

Faculty of Bioengineering

Identification of the drivers of water quality spatio-temporal variability in the estuarine areas of Lake Tanganyika

Author :	Aurélien Lengrand
Supervisors :	Marnik Vanclooster Alice Alonso
Reviewers :	Pierre-Denis Plisnier Julien Radoux
Academic Year :	2023 - 2024
Bioengineer :	Environmental science and technology

Acknowledgements

I would like to express my sincere gratitude to Professor Marnik Vanclooster and Dr. Alice Alonso for their invaluable guidance and support throughout this project. Their expertise, encouragement, and insightful feedback have been instrumental in shaping this research. In particular, I am deeply grateful to Dr. Alonso for her guidance in navigating the new world of research and for dedicating her time to mentorship.

I am also grateful to PhD student François Toussaint for the time he dedicated to me and for the advice he provided.

I extend my heartfelt thanks to my family, particularly my mother and sister, for their unwavering support, valuable advice, and endless encouragement.

A special acknowledgment goes to Thomas Lenfant for his collaboration and support throughout this five-year adventure.

Lastly, I would like to express my gratitude to Chloé Streel for her support during moments of doubt and the countless sleepless nights spent working. Your encouragement and belief in me have been invaluable, especially during challenging times.

Contents

1 Introduction	1
2 State of the art	3
2.1 Optical Remote sensing	3
2.1.1 Reflectance and spectral signature	3
2.1.2 Sensors characteristics	3
2.2 Water quality and Remote Sensing	4
2.2.1 Water quality variables	4
2.2.1.1 Turbidity	5
2.2.1.2 Chlorophyll	5
2.2.2 Optical characteristics of water bodies	8
2.2.3 Atmospheric correction for aquatic environment	9
2.3 Water Quality Parameters Retrieval Modes	10
2.3.1 Turbidity Algorithms	11
2.3.2 Chlorophyll-A Algorithms	12
2.4 Lake Tanganyika	14
2.4.1 Limnological cycle of Lake Tanganyika	14
2.4.2 Chlorophyll-A based Lake segmentation	17
2.4.3 Impact of the sediments brought by Ruzizi river	18
3 Materials and methods	20
3.1 Overall view	20
3.2 Data used	21
3.2.1 Landsat 8-9 Provisional Aquatic Reflectance	21
3.2.1.1 Landsat Provisional Aquatic Reflectance algorithm	21
3.2.1.2 Performance of the Landsat Provisional Aquatic Reflectance	23
3.2.1.3 Characteristics of Landsat 8/9 Provisional Aquatic Reflectance	23
3.2.2 CCI Lakes	24
3.2.3 Flow of the Ruzizi and Malagarasi rivers	27
3.3 Region of interest	28
3.3.1 Bujumbura	28
3.3.2 Kigoma	29
3.4 Downscaling approach	31
3.4.1 Data harmonization	31
3.4.1.1 Matching dates	31
3.4.1.2 Reprojection	31
3.4.1.3 Resampling	32
3.4.1.4 Daily mosaic	32
3.4.2 Chlorophyll-A and Turbidity Algorithms selection	32
3.4.2.1 Turbidity	33
3.4.2.2 Chlorophyll-A	33

3.4.3	Linear regression method	33
3.5	Semi-Empirical algorithm and Analytical index	35
3.5.1	OC3 algorithm for Landsat 8	35
3.5.2	Normalized Difference Turbidity Index (NDTI) and Total Suspended Matter index (TSMi)	35
3.6	Time series analysis	36
3.6.1	Empirical Orthogonal Function (EOF)	36
3.6.2	Kmean clustering	36
3.6.3	Correlation analysis between ChloA and Turbidity dynamic in each cluster and the river flow	37
4	Results	38
4.1	Downscaling approach	38
4.1.1	Turbidity algorithms optimization	38
4.1.2	Chlorophyll-A algorithms optimization	40
4.1.3	Investigating the downscaling failure	42
4.1.3.1	Band comparison	43
4.1.3.2	Impact of the band used on the turbidity computed	45
4.2	Semi-Empirical algorithm and Analytical indexes	46
4.3	Chlorophyll-A and Turbidity Time Series analysis	48
4.3.1	Global dynamic of the datasets	48
4.3.1.1	Bujumbura	48
4.3.1.2	Kigoma	50
4.3.2	Empirical Orthogonal Function	52
4.3.2.1	Bujumbura region	52
4.3.2.2	Kigoma region	54
4.3.3	Kmean clustering	55
4.3.3.1	Bujumbura	56
4.3.3.2	Kigoma	57
4.3.4	Correlation analysis between Chlorophyll-A and Turbidity daily median in each cluster and with the river flow	57
4.3.4.1	Bujumbura	58
4.3.4.2	Kigoma	59
4.3.5	Correlation analysis between Chlorophyll-A and Turbidity monthly median in each cluster and with the river flow	60
4.3.5.1	Bujumbura	60
4.3.5.2	Kigoma	62
5	Discussion	64
5.1	Investigating the downscaling failure	64
5.1.1	Band Comparison	64
5.1.2	Turbidity Algorithms Optimization	64

5.1.3 Chlorophyll-A Algorithms Optimization	65
5.2 Semi-Empirical Algorithm and Analytical Indexes	65
5.3 Chlorophyll-a and Turbidity time series analysis	66
5.3.1 Global dynamic of the datasets	66
5.3.2 Empirical Orthogonal Function	67
5.3.3 Kmean clustering	67
5.3.4 Correlation analysis between Chlorophyll-A and Turbidity daily and monthly median in each cluster and with the river flow	68
5.4 Future prospect	69
6 Conclusion	70
A Appendix	81

List of Figures

1	Reflectance versus wavelength is depicted here, showcasing the impact of suspended sediment concentration on Reflectance. The data was gathered from a boat-mounted high-resolution spectroradiometer, positioned 50 cm above the lake surface (Ritchie, 1976).	5
2	Absorption spectrum of chlorophyll A and B (Panawala, 2017).	6
3	Spectral signature of the water components (clear water in blue, water with chlorophyll content in green, water with CDOM in black, and sedimented water in orange) (Palacios, 2016).	7
4	Spectral signature of water with constant chlorophyll-A concentration (57µg/l) and varying suspended sediment concentrations (SESTON), illustrating the relative contributions of chlorophyll and suspended sediment to the reflectance spectra of surface water. Data based on <i>in situ</i> laboratory measurements taken 1 meter above the water surface. Source: (Schalles et al., 1997).	7
5	A schematic diagram of the various processes that contribute to the signal as measured by a remote sensor in optically shallow water where the substrate has a significant effect on the water leaving radiance at the surface (A. G. Dekker et al., 2001).	8
6	Contributions to the total upwelling radiance above the sea surface, Lu. Yellow arrows are the Sun's unscattered beam; blue arrows represent lights that are transmitted through the surface into the ocean L_t ; orange arrows are atmospheric path radiance L_a ; red is surface-reflected radiance L_r ; green is water-leaving radiance L_w . Thick arrows represent single-scattering contributions; thin arrows illustrate multiple-scattering contributions (Mobley, 2022).	9
7	Cross section of a thermally stratified lake showing the location of the epilimnion and hypolimnion and associated temperature changes (Adapted from Brusseau et al., 2019).	14
8	Main events in the limnology of Lake Tanganyika during a yearly cycle (Plisnier et al., 1996).	15
9	Schematic profile of water layers during an annual cycle. States 2 and 3 alternate before reaching stage 4. Internal waves are not indicated. (Plisnier et al., 1996).	17
10	Regionalisation of Lake Tanganyika in 13 subregions (R1–R13) of co-varying chlorophyll-A concentrations with main rivers, Malagarasi (A), Ruzizi (B) and Lukuga (C) (Bergamino et al., 2010).	17
11	Mean Secchi depth and pulled results for suspended solids and chlorophyll-A along the transect of the Ruzizi delta (Vandelannoote et al., 1999).	18
12	Methodology flowchart	20
13	Study Area map	28
14	ROI Ruzizi map	29

15	Mean annual fluctuations in the flow rate of River Rusizi at Gatumba and in rainfall at Bujumbura airport with standard deviation (unpublished IGEBU data from 1982 to 1989) (Vandelannoote et al., 1999).	29
16	ROI Malagarasi map	30
17	Discharge of the Malagarasi river in 1975-80 (Huttula, 1997).	30
18	Monthly rainfall from 1998 to 2012 (data were obtained from the Tanzania Meteorological Agency in Kigoma (Gao et al., 2015)).	31
19	Median map of Turbidity from CCI Lakes between 2016 and 2020	38
20	Linear regression analysis between CCI Lakes Turbidity and the turbidity obtained with Landsat PAR and the algorithms optimized (cf Table 13). Individual points are color-coded based on the number of neighboring points, with the red line representing the identity line. The different algorithms are (a) NDTI, (b) TSMi, (c) first polynomial of B4, (d) second polynomial of Band 4, (e) third polynomial of Band 4, (f) $B4^x$ and (g) $(B3 + B4) \frac{B2}{3}$.	40
21	Median map of ChloA from CCI Lakes between 2016 and 2020	41
22	Linear regression analysis between CCI Lakes ChloA and the ChloA obtained with Landsat PAR and the algorithms optimized (cf Table 14). Individual points are color-coded based on the number of neighboring points, with the red line representing the identity line. The different algorithms are : (a) OC2 equation optimized, (b) OC3 equation optimized and (c) OC3 equation with coefficients fitted by J. E. O'Reilly and Werdell, 2019).	42
23	Comparison of the median R_{rs} conducted between CCI Lakes and Landsat PAR (1 km) for four different wavelengths between 2016 and 2020 : Ultra-Blue corresponds to CCI Lakes $R_{rs}(443)$ and Landsat PAR $R_{rs}(443)$, Blue corresponds to CCI Lakes $R_{rs}(490)$ and Landsat PAR $R_{rs}(482)$, Green corresponds to CCI Lakes $R_{rs}(560)$ and Landsat PAR $R_{rs}(561.5)$ and Red corresponds to CCI Lakes $R_{rs}(665)$ and Landsat PAR $R_{rs}(654.5)$.	44
24	Linear regression analysis between CCI Lakes and Landsat PAR for different R_{rs} . Individual points are color-coded based on the number of neighboring points, with the red line representing the identity line. The bands presented are (a) CCI Lakes $R_{rs}(443)$ and Landsat PAR $R_{rs}(443)$, (b) CCI Lakes $R_{rs}(490)$ and Landsat PAR $R_{rs}(482)$, (c) CCI Lakes $R_{rs}(560)$ and Landsat PAR $R_{rs}(561.5)$ and (d) CCI Lakes $R_{rs}(665)$ and Landsat PAR $R_{rs}(654.5)$.	45
25	Representation of the different CCI Lakes bands acquired the 10/05/2020. (a) R_{rs} (665 nm), (b) R_{rs} (754 nm), (c) Turbidity concentration (d) TSMi retrieved using CCI Lakes $R_{rs}(560nm)$ and $R_{rs}(665nm)$.	46
26	Median maps of the NDTI and TSMi time series using Landsat PAR between 2016 and 2020	47
27	Median map of the ChloA Time Serie using Landsat PAR between 2016 to 2020	48
28	Monthly median maps of the Bujumbura region between 2013 and 2023 for TSMi	49
29	Monthly median maps of the Bujumbura region between 2013 and 2023 for ChloA	50
30	Monthly median maps of the Kigoma region between 2013 and 2023 for TSMi	51

31	Monthly median maps of the Kigoma region between 2013 and 2023 for ChloA .	52
32	Spatial dimension of the first component of the EOF for the Bujumbura region for (a) Turbidity and (b) Chlorophyll-A	53
33	Monthly median, 25th and 75th percentiles of the temporal dimension of the first component of the EOF for the Bujumbura region for (a) Turbidity and (b) Chlorophyll-A	54
34	Spatial dimension of the first component of the EOF for the Bujumbura region for (a) Turbidity and (b) Chlorophyll-a	55
35	Monthly median, 25th and 75th percentiles of the temporal dimension of the first component of the EOF for the Kigoma region for (a) Turbidity and (b) Chlorophyll-a	55
36	Clustering of the Bujumbura region based on the ChloA and Turbidity dynamics	56
37	Clustering of the Kigoma region based on the ChloA and Turbidity dynamics . .	57
38	Daily median, 5th and 95th percentiles of TSMi and ChloA between 2013 and 2020 for the Coast (red) and Intermediate (blue) cluster of the region of Bujum- bura compared with the Ruzizi river flow derived from Wongchuig et al., 2023 between 2013 and 2020.	59
39	Daily median, 5th and 95th percentiles of TSMi and ChloA between 2013 and 2020 for the Coast (red) and Intermediate (blue) cluster of the region of Kigoma compared with the Malagarasi river flow derived from Wongchuig et al., 2023 between 2013 and 2020.	60
40	Monthly median of TSMi and ChloA between 2013 and 2023 for the Coast and Intermediate cluster of the region of Bujumbura compared with the Ruzizi river flow derived from Wongchuig et al., 2023 between 2013 and 2020.	62
41	Monthly median of TSMi and ChloA between 2013 and 2023 for the Coast and Intermediate cluster of the region of Kigoma compared with the Malagarasi river flow derived from Wongchuig et al., 2023 between 2013 and 2020.	63

List of Tables

1	Inversion models of TSM. In these models, C_{TSM} denotes the total suspended matter concentration, $R_{rs}(\lambda)$ represents the remote sensing reflectance at wavelength λ	11
2	Inversion models of ChloA. C_{ChloA} represents the concentration of chlorophyll-A, and $R_{rs}(\lambda)$ denotes the R_{rs} at wavelength λ	12
3	Description of Aquatic Reflectance Bands (Sayler, 2023)	24
4	Landsat Provisional AR WATER MASK Band Interpretations (“Landsat Provisional Aquatic Reflectance ADD U.S. Geological Survey”, 2020)	24
5	Instrument Satellite Dates used (dd/mm/yyyy) and Typical local time of observation (Carrea et al., 2023)	25
6	CCI Lakes variables used (“Lakes CCI project Key Documents”, n.d.)	25
7	Dominant characteristics of OWTs in inland waters (Spyrakos et al., 2018)	26
8	Chlorophyll-A algorithms per sensor and Optical Water Type (OWT) (Carrea et al., 2023)	26
9	Turbidity algorithms per sensor and Optical Water Type (OWT)(Carrea et al., 2023)	27
10	Turbidity algorithms tested	33
11	Chlorophyll-A algorithms tested	33
12	Coefficients of the OC3 algorithm for Landsat 8 (J. E. O’Reilly and Werdell, 2019)	35
13	Optimization results of Turbidity algorithms described in Table 10 between CCI Lakes and Landsat PAR resampled at 1km. B2 represents R_{rs} of Landsat PAR at 490nm, B3 at 560 nm, and B4 at 665nm.	39
14	Optimization results of Chlorophyll-A algorithms described in Table 11 between CCI Lakes and Landsat PAR resampled at 1km.	42
15	Linear regression results of Landsat PAR R_{rs} resampled at 1km and CCI Lakes R_{rs} for four bands: Ultra-Blue, Blue, Green and Red	45
16	Percentage of variance explained by each component of the Empirical Orthogonal Function for the Bujumbura region	53
17	Percentage of variance explained by each component of the Empirical Orthogonal Function for the Kigoma region	54

List of Acronyms

- CCI Lakes - Lakes Climate Change Initiative project
- CDOM - Colored Dissolved Organic Matter
- ChloA - Chlorophyll-A
- EOF - Empirical Orthogonal Function
- Landsat PAR - Landsat 8/9 Provisional Aquatic Reflectance
- LWLR - Lake Water-Leaving Reflectance
- MAE - Mean Absolute Error
- MODIS - Moderate Resolution Imaging Spectroradiometer
- NAP - Non-Algal Particles
- NASA - National Aeronautics and Space Administration
- NDTI - Normalized Difference Turbidity Index
- NIR - Near-Infrared
- OAC - Optically Active Components
- OBPG - Ocean Biology Processing Group
- OC - Ocean Color
- OLCI - Ocean and Land Color Imager
- OLI - Operational Land Imager
- RMSE - Root Mean Square Error
- ROI - Region of Interest
- R_{rs} - Remote sensing reflectance
- TSM - Total Suspended Matter
- TSMi - Total Suspended Matter Index
- USGS - United States Geological Survey

1 Introduction

The study and monitoring of water quality are paramount, particularly in the Sustainable Development Goals (SDGs) context. These global objectives underscore the need to ensure access to clean water and sanitation for all, which also entails monitoring and preserving the quality of freshwater resources. Among these resources, Lake Tanganyika stands out for its vital role in supporting several life functions for the surrounding populations.

Lake Tanganyika is an expansive rift valley lake stretching approximately 650 kilometers in length and an average width of 50 kilometers. Its impressive depth, averaging 570 meters and reaching a maximum of 1,470 meters, hosts one of the globe's most prolific pelagic fisheries. This thriving aquatic ecosystem serves as a fundamental source of sustenance and economic prosperity for local communities and commercial ventures, highlighting the lake's profound ecological and socioeconomic importance (Mölsä et al., 1999). However, historical and paleolimnological data combined offer compelling evidence indicating that climate change plays a role in reducing productivity in Lake Tanganyika over the last 80 years (C. M. O'Reilly et al., 2003).

Lakes face a continuous interplay of internal pressures and external forces, underscoring the need for consistent scientific data collection to assess trends and spatial dynamics in response to environmental shifts. Without such data, managing these intricate ecosystems becomes reliant on conjecture or semi-scientific methods, which are insufficient for addressing threats effectively (Plisnier et al., 2023).

Deforestation on a large scale, urban expansion, and inappropriate agricultural practices have worsened erosion, resulting in sediment accumulation in the littoral zone, which is a crucial habitat for numerous organisms. Increased turbidity and changes in substrates can alter these habitats, disrupting food chains and primary productivity, and thereby reducing species diversity. Additionally, the supply of phosphorus and nitrates from water runoff, erosion, and leaching of fertilized agricultural land increases the nutrient stock in Lake Tanganyika. This nutrient influx can further degrade water quality and affect the lake's ecological balance (Niyoyitungiye, 2020). These observations highlight the need for monitoring the impact of rivers flowing into the lake on water quality in the estuaries.

For efficient water quality monitoring and environmental analysis, remote sensing data offer substantial advantages over traditional *in-situ* methods. Remote sensing provides superior spatial and temporal resolution, allowing for more precise and frequent observation of hydrological and environmental phenomena. A prior study has already examined Lake Tanganyika using satellite products with 1 km spatial resolution, yielding valuable insights into the lake's overall dynamics (Bergamino et al., 2010). Nevertheless, these datasets face constraints when assessing regions with pronounced spatial heterogeneity, like estuaries and coastal zones. Hence, it is crucial to characterize the spatiotemporal variability of these regions using remote sensing data with fine spatial resolution.

This master thesis will concentrate on examining the spatiotemporal variability of water qual-

ity parameters, specifically chlorophyll-A (ChloA) and turbidity, in the estuaries of the Ruzizi and Malagarasi rivers, the primary tributaries of Lake Tanganyika. However, no datasets offering water quality variables at fine spatial resolution currently exist. This is primarily because higher-resolution satellite remote sensing products such as Landsat and Sentinel-2 are designed for terrestrial studies rather than aquatic ones. This characteristic involves challenges associated with atmospheric correction and the lower signal-to-noise ratio in aquatic environments compared to terrestrial environments. Nevertheless, some studies have successfully utilized higher-resolution products for aquatic studies, demonstrating their potential. Furthermore, the recent release of Landsat 8/9 Provisional Aquatic Reflectance (Landsat PAR), a satellite product specifically tailored for aquatic environment analysis with a spatial resolution of 30 meters, holds promise for overcoming these challenges. Therefore, the primary objective of this study is to construct time series of ChloA and turbidity in the estuarine areas of Lake Tanganyika using Landsat PAR alongside ChloA and turbidity algorithms.

The objectives can be summarized as follows :

Objective 1: Identify and evaluate the reliability of several chlorophyll-A and turbidity algorithms for Landsat PAR.

Objective 2: Characterize the intra and interannual dynamics of chlorophyll-A and turbidity in the estuaries.

Objective 3: Identify the drivers of chlorophyll-A and turbidity in the estuaries.

This master's thesis will begin with an in-depth review of remote sensing technology, exploring its applications in monitoring water quality variables and the algorithms employed for their detection. Additionally, it will examine the limnological cycle of Lake Tanganyika and the impact of the Ruzizi river water in the lake to provide context for the study. The methodology section will outline the approach taken to construct time series of chlorophyll-A and turbidity, as well as the identification of their spatiotemporal variability in the estuaries. Subsequently, the study will present the results obtained through this methodology, followed by a comprehensive discussion of these findings.

2 State of the art

2.1 Optical Remote sensing

Earth remote sensing is primarily defined as using electromagnetic radiation to gather data about the ocean, land, and atmosphere without direct physical contact with the object, surface, or phenomenon being studied (Martin, 2014). This technology has revolutionized several scientific sectors by enabling easier, faster, and more cost-effective field data recording. Satellite data are particularly utilized for studying large or hard-to-access environments

2.1.1 Reflectance and spectral signature

When light interacts with a material, specific wavelengths are selectively absorbed, and others are transmitted through the material. Reflectance, denoting the ratio of the light reflected from a sample to the incident light's intensity, is gauged using reflection spectrophotometers. These instruments, comprising a light source and a prism for wavelength separation, direct the light beam onto the sample. A detector then measures the intensity of reflected light at different wavelengths compared to a reference standard with known reflectance. Consequently, a continuous reflectance spectrum of the sample is generated within the measured wavelength range (Meer, 2001). Each substance on Earth's surface possesses distinct values for spectral reflectance characteristics. Reflectance is closely tied to the object's color and tone in an image. An object's color is determined by the wavelength it reflects. The spectral reflectance of an object is the average over specifically defined wavelengths, aiding in its differentiation from other objects. Reflectance ($R(\lambda)$) values vary with wavelength and terrain features and can be expressed mathematically as:

$$R(\lambda) = \frac{\text{Reflected Radiation Energy}}{\text{Incident Radiation Energy}} \quad (1)$$

This formula encapsulates the variation in reflectance across different wavelengths and terrain attributes (Gupta et al., 2022).

2.1.2 Sensors characteristics

A sensor is a device that detects and records electromagnetic energy. The range of wavelengths to which a sensor is sensitive is termed spectral sensitivity. This property determines the sensor's ability to detect and interpret electromagnetic radiation across different wavelengths. On top of spectral sensitivity, sensors can also be described by three different types of resolution :

- **Spatial resolution:** Spatial resolution refers to a sensor's capability to distinguish and record small adjacent objects within an image. It quantifies the level of detail that can be captured and is often expressed in terms of line pairs per millimeter (lp/mm) for aerial photography. A higher spatial resolution indicates finer detail and better clarity of objects within the image. Several factors influence the spatial resolution of an image, including

the scale factor, optical system quality, the grain structure of the photographic film, the contrast of the original objects, atmospheric scattering effects, and image motion. The atmospheric condition, in particular, is highly variable and can impact resolution significantly, with changes occurring between and even during missions.

- **Radiometric resolution:** The distinction in input levels that a sensor can differentiate is referred to as radiometric resolution. For instance, let's consider a dynamic range of energy spanning from 0.5 to 3 W. If this range is represented by 100 digital number (DN) values, the radiometric resolution would be 25 mW per DN. Conversely, if the same range is represented by 250 DN values, the radiometric resolution improves to 10 mW per DN. This demonstrates how a finer radiometric resolution allows for a more precise representation of energy levels within the dynamic range.
- **Temporal resolution:** Temporal resolution refers to the frequency at which a sensor or system can acquire data over time. It measures the interval between successive observations or measurements of a given location or area. A higher temporal resolution means that data is collected more frequently, providing a more detailed and continuous view of temporal changes in the observed phenomena. Conversely, a lower temporal resolution indicates longer intervals between data acquisitions, resulting in less frequent updates and potentially missing short-term variations or events. Temporal resolution is crucial in monitoring dynamic processes such as vegetation growth, land use changes, and natural disasters, as it influences the ability to capture and understand temporal dynamics over time.

The selection of a sensor for a specific application is crucial as its characteristics, including spectral sensitivity, spatial resolution, radiometric resolution, and temporal resolution, all have a significant impact on the resulting dataset. Therefore, choosing the appropriate sensor tailored to the desired usage is essential for obtaining accurate and meaningful data for analysis (Janssen et al., 2001).

2.2 Water quality and Remote Sensing

Watercolor remote sensing, initially designed for oceans, emerged as a crucial method due to its effectiveness in evaluating extensive aquatic areas. This approach primarily focuses on determining optical water quality variables. In the context of oceans, it serves as one of the few viable methods for assessing such vast expanses of water. The optical properties of ocean water are primarily influenced by phytoplankton and its decomposition by-products. This method proves invaluable in comprehending and monitoring the dynamic optical characteristics associated with marine environments.

2.2.1 Water quality variables

Various factors influence water quality, including suspended sediments, turbidity, chlorophylls, chemicals, colored dissolved organic matter (CDOM), nutrients, pesticides, thermal releases,

and oils. These substances can induce visible and thermal changes in surface waters, altering the energy spectra of reflected solar and emitted thermal radiation, which are detectable through remote sensing techniques. While most chemicals do not directly impact spectral or thermal properties, their effects on water properties can be indirectly inferred from remotely sensed measurements of other water quality parameters (Ritchie and Schiebe, 2000). In this study, we are interested in two water quality variables: turbidity and chlorophylls.

2.2.1.1 Turbidity

Turbidity refers to a decrease in water clarity due to the presence of suspended matter, which absorbs or scatters downwelling light. When suspended particles become visibly apparent, water is considered turbid. Inorganic suspended materials, known as suspenoids or tripton, diminish light penetration, create adsorption and desorption surfaces, and can aggregate with dissolved substances, bacteria, and algae. These adsorbed substances become available to bacteria, algae, and zooplankton ("Encyclopedia of Inland Waters", 2009). Total suspended matter (TSM) concentration is a critical parameter in assessing water quality, particularly in estuaries and inland lakes section (Yang et al., 2022), where high TSM concentrations can disrupt the optical properties of water. This disturbance affects factors such as the depth of the true light layer, thereby influencing the growth of aquatic plants and the primary productivity of water bodies (K. A. Moore et al., 1997; Havens, 2003; Hou et al., 2017).

Suspended sediments cause an increase in the reflectance emerging from surface waters, particularly in the visible and near-infrared portions of the electromagnetic spectrum (Ritchie, 1976). The variation in spectral signature corresponding to different concentrations of suspended sediments is illustrated in Figure 1.

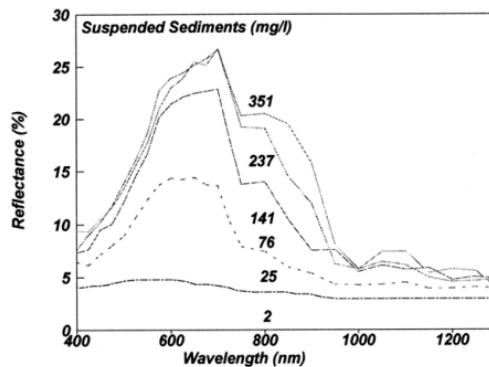


Figure 1: Reflectance versus wavelength is depicted here, showcasing the impact of suspended sediment concentration on Reflectance. The data was gathered from a boat-mounted high-resolution spectroradiometer, positioned 50 cm above the lake surface (Ritchie, 1976).

2.2.1.2 Chlorophyll

Lakes and other water bodies rely on their catchments for nutrients and other substances to sustain biological activities. However, an excess of these inputs leads to nutrient enrichment and eutrophication, resulting in the "aging" of the lake. The rate of eutrophication depends on topography, soils, land use, and runoff in the contributing catchment. Eutrophication can

be quantified in terms of trophic level or the concentration of chlorophyll contained in algal/plankton cells. While the aging of water bodies is a natural process, better information on eutrophication (chlorophyll content) in a lake enables the development of more effective management plans to control nutrient inputs from the catchment and mitigate the rate of "aging" in lakes. Monitoring the trophic level or chlorophyll concentration in lakes is crucial for managing eutrophication. Excessive algal populations can be problematic, and their concentrations are typically monitored by collecting samples, extracting chlorophyll, and measuring concentrations using photometric techniques in the laboratory. Remote sensing has also been utilized to measure chlorophyll concentrations and patterns, primarily based on empirical relationships between radiance/reflectance in narrow bands or band ratios and chlorophyll concentration (see Section 2.3) (Ritchie and Schiebe, 2000).

Chlorophyll-A (ChloA) and B are two major types of chlorophyll found in plants and green algae. Both are involved in the process of photosynthesis.

ChloA is the primary pigment responsible for absorbing light and providing energy for oxygenic photosynthesis. It is present in all plants, green algae, and cyanobacteria. ChloA absorbs most effectively at wavelengths of 429 nm (violet-blue) and 659 nm (orange-red). It reflects a blue-green color, giving plants their characteristic green appearance (cf Figure 2).

On the other hand, chlorophyll-B is a secondary pigment responsible for collecting light energy and transferring it to ChloA during photosynthesis. It is also found in plants and green algae. Chlorophyll-B absorbs most effectively at wavelengths of 455 nm (violet) and 642 nm (red), reflecting a yellow-green color (cf Figure 2).

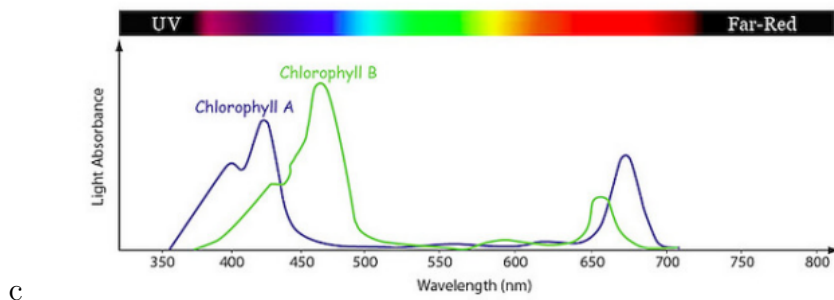


Figure 2: Absorption spectrum of chlorophyll A and B (Panawala, 2017)

Different studies have highlighted the potential for measuring chlorophyll using remote sensing techniques. However, they have also pointed out a significant limitation: the broad wavelength spectral data provided by current satellites often fail to distinguish between chlorophyll and suspended sediments (A. Dekker and Peters, 1993; Ritchie et al., 1994). This limitation stems from the dominance of spectral signals emanating from suspended sediment as shown in Figure 3.

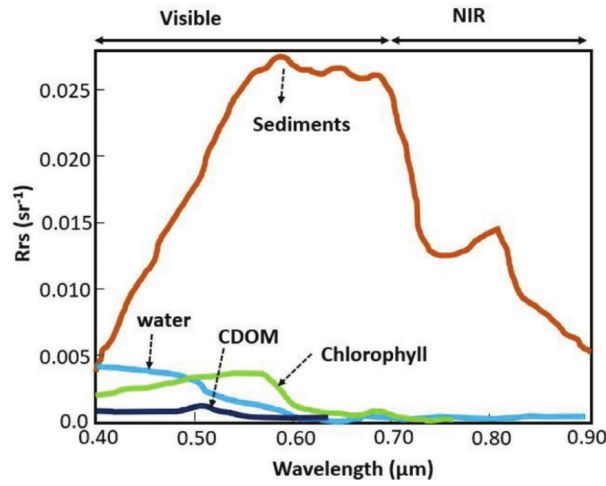


Figure 3: Spectral signature of the water components (clear water in blue, water with chlorophyll content in green, water with CDOM in black, and sedimented water in orange) (Palacios, 2016)

To overcome this challenge, research on the relationship between chlorophyll and narrow-band spectral details at the "red edge" (680-750 nm) of the visible spectrum has shown a linear correlation between ChloA concentration and the difference in emergent energy between the chlorophyll scattering range (700–705 nm) and the chlorophyll absorption range (675–680 nm). This relationship holds even in the presence of high suspended sediment concentrations that dominate other parts of the spectrum, as illustrated in Figure 4 (Ritchie and Schiebe, 2000; Schmugge et al., 2002; Yacobi et al., 1995).

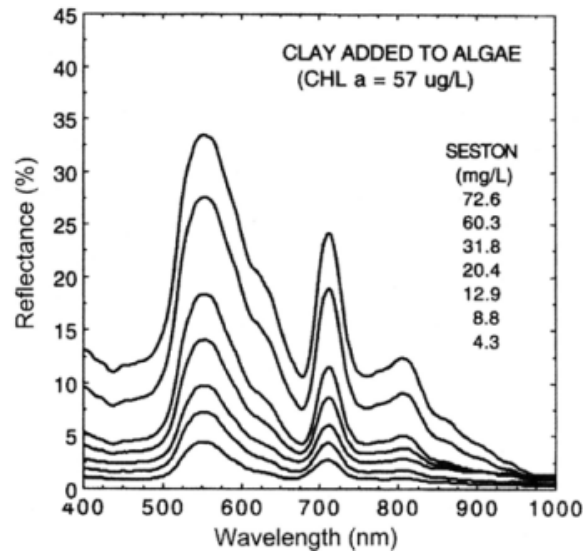


Figure 4: Spectral signature of water with constant chlorophyll-A concentration (57 μg/l) and varying suspended sediment concentrations (SESTON), illustrating the relative contributions of chlorophyll and suspended sediment to the reflectance spectra of surface water. Data based on *in situ* laboratory measurements taken 1 meter above the water surface. Source: (Schalles et al., 1997).

2.2.2 Optical characteristics of water bodies

The color of the water is a multifaceted optical characteristic shaped by scattering, absorption processes, emission from the water column, and reflectance from the substrate (see Figure 5). The reflectance of the substrate (e.g., seagrass, macro-algae, corals, sand, etc.) is similarly influenced by absorption, scattering, and to a lesser extent, emission of substrate materials. Color variations are primarily dictated by particulate and dissolved substances that absorb and scatter solar and sky radiation penetrating the water surface.

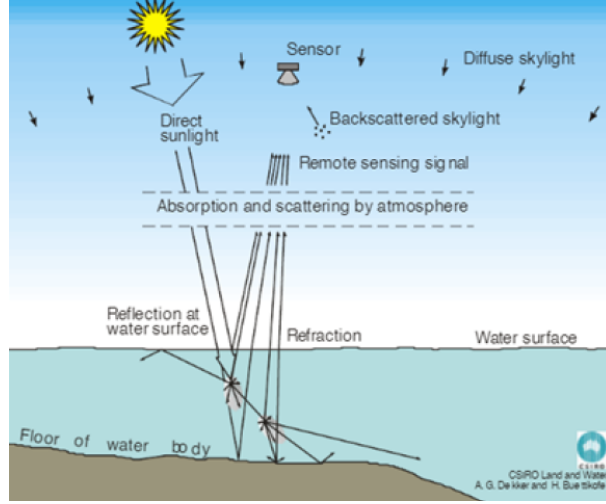


Figure 5: A schematic diagram of the various processes that contribute to the signal as measured by a remote sensor in optically shallow water where the substrate has a significant effect on the water leaving radiance at the surface (A. G. Dekker et al., 2001)

The light absorption (a) of a water mass can be defined as :

$$a = a_{ph} + a_{NAP} + a_{CDOM} + a_w \quad (2)$$

with a_{ph} referring to the light absorption by phytoplankton, a_{NAP} referring to the light absorption by the non-algal particles (NAP) also called sediments, a_{CDOM} referring to the light absorption by the CDOM and a_w referring to the light absorption by the water itself.

The light scattering (b) can be defined as :

$$b = b_f + b_b \quad (3)$$

with b_f referring to the light scattering by particles in the forward direction and b_b referring to the light scattering by particles in the backward direction.

The remote sensing reflectance (R_{rs}) of a water mass, captured at the water's surface (referred to as water-leaving reflectance), can be defined as:

$$R_{rs}(\lambda, 0^+) = C \frac{b_b(\lambda)}{a(\lambda) + b_b(\lambda)} \quad (4)$$

with $Rrs(\lambda, 0^+)$ referring to the surface water reflectance at the wavelength λ under zero-angle incidence (meaning the reflectance when light arrives perpendicularly to the water surface) and C a normalization coefficient that depends on the measurement system and is used to adjust the scale of reflectance (Palacios, 2016).

2.2.3 Atmospheric correction for aquatic environment

A remote sensing instrument detects upwelling radiances L , incorporating contributions from the atmosphere, water surface, and water column. The atmospheric contribution L_a emerges from solar radiance scattered by atmospheric gases and aerosols one or more times in the direction of the sensor. Surface-reflected radiance (sun and sky glint) L_r represents upwelling solar radiance reflected toward the sensor by the water surface. Water-leaving radiance L_w originates from light transmitted through the surface into the ocean L_t , where it undergoes modifications due to the absorbing and scattering components in the water. Subsequently, it is scattered upward L_u and ultimately exits the sea surface in the direction of the sensor (Mobley, 2022). Those contributions are represented in Figure 6.

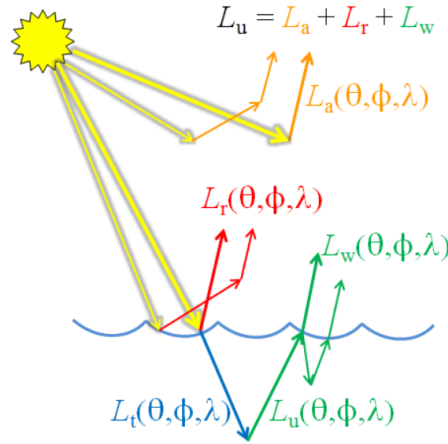


Figure 6: Contributions to the total upwelling radiance above the sea surface, L_u . Yellow arrows are the Sun’s unscattered beam; blue arrows represent lights that are transmitted through the surface into the ocean L_t ; orange arrows are atmospheric path radiance L_a ; red is surface-reflected radiance L_r ; green is water-leaving radiance L_w . Thick arrows represent single-scattering contributions; thin arrows illustrate multiple-scattering contributions (Mobley, 2022).

The radiance scattered by the atmosphere along the path between the sea surface and the sensor carries information about atmospheric aerosols and other atmospheric parameters. However, it’s essential to note that only the water-leaving radiance contains information relevant to the characteristics of the water column itself. A downward-looking sensor measures the total radiance $L_u = L_a + L_r + L_w$ and cannot separate the various contributions to the total radiance. The atmospheric correction refers to the process of removing the contributions by surface glint L_r and atmospheric scattering L_a so the total radiance only depends on the water leaving radiance L_w (Mobley, 2022).

