Deliverable 2.3 v2.0- Waterbody-level classification using the EO-based indicator for the pilot study.

Lead: 3edata
Contributors: USTIR

Authors: Carmen Cillero (3edata); Yago Alonso(3edata); Federico Cheda (3edata);Harriet Wilson (USTIR); […]

Issue	Version	Date	Comments
1	1.0	13/09/2024	First version of the deliverable for internal submission
1	2.0	16/09/2024	Internal edits approved

EO data disclaimer: Copernicus Land Monitoring Service (CLMS) Lake Water Quality (LWQ) early release *(Calimnos v2.1 L3 test dataset (July 2024)* of Lake Tanganyika produced by Plymouth Marine Laboratory within the Copernicus Land Monitoring Service) and Lake Surface Water Temperature (LSWT).

Table of contents

1. Purpose of the document

This document is a summary of the initial development and preliminary results of the standalone Earth Observation (EO)-based indicator specifically designed for SDG indicator 6.3.2. The EO indicator is a 'proof-of-concept' demonstrated for Lake Tanganyika as a pilot waterbody. This document is a comprehensive document describing the design of the EO-based indicator, while a shorter user-focused document is available in D2.4. Together, D2.3 and D2.4 are considered the final outputs of Activity 2 within the UNEP funded seed project "Earth Observation pathway for SDG indicator 6.3.2".

2. Introduction

The objective of the project "**Earth Observation pathway for SDG indicator 6.3.2"** is to define the pathway to SDG-ready EO data for indicator 6.3.2 reporting through a transboundary pilot in Lake Tanganyika. Currently, many countries are struggling to report on SDG indicator 6.3.2, and Earth observation (EO) is an unrealized tool to increase spatial coverage and temporal consistency in waterbody level reporting. The project proposes three parallel activities:

1. Co-design and capacity building with decision makers and local focal points,
2. **Demonstrate an EO-based indicator tailored for SDG 6.3.2,**
3. Design User Dashboard for SDG-ready EO data.

At a global scale, many countries (especially low-income countries) have sparse in-situ measurements across water bodies and inconsistent time series with which to report on SDG 6.3.2. Evolving EO capabilities allow the opportunity to consider water quality on longer time series, establish baseline conditions and add an explicitly spatial component which is necessary for source-to-sea analysis.

This document is the third deliverable of Activity 2 (Del 2.3) focused on the description of the final formulation and results of the EO based indicator for Level 2 reporting of SDG 6.3.2. The indicator is intended to improve the capacity of water managers to identify problems and make informed decisions based upon data.

3. EO composite Indicator

The specific SDG indicator 6.3.2 is formulated as: **Proportion of bodies of water with good ambient water quality.**

Good ambient water quality is water of a certain standard that flows in our rivers, lakes and aquifers without causing harm to human or ecosystem health (UNEP, 2023). The SDG 6.3.2 indicator enables the impact of human development on ambient water quality to be evaluated over time and it provides an indication of the services that can be obtained from the aquatic ecosystems, such as clean water for drinking, preserved biodiversity, sustainable fisheries, and water for irrigation. This indicator provides a mechanism for determining whether water quality management measures are contributing to the improvement of water quality in inland water bodies (Pahlevan et al. 2022).

The indicator integrates 3 water quality parameters into a composite index, classifying the waterbody's status and being complementary to SDG 6.3.2 Level 1 parameters derived from in situ measurements. The EO-based indicator is intended to be an alternative option for the quantitative assessment of the water quality status but will also enable the identification of dynamic patterns and trends both at water body and global scales, offering valuable information for management purposes.

The water quality parameters measurable through EO satellite technology or derived from EO data and accessible through products meeting all the requirements (see Del 2.1 for justification), that were finally included in the formulation of the indicator were:

Chl-a

- It is commonly used as a proxy of phytoplankton biomass. Phytoplankton abundance and biomass increases due to increased eutrophication, which can be a result of increased nutrient concentrations.
- It is common to almost all taxonomic groups, so chl-a maps can be suitable for identifying and monitor blooms (No species or toxicity)

Turbidity

- Is the measure of relative clarity of a liquid. It is a measurement of the amount of light that is scattered by material in the water

- High concentrations of particulate matter affect light penetration and ecological productivity, recreational values, and habitat quality (for fish and other aquatic life), and cause lakes to fill in faster.

Lake Surface Temperature

- It affects biological activity and growth.
- It regulates water chemistry and rate of chemical reactions
- Affects Oxygen availability

3.1. Final formulation of the indicator

The final formulation of the indicator was inspired by the CCME Water Quality Index (CCME WQI) (Canadian Council of Ministers of the Environment, 2001; 2017) with the aim of summarising water quality data from multiple variables into a single metric and facilitating its communication to a general audience.

The original CCME WQI Index incorporates three elements: *scope* - the number of parameters not meeting water quality guidelines; *frequency* - the number of times these guidelines are not met; and *amplitude* - the amount by which the guidelines are not met. The index produces a number between 0 (worst water quality) and 100 (best water quality).

The CCME formulation requires a minimum number of 4 water quality parameters, but is more robust when this number is equal to 8 or above. Currently there are insufficient global EO water quality products to reach the number of parameters suggested (see Del 2.1) however the CCME formulation offers the possibility of easily integrating new parameters when they become available or when they are ready at a national/regional/local scale.

