MATCHING SATELLITE AND GROUND-BASED DATA FOR CHLOROPHYLL-A SENSING IN RESERVOIRS

By

PRIYA KAYASTHA

Bachelor of Science in Environmental Science

Kathmandu University

Dhulikhel, Kavre, Nepal

2017

Submitted to the Faculty of the
Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of
MASTER OF SCIENCE
July, 2021

MATCHING SATELLITE AND GROUND-BASED DATA FOR CHLOROPHYLL-A SENSING IN RESERVOIRS

Thesis Approved:

Abubakarr Mansaray

Thesis Adviser

Andrew Dzialowksi

Scott Stoodley

ACKNOWLEDGEMENTS

I am thankful to my wonderful advisor and committee members Dr. Abu Mansaray, Dr. Andrew Dzialowski, and Dr. Scott Stoodley for their guidance and support throughout my research. The completion of this study could not have been possible without their expertise. The journey towards the completion of my master's degree has been a series of learning from mistakes and improving. I got the opportunity to learn and explore a lot in the field of research through their guidance. I am very grateful that I got this golden opportunity to work with them.

I am very grateful to Oklahoma Water Resources Board (OWRB) for providing me data to carry out my research. I want to specially thank Julie Chambers from OWRB for helping me gather data necessary for my study.

I am indebted to my family and friends in the USA and back in my home country, Nepal for sending positive thoughts, love, and encouragement. My mother, father, and brother have been my support system. Their constant love and motivation has inspired me a lot.

Special thanks to my friends and classmates Meghan, Molly, and Hailey here in Oklahoma State University for sharing this journey with me towards the completion of my master's degree. I would also like to extend my gratitude to Kavina Eksteen, Program coordinator of Environmental Science Graduate Program. She has always supported me encouraged me to do better.

Name: PRIYA KAYASTHA

Date of Degree: JULY, 2021

Title of Study: MATCHING SATELLITE AND GROUND-BASED DATA FOR CHLOROPHYLL-A SENSING IN RESERVOIRS

Major Field: ENVIRONMENTAL SCIENCE

Abstract

Algal blooms are a major concern due to the adverse health and economic impacts they have on humans, animals, and aquatic ecosystems. Chlorophyll-a, a photosynthetic pigment present in all major groups of algae, has been widely used as an indicator of algal blooms. Studies have used satellite remote sensing to develop Chl-a algorithms by pairing ground-based Chl-a and satellite spectral data that are temporally and spatially coincident. The problem with the development of satellite based Chl-a algorithms is that temporally coincident satellite spectra are not always available due to the mismatch between ground-based sampling and image acquisition dates. It is a common practice to use imagery acquired days before or after ground-based sampling assuming that negligible water quality changes would have occurred if no major limnetic and hydroclimatic activities happened. The literature suggests one day as the ideal temporal disparity between ground-based and satellite data, with some studies suggesting up to 7 days as acceptable. Previous studies have used multiple reservoirs to arrive at these conclusions. It is not clear if the variations observed were due the increasing temporal disparity or the multiple reservoirs. The objective of this study was to delineate the variations caused by increasing time windows and those caused by single versus multiple reservoirs. This study developed regression models using Landsat-5, Landsat-8, and Sentinel-2 for time windows ranging from the same day of satellite and ground-based data collection to eight days apart. We also developed and compared regression statistics for a single reservoir to those of multiple reservoirs. Single bands, band ratios, and spectral derivatives were regressed against Chl-a data from 10 reservoirs of Oklahoma. For the time window, the R^2 values for all three satellites decreased as the time window increased, with Sentinel-2 performing better than Landsat-8 and Landsat-5. The spectral bands present in Sentinel-2 are important predictors of Chl-a. The analysis for single versus multiple reservoirs revealed that the number of reservoirs was not as important as the time window and the satellite resolution for Chl-a sensing. Sentinel-2 and Landsat-8 provided high R^2 values when multiple reservoirs were used.