2.3 Water Quality Parameters Retrieval Modes

Water quality parameters such as chlorophyll-A and turbidity can be retrieved from remote sensing data using four different approaches (Yang et al., 2022) :

- **Empirical Mode:** The empirical mode (EM) is a simple statistical regression formula used to correlate ground-measured water quality parameter values with spectral reflectance data (Ouma et al., 2018). While easy to use and refine for increased accuracy by selecting precise spectral channels, EM has limitations (Li et al., 2017). It struggles to meet accuracy requirements due to the complex composition of water quality variables, leading to uncertainty in parameter concentration estimation (Palmer et al., 2015). Additionally, EM lacks generalizability across regions and time periods, requiring extensive field data for model development and validation (Gao et al., 2015).
- **Analytical Mode:** The analytical mode (AM) utilizes bio-optical and radiation transmission models to simulate light propagation in the atmosphere and water bodies, establishing a relationship between water quality components and radiance or reflection spectra. Gordon et al., 1975 proposed a general relationship between optical reflectance and water properties, forming the basis for AM. However, the complexity of water composition and radiation transmission poses challenges in model development, requiring extensive measurements of factors such as water's optical characteristics and quality variables. Discrepancies in spectral resolution between satellite and ground-based sensors further complicate model research, limiting the practical application of quantitative remote sensing inversion based on AM (Palmer et al., 2015; C. F. Le et al., 2009).
- **Semi-Empirical Mode:** The semi-empirical mode (SEM) combines aspects of both the empirical and analytical modes. It utilizes a paradigm derived from rigorous analytical models, allowing for the calculation of water quality variables by incorporating a small amount of *in situ* data along with reflectance or radiance values. This approach has been employed to retrieve various water quality components with higher accuracies, including suspended matter (SM), ChloA, and colored dissolved organic matter (CDOM), as well as visibility and turbidity (Hunter et al., 2010; Zhou et al., 2014). For instance, Lathrop, 1989 demonstrated that the SPOT-1 Red and Near-Infrared (NIR) bands were particularly sensitive to suspended matter concentration, enabling the calculation of different SM concentrations using the Red-Green band ratio and NIR band within the SEM framework. The semi-empirical method accounts for the optical characteristics of water quality parameters and offers greater flexibility compared to the purely empirical approach.
- **Artificial Intelligence Mode:** The Artificial Intelligence (AI) mode (AIM) represents a distinct approach from the previously mentioned modes (EM, AM, and SEM). Due to complicated and various background effects from the water surface, a combination of water quality parameters, and sediment deposits, AI mode can capture both linear and nonlinear relationships compared with traditional statistical or physical modes (Chebud et al., 2012; Becker et al., 2019).

2.3.1 Turbidity Algorithms

Researchers have developed various methods for estimating TSM concentrations using remote sensing data, achieving successful practical applications. For example, A. G. Dekker et al., 2002 studied turbid lakes with TSM concentrations ranging from 5 to 50 g/m³ and confirmed a strong exponential relationship between TSM concentration and the main values of the Green (495-570 nm) and Red (625-740 nm) bands.

Several models for TSM inversion are presented in Table 1. In these models, C_{TSM} denotes the total suspended matter concentration, $R(\lambda)$ represents the R_{rs} at wavelength λ .

Table 1: Inversion models of TSM. In these models, C_{TSM} denotes the total suspended matter concentration, $R_{rs}(\lambda)$ represents the remote sensing reflectance at wavelength λ

Equation	Modes	Reference
$C_{TSM} = a \times R_{rs}(654.5) + b$	EM	Miller and McKee, 2004
$C_{TSM} = a \times R_{rs}(654.5)^3 + b \times R_{rs}(654.5)^2 + c \times R_{rs}(654.5) + d$	SEM	Ondrusek et al., 2012
$C_{TSM} = a \times R_{rs}(654.5)^b$	EM	Chen et al., 2007
$C_{TSM} = a \times R_{rs}(654.5)^2 + b \times R_{rs}(654.5) + c$	SEM	Petus et al., 2010
$\log_{10}(C_{TSM}) = a \times R_{rs}(561.5) + b \times R_{rs}(654.5) + c \times \frac{R_{rs}(482)}{R_{rs}(561.5)} + d$	EM	M. Zhang et al., 2010
$C_{TSM} = \frac{a \times R_{rs}(825)}{1 - \frac{R_{rs}(1251)}{b}} + c$	SEM	Caballero et al., 2014
$\log(C_{TSM}) = \frac{-b \pm \sqrt{b^2 - 4a(c - \frac{\log R_{rs}(565)}{\log R_{rs}(450)})}}{2a}$	SEM	Wang et al., 2018

Analytical algorithms can also be used to retrieve turbidity. The first one is the Normalized Difference Turbidity Index (NDTI) that has been introduced in Lacaux et al., 2007. This index relies on the fact that pure water exhibits a specific radiometric response, with its signal being weak in the green wavelength (less than 10%), becoming very small in the red wavelength, and nearly negligible in the NIR (Robin, 1995). However, the increase in turbidity alters the radiometric responses of the environment, making it behave similarly to bare soils (Guyot and Seguin, 1988). This implies that the values of the red radiometric responses are much larger than those of the green ones, making the relationship between the green and red wavelengths reversed (Campbell, 1996; Verbyla, 1995).

The NDTI equation is defined as:

$$NDTI = \frac{R_{rs}(red) - R_{rs}(green)}{R_{rs}(red) + R_{rs}(green)} \quad (5)$$

Here, $R_{rs}(red)$ represents the radiometric response in the red wavelength, while $R_{rs}(green)$ denotes the radiometric response in the green wavelength.

The second analytical algorithm is the Total Suspended Matter Index (TSMi). This algorithm,

introduced by A. G. Dekker et al., [2002], is based on the green and red Landsat bands as they were identified as the most suitable for estimating variations in TSM.

The TSMi equation is defined as :

$$TSMi = \frac{R_{rs}(green) + R_{rs}(red)}{2} \quad (6)$$

Here, $R_{rs \text{ red}}$ represents the radiometric response in the red wavelength, while $R_{rs \text{ green}}$ denotes the radiometric response in the green wavelength.

2.3.2 Chlorophyll-A Algorithms

Some inversion models for ChloA concentration are presented in Table 2. C_{ChloA} represents the concentration of ChloA, and $R_{rs}(\lambda)$ denotes the R_{rs} at wavelength λ . Empirical and semi-analytical algorithms, while widely used, are often limited in geographical applicability and accuracy due to challenges in atmospheric correction, leading to uncertainties (Yang et al., [2022]).

Table 2: Inversion models of ChloA. C_{ChloA} represents the concentration of chlorophyll-A, and $R_{rs}(\lambda)$ denotes the R_{rs} at wavelength λ

Equation	Modes	Reference
$C_{ChloA} = a \times [R_{rs}^{-1}(665) - R_{rs}(715)] * R_{rs}(750) + b$	SEM	Gitelson et al., [2008]
$C_{ChloA} = a \times [R_{rs}^{-1}(667) \times R_{rs}(748)]$	SEM	Gitelson et al., [2009]
$C_{ChloA} = a \times \frac{-b \pm \sqrt{b^2 - 4a[c - (R_{rs}^{-1}(670) - R_{rs}^{-1}(710)) \times R_{rs}(750)]}}{2a}$	SEM	C. Le et al., [2009]
<i>Convolutional neural network</i>	AIM	Pyo et al., [2019]

Another intriguing algorithm is offered by a group associated with NASA known as the Ocean Biology Processing Group (OBPG). The OBPG produces standardized ocean color products for compatible sensors at Level-2 and Level-3. This algorithm calculates ChloA concentration in mgm^{-3} using empirical relationships from *in situ* ChloA measurements and R_{rs} across three or more bands spanning the 440 - 670 nm spectral bands. It is applicable to current ocean color sensors and included in NASA's Level-2 OC and Level-3 CHL product suites ("Viewing Chlorophyll a — Algorithm Publication Tool", [n.d.]).

This algorithm is divided into three different algorithms based on the ChloA concentration range :

- **OCx Algorithm:** The Ocean Chlorophyll algorithm is widely used to compute ChloA concentration from band R_{rs} . This algorithm is called OCx with x being the number of bands R_{rs} used. The OC3 algorithm is used to compute ChloA concentration greater than 0.35 mg/m^3 in the MODIS Aqua product (J. E. O'Reilly et al., [1998]). This algorithm is

presented in Equation 7 and 8:

$$\log_{10}(\text{ChloA}) = a_0 + a_1 \times x + a_2 \times x^2 + a_3 \times x^3 + a_4 \times x^4 \quad (7)$$

$$x = \log_{10} \left(\frac{\max [R_{rs}(443), R_{rs}(482)]}{R_{rs}(561)} \right) \quad (8)$$

with $a_{0 \rightarrow 4}$ being polynomial coefficients specific for a satellite.

- **CI Algorithm:** The CI equation developed in Hu et al., 2019 is used to compute ChloA concentration smaller than 0.25 mg/m³. This equation uses band R_{rs} and polynomial coefficients specific to each satellite instrument. This algorithm uses the two equations presented in Equation 9 and 10.

$$CI = R_{rs}(\lambda_{\text{green}}) - [(R_{rs} \lambda_{\text{blue}}) + \frac{(\lambda_{\text{green}} - \lambda_{\text{blue}})}{(\lambda_{\text{red}} - \lambda_{\text{blue}})} \times (R_{rs}(\lambda_{\text{red}}) - R_{rs}(\lambda_{\text{blue}}))] \quad (9)$$

$$\text{ChloA} = 10^{(a_0 + a_1 \times CI)} \quad (10)$$

with λ being the wavelength values in nanometers (nm).

- **Mixed Algorithm:** For ChloA concentration between 0.25 and 0.35 mg/m³ the CI and OCx algorithm are blended using a weighted approach. This algorithm is presented in Equation 11.

$$\text{ChloA} = \frac{\text{ChloA}_{CI} \times (0.35 - \text{ChloA}_{CI})}{0.35 - 0.25} + \frac{\text{ChloA}_{OCx} \times (\text{ChloA}_{CI} - 0.25)}{0.35 - 0.25} \quad (11)$$

The last algorithm aborded in this study is the Normalized Difference Chlorophyll Index (NDCI) algorithm, introduced by Mishra and Mishra, 2012 which addresses the challenge of accurately estimating ChloA concentration in turbid productive waters. Unlike traditional empirical algorithms like OC4v4, which struggle in such conditions due to the presence of optically active constituents like CDOM, NDCI offers improved accuracy (J. E. O'Reilly et al., 1998). It utilizes R_{rs} data at 665 nm and 708 nm, mimicking MERIS sensor bands, to minimize confounding influences. By normalizing the spectral band difference at these wavelengths, NDCI effectively accounts for uncertainties in atmospheric contributions and seasonal solar azimuth differences. This approach demonstrates high sensitivity to ChloA concentration variations and shows promise for widespread applicability in optically complex waters. The NDCI equation is formulated as :

$$\text{ChloA} \propto \frac{[R_{rs}(708) - R_{rs}(665)]}{[R_{rs}(708) + R_{rs}(665)]} \quad (12)$$

2.4 Lake Tanganyika

Lake Tanganyika is a vast north-south trending rift valley lake, with a mean width of 50 km and a mean length of 650 km. Its depth averages 570 m, reaching a maximum depth of 1,470 m. Situated on the borders of Burundi, Tanzania, Zambia, and the Democratic Republic of Congo, Lake Tanganyika holds significant importance as a vital source of sustenance and economic revenue for these nations. Renowned for its exceptional productivity, the lake has long sustained thriving pelagic fisheries, making it one of the world's most prolific aquatic ecosystems (Nixon, 1988). Among the notable rivers flowing into it are the Malagarasi, the Ruzizi, and the Kalambo. The Lukuga River serves as its outlet, connecting to the Lualaba River ("Lake Tanganyika | Size, Geographical Location & Wildlife | Britannica", 2023).

2.4.1 Limnological cycle of Lake Tanganyika

The lake encompasses distinct layers within a body of water, each with its unique characteristics and roles in ecosystem dynamics. At the surface lies the epilimnion, characterized by warmer temperatures conducive to photosynthesis. This uppermost layer typically maintains higher oxygen levels compared to deeper layers. Below the epilimnion lies the metalimnion, which hosts the thermocline—a zone where the temperature gradient is most pronounced. Surrounding the thermocline, the metalimnion exhibits steep temperature changes compared to adjacent layers. While the thermocline occupies a minimal space, the metalimnion encompasses a broader area with a substantial rate of temperature change. Beneath these layers lies the hypolimnion, the coldest and densest region of the water body. Secluded from atmospheric influences, the hypolimnion often remains stagnant in thermally stratified lakes or reservoirs. Organic matter from the epilimnion settles into the hypolimnion, fueling bacterial decomposition processes. As bacteria respire and decompose organic material, oxygen consumption outpaces replenishment from photosynthesis or atmospheric sources. The epilimnion functions as the trophogenic zone, where mixing and photosynthesis exceed respiratory processes. Conversely, the hypolimnion serves as the tropholytic region, where organic matter breakdown and bacterial mineralization dominate. Understanding these distinct layers and their interconnected processes is essential for comprehending the dynamics of freshwater ecosystems and their responses to environmental changes (Brusseau et al., 2019). A schematic visualization of the different layers is presented in Figure 7.

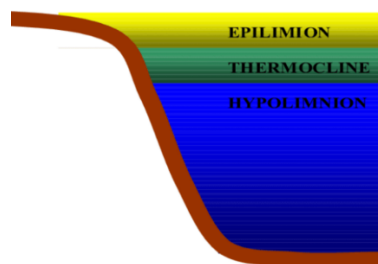


Figure 7: Cross section of a thermally stratified lake showing the location of the epilimnion and hypolimnion and associated temperature changes (Adapted from Brusseau et al., 2019)

Plisnier et al., [1996] addressed the topic of the limnological cycle of Lake Tanganyika, outlining its significance and providing insights into its complex dynamics. The limnological cycle encompasses the recurring patterns of physical, chemical, and biological processes within aquatic ecosystems like lakes. The following steps elaborated below, are succinctly outlined in Figure 8.

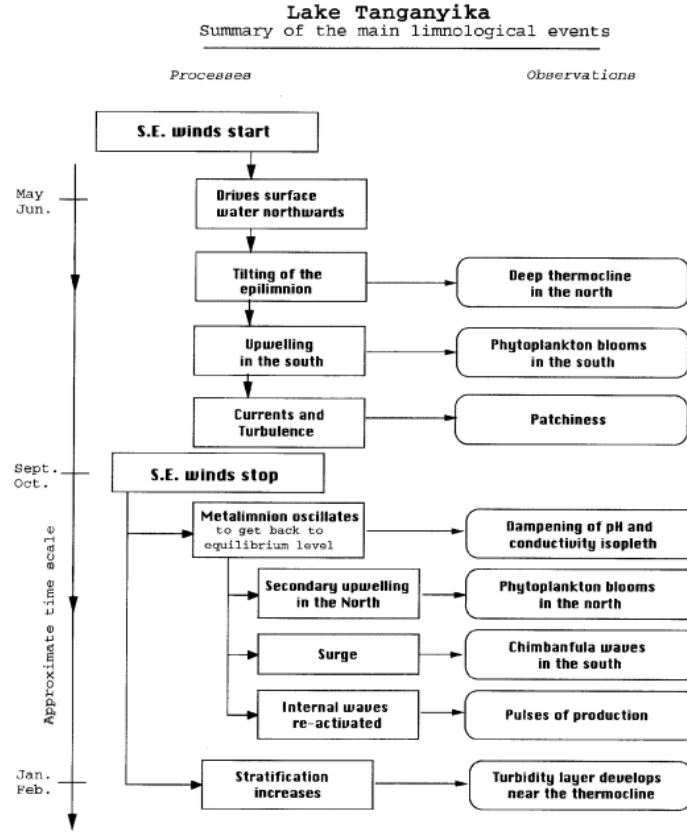


Figure 8: Main events in the limnology of Lake Tanganyika during a yearly cycle (Plisnier et al., [1996])

In the months of May and June, prevailing southeast winds propel warm epilimnion water towards the northern reaches of the lake. This action results in a noticeable tilting of the epilimnion (see Figure 9), with the extent of warm water accumulation in the north contingent upon wind strength.

The accumulation of warm water in the northern sector predominantly occurs at the surface. Concurrently, southward-flowing deep currents trigger upwelling along the southern coast. This phenomenon is attributed to both cooling from the deep, return currents and from lake-atmosphere interactions, leading to minimal stratification. During upwelling events, mixing of water near Mpulungu is observed, resulting in elevated concentrations of nitrogen and phosphorus in the epilimnion. Moreover, decreased water transparency and increased phytoplankton biomass (Beauchamp, [1939]; Coulter, [1968]) are observed, particularly during the dry season.

The upwelling process exhibited notable mixing, especially when deep internal waves reached their peak. Mixing intensified during the daily cycle, particularly in the early morning when

strong winds and low air temperatures coincided. Turbulence during the dry season extends to considerable depths, with vortices and water exchange detected up to 200 meters below the thermocline. Following the cessation of southeast winds at the dry season's end, the metalimnion transitions towards a horizontal orientation but continues to oscillate over several months. This period is characterized by turbulent water movements, as evidenced by fluctuations in metalimnion and thermocline depth, as well as variations in pH and conductivity.

Secondary upwelling events in the northern sector, noted in October 1993, are likely linked to metalimnion oscillations. These events coincide with rises in deep conductivity isopleths towards the surface and reduced transparency near Bujumbura/Uvira. Although weaker than southern upwelling, they exhibit a yearly recurrence and contribute to phytoplankton blooms (Symoens, 1995a; Symoens, 1995b; Dubois, 1958; Hecky and Kling, 1981). The northern bloom typically occurs alongside the onset of rain, which is commonly attributed to an increased nutrient input. However, although rainfall likely contributes to phytoplankton blooms, it is proposed that secondary upwelling during this period primarily brings nutrients to the surface. This phenomenon elucidates the usual lag between intense algae production in the south (main upwelling) and subsequent heightened productivity in the north (secondary upwelling).

Strong surface waves, locally called "Chimbanfula" waves, annually observed near Mpulungu mark the onset of the wet season (in October) and may be associated with metalimnion oscillations after the S.E. winds ceased and traveling waves. These waves are linked to fish kills and plankton blooms. Internal waves persist throughout the year, inferred from fluctuations in various parameters. The pulses generated by internal waves manifested as nutrient waves in the water column, resulting in a vertical displacement of nutrient concentrations of approximately 20 meters. These long-period waves likely triggered regular pulses of phytoplankton production, as the rhythmic movements of internal waves facilitated the movement of deep eutrophic water to the biotic and photic zones, which typically exhibit oligotrophic characteristics. These "nutrient waves" likely influenced organisms adapted to capitalize on favorable yet episodic conditions.

As the lake stabilizes from February to May, increased turbidity near the thermocline at Bujumbura/Uvira fosters the development of a deep-living organismal community, potentially akin to the bacterial 'plate' observed in Lake Kivu (Haberyan and Hecky, 1987). This community thrives on nutrients and reduced matter brought up by internal waves, particularly during periods of calmness in the lake. River inflows have minimal influence on the pelagic region of Lake Tanganyika, given that these waters tend to be colder and swiftly descend beneath the thermocline (Plisnier et al., 1996).

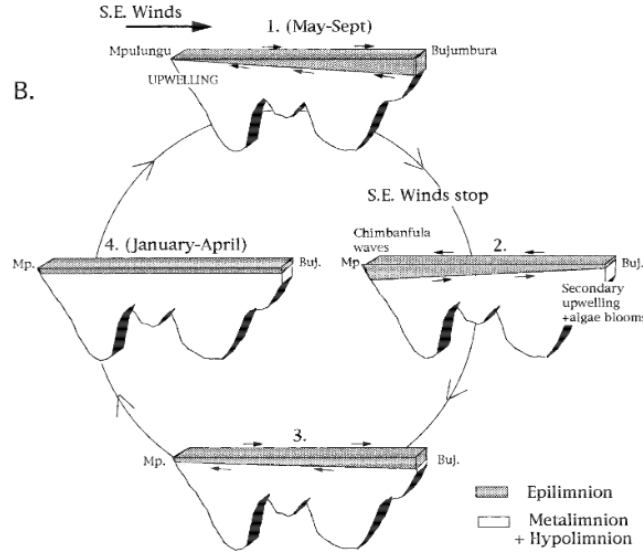


Figure 9: Schematic profile of water layers during an annual cycle. States 2 and 3 alternate before reaching stage 4. Internal waves are not indicated. (Plisnier et al., 1996)

2.4.2 Chlorophyll-A based Lake segmentation

Utilizing daily MODIS AQUA satellite data, Bergamino et al., 2010 investigated the variability of ChloA concentration in Lake Tanganyika from July 2002 to November 2005. This study partitioned the lake into 11 spatially coherent regions, complemented by two delocalized coastal regions. A visualization of the spatial distribution of the 13 regions characterized by comparable and co-varying levels of phytoplankton is illustrated in Figure 10.

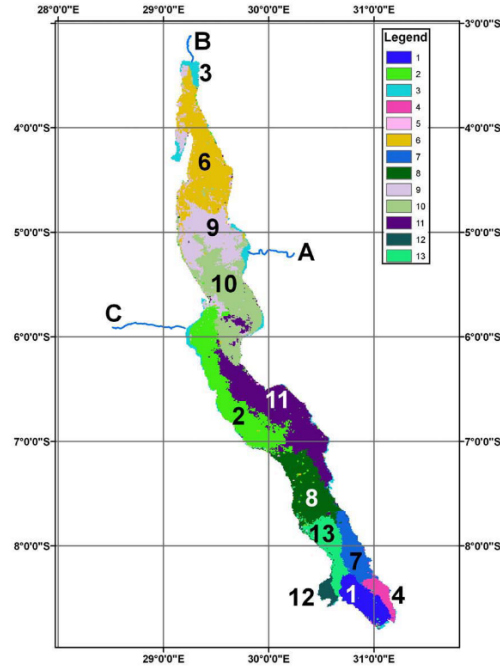


Figure 10: Regionalisation of Lake Tanganyika in 13 subregions (R1–R13) of co-varying chlorophyll-A concentrations with main rivers, Malagarasi (A), Ruzizi (B) and Lukuga (C) (Bergamino et al., 2010)

In the case of this study region R3 corresponds to both the estuarine areas of the Ruzizi River and the Malagarasi River.

2.4.3 Impact of the sediments brought by Ruzizi river

The River Ruzizi, the only outlet from Lake Kivu, flows into the northern end of Lake Tanganyika. It accounts for approximately 30 percent of the riverine inflow to Lake Tanganyika (Hecky et al., 1991). At the Ruzizi delta, a clear boundary separates the blue-green lake water from the brown river water. This observation has led to the widely accepted assumption that the Ruzizi River merges with the lake as an underflow, without an immediate impact on the surface water (Beauchamp, 1939; Coulter, 1968).

Vandelannoote et al., 1999 assessed the fate of the Ruzizi river flow in Lake Tanganyika. It evaluated two different fates of the river flow: offshore transport and horizontal spread by shorebound currents.

Offshore transport of Ruzizi water : The surface water of Lake Tanganyika is widely recognized for its extreme nutrient-poor characteristics (Van Meel, 1987; Edmond et al., 1993). Consequently, nutrient inputs from the Ruzizi River are rapidly assimilated into the lake. Similar observations were made at the mouth of the River Ntahangwa, which is known as Burundi's most polluted affluent of Lake Tanganyika (Vandelannoote et al., 1996; Vandelannoote et al., 1999). This assimilation process results in a rapid decrease in ChloA and suspended sediment concentrations near the river mouth, as illustrated in Figure 11. The impact of the river flow on surface concentrations is assumed to be very located, reaching an average distance of 100 meters off the river mouth for suspended sediments and ChloA.

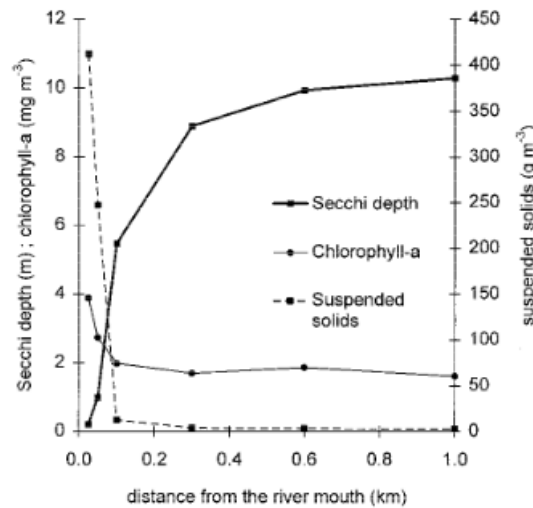


Figure 11: Mean Secchi depth and pulled results for suspended solids and chlorophyll-A along the transect of the Ruzizi delta (Vandelannoote et al., 1999)

Horizontal spread by shorebound currents : In general, there is a strong tendency for inflows to a lake to adhere to the shoreline (Horne and Goldman, 1994). In Lake Tanganyika, this behavior corresponds with the description provided by Well and Chapman, 1976 and supported

by Vandelannoote et al., [1999] of a persistent and steady clockwise coastal circulation. This circulation can lead to the transport of the Ruzizi water to the east as noticed by Ntakimazi, [1995] with Ruzizi water found at around 1km east to the river mouth).

3 Materials and methods

3.1 Overall view

Figure 12 presents the methodology used in this study.

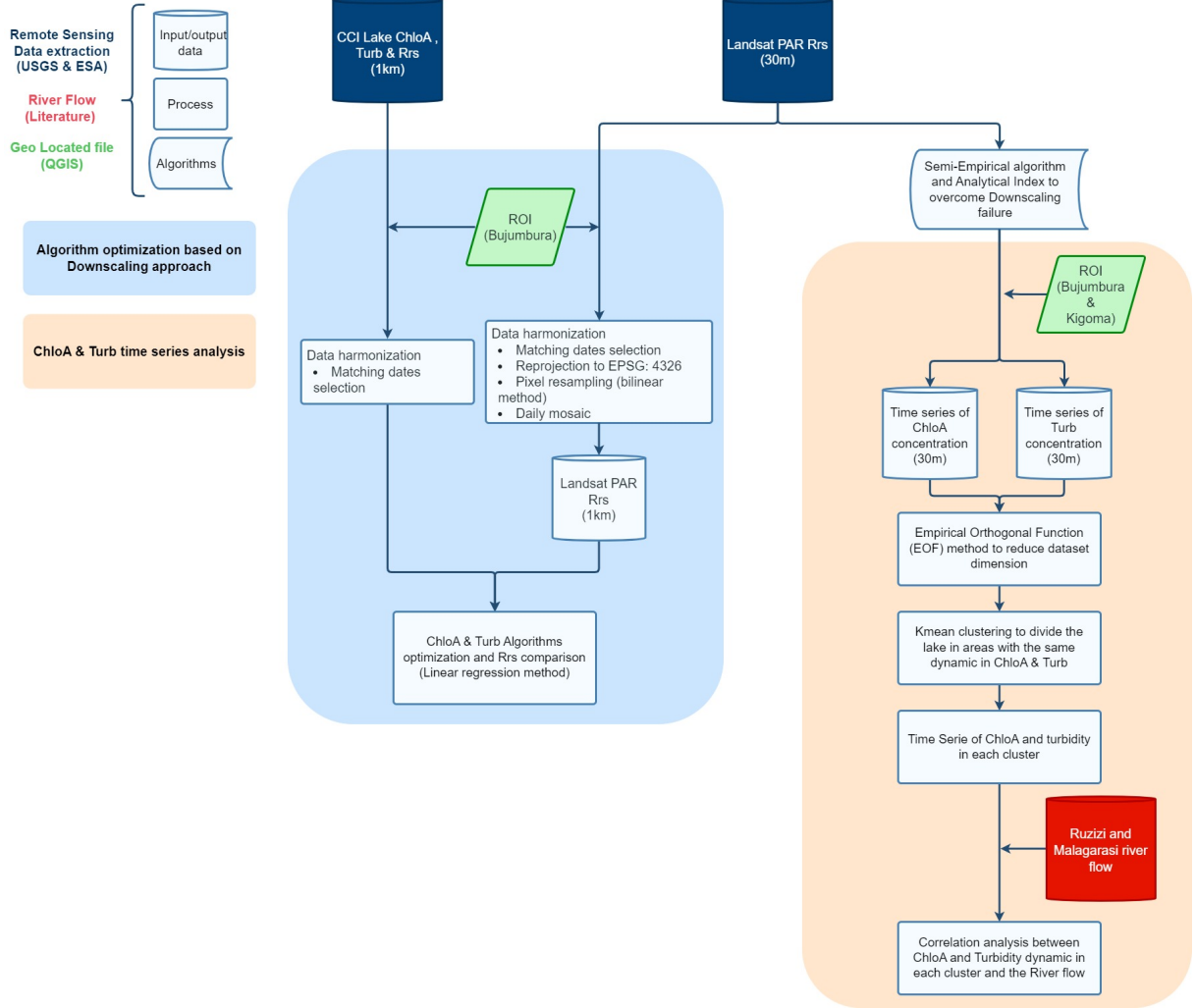


Figure 12: Methodology flowchart

This work is based on a remote sensing product called Landsat Provisional Aquatic Reflectance (PAR). The first step focused on attempting to create a dataset of ChloA and turbidity concentrations using the Landsat PAR R_{rs} from 2013 to 2023. The employed method is known as downscaling, which aims to reduce the spatial resolution of a product (in this case, the ChloA and turbidity dataset from CCI Lakes) by leveraging the relationship between this product and another with a finer spatial resolution (Landsat PAR).

The second step involved constructing datasets for ChloA and turbidity in the estuarine areas of the Ruzizi and Malagarasi Rivers using a semi-empirical algorithm and an analytical index. However, the validity of these datasets could not be verified due to a lack of *in situ* data.

The third step focused on analyzing the spatiotemporal dynamics of the previously created datasets. Each step was applied independently to each region. This began with reducing

the dimensionality of the datasets using Empirical Orthogonal Function (EOF) analysis. This method generated a few components for each variable (ChloA and turbidity) that represented the global spatial dynamics over 10 years. The first two components (one for ChloA and one for turbidity) were selected to cluster regions (thanks to the Kmean clustering function) with similar dynamics in both ChloA and turbidity. The daily median and monthly median concentrations of ChloA and turbidity were examined for each cluster to assess their dynamic variations. These dynamics were then compared with the river flow from the nearest river to evaluate the impact of this river on ChloA and turbidity dynamics.

3.2 Data used

3.2.1 Landsat 8-9 Provisional Aquatic Reflectance

The Operational Land Imager (OLI) on Landsat 8 is the most recent optical sensor in the Landsat satellite series. While its primary purpose was land use and land cover monitoring, OLI images have been utilized for various studies related to lakes. These include estimating water transparency, colored dissolved organic matter, identifying algal blooms, measuring total suspended sediments (TSS), ChloA concentration, Secchi disk depth, and turbidity. One significant advantage of OLI over other ocean color missions for inland waters is its spatial resolution of 30 meters, which allows for mapping optically active components (OACs) in small aquatic systems ($<5 \text{ km}^2$). However, in aquatic environments, the above-water remote sensing reflectance (R_{rs}) can be as low as 1 percent at the peak of reflectance. Due to this low signal intensity from aquatic environments, atmospheric correction routines for R_{rs} retrieval require extremely high accuracy. This task becomes even more complex in highly heterogeneous inland waters due to additional factors affecting R_{rs} retrieval, such as water turbidity (which can increase the signal in the near-infrared spectrum), whitecaps, sun-glint, and adjacency effects from surrounding land. To assist in the retrieval of R_{rs} from Landsat 8/OLI imagery in inland waters, several atmospheric correction routines have been developed recently such as the Landsat Provisional Aquatic Reflectance developed by the United States Geological Survey (USGS) (Ogashawara et al., 2020).

3.2.1.1 Landsat Provisional Aquatic Reflectance algorithm

The Landsat provisional aquatic reflectance algorithm is derived directly from the Sea-viewing Wide Field-of-View Sensor (SeaWiFS) Data Analysis System (SeaDAS) package. This originates from the distribution by the National Aeronautics and Space Administration (NASA) Ocean Biology Processing Group (OBPG) (see Mobley, 2022). The objective of this algorithm is to provide water-leaving radiance as seen in Section 2.2.3 from Landsat 8/9 Level 1 (L1) Top of Atmosphere (TOA) reflectances.

The Landsat L1 TOA water-leaving radiances, $L_w(\lambda)$, corresponding to each sensor wavelength, λ , undergo normalization to eliminate any residual impact of solar orientation and atmospheric attenuation on the downwelling radiation. This process results in the generation of normalized water-leaving radiance, $L_w(\lambda)$, commonly represented as radiance reflectance, $Rrs(\lambda)$, or

Remote Sensing Reflectance. This representation is achieved by straightforwardly dividing the average extraterrestrial solar irradiance, $F_0(\lambda)$. Within this algorithm, the Top of Atmosphere (TOA) radiance is presumed to be linearly divided into distinct physical contributions, as illustrated below:

$$L_t(\lambda) = [L_r(\lambda) + L_a(\lambda) + t_{dv}(\lambda)L_f(\lambda) + t_{dv}(\lambda)L_w(\lambda)]t_{gv}(\lambda)t_{gs}(\lambda)f_p(\lambda) \quad (13)$$

where

$L_t(\lambda)$ is the radiance captured by the sensor at each wavelength

$L_r(\lambda)$ is the radiance contribution due to Rayleigh scattering by air molecules

$L_a(\lambda)$ is the contribution due to scattering by aerosols, including multiple scattering interactions with the air molecules

$L_f(\lambda)$ is the contribution from surface whitecaps and foam

$L_w(\lambda)$ is the water-leaving component

$t_{dv}(\lambda)$ is the transmittance of diffuse radiation through the atmosphere in the viewing path from surface to sensor

$t_{gv}(\lambda)$ is the transmittance loss due to absorbing gases for all upwelling radiation traveling along the sensor view path

$t_{gs}(\lambda)$ is the transmittance to the downwelling solar radiation due to the presence of absorbing gases along the path from the Sun to the surface

$f_p(\lambda)$ is an adjustment for the effects of polarization

The atmospheric correction algorithm retrieves the water-leaving radiance, $L_w(\lambda)$, by estimating and subtracting the terms on the right-hand side of the equation mentioned above from $L_t(\lambda)$. Subsequently, the Remote Sensing Reflectance, $Rrs(\lambda)$, is computed as shown in Equation [14](#):

$$Rrs(\lambda) = \frac{L_w(\lambda)}{F_0 f_s \cos(\theta_s) t_{ds}} f_b(\lambda) f_\lambda \quad (14)$$

where

F_0 is the extraterrestrial solar irradiance

f_s is the adjustment if F_0 for variation in Earth-Sun distance

f_b is the bidirectional reflectance correction

f_λ is the correction for out-of-band response

θ_s is the solar zenith angle, which is the angle between the direction of the sun and the point directly overhead (zenith).

$t_{ds}(\lambda)$ is the transmittance of downwelling solar radiation due to the presence of absorbing gases along the path from Sun to surface.

The majority of the variables in the Equations [13] and [14] are derived through precomputed radiative transfer simulations or models, which rely exclusively on factors like the sensor’s spectral response, solar and sensor viewing angles, and supplementary data such as atmospheric gas concentrations, surface windspeeds, and surface pressure (“Landsat Provisional Aquatic Reflectance ADD | U.S. Geological Survey”, [2020]).

3.2.1.2 Performance of the Landsat Provisional Aquatic Reflectance

Ogashawara et al., [2020] evaluates the performance of the Landsat Provisional Aquatic Reflectance (Landsat PAR) by studying its behavior. The study aimed to assess how the Landsat PAR product performs for different small, turbid, and eutrophic lakes in Northern Germany compared to *in situ* measurements of above-water R_{rs} . The study highlighted several issues with the Landsat PAR product, such as the misidentification of cloud cover, leading to the overmasking of some images. Another identified problem was the overcorrection of aerosols, resulting in negative R_{rs} values. Landsat PAR strongly underestimated R_{rs} for most data points, particularly for bands 2 and 3, while performance for band 4 was relatively better. Despite its potential for inland water monitoring, the Landsat PAR product has significant limitations in accurately estimating water R_{rs} . Therefore, this product should be used with caution, considering these limitations.

The USGS Landsat PAR team responded regarding the issue of negative R_{rs} values, stating: "This is a relatively common/ known issue with the Aquatic products. In order to generate these images, they pull in an ocean algorithm (SeaDAS). The algorithm is really good for open water, but when applied to inland bodies of water it often fails. It overcorrects for aerosols, and then masks pixels where the atmospheric correction failed causing extremely negative values. The bands you are working with (Bands 1-4) are some of the most affected by this flaw in the algorithm, as it causes a not insignificant percent of the values of each of these bands to be incorrectly masked out." (USGS Landsat PAR responsible team, personal communication, April 24, 2024)

3.2.1.3 Characteristics of Landsat 8/9 Provisional Aquatic Reflectance

Landsat 8/9 satellite data offers a spatial resolution of 30 meters. Temporally, the revisit period has evolved: from 2013 to September 2021, it was 26 days, reduced to 13 days since then with the addition of the Landsat 9 satellite. Data availability spans from 2013 to the present. The Landsat Collection 2 Provisional AR product encompasses a range of information crucial for understanding and utilizing the data effectively, a list of all the bands available is presented in Table [18] available in the Appendix. The bands used in this study are presented in Table [3].

Table 3: Description of Aquatic Reflectance Bands (Sayler, 2023)

Name	Wavelength	Valid Range	Scale Factor
AR B1 (Coastal aerosol)	0.435-0.451 μm	0 to 10000	0.00001
AR B2 (Blue)	0.452-0.512 μm	0 to 10000	0.00001
AR B3 (Green)	0.533-0.590 μm	0 to 10000	0.00001
AR B4 (Red)	0.636-0.673 μm	0 to 10000	0.00001
AR B5 (NIR)	0.851-0.879 μm	0 to 10000	0.00001
Water Mask	/	0 to 4	/

The Water Mask band is generated and included in the AR product package to indicate the classification of pixels into categories such as land, water, cloud, cloud shadow, and snow (“Landsat Provisional Aquatic Reflectance ADD | U.S. Geological Survey”, 2020). The interpretation of this band is presented in Table 4

Table 4: Landsat Provisional AR WATER MASK Band Interpretations (“Landsat Provisional Aquatic Reflectance ADD | U.S. Geological Survey”, 2020)

Value	Description
0	Land/Fill (based on Level 2 product)
1	Water (based on Level 2 and Level 1 QA Pixel products)
2	Cloud (based on Level 1 QA Pixel product)
3	Cloud Shadow (based on Level 1 QA Pixel product)
4	Snow (based on Level 1 QA Pixel product)

In this study, only the pixels representing water (i.e., those with a value of 1 in the Water Mask band) were selected for further analysis or processing. Other pixel values corresponding to land, clouds, cloud shadows, or snow were excluded from consideration.

