ISPRS Journal of Photogrammetry and Remote Sensing 228 (2025) 323–339

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

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

Colour-informed ecoregion analysis highlights a satellite capability gap for spatially and temporally consistent freshwater cyanobacteria monitoring

Davide Lomeo [a,*], Stefan G.H. Simis [b], Nick Selmes [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

ARTICLE INFO

Keywords:
Cyanobacteria
Optical water types
Inland waters
Ocean color
Ecoregions
Self-organising map

ABSTRACT

Cyanobacteria blooms pose significant risks to water quality in freshwater ecosystems worldwide, with implications for human and animal health. Constructing consistent records of cyanobacteria dynamics in complex inland waters from satellite imagery remains challenged by discontinuous sensor capabilities, particularly with regard to spectral coverage. Comparing 11 satellite sensors, we show that the number and positioning of wavebands fundamentally alter bloom detection capability, with wavebands centred at 412, 620, 709, 754 and 779 nm proving most critical for capturing cyanobacteria dynamics. Specifically, analysis of observations from the Medium Resolution Imaging Spectrometer (MERIS) and Ocean and Land Colour Instrument (OLCI), coincident with the Moderate Resolution Imaging Spectroradiometer (MODIS) demonstrates how the spectral band configuration of the latter affects bloom detection. Using an Optical Water Types (OWT) library understood to capture cyanobacterial biomass through varying vertical mixing states, this analysis shows that MODIS can identify optically distinct conditions like surface accumulations but fails to resolve initial bloom evolution in well-mixed conditions, particularly in optically complex regions. Investigation of coherent ecoregions formed using Self-organising Maps trained on OWT membership scores confirm that MODIS captures broad spatial patterns seen with more capable sensors but compresses optical gradients into fewer optical types. These constraints have significant implications for interpreting spatial–temporal dynamics of cyanobacteria in large waterbodies, particularly during 2012–2016 when MERIS and OLCI sensors were absent, and small waterbodies, where high spatial resolution sensors not originally design to study water are used. In addition, these findings underscore the importance of key wavebands in future sensor design and the development of approaches to maintain consistent long-term records across evolving satellite capabilities. Our findings suggest that attempts at quantitatively harmonising cyanobacteria bloom detection across sensors may not be ecologically appropriate unless these observation biases are addressed. For example, analysing the frequency and intensity of surfacing blooms, while considering the meteorological factors that may drive these phenomena, could be considered over decadal timescales, whereas trend analysis of mixed-column biomass should only concern appropriate sensor observation periods.

1. Introduction

Generating consistent long-term satellite records of cyanobacteria dynamics in inland waters remains a fundamental challenge to understanding the compound impacts of land use and climate change. Cyanobacteria blooms can rapidly alter water quality through cascading biogeochemical changes, and certain species can produce toxins that pose a health concern for humans and animals (Huisman et al., 2005; Paerl and Otten, 2013; Chorus et al., 2021). Cyanobacteria blooms manifest as vertically mixed populations as well as (near) surface accumulations, associated with varying characteristic optical features (Kutser et al., 2006; Kutser, 2009). Capturing bloom dynamics from their onset to their deterioration requires specific spectral information that is predominantly collected with global-scale coverage by ocean-colour sensors operated within the last two decades, but with significant gaps.

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

https://doi.org/10.1016/j.isprsjprs.2025.07.030
Received 1 April 2025; Received in revised form 12 July 2025; Accepted 20 July 2025
Available online 22 July 2025

Remote sensing approaches for monitoring cyanobacteria have evolved thanks to the understanding of their bio-optical characteristics in foundational studies. Early work to determine cyanobacteria biomass from water colour focussed on the absorption features of the prominent photosynthetic pigments phycocyanin and chlorophyll-*a*, quantified through specific reflectance waveband combinations (Dekker, 1993; Jupp et al., 1994). The phycocyanin absorption peak at 615 nm is captured in the 620 nm wavebands of recent ocean colour sensors, and can be quantified in turbid water types in the same way that chlorophyll-*a* absorption centred on 675 nm is quantified against the 709 nm waveband where pigment absorption is largely absent (Dall'Olmo et al., 2003; Simis et al., 2005; Ruiz-Verdú et al., 2008; Gilerson et al., 2010). Algorithm development, particularly in optically complex inland waters, has progressed from reflectance line height formulations to spectral decomposition, reflectance ratios (Dekker, 1993; Gower et al., 2005; Simis et al., 2007) and second derivatives (Wynne et al., 2008; Matthews et al., 2012). Advances in detecting and quantifying cyanobacteria presence have highlighted fundamental challenges in implementing these approaches across different sensor capabilities, particularly when specific spectral information needed to track bloom evolution is not consistently available across satellite missions. Machine learning models have further proven promising to retrieve a range of biogeochemical and optical properties from multispectral sensors that would otherwise lack the appropriate waveband configurations (Pahlevan et al., 2020; Smith et al., 2021; Werther et al., 2022), including phycocyanin (O'Shea et al., 2021). However, despite these successes, limited spectral resolution remains a primary limiting factor to resolve optically complex conditions (Balasubramanian et al., 2025), and therefore to track cyanobacteria growth and blooms.

Several satellite sensors designed for ocean colour applications have provided capabilities to match the requirements for observing water quality from medium to large eutrophic inland waterbodies. These include the Medium Resolution Imaging Spectrometer (MERIS) and its successor, the Ocean and Land Colour Instrument (OLCI), both with waveband configurations that have proven relevant to cyanobacteria biomass detection. However, between these missions (2012–2016), key wavebands centred at 620 and 709 nm are lacking from remaining ocean colour sensors such as the Moderate Resolution Imaging Spectroradiometer (MODIS). The 620, 681, and 709 nm wavebands are key examples of sensor design dichotomy, with MERIS and OLCI designs, and more recently the Geostationary Ocean Color Imager II (GOCI-II), offering more bands to resolve additional optical complexity, while the Visible Infrared Imaging Radiometer Suite (VIIRS) follows on from the design of MODIS and predecessors (Fig. 1). This division particularly impacts the retrieval of key phytoplankton pigment absorption and fluorescence behaviours from low biomass (when chlorophyll-a fluorescence is seen in the 681 nm waveband) to high biomass (when light absorption and particle scattering dominate the red to near infrared signal), and with increasing presence of cyanobacteria in eutrophic conditions, showing phycocyanin absorption at 620 nm, usually as the dominant accessory light-harvesting pigment (Simis et al., 2005, 2007; Matthews et al., 2012). Long-term records of lake water quality from

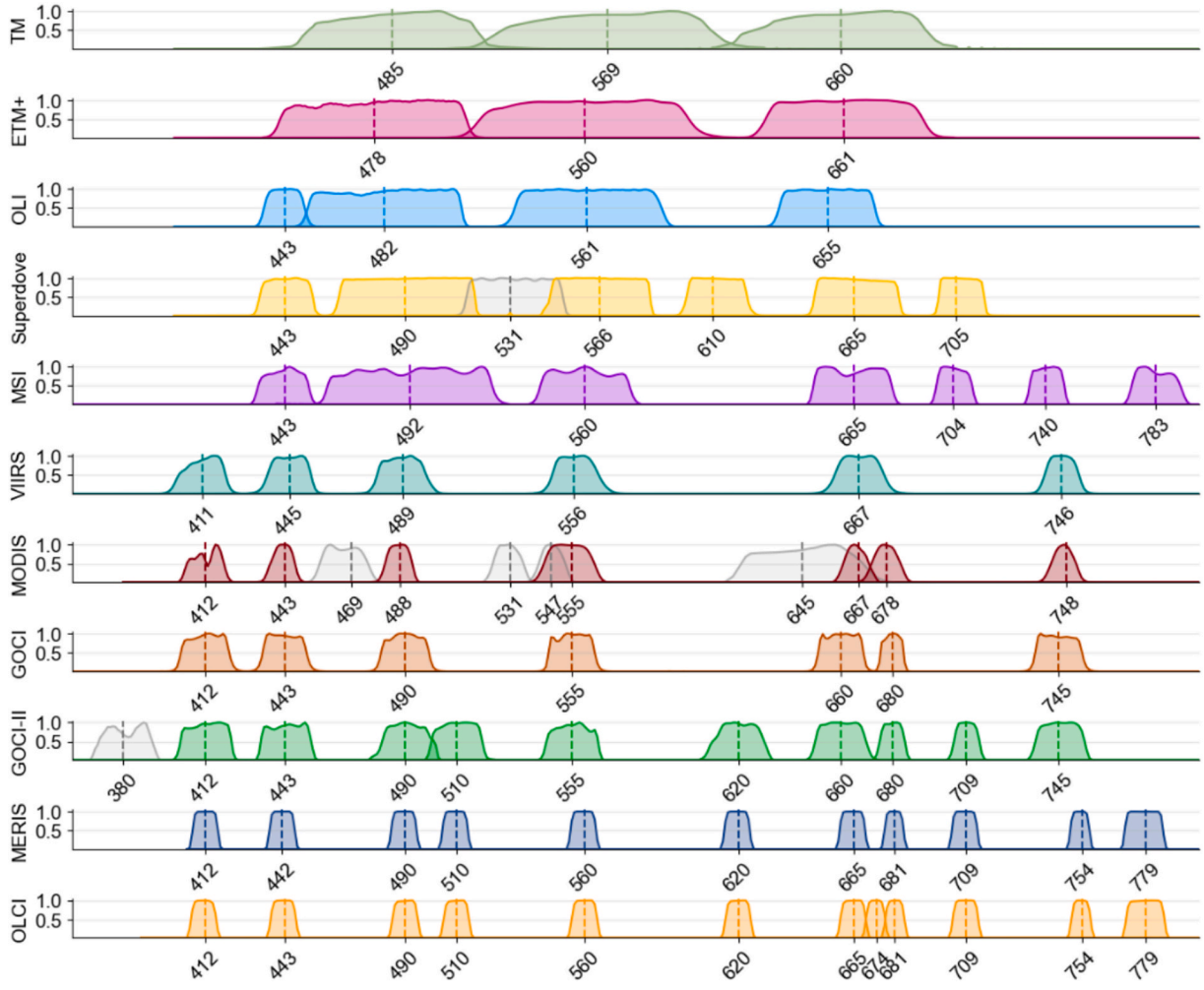

Fig. 1. Spectral response functions (SRF) of eleven optical sensors for the wavebands available in the range 380–779 nm, excluding the oxygen bands for MERIS (761 nm) and OLCI (761, 764 and 767 nm). Curves in grey represent the wavebands that are not present in OLCI.

optical satellite sensors tend to be either based on analysing single-sensors which have limited capability (Shi et al., 2017; Wang et al., 2020) or combining multiple sensors (Liu et al., 2021, 2024) by optimising per-sensor algorithm sets. There is this scope for further improving consistency, for example by prioritising the most capable sensors (such as MERIS or OLCI), while carefully defaulting to less capable sensors (such as MODIS) only when needed, but always emphasising consistency in the retrieval methods.

Optical Water Types (OWT) frameworks have brought significant advances in handling the wide optical diversity of natural waters, over which individual algorithms are unlikely to perform consistently (Morel and Prieur, 1977; Kent and Mardia, 1988; Moore et al., 2001; Spyrakos et al., 2018). Predetermined OWT sets can be used to determine the similarity of new observations to the reference set, assigning weights to algorithms known to perform adequately against samples belonging to that OWT (Moore et al., 2014; Liu et al., 2021). This approach ideally leverages global reference datasets to determine the suitability of individual algorithms to each OWT, subsequently allowing geospatial extrapolation of the algorithm and its associated uncertainty to other waterbodies that express similar optical behaviours. OWT classifications are also constrained by the spectral capabilities of the sensors for which they are defined: the mathematical space used to define an OWT set with a given number of wavebands becomes compressed as the number of available wavebands decreases. Water types that appear distinct given many wavebands, may thus become mathematically similar with fewer spectral dimensions, even when representing different ecological conditions.

Understanding the systematic limitations to cyanobacteria monitoring introduced by discontinuous sensor capabilities requires a framework to evaluate how well sensors capture various cyanobacteria growth phases, from initial development through peak biomass and eventual decline. OWTs form one part of this framework, with multiple sensors ideally observing the same distribution of OWT membership for a given set of observations. Partitioning natural waters into spatially distinct regions of similar biogeochemical characteristics, also referred to as ecoregions, provides another route to analyse observation consistency. Ecoregion analysis has widely been used to study broad-scale biogeochemical patterns in ocean colour (Platt and Sathyendranath, 1988; Longhurst, 1995). This has included studies of global phytoplankton dynamics using factors such as wind, euphotic depth, surface irradiance, nutrients availability and chlorophyll-a concentrations (D'Ortenzio, 2009; D'Ortenzio et al., 2012; IOCCG, 2009; Ardyna et al., 2017). Ecoregion partitioning is achieved using several techniques, with Self-Organising Maps (SOM; Kohonen, 1982, 2001) proving especially useful in dealing with the complexity of natural waters observed using remote sensing data (Yacoub et al., 2001; Niang, 2003). SOMs are neural network-based methods that map high-dimensional data onto lower dimensions (typically two) preserving the topological relationship of the input data (Liu et al., 2006). As such, SOMs are both a projection and a clustering method that order similar patterns onto neighbouring SOM units or nodes (Richardson et al., 2003; Liu and Weisberg, 2011). The use of SOMs has demonstrated useful in several applications, such as meteorology (Richardson et al., 2003; Liu et al., 2006), in situ water quality analysis (Richardson, 2002; Chazottes et al., 2007), land cover and ocean colour classifications (Yacoub et al., 2001; Niang, 2003; Villmann et al., 2003), and satellite-derived chlorophyll-*a* analysis (Diouf et al., 2013; Fendereski et al., 2014; El Hourany et al., 2019; Liu et al., 2022), proving more advantageous than other commonly used dimensionality reduction and clustering techniques (Astel et al., 2007). The ability of SOMs to capture and encapsulate the complex relationship between biogeochemical characteristics and seasonal cycles into ecoregions has the potential to provide useful information on sensor capabilities, including the sensitivity of different sensors to various phases of cyanobacteria growth over space and time. We may further consider that SOMs trained on OWT memberships have the potential to reveal additional underlining biases in detecting cyanobacteria dynamics that other strategies would be less likely to capture.

While hyperspectral sensors will not suffer the spectral limitations described in this work, multispectral sensors offer baseline observations from which changes in ecosystem health can be determined and still provide the only global coverage with up to daily revisit. As more efforts to bridge the gaps between past and present satellite missions are undertaken, it is hence useful to understand how different multispectral sensor capabilities affect the characterisation of cyanobacteria bloom development, thus to what extent OWT classifications remain reliant on the spectral resolution of sensors. By examining how well different sensor configurations distinguish between optical features classified into a given OWT set, it is possible to assess their ability to consistency resolve optical transitions. In principle, this can be shown by exploring theoretical sensor capabilities that assume equivalent measurement conditions, radiometric sensitivities, and atmospheric correction biases, examining the impacts of different waveband configurations and individual waveband contributions to the sensitivity to different bloom conditions (IOCCG, 2012).

