ISPRS Journal of Photogrammetry and Remote Sensing 223 (2025) 58–77

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

journal homepage: www.elsevier.com/locate/isprsjprs

A novel cyanobacteria occurrence index derived from optical water types in a tropical lake

Davide Lomeo [a,*], Stefan G.H. Simis [b], Xiaohan Liu [b], Nick Selmes [b], Mark A. Warren [b], Anne D. Jungblut [c], Emma J. Tebbs [a]

[a] Department of Geography, King's College London, London, United Kingdom
[b] Plymouth Marine Laboratory, Plymouth, United Kingdom
[c] Department of Life Sciences, Natural History Museum, London, United Kingdom

A R T I C L E I N F O

Keywords:
Cyanobacteria
Cyanobacteria Occurrence Index
Optical Water Types
Inland Waters
Tropical Lakes
Ocean Color

A B S T R A C T

Cyanobacteria blooms are a threat to water quality of lakes and reservoirs worldwide, requiring scalable monitoring solutions. Existing approaches for remote sensing of cyanobacteria focus on quantifying (accessory) photosynthetic pigment to map surface accumulations. These approaches have proven challenging to validate against in situ observations, limiting uptake in water quality management. Optical Water Types (OWTs) have been used in inland and ocean waters to dynamically select suitable algorithms over optical gradients, thereby helping to limit out-of-scope application of individual algorithms. Here, we present a proof-of-concept study in Winam Gulf, Lake Victoria, extending an existing OWT framework using a hybrid approach combining in situ and satellite-derived water types. This extended OWT set of 25 water types, obtained from K-means clustering $>$ 18 million Sentinel-3 Ocean and Land Colour Instrument (OLCI) spectra, was found to better capture the optical diversity of cyanobacteria bloom phases compared to the original OWT set. We translate this framework into a novel Cyanobacteria Occurrence Index (COI) by assigning weights to key optical features observed in the OWT set, such as phycocyanin absorption and surface accumulation. COI was strongly correlated with established algorithms for chlorophyll-*a* (Maximum Peak Height; r = 0.9) and phycocyanin (Simis07; r = 0.84), while potentially capturing various bloom phases in optically mixed conditions. We demonstrate how COI could be mapped onto a three-category risk classification to facilitate communication of cyanobacteria occurrence risk. Initial tests across diverse waterbodies suggest potential for wider application, though further validation across different environmental conditions is needed. This work provides a foundation for improved cyanobacteria monitoring in optically complex waters, particularly where conventional sampling approaches face limitations.

1. Introduction

Cyanobacteria blooms are a growing global concern due to their significant impact on water quality in freshwater ecosystems. Blooms of cyanobacteria have increased in frequency and intensity worldwide due to eutrophication, enhanced by climate change and intensified anthropogenic activities (Taranu et al., 2015; Huisman et al., 2018; Fang et al., 2022). This is problematic because cyanobacteria blooms can have significant ecological impacts. They can alter the structure of aquatic food webs (Paerl et al., 2011), create unfavourable conditions for other non-buoyant aquatic organisms through shading (Paerl & Otten, 2013; Huisman et al., 2018), and their collapse can lead to oxygen depletion in water, potentially resulting in fish kills (Paerl & Paul, 2012). Furthermore, some species of cyanobacteria produce potent toxins that pose significant health risks to animals, humans, and aquatic life (Paerl & Paul, 2012; Merel et al., 2013; Christensen & Khan, 2020).

Remote sensing has emerged as a complementary method for monitoring cyanobacteria in freshwater ecosystems. Several remote sensing studies focus on detecting diagnostic pigments, such as phycocyanin (PC) and chlorophyll-a (Chla), to map cyanobacteria occurrence. These approaches quantify the absorption or fluorescence signatures of these pigments (Dekker, 1993; Kutser et al., 2006; Hunter et al., 2010; Matthews, 2011; Pahlevan et al., 2020; O'Shea et al., 2021; Pahlevan et al., 2021b; Smith et al., 2021). In freshwater environments, PC is

* Corresponding author.
E-mail addresses: davide.lomeo@kcl.ac.uk (D. Lomeo), stsi@pml.ac.uk (S.G.H. Simis), liux@pml.ac.uk (X. Liu), nse@pml.ac.uk (N. Selmes), mark1@pml.ac.uk (M.A. Warren), a.jungblut@nhm.ac.uk (A.D. Jungblut), emma.tebbs@kcl.ac.uk (E.J. Tebbs).

https://doi.org/10.1016/j.isprsjprs.2025.03.006
Received 21 October 2024; Received in revised form 8 March 2025; Accepted 10 March 2025
Available online 12 March 2025

typically only associated with the presence of cyanobacteria and red algae, and a useful indicator of cyanobacteria in the presence of eutrophication pressure. Drawbacks of using PC as a water quality indicator are a lack of standardised pigment extraction techniques, and a high degree of cellular pigment variability due to regulation of accessory pigment production (Tandeau De Marsac, 1977). Optical techniques to quantify PC use wavebands in the 610–650 nm range, where the pigment has distinct absorption and fluorescence features. Chla, found in all plants, algae, and cyanobacteria, is used in remote sensing of water as a proxy of phytoplankton biomass, albeit noted that the cellular concentration of Chla is typically lower in cyanobacteria compared to eukaryotic phototrophs (Johnsen & Sakshaug, 1996). Depending on the presence of other optically active substances, Chla detection typically relies on absorption properties in the blue region around 443 nm (in relatively clear waters), or in the red to near-infrared (NIR) regions between 670 and 700 nm (in waters with higher turbidity or dissolved organic matter).

Remote sensing pigment concentration retrieval methods for optically complex water (including all inland and most coastal water bodies) are typically based on waveband combinations. In ocean colour sensors, like the Medium Resolution Imaging Spectrometer (MERIS) and the Ocean and Land Colour Instrument (OLCI), the presence of cyanobacteria via PC light absorption can be captured in the 620 nm waveband, referenced against 709 nm where pigment absorption is low (Simis et al., 2005) or other neighbouring wavebands in a line height approach (Dekker, 1993). The waveband ratio of 709 nm and 665 nm is sensitive to Chla biomass (Gons, 1999, 2002; Gilerson et al., 2010). Other algorithms use the 681 nm waveband to quantify the Chla fluorescence peak, or the shape of the reflectance spectrum in the same spectral region, to define indicators sensitive to both lower Chla fluorescence yield per unit pigment, and higher light scattering efficiency by cyanobacteria (Wynne et al., 2008; Matthews et al., 2012; Matthews & Odermatt, 2015). The above set of waveband combinations are typically modelled with in situ data to create diagnostic pigment concentration retrieval algorithms, with some limitations. Firstly, PC algorithms based in light absorption can be confounded by the presence of other pigments and dissolved organic matter (Ruiz-Verdú et al., 2008). This inherently leads to inaccuracies in cyanobacteria biomass estimates, which can be overestimated in the presence of other phytoplankton groups (Dekker et al., 2001; Dall'Olmo & Gitelson, 2005; Simis et al., 2007). Additionally, although some of these algorithms provide realistic estimates of (near) surface bloom extent, they become less reliable in optically mixed waters where cells are less concentrated in the first optical depth (typically 1 m in clear waters) and mixed throughout the water column (Wynne et al., 2010). Therefore, whilst the underlying mechanism of empirical or semi-analytical band ratio algorithms provides a reliable indication of pigment presence and biomass, they tend to lack robust application to optically varying water conditions.

The effectiveness of pigment concentration retrieval algorithms is further challenged by atmospheric effects. Accurate retrieval of water-leaving reflectance relies on the performance of atmospheric correction algorithms, directly affecting the reliability of downstream products (Pahlevan et al., 2021a). Large uncertainty associated with atmospheric correction is typical in inland and coastal waters. Notably, ineffective near infrared light absorption by water in the presence of dense surface biomass accumulations or high turbidity challenges the ability of atmospheric correction algorithms to separate water leaving radiance from atmospheric effects (Moses et al., 2009; IOCCG, 2010). The inherent variability in atmospheric conditions and water optical properties continues to pose challenges for consistent and accurate atmospheric correction across optically diverse waterbodies. This variability, combined with the limitations of pigment-specific algorithms, has prompted more flexible approaches that can adapt to varying optical conditions as well as systematic biases.

Spectral end-member mapping techniques and the development of spectral libraries have been used to address the wide optical complexity of natural waters (Kent & Mardia, 1988; Wang, 1990; Moore et al., 2009, 2014). These methods acknowledge that individual algorithms are unlikely to be valid across the full optical diversity of natural waters. *A priori* classification of remotely sensed observations into Optical Water Types (OWTs) has brought further solutions to deal with the optical complexity of water bodies from the local to the global scale, from oceans to inland water observations (Moore et al., 2001; Eleveld et al., 2017; Jackson et al., 2017; Liu et al., 2021). Fuzzy classification of observations to OWTs can facilitate the selection and subsequent weighted averaging of individual algorithm predictions, each performing well within a given range of biogeochemical properties but with larger uncertainties beyond this scope. Consequently, this dynamic approach enhances the reliability of remote sensing products, providing more accurate assessments of water quality indicators such as phytoplankton biomass and suspended sediments (Neil et al., 2019; Kravitz et al., 2021), and even providing means to propagate algorithmic uncertainty to these products (Liu et al., 2021).

Significant progress has been made in developing globally representative OWT frameworks for inland waters, with Spyrakos et al. (2018) establishing a comprehensive set of 13 water types derived from in situ hyperspectral measurements from temperate systems. This library has proven valuable for algorithm selection and blending globally, and it is currently used as the baseline OWT classification for the European Space Agency (ESA) Lakes Climate Change Initiative (Simis et al., 2022, Liu et al. 2021). A current limitation of this framework is that its definitions of OWTs that make reference to cyanobacteria are based on a limited geographic sample, potentially underrepresenting cyanobacteria dynamics at lower latitudes. This is problematic, because light and nutrient conditions influence accessory photosynthetic pigment regulation in cyanobacteria (Tandeau De Marsac, 1977; Wyman & Fay, 1986b, 1986a; Grossman et al., 1993); the intracellular pigment concentrations may therefore be expected to vary regionally and with latitude.

The definition of OWTs from in situ hyperspectral reflectance typically provides detailed biogeochemical attributions, but there are inherent challenges in obtaining representative in situ measurements of cyanobacteria blooms. Near- or at-surface accumulations are easily disturbed during sampling, with ship movement and wind action rapidly altering the vertical distribution of biomass. The short-lived nature of many bloom events (often lasting days to weeks) further complicates the collection of comprehensive optical measurements. Moreover, any given atmospheric correction algorithm must accurately represent these measurements across all optical manifestations of cyanobacteria, including near- or at-surface biomass accumulation, which introduces additional complexity. These sampling and measurement challenges may partly explain why the current OWT framework for inland waters by Spyrakos et al. (2018) may not fully capture cyanobacteria bloom dynamics. While three of the thirteen OWTs are associated with high concentrations of Chla and PC (and one explicitly mentions cyanobacteria blooms), and four others exhibit varying pigment concentrations, these types are often dominated by other optical features such as suspended sediments or coloured dissolved organic matter.

There is an opportunity to leverage the extensive spatial and temporal coverage of satellite observations to expand the current OWT definitions, particularly where obtaining representative in situ measurements of cyanobacteria blooms is challenging. Here, we extend the 13 OWTs for inland waters by Spyrakos et al. (2018) to Sentinel-3 OLCI-derived spectra of Winam Gulf, a shallow and highly eutrophic basin of Lake Victoria, the world's largest tropical lake. Lake Victoria has vast social and economic importance in the region (e.g. through fisheries, drinking water), and has documented history of recurring cyanobacteria blooms and cyanotoxins (Hecky et al., 2010; Sitoki et al., 2010; Simiyu et al., 2018; Mchau et al., 2019; Brown et al., 2024). By combining these established in situ derived OWTs with new types obtained directly from satellite observations, we test whether cyanobacteria occurrence can consistently be recorded within atmospherically corrected OLCI imagery, and whether the definition of complementary and optimised OWTs successfully recognise the various stages of cyanobacteria bloom. This approach inevitably introduces uncertainty in the biogeochemical attribution of satellite-derived OWTs, particularly for surface accumulations that are difficult to sample representatively. However, we may expect better recognition of bloom conditions that are challenging to capture. Where current gaps in OWT classification exist, we hypothesise that an extended set may enhance cyanobacteria monitoring capabilities in similar tropical environments, and potentially in other regions where traditional sampling approaches are limited.

Given the challenges and limitations of diagnostic pigment concentrations retrieval algorithms in optically varying water conditions, an alternative strategy towards monitoring cyanobacteria occurrence drawing further on the OWT classification is explored. This work applies a ranking approach to assign weights to OWT classes where one or more optical features linked to cyanobacteria presence are evident. Particularly, the behaviour of PC absorption captured in the 709 over 620 nm waveband ratio, and the shape of the spectrum around the Chla fluorescence peak described in the 709 over 681 nm waveband ratio, are exploited to distinguish OWT classes that represent a higher risk of cyanobacteria occurrence.

While this approach enhances the detection of cyanobacteria by targeting specific optical features, there remains a need for more integrated and comprehensive assessment to encapsulate the full range of cyanobacteria bloom dynamics, moving beyond traditional pigment concentration retrieval methods where these have thus far proven too variable, unreliable, or lacking validation. Combining past and present efforts, we present a novel Cyanobacteria Occurrence Index (COI) built on our extended water types library and their association with cyanobacteria optical features, into a single metric for cyanobacteria occurrence risk. We test the sensitivity of COI against established pigment concentration retrieval algorithms, demonstrating its ability to capture a broader spectrum of cyanobacteria bloom features and dynamics that are typically individually described by these algorithms.

2. Methods

2.1. Study site

Winam Gulf, also known as Nyanza Gulf or Kavirondo Gulf, is the north-easternmost basin of Lake Victoria (Fig. 1). The Gulf covers an area of approximately 1,400 km^2 and has an average depth of ~ 8 m, compared to 40 m in Lake Victoria proper. The largest river inflows are from Nyando, Sondu and Kibuon to the south-east (Romero et al., 2005; Alexander & Imberger, 2013; Simiyu et al., 2022). Winam Gulf is located in west Kenya, a region that experiences a long and a short rainy season, and a dry season (Dosio et al., 2022). The average annual rainfall is between 600 mm and 2,000 mm, and average daily temperature is between 17 °C and 30 °C (Humphrey et al., 2022). The shallow Winam Gulf experiences full wind mixing which leads to well oxygenated but highly turbid waters, in contrast with the rest of Lake Victoria, where waters are stratified, clearer, and anoxic at depth. These distinct environments are separated by the narrow Rusinga Channel which limits dilution between Winam Gulf and Lake Victoria (Gikuma-Njuru & Hecky, 2005; Alexander & Imberger, 2013; Simiyu et al., 2022). The area around Winam Gulf is one of the most highly populated and intensively cultivated lands in Kenya and around Lake Victoria (Gikuma-Njuru et al., 2013). Tributaries transport large quantities of sediments, pollutants, and nutrients, affecting water quality (Sitoki et al., 2010, 2012). Cyanobacteria are recurrently found in this portion of the lake, with the most common belonging to the genera *Microcystis*, *Cylindrospermopsis*, *Dolichospermum.* The most commonly detected