To offer a stand-alone indicator for SDG 6.3.2 level 2 reporting, which has potential global scalability, a methodology was developed to also provide threshold values, which are based on the baseline conditions. As it is described in Del 2.1, we used the Copernicus Land Monitoring Service (CLMS) Lake Water Quality (LWQ) early release *(Calimnos v2.1 L3 test dataset (July 2024)* of Lake Tanganyika produced by Plymouth Marine Laboratory within the Copernicus Land Monitoring Service) and the Lake Surface Water Temperature (LSWT) Products for the calculation of the indicator, using the 2002-2012 dataset as the baseline period and for defining the threshold values and the 2017-2019 dataset to calculate and report the indicator for the SDG 6.3.2 Reporting Period 2 (RP2).

The original CCME formulation was further reduced after user and expert consultation to reduce the complexity and align closer to the Level 1 formulation of the S.D.G 6.3.2. The steps taken to simplify the indicator resulted in the selection of only one of the 3 original elements: frequency. This final approach allowed us to maintain the simplicity and flexibility of the indicator while aligning with the SDG 6.3.2 Level 1 reporting methodology.

Figure 1 shows the workflow of the EO indicator and the following sections will enter into more detail regarding the calculations and results for the RP2 in Lake Tanganyika.

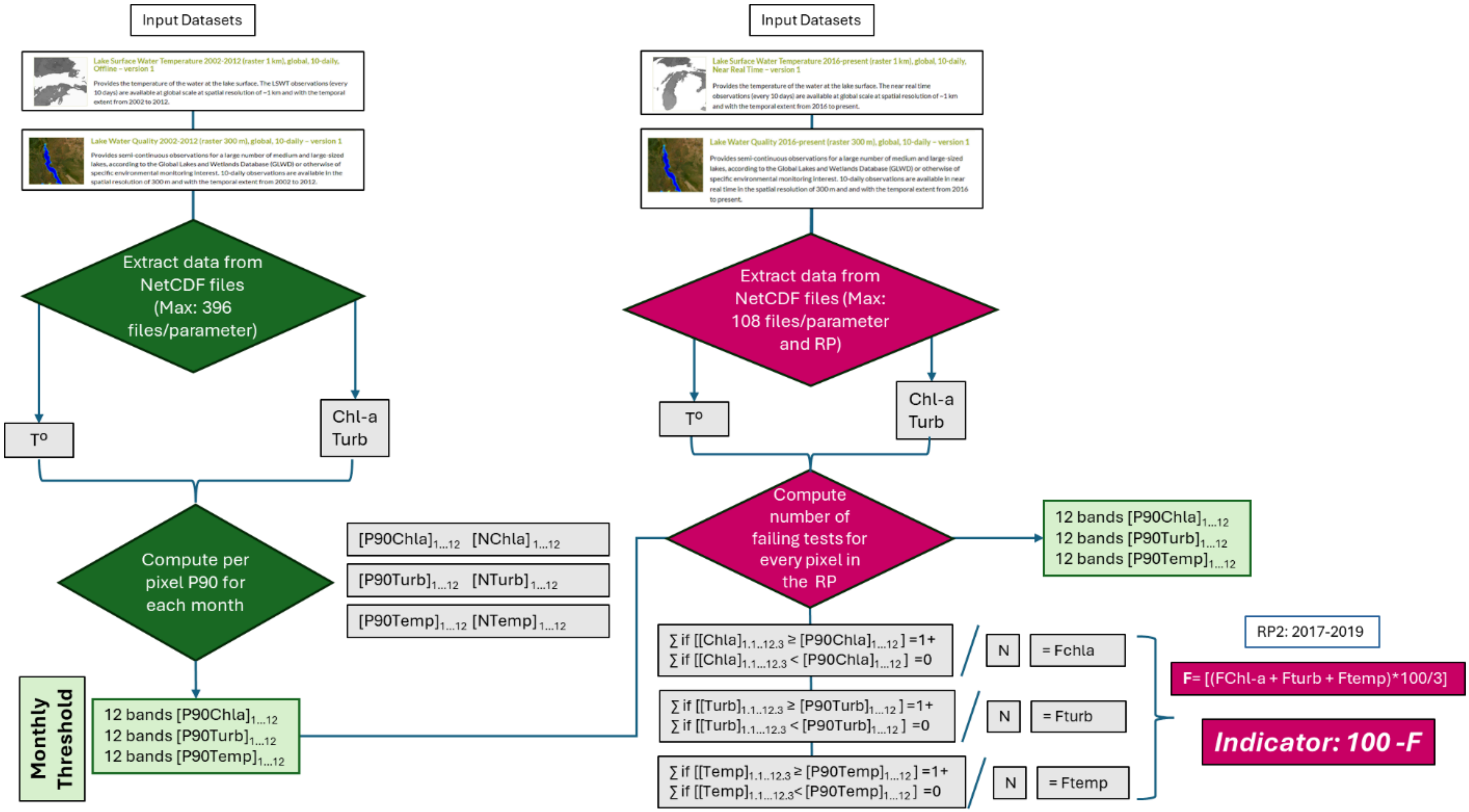


Figure 1: Workflow design for the computation of the EO Indicator for SDG 6.3.2.

3.1.1.- Definition of baseline and threshold values.

Aiming to develop a stand-alone indicator based on EO data, independent of the availability of data for any specific waterbody, we decided to set a threshold value for the 3 parameters based on historical data. These EO data should meet the requirements of being fully comparable and provide a time series of at least 10 years.

As it was described in Del 2.1, the CLMS LWQ (early release version *Calimnos v2.1 L3 test dataset)* and CLMS LSWT products offered this possibility (Table 1).