TABLE OF CONTENTS

Chapter Page

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

CHAPTER I

INTRODUCTION

Algae can form blooms that result from excess nutrients, increased sunlight, elevated temperature, and calm conditions in water bodies (Klemas, 2012; Boyer et al., 2015; Kislik et al., 2018). Some species can produce dangerous toxins that have adverse impacts on human and animal health (Klemas, 2012; Boyer et al., 2015; Song et al., 2021). Non-toxic species of algae can also be detrimental to the health of aquatic ecosystems when they bloom and die (Smolen, 2007; Kislik et al., 2018; Song et al., 2021). Dead algae undergo microbial decomposition and create anoxic conditions in the water. This may result in fish kills and death of other aquatic animals (Klemas, 2012; Boyer et al., 2015;Song et al., 2021). Algal blooms can also cause economic losses due to beach closures that impact businesses that depend on recreational activities associated with the affected waters (Klemas, 2012; Brooks et al., 2015). Chlorophyll a (Chl-a) is one of the pigments measured to determine the severity of algal blooms.

Chlorophyll a (Chl-a) is one of the pigments measured to determine the severity of algal blooms. It is a photosynthetic pigment that is present in the leaves of green plants and major groups of algae (Dodds and While, 2010; Kirk, 2011; Klemas, 2012; Yan et al., 2018). The presence of Chl-a is a key indicator of water quality and the trophic status of lakes and reservoirs (Dodds and While, 2010; Kirk, 2011; Yan et al., 2018). The concentration ranges of Chl-a in micrograms per liter (µg/l) can be used to classify water bodies as oligotrophic, mesotrophic, eutrophic, and hypereutrophic. Oligotrophic lakes have a Chl-a range of 0-2.6 µg/l, very little nutrients, deep clear water, and very little algae. The low algal concentration allows deeper light penetration and low decomposition rate of dead algae. Organisms in this type of lakes supports high oxygen levels throughout the water column. Mesotrophic lakes have a Chl-a concentration range of 2.6 -7.3 µg/l, medium amount of nutrients, and clear water with some algal blooms. Eutrophic and hypereutrophic lakes have a Chl-a range of 7.3-56 µg/l and 56-155+ µg/l, respectively. These lakes have very high nutrient contents, and murky and turbid water. The decomposition rate is very high and is oxygen deprived throughout the water column (Dodds and While, 2010).

Chlorophyll-a concentrations can be measured in the laboratory using various extraction methods, spectrophotometry, and fluorometry (Lorenzen, 1965; Leeuwe, 2006; Gohin et al., 2008; Tebbs et al., 2013; Zeng and Li, 2015). In-situ probes are also used to measure Chl-a concentrations in the field. These ground-based Chl-a monitoring programs target specific locations within water bodies at specific times. These ground-based Chl-a monitoring programs target specific locations within water bodies at specific times. A major problem with ground-based Chl-a monitoring is that algal

blooms can occur at spatial and temporal scales that do not match sampling frequencies (Tan et al., 2017). Ground-based monitoring may miss major bloom events creating potential exposure risk for human and animals.

Methods have been developed that fill these temporal and spatial gaps and complement ground-based monitoring. Research has utilized remote sensing with platforms such as Unmanned Aerial Vehicles (UAV), airplanes, and satellites to monitor Chl-a (Klemas et al., 2012; Gholizadeh et al., 2016; Kislik et al., 2018). Satellite remote sensing offers greater temporal and spatial resolutions than ground-based sampling alone (Klemas et al., 2012; Gholizadeh et al., 2016; Tan et al., 2017; Ansper and Alikas, 2018). Satellite remote sensing offers greater temporal and spatial resolutions than ground-based sampling alone (Klemas et al., 2012; Gholizadeh et al., 2016; Tan et al., 2017; Ansper and Alikas, 2018).

Chlorophyll-a has spectral properties that allow it to interact with solar radiation (Ho & Michalak, 2015; Gholizadeh et al., 2016; Tan et al., 2017). It absorbs energy from the violet-blue and orange-red wavelengths of light and reflects green and near-green portions of the electromagnetic spectrum. Chl-a also has a peak reflection near 700 nm in the near infrared (NIR) region (Yacobi et al., 2011; Lopez et al., 2020). These optical properties allow for the development of various methods that can be used to estimate Chl-a concentrations using spectral data obtained from satellites. Examples of such methods are band ratios and band algorithms (Tan et al., 2017), spectral derivatives (Becker et al., 2005), and color space transformations (Liu et al., 2017).

