the species in the model training data. We divided the data over 10 folds (i.e. groups of species) using this approach (Figure 3B). Each fold is used for test data, while the rest is used to train the model. For the model testing, we removed observations if data was not available on one of the species traits, or if a species was not present in the phylogenetic tree. This resulted in a total of 2,406 observations across 334 species.

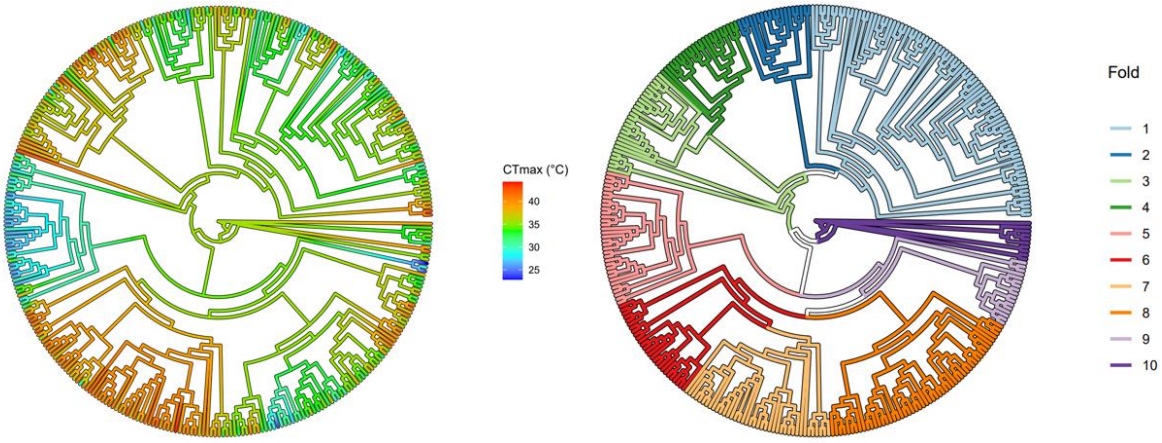



Figure 3 The phylogeny of the 334 fish species in the experimental dataset. A) Colours reflect variation in heat tolerance (CTmax; species average). The variation in heat tolerance shows a phylogenetic signal (lambda=0.89). B) The division of species over the folds for cross validation.

From the cross-validation results, the model with the lowest RMSE averaged across the folds was chosen (Table 3). This model calculates heat tolerance using species-specific

information (maximum habitat temperature and random effects group-level value) of the species occurring at a site and acclimation temperature. As expected, the model performs better for species included in model training (RMSE = 1.79, MAE = 1.27), compared to species not included in model training (RMSE = 3.16, MAE = 2.50). It can explain 87.9% of the variation in the experimental heat tolerance data. The fixed effects alone explain 48.5% of the variation. The model was then trained using all experimental data to predict heat tolerance of fish species. Using this model, heat tolerance can be predicted for 10,543 freshwater fish species.

Table 3 Average estimates and 95% posterior credible intervals of Bayesian posterior distributions for fitted parameters of the heat tolerance model. Bayesian p-value based on the density at the maximum a posteriori (MAP) are also given.

| **Parameter** | **Mean estimate** | **2.5%** | **97.5%** | ***p*(MAP)** |
|---|---|---|---|---|
| Random effects | | | | |
| Phylogeny | 4.44 | 3.08 | 6.08 | |
| Species | 1.31 | 1.11 | 1.51 | |
| Fixed effects | | | | |
| Intercept | 20.89 | 16.80 | 25.06 | <0.001 |
| Maximum habitat temperature | 0.31 | 0.30 | 0.33 | <0.001 |
| Acclimation temperature | 0.24 | 0.18 | 0.30 | <0.001 |

### 4.2.3 River Connectivity

We studied historical and upcoming changes in river connectivity following a procedure co-developed by members of BIOMONDO and described in Barbarossa *et al.* (2020). This procedure results in an assessment of the degree of geographic range fragmentation across the entire Mekong basin, expressed as a connectivity index (CI) (range 0-1) where 1 represents a range that is fully connected and smaller values indicating less connectivity. To calculate the connectivity index, species range data and dam locations are required as inputs. We used species range data from IUCN and point occurrence records from multiple sources (see Barbarossa et al., 2020). For the Mekong basin this included the geographical ranges of 783 lotic fish species. We used the dams data (n=107) from Schmitt et al. (2019), which includes info on location, status (Existing, Planned or under Construction), commercial operation date, installed capacity, and dam height. The species ranges were referenced to HydroBASINS subbasins (Pfaffstetter level 12). With the Pfaffstetter level 12 we used the highest level of spatial definition available, i.e., the smallest sub-basin units. Each of the sub-basins carries information on the connectivity to the next downstream sub-basin, which allows to determine the total connected area within a main hydrologic basin. Dams falling within a sub-basin were georeferenced to the downstream boundary of that sub-basin so that isolated patches were a collection of HB sub-basin units.

The connectivity index is calculated based on the approach by Cote et al. (2009). First subbasin area is converted to stream length using Hack's law. Then the total amount of stream length occupied by a species in each isolated patch is calculated. The CI is calculated by squaring these values and summing them and dividing this by the squared total amount of stream length a species occupied in the basin. For details and equations see Barbarossa et al. (2020). In general, the more isolated fragments, the lower the connectivity (Figure 4). We differentiated between the effect on fragmentation on non-diadromous and diadromous fish species (i.e. fish that travel between salt and freshwater), as we assume the most downstream dam to affect the connectivity of diadromous fish species to a much higher extent. Therefore, the CI for diadromous species is dependent on the geographic range belonging to the most downstream isolated patch connected to the ocean (Figure 4). For diadromous fish species the CI is calculated by dividing the total amount of occupied stream length in the patch connected to the ocean by the total amount of occupied stream length in the basin. We aggregated the species-specific CI values over all species or species groups (diadromous/non-diadromous) to give an overall indication of connectivity in the basin.

Furthermore, we explored the change in connectivity by using different dam subsets as input. We calculated the change in CI over time based on the commercial operation date of the dams. We also explored the effect of the removal of each individual dam on the CI.

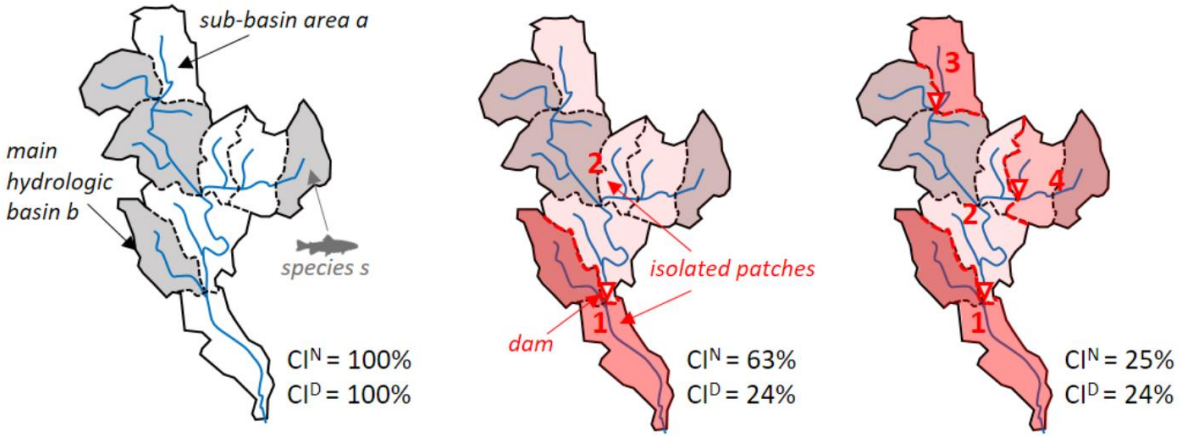



Figure 4 CI calculation for a hypothetical species s (occupying the gray areas) in a fictitious basin b (external solid line) partitioned in Hydrobasins subbasins (internal boundary dashed lines). The addition of dams fragments the basin in isolated patches (red hues with numbers). For each configuration, the CI is given for species s being either diadromous or non-diadromous. Note that the CI would not change for a diadromous species between the center and the right panel, even though the right panel contains more dams, as the connectivity for diadromous species is controlled by the most downstream dam (Barbarossa et al., 2020).

## 4.3 Data Cube – Model Loop

Biodiversity modelling is key to understand cause-effect relationships as well as analysing potential future trends through scenario experiments. We are linking models to primary EO data, by using those as driving forces and for parameterization.

We combine the primary data from Earth Observation and the model in the BIOMONDO Freshwater Laboratory. The Data Cube – Model Loop is the novel method described in this

ATBD functioning as the interactive part between the Data Cube storing the data and making the needed variables available to the Model. After running the model, the output variables will be implemented back into the Data Cube.

### 4.3.1 Loop Architecture

The scope of this precursor project is to implement, provide and test an initial BIOMONDO Freshwater Lab. We are using existing software, namely the xcube ecosystem of python tools and services, to implement basic functionalities, e.g. model and visualization interfaces.

Jupyter Notebooks are an interactive and browser-based tool to process and visualize data via a programming language. The good structurability of the Jupyter Notebooks with code cells and additional text cells for documentation enables expert users to carry out data analyses. The programming language we use for the notebooks is Python 3.8. The BIOMONDO data cubes can be accessed via an interface from the Jupyter Notebooks to access the data and use it as model inputs. The outputs of the models are stored in separate data cubes or are extending existing data cubes. This depends on the model output and the later use of the data cube. The final data cubes will be made available via the BIOMONDO Freshwater Lab.

Figure 5 shows an example of the loop architecture for the heat tolerance model. The complete loop is run in a Python Jupyter Notebook environment and a daily LSWT cube is used as input to the model. After the model is successfully run for various species the ouput is the input cube including the heat tolerance datasets for all available species in a data cube.

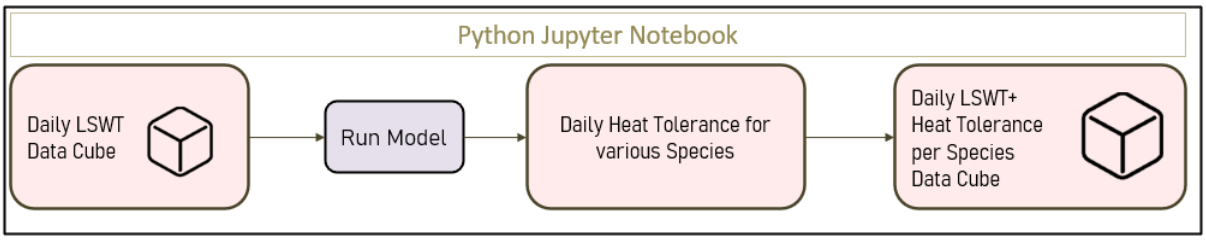



Figure 5 Example of the Data Cube Model Loop

## 4.4 Thematic Ecosystem Change Index

Freshwater biodiversity is a key indicator of the overall health of freshwater ecosystems. Changes in biodiversity can be a sign of broader ecological changes, such as changes in water quality or habitat loss, which can affect the entire ecosystem. By monitoring changes in freshwater biodiversity, we can gain insights into the health and resilience of freshwater ecosystems and develop strategies for their long-term conservation. Quantifying changes in freshwater biodiversity is important for understanding the impacts of human activities on freshwater ecosystems and developing effective conservation and management strategies. It is also critical for tracking progress towards global biodiversity targets and ensuring the long-term health and resilience of freshwater ecosystems.

Thematic ecosystem change indices (TECIs) provide information on the extent and intensity of changes in ecosystems. The developed TECIs in BIOMONDO are based on the

analysis of various data collected within BIOMONDO, e.g. EO data, models and ground monitoring, and are designed to capture changes in water quality, habitat conditions of fish or land cover that are associated with specific ecosystem processes, including eutrophication or urbanization. TECIs can provide valuable information for understanding the drivers and impacts of ecosystem change, as well as for monitoring progress towards biodiversity, conservation, and sustainable development goals. Some examples of TECIs include the habitat conditions for fish, which tracks changes in the heat tolerance of fish, cyanobacteria abundance and the lake surface water temperature. Other TECIs focus on specific regions or ecosystems, such as the Mekong River Basin, and are tailored to capture the unique drivers and dynamics of change in these areas. Overall, thematic ecosystem change indices are an important tool for monitoring and understanding ecosystem change, and can support informed decision-making for biodiversity monitoring, conservation, and sustainable development.

Within BIOMONDO the experimental dataset outputs are made available in the BIOMONDO Lab, and the TECIs are derived by combining them to assess change in the freshwater ecosystem, including its local and wider surrounding. This also includes the riparian zone and the watershed, as well as the water body and the lake/riverbed. It is a very comprehensive characterisation of the ecosystem by a large number of different variables. However, the data comes at different times and contains "holes" in space where observations are not possible. We are using Machine Learning for deriving the TECIs.

Machine Learning offers best options to

- interpret phenomenon
- identify the components of a trend
- establish relationships between variables to identify causes of change
- detect outliers
- detect broken trends (e.g. due to disruptive anthropogenic change)
- identify and characterise cyclicity and seasonality
- predict its future values

For freshwater ecosystem various TECIs are proposed and are shown in Table 4. All of them can be mapped to the original Essential Biodiversity Variable (EBV), biodiversity change driver and ecosystem function indicators (see Table 7 in Appendix).

Table 4 Proposed TECIs for Biodiversity

| **Novel EO Product (Thematic Ecosystem Change Index TECI)** | | **Reference to biodiversity requirements (column "indicator")** |
|---|---|---|
| 1 | LCLU and its change in the riparian zone and catchment area | 1, 3, 11, 21, 22, 41 |
| 2 | Water surface characterisation and evolution (status = permanent or ephemeral, growing, shrinking, ice cover) | 1, 2, 5, 6, 27, 35, 37, 38, 42, 43, 44, 47 |
| 3 | Water vegetation evolution | 3, 15 |
| 4 | Water quality (Chl, TSM, TUR, Kd, CDOM) and lake water temperature evolution | 7, 11, 13, 15, 18, 20, 23, 31, 32, 33, 34, 36, 39, 40, 45, 46, 48 |
| 5 | Phenology (land) | 11, 12, 19 |
| 6 | Phenology (water) | 12, 19 |
| 7 | Bottom characterisation (macrophytes, sand) | 8, 25 |

| | | |
|---|---|---|
| 8 | Habitat condition for fish | 3, 4, 7 ,9 ,10 ,15 ,20 ,33 |
| 9 | Vitality ecosystem condition | 4 ,7 ,9 ,10 ,14 ,18 ,20 ,32 ,33 |

The TECIs describe important characteristics of aquatic ecosystems and, several of these characteristics are likely to correlate with global patterns of biodiversity in aquatic ecosystems. While a high primary productivity in terrestrial ecosystems is, generally, assumed to promote biodiversity, eutrophication, and algae blooms (also associated with a high productivity) are generally assumed to have a negative effect on aquatic biodiversity. To facilitate studies into such macro-scale relationships between the characteristics of aquatic ecosystems (as described by existing and the here proposed novel EO products) we developed novel products that contain information relevant for assessment of biodiversity and monitoring of changes. In addition, we expect observed environmental changes to be perceived as being particularly relevant when they are observed in areas of a high conservation value, e.g. when occurring in 'hot spots' of biodiversity and/or with a large number of rare species. An example of a study where EO data is used to study macroscale patterns in terrestrial ecosystems can be found in Tuck et al. 2014.

Anomaly detection, which involves identifying and analysing unusual patterns or events in data, can be a useful tool for monitoring and understanding changes in freshwater biodiversity. By detecting anomalous events, such as changes in species composition or distribution, we can better understand the drivers of change and develop more effective management strategies to protect these important ecosystems. Figure 6 shows that unsupervised anomaly detection is a key element in the workflow for the TECI estimation. The consecutive steps will be explained separately.

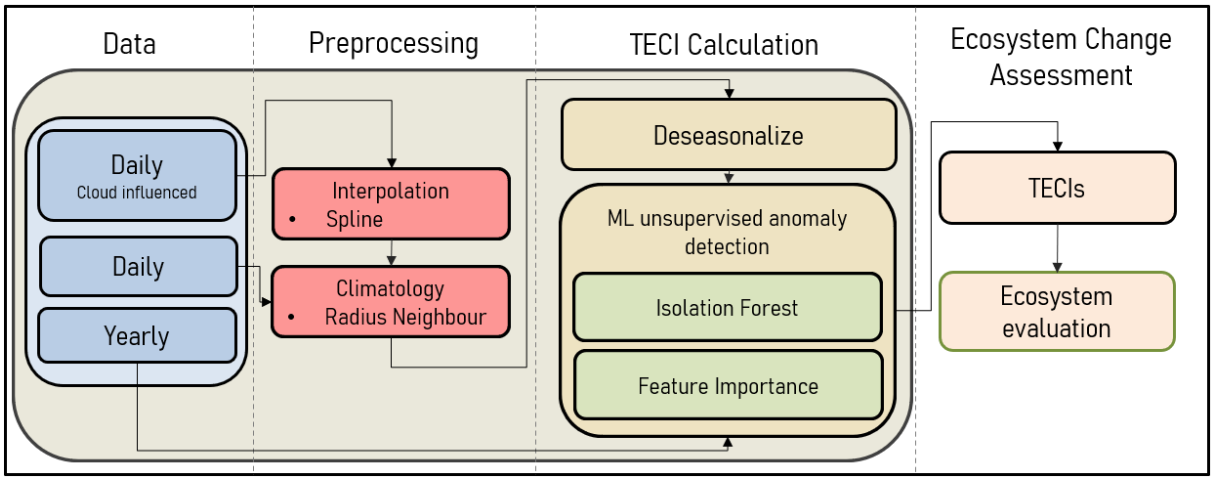



Figure 6 Workflow for TECI calculation and ecosystem change assessment.

## Data

Various experimental datasets including EO and modeled data were compiled within the BIOMONDO project. A description of all experimental datasets can be found in the D.2.4 (Table 1).

The experimental datasets are the foundation of the TECIs and they have different origins, which indicates that the spatial and temporal extend of the datasets are not the same and can vary within each pilot site. Within the datasets, daily values which are

cloud influenced and therefore have spatial and temporal gaps, daily interpolated values and various temporal aggregations (monthly, yearly) are available. TECIs represent the change within an ecosystem and therefore we spatially aggregate the source data based on waterbodies or on a subbasin level for each waterbody defined by experts. As shown in Figure 6 each of the datasets take different subsequent steps.