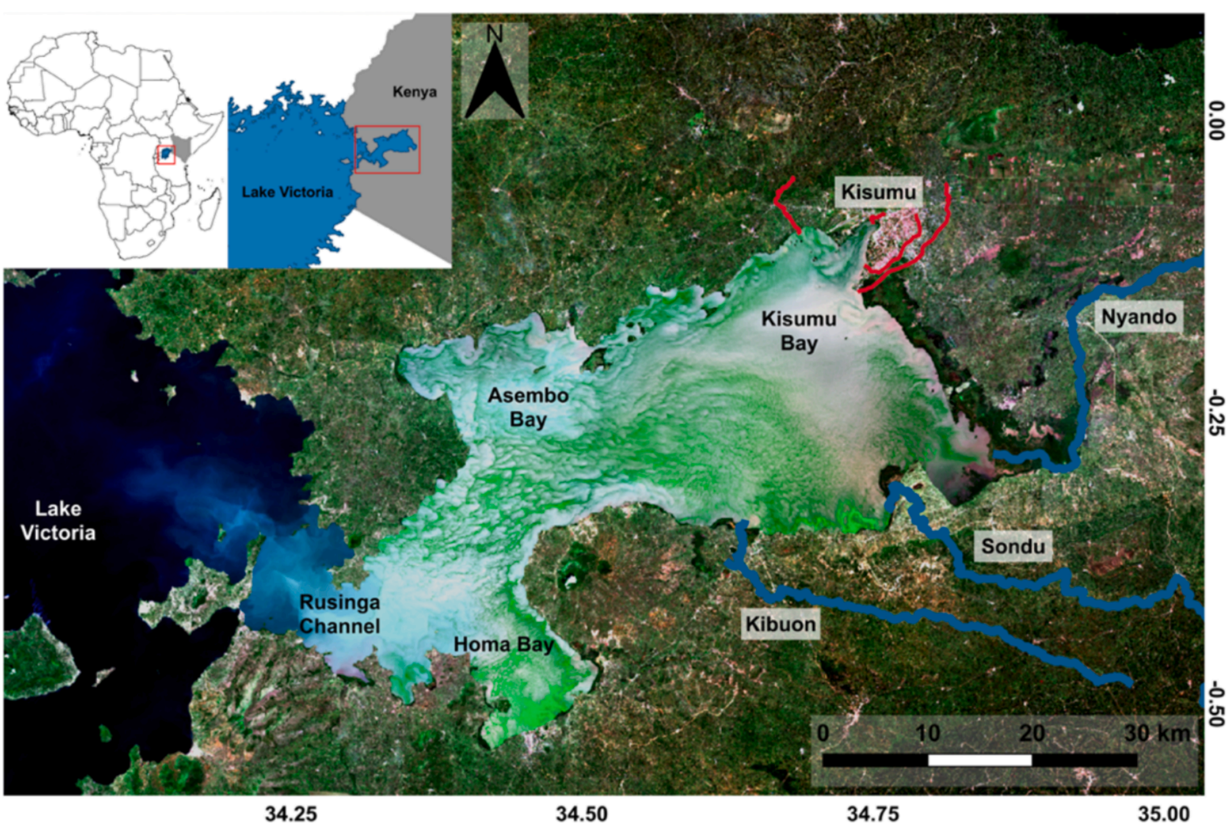

Fig. 1. Level-2A True Colour Composite of Winam Gulf captured by Sentinel-2 Multispectral Instrument (MSI) A on 12 Dec2020. Blue lines show the three major rivers Nyando, Sondu and Kibuon. Red lines show smaller tributaries around Kisumu Bay, three of which flows through Kisumu, the largest city in the area. Bright areas may be due to surface biomass and / or sediments. In these conditions, the Sen2Cor atmospheric correction algorithm used to produce the image may fail to accurately reproduce reflectance.

cyanobacterial toxin is *Microcystin*, with recent findings of *Cylindrospermopsin* (Sitoki et al., 2012; Simiyu et al., 2018; Roegner et al., 2020; Simiyu et al., 2022; Brown et al., 2024).

2.2. Satellite imagery

A total of 1,795 Level 3C 300 m resolution Sentinel-3 Ocean and Land Colour Instrument (OLCI)-A/B images for Winam Gulf were provided by the Natural Environment Research Council Earth Observation Data Analysis and Artificial-Intelligence Service (NEODAAS, UK) for the period 2016–2023. NEODAAS processed the scenes through the *Calimnos* processing chain using the candidate configuration for version 3.0 of the ESA Lakes Climate Change Initiative v2.1 (Lakes_cci – Liu et al., 2021; Simis et al., 2023a, 2023b; Carrea et al., 2024). This new configuration uses Polymer v4.17b for atmospheric correction, which employs an extended range of initialisation conditions compared to previous versions that improves retrieval of turbid water conditions, including near or at-surface blooms. In addition, 331 Level 2A 10-m resolution Sentinel-2 Multispectral Instrument (MSI)-A/B images coinciding with OLCI-A/B images, acquired within 30 min of each other, were obtained from Google Earth Engine Python API for the period 2016–2023. These were provided by ESA and atmospherically corrected using Sen2Cor. The analysis primarily focused on Sentinel-3 OLCI images, with Sentinel-2 MSI images exclusively used for visual referencing of bloom conditions.

2.3. Optical water types classification

The 13 optical water types (OWTs) for inland waters developed by Spyrakos et al. (2018), convolved to OLCI wavebands, were used to map water conditions in Winam Gulf over time. These types were originally formulated using a k-means classifier applied to in situ hyperspectral remote-sensing reflectance originally collated from the Lake Bio-optical Measurements and Matchup Data for Remote Sensing (LIMNADES) community repository. The optimal number of clusters was determined using gap-statistics (Tibshirani et al., 2001). These 13 OWTs are used for algorithm selection and blending in the Lakes_cci Lake Water-Leaving Reflectance (R_w) product set, following atmospheric correction with Polymer as described in Simis et al. (2022) and Liu et al. (2021). A further two optical types, developed to flag pixels affected by land adjacency (Jiang et al., 2023), were also included. Membership similarity score (S_{owt}) values are calculated at each pixel based on the spectral angle metric (Kruse et al., 1993):

$$a = \cos^{-1}\frac{\sum_{i=1}^{n} p_i r_i}{\sqrt{\sum_{i=1}^{n} p_i^2}\sqrt{\sum_{i=1}^{n} r_i^2}} \quad (1)$$

$$S_{owt} = 1 - a/\pi \quad (2)$$

where p_i and r_i are the standardised pixel and reference spectra in band i, respectively. The resultant S_{owt} is a number between 0 and 1, where 1 indicates identical spectral shape. Only the 12 OLCI wavebands in the range 412–779 nm, excluding the oxygen bands at 761, 764 and 767 nm were used. Both OLCI-derived and OWT spectra were standardised prior the calculation of the S_{owt} by dividing over their integrals to reduce the influence of varying reflectance amplitudes and to allow focusing on the similarity of their shapes.

The highest S_{owt}, also referred to as dominant water type (Moore et al., 2014), was used to provide an initial end-member to each observation (pixel). This allowed to broadly differentiate between spectra highly similar to OWTs not typically associated with the presence of cyanobacteria (non-cyano OWT; 2, 3, 5, 9, 10, and 13), and those that are (cyano-OWT; 1, 4, 6, 7, 8, 11, 12), as described in Spyrakos et al. (2018). Approximately, 18 out of 26 million total OLCI-derived spectra had one of the seven cyano-OWT as their dominant member. Spectra for which the dominant OWT was either of the two land effected flags were removed. A summary of the workflow, including the steps outlined in the next sections, is shown in Fig. 2.

2.4. Candidate optical water types

Spectra for which the dominant type was one of the seven cyano-OWT were clustered using unsupervised k-means clustering to determine what spectra captured by OLCI were not represented by the original OWT set. Principal Component Analysis (PCA) was used to reduce the dimensionality of the multispectral data and flag outliers, and only the first components with a combined explained variance exceeding 90 % were selected. K-means clustering was performed on the 18 million PCA values using the gap statistic to determine the optimal number of clusters. The k-means classifier was provided with 50 starting points (i.e., starting from 50 different cluster centres). PCA points outside the 90th percentile Euclidian distance from cluster centres were removed to obtain well-defined clusters.

Spectra belonging to each of the clusters were named after the dominant cyano-OWT they were originally assigned to, alongside the cluster number they belonged to (i.e., 1.1, 7.2, 11.4, etc.). This library of mean standardised 'cyano subtypes' spectra was merged with the convolved in situ library of non-cyano OWT spectra, left untouched from their original formulation, to produce an extended candidate water types library.

S_{owt} values for each combination of OLCI spectra and the OWT set were calculated using Eq. (1) and (2) to determine spectral curve similarities, and then mapped onto a phylogenetic tree. The tree was constructed using the Euclidian distance between S_{owt} cluster centres. From this first iteration, any branch containing fewer than three members was combined in the next iteration, unless visual inspection suggested the types were functionally different (examples are given further below).

S_{owt} values were iteratively re-calculated after combining subtypes, and the membership sum (sum of the S_{owt} set of each observation), normalised to the number of water types, calculated. This metric defined the ability of the new water types set to describe optical variability in the observation dataset. The final iteration of the OWT set was considered reached when no new similar pairs arose during the process. The OWT library was considered efficiently reduced and complete at the point where further reduction of types cause a substantial drop in the membership sum from previous iterations.

2.5. Owt-based cyanobacteria occurrence

A new metric, the weighted S_{owt} sum (W_{sum}), was calculated to determine cyanobacteria occurrence:

$$W_{sum,i} = \sum_j (S_{owt_{i,j}} \times w_{i,j} \times P_{i,j}) \quad (3)$$

with, $.P_{i,j} = \frac{S_{owt_{i,j}}}{\sum_j S_{owt_{i,j}}}$

where $S_{owt_{i,j}}$ is the membership score of water type j at pixel i. $w_{i,j}$ is a subjective ranking informed by the presence of optical indicators of cyanobacteria presence in the OWT spectra. $P_{i,j}$ is a proportionality factor to account for uneven distributions of similarity of a given observations to the OWT set, i.e., the covariance between subtly different OWTs resulting in multiple similarity scores of similar values. The ranking factor was obtained by considering the sum of the 709/620 and 709/681 waveband ratios of the candidate water types library. Visual interpretation of MSI images guided the identification of three threshold values for this sum, whereby water types with a sum of ratios < 2 were assigned $w = 0$, those $>= 3$ were assigned $w = 1000$, and those with intermediate values were assigned $w = 100$. This arbitrary scaling applies an order-of-magnitude differentiation to the ranking factors to reflect the increasing likelihood of cyanobacteria presence with increasing strength of the indicator waveband ratios. The zero rank eliminates any contribution to the combined score in the absence of

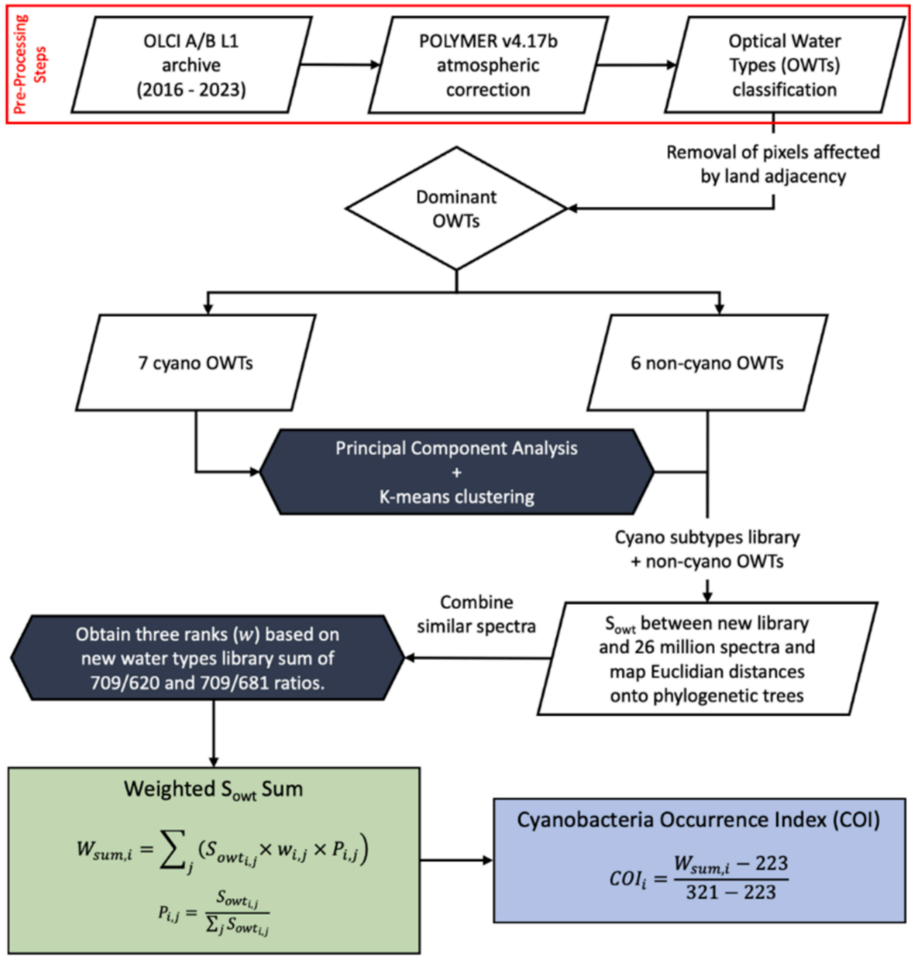

Fig. 2. Flow chart of the generation of a new candidate water types library and translation into a cyanobacteria occurrence index (COI).

likely cyanobacteria indicators, while the higher weights (100, 1000) ensure that high-ranking sum of ratios dominated the final metric.

W_{sum} was finally scaled to a new Cyanobacteria Occurrence Index (COI) to provide a single, reproducible, and easily interpretable metric, defined as:

$$COI_i = \frac{W_{sum,i} - 223}{321 - 223} \tag{4}$$

where $W_{sum,i}$ is the weighted S_{owt} sum at pixel i defined in Eq. (3). The constants 223 and 321 are the minimum and maximum W_{sum} obtained by applying Eq. (3) on the candidate water types library, giving likely boundary values. The resulting COI ranges from 0 to 1, with increasing values indicating increasing likelihood of cyanobacteria occurrence and optical features indicative of productive bloom. It is possible that a different dataset provides W_{sum} values that fall outside the provided boundaries, particularly in the low-end, in which case COI may be set to 0 if $W_{sum} < 223$ and 1 if $W_{sum} > 321$.

3. Results

3.1. OWT definitions

K-means clustering of cyano-OWT derived PCA values of the Winam Gulf observation data set identified 25 candidate spectral clusters (or cyano-subtypes). These subtypes, along with 6 non-cyano OWT which were not re-analysed, formed the initial candidate water types library (Fig. 3). The most notable feature in spectra that were most closely associated with non-cyano OWT is the presence of pigment absorption in the red spectral region, and a NIR backscattering peak within the OWT 13 cluster, indicating that some pixels were classified as this clear water type, yet presented features commonly associated with phytoplankton presence (Fig. 3F). Two to five new sub clusters were identified per cyano-OWT included in the original definitions, confirming a wide range of optical variability currently not captured by the OWT definitions in Spyrakos et al. (2018) (Fig. 3G-M).

The Euclidian distances between cluster centres, mapped onto a first phylogenetic tree revealed nine pairs within the distance threshold of 138 (Fig. 4A). Visual comparison of these pairs revealed no meaningful difference in shapes (e.g., around pigment absorption features) and were combined, whereas two pairs were considered dissimilar and not combined, namely 6.5 and 12.1, and 6.2 and 12.3 (Fig. 4D and E). These two pairs differed in the magnitude of absorption and backscattering at 681 nm and 709 nm, respectively, which could be interpreted as different phases of bloom development, or as optically mixed waters across a bloom gradient. Another pair with similar cluster centres, 11.2 and OWT 5, was not combined, to maintain separation between top-level cyano and non-cyano spectral classes (Fig. 4C). A tenth pair consisting of 4.2 and 6.3 was combined in the second iteration of the analysis, resulting from reclassification after the initial reduction in classes (not shown). The combined cyano subtypes were assigned the name of the first cluster if both originated from the same cyano-OWT (e.g., 7.2 and 7.3), whereas they were assigned the name of both clusters they belonged to if they originated from different cyano-OWT (e.g., 4.1_11.1).