Satellite based spectral data and ground-based water quality data are used to develop algorithms to establish empirical or semi-analytical relationships (Tan et al., 2017; Yan et al., 2018). Empirical models explain the direct relationships between the spectral data and water quality parameters. The empirical approach uses bands, band ratios, spectral derivatives, and band indices as the independent variables and sampled Chl-a concentrations as the dependent variable (Kloiber et al., 2001; Deutsch et al, 2014; Tan et al., 2017; Yan et al., 2018; Buma and Lee, 2020). Semi-analytical algorithms consider the empirical relationships between the independent variable (spectral data) and the dependent variable (Chl-a) as well as the inherent optical properties of the parameters under study. Inherent properties such as the backscattering coefficient, the absorption coefficient of water, and the biochemical properties of Chl-a that give it unique absorption and reflectance tendencies at different wavelengths are considered for developing semi-analytical Chl-a models (Yan et al., 2018).

Examples of satellite platforms that have been used to develop such empirical and semi-analytical relationship include the Landsat series (Landsat 1-8), Medium Resolution Imaging Spectroradiometer (MERIS), Moderate Resolution Imaging Spectroradiometer (MODIS), the Sentinel series (Sentinel 1-3), and commercial satellites (PlanetScope, Quickbird, Worldview, RapidEye). Each of these satellites varies with respect to temporal, spectral, radiometric, and spatial resolution as well as image cost and availability (Torbick and Corbiere, 2015; Shi et al., 2019; Buma and Lee, 2020; Mansaray et al., 2021).

Two of the most commonly used satellites are Landsat and Sentinel. Landsat provide open-source images with a spatial resolution of 30 m and temporal resolution of 16 days. Landsat has been used to remotely sense Chl-a in water bodies (Kloiber et al., 2001; Tan et al., 2017; Ouma et al., 2020). A ratio of Red to NIR is a common band combination that is used to estimate Chl-a concentrations (Han and Rundquist 1997; Moses, et al. 2009; Yacobi, et al. 2011; Tebbs et al., 2013; Deutsh et al., 2014). Other examples of band combinations include a ratio of Blue to Green (Torbick, et al. 2008), a normalized difference between Near-Infrared (NIR) and Red at specific wavelengths (Mishra and Mishra 2012), a three-band model (TM^{-1}-TM^{-1})*TM4 (Tan et al., 2017), a ratio of Blue to Red (Laili et al., 2015). Single bands have also been used to develop Chl-a models. An example of such single band is the Landsat-8 Operational Land Imager's (OLI) Green band derived linear model (Ouma et al., 2020).

The Sentinel-2 European multi-spectral instrument (MSI) has been used to improve the detection of Chl-a in inland water bodies (Xu et al., 2019; Buma and Lee, 2020; Pirasteh et al., 2020; Saberioon et al., 2020). The Sentinel-2 has additional bands in the Red-Edge (RE) region that uniquely detect Chl-a in water bodies. The RE band is significant as algae cause a peak reflectance near 700 nm making it suitable for the development of a wide range of Chl-a algorithms (Bramich et al., 2021). Sentinel-2 also has other bands that are similar to the Landsat and they have also been used to detect Chl-a. The MSI's combination of Green and Shortwave near-infrared-1 (SWIR-1) (Ouma et al., 2020), NIR-Red band ratios (Watanabe et al., 2017), a three-band NIR-Red model ($R^{-1} - RE^{-1}) \times NIR$), a NIR-Red ratio $(R^{-1} - RE^{-1})/(NIR^{-1} - RE^{-1})$ (Ansper and Alikas,

2018), and a Green-Red band ratio (Ha et al., 2017) are examples of the predictors for Chl-a detection in inland water bodies.

Spectral derivative techniques are used to detect Chl-a in water bodies. They eradicate background signals, distinguish overlapping spectral features, and reduces interferences such as the impact of turbidity in the estimation of Chl-a in aquatic systems (Demetriades-Shah et al., 1990; Becker et al., 2005). The derivatives of reflectance with different orders highlight the absorption and scattering characteristics of photosynthetic pigments. The spectral regions 630–645 nm, 660–670 nm, 680–687 nm, and 700– 735 nm were observed to be important wavelength ranges where the first derivatives can be applied for the estimation of Chl-a (Han, 2005). Equations (1) and (2) are used to estimate the first and second derivatives of spectral data (Becker et al., 2005).

$$d^{1st} = (\rho_{n+1} - \rho_n)/(\lambda_{n+1} - \lambda_n) \quad (1)$$

$$d^{2nd} = (d_{n+1}{}^{1st} - d_n{}^{1st})/(\lambda_{n+2} - \lambda^n) \quad (2)$$

where d^{1st} and d^{2nd} are first and second derivatives, n is the band number, ρ is the surface reflectance, and λ is the wavelength.