Here, we use a set of OWTs that can represent the range of optical conditions encountered in freshwater bodies (Spyrakos et al., 2018), currently used in global-scale lake water quality services (e.g., Copernicus, ESA CCI), extended with the OLCI-derived OWTs developed by Lomeo et al. (2025), for enhanced cyanobacteria sensitivity and understood to capture a wide range of cyanobacterial population growth through varying mixing states. We hypothesise that sequentially removing wavebands from the reference OWT set will determine the degree to which each contributes to estimating cyanobacteria occurrence. Comparative analysis of current ocean colour sensor capabilities is then possible. In previous work, Lomeo et al., (2025) have shown that a weighted OWT membership score sum (W_{sum}) provides a useful single metric to characterise the risk of cyanobacteria occurrence, further reducing the complexity of the underlying OWT membership. This experiment is expected to provide a baseline understanding of the implications for monitoring capabilities, as well as for OWT definitions and related downstream products. Additionally, we hypothesise that examining the full OWT membership distribution over space and time can provide deeper insights into the fundamental constraints in how cyanobacteria growth through varying mixing states is detected.

Extending OWT analysis of sensor biases to ecoregion analysis is done here for Lake Victoria, an economically important water body with a long history of cyanobacteria blooms (Sitoki et al., 2010). We compare OWT membership scores and W_{sum} values from coincident observations between MODIS-MERIS (2011) and MODIS-OLCI (2019) to understand how reduced spectral information affects OWT classification in practice. By examining both the magnitude and pattern of differences in OWT membership, specific optical conditions, where MODIS systematically over- or underestimates similarity to reference water types, can be identified. We further compare OWT retrieval from matching sets of water-leaving reflectance wavebands across the three sensors to elucidate on individual sensor sensitivity to observe a given OWT when biases resulting from atmospheric correction are accounted for. OWT membership scores for coincident observations are subsequently used to train SOMs, partitioning Lake Victoria into spatial-temporally coherent ecoregions. The distribution of W_{sum} across the identified ecoregions is used to evidence individual sensor capabilities in characterising cyanobacteria bloom development. Key waveband ratios representative of phytoplankton pigment in the blue-green (443/ 560 nm), and NIR-red (709/ 665 and 754/ 665 nm) are used to provide evidence of the ecological validity of the identified ecoregions.

Ultimately, this work is intended to establish whether MODIS observations during 2012–2016 can be reliably harmonised with MERIS and OLCI for long-term monitoring of cyanobacteria growth, bloom and decay, or whether fundamental limitations result in systematic gaps in our understanding of inland water responses to environmental change. These insights provide crucial context for interpreting historical satellite records and developing robust approaches for maintaining consistent

long-term monitoring capabilities across sensor technologies.

2. Methods

2.1. Study site

Lake Victoria is the largest tropical lake in the world and the second largest inland waterbody globally. Its shoreline is shared between Tanzania, Uganda, and Kenya (Fig. 2). Approximately 42 million people live within the lake catchment and depend on it for drinking water and commercial activities like fishery (Olokotum et al., 2020). The lake has a surface area extending to approximately 68,000 km^2, mean depth of 40 m, and maximum depth of 80 m (Johnson et al., 2000; Bootsma and Hecky, 2003), with lake water levels fluctuating considerably over time (Vanderkelen et al., 2018). The lake is subject to a bi-modal rainfall seasonality. The 'long rains' typically occur between March and May due to the northward movement of the Intertropical Convergence Zone (ITCZ), while 'short rains' occur between September and November, when the ITCZ progresses southward (Yang et al., 2015). Precipitation typically occurs at night and during early morning hours (Thiery et al., 2015). Since the 1950 s the lake has experienced increasing eutrophication due to anthropogenic activities within its basin, with consequent implications to water quality (Sitoki et al., 2010). Some of the most optically dynamic regions are found along the shorelines and within semi-enclosed basins like Winam Gulf (Kenya), Murchison Bay (Uganda) and Mwanza and Speke Gulfs (Tanzania) (Frank et al., 2023; Nakkazi et al., 2024).

2.2. Datasets

Satellite imagery was processed to level 3C (L3C) by the Natural Environment Research Council Earth Observation Data Analysis and Artificial-Intelligence Service (NEODAAS, UK) for the years 2003–2023. The dataset included 981 MERIS scenes in the period 2003–2012, 1389 MODIS Aqua scenes in 2012–2016, and 2372 OLCI A/B scenes in 2016–2023. All images were projected onto a 1/448 degree grid, which approximates to 250 m resolution at the equator. A further two years of MODIS Aqua observations for the years 2011 and 2019 were used for sensor intercomparison with MERIS (2011) and OLCI (2019), these being the nearest years with full matching sensor coverage. The choice of using MODIS on the Aqua platform, with a slightly later overpass (1:30 pm) compared to OLCI and MERIS (10:30 am), was driven by the documented higher water-leaving radiance retrieval stability compared to MODIS Terra, for which the radiometric response in certain wavebands has degraded substantially since launch (Franz, 2008). MODIS Aqua (MODIS hereafter) has been used in the ESA Lakes Climate Change Initiative (Lakes_cci) to derive multi-decadal chlorophyll-a concentration for large lakes. Images were processed by NEODAAS through the *Calimnos* processing chain, using the candidate configuration for version 3.0 of the ESA Lakes_cci Climate Research Data Package (Simis et al., 2022a; Liu et al., 2021). This new configuration uses Polymer v4.17b for atmospheric correction, which employs an extended range of initialisation conditions compared to previous versions, improving retrieval of turbid water conditions including near or at-surface blooms.

2.3. Sampling strategy

A random stratified spatial sampling strategy was used to identify suitable areas for spatial–temporal intercomparison between MERIS, MODIS, and OLCI, and to reduce computational requirements for the analysis. Specifically, 2,100 windows of 3 x 3 pixels were selected, with the number of windows exponentially decreasing as a function of distance from land. This strategy captures the (expected) greater variability of water colour in nearshore areas and semi-enclosed basins. 1,333 windows were thus randomly selected within the first 20 km from land, 428 in the range 20–40 km, 182 in the range 40–60 km, and 94 beyond 60 km. Areas within 2 km from land were masked to account for the higher likelihood of the influence of land adjacency, such as mixed land/water pixels and atmospheric mixing of land and water-reflected signals. The selected 2,100 windows were further inspected for consistency, and windows for which any of their 9 pixels had been masked by atmospheric correction (e.g., due to presence of cloud or cloud-shadow) were removed from analysis. The remaining pixels were treated as individual observations and included for subsequent analysis. This approach resulted in a reduction of observations from approximately 11 billion across the three sensors to 4 million observations for MERIS, 9.5 million for MODIS, and 12 million for OLCI. Coincident observations between MODIS and MERIS in 2011 were 192,665, whereas for MODIS and OLCI in 2019 were 804,287.

An additional stratified sampling was implemented on the reduced observation sets for later training of neural networks and to enable a comparable water colour partitioning across the lake, further discussed below. MERIS, MODIS and OLCI observations were harmonised across matching days of year, ranging 1–365, disregarding the year of overpass and records on the 366th day of leap years. For each day, observations were proportionally sampled within the previously identified windows

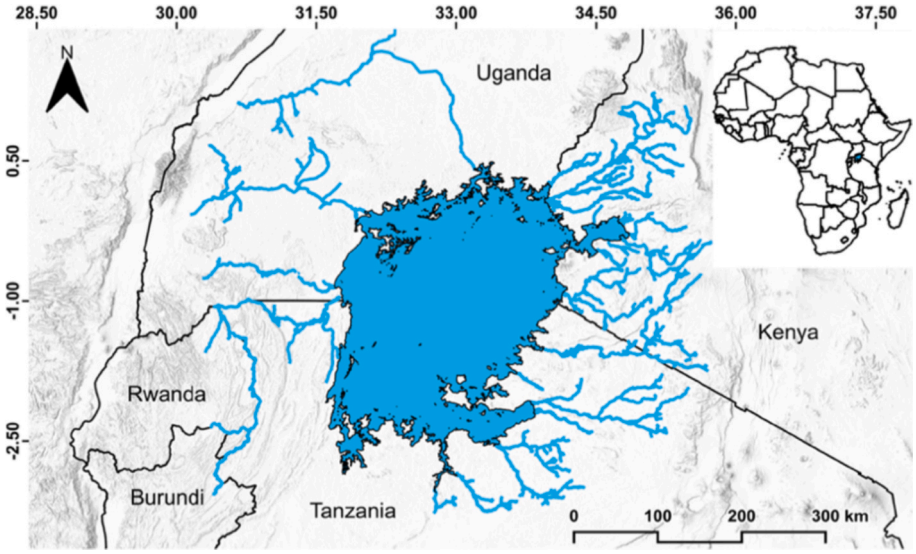

Fig. 2. Map of Lake Victoria and its major rivers.

to ensure a balanced spatial representation, with a minimum threshold of 100 observations per day. This approach yielded approximately 3 million observations for each sensor, or 10,000 observations per day across 322 days, maintaining comparable spatial–temporal distributions.

2.4. Optical water types classification

Water colour in Lake Victoria was mapped using a library of 25 optical water types (OWT) for inland waters (Lomeo et al., 2025), an extension of the 13 OWT for inland waters by Spyrakos et al. (2018). This extended library was formulated using OLCI-derived spectra to better determine cyanobacteria occurrence, including mixed and surface accumulated bloom conditions (Lomeo et al., 2025). A further two type spectra developed to flag pixels affected by land adjacency (Jiang et al., 2023) were also used. OWT similarity scores (S_{owt}) were calculated for each pixel using the spectral angle metric (Kruse et al., 1993), emphasising similarity in spectral shape rather than amplitude, as recommended by Liu et al. (2021), using the following equations:

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

$$S_{owtj} = 1 - a_j/\pi \quad (2)$$

where p_i and r_i are the standardised pixel and reference spectra in band i, respectively. The resultant S_{owtj} is the membership score for OWT j, represented as a number between 0 and 1, where 1 indicates identical spectral shapes. The S_{owt} values were calculated between the 25 OWTs, and the sensor-derived standardised spectra in the range 412–779 nm for MERIS, MODIS and OLCI, excluding the oxygen bands at 761, 764 and 767 nm for OLCI, and 762 nm for MERIS. Satellite-derived spectra, like the OWTs, were standardised by dividing over their integrals to reduce the influence of varying reflectance amplitudes and instead focus on the similarity of their shapes (Liu et al., 2021). Observations for which the highest S_{owt} value, also referred to as dominant water type (Moore et al., 2001, 2014), was associated to either of the two land effected types, were removed.

It is worth adding that the library of 25 OWTs by Lomeo et al., (2025) was originally formulated by clustering OLCI-derived spectra for which the dominant OWT was associated to the presence of cyanobacteria, according to the definitions by Spyrakos et al. (2018), into 'subtypes'. These new subtypes were named after the dominant OWT they originated from (e.g., OWT 1, 7), together with the cluster number they belonged to (e.g., OWT 1.2 or 7.2). Some subtypes originating from different dominant OWTs were merged if they spectral similarity was high, hence the naming 1.1_7.1 or 4.1_11.1. We invite the reader to refer to Lomeo et al., (2025) for an in-depth explanation of the clustering process.

2.5. OWT-based cyanobacteria occurrence

The weighted membership score sum (W_{sum}) metric (Lomeo et al., 2025) was used to evaluate the spatial–temporal dynamics of cyanobacteria bloom occurrence risk across Lake Victoria. W_{sum} combines S_{owt} values obtained with Eq. (2) with a subjective ranking informed by typical optical indicators of cyanobacteria presence in the reference spectral library of 25 OWT using the sum of 709/620 nm and 709/681 nm waveband ratios (Lomeo et al., 2025). The W_{sum} is calculated as follows:

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

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 an OWT ranking obtained by converting the sum of the 709/620 and 709/681 waveband ratios of the 25 OWT library into an order-of-magnitude differentiation. The ranking categorises cyanobacteria occurrence risk into low ($w = 0$), medium ($w = 100$), and high ($w = 1000$), reflecting the increasing likelihood of cyanobacteria presence based on either pigment absorption indicators and/or the formation of (sub)surface layers. $P_{i,j}$ is a proportionality factor that accounts for the uneven distributions of S_{owt} values of a given observation across the set of OWTs, inherently capturing the covariance of similarity scores across the OWT set. The resulting W_{sum} is a number expected to range between 223 (absent to low cyanobacteria occurrence risk) and 321 (high to very high risk) in productive turbid waters, though the range may extend in the low-end for less productive, clear waters (Lomeo et al., 2025).

2.6. Sensor water-leaving reflectance and W_{sum} intercomparison and uncertainty evaluation

The water-leaving reflectance (R_w) stability between MERIS, MODIS, and OLCI was assessed using all available wavebands within 6 nm distance for coincident observations, i.e., observations with aligning date, latitude, and longitude within shared valid windows (as described in section 2.3), in 2011 for MODIS-MERIS, and in 2019 for MODIS-OLCI. These were the wavebands centred at 412, 443, 490 (488), 560 (555), 665 (667), 681 (678), and 754 (748) nm, where the centre wavelength in brackets refers specifically to MODIS (Fig. 1).

S_{owt} and W_{sum} values obtained from the set of 25 OWTs for coincident observations were compared to evaluate the ability of each sensor to capture different OWT definitions and associated cyanobacteria determination, respectively. Here, OLCI was used as the reference sensor because of its appropriate spectral configuration for cyanobacteria occurrence monitoring and its superior radiometric sensitivity compared to its predecessors. S_{owt} and W_{sum} values were calculated on the full range of wavebands available to each sensor, as well as for MERIS and OLCI waveband subsets defined to match only MODIS wavebands ($MERIS_M$ and $OLCI_M$, respectively). Using the full range available to the three sensors helped to establish the degree of 'spectral information loss effect' introduced by MODIS against two more capable ocean colour sensors. Using the subsets ranges substantiated OLCI and MERIS 'upstream bias', i.e., their sensitivity to cyanobacteria presence compared to MODIS using 'equivalent' waveband configurations.

Additionally, W_{sum} values were iteratively calculated using the reference set of 25 OWTs after sequentially dropping wavebands. This exercise revealed the contribution of each OLCI waveband in the overall sensitivity of the sensor to cyanobacteria presence. Furthermore, the 25 OWT set was adapted to 'match' the waveband configurations of ten widely used optical sensors (Fig. 1) to evaluate how cyanobacteria detection is affected when assuming virtually equivalent atmospheric conditions and radiometric sensitivities.