	Parameters	**Spatial Resolution**	**Temporal resolution**	**Historical dataset**	**SDG 6.3.2 RP**
CLMS LWQ V.2	Chl a, Turbidity	300 m.	10 daily aggregates	2002-2012	2017-2019 2020-2022
CLMS LSWT	Surface Temperature	1 km	10 daily aggregates	2002-2012	2017-2019

Table 1: Specifications of the EO data used for the development of the EO indicator for SDG 6.3.2.

The thresholds for the 3 parameters were calculated as a dynamic value, to account for seasonality in the data and better reflect the deviations from the baseline. Threshold values were calculated for each pixel in the lake, accounting for spatial heterogeneity across the lake. The **value of 90th percentile was chosen as the threshold** for this exercise as it has been used for Chlorophyll-a to evaluate the ecological quality status according to the Water Framework Directive in the EU in transitional environments ((e.g. Brito et al., 2012) and considered as extreme event threshold for Chl-a and Temperature in inland lakes in global exercises (Woolway, et al., 2021). In simple terms, the 90th percentile describes the value in a distribution under which 90 percent of the values occur.

Chlorophyll-a 90th percentile (P90Chla), Turbidity (P90Turb) and Temperature (P90Temp) 90th percentile were calculated using continuous remote sensing datasets over 11 years of 10 daily aggregates data. The thresholds were conceived as dynamic thresholds to account for the seasonality. They were calculated separately for every month and every pixel (Figure 1).

3.1.2 .- Indicator calculations.

The Frequency term of the indicator is defined as per parameter:

$$F = \frac{Number\ of\ data\ above\ the\ monthly\ P90\ threshold}{Total\ number\ of\ Valid\ data\ (N)}$$

This term gives a number between 0 and 1 for every parameter (and per pixel) indicating the proportion of data that surpassed the 90th percentile threshold over the Reporting period related to the total amount of available data.

The final formulation of the composite Indicator is:

$$EO\ Indicator\ SDG\ 632 = 100 - \left[\left(\frac{Fchla+FTurb+FTemp}{3}\right)x\ 100\right]$$

This computation gives a final result per pixel and RP that ranges from 0 to 100, being 0 the lower water ambient quality and 100 the highest. The result can be interpreted as a class value divided into **5 descriptive categories**, allowing the analysis of the spatial distribution of the ambient water quality:

Value	Description
90-100	Conditions better or close to baseline state
80-90	Conditions rarely depart from baseline state
60-80	Conditions sometimes depart from baseline state
40-60	Conditions often depart from baseline state
0-40	Conditions usually depart from baseline state

Table 2: Value of the indicator categories and description of the interpretation.

A **binary outcome of good or bad** water quality, as found in the Level 1 methodology, can also be derived for the EO indicator. In the EO indicator, pixels will be considered the same as single measurements, meaning that if more than 80% of the pixel values meet the target value (defined by the P90 threshold), then the water bodies are considered to be in a good state.

In the case of needing to **report a single value per lake**, the median value of all pixels available for the lake will be computed. This result can be interpreted following the categories description or the binary good/bad denomination.

3.2. Links to SDG 6.3.2 Level 1 indicator

SDG 6.3.2 can be reported in 2 levels (Figure 2); Level 1 reporting covers the core pressures on water quality that can be compared globally, while Level 2 offers more flexibility, providing the opportunity to include information that relates to pressures of national or sub-national relevance (UNEP. 2023). The EO Indicator developed in this project can be used for Level 2.

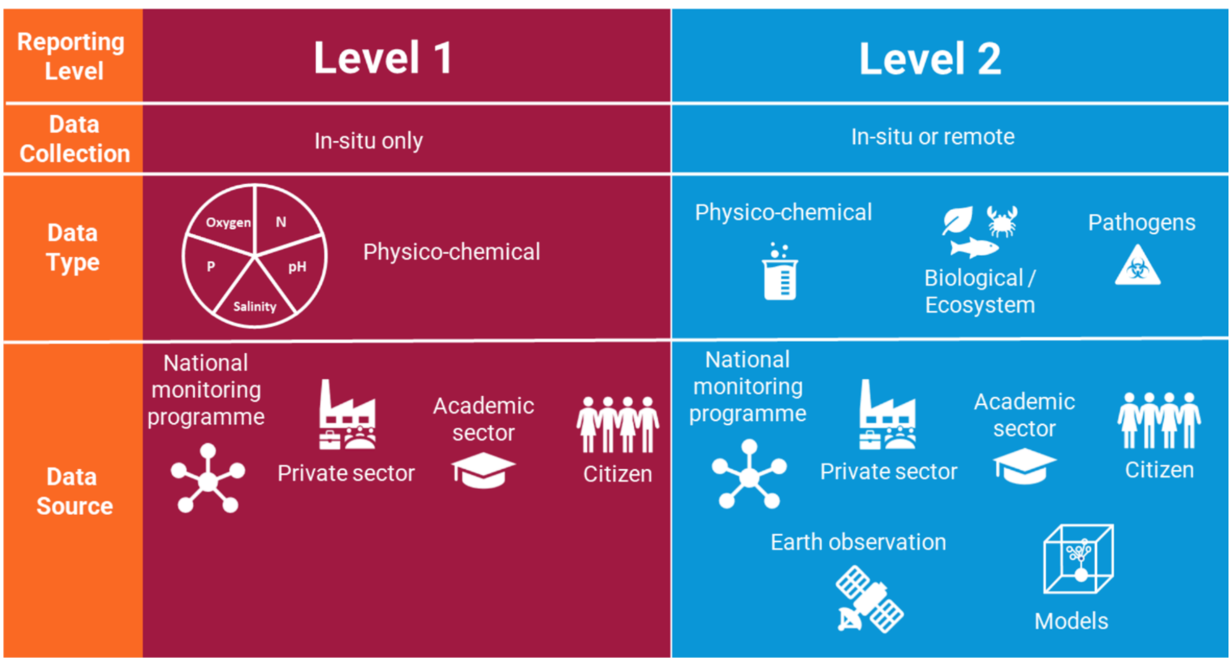


Figure 2.- Schematic of similarities and differences between mandatory Level 1 and optional Level 2 reporting in terms of data collection, data type and data source that can be used (UNEP, 2023)

3.3.1 .- Target-based approach.

