# 3. Materials and methods

## 3.1 Study area

Lake Victoria, with a surface area of about 68,800 KM2 and an average depth of 40m at a maximum depth of 79m ranks the second largest fresh water lake in the world after Lake Superior and the Largest in Africa. Lying between 3o S to 0o 30`N latitude and 31o 40`E to 34o 50`E is distributed among these three East African countries viz Tanzania 51%, Uganda 43% and Kenya the remaining 6% (africangreatlakes.org).

That in place, the lake is privileged to serve as economical home of about 40 million residents (Dorothy et. Al 2020) in those riparian reserves. These millions of individuals solely bank on the lake for all aspects of their daily economic livelihood ranging from, fishing, agriculture, and industrial applications just to barely highlight but a few. That in place, it’s ecological monitoring should be of great geoscientific interest.

Being located in Equatorial regions of the globe, the lake has an alternating climatic condition varying from tropical rain forest with rainfall over the lake for a better portion of the year to a semi dry climate with sporadically discontinuous droughts over some locations.

This provides ambient temperatures varying between 12-26°C it therefore provided optimum host conditions for the growth and development of the Cyanobacteria in this scope.

Figure 2 shows the location and extent of the study area as discussed above.

## 3.2 Data

*Table 1: Data Sources and their roles*

|  |  |  |
| --- | --- | --- |
| Data Type | Source | Role/Use |
| Landsat 8 OLI  (30m, 16 days) | Google Earth Engine  (2015-2021) | Spatiotemporal HAB Monitoring |
| Landsat 8 TIR  (100m, 16 days) | Google Earth Engine  (2015-2021) | Lake Surface Water Temperature Monitoring (LSWT) |
| Meteorological Data | Kenya Marine & Fisheries Research Institute-KMFRI (2015-2021) | Water Quality assessment |
| Shapefiles | Geodatabase of Global Administrative areas- GADM | Delineate the Study area |
| In-Situ Data | In-situ Sensors 2021 Onwards | Continued In-Situ Algal Monitoring |

Table 1: Tools and Materials used in the study

|  |  |  |
| --- | --- | --- |
| **Tool/Material** | **Role** | **Availability** |
| **Google Earth Engine (GEE)** | **Geocomputation & Processing** | **Freely Available** |
| **QGIS, R & Python** | **Further Analysis & Maps** | **Free** |
| **Microcontroller & Sensors** | **In-Situ data Monitoring** | **Local Purchase** |
| **KiCAD** | **Design the Schematics & basic Circuits** | **Free & Open source** |

## 3.3 Methodology

chlorophyll-a concentration from the OC3 algorithm (chl-a, mg/m3) after atmospheric correction with ACOLITE for the same scene.

***1. A review of ocean color remote sensing methods and statistical techniques for the detection, mapping and analysis of phytoplankton blooms in coastal and open oceans***

**3. Ocean color remote sensing algorithms**

The need for accurate retrievals of Chl-a concentrations in open and coastal ocean waters from ocean color data has driven most of the research in algorithm development over the past thirty years. Other algorithms have also been developed, such as those that use specific spectral features of the reflectance spectrum to detect phytoplankton with surface expressions. There is a wide variety of operational ocean color satellite sensors and algorithms to assist in the detection and monitoring of phytoplankton blooms, and this section explores the various forms currently available, which specifically includes reflectance-based classification algorithms, spectral band-ratios, spectral band-difference algorithms and bio-optical models. The limitations and advantages associated with their application in the detection and mapping of algal blooms are discussed.

**3.1. Reflectance classification algorithms**

It has long been recognized that information about optically active constituents present within a parcel of water can be obtained from its spectral reflectance spectrum (e.g., Steemann Nielsen, 1963, 1937; Steemann Nielsen and Jensen, 1957). Spectral bands located in the blue, green, yellow, red or near-infrared (NIR) portion of the reflectance spectrum can be used in many ways to detect algal blooms Algorithms that rely mostly on the detection of specific spectral features are often well suited for algal blooms with surface expressions. This approach can be sufficient for the discrimination of algal blooms from other naturally occurring phenomena (e.g., Siegel et al., 2007), but the sole use of the reflectance spectrum can often only provide qualitative estimates. The reliability of the measured reflectance is hampered by its sensitivity to, e.g., the thickness of the floating algal layer, suspended particulates, bottom reflectance (although novel correction techniques now exist (e.g., Barnes et al., 2013)) and atmospheric correction errors. This reliability is even more questionable when dealing with coastal waters, *where other optically active substances affect the water-leaving radiance.* To ensure the validity of the algal bloom information retrieved from reflectance classification algorithms, it is recommended that knowledge of the study region be taken into account. A detailed analysis of the reflectance spectra of the flagged pixels is also required.

**3.2. Reflectance band-ratio algorithms**

In open ocean Case 1 waters, phytoplankton is the primary water constituent (Morel, 1980; Morel and Prieur, 1977); thus, Chl-a concentrations can be empirically related to the water-leaving reflectance using relationships of various forms (e.g., Matthews, 2011; Dierssen, 2010). These empirical relationships are often derived using large, sometimes global (e.g., Fargion and McClain, 2003), in situ datasets of coincident Chl-a and reflectance measurements. *Empirical blue–green (440–550 nm) spectral band-ratios are the most common types of ocean color algorithms used for Chl-a retrievals because most of the phytoplankton absorption occurs within this portion of the visible spectrum*. However, the use of visible wavelengths can be unreliable in coastal waters. In optically complex, Case 2 waters, blue–green reflectance band-ratios become less sensitive to changes in Chl-a concentrations because increasing concentrations of color dissolved organic matter (CDOM) and total suspended matter (TSM) (e.g., Bowers et al., 1996) require the use of other spectral bands located in the red (620–700 nm) and NIR (>700 nm) (e.g., Gitelson et al., 2009).