3.2.2 CCI Lakes

The Lakes Climate Change Initiative project (CCI Lakes) is an extensive dataset spanning from 1992 to 2020, encompassing crucial variables concerning lake dynamics such as surface water temperature, ice cover, water-leaving reflectance, water level, and extent. It includes observations for over 2000 relatively large lakes, representing a significant portion of the global freshwater surface, despite covering only a fraction of the world’s lakes.

The dataset utilizes visible and near-infrared optical imagery, thermal imagery, and microwave radar data from satellites. It provides observations in a common grid format with a spatial resolution of 1/120 degrees (approximately 1 km at the equator), presented in daily files. To ensure reliability, the data and algorithms underwent validation against *in situ* measurements whenever possible. Additionally, consistency analysis among the variables was conducted during the dataset’s development, resulting in the most comprehensive and consistent collection of satellite observations for lakes Essential Climate Variables currently available.

In the case of this study, we were interested in Lake Water-Leaving reflectance (LWLR). The primary input data for LWLR products consisted of Level-1b data sourced from the MERIS sensor on Envisat (3rd reprocessing) and the Ocean and Land Colour Instrument (OLCI) sensor on Sentinel-3A and Sentinel-3B. These datasets were obtained from the European Space Agency (ESA) from 2002 to 2012 and 2016–2020, respectively. Additionally, MODIS (Moderate Resolution Imaging Spectroradiometer) Level-1b data were acquired from the National Aeronautics and Space Administration (NASA) for the period spanning 2009 to 2019. This inclusion aimed to ensure overlap with and continuity between the observation periods of MERIS and OLCI (Carrea et al., 2023). The input data streams are summarized in Table 5.

Table 5: Instrument Satellite Dates used (dd/mm/yyyy) and Typical local time of observation (Carrea et al., 2023)

Instrument	Satellite	Dates used (dd/mm/yyyy)
MERIS	Envisat	29/04/2002 to 07/04/2012
MODIS	Aqua	08/04/2012 to 24/04/2016
OLCI	Sentinel-3A/B	25/04/2016 to 31/12/2020

The variables used in this study are presented in Table 6.

Table 6: CCI Lakes variables used (“Lakes CCI project Key Documents”, n.d.)

Name	Wavelength	Units
$R_{rs}(443)$	0.443 μm	/
$R_{rs}(490)$	0.490 μm	/
$R_{rs}(560)$	0.560 μm	/
$R_{rs}(665)$	0.665 μm	/
ChloA mean	/	$\mu\text{g/l}$
ChloA uncertainty	/	/
Turb mean	/	NTU
Turb uncertainty	/	/

The optical diversity and complexity of inland water bodies can often result in ambiguous interpretations of LWLR (Lake Water-Leaving Reflectance) concerning lake biogeochemical properties. To address this challenge, algorithms for deriving substance concentrations from LWLR were carefully selected and applied within predefined and validated scopes. Their assignment was based on a set of 13 lake Optical Water Types (OWT) (Spyrakos et al., 2018) following fuzzy classification methodologies (T. Moore et al., 2001).

To retrieve ChloA and Turbidity, various algorithms were employed, depending on the OWT. An OWT is a category that identifies and characterizes different types of water based on their optical properties. These optical properties include spectral reflectance, light transmission, and light absorption at various wavelengths. This classification helped to better understand water

characteristics and interpret data collected by optical sensors, which is crucial for monitoring water quality, modeling biogeochemical processes, and managing aquatic ecosystems. The resulting classification identified 13 spectrally distinct clusters of measurements in inland water (Spyrakos et al., 2018). They are presented in Table 7.

Table 7: Dominant characteristics of OWTs in inland waters (Spyrakos et al., 2018).

OWT	Dominant characteristics
1	Hypereutrophic waters with scum of cyanobacterial bloom and vegetation-like R_{rs}
2	Common case waters with diverse reflectance shape and marginal dominance of pigments and CDOM over inorganic suspended particles
3	Clear waters
4	Turbid waters with high organic content
5	Sediment-laden waters
6	Balanced effects of optically active constituents at shorter wavelength
7	Highly productive waters with high cyanobacteria abundance and elevated reflectance at red/near-infrared spectral region
8	Productive waters with cyanobacteria presence and with R_{rs} peak close to 700 nm
9	Optically neighboring to OWT2 waters but with higher R_{rs} at shorter wavelengths
10	CDOM-rich waters
11	Waters high in CDOM with cyanobacteria presence and high absorption efficiency by NAP
12	Turbid, moderately productive waters with cyanobacteria presence
13	Very clear blue waters

However, the CCI Lake dataset does not include information about the OWT of a pixel. A list of the algorithms used depending on the OWT is presented in Table 8 and 9.

Table 8: Chlorophyll-A algorithms per sensor and Optical Water Type (OWT) (Carrea et al., 2023)

Sensor	OWT	Wavelengths [μm]	Algorithm Reference
MERIS / OLCI	3, 9, 10, 13	0.490 , 0.560	OC2 (J. E. O'Reilly et al., 1998)
	2, 8, 11, 12	0.709, 0.665	709/665 empirical band ratio (Gileron et al., 2010).
	1, 4, 5, 6	0.709, 0.665	Semi-analytical NIR-Red band algorithm for MERIS (Gons et al., 2005)
	7	0.665	Adapted QAA algorithm (Mishra et al., 2013)
MODIS	1, 5, 7, 9, 12, 13	0.490, 0.560	OC2M (J. E. O'Reilly et al., 1998)
	2, 3, 8	0.443, 0.490, 0.560	OC3M (J. E. O'Reilly et al., 1998)
	4, 6, 11	0.667, 0.748	748/667 empirical band ratio (Dall'Olmo et al., 2005)
	10	0.469, 0.560	OC2-HI (J. E. O'Reilly et al., 1998)

Table 9: Turbidity algorithms per sensor and Optical Water Type (OWT)(Carrea et al., 2023)

Sensor	OWT	Wavelengths [μm]	Algorithm Reference
MERIS / OLCI	1, 7, 10	0.754	Y. Zhang et al., 2014
	2, 4, 6, 8, 12	0.665	Based on Vantrepotte et al., 2011.
	3, 5, 9, 11, 13	0.709	Binding et al., 2010
MODIS	1, 12	0.645	Miller and McKee, 2004
	2, 6, 11, 13	0.645	Ondrusek et al., 2012
	3, 5, 9	0.645	Chen et al., 2007
	4, 8, 10	0.645	Petus et al., 2010
	7	0.488, 0.555, 0.645	M. Zhang et al., 2010

The applicability of each selected algorithm for ChloA and turbidity to global lake observations, corresponding to a specific subset of OWTs, was established. This was achieved through the utilization of *in situ* observation data from LIMNADES (Lake Bio-optical Measurements and Matchup Data for Remote Sensing) to fine-tune algorithm coefficients, define their applicable concentration range, and characterize their respective uncertainty (Carrea et al., 2023).

3.2.3 Flow of the Ruzizi and Malagarasi rivers

Wongchuig et al., 2023 presents an improved 40-year (1981–2020) hydrological reanalysis of daily Congo River Basin (CRB) discharge. It analyzes the spatiotemporal dynamics of recent major floods and droughts within the basin, as well as their teleconnections with the El Niño–Southern Oscillation (ENSO), the dominant driver of tropical precipitation.

To conduct this analysis, a large-scale hydrologic-hydrodynamic model (MGB) was employed, which incorporates lake storage dynamics representation. Additionally, a data assimilation (DA) technique was used, integrating both *in situ* and remote sensing observations to enhance the accuracy of the hydrological reanalysis.

The article presents daily flow data for the Ruzizi and Malagarasi rivers spanning from 1981 to 2020. This dataset was utilized to juxtapose the daily median concentrations of ChloA and turbidity with the river flow of the corresponding day. Additionally, a monthly median of the river flow from 2013 to 2020 was established to compare this pattern with the intra-annual dynamic of ChloA and turbidity in the estuarine areas.

3.3 Region of interest

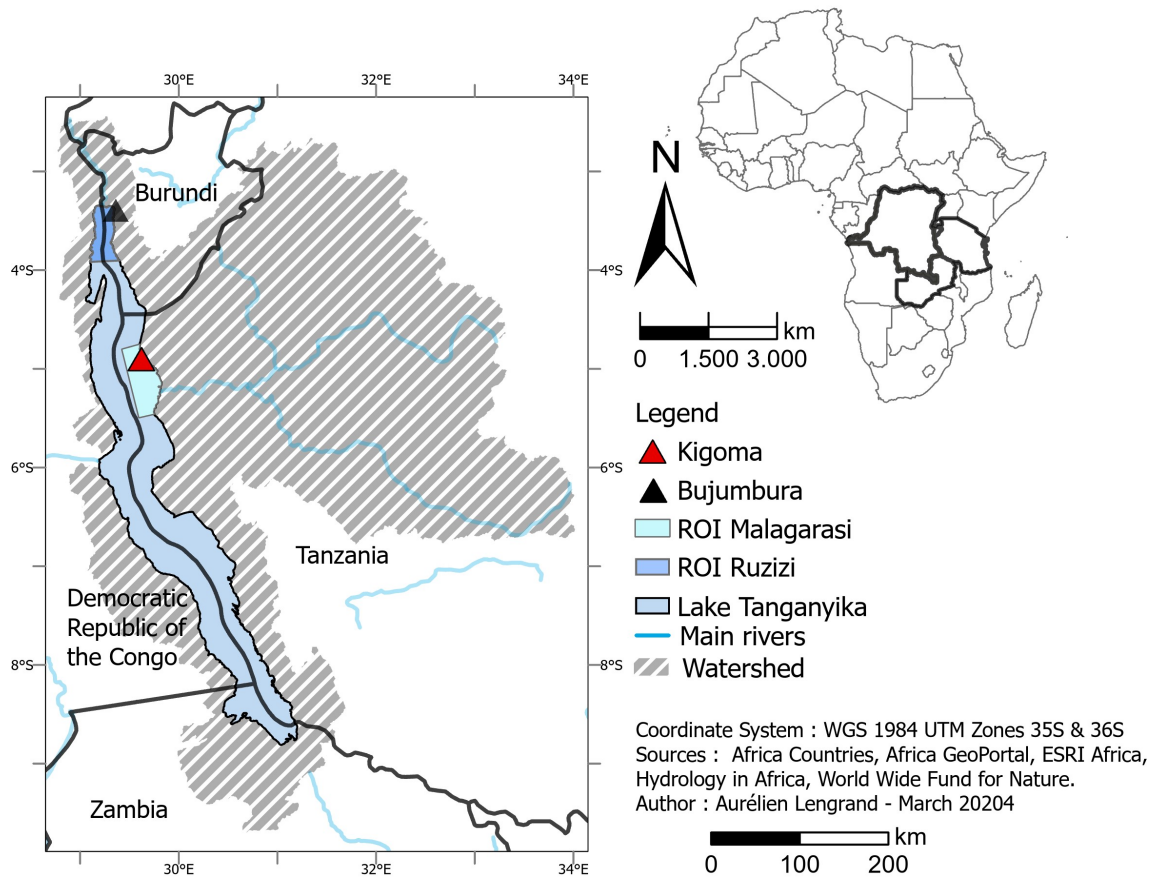


Figure 13: Study Area map

In this study, two estuarine areas will be examined: the Ruzizi river mouth region near the city of Bujumbura (defined by ROI Ruzizi), and the Malagarasi river mouth region near the city of Kigoma (defined by ROI Malagarasi).

3.3.1 Bujumbura

The Bujumbura region is characterized by the Ruzizi River and the coastal area of the city of Bujumbura. These two regions exhibit high concentrations of ChloA and turbidity, emphasizing the need for spatial visualization of those areas as presented in Figure 14.

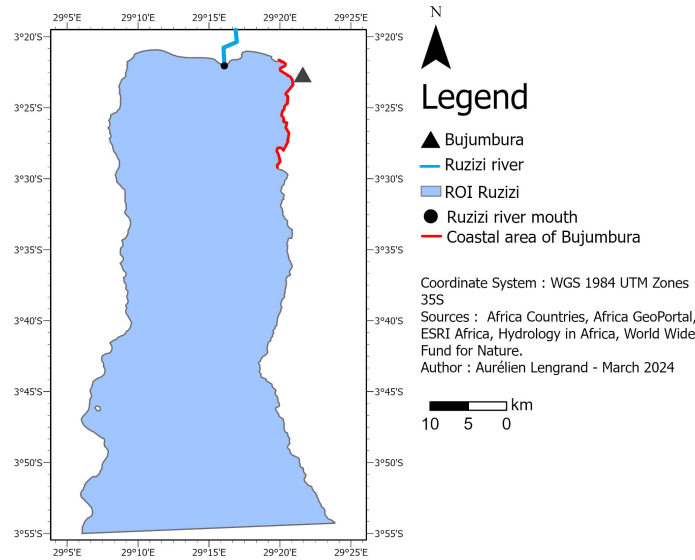


Figure 14: ROI Ruzizi map

This region is also influenced by four rivers crossing the city of Bujumbura, as depicted in the Appendix section [A.3](#).

Vandelannoote et al., [1999](#) presents the flow of the Ruzizi river (recorded at Gatumba) along with monthly precipitation data recorded at Bujumbura airport from 1982 to 1989. The river flow exhibits a peak during April/May and a minimum during September. The monthly precipitation shows a slightly different pattern with a peak in April and a minimum in July.

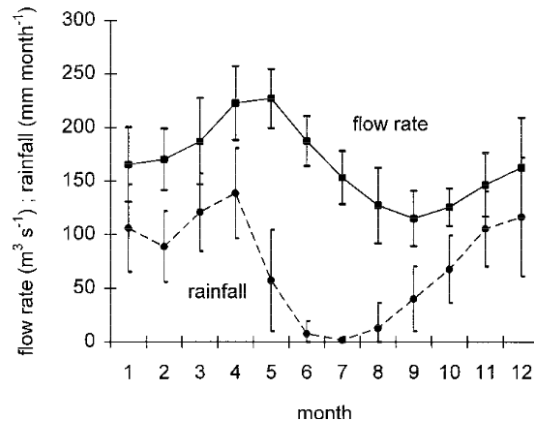


Figure 15: Mean annual fluctuations in the flow rate of River Rusizi at Gatumba and in rainfall at Bujumbura airport with standard deviation (unpublished IGEU data from 1982 to 1989) (Vandelannoote et al., [1999](#)).

3.3.2 Kigoma

The Kigoma region is defined by the mouth of the Malagarasi River and the coastal area of the city of Kigoma. Similar to the Bujumbura Region, these two areas demonstrate elevated concentrations of ChloA and turbidity. This underscores the importance of spatial visualization for accurately depicting these regions, as illustrated in Figure [16](#).

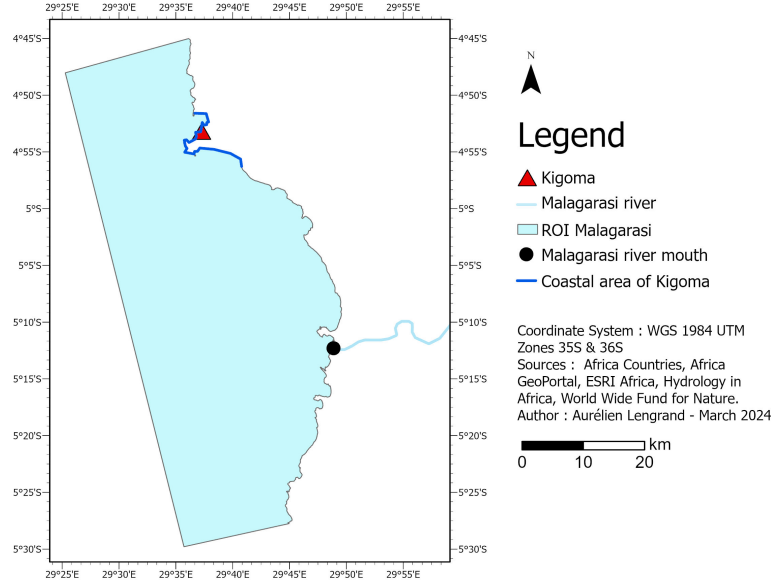


Figure 16: ROI Malagarasi map

Huttula, [1997] offers an overview of the discharge flow of the Malagarasi River into Lake Tanganyika, utilizing *in situ* data collected from 1975 to 1980. This flow pattern is depicted in Figure [17]. The flow exhibits a sinusoidal shape, with peak values occurring in May (approximately $370 \text{ m}^3/\text{s}$) and reaching a minimum in October (around $25 \text{ m}^3/\text{s}$).

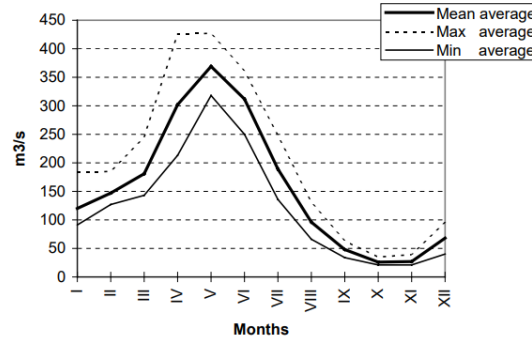


Figure 17: Discharge of the Malagarasi river in 1975-80 (Huttula, [1997]).

Gao et al., [2015] presents precipitation data recorded in Kigoma from 1998 to 2012 (obtained from the Tanzania Meteorological Agency in Kigoma). The monthly precipitation shows similarities to that recorded in Bujumbura, with a peak in March and a minimum in July.

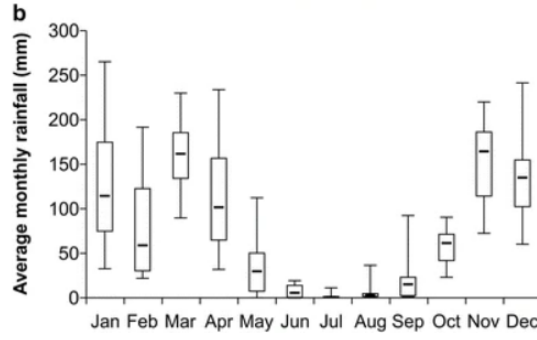


Figure 18: Monthly rainfall from 1998 to 2012 (data were obtained from the Tanzania Meteorological Agency in Kigoma (Gao et al., 2015)).

3.4 Downscaling approach

In the context of remote sensing, downscaling refers to the process of reducing the pixel size or spatial resolution of remotely sensed images (Atkinson, 2013). In this case, the objective is to reduce the spatial resolution of the CCI Lakes ChloA and turbidity dataset through the correlation between those variables and the R_{rs} captured by Landsat PAR. Data harmonization is necessary to study this relationship.

3.4.1 Data harmonization

Data harmonization involves the process of reconciling diverse types, levels, and sources of data into compatible and comparable formats, facilitating more effective decision-making and analysis (Cheng et al., 2024). The objective here is to make the Landsat PAR dataset comparable with the CCI Lake dataset. In this context, the objective is to ensure comparability between the Landsat PAR dataset and the CCI Lake dataset, which involves aligning their temporal resolution, spatial resolution, projection, and origin.

3.4.1.1 Matching dates

The initial step in data harmonization involved meticulously selecting images from both the CCI Lakes and Landsat PAR datasets, ensuring they were captured on the same day. This selection, based on thorough metadata analysis, yielded 113 images taken between 2016 and 2020 within the specified Region of Interest.

3.4.1.2 Reprojection

Reprojection involves altering the projection of an image from one coordinate reference system (CRS), also known as spatial reference system (SRS), to another.

A coordinate reference system delineates the position of a point on a flat or spherical surface. A common example is a graph with x and y axes. A geographic coordinate system is a three-dimensional reference system pinpointing Earth's surface points using projections. Projections are techniques used to represent a spherical surface on a flat one, facilitating the overlay of Cartesian coordinate systems ("Reprojection", n.d.).

In this study, Landsat PAR data were reprojected from *EPSG* : 32636 – *WGS84* to *EPSG* : 4326 – *WGS84* to align with the projection used for CCI Lakes data. This step was proceeded using the *projectRaster* function with the bilinear method from the R package *raster*.

3.4.1.3 Resampling

The process of transforming a sampled image from one coordinate structure to another is referred to as image resampling (Patil, 2018). This step is crucial for aligning two datasets that have pixels with varying sizes, positions, and other characteristics. Through resampling, these datasets are transformed into versions where the pixels precisely correspond in spatial locations.

Three different resampling methods are commonly used: Nearest-neighbor resampling involves assigning the value of the closest pixel in the original image to the new pixel location in the adjusted image, maintaining original values but potentially leading to blocky appearances and position errors. The bilinear method utilizes bilinear interpolation to calculate new pixel values by averaging the values of the four nearest pixels in the original image, resulting in smoother images but potentially altering original pixel values. Cubic convolution calculates image gray levels by averaging the values of the 16 nearest pixels to the input coordinates, yielding smoother results compared to nearest-neighbor interpolation and slightly superior results to bilinear interpolation. However, it requires approximately 10 times more computational time than the nearest neighbor method. This technique computes a distance-weighted average of a block of sixteen pixels from the original image surrounding the new output pixel location (Baboo and Devi, 2010).

In this study, the bilinear method was chosen due to its balance between visual quality, computational efficiency, and implementation simplicity. It provided improved visual quality compared to simpler methods like the nearest-neighbor method, reducing blocky artifacts and producing smoother images. While more complex techniques such as cubic convolution may offer slightly better results, they come with increased computational costs. This process was made with the *resample* function from the R package *raster*

3.4.1.4 Daily mosaic

Image mosaicing is an efficient method for creating a unified and continuous image by aligning multiple partially overlapping images (Pandey and Pati, 2019). This method efficiently combines images captured on the same day by aligning and merging them into a single cohesive and continuous image. This consolidation simplifies manipulation and facilitates comparison with the CCI Lakes dataset. This step was made using the *merge* function from the Python package *rasterio*.

3.4.2 Chlorophyll-A and Turbidity Algorithms selection

This section aims to define the optimal algorithms for retrieving ChloA and Turbidity concentrations from the CCI Lakes dataset using Landsat PAR R_{rs} . To achieve this goal, various

algorithms were optimized through the application of linear regression techniques.

3.4.2.1 Turbidity

As hypothesized in Hannouche et al., 2011 and utilized in the CCI Lakes product (Carrea et al., 2023), TSM algorithms were employed to retrieve turbidity. As described in Section 2.3.1 (see Table 1) and Section 3.2.2 (see Table 9), a range of algorithms are employed for the retrieval of turbidity from R_{rs} . However, our study is bound by the bands available in the Landsat PAR dataset (see Table 3). Given these constraints, we tested seven distinct algorithms, for which the equations are presented in Table 10.

Table 10: Turbidity algorithms tested

Equation	Reference
$Turb = a \ NDTI + b$	Lacaux et al., 2007
$Turb = a \ TSMi + b$	A. G. Dekker et al., 2002
$Turb = a \ B4 + n$	Miller and McKee, 2004
$Turb = a \ B4^3 + b \ B4^2 + c \ B4 + d$	Ondrusek et al., 2012
$Turb = a \ B4^b$	Chen et al., 2007
$Turb = a \ B4^2 + b \ B4 + c$	Petus et al., 2010
$\log_{10}(Turb) = a \ (B3 + B4) + b \ (\frac{B2}{B3}) + c$	M. Zhang et al., 2010

3.4.2.2 Chlorophyll-A

As delineated in Section 2.3.2 (see Table 2) and Section 3.2.2 (see Table 8), a variety of algorithms are employed to retrieve ChloA from R_{rs} . As explained with the selection of turbidity algorithms, our study is bound by the bands available in the Landsat PAR dataset (see Table 3). Given these constraints, we tested two distinct algorithms, for which the equations are presented in Table 11.

Table 11: Chlorophyll-A algorithms tested

Name	Equation	Reference
OC2	$\log_{10}(\text{ChloA}) = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + a_4 x^4$, $x = \log_{10} \left(\frac{Rsr_{482}}{Rrs_{561}} \right)$	J. E. O'Reilly et al., 1998
OC3	$\log_{10}(\text{ChloA}) = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + a_4 x^4$, $x = \log_{10} \left(\frac{\max[Rsr_{443}, Rsr_{482}]}{Rrs_{561}} \right)$	J. E. O'Reilly et al., 1998

3.4.3 Linear regression method

Linear regression is utilized to explore the linear correlation between a dependent variable Y and one or more independent variables X. Before conducting linear regression, it's crucial to assess the relationship between variables using a scatter plot visually. If the relationship appears linear, linear regression may be suitable. However, if the relationship appears nonlinear, alternative methods should be considered for studying the relationship (Schneider et al., 2010). In this

study, linear regression was employed for two distinct purposes: first, to analyze the correlation between Landsat PAR and CCI Lakes R_{rs} ; second, to optimize ChloA and turbidity algorithms based on PAR and CCI Lakes ChloA and turbidity concentrations. Those regressions were evaluated regarding three different indicators :

- **R^2 :**

Let

- n be the number of observations
- $\hat{y}_i$ be the estimated value of the dependent variable for the i^{th} observation, as computed with the regression equation
- y_i be the observed value of the dependent variable for the i^{th} observation
- $\bar{y}$ be the mean of all n observations of the dependent variable

The coefficient of determination is then defined as follows:

$$R^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} = \frac{\text{explained variance}}{\text{total variance}} \quad (15)$$

R^2 is the fraction of the overall variance that is explained. The closer the regression model's estimated values $\hat{y}_i$ lie to the observed values y_i , the nearer the coefficient of determination is to 1 and the more accurate the regression model is (Schneider et al., 2010).

- **RMSE and MAE:**

The root mean square error (RMSE) and the Mean Absolute Error (MAE) are commonly employed as a standard statistical metric for assessing model performance (Chai and Draxler, 2014). They are calculated as :

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{n}} \quad (16)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (17)$$

where

- $\hat{y}_1, \hat{y}_2, \dots, \hat{y}_n$ are predicted values
- $y_1, y_2, \dots, y_n$ are observed values
- n is the number of observations

RMSE is the square root of the average squared differences between predicted and actual values, making it sensitive to large errors and suitable for applications where larger errors

are undesirable. MAE, on the other hand, is the average of the absolute differences between predicted and actual values, providing a straightforward measure that is less affected by outliers (Chai and Draxler, 2014).

3.5 Semi-Empirical algorithm and Analytical index

The second approach considered for building ChloA and turbidity datasets with fine spatial resolution involves using analytical indexes and pre-fitted semi-empirical algorithms. However, it's important to exercise caution with data obtained through this method due to the unknown uncertainty.

One already-fitted Semi-Empirical algorithm for computing ChloA will be described in Section 3.5.1, and two Analytical Turbidity indexes will be described in Section 3.5.2.

3.5.1 OC3 algorithm for Landsat 8

As described in Section 2.3.2 the OCx Algorithm is used to compute ChloA concentration based on R_{rs} .

$$\log_{10}(\text{ChloA}) = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + a_4 x^4 \quad (18)$$

$$x = \log_{10} \left(\frac{\max [R_{sr443}, R_{sr482}]}{R_{rs561}} \right) \quad (19)$$

To use this algorithm, polynomial coefficients must be fitted for a specific satellite. The polynomial coefficients from Equation 7 have been fitted for 25 satellite instruments including Landsat 8 with 2720 *in situ* ChloA measurements with a concentration going from 0.012 to 77.9 mg/m³. The optimization for the Landsat 8 instrument gave an R² of 0.8466 (J. E. O'Reilly and Werdell, 2019). The values of those coefficients are given in Table 12.

Table 12: Coefficients of the OC3 algorithm for Landsat 8 (J. E. O'Reilly and Werdell, 2019)

a_0	a_1	a_2	a_3	a_4
0.30963	-2.40052	1.28932	0.52802	-1.33825

3.5.2 Normalized Difference Turbidity Index (NDTI) and Total Suspended Matter index (TSMi)

As stated in Section 2.3.1, the NDTI Algorithm is described by the following equation:

$$NDTI = \frac{R_{\text{red}} - R_{\text{green}}}{R_{\text{red}} + R_{\text{green}}} \quad (20)$$

This algorithm yields results ranging from -1 to 1, with -1 indicating clear water and 1 indicating highly turbid water.

The TSMi Algorithm is described by the following equation :

$$TSMi = \frac{R_{green} + R_{red}}{2} \quad (21)$$

Here, R_{red} represents the radiometric response in the red wavelength, while R_{green} denotes the radiometric response in the green wavelength.

This algorithm yields results ranging from -1 to 1, with -1 indicating clear water and 1 indicating highly turbid water.

Even though those two algorithms are based on the same principle: the increase in the Green and Red bands R_{rs} with an increase in Turbidity concentration (see 2.2.1.1) the NDTI algorithm is more suited to compute Turbidity concentration in very Turbid water (Lacaux et al., 2007). Given that Lake Tanganyika has clear waters and considering the misrepresentation of the lake's global turbidity dynamics by the NDTI (see section 4.2), the TSMi algorithm was preferred over the NDTI.

One hypothesis here is that the observations of the downscaling failure and the misrepresentation of the global turbidity dynamic by the NDTI are also true for the Kigoma region.

3.6 Time series analysis

Once the datasets are created, the following steps aim to determine the spatiotemporal variability of ChloA and turbidity in the estuarine areas of the Ruzizi and Malagarasi Rivers. The steps presented in this section were applied independently to each region.

3.6.1 Empirical Orthogonal Function (EOF)

Empirical Orthogonal Function (EOF) reduces a dataset containing a large number of variables to a dataset with fewer (hopefully much fewer) new variables. These new variables are linear combinations of the original ones, chosen to represent the maximum possible fraction of the variability contained in the original data (Wilks, 2006). This process was made thanks to the *irlba* function from the R package *irlba* (Baglama et al., 2022).

3.6.2 Kmean clustering

In this study, we utilized the dominant spatial modes of ChloA and turbidity obtained through the EOF method to cluster regions exhibiting similar dynamics in both variables. The K-means clustering function was employed to spatially delineate areas with comparable dynamics. This analysis involved grouping objects based on their differences using distance functions (Li and Wu, 2012).

The success of clustering depends on the desired number of clusters. While statistical methods could identify the optimal number of clusters, they were not used here due to memory

constraints stemming from large dataset sizes. Instead, the optimal number of clusters was determined through a trial-and-error approach.

For this analysis, we utilized the *kmeans* function from the R package *stats* (“kmeans function - RDocumentation”, [n.d.](#)).

3.6.3 Correlation analysis between ChloA and Turbidity dynamic in each cluster and the river flow

To comprehensively analyze the temporal dynamics of each cluster, median concentrations of ChloA and turbidity were calculated for each available image, resulting in a single daily concentration value for each cluster. These concentrations were then compared with the daily river flow to assess the influence of river flow on each cluster.

Subsequently, a monthly median for each year was constructed using the previously calculated daily medians. The monthly median representing intra-annual dynamics between 2013 and 2023 was derived by aggregating the monthly medians of each year. For example, the median for January was derived from the monthly median of January for each year between 2013 and 2023. This overall monthly median was visually compared with a monthly median of the river flow.

4 Results

4.1 Downscaling approach

This section presents the results corresponding to the downscaling approach in three subsections: the first subsection contains the results of the optimization of turbidity algorithms. In the second subsection, the results of the optimization of ChloA algorithms are presented. Finally, in the last subsections, we discuss the details of the approach.

4.1.1 Turbidity algorithms optimization

Different turbidity algorithms using Landsat PAR R_{rs} have been optimized based on the turbidity dataset provided by CCI Lakes from 2016 to 2020.

In Figure 19, the median map of the turbidity time series is analyzed showing that the global dynamic of the CCI Lakes turbidity dataset is represented by higher turbidity concentrations (around 4 NTU) near the coast and lower turbidity concentrations in the middle of the lake (around 1 NTU).

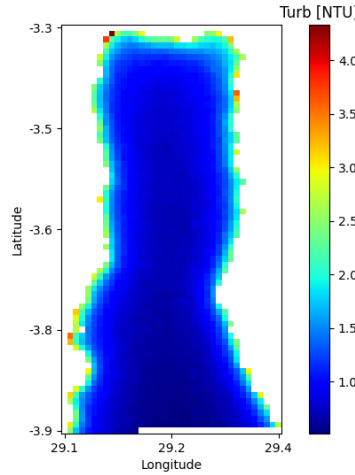


Figure 19: Median map of Turbidity from CCI Lakes between 2016 and 2020

The coefficients of the different Turbidity algorithms presented in Table 10 were optimized using the CCI Lakes turbidity dataset. The optimization results are presented in Table 13 and Figure 20.

The two different indices tested provide very different results. The NDTI shows an R^2 of 0.01, an RMSE of 0.8 NTU and a MAE of 0.5 NTU. As shown in Figure 20, there is no clear distinction between high and low turbidity values provided by CCI Lakes. This observation is also evident in the optimized equation, where the NDTI has a negligible and negative contribution (-0.228), and the intercept is 0.908. The TSMi shows better results with an R^2 of 0.12, an RMSE of 0.7 NTU, and a MAE of 0.4 NTU. The plot shows a scattering of observations close to the 0

value. There is no clear linear relationship between the CCI Lakes Turbidity dataset and the predicted observed value.

Three polynomial forms of the Green band have been tested. The optimization of all these forms shows very similar results, with an R^2 of around 0.04 for the first two polynomial forms and an R^2 of 0.07 for the third polynomial. RMSE and MAE are similar for the three forms with respectively a value of 0.8 NTU and 0.5 NTU. As for the TSMi, the plots show a scattering of observations close to the 0 value, with more predicted values close to 0.

The exponential form of the Green band shows results similar to NDTI, with an R^2 of 0.01, an RMSE of 0.8 NTU, and a MAE of 0.5 NTU. The plot shows no clear distinction between high and low turbidity values provided by CCI Lakes.

The last algorithm, using the Blue, Green, and Red bands, shows an R^2 of 0.08, an RMSE of 0.9 NTU, and a MAE of 0.5 NTU. The plot shows almost homogeneous predicted values for both high and low values of observed turbidity, with predictions ranging from 0 to 2.5 and observed values ranging from 0 to 15.

The same approach was applied, taking into account the uncertainty related to the CCI Lakes Turbidity dataset. The results of this analysis are presented in the Appendix section [A.4](#) and are very similar to the ones presented here.

Table 13: Optimization results of Turbidity algorithms described in Table [10](#) between CCI Lakes and Landsat PAR resampled at 1km. B2 represents R_{rs} of Landsat PAR at 490nm, B3 at 560 nm, and B4 at 665nm.

Optimized equation	R^2	RMSE	MAE
$Turb = -0.228 \text{ NDTI} + 0.908$	0.01	0.8	0.5
$Turb = 51.912 \text{ TSMi} + 0.617$	0.12	0.7	0.4
$Turb = 36.133 \text{ B4} + 0.847$	0.04	0.8	0.5
$Turb = 21.606 \text{ B4} + 8.665 \text{ B4}^2 + 1.009$	0.05	0.8	0.5
$Turb = 21.606 \text{ B4} + 8.665 \text{ B4}^2 - 15.982 \text{ B4}^3 + 1.009$	0.07	0.8	0.5
$Turb = 1.331 \text{ B4}^{0.054}$	0.01	0.8	0.5
$Turb = 10^{(5.908(B3+B4) - 0.041(\frac{B2}{B3}) - 0.107)}$	0.08	0.9	0.5

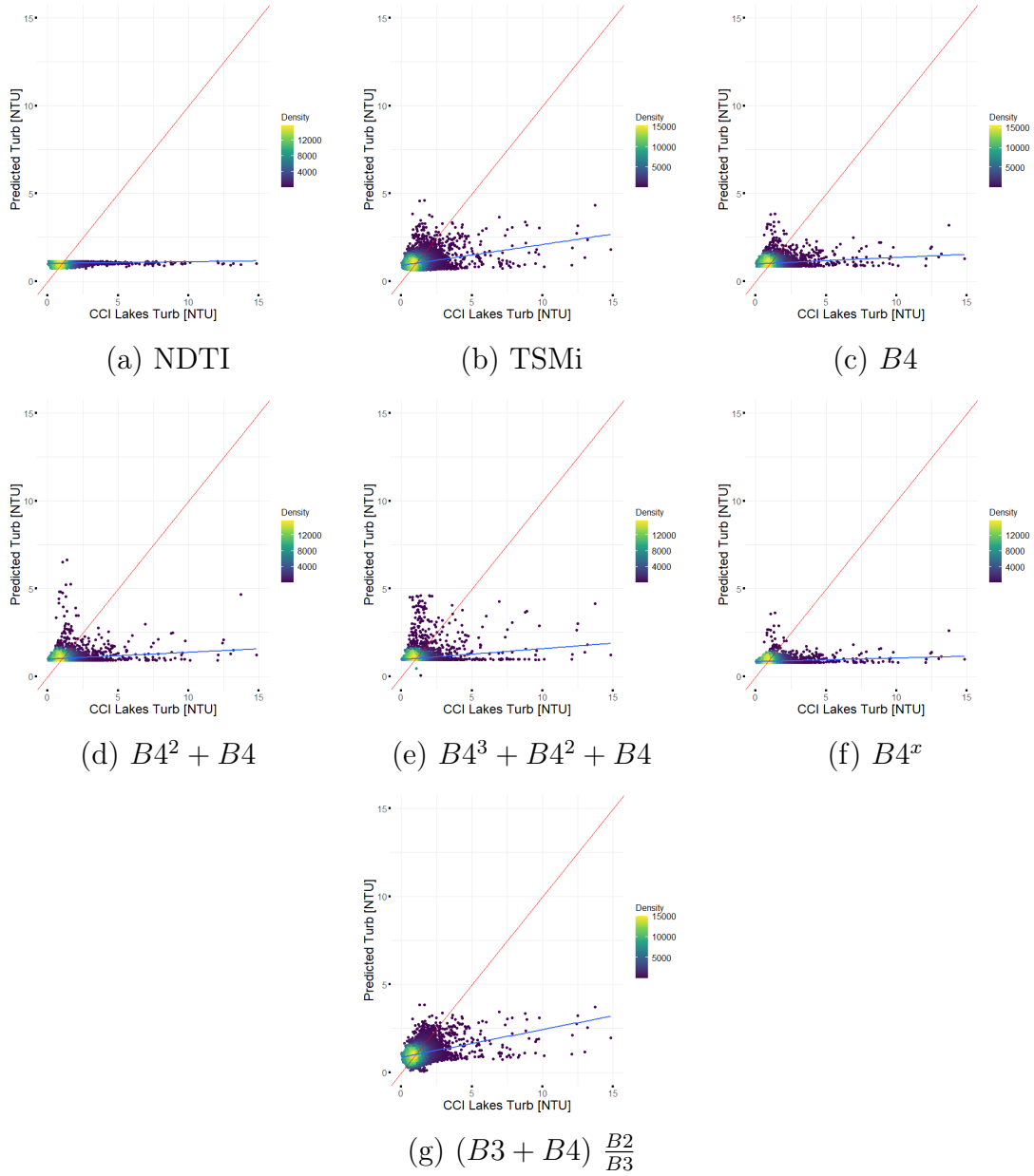


Figure 20: Linear regression analysis between CCI Lakes Turbidity and the turbidity obtained with Landsat PAR and the algorithms optimized (cf Table 13). Individual points are color-coded based on the number of neighboring points, with the red line representing the identity line. The different algorithms are (a) NDTI, (b) TSMi, (c) first polynomial of B4, (d) second polynomial of Band 4, (e) third polynomial of Band 4, (f) $B4^x$ and (g) $(B3 + B4) \frac{B2}{B3}$.