As it is described in the methodology of SDG 6.3.2 (UNEP, 2023), the indicator 6.3.2 uses a target-based approach to classify water quality. This means that the measured values are compared with numerical values that represent *"good ambient* *water quality"*. These targets may be either water quality standards that are defined by national legislation, derived from knowledge of the natural world or derived from a calculation of the baseline status of water bodies.

In the EO indicator target values are derived from the P90 threshold calculated from historical data. The idea is to define the baseline status maximum value that should not be frequently surpassed. Alternatively, however, users are able to use their national target values, as is an option in Level 1 reporting.

3.3.2 .- Binary classification.

The final outcome of SDG 6.3.2 Level 1 reporting is a binary classification, i. e. the bodies of water are defined as having either Good or Bad ambient water quality. To classify if a body of water is of "good ambient water quality", a threshold is applied where 80 per cent or more of monitoring values must meet their targets. This target has also been applied in the EO indicator, as it was explained in the previous section.

This approach tries to make Level 1 and Level 2 as complementary and comparable as possible.

4. Results of the EO Indicator for Lake Tanganyika

In the following sections we will explain briefly the results of the work performed to obtain the final evaluation of Lake Tanganyika for the indicator SDG 6.3.2. We will both describe the results of the dynamic per pixel threshold and the final results of the evaluation obtained for the Reporting Period 2 (2017-2019).

We will show how the spatially explicit component provided by EO data adds value in determining whether water quality management measures are contributing to the improvement, maintenance or decline of water quality in inland water bodies.

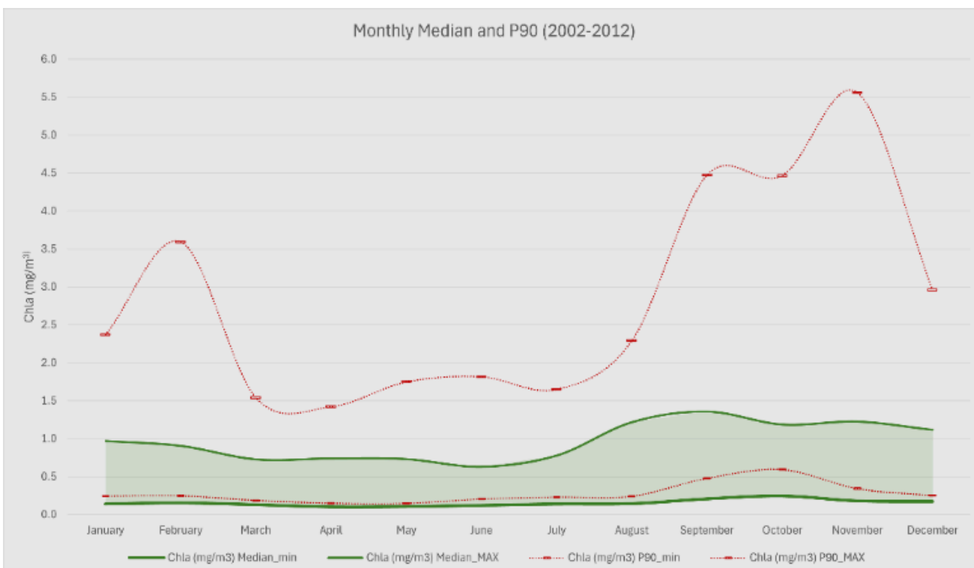
4.1.- Seasonal per-pixel threshold values

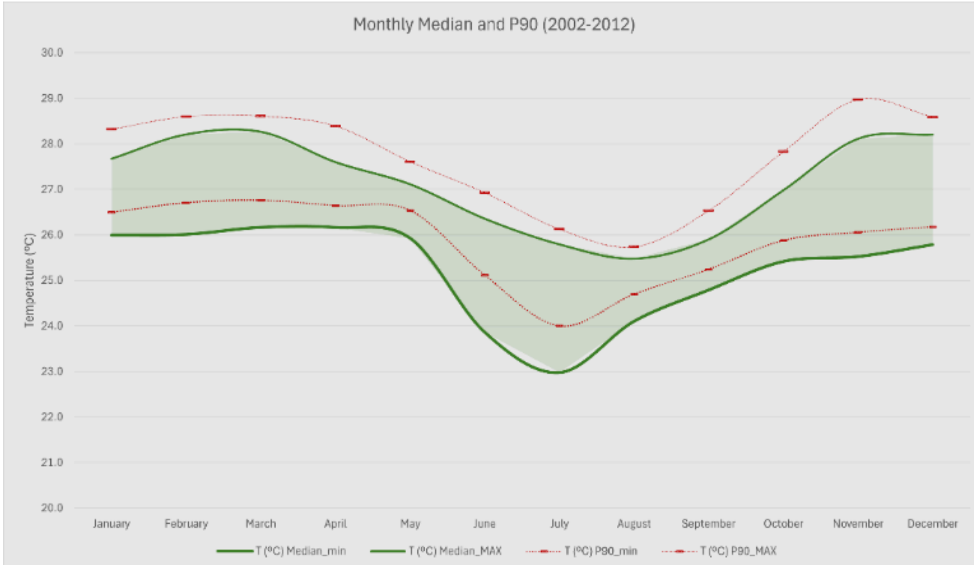
The first step in the formulation of the EO Indicator for SDG 6.3.2 was the computation of the dynamic threshold, defined as a pixel-level monthly value per parameter corresponding to the 90th percentile of every month using 11 years of data (2002-2012). We also computed the median values as descriptive statistics.