## Preprocessing

Daily cloud influenced datasets are interpolated after the spatial aggregation within the waterbody. For the interpolation a spline function constructed of piecewise third-order polynomials is used. Spline interpolation is a mathematical technique that involves fitting a smooth curve to a set of data points, which can be used to estimate values between those points. This method is particularly useful in remote sensing applications for biodiversity analysis because it can help to fill in missing data, smooth out noise and gaps in the data. By using spline interpolation, we can gain a better understanding of the spatial and temporal patterns of biodiversity in remote sensing data, which is essential for effective conservation and management of these ecosystems.
Daily datasets with already interpolated values, e.g. the warming tolerance model outputs described in Chapter 4.2 skip this step.

As a second preprocessing step the climatologies of these datasets are calculated. The Radius Neighbour (RN) methodology is a technique used to calculate climatologies, which are long-term averages of various variables. This method involves selecting a target location and calculating the average value of the variable at that location based on data from its surrounding neighboring locations within a specific radius. By using this methodology, we can account for spatial variability in variables and obtain a more accurate representation of the conditions at a given location.

## TECI Calculation

The central methodology of the TECI estimation is the anomaly detection of the experimental datasets. Datasets with calculated climatologies are deseasonalized by subtracting the climatology from the actual timeseries of each variable.
Each TECI estimation is calculated with Isolation Forest, a multivariate unsupervised machine learning algorithm, calculating the anomalies based on the relationships between all available variables within the TECI (Table 4).

The Isolation Forest is a popular machine learning algorithm for anomaly detection, which works by isolating observations that are anomalous from the rest of the data. It is a tree-based algorithm that randomly selects a feature and a random split value to create a decision tree. The process is then repeated until each observation in the data is isolated into its own tree.

The steps of the Isolation Forest methodology are as follows:

1. Random selection of a feature: At each step of the algorithm, a random feature is selected from the dataset.

2. Random selection of a split value: A random value within the range of the selected feature is chosen as the split value.

3. Creation of a decision tree: A decision tree is created based on the selected feature and the split value.

4. Recursive partitioning of data: The algorithm then recursively partitions the data into smaller subsets based on the split value until each observation is isolated into its own tree.

5. Scoring of anomalous data: The anomaly score for each observation is calculated by averaging the path length required to isolate the observation across all the decision trees.

6. Identification of anomalies: The observations with the highest anomaly scores are considered to be the most anomalous and are identified as such.

The key advantage of the Isolation Forest algorithm is that it can efficiently detect anomalies in large datasets without requiring extensive computation, making it suitable for real-time applications. Additionally, it can detect anomalies in both continuous and categorical data and is robust to outliers and noise.

The Isolation Forest returns an anomaly score for each data point without the information on which variable had the highest impact on the anomaly score. The anomaly score for a data point represents its degree of abnormality or outlierness relative to the rest of the dataset. The anomaly score ranges from 0 to 1, with higher scores indicating greater abnormality or outlierness. For biodiversity assessments it is important to know which variable is mainly affecting the anomaly. SHAP (SHapley Additive exPlanations) values are a type of feature importance measure in machine learning that provides a way to explain the output of a model in terms of the contribution of individual features to the predicted outcome. The SHAP value for a particular feature represents the average marginal contribution of that feature to the predicted outcome across all possible subsets of features in the model. In other words, SHAP values provide a way to quantify the relative importance of each feature in the model by measuring how much each feature contributes to the predicted outcome, taking into account the interactions between features. The SHAP value, from now on called feature importance measure, provided by the Isolation Forest algorithm is based on the number of times a feature is selected for splitting across all the trees. The more frequently a feature is selected, the higher its importance score. Therefore, the feature importance measure can be used to identify the most relevant features that contribute to the detection of anomalies in the dataset. The range of feature importance measure depends on the specific machine learning model and feature set being analyzed. The magnitude and sign of the feature importance measure for a particular feature indicate the extent to which that feature contributes to the prediction made by the model. It's important to note that feature importance measures are not limited to any specific range or scale, as they are calculated based on the unique features and structure of the model being analyzed. However, feature importance measures are

typically normalized to make them more interpretable and easier to compare across different features and models.

## Ecosystem Change Assessment and Interpretation

The final step of the TECI estimation is the ecosystem change assessment where the TECIs can be used for ecosystem change evaluation by calculating the feature importance. The feature importance measures (SHAP values) are stacked and are called the TECI score. The TECI scores are visualized in a bar plot showing the change index over time for each variable. Interpreting these TECI scores is challenging, and need to be supported by expert interpretations as the TECI score is based on stacked values of multiple features important measures. The more variables a TECI has the higher the overall TECI score can be. Although the feature importance measure of each variable is normalized and comparable towards other variables within the same TECI, there is no fixed range a TECI score can adopt. The TECIs are an innovative method and therefore more knowledge is gained throughout the process and during the next phases of BIOMONDO.

Figure 7 shows the TECI Score of TECI 8 – Habitat Conditions for Fish in Granfjärden Mälaren for the summer months. Four variables are included, and each variable shows a unique feature importance. All scores stacked together make up the overall TECI scores showing that the years 2018 and 2020 have the highest probability for anomalies. The overall focus is based on data for the fish Osmerus eperlanus, indicating changes in the habitat condition for this species.

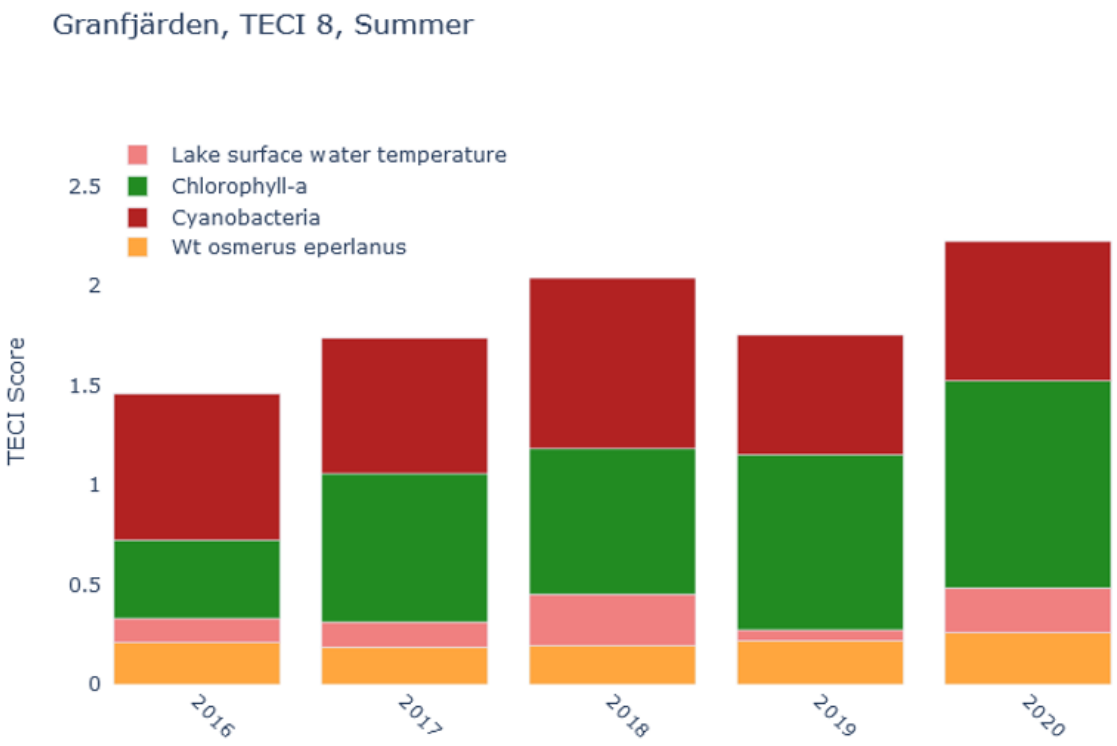


Figure 7 Score of TECI 8 – Habitat Conditions for Fish, Granfjärden Mälaren

Figure 8 shows the TECI score for TECI 4 – Water Quality in the Markermeer for the spring months. The available parameters for this TECI are the chlorophyll-a concentration, turbidity, and the lake surface water temperature. The score shows that the years 2013, 2016 and 2018 show the highest probability of anomalies within the water quality. The single TECI value of each variable show that especially the lake surface water temperature influences the overall TECI score in 2013 and 2018. For 2016 the TECI value shows that mainly turbidity is the anomaly driver.

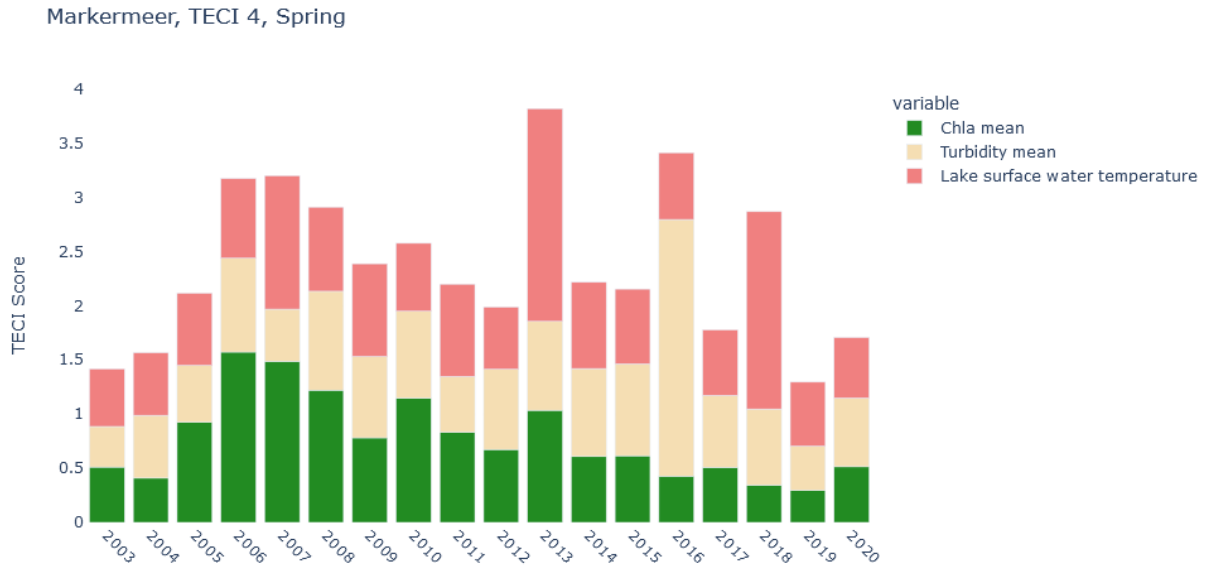



Figure 8 Score of TECI 4 – Water Quality, Markermeer

The validation of such highly synthetic TECIs is challenging because they are integrative values describing a temporal pattern in a multi-dimensional dataspace. The traditional concept, where the same quantity is measured from space and in-situ, and where validation means to compare those coincidently made measurements, does not apply here. The TECIs indicate a certain quality of an ecosystem, and we propose a certain mathematical method to derive it. Algorithm validation is applied at two levels: (1) we ensure that our experimental datasets are validated appropriately. The validation performed on the datasets are descripted in D2.3 (Table 1). (2) Expert knowledge can be used as a validation method for the TECIs. The judgment of an expert is needed to confirm the accuracy or validity of a particular result or prediction. Experts of each pilot site can assess the TECI scores of the algorithm and confirm whether they are consistent with their knowledge and experience of the ecosystem changes being studied. This can help to identify any errors or biases in the algorithm and improve the accuracy of the analysis. Expert validation also helps to identify any areas where further investigation or data collection is needed to improve the analysis.

While indicators can be useful tools for assessing environmental conditions and informing management decisions, they should be interpreted with caution.

1. The TECI value may not accurately reflect the full complexity of the system being monitored. Simplifying complex environmental processes into a single indicator can lead to oversimplification and loss of important information.
2. TECI values may be influenced by factors that are not directly related to the environmental condition being measured. For example, changes in sampling frequency, location, or methodology can affect indicator values, even if the underlying environmental condition remains unchanged.
3. TECI values may not be appropriate for all contexts or ecosystems. Indicators developed for one ecosystem may not be relevant or appropriate for another, and may not capture the unique ecological processes and stressors present in the new system.

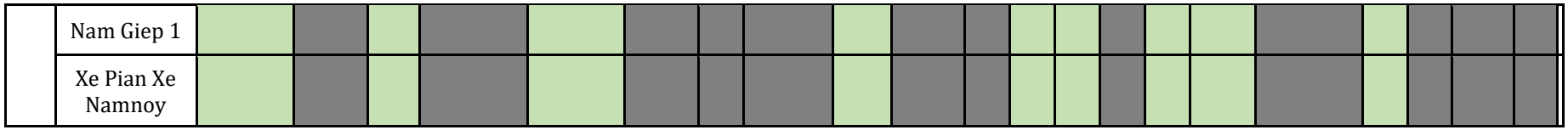
Currently the TECIs are derived on a set of variables retrieved within the BIOMONDO project. These variables depend on the availability of operational EO datasets and models described in this document. Increasing the number of variables within TECIs can improve estimations for several reasons. Incorporating more variables allows for a more comprehensive representation of biodiversity change in an ecosystem and the environmental factors that influence it. This can lead to a more accurate and nuanced understanding of how biodiversity is changing over time. More variables can increase the sensitivity of the index to changes in biodiversity, making it more likely to detect changes even when they are subtle or occur over short time periods. Including more variables can help to reduce bias in the index by accounting for a wider range of environmental factors that may be influencing biodiversity, rather than relying on a limited set of variables that may not fully capture the complexity of the ecosystem.

The status of the TECI data readiness can be seen in Table 5. For each TECI a set of available and reliable indicators was set and the readiness of the data covering the indicator was colored. Two states are mentioned:

- Ready-to-use (green): Data is processed and stored in a data cube or vector file.
- Not planned (grey): Indicator is not selected for any pilot study or was discarded due to missing algorithms or data sources.

Table 5 Data Readiness for the TECIs

| | | TECIs | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | **1 LCLU and its change in the riparian zone and catchment area** | | | | 2 Water surface characterisation and evolution | | | | **4 Water quality and LSWT evolution** | | | | | | | **6 Phenology (water)** | **8 Habitat Condition for Fish** | | | |
| Indicator | | Land use and land cover change in the water- | Species populations Population abundance | Map of major river dams | Supporting habitats - Habitat extent - Landcover/forest cover | Land use and land cover change in the water- | Water body extent and retention time | Hydro-period | Waterlevel | Water column trophic status | Species populations Population abundance | Water temperature | Algal blooms in lakes | TSM | Primary Production | Chlorophyll-a | Kd | Ecosystem function Ecosystem phenology | Chlorophyll –a | Cyanobacteria | Water Temoerature | Warming Tolerance |
| Pilot 1 | Markermeer | | | | | | | | | | | | | | | | | | | | |
| Pilot 2 | Markermeer | | | | | | | | | | | | | | | | | | | | |
| | Lake Mälaren | | | | | | | | | | | | | | | | | | | | |
| | Lake Balaton | | | | | | | | | | | | | | | | | | | | |
| Pilot 3 | Xekaman 1 | | | | | | | | | | | | | | | | | | | | |
| | Lower Se San 2 | | | | | | | | | | | | | | | | | | | | |


Nam Giep 1
Xe Pian Xe Namnoy

# 5 Pilot Applications

The aim of the BIOMONDO Freshwater Pilots is to address scientific and policy priorities in freshwater biodiversity and vulnerable ecosystems by demonstrating how novel Earth Observation and Biodiversity modelling products can be integrated to address the knowledge gaps in biodiversity monitoring, modelling and assessment and to enhance decision support systems. Three pilots have been developed and are described in chapters 5.1-5.3.

## 5.1 Pilot 1 – Eutrophication

Nutrient concentrations have increased substantially in lakes and rivers throughout the world, resulting in eutrophication, harmful algal blooms, loss of submerged macrophytes affecting sedimentation and turbidity, and biodiversity loss.

In this BIOMONDO pilot we explore the possibilities of integrating EO data into the Delft3D model suite to investigate the potential contribution by EO data to the model performance. We also demonstrate the use of EO based phenology metrics to study effects of eutrophication.

### 5.1.1 Use of EO data for model input

For pilot site Markermeer, the water quality module (Delft3D-WAQ) is connected with the hydrodynamic module (Delft3D-FLOW) to simulate algae dynamics and transport. The hydrodynamic model for the Markermeer was run for the year 2016. The hydrodynamic results for the year 2016 have also been used as input for the other years (2020 and 2021) for which the water quality model was run. The hydrodynamic calculations were carried out using a 3D model. However, to speed up the computation of the water quality model, the hydrodynamic results were aggregated to a 2D grid, which has only a limited impact on the results since vertical stratification events hardly occur in this shallow lake. The 2D Delft3D-WAQ model was run for the (entire) years 2016, 2020 and 2021. Meteorological input data was obtained from the nearest meteorological station for the same years.

The model results were reprojected to the data cube grid latitude and longitude using the nearest neighbour method. For every latitude, longitude and day the parameters in Table 2 were generated. Results regarding chlorophyll-a and total net primary production have been used to assess the model performance (see D2.3 PVR document) and the impact of using an EO-based temperature forcing (see section below).

In the early stages of the pilot development, it was assessed which EO based products would be suitable to improve the input data. Although chlorophyll-a might seem like a

good candidate, this is not a state variable in the model but rather a postprocessed output derived from the modelled algal biomass and species composition (which themselves are not available as EO-products). EO based chlorophyll-a was however used to validate the model. Another promising candidate to provide EO based input data would be suspended sediment concentration, but this was not part of the BIOMONDO EO products for Lake Markermeer. Finally, it was decided to use the EO based surface water temperature as input to the model. In the original set up of the Markermeer model, the water temperature was based on in situ measured air temperatures from the nearest station of the Royal Dutch Meteorological Institute (KNMI). Supposedly, the EO based surface water temperature is closer to the actual water temperature than the air temperature measured in (semi) distant onshore location. Also, the EO-data provides information on the spatial variability that is not provided by measurements in a single location. Hence, using the EO based temperature forcing is expected to improve the model performance.

The components of all methods use in the pilot are demonstrated in Figure 9.