A major challenge to satellite based Chl-a monitoring is that most open source satellites have revisit flyover schedules that are not frequent enough to overlap with the erratic nature of algal blooming and dissipation. The 16-day revisit schedule of Landsat and the 5-day revisit schedule of Sentinel-2 may miss significant bloom events between flyover dates and times. These temporal gaps can be extended when cloud cover is high and impede the use of imagery acquired on schedule. This creates the problem of a potential mismatch between satellite and ground-based Chl-a data which makes it

challenging to apply the Central Limit Theorem (CLT) for algorithm development (McCullough et al., 2012, Boucher et al., 2018).

The CLT stipulates that the means of random sample size (n), from a population with mean, μ, and variance, σ^2, spread normally with mean (μ), and variance (σ^2 /n). The residuals of a regression model with goodness of fit should be normally distributed (McLeod, 2019). The residuals will be closer to the normality when larger data set ($n \geq 30$) is used. The sample mean and standard deviation will be closer in value with the population mean (μ) and standard deviation (σ) when the sample size is increased. It can also reduce the sampling errors in the dataset (McLeod, 2019).

Previous studies have paired satellite and ground-based data that were collected on different days to address the issue of limited data size for the development of models. Kloiber et al. (2001) determined the maximum acceptable temporal disparity between satellite overpass date and ground-based data collection for Secchi Disk Transparency (SDT) in lakes of the Twin Cities Metropolitan Area in Minnesota. The empirical relationships between Landsat-5 TM3/TM1 and SDT were recorded for an increasing time window from the same day to ±7 days between satellite image acquisition and ground-based sampling. A ±7-day time window means ground-based sampling was carried out seven days before or after the satellite acquired an image. They repeated the analysis by using a constant set of data (N=20) which was randomly selected for different time windows of 0-7 days between samplings. Kloiber et al. (2001) observed decreasing R^2 values and increasing standard error of estimate (SEE) with increasing time window between ground-based sampling and satellite overpass. They concluded that ±1 day was the idea time window for pairing satellite and ground-based SDT data. The study also

asserted that when SDT data were limited, a time window of ±7 days would allow a strong linear relationship between satellite and ground-based data especially if SDT did not significantly change during that period.

The Kloiber et al (2001) study has provided an important reference for determining acceptable temporal disparities in the collection of satellite and ground-based data (Olmanson et al., 2008; Brezonick et al., 2009; McCullough et al., 2012; Urbanski et al., 2016; Tan et al., 2017; Boucher et al., 2018). However, other studies have suggested an even longer period of up to a ten-day window between ground-based sampling and image acquisition as acceptable for pairing water quality data (Kloiber et al., 2001; Olmanson et al., 2008; Brezonick et al., 2009; McCullough et al., 2012; Urbanski et al., 2016; Tan et al., 2017; Boucher et al., 2018). The assumption is that water quality conditions will not significantly change over this time period as long as no major events such as flooding, severe weather, and limnetic mixing take place in the lake (Torbick and Corbiere, 2015; Kuhn et al., 2019).

Previous studies have utilized data from single (Brivio et al., 1999; Tebbs et al., 2013; Yip et al., 2014; ; Bonansea et al., 2015 Tan et al., 2017; Watanabe et al., 2017) and multiple (Kloiber et al., 2001; Chipman et al., 2004; Olmanson et al., 2007; Keith et al., 2018; Watanabe et al., 2019) water bodies to determine the number of days we can go before or after satellite overpass to develop predictive models. These studies have successfully illustrated the application of remote sensing in retrieving water quality parameters. Most of the studies (Kloiber et al., 2001; Olmanson et al., 2007; Mancino et al., 2009; Keith et al., 2018) that have used multiple waterbodies for predicting water quality parameters have used a specified time difference between satellite overpass and

ground-based sampling to obtain data for developing predictive models. The predictive models developed in those studies have illustrated strong relationships between spectral and ground-based water quality data. However, these studies have not discussed how the data from multiple water bodies or temporal disparity between spectral and ground-based data could cause variation in Chl-a estimation. This ambiguity makes it important to determine if other factors also contribute to the variations obtained in addition to the temporal disparities between satellite and ground-based data.