The chlorophyll-a monthly median per pixel showed low values the whole year, with a maximum monthly median value of 1.35 mg/m^3 corresponding to September. In general, the lake shows lower values from March to July (Figure 3).

The Temperature median and dynamic threshold also show a clear and coincident seasonal pattern with lower temperatures from May to September (Figure 3).

Turbidity is usually low in the lake, with monthly median values always below 3 NTU and a seasonality showing a slight and continuous rise starting with the wet season (October) and reaching the maximum in February, with parts of the lake showing very turbid water (Max P90th: 31.8 NTU) (Figure 3).





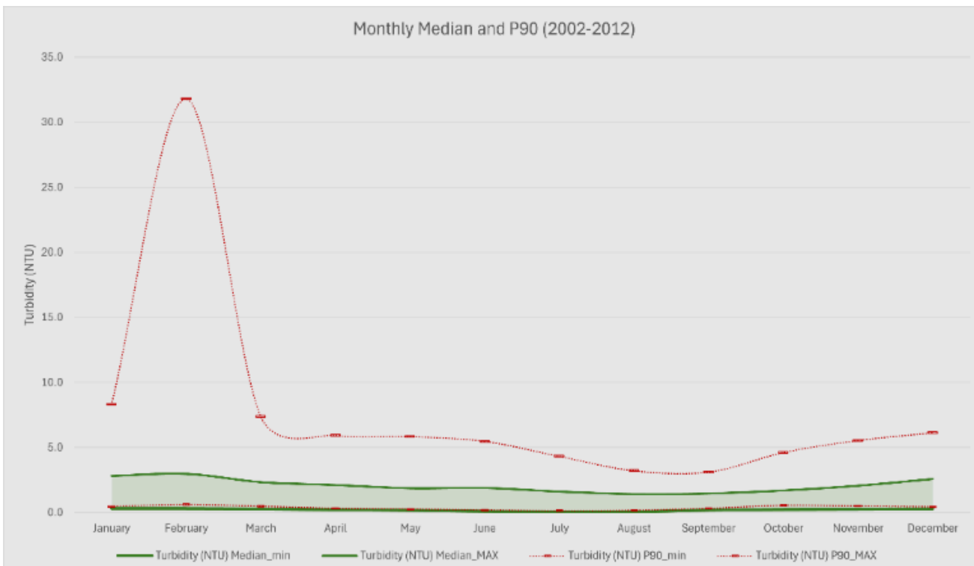
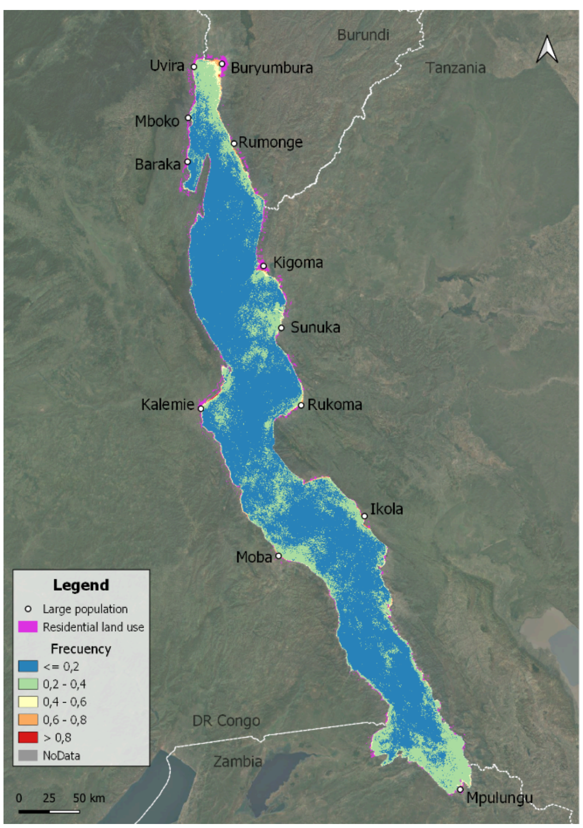


Figure 3: Monthly median (green shadow), maximum and minimum 90th percentile pixel value in Lake Tanganyika for the period 2002 – 2012. From top to bottom: Chlorophyll-a, Temperature and Turbidity.

4.3.- Results for the Reporting Period 2 (2017-2019) in Lake Tanganyika

4.3.1.-Chlorophyll a

Fchla is the frequency of observations for a given pixel, which exceed the baseline ChlaP90 threshold value, where higher values indicate lower water quality. *Fchla* for RP2 showed that, for most of the lake, around 20% of the available data were above the ChlaP90 threshold value and therefore exceeding the target water quality for chl-a (Figure 4.). The lake's overall median *Fchla* was 0.16, indicating a tendency toward a steady state compared to the baseline period. The southern and northern extremities of the lake had a greater *Fchla*, which also correspond to areas with greater population concentration. Higher *Fchla* were also found in the east and west coasts surrounding the cities of Buryumbura (Burundi), Kigoma, Sunuka, Ikola (Tanzania), Kalemie, Moba (DRC).