Each time a pair of similar type spectra was combined, S_{owt} values were re-calculated, and the normalised sum of S_{owt} was used to evaluate the performance of the library (Fig. 4B). Each updated library showed

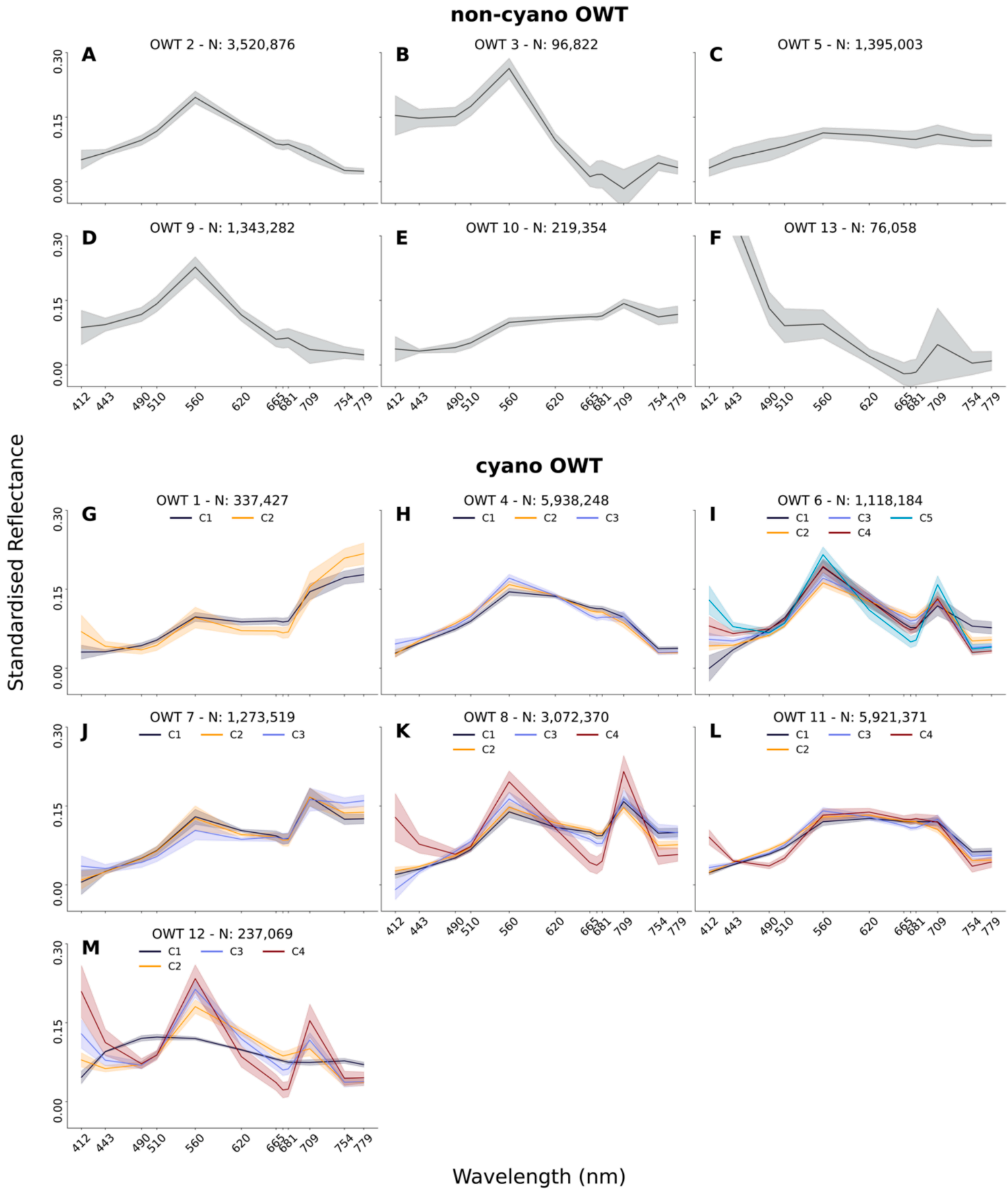

Fig. 3. Standardised spectra of Optical Water Types not associated to cyanobacteria presence (non-cyano OWT) (A-F), and those that are (cyano-OWT) (G-M) after k-means clustering. Clusters are shown as C1, C2, etc. Spectra are grouped by their similarity to the original OWT definitions of Spyrakos et al. (2018). Continuous lines describe mean spectra, shaded areas refer to the standard deviation. Wavebands > 779 nm, not used in the clustering process, are not shown.

better performance than the original library of 13 OWTs, up to and including the 7th iteration, beyond which the sum of membership scores dropped markedly (Fig. 4B). Consequently, the 25 types at iteration 7 were considered the 'optimal' number resulting from this step, comprising 6 non-cyano OWT and 19 cyano subtypes.

3.2. Assignment of risk of cyanobacteria occurrence

Ranking thresholds were defined as $w = 0$ for the group of water types with an average sum of band ratios $= 1.5$ ($\pm$ 0.3), $w = 100$ for water types with an average $= 2.5$ ($\pm$ 0.3), and $w = 1000$ to water types with an average $= 4.9$ ($\pm$ 1.9). The latter was highly skewed by the band ratio values of water types 12.4 and 8.2, at 8.2 and 6.8, respectively.

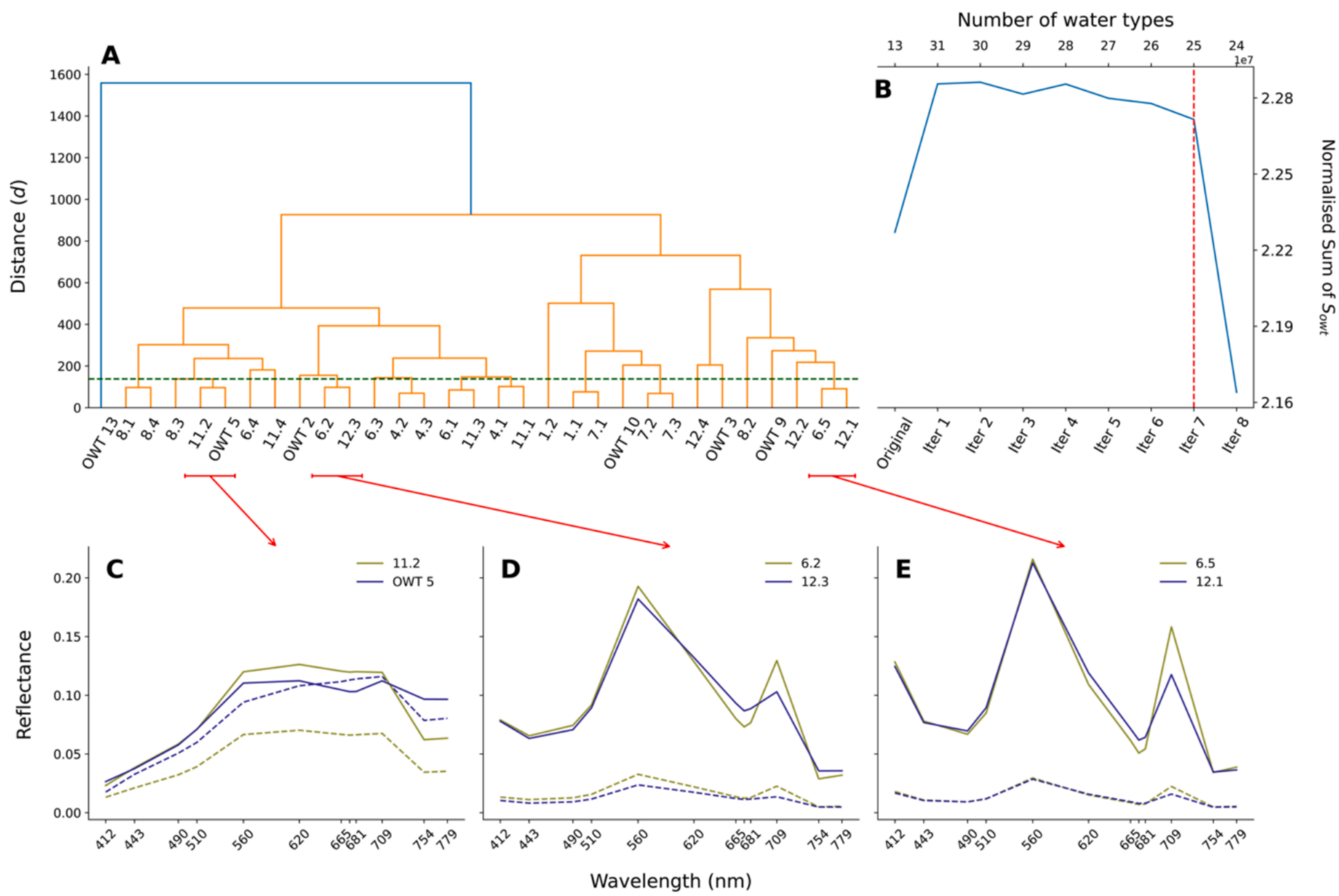

Fig. 4. (A) Euclidian distance tree showing the similarity structure of the OWTs at the 1st iteration. The green dotted line at d = 138 marks the threshold below which types were evaluated for dissolution. (B) Normalised membership sum calculated after combining each pair of similar spectra. The vertical red dotted line shows the iteration (bottom axis) with the optimal number of candidate water types (top axis) before the performance dropped. (C – E) The three pairs of spectra that were not combined despite connected by a single branch in the phylogenetic tree. The dotted lines are the spectra in their original non-standardised form. The labels for the waveband 400 and 674 on the x-axis were omitted to aid visualisation.

Visual inspection of MSI-A/B images largely confirmed the separation between the three groups, except for water type 12.2, which was associated to both surface scum (intense light bright green) (Fig. 5A), and macrophytes (dark green colour with raft-like shapes indicative of wind dispersal) (Fig. 5B). Visual inspections showed macrophyte presence in 2019 and 2020 within type 12.2. Consequently, despite a low sum of band ratios at = 1.75, w = 1000 was assigned to water type 12.2 in consideration of the potential for cyanotoxin to accumulate within and below the scum at the surface. Other examples shown include water type 12.4 (having the highest sum of band ratios) where bloom presence was apparent through intense green/ bright green water (Fig. 5A-C) without clear evidence of accumulation at the surface. At-surface accumulation was visible in types 12.2 and 1.2 (Fig. 5A). Water type 8.3, assigned an intermediate rank (w = 100), was present in areas with patchy (or stripy) green material not floating at the surface (Fig. 5A-B).

The comparison of spectra by dominant water types distinguishes highly turbid waters, including presence of scum or near-surface accumulation, from conditions with presence of cyanobacteria and other phytoplankton mixed within the water column (Fig. 6). Most spectra in the former displayed the characteristic red edge followed by a NIR reflectance plateau, indicating high phytoplankton biomass at surface masking the absorption of light by water itself. These water types were all assigned w = 1000, except for OWT 5 which had a low sum of band ratios indicative of cyanobacterial optical features. Spectra indicative of presence of cyanobacteria and other phytoplankton mixed within the water column showed a range of sum of band ratios. Those with the highest sum of band ratio values showed the characteristic red edge of high phytoplankton (sub-surface) cells accumulation and were assigned w = 1000. Unlike the surface bloom (scum) types, these water types showed efficient light absorption by water in the NIR and SWIR. The rank assigned to this group of water types was set to w = 0 if the red edge disappeared or backscattering by particles (sediments) began to dominate the spectral shape in the red and NIR. A qualitative interpretation of the spectra of the 25 water types in shown in Table 1.

3.3. Cyanobacteria occurrence analysis

The ability of the extended water types library to capture the full optical diversity of the study area was confirmed by relatively low variance of average S_{owt} values over time (Fig. 7A). In particular, 32.4 % of observations had an average S_{owt} value lower than the 7-year average of 0.86. 2.8 % of observations had an average S_{owt} value lower than 0.8.

The COI in Winam Gulf showed considerable interannual variability, with an average of 0.66 (± 0.14), suggesting persistent cyanobacteria occurrence (Fig. 7B). April, June, and October had the highest average COI (0.68 ± 0.14). The highest average across the eight years examined occurred in 2020 (0.74 ± 0.16). COI dropped sharply in 2022 but increased again to average 0.62 (± 0.14) during 2023. Areas of lower COI were typically found in the Rusinga Channel, where water is deeper than the rest of Winam Gulf, ranging 15——25 m, and less turbid (Alexander & Imberger, 2013; Gikuma-Njuru et al., 2013). Areas of higher COI included semi-enclosed bays, such as Homa Bay in the south,

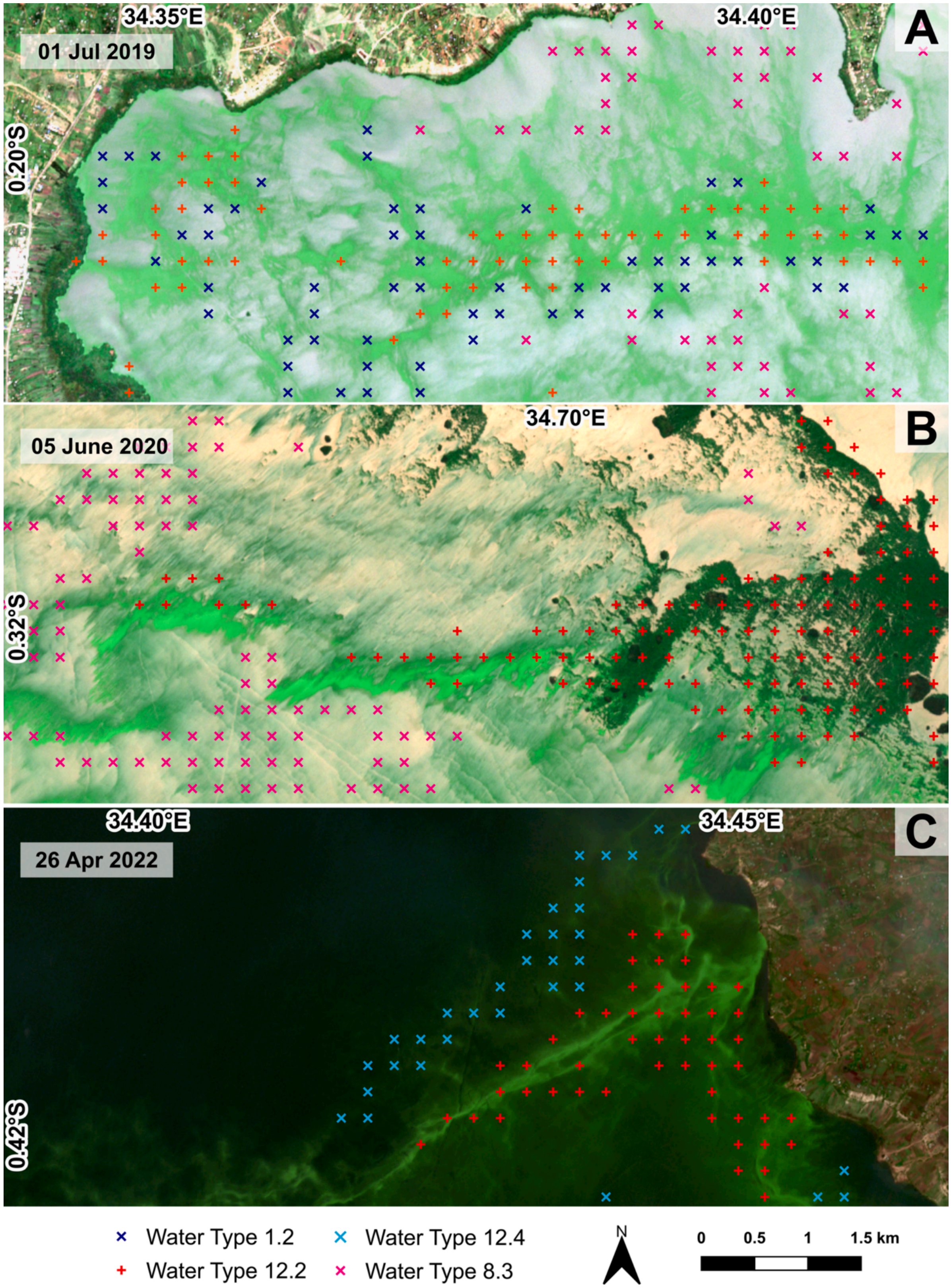

Fig. 5. (A-C) Sentinel-2 Multispectral Instrument (MSI)-A images corresponding to Ocean Land Colour Instrument (OLCI)-A/B overpasses. The coloured symbols in each panel correspond to the centroid of (300 m) OLCI pixels, colour-coded to the corresponding dominant water types in the legend.

Asembo Bay in the north, and Kisumu Bay in the northeast, as well as areas with large riverine inputs like Sondu, Kibuon, and Nyando to the southeast (Fig. 7C). Although COI peaked in 2020 and subsequently decreased, certain areas, particularly semi-enclosed bays, showed consistently higher cyanobacteria occurrence.