To quantify the differences between sensors, five statistical metrics were used: Root Mean Squared Difference (RMSD), Median Absolute Percentage Difference (MAPD), slope, bias, and R^2. These metrics were selected to capture both the magnitude and the nature of sensor differences, while characterising the systematic relationships between sensor measurements. These were calculated as follows:

$$RMSD = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - x_j)^2} \quad (4)$$

$$MAPD = \frac{1}{n}\sum_{i=1}^{n} med\left\{\frac{|x_i - x_j|}{x_j}\right\} \times 100\% \quad (5)$$

$$Bias = x_i - x_j \quad (6)$$

where x_i are the R_w, S_{owt} or W_{sum} of MODIS, and x_j are those from either MERIS or OLCI. The slope and R^2 were obtained from linear regressions

of S_{owt} and W_{sum} values for coincident observations across the three sensors. Observations for which the highest S_{owt} value was < 0.9 were removed from the statistical analysis to reduce the influence of outliers, followed by the removal of the top and bottom 1 % of the remaining observations. Observations with the highest $S_{owt} < 0.9$ may not be well-captured within the OWT set for several reasons, including atmospheric artifacts introduced by atmospheric correction, or spectra highly influenced by land adjacency. In addition, the 443/560, 709/665 and 754/665 waveband ratios were used to provide context to the distribution of W_{sum} values across optically similar regions in the lake (described in more details below).

2.7. Self-organising maps

The Self-organising Map (SOM) technique, also known as Kohonen map (Kohonen, 1982, 2001), was used to partition Lake Victoria into regions (ecoregions) of similar spatial–temporal water colour patterns. Ecoregions were used to classify the biogeochemical and seasonal characteristics of spatial patterns in the lake, and to assess sensor-specific biases in cyanobacteria detection.

The Python implementation 'MiniSom' (Vettigli, 2018) was used to train SOMs. Typically, a grid of arbitrary size is defined, where each grid element is a node that stores the weights generated during the training process. While it has been suggested that the number of required nodes needed is linked to sample size (Vesanto and Alhoniemi, 2000), it is understood that the grid size should be dependent on the required granularity of information extracted during training, as well as the complexity (degrees of freedom) of the data (Richardson et al., 2003; Liu et al., 2006; Chazottes et al., 2007; Wang et al., 2019). It may be tempting to set the number of nodes to the required number of end-classes, but a (too) small grid may lead to pattern overgeneralisation (i.e., underfitting), whereas a (too) large grid may lead to pattern blending (i.e., overfit) (Reusch et al., 2005). Several preliminary configurations were attempted, and a grid of 10 x 10 size (Fig., 3) was deemed the most appropriate to balance between the need to capture enough complexity and avoid model under/ overfitting. Prior to training, weights are randomly initialised on each node. The shape of the nodes in grid was set to hexagonal, such that each node was connected to a minimum of three neighbours (at the boundaries of the grid), and up to six (Fig. 3). This shape is usually preferred to using square nodes in remote sensing studies that use large, complex, datasets, because it is able to capture shared patterns more effectively (Liu et al., 2006). The training was run for 1 million iterations to allow the inherent complexity of the input data to be fully represented within the grid. At each iteration, each observation vector (or row vector) is compared to all nodes using an 'activation distance', typically Euclidian, and the node with the smallest difference is the 'winner', becoming the centre of the updated 'nodes neighbourhood'. This approach allows the data to automatically converge towards similar patterns. Individual observations (i.e., pixels) were selected at random from the input dataset, so that each observation would be used for training multiple times (given the observation count was smaller than 1 million as described above). The rate at which the input converges is controlled by a dynamic learning rate, which was set to 0.01 at the start of training, and that decreased as the algorithm learned to more accurately converge and topologically sort nodes in the map (Liu et al., 2006). Additionally, as the training progresses, the area considered for convergence (or sigma) reduces, which ensures that patterns are progressively refined from an initial, more generalised map (grid). Again, several preliminary configurations were attempted, and using a sigma of 4 was chosen because it provided the most balanced output, whereby a smaller sigma led to patchier node sorting, and a larger sigma led to two patterns to dominate the grid.

The input variables for SOM training were the 25 S_{owt} values obtained with Eq. (2), the maximum R_w amplitude in the range 412–779 nm, and the sine and cosine deconstruction of the day of year (1–365). The 25 S_{owt} values are directly related to the optical characteristic of water against the reference library of 25 OWTs. These are preferred to using individual R_w wavebands because they represent an informed dimensionality reduction from raw R_w, capturing meaningful optical relationships rather than treating each waveband as an independent variable. The S_{owt} values inherently account for the covariance between wavebands while emphasising the spectral features most relevant to distinguishing water types, enabling more robust pattern recognition of biogeochemical conditions. The maximum R_w amplitude, or the peak reflectance across the target sensor wavebands range, complements the 25 S_{owt} values by preserving information about particle scattering that would otherwise be lost in standardised spectra. Since absorption is more featured than scattering and is what primarily determines spectral shape, including the peak amplitude helps distinguish between waters with similar spectral shapes but different particle characteristics. The sine and cosine of day-of-year provide a sinusoidal representation of time, capturing seasonal patterns of water colour while retaining the

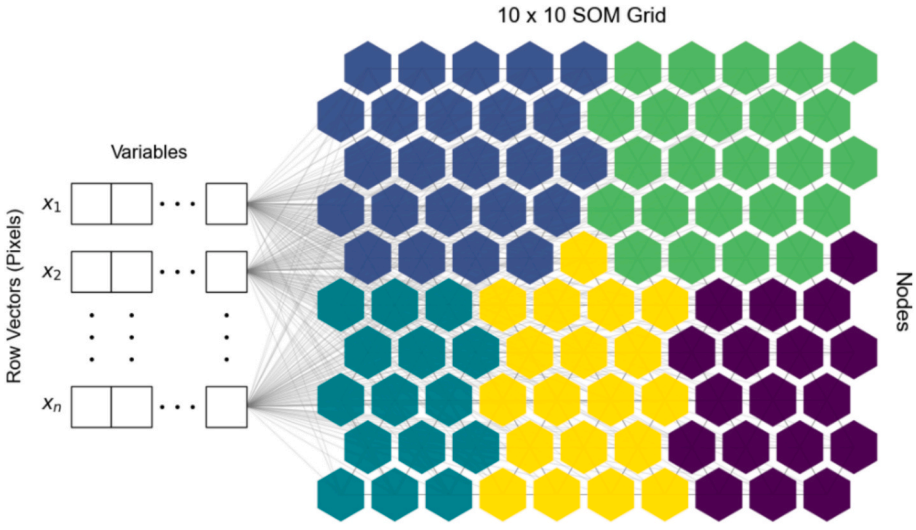

Fig. 3. Self-organising map (SOM) design. The input variables for the SOM training were the 25 optical water types (OWT) membership scores (S_{owt}), the max amplitude (or the highest water-leaving reflectance value at each pixel), and the sin and cos decomposition of time (day of year). The input row vectors (pixels) are matched to each node in the 10 x 10 SOM grid (map), and the weights in the nodes are updated. The nodes in the figure are colour-coded to (hypothetical) clusters formed using the weights of the nodes after training.

cyclical nature of these processes without discontinuities at the turn of the year. The 25 S_{owt} values and peak R_w amplitude were normalised across the full dataset using a z-score standardisation to ensure all variables contributed equally to the SOM regardless of their original units and ranges. This scaling was essential as SOM training uses distance measures to determine similarity, and without normalisation, variables with larger absolute ranges would disproportionately influence the clustering.

Five SOMs were trained using the harmonised datasets previously described. Three were trained using sets with S_{owt} values calculated on native sensor waveband configurations for MERIS, MODIS, and OLCI. Two SOMs were trained for MERIS and OLCI sets with S_{owt} values calculated on the same waveband configuration as MODIS ($MERIS_M$ and $OLCI_M$).

2.8. Identification of ecoregions

Although trained nodes are already organised in the SOM maps by shared characteristics, the number of patterns they underpin is related to the grid size. Therefore, it is useful to further reduce the information space to a meaningful number of groups (clusters), which in this case translates into the number of ecoregions (Fig. 3). The trained nodes of the five SOM models were clustered using a Hierarchical Agglomerative Clustering (HAC) (Jain and Dubes, 1988). HAC is an unsupervised classification method that provides a structured clustering relative to the distance (e.g., Euclidian) between clusters, and that does not require a priori indication of cluster numbers. These can be identified at a later stage using a linkage matrix (or dendrogram).

The SOM trained on OLCI data was used as the reference classification and divided into five ecoregions. While not directly supplemented by in-situ biological validation, this was expected to provide reasonable distinction between nearshore and (deeper) open water processes as well as any phytoplankton succession, that could still be captured by sensors with somewhat reduced capacity to distinguish all OWTs. The centroids of these ecoregions were obtained averaging the weight vectors across nodes belonging to the same cluster. These represented the (average) optical characteristic of each ecoregion in multidimensional space. The nodes in the remaining four SOMs were assigned to the nearest reference OLCI ecoregion centroid in terms of Euclidean distance. This approach ensured that ecoregions were numbered consistently across sensors, whilst allowing for sensor intercomparison.

3. Results

3.1. Influence of waveband presence or absence on W_{sum}

Sequentially removing individual wavebands from the reference library of 25 OLCI OWT spectra revealed that W_{sum} values are primarily affected by the presence and absence of wavebands centred at 412, 560, 620, 709, 754, and 779 nm (Fig. 4A). For OWTs representing the highest risk of cyanobacteria occurrence (i.e., w = 1000), removing the waveband centred at 412 nm led to positive W_{sum} biases, overestimating the risk of cyanobacteria occurrence. Removing wavebands centred at 754 and 779 nm primarily affected W_{sum} values for OWTs that are not typically linked to high cells backscattering in these regions, which is often indicative of surface accumulation of material (like OWT 1.2, 7.2). The omission of wavebands centred at 709 nm and 620 nm yielded negative W_{sum} biases, particularly in OWTs associated with cyanobacterial dominance. The influence of the 709 nm waveband diminished with decreasing risk of cyanobacteria occurrence. In clear waters, the effect of removing this waveband reversed, leading to positive biases. Similar patterns were repeated for OWTs belonging to the other two

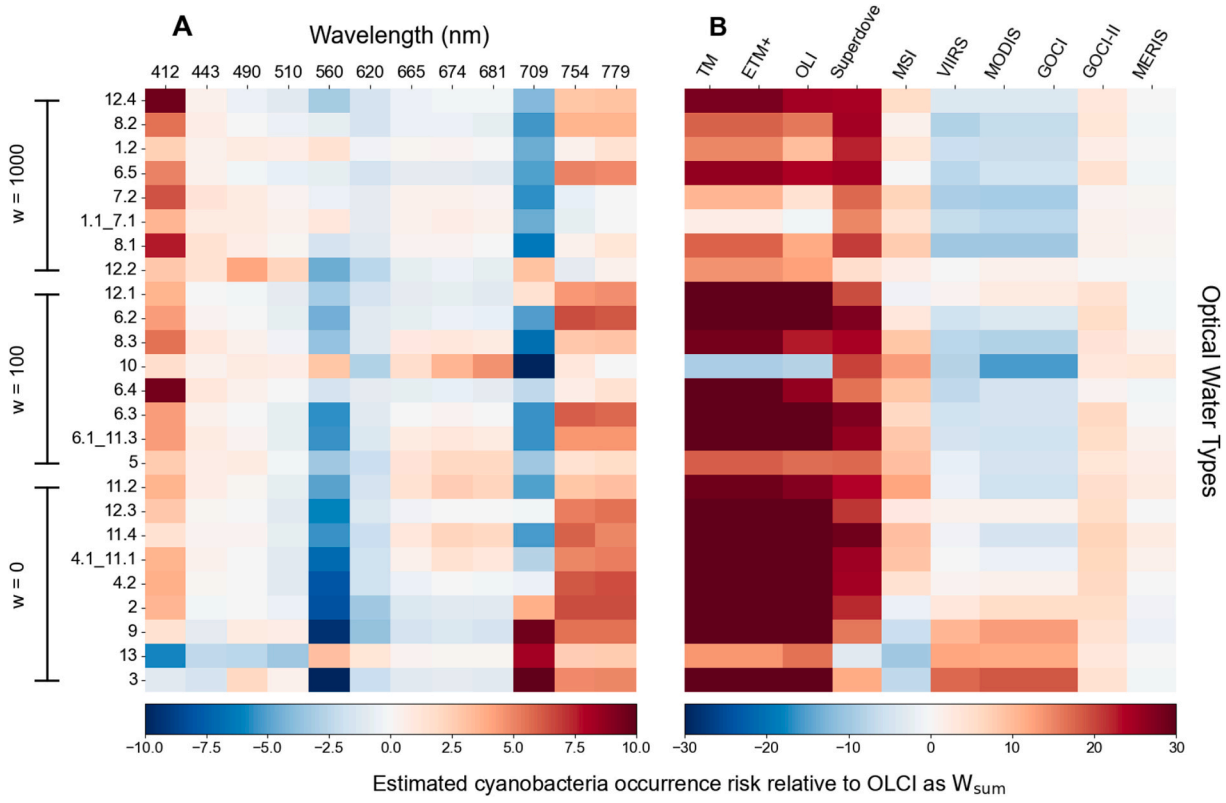

Fig. 4. (A) Bias between the weighted optical water type (OWT) membership score sum (W_{sum}) values calculated using the full OLCI waveband configuration and those calculated after sequentially removing each waveband. Negative W_{sum} bias suggests underestimation of the occurrence of cyanobacteria, and vice versa. (B) W_{sum} bias when comparing ten theoretical optical sensor configurations with OLCI. Bias expresses the difference in W_{sum} compared to using the full OLCI waveband configuration.

groups of cyanobacteria occurrence risk (w = 100 and w = 0), with increasing impact of the absence of the waveband centred at 560 nm and decreasing influence of the waveband centred at 709 nm. The removal of the 560 nm band led to (high) negative W_{sum} biases in OWTs associated with a lower cyanobacteria occurrence risk (for example, OWT 9 and 3).

Optical sensors originally designed for ocean colour applications showed the closest alignment with OLCI in the prediction of cyanobacteria occurrence (as W_{sum}). These sensors include MERIS and GOCI-II, while others generally overestimated W_{sum} across OWTs (Fig. 4B). Landsat sensor configurations exhibited highly positive W_{sum} bias across most OWTs, although they performed slightly better for OWTs associated with surface bloom conditions. Despite including wavebands centred near 620 and 709 nm, Superdove configurations showed consistently poor agreement with OLCI across all OWTs. MSI primarily yielded positive W_{sum} bias, particularly in OWTs associated with medium cyanobacteria occurrence risk (i.e., w = 100), while showing negative bias for OWTs associated to clear water (e.g., 13, 3), and minor positive bias in OWTs associated to cyanobacteria dominated water. VIIRS, MODIS, and GOCI displayed similar patterns, characterised by negative W_{sum} biases in OWTs associated with high cyanobacteria occurrence risk, transitioning to positive biases for OWTs associated with the lowest risk of cyanobacteria occurrence/ clear waters.