4.1.2 Chlorophyll-A algorithms optimization

Different ChloA algorithms using Landsat PAR R_{rs} have been optimized based on the ChloA dataset provided by CCI Lakes from 2016 to 2020. The median map of this time series is shown in Figure 21.

The global dynamic of the CCI Lakes ChloA dataset is represented by high concentrations near the coastal area of Bujumbura and the mouth of the Ruzizi River (around 8 $\mu\text{g/l}$). The rest of the lake is relatively homogeneous, with low concentration values of around 1 $\mu\text{g/l}$.

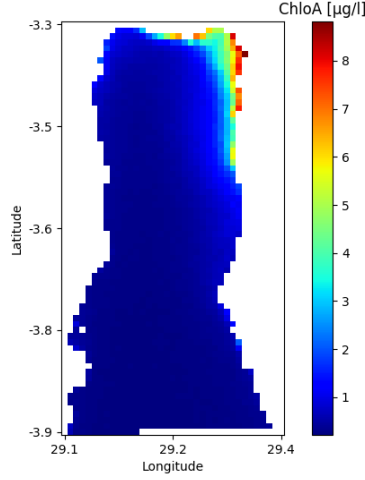


Figure 21: Median map of ChloA from CCI Lakes between 2016 and 2020

The coefficients of the two ChloA algorithms described in Table 11 were optimized using the CCI Lakes ChloA dataset. Additionally, the results obtained with the OC3 equation using coefficients optimized for Landsat 8 in J. E. O'Reilly and Werdell, 2019 were compared with the CCI Lakes ChloA dataset. The results are presented in Table 14 and Figure 22.

The OC2 and OC3 optimized algorithms show very similar results, with an R^2 of 0.02 for both algorithms, an RMSE of 7.9 $\mu\text{g/l}$, and an MAE of 1 $\mu\text{g/l}$. The plots also exhibit similar patterns, with observations scattered around the 0 value. Interestingly, despite the close similarity in results between the two algorithms, the optimized coefficients differ noticeably between them.

The significant difference between the RMSE and MAE indicates that, while the average predictions are relatively close to the actual values (as indicated by the low MAE), there are also considerable prediction errors that have a significant impact on the overall model performance (reflected in the high RMSE). This could indicate that there are predictions that are extremely far from the actual values, affecting the overall accuracy of the model despite a relatively low average error.

The OC3 algorithm, used with coefficients optimized by J. E. O'Reilly and Werdell, 2019 and compared with the CCI Lakes ChloA dataset shows an R^2 of 0.02, an RMSE of 7.9 $\mu\text{g/l}$, and an MAE of 1.3 $\mu\text{g/l}$. The plot indicates no differentiation in the predicted values between high and low observed ChloA.

It is interesting to note that the results between the OC2/3 algorithms and the OC3 algorithm used with the O'Reilly coefficients are very similar in terms of R^2 , RMSE, and MAE, even though their plots differ significantly.

The same approach was applied, taking into account the uncertainty related to the CCI Lakes ChloA dataset. The results of this analysis are presented in the Appendix section A.4 and are very similar to the ones presented here.

Table 14: Optimization results of Chlorophyll-A algorithms described in Table 11 between CCI Lakes and Landsat PAR resampled at 1km.

Optimized equation	R ²	RMSE	MAE
$ChloA = 10^{(-31.212 x - 3.348 x^2 + 15.255 x^3 + 2.269 x^4 - 0.280)}$, $x = \log_{10} \left(\frac{Rrs_{482}}{Rrs_{561}} \right)$	0.02	7.9	1.0
$ChloA = 10^{(-26.356 x + 22.943 x^2 - 2.642 x^3 - 11.721 x^4 - 0.280)}$, $x = \log_{10} \left(\frac{\max[Rrs_{443}, Rrs_{482}]}{Rrs_{561}} \right)$	0.02	7.9	1.0
$ChloA = 10^{(OC3_{O'Reilly})}$	0.02	7.9	1.3

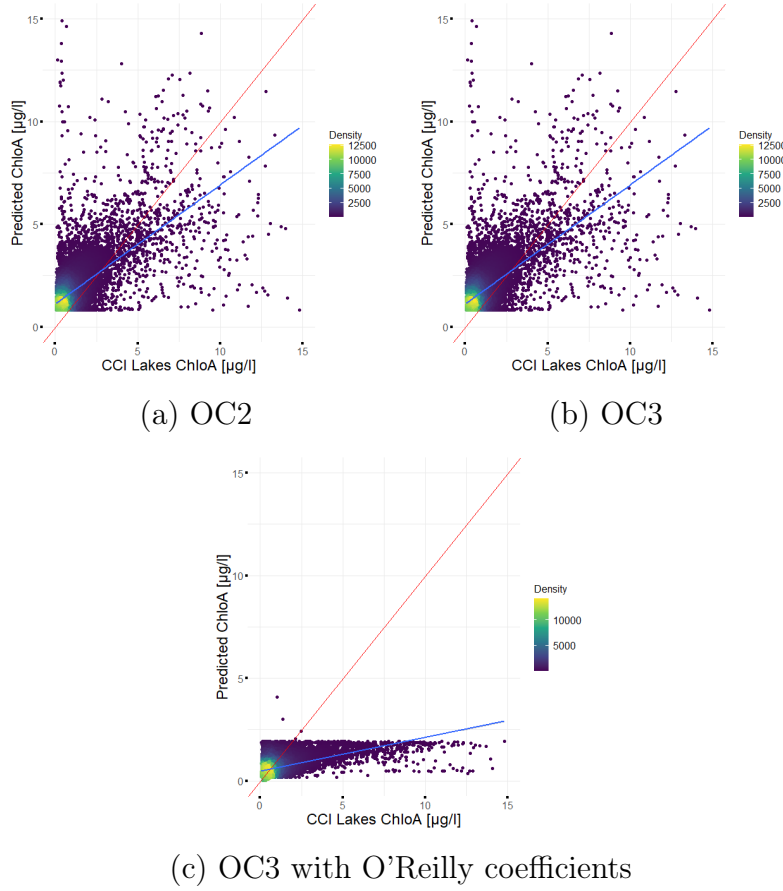


Figure 22: Linear regression analysis between CCI Lakes ChloA and the ChloA obtained with Landsat PAR and the algorithms optimized (cf Table 14). Individual points are color-coded based on the number of neighboring points, with the red line representing the identity line. The different algorithms are : (a) OC2 equation optimized, (b) OC3 equation optimized and (c) OC3 equation with coefficients fitted by J. E. O'Reilly and Werdell, 2019.

4.1.3 Investigating the downscaling failure

This section aims to investigate the reasons for the non-conclusive results observed in the optimization of Turbidity and ChloA algorithms. The first part will focus on the band R_{rs} comparison between Landsat PAR and CCI Lakes, and the second part will investigate the differences in the turbidity dynamic obtained when using different algorithms.

4.1.3.1 Band comparison

Figure 23 and Figure 24 illustrate the comparison of the median and the pixel-by-pixel linear regression between Landsat PAR (1 km) R_{rs} and CCI Lakes R_{rs} for four different wavelengths. The first band corresponds to Ultra Blue, the second to Blue, the third to Green, and the fourth to Red (see Table 3 and Table 6 for the respective wavelengths). This analysis used data from 2016 to 2020 in the Bujumbura ROI. To maintain the valid range of R_{rs} as advised by the USGS (cf Table 3), negative pixel values and pixels with values greater than 0.1 were masked.

The median R_{rs} maps between the two products depicted in Figure 23 reveal notable differences. The Ultra Blue band R_{rs} in the CCI Lakes product exhibits low values in the middle of the lake (around 0.0025) and higher values near the coast (around 0.015). This pattern is particularly noticeable near Bujumbura, with higher values (around 0.02) farther offshore than the rest of the lake.

On the other hand, the Landsat PAR product shows a different pattern, with higher values in the middle of the lake (around 0.015) and lower values along the coastal areas (around 0.0025). Similarly, higher values are also observed near the city of Bujumbura.

For CCI Lakes, the observations show similarities between the Ultra Blue band and the Blue band with low R_{rs} in the middle of the lake (around 0.0025) and high values next to Bujumbura (around 0.02). The Landsat PAR product, like the Ultra Blue band, shows high values (around 0.015) in the middle of the lake and low values in the coastal areas (around 0.0025). Higher values are also present next to Bujumbura with values around 0.02.

The Green band presents more matching median maps between CCI Lakes and Landsat PAR. Both results show low values (around 0.01) in the middle of the lake and high values next to Bujumbura (around 0.03 with peak values at 0.05). It's interesting to note that the Landsat PAR product shows less homogeneous values in the middle of the lake than the CCI Lakes product.

Finally, the observations for the Red band are similar to the ones for the Green band with low values (around 0.005) in the middle of the lake and high values (around 0.025) next to Bujumbura. The homogeneity difference in the middle of the lake between CCI Lakes and Landsat PAR is also present. The main difference between the Green and the Red band is the extent of the high values pattern next to Bujumbura with a less extensive pattern in the coastal area for the Red band.

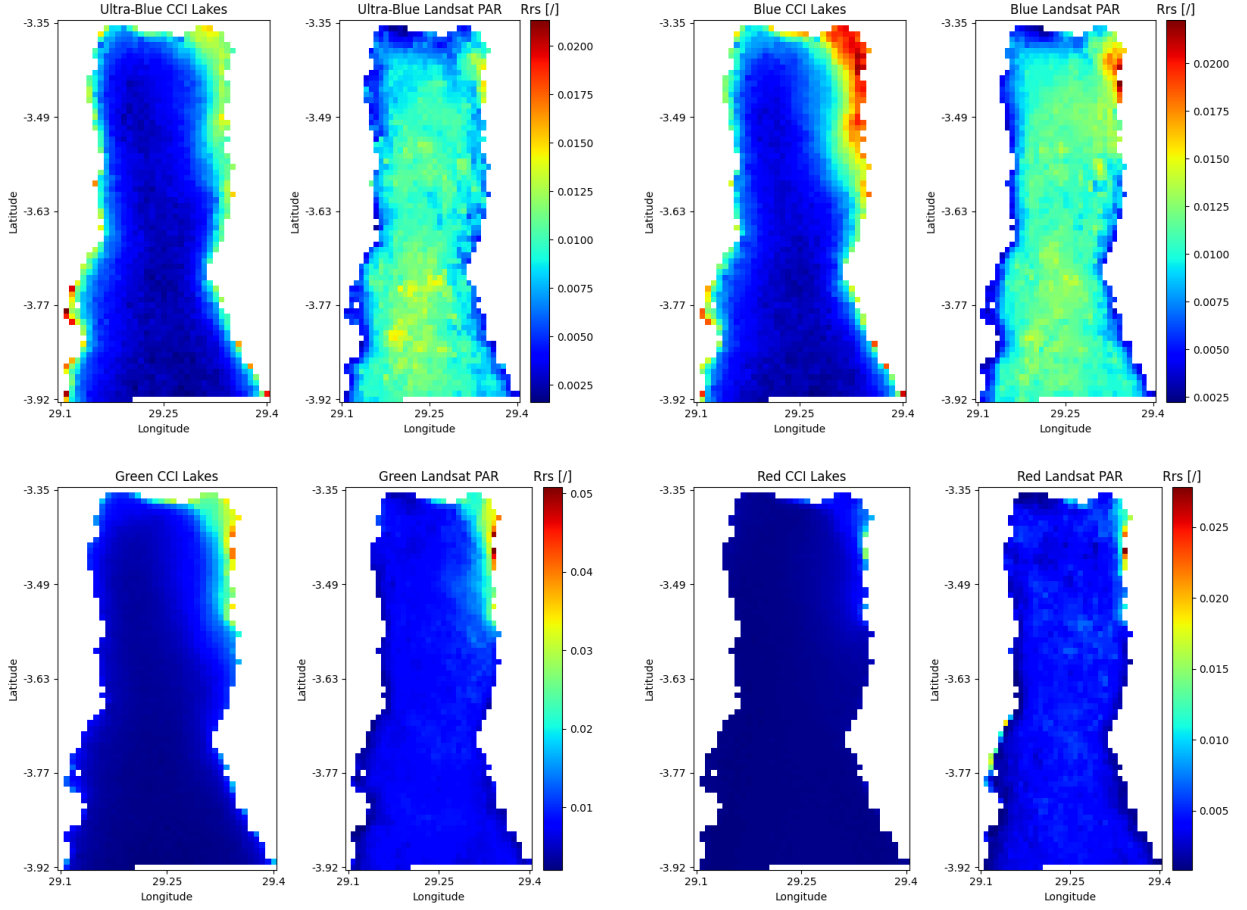


Figure 23: Comparison of the median R_{rs} conducted between CCI Lakes and Landsat PAR (1 km) for four different wavelengths between 2016 and 2020 : Ultra-Blue corresponds to CCI Lakes $R_{rs}(443)$ and Landsat PAR $R_{rs}(443)$, Blue corresponds to CCI Lakes $R_{rs}(490)$ and Landsat PAR $R_{rs}(482)$, Green corresponds to CCI Lakes $R_{rs}(560)$ and Landsat PAR $R_{rs}(561.5)$ and Red corresponds to CCI Lakes $R_{rs}(665)$ and Landsat PAR $R_{rs}(654.5)$.

To highlight the differences between CCI Lakes and Landsat PAR, pixel-by-pixel linear regressions between the two products for four bands are presented in Figure 24.

The linear relationships of the Ultra Blue and Blue bands indicate no correlation ($R^2 < 0.005$) between Landsat PAR (1 km) R_{rs} and CCI Lakes R_{rs} . On the other hand, the linear regressions for the Green and Red bands show better correlations (R^2 of 0.26 and 0.13, respectively). However, these results need to be balanced by the reduced number of observations for the Red band ($N = 13\,333$) compared to the Ultra Blue ($N = 22\,850$), Blue ($N = 26\,338$), and Green ($N = 26\,315$) bands.

The RMSE values for the Ultra Blue and Blue bands are 0.01 and 0.009, respectively, indicating a consistent error between the CCI Lakes and Landsat PAR products. The RMSE values of the Green and Red Bands are 0.06 and 0.04, respectively. The MAE values are very similar to the RMSE values, with values of 0.007, 0.007, 0.006, and 0.004, respectively.

Those results are consistent with the observations derived from the R_{rs} median map comparison, showing more similarities between CCI Lakes and Landsat PAR for the Green and Red bands

than for the Ultra Blue and Blue bands.

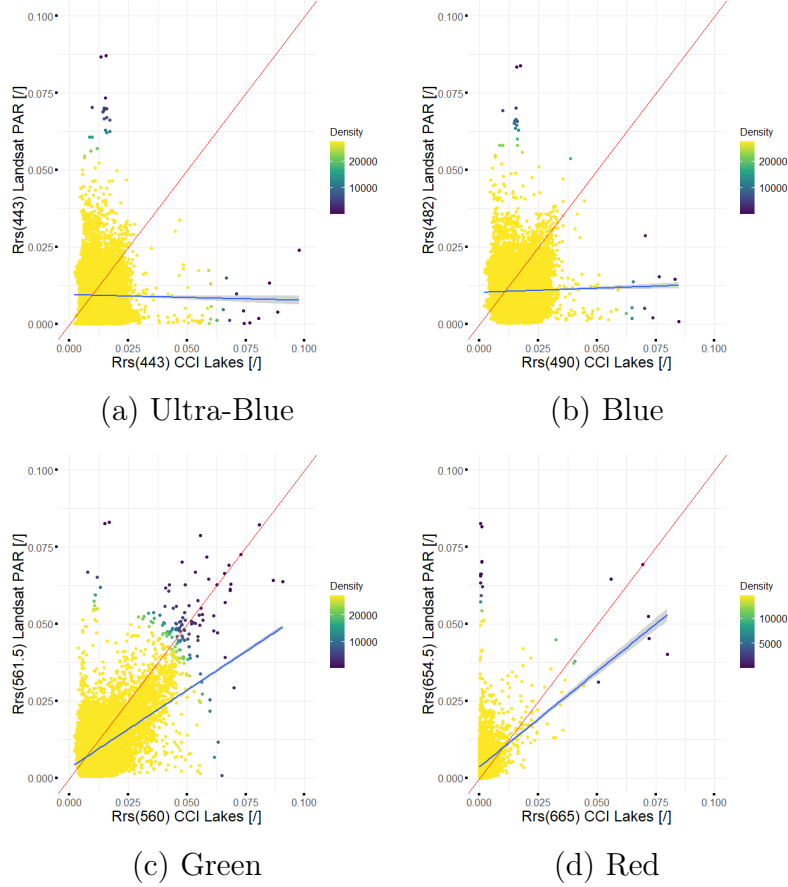


Figure 24: Linear regression analysis between CCI Lakes and Landsat PAR for different R_{rs} . Individual points are color-coded based on the number of neighboring points, with the red line representing the identity line. The bands presented are (a) CCI Lakes $R_{rs}(443)$ and Landsat PAR $R_{rs}(443)$, (b) CCI Lakes $R_{rs}(490)$ and Landsat PAR $R_{rs}(482)$, (c) CCI Lakes $R_{rs}(560)$ and Landsat PAR $R_{rs}(561.5)$ and (d) CCI Lakes $R_{rs}(665)$ and Landsat PAR $R_{rs}(654.5)$.

Table 15: Linear regression results of Landsat PAR R_{rs} resampled at 1km and CCI Lakes R_{rs} for four bands: Ultra-Blue, Blue, Green and Red

Band	R^2	RMSE	MAE
Ultra-Blue	0.00	0.010	0.007
Blue	0.00	0.009	0.007
Green	0.26	0.007	0.006
Red	0.13	0.005	0.004

4.1.3.2 Impact of the band used on the turbidity computed

Figure 25 presents the R_{rs} of three CCI Lakes bands obtained on 10-05-2020. The turbidity concentration image of the same day is also presented, as well as the TSMi constructed using the 560 nm and 665 nm bands from CCI Lakes. The most significant difference observed between the 754 nm and 665 nm bands is the increase in R_{rs} in the coastal areas. The turbidity concentration image provided by CCI Lakes appears to be similar to the 754 nm band, which

moderates the high concentration pattern observed near the city of Bujumbura and results in higher values in this coastal area. Depending on the algorithm used, the turbidity dynamics can vary significantly. The fact that no band around 750 nm is available on the Landsat PAR product thus prevents the use of similar algorithms and can explain part of the downscaling failure observed for turbidity.

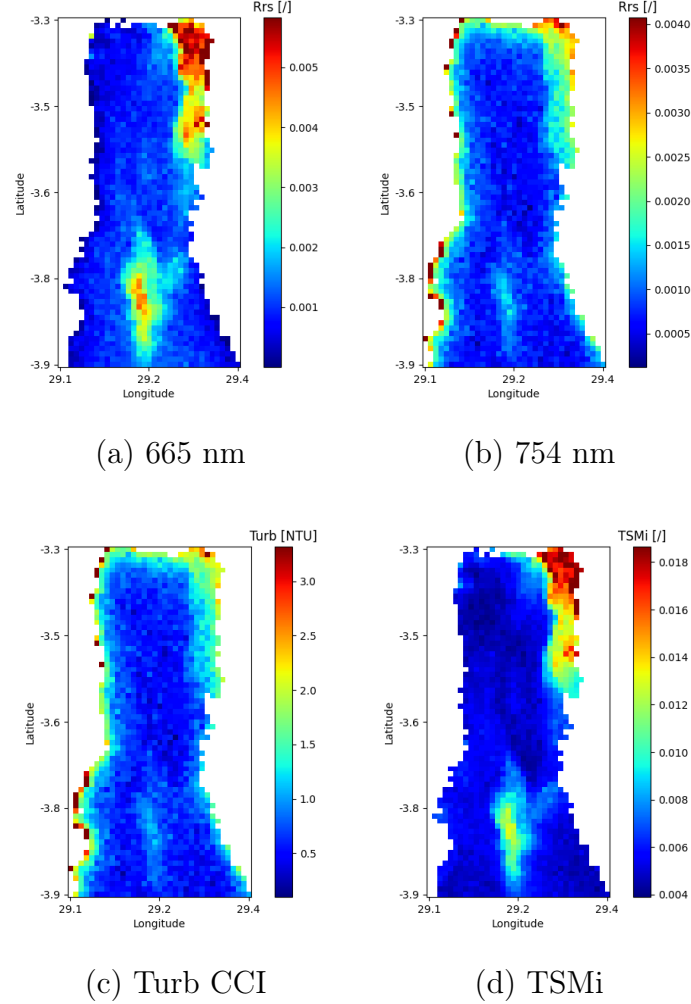


Figure 25: Representation of the different CCI Lakes bands acquired the 10/05/2020. (a) R_{rs} (665 nm), (b) R_{rs} (754 nm), (c) Turbidity concentration (d) TSMi retrieved using CCI Lakes R_{rs} (560nm) and R_{rs} (665nm).

4.2 Semi-Empirical algorithm and Analytical indexes

To overcome the downscaling failure, one approach is to use analytical indexes and an already-fitted Semi-Empirical algorithm. This section presents the median of the datasets obtained when applying the OC3 algorithm as presented in Section 3.5.1 to the Landsat PAR R_{rs} from 2016 to 2020, along with two indices aiming to retrieve the turbidity dynamics as presented in Section 3.5.2: NDTI and TSMi. This approach allows us to evaluate if the global dynamic of the CCI Lakes dataset is well represented for each created time series.

Figure 26 presents the turbidity dynamic obtained with the NDTI. The figure shows slightly

higher values in the middle of the lake and lower values in the coastal areas. One exception is the coastal area of Bujumbura, where the values are similar to those in the middle of the lake. However, this dynamic does not represent the turbidity dynamic of the CCI Lakes dataset (see Figure 19), which shows higher concentrations in the coastal areas and lower concentrations in the middle of the lake.

On the other hand, the turbidity dynamic obtained with the TSMi shows low values in the middle and coastal areas of the lake and higher values next to the city of Bujumbura. Compared to the CCI Lakes turbidity dataset, the middle of the lake is well represented. Higher values in the coastal areas are not present except for the coastal areas of Bujumbura.

In general, even if some patterns are not present in the TSMi dataset, this index seems more suited than the NDTI, with less contradictory dynamics. Therefore, this index will be used for turbidity time series with fine spatial resolution.

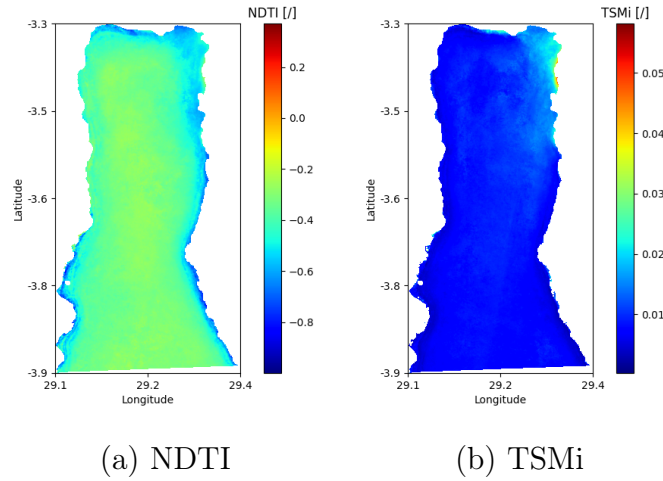


Figure 26: Median maps of the NDTI and TSMi time series using Landsat PAR between 2016 and 2020

Figure 27 presents the ChloA dynamic obtained with the OC3 algorithm used with optimized coefficients for Landsat 8 by J. E. O'Reilly and Werdell, 2019. The global dynamic of the created time series closely resembles that of the CCI Lakes ChloA dataset (see Figure 21). We observe relatively constant values in the middle of the lake, with higher ChloA concentrations near the city of Bujumbura and along the east and west coasts of the Ruzizi river mouth. However, there is a difference in the ChloA concentration range between the two time series. For the CCI Lakes dataset, ChloA concentrations range between 0 and 9, whereas for the created time series, ChloA concentrations range between 0 and 45.

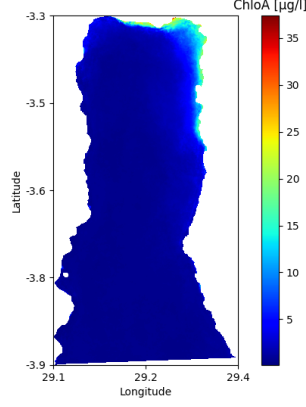


Figure 27: Median map of the ChloA Time Serie using Landsat PAR between 2016 to 2020

4.3 Chlorophyll-A and Turbidity Time Series analysis

ChloA and turbidity datasets were built using the OC3 equation with the optimized coefficients for Landsat 8 and the TSMi on Landsat PAR R_{rs} from 2013 to 2023. This analysis aimed to determine the spatiotemporal variability of ChloA and turbidity in the estuarine areas of the Ruzizi and Malagarasi Rivers. The analysis presented in this section was applied independently to each region.

It's important to note that the precision of the created datasets cannot be estimated due to the lack of *in situ* data and the observed failure in downscaling. Additionally, the TSMi values are not expressed in terms of turbidity concentration. One hypothesis made on the following results is that there is a linear relationship between the TSMi values and turbidity concentration. For ChloA, the hypothesis is that the dynamics captured by the OC3 algorithm with Landsat 8 fitted coefficients are similar to the ChloA dynamics of the lake. However, the retrieved ChloA concentration should not be interpreted as accurate but rather used to capture the overall dynamics.

4.3.1 Global dynamic of the datasets

This section will present the global dynamic of the time series created previously with the TSMi and the OC3 algorithm with coefficients fitted for Landsat 8 between 2013 and 2023 for the two ROI: Bujumbura and Kigoma.

4.3.1.1 Bujumbura

The monthly median maps of TSMi, depicted in Figure 28, underscore the disparities in observations between the wet season (from May to September) and the dry season (from October to April). This observation is further elucidated in Figure 47 in the Appendix section A.5, which quantifies the phenomenon. Higher TSMi values are noted during the wet season, characterized by a more diffuse plume of elevated values in the coastal area of Bujumbura.

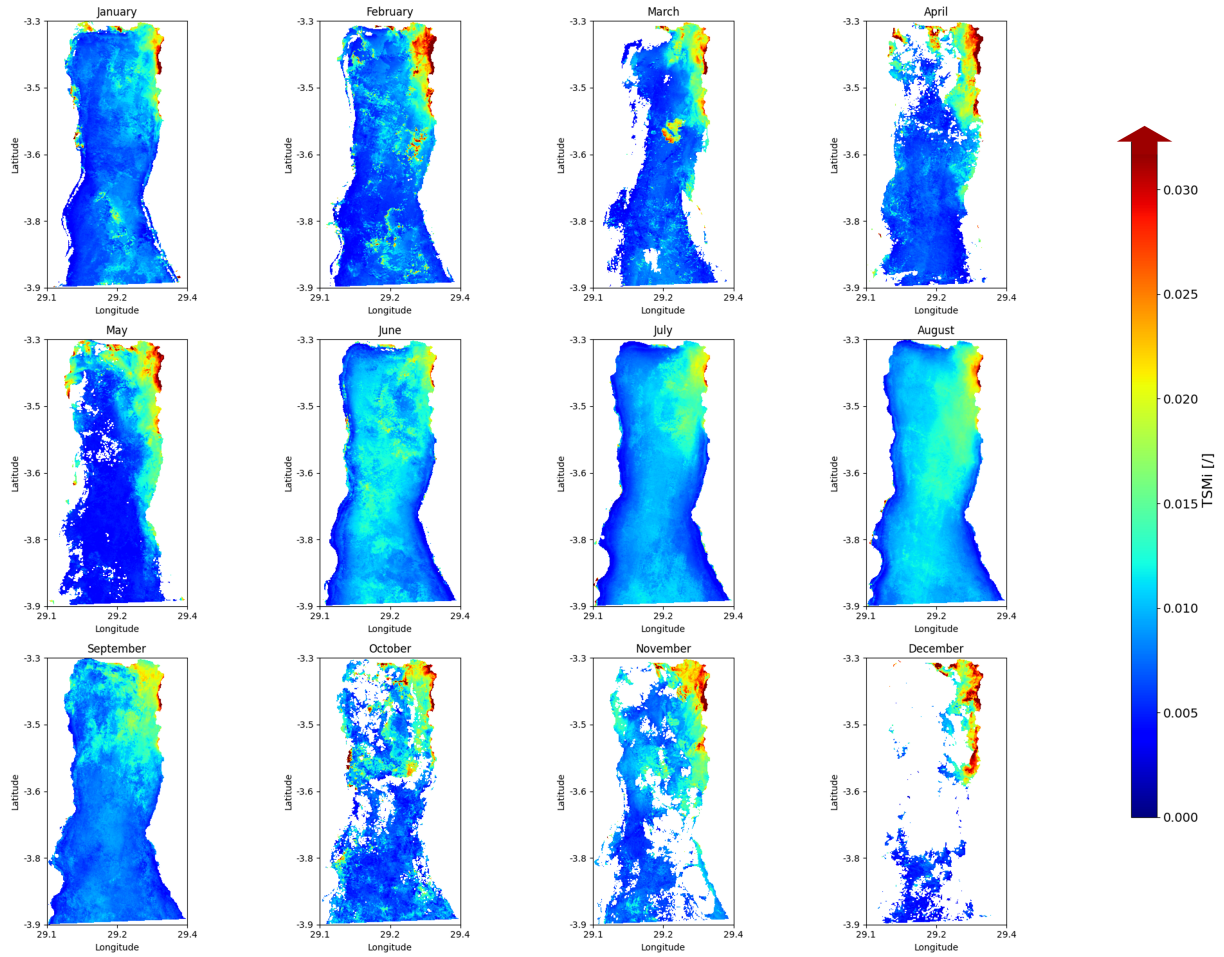


Figure 28: Monthly median maps of the Bujumbura region between 2013 and 2023 for TSMi

Conversely, the monthly median maps of ChloA, as depicted in Figure 29, exhibit differences compared to the TSMi dataset. Indeed, the ChloA dataset demonstrates fewer no-value pixels throughout the year. This observation is also quantified in Figure 48. The overall dynamics mirror those of the TSMi dataset, with higher concentrations observed from September to April (as opposed to October to April for the TSMi dataset) and a more diffuse plume of high values in the coastal areas of Bujumbura during this season.

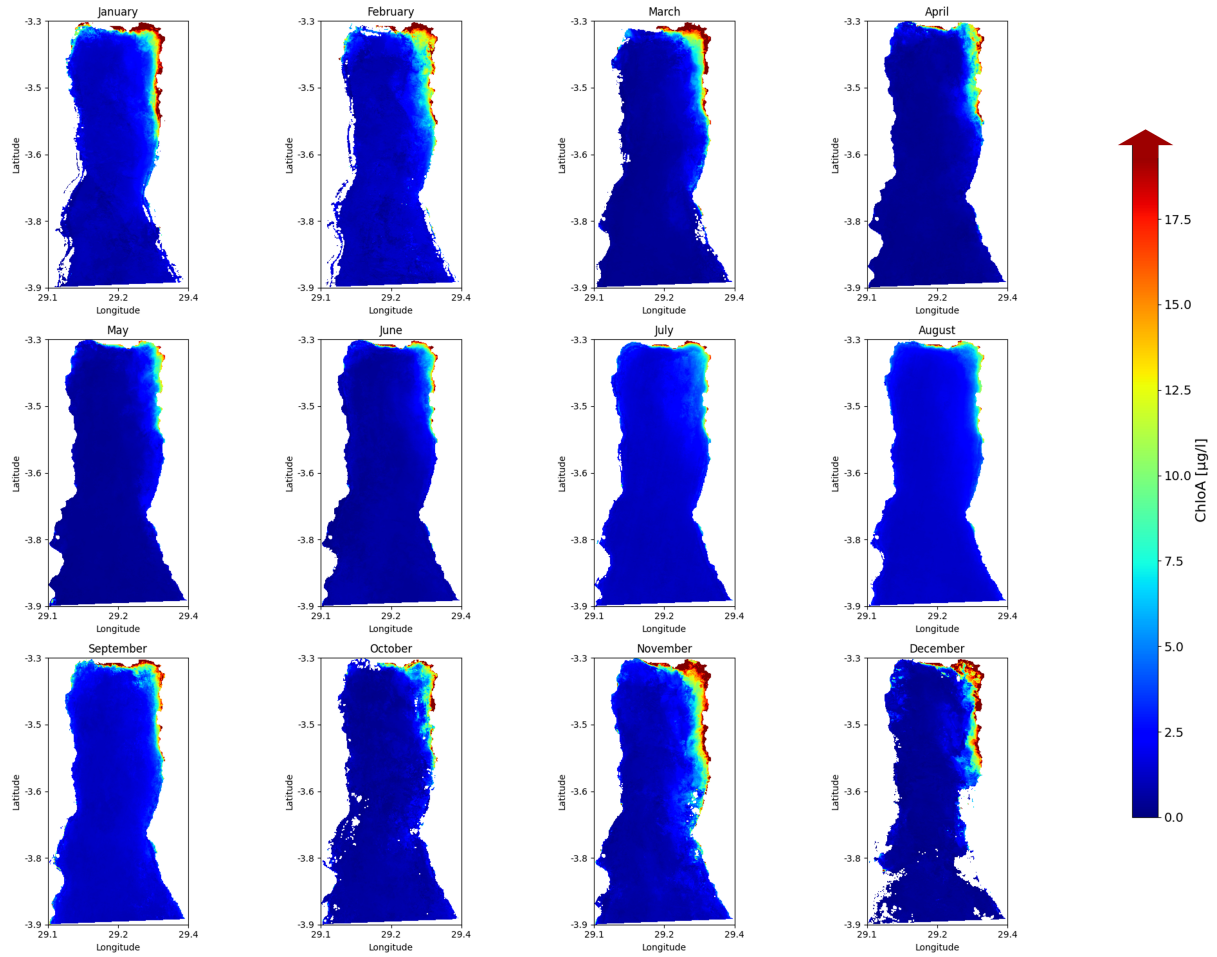


Figure 29: Monthly median maps of the Bujumbura region between 2013 and 2023 for ChloA

4.3.1.2 Kigoma

For the Kigoma region, fewer observations occur during the wet season compared to the dry season for the TSMi dataset. This observation is quantified in Figure 49. The overall trend demonstrates higher concentrations between November and April, with a plume of high values observed near the estuarine area of the Malagarasi River and the coastal area of Kigoma.

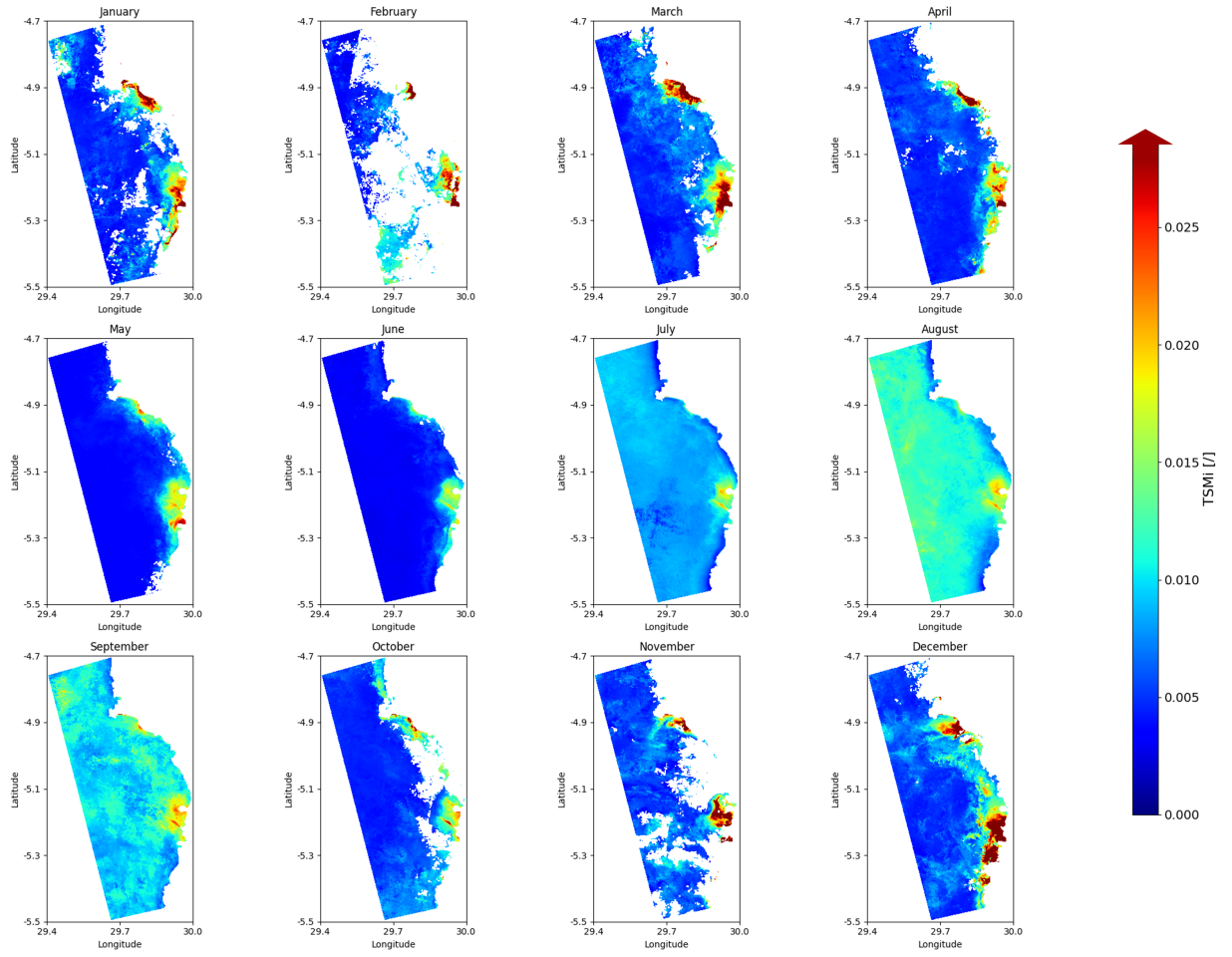


Figure 30: Monthly median maps of the Kigoma region between 2013 and 2023 for TSMi

Again, for the ChloA dataset less observations are obtained for the wet season compared to the dry season (quantified in Figure 50). Higher concentrations are observed next to the estuarine of the Malagarasi River and in the coastal area of Kigoma from December to May.