COI was compared to the Trophic State Index (TSI) as defined by Carlson (1977) to determine its correspondence to trophic levels. TSI was calculated as:

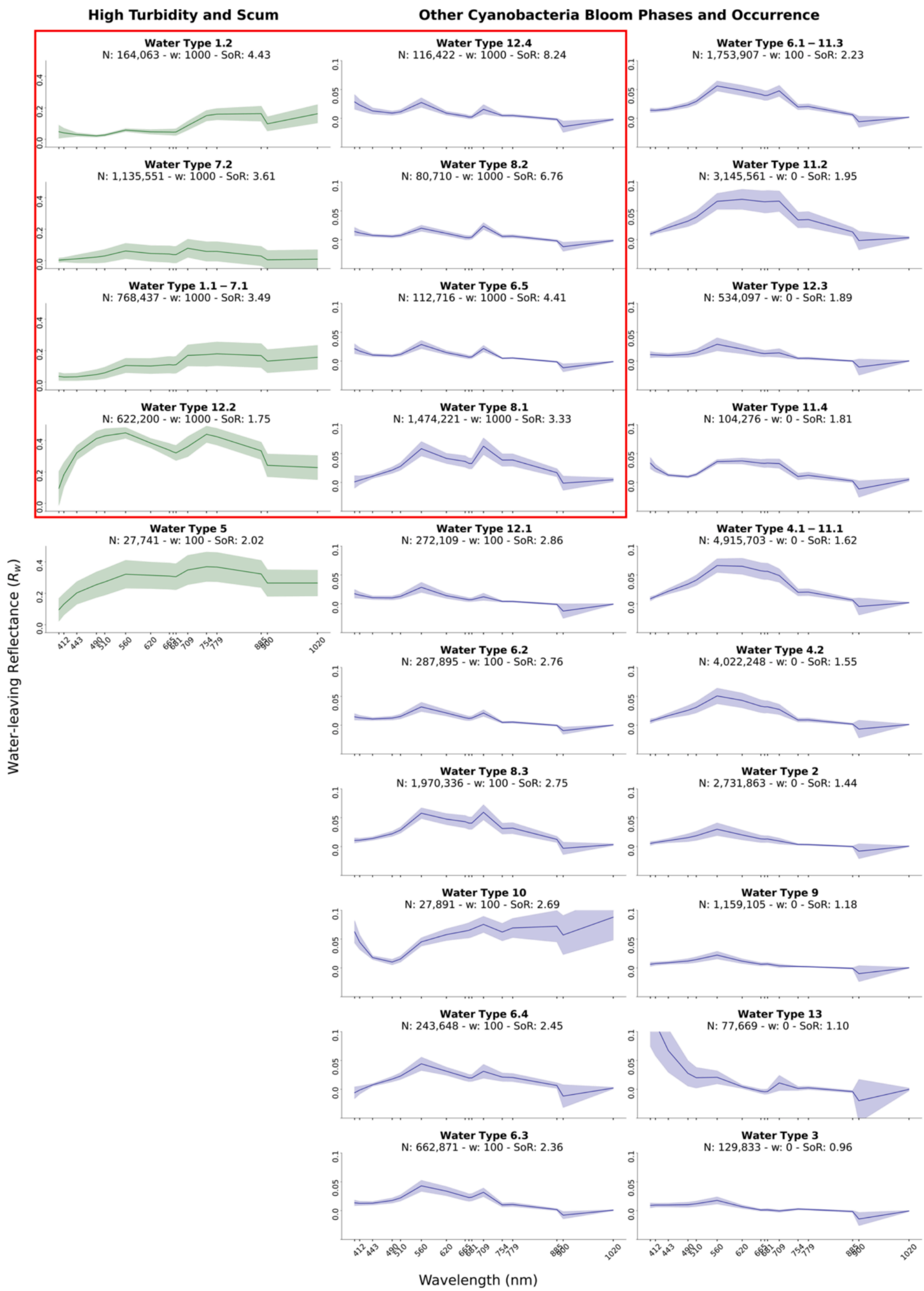

Fig. 6. Water-leaving reflectance (R_w) spectra grouped by dominant water type, ordered by rank (w) and the sum of 709/620 and 709/681 wavebands ratios (SoR), from top to bottom, left to right. N refers to the number of spectra associated with each dominant water type. Spectra coloured in green in the left panels are associated with high turbidity and scum and are on a different scale than the blue spectra in the right panels. The red box grouping the first panels groups all water types associated with high risk of cyanobacteria occurrence ($w = 1000$). Labels for wavebands centred on 400 and 674 nm omitted to aid visualisation.

Table 1
Original Optical Water Types (OWTs) definitions by Spyrakos et al. (2018) alongside a qualitative interpretation of the candidate library of 25 OWTs that originated from them. The interpretation focuses on the absorption features of phycocyanin (PC) at 620 nm and Chlorophyll-a (Chla) at 665 nm, Chla fluorescence at 681 nm, and the likelihood of surface-accumulated biomass from the relative magnitude of absorption by water in the NIR between 709 and 779 nm. This is distinguished between as 'at-surface' and 'near-surface' to ensure separation between OWTs associated to surface accumulation and mixed conditions. The OWTs left unmodified from their original definitions during the clustering process are reported as 'OWTn' under the 'Extended OWTs' column. The column w (weights) shows the arbitrary weights assigned to each library entry.

Original OWTs	Dominant characteristics	Extended OWTs	Qualitative Interpretation	w
OWT1	Hypereutrophic waters with scum of cyanobacterial bloom and vegetation-like R_{rs}	1.1_7.1*	Moderate PC and Chla absorption, with high probability of at-surface accumulation	1000
		1.2	Similar to 1.1_7.1, but with a less pronounced red-edge	1000
OWT2	Waters with diverse reflectance shape and marginal dominance of pigments and CDOM over inorganic suspended particles	OWT2	Very low PC and Chla absorption, low Chla fluorescence, and low probability of near-surface accumulation	0
OWT3	Clear waters	OWT3	Very low PC and Chla absorption, and low probability of near-surface accumulation	0
OWT4	Turbid waters with high organic content	4.1_11.1*	Very Low PC and Chla absorption in highly turbid conditions and low Chla fluorescence peak, with low probability of near-surface accumulation	0
		4.2	Similar to 4.1_11.1, but with lower turbidity	0
OWT5	Sediment-laden waters	OWT5	Low PC and moderate Chla absorption, with high probability of at-surface accumulation	100
OWT6	Balanced effects of optically active constituents at shorter wavelength	6.1_11.3*	Low PC and moderate Chla absorption, with high probability of near-surface accumulation	100
		6.2	Moderate PC and Chla absorption, with high probability of near-surface accumulation	100
		6.3	Low PC and moderate Chla absorption, with high probability of near-surface accumulation	100
		6.4	Moderate PC and Chla absorption, with medium probability of at-surface accumulation	100
		6.5	High PC and Chla absorption, with high probability of near-surface accumulation	1000
OWT7	Highly productive waters with high cyanobacteria abundance and elevated reflectance at red/near-infrared spectral region	7.2	High PC and moderate Chla absorption, with high probability of at-surface accumulation	1000
OWT8	Productive waters with cyanobacteria presence and with R_{rs} peak close to 700 nm	8.1	High PC and Chla absorption, with high probability of near-surface accumulation	1000
		8.2	High PC and moderate Chla absorption, with high probability of near-surface accumulation	1000
		8.3	Moderate PC and Chla absorption, with high probability of near-surface accumulation	100
OWT9	Optically neighbouring to OWT2 waters but with higher R_{rs} at shorter wavelength	OWT9	Low PC and Chla absorption, medium probability of Chla fluorescence, with low probability of near-surface accumulation	0
OWT10	CDOM-rich waters	OWT10	Low PC and moderate Chla absorption, with moderate probability of at-surface accumulation	100
OWT11	Waters high in CDOM with cyanobacteria presence and high absorption efficiency by Non-algal particles	11.2	Low PC and Chla absorption in highly turbid conditions, with low probability of near-surface accumulation	0
		11.4	Low PC and Chla absorption, moderate Chla fluorescence, in turbid conditions, with low probability of near-surface accumulation	0
OWT12	Turbid, moderately productive waters with cyanobacteria presence	12.1	High PC and moderate Chla absorption, with medium probability of near-surface accumulation	100
		12.2	Moderate PC and Chla absorption, with high probability of at-surface accumulation, likely associated to surface scum	1000
		12.3	Low PC and Chla absorption, with low	0

(continued on next page)

Table 1 (*continued*)

Original OWTs	Dominant characteristics	Extended OWTs	Qualitative Interpretation	w
			probability of near-surface accumulation	
		12.4	High PC and Chla absorption, with high probability of near-surface accumulation	1000
OWT13	Very clear blue waters	OWT13	Low PC and Chla absorption, with medium probability of near-surface accumulation	0

*New OWTs obtained from the combination of two clusters originating from different dominant OWTs. For example, OWT 1.1_7.1 arise from the combination of OWT 1.1 (i.e., cluster 1 originated from OWT1) and OWT 7.1 (i.e., cluster 1 originated from OWT7).

$$TSI(Chl) = 10\left(6 - \frac{2.04 - 0.68\ln Chl}{\ln 2}\right) \quad (5)$$

where Chla was derived applying the Mixture Density Network algorithm (MDN) of Pahlevan et al., 2020 to the OLCI R_w spectra. The trophic levels were obtained by binning the resultant numbers between 0 and 100 associated to each observation into four categories (Carlson, 1977). The distribution of COI by TSI (Fig. 8) revealed broad correspondence. Observations classed as mesotrophic and eutrophic appeared to follow a joint distribution, whereas hyper-eutrophic conditions formed a largely isolated class in the highest COI range.

An alternative interpretation of COI, particularly intended for management purposes, divides the COI along the apparent discontinuities in the COI distribution by TSI, with thresholds for low risk at COI < 0.5 and the transition from medium to high risk at COI > 0.8 (Figs. 8 and 9C). Using this thresholding, it was found that, on average, low risk waters accounted for 9.5 % of the area annually, with the lowest percentage in October (5.25 %) and the highest in December (13.78 %) (Fig. 9A-B). Medium and high risk made up 69.8 % and 20.7 % of the area, respectively. High risk conditions were most prevalent in April (27.3 %), followed by June (24 %) and October (23.5 %). The highest count of high risk days was in 2020 (209 days) with an average daily area of 687 km^2 classified as high risk. This was followed by 2019 (184 days

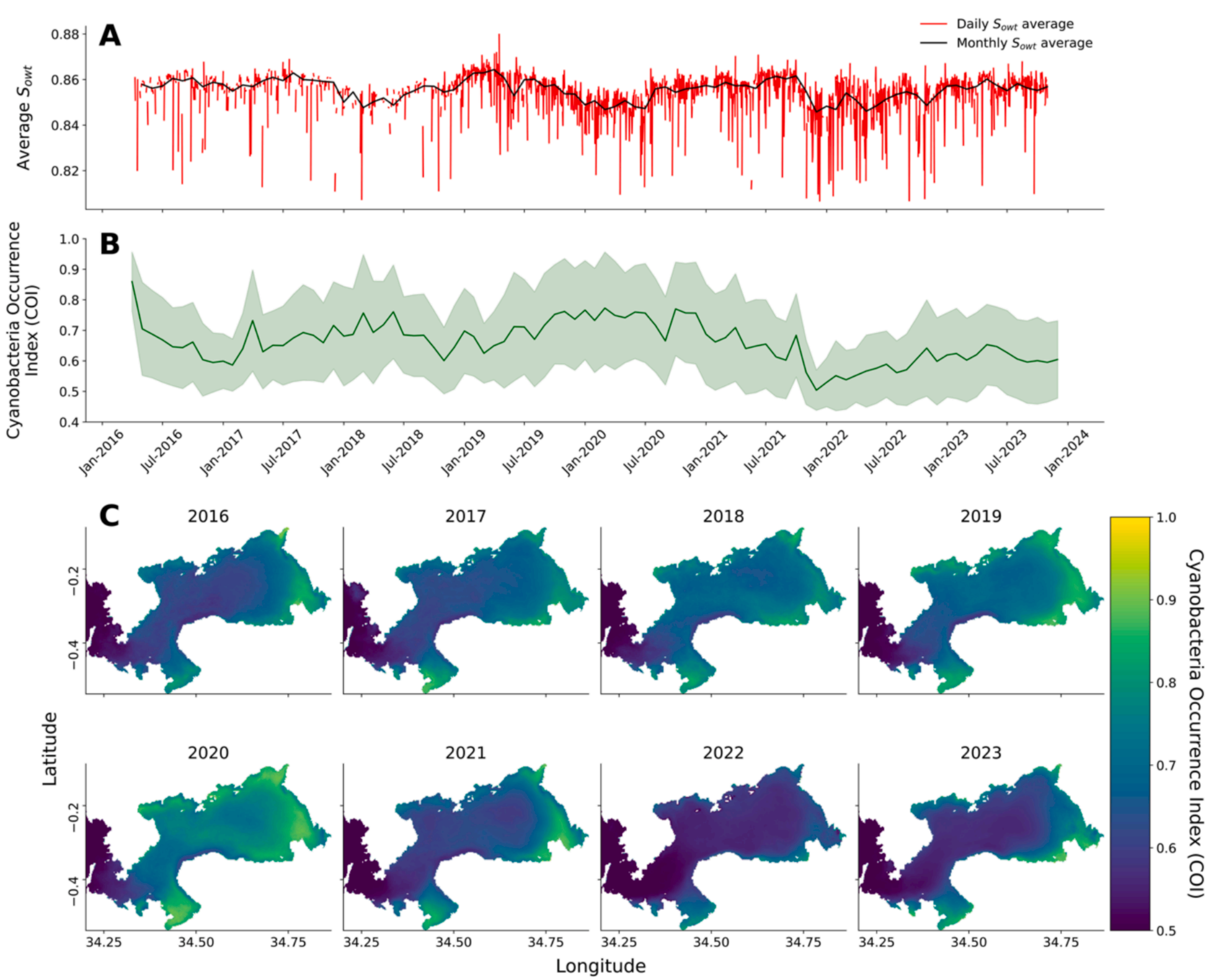

Fig. 7. (A) Daily (red line) and monthly (black line) average S_{owt} values. (B) Monthly average cyanobacteria Occurrence Index (COI) (solid green line) and relative monthly variances (shaded green areas). (C) Spatial distribution of average COI. The COI colour scale is stretched between 0.5 and 1 to enhance spatial differences.

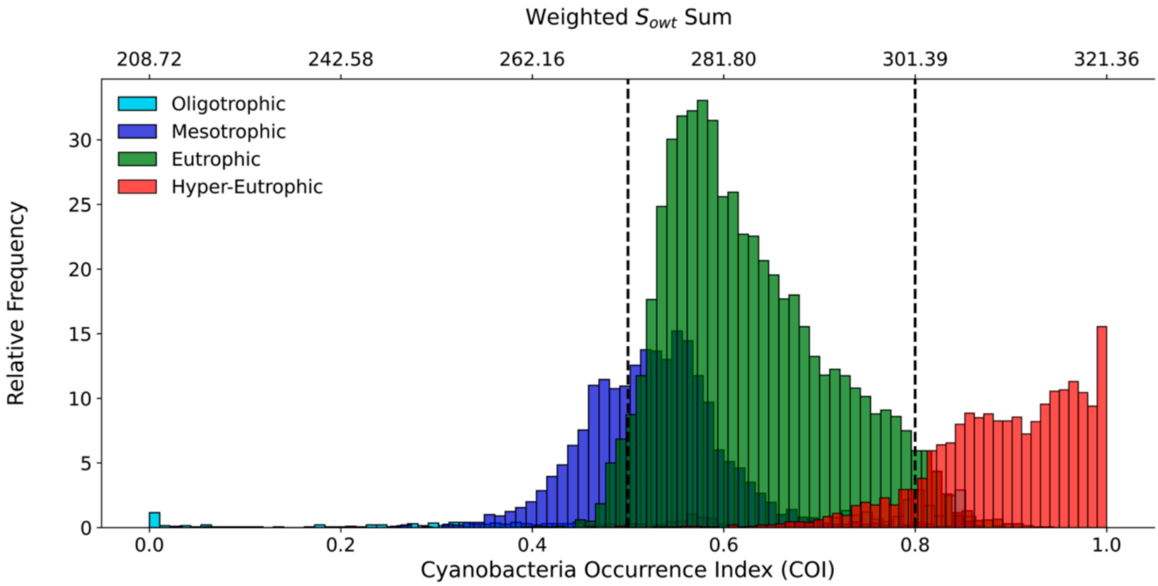

Fig. 8. Distribution of the COI by Trophic State Index (TSI), which identifies oligotrophic, mesotrophic, eutrophic, and hyper-eutrophic conditions. The horizontal axis at the top shows the weighted S_{owt} sums corresponding to COI in the bottom horizontal axis The vertical dotted lines mark the two arbitrary thresholds (COI = 0.5, and COI = 0.8) used to differentiate between low, medium, and high risk waters.