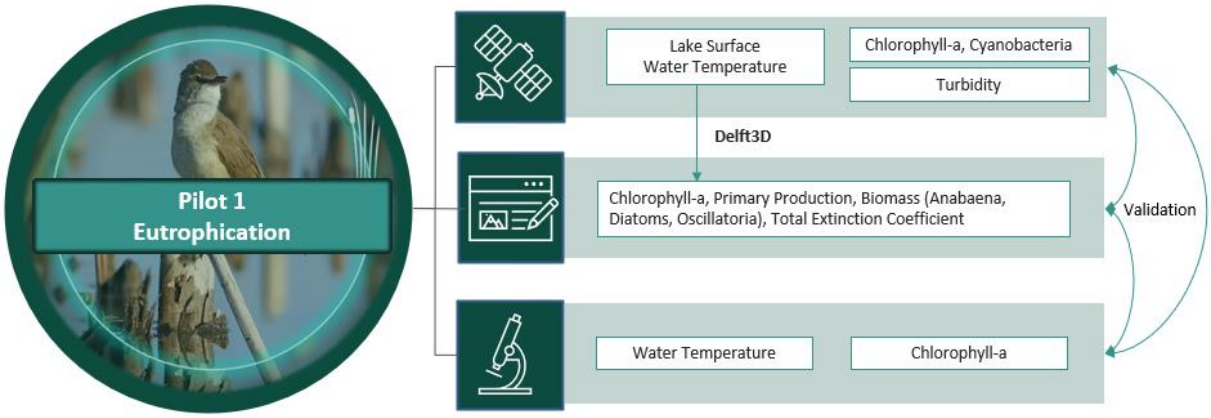



Figure 9 Overview of Pilot 1 components and workflow

To investigate the congruence between model results and observations (EO and in situ data), a comparison of the modelled chlorophyll a with the EO chlorophyll products was made for the years 2016 and 2020 and with in situ primary production measurements for 2020 and 2021. EO and model data were plotted against time to show the calculated and observed chlorophyll-a concentrations. Validation results with in situ data are discussed in the D2.3 PVR document.

EO data corresponding to chlorophyll a was used to compare with the model data for the whole year at two sites: at the central station of the lake and in the south at Pampus Oost (Figure 10). The modelled chlorophyll a concentration showed a reversed pattern in time as compared to EO data (Figure 11). This mismatch is thought to be due to the relatively high inorganic matter concentrations in this lake since growing algae in spring and summer attach to these suspended solids, they then sink to the bottom. This continues during the whole spring and summer period, leading to an accumulation of algae at the bottom

of the lake. They remain there until they are resuspended in the winter due to high wind conditions. In situ data of chlorophyll-a shows the same pattern as the EO data. However, when looking at the primary production flux (no EO data for this), the model is in quite good agreement with in situ data, suggesting that the observed chlorophyll is not active and/or not contributing to primary production.


Figure 10 Locations in the Markermeer used to compare EO and model data (Pampus Oost and Central Station).

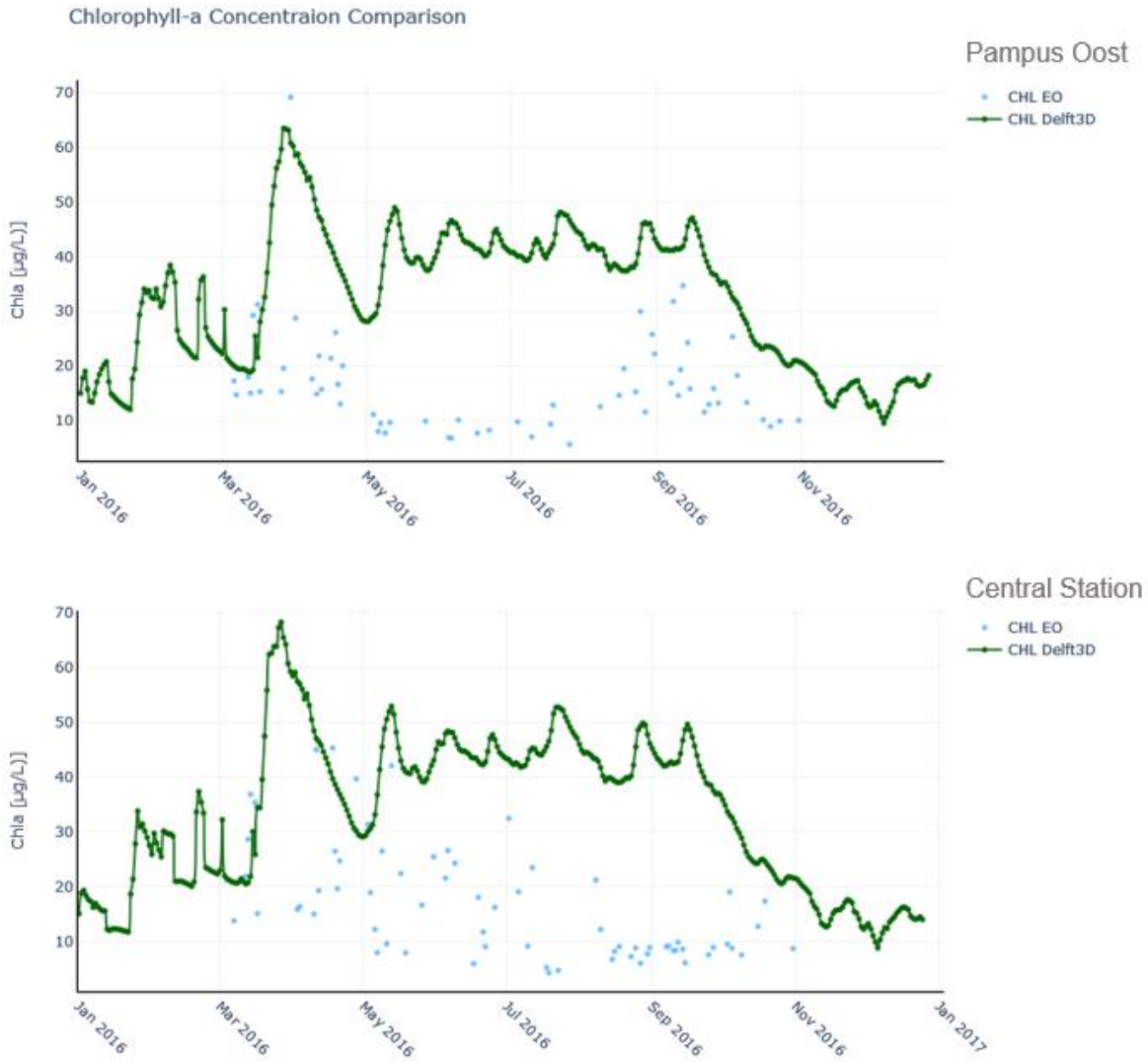



Figure 11 Comparison of CHL-a concentrations in 2016 between EO and model data for two locations (middle of the lake (upper figure) and south (lower figure) in the Markermeer.

As explained above, to improve the model performance regarding the modelled primary production, EO temperature products were used as a direct input into the Delft3D-WAQ model. To this end, the spatially and temporally varying EO temperature product is used to force the water temperatures in the model, while in the original set up of the model, the water temperature was based on in situ measured air temperatures from the nearest station of the Royal Dutch Meteorological Institute (KNMI). Note that the EO-based forcing was only available up to September 2020. In contrast to the spatially and temporally varying EO-based forcing, the original in situ based forcing consisted of a temporally varying but spatially homogeneous forcing function. A comparison of the two forcing functions showed that the two forcings show similar temperature ranges as well as similar seasonal patterns (Figure 12). However, the EO-based temperature field includes less temporal variation than the situ-based temperature field and shows slightly higher temperatures in summer. Also, in 2020 it shows slightly lower temperatures in winter. These

differences may be caused by the inertia of the water system, which may make it slow to respond to changes in ambient air temperatures.

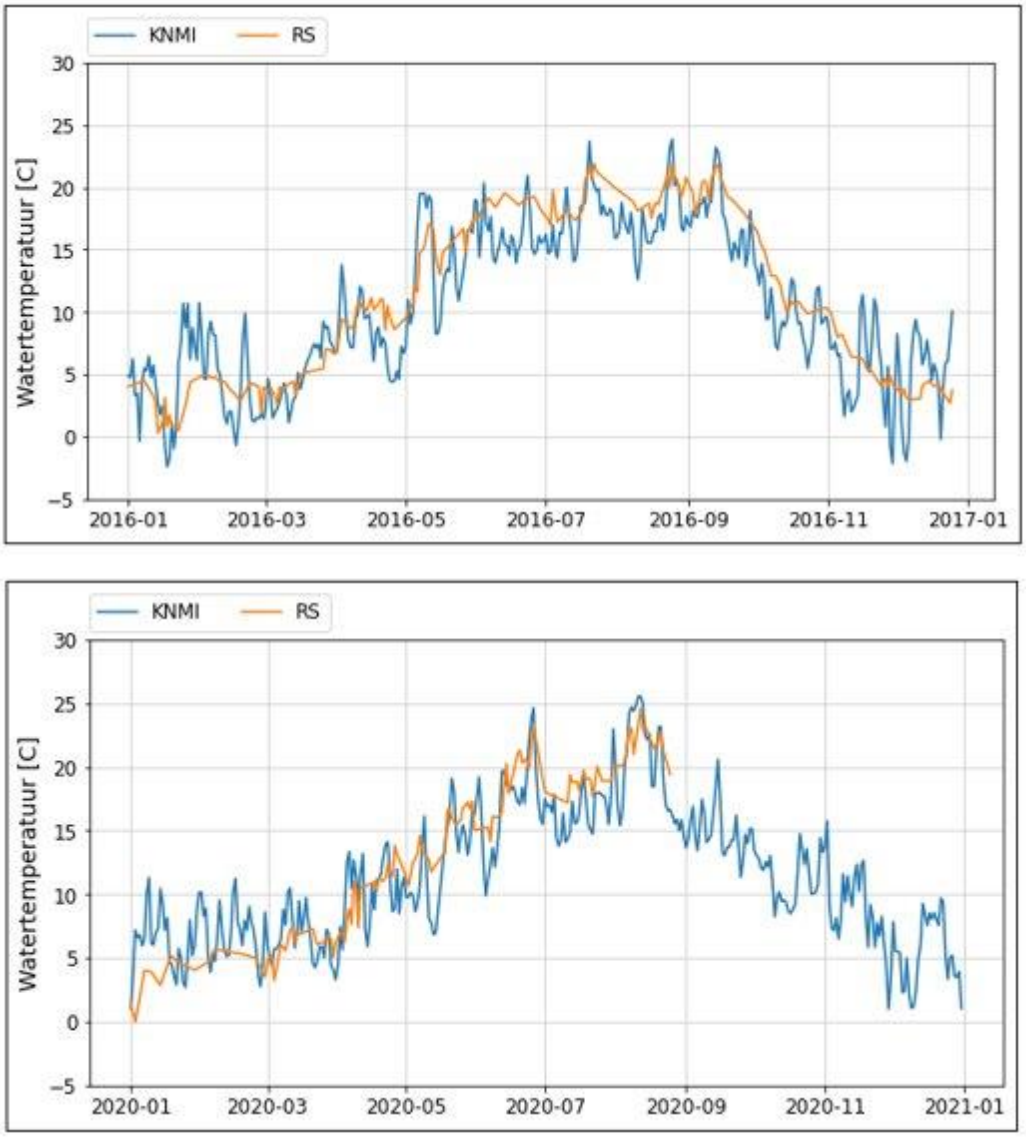
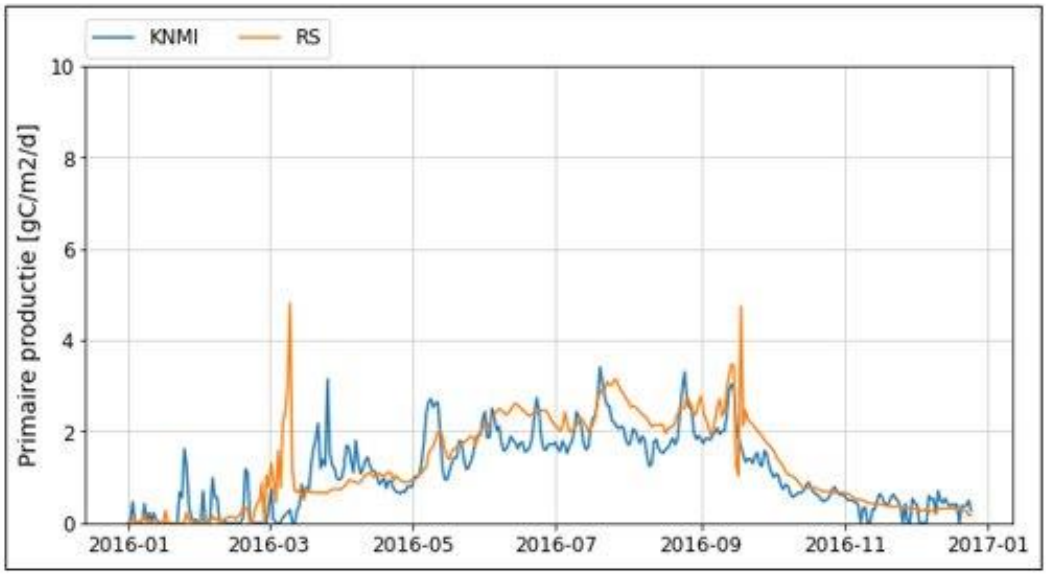



Figure 12 Water temperature as included into the Delft3D-WAQ model of the Markermeer based on in situ data (blue) or EO products (yellow) in the year 2016 (upper panel) and 2020 (lower panel), both spatially averaged for the area 'Markermeer Midden'.

Next, the impact of using EO temperature products as a forcing for water temperature on the modelled primary production was assessed. Model results show that using an EO-based temperature forcing in the model resulted in a slightly higher primary production in summer, and an overall decrease of the temporal variation in the calculated net primary production (Figure 13). This is a direct effect of the higher temperatures in summer and the smaller temporal variation in the EO-based temperature forcing. However, at a few moments in time, the modelled PP based on EO forcing shows peaks of high temporal variability that are not occurring in the temperature forcing. Also, the EO-based forcing causes the spring bloom to start earlier (most clearly visible in 2016). This is caused by

the slightly higher temperatures occurring in early spring when sufficient light is available to start the bloom. Clearly, small differences in forced temperatures may cause indirect and/or disproportional changes in primary production, which may be very relevant for the higher trophic food web. Because the temporal overlap of the modelled PP on basis of EO forcing and the in situ measured PP is too small, we cannot say whether the EO-based forcing improves the performance of the modelled primary production.



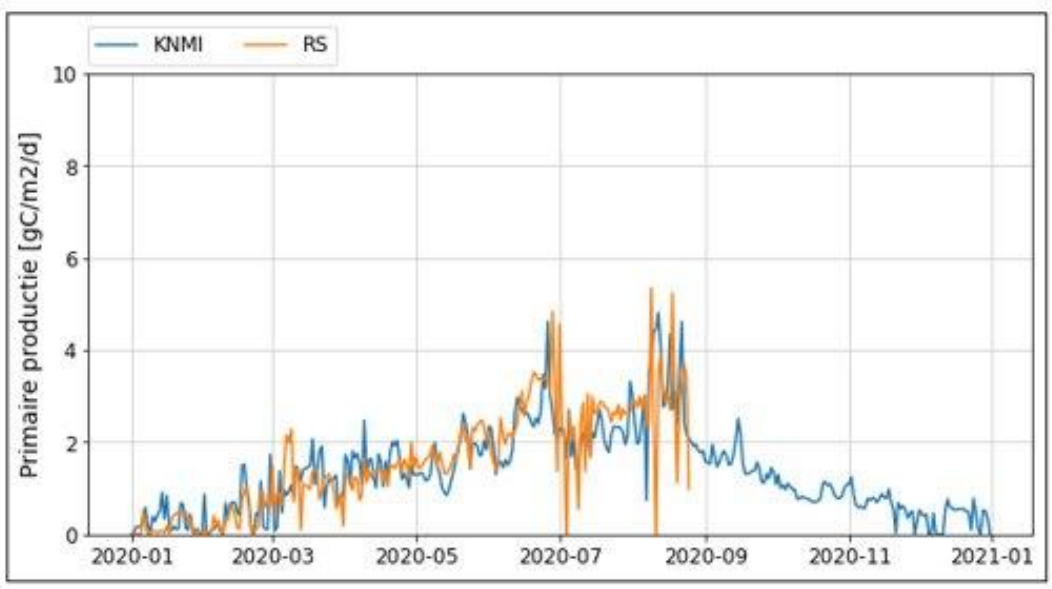



Figure 13 Net primary production as modelled by the Delft3D-WAQ model for the Markermeer in the area 'Markermeer Midden' for 2016 (upper panel) and 2020 (lower panel) using in situ data (blue) and EO products (yellow) as a forcing for water temperature in the model.

### 5.1.2 Use of EO data for phenology assessments

One of the most visible impacts of lake eutrophication is a change in the size and duration of algal blooms which can be assessed using the phenology metrics described in section 4.1.3. Very large algal blooms that covered more than 80% of the lake (i.e. the fraction of pixels in which a peak was detected) were observed in February-March over the years 2001-2007, while this recurring pre-spring bloom and started earlier (January-February) and declined in size over the years 2008-2010. At the same time, a large summer bloom (June) appeared in the year 2010 which covered nearly 70% of the lake (see Figure 14). This analysis will be further extended over the years 2001-2021 once the new ESA Lakes v.2.1 is made available to us. This new release includes improved quality control flags that may benefit our analysis and combines MERIS, MODIS-Aqua, and OLCI data and will help us to compare results with those in section 5.1.1.