**3.2.1. Blue–green band-ratios for open and coastal ocean waters**

Gordon et al. (1983, 1980) and Feldman et al. (1984) were among the first to use empirical band-ratios from CZCS spectral bands for the study of the near-surface distribution of phytoplankton blooms in the open ocean and to explore their relationships with oceanographic conditions. Many other studies used their initial work to derive global phytoplankton maps from CZCS Chl-a imagery (e.g., Banse and English, 2000, 1997, 1994; Nezlin et al., 1999; Tang et al., 1999; Fuentes-Yaco et al., 1997b). The second and third generation of ocean color sensors addressed the need for more spectral bands, thereby enabling the development of more sophisticated atmospheric correction schemes and in-water constituent retrieval algorithms, which are required for both improved retrieval accuracy for water quality variables and algal bloom proxies in coastal ocean waters. SeaWiFS OC4 (O’Reilly et al., 2000, 1998) and MODIS OC3M (Campbell and Feng, 2005b, 2005a) are switching band-ratio algorithms that use spectral bands in the blue and green regions of the visible spectrum to estimate Chl-a concentrations. The MODIS OC3M (e.g., Chen and Quan, 2013) is extended from the SeaWiFS OC4 and adapted to the MODIS spectral bands. The use of global standard ocean color band-ratios has been demonstrated to significantly overestimate Chl-a. Moore et al. (2009) have shown that the nominal uncertainty of 35% for Chl-a retrievals is true in ocean gyres, but the OC3M relative Chl-a error is >50%outside those gyres and can be >100% in coastal waters. Komick et al. (2009) found that MODIS OC3M systematically overestimated Chl-a when lower than 0.13 mg m3 in Western Canadian waters. Similar results were also found by Radenac et al. (2013) for the equatorial Pacific warm pool. For SeaWiFS OC4, Volpe et al. (2007) found that Chl-a was overestimated by 70% for Chl-a levels lower than 0.2 mg m3 in the Mediterranean Sea. Such low Chl-a concentrations are encountered in the vast majority of the global ocean (Hu et al., 2012).

**3.2.2. The relevance of the red-NIR spectral regions in coastal waters**

The in vivo absorption peak near 676 nm is minimally affected by the influence of CDOM and TSM when the two are in low concentrations. Spectral bands near 676 nm have been widely used for the retrieval of Chl-a in coastal waters (Odermatt et al., 2012; Gurlin et al., 2011). Gitelson et al. (1999) have shown that reflectance increases in the NIR beyond 700 nm due to increased scattering from algal biomass, correlated to an increase in Chl-a for most phytoplankton groups. The sensitivity analysis conducted by Ruddick et al. (2001) on two red-NIR band-ratio algorithms revealed that the relative error on Chl-a retrievals became more significant at low Chl-a concentrations (<10mg-3) and in low backscatter conditions but also that the choice of paired wavelengths was very important. *Their study showed that a band-ratio algorithm that uses the red-NIR band pair 672 and 704 nm would perform best at Chl-a 10 mg m3, whereas a wider red-NIR band pair spreading further apart (e.g., 667 and 748 nm) would perform best at Chl-a 100 mg m3*. The ‘‘red-edge’’ is technically defined as an increase in spectral reflectance in the red-NIR (680–750 nm) and often results from the presence of partly submersed vegetation (e.g., Dierssen et al., 2007, 2006; Bostater et al., 2003; Gitelson, 1992) or algal bloom surface expressions (e.g., Shen et al., 2012; Ruddick et al., 2008) (Fig. 5). Only a few ocean color sensors have the spectral requirements that enable the detection of those reflectance features. For SeaWiFS, two of the nine spectral bands are positioned in red-NIR region of the spectrum (namely 670 nm and 765 nm), and these two bands have limited use for the detection of Chl-a. In contrast, MODIS and MERIS provide more spectral bands between 600 and 800 nm (Figs. 5 and 6).

**3.2.3. Band-ratio algorithms: Limitations and challenges**

Most reflectance band-ratios are designed for global applications over optically deep ocean waters (Odermatt et al., 2012).

* The use of band-ratios often leads to erroneous retrievals in coastal waters, where the optical complexity is highly variable (Dierssen, 2010; Blondeau-Patissier et al., 2004).
* Additional limitations in the use of band-ratios are regional differences in optical properties and concentrations; the generalized global parameterization of some algorithms is inapplicable in some of the world’s ocean regions (e.g., Volpe et al., 2007; Claustre and Maritorena, 2003; Sathyendranath et al., 2001; Dierssen and Smith, 2000).
* Many studies have shown that the retrieval accuracy of Chl-a by satellite ocean color sensors, aimed to be within ±35% in oceanic waters, cannot always be met when using band-ratio algorithms (e.g., Moore et al., 2009; Hu et al., 2000).
* In coastal waters, the quality of this retrieval significantly degrades and is often considered unreliable.
* The use of blue–green spectral bands for the specific detection of Chl-a in coastal waters is affected by the absorption signal of CDOM and TSM (e.g., Dierssen, 2010; Gower, 2000; Joint and Groom, 2000).

To overcome this limitation, other studies suggested the use of red-NIR band-ratios for Chl-a retrieval in coastal waters (e.g., Moses et al., 2012; Shanmugam, 2011).

**3.3. Spectral band difference algorithms**

Spectral band difference algorithms exploit spectral regions that feature significant changes in the reflectance spectrum due to the presence of an algal bloom, compared to the nearby bloom-free water. Given that absorption tends to vary more rapidly with wavelength than scattering, two adjacent reflectance spectral bands may have similar backscattering properties but will differ significantly in absorption. Hence, this absorption feature can be quantified by spectral difference. *The various forms of spectral band difference algorithms use band triplets from the red-NIR or the blue–green spectral regions (Table 2) depending on whether the algorithm is designed to be sensitive to an algal group, high chlorophyll concentrations or surface bloom expressions* (Fig. 7). *One of the most used ocean color spectral band difference algorithms is the Fluorescence Line Height (FLH) (see review of Xing et al. (2007)), an index for quantifying solar-induced chlorophyll fluorescence*. Other similar mathematical expressions are used to derive algal bloom indices from SeaWiFS, MERIS and MODIS (Table 2) and are discussed in this section.