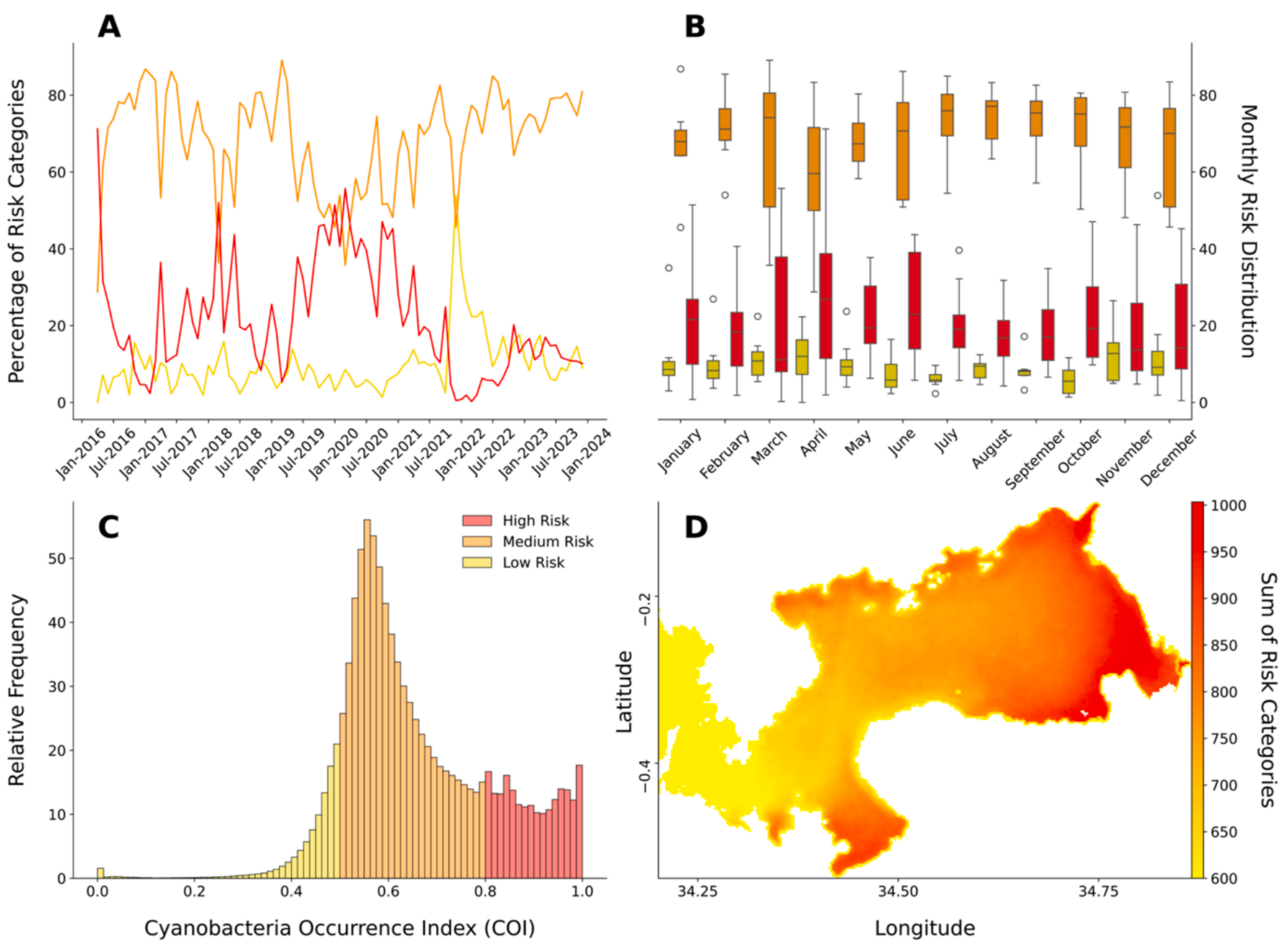

Fig. 9. (A) Monthly average percentage of pixels for each of the three risk categories (low = yellow, medium = orange, high = red). (B) Monthly distribution of the risk categories. (C) Distribution of Cyanobacteria Occurrence Index (COI) coloured by risk category. (D) Map of the sum of risk categories from 2016 to 2023. The numbers on the colour bar are negligible since they do not correspond to true risk values, but they help visualise areas that tend to experience higher (lower) risk over time. All plots follow the same colouring scheme in the legend in panel C. The three risk categories were obtained by setting COI < 0.5 as low risk, COI >= 0.8 as high risk, and intermediate values as medium risk.

averaging 530 km^2) and 2023 (180 days averaging 258 km^2). The month of January recorded the highest occurrence of high risk days with an average of 13.4, followed by May (13.1 days) and July, September, and October (all 11.8 days). December exhibited the highest average area covered with 573 km^2, followed by November and April, both with 523 km^2. Across the entire dataset, the daily high risk cover ranged from 52 km^2 to 1465 km^2 (a slightly larger area than Winam Gulf due to the inclusion of a small portion of Lake Victoria in our dataset). Spatially, the highest risk was recorded at the estuaries of Sondu and Nyando to the east, followed by Kisumu and Homa Bays (Fig. 9D). Waters in the Rusinga Channel showed the lowest risk across the study site.

3.4. Comparison with pigment concentration retrieval algorithms

To evaluate the relative response of COI versus Chla and PC estimates, it was compared to the established algorithms Maximum Peak-Height (MPH, Matthews et al., 2012), MDN and Simis07 (Simis et al., 2007) (Fig. 10). COI showed a strong Pearson's r correlation coefficient with MPH Chla ($r = 0.9$) and Simis07 PC estimates ($r = 0.84$), showing varying degrees of linearity and sensitivity along the pigment concentration ranges. The correlation between MPH and Simis07 was also high ($r = 0.8$, not shown). In contrast, the correlation between COI and MDN Chla estimates was lower ($r = 0.5$), similar to the correlation of MDN with MPH ($r = 0.44$) and Simis07 PC ($r = 0.37$).

Spatial comparison between COI and the three pigment concentration retrieval algorithms revealed broad alignment, while also highlighting some key differences. Generally, higher COI values corresponded with areas where at least one algorithm detected elevated Chla or PC concentrations (Fig. 11). However, the correspondence varied across different dates and locations. For instance, on the 23rd of March 2018, MPH detected high Chla concentrations on the east side of Winam Gulf, with concentrations decreasing towards the middle of the Gulf. In contrast, Simis07 and MDN showed an opposing trend, indicating low concentrations of PC and Chla in the same area where MPH detected high Chla. COI values in this region appeared to mediate between these conflicting algorithm outputs, acknowledging the presence of this optically distinct area but assigning it lower values compared to surrounding waters, aligning more closely with Simis07 results. On the 22nd of May 2023, MDN was the only algorithm that detected moderately high Chla concentrations in the central part of Winam Gulf. MPH and Simis07 primarily detected high concentrations of Chla and PC along the eastern and southern coastlines. COI identified the same subtle blooms as MDN in the central area, while also providing a more comprehensive representation of cyanobacteria presence and abundance compared to any single pigment concentration retrieval algorithm.

The distribution of COI in relation to different trophic states confirms a marked separation between eutrophic and hyper-eutrophic conditions, with larger overlaps between mesotrophic and eutrophic. The three risk categories (obtained using the same COI thresholding as in Fig. 8), although less granular than COI maps, show a marked distinction between the different trophic levels. Low risk areas were primarily concentrated within and beyond the Rusinga Channel to the west, becoming gradually more eutrophic and with higher COI towards the east of Winam Gulf. Semi-enclosed areas were again identified as the areas with higher occurrence risk.

3.5. Assessment in other waterbodies

When applied to other waterbodies (maintaining the algorithm configuration for Winam Gulf), the broad alignment of COI with the three pigment retrieval algorithms under investigation held, despite optically different dynamics of the lakes in question (Fig. 12). For instance, Lake Erie on the 25th of July 2019 showed a higher risk of cyanobacteria occurrence in the west, where blooms are typically observed, and lower COI to the east, lower than most observations in Winam Gulf. The distribution of TSI showed a higher count of oligotrophic and mesotrophic waters, which aligned with the lowest COI values. The MPH, MDN and Simis07 algorithms generally produced low values of Chla and PC concentration in the east, with the latter possibly affected by atmospheric correction artifacts. COI provided a much more detailed map, revealing cyanobacteria eddies not produced by any of the other algorithms. Lake Taihu on the 22nd of May 2019 exhibited hyper-eutrophic and eutrophic characteristics, aligning with high COI values. The three pigment retrieval algorithms had some disagreement in high-biomass areas on the west side of the lake, with the MDN visibly saturating in areas of surface accumulations. The COI mostly showed values in the range of 0.6 – 1, revealing a detailed cyanobacteria occurrence map, especially in areas of high biomass accumulation at the surface. The picture is more complex for the small Hartbeespoort Reservoir, where TSI has an unexpected distribution, possibly to limitations of MDN and adjacency effect.

4. Discussion

Cyanobacteria present a wide diversity of optical behaviours related to their vertical distribution in the photic zone, adaptive pigmentation, and colony formation. Previous attempts to estimate cyanobacteria biomass from pigment absorption have proven challenging to interpret in water management practices. Our results show that these dynamics can be captured with an extended set of Optical Water Types (OWTs) based on Sentinel-3 OLCI-derived spectra, from which further algorithm and risk model development may start. The OWT classification complexity can be further reduced into a single index of cyanobacteria

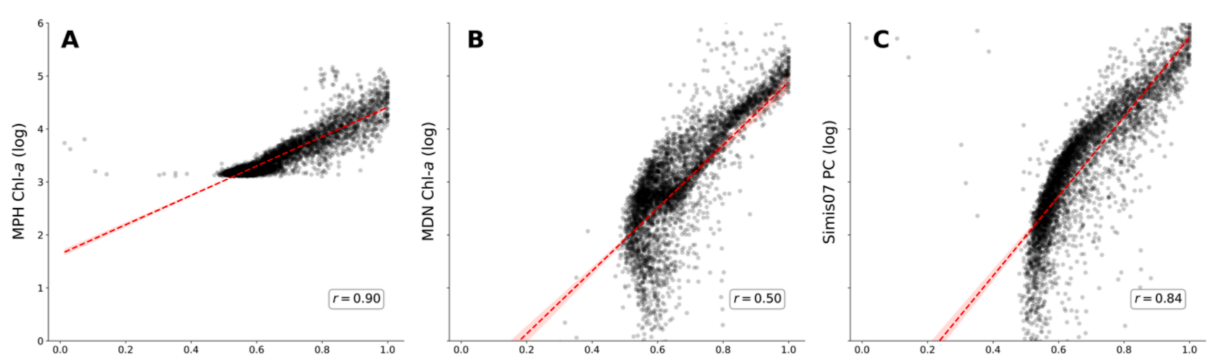

Fig. 10. Comparison between the Cyanobacteria Occurrence Index (COI) versus (A) the Maximum Peak-Height (MPH) and (B) Mixture Density Network (MDN) chlorophyll-a (Chl-a) estimates, and (C) Simis07 algorithm phycocyanin (PC) estimates. MPH, MDN and Simis07 were log-transformed to enable the comparison. The plots were generated using a sample of 100,000 datapoints selected at random to facilitate visualisation. Pearson's r showed at the bottom of each plot was obtained from the full dataset of 26 million datapoints.

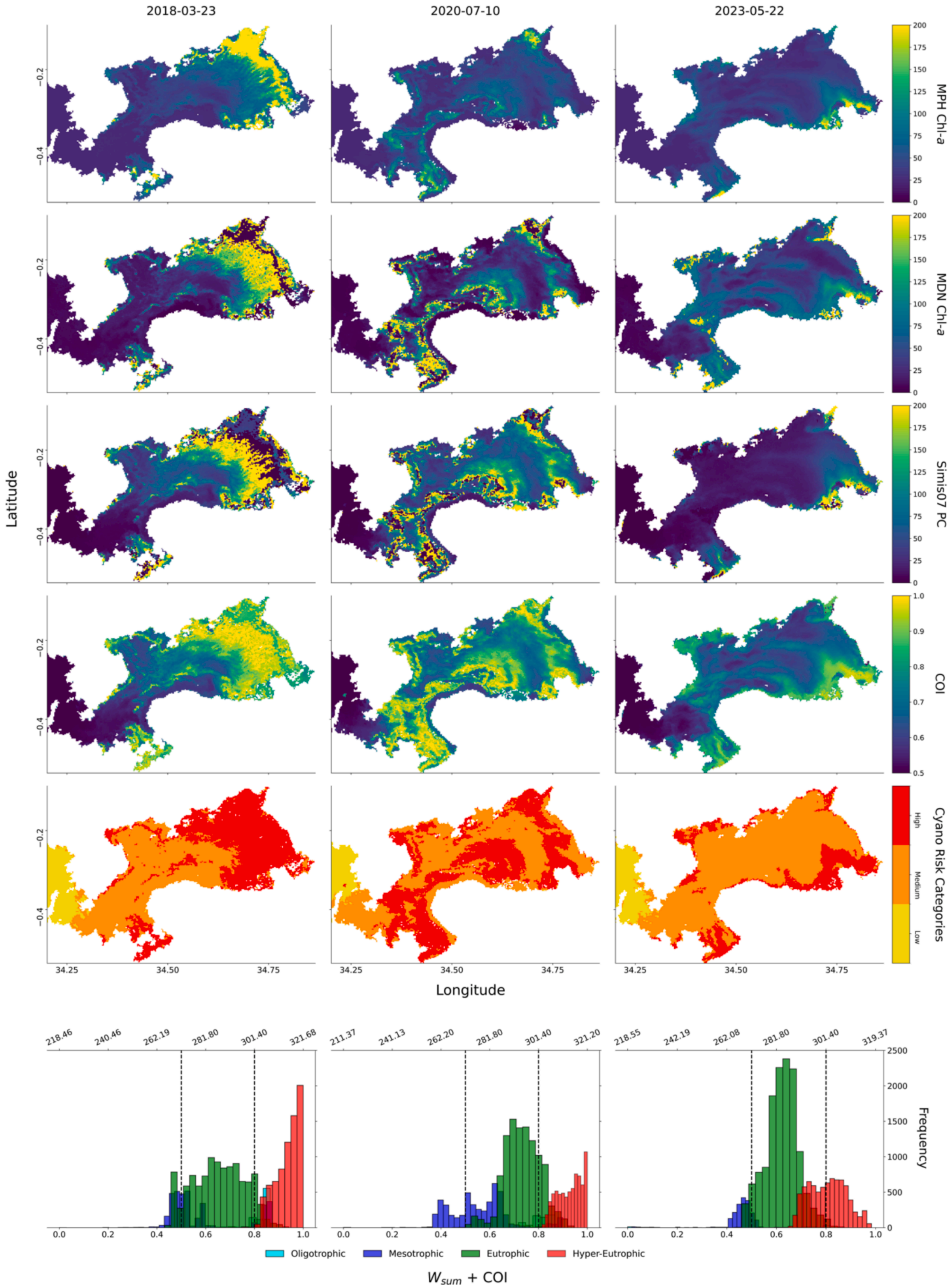

Fig. 11. Maps of Maximum Peak-Height (MPH) Chlorophyll-*a* (Chl-*a*) estimates, Mixture Density Network (MDN) Chl-*a* estimates, Simis07 phycocyanin (PC) algorithm estimates, Cyanobacteria Occurrence Index (COI), the three risk categories, and the distribution of COI and Weighted S_{owt} sum by trophic status, from top to bottom. Figures in each column refer to the same timestamp at the top of the columns. The cyano risk categories were obtained by considering COI $<$ 0.5 as low risk, COI $>$ 0.8 as high risk, and medium risk COI values in between. The thresholds are shown as vertical dotted lines in the histograms at the bottom of the figure. COI and pigment estimate values were constrained in the range 0.5–1 and 0–200 (μgL^{-1}), respectively, to enhance spatial visualisation.