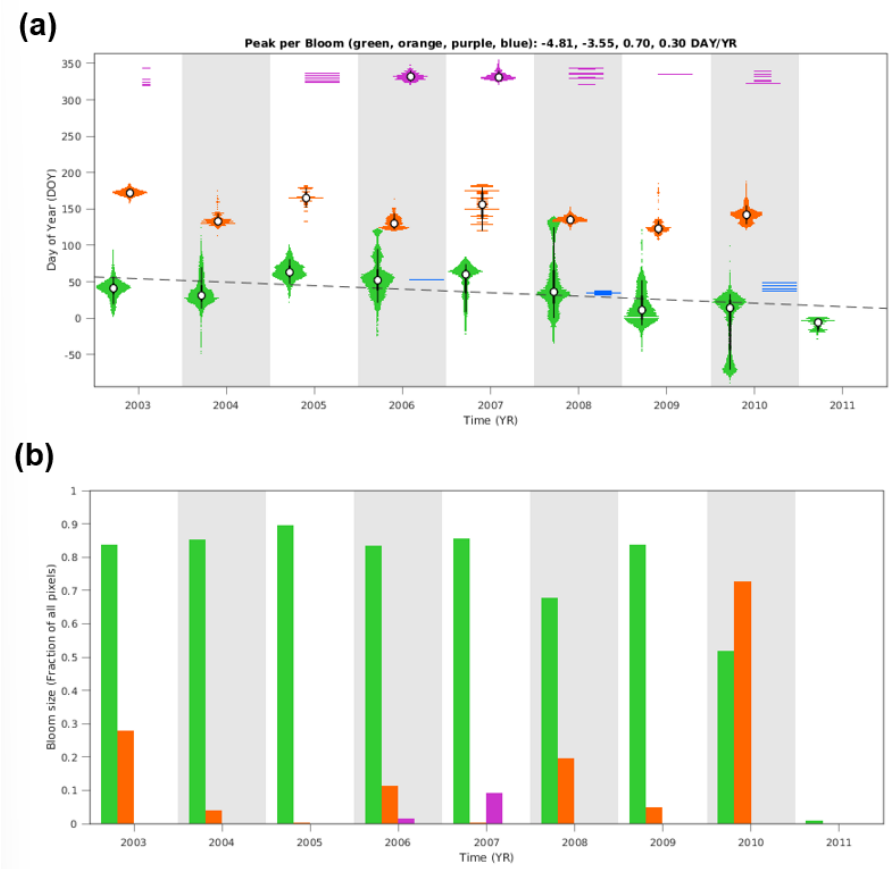



Figure 14 Trends of change in the timing of Algal blooms (as in Fig. 2.d) as observed in lake Marken using OLCI data. Trends of change in the timing of algal blooms (as in Fig. 2.d). The early-spring bloom (January-March, approximate day of year 10-90) is indicated in green, while the summer bloom is indicated in orange (June, approximate day of Year 140-160). **(b)** The size/area covered by these algal blooms in terms of the fraction of pixels in which a peak is detected.

## 5.2 Pilot 2 – Heat Tolerance of Fish

Lake surface water temperatures have been rising rapidly globally. Additionally, lake heat waves intensity and duration are expected to increase with future climate change, exacerbating the effects of long-term warming. Lake ecosystems are vulnerable to these temperature changes: directly by pushing to or exceeding species and ecosystems limits of resilience, and indirectly through for example decreasing amount of oxygen in the water, altering stratification or algae blooms altering oxygen availability.

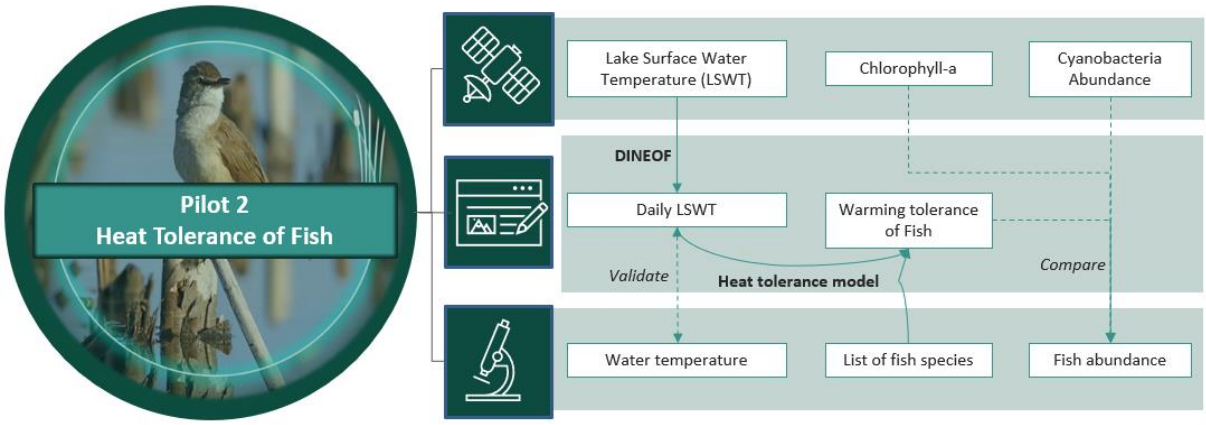



Figure 15 Overview of Pilot 2 components and workflow

The objective of this BIOMONDO pilot is to explore the possibilities of using a combination of EO data on LSWT and thermal tolerance of freshwater fish species to quantify the impacts of increases in temperature and heat waves on freshwater fish diversity. Heat tolerance is calculated using species-specific information (maximum habitat temperature and random effects group-level value) of the species occurring at a site and acclimation temperature. Acclimation temperature is taken as the average LSWT 14 days prior. The heat tolerance of a certain day is then compared to the LSWT of that day. The difference between these values shows whether the LSWT exceeds or approaches the heat tolerance, and is called the warming tolerance (i.e. how close current conditions are to their thermal limits; Clusella-Trullas et al., 2021).

### 5.2.1 Lake Marken

For Lake Marken daily LSWT data was generated (see Chapter 4.1.2) from September 1999 up to and including August 2020. We retrieved the warming tolerance across time and space for 28 freshwater fish species occurring in Lake Marken. The warming tolerance is defined as the difference between the LSWT and the upper thermal limit (heat tolerance) of a fish. The warming tolerance shows whether the LSWT exceeds (negative

values) or approaches (low values) the heat tolerance, i.e. how close current conditions are to their thermal limits (Clusella-Trullas et al., 2021).

Across the years and throughout the lake, fish species experienced different values of warming tolerances. The lower the warming tolerance, the closer the temperature is to the heat tolerance, and the more vulnerable a species is to heat exposure. Comparing the vulnerability among species (Figure 16), the European smelt (Osmerus eperlanus) is most susceptible to heat. This corresponds to findings of local experts and fishermen (Joep de Leeuw, 2022; Noordhollands Dagblad, 2018). We can also see that for all species the heat tolerance was not exceeded during the studied period (no negative values of the warming tolerance). Although this would mean no fish were killed directly due to high temperature, approaching the heat tolerance can still be problematic. When water temperatures approach a fish's heat tolerance, this can affect growth, reproduction, immunity, and the ability of an individual to cope with additional stressors (Sadoul & Vijayan, 2016; McArley et al., 2017; Alfonso et al., 2021).

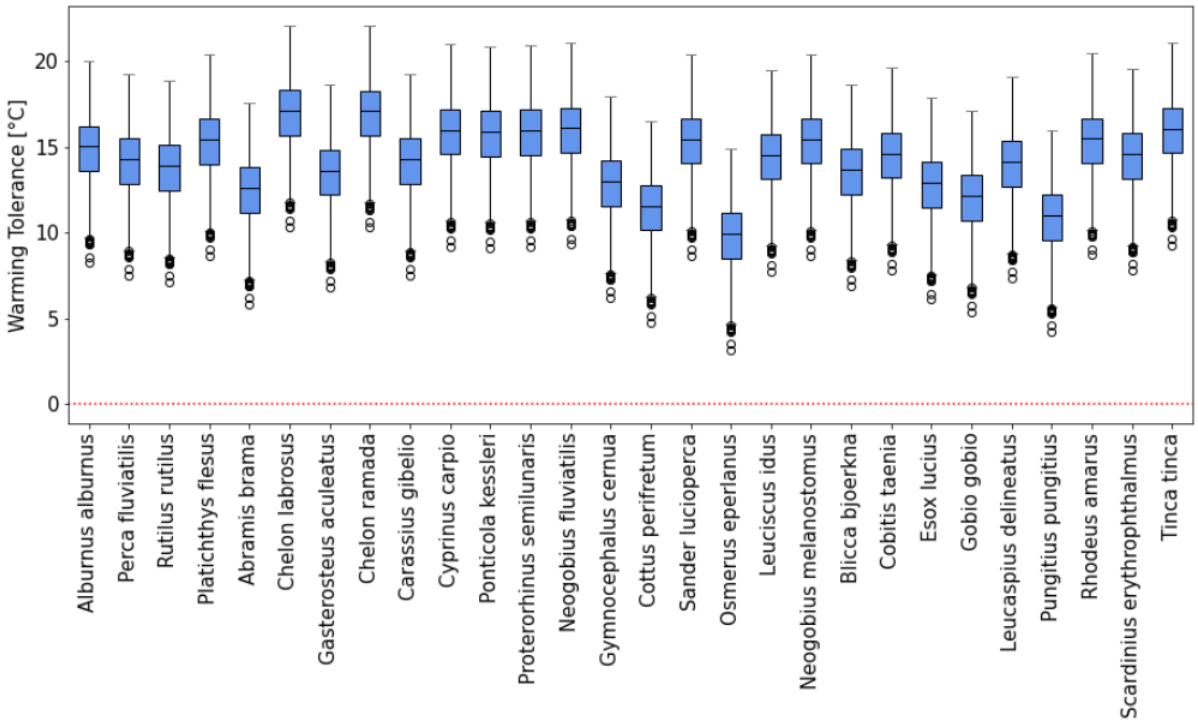



Figure 16 Warming Tolerance in the summer months 2000 – 2020 (JJA) for all species abundant in Lake Marken (central station)

Focussing on the European smelt, we can see the average warming tolerances per month across the lake in Figure 17. As expected, especially in the summer months the heat tolerance is approached by the LSWT (warming tolerance is low). Throughout the years, the period of high vulnerability seems to extend from 2 to 3 months (June-August) to 3 to 4 months (May-September). However, these are lake averages, so extremes in warming tolerance across the lake and throughout the month could cancel each other out.

In a report by IMARES (Institute for Marine Resources and Ecosystem Services; 2007), summer fish die-off of the European smelt was described. There are no documented

observations, but interviews mention disappearance of the species in August 2003 during high LSWT (>25 °C), July/August of 2006, and June 2007 (though it was relatively cool summer). Although in Figure 17 we can see that the warming tolerance of smelt is small during the summers of 2003 and 2006, these years do not stand out compared to other years. IMARES imply that summer die-offs occur at high temperatures when, especially in combination with algae blooms, low oxygen levels occur, but high temperatures do not directly cause the fish deaths. This can explain why we do not see a direct or noticeable threat of temperature in relation to heat tolerance during these summers.

There has not been any recent documentation of smelt fish kills, primarily due to the low smelt abundance during recent years so it would be less noticeable. On the other hand, summer die-offs might occur less often due less algae blooms and therefore less decrease in oxygen levels. Moreover, smelt can escape to relatively cool deep wells, which increased in number due to sand extraction (de Leeuw, 2022). The possibility to escape to deeper areas is the most plausible explanation, as the lake surface temperatures across the lake do not provide refuge, as for example can be seen by the homogeneous distribution of warming tolerance in 2003 in Figure 18.

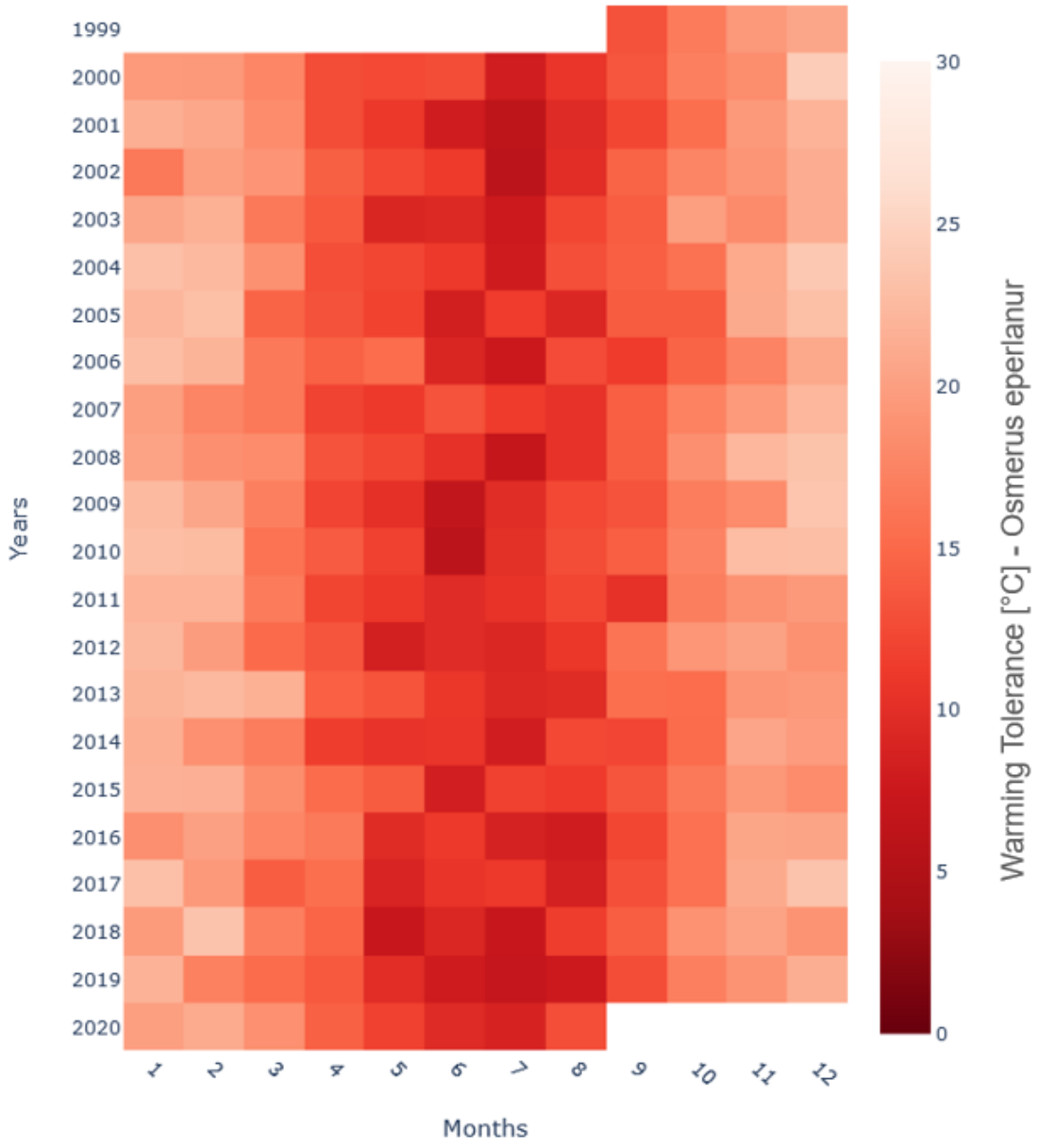



Figure 17 Average monthly warming tolerance (average across the lake) for Osmerus eperlanus in Lake Marken.

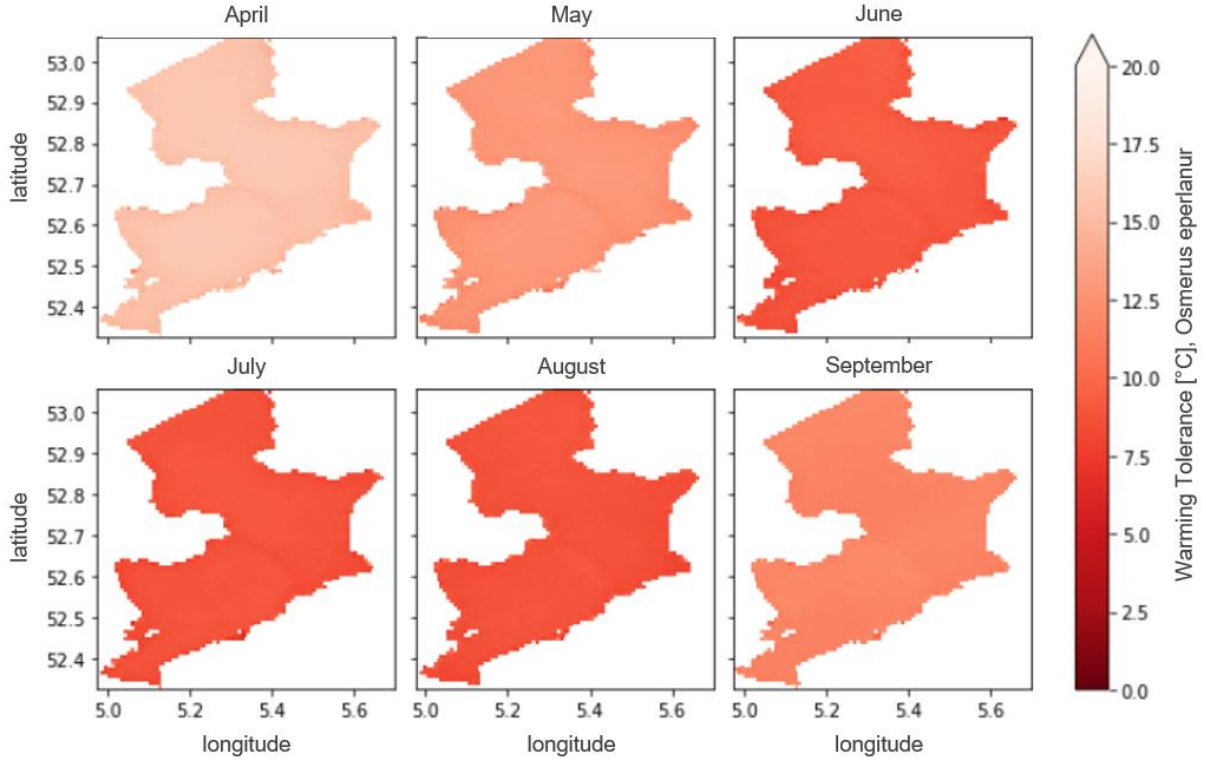



Figure 18 Spatial variability of warming tolerance for Osmerus eperlanur in Lake Marken (2003), average warming tolerance for the spring and summer months.

### 5.2.2 Lake Mälaren

Figure 19 shows species warming tolerances experienced in (central) Lake Mälaren throughout September 1999-October 2020. According to local experts smelt, vendace and perch would be most susceptible to high temperatures (Axenrot & Sandström, 2022).

Vendace (Coregons Albula) and smelt (Osmerus eperlanus) also show the lowest warming tolerance in our results. The heat tolerance was not exceeded by the LSWT for any fish species at this chosen central point in the lake.