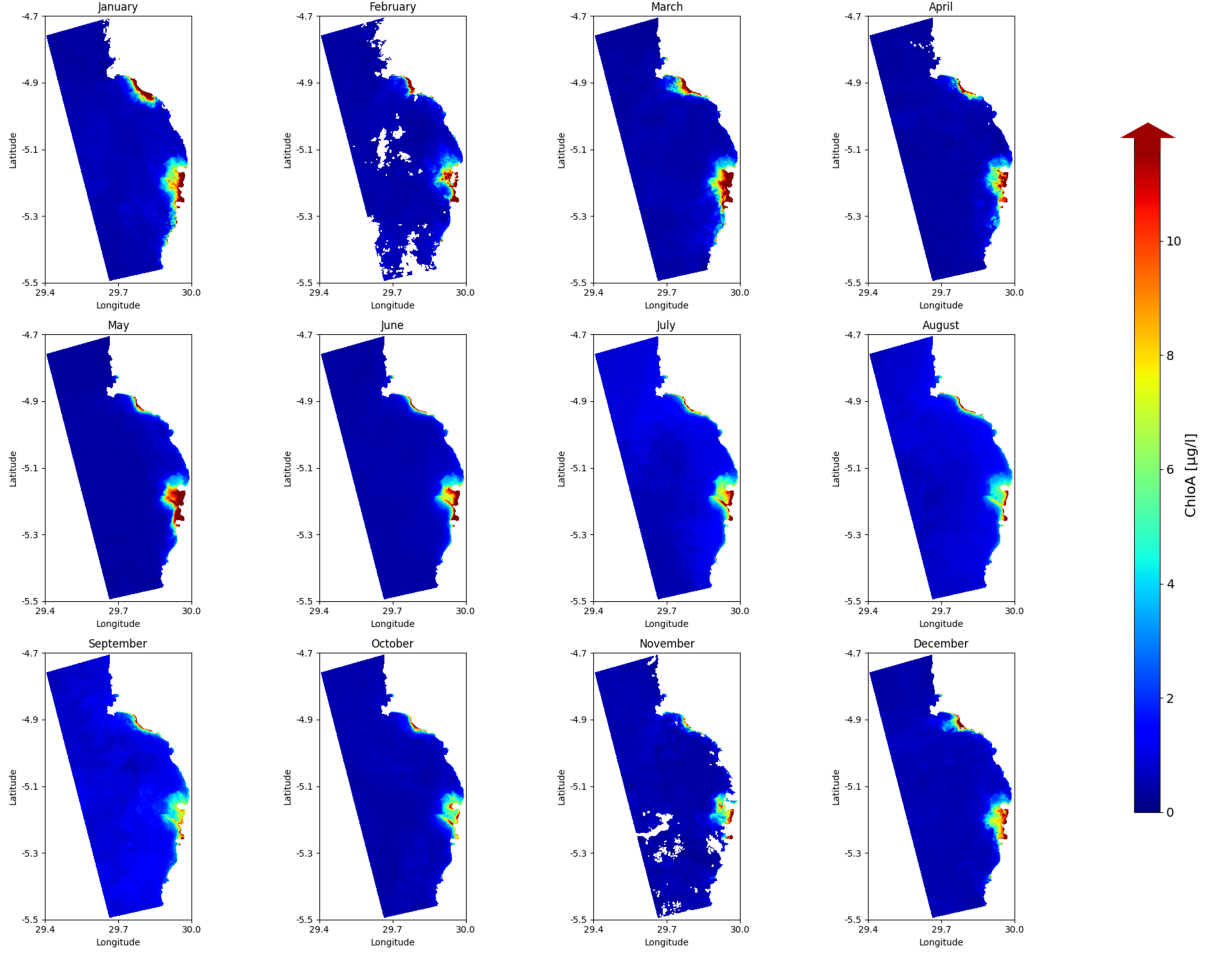


Figure 31: Monthly median maps of the Kigoma region between 2013 and 2023 for ChloA

4.3.2 Empirical Orthogonal Function

The EOF, as explained in Section 3.6.1, is used to identify the dominant modes of spatial and temporal variation for each variable in each ROI. In this study, four components were used to capture the maximum variation in the dataset. The selection of components was based on a trial-and-error approach, guided by visual interpretation of the spatial and temporal dimensions of the components for each variable and each ROI.

4.3.2.1 Bujumbura region

For ChloA and turbidity in the Bujumbura region, the results are quite similar between variables (see Table 16). The first component of each variable explains most of the variance (around eighty percent). The second component explains around ten percent, and the last two components explain five percent each.

Table 16: Percentage of variance explained by each component of the Empirical Orthogonal Function for the Bujumbura region

Variable \ Component	1 st	2 nd	3 rd	4 th	Total
ChloA	79.66	8.56	6.93	4.85	100
Turbidity	78.05	10.37	6.94	4.63	100

Figure 32 presents the spatial mode of the first component for turbidity and ChloA. The spatial mode of the other components is presented in the Appendix section A.6.1.1.

There are similarities between the spatial dimensions of the first components of ChloA and turbidity and the median of those variables between 2016 and 2020 presented in Figures 26 and 27. The spatial distribution of turbidity exhibits variations in dynamics observed within the coastal area of Bujumbura. However, for the spatial distribution of ChloA, these differences are evident not only within the coastal area of Bujumbura but also along the east and west coasts of the Ruzizi river mouth.

The values presented in Figure 32 correspond to the spatial patterns that, when combined with the temporal dimension presented in Figure 33, will reconstruct the time series images as closely as possible. The closer the pixel value is to 0, the less it will be influenced by the temporal dimension. Areas with pixel values close to 0 will therefore remain relatively constant over time.

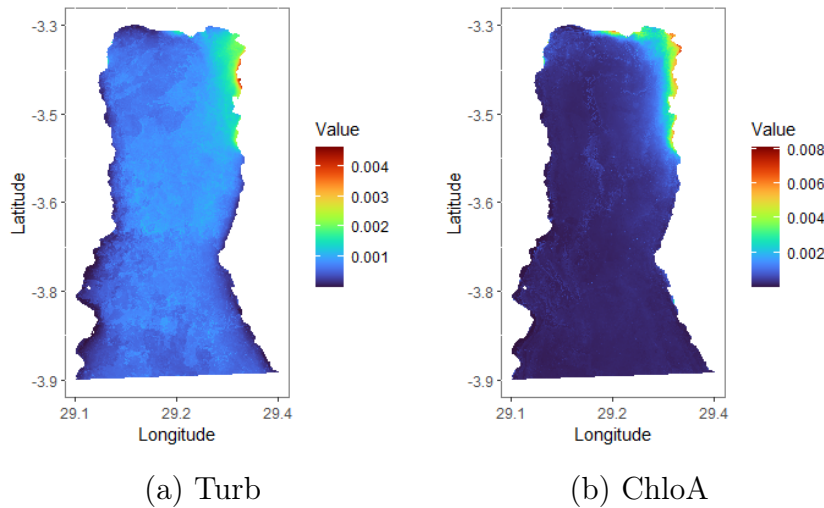


Figure 32: Spatial dimension of the first component of the EOF for the Bujumbura region for (a) Turbidity and (b) Chlorophyll-A

Figure 33 displays the monthly median values, as well as the 25th and 75th percentiles, of the temporal dimension for both ChloA and turbidity's first component. Similar to the spatial dimension, values closer to 0 indicate images where the spatial pattern is less pronounced, while values farther from 0 suggest a more prominent pattern. It can be observed that for turbidity,

the spatial pattern depicted in Figure 32 is more evident during the dry season (from May to September), peaking in August, with minimum values occurring from October to December. In contrast, for ChloA, the spatial pattern is more pronounced in November, January, and March, with a minimum value in February and a consistent representation during the dry season. The monthly median values of the temporal pattern for the last three components are provided in the Appendix, Section A.6.2.

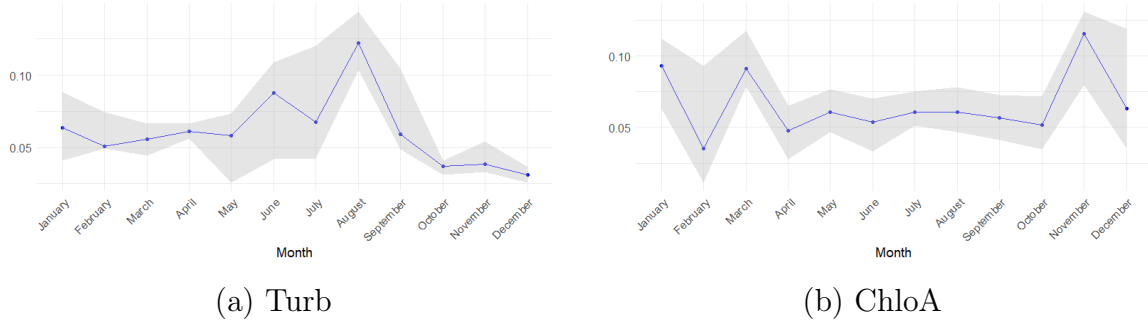


Figure 33: Monthly median, 25th and 75th percentiles of the temporal dimension of the first component of the EOF for the Bujumbura region for (a) Turbidity and (b) Chlorophyll-A

4.3.2.2 Kigoma region

The results presented in Table 17 are very similar between the Bujumbura region and the Kigoma region. In both regions, the first components of both variables explain most of the variance (around eighty percent), the second components explain around ten percent, and the third and fourth components explain around five percent each.

Table 17: Percentage of variance explained by each component of the Empirical Orthogonal Function for the Kigoma region

Variable \ Component	1 st	2 nd	3 rd	4 th	Total
ChloA	79.56	8.43	6.65	5.36	100
Turbidity	81.33	9.48	5.09	4.1	100

The spatial dimensions of the first components of ChloA and turbidity are presented in Figure 34. The spatial dimensions of the last three components are presented in the Appendix (Section A.6.1.2).

The spatial dimensions of the first components reveal distinct dynamics observed near the mouth of the Malagarasi River and in the coastal area of the city of Kigoma, contrasting with the rest of the ROI. While the dynamics appear similar between the two variables, turbidity exhibits a less homogeneous dynamic in the middle of the lake compared to ChloA.

The interpretation of pixel values is the same as for the Bujumbura region, with values close to 0 indicating almost no temporal variation.

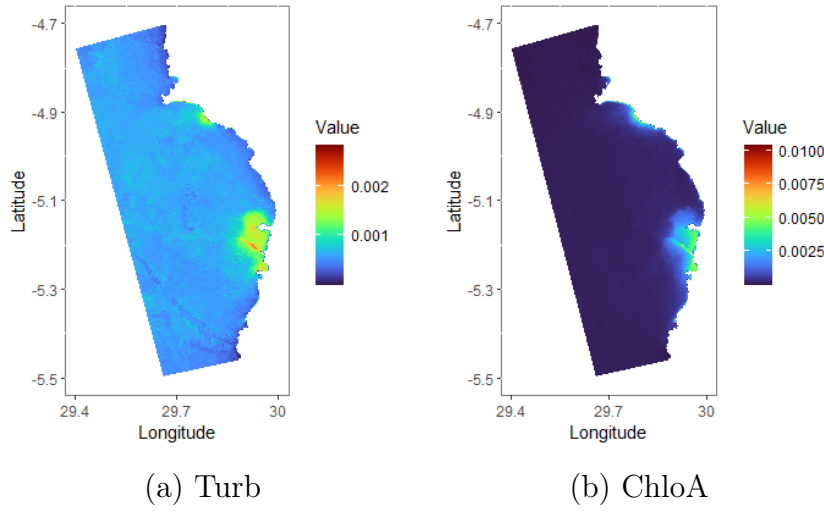


Figure 34: Spatial dimension of the first component of the EOF for the Bujumbura region for (a) Turbidity and (b) Chlorophyll-a

In Figure 35, the monthly median values, as well as the 25th and 75th percentiles, of the temporal dimension for the first component of both ChloA and turbidity are depicted. Similar to the spatial dimension, values closer to 0 signify images where the spatial pattern is less pronounced, whereas values farther from 0 suggest a more prominent pattern. Regarding turbidity, the spatial pattern illustrated in Figure 34 is more prevalent from July to September, peaking in August. Conversely, for ChloA, this pattern is more pronounced in May, with values fluctuating throughout the year and minimum values occurring in April.

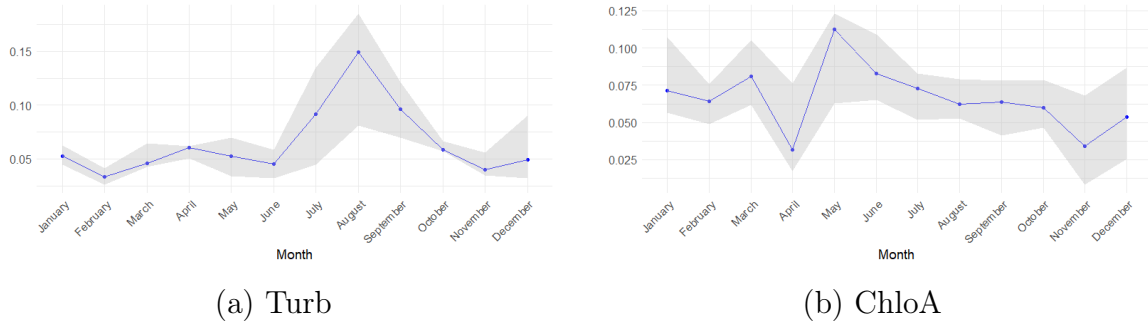


Figure 35: Monthly median, 25th and 75th percentiles of the temporal dimension of the first component of the EOF for the Kigoma region for (a) Turbidity and (b) Chlorophyll-a

4.3.3 Kmean clustering

A clustering of both regions was performed based on the spatial dimension of the first components of ChloA and turbidity. The number of clusters was determined through visual interpretation for both regions. The clustering results for two and four clusters can be found in Appendix section A.7.1 for the Bujumbura region and in Appendix section A.7.2 for the Kigoma region. These clusters represent areas with similar dynamics in ChloA and turbidity.

4.3.3.1 Bujumbura

For the Bujumbura region, three different clusters have emerged:

- **Coastal cluster:** This cluster represents three different areas:
 - The first area covers a 20 km stretch along the coast of the city of Bujumbura and southern coastal areas. The offshore extent of this area varies from 4 km to a few meters, depending on the location.
 - The second area is located in the west region of the Ruzizi River mouth. This area is approximately 3.5 km long, and the offshore extent varies from around 1200m to a few meters.
 - The last area is situated in the east region of the Ruzizi River mouth, covering a region around 2 km long. The offshore extent varies between 1.5 km and a few meters.
- **Intermediate cluster:** This cluster is mainly defined by two different areas:
 - The first area covers the offshore region from Bujumbura to the offshore area of the Ruzizi River mouth. This area is around 35 km long, and its offshore extent varies from around 5.5 km to a few meters.
 - The second area covers the coastal area of Uvira, a city located at the same latitude as Bujumbura but on the west coast of the lake. This region is around 5 km long, and its offshore extent varies from 500 m to a few meters.
 - Some isolated pixels are also present on the east and west coasts of the lake, at a latitude of around 3.8°S.
- **Middle cluster:** The last cluster represents all the areas not contained in the other clusters.

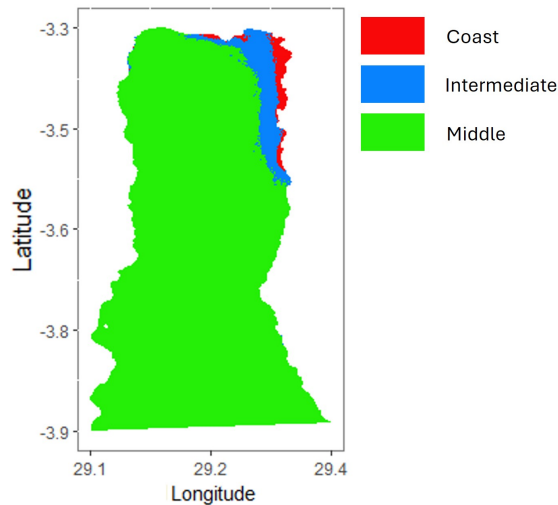


Figure 36: Clustering of the Bujumbura region based on the ChloA and Turbidity dynamics

4.3.3.2 Kigoma

For the Kigoma region three clusters have emerged :

- **Coastal cluster:** This cluster represents two distinct areas:
 - The first area corresponds to the north and south coasts of the Malagarasi River mouth, with a length of 10 km and an offshore extent ranging from around 5 km to a few meters.
 - The second area corresponds to the coastal area of Kigoma, with a length of around 13 km and an offshore extent ranging from 1 km to a few meters. A few pixels are also present north of the city of Kigoma.
- **Intermediate cluster:** This cluster represents two different areas:
 - The first area corresponds to the offshore area of the Malagarasi River mouth, with a length of 25 km and an offshore extent ranging from around 6km to a few meters. This area also contains a few pixels on the coast of the Malagarasi river mouth.
 - The second area corresponds to the offshore area of the coast of Kigoma, with a length of around 13 km and an offshore extent ranging from around 3 km to a few meters. This area also includes the northern part of the coastal area of Kigoma, with a length of 1.5 km and an offshore extent of around 500 m.
- **Middle cluster:** The last cluster, called "middle," represents all areas not included in the first two clusters.

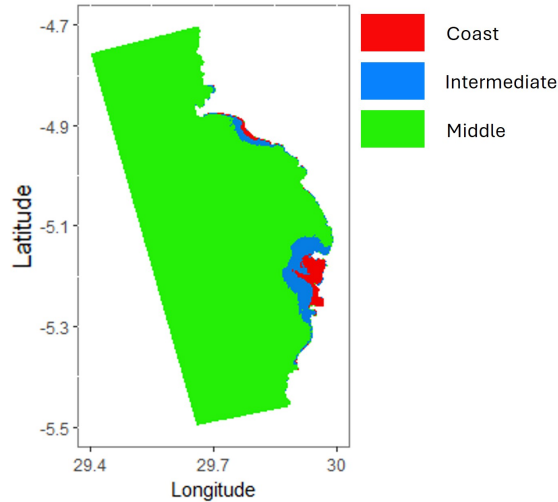


Figure 37: Clustering of the Kigoma region based on the ChloA and Turbidity dynamics

4.3.4 Correlation analysis between Chlorophyll-A and Turbidity daily median in each cluster and with the river flow

As detailed in Section [3.6.3](#), this section aims to examine the correlation within each cluster between the dynamics of ChloA and turbidity, as well as the correlation between each dynamic

and the river flow. The first stage included analyzing the daily median of the coast and intermediate clusters and contrasting them with the river flow on the corresponding day.

4.3.4.1 Bujumbura

Figure 38 illustrates the daily median, along with the 5th and 95th percentiles of each variable, juxtaposed with the river flow. A linear regression analysis was performed between the ChloA and TSMi concentrations and the river flow. The plots of these linear regressions can be found in the Appendix, specifically in Section A.8.

Generally, when observing the plots of TSMi and ChloA concentration for the region of Bujumbura in both the coast and intermediate clusters, no clear pattern emerges. Indeed, the disparity in the data makes it challenging to identify seasonality from these time series. However, it can be observed that peaks in the river flow are not consistently accompanied by increases in TSMi or ChloA concentrations for either cluster. In early 2016, a maximum river flow (around $220 \text{ m}^3/\text{s}$) appears to precede a peak in ChloA concentration (around $110 \text{ }\mu\text{g/l}$). However, these events are separated by more than a month (2016-01-11 for the maximum river flow and 2016-02-28 for the peak in ChloA concentration).

For ChloA, the intermediate cluster exhibits lower concentrations than the coast cluster, with values around $10 \text{ }\mu\text{g/l}$ for the intermediate cluster and $20 \text{ }\mu\text{g/l}$ for the coast cluster. The fluctuations are also less pronounced in the intermediate cluster compared to the coast cluster. This observation is also applicable to TSMi, though less distinctly. The coast cluster generally shows higher values than the intermediate cluster, with fluctuations around 0.015 for the intermediate cluster compared to 0.025 for the coast cluster. The intermediate cluster shows more fluctuations in TSMi compared to ChloA dynamics within the same cluster.

The TSMi dynamics in the coast cluster showed the highest correlation with the Ruzizi river flow, with an R^2 value of 0.16. In contrast, the TSMi dynamics in the intermediate cluster exhibited a lower correlation, with an R^2 value of 0.02. The correlation coefficients for ChloA and the Ruzizi river flow were 0.04 for the coast cluster and 0.05 for the intermediate cluster.

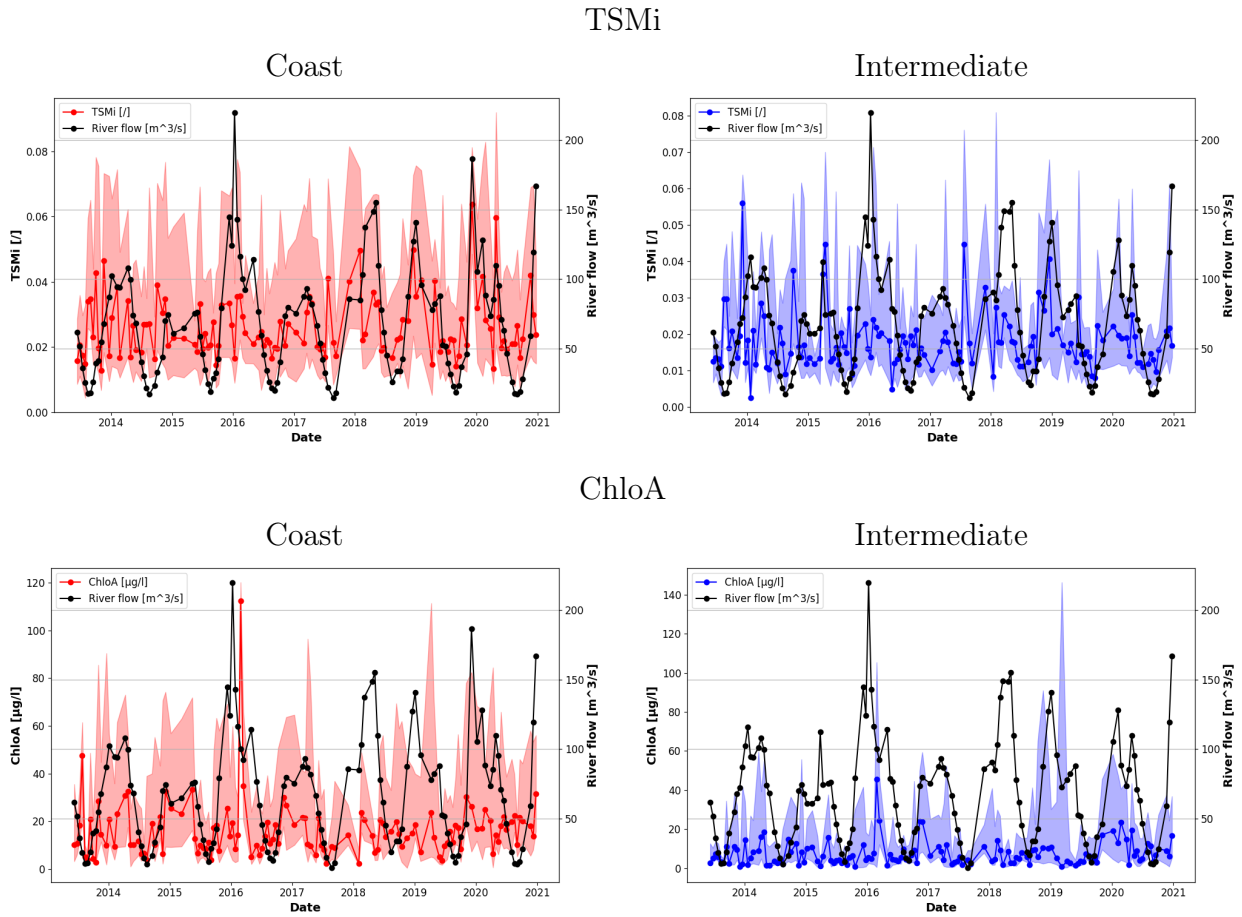


Figure 38: Daily median, 5th and 95th percentiles of TSMi and ChloA between 2013 and 2020 for the Coast (red) and Intermediate (blue) cluster of the region of Bujumbura compared with the Ruzizi river flow derived from Wongchuig et al., 2023 between 2013 and 2020.

4.3.4.2 Kigoma

Similarly to the region of Bujumbura, based on the analysis of the ChloA and TSMi daily median it is hard to derive a seasonality for the two variables due to the disparity in the data. As for Bujumbura, peaks in the river flow are not consistently accompanied by increases in TSMi or ChloA concentrations for either cluster. In general, concentrations are lower in the intermediate than in the coast cluster for both variables (for ChloA: around 15µg/l for the coast cluster compared to around 5 µg/l for the intermediate cluster. For TSMi: around 0.02 for the coast cluster compared to around 0.01 for the intermediate cluster).

The ChloA dynamics in the coast cluster showed the highest correlation with the Malagarasi river flow, with an R^2 value of 0.16. In contrast, the ChloA dynamics in the intermediate cluster, as well as the TSMi dynamics in both clusters, exhibited similar correlations with the Malagarasi river flow, with an R^2 value of 0.02.

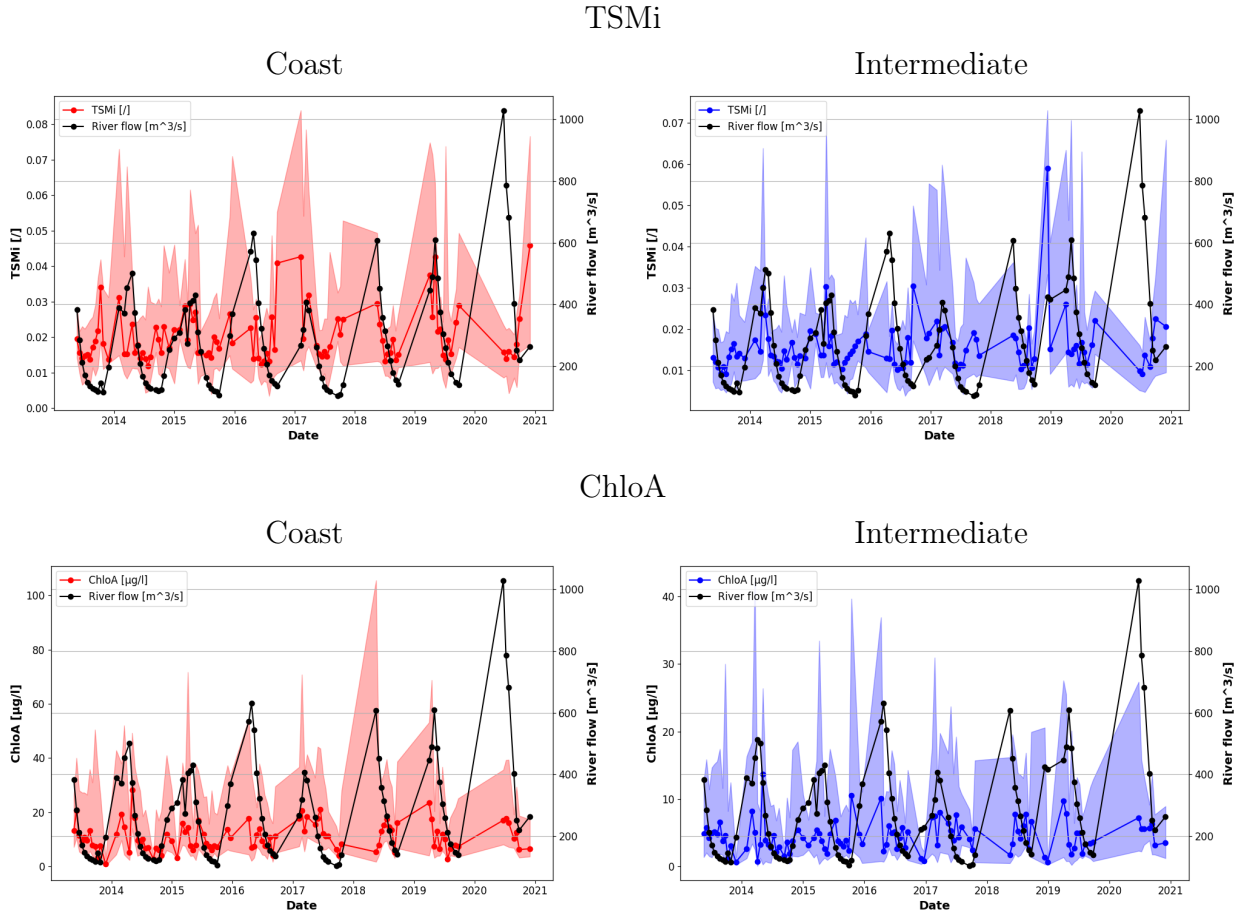


Figure 39: Daily median, 5th and 95th percentiles of TSMi and ChloA between 2013 and 2020 for the Coast (red) and Intermediate (blue) cluster of the region of Kigoma compared with the Malagarasi river flow derived from Wongchuig et al., 2023 between 2013 and 2020.

4.3.5 Correlation analysis between Chlorophyll-A and Turbidity monthly median in each cluster and with the river flow

To study the intra-annual dynamics of ChloA and Turbidity in each cluster, the daily medians were used to create monthly medians for each year. The monthly medians from 2013 to 2023 were then averaged to create a single median value for each month, representing the average concentration per month over the ten-year period.

4.3.5.1 Bujumbura

Figure 40 presents the monthly median per year and a global monthly median for TSMi and ChloA in the coast and intermediate cluster of the Bujumbura region. The TSMi dataset exhibits variations across months and clusters. Both the Coast and Intermediate clusters show similarities, with higher values from October to April and lower values from May to September. However, this pattern is less pronounced in the Intermediate cluster, where the differences between high-value and low-value months are smaller (approximately 0.28 to 0.02 for the Coast cluster and around 0.018 to 0.015 for the Intermediate cluster).

Both clusters show some similarities with the Ruzizi river flow, with TSMi concentrations

decreasing from May, paralleling the river flow decrease. However, variations in high river flow are not reflected in TSMi values, which remain almost constant from October to December, despite a significant increase in river flow during these months.

The ChloA dynamics in the Coast cluster show higher concentrations from November to March and lower concentrations from May to August. The Intermediate cluster has higher concentrations from November to April and lower concentrations from May to October. The variation between high and low-concentration months is more pronounced in the Coast cluster than in the Intermediate cluster (around 20 $\mu\text{g/l}$ to 10 $\mu\text{g/l}$ for the Coast cluster and from 10 $\mu\text{g/l}$ to 5 $\mu\text{g/l}$ in the Intermediate cluster). There are some differences between the clusters, with higher concentrations in September in the Coast cluster and in April in the Intermediate cluster. The dynamic in November and December is also slightly different.

Comparison with the Ruzizi river flow is less straightforward for ChloA than for TSMi. Although higher concentrations of ChloA are observed during months with high river flow, the timing of the increase and decrease in ChloA concentrations does not align directly with the periods of rising and falling Ruzizi river flow, as it does for TSMi.

For both clusters, notable differences exist between the ChloA and TSMi dynamics, particularly during the months of increase and decrease (October and April). A peak value in TSMi is observed in February, which is not reflected in the ChloA data.

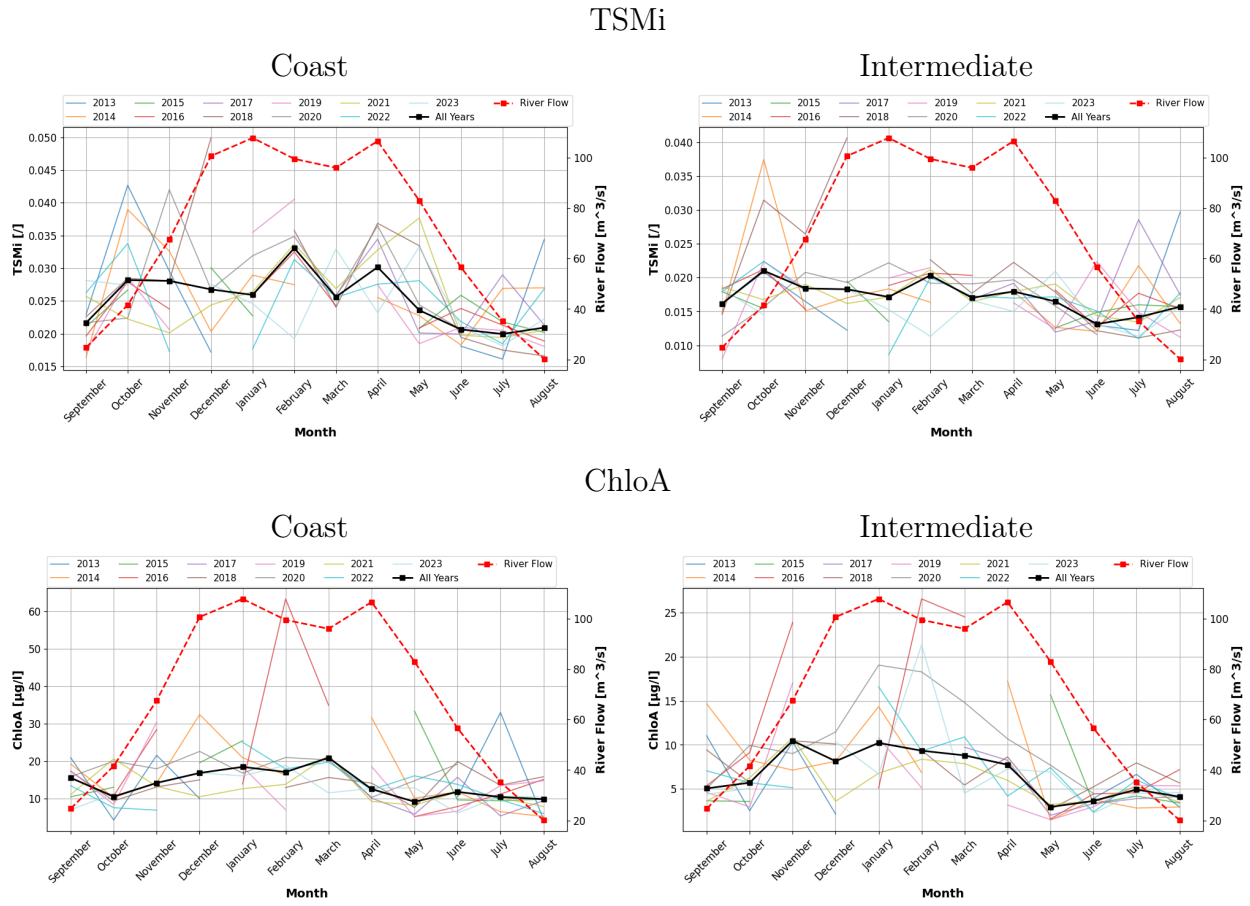


Figure 40: Monthly median of TSMi and ChloA between 2013 and 2023 for the Coast and Intermediate cluster of the region of Bujumbura compared with the Ruzizi river flow derived from Wongchuig et al., [2023](#) between 2013 and 2020.

4.3.5.2 Kigoma

Figure [41](#) presents the monthly median per year and a global monthly median for TSMi and ChloA in the coast and intermediate cluster of the Kigoma region. The TSMi dynamics exhibit higher values from November to April and lower values from June to August. Both the Coast and Intermediate clusters display a similar trend, but discrepancies emerge in December and March where a higher increase is observed in December than in March in the coast cluster while this is the inverse in the intermediate cluster. The range of difference between high and low-value months is narrower in the Intermediate cluster compared to the Coast cluster. This difference ranges from around 0.025 to 0.015 for the Coast cluster, whereas it's approximately 0.018 to 0.012 for the Intermediate cluster.

While there are similarities between the TSMi dynamics and the Malagarasi river flow, such as a decrease in TSMi values from May to July and an increase from October to December, there are notable differences. For instance, lower TSMi values in April coincide with a maximum in river flow, and the increase in TSMi values in August and September contrasts with a minimum in river flow.

The ChloA dynamics in the Coast cluster show high concentrations from January to May and

lower concentrations from August to October. A sudden decrease in ChloA concentration is observed in April, with levels around 12 $\mu\text{g/l}$, while March and April present concentrations around 14 $\mu\text{g/l}$. This pattern is surprising given the maximum river flow in April. Except for February, the Coast cluster generally follows the Malagarasi river flow dynamics, with a decrease from May to August and an increase in November and December.

The Intermediate cluster generally follows the same dynamics as the Coast cluster, except in July and August. During these months, a decrease in ChloA concentration is observed in the Coast cluster, while the Intermediate cluster maintains constant concentrations.

Differences are observed in both clusters between ChloA and TSMi, particularly from January to May, where different dynamics are observed. While ChloA shows a constant increase from January to March, TSMi exhibits a decrease in January and a slight increase in March. In May, high concentrations are observed in ChloA but not in TSMi.

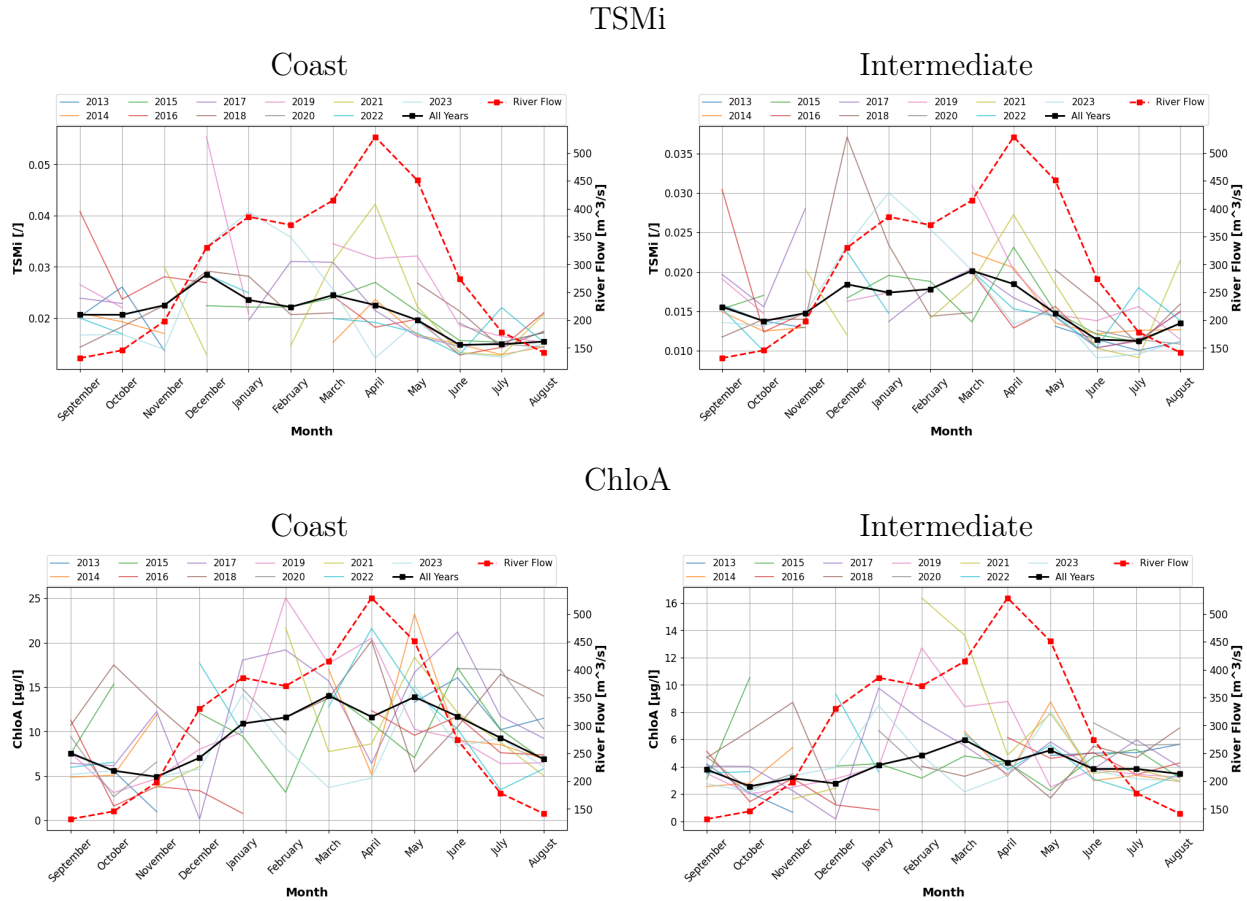


Figure 41: Monthly median of TSMi and ChloA between 2013 and 2023 for the Coast and Intermediate cluster of the region of Kigoma compared with the Malagarasi river flow derived from Wongchuig et al., [2023](#) between 2013 and 2020.