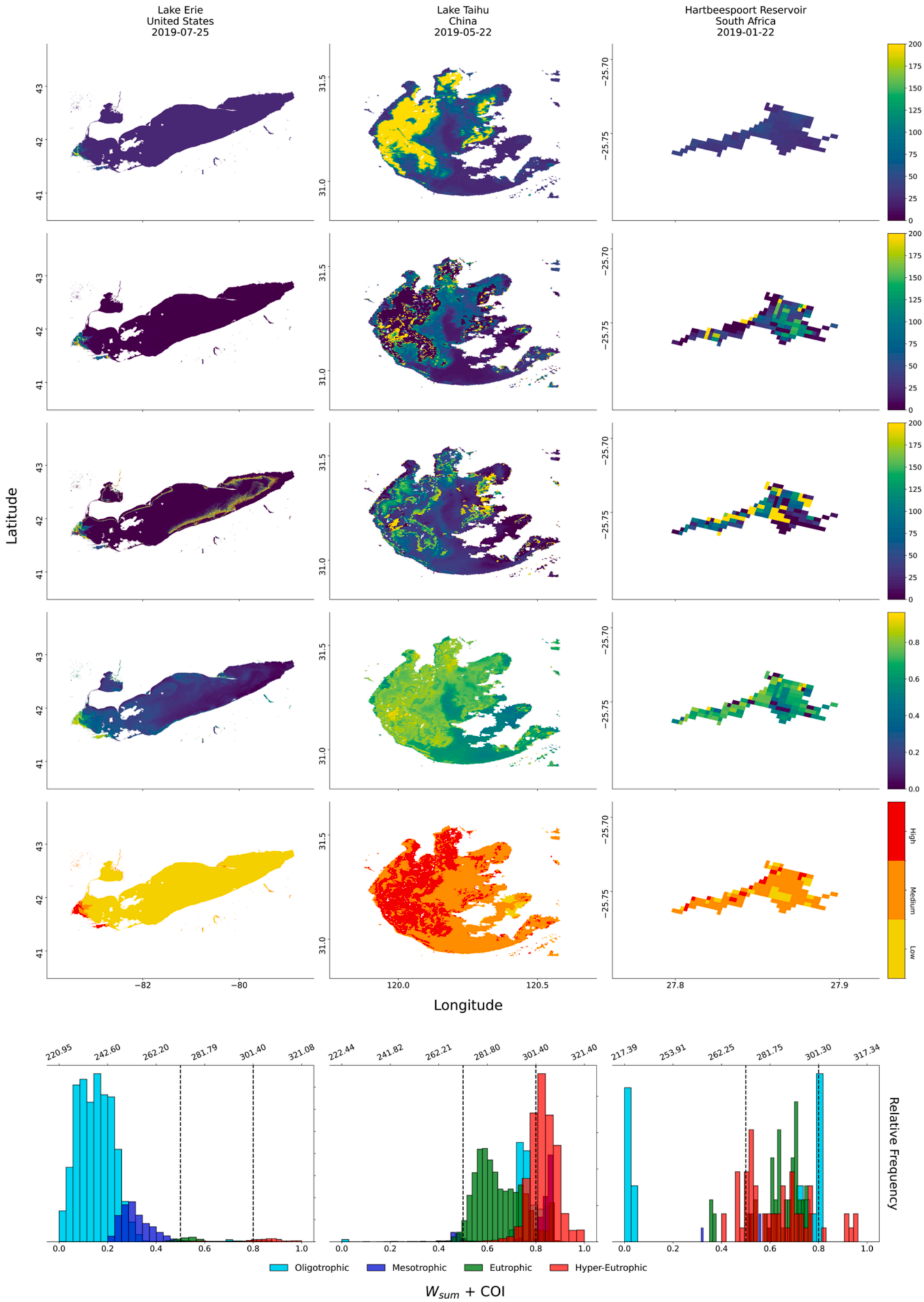

Fig. 12. Maps of Maximum Peak-Height (MPH) Chlorophyll-*a* (Chl-*a*) estimates, Mixture Density Network (MDN) Chl-*a* estimates, Simis07 phycocyanin (PC) algorithm estimates, Cyanobacteria Occurrence Index (COI), the three risk categories, and the distribution of COI and Weighted S_{owt} sum by trophic status, from top to bottom. Figures in each column refer to the lake and timestamp at the top of the columns. The cyano risk categories were obtained by considering COI < 0.5 as low risk, COI > 0.8 as high risk, and medium risk COI values in between. The thresholds are shown as vertical dotted lines in the histograms at the bottom of the figure. Pigment concentration estimate values were constrained in the range 0–200 μgL^{-1} to enhance spatial visualisation.

occurrence, the COI, which we propose could serve as an indicator for satellite-derived inference of the risk of cyanobacteria occurrence. In this context, the risk of cyanobacteria occurrence relates to the presence of optical features typically associated with abundance of cyanobacteria. This occurrence risk should not be confused with health risks to humans and animals, associated with cyanotoxin production and accumulation or rapid biogeochemical change during hypoxia events. Instead, the COI could be used in management to guide sampling strategies to determine the species composition and health risk posed by observed blooms.

Whilst this approach requires wider evaluation across water bodies, it provides proof-of-concept that subtle features in water colour stemming from variation in phytoplankton composition, mixing conditions and cell physiology are likely to be recurrent and may be captured from current satellite sensors. This builds on previous work where OWTs defined from in situ observations alone is likely to have missed less commonly observed conditions, including at-surface blooms that are challenging to record in situ without disturbing their appearance. The development of this approach in the tropical setting of Winam Gulf, experiencing frequent cyanobacterial blooms and variable mixing conditions, provides a suitable testbed, despite a general lack of equivalent in situ observations. Bloom phenology is largely driven by varying nutrient conditions, grazing pressure, and species composition, highlighting the importance of validating the approach across systems and seasons with different controlling mechanisms. The lack of comprehensive biogeochemical measurements, including pigment concentrations, nutrient levels, and cell counts, limits our ability to definitively attribute optical features to specific physiological states, in contrast to deriving optical classifications from detailed in situ investigations. Here, we show that a hybrid approach is feasible.

It is important to recognise that defining OWTs from atmospherically corrected observations results in a water types library that includes any biases from atmospheric correction. This, and the specific sensor waveband configuration, may result in challenges to apply the library to other sensors. Similar efforts to generate OWTs from satellite-derived ocean colour spectra already exist for the global ocean (Jackson et al., 2017) and transitional and coastal water bodies (Atwood et al., 2024). The optical diversity and atmospheric correction challenges of inland waters are nevertheless expected to exceed those. Thus, the present extension of OWTs derived from OLCI should be considered specific to the current Polymer atmospheric correction and may include sensor-specific biases. Nevertheless, the gradual spatiotemporal dynamics observed in our results, along with correlation to established algorithms and trophic state indicators, suggests the extended water types library captures meaningful variations in bloom conditions.

The primary proof-of-concept in this work is an improved OWT membership sum resulting from extension from 13 in situ reflectance derived OWTs to a total of 25 types. K-means clustering revealed significant intra-OWT variability (Fig. 3) indicating that Winam Gulf presented optical features outside the scope of the 13 OWTs for inland waters formulated by Spyrakos et al. (2018). As discussed above, it is not known whether this additional variability stems from conditions of the water, or from the remote sensing approach. Nevertheless, the process of expanding the initial types indicative of cyanobacteria into a wider diversity of subtypes proved effective. While it is expected for a greater number of water types to 'better represent' water colour variability, the combination of subtypes based on the Euclidian distance between OWT membership score (S_{owt}) clusters avoided 'overfitting' the candidate set to our study site. This process allowed for a more generalised candidate library that improved the membership sum compared to the original set of 13 OWTs (Fig. 5B). Furthermore, since the membership sum did not show significant drops over time (Fig., 7A), it can be concluded that all optical variations in the data set were captured, and that cyanobacteria bloom phases in Winam Gulf appeared to be described in our extended candidate water types library.

The clustering and classification of spectra into dominant members using the candidate water types set revealed cases that have not been previously described or not fully characterised. For example, the (non-cyanobacteria) OWT 5 showed the largest variance of Chla and inorganic suspended matter in the OWT library for inland waters by Spyrakos et al. (2018), suggesting that its original formulation was potentially incomplete. This is evident by the much lower observation count within our candidate library (N < 30,000) (Fig. 6) compared to its observation count prior the definition of the new types (N > 1.3 millions) (Fig. 3C). Additionally, this indicates that the new candidate library absorbed some of the optical characteristics originally associated to OWT 5. Another case not described in the original set of 13 OWTs for inland waters by Spyrakos et al. (2018) is what we identified as water type 12.2. Visual interpretation of MSI images showed that spectra with this water type as their dominant member were associated to intense surface scum and sometimes macrophyte mats (Fig. 5). While the detection of macrophytes requires a separate assessment beyond the scope of this work, it is important to note that strategies to distinguish macrophytes from cyanobacteria blooms and surface scum exist. For instance, Matthews & Odermatt (2015) proposed a method that combines the Normalised Difference Vegetation Index (NDVI) and the MPH algorithm in a flagging system. This approach uses an MPH threshold to identify legitimate floating vegetation, while an NDVI threshold helps differentiate water from submerged and floating aquatic vegetation. Their combination allows for the separation of floating cyanobacteria scum from emergent macrophyte vegetation. Despite a high likelihood of cyanobacteria presence for water type 12.2, the band ratios used to determine COI weights were relatively low, which can be explained by higher R_w values at 620 nm derived by the different optical characteristics of surface scum, especially when bleached (Tebbs et al., 2015). While macrophytes have been documented in Winam Gulf (Romero et al., 2005; Kiage & Obuoyo, 2011; Otieno et al., 2022), visual inspection of 331 MSI observations coinciding with OLCI revealed that their presence was primarily confined to the Nyando estuary to the east, with more frequent appearances throughout the Gulf only observed in 2019 and 2020. The clustering process did not clearly separate surface accumulations of cyanobacteria from floating macrophytes within water type 12.2, as evidenced by its large observation count (N > 600,000). This suggests that at the spatial and spectral resolution of OLCI, these features may appear similar when using standardised spectra, and may not be consistently distinguishable. However, while this aspect remains untested, we would expect macrophytes to influence the COI, and separate algorithms may be needed in the future to mask them out. Finally, there is evidence to suggest that the remaining, non-cyanobacteria, OWT set also deserves further scrutiny. Particularly, spectra associated with OWT 13 showed a backscattering peak at 709 nm (Fig. 6), indicating phytoplankton presence and a deviation from the clear water case described by Spyrakos et al. (2018).

The candidate water types library likely offers a more detailed representation of cyanobacteria bloom phases than the original 13 OWTs for inland waters, capturing bloom formation, accumulation, and breakdown (Fig. 6). One of the two main spectral groups was associated primarily with surface bloom conditions (e.g., 1.2, 7.2, 1.1_7.1), exhibiting a reflectance peak around 709 nm and plateau in the NIR region. These spectra occur when cells accumulate at the water surface. Light is efficiently reflected in a shallow layer of water near the surface, thus reducing the path length of light absorption by water which in mixed conditions would greatly reduce the reflectance in the NIR. Increased R_w in the NIR may also be caused by a higher degree of atmospheric path radiance from adjacent non-water surfaces (Jiang et al., 2023). If exposed to the sun for long enough, cells accumulated at the surface may start bleaching (i.e., 12.2) (Fig. 5A-B). The second group of spectra was associated to various cyanobacteria bloom phases within the water column, as evidenced by a backscattering peak in the NIR, and an efficient absorption by water beyond the NIR. Water types 12.4, 8.2, 6.5 and 8.1 may indicate fully developed blooms. As blooms intensify and pigment absorption per unit area increases, absorption features in the red spectral region intensify (Simis et al., 2007; Matthews & Odermatt, 2015; Stumpf et al., 2016). A lower peak at 709 may indicate early to very early cyanobacteria blooms phases (e.g., 12.1, 6.2, 8.3, 6.4, 6.3), when cells have not yet aggregated and become the dominant signal. During what may indicate the decaying cyanobacteria bloom phase, spectral characteristics change significantly. As cells break down, decaying material adds to browning of the water (Tebbs et al., 2013) and a flattening of the remote sensing signal, such as represented by water type 12.3. Finally, the remainder of water types present mixed conditions, where cyanobacteria and other phytoplankton are present in varying concentrations throughout the water column (Table 1). The spectral signature becomes more complex, due to the concurrent contribution from different pigments and varying backscattering due to suspended solids (e.g., 11.2, 4.1_11.1, 4.2).

The interpretation of OWT membership in terms of optical biogeochemical dynamics will likely remain challenging to users without bio-optical knowledge. Given this consideration, it is desirable to generate an intuitive index such as the COI. The COI, generated by assigning a weight to each OWT to indicate cyanobacteria occurrence risk or likelihood, is a subjective ranking. Objective ranking is not feasible due to the lack of pre-requisite in situ data. Typical optical features are instead used to assign order-of-magnitude ranked classes. Notably, the sum of the 709/620 and 709/681 nm waveband ratios serves as a useful indicator of cyanobacteria presence and abundance, capturing both pigment absorption features and their expression under different bloom conditions. When cyanobacteria concentrate near the surface, the optical path length through the water column is altered, changing how light interacts with water and other particles at these wavelengths (Gordon & Morel, 1983). Consequently, the single metric W_{sum}, and by extension COI, recognises that multiple biogeochemical processes can take place simultaneously through the combined use of all membership scores and are accounted for by the ranking scheme based on typical optical features. It is further useful to note that the weighting of OWT memberships underpinning the COI, being based on these optical features associated with cyanobacteria presence, is likely behind the positive and strong correlation with pigment retrieval algorithms demonstrated here. A major difference between COI and these algorithms is that the COI does not include any attempt to calibrate the response to pigment concentration, noting difficulties to achieve linearity in past efforts, and significant biological variability in pigment production within cyanobacteria populations related to nutrient availability, and between phytoplankton groups.

The strong correlation of COI with diagnostic pigment-specific algorithms like MPH and Simis07 confirmed its sensitivity to cyanobacteria in dynamic water conditions (Fig. 10). MPH was designed to capture Chla concentrations linearly by leveraging the reflectance peak around 709 nm, which is sensitive to changes in Chla across different trophic levels (Matthews et al., 2012). However, MPH is less effective in distinguishing variations of Chla when the 709 peak diminishes, leading to large uncertainties at lower concentrations (Fig. 10A). The relationship with Simis07 is less straightforward due to the optical complexity of PC (Fig. 10C). PC shares its absorption region with other pigments like Chlorophyll-*b* and Chlorophyll-*c*, complicating its isolation (Simis et al., 2007). The Simis07 algorithm incorporates spectral deconvolution techniques to separate the absorption signals of these pigments. However, this process is less effective at lower concentrations where the relative contribution of PC absorption decreases (Tandeau De Marsac, 1977; Wyman & Fay, 1986b, 1986a; Grossman et al., 1993). The comparison with MDN showed the advantages of COI in detecting cyanobacteria presence and abundance (Fig. 10B). MDN leverages the full range of wavebands available to estimate Chla across diverse bio-optically variable waterbodies, making it well-suited to detect subtler blooms mixed within the water column that are yet to surface (Pahlevan et al., 2020). However, its reliance on the fluorescence peak at 681 nm makes it less suitable to track cyanobacteria. Additionally, MDN may struggle in especially turbid water conditions, which may divert from the original limited dataset that the model was trained on. Consequently, MDN may overestimate higher Chla concentrations and underestimate lower concentrations, especially in the presence of cyanobacteria. Spatially, the COI aligned with both high Chla estimated by MPH and high PC concentrations measured by Simis07, reliably identifying cyanobacteria activity both in Winam Gulf (Fig. 11), and in other waterbodies with different optical characteristics (Fig. 12). While MPH and Simis07 identify areas with high concentrations of cyanobacteria, they can miss cyanobacteria presence at lower concentrations in well-mixed conditions that COI appears to detect, as evidenced by its agreement with MDN. Additionally, the COI showed sensitivity to surface (bleached) scum that MDN and Simis07 do not resolve and relate to low concentrations of Chla and PC, respectively. This is particularly significant since the risk posed by cyanobacteria in surface scum conditions can increase substantially, highlighting the limitations of traditional pigment-retrieval algorithms in these scenarios. The technical differences of the pigment-retrieval algorithms — MPH focusing on peak bloom conditions, Simis07 on overall PC concentration using spectral deconvolution, and MDN on Chla levels leveraging fluorescence — lead to variations in their responses to different water conditions. The COI demonstrated consistent behaviour across varying optical conditions, suggesting potential for broader application. The preservation of these relationships across different waterbodies, from the shallow tropical Winam Gulf to temperate systems like Lake Erie and the highly eutrophic Lake Taihu, provides encouraging evidence for the transferability of the approach, though further validation is needed to establish robust thresholds for different ecological contexts.