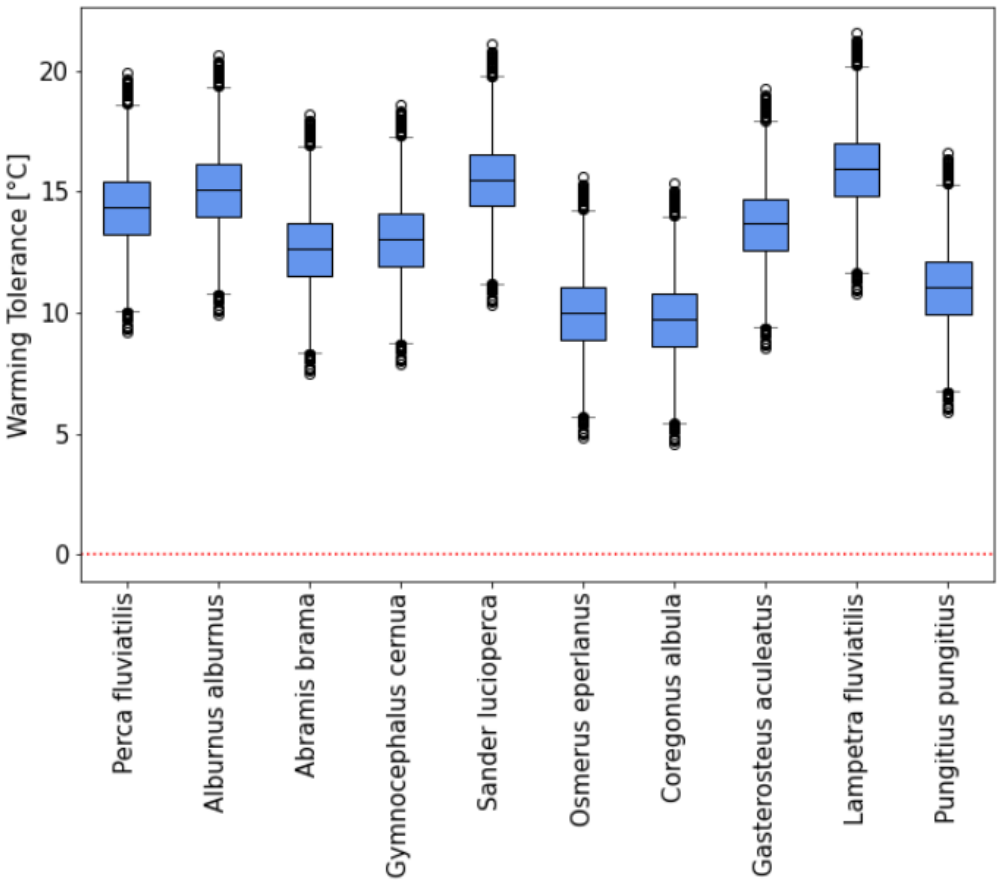



Figure 19 Warming Tolerance in the summer months 2000 – 2020 (JJA) for all species abundant in Lake Mälaren (central station)

Looking at the average monthly warming tolerance averaged across the lake for smelt, (Figure 20) we can see that during the summer months, especially June and July, the warming tolerance is lowest and the threat of high temperatures is highest. In the summer of 2018 massive fish-death due to high temperatures occurred across Swedish lakes, including Lake Mälaren (Mitt i, 2018; Axenrot & Sandström, 2022). Although this summer shows high temperatures with relatively low warming tolerance value (see Figure 20 and Figure 21 for smelt as an example), it does not stand out compared to other summers. The fish deaths are suspected to be caused by high temperatures and the accompanying lack of oxygen as the water gets warmer. Mikael Svensson (SLU) explains in Mitt i (2018): ´The fish get stressed, the metabolism goes up and they want to consume more oxygen than is available." Though the fish deaths may not be directly caused by an exceedance of the heat tolerance, the water temperatures approaching a fish's heat tolerance will cause stress and affect the ability of an individual to cope with additional stressors (Alfonso et al, 2021). This is probably what occurred in 2018, with the lower oxygen levels as additional stressor.

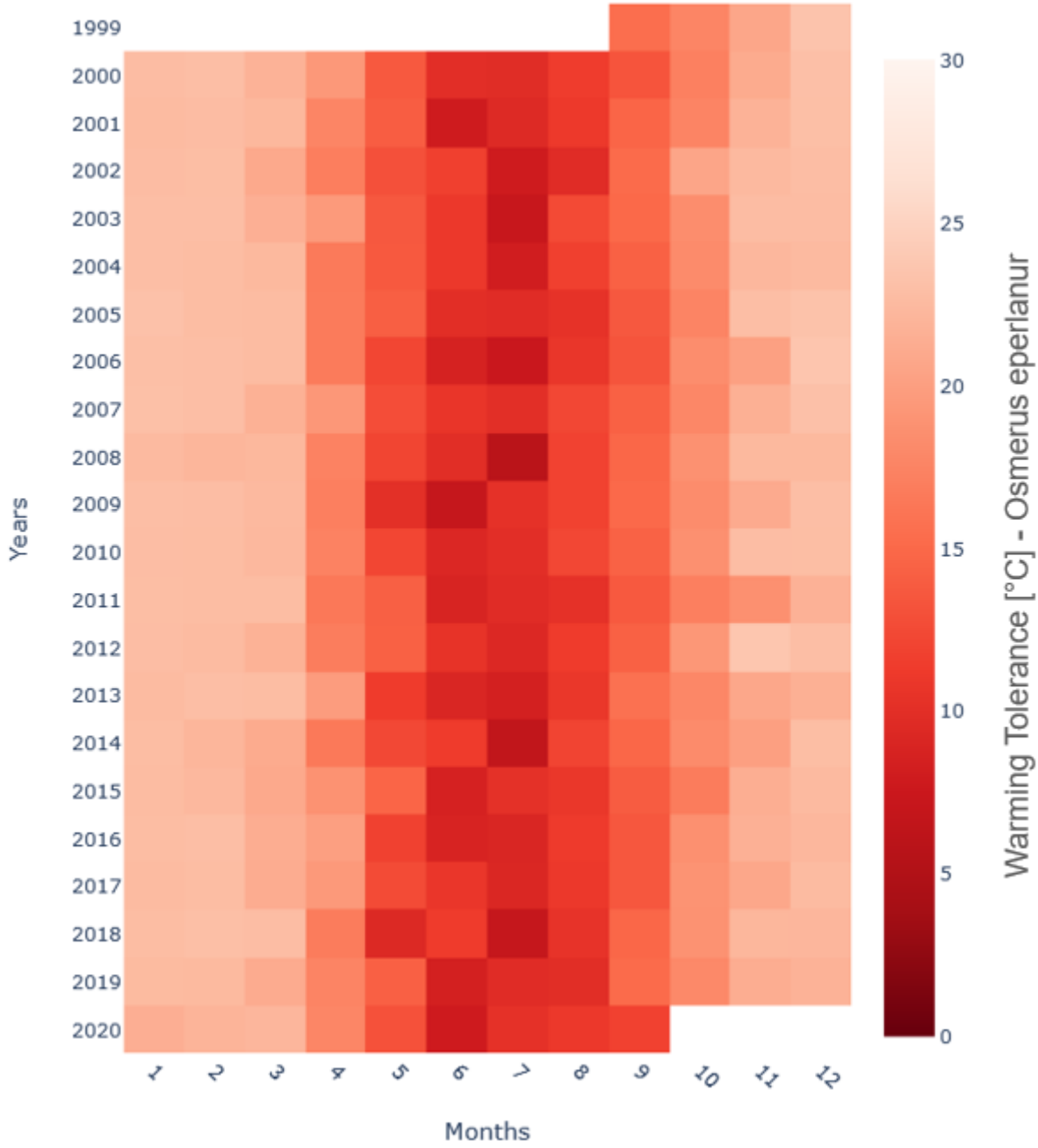



Figure 20 Average monthly warming tolerance (average across the lake) for Osmerus eperlanur in Lake Mälaren.

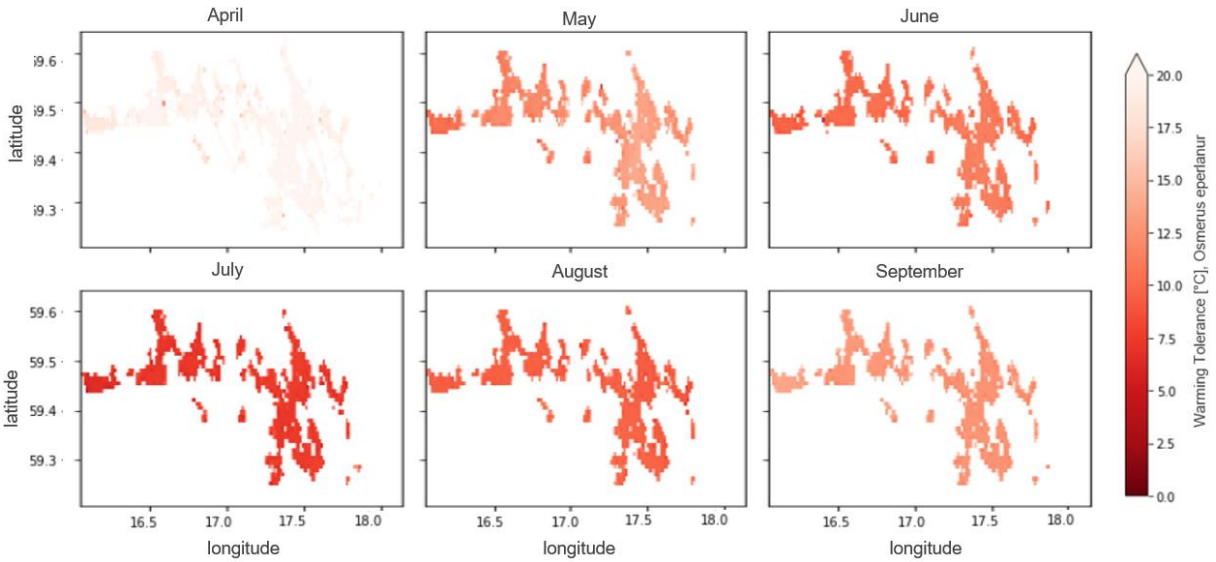



Figure 21 Spatial variability of warming tolerance for Osmerus eperlanur in Lake Mälaren (2018), average warming tolerance for the spring and summer months.

We received abundance data from fish surveys at three locations in the lake (Figure 22) taken yearly in September (Axenrot, T. (2020)). We compared these results to both the warming tolerance and remotely sensed chlorophyll-a measurements and cyanobacteria abundance (Schaeffer et al. 2022, Matthews et al. 2015) throughout the summer. We included chlorophyll a and cyanobacteria here, as these are known to be affected by temperature change and blooms can decrease the amount of available oxygen. Overgrowth of algae consumes oxygen and blocks sunlight from plants under water. When the algae die, the decay decreases the amount of dissolved oxygen further. This can cause hypoxic zones, in which fish may not be able to survive.

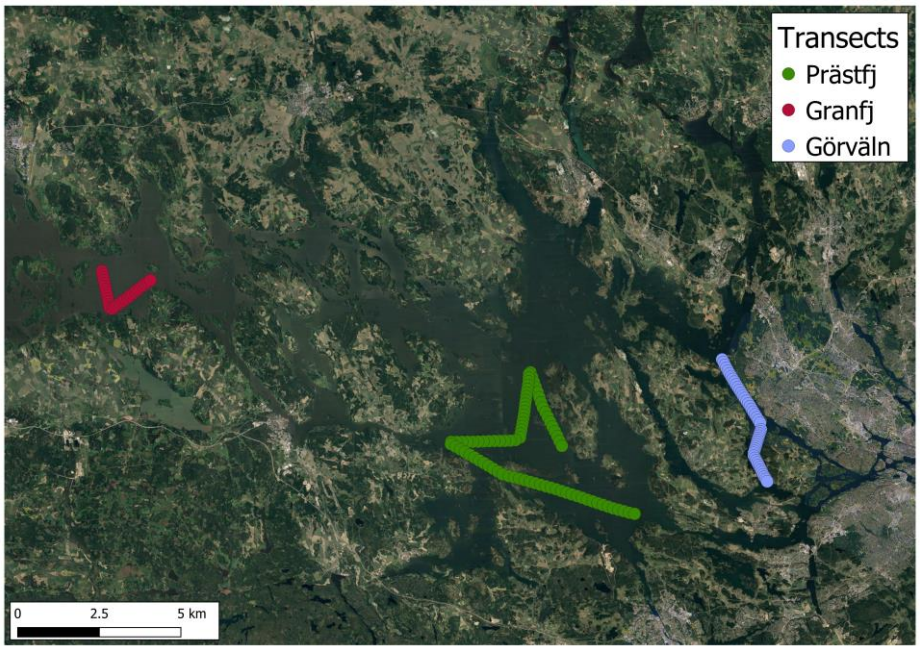



Figure 22 Location of the three transects of fish surveys in Lake Mälaren. At Granfjärden (Granfj) and Görväln the lake is shallower compared to Prästfjärden. At Granfjärden high Chl a concentrations and blooms are regular phenomenon.

We now take a closer look at the results for smelt (Osmerus eperlanus) at the Granfjärden transect (Figure 22, Figure 23). We see a decrease in fish abundance between October 2017 and September 2018, an increase towards September 2019, followed by a decrease in the next year. The summers of 2018 and 2020 are characterized by many alerts of cyanobacteria, relatively low values of warming tolerance during large periods of the summer, and high concentrations of chlorophyll a compared to the summers of 2017 and 2019. In the summer of 2020, we can see warming tolerance and chlorophyll a show a clear opposite trend, which is possible as they both are influence by water temperatures. During periods of high temperatures, the warming tolerance is lower, and chlorophyll-a and cyanobacteria can bloom resulting in higher concentrations and alerts respectively. The results may indicate an influence of temperature dependent threats (i.e. absolute water temperature and oxygen depletion) on species abundance during the summers of 2018 and 2020.

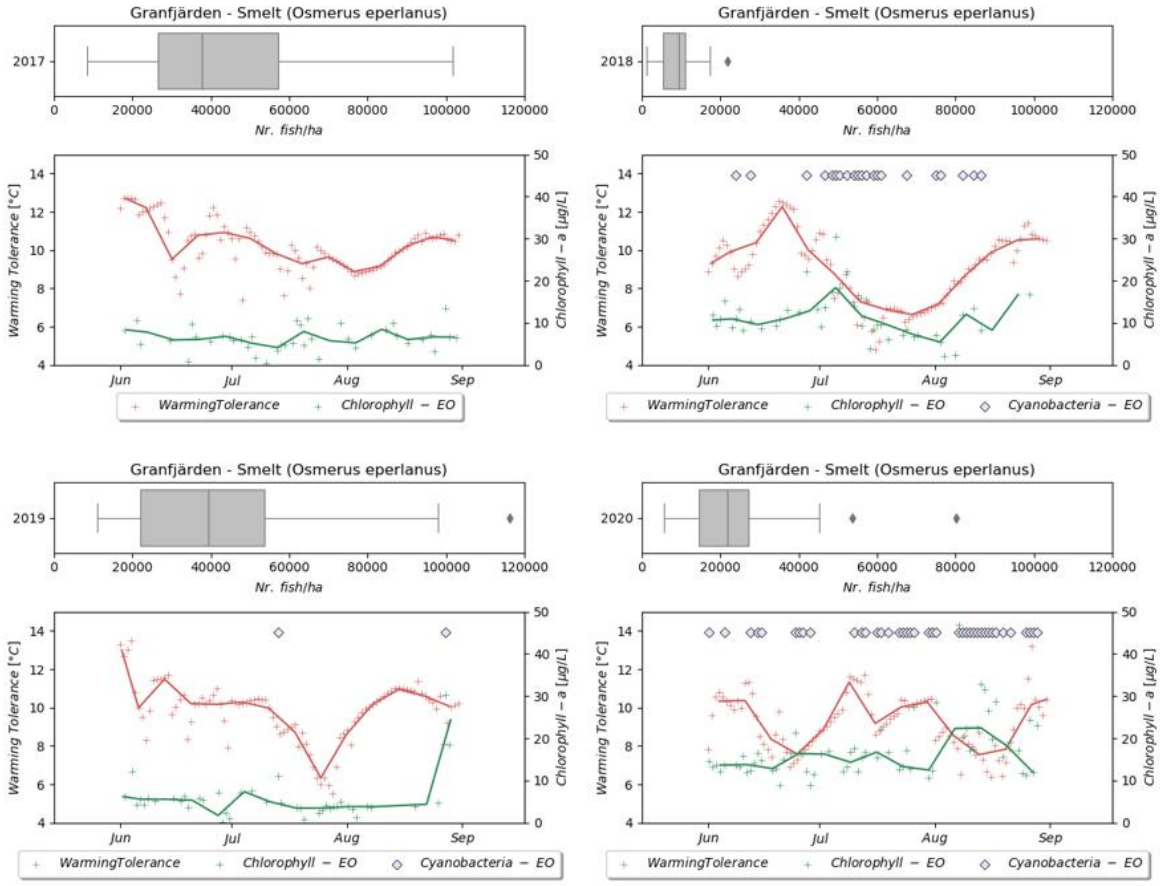



Figure 23 Chlorophyll-a, cyanobacteria abundance, fish abundance of smelt and warming tolerance of smelt at the Granfjärden transect for the summers of 2017, 2018, 2019 and 2020. Fish abundance data boxplots show measurements across the transects in September of the specified year. Warming tolerance (in red) and chlorophyll-a concentrations (in green) are given per observation (average across the transect, plus symbol) and as a weekly average (line). The occurrence of a cyanobacteria abundance is indicated with diamonds.

The fish survey data indicated lower abundances of Osmerus eperlanus in 2018 and 2020. To explore if these observations can be detected by the developed TECI's, TECI 8 – Habitat Conditions for Fish was generated for Granfjärden and plotted in Figure 23. These observations are also visible in the TECI scores (Figure 24), as the summers of 2018 and 2020 corresponds to show higher TECI scores compared to the other years, indicating high probabilities that these years have higher certainties of anomalies and also that the chlorophyll-a concentration and cyanobacteria abundance contribute the most to this higher TECI score.