**3.3.1. Fluorescence Line Height (FLH)**

The literature published on this topic since the 1960s (Yentsch and Menzel, 1963) has shown that estimating fluorescence is greatly beneficial to studies of phytoplankton biomass (Falkowski and Kiefer, 1985), physiology (e.g., Westberry et al., 2013; Behrenfeld et al., 2009), and composition (e.g., Hu et al., 2005). The remote sensing approach used to retrieve FLH was originally developed by Neville and Gower (1977), and its first application to an ocean color sensor was on MODIS-Terra (Abbott and Letelier, 1999; Letelier and Abbott, 1996). The spectral band positions of MERIS (Gower et al., 1999) and MODIS (Hoge et al., 2003) allow for the computation of FLH, but this product cannot be derived from CZCS, SeaWiFS and VIIRS because of the lack spectral bands in the 670–690 nm range. The use of MODIS FLH to detect HAB has been successfully used by many (Frolov et al., 2013; Cannizzaro et al., 2008; Hu et al., 2005), providing more reliable information than a standard Chl algorithm. MERIS FLH was found to be successful at detecting high biomass phytoplankton in sediment-dominated coastal waters (e.g., Gower et al., 2005). However, others studies have led to inconclusive results on the benefits of FLH in the detection of algal blooms (e.g., Tomlinson et al., 2008).

**3.3.2. Maximum Chlorophyll Index (MCI)**

The Maximum Chlorophyll Index (MCI) can only be applied to MERIS because of its use of the 708.75 nm band. This band is more responsive to strong reflectance in the NIR, and the lack of similar bands in MODIS and VIIRS may hamper the detection of high concentration bloom events. The MCI is mainly designed for the detection of high-concentration algal blooms, and it was successfully used to globally monitor phytoplankton blooms in the world’s oceans by Gower et al. (2008). The minimum Chl-a concentration required for a phytoplankton bloom to be detected by the MCI is 30 mg m3 (Gower et al., 2005), but phytoplankton blooms can have much higher concentrations, with some studies reporting Chl-a > 200 mg m3 (e.g., Gower and King, 2007a; Sasamal et al., 2005).

**3.3.3. Floating Algae Index (FAI) and Scaled Algae Index (SAI)**

Hu et al. (2010c) and Hu (2009) proposed the Floating Algae Index (FAI) to detect large (>4000 km2) surface-floating algae from MODIS 250 and 500 m bands in both fresh and marine environments. The FAI uses a functional form similar to FLH and MCI where the height of the NIR peak is estimated relatively to a linear baseline from adjacent bands in the red and short-wave infrared (SWIR) wavelengths (Table 2). Thus, the FAI is sensitive to the red-edge and is robust to the influence of CDOM, aerosols and sun glint because of the use of NIR bands. *However, similarly to MCI, the FAI is likely sensitive to turbid waters and shallow depths. It is used in combination with pre-determined thresholds to help separate land, cloud and high concentrations of submersed algae or sediments from pixels associated with surface algae scums*. Its global applicability remains untested. Building on this research, Garcia et al. (2013) developed an automatized image processing algorithm, the Scaled Algae Index (SAI), which is a necessary intermediate product for quantifying the spatial coverage of the floating macro-algae observed in satellite imagery based on FAI.

***3.3.4. Color Index Algorithm (CIA)***

*A recent development in algal bloom indices is the Color Index Algorithm (CIA) proposed by Hu et al. (2012). This empirical algorithm was originally developed to estimate surface Chl-a concentrations in oligotrophic (60.25 mg m3 ) waters. The CIA is a three-band reflectance difference algorithm (443 nm, 555 nm and 670 nm bands), making it applicable to SeaWiFS, MODIS and MERIS. Its accuracy has not yet been fully validated because of the lack of low-concentration, high-quality in situ Chl-a data from the world’s oceans. The CIA was successfully applied to the waters of the Red Sea by Brewin et al. (2013), where it was found to perform better than the MODIS OC3 at retrieving Chl-a because of the low concentrations typically encountered in those waters.*

**3.3.5. Spectral band difference algorithms:** Limitations and challenges.

The robust application of FLH for the detection of algal blooms remains under discussion. Gilerson et al. (2007, 2008) found that MODIS FLH retrieved fluorescence with reasonable accuracy only for waters with Chl < 4 mg m3 but that the signal was masked by particulate backscattering in turbid waters with high CDOM and TSM concentrations.

The authors also questioned the performance of the MERIS FLH algorithm in coastal waters where large errors could be introduced via the linear baseline between the 665 nm and 708.75 nm bands.

The use of a linear baseline between the two NIR MERIS bands for the computation of FLH has been demonstrated to work in coastal waters with Chl-a concentrations of up to 20 mg m3 (Gower and King, 2007b) (Fig.

For higher chlorophyll concentrations however, due to the combined effects of water absorption and Chl-a distortion, the Chl-a spectrum and a linear baseline can no longer be used.

Gower et al. (1999) also challenged the theory that the scattering by TSM had a significant reducing effect on the relative fluorescence height above the baseline. The influence of CDOM on the FLH is often considered small in relatively low concentrations due to its negligible absorption in the NIR. The author suggested the use of true-color FAI-paired images to separate the clouds from the ocean surface features.

**6. Conclusions and future directions**

For simplistic approaches, such as the sole use of the reflectance spectrum, it is recommended that knowledge of the study area and a further detailed analysis of the pixels’ reflectance spectra are taken into account to ensure the validity of the algal bloom information retrieved. *Band-ratio algorithms are shown to be adequate for open ocean waters,* but their use in complex coastal waters is limited, particularly when band-ratios use blue and green bands because the influence of CDOM and TSM at those wavelengths affects their retrievals.

Alternatively, *many studies have shown the great potential of band-ratios using red and NIR bands to detect Chl-a and algal blooms in coastal waters (e.g., Le et al., 2013) because this spectral region is less affected by those two optically active constituents.*

Spectral band difference ocean color indices have also been shown to provide reliable information on algal blooms in both open and coastal ocean waters. Ocean color indices are often used in combination with other descriptors to enhance the interpretation of satellite imagery that possibly features algal bloom events (e.g., Yuan et al., 2005). *The MERIS MCI in particular is ‘‘a versatile tool’’ (Binding et al., 2012, 2013), but the FLH and FAI, as well as other indices, can be used in conjunction with satellite estimates of Chl-a and SST to improve our image analysis in a very efficient way*.