The COI suggested major shifts in cyanobacteria occurrence in Winam Gulf. For example, the COI revealed higher than average cyanobacteria presence in 2020 (Figs. 7 and 9). This was the year with the highest water levels ever recorded in Lake Victoria and Rift Valley Lakes, primarily attributed to high precipitation anomalies (Herrnegger et al., 2021; Pietroiusti et al., 2024). Heavier rainfall is likely to have delivered sediment and nutrients into the lake, increasing turbidity and leading to further eutrophication, conditions in which cyanobacteria are known to outcompete other phytoplankton species (Paerl & Paul, 2012; Huisman et al., 2018). In successive years, COI gradually decreased, showing a low towards the end of 2021. This could be in part linked to a higher dilution driven by higher water levels, which may have suppressed the activity of certain bloom-forming species of phytoplankton and cyanobacteria as observed in other lakes in Kenya (Byrne et al., 2024). However, other environmental factors may have played a role in the suppressed cyanobacteria activity. Given the prevalence of surfacing blooms of cyanobacteria in the Gulf, higher than average COI in coastal bays provide further indication that wind-driven biomass is appropriately accounted for in the COI.

For management purposes, identifying risk levels from COI may be particularly useful, given the popular use of traffic light systems in water quality risk assessment, including those in existence for cyanobacteria (surface) blooms (EPA, 2015; Loisa et al., 2015). The translation of COI values into risk categories provides an accessible framework for water quality managers while maintaining the underlying detail of the index. COI climatology may further help to identify areas and periods that require more careful management (Fig. 8), allowing for proactive monitoring strategies. For instance, it was shown that the highest count of high risk pixels in Winam Gulf consistently occurs during the long-rainy season (Mar-Jun), which may be a direct effect of rainfall contributing to the transport of nutrients into the lake from surrounding land, promoting cyanobacterial growth. Such temporal patterns, when combined with local environmental knowledge, can help optimise monitoring resources and inform management decisions. High risk areas in enclosed bays and areas with significant riverine inputs (like Homa, Asembo, and Kisumu Bay) would then be prioritised. Recent genetic studies have found cylindrospermopsins and microcystins in a range of concentrations especially in these areas (Brown et al., 2024), providing biological validation to the risk assessment approach. While these specific patterns reflect the conditions in Winam Gulf, the ability of the framework to identify areas of concern based on optical characteristics suggests its potential utility for other water bodies. Of course, the thresholds explored in this work are based on visual interpretation of the distribution of TSI, inherently linked to Chla concentration (Fig. 8), but could be redrawn based on the management targets of individual water bodies, which is also common practise, e.g., in the European Water Framework Directive and monitoring for Sustainable Development Goal 6.3.2 (Halleux & Pichon, 2023).

Further validation and refinement of both the candidate water types library and the COI are essential to establish their broader applicability. While deriving OWTs directly from satellite observations helps capture dynamic bloom conditions, this approach faces specific validation challenges. Surface accumulations are particularly difficult to sample representatively due to disturbance from vessel movement and wind action, compounded by their rapid temporal evolution. This requires carefully designed sampling strategies that can capture both optical and biochemical characteristics while minimising disturbance. The validation strategy should address three key aspects. First, systematic testing of the candidate library across diverse geographical and climatic regions using the available satellite record will determine whether the bloom phases identified in Winam Gulf are consistently detected by different sensors. This cross-sensor validation is particularly important given that the current library is specific to the waveband configuration of OLCI. Second, the robustness of COI should be evaluated against a range of water conditions potentially underrepresented in our dataset, including validation against in situ Chla and PC measurements across different bloom stages. This validation effort is hampered by the current lack of standardised methods for PC extraction (Simis et al., 2007; Stumpf et al., 2016) and may alternatively rely on non-optical indicators of cyanobacteria biomass, such as cell counts, or those of less quantitative nature such as from fluorescence probes. Third, the sensitivity of both the water types library and COI to different atmospheric correction methods requires systematic assessment, given their significant influence on retrieved reflectance spectra in optically complex waters. While the underlying mechanism of COI should theoretically be robust to atmospheric correction biases, as these are propagated consistently through the wavebands of the used sensor, empirical validation through multi-algorithm comparisons would strengthen confidence in the approach.

Looking forward, this present study suggests new opportunities to study cyanobacteria dynamics and their management across spatial and temporal scales. Long-term satellite records processed with our approach could reveal valuable insights into climate change impacts on bloom patterns in lakes, including currently underrepresented tropical and subtropical systems. Analysis of these time series could help uncover relationships between bloom phenology and environmental gradients, supporting predictive modelling efforts. Such information would be crucial for adaptive water resource management, particularly in regions where traditional monitoring approaches face logistical or resource constraints. By systematically addressing these research priorities through coordinated validation efforts and stakeholder engagement, we can enhance the reliability of both the candidate water types library and COI, developing them into robust tools for characterising and monitoring cyanobacteria dynamics worldwide.

CRediT authorship contribution statement

Davide Lomeo: Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Stefan G.H. Simis:** . **Xiaohan Liu:** Writing – review & editing. **Nick Selmes:** Software, Resources, Data curation. **Mark A. Warren:** . **Anne D. Jungblut:** Writing – review & editing, Supervision. **Emma J. Tebbs:** .

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors acknowledge funding to DL from the Natural Environmental Research Council (NERC) through the London NERC DTP (NE/S007229/1) and the Natural Environment Research Council Earth Observation Data Analysis and Artificial-Intelligence Service (NEODAAS) for providing satellite data.

References

Alexander, R., Imberger, J., 2013. Phytoplankton patchiness in Winam Gulf, Lake Victoria: a study using principal component analysis of in situ fluorescent excitation spectra. *Freshwater Biology*. 58 (2), 275–291. https://doi.org/10.1111/fwb.12057.

Atwood, E.C., Jackson, T., Laurenson, A., Jönsson, B.F., Spyrakos, E., Jiang, D., Sent, G., Selmes, N., Simis, S., Danne, O., Tyler, A., Groom, S., 2024. Framework for Regional to Global Extension of Optical Water Types for Remote Sensing of Optically Complex Transitional Water Bodies. *Remote Sensing*. 16 (17), 3267. https://doi.org/10.3390/rs16173267.

Brown, K.M., Barker, K.B., Wagner, R.S., Ward, C.S., Sitoki, L., Njiru, J., Omondi, R., Achiya, J., Getabu, A., McKay, R.M., Bullerjahn, G.S., & the NSF-IRES Lake Victoria Research Consortium (2024). Bacterial community and cyanotoxin gene distribution of the Winam Gulf, Lake Victoria, Kenya. *Environmental Microbiology Reports*. 16 (3), e13297. doi:10.1111/1758-2229.13297.

Byrne, A., Tebbs, E.J., Njoroge, P., Nkwabi, A., Chadwick, M.A., Freeman, R., Harper, D., Norris, K., 2024. Productivity declines threaten East African soda lakes and the iconic Lesser Flamingo. *Current Biology*. 34 (8), 1786–1793.e4. https://doi.org/10.1016/j.cub.2024.03.006.

Carlson, R.E., 1977. A trophic state index for lakes. *Limnology and Oceanography*. 22 (2), 361–369. https://doi.org/10.4319/lo.1977.22.2.0361.

Carrea, L., Crétaux, J.-F., Liu, X., Wu, Y., Bergé-Nguyen, M., Calmettes, B., Duguay, C., Jiang, D., Merchant, C.J., Mueller, D., Selmes, N., Simis, S., Spyrakos, E., Stelzer, K., Warren, M., Yesou, H. & Zhang, D. (2024). *ESA Lakes Climate Change Initiative (Lakes_cci): Lake products, Version 2.1*.p.11057 Files, 824251210119 B. doi:10.5285/7FC9DF8070D34CACAB8092E45EF276F1.

Christensen, V.G., Khan, E., 2020. Freshwater neurotoxins and concerns for human, animal, and ecosystem health: A review of anatoxin-a and saxitoxin. *Science of the Total Environment*. 736, 139515. https://doi.org/10.1016/j.scitotenv.2020.139515.

Dall'Olmo, G. & Gitelson, A.A. (2005). Effect of bio-optical parameter variability on the remote estimation of chlorophyll-a concentration in turbid productive waters: experimental results. *Applied Optics*. 44 (3), 412. doi:10.1364/AO.44.000412.

Dekker, A.G., Vos, R.J. & Peters, S.W.M. (2001). *Comparison of remote sensing data, model results and in situ data for total suspended matter žTSM/ in the southern Frisian lakes*.

Dekker, A.G. (1993). Detection of optical water quality parameters for entrophic waters by high resolution remote sensing. Free University, Amsterdam.

Dosio, A., Lennard, C., Spinoni, J., 2022. Projections of indices of daily temperature and precipitation based on bias-adjusted CORDEX-Africa regional climate model simulations. *Climatic Change*. 170 (1–2), 13. https://doi.org/10.1007/s10584-022-03307-0.

Eleveld, M., Ruescas, A., Hommersom, A., Moore, T., Peters, S., Brockmann, C., 2017. An Optical Classification Tool for Global Lake Waters. *Remote Sensing*. 9 (5), 420. https://doi.org/10.3390/rs9050420.

EPA (2015). *Recommendations for Public Water Systems to Manage Cyanotoxins in Drinking Water*.

Fang, C., Song, K., Paerl, H.W., Jacinthe, P., Wen, Z., Liu, G., Tao, H., Xu, X., Kutser, T., Wang, Z., Duan, H., Shi, K., Shang, Y., Lyu, L., Li, S., Yang, Q., Lyu, D., Mao, D., Zhang, B., Cheng, S., Lyu, Y., 2022. Global divergent trends of algal blooms detected by satellite during 1982–2018. *Global Change Biology*. 28 (7), 2327–2340. https://doi.org/10.1111/gcb.16077.

Gikuma-Njuru, P., Hecky, R.E., 2005. Nutrient concentrations in Nyanza Gulf, Lake Victoria, Kenya: light limits algal demand and abundance. *Hydrobiologia*. 534 (1–3), 131–140. https://doi.org/10.1007/s10750-004-1418-9.

Gikuma-Njuru, P., Hecky, R.E., Guildford, S.J., MacIntyre, S., 2013. Spatial variability of nutrient concentrations, fluxes, and ecosystem metabolism in Nyanza Gulf and Rusinga Channel, Lake Victoria (East Africa). *Limnology and Oceanography*. 58 (3), 774–789. https://doi.org/10.4319/lo.2013.58.3.0774.

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*.

Gons, H.J., 1999. Optical Teledetection of Chlorophyll *a* in Turbid Inland Waters. *Environmental Science & Technology*. 33 (7), 1127–1132. https://doi.org/10.1021/es9809657.

Gons, H.J., 2002. A chlorophyll-retrieval algorithm for satellite imagery (Medium Resolution Imaging Spectrometer) of inland and coastal waters. *Journal of Plankton Research*. 24 (9), 947–951. https://doi.org/10.1093/plankt/24.9.947.

Gordon, H.R. & Morel, A.Y. (1983). *Remote Assessment of Ocean Color for Interpretation of Satellite Visible Imagery: A Review*. Lecture Notes on Coastal and Estuarine Studies. American Geophysical Union, Washington, D. C. doi:10.1029/LN004.

Grossman, A.R., Schaefer, M.R., Chiang, G.G., Collier, J.L., 1993. Environmental effects on the light-harvesting complex of cyanobacteria. *Journal of Bacteriology*. 175 (3), 575–582. https://doi.org/10.1128/jb.175.3.575-582.1993.

Halleux, V. & Pichon, E. (2023). *UN Sustainable Development Goal 6 on clean water and sanitation (SDG 6)*. https://www.europarl.europa.eu/RegData/etudes/BRIE/2023/751404/EPRS_BRI(2023)751404_EN.pdf.

Hecky, R.E., Mugidde, R., Ramlal, P.S., Talbot, M.R., Kling, G.W., 2010. Multiple stressors cause rapid ecosystem change in Lake Victoria. *Freshwater Biology*. 55, 19–42. https://doi.org/10.1111/j.1365-2427.2009.02374.x.

Herrnegger, M., Stecher, G., Schwatke, C., Olang, L., 2021. Hydroclimatic analysis of rising water levels in the Great rift Valley Lakes of Kenya. *Journal of Hydrology: Regional Studies*. 36, 100857. https://doi.org/10.1016/j.ejrh.2021.100857.

Huisman, J., Codd, G.A., Paerl, H.W., Ibelings, B.W., Verspagen, J.M.H., Visser, P.M., 2018. Cyanobacterial blooms. *Nature Reviews Microbiology*. 16 (8), 471–483. https://doi.org/10.1038/s41579-018-0040-1.

Humphrey, O.S., Osano, O., Aura, C.M., Marriott, A.L., Dowell, S.M., Blake, W.H., Watts, M.J., 2022. Evaluating spatio-temporal soil erosion dynamics in the Winam Gulf catchment, Kenya for enhanced decision making in the land-lake interface. *Science of the Total Environment*. 815, 151975. https://doi.org/10.1016/j.scitotenv.2021.151975.

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.

IOCCG (2010). *Atmospheric Correction for Remotely-Sensed Ocean-Colour Products.* Reports of the International Ocean-Colour Coordinating Group. Dartmouth, Canada.

Jackson, T., Sathyendranath, S., Mélin, F., 2017. An improved optical classification scheme for the Ocean Colour Essential Climate Variable and its applications. *Remote Sensing of Environment*. 203, 152–161. https://doi.org/10.1016/j.rse.2017.03.036.

Jiang, D., Scholze, J., Liu, X., Simis, S.G.H., Stelzer, K., Müller, D., Hunter, P., Tyler, A., Spyrakos, E., 2023. A data-driven approach to flag land-affected signals in satellite derived water quality from small lakes. *International Journal of Applied Earth Observation and Geoinformation*. 117, 103188. https://doi.org/10.1016/j.jag.2023.103188.

Johnsen, G., Sakshaug, E., 1996. Light Harvesting in bloom-forming marine phytoplankton: species-specificity and photoacclimation. *Scientia Marina*. 60 (1), 47–56.

Kent, J.T., Mardia, K.V., 1988. Spatial classification using fuzzy membership models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 10 (5), 659–671. https://doi.org/10.1109/34.6774.

Kiage, L.M., Obuoyo, J., 2011. The Potential Link Between El Nino and Water Hyacinth Blooms in Winam Gulf of Lake Victoria, East Africa: Evidence from Satellite Imagery. *Water Resources Management*. 25 (14), 3931–3945. https://doi.org/10.1007/s11269-011-9895-x.

Kravitz, J., Matthews, M., Lain, L., Fawcett, S., Bernard, S., 2021. Potential for High Fidelity Global Mapping of Common Inland Water Quality Products at High Spatial and Temporal Resolutions Based on a Synthetic Data and Machine Learning Approach. *Frontiers in Environmental Science*. 9, 587660. https://doi.org/10.3389/fenvs.2021.587660.

Kruse, F.A., Lefkoff, A.B., Boardman, J.W., Heidebrecht, K.B., Shapiro, A.T., Barloon, P.J. & Goetz, A.F.H. (1993). The spectral image processing system (SIPS)-interactive visualization and analysis of imaging spectrometer data. In: *AIP Conference Proceedings*. 1993 AIP, Pasadena, California (USA). pp. 192–201. doi:10.1063/1.44433.