3.2. MERIS, MODIS and OLCI radiometric stability

Comparison of aligning R_w wavebands for coincident observations in the MERIS, MODIS and OLCI L3C datasets showed fair agreement between sensors (Fig. 5 and Table 1). The largest consistency between MODIS and MERIS was found for the wavebands centred at 560 (555) nm, 665 (667) nm, and 681 (678) nm, with MAPD = 41.7 %, 27.6 % and 20.1 % respectively, and average R^2 = 0.80. The largest disagreement between the sensors was found for the waveband centred at 412 nm, with MAPD = 76.7 % and slope = 1.61. R_w stability between MODIS and OLCI followed a similar pattern, with agreement across wavebands in the range 560 (555) – 681 (678) nm, and the strongest disagreement in the waveband centred at 412 nm, with MAPD = 93.3 %. Overall, MODIS showed better stability with MERIS compared to OLCI, although uncertainty in the optical region associated to the chlorophyll-*a* fluorescence red peak was relatively similar in both.

3.3. Stability of OWT similarity metrics between MODIS and MERIS or OLCI

Coincident observations of MODIS with MERIS and OLCI revealed that MODIS overestimated the majority of S_{owt} values compared to the other sensors (Fig. 6, Table 2). The magnitude of overestimation varied between OWTs, with the largest positive biases in S_{owt} observed for OWTs associated to high cyanobacteria occurrence risk (i.e., w = 1000). For example, for OWT 8.2 MODIS had biases of 0.116 and 0.113 against MERIS and OLCI, respectively, whereas for OWT 11.2, which is associated to low likelihood of cyanobacteria occurrence, the bias was 0.009 and −0.002 against MERIS and OLCI, respectively. Low regression slopes (0.26–0.69) and high intercepts (0.4–0.64) indicated that MODIS spectra tend to overestimate low similarity.

When the comparison was restricted to matching wavebands, MODIS underestimated S_{owt} values. Bias ranged consistently from −0.01 to −0.03, with slopes between 0.5 and 0.7 and lower intercepts (0.2–0.4). Comparing this result to the overestimation previously observed when using the complete set of wavebands of each sensor, therefore shows that this difference stems from the spectral angle calculated using different numbers of wavebands (7 for MODIS versus 11 for MERIS and 12 for OLCI). When evaluated in the same seven-dimensional spectral space, MODIS showed consistently larger angles with the reference spectra compared to MERIS and OLCI, leading to lower S_{owt} values across most OWTs. Expectedly, S_{owt} bias propagated to W_{sum}, with MODIS systematically overestimating by 23.6 compared to MERIS and by 20.8 compared to OLCI, on average (Table 2). Bias reduced significantly when using matching wavebands, giving 2.3 against MERIS and 0.6 against OLCI (Table 3).

Spatially, the distribution of W_{sum} across Lake Victoria showed higher values in dendritic and semi-enclosed basins across all sensors in both 2011 and 2019 (Fig. 7A, B, E, F). MODIS yielded higher average W_{sum} values across most of the lake compared to MERIS and OLCI, except in dendritic areas and in Winam Gulf, in the north-east sector of the lake, where values were often lower (Fig. 7I and 7J). These areas are those that see the highest median W_{sum}, suggesting that where

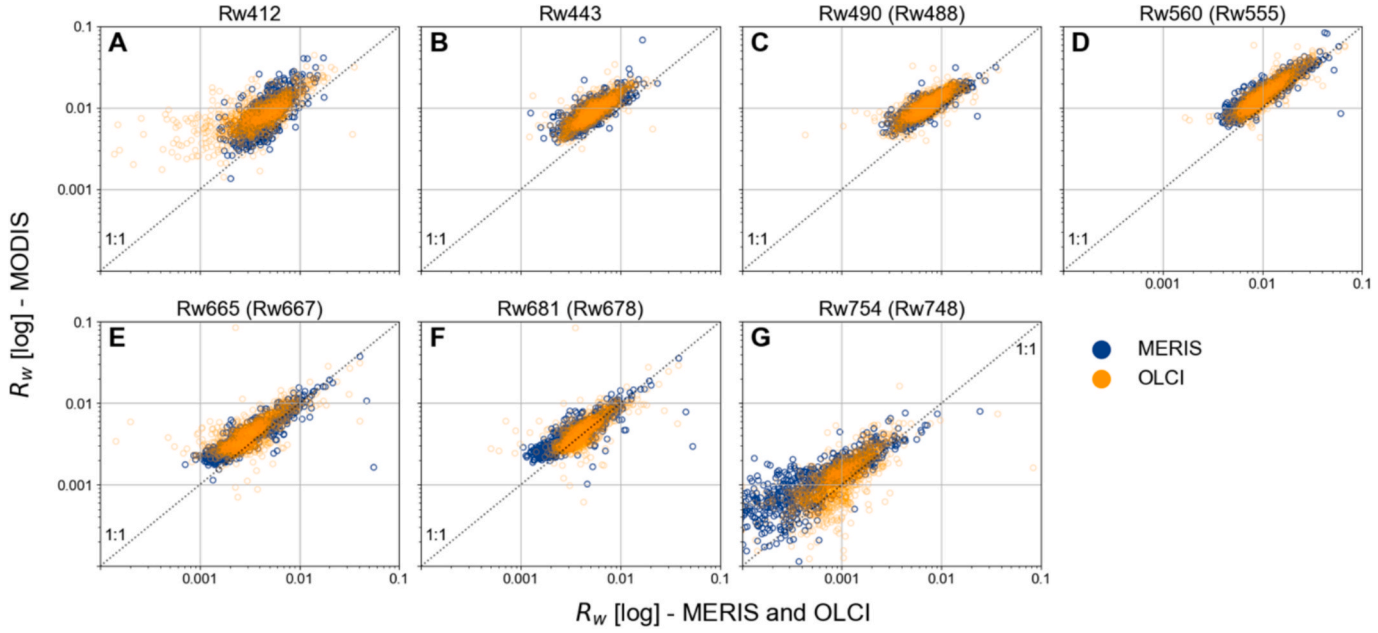

Fig. 5. Comparison of fully normalised water-leaving reflectance (R_w) wavebands of coincident observations for MODIS and MERIS/ OLCI (A-G) in Lake Victoria. The overlapping year are 2011 for MERIS and 2019 for OLCI. MODIS wavebands are indicated in brackets where their centre does not match with MERIS and OLCI wavebands. Both axes were log-scaled to enable visual comparison across all wavebands.

Table 1
Descriptive statistics of coincident MODIS and MERIS/ OLCI observations. Wavebands in brackets are the MODIS bands used for the comparison.

Sensors		MODIS vs MERIS (Average N = 186,219)						MODIS vs OLCI (Average N = 776,825)				
Wavebands	**R^2**	**Slope**	**Intercept**	**Bias**	**RMSD**	**MAPD (%)**	**R^2**	**Slope**	**Intercept**	**Bias**	**RMSD**	**MAPD (%)**
412 nm	0.48	1.61	0.0010	0.0034	0.0049	76.7	0.54	1.08	0.0040	0.0040	0.0052	93.3
443 nm	0.65	1.15	0.0024	0.0030	0.0037	59.4	0.69	1.21	0.0026	0.0035	0.0041	66.2
490 (488) nm	0.75	0.93	0.0035	0.0028	0.0033	41.3	0.67	0.97	0.0037	0.0034	0.0039	47.5
560 (555) nm	0.82	1.02	0.0038	0.0037	0.0048	41.7	0.79	1.12	0.0028	0.0038	0.0052	37.9
665 (667) nm	0.80	0.82	0.0013	0.0007	0.0012	27.6	0.67	0.86	0.0014	0.0009	0.0016	30.5
681 (678) nm	0.77	0.77	0.0014	0.0006	0.0011	20.1	0.62	0.78	0.0011	0.0001	0.0012	13.8
754 (748) nm	0.75	0.93	0.0005	0.0004	0.0006	58.3	0.60	1.01	0.0002	0.0002	0.0006	34.7

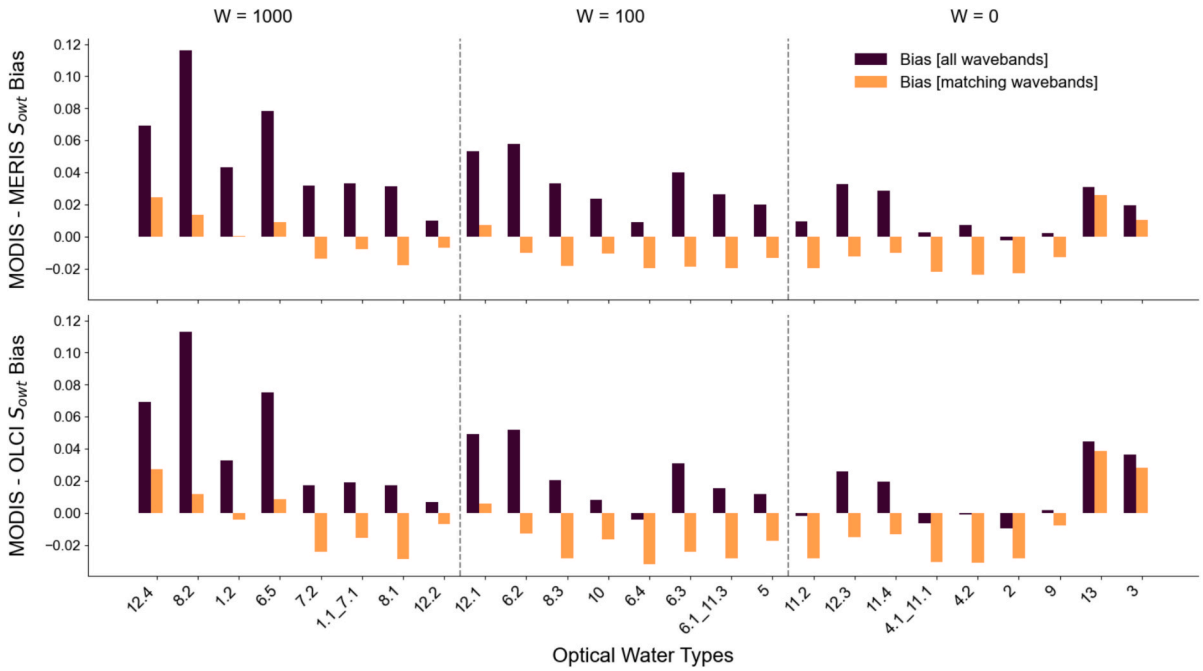

Fig. 6. Comparison of Optical Water Types (OWT) membership scores (S_{owt}) bias of MODIS observations coincident with MERIS or OLCI. S_{owt} bias was calculated using native waveband configurations of the sensors (purple bars), as well as for the subset of wavebands available on each sensor (orange bars). The OWT are grouped by ranking (W) using the vertical dotted lines. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

cyanobacteria occurrence risk is expected to be higher, MERIS and OLCI tend to have higher values than MODIS. In general, areas of lower W_{sum} values coincided with reduced observation counts in MODIS (Fig. 7D and 7H).

3.4. Ecoregion consistency across sensors

Partitioning of water colour in Lake Victoria using SOMs revealed consistent spatial patterns of ecoregions between sensors, particularly between MERIS and OLCI (Fig. 8). Ecoregion 1 was consistently associated with dendritic and semi-enclosed basins, ecoregion 5 primarily with areas alongside the coastlines and as a transition from ecoregion 1. Ecoregions 2, 3, and 4 where mostly associated to water colour dynamics in the centre of the lake, with high seasonal variability across all sensors. MODIS generally aligned with MERIS and OLCI in assigning ecoregions with the highest risk of cyanobacteria occurrence (i.e., ecoregions 1 and 5). However, MODIS was less sensitive to gradual seasonal evolution of cyanobacteria dynamics in the centre of the lake, and especially the transitioning between ecoregions.

The distribution of W_{sum} values across ecoregions confirmed that the SOMs successfully partitioned the spatial–temporal optical variability of the lake (Fig. 9A). OLCI and MERIS followed almost identical patterns, with ecoregion 1 yielding the highest risk of cyanobacteria occurrence, followed by ecoregion 5. Ecoregion 4, 3, and 2, showed gradually decreasing W_{sum} ranges, with ecoregion 2 yielding the lowest. The classification into the five ecoregions was also captured by MODIS, but W_{sum} ranges were narrower and with higher absolute values, and with small difference between ecoregions, especially 1 and 5. The lake partitioning performed by $MERIS_M$ followed almost equivalent patterns to MODIS, although with lower absolute values, whereas $OLCI_M$, despite aligning with the magnitude of W_{sum} values from MODIS, failed to discern between ecoregions 2–4.

The blue-green waveband ratio captured the optical differences between ecoregions across the three sensors, aligning with expectations that a higher ratio describes clearer water types with lower risk of cyanobacteria occurrence and vice versa (Fig. 9B). The NIR-red waveband ratio using the waveband centred at 709 nm (not available for MODIS) showed agreement between MERIS and OLCI, successfully discerning between ecoregions. This aligned with the expectation of increasingly higher values with increasing phytoplankton biomass, and with ecoregions associated to increasingly higher risk of cyanobacteria occurrence, although the difference between ecoregions 3 and 4 in OLCI was minimal (Fig. 9C). The NIR-red waveband ratio using the waveband centred at 754 (748) nm only captured differences between ecoregions in MERIS, with OLCI and MODIS failing to discern between ecoregions, showing a wide range for ecoregion 2 (Fig. 9D).

Table 2
Descriptive statistics of Optical Water Types (OWT) membership scores (S_{owt}) comparison between MODIS and MERIS/ OLCI, alongside the weighted S_{owt} sum (W_{sum}).