Studies that have used single as well as multiple water bodies to estimate Chl-a have observed factors such as weather conditions which includes windy days, sunny days, and precipitation events, lake water temperature (Tan et al.,2017), activities in watershed such as irrigation, recreation, industrialization (Mamun et al., 2021), nutrient concentrations (Watanabe et al., 2018) contribute to the variation in Chl-a. Mixing events, turbidity, Biological Oxygen Demand (BOD) are also important factors that could cause variations in Chl-a estimation (Watanabe et al., 2018). Some studies have also observed variation in Chl-a concentration across different zones of the reservoirs (Li et al., 2017). These factors are out of the scope of study.

The first objective of this research was to determine how many days before or after the satellite overpass, can we utilize the ground-based data for developing regression models and get good statistical results. The second objective was to compare the statistical results of Chl-a regression models using data from a single reservoir to those that were developed using data from multiple reservoirs within a specific time window between ground-based sampling and satellite acquisition.

The key significance of this study is to further understand the extent to which satellites may be useful to address the problems faced in ground-based for better management and monitoring of bloom events. Satellites provide synoptic observation of the Earth at low cost, and they have been widely used to study algal blooms in inland water bodies. The combination of satellite remote sensing with ground-based sampling can offset the limitations of ground-based monitoring such as problems related to accessibility, time consumption, requirement of human and economic resources.

CHAPTER II

MATERIALS AND METHODS

2.1 Study Area

Oklahoma has over 200 reservoirs that were constructed for flood control, water supply, recreational activities, wildlife management and protection, and hydroelectric power (Kenneth, 1998). These reservoirs are monitored under the Beneficial Use Monitoring Program (BUMP), which is run by the Oklahoma Water Resources Board (OWRB). The BUMP program monitors a total of 130 reservoirs on a quarterly basis and a five-year rotational routine. This study is focused on 10 of the BUMP reservoirs, which were selected because they had enough available Chl-a data. Most of these reservoirs have track records of cyanobacterial algal bloom events (Mason and Triplett, 2016; Boyer et al., 2017). Figure 1 shows a map of Oklahoma and the BUMP reservoirs including the study reservoirs in colors. The characteristics of study reservoirs are presented in Table 1.

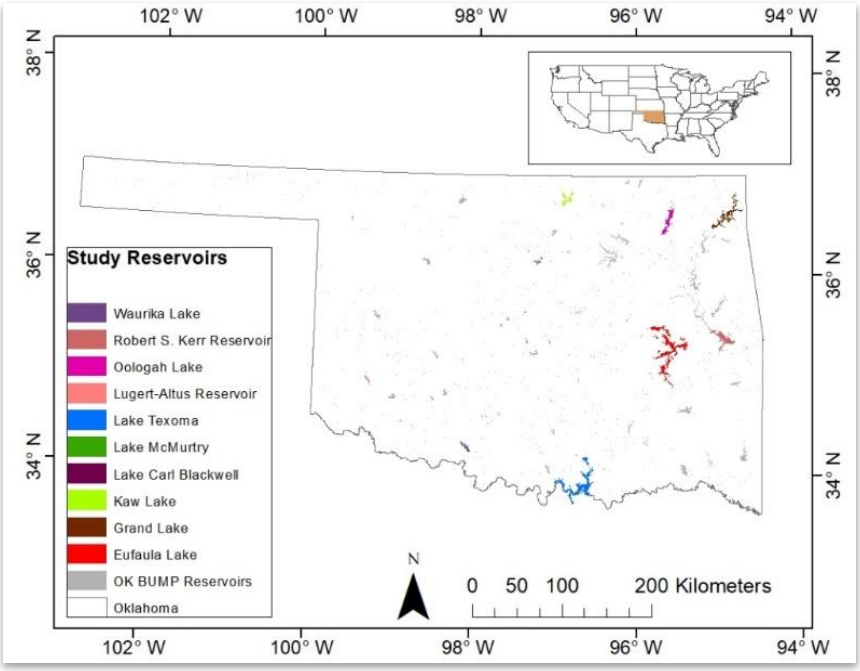

Figure 1 Map of Oklahoma showing the BUMP reservoirs across the state and the 10 study reservoirs used in this study.