Kutser, T., Metsamaa, L., Strömbeck, N., Vahtmäe, E., 2006. Monitoring cyanobacterial blooms by satellite remote sensing. *Estuarine, Coastal and Shelf Science*. 67 (1–2), 303–312. https://doi.org/10.1016/j.ecss.2005.11.024.

Liu, X., Steele, C., Simis, S., Warren, M., Tyler, A., Spyrakos, E., Selmes, N., Hunter, P., 2021. Retrieval of Chlorophyll-a concentration and associated product uncertainty in optically diverse lakes and reservoirs. *Remote Sensing of Environment*. 267, 112710. https://doi.org/10.1016/j.rse.2021.112710.

Loisa, O., Kääriä, J., Saario, J., Niemi, J., Laaksonlaita, J., Sarvala, J., 2015. Real-time applications in accurate monitoring of harmful cyanobacteria blooms: possibilities and benefits to the management and the public. *Water Practice and Technology*. 10 (4), 699–703. https://doi.org/10.2166/wpt.2015.085.

Matthews, M.W., 2011. A current review of empirical procedures of remote sensing in inland and near-coastal transitional waters. *International Journal of Remote Sensing*. 32 (21), 6855–6899. https://doi.org/10.1080/01431161.2010.512947.

Matthews, M.W., Bernard, S., Robertson, L., 2012. An algorithm for detecting trophic status (chlorophyll-a), cyanobacterial-dominance, surface scums and floating vegetation in inland and coastal waters. *Remote Sensing of Environment*. 124, 637–652. https://doi.org/10.1016/j.rse.2012.05.032.

Matthews, M.W., Odermatt, D., 2015. Improved algorithm for routine monitoring of cyanobacteria and eutrophication in inland and near-coastal waters. *Remote Sensing of Environment*. 156, 374–382. https://doi.org/10.1016/j.rse.2014.10.010.

Mchau, G.J., Makule, E., Machunda, R., Gong, Y.Y., Kimanya, M., 2019. Harmful algal bloom and associated health risks among users of Lake Victoria freshwater: Ukerewe Island. Tanzania. *Journal of Water and Health*. 17 (5), 826–836. https://doi.org/10.2166/wh.2019.083.

Merel, S., Walker, D., Chicana, R., Snyder, S., Baurès, E., Thomas, O., 2013. State of knowledge and concerns on cyanobacterial blooms and cyanotoxins. *Environment International*. 59, 303–327. https://doi.org/10.1016/j.envint.2013.06.013.

Moore, T.S., Campbell, J.W., Feng, H., 2001. A fuzzy logic classification scheme for selecting and blending satellite ocean color algorithms. *IEEE Transactions on Geoscience and Remote Sensing*. 39 (8), 1764–1776. https://doi.org/10.1109/36.942555.

Moore, T.S., Campbell, J.W., Dowell, M.D., 2009. A class-based approach to characterizing and mapping the uncertainty of the MODIS ocean chlorophyll product. *Remote Sensing of Environment*. 113 (11), 2424–2430. https://doi.org/10.1016/j.rse.2009.07.016.

Moore, T.S., Dowell, M.D., Bradt, S., Ruiz Verdu, A., 2014. An optical water type framework for selecting and blending retrievals from bio-optical algorithms in lakes and coastal waters. *Remote Sensing of Environment*. 143, 97–111. https://doi.org/10.1016/j.rse.2013.11.021.

Moses, W.J., Gitelson, A.A., Berdnikov, S., Povazhnyy, V., 2009. Estimation of chlorophyll- *a* concentration in case II waters using MODIS and MERIS data—successes and challenges. *Environmental Research Letters*. 4 (4), 045005. https://doi.org/10.1088/1748-9326/4/4/045005.

Neil, C., Spyrakos, E., Hunter, P.D., Tyler, A.N., 2019. A global approach for chlorophyll-a retrieval across optically complex inland waters based on optical water types. *Remote Sensing of Environment*. 229, 159–178. https://doi.org/10.1016/j.rse.2019.04.027.

O'Shea, R.E., Pahlevan, N., Smith, B., Bresciani, M., Egerton, T., Giardino, C., Li, L., Moore, T., Ruiz-Verdu, A., Ruberg, S., Simis, S.G.H., Stumpf, R., Vaičiūtė, D., 2021. Advancing cyanobacteria biomass estimation from hyperspectral observations: Demonstrations with HICO and PRISMA imagery. *Remote Sensing of Environment*. 266, 112693. https://doi.org/10.1016/j.rse.2021.112693.

Otieno, D., Nyaboke, H., Nyamweya, C.S., Odoli, C.O., Aura, C.M., Outa, N.O., 2022. Water hyacinth (*Eichhornia crassipes*) infestation cycle and interactions with nutrients and aquatic biota in Winam Gulf (Kenya), Lake Victoria. *Lakes & Reservoirs: Science, Policy and Management for Sustainable Use*. 27 (1), e12391. https://doi.org/10.1111/lre.12391.

Paerl, H.W., Hall, N.S., Calandrino, E.S., 2011. Controlling harmful cyanobacterial blooms in a world experiencing anthropogenic and climatic-induced change. *Science of the Total Environment*. 409 (10), 1739–1745. https://doi.org/10.1016/j.scitotenv.2011.02.001.

Paerl, H.W., Otten, T.G., 2013. Harmful Cyanobacterial Blooms: Causes, Consequences, and Controls. *Microbial Ecology*. 65 (4), 995–1010. https://doi.org/10.1007/s00248-012-0159-y.

Paerl, H.W., Paul, V.J., 2012. Climate change: Links to global expansion of harmful cyanobacteria. *Water Research*. 46 (5), 1349–1363. https://doi.org/10.1016/j.watres.2011.08.002.

Pahlevan, N., Smith, B., Schalles, J., Binding, C., Cao, Z., Ma, R., Alikas, K., Kangro, K., Gurlin, D., Hà, N., Matsushita, B., Moses, W., Greb, S., Lehmann, M.K., Ondrusek, M., Oppelt, N., Stumpf, R., 2020. Seamless retrievals of chlorophyll-a from Sentinel-2 (MSI) and Sentinel-3 (OLCI) in inland and coastal waters: A machine-learning approach. *Remote Sensing of Environment*. 240, 111604. https://doi.org/10.1016/j.rse.2019.111604.

Pahlevan, N., Mangin, A., Balasubramanian, S.V., Smith, B., Alikas, K., Arai, K., Barbosa, C., Bélanger, S., Binding, C., Bresciani, M., Giardino, C., Gurlin, D., Fan, Y., Harmel, T., Hunter, P., Ishikaza, J., Kratzer, S., Lehmann, M.K., Ligi, M., Ma, R., Martin-Lauzer, F.-R., Olmanson, L., Oppelt, N., Pan, Y., Peters, S., Reynaud, N., Sander De Carvalho, L.A., Simis, S., Spyrakos, E., Steinmetz, F., Stelzer, K., Sterckx, S., Tormos, T., Tyler, A., Vanhellemont, Q., Warren, M., 2021a. ACIX-Aqua: A global assessment of atmospheric correction methods for Landsat-8 and Sentinel-2 over lakes, rivers, and coastal waters. *Remote Sensing of Environment*. 258, 112366. https://doi.org/10.1016/j.rse.2021.112366.

Pahlevan, N., Smith, B., Binding, C., Gurlin, D., Li, L., Bresciani, M., Giardino, C., 2021b. Hyperspectral retrievals of phytoplankton absorption and chlorophyll-a in inland and nearshore coastal waters. *Remote Sensing of Environment*. 253, 112200. https://doi.org/10.1016/j.rse.2020.112200.

Pietroiusti, R., Vanderkelen, I., Otto, F.E.L., Barnes, C., Temple, L., Akurut, M., Bally, P., Van Lipzig, N.P.M., Thiery, W., 2024. Possible role of anthropogenic climate change in the record-breaking 2020 Lake Victoria levels and floods. *Earth System Dynamics*. 15 (2), 225–264. https://doi.org/10.5194/esd-15-225-2024.

Roegner, A., Sitoki, L., Weirich, C., Corman, J., Owage, D., Umami, M., Odada, E., Miruka, J., Ogari, Z., Smith, W., Rejmankova, E., Miller, T.R., 2020. Harmful Algal Blooms Threaten the Health of Peri-Urban Fisher Communities: A Case Study in Kisumu Bay, Lake Victoria. Kenya. *Exposure and Health*. 12 (4), 835–848. https://doi.org/10.1007/s12403-019-00342-8.

Romero, J.R., Imberger, J., Antenucci, J.P., Ewing, T.C., Khisa, P. & Njuguna, H. (2005). Management implications of the physical imnological studies of Rusinga Channel and Winam Gulf in Lake Victoria. In: *ILEC*. 2005 Nairobi, Kenya. p.

Ruiz-Verdú, A., Simis, S.G.H., De Hoyos, C., Gons, H.J., Peña-Martínez, R., 2008. An evaluation of algorithms for the remote sensing of cyanobacterial biomass. *Remote Sensing of Environment*. 112 (11), 3996–4008. https://doi.org/10.1016/j.rse.2007.11.019.

Simis, S., Crètaux, J.-F., Yèsou, H., Malnes, E., Vickers, H., Pablo, B., Merchant, C., Carrea, L., Duguay, C. & Wu, Y. (2022). *D2.2: Algorithm Theoretical Basis Document (ATBD)*.

Simis, S.G.H., Calmettes, B., Crétaux, J.-F., Liu, X., Yésou, H., Carrea, L., Duguay, C. & Wu, Y. (2023a). *Lakes_CCI+ - Phase 2 – D2.1. Product Validation and Algorithm Selection Report (PVASR)*. (2).

Simis, S.G.H., Peters, S.W.M., Gons, H.J., 2005. Remote sensing of the cyanobacterial pigment phycocyanin in turbid inland water. *Limnology and Oceanography*. 50 (1), 237–245. https://doi.org/10.4319/lo.2005.50.1.0237.

Simis, S.G.H., Ruiz-Verdú, A., Domínguez-Gómez, J.A., Peña-Martinez, R., Peters, S.W.M., Gons, H.J., 2007. Influence of phytoplankton pigment composition on remote sensing of cyanobacterial biomass. *Remote Sensing of Environment*. 106 (4), 414–427. https://doi.org/10.1016/j.rse.2006.09.008.

Simis, S.G.H., Carrea, L., Calmettes, B., Crétaux, J.-F., Duguay, C., Liu, X., Mangili, A., Müller, D., Yésou, H., 2023b. *Lakes_cci+ - Phase 2. D2.2. Algorithm Theoretical Basis Document* ATBD). (2.

Simiyu, B.M., Amukhuma, H.S., Sitoki, L., Okello, W., Kurmayer, R., 2022. Interannual variability of water quality conditions in the Nyanza Gulf of Lake Victoria. Kenya. *Journal of Great Lakes Research*. 48 (1), 97–109. https://doi.org/10.1016/j.jglr.2021.10.017.

Simiyu, B., Oduor, S., Rohrlack, T., Sitoki, L., Kurmayer, R., 2018. Microcystin Content in Phytoplankton and in Small Fish from Eutrophic Nyanza Gulf, Lake Victoria. Kenya. *Toxins*. 10 (7), 275. https://doi.org/10.3390/toxins10070275.

Sitoki, L., Gichuki, J., Ezekiel, C., Wanda, F., Mkumbo, O.C., Marshall, B.E., 2010. The Environment of Lake Victoria (East Africa): Current Status and Historical Changes. *International Review of Hydrobiology*. 95 (3), 209–223. https://doi.org/10.1002/iroh.201011226.

Sitoki, L., Kurmayer, R., Rott, E., 2012. Spatial variation of phytoplankton composition, biovolume, and resulting microcystin concentrations in the Nyanza Gulf (Lake Victoria, Kenya). *Hydrobiologia*. 691 (1), 109–122. https://doi.org/10.1007/s10750-012-1062-8.

Smith, B., Pahlevan, N., Schalles, J., Ruberg, S., Errera, R., Ma, R., Giardino, C., Bresciani, M., Barbosa, C., Moore, T., Fernandez, V., Alikas, K., Kangro, K., 2021. A Chlorophyll-a Algorithm for Landsat-8 Based on Mixture Density Networks. *Frontiers in Remote Sensing*. 1, 623678. https://doi.org/10.3389/frsen.2020.623678.

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., Ogashawara, I., Ruiz-Verdú, A., Schalles, J.F., Tebbs, E., Zhang, Y. & Tyler, A.N. (2018). Optical types of inland and coastal waters: Optical types of inland and coastal waters. *Limnology and Oceanography*. 63 (2), 846–870. doi:10.1002/lno.10674.

Stumpf, R.P., Davis, T.W., Wynne, T.T., Graham, J.L., Loftin, K.A., Johengen, T.H., Gossiaux, D., Palladino, D., Burtner, A., 2016. Challenges for mapping cyanotoxin patterns from remote sensing of cyanobacteria. *Harmful Algae*. 54, 160–173. https://doi.org/10.1016/j.hal.2016.01.005.

Tandeau De Marsac, N., 1977. Occurrence and nature of chromatic adaptation in cyanobacteria. *Journal of Bacteriology*. 130 (1), 82–91. https://doi.org/10.1128/jb.130.1.82-91.1977.

Taranu, Z.E., Gregory-Eaves, I., Leavitt, P.R., Bunting, L., Buchaca, T., Catalan, J., Domaizon, I., Guilizzoni, P., Lami, A., McGowan, S., Moorhouse, H., Morabito, G., Pick, F.R., Stevenson, M.A., Thompson, P.L. & Vinebrooke, R.D. (2015). Acceleration of cyanobacterial dominance in north temperate-subarctic lakes during the Anthropocene P. Jeyasingh (ed.). *Ecology Letters*. 18 (4), 375–384. doi:10.1111/ele.12420.

Tebbs, E.J., Remedios, J.J., Harper, D.M., 2013. Remote sensing of chlorophyll-a as a measure of cyanobacterial biomass in Lake Bogoria, a hypertrophic, saline–alkaline, flamingo lake, using Landsat ETM+. *Remote Sensing of Environment*. 135, 92–106. https://doi.org/10.1016/j.rse.2013.03.024.

Tebbs, E.J., Remedios, J.J., Avery, S.T., Rowland, C.S., Harper, D.M., 2015. Regional assessment of lake ecological states using Landsat: A classification scheme for alkaline–saline, flamingo lakes in the East African Rift Valley. *International Journal of Applied Earth Observation and Geoinformation*. 40, 100–108. https://doi.org/10.1016/j.jag.2015.03.010.

Tibshirani, R., Walther, G., Hastie, T., 2001. Estimating the Number of Clusters in a Data Set Via the Gap Statistic. *Journal of the Royal Statistical Society Series b: Statistical Methodology*. 63 (2), 411–423. https://doi.org/10.1111/1467-9868.00293.

Wang, F., 1990. Fuzzy classification of remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*. 28 (2), 194–201. https://doi.org/10.1109/36.46698.

Wyman, M. & Fay, P. (1986a). Underwater light climate and the growth and pigmentation of planktonic blue-green algae (Cyanobacteria) I. The influence of light quantity. *Proceedings of the Royal Society of London. Series B. Biological Sciences*. 227 (1248), 367–380. doi:10.1098/rspb.1986.0027.

Wyman, M. & Fay, P. (1986b). Underwater light climate and the growth and pigmentation of planktonic blue-green algae (Cyanobacteria) II. The influence of light quality. *Proceedings of the Royal Society of London. Series B. Biological Sciences*. 227 (1248), 381–393. doi:10.1098/rspb.1986.0028.

Wynne, T.T., Stumpf, R.P., Tomlinson, M.C., Warner, R.A., Tester, P.A., Dyble, J., Fahnenstiel, G.L., 2008. Relating spectral shape to cyanobacterial blooms in the Laurentian Great Lakes. *International Journal of Remote Sensing*. 29 (12), 3665–3672. https://doi.org/10.1080/01431160802007640.

Wynne, T.T., Stumpf, R.P., Tomlinson, M.C., Dyble, J., 2010. Characterizing a cyanobacterial bloom in Western Lake Erie using satellite imagery and meteorological data. *Limnology and Oceanography*. 55 (5), 2025–2036. https://doi.org/10.4319/lo.2010.55.5.2025.