Sensors			MODIS vs MERIS (Average N = 181,029)					MODIS vs OLCI (Average N = 739,896)					
OWT	**W**	**R^2**	**Slope**	**Intercept**	**Bias**	**RMSD**	**MAPD (%)**	**R^2**	**Slope**	**Intercept**	**Bias**	**RMSD**	**MAPD (%)**
12.4	1000	0.14	0.44	0.521	0.069	0.072	8.6	0.28	0.41	0.539	0.069	0.074	8.6
1.2	1000	0.68	0.41	0.431	0.043	0.047	6.6	0.63	0.44	0.408	0.033	0.035	5.0
6.5	1000	0.48	0.48	0.505	0.078	0.084	9.5	0.50	0.54	0.455	0.075	0.077	9.1
7.2	1000	0.56	0.48	0.402	0.032	0.041	4.6	0.45	0.49	0.384	0.017	0.027	2.7
1.1_7.1	1000	0.62	0.45	0.406	0.033	0.040	4.9	0.54	0.48	0.382	0.019	0.025	2.8
8.1	1000	0.54	0.49	0.408	0.032	0.043	4.5	0.42	0.51	0.393	0.018	0.030	2.7
12.2	1000	0.16	0.26	0.644	0.010	0.027	1.8	0.16	0.35	0.567	0.007	0.019	1.3
8.2	1000	0.37	0.38	0.590	0.116	0.120	15.1	0.41	0.47	0.524	0.113	0.114	14.6
6.4	100	0.52	0.50	0.405	0.009	0.032	2.4	0.38	0.49	0.407	−0.004	0.027	1.9
5	100	0.51	0.43	0.472	0.020	0.034	2.9	0.47	0.47	0.438	0.012	0.024	1.9
6.1_11.3	100	0.57	0.51	0.416	0.026	0.040	3.7	0.47	0.53	0.397	0.015	0.029	2.4
6.3	100	0.57	0.52	0.429	0.040	0.050	5.0	0.47	0.55	0.404	0.031	0.039	3.8
10	100	0.57	0.43	0.419	0.024	0.036	3.7	0.55	0.46	0.390	0.008	0.022	1.9
8.3	100	0.56	0.51	0.412	0.033	0.044	4.5	0.46	0.52	0.392	0.020	0.032	2.9
6.2	100	0.57	0.54	0.446	0.058	0.065	6.9	0.46	0.56	0.422	0.052	0.056	6.1
12.1	100	0.53	0.53	0.454	0.053	0.060	6.3	0.53	0.55	0.435	0.049	0.053	5.8
11.2	0	0.56	0.48	0.405	0.009	0.031	2.4	0.47	0.50	0.393	−0.002	0.026	1.9
12.3	0	0.62	0.54	0.420	0.033	0.043	4.0	0.51	0.55	0.407	0.026	0.034	3.1
11.4	0	0.63	0.49	0.427	0.029	0.039	3.8	0.57	0.49	0.418	0.020	0.028	2.7
4.1_11.1	0	0.55	0.49	0.410	0.003	0.031	2.2	0.46	0.50	0.404	−0.007	0.028	2.0
4.2	0	0.54	0.51	0.421	0.007	0.033	2.3	0.43	0.51	0.416	−0.001	0.028	1.8
2	0	0.53	0.53	0.414	−0.002	0.032	2.0	0.37	0.53	0.413	−0.009	0.029	1.9
9	0	0.30	0.48	0.489	0.002	0.028	1.9	0.11	0.32	0.642	0.002	0.028	1.7
13	0	0.58	0.69	0.266	0.031	0.040	4.2	0.26	0.41	0.477	0.045	0.058	6.2
3	0	0.56	0.49	0.487	0.020	0.030	2.3	0.51	0.44	0.547	0.036	0.047	4.0
W_{sum}	–	0.58	0.38	181.870	23.638	25.418	9.3	0.54	0.43	169.633	20.816	21.354	8.0

Table 3
Descriptive statistics of Optical Water Types (OWT) membership scores (S_{owt}) comparison between MODIS and $MERIS_M$/ $OLCI_M$ (i.e., using the same waveband configuration as MODIS), alongside the Weighted S_{ow} sum (W_{sum}).

Sensors			MODIS vs $MERIS_M$ (Average N = 191,177)					MODIS vs $OLCI_M$ (Average N = 787,682)					
OWT	**W**	**R^2**	**Slope**	**Intercept**	**Bias**	**RMSD**	**MAPD (%)**	**R^2**	**Slope**	**Intercept**	**Bias**	**RMSD**	**MAPD (%)**
12.4	1000	0.13	0.54	0.412	0.025	0.033	3.0	0.25	0.31	0.612	0.027	0.043	3.4
1.2	1000	0.76	0.72	0.199	0.001	0.008	0.7	0.48	0.59	0.283	−0.004	0.011	0.8
6.5	1000	0.47	1.01	0.003	0.009	0.019	1.4	0.35	0.42	0.521	0.008	0.026	1.5
7.2	1000	0.68	0.65	0.249	−0.014	0.023	2.0	0.37	0.49	0.370	−0.024	0.034	3.2
1.1_7.1	1000	0.74	0.67	0.232	−0.008	0.015	1.3	0.46	0.56	0.310	−0.015	0.022	2.1
8.1	1000	0.66	0.64	0.263	−0.018	0.028	2.4	0.35	0.47	0.397	−0.029	0.040	3.6
12.2	1000	0.35	0.73	0.227	−0.007	0.016	1.0	0.15	0.41	0.509	−0.007	0.018	1.2
8.2	1000	0.39	0.98	0.029	0.014	0.021	1.7	0.33	0.38	0.554	0.012	0.028	1.8
6.4	100	0.60	0.63	0.280	−0.020	0.032	2.6	0.28	0.43	0.448	−0.032	0.044	3.9
5	100	0.67	0.67	0.256	−0.013	0.021	1.7	0.44	0.58	0.333	−0.017	0.025	2.2
6.1_11.3	100	0.69	0.65	0.272	−0.020	0.031	2.4	0.43	0.55	0.355	−0.028	0.038	3.4
6.3	100	0.69	0.67	0.266	−0.019	0.029	2.2	0.44	0.61	0.322	−0.024	0.034	2.8
10	100	0.71	0.65	0.244	−0.011	0.019	1.7	0.51	0.55	0.309	−0.016	0.024	2.3
8.3	100	0.69	0.65	0.265	−0.018	0.029	2.4	0.41	0.53	0.363	−0.028	0.039	3.5
6.2	100	0.66	0.77	0.193	−0.010	0.022	1.4	0.37	0.60	0.351	−0.013	0.025	1.7
12.1	100	0.52	1.02	−0.014	0.007	0.018	1.2	0.34	0.43	0.515	0.006	0.025	1.4
11.2	0	0.67	0.64	0.267	−0.020	0.030	2.5	0.42	0.51	0.371	−0.028	0.039	3.6
12.3	0	0.69	0.75	0.212	−0.012	0.023	1.6	0.42	0.63	0.314	−0.015	0.026	1.9
11.4	0	0.70	0.70	0.234	−0.010	0.020	1.5	0.48	0.60	0.314	−0.013	0.023	1.8
4.1_11.1	0	0.65	0.63	0.280	−0.022	0.033	2.7	0.40	0.50	0.389	−0.030	0.041	3.7
4.2	0	0.64	0.63	0.294	−0.024	0.035	2.8	0.38	0.52	0.396	−0.031	0.042	3.6
2	0	0.59	0.64	0.297	−0.023	0.035	2.6	0.31	0.53	0.399	−0.028	0.039	3.2
9	0	0.33	0.58	0.380	−0.013	0.028	1.9	0.06	0.24	0.714	−0.008	0.030	2.0
13	0	0.59	0.63	0.304	0.026	0.037	3.6	0.22	0.34	0.524	0.039	0.056	5.3
3	0	0.60	0.52	0.457	0.010	0.023	1.6	0.41	0.39	0.586	0.028	0.043	3.1
W_{sum}	–	0.66	0.94	19.270	2.341	3.853	1.0	0.47	0.65	99.656	0.600	3.819	0.8

The seasonal distribution of W_{sum} values by ecoregion for four selected areas of Lake Victoria showed that MODIS followed similar patterns to MERIS and OLCI, particularly in regions where water colour was less variable across seasons, while it consistently deviated in areas of greater seasonal variability (Fig. 10). For example, in the area covering the Ssese Islands and Murchison Bay (Fig. 10 A-B), MODIS-derived ecoregions showed a relatively good agreement with both MERIS and OLCI (Fig. 10 C-D). In open waters, on the other hand, particularly during seasons when the lowest cyanobacteria occurrence risk was predominant (i.e., ecoregion 2), MODIS systematically assigned these to ecoregions 3 and 4 (which represent a higher risk of cyanobacteria occurrence). When this portion of the lake experienced the predominance of ecoregion 1 in June – August 2019 (i.e., the ecoregion associated to the highest risk of cyanobacteria occurrence), MODIS but largely misclassified this to ecoregion 5 (Fig. 10 C-D). In Winam and Speke Gulfs (Fig. 10 E-H), where ecoregions associated to the highest

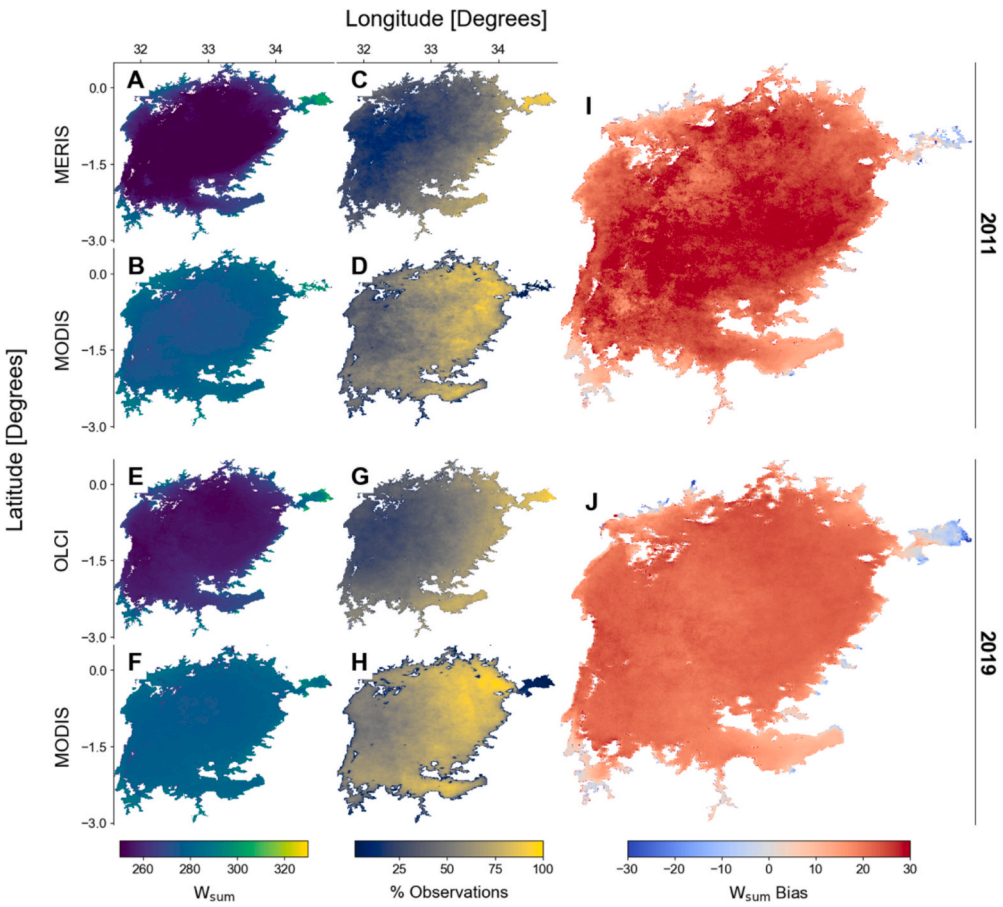

Fig. 7. Median weighted optical water types membership score sum (W_{sum}) of observations in 2011 for MERIS (A) and MODIS (B) with respective percentage of observations (C and D). Median W_{sum} of observations in 2019 for OLCI (E) and MODIS (F) with respective percentage of observations (G and H). Median W_{sum} bias between MODIS and MERIS in 2011 (I) and between MODIS and OLCI in 2019 (J).

cyanobacteria occurrence risk was predominant, MODIS captured the most severe conditions. However, when other ecoregions presented in MERIS and OLCI (e.g., ecoregions 4 and 5) MODIS assigned these to ecoregion 1.

4. Discussion and conclusion

Cyanobacteria blooms manifest through a wide range of optical conditions as cells transition between light and nutrient regimes, influencing the expression of characteristic pigmentation, colony formation, buoyancy, and sensitivity to grazing, photodegradation and other loss factors. While relationships between sensor spectral properties and monitoring capabilities are generally well established from the bio-optical literature, our analysis shows that reduced spectral information leads to a fundamental loss in the ability to determine the likelihood of cyanobacteria presence in different mixing states. Multispectral configurations, as found in current or simulated ocean colour sensors (recent hyperspectral instruments excluded), can distinguish populations blooms near or accumulated at the surface in most waveband configurations. However, there are broad differences between multispectral sensors equipped with specific channels in the yellow to red and infra-red region and those that lack one or more of these, particularly in their ability to capture conditions common to the onset of cyanobacteria population growth in the water column. Such a systematic limitation has important implications for the 2012–2016 period where MERIS and OLCI missions leave a capability gap and MODIS is the most obvious sensor to fall back on for large waterbodies, as well as for high spatial resolution sensors designed to monitor land that are used to study water quality in smaller waterbodies. Our analysis of coincident observations of MODIS, MERIS and OLCI, through the lens of Optical Water Types (OWT), demonstrates that constructing consistent long-term records of cyanobacteria dynamics requires careful consideration of how reduced spectral information fundamentally limits the detection of bloom dynamics. Our recommendation based on this research is that long-term cyanobacteria trend detection from satellite sensors relying on distinctive optical features should either work around the four-year gap, or focus only on the presence/absence of surfacing blooms. The latter is not without risk, as changes in cyanobacteria species composition and meteorological factors, particularly wind-induced mixing or lack thereof, should be taken into account to determine the cause of any trends. Biomass estimates for surfacing blooms are also subject to higher estimation uncertainty, as the depth of the visible surface layer cannot be assessed with sufficient accuracy as is the case with semi-analytical reflectance inversion algorithms.

OWTs are particularly useful to examine capability differences between multispectral sensors because they directly show how spectral shape information is affected by varying waveband configurations. While modern machine learning methods that leverage reflectance amplitude might capture different aspects of water quality (Pahlevan et al., 2020; Smith et al., 2021; Werther et al., 2022; O'Shea et al., 2023), OWTs offer insights into how specific optical features contribute to water type differentiation. This makes them suitable for identifying which aspects of bloom determination are most affected by reduced spectral information.