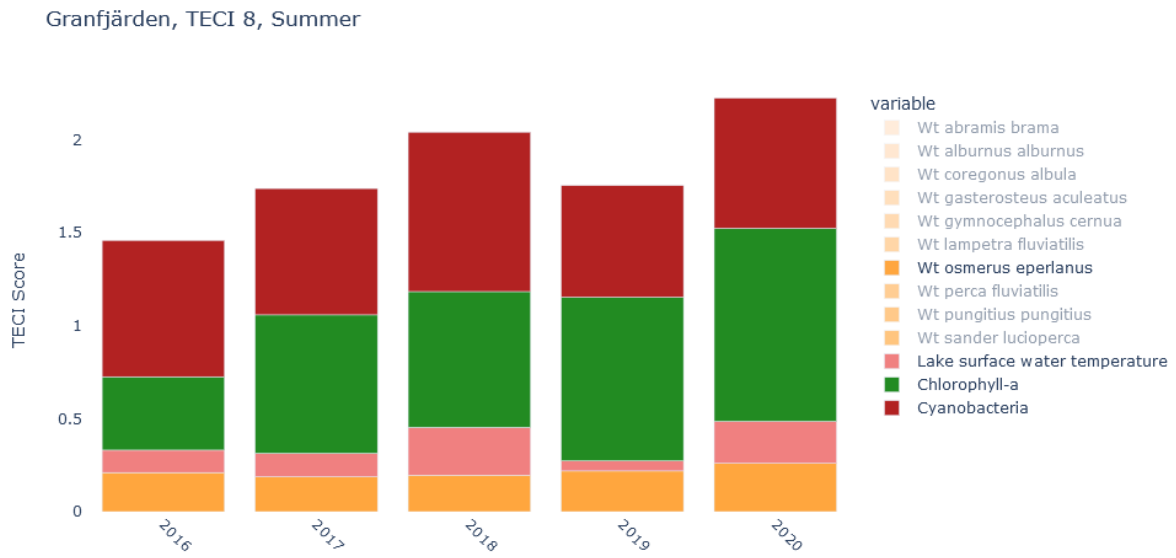



Figure 24 TECI results at Granfjärden for the summers of 2016, 2017, 2018, 2019 and 2020. Shown are the stacked TECI results of Cyanobacteria (red), Chlorophyll-a (green), warming tolerance of smelt (orange) and LSWT itself (pink).

### 5.2.3 Lake Balaton

The warming tolerance was also calculated from September 1999 up to and including October 2020 for 33 fish species in Lake Balaton (Specziár & Erős, 2020; species scientifically sampled or recorded on a photograph by citizens in 2018). The results on warming tolerance are available.

## 5.3 Pilot 3 – River Dams

In this pilot, we focus on the impact of river dams on freshwater ecosystems at river catchment scale. More specifically, the objective of this BIOMONDO pilot is to explore the possibilities for combining EO data and biodiversity modelling for monitoring and assessing the impact of dam construction and removal on biodiversity.

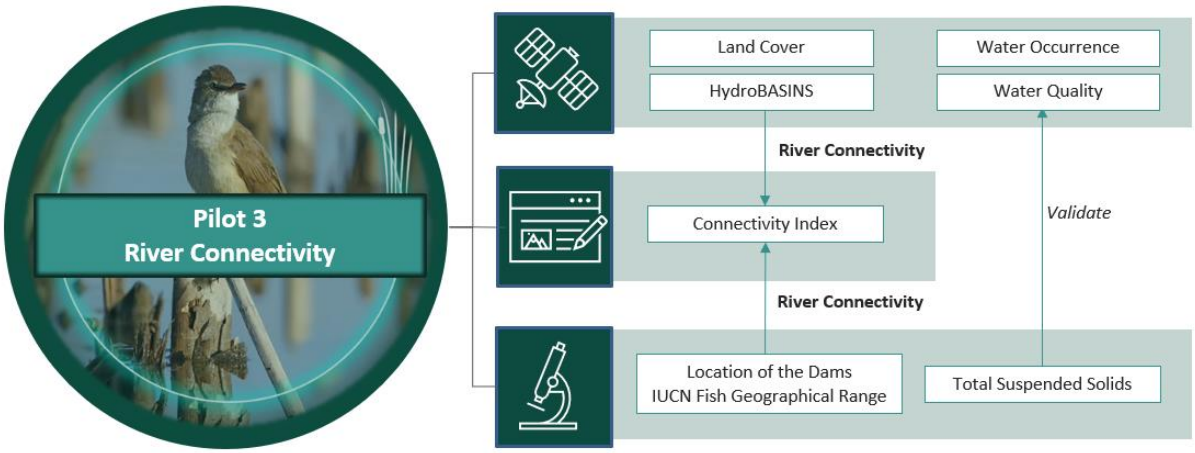



Figure 25 Overview of Pilot 2 components and workflow

Obstacles such as dams and other human-made waterworks heavily alter and interrupt dispersal routes of many species. Freshwater ecosystems are particularly sensitive to such a 'loss of connectivity' because: 1) there is little exchange of organisms between river basins, and 2) dispersal is constrained by the dendritic (tree-like) structure and directional flow of river networks. In addition, river dams and other waterworks affect water quality and disturb the natural flow regimes and habitat extent of aquatic and semi-aquatic species. Within this context, a reduction of sediment flows that may reach river deltas is of particular importance, because they may disappear due to ongoing processes of subsidence and erosion without the ongoing inflow of sediment from the wider catchment area. We aim to study these multiple, simultaneous effects of river dams, including differences in these effects between different types of dams. Ultimately, our findings should contribute to a decision framework that helps to weigh the pros (i.e. in terms of energy production) and cons of (different types of) individual river dams and their placement within a river basin (e.g. as in Schmitt *et al.* 2018).

We chose the Mekong delta as the primary site for this pilot because a lot of relevant changes have occurred in this delta in the period for which EO data are available.

### 5.3.1 Fragmentation of the Mekong basin

We studied historical and upcoming changes in river connectivity following a procedure co-developed by members of BIOMONDO and described in Barbarossa *et al.* (2020). This procedure results in an assessment of the degree of geographic range fragmentation

across the entire Mekong basin, expressed as a connectivity index (CI) (range 0-1) where 1 represents a range that is fully connected and 0.5 results from a dam dividing a range into two equally sized fragments. Changes in the geographic range connectivity as new dams were built since the 1960s were estimated (see Figure 26).

While the cumulative installed capacity has increased in a relatively steady manner, there were relatively large drops in connectivity in e.g. 1994 and 2019. Drops in mean connectivity were from 94 to 85 and from 73 to 60, respectively, while the gains in installed capacity were relatively small, i.e. 136 and 3670 MW (see Figure 26). This suggests that the impacts of some dams on connectivity is much larger relative to their energy production when compared to other dams.

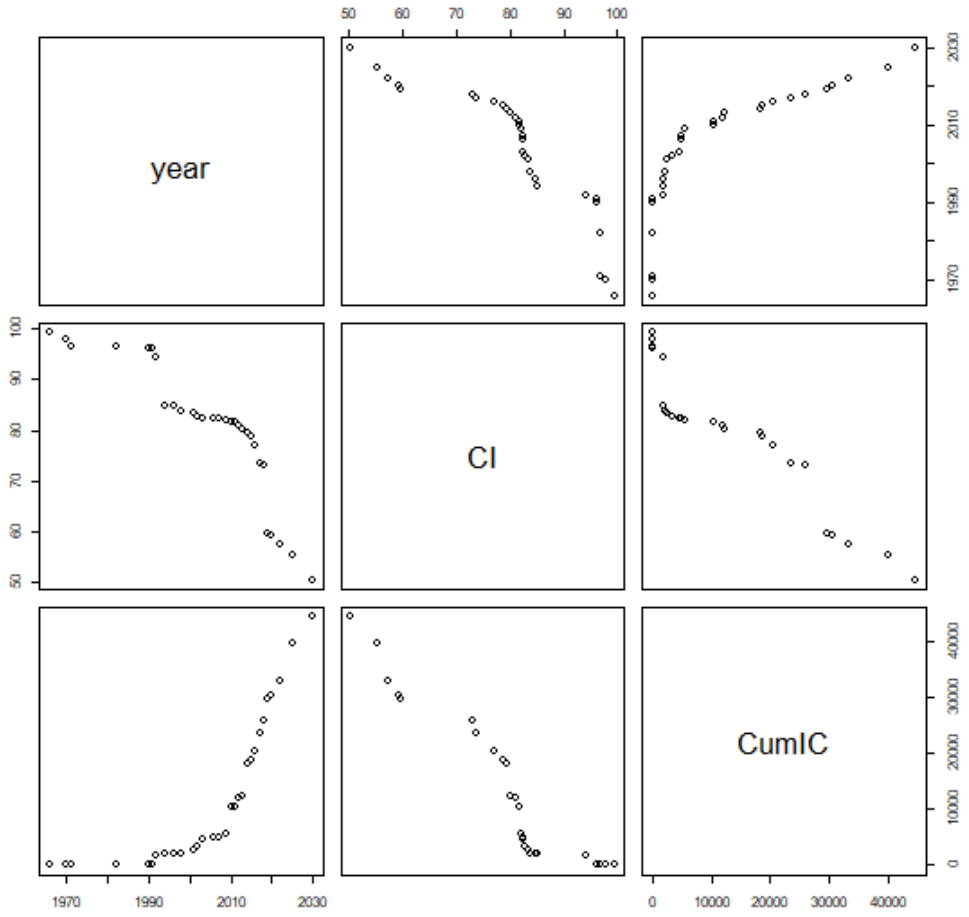



Figure 26 Mean Connectivity Index (CI, in percent) and the total energy production by river dams (i.e. the Cumulative Installed Capacity, CumIC) over the years in the Mekong delta.

This finding was further explored by an analysis of the impact of individual dams on the connectivity of the Mekong basin, which was studied by determining the impact of the removal of a single dam (from a set including existing dams and dams that are currently under construction) on the average connectivity of all, diadromous, and non-d diadromous fish species. We found that, in particular, the removal of the Don Sahong dam would

increase the average connectivity substantially: across all species from 0.611 to 0.686, from 0.767 to 0.852 for diadromous species and from 0.599 to 0.675 for non-diadromous species. For example, there may also be substantial differences in impact between diadromous and non-diadromous for particular dams. For example, removal of the Pak Mun dam would increase average connectivity to 0.648, 0.768, and 0.639, for all, diadromous, and non-diadromous species respectively. Removal of the Pak Mun dam would thus mainly affect the average range connectivity of non-diadromous species while it has hardly any effect on the connectivity of diadromous species. On average, the removal of a single dam would lead to an increase in average connectivity to 0.614, 0.769, and 0.603 for all, diadromous, and non-diadromous fish species. In most cases, the removal of a single dam thus has only a limited impact on the range connectivity of fish species.

### 5.3.2 Impact of river dams on water quality within the Mekong Basin

Remote sensing is a valuable tool that can be used to monitor changes in sediment transport caused by the construction of dams. In this pilot EO data is used to analyse changes in the colour and turbidity of the water. Sediment-laden water is typically more turbid and has a different colour than clear water. By analysing satellite images of the river it is possible to detect changes in water colour and turbidity that may be indicative of changes in sediment transport. Four dams were selected as a primary case study based on Schmitt et al., 2018 for further investigation (Table 6).

Table 6 Case study dams

| **Dam** | **Operating since** | **Height** | **River** | ***country*** |
|---|---|---|---|---|
| Xe Kaman 1 | Dec. 2017 | 120m | Xe Kaman | Laos |
| Lower Se San 2 | Dec. 2018 | 75m | Se San | Cambodia |
| Xe-Pian,Xe-Namnoy | Dec. 2019 | 74m | Sekong | Laos |
| Nam Giep 1 | Sep. 2019 | 167m | Ngiep | Laos |

HydroSHEDS (Hydrological data and maps based on SHuttle Elevation Derivatives) is a mapping project developed by the World Wildlife Fund (WWF) and the US Geological Survey (USGS) that provides a comprehensive and consistent set of global hydrological data and maps. One of the main outputs of HydroSHEDS is the delineation of hydrologic basins or watersheds across the globe. The basins for the pilot 3 analysis were retrieved from this database (Wickel et al. 2007). Within the BIOMONDO team we decided to create subbasins based on the Level 12 HydroSHEDS basins which touched by the reservoir of each dam. Figure 27 shows an example of the created subbasin for Xe Pian Xe Namnoy.

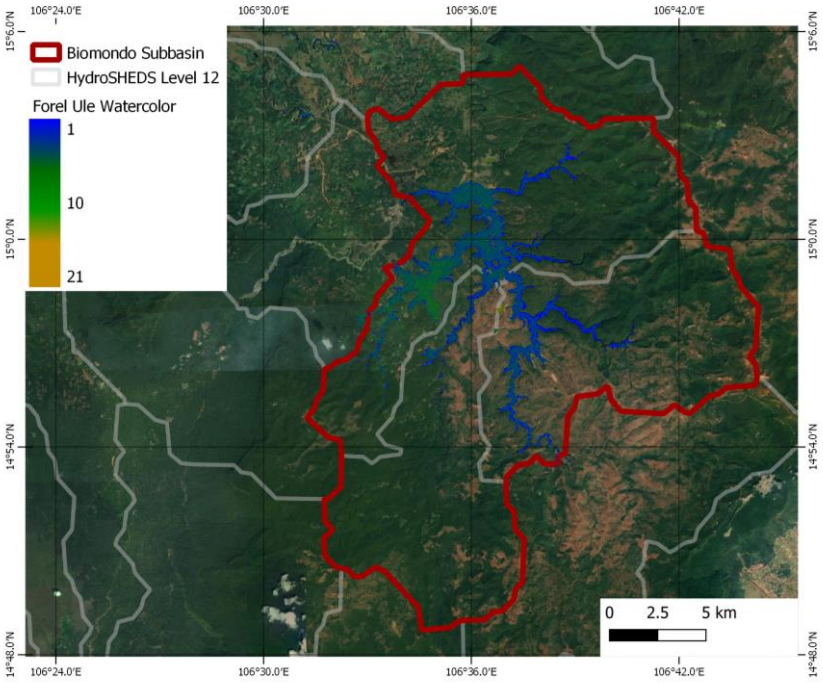



Figure 27 BIOMONDO subbasin (red) of the Xe Pian Xe Namnoy dam based on the HydroSHEDS Level 12 basins (grey). All basins touched by the reservoir are considered.

To investigate the impact of river dams on the river water quality we analysed the EO based Forel Ule Value (FUV, Ch. 4.1.1.2). Figure 28 shows the water colour (FUV) of the Lower Se San2 downstream river water. The heatmaps shows that for the years 2018-2020 the colour was more blueish compared to the earlier year. The first turbine for the Lower Se San 2 began producing electricity in December 2017. It is assumed that with the operational start of the dam the sediment transport decreased, which also decreased the brownish colour of the water. The Forel Ule values of 0-5 correspond to turbidity values of 0-20 FNU and is increasing with Forel Ule values from 16-21 corresponding to values of >100 FNU.

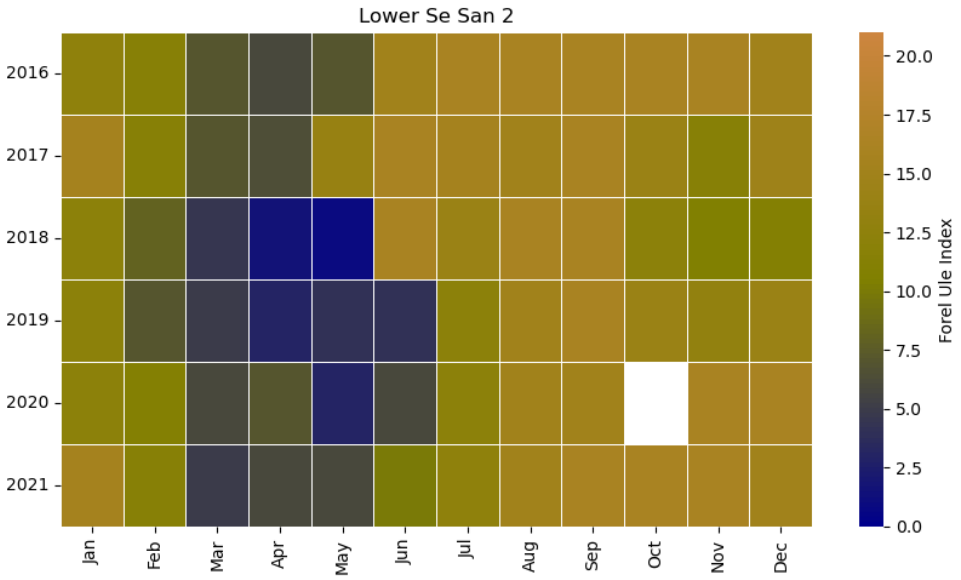



Figure 28 Forel Ule Value of the Lower Se San 2 BIOMONDO subbasin

Similar results can be investigated within the subbasin of the Nam Giep 1 dam. The Forel Ule value prior to September 2018 show that there is a drastic change in the water color of all subbasin water pixel. The Nam Giep reservoir has FU values around 2-4 showing mainly clear water pixel.

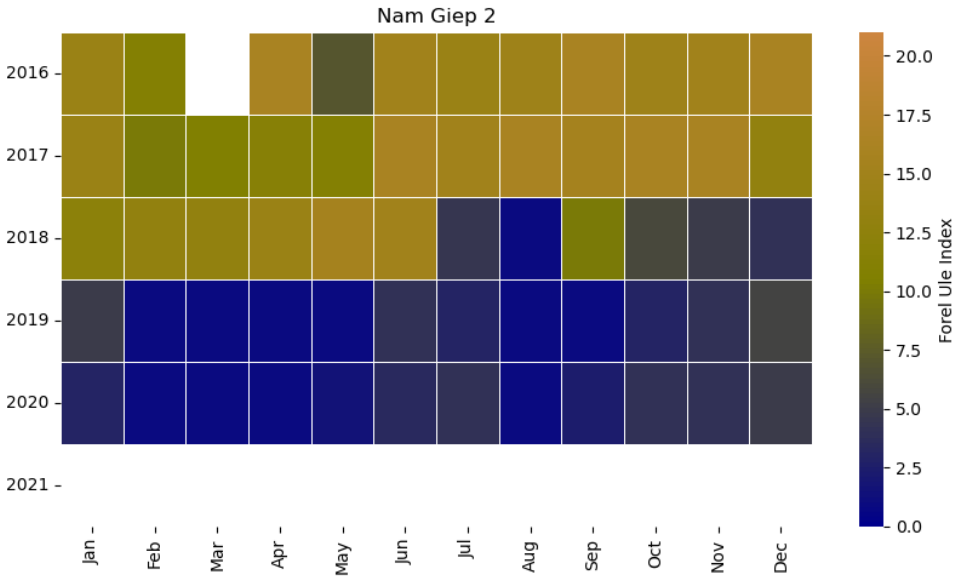



Figure 29: Forel Ule Value of the BIOMONDO subbasin of Nam Giep 2

Remote sensing transect analysis of rivers is a powerful technique that can be used to study changes in sediment transport over time. The transect analysis is particularly useful for studying changes in sediment transport caused by human activities, such as the construction of dams or changes in land use. These activities can alter the natural flow of water and sediment in a river, leading to changes in channel morphology and sediment deposition pattern.