5 Discussion

5.1 Investigating the downscaling failure

This section discusses all the results presented in Section 4.1. This part is separated into three subsections: the first subsection focuses on the band comparison between Landsat PAR and CCI Lakes, the second discusses the optimization of turbidity algorithms, and the third discusses the optimization of ChloA algorithms.

5.1.1 Band Comparison

Several hypotheses could explain the differences observed between the R_{rs} of the CCI Lakes products and the Landsat PAR upscaled to 1 km resolution. Firstly, when upscaling the Landsat PAR product from 30 m to 1 km resolution, the bilinear resampling method was used. This method, as detailed in Section 3.4.1.3, uses the four closest pixels to the new pixel center to interpolate the value of the new pixel. Although bilinear or nearest-neighbor resampling methods have been used in some studies for upscaling Landsat 8 to 4 km resolution (cf. Fu et al., 2018; Mohebzadeh et al., 2020), these methods can introduce uncertainties, especially in areas with strong spatial heterogeneity, such as estuarine environments. The interpolation process may not adequately capture the high variability in R_{rs} within these regions, leading to discrepancies when compared to higher-resolution products.

Secondly, as described in Section 3.2.1, Landsat PAR is a provisional product. The atmospheric correction applied to this product is not considered reliable, and the data provided have several weaknesses, as identified by Ogashawara et al., 2020 and presented in Section 3.2.1.2. One of the main issues is the overcorrection of aerosols, leading to negative R_{rs} values and an underestimation of R_{rs} , especially for Bands 2 and 3, while Band 4 shows slightly better results. These observations may explain the differences observed between Landsat PAR and CCI Lakes R_{rs} .

The better performance observed for Band 3 could be associated with higher changes between the middle of the lake and the region of Bujumbura compared to Bands 1 and 2. This observation suggests that Landsat PAR is only able to capture significant variations in R_{rs} and does not capture small changes.

Furthermore, the better results presented for Band 4 can be associated with a better estimation of R_{rs} by the Landsat PAR product, as described by Ogashawara et al., 2020. The reduced number of observations due to the masking of numerous negative R_{rs} values may also influence the results obtained for this band.

5.1.2 Turbidity Algorithms Optimization

The first notable observation of the CCI Lakes Turbidity dataset is the differences between the median map of Turbidity (see Figure 19) and the median map of the CCI Lakes Red band (665 nm) R_{rs} (see Figure 23). Given that the majority of the algorithms used in CCI Lakes are

based on the Red band R_{rs} (cf. Table 13), more similarities were expected between the two median maps. From 2016 to 2020, the CCI Lakes Turbidity dataset was derived from OLCI. This remote sensing product utilizes various algorithms, as shown in Table 9 with one of them (Y. Zhang et al., 2014, used for the OWTs 1, 7, and 10) employing a wavelength of 754 nm instead of 665 nm.

The differences observed between these two bands (see section 4.1.3.2) may explain the discrepancies between CCI Lakes turbidity and the 665 nm band. However, when examining the spectral properties of turbidity, it is challenging to directly attribute these differences solely to turbidity concentration. The peak absorbance around 665 nm for ChloA (see Figure 2) could explain part of this pattern. The presence of ChloA in these areas could mask higher turbidity concentrations. Since this peak is not present around 750 nm, it explains the observed differences. This hypothesis is mitigated by the spectral signature of water with constant ChloA concentration and increasing suspended sediment concentrations (see Figure 4), where the difference in R_{rs} for varying turbidity concentrations is smaller at 665 nm than at 700 nm but appears quite similar at 750 nm. However, these observations clearly demonstrate the interactions between turbidity and ChloA, highlighting the complexity of distinguishing between the two using remote sensing products with a limited number of bands.

Considering the observed differences between the CCI Lakes and Landsat PAR band R_{rs} and the different dynamics between the CCI Lakes 665 nm band and the turbidity dataset, the results from optimizing turbidity algorithms using the Landsat PAR R_{rs} dataset were unsurprising.

5.1.3 Chlorophyll-A Algorithms Optimization

The optimization results obtained for ChloA could be attributed to the differences between the Ultra Blue and Blue bands R_{rs} of CCI Lakes and Landsat PAR. Although a slightly better correspondence was found for the Green band, its use with the other two bands may obscure this improvement. The global underestimation of ChloA concentration when using the OC3 equation with coefficients fitted for Landsat 8 by J. E. O'Reilly and Werdell, 2019 can be linked to the overall underestimation of R_{rs} by the Landsat PAR product, as identified by Ogashawara et al., 2020.

5.2 Semi-Empirical Algorithm and Analytical Indexes

This section discusses the visual comparison between the overall dynamics captured by the Semi-Empirical algorithm and the two Analytical indexes and those observed with CCI Lakes data.

For the NDTI, the high values observed in the middle of the lake can be associated with R_{rs} values that are close. When R_{rs} values are similar, they yield NDTI values close to 0, which indicates more turbid water compared to coastal areas (coastal areas present negative NDTI values), even if R_{rs} values are themselves very low. This pattern may be linked to an underestimation of R_{rs} in the green band, whereas this underestimation is less pronounced in

the red band, as noted by Ogashawara et al., [2020].

The advantage of using the TSMi compared to the NDTI is that TSMi values will be close to 0 for small R_{rs} values, even if these values are close to each other.

However, the global increase in turbidity concentration in coastal areas is not represented by either NDTI or TSMi. This pattern arises due to the same reasons that the red band of CCI Lakes does not correspond to the observed turbidity pattern. These differences underscore the importance of the algorithm used and caution in interpreting results derived from this dataset. Indeed, the precision of the dataset created cannot be estimated due to the lack of in-situ data and the observed failure in downscaling. On top of that, the TSMi values are not expressed in terms of turbidity concentration. One hypothesis made on further interpretations is that there is a linear relationship between the TSMi values and turbidity concentration.

The overall dynamics observed for the OC3 algorithm with Landsat 8 fitted coefficients show similarities with the CCI Lakes ChloA dataset, indicating that despite the overall underestimation of R_{rs} by Landsat PAR, the algorithm still captures the broad spatial patterns of ChloA concentration.

As for the TSMi, the accuracy of the ChloA concentration retrieved by this algorithm is not guaranteed due to the lack of in-situ data and the failure in downscaling. The underlying hypothesis is then that the dynamics captured by the OC3 algorithm with Landsat 8 fitted coefficients are similar to the ChloA dynamics of the lake. However, the ChloA concentration retrieved should not be interpreted as accurate but rather used to capture the overall dynamics.

5.3 Chlorophyll-a and Turbidity time series analysis

5.3.1 Global dynamic of the datasets

The monthly median maps presented in Figures [28], [29], [30], and [31], combined with the figures in Appendix Section A.5, underscore the challenges posed by the reduced number of observations during the wet season. This limitation arises from the prevalent cloud cover during this period, rendering it impossible to capture dynamics effectively. Consequently, there are instances of non-observation for certain pixels from October to May over the span of the available 10 years of data.

The higher number of masked pixels in the ChloA dataset compared to the TSMi dataset can be attributed to the spectral signature of water, as elucidated in Figure 5. In clear water, the R_{rs} in the Red band is very low. Considering the underestimation of the Landsat PAR product as described by Ogashawara et al., [2020], this leads to the masking of several low turbidity pixels.

The observed pattern in the TSMi monthly median maps, particularly the isolated high turbidity values observed in the middle of the lake, raises concerns about the reliability of this dataset. Such anomalies can indeed cast doubt on the accuracy and consistency of the data.

5.3.2 Empirical Orthogonal Function

The EOF was used in order to identify the dominant modes of both spatial and temporal variations.

In both regions, the four components generated do not adequately capture the dynamics of TSMi during the wet season. This limitation is directly attributed to the TSMi dataset, which suffers from a scarcity of available images during this period. Conversely, the dynamics of ChloA in the wet season are well represented by the first two components in both Bujumbura and Kigoma.

The pattern observed is validated for both regions: the variance explained by the first two components indicates that the spatial patterns capture most of the data in the TSMi and ChloA datasets, while the last three components add negligible variance. The four spatial components of ChloA for both regions align with the observed ChloA dynamics, whereas the last two spatial components of TSMi seem to be associated with noise and exhibit incoherent dynamics.

5.3.3 Kmean clustering

The decision to select only the spatial dimension of the first components of ChloA and turbidity was based on the following observations for both regions: the visual interpretation of the second mode of ChloA and turbidity was very similar to the first mode, and it seems not worthwhile taking into account those components regarding some irregularities (high concentration pixels in the middle of the lake for example). Additionally, the variance explained by the second components was considered negligible compared to the variance explained by the first component. However, these assumptions could impact the clusters created, as they represent only the temporal patterns explained by the first two components. Consequently, these clusters might not fully capture the global dynamics in these areas over the year.

For Bujumbura and Kigoma, the presence of similar clusters next to the estuary and along the coastal areas of major cities can be explained by the consistently high concentrations of both ChloA and turbidity observed throughout the year in these regions. The intermediate cluster acts as a buffer zone between these areas and the rest of the lake, representing regions less impacted by the coastal areas and the estuary.

It's important to note the influence of the coastal area of Bujumbura on water quality dynamics. Although it is not possible to definitively assess the impact of the Ruzizi River on the high ChloA and turbidity concentrations observed in the coastal area of Bujumbura due to the presence of a clockwise circulation pattern near the shore (as described by Well and Chapman, 1976), the hypothesis that this pattern is solely caused by the Ruzizi River seems improbable considering the non-continuity of the coast cluster between the east of the Ruzizi River mouth and the coastal areas of Bujumbura.

Therefore, it seems likely that the four rivers crossing the city of Bujumbura and flowing into the lake play a significant role in the elevated ChloA and turbidity levels in this area. However,

this hypothesis seems to be invalidated by Vandelannoote et al., [1996] who indicate that the complete mixing of these rivers with lake water occurred within 100 to 300 meters from shore.

In the Bujumbura region, the inclusion of Uvira's coastal region in the intermediate cluster suggests that this city also influences water quality variables, albeit to a lesser extent than Bujumbura. Similarly, in the Kigoma region, the coastal area of Kigoma is contained in the coast cluster, indicating its influence on water quality. These observations raise concerns about the impact of urban expansion along the lake's coast and its effects on water quality.

5.3.4 Correlation analysis between Chlorophyll-A and Turbidity daily and monthly median in each cluster and with the river flow

Firstly, it's important to note that the different clusters in Bujumbura and Kigoma will not be influenced in the same way by the rivers. When observing the cluster distribution between the two ROIs, the coast cluster in the Bujumbura region will be more influenced by the four rivers crossing the city of Bujumbura than by the Ruzizi River directly. However, it's expected that these rivers will present a similar flow dynamic as they will be influenced similarly by precipitation. For the Kigoma region, the coast cluster will be mainly driven by the Malagarasi River flow due to the reduced extent of this cluster in the coastal area of Kigoma.

The disparities noted in the monthly median analyses underscore that ChloA and turbidity display distinct dynamics in estuarine and coastal regions. On top of the differences induced by the dataset uncertainties, this divergence could be explained by varied influencing factors for each variable. Indeed factors such as the upwelling events (Plisnier et al., [1996]) or the dependence of the ChloA on nutrients and the water temperature (Huang et al., [2022]) could explain such differences.

The nonlinear correlation between river flow and the ChloA and TSMi dynamic could also be attributed to the nonlinear relationship between sediment concentration in river water and seasonal variations. Indeed, erosion might be more intense at the onset of rapid flow and decrease over time, even if the river flow does not decrease accordingly (Vercruysse et al., [2017]). Surprisingly, the decrease in precipitation in February in both Bujumbura and Kigoma is not accompanied by a decrease in TSMi in Kigoma and is even accompanied by an increase in TSMi in Bujumbura.

For the ChloA dynamics in the region of Bujumbura, it is challenging to definitively attribute the observed patterns to either river flow or the limnological cycle described by Plisnier et al., [1996]. The differences in dynamics between the months of the metalimnion oscillation (on average from September to January) are not notably distinct from other months of high concentration. However, the observation of high ChloA concentrations during months where no upwelling is expected, coinciding with periods of high river flow, strengthens the hypothesis of the influence of river flow on ChloA concentration.

For certain months, the information captured by the current datasets makes it challenging to explain the observed dynamics (for example, the drop in ChloA concentration in April in

the Kigoma region). These results underscore the need for a comprehensive study of these dynamics, incorporating additional water quality variables such as water temperature and nutrient concentrations. Furthermore, there is a critical need for reliable data to ensure that the observed patterns are not artifacts of the dataset.

Regarding the results presented here, it is hard to attribute only the ChloA and turbidity dynamic to the river flow. However, it is likely that this variable plays an important role in the ChloA and turbidity dynamic in the estuarine areas of Lake Tanganyika.

5.4 Future prospect

One significant challenge addressed in this study is the quality of data provided by the Landsat PAR product. New versions of Landsat PAR aim to tackle the issue of aerosol overcorrection, especially for inland waters (USGS, personal communication, April 24, 2024). These advancements could greatly enhance the accuracy of water quality assessments derived from satellite data, offering more reliable information for regions such as Bujumbura and Kigoma.

Another challenge in these regions is frequent cloud cover, which limits the availability of clear satellite images. Increasing the temporal resolution of observations is essential to overcoming this limitation. Sentinel-2 and Sentinel-3 satellites offer potential solutions. Sentinel-2 provides higher temporal resolution and spatial detail, while Sentinel-3, despite its 300-meter resolution, offers more frequent observations. These satellites can help fill the gaps left by cloud cover and improve overall temporal coverage of water quality monitoring. Additionally, they feature extra spectral bands that can provide more detailed information on various water quality parameters.

To ensure the quality and reliability of satellite-derived datasets, integrating *in situ* data is essential. Ground truth measurements can validate satellite observations, helping to assess and improve the accuracy of remote sensing data. This integration is critical for understanding the local dynamics of water quality variables and for making informed decisions based on satellite data.

The integrated examination of ChloA and turbidity dynamics alongside other water quality variables, such as water temperature and nutrient concentrations, can greatly enhance our comprehension of the dynamics in these areas. By considering multiple factors simultaneously, we can gain deeper insights into the complex interactions driving changes in water quality over time.

6 Conclusion

The main objectives of this study were to analyze the spatio-temporal dynamics and identify the drivers of ChloA and turbidity in the estuarine areas of Lake Tanganyika, specifically focusing on the Malagarasi and Ruzizi river mouths. These objectives highlighted the need to build high-resolution time series for ChloA and turbidity. This study used Landsat PAR for its 30 m spatial resolution, long-range availability (from 2013), and specific atmospheric correction for aquatic environments. CCI Lakes ChloA and turbidity datasets were used to identify and evaluate the reliability of different ChloA and turbidity algorithms for Landsat PAR.

The failure to optimize ChloA and turbidity algorithms using Landsat PAR R_{rs} resampled at 1 km with the CCI Lakes ChloA and turbidity data underscored three important points: first, it highlighted the weakness and need for improvement in the atmospheric correction provided by Landsat PAR; second, it emphasized the importance of the algorithm choice and the bands used to compute turbidity; and finally, it drew attention to the general method of downscaling and the noise that can be introduced when resampling a product from 30 m resolution to 1 km.

To address the downscaling issue and the lack of data, time series for ChloA and turbidity in the estuaries were constructed using the OC3 algorithm with optimized coefficients for Landsat 8 for ChloA, and the TSMi for turbidity. However, it's important to note that the accuracy of these datasets cannot be verified, so the interpretation of those data should be interpreted with caution.

The spatial analysis of ChloA and turbidity dynamics reveals distinct patterns depending on the region. In the Bujumbura region, the influence of the Ruzizi River is mainly localized on both the east and west coasts of the estuary, with minimal offshore extent, while the coastal area of Bujumbura reflects the city's impact on water quality. In contrast, in the Kigoma region, the influence of the Malagarasi River extends further, with the coastal area of Kigoma showing a more localized impact on water quality.

The temporal analysis reveals a disparity in the dynamics between ChloA and turbidity in both regions, suggesting the necessity for additional water quality variables, such as water temperature, to better comprehend their drivers. Comparing the intra-annual dynamics of ChloA and turbidity with river flow indicates similarities, with higher concentrations in both variables observed during periods of increased river flow. However, this relationship does not appear to be linear. Concerns arise about the reliability of the dynamics captured by the time series due to reduced observations during certain months caused by cloud cover.

The prospects resulting from this study are diverse. Exploring coastal and estuarine areas using remote sensing satellites with reliable atmospheric correction for aquatic environments, suitable spectral bands for studying aquatic environments, and higher temporal resolution is crucial. Sentinel-2 and Sentinel-3 emerge as promising options in this regard. Additionally, conducting parallel studies of ChloA and turbidity alongside other water quality variables, such as water temperature and nutrient concentration, could enhance our understanding of the dynamics occurring in estuaries.

References

- Atkinson, P. M. (2013). Downscaling in remote sensing. *International Journal of Applied Earth Observation and Geoinformation*, 22, 106–114. <https://doi.org/10.1016/j.jag.2012.04.012>
- Baboo, D. S. S., & Devi, M. R. (2010). An Analysis of Different Resampling Methods in Coimbatore, District.
- Baglama, J., Reichel, L., & Lewis, B. W. (2022, October). Irlba: Fast Truncated Singular Value Decomposition and Principal Components Analysis for Large Dense and Sparse Matrices. Retrieved May 25, 2024, from <https://cran.r-project.org/web/packages/irlba/index.html>
- Beauchamp, R. S. A. (1939). Hydrology of Lake Tanganyika [_eprint: <https://onlinelibrary.wiley.com/doi/p> *Internationale Revue der gesamten Hydrobiologie und Hydrographie*, 39(3-4), 316–353. <https://doi.org/10.1002/iroh.19390390303>
- Becker, R. H., Sayers, M., Dehm, D., Shuchman, R., Quintero, K., Bosse, K., & Sawtell, R. (2019). Unmanned aerial system based spectroradiometer for monitoring harmful algal blooms: A new paradigm in water quality monitoring. *Journal of Great Lakes Research*, 45(3), 444–453. <https://doi.org/10.1016/j.jglr.2019.03.006>
- Bergamino, N., Horion, S., Stenuite, S., Cornet, Y., Loiselle, S., Plisnier, P. .-, & Descy, J.-P. (2010). Spatio-temporal dynamics of phytoplankton and primary production in Lake Tanganyika using a MODIS based bio-optical time series. *Remote Sensing of Environment*, 114(4), 772–780. <https://doi.org/10.1016/j.rse.2009.11.013>
- Binding, C. E., Jerome, J. H., Bukata, R. P., & Booty, W. G. (2010). Suspended particulate matter in Lake Erie derived from MODIS aquatic colour imagery [Publisher: Taylor & Francis _eprint: <https://doi.org/10.1080/01431160903302973>]. *International Journal of Remote Sensing*, 31(19), 5239–5255. <https://doi.org/10.1080/01431160903302973>
- Brusseau, M. L., Walker, D. B., & Fitzsimmons, K. (2019, January). Chapter 3 - Physical-Chemical Characteristics of Water. In M. L. Brusseau, I. L. Pepper, & C. P. Gerba (Eds.), *Environmental and Pollution Science (Third Edition)* (pp. 23–45). Academic Press. <https://doi.org/10.1016/B978-0-12-814719-1.00003-3>
- Caballero, I., Morris, E. P., Ruiz, J., & Navarro, G. (2014). Assessment of suspended solids in the Guadalquivir estuary using new DEIMOS-1 medium spatial resolution imagery. *Remote Sensing of Environment*, 146, 148–158. <https://doi.org/10.1016/j.rse.2013.08.047>
- Campbell, J. B. (1996). *Introduction to Remote Sensing* [Google-Books-ID: 1KfQxsN0vp8C]. CRC Press.
- Carrea, L., Crétaux, J.-F., Liu, X., Wu, Y., Calmettes, B., Duguay, C. R., Merchant, C. J., Selmes, N., Simis, S. G. H., Warren, M., Yesou, H., Müller, D., Jiang, D., Embury, O., Bergé-Nguyen, M., & Albergel, C. (2023). Satellite-derived multivariate world-wide lake physical variable timeseries for climate studies [Publisher: Nature Publishing Group]. *Scientific Data*, 10(1), 30. <https://doi.org/10.1038/s41597-022-01889-z>

- Chai, T., & Draxler, R. (2014). Root mean square error (RMSE) or mean absolute error (MAE)?— Arguments against avoiding RMSE in the literature. *Geoscientific Model Development*, 7, 1247–1250. <https://doi.org/10.5194/gmd-7-1247-2014>
- Chebud, Y., Naja, G. M., Rivero, R. G., & Melesse, A. M. (2012). Water Quality Monitoring Using Remote Sensing and an Artificial Neural Network. *Water, Air, & Soil Pollution*, 223(8), 4875–4887. <https://doi.org/10.1007/s11270-012-1243-0>
- Chen, Z., Hu, C., & Muller-Karger, F. (2007). Monitoring turbidity in Tampa Bay using MODIS/Aqua 250-m imagery. *Remote Sensing of Environment*, 109(2), 207–220. <https://doi.org/10.1016/j.rse.2006.12.019>
- Cheng, C., Messerschmidt, L., Bravo, I., Waldbauer, M., Bhavikatti, R., Schenk, C., Grujic, V., Model, T., Kubinec, R., & Barceló, J. (2024). A General Primer for Data Harmonization. *Scientific Data*, 11, 152. <https://doi.org/10.1038/s41597-024-02956-3>
- Coulter, G. W. (1968). Hydrological processes and primary production in Lake Tanganyika. *Proc. 11th Conf. Great Lakes Res. Int. Assoc. Great Lakes Res.* 609–626.
- Dall’Olmo, G., Gitelson, A. A., Rundquist, D. C., Leavitt, B., Barrow, T., & Holz, J. C. (2005). Assessing the potential of SeaWiFS and MODIS for estimating chlorophyll concentration in turbid productive waters using red and near-infrared bands. *Remote Sensing of Environment*, 96(2), 176–187. <https://doi.org/10.1016/j.rse.2005.02.007>
- Dekker, A. G., Vos, R. J., & Peters, S. W. M. (2002). Analytical algorithms for lake water TSM estimation for retrospective analyses of TM and SPOT sensor data. *International Journal of Remote Sensing*, 23(1), 15–35. <https://doi.org/10.1080/01431160010006917>
- Dekker, A., & Peters, S. (1993). The Use of the Thematic Mapper for the Analysis of Eutrophic Lakes: A Case Study in The Netherlands. *International Journal of Remote Sensing*, 14, 799–822. <https://doi.org/10.1080/01431169308904379>
- Dekker, A. G., Brando, V. E., Anstee, J. M., Pinnel, N., Kutser, T., Hoogenboom, E. J., Peters, S., Pasterkamp, R., Vos, R., Olbert, C., & Malthus, T. J. (2001). Imaging Spectrometry of Water. In F. D. v. d. Meer & S. M. D. Jong (Eds.), *Imaging Spectrometry: Basic Principles and Prospective Applications* (pp. 307–359). Springer Netherlands. https://doi.org/10.1007/978-0-306-47578-8_11
- Dubois, J. T. (1958). Evolution de la temperature de l’oxygene dissous et de la transparence dans la baie Nord du lac Tanganika. *Hydrobiologia*, 10(1), 215–240. <https://doi.org/10.1007/BF00142188>
- Edmond, J. M., Stallard, R. F., Craig, H., Craig, V., Weiss, R. F., & Coulter, G. W. (1993). Nutrient chemistry of the water column of Lake Tanganyika [eprint: <https://onlinelibrary.wiley.com/doi/10.4319/lo.1993.38.4.0725>]. *Limnology and Oceanography*, 38(4), 725–738. <https://doi.org/10.4319/lo.1993.38.4.0725>
- Encyclopedia of Inland Waters [ISBN: 9780123706263]. (2009). Retrieved April 2, 2024, from <http://www.sciencedirect.com:5070/referencework/9780123706263/encyclopedia-of-inland-waters>
- Fu, Y., Xu, S., Zhang, C., & Sun, Y. (2018). Spatial downscaling of MODIS Chlorophyll-a using Landsat 8 images for complex coastal water monitoring. *Estuarine, Coastal and Shelf Science*, 209, 149–159. <https://doi.org/10.1016/j.ecss.2018.05.031>

- Gao, Y., Gao, J., Yin, H., Liu, C., Xia, T., Wang, J., & Huang, Q. (2015). Remote sensing estimation of the total phosphorus concentration in a large lake using band combinations and regional multivariate statistical modeling techniques. *Journal of Environmental Management*, 151, 33–43. <https://doi.org/10.1016/j.jenvman.2014.11.036>
- Gilerson, A. A., Gitelson, A. A., Zhou, J., Gurlin, D., Moses, W., Ioannou, I., & Ahmed, S. A. (2010). Algorithms for remote estimation of chlorophyll-a in coastal and inland waters using red and near infrared bands [Publisher: Optica Publishing Group]. *Optics Express*, 18(23), 24109–24125. <https://doi.org/10.1364/OE.18.024109>
- Gitelson, A. A., Dall’Olmo, G., Moses, W., Rundquist, D. C., Barrow, T., Fisher, T. R., Gurlin, D., & Holz, J. (2008). A simple semi-analytical model for remote estimation of chlorophyll-a in turbid waters: Validation. *Remote Sensing of Environment*, 112(9), 3582–3593. <https://doi.org/10.1016/j.rse.2008.04.015>
- Gitelson, A. A., Gurlin, D., Moses, W. J., & Barrow, T. (2009). A bio-optical algorithm for the remote estimation of the chlorophyll-a concentration in case 2 waters. *Environmental Research Letters*, 4(4), 045003. <https://doi.org/10.1088/1748-9326/4/4/045003>
- Gons, H. J., Rijkeboer, M., & Ruddick, K. G. (2005). Effect of a waveband shift on chlorophyll retrieval from MERIS imagery of inland and coastal waters. *Journal of Plankton Research*, 27(1), 125–127. <https://doi.org/10.1093/plankt/fbh151>
- Gordon, H. R., Brown, O. B., & Jacobs, M. M. (1975). Computed Relationships Between the Inherent and Apparent Optical Properties of a Flat Homogeneous Ocean [Publisher: Optica Publishing Group]. *Applied Optics*, 14(2), 417–427. <https://doi.org/10.1364/AO.14.000417>
- Gupta, D. K., Prashar, S., Singh, S., Srivastava, P. K., & Prasad, R. (2022, January). Chapter 1 - Introduction to RADAR remote sensing. In P. K. Srivastava, D. K. Gupta, T. Islam, D. Han, & R. Prasad (Eds.), *Radar Remote Sensing* (pp. 3–27). Elsevier. <https://doi.org/10.1016/B978-0-12-823457-0.00018-5>
- Guyot, G., & Seguin, B. (1988). Possibilit  s d’  utilisation de la t  l  d  tection satellitaire en agrom  t  orologie [Publisher: EDP Sciences]. *Agronomie*, 8(1), 1–13. Retrieved April 3, 2024, from <https://hal.science/hal-00885063>
- Haberyan, K. A., & Hecky, R. E. (1987). The late Pleistocene and Holocene stratigraphy and paleolimnology of Lakes Kivu and Tanganyika. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 61, 169–197. [https://doi.org/10.1016/0031-0182\(87\)90048-4](https://doi.org/10.1016/0031-0182(87)90048-4)
- Hannouche, A., Chebbo, G., Ruban, G., Tassin, B., Lemaire, B. J., & Joannis, C. (2011). Relationship between turbidity and total suspended solids concentration within a combined sewer system. *Water Science and Technology*, 64(12), 2445–2452. <https://doi.org/10.2166/wst.2011.779>
- Havens, K. E. (2003). Submerged aquatic vegetation correlations with depth and light attenuating materials in a shallow subtropical lake. *Hydrobiologia*, 493(1), 173–186. <https://doi.org/10.1023/A:1025497621547>
- Hecky, R. E., & Kling, H. J. (1981). The phytoplankton and protozooplankton of the euphotic zone of Lake Tanganyika: Species composition, biomass, chlorophyll content, and spatio-temporal distribution1 [_eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.4319/lo.1981.26.3.0548>

- Limnology and Oceanography*, 26(3), 548–564. <https://doi.org/10.4319/lo.1981.26.3.0548>
- Hecky, R. E., Spigel, R. H., & Coulter, G. W. (1991, April). The nutrient regime. In J.-J. Tiercelin, A. Mondegeur, R. E. Hecky, R. H. Spigel, & G. W. Coulter (Eds.), *Lake Tanganyika and its Life* (p. 0). Oxford University Press. <https://doi.org/10.1093/oso/9780198585251.003.0004>
- Horne, A. J., & Goldman, C. R. (1994). *Limnology* [Google-Books-ID: 3Iys5rnwkWYC]. McGraw-Hill.
- Hou, X., Feng, L., Duan, H., Chen, X., Sun, D., & Shi, K. (2017). Fifteen-year monitoring of the turbidity dynamics in large lakes and reservoirs in the middle and lower basin of the Yangtze River, China. *Remote Sensing of Environment*, 190, 107–121. <https://doi.org/10.1016/j.rse.2016.12.006>
- Hu, C., Feng, L., Lee, Z., Franz, B. A., Bailey, S. W., Werdell, P. J., & Proctor, C. W. (2019). Improving Satellite Global Chlorophyll a Data Products Through Algorithm Refinement and Data Recovery [_eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1029/2019JC014941>]. *Journal of Geophysical Research: Oceans*, 124(3), 1524–1543. <https://doi.org/10.1029/2019JC014941>
- Huang, H., Wang, W., Lv, J., Liu, Q., Liu, X., Xie, S., Wang, F., & Feng, J. (2022). Relationship between Chlorophyll a and Environmental Factors in Lakes Based on the Random Forest Algorithm [Number: 19 Publisher: Multidisciplinary Digital Publishing Institute]. *Water*, 14(19), 3128. <https://doi.org/10.3390/w14193128>
- Hunter, P. D., Tyler, A. N., Carvalho, L., Codd, G. A., & Maberly, S. C. (2010). Hyperspectral remote sensing of cyanobacterial pigments as indicators for cell populations and toxins in eutrophic lakes. *Remote Sensing of Environment*, 114(11), 2705–2718. <https://doi.org/10.1016/j.rse.2010.06.006>
- Huttula, T. (1997, January). *Flow, thermal regime and sediment transport studies in Lake Tanganyika*. Kuopio University (Vol. 73) [Journal Abbreviation: Publications C. Natural and Environmental Sciences Publication Title: Publications C. Natural and Environmental Sciences].
- Janssen, L. L. F., Bakker, W. H., & International Institute for Aerospace Survey and Earth Sciences (Eds.). (2001). *Principles of remote sensing: An introductory textbook* (2nd ed). International Institute for Aerospace Survey; Earth Sciences.
- Kmeans function - RDocumentation. (n.d.). Retrieved May 25, 2024, from <https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/kmeans>
- Lacaux, J., Tourre, Y., Vignolles, C., Ndione, J., & Lafaye, M. (2007). Classification of ponds from high-spatial resolution remote sensing: Application to Rift Valley Fever epidemics in Senegal. *Remote Sensing of Environment*, 106(1), 66–74. <https://doi.org/10.1016/j.rse.2006.07.012>
- Lake Tanganyika | Size, Geographical Location & Wildlife | Britannica. (2023, December). Retrieved April 5, 2024, from <https://www.britannica.com/place/Malagarasi-River>
- Lakes CCI project Key Documents. (n.d.). Retrieved April 7, 2024, from <https://climate.esa.int/en/projects/lakes/>

- Landsat Provisional Aquatic Reflectance ADD | U.S. Geological Survey. (2020). Retrieved December 12, 2023, from <https://www.usgs.gov/media/files/landsat-provisional-aquatic-reflectance-add>
- Lathrop, R. G. (1989). Monitoring Water Quality and River Plume Transport in Green Bay, Lake Michigan with SPOT-1 Imagery.
- Le, C. F., Li, Y. M., Zha, Y., Sun, D., & Yin, B. (2009). Validation of a Quasi-Analytical Algorithm for Highly Turbid Eutrophic Water of Meiliang Bay in Taihu Lake, China [Conference Name: IEEE Transactions on Geoscience and Remote Sensing]. *IEEE Transactions on Geoscience and Remote Sensing*, 47(8), 2492–2500. <https://doi.org/10.1109/TGRS.2009.2015658>
- Le, C., Li, Y., Zha, Y., Sun, D., Huang, C., & Lu, H. (2009). A four-band semi-analytical model for estimating chlorophyll a in highly turbid lakes: The case of Taihu Lake, China. *Remote Sensing of Environment*, 113(6), 1175–1182. <https://doi.org/10.1016/j.rse.2009.02.005>
- Li, Y., & Wu, H. (2012). A Clustering Method Based on K-Means Algorithm. *Physics Procedia*, 25, 1104–1109. <https://doi.org/10.1016/j.phpro.2012.03.206>
- Li, Y., Zhang, Y., Shi, K., Zhu, G., Zhou, Y., Zhang, Y., & Guo, Y. (2017). Monitoring spatiotemporal variations in nutrients in a large drinking water reservoir and their relationships with hydrological and meteorological conditions based on Landsat 8 imagery. *Science of The Total Environment*, 599-600, 1705–1717. <https://doi.org/10.1016/j.scitotenv.2017.05.075>
- Martin, S. (2014). *An introduction to ocean remote sensing* (Second edition). Cambridge University Press.
- Meer, F. D. v. d. (2001). Basic Physics of Spectrometry. In F. D. v. d. Meer & S. M. D. Jong (Eds.), *Imaging Spectrometry: Basic Principles and Prospective Applications* (pp. 3–16). Springer Netherlands. https://doi.org/10.1007/978-0-306-47578-8_1
- Miller, R., & McKee, B. (2004). Using MODIS Terra 250 m imagery to map concentrations of total suspended matter in coastal waters. *Remote Sensing of Environment*, 93, 259–266. <https://doi.org/10.1016/j.rse.2004.07.012>
- Mishra, S., & Mishra, D. R. (2012). Normalized difference chlorophyll index: A novel model for remote estimation of chlorophyll-a concentration in turbid productive waters. *Remote Sensing of Environment*, 117, 394–406. <https://doi.org/10.1016/j.rse.2011.10.016>
- Mishra, S., Mishra, D. R., Lee, Z., & Tucker, C. S. (2013). Quantifying cyanobacterial phyco-cyanin concentration in turbid productive waters: A quasi-analytical approach. *Remote Sensing of Environment*, 133, 141–151. <https://doi.org/10.1016/j.rse.2013.02.004>
- Mobley, C. (2022). The Oceanic Optics Book [Medium: 924pp. Publisher: [object Object]]. <https://doi.org/10.25607/OBP-1710>
- Mohebzadeh, H., Yeom, J., & Lee, T. (2020). Spatial Downscaling of MODIS Chlorophyll-a with Genetic Programming in South Korea [Number: 9 Publisher: Multidisciplinary Digital Publishing Institute]. *Remote Sensing*, 12(9), 1412. <https://doi.org/10.3390/rs12091412>

- Mölsä, H., Reynolds, J. E., Coenen, E. J., & Lindqvist, O. V. (1999). Fisheries research towards resource management on Lake Tanganyika. *Hydrobiologia*, 407(0), 1–24. <https://doi.org/10.1023/A:1003712708969>
- Moore, K. A., Wetzel, R. L., & Orth, R. J. (1997). Seasonal pulses of turbidity and their relations to eelgrass (*Zostera marina* L.) survival in an estuary. *Journal of Experimental Marine Biology and Ecology*, 215(1), 115–134. [https://doi.org/10.1016/S0022-0981\(96\)02774-8](https://doi.org/10.1016/S0022-0981(96)02774-8)
- Moore, T., Campbell, J., & Feng, H. (2001). A fuzzy logic classification scheme for selecting and blending satellite ocean color algorithms. *IEEE Transactions on Geoscience and Remote Sensing*, 39, 1764–1776. *Geoscience and Remote Sensing, IEEE Transactions on*, 39, 1764–1776. <https://doi.org/10.1109/36.942555>
- Nixon, S. W. (1988). Physical energy inputs and the comparative ecology of lake and marine ecosystems [_eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.4319/lo.1988.33.4part2.1005>]. *Limnology and Oceanography*, 33(4part2), 1005–1025. <https://doi.org/10.4319/lo.1988.33.4part2.1005>
- Niyoyitungiye, L. (2020, June). *Diagnostic analysis of the major threats of Lake Tanganyika and proposals for improving its ecological status*. Retrieved June 1, 2024, from <https://hal.science/hal-02882704> working paper or preprint.
- Ntakimazi, G. (1995). Le rôle de Ecotones terre-eau dans la diversité biologique et les ressources du lac Tanganyika - UNESCO Archives AtOM catalogue. Retrieved May 24, 2024, from <https://atom.archives.unesco.org/le-rolle-de-ecotones-terre-eau-dans-la-diversite-biologique-et-les-ressources-du-lac-tanganyika>
- Ogashawara, I., Jechow, A., Kiel, C., Kohnert, K., Berger, S. A., & Wollrab, S. (2020). Performance of the Landsat 8 Provisional Aquatic Reflectance Product for Inland Waters. *Remote Sensing*, 12(15). <https://doi.org/10.3390/rs12152410>
- Ondrusek, M., Stengel, E., Kinkade, C., Vogel, R., Keegstra, P., Hunter, C., & Kim, C. (2012). The development of a new optical total suspended matter algorithm for the Chesapeake Bay. *Remote Sensing of Environment*, 119, 243–254. <https://doi.org/10.1016/j.rse.2011.12.018>
- O'Reilly, C. M., Alin, S. R., Plisnier, P.-D., Cohen, A. S., & McKee, B. A. (2003). Climate change decreases aquatic ecosystem productivity of Lake Tanganyika, Africa [Publisher: Nature Publishing Group]. *Nature*, 424(6950), 766–768. <https://doi.org/10.1038/nature01833>
- O'Reilly, J. E., Maritorena, S., Mitchell, B. G., Siegel, D. A., Carder, K. L., Garver, S. A., Kahru, M., & McClain, C. (1998). Ocean color chlorophyll algorithms for SeaWiFS [_eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1029/98JC02160>]. *Journal of Geophysical Research: Oceans*, 103(C11), 24937–24953. <https://doi.org/10.1029/98JC02160>
- O'Reilly, J. E., & Werdell, P. J. (2019). Chlorophyll algorithms for ocean color sensors - OC4, OC5 & OC6. *Remote Sensing of Environment*, 229, 32–47. <https://doi.org/10.1016/j.rse.2019.04.021>
- Ouma, Y. O., Waga, J., Okech, M., Lavisa, O., & Mbuthia, D. (2018). Estimation of Reservoir Bio-Optical Water Quality Parameters Using Smartphone Sensor Apps and Landsat