Our experiments consistently reveal that the absence of key wavebands fundamentally alters how satellite sensors determine the likelihood of cyanobacteria presence across different mixing states. This limitation stems from the underlying mathematical principles behind

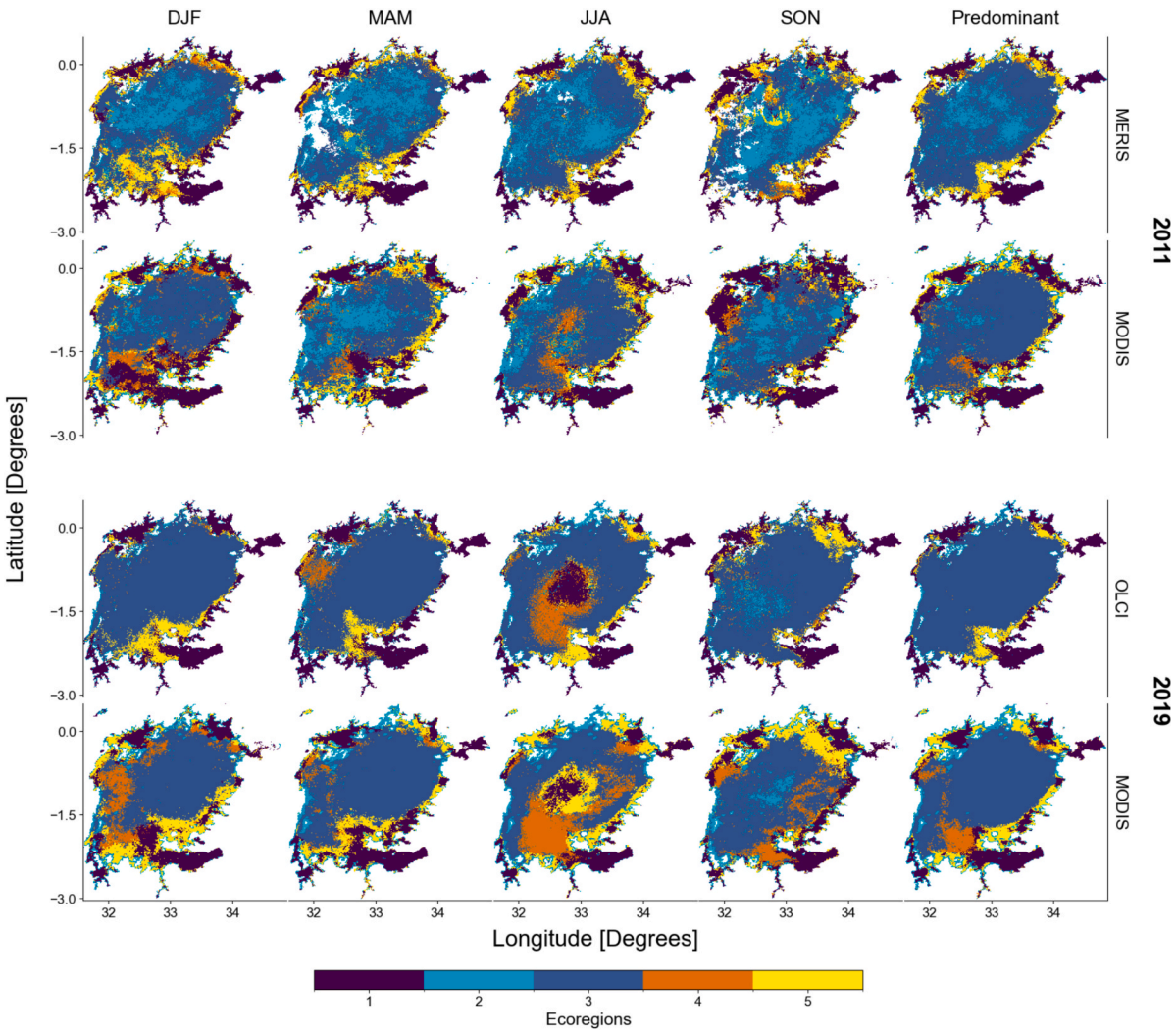

Fig. 8. Most the frequent ecoregions by season (DJF, MAM, JJA, SON), and the most frequent, or predominant, for coincident observations for MERIS and MODIS in 2011, and for OLCI and MODIS in 2019.

similarity metrics used to define OWTs. Among these, the spectral angle has been shown to provide robust similarity measurements across the wide optical diversity of inland waters, effectively capturing spectral shapes even when the covariance between adjacent wavebands is low (Liu et al., 2021). In higher-dimensional space, the spectral angle effectively captures subtle differences between OWTs from the inherent high number of degrees of freedom. With fewer wavebands, the mathematical space becomes constrained and the angular separation between different optical conditions compresses, causing otherwise distinct spectra (and possibly pertaining to different ecological processes) to appear more similar. This compression creates systematic biases that vary with both the optical complexity of water conditions and with the selection of remaining wavebands. Biases manifest non-linearly as either an over- or underestimation of the likelihood of cyanobacteria presence depending on the mixing state and dominant optical features present (Figs. 5, 6). Crucially, while optically well-defined conditions like surface accumulations or clear water conditions maintain their distinctiveness even in reduced spectral dimensions due to their large deviation from reflectance shapes of well-mixed water, the subtle spectral features that are diagnostic of cyanobacteria become increasingly difficult to differentiate without key wavebands in the yellow to infrared region. This pattern is consistent across our theoretical waveband sensitivity analysis, direct sensor comparison, and ecoregion-based evaluation, evidencing the fundamental limitation in detecting cyanobacteria bloom evolution that transcends specific algorithms or waterbodies.

The theoretical waveband sensitivity analysis (Fig. 4A) demonstrates that the loss of spectral regions expected to have high-magnitude R_w in certain optical conditions, e.g., when a high energy under the curve in the NIR region is expected in the presence of surface accumulations, or in the blue region in clear water conditions, leads to an underestimation of the likelihood of cyanobacteria presence. On the other hand, the loss of regions expected to have low-magnitude R_w, e.g., weak reflectance in the NIR region in relatively clear water conditions or in in the blue region when cyanobacteria are present, leads to overestimation. This behaviour can be effectively described using the weighted OWT membership score sum (W_{sum}) metric, which captures covarying OWT properties of cyanobacteria presence (Lomeo et al., 2025). For instance, removing the waveband centred at 709 nm in the set of 25 OLCI OWT definitions led to the largest (negative) W_{sum} bias compared to using the full range of wavebands, especially for OWTs typically associated with the highest cyanobacteria occurrence risk (Fig. 4A). This is expected given the important contribution of the 709 nm waveband to the shape of R_w and OWT spectra when phytoplankton biomass is high. As the concentration of cells, and thereby the contribution of particle backscattering and pigment absorption relative to water absorption in this region reduces, the lack of the 709 nm waveband led to the largest (positive) W_{sum} bias compared to using the full range of wavebands in OWTs typically associated with the lowest cyanobacteria occurrence

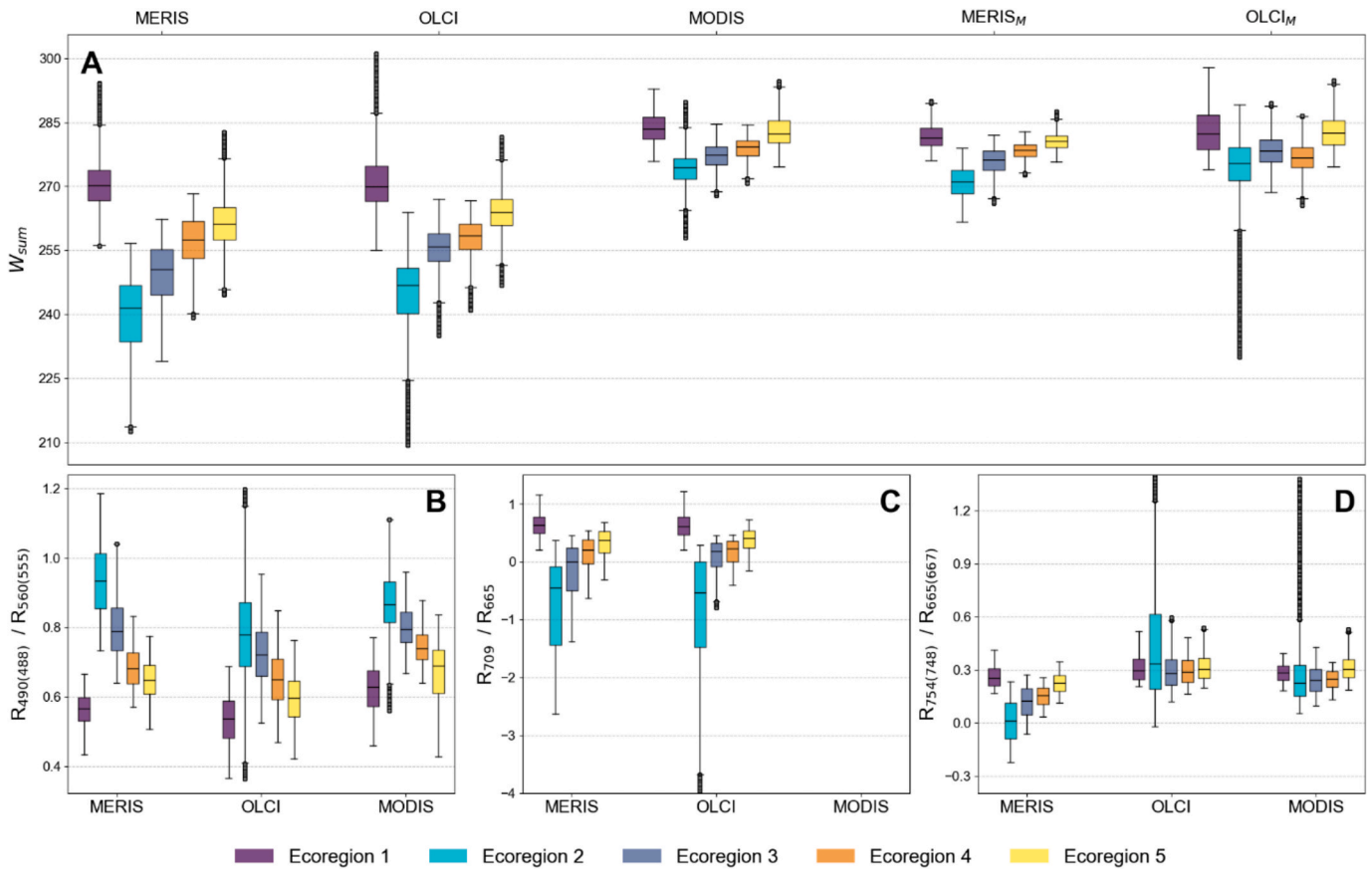

Fig. 9. (A) Comparison of weighted optical water type membership score sum (W_{sum}) for MERIS (2011), MODIS (2011 and 2019 combined), OLCI (2019), $MERIS_M$ (2011), and $OLCI_M$ (2019), by ecoregions. $MERIS_M$ and $OLCI_M$ refer to MERIS and OLCI sensors using the same waveband configuration as MODIS. (B) $R_{490\ (488)}$ / $R_{560\ (555)}$, (C) R_{709} / R_{665} (not available for MODIS), (D) $R_{754\ (748)}$ / $R_{665\ (667)}$ waveband ratios for MERIS, OLCI, and MODIS by ecoregions. The wavelengths in brackets refer to MODIS wavebands. The numbers on the vertical axes in bracket are the MODIS wavebands.

risk (Fig. 4A). The absence of a waveband centred at 620 nm affects the definition of all OWTs. In some cases, secondary absorption features of algal accessory pigments like chlorophylls *b* and *c* likely lead to overestimating cyanobacteria presence (Simis et al., 2007). However, the 620 nm waveband remains crucial for distinguishing cyanobacteria from other phytoplankton groups across varying optical conditions (Kutser et al., 2006). The combined critical importance of the 620 and 709 nm wavebands to detect cyanobacteria presence is well established (Dekker, 1993; Simis et al., 2005; Matthews et al., 2012), and our analysis shows that sensors lacking either of these wavebands will particularly underestimate OWTs associated to medium ($w = 100$), and high ($w = 1000$), cyanobacteria occurrence risk. W_{sum} bias resulting from removing the 412 nm waveband from OWT definitions, especially those associated with the highest cyanobacteria occurrence risk, corresponds to efficient light absorption by pigments in cyanobacteria-dominated waters (Ruiz-Verdú et al., 2008). Thus, in the presence of cyanobacteria, when higher absorption in this spectral region is expected, the absence of the 412 nm waveband reduces the ability of sensors to distinguish between different cyanobacteria growth phases through varying mixing conditions, leading to an overestimation of W_{sum}. Conversely, at the lowest cyanobacteria cooccurrence risk (i.e., clearer water conditions), the bias becomes negative. Similarly, positive W_{sum} bias linked to the removal of 754 and 779 nm wavebands confirms the importance of capturing the relative contribution of absorption by water to beam attenuation which is primarily modulated by particle scattering. When cells accumulate at the surface, the shape of the spectrum in the NIR becomes highly distinguishable from other mixing conditions, much like the case of floating vegetation. This reduces the individual influence of these wavebands in the spectral angle calculation, resulting in smaller W_{sum} bias. Conversely, in OWT definitions not associated to material accumulating at the water surface, the presence of either of these wavebands in the spectral angle calculation becomes fundamental to distinguish between subtle variations of mixing states. The contribution of the wavebands described above only diverge for OWTs with distinct optical characteristics, whereby the spectral angle metric become sensitive to changes in other parts of the spectrum compared to (most) OWTs. These include OWT 12.2 (surface scum), which present a unique M−like shape, showing reflectance peaks at bands 560 and 754 nm, and throughs at 412, 681 and 885 nm (Lomeo et al., 2025), and OWT 10 and 13, representing colour dissolved organic matter rich waters primarily found at riverine estuaries and clear water conditions, respectively (Spyrakos et al., 2018).