Table 1 Characteristics of study reservoirs acquired from BUMP Lake Data, OWRB 2018

S.N	**Name of Lake**	**Trophic Class**
1.	Lake Carl Blackwell(4 sample sites)	Eutrophic
2.	Eufaula Lake (17 sample sites)	Mesotrophic- Eutrophic
3.	Grand Lake (13 sample sites)	Eutrophic
4.	Kaw Lake (5 sample sites)	Hypereutrophic
5.	Lugert-Altus (4 sample sites)	Eutrophic
6.	Lake McMurtry (3 sample sites)	Eutrophic
7.	Oologah (7 sample sites)	Mesotrophic
8.	Robert S. Kerr Reservoir (6 sample sites)	Eutrophic
9.	Texoma (13 sample sites	Eutrophic -Hypereutrophic
10.	Waurika Lake (5 sample sites)	Eutrophic

2.2 Water quality data

Chlorophyll-a data for the 10 study reservoirs were obtained from the OWRB's BUMP with data collected from 2006 to 2020. The Chl-a samples were collected below the surface at a depth of 0.5 m and preserved on ice. They were taken to the laboratory for filtration, extraction, and measurement of Chl-a. Chl-a analysis was carried out using the American Public Health Association (APHA) 10200-H guidelines (OWRB, 2018).

2.3 Image acquisition and processing

Landsat-5 and Landsat-8 data were downloaded from the U.S Geological Survey Earth Explorer website. The data were subjected to geometric, radiometric, and atmospheric corrections using the Land Surface Reflectance Code (LaSRC) algorithm (Version 1.5.0) for Landsat 8 and Landsat Ecosystem Disturbance Adaptive Processing System (LEDAPS) algorithm (Version 3.4.0) for Landsat-5 products. The surface reflectance images were stored with a scaling factor of 10,000 and made available at the Earth Explorer website for open-source use.

Landsat has a temporal resolution of 16 days and a spatial resolution of 30 m. The Landsat-5 Thematic Mapper (TM) has an image tile of 170 *185 km^2 , a radiometric resolution of 8-bits, and seven spectral bands. The Landsat-8 Operational Land ImaBanger (OLI) has an image tile of 185*180 km^2, a radiometric resolution of 16-bits, and eleven spectral bands. The U.S Landsat products are generated in the Albers Equal Area (AEA) Conic map projection. They possess a WGS84 datum and are provided in Georeferenced Tagged Image File Format (GeoTIFF)(Dwyer et al., 2018).

Sentinel-2 Level 2A images were obtained from the official website of Sentinel Hub EO Browser. The Level 2A images were generated through the Payload Data Ground Segment (PDGS) utilizing the Sen2Cor processor and made available for open-source use. The Level 2A images are atmospherically corrected Bottom-of-Atmosphere (BOA) products. Each Level-2A product constitutes 100*100 km^2 tiles in cartographic geometry. It has a revisit time of 5 days. Sentinel-2 has 13 spectral bands, three possible spatial resolutions of 10m (Band 2, Band 3, Band 4, and Band 8), 20m (Band 5, Band 6, Band 7, Band 8a, Band 11, and Band 12), and 60m (Band 1, Band 9, and Band 10), and a radiometric resolution of 12-bit. The image tiles are projected in UTM/WGS84 projection. The Sentinel-2 image format used for the study was a 32-bit float with *.tiff* extension.

Images with cloud cover less than 10% or those with no cloud cover at the sample sites were selected for spectral data extraction. The downloaded satellite images were imported to ArcGIS 10.8 and pixel values for the sampling points were extracted using the extraction tool in the Spatial analyst toolset. The pixel values obtained for each of the sampling points were divided by the scaling factor (10,000) that was used for image storage in Landsat-5 and Landsat-8. The Sentinel-2 32-bit float pixel values were stored as surface reflectance values ranging from 0-1. Appendix II shows details of the dates of image acquisition and ground-based sampling for the study reservoir for Landsat-5, Landsat-8, and Sentinel-2.

2.4 Bands, Band Combinations, and Spectral Derivatives

This study used band combinations and selections following Mansaray et al. (2021). Different spectral bands and band combinations of Landsat-5 TM, Landsat-8

OLI, and Sentinel-2 were utilized for the development of empirical Chl-a models. A total of 27, 30, and 58 bands, band ratios, and spectral derivatives were used for Landsat-5, Landsat-8, and Sentinel-2, respectively.