*There are currently few optical and biogeochemical in situ data characterizing pre- and post-blooms conditions. Automated in situ sensors, such as Autonomous Underwater Vehicles (AUV) equipped with bio-optical sensors, may provide a solution (IOCCG Report 11, 2011).*

In addition to parameters derived from ocean color, today’s satellite-derived variables also include Photosynthetic Active Radiations (PAR), wind speed and direction, rainfall, salinity and sea surface height (SSH). These variables can all be combined to provide a more complete assessment of the underlying factors of algal bloom events (e.g., Srokosz et al., 2004; Urquhart et al., 2013). The use of multi-source datasets from in situ data, satellite products or ecosystem models and their analysis with statistical methods is a prerequisite for fully understanding algal blooms’ onset mechanisms and dynamics.

Looking back to the potential of ocean color remote sensing, Cracknell et al. (2001) noted that ‘‘operational real-time monitoring of the location, extent, movement and growth rate of a phytoplankton bloom is an important challenge at present’’ (p. 221). The capabilities of ocean color remote sensing in providing operational real-time monitoring of a phytoplankton bloom are progressing but continue to remain a challenge in coastal waters (e.g., Malone, 2008). Remote sensing ocean color imagery has provided, and still provides, an invaluable source of frequent, synoptic information. Combined satellite datasets from SeaWiFS, MERIS and MODIS ocean color sensors are equivalent to almost two decades of ocean color imagery for the open and coastal ocean from 1997 to the present, with a brief window of earlier data provided by CZCS from 1979 to 1986.

***2. 29. A Statistical Algorithm for Estimating Chlorophyll Concentration in the New Caledonian Lagoon***

**2. Material and Methods**

**2.1. Data**

Two databases are used in this study: world data from SeaWIFS Bio-optical Archive and Storage System (SeaBASS:[40]) and data collected in the New Caledonia area (NCDataBase). Each database contains in situ and MODIS Rrs values in several spectral bands centered on 412 nm, 443 nm, 488 nm, 531 nm, 555 nm, and 667 nm for NCDataBase [29,31,38] and 547 instead of 555 nm for SeaBASS [40,41].

All MODIS Rrs over New Caledonia in the NCDataBase were extracted from 2002 to 2010 [42].

**2.3. Algorithm Steps**

The steps to get an algorithm adapted to New Caledonia are the following:

(1) using the NCDataBase, determine a model for low [chl-a] (AFLC), i.e. a well-suited model for waters having low [chl-a];

(2) using the SeaBASS database, determine a model for waters with high [chl-a] (AFHC);

(3) using the two merged databases, determine a criterion to distinguish low and high [chl-a]; and

(4) implement a continuous connection between the models for low and high [chl-a].

The number of all the combinations with six variables (Rrs(412), Rrs(443), Rrs(488), Rrs(531), Rrs(555) and Rrs(667)) is 63



When a model formed with many variables gave results equivalent to a model formed with fewer explanatory variables, the model with fewer variables was chosen. For each of these 63 models, 50 RMSE values, one per sample, were computed. The predictive variables are the Rrs in the five spectral bands centered on 412, 443, 488, 531, and 555 nm. This SVM model was compared to OC3.

Step 3 consists in determining from MODIS Rrs if the [chl-a] is high or low. In this step, two methods were tested to determine what MODIS color ranges are linked to a high or a low [chl-a]: SVM (as a classifier) and decision tree. As explained in more detail later (Section 4.1), the decision tree was preferred to the SVM because of its practicality. Indeed, only the ratio Rrs p488q {Rrs p555q is used to determine which group of [chl-a] should be linked to a MODIS color. For Step 4, several kinds of continuous connections, with weight functions, between the AFHC and the AFLC were tried: linear, quadratic, root squared, logarithmic, exponential, and arc-tangential. Equations (5.1)–(5.4) describe some weight functions with s the threshold determining the limit between high and low [chl-a], ε P s0;sr the tolerance used to set the transition interval width, a “ s ´ ε the inferior bound of the transition interval, b “ s ` ε the superior bound of the transition interval, and x is the variable which represents the ratio Rrs p488q {Rrs p555q.

**2.4. Statistical Tests/Acc. Assessment.**

In order to verify the effectiveness of an algorithm without an overtraining effect, data were systematically divided into two samples: one learning sample to build the model, and one test sample on which the built model was applied and checked with indicators (specified after). The learning sample was constructed with 70% of the data and the test sample was formed with the remaining 30%.

**Article 3: Evaluation of Sentinel-2 and Landsat 8 Images for Estimating Chlorophyll-a Concentrations in Lake Chad, Africa**

**2.2.2. Landsat 8**

The L8 satellite supplies multispectral images comprising of 11 bands, the majority of which have a 30-m spatial resolution. The red, near-infrared (NIR) and shortwave infrared (SWIR) bands in the L8 satellite images have a narrower bandwidth than those from previous Landsat missions. The L8 radiation resolution was increased to 16 bits, and the SNR was increased significantly. These advances improved L8’s pigment discrimination ability. Though built for terrestrial applications, these bands have proven useful for estimating concentrations of Chl a in water bodies [54,55].

**2.3. Methods 2.3.1.**

**Satellite Data Preprocessing**

When radiation travels from a reflected surface through the atmosphere to a satellite sensor, absorption and scattering from different molecules modifies the direction and intensity of the emitted radiation. The SNR ratio of the ToA radiance of a waterbody varies depending on the atmospheric aerosol concentrations. This makes controlling these atmospheric effects a major challenge when dealing with remote sensing of inland water bodies. An accurate estimation of any water quality parameter requires atmospheric correction (AC) for any given atmospheric effects [56]. As such, before image processing in this study, rigorous AC was conducted across all the images. To achieve uniformity in the processing steps, all the images used in this study were Level 1 products. Because we will be estimating Chl-a concentration from satellite imagery without reference to in situ estimates, it is essential to remove any intervening atmospheric effects from the acquired Landsat images. This reduces error sources from the resulting estimated Chl-a concentrations in the lake. For this purpose, two conventional AC algorithms were applied:

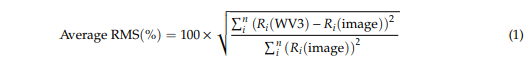
Fast Line-of-Sight Atmospheric Analysis of Hypercubes (FLAASH) [57]

Quick Atmospheric Correction (QUAC) [58].