Sensors with waveband configurations that are similar to OLCI, like GOCI-II and MERIS, detect cyanobacteria population dynamics with minimal bias relative to OLCI (Fig. 4B). Waveband configurations such as those of MODIS, VIIRS, and GOCI show departures from OLCI, particularly in OWTs associated with high cyanobacteria occurrence risk and clear waters. Land-focused sensors like the Landsat series show substantial biases across most OWTs, with marginally better performance in OWTs associated to surface accumulations, which present the strongest optical contrasts with most other water types. Even MSI, despite a more appropriate waveband configuration for studying water colour than Landsat sensors, struggles to resolve varying cyanobacteria presence, particularly for OWTs associated to medium cyanobacteria occurrence risk, likely to include the initial (or final) stages of cyanobacterial bloom (Lomeo et al., 2025). Superdove, while possessing wavebands expected (in theory) to help resolve cyanobacteria, such as a yellow band overlapping with phycocyanin absorption, shows

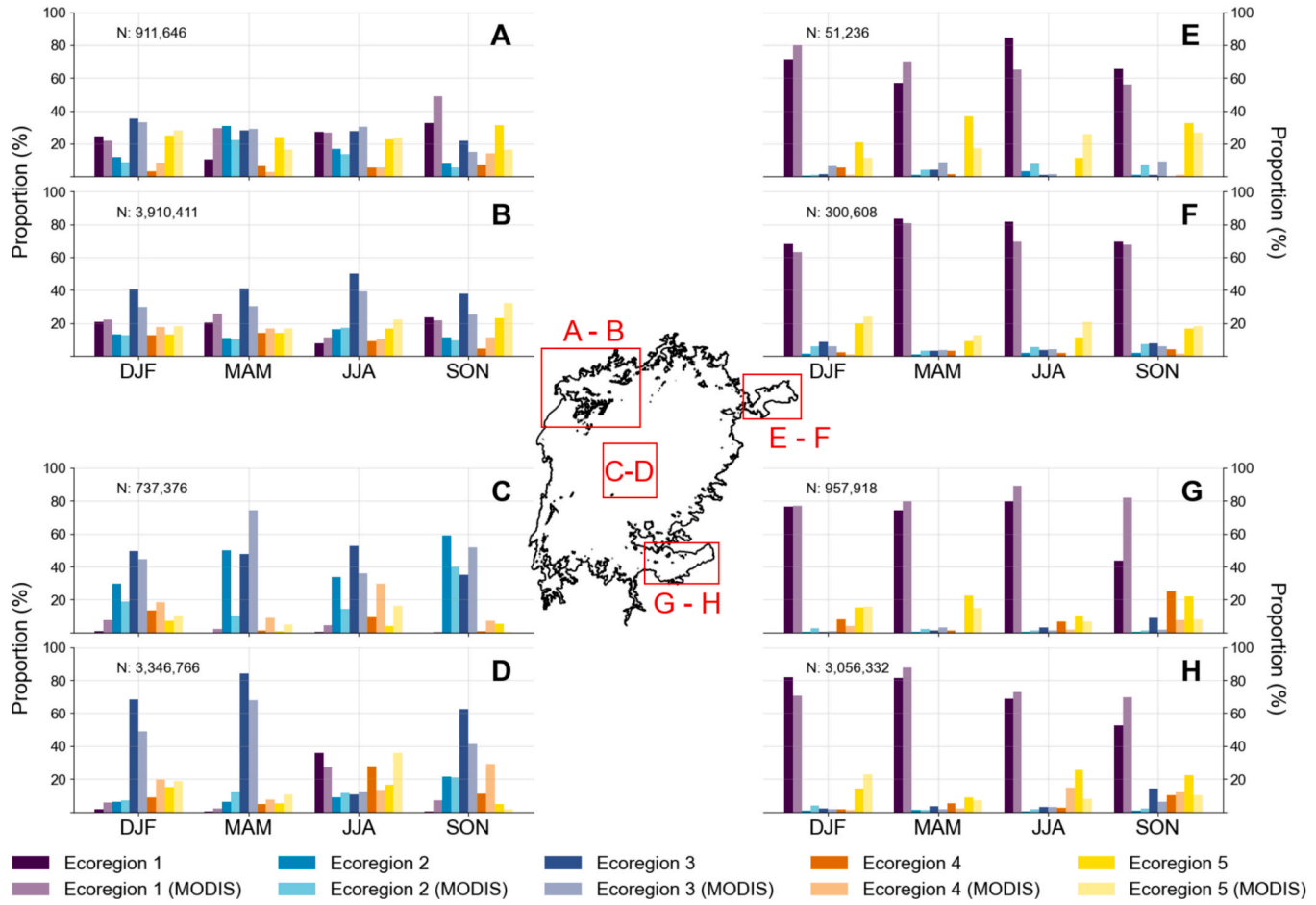

Fig. 10. Weighted membership score sum (W_{sum}) relative proportion by ecoregion by season for coincident observations for MERIS and MODIS in 2011 (A, C, E, G) and OLCI and MODIS in 2019 (B, D, F, H) in four different regions of the lake. The location of these regions is shown in the central plot of the boundaries of Lake Victoria: the Ssese islands and Murchison Bay (A – B), a central portion of the lake (C – D), Winam Gulf (E – F), and Speke Gulf (G – H).

substantial biases, likely due to the lack of wavebands in the blue and in the NIR spectral regions to help distinguish optical water types. In all these results, sensor capabilities pertaining to accurate atmospheric correction are not taken into account, and we assume that R_w is reproduced to the same accuracy as the data underlying the definition of the OWT library. In practice, atmospheric correction for optically complex water greatly benefits from wavebands spanning the light spectrum and additional wavebands in the near infrared and shortwave infrared.

Comparison of coincident observations of MODIS, MERIS, and OLCI, confirm that the reduced spectral resolution and waveband configuration in MODIS lead to artificially assigning higher OWT membership score (S_{owt}) values compared to MERIS and OLCI, collapsing the continuous gradient of bloom evolution into fewer types. This occurs despite key spectral regions for ocean colour monitoring, such as the wavebands centred at 560 (555), 665 (667), and 681 (678) nm, show a fair stability between the three sensors and their atmospheric correction (Fig. 6; Table 1), as also found in other studies (Liu et al., 2024; Pahlevan et al., 2024). The greater disagreement in other shared wavebands such as those centred at 412 and 754 (748) nm likely reflects challenges for atmospheric correction, particularly given the reduced spectral information and different radiometric characteristics in MODIS (Moses et al., 2009; IOCCG, 2010; Simis et al., 2022b). However, our results substantiate that the non-linear responses of MODIS to optical features associated to cyanobacteria bloom development make attempts to harmonise with other sensors non-trivial, and likely not always ecologically appropriate.

MODIS-derived S_{owt} values are systematically higher across the OWT set compared to both MERIS and OLCI, especially for OWTs associated with high cyanobacteria occurrence risk (Fig. 6). Spectra that represent distinct cyanobacteria mixing states appear mathematically more similar with MODIS. When the difference to a given reference OWT spectrum is large (i.e., low S_{owt} values), the loss of spectral detail artificially reduces the measured angles, leading to overestimations in MODIS. Conversely, when spectra are similar (i.e., high S_{owt} values), the reduced spectral configuration of MODIS slightly increases the measured angle, producing lower S_{owt} values (Table 2). When restricting the comparison to matching wavebands, MODIS shows consistently lower S_{owt} values compared to $MERIS_M$ and $OLCI_M$ (Table 3), reflecting its tendency to yield larger spectral angles. This difference likely stems from a combination of factors including its broader spectral response functions and different waveband positioning (Fig. 4), differences in radiometric characteristics, and later overpass time of the Aqua sensor compared to MERIS and OLCI, likely leading to less consistent comparisons.

The systematic compression into fewer water types by MODIS clearly results in limited monitoring capabilities across the diverse optical conditions observed in Lake Victoria. While MODIS generally overestimates W_{sum} across the lake, it systematically underestimates W_{sum} in dendritic regions and semi-enclosed basins (Fig. 7I, J), where it also consistently shows lower observation counts (Fig. 7D, I). This is particularly evident in Winam Gulf (Fig. 10E-F), where MODIS provides 91 % to 96 % fewer valid estimates compared to MERIS and OLCI. These are also areas known to have a recurrent issue of persistent cyanobacteria prevalence (Sitoki et al., 2012; Mbonde et al., 2015; Simiyu et al.,

2018; Olokotum et al., 2020; Brown et al., 2024). The coarser native spatial resolution of MODIS leads to larger masked areas during atmospheric correction near shorelines, while its radiometric characteristics cause well-understood saturation over bright targets like surface blooms and high suspended sediment concentrations (Wynne et al., 2013). The coincidence of these biases with areas of reduced observations creates systematic gaps in bloom monitoring precisely where detailed optical information would be most valuable, underscoring how the fundamental limitations of MODIS compound in regions where cyanobacteria blooms frequently develop and impact water quality.

The ecoregion analysis demonstrates that while MODIS captures broad spatial patterns of water colour, particularly in areas of high cyanobacteria occurrence risk, it fails to resolve the subtle transitions that characterise cyanobacteria bloom development through mixing states. Self-organising Maps (SOM) captured typical optical characteristics leveraging S_{owt} values within shared seasonal cyclin, proving an effective method to evaluate monitoring capabilities irrespective of waveband configurations and overpass times (Fig. 8), and this is evidenced by the distribution of W_{sum} across the identified ecoregions. While MERIS and OLCI show comparable and consistent W_{sum} ranges, MODIS yields consistently higher values with notably compressed ranges (Fig. 9A). This pattern persists even when MERIS and OLCI are constrained to the waveband configuration of MODIS, though with lower absolute values, directly suggesting that the higher average W_{sum} and narrower ranges are linked to the reduced spectral space rather than individual sensor characteristics. The spectral basis for these differences manifests clearly in commonly used waveband ratios across different optical conditions. The 443/ 560 nm waveband ratios provide consistent distinction between ecoregions across the three sensors (Fig. 9B), reflecting the relative radiometric stability previously described (Fig. 5). Additionally, while MERIS and OLCI clearly distinguished between ecoregions using the 709/ 665 nm waveband ratio (Fig. 9C), this behaviour was not replicated when using the 754/ 665 nm waveband ratio (Fig. 9D), highlighting the lower sensitivity of wavebands at longer wavelengths to changes in the vertical distribution of cells, unless accumulated at the surface (Kutser et al., 2006). This means that the loss of the waveband at 709 nm in MODIS fundamentally limits its ability to capture the expected optical transitions between mixing states, which is ultimately why W_{sum} magnitudes and ranges diverge from MERIS and OLCI.

Regional variations in optical complexity across Lake Victoria expose the practical consequences of the reduced spectral coverage of MODIS. In the area comprising the Ssese Islands and Murchison Bay in the northwest, MODIS largely aligned with both MERIS and OLCI observations, showing only minor disagreements in discriminating between optical conditions (Fig. 10A-B). This region exhibited less complex optical gradients and fewer transitional mixing states during the study period (at least in the years under examination). The limitations of MODIS become pronounced in open waters, where it consistently detected higher cyanobacteria occurrence risk in clear water conditions. Even when capturing the presence of optical characteristics associated to blooms of varying severity, as observed during JJA 2019, MODIS failed to distinguish subtle transitions that OLCI resolved more distinctly (Fig. 10C-D). These constraints become most evident in regions like Winam and Speke Gulfs. Although MODIS successfully identified optical conditions associate with higher cyanobacteria occurrence risk, it systematically compressed the optical variability into the ecoregion associated with the highest (average) risk of cyanobacteria occurrence (Fig. 10E-H). Where MERIS and OLCI detected subtle gradients between different ecoregions, indicating evolving mixing states, MODIS defaulted to classifying these to the ecoregion associated to the highest risk of cyanobacteria occurrence, loosing granular information that may be crucial for proactive water quality management.

In conclusion, while OWT frameworks provide practical solutions to translate optical complexity in water bodies into realistic estimates of biogeochemical properties, through dynamic algorithm selection across optical gradients (Moore et al., 2001; Liu et al., 2021), our analysis reveals important constraints imposed by sensor-specific spectral configurations. These constraints significantly impact how we interpret satellite records, particularly during the 2012–2016 when ocean colour mission capabilities were limited to sensors like MODIS, as well as for observations from high-spatial resolution sensors commonly used to study smaller waterbodies. While the study focused on Lake Victoria, the findings are expected to find global relevance and applicability, especially in light of previous testing of the cyanobacteria occurrence index in other lakes (Lomeo et al., 2025). Emerging machine learning approaches have demonstrated remarkable capabilities in retrieving water quality parameters even from sensors with limited spectral coverage (Pahlevan et al., 2020; Smith et al., 2021), which likely leverage covariation across the reflectance spectrum that may not be explicitly captured in discrete OWT classifications. The systematic compression of optical gradients demonstrated across our experiments indicates that apparent trends in MODIS observations may reflect changes in detection capability rather than true ecological dynamics. These findings underscore the importance of spectral capability in mission design, while highlighting that complementary approaches incorporating both spectral shape (as emphasised in OWTs) and amplitude information may provide more robust characterisation of cyanobacteria presence across varying mixing states. As emerging ocean colour missions like the Plankton, Aerosol, Cloud ocean Ecosystem (PACE) offer enhanced spectral capabilities that are expected to improve cyanobacteria monitoring globally, understanding the specific spectral limitations we have identified provides essential context for analysing decades of observations that underpin our understanding of inland water responses to environmental change. This context becomes particularly critical as we increasingly rely on historical records to understand climate-driven changes in phytoplankton communities and develop evidence-based water quality management strategies for these vital ecosystems.

CRediT authorship contribution statement

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

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.

Data availability

The datasets used in this work will be made available upon request.

References

Ardyna, M., Claustre, H., Sallée, J., D'Ovidio, F., Gentili, B., Van Dijken, G., D'Ortenzio, F., Arrigo, K.R., 2017. Delineating environmental control of phytoplankton biomass and phenology in the Southern Ocean. Geophys. Res. Lett. 44 (10), 5016–5024. https://doi.org/10.1002/2016GL072428.

Astel, A., Tsakovski, S., Barbieri, P., Simeonov, V., 2007. Comparison of self-organizing maps classification approach with cluster and principal components analysis for large environmental data sets. Water Res. 41 (19), 4566–4578. https://doi.org/10.1016/j.watres.2007.06.030.

Balasubramanian, S.V., O'Shea, R.E., Saranathan, A.M., Begeman, C.C., Gurlin, D., Binding, C., Giardino, C., Tomlinson, M.C., Alikas, K., Kangro, K., Lehmann, M.K., Reed, L., 2025. Mixture density networks for re-constructing historical ocean-color products over inland and coastal waters: demonstration and validation. Front. Remote Sens. 6, 1488565. https://doi.org/10.3389/frsen.2025.1488565.

Bootsma, H.A., Hecky, R.E., 2003. A Comparative Introduction to the Biology and Limnology of the African Great Lakes. J. Great Lakes Res. 29, 3–18. https://doi.org/10.1016/S0380-1330(03)70535-8.

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.

Chazottes, A., Crépon, M., Bricaud, A., Ras, J., Thiria, S., 2007. Statistical analysis of absorption spectra of phytoplankton and of pigment concentrations observed during three POMME cruises using a neural network clustering method. Appl. Opt. 46 (18), 3790. https://doi.org/10.1364/AO.46.003790.

Chorus, I., Welker, M., Chorus, I., Welker, M., 2021. Toxic Cyanobacteria in Water; A Guide to Their Public Health Consequences, Monitoring and Management; Second Edition.

Dall'Olmo, G., Gitelson, A.A., Rundquist, D.C., 2003. Towards a unified approach for remote estimation of chlorophyll-a in both terrestrial vegetation and turbid productive waters. Geophys. Res. Lett. 30 (18), 2003GL018065. https://doi.org/10.1029/2003GL018065.