- ETM+: Review and Comparative Experimental Results [Publisher: Hindawi]. *Journal of Sensors*, 2018, e3490757. <https://doi.org/10.1155/2018/3490757>
- Palacios, S. L. (2016). ARSET Fundamentals Trainings | NASA Applied Sciences. Retrieved April 2, 2024, from <http://appliedsciences.nasa.gov/what-we-do/capacity-building/arset/arset-fundamentals>
- Palmer, S. C. J., Kutser, T., & Hunter, P. D. (2015). Remote sensing of inland waters: Challenges, progress and future directions. *Remote Sensing of Environment*, 157, 1–8. <https://doi.org/10.1016/j.rse.2014.09.021>
- Panawala, L. (2017). Home » Science » Biology » Cell Biology » Difference Between Chlorophyll A and B Difference Between Chlorophyll A and B Main Difference – Chlorophyll A vs Chlorophyll B.
- Pandey, A., & Pati, U. C. (2019). Image mosaicing: A deeper insight. *Image and Vision Computing*, 89, 236–257. <https://doi.org/10.1016/j.imavis.2019.07.002>
- Patil, M. (2018). Interpolation Techniques in Image Resampling. *International Journal of Engineering and Technology*, 7, 567–570. <https://doi.org/10.14419/ijet.v7i3.34.19383>
- Petus, C., Chust, G., Gohin, F., Doxaran, D., Froidefond, J.-M., & Sagarminaga, Y. (2010). Estimating turbidity and total suspended matter in the Adour River plume (South Bay of Biscay) using MODIS 250-m imagery. *CONTINENTAL SHELF RESEARCH*, 30(5), 379–392. <https://doi.org/10.1016/j.csr.2009.12.007>
- Plisnier, P.-D., Langenberg, V., Mwape, L., Chitamwembwa, D., Tshibangu, K., & Coenen, E. (1996). Limnological sampling during an annual cycle at three stations on Lake Tanganyika (1993-1994). Retrieved April 4, 2024, from <https://agris.fao.org/search/en/providers/122621/records/647228b969d6cbfdd4a36863>
- Plisnier, P.-D., Kayanda, R., MacIntyre, S., Obiero, K., Okello, W., Vodacek, A., Cocquyt, C., Abegaz, H., Achieng, A., Akonkwa, B., Albrecht, C., Balagizi, C., Barasa, J., Abel Bashonga, R., Bashonga Bishobibiri, A., Bootsma, H., Borges, A. V., Chavula, G., Dadi, T., ... Lawrence, T. (2023). Need for harmonized long-term multi-lake monitoring of African Great Lakes. *Journal of Great Lakes Research*, 49(6), 101988. <https://doi.org/10.1016/j.jglr.2022.01.016>
- Pyo, J., Duan, H., Baek, S., Kim, M. S., Jeon, T., Kwon, Y. S., Lee, H., & Cho, K. H. (2019). A convolutional neural network regression for quantifying cyanobacteria using hyperspectral imagery. *Remote Sensing of Environment*, 233, 111350. <https://doi.org/10.1016/j.rse.2019.111350>
- Reprojection. (n.d.). Retrieved April 7, 2024, from <http://doc.lizardtech.com/docs/GeoExpress/9.5/WebHelpCL/Content/Reprojection.htm>
- Ritchie, J. C. (1976). Remote Sensing of Suspended Sediments in Surface Waters. *PHOTOGRAM-METRIC ENGINEERING*.
- Ritchie, J. C., & Schiebe, F. R. (2000). Water Quality. In G. A. Schultz & E. T. Engman (Eds.), *Remote Sensing in Hydrology and Water Management* (pp. 287–303). Springer. https://doi.org/10.1007/978-3-642-59583-7_13
- Ritchie, J. C., Schiebe, F. R., Cooper, C. M., & Harrington, J. A., Jr. (1994). Chlorophyll Measurements in the Presence of Suspended Sediment Using Broad Band Spectral Sensors

- Aboard Satellites [ADS Bibcode: 1994JFEco...9..197R]. *Journal of Freshwater Ecology*, 9, 197–206. <https://doi.org/10.1080/02705060.1994.9664887>
- Robin, M. (1995). *La télédétection : Des satellites aux systèmes d'information géographiques*. Nathan.
- Sayler, K. (2023). Landsat 8-9 Collection 2 Level-2 Provisional Aquatic Reflectance Product Guide.
- Schalles, J. F., Schiebe, F. R., & Starks, P. J. (1997). Estimation of algal and suspended sediment loads (singly and combined) using hyperspectral sensors and integrated mesocosm experiments [Number: CONF-970319- Publisher: Environmental Research Institute of Michigan, Ann Arbor, MI (United States)]. Retrieved April 2, 2024, from <https://www.osti.gov/biblio/478136>
- Schmugge, T. J., Kustas, W. P., Ritchie, J. C., Jackson, T. J., & Rango, A. (2002). Remote sensing in hydrology. *Advances in Water Resources*, 25(8-12), 1367–1385. [https://doi.org/10.1016/S0309-1708\(02\)00065-9](https://doi.org/10.1016/S0309-1708(02)00065-9)
- Schneider, A., Hommel, G., & Blettner, M. (2010). Linear Regression Analysis. *Deutsches Ärzteblatt International*, 107(44), 776–782. <https://doi.org/10.3238/arztebl.2010.0776>
- Sibomana, C., Buhungu, S., Ntakirutimana, D., & Nahimana, D. (2024). Impact of anthropogenic Pollution on Bujumbura Rivers: Toward Rivers and Lake Tanganyika Biodiversity Conservation, Burundi. *Journal of Applied Sciences and Environmental Management*, 28(1), 253–262. Retrieved May 17, 2024, from <https://www.ajol.info/index.php/jasem/article/view/263792>
- Spyrakos, E., O'Donnell, R., Hunter, P. D., Miller, C., Scott, M., Simis, S. G. H., Neil, C., Barbosa, C. C. F., Binding, C. E., Bradt, S., Bresciani, M., Dall'Olmo, G., Giardino, C., Gitelson, A. A., Kutser, T., Li, L., Matsushita, B., Martinez-Vicente, V., Matthews, M. W., ... Tyler, A. N. (2018). Optical types of inland and coastal waters [eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/lno.10674>]. *Limnology and Oceanography*, 63(2), 846–870. <https://doi.org/10.1002/lno.10674>
- Symoens, J. J. (1995a). Observation d'une fleur d'eau à Cyanophycées au Lac Tanganika. *Folia Sci. Afr. Centr.* 1: 17.
- Symoens, J. J. (1995b). Sur le maximum planctonique observé en fin de saison sèche dans le bassin nord du Lac Tanganika. *Folia Sci. Afr. Centr.* 1:12.
- Van Meel, L. (1987). *Contribution à la limnologie de quatre grands lacs du Zaïre oriental: Tanganika, Kivu, Mobutu Sese Seko (ex Albert), Idi Amin Dada (ex Edouard) : Les paramètres chimiques* [Google-Books-ID: NXIqAAAAMAAJ]. Edition de l'Institut royal des sciences naturelles de Belgique.
- Vandelannoote, A., Deelstra, H., & Ollevier, F. (1999). The inflow of the Rusizi River to Lake Tanganyika. *Hydrobiologia*, 407(0), 65–73. <https://doi.org/10.1023/A:1003701901217>
- Vandelannoote, A., Robberecht, H., Deelstra, H., Vyumvuhore, F., Bitetera, L., & Ollevier, F. (1996). The impact of the River Ntahangwa, the most polluted Burundian affluent of Lake Tanganyika, on the water quality of the lake. *Hydrobiologia*, 328(2), 161–171. <https://doi.org/10.1007/BF00018713>

- Vantrepotte, V., Loisel, H., Meriaux, X., Neukermans, G., Dessailly, D., Jamet, C., Gensac, E., & Gardel, A. (2011). Seasonal and inter-annual (2002-2010) variability of the suspended particulate matter as retrieved from satellite ocean color sensor over the French Guiana coastal waters. *Journal of Coastal Research, SI 64*.
- Verbyla, D. L. (1995). *Satellite Remote Sensing of Natural Resources*. CRC Press. <https://doi.org/10.1201/9780138740191>
- Vercruyssen, K., Grabowski, R. C., & Rickson, R. (2017). Suspended sediment transport dynamics in rivers: Multi-scale drivers of temporal variation. *Earth-Science Reviews, 166*, 38–52. <https://doi.org/10.1016/j.earscirev.2016.12.016>
- Viewing Chlorophyll a — Algorithm Publication Tool. (n.d.). Retrieved April 3, 2024, from https://www.earthdata.nasa.gov/apr/documents/chlor-a/v1.0#additional_information
- Wang, C., Li, W., Chen, S., Li, D., Wang, D., & Liu, J. (2018). The spatial and temporal variation of total suspended solid concentration in Pearl River Estuary during 1987–2015 based on remote sensing. *Science of The Total Environment, 618*, 1125–1138. <https://doi.org/10.1016/j.scitotenv.2017.09.196>
- Well, v., P., & Chapman, D. W. (1976). Summary of limnological observations in Lake Tanganyika near Kigoma 1974-1975. *FI:DP/URT/71/012/39, F.A.O., Rome, 17 pp*.
- Wilks, D. S. (2006). *Statistical methods in the atmospheric sciences* (2nd ed). Elsevier.
- Wongchuig, S., Kitambo, B., Papa, F., Paris, A., Fleischmann, A. S., Gal, L., Boucharel, J., Paiva, R., Oliveira, R. J., Tshimanga, R. M., & Calmant, S. (2023). Improved modeling of Congo's hydrology for floods and droughts analysis and ENSO teleconnections. *Journal of Hydrology: Regional Studies, 50*, 101563. <https://doi.org/10.1016/j.ejrh.2023.101563>
- Yacobi, Y. Z., Gitelson, A., & Mayo, M. (1995). Remote sensing of chlorophyll in Lake Kinneret using highspectral-resolution radiometer and Landsat TM: Spectral features of reflectance and algorithm development. *Journal of Plankton Research, 17*(11), 2155–2173. <https://doi.org/10.1093/plankt/17.11.2155>
- Yang, H., Kong, J., Hu, H., Du, Y., Gao, M., & Chen, F. (2022). A Review of Remote Sensing for Water Quality Retrieval: Progress and Challenges [Number: 8 Publisher: Multidisciplinary Digital Publishing Institute]. *Remote Sensing, 14*(8), 1770. <https://doi.org/10.3390/rs14081770>
- Zhang, M., Tang, J., Dong, Q., Song, Q., & Ding, J. (2010). Retrieval of total suspended matter concentration in the Yellow and East China Seas from MODIS imagery. *Remote Sensing of Environment, 114*, 392–403. <https://doi.org/10.1016/j.rse.2009.09.016>
- Zhang, Y., Shi, K., Liu, X., Zhou, Y., & Qin, B. (2014). Lake Topography and Wind Waves Determining Seasonal-Spatial Dynamics of Total Suspended Matter in Turbid Lake Taihu, China: Assessment Using Long-Term High-Resolution MERIS Data [Publisher: Public Library of Science]. *PLOS ONE, 9*(5), e98055. <https://doi.org/10.1371/journal.pone.0098055>
- Zhou, L., Roberts, D. A., Ma, W., Zhang, H., & Tang, L. (2014). Estimation of higher chlorophylla concentrations using field spectral measurement and HJ-1A hyperspectral satellite

data in Dianshan Lake, China. *ISPRS Journal of Photogrammetry and Remote Sensing*, 88, 41–47. <https://doi.org/10.1016/j.isprsjprs.2013.11.016>

A Appendix

A.1 Limnological sampling

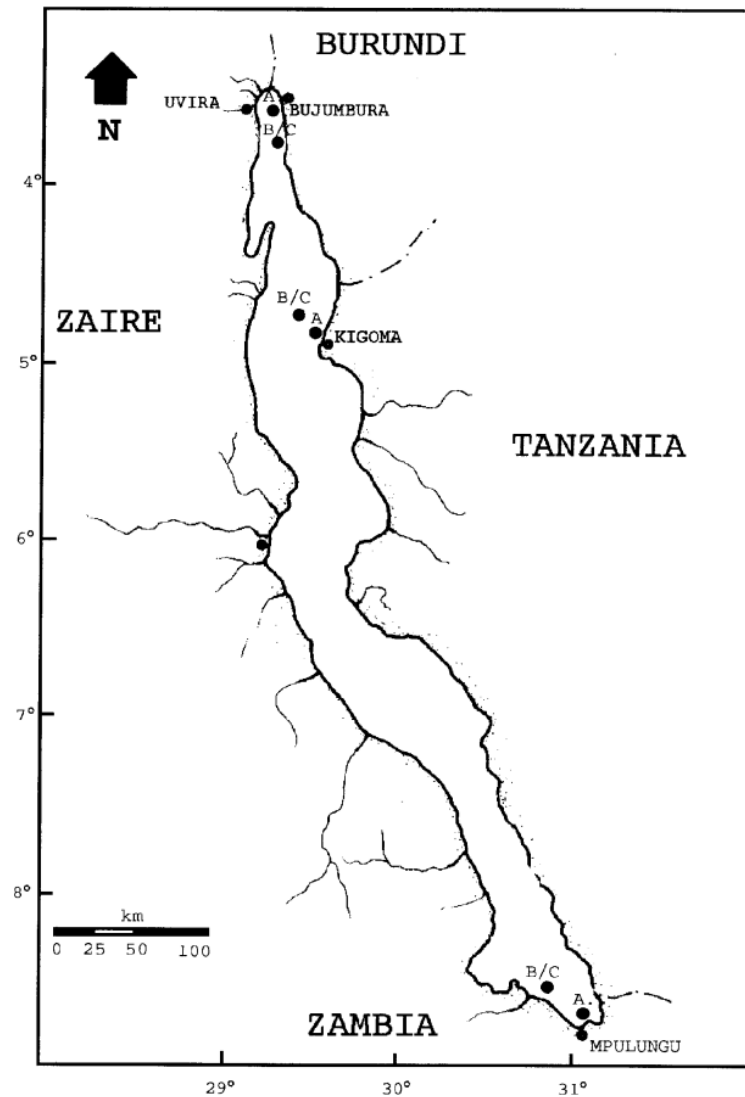


Figure 42: Location of sites A, B, and C for limnological sampling near each station (A=regular sampling down to 100m, B = intensive sampling every 6 weeks (24h cycle), C = seasonal sampling (every 3 months) at the same sites as B) (Plisnier et al., 1996)

Location	Sampling sites	Latitude	Longitude
Buj-Uvi.	A/H1	3°28.00' S	29°17.00'E
	B/C	3°45.00' S	29°15.00 E
	V	3°28.68' S	29°16.18 E
Kigoma	A/H1	4°51.26' S	29°35.54'E
	B	4°50.69' S	29°34.65'E
	V	4°51.50' S	29°34.50'E
Mpulungu	A/H1	8°43.98' S	31°02.43'E
	B/C	8°34.45' S	30°50.10'E
	V	8°41.22' S	30°59.80'E

Figure 43: Coordinates of sampling sites as recorded by GPS (Plisnier et al., 1996)

A.2 Files included in Landsat PAR

Table 18: Files included in Landsat C2 Provisional AR products

Data Type	Description
Aquatic Reflectance (AR)	Represents the nondimensional normalized water-leaving remote sensing reflectance, assuming a perfectly Lambertian surface.
Rayleigh-corrected Reflectance (RHORC)	Represents the intermediate Rayleigh-corrected reflectance, which is used to generate AR.
Atmospheric auxiliary bands	Represents water vapor, surface pressure, wind speed, tropospheric nitrogen dioxide, and vertical columnar ozone used as input information for atmospheric correction calculations.
Elevation (HEIGHT)	Represents the auxiliary digital elevation model used as input for atmospheric correction calculations.
Processing Flags (L2_FLAGS)	A bit-packed band that provides per-pixel information about the success or failure of processing and the validity of sun glint, view angle, solar angle, polarization, chlorophyll-a, etc.
Level-1 Pixel Quality Band (QA_PIXEL)	The bit combinations that define certain quality conditions. More information about Pixel Quality Assessment can be found in the Landsat 8-9 Collection 2 Level-2 Science Product Guide.
Water Mask (WATER_MASK)	Classified raster consisting of pixels identified as land (0), water (1), cloud (2), cloud shadow (3), and snow (4) used for masking non-water pixels in the RHORC and AR bands.
Solar and Sensor angles (SAA, SZA, VAA, VZA, SCATTANG)	Per-pixel Level-1 solar/viewing zenith and azimuth angles in addition to the scattering angle based on these angles.
Angle Band Coefficients File (ANG)	Text (.txt) document containing input information to generate the per-pixel sensor and sun viewing angle bands.
Metadata	Includes Aquatic Reflectance Science Product information in XML format (Product_ID.xml) and Level-1 metadata both in .txt and XML format.

A.3 Bujumbura rivers

Four main rivers, flowing from north to south—Kinyankonge, Ntakangwa, Muha, and Kanyosha—pass through the urban area before emptying into Lake Tanganyika (Sibomana et al., 2024)

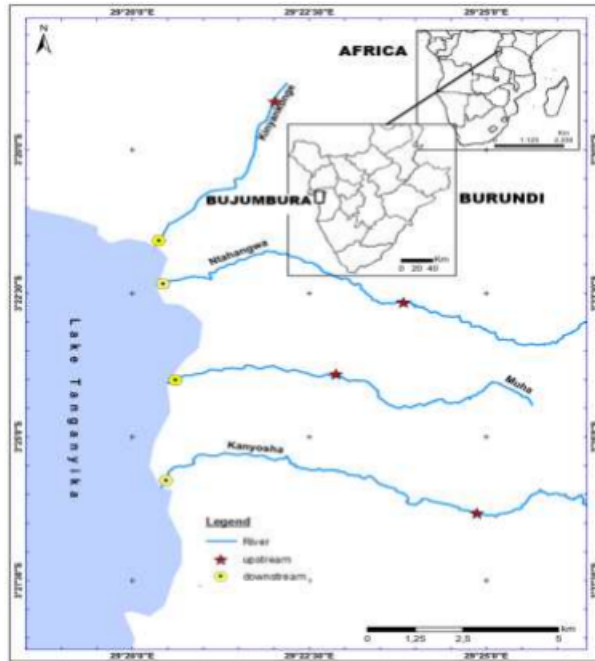


Figure 44: Map of the rivers in Bujumbura (Sibomana et al., 2024)

A.4 Algorithms optimization considering CCI Lakes uncertainty

In this section, the optimization process was conducted while considering the uncertainty associated with the CCI Lakes products. For turbidity, a threshold of 70 percent was chosen to balance the quality of the observations with the number of available observations. With this threshold, only 23 percent of the global dataset was considered, corresponding to 6904 pixels.

For ChloA, a threshold of 60 percent was chosen. With this threshold, 71.2 percent of the dataset was considered, corresponding to 20643 pixels.

Table 19: Optimization results and optimized equation of Turbidity algorithms (see Table 10) between CCI Lakes with uncertainty < 70 and Landsat PAR resampled at 1km

Equation	R ²	RMSE	MAE
$Turb = 0.195 NDTI + 0.940$	0.01	0.4	0.3
$Turb = 79.108 TSMi + 0.375$	0.04	0.1	0.3
$Turb = 28.846 B4 + 0.721$	0.04	0.7	0.5
$Turb = 4.261 B4 - 2.592 B4^2 + 0.921$	0.02	0.4	0.3
$Turb = 4.261 B4 - 2.592 B4^2 - 0.375 B4^3 + 0.921$	0.02	0.4	0.3
$Turb = 1.504 B4^{0.052}$	0.01	0.4	0.3
$Turb = 10^{(113.946(B3+B4) - 0.708(\frac{B2}{B3}) - 0.154)}$	0.04	7.8	7.6

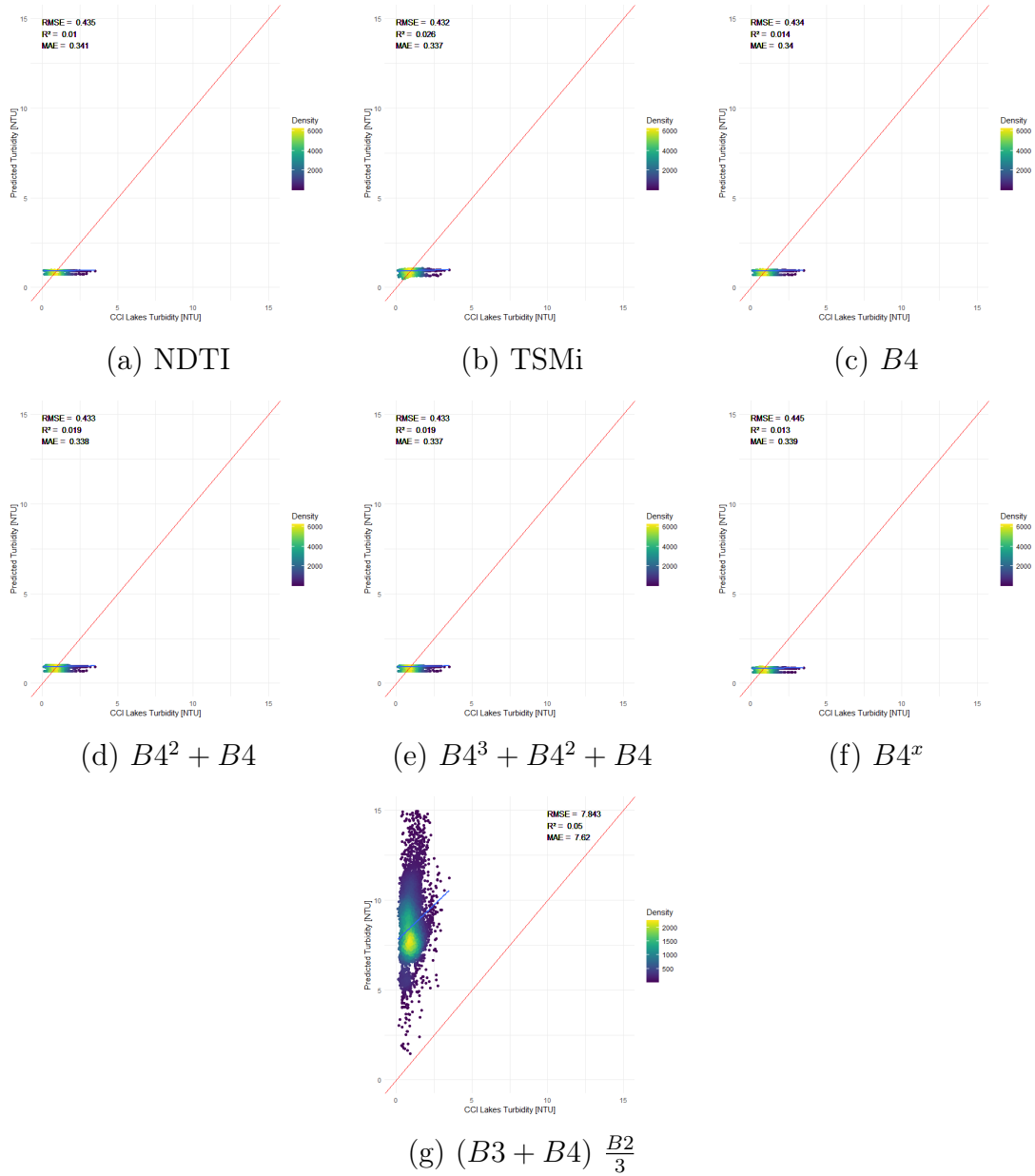


Figure 45: Linear regression analysis between CCI Lakes Turbidity with uncertainty < 70 and retrieved Landsat 8/9 PAR Turbidity using different algorithms. Individual points are color-coded based on the number of neighboring points, with the red line representing the identity line : (a) Linear regression of NDTI. (b) Linear regression of TSMi. (c) Linear regression of Band 4. (d) Linear regression of second polynomial of Band 4. (e) Linear regression of third polynomial of Band 4. (f) Linear regression of $B4^x$. (g) Linear regression of $(B3 + B4) \frac{B2}{3}$.

Table 20: Optimization results and optimized equation of Chlorophyll-A algorithms (see Table 11) between CCI Lakes with uncertainty < 70 and Landsat PAR resampled at 1km

Equation	R ²	RMSE	MAE
$\log_{10}(ChloA) = (-19.527 x + 14.678 x^2 - 6.758 x^3 - 1.447 x^4 - 0.143)$, $x = \log_{10}\left(\frac{Rrs_{482}}{Rrs_{561}}\right)$	0.03	5.8	0.8
$\log_{10} = (-19.527 x + 14.678 x^2 - 6.758 x^3 - 1.447 x^4 - 0.143)$, $x = \log_{10}\left(\frac{\max[Rsr_{443}, Rrs_{482}]}{Rrs_{561}}\right)$	0.03	5.8	0.8
$\log_{10}(ChloA) = (OC3_{O'Reilly})$	0.02	6	1.1

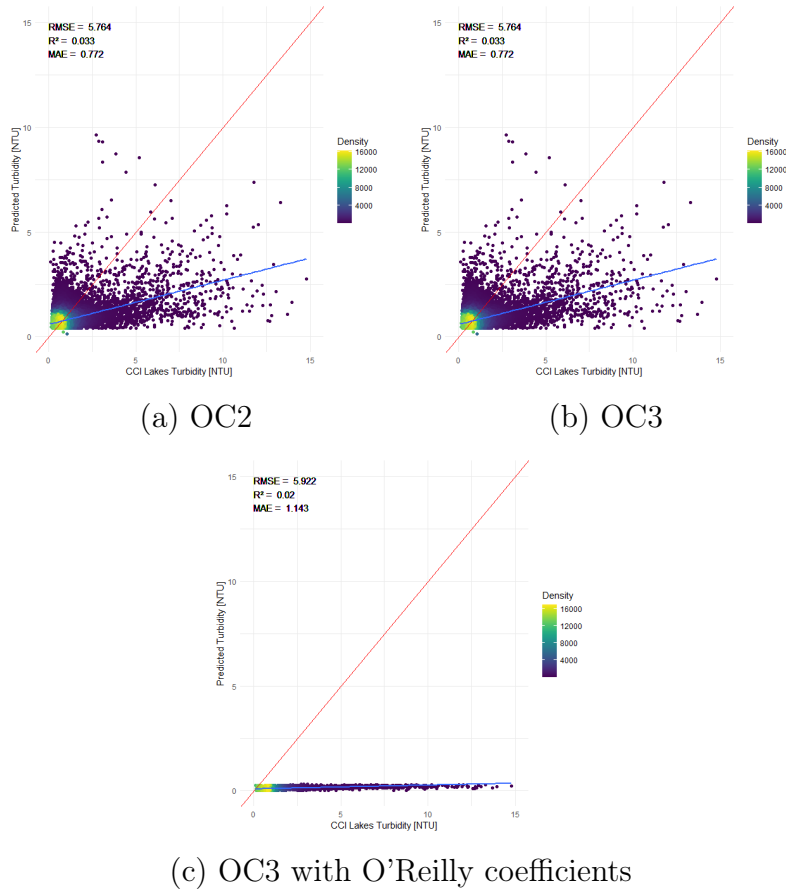


Figure 46: Linear regression analysis conducted between CCI Lakes ChloA with uncertainty < 60 and retrieved Landsat 8/9 PAR ChloA using different algorithms. Individual points are color-coded based on the number of neighboring points, with the red line representing the identity line : (a) Linear regression between ChloA CCI Lakes and the predicted ChloA using the OC2 equation optimized. (b) Linear regression between ChloA CCI Lakes and the predicted ChloA using the OC3 equation optimized. (c) Linear regression between ChloA CCI Lakes and the predicted ChloA using the OC3 equation with coefficients fitted by J. E. O'Reilly and Werdell, 2019

A.5 Number of observations for each variable and each region

A.5.1 Bujumbura

In general, for TSMi, there are fewer observations during the wet season (October to April) compared to the dry season (May to September). Additionally, there is a spatial pattern observed, with more observations located near the coastal area of Bujumbura.

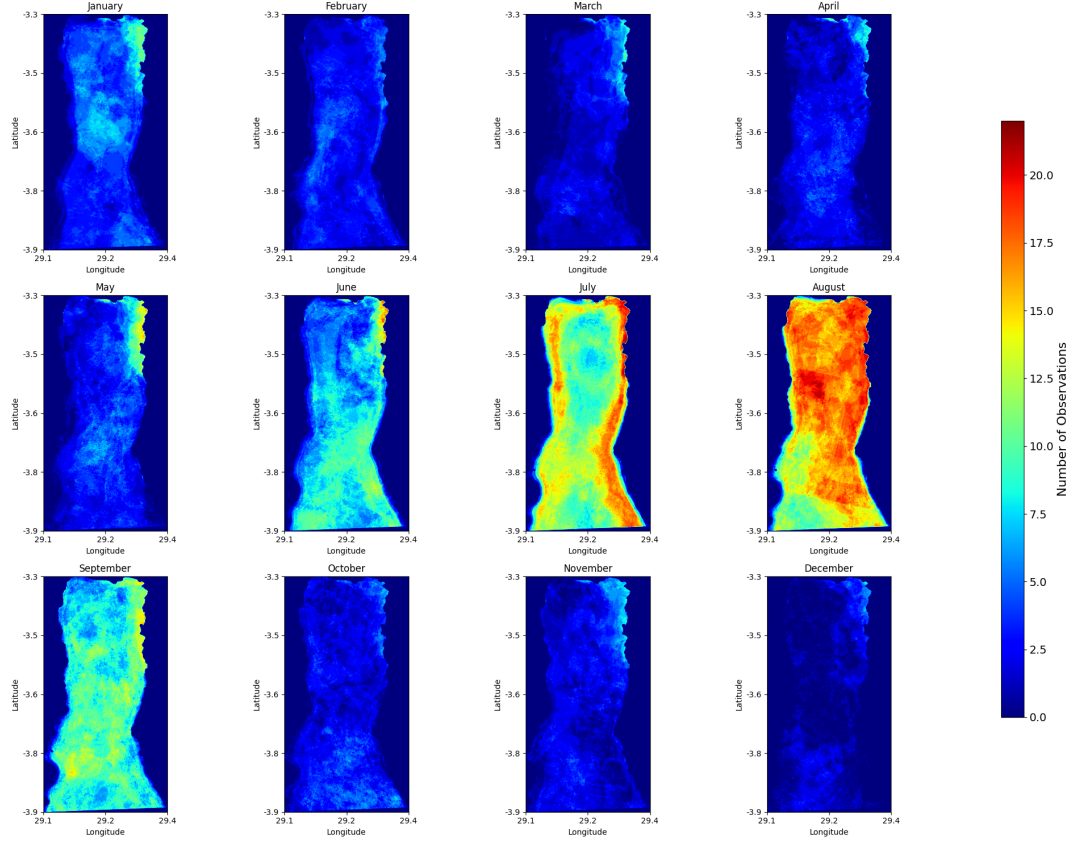


Figure 47: Number of observations by pixel for each month based on the TSMi dataset from 2013 to 2023 for the region of Bujumbura

The ChloA dataset demonstrates a temporal pattern akin to that of the TSMi dataset, displaying more observations during the dry season compared to the wet season. Spatially, the ChloA dataset shows a more uniform distribution, with fewer pixels lacking observations. Additionally, apart from January, there isn't a significantly higher number of observations in the coastal area of Bujumbura.

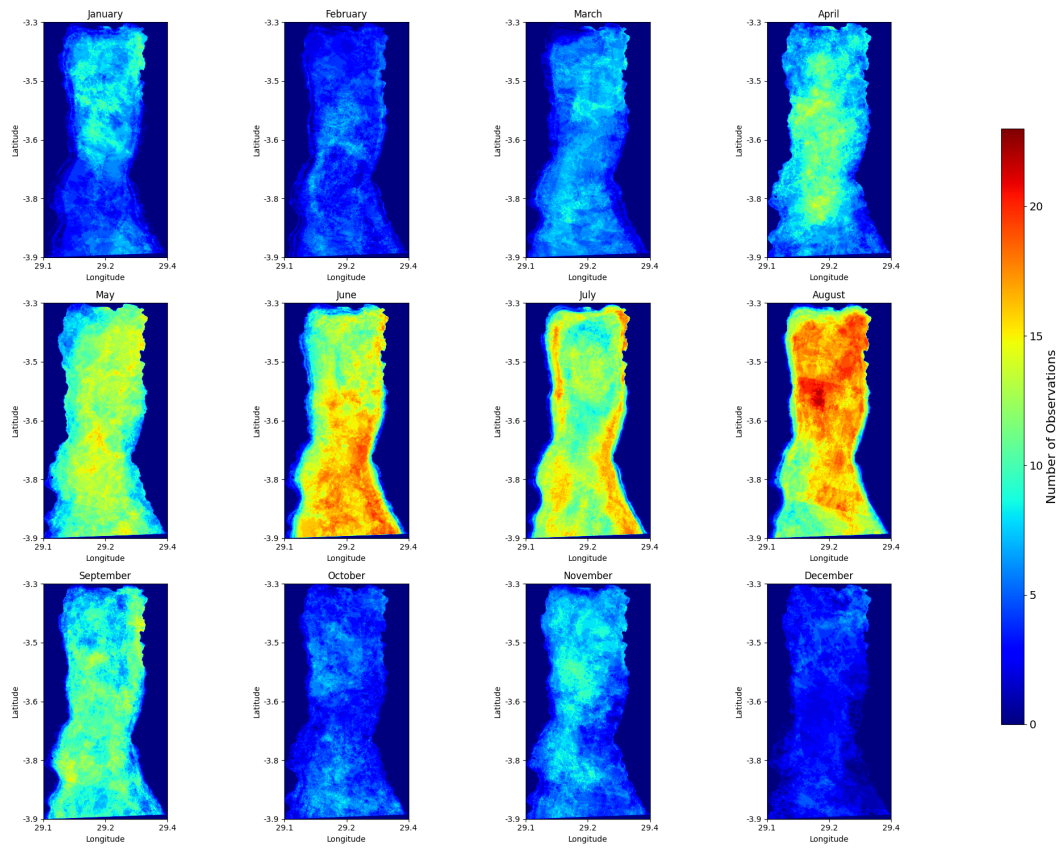


Figure 48: Number of observations by pixel for each month based on the ChloA dataset from 2013 to 2023 for the region of Bujumbura

A.5.2 Kigoma

In general, the TSMi dataset exhibits fewer observations during the wet season (October to April) compared to the dry season (May to September). During the dry season, there are more observations observed near the estuarine area of the Malagarasi River. This pattern is also noticeable in January, March, and October, albeit to a lesser extent.

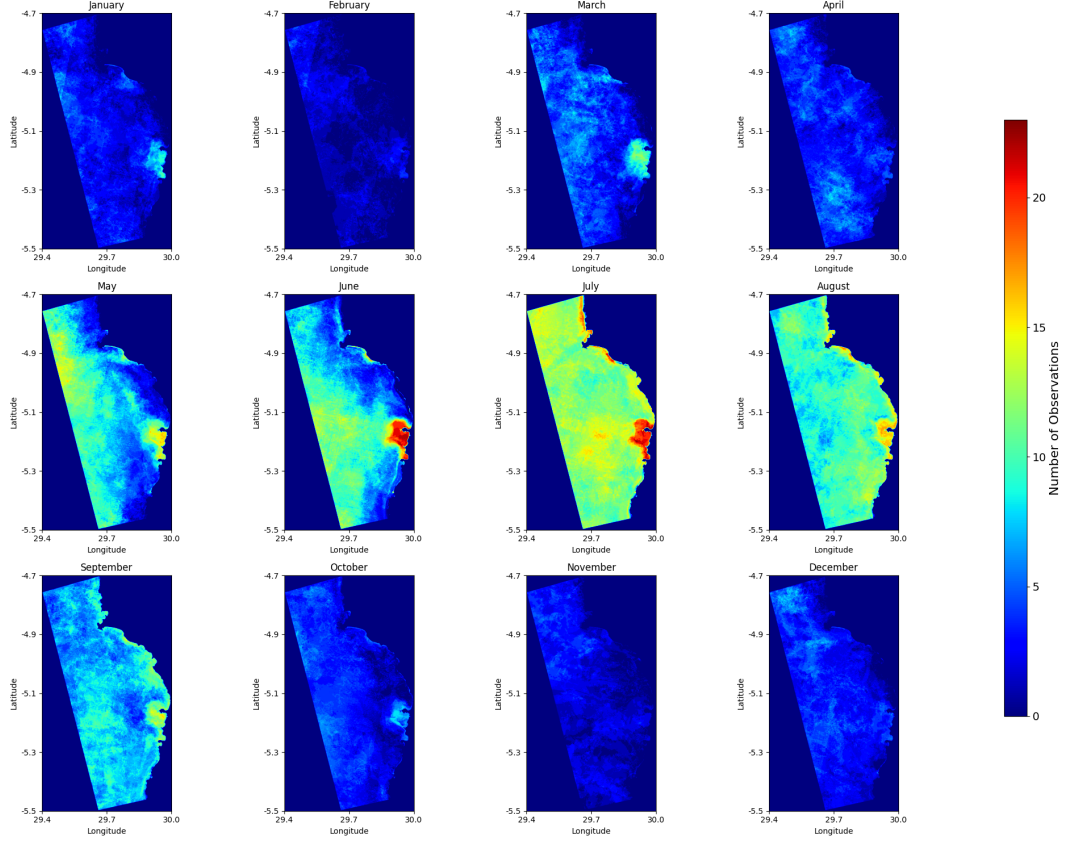


Figure 49: Number of observations by pixel for each month based on the TSMi dataset from 2013 to 2023 for the region of Kigoma

The ChloA dataset follows a temporal pattern akin to the TSMi dataset, indicating more observations during the dry season than the wet season. Spatially, ChloA exhibits a more uniform distribution, with fewer pixels lacking observations. Moreover, there isn't a notably higher number of observations in the estuarine area of the Malagarasi River.