The FLAASH method is used to estimate the scattering effects caused by the atmosphere. FLAASH needs additional input from each image in order to operate. These specific inputs are found in header files which are usually downloaded alongside their respective multispectral images. The QUAC is a simpler method which establishes a relationship between the surface reflectance and the observed radiance signal. QUAC can be applied to VIS and NIR-SWIR bands, and does not require additional inputs from image header files. FLAASH and QUAC can be conducted using the built-in functions in ENVI®. The downloaded S2 satellite images (L1C) included previously-orthorectified ToA images. Using the two AC methods (FLAASH and QUAC), we processed the acquired L1C images and obtained bottom-of-atmosphere (BoA) 2A-level products. Similar ACs were performed for L8 images, whereby their digital numbers (DNs) were converted to sensor radiance values.

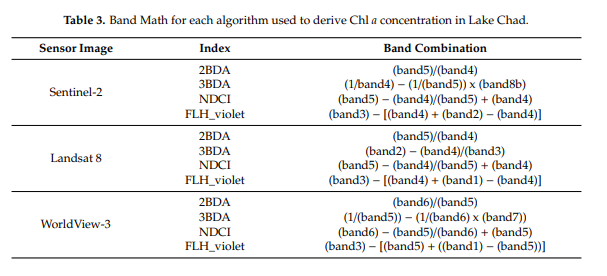
The radiance values were then converted to bottom surface reflectance, which represents surface reflectance from the water body. WV3 spectral imagery was used as a reference to evaluate which AC methods performed most accurately for both Sentinel and Landsat sensors.

The average root mean square percentage (RMS (%)), defined in equation 1, was used to assess this preprocessing step.



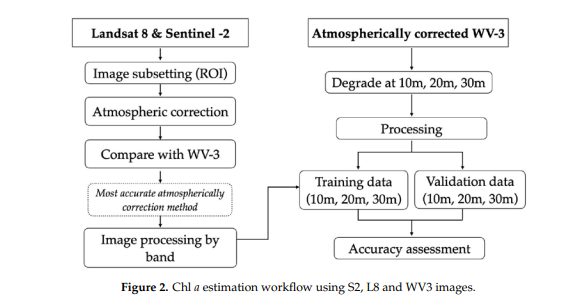
**2.3.2. Estimating Chl a Concentration**

Research studies have shown that atmospheric and aerosol effects on satellite products can be reduced using spectral ratios. *These spectral ratios have also been frequently used to estimate Chl-a [56,59]. Chl-a is known to have prominent scattering–absorption patterns between certain wavelengths. For instance, around the blue region of the electromagnetic (em) spectrum (between 450 and 475 nm), Chl a exhibits high absorption tendency*. *This is also experienced around the red region at 670 nm. At the green and NIR regions of the em spectrum, Chl a exhibits high reflectance values that could reach 500 and 700 nm, respectively.* This *information has extensively been used by researchers to develop Chl-a quantification algorithm [31,51].* Numerous water quality satellite reflectance algorithms have been used for retrieving Chl a concentration [59,60]. In this study, we selected four algorithms: 2BDA [27,61], 3BDA [28,62], NDCI [22] and FLH [29]. These algorithms were selected based on reviews and the Chl a estimation accuracy for lakes [63,64]. These accuracies relied on the degree of agreement between derived Chl a estimates from satellite images, and Chl a estimates obtained on site and tested using laboratory methodologies. Band position, spacing and width were also considered while selecting the four algorithms for this study. Another selection criterion was the ease with which these algorithms could be implemented by resource managers in the future Using the Band Math function in ENVI®, we first established the performance baseline for this study by applying 2BDA, 3BDA, NDCI and FLH algorithms to an atmospherically corrected WV3 reflectance image. The same algorithms were then applied to Sentinel-2 and L8 reflectance products (Table 3). Band Resampling was performed before processing the S2 image using 2BDA, 3BDA and NDCI index, as shown in Table 3. S2 Band 4 (10-m resolution) was resampled to match the resolutions of bands 5 and 8b (20-m) using the nearest neighbor method, which is an easy-to-compute resampling method that preserves the input data values during processing. The resulting S2 2BDA, 3BDA and NDCI index products had a 20-m resolution, while that for FLH was maintained at its original 10-m resolution.

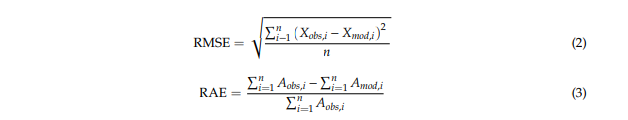


**2.3.3. Accuracy**

The accuracies of modeled Chl a data from S2 and L8 images, and comparisons between them, were evaluated using data developed from WV3. For this study, we are interested in quantifying Chl a concentrations from S2 and L8 images (Figure 2). Accordingly, water pixels from the WV3 images were extracted for all the bands. This was followed by manually removing uncertain classes, which appeared predominantly around the boundary edges. The obtained water pixels were upscaled to 10-, 20- and 30-m resolutions using an area-weighted spatial model. The model simply averages neighboring smaller pixels into large ones. Consequently, the WV3 pixels could represent the percentage of Chl a concentrations at scales that matched the Landsat and Sentinel images. The algorithms in Table 3 were subsequently applied to the surface reflectance from the WV3 image.



From the newly derived Chl a data, we randomly selected training and validation datasets for the 10-, 20- and 30-m spatial resolution images. These datasets were sampled equally across the test site to ensure the representation of the entire range of Chl a. The training and validation sampling points were used to map Chl a across the S2 and L8 images. The S2 and L8 images were also resampled from the training and validation datasets so that their sampling points and grids would overlap at exactly 10 and 20 m for S2, and 30 m for L8 pixels. After applying the Chl a extraction algorithm to the surface reflectance of the S2 and L8 images, their respective resulting indexed products were used as variables for Chl a mapping comparison.