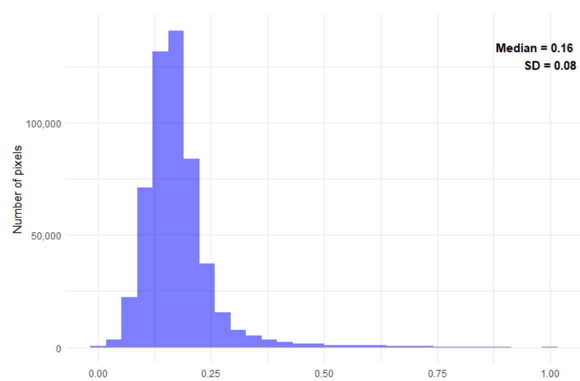


Figure 4: *Fchla* (frequency that chl-a was above the baseline P90 value) for Lake Tanganyika during RP2 (2017-2019)

4.3.2.- Turbidity

Fturb is the frequency of observations for a given pixel, which exceed the baseline TurbP90 threshold value, where higher values indicate lower water quality. *Fturb* results for RP2 showed that, for most of the lake, between 20% to 60% of the available data exceeded the baseline-derived Turb90 threshold value (Figure 4.). The lake's overall median *Fturb* was 0.43, indicating that the lake was regularly more turbid than the baseline period. The spatial pattern for turbidity showed a higher increase in turbidity in the southern part of the lake, similarly to what was described for chlorophyll-a, coinciding with shoreline areas of high population density, most notable in Rumonge (Burundi), Kalemie and Moba (DRC).

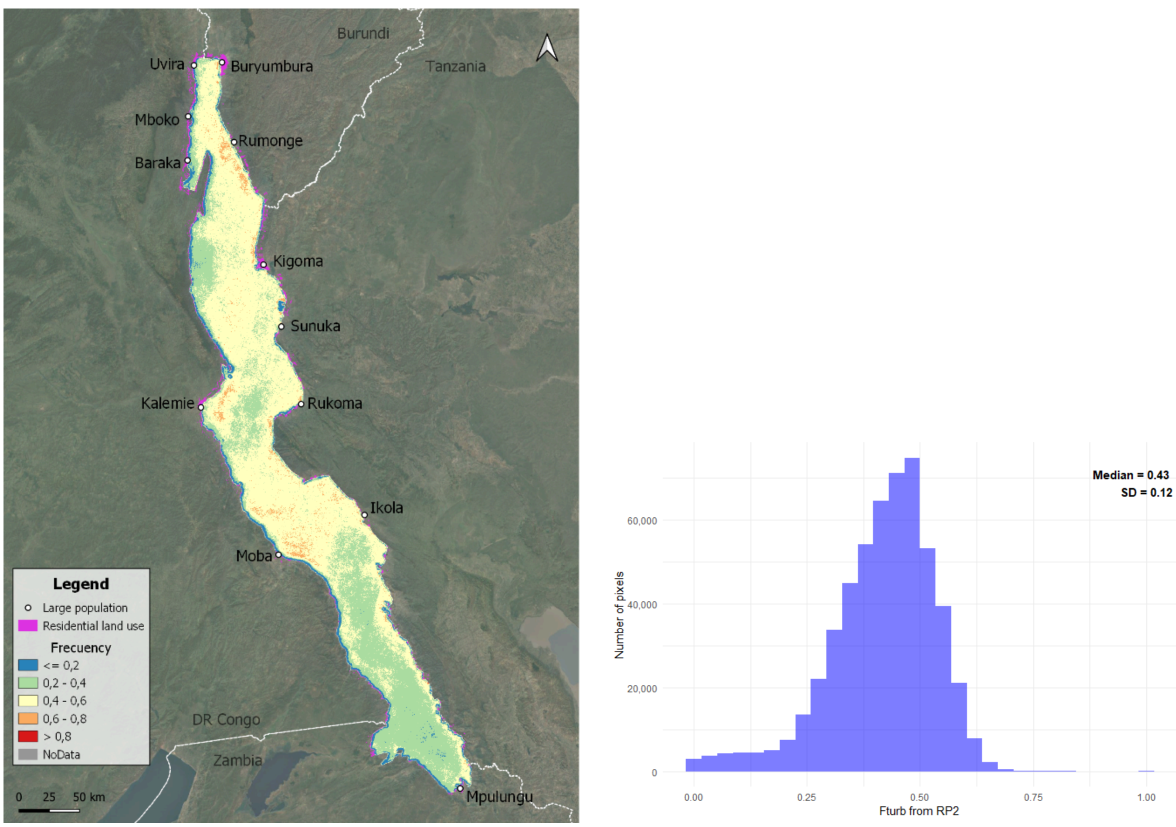


Figure 5: *Fturb* values for Lake Tanganyika during RP2 (2017-2019)

4.3.3.- Temperature

Ftemp is the frequency of observations for a given pixel, which exceed the baseline TempP90 threshold value, where higher values indicate lower water quality. The results for RP2 show a clear spatial pattern, where the Northern part of the lake has higher values (greater increase in water temperature) than the Southern part of the lake (stable or smaller increase in water temperatures compared to baseline). The lake's overall median *Ftemp* was 0.37, indicating that the lake surface water was regularly warmer than the baseline period. However, Ftemp values for the northern part of the lake reached over 0.60, indicating that around 60% of available data surpassed the baseline Temp90 threshold, while the southern part of the lake was typically under 20%.