A transect analysis was performed on the studied dams in Table 6. The transect included river parts down and upstream of the dam. The changes in the turbidity and the water-color were analysed comparing different years. Figure 30 shows both parameters over the longitude of the transect and the different years. Both parameters show that earlier years had higher turbidity values and the water colour was around 16 along the whole transect showing a brownish colour. In 2018 the dam was operational, and the turbidity values decreased for the whole transect. Especially upstream of the dam (dam reservoir) shows very low turbidity value and a very blueish water colour (values around 2-5 FU). Downstream of the dam the turbidity value increased with increasing distance to the dam. The same can be observed for the water colour value. The transect analysis shows that the Nam Giep 1 dam has an impact on the sediment transport of the river and that not only upstream bot also downstream of the dam the sediment load is decreased.


Figure 30 Transect analysis of Nam Giep 1. Turbidity (top) and water color (bottom) were analysed for different years. Grey box shows the area of the dam where no values were extracted.

Furthermore, we investigated the Land Cover change of the dam surroundings. For the analysis the Copernicus GLOBAL LAND COVER product is used to assess the change within the land cover classes around the dams. This dataset was selected based on the availability of the critical years between 2015 and 2019. Further Land Cover/Land Use products were evaluated but did not meet the temporal criteria needed. The changes of the land cover were analysed in the BIOMONDO subbasins of each river. Figure 31 shows the changes of the Xekaman 1 within the years 2015-2019. The increase of 7206 pixels in the water body class show that the building of the dam led to the development of a dam reservoir. The dominant land cover prior to the dam was evergreen forest and cropland.

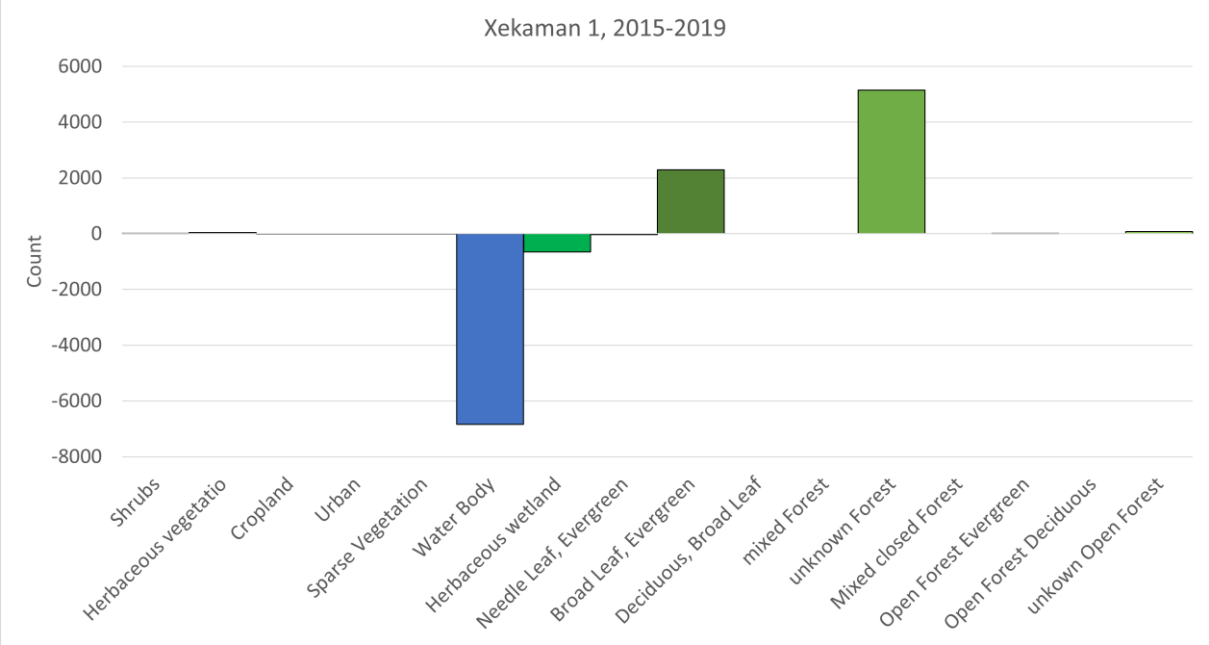


Figure 31 Land cover changes within 2015-2019 in the Xekaman 1 BIOMONDO subbasin

Analysis of the TECI 4 – Water Quality of the Xe-Pian Xe-Namnoy in Summer showed a increased TECI score for 2018 driven mainly by turbidity (Figure 32). This is in line with the dam collapse in the summer of 2018, the Xe-Pian Xe-Namnoy dam in southern Laos suffered a catastrophic failure that resulted in severe flooding and widespread damage. On July 23, heavy rainfall caused the dam to overflow and collapse, releasing a massive amount of water downstream. The flooding caused significant damage to nearby villages and infrastructure, including homes, roads, and bridges. According to official reports, more than 40 people were killed, and thousands were displaced. This demonstrates that the TECI score can indicate high impacts on biodiversity due to extreme events and which parameters is the driver.

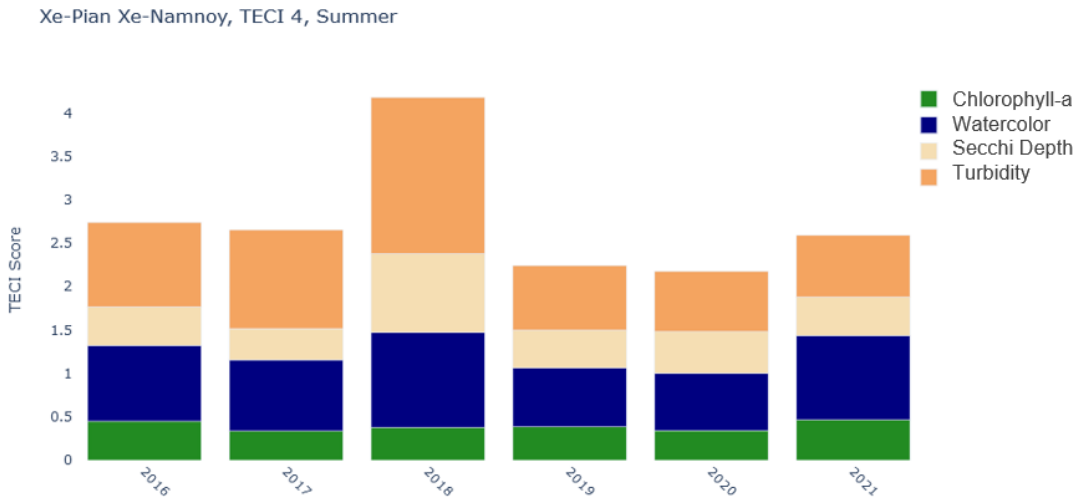



Figure 32 TECI 4 Water Quality score of the Xe-Pian Xe Namnoy dam reservoir in summer

### 5.3.3 The multiple simultaneous effects of river dams

Decision making regarding the placement of river dams, ultimately, involves a consideration of the multiple simultaneous impacts of river dams, e.g. on green energy production, habitat connectivity, and sedimentation processes.

One way to approach this problem is by determining the extent to which existing or planned river dams deviate from the pareto optimal set (Giagkiozis & Fleming, 2014). This is a model generated optimal distribution (in our case of dams across the Mekong River basin) that takes multiple objectives into account. We determined such an optimal set when considering the energy produced and the impact of river dams on habitat connectivity. The ideal dam would have a minimal impact on connectivity while producing the maximum possible amount of green energy. The extent to which existing and planned river dams correspond to this ideal was determined by determining each dam's inclusion probability in the optimal pareto front. A high inclusion probability corresponds, in this case, to dams that (likely) are close, while a low inclusion probability to dams that are far away from this ideal situation. When planning new dams you would thus want to choose a location where the inclusion probability is high, while dam removal would be most beneficial at locations where the inclusion probability is low.

To facilitate the assessment of these multiple simultaneous effects of river dams the inclusion probability as well as information on a variety of other river-dam impacts were made available in a viewer (see Figure 33). More about the viewer and the capabilities are documented in the D2.4 document (Table 1). In addition to facilitating a more comprehensive understanding of the various, simultaneous impacts of river dams this viewer also allows us to critically evaluate our results with the help of stakeholders and experts. One outcome of this evaluation could be that the EO could be helpful when improving the calculation of habitat connectivity which currently uses HydroBASINS as its main unit of calculation. Updates to the shape of HydroBASINS after dam placement (as in Figure 27) and/or a more detailed mapping of (alternative) river paths may help to improve this calculation and may become a valuable input for the BIOMODO's roadmap on how to improve freshwater biodiversity monitoring from space.

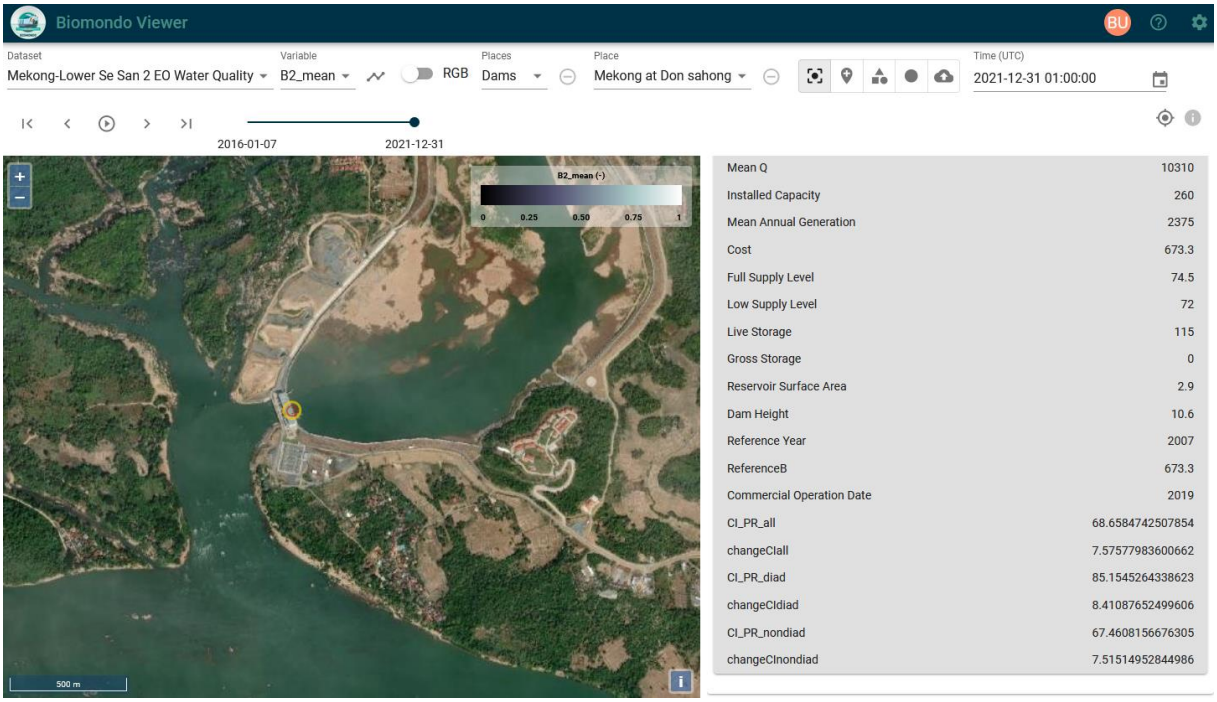


Figure 33 To allow for a further exploration of the multiple simultaneous effects of individual dams key information was made available in a viewer. CI_PR = Connectivity Index_Post Removal.

# 6 References


Alfonso, S., Gesto, M., & Sadoul, B. (2021). Temperature increase and its effects on fish stress physiology in the context of global warming. Journal of Fish Biology, 98(6), 1496-1508.

Alvera-Azcárate, A., Barth, A., Rixen, M., & Beckers, J. M. (2005). Reconstruction of incomplete oceanographic data sets using empirical orthogonal functions: application to the Adriatic Sea surface temperature. *Ocean Modelling*, *9*(4), 325-346.

Alvera-Azcárate, A., Barth, A., Beckers, J. M., & Weisberg, R. H. (2007). Multivariate reconstruction of missing data in sea surface temperature, chlorophyll, and wind satellite fields. *Journal of Geophysical Research: Oceans*, *112*(C3).

Alvera-Azcárate, A., Vanhellemont, Q., Ruddick, K., Barth, A., & Beckers, J. M. (2015). Analysis of high frequency geostationary ocean colour data using DINEOF. *Estuarine, Coastal and Shelf Science*, *159*, 28-36.

Axenrot, T. (2020). Hydroakustik i sötvatten. Ett verktyg i fisk- och miljöövervakning. Aqua reports 2020:12. Department of Aquatic Resources (SLU Aqua), Swedish University of Agricultural Sciences. axenrot_t_200924.pdf (slu.se)

Axenrot. T. & Sandström, A. (SLU) (2022). Personal correspondence.

Barbarossa, V., Bosmans, J., Wanders, N., King, H., Bierkens, M. F., Huijbregts, M. A., & Schipper, A. M. (2021). Threats of global warming to the world's freshwater fishes. Nature communications, 12(1), 1-10.

Barbarossa, V., Schmitt, R. J., Huijbregts, M. A., Zarfl, C., King, H., & Schipper, A. M. (2020). Impacts of current and future large dams on the geographic range connectivity of freshwater fish worldwide. Proceedings of the National Academy of Sciences, 117(7), 3648-3655.

Barth, A., Alvera-Azcárate, A., Licer, M., & Beckers, J. M. (2020). DINCAE 1.0: a convolutional neural network with error estimates to reconstruct sea surface temperature satellite observations. *Geoscientific Model Development*, *13*(3), 1609-1622.

Bennett, S., Wernberg, T., Arackal Joy, B., De Bettignies, T., & Campbell, A. H. (2015). Central and rear-edge populations can be equally vulnerable to warming. Nature Communications, 6(1), 1-7.

Blaas, M., P. Boderie, R. Noordhuis & S. Gaytan Aguilar (2016) Water quality model Lake Marken. FP7 INFORM deliverable D6.2.

Brockmann, C., Doerffer, R., Peters, M., Kerstin, S., Embacher, S., & Ruescas, A. (2016, August). 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. In *Living Planet Symposium* (Vol. 740, p. 54).

Comte, L., & Olden, J. D. (2017b). Climatic vulnerability of the world's freshwater and marine fishes. Nature Climate Change, 7(10), 718–722.

Cote, D., Kehler, D. G., Bourne, C., & Wiersma, Y. F. (2009). A new measure of longitudinal connectivity for stream networks. Landscape Ecology, 24, 101-113.

Clusella-Trullas, S., Garcia, R. A., Terblanche, J. S., & Hoffmann, A. A. (2021). How useful are thermal vulnerability indices?. *Trends in Ecology & Evolution*, *36*(11), 1000-1010.

Doerffer, R., & Schiller, H. (2007). The MERIS Case 2 water algorithm. *International Journal of Remote Sensing*, *28*(3-4), 517-535.

Froese, R. and D. Pauly. Editors. 2022. FishBase. World Wide Web electronic publication. www.fishbase.org, ( 06/2022 )

Geller, G. N., Halpin, P. N., Helmuth, B., Hestir, E. L., Skidmore, A., Abrams, M. J., … & Williams, K. (2017). Remote sensing for biodiversity. In *The GEO handbook on biodiversity observation networks* (pp. 187-210). Springer, Cham.

Giagkiozis, I., & Fleming, P. J. (2014). Pareto front estimation for decision making. Evolutionary computation, 22(4), 651-678.

Goolsby, E. W., Bruggeman, J., & Ané, C. (2017). Rphylopars: fast multivariate phylogenetic comparative methods for missing data and within-species variation. Methods in Ecology and Evolution, 8(1), 22-27.

Gray, J., Sulla-Menashe, D. & Friedl, M. A. User guide to collection 6 modis land cover dynamics (mcd12q2) product. NASA EOSDIS Land Processes DAAC: Missoula, MT, USA (2019).

Kritzberg, E. S., Hasselquist, E. M., Škerlep, M., Löfgren, S., Olsson, O., Stadmark, J., ... & Laudon, H. (2020). Browning of freshwaters: Consequences to ecosystem services, underlying drivers, and potential mitigation measures. *Ambio*, *49*(2), 375-390.

IMARES (2007). Zomersterfte spiering in het IJsselmeer en Markermeer. Available from: https://edepot.wur.nl/146639

de Leeuw, J. (2022). Personal correspondence.

Lesser, G.R., J.A. Roelvink, J.A.T.M., van Kester & G.S. Stelling (2004) Development and validation of a three-dimensional model. Coastal Engineering 51: 883-915.

Leiva, F. P., Calosi, P., & Verberk, W. C. E. P. (2019). Scaling of thermal tolerance with body mass and genome size in ectotherms: a comparison between water- and air-breathers. Phil. Trans. R. Soc. B374(1778), 20190035.

McArley, T. J., Hickey, A. J. R., & Herbert, N. A. (2017). Chronic warm exposure impairs growth performance and reduces thermal safety margins in the common triplefin fish (Forsterygion lapillum). Journal of Experimental Biology, 220, 3527–3535.

Meyer-Jacob, C., Michelutti, N., Paterson, A. M., Cumming, B. F., Keller, W. B., & Smol, J. P. (2019). The browning and re-browning of lakes: Divergent lake-water organic carbon trends linked to acid deposition and climate change. *Scientific reports*, *9*(1), 1-10.