The evaluation of the derived Chl a estimates was carried out using similar estimates derived from the WV3 images. The root mean square error (RMSE) and relative area error (RAE) were used as the statistical metric for this evaluation purpose. They can be calculated using the following formula  where Xobs,i is the referenced chlorophyll concentration at point i, Xmod,i is the modeled estimate of Chl a concentration at point i, n represents the total number of validation points, and Aobs,i and Amod,i represent areas of the referenced and modeled Chl a estimates for point *i*

*Search for this Article(s): Watanabe, Remote sensing of Chlorophyll-a based on Landsat 8 and S2 MSI in Barra Bonita reservoir in Brazil.*

*Landsat 8 OLI two band algorithm for chlorophyll distribution mapping in hypertrophic waters.*

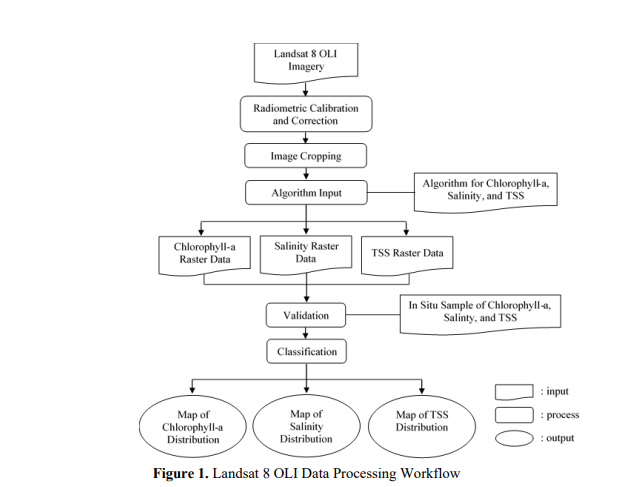
**Article 4: A machine learning approach to estimate chlorophyll-a from Landsat-8 measurements in inland lakes**

2.3. Satellite data acquisition and processing 1262 OLI scenes (radiance data) in MLRY and RHR were obtained from the United States Geological Survey (USGS) portal (Fig. 2). Radiance data were used instead of reflectance because reflectance was processed using a model developed for land applications, and radiance is not prone to error over water (Ilori et al., 2019). Only images with < 70% cloud cover, based on visual examination, were downloaded. A full atmospheric correction test (NIR-SWIR and MUMM) through the SeaWiFS Data Analysis System (SeaDAS, version 7.5) often failed in most of pixels in the lakes because the turbid waters caused the assumptions to fail. The dark-spectrumfunction method in ACOLITE (Vanhellemont and Ruddick, 2018) may result in large uncertainties in MLRY lakes (e.g., Lake Taihu) due to excessive algae particles and strongly absorptive aerosols (Wang et al., 2019). Therefore, the following procedure was used to create pseudoreflectance products (Cao et al., 2017; Feng et al., 2012). First, Rayleigh-corrected reflectance (Rrc) was derived after correction for Rayleigh scattering and gaseous absorption effects using SeaDAS 7.5 with ancillary data (such as meteorological data) (Franz et al., 2015). To remove, at least partially, the aerosol signal, the following algorithm was used: R' rc(λ) = Rrc(λ) - Rrc(2201). This method may retain residual aerosol signals in other bands; however, it partially removes the bulk aerosol, haze, or glint signal (Cao et al., 2017; Feng et al., 2012). Cloud-contaminated pixels were removed via a threshold set on the SWIR reflectance (Rrc(2201) > 0.018) (Aurin et al., 2013). Waterbody boundaries were extracted using a scheme of normalized difference water index (NDWI) threshold segmentation (Li and Sheng, 2012), which is an automated mapping algorithm based on hierarchical image segmentation and delineates each waterbody using a local segmentation threshold. Subsequently, the segmentation-based water boundary of OLI was screened using the Chinese lake boundaries (Ma et al., 2011). Reservoirs, rivers, and ponds were excluded, and only lakes larger than 1 km2 were considered. OLI may not observe narrow sections in some thin lakes (< 30 m) and regions covered by macrophytes. Because algal blooms common to our study lakes are usually caused by cyanobacteria, surface scums can be present; hence, a threshold −0.004 on the floating-algae-index (FAI) was used to exclude pixels with algal blooms (Hu et al., 2010)

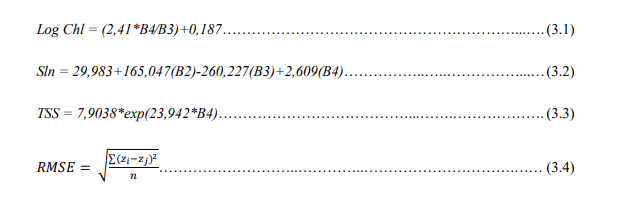
**Article 5: 24. Identifying Distribution of Chlorophyll-a Concentration Using Landsat 8 OLI on Marine Waters Area of Cirebon**

**2. Methodology**

The study area of this research, is done in parts of seawater that exist in the coast of a part of a district that is in the city and regency administrative border of Cirebon next to Java Sea. The districts mentioned in this research, are Mundu, Astanajapura, and Pangenan district in Cirebon regency along with Kejaksan district, and Lemahwungkuk in Cirebon city. The research area is in the coordinate location of 108°34’00”-108°39’30” E and 6°41’30”-6°47’00” S. The variables used in this study are the distribution of chlorophyll-a concentration, salinity, and TSS. These three variables are obtained from several data which are rainfall data and reflectance data on Landsat 8 OLI images. Materials needed to gather data are Landsat 8 OLI images and in-situ samples from the three variables. The rainfall data used to determine wet months and dry months, are rainfall >200 mm/month are categorised as wet month and rainfall <100mm/month are categorised as dry month [10]. Landsat 8 OLI images are used as a representation of the distribution of chlorophyll-a variables, salinity, and TSS. In-situ data samples are used to validate images that have been processed. The processing of Landsat 8 OLI imagery (Figure 1) are done using ENVI 5.1 and ArcGIS 10.1 software. ENVI 5.1 is used to process radiometric correction and calibration, cropping, and algorithm input. Radiometric calibration is used to convert image pixels into reflectance value. The radiometric corrections that are done are sun angle corrections to fix error reflectance values caused by sun positioning [11]. Cropping is done to minimize the image coverage according to the study area. Algorithm inputs are done to convert reflectance values of images into chlorophyll-a, salinity, and TSS values using a certain equation. The data processing uses ENVI 5.1 software to produce three raster data, which are raster with the three variable values. After the three data of raster are produced, ArcGIS 10.1 is used to process said data for image validation and classification. A detailed explanation regarding validation is provided in the next paragraph. Classification is done to divide three variable values into four or five classes to simplify the spatial analysis of the three variables. Data processing using ArcGIS 10.1 software will produce the chlorophyll-a distribution map, salinity map, and TSS map.