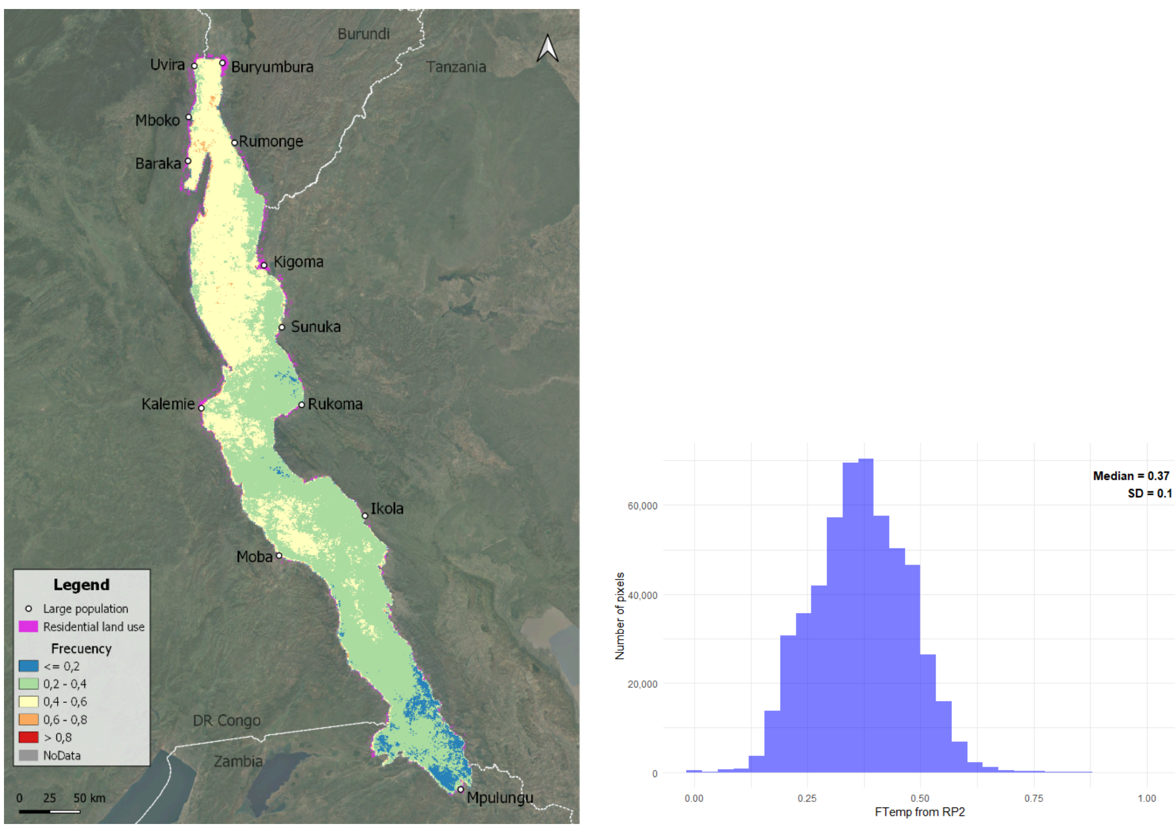


Figure 6: *Ftemp* values for Lake Tanganyika during RP2 (2017-2019)

4.3.4.- EO Indicator results for Lake Tanganyika

The EO indicator (Section 3.1.2) is the frequency of water quality observations (chl-a, turbidity and water quality in this case) that meet their target values. The EO indicator is normalised for a scale between 0 and 100, where higher values indicate better water quality (see categories in Table 2).

The EO indicator results for Lake Tanganyika show that the Southern tip of the lake, roughly below Moba, were relatively stable with the baseline conditions, with values around 70 (green) to 80 (blue) (Figure 7). In contrast, the northern majority of the lake recorded typically lower water quality and a greater departure from the baseline conditions, with values ranging from 70 (yellow) to less than 50 (red).

Proximity to known populated areas was noted as a likely factor on water quality, since the parts of the lake in worst quality class were those closest to the populated areas. The part of the lake near Buryumbura has particularly poor water quality in RP2.

To provide a single water quality assessment for the whole waterbody, the median of all the values was calculated as 67.9. The translation of this number into the binary classification, it would be reported that Lake Tanganyika had “not good” ambient water quality during RP2. Evaluating the proportion of the lake in each class we can see that most of the lake is in the conditions that sometimes depart from the baseline (Figure 8).