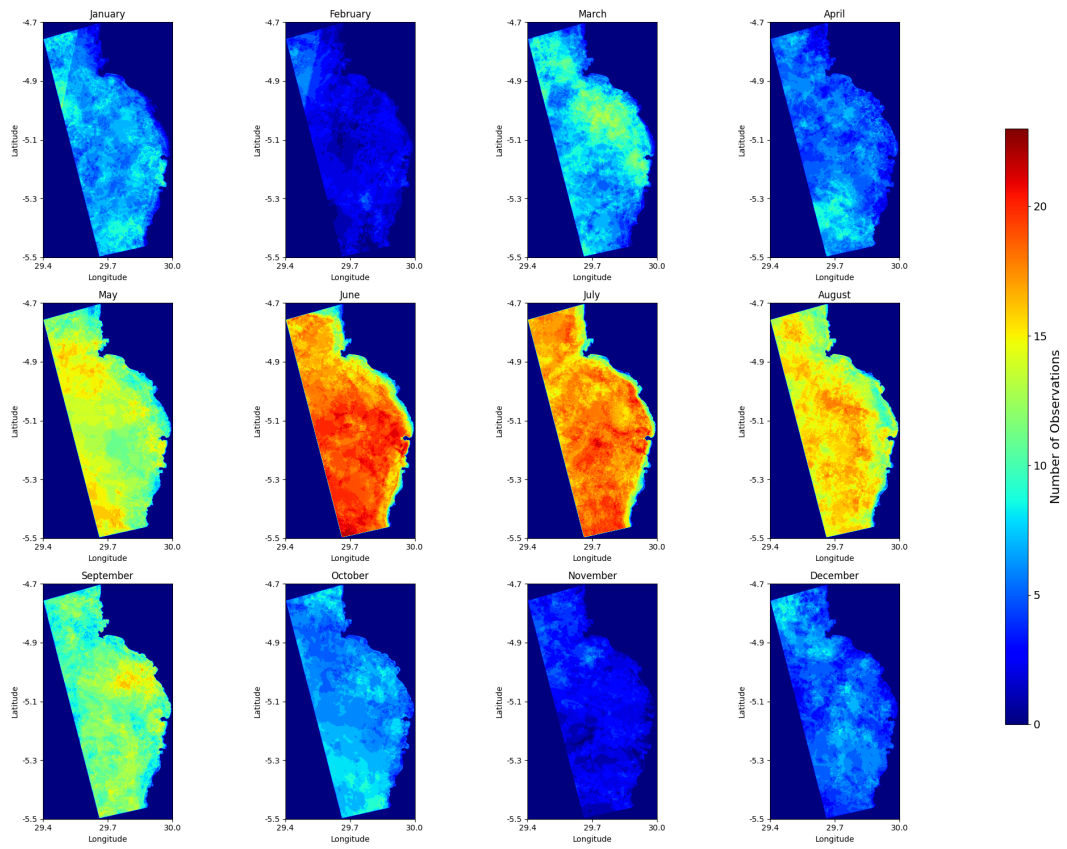


Figure 50: Number of observations by pixel for each month based on the ChloA dataset from 2013 to 2023

A.6 EOF

A.6.1 Spatial modes

A.6.1.1 Bujumbura

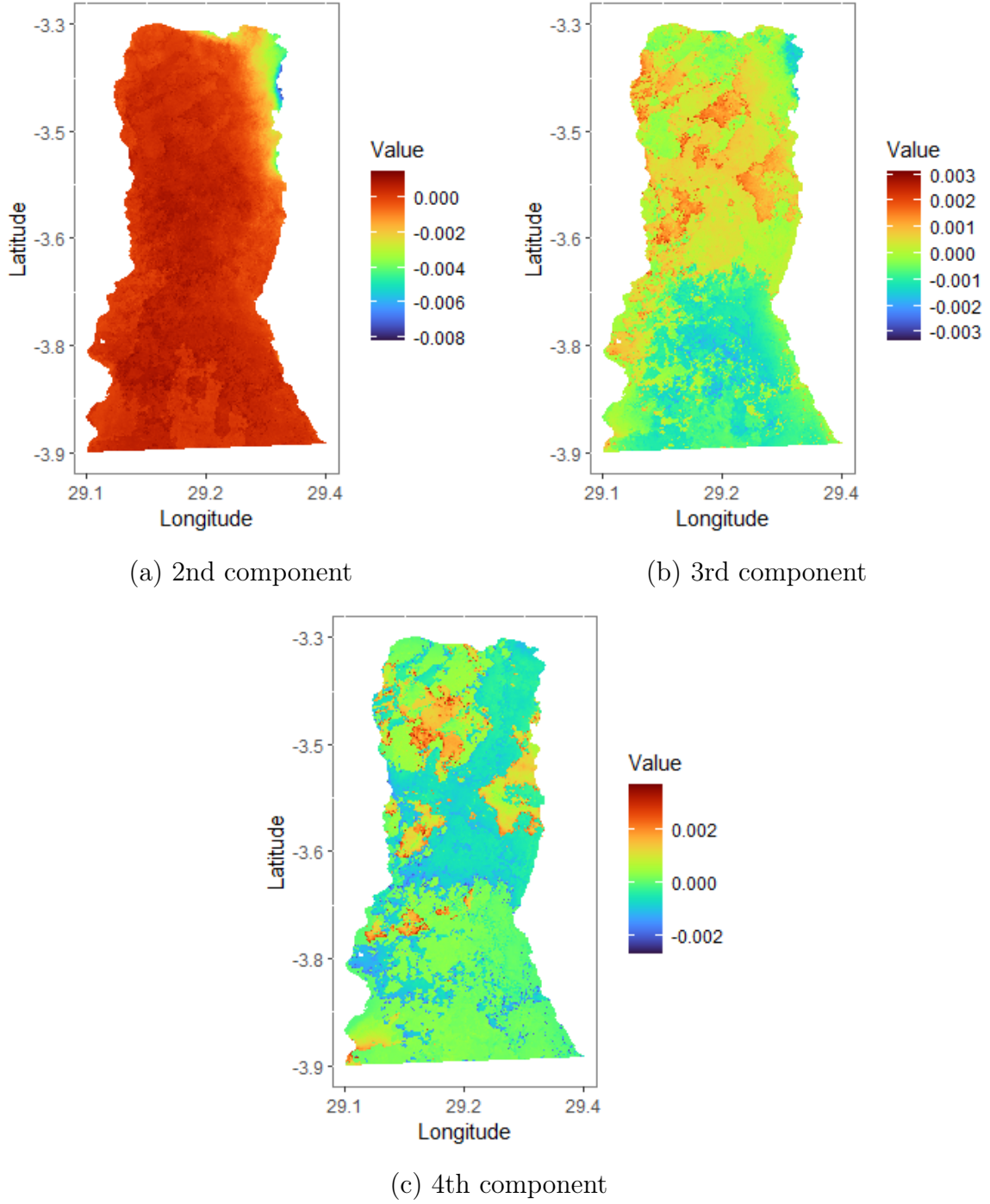
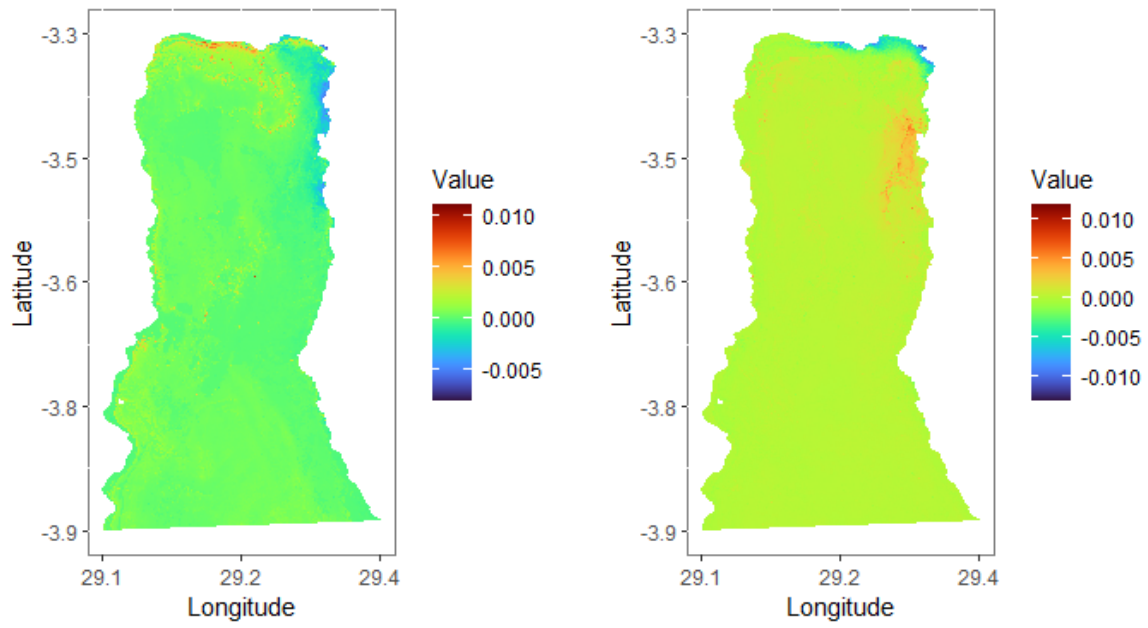
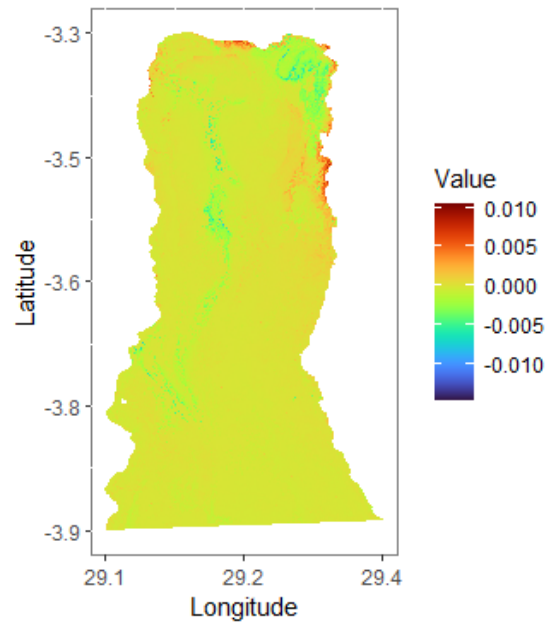


Figure 51: Spatial mode of the (a) 2nd, (b) 3rd and (c) 4th components of the turbidity dataset in the region of Bujumbura



(a) 2nd component

(b) 3rd component



(c) 4th component

Figure 52: Spatial mode of the (a) 2nd, (b) 3rd and (c) 4th components of the ChloA dataset in the region of Bujumbura

A.6.1.2 Kigoma

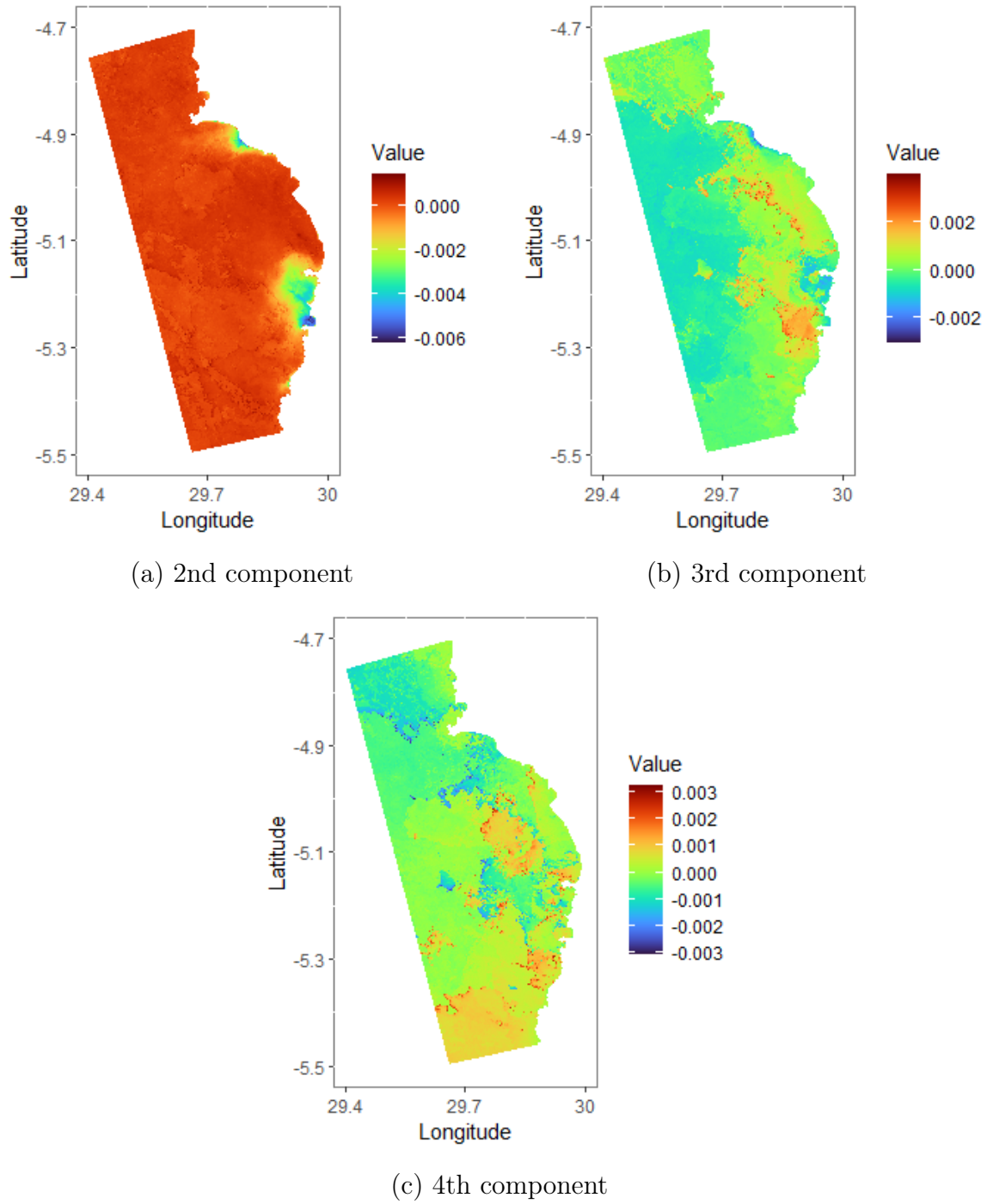
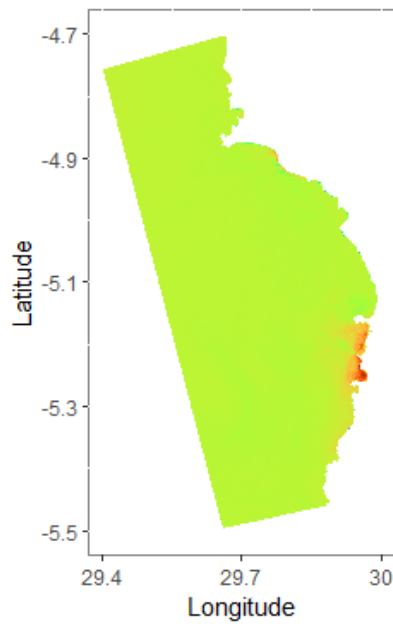
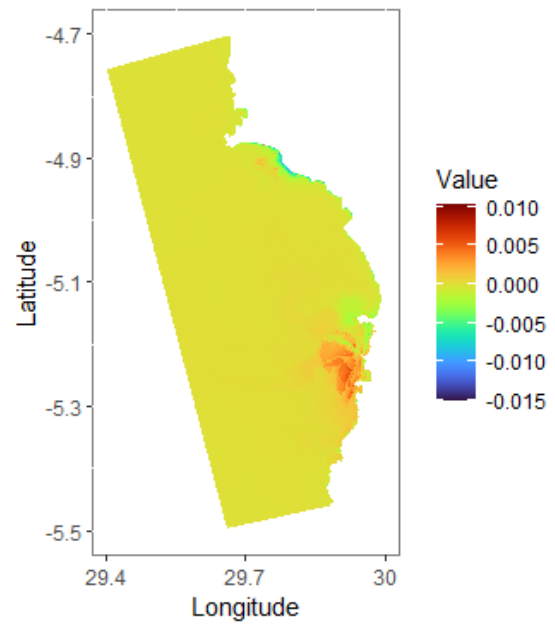


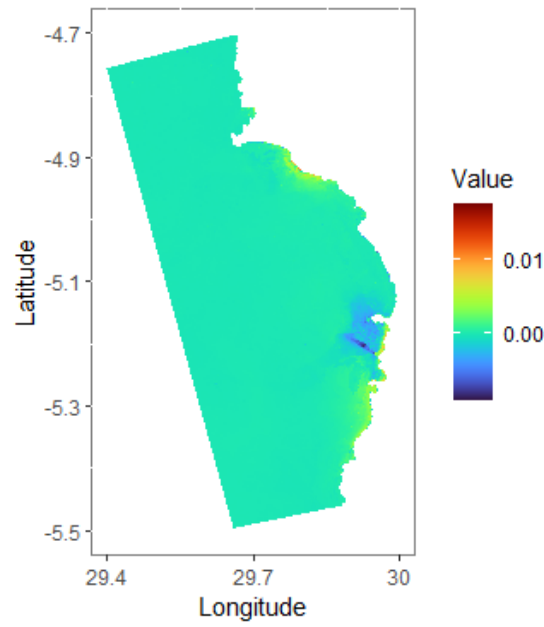
Figure 53: Spatial mode of the (a) 2nd, (b) 3rd and (c) 4th components of the turbidity dataset in the region of Kigoma



(a) 2nd component



(b) 3rd component

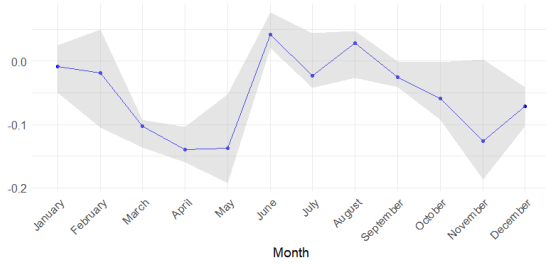


(c) 4th component

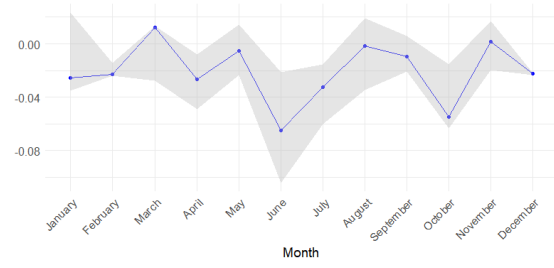
Figure 54: Spatial mode of the (a) 2nd, (b) 3rd and (c) 4th components of the ChloA dataset in the region of Kigoma

A.6.2 Temporal mode

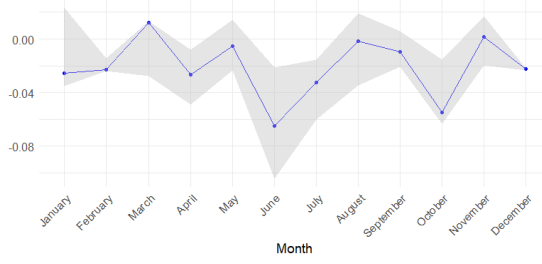
A.6.2.1 Bujumbura



(a) 2nd component

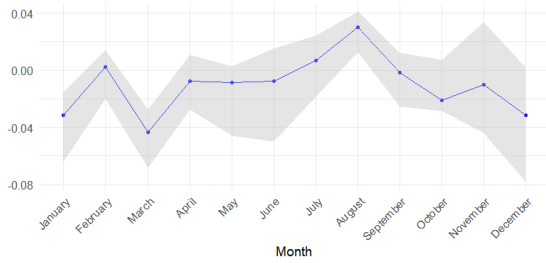


(b) 3rd component

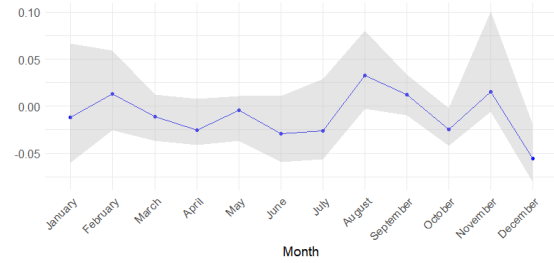


(c) 4th component

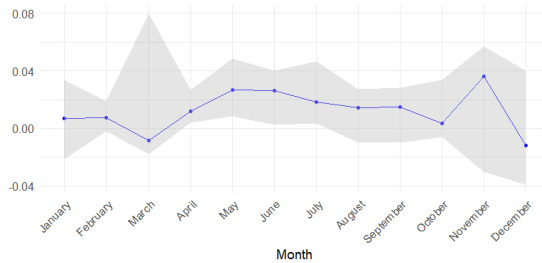
Figure 55: Monthly median, 5th and 95th percentiles of the temporal mode of the (a) 2nd, (b) 3rd and (c) 4th components of the TSMi dataset in the region of Bujumbura



(a) 2nd component



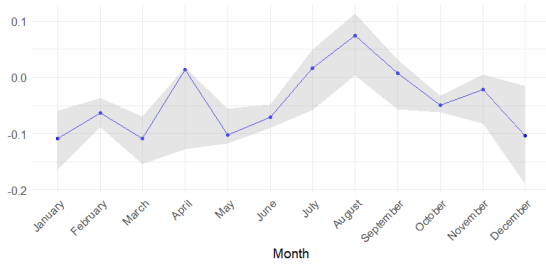
(b) 3rd component



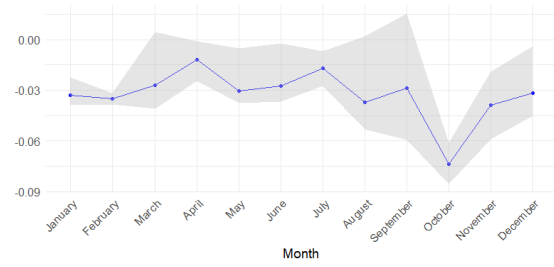
(c) 4th component

Figure 56: Monthly median, 5th and 95th percentiles of the temporal mode of the (a) 2nd, (b) 3rd and (c) 4th components of the ChloA dataset in the region of Bujumbura

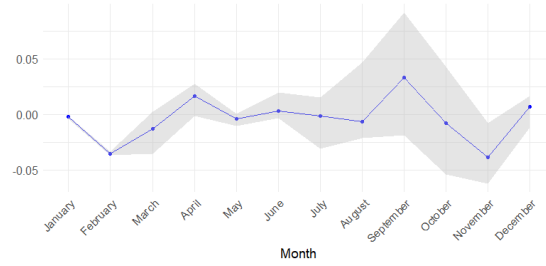
A.6.2.2 Kigoma



(a) 2nd component

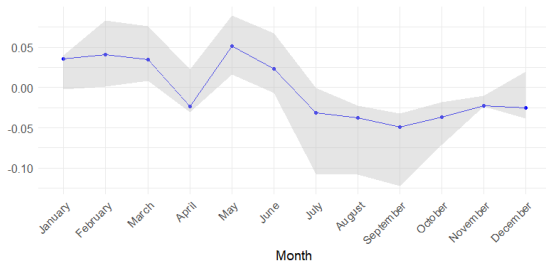


(b) 3rd component

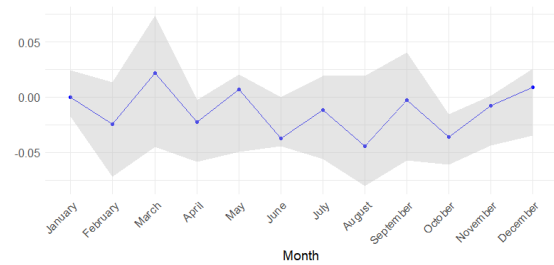


(c) 4th component

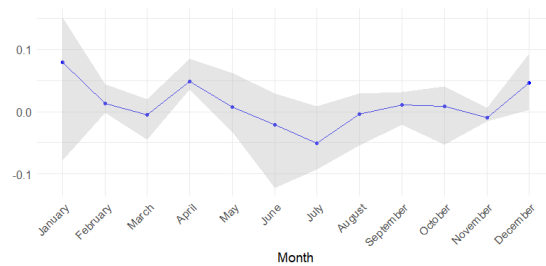
Figure 57: Monthly median, 5th and 95th percentiles of the temporal mode of the (a) 2nd, (b) 3rd and (c) 4th components of the TSMi dataset in the region of Kigoma



(a) 2nd component



(b) 3rd component



(c) 4th component

Figure 58: Monthly median, 5th and 95th percentiles of the temporal mode of the (a) 2nd, (b) 3rd and (c) 4th components of the ChloA dataset in the region of Kigoma

A.7 Kmeans clustering

A.7.1 Bujumbura

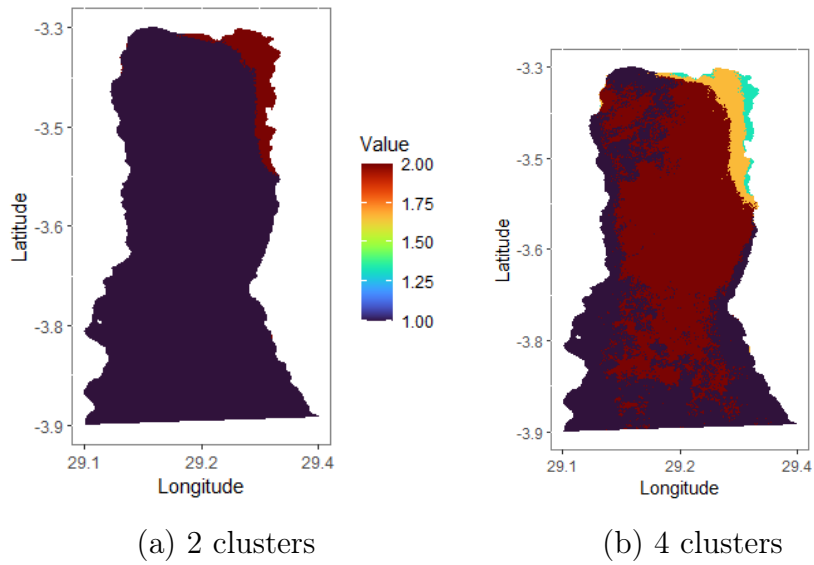


Figure 59: Cluster of the Bujumbura region with (a) Two clusters and (b) Four clusters

A.7.2 Kigoma

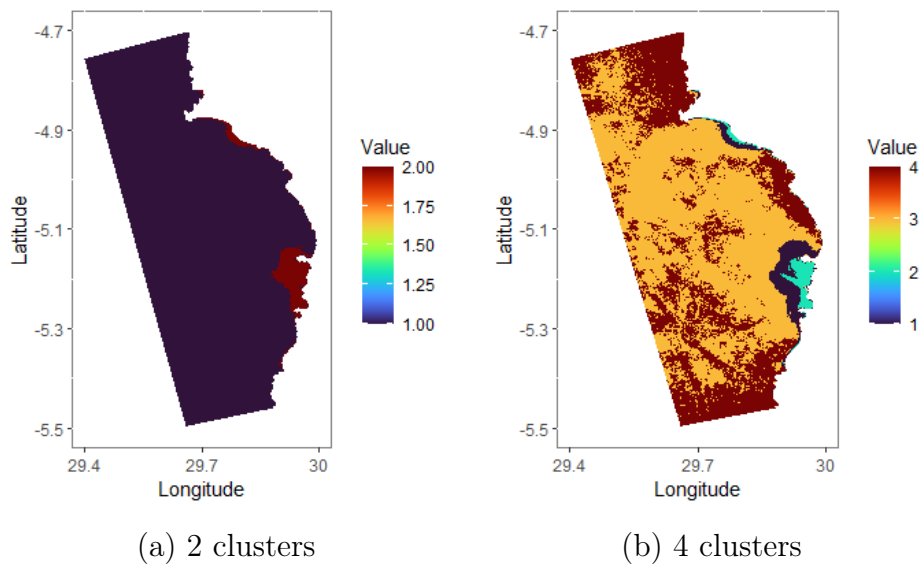


Figure 60: Cluster of the Kigoma region with (a) Two clusters and (b) Four clusters

A.8 Linear regression between the daily median of ChloA and TSMi with the river flow

A.8.1 Bujumbura

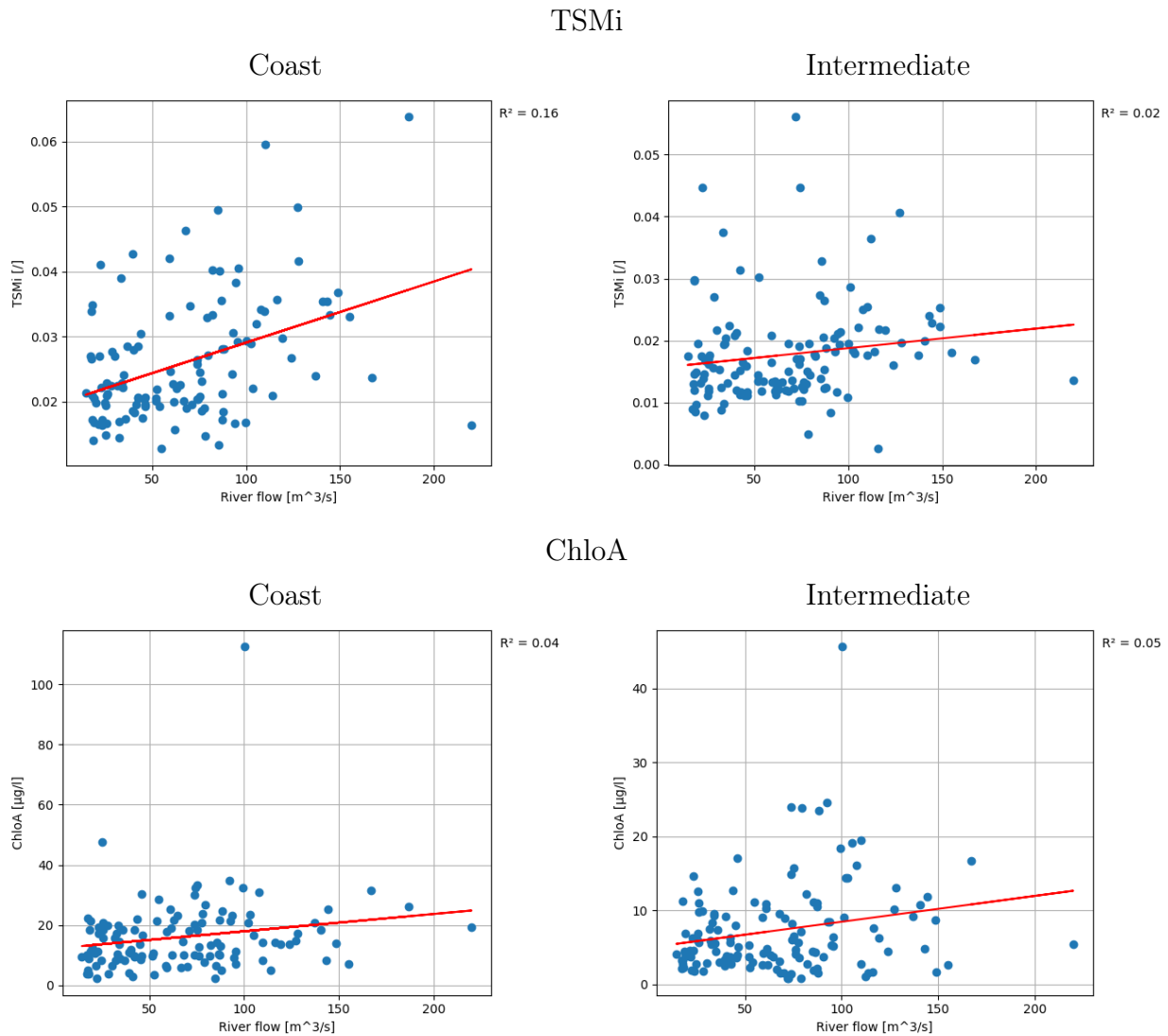


Figure 61: Linear regression between the daily median of ChloA and Turbidity in the Coast and Intermediate cluster of the Bujumbura region and the Ruzizi river flow of the same day derived from Wongchuig et al., [2023](#).

A.8.2 Kigoma

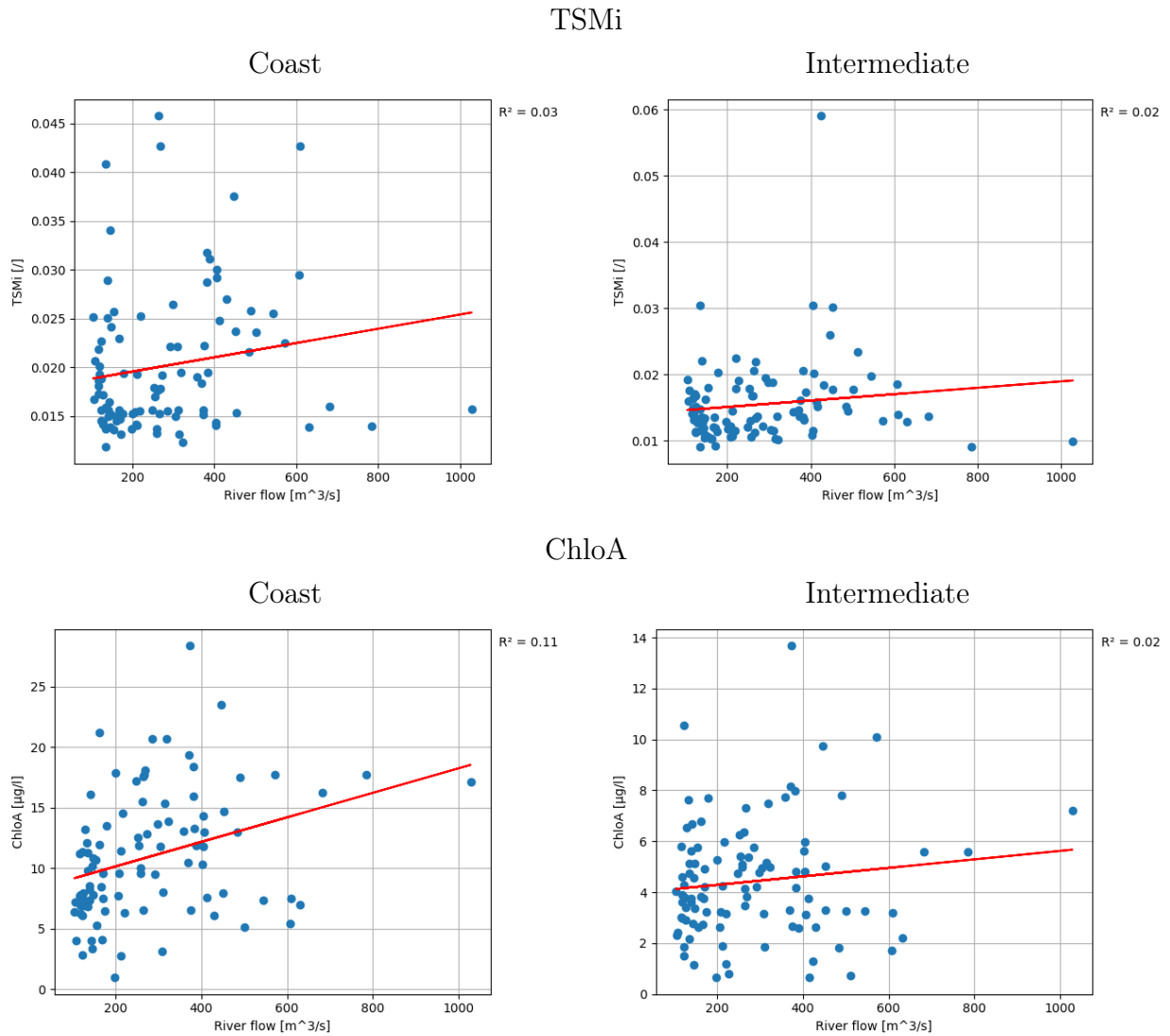


Figure 62: Linear regression between the daily median of ChloA and Turbidity in the Coast and Intermediate cluster of the Kigoma region and the Malagarasi river flow of the same day derived from Wongchuig et al., 2023.

Identification of the drivers of water quality spatio-temporal variability in the estuarine areas of Lake Tanganyika

Aurélien Lengrand

Large-scale deforestation, urban expansion, and inappropriate agricultural practices in Lake Tanganyika's watershed lead to sediment and nutrient influx into the lake, degrading water quality. Monitoring water quality in the lake's estuarine areas is crucial for understanding ecosystem health and promoting sustainable management practices. This study aimed to analyze the spatio-temporal dynamics and identify drivers of chlorophyll-A and turbidity in the estuaries of the Malagarasi and Ruzizi rivers, key tributaries of Lake Tanganyika. Using Landsat 8/9 Provisional Aquatic Reflectance with a fine spatial resolution (30m) alongside CCI Lakes Chlorophyll-A and turbidity datasets (1km spatial resolution), efforts were made to optimize chlorophyll-A and turbidity algorithms. However, no conclusive results were obtained. This underscores the need for improved atmospheric correction and algorithm choice. To address this challenge, chlorophyll-A and turbidity time series were built using an already optimized algorithm for chlorophyll-A and a turbidity index. Spatial analyses revealed distinct patterns of chlorophyll-A and turbidity dynamics, emphasizing the local impacts of river mouths and coastal areas on water quality. Temporal analyses showed a seasonal correlation between river flow and the dynamics of both variables, with higher concentrations observed during the wet season. Disparities between chlorophyll-A and turbidity dynamics and river flow highlighted the need for additional water variables for a comprehensive understanding. This study emphasizes the necessity of remote sensing satellites like Sentinel-2 and Sentinel-3 to further explore and monitor estuarine water quality dynamics.

Université catholique de Louvain

Faculté des bioingénieurs

Croix du Sud, 2bte L7.05.01, 1348 Louvain-La-Neuve, Belgique | www.uclouvain.be/agro