Landsat 8 OLI imagery data that have been corrected and calibrated are then converted into chlorophyll-a concentration values, salinity, and TSS. There are three algorithms which are used to identify the concentration of chlorophyll-a, salinity, and TSS. Chlorophyll-a concentration using algorithm of Wibowo et al. [12], salinity using the algorithm of Supriatna et al. [13], and TSS using the algorithm of Budhiman [14], where these three researches have done the mapping of variables related to Indonesian seawaters. These three algorithms can be seen in equations 3.1, 3.2, and 3.3. The imagery that has been processed into the three variables, can then be validated by counting the Root Mean Square Error (RMSE) between specified sample data from image processing with in-situ measurements which obtained by field survey. Acquiring sample points are based on distance from coastline each 600 meters until it reaches the distance approximated about 6 km as many as 20 points (Figure 2). RMSE uses the equation 3.4 below

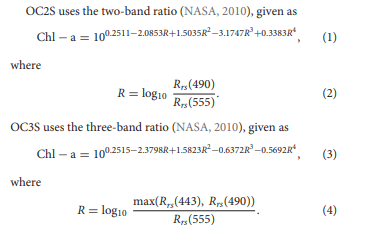


Data analysis used in this research are spatial analysis and descriptive statistics. Spatial analysis is done by overlaying the three map variables. Statistical analysis is done with testing the relationship between the sample values of the three variables during wet season and dry season with regression method. Temporally, chlorophyll-a concentrations are discussed according to wet month and dry month on the year 2014-2015. In regard to spatial distribution, the concentration of chlorophyll-a is discussed with salinity variables and TSS to discover its relationship.

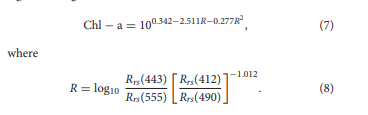
**Article 6: 30 Assessment of Satellite-Based Chlorophyll-a Algorithms in Eutrophic Korean Coastal Waters.**

**Ocean Color Chl-a Algorithms**

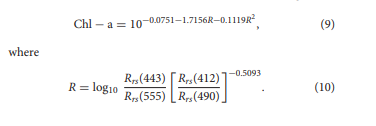
Six algorithms (two OCxS algorithms for the open ocean, one GOCI-standard algorithm, and three Tassan’s algorithms regionally modified for Korean waters) were selected to assess the GOCI-derived Chl-a estimates in Jinhae Bay (Moon et al., 2010; NASA, 2010; Siswanto et al., 2011; Kim W. et al., 2016). The OCxS algorithms were developed as the operational algorithms for NASA’s Sea-Viewing Wide Field-of-view Sensor (SeaWiFS). They use a fourth-order polynomial relationship between the ratio of Rrs (i.e., blue-green band ratio) and Chl-a (NASA, 2010). In this study, we tested two OCxS algorithms: OC2S and OC3S. OC2S uses the two-band ratio (NASA, 2010), given as



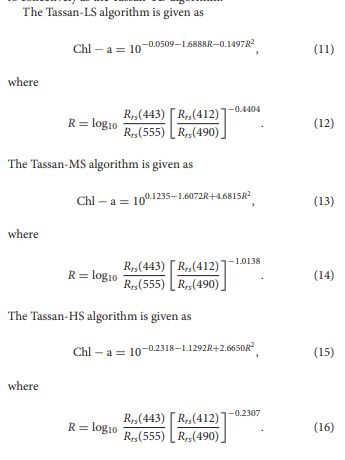
As the regional algorithm for Korean waters, we first tested the GOCI-standard Chl-a algorithm, developed for GOCI products using in situ Chl-a data collected around the Korean Peninsula from 1998 to 2009 (Moon et al., 2010). The GOCI-standard is given as:



Kim W. et al. (2016) modified Tassan’s algorithm for the GOCI product by using in situ data (2010–2014) collected from coastal areas around the Korean Peninsula, East China Sea, and Tsushima Strait. This modification, termed Tassan-All, is given as:



Kim W. et al. (2016) also modified regionally the coefficients of Tassan’s algorithm, applying different coefficients for three TSM levels: low TSM of 0–0.5 g m−3 or Rrs(555) 10 g m−3 or Rrs(555) >0.015 sr−1 (Tassan-HS). The three modifications are referred to collectively as the Tassan-TD algorithm. The Tassan-LS algorithm is given as:



**Article 7: 31. Estimation of Chlorophyll-a in Northern Coastal Bay of Bengal Using Landsat-8 OLI and Sentinel-2 MSI Sensors**

Methods Landsat-8 OLI level-1 data product consists of quantized and calibrated scaled DN values while Level-1 data obtained from Sentinel-2 MSI is the TOA Reflectance. Retrieval of Chl-a from the satellite sensors over the study region involves four steps, (i) Obtaining absolute TOA Reflectance from scaled DN values in case of Landsat-8 OLI and scaled TOA Reflectance for Sentinel-2 MSI, respectively, for all the required bands, (ii) Conversion of TOA Reflectance to Surface Reflectance (actually originating from the water surface), (iii) Conversion of the Surface Reflectance to corresponding Remote Sensing Reflectance (Rrs) at these bands, and finally (iv) Retrieval of Chl-a from the Rrs utilizing Ocean Chlorophyll (OC) algorithms (2- band: OC-2, and/or 3-bands: OC-3). The scaled DN values from Landsat-8 OLI and scaled TOA Reflectance from Sentinel-2 MSI for the required bands were converted to corresponding Surface Reflectances using the special-purpose inbuilt calculator in QGIS software. It is noted that the Surface Reflectance mentioned refers to that originating from the water surface. For Landsat-8 OLI, TOA Reflectance (ρp), which is the unitless ratio of reflected vs. total power energy (NASA, 2011), is calculated using the formulation:



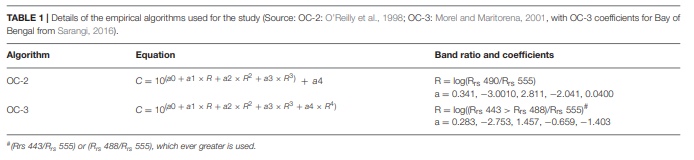
where, ρp is the TOA Reflectance, Lλ is the spectral radiance at the sensor’s aperture (at-satellite radiance), d is the EarthSun distance in astronomical units (provided in Landsat-8 metadata file available from: https://earth.esa.int/documents/ 10174/679851/LANDSAT\_Products\_Description\_Document. pdf), ESUNλ is the mean solar exo-atmospheric irradiances, θs is the Solar zenith angle in degrees, which is equal to θs = 90◦–θe, where θe is the Sun elevation, and π = 3.142. Sentinel-2 data which are already scaled TOA Reflectance were converted to absolute TOA Reflectance from the quantification value provided in the metadata. Surface Reflectance (ρ) are then determined from TOA Reflectance for the two sensors at the necessary bands (λs) following Moran et al. (1992), as:



where, Lp is the path radiance, Tv is the atmospheric transmittance in the viewing direction, Tz is the atmospheric transmittance in the illumination direction, Edown is the downwelling diffuse irradiance. Following atmospheric correction and Dark Object Subtraction (DOS), Rrs are obtained from the Land Surface Reflectance following Moses et al. (2015), given by:



The Rrs obtained for the bands are then used in the bio-optical algorithms for retrieval of Chl-a in ArcGIS software for OLI and MSI sensors. Similarly, Level-1B MODIS-A data is processed in SeaDAS software to obtain Chl-a using OC-2 and OC-3 algorithms through L2gen. Many different algorithms like OC2 algorithm version-2 (OC2v2), OC3, Global Processing (GPs), Morel-1, 2, 3, and 4 have been developed over the past several years to estimate Chl-a from the reflectance of specific bands obtained from various satellite sensors (O’Reilly et al., 1998, 2000). However, algorithms to estimate Chl-a can be broadly categorized into two, empirical algorithms and semi-analytic models. Some of the algorithms require Rrs values of specific bands obtained from various satellite sensors while others require normalized Water Leaving Radiance (Lwn) at specific bands again obtained from various satellite sensors. Given the dynamic nature of the study region, OC-2 and OC-3 algorithms have formed the obvious choice in our analysis since they are well-suited for case-2 waters among a number of existing bio-optical algorithms (O’Reilly et al., 2000). These algorithms are based on the non-linear relationship between oceanic reflectance and in situ measured Chl-a, more precisely the ratios of reflectance in blue and green bands or their combinations. OC-2 is a modified cubic polynomial algorithm which was originally developed for the SeaWiFS data and tuned to the SeaBAM data (O’Reilly et al., 1998). OC3 algorithm has been used to retrieve low as well as high.



Chl-a hence making it useful for case-2 waters like estuarine regions (Morel and Maritorena, 2001). In the present work, OC-2 and OC-3 algorithms were used with the suitable nearest band ratio combination of the two sensors. Table 1 shows the two algorithms used in the present study, their band ratios and coefficients. For OC-2, ratio of Rrs at 490 and 555 nm are used. For OC-3, the higher ratio between Rrs at 443 and 555 or 448 and 555 nm are used. Coastal Ocean Monitoring and Prediction System (COMAPS) data used in the present work were available only till the period 2014 for the study area. MSI being a relatively new sensor with data availability only from 2016, the model validation exercise is carried out only for Landsat-8 OLI derived Chl-a. Data used for validation are taken from the samples collected at four stations along Hooghly and Sandheads. The study area map comprising of northern coastal Bay of Bengal is shown in Figure 1. Chl-a averages of the pre-monsoon periods during the years 2014–2017 as retrieved from Landsat-8 OLI using the OC-2 algorithm are overlaid in the background along with the locations of the COMAPS data stations (Figure 1b) The details of the station name, station id, longitude, and latitude are given in Table 2. For the validation exercise, Landsat-8 OLI data geographically collocated with the corresponding in situ observations during the periods December 2013–March 2014 are extracted and converted to Rrs to apply OC-2 and OC-3 algorithms and TABLE 2 | Details of the COMAPS station locations used for validation of Chl-a estimated from Landsat-8 OLI. Station name Station ID Latitude (◦N) Longitude (◦E) Hooghly Estuary HE-00 21.7047 88.0270 Sandheads SH-0.5 21.6257 88.0746 Sandheads SH-05 21.5855 88.0800 Sandheads SH-02 21.6124 88.0771 derive Chl-a. The statistical metrics used for validation were correlation coefficient (r), Root Mean Square Error (RMSE) and bias. The formula for RMSE and bias are given in Equations (4) and (5).

**Article 8: Remotely sensing harmful algal blooms in the Red Sea**

Remote sensing parameters for monitoring HABs in the Red Sea Several different parameters have been used to investigate occurrence of HABs, including false color composite imagery, satellite-derived Chl-a maps, and spectral analysis-based bloom maps [51,52]. These parameters have enabled us to detect the presence/absence of HABs and map the spatial extent of different HAB species that were previously reported by various field programs in the Red Sea. A detailed synopsis of the HAB analysis is provided as follows. The first step of this analysis was to visually interpret the abnormality of ocean color in the Red Sea during HAB events. To achieve this, we processed false-color composite MODIS images by combining Rrs in the NIR, green and blue wavelength bands (also known as the RGB band). Hu et al. [53] demonstrated that false color imagery can easily distinguish between dark features caused by high absorption of light related to the presence of Chl-a, and bright features caused by non-pigment materials such as sediment, corals and shallow bathymetry. Secondly, to investigate Chl-a concentrations during a HAB event, we examined satellite-derived Chl-a generated using the ABI algorithm. Gokul and Shanmugam [19] demonstrated that satellite-derived ABI Chl-a can effectively estimate Chl-a values associated with HABs, and discriminate algal bloom patches from other non-algal particles in optically complex coastal waters. In addition, based on the satellite derived Rrs spectra (training dataset) and its second-order derivatives of different phytoplankton functional types (PFTs), we produced bloom maps in order to detect the presence/absence and map the spatial distribution of different phytoplankton species that were known for HAB outbreaks in the Red Sea (see Fig 2). Finally, in situ measurements of the different phytoplankton cell counts (related to these HAB species) were used for the validation of the model results



Objective 3

**References**