Mitt i (2018). Småfiskar dör i Mälaren på grund av hettan. Available from: https://www.mitti.se/nyheter/smafiskar-dor-i-malaren-pagrund-av-hettan/lmrhf!4157697/

Noordhollands Dagblad (2018). Herstel van de paling: meer jonge aal in IJsselmeer. Available fromwww.noordhollandsdagblad.nl/cnt/dmf20180924_63215512?utm_source=google&utm_medium=organic

Roberts, D. R., Bahn, V., Ciuti, S., Boyce, M. S., Elith, J., Guillera-Arroita, G., ... & Dormann, C. F. (2016). Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. Ecography, 40(8), 913-929.

Sadoul, B., & Vijayan, M. M. (2016). Stress and growth. In Fish physiology (Vol. 35, pp. 167-205). Academic Press.

Schmitt, R. J., Bizzi, S., Castelletti, A., & Kondolf, G. M. (2018). Improved trade-offs of hydropower and sand connectivity by strategic dam planning in the Mekong. *Nature Sustainability*, *1*(2), 96-104.

Schmitt, R. J., Bizzi, S., Castelletti, A., Opperman, J. J., & Kondolf, G. M. (2019). Planning dam portfolios for low sediment trapping shows limits for sustainable hydropower in the Mekong. Science advances, 5(10), eaaw2175.

Specziár, A., & Erős, T. (2020). Development of a fish-based index for the assessment of the ecological status of Lake Balaton in the absence of present day reference condition. Knowledge & Management of Aquatic Ecosystems, (421), 11.

Van Duin, E.H.S., 1992. Sediment transport, light and algal growth in the Markermeer : a two-dimensional water quality model for a shallow lake. PhD thesis Wageningen University.

Wernand, M. R., & Van der Woerd, H. J. (2010). Spectral analysis of the Forel-Ule Ocean colour comparator scale. *Journal of the European Optical Society-Rapid Publications*, *5*.

Wickel, B. A., Lehner, B., & Sindorf, N. (2007, December). HydroSHEDS: A global comprehensive hydrographic dataset. In AGU Fall Meeting Abstracts (Vol. 2007, pp. H11H-05).

# Appendix 1

Table 7 List of indicators for a freshwater biodiversity laboratory

| **1 – Biodiversity Change Drivers** | | | | | |
|---|---|---|---|---|---|
| **1.1 – Habitat function & structure** | | | | | |
| **Indicator** | | | **Data source** | **Readiness/Processing** | **iiEES potential** |
| 1 | Land use and land cover change in the watershed (catchment), including deforestation | Land use/land cover change and forest cover products are readily available globally or for most regions of the world, and products should be selected based on the suitability for a given region and the application. As mentioned earlier, care should be taken when using such wide-area products to ensure they are sufficiently accurate and appropriate for local conditions; remember that land use and land cover products often have average accuracies hovering around 80 % (local accuracies can be less). | LCLU: Copernicus Corine Land Cover (CLC), European coverage, 25ha MMU, 5 time steps since 1990, including a change layer. CLMS-CGLS LandCover from ProbaV, global, 100m, 2015-19<br>ESA Worldcover, 10m, 1 time step (product expected end of 2021)<br>Also from CLMS: European settlement map, high resolution layers: forests, water+wetness (2 time steps 2015 and 2018)<br>C3S LC, long consistent time series at 300m.<br><br>Deforestation: Sen4REDD+, but available: Univ. Maryland Global Forest Change Results from time-series analysis of Landsat images characterizing forest extent and change.<br><br>Otherwise: S2 and S1 for own classification. | Data Access<br>We use available products where available and appropriate. | yes |
| 2 | Area and location of rivers, lakes, impoundments and wetlands | Remote sensing can also directly map the area and extent of habitat, typically by mapping wetland, floodplain and riparian vegetation, or by mapping the area of lakes. | Freshwater ecosystem explorer (UN)<br>Contains "Permanent and seasonal surface water dynamics data"<br>JRC Global Surface Water product maps the location and temporal distribution of water surfaces at the global scale over the past 3.6 decades and provides statistics on their extent and change to support better informed water-management decision-making. | | yes |

| | | | | | |
|---|---|---|---|---|---|
| 3 | Habitats (of rivers, lakes ...) such as submerged or emergent macrophytes and riparian forests | same as above | Macrophytes: from raw satellites data Riparian zone Europe: Copernicus Service, Riparian LCLU; no change but only 2012 | Data Access for CLMS Riparian zone product | yes |
| 4 | Habitat connectivity along the water body and to adjacent water bodies and terrestrial ecosystems | The lateral connectivity of a river—its connection to the floodplain and wetlands—is observed directly from identification of the riparian buffer and from water inundation mapping.<br>Longitudinal connectivity can be derived from number and location of dams in the river | Possibly, this can be derived from 2 variables, namely "identification of riparian buffer" in CLMS, and "water inundation mapping".<br><br>Available data sets: GRaND, GOODD, FHRED (=planned), several regional data sets. Also several projects going on to complete these sets. O.a. AMBER project | Data Access to the two base products and the own processing to derive the habitat connectivity | maybe (interesting but maybe too much work) |
| **1.2 - Drivers - Biophysical/hydrological** | | | | | |
| **Indicator** | | | **Data source** | **Readiness/Processing** | **iiEES potential** |
| 5 | Water body extent (a proxy for volume) and retention time | Inundation mapping over a time-series of satellite observations provides information about the water extent and retention time and the hydro-period for a freshwater system. | Same as #2 | | yes |
| 6 | Hydro-period (the temporal pattern of high and low water) | same as above | Is included in JRC Global Water product see #2. It shows the typical temporal pattern, and allows also to go into each year individually | | yes |
| 7 | Water column trophic status, especially eutrophication and sediment load | Water column trophic status can be determined from estimates of chlorophyll, sediment and coloured dissolved organic matter concentration, or from estimates of water clarity/Secchi depth using optical RS. | Copernicus CGLS (produced by BC/PML). If not appropriate, we can start from S3 and S2 data. Note: a 100m S2 based product is not yet included in CGLS NRT service but will come. | Data Access<br>CGLS inland water products at 300m<br>Own processing for HR WQ product from S2 | yes |
| 8 | Submerged vegetation | Vegetation community identification is typically approached through classification procedures. These procedures work because they take advantage of a physical characteristic of green vegetation: strong absorption of red and blue wavelengths by the chlorophyll in the surface layers, and reflectance in the near infrared from the inner cell structure. Measuring reflectance in those wavelengths can be related to vegetation properties such as biomass or stress, which are the first order properties used for mapping the | There is no operational service producing this product. Algorithms are available in scientific literature. | Own processing:<br>Submerged macrophytes based on S2 and L8, should be done like to #3 | yes |

| | | | | | |
|---|---|---|---|---|---|
| | | specialised vegetation communities that occupy different wetland zones and are good proxies for habitat diversity. | | | |
| 9 | Species and Ecosystem Services | Direct observation of species and habitats is possible in limited cases—with high spatial and/or spectral resolution data it is possible to directly observe riparian, wetland, and submerged macrophytes and IAS. Such images tend to be expensive, however. To further detail classification to species level or to identify intrinsic species characteristics or processes, airborne hyperspectral sensors may be needed since current spaceborne systems do not have sufficient spatial or spectral resolution for this, and the small size of many wetland communities can make airborne monitoring practical. In submerged aquatic plant communities, species differentiation may be possible because the fine spectral bands measured by a hyperspectral sensor allow for more precise characterisation of individual plant species reflectance. This type of data can then be linked to intrinsic plant physiological processes. For example, it is possible to use hyperspectral reflectance characteristics and stable isotope markers to distinguish native submerged plant species from submerged IAS because they use different photosynthetic pathways. This information can provide insight into IAS adaptation traits for freshwater ecosystems (Santos et al. 2012). Using hyperspectral data, it is also possible to measure the foliar chemistry of inundated plants or other biotic communities such as the cyanobacteria commonly associated with harmful algal blooms (HABs). These allow inferences about the status of freshwater ecosystem services such as safe drinking water, nutrient cycling/eutrophication and carbon cycling. However, additional datasets are usually required for these inferences, such as laboratory samples of foliar chemistry, photosynthetic rates, respiration rates, stable isotope concentrations, biomass, and other measurable properties of plant species. | It would require buying VHR data, or making own measurements by drones. Use simulated hyperspectral data and test with PRISMA.<br>Greatest potential is to address with information by combining other drivers and EBVs with ecosystem models. | | something for the longer term, Roadmap |
| 10 | Invasive alien species (IAS) | same as #9 | | | long term |
| **2 - Essential Biodiversity Variables** | | | | | |

| Indicator | | | Data source | Readiness/Processing | iiEES potential |
|---|---|---|---|---|---|
| 11 | Species populations - Population abundance | Species abundances: Predicted count of individuals over contiguous spatial and temporal units addressing the global extent of a species group. | VI in the riparian zone: CLMS, combination of classification from Riparian product and VI from Phenology service<br><br>Chl in the lakes: same as #7 | Data Access<br>LC and VI from CLMS CHL from CGLS (S3)<br>Own processing for HR WQ product from S2 | yes |
| 12 | Species traits - Phenology | Phenology: Presence, absence, abundance or duration of seasonal activities of organisms. | Riparian zone, wetlands: Copernicus LMS, Pan European High Resolution Vegetation Phenology, data going back to 2017 only!<br><br>In-water, Chl: deriving it directly from S3 and S2; based on CIWAWA (BC) | Data Access<br>Phenology from CLMS (available Q2/2021)<br>Own processing: Chl Phenology derived from analysing the time series of CHL, from #11 | yes |
| 13 | Species traits - Body mass / biomass | Not GEOBON EBV but added by Vihervaara et al. (2017); for inland waters, this is to be understood as algae biomass | same as #7 | Data Access<br>CGLS inland water products at 300m<br>Own processing for HR WQ product from S2 | yes |
| 14 | Species traits - Physiological traits | Physiology: Chemical or physical functions promoting organism fitness and responses to environment. Vihervaara: Inland waters - state of streams and brooks | not known | - | no |
| 15 | Ecosystem structure - Habitat structure / condition | Not named in GEOBON; GEOBON: The spatial arrangement of ecosystem units collectively defined by organisms forming these units. EBV Live cover fraction = The horizontal (or projected) fraction of area covered by living organisms, such as vegetation, macroalgae or live hard coral. EBV Ecosystem distribution = The horizontal distribution of discrete ecosystem units. (there is also EBV ecosystem vertical profile)<br><br>Vihervaara Inland Water: algae, organic matter, inland water breeding birds, inland water fish stocks, state of streams and brooks | Copernicus CGLS provides LSWT, CHL, TUR at medium resolution.<br>S2 are suitable to derive bottom vegetation or emerged vegetation | Data Access<br>CGLS inland water products at 300m<br>Own processing for HR WQ product from S2 | maybe |
| 16 | Ecosystem structure - Ecosystem extent and fragmentation | Not in GEOBON<br>Vihervaara: Inland water breeding birds | - | - | no |

| | | | | | |
|---|---|---|---|---|---|
| 17 | Ecosystem structure - Ecosystem composition by functional type | Not in GEOBON<br>Vihervaara: Inland water algae | Species composition is possible – to some extend – with OLCI. But overall this should be left for hyperspectral data and thus for the Roadmap | - | no |
| 18 | Ecosystem function - Net primary productivity | GEOBON: Ecosystem primary productivity: The rate at which energy is transformed into organic matter primarily through photosynthesis. | Needs CHL, PAR, PE curves. I.e. we could use Chl from CGLS, but PAR we would calculate from S3 and S2 directly. Then we can also use our own CHL. A difficulty will be to get PE curves for the specific water bodies. | Data Access<br>CGLS inland water products at 300m<br>Own processing for HR product from S2 | yes |
| 19 | Ecosystem function - Ecosystem phenology | Not in Vihervaara but in GEOBON!<br><br>GEOBON: Duration and magnitude of cyclic processes observed at the ecosystem level, such as in vegetation activity, phytoplankton blooms, etc. | same as #12 | Data Access<br>Phenology from CLMS (available Q2/2021)<br>Own processing<br>CHL Phenology derived from analysing the time series of CHL, from #11 | yes |
| **3 – Model input and output** | | | | | |
| **3-1 –GLOBIO inputs** | | | | | |
| **Indicator** | | | **EO Data source** | **Readiness/Processing** | **iiEES potential** |
| 20 | Water temperature | | CGLS LSWT<br>No product for higher spatial resolution. Explore L8 and point to future LSTM | Data Access<br>CGLS inland water products at 300m, CCI Lakes at 1000m | yes |
| 21 | LCLU | | same as #1 | Data Access<br>CLMS and others, see #1 | yes |
| 22 | Map of major river dams | | part of #1 Available data sets: GRaND, GOODD, FHRED (=planned), several regional data sets. Also several projects going on to complete these sets. O.a. AMBER project. | see #1 | yes |
| **3.2 –Validation of GLOBIO outputs** | | | | | |
| **Indicator** | | | **EO Data source** | **Readiness/Processing** | **iiEES potential** |
| 23 | Algal blooms in lakes | Concentration of harmful algae in lakes | same as #7 | Data Access CGLS inland water products at 300m and, CCI Lakes at 1000m | yes |

| | | | | | |
|---|---|---|---|---|---|
| | | | | Own processing for WQ S2 | |
| 23a | Aquatic vegetation | | Can be done | | yes |
| 23b | MSA | | Not possible to detect from EO | | no |
| **3.3 –DELWAQ inputs** | | | | | |
| **Indicator** | | | **EO Data source** | **Readiness/Processing** | **iiEES potential** |
| 24 | bathymetry | | | | no |
| 25 | bed composition | | S2 own processing to map the bottom of the water body; algorithms are available in literature | | yes |
| 26 | wind | | CAMS | Data Access<br>CAMS | yes |
| 27 | water level | | available from CGLS, CCI Lakes | Data Access<br>CCI Lakes at 1000m | yes |
| 28 | stream velocity | | not possible from EO for rivers and lakes (to be confirmed) | - | no |
| 29 | wave characteristics | | not possible from EO for rivers and lakes (to be confirmed) | - | no |
| 30 | nutrients | | not possible from EO | - | no |
| **3-4 – Validation of DELWAQ outputs** | | | | | |
| **Indicator** | | | **EO Data source** | **Readiness/Processing** | **iiEES potential** |
| 31 | Total suspended matter | | CGLS & S2 own processing, same as #7 | Data Access<br>CGLS S3, CCI Lakes<br>Own processing for WQ S2 | yes |
| 32 | Primary Production | | same as #18 | | yes |
| 33 | Chlorophyll-a | | CGLS & S2 own processing, same as #7 | Data Access<br>CGLS S3, CCI Lakes<br>Own processing for WQ S2 | yes |
| 34 | Kd | | CGLS & S2 own processing, same as #7 | Own processing for WQ S2 | yes |

| 4 – Ecosystem functions | | | | | |
|---|---|---|---|---|---|
| **Indicator** | | | **EO Data source** | **Readiness/Processing** | **iiEES potential** |
| 35 | Disturbance regulation - Flood occurance - standing water | Petronelli: Global Flood Monitoring System, Global Inundation Extent from Multi-Satellites (1993–2007) | same as #2<br>JRC https://global-surface-water.appspot.com/ | | yes |
| 36 | Disturbance regulation - Eutrophication of water bodies - ocean colour | Petronelli: usual WQ parameters | CGLS & S2 own processing, same as #7 | | yes |
| 37 | Water regulation - Inland water dynamic - Change in water stage | Petronelli: Sentinel 3 altimetry (lakes level) | same as #27<br>available from CGLS | | yes |
| 38 | Water regulation - Inland water dynamic - Water body distribution | Petronelli: Global Surface Water (1984–2015) | same as #2<br>JRC https://global-surface-water.appspot.com/ | | yes |
| 39 | Soil/Sediment retention - Sediment plumes - Turbidity | Petronelli: ocean colour sensors | CGLS & S2 own processing, same as #7 | Data Access<br>CGLS S3, CCI Lakes<br>Own processing for WQ S2 | yes |
| 40 | Nutrient regulation - Nutrient availability - Chl-a & algae blooms | Petronelli: ocean colour sensors; interesting: Huang et al. (2014) used Landsat and MODIS data to characterize eutrophication in response to land use change in a lake's catchment | CGLS & S2 own processing, same as #7 | Data Access<br>CGLS S3, CCI Lakes<br>Own processing for WQ S2 | yes |
| 41 | Supporting habitats - Habitat extent - Land cover/forest cover | Petronelli: ESA global maps | same as #1 | | yes |
| 42 | Supporting habitats - Habitat extent - Water body distribution | | same as #2<br>JRC https://global-surface-water.appspot.com/ | | yes |
| 43 | Supporting habitats - Habitat extent - Inland Water dynamic | Wetland Trend Index | same as #5 (which is same as #2)<br>JRC https://global-surface-water.appspot.com/ | | yes |

| | | | | | |
|---|---|---|---|---|---|
| 44 | Supporting habitats - Habitat extent - Lake ice | | GCLOPS inland water, cryosphere products | | yes |
| 45 | Food - Production of fish stocks - Chl-A | | CGLS & S2 own processing, same as #7 | | yes |
| 46 | Food - Production of fish stocks - Water temperature | | CGLS inland water | | yes |
| 47 | Water supply - water provision - water body distribution | | same as #2<br>JRC https://global-surface-water.appspot.com/ | | yes |
| 48 | Water supply - water quality - SRS-based estimates of water quality | Petronelli: EnMAP | CGLS & S2 own processing, same as #7<br>For the roadmap: hyperspectral | | yes |

Explanation of the columns of Table 7

- The “Indicator columns” include an identifier (first column), a short name and a description according to the source publication (second and third columns).
- The “data source” column contains satellite data sources or service providing the required information. These data sources have been checked to be available, suitable, and accessible.
- The column “Readiness/Processing” describes the actions to be undertaken in order to transform the data from its source into a variable which is required by the indicator. This can be “data access” which means download and prepropress (subsetting, projection, …) or “own processing” where a thematic algorithm needs to be run on raw data.
- The column “iiEES potential” is an initial assessment if the indicator is possible to be included in an Earth System Science biodiversity laboratory.

Further explanations:

- Biodiversity Change drivers are according to Geller G.N. et al. (2017)

- EBVs are according to Viheraava (2017) and aligned with GEOBON EBV description
- Ecosystem Services are according to Petronelli (2017), Table 3 & 4

The colours are used to highlight “Indicators” (blue), “Data access” (green) and “own processing” (orange). No further meaning is associated with the colours.