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

Diouf, D., Niang, A., Brajard, J., Crepon, M., Thiria, S., 2013. Retrieving aerosol characteristics and sea-surface chlorophyll from satellite ocean color multi-spectral sensors using a neural-variational method. Remote Sens. Environ. 130, 74–86. https://doi.org/10.1016/j.rse.2012.11.002.

D'Ortenzio, F., 2009. On the trophic regimes of the Mediterranean Sea: a satellite analysis.

D'Ortenzio, F., Antoine, D., Martinez, E., Ribera d'Alcalà, M., 2012. Phenological changes of oceanic phytoplankton in the 1980s and 2000s as revealed by remotely sensed ocean-color observations. Global Biogeochem. Cycles 26 (4), 2011GB004269. https://doi.org/10.1029/2011GB004269.

El Hourany, R., Abboud-Abi Saab, M., Faour, G., Mejia, C., Crépon, M., Thiria, S., 2019. Phytoplankton diversity in the mediterranean sea from satellite data using self-organizing maps. J. Geophys. Res. Oceans 124 (8), 5827–5843. https://doi.org/10.1029/2019JC015131.

Fendereski, F., Vogt, M., Payne, M.R., Lachkar, Z., Gruber, N., Salmanmahiny, A., Hosseini, S.A., 2014. Biogeographic classification of the Caspian Sea. Biogeosciences 11 (22), 6451–6470. https://doi.org/10.5194/bg-11-6451-2014.

Frank, T.H., Cornelissen, I.J.M., Vijverberg, J., Nagelkerke, L.A.J., 2023. Spatial and seasonal variation in the phytoplankton community of Lake Victoria's Mwanza Gulf, compared to northern parts of the lake. J. Great Lakes Res. 49 (2), 453–462. https://doi.org/10.1016/j.jglr.2023.02.002.

Franz, B.A., 2008. Moderate Resolution Imaging Spectroradiometer on Terra: limitations for ocean color applications. J. Appl. Remote Sens. 2 (1), 023525. https://doi.org/10.1117/1.2957964.

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.

Gower, J., King, S., Borstad, G., Brown, L., 2005. Detection of intense plankton blooms using the 709 nm band of the MERIS imaging spectrometer. Int. J. Remote Sens. 26 (9), 2005–2012. https://doi.org/10.1080/01431160500075857.

Huisman, J., Matthijs, H.C.P., Visser, P.M., Huisman, J., Matthijs, H.C.P., Visser, P.M., 2005. Harmful cyanobacteria. Aquatic ecology series v. 3. Springer, Dordrecht ; Norwell, MA.

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

IOCCG, 2012. Mission Requirements for Future Ocean-Color Sensors. Reports of the International Ocean-Colour Coordinating Group, Dartmouth, Canada.

IOCCG (2009). Partition of the Ocean into Ecological Provinces: Role of Ocean-Colour Radiometry. Reports of the International Ocean-Colour Coordinating Group 9. Dartmouth, Canada.

Jain, A.K., Dubes, R.C., 1988. Algorithms for clustering data. Prentice Hall advanced reference series, Prentice-Hall, Englewood Cliffs, NJ.

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. Int. J. Appl. Earth Obs. Geoinf. 117, 103188. https://doi.org/10.1016/j.jag.2023.103188.

Johnson, T.C., Kelts, K., Odada, E., 2000. The holocene history of Lake Victoria. AMBIO J. Hum. Environ. 29 (1), 2–11. https://doi.org/10.1579/0044-7447-29.1.2.

Jupp, D., Kirk, J., Harris, G., 1994. Detection, identification and mapping of cyanobacteria — using remote sensing to measure the optical quality of turbid inland waters. Mar. Freshw. Res. 45 (5), 801. https://doi.org/10.1071/MF9940801.

Kent, J.T., Mardia, K.V., 1988. Spatial classification using fuzzy membership models. IEEE Trans. Pattern Anal. Mach. Intell. 10 (5), 659–671. https://doi.org/10.1109/34.6774.

Kohonen, T., 1982. Self-organized formation of topologically correct feature maps. Biol. Cybern. 43 (1), 59–69. https://doi.org/10.1007/BF00337288.

Kohonen, T., 2001. Self-organizing maps. Springer series in information sciences 30, 3rd ed. Springer, Berlin.

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., 2009. Passive optical remote sensing of cyanobacteria and other intense phytoplankton blooms in coastal and inland waters. Int. J. Remote Sens. 30 (17), 4401–4425. https://doi.org/10.1080/01431160802562305.

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

Liu, S., Glamore, W., Tamburic, B., Morrow, A., Johnson, F., 2022. Remote sensing to detect harmful algal blooms in inland waterbodies. Sci. Total Environ. 851, 158096. https://doi.org/10.1016/j.scitotenv.2022.158096.

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 Sens. Environ. 267, 112710. https://doi.org/10.1016/j.rse.2021.112710.

Liu, X., Warren, M., Selmes, N., Simis, S.G.H., 2024. Quantifying decadal stability of lake reflectance and chlorophyll-a from medium-resolution ocean color sensors. Remote Sens. Environ. 306, 114120. https://doi.org/10.1016/j.rse.2024.114120.

Liu, Y., Weisberg, H.R., 2011. A Review of Self-Organizing Map Applications in Meteorology and Oceanography. In: J.I. Mwasiagi (ed.). Self Organizing Maps - Applications and Novel Algorithm Design. InTech. p. doi:10.5772/13146.

Liu, Y., Weisberg, R.H., Mooers, C.N.K., 2006. Performance evaluation of the self-organizing map for feature extraction. J. Geophys. Res. Oceans 111 (C5), 2005JC003117. https://doi.org/10.1029/2005JC003117.

Lomeo, D., Simis, S.G.H., Liu, X., Selmes, N., Warren, M.A., Jungblut, A.D., Tebbs, E.J., 2025. A novel cyanobacteria occurrence index derived from optical water types in a tropical lake. ISPRS J. Photogramm. Remote Sens. 223, 58–77. https://doi.org/10.1016/j.isprsjprs.2025.03.006.

Longhurst, A., 1995. Seasonal cycles of pelagic production and consumption. Prog. Oceanogr. 36 (2), 77–167. https://doi.org/10.1016/0079-6611(95)00015-1.

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 Sens. Environ. 124, 637–652. https://doi.org/10.1016/j.rse.2012.05.032.

Mbonde, A.S., Sitoki, L., Kurmayer, R., 2015. Phytoplankton composition and microcystin concentrations in open and closed bays of Lake Victoria Tanzania. Aquat. Ecosyst. Health Manag. 18 (2), 212–220. https://doi.org/10.1080/14634988.2015.1011030.

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

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 Sens. Environ. 143, 97–111. https://doi.org/10.1016/j.rse.2013.11.021.

Morel, A., Prieur, L., 1977. Analysis of variations in ocean color1. Limnol. Oceanogr. 22 (4), 709–722. https://doi.org/10.4319/lo.1977.22.4.0709.

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. Environ. Res. Lett. 4 (4), 045005. https://doi.org/10.1088/1748-9326/4/4/045005.

Nakkazi, M.T., Nkwasa, A., Martínez, A.B., Van Griensven, A., 2024. Linking land use and precipitation changes to water quality changes in Lake Victoria using earth observation data. Environ. Monit. Assess. 196 (11), 1104. https://doi.org/10.1007/s10661-024-13261-2.

Niang, A., 2003. Automatic neural classification of ocean colour reflectance spectra at the top of the atmosphere with introduction of expert knowledge. Remote Sens. Environ. 86 (2), 257–271. https://doi.org/10.1016/S0034-4257(03)00113-5.

Olokotum, M., Mitroi, V., Troussellier, M., Semyalo, R., Bernard, C., Montuelle, B., Okello, W., Quiblier, C., Humbert, J.-F., 2020. A review of the socioecological causes and consequences of cyanobacterial blooms in Lake Victoria. Harmful Algae 96, 101829. https://doi.org/10.1016/j.hal.2020.101829.

O'Shea, R.E., Pahlevan, N., Smith, B., Boss, E., Gurlin, D., Alikas, K., Kangro, K., Kudela, R.M., Vaičiūtė, D., 2023. A hyperspectral inversion framework for estimating absorbing inherent optical properties and biogeochemical parameters in inland and coastal waters. Remote Sens. Environ. 295, 113706. https://doi.org/10.1016/j.rse.2023.113706.

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 Sens. Environ. 266, 112693. https://doi.org/10.1016/j.rse.2021.112693.

Paerl, H.W., Otten, T.G., 2013. Harmful cyanobacterial blooms: causes, consequences, and controls. Microb. Ecol. 65 (4), 995–1010. https://doi.org/10.1007/s00248-012-0159-y.

Pahlevan, N., Balasubramanian, S., Begeman, C.C., O'Shea, R.E., Ashapure, A., Maciel, D. A., Hall, D.K., Odermatt, D., Giardino, C., 2024. A retrospective analysis of remote-sensing reflectance products in coastal and inland waters. IEEE Geosci. Remote Sens. Lett. 21, 1–5. https://doi.org/10.1109/LGRS.2024.3351328.

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 Sens. Environ. 240, 111604. https://doi.org/10.1016/j.rse.2019.111604.

Platt, T., Sathyendranath, S., 1988. Oceanic primary production: estimation by remote sensing at local and regional scales. Science 241 (4873), 1613–1620. https://doi.org/10.1126/science.241.4873.1613.

Reusch, D.B., Alley, R.B., Hewitson, B.C., 2005. Relative performance of self-organizing maps and principal component analysis in pattern extraction from synthetic climatological data. Polar Geogr. 29 (3), 188–212. https://doi.org/10.1080/789610199.

Richardson, A.J., 2002. Identifying characteristic chlorophyll a profiles in the coastal domain using an artificial neural network. J. Plankton Res. 24 (12), 1289–1303. https://doi.org/10.1093/plankt/24.12.1289.

Richardson, A.J., Risien, C., Shillington, F.A., 2003. Using self-organizing maps to identify patterns in satellite imagery. Prog. Oceanogr. 59 (2–3), 223–239. https://doi.org/10.1016/j.pocean.2003.07.006.

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 Sens. Environ. 112 (11), 3996–4008. https://doi.org/10.1016/j.rse.2007.11.019.

Shi, K., Zhang, Y., Zhou, Y., Liu, X., Zhu, G., Qin, B., Gao, G., 2017. Long-term MODIS observations of cyanobacterial dynamics in Lake Taihu: responses to nutrient enrichment and meteorological factors. Sci. Rep. 7 (1), 40326. https://doi.org/10.1038/srep40326.

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., Liu, X., Calmettes, B., Yésou, H., Merchant, C., Carrea, L., Duguay, C., Wu, Y., 2022. D4.1: Product Validation and Intercomparison Report. (2.1).

Simis, S.G.H., Peters, S.W.M., Gons, H.J., 2005. Remote sensing of the cyanobacterial pigment phycocyanin in turbid inland water. Limnol. Oceanogr. 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 Sens. Environ. 106 (4), 414–427. https://doi.org/10.1016/j.rse.2006.09.008.

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. Int. Rev. Hydrobiol. 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. Front. Remote Sens. 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., 2018. Optical types of inland and coastal waters: Optical types of inland and coastal waters. Limnol. Oceanogr. 63 (2), 846–870. https://doi.org/10.1002/lno.10674.

Thiery, W., Davin, E.L., Panitz, H.-J., Demuzere, M., Lhermitte, S., Van Lipzig, N., 2015. The Impact of the African Great Lakes on the regional climate. J. Clim. 28 (10), 4061–4085. https://doi.org/10.1175/JCLI-D-14-00565.1.

Vanderkelen, I., Van Lipzig, N.P.M., Thiery, W., 2018. Modelling the water balance of Lake Victoria (East Africa) – part 1: observational analysis. Hydrol. Earth Syst. Sci. 22 (10), 5509–5525. https://doi.org/10.5194/hess-22-5509-2018.

Vesanto, J., Alhoniemi, E., 2000. Clustering of the self-organizing map. IEEE Trans. Neural Netw. 11 (3), 586–600. https://doi.org/10.1109/72.846731.

Vettigli, G., 2018. MiniSom: minimalistic and NumPy-based implementation of the Self Organizing Map. https://github.com/JustGlowing/minisom.

Villmann, T., Merényi, E., Hammer, B., 2003. Neural maps in remote sensing image analysis. Neural Netw. 16 (3–4), 389–403. https://doi.org/10.1016/S0893-6080(03)00021-2.

Wang, S., Li, J., Zhang, B., Lee, Z., Spyrakos, E., Feng, L., Liu, C., Zhao, H., Wu, Y., Zhu, L., Jia, L., Wan, W., Zhang, F., Shen, Q., Tyler, A.N., Zhang, X., 2020. Changes of water clarity in large lakes and reservoirs across China observed from long-term MODIS. Remote Sens. Environ. 247, 111949. https://doi.org/10.1016/j.rse.2020.111949.

Wang, S., Shen, M., Ma, Y., Chen, G., You, Y., Liu, W., 2019. Application of remote sensing to identify and monitor seasonal and interannual changes of water turbidity in Yellow River Estuary China. J. Geophy. Res.: Oceans. 124 (7), 4904–4917. https://doi.org/10.1029/2019JC015106.

Werther, M., Odermatt, D., Simis, S.G.H., Gurlin, D., Lehmann, M.K., Kutser, T., Gupana, R., Varley, A., Hunter, P.D., Tyler, A.N., Spyrakos, E., 2022. A Bayesian approach for remote sensing of chlorophyll-a and associated retrieval uncertainty in oligotrophic and mesotrophic lakes. Remote Sens. Environ. 283, 113295. https://doi.org/10.1016/j.rse.2022.113295.

Wynne, T.T., Stumpf, R.P., Briggs, T.O., 2013. Comparing MODIS and MERIS spectral shapes for cyanobacterial bloom detection. Int. J. Remote Sens. 34 (19), 6668–6678. https://doi.org/10.1080/01431161.2013.804228.

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. Int. J. Remote Sens. 29 (12), 3665–3672. https://doi.org/10.1080/01431160802007640.

Yacoub, M., Badran, F., Thiria, S., 2001. A Topological Hierarchical Clustering: Application to Ocean Color Classification. In: G. Dorffner, H. Bischof, & K. Hornik (eds.). *Artificial Neural Networks — ICANN 2001*. Lecture Notes in Computer Science. Springer Berlin Heidelberg, Berlin, Heidelberg. pp. 492–499. doi:10.1007/3-540-44668-0_69.

Yang, W., Seager, R., Cane, M.A., Lyon, B., 2015. The annual cycle of east african precipitation. J. Clim. 28 (6), 2385–2404. https://doi.org/10.1175/JCLI-D-14-00484.1.