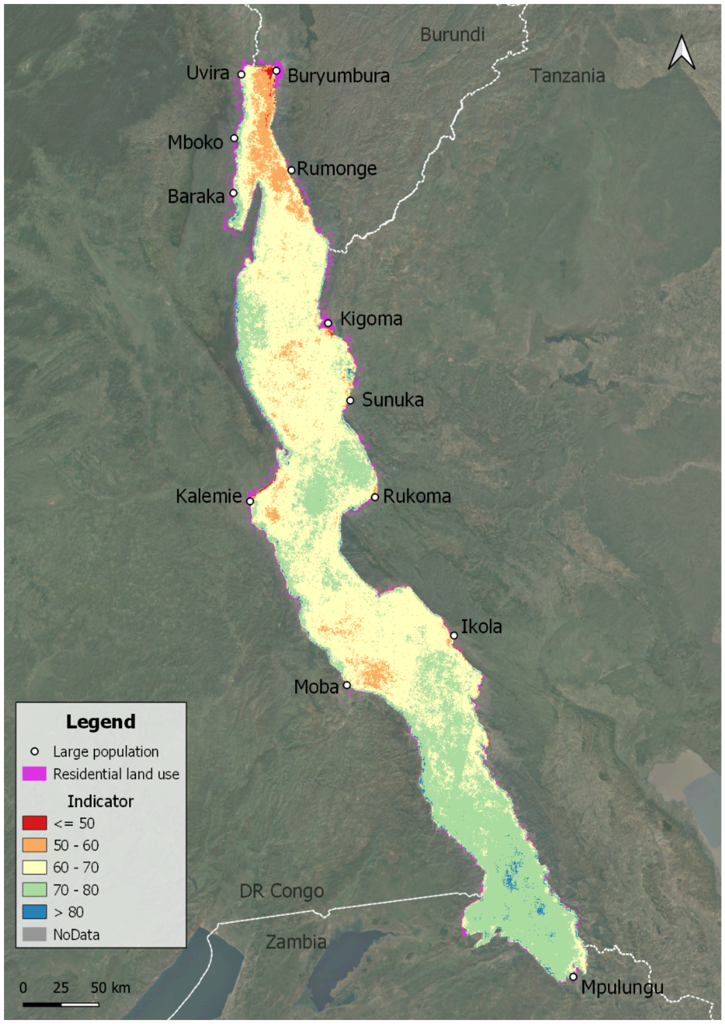
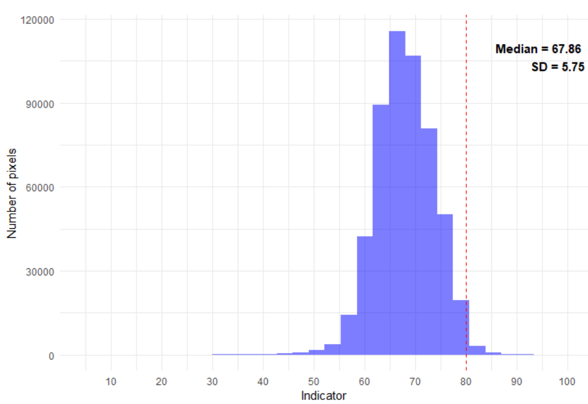


Figure 7: EO Indicator Value for Lake Tanganyika during RP2 (2017-2019)



Value	Description	Proportion of lake
90 - 100	Conditions exceed or very close to the baseline state.	<1%
80 - 90	Conditions rarely depart from the baseline state.	1%
60 - 80	Conditions sometimes depart from the baseline state.	92%
40 - 60	Conditions often depart from the baseline data.	7%
0 - 40	Conditions usually depart from the baseline state.	<1%

Figure 8: EO Indicator Value for Lake Tanganyika during RP2 (2017-2019): Frequency diagram and categories for the Lake.

5. Conclusions

The EO Indicator presented in this project has shown the added value of EO data for the reporting of SDGs.

The current formulation allows for the required flexibility and future improvements as it can be designed with more parameters, if these are available, and also allows the establishment of National thresholds for the parameters.

It has also shown the added value of EO data for management purposes, allowing to find spatial patterns related to the population density, which can be useful in the case of organising coordinated efforts for improving ambient water quality.

We have developed a stand-alone EO Indicator that enables the evaluation of ambient water quality for a waterbody without requiring any other source of data. This, together with the fact that we are using operational and global products from the Copernicus Land Monitoring Service, opens the door for the evaluation and reporting of SDG 6.3.2 at a global scale, with no need of national data availability. An EO-indicator of this nature could enable low income countries to report more widely on SDG 6.3.2, and have information about the ambient water quality of their water bodies for management decision making.

6. Next steps

The work presented in this project, conducted by members of the World Water Quality Alliance Earth Observation workstream during a 5-month UNEP seed-funded initiative, marks the initial phase of developing the EO indicator. Further steps are planned to expand its applicability for global use, with several key actions already identified for the next stage of development. These include working closely with the SDG focal points of the riparian countries in Lake Tanganyika to qualitatively and quantitatively validate the indicator and including a measure of uncertainty that takes into account the amount of data per pixel and the compounding uncertainty of input water quality parameters.

7. Bibliography

Brito, A.C., Vanda Brotas, Miguel Caetano, Teresa P. Coutinho, Adriano A. Bordalo, John Icely, João M. Neto, João Serôdio, Teresa Moita (2012). Defining phytoplankton class boundaries in Portuguese transitional waters: An evaluation of the ecological quality status according to the Water Framework Directive, Ecological Indicators, Volume 19,5-14. https://doi.org/10.1016/j.ecolind.2011.07.025.

Canadian Council of Ministers of the Environment. 2001. Canadian water quality guidelines for the protection of aquatic life: CCME Water Quality Index 1.0, Technical Report. In: Canadian environmental quality guidelines, 1999, Canadian Council of Ministers of the Environment, Winnipeg.

Canadian Council of Ministers of the Environment. 2017. Canadian water quality guidelines for the protection of aquatic life: CCME Water Quality Index, User's Manual – 2017 Update. In: Canadian environmental quality guidelines, 1999, Canadian Council of Ministers of the Environment, Winnipeg.

Pahlevan, N; Dekker, A. Greb, S. (2022). Earth Observation in Support of SDG6.3.2/6.6.1: Reporting Surface Water Quality. In Earth Observation Applications and Global Policy Frameworks. https://earth-space-sci.pubpub.org/pub/dqh01p03.

UNEP (2023). An introduction to SDG indicator 6.3.2: Proportion of bodies of water with good ambient water quality. SDG632_Introduction to SDG Indicator 6.3.2_ Version_20230420.

Woolway, R. I., Kraemer, B.M., Zscheischler, J., & Albergel, C. (2021) Environ. Res. Lett. 16 124066 DOI 10.1088/1748-9326/ac3d5a.

[end of document]