The spectral derivative technique was utilized to obtain first and second order derivatives for the respective spectral bands of Landsat-5, Landsat-8, and Sentinel-2. These derivatives were used as independent variables for the regression analysis along with bands, bands combinations, and ratios. The first derivative denoted by (d_n) was calculated using Eq (1) (Becker et al., 2005).

$$d_n = (\rho_{n+1} - \rho_n) / (\lambda_{n+1} - \lambda_{n)} \quad \text{Eq. (1)}$$

where n= band number, λ= center wavelength, ρ= surface reflectance. The second derivative was designated as ($d_{2n)}$ and calculated using Eq (2) (Becker et al., 2005).

$$d_{2n} = (d_{n+1} - d_n)/0.5*(\lambda_{n+2} - \lambda_n) \quad \text{Eq. (2)}$$

2.5 Best-fit model development and validation for varying time windows

Multiple linear regression analysis was carried out for the two factors considered in the study: time window and single vs multiple reservoirs. This study utilized data for time windows ranging from zero to ±8 days. The time window in this study is defined as the temporal gap between the date of satellite overpass and the date of ground-based Chl-a sample visit. The time window was recorded as zero day (0-day) when sampling took place on the same day as a satellite overpass. This time window was considered to be temporally coincident. The time window was recorded as 1-day (±1 day of satellite overpass) when sampling was done one day before or after satellite overpass. An eight-day time lapse of ground-based sampling before or after satellite overpass gave an 8 -day

time window (±8 days of satellite overpass). The cloud-free satellite images closest to the sampling date was considered for pairing when multiple images were available.

This study developed regression models for all the nine time windows (0, ±1, ±2, ±3, ±4, ±5, ±6, ±7, and ±8 days). Studies have found that a time-window of up to ten days between satellite image acquisition and ground-based sampling could be used to develop regression models (Olmanson et al., 2007). The independent variables were the spectral bands, band combinations, and spectral derivatives. The dependent variables were the measured concentrations of Chl-a, the natural logarithms of those concentrations (LN(Chl-a)), and their square roots (SQRT(Chl-a)). It is a common practice in remote sensing to use the power transformations of the measured concentration for model development (Tan et al., 2017; Mansaray et al., 2021). Increasing the time window gave more data points because the data from the previous time window become part of the data for the successive time windows as illustrated in the blue boxes in Table 2.

Table 2 Increasing data size per time window with the blue boxes representing the data used to develop the nine regression models

Serial numbers of Regression Models	Time window for which data were included								
	0	±1	±2	±3	±4	±5	±6	±7	±8
1									
2									
3									
4									
5									
6									
7									
8									
9									

Regression analysis was done using the stepwise selection of terms through the Forward Information Criteria (FIC). The FIC involves a step-by-step addition of terms to

the regression model. It evaluates how the addition of terms best explains the variation in the dependent variable. The model that gave the best statistics in terms of the coefficient of determination ($R^2 \geq 0.5$), significance of the relationship at a 0.05 significance level (p-value$\leq$0.05), multi-collinearity of the independent variables (Variance Inflation Factor (VIF) <10), minimum number of data points ($N \geq 30$) and low root mean square error (RMSE) relative to the concentration range was considered as the best-fit model.

The train-test and K-folds validation methods were used to test the best-fit models for each dataset. The data were divided into a training and a testing set in the train-test validation. Using the training test, the best-fit model was reproduced and then used to calculate Chl-a concentrations in the testing set. The model-derived Chl-a concentrations were regressed against the measured Chl-a concentrations of the testing set and the R^2 was recorded.

The best-fit model was also validated using the K-folds Cross Validation method. The original sample data was divided randomly into K subsamples. This study divided the data into 10 subsamples. The value of K was chosen as 10 as this is a typical choice (Kuhn et al., 2013; Kuppssinsku et al., 2020). Minitab 20 attempted to reproduce the best-fit model using the training set and then tested it using th.e testing set for each subsample. The R^2 and the RMSE were recorded for each round as shown below.

- Model-1: Trained on Fold-2 + Fold-3 +…. Fold-10; Test on Fold-1; Record R^2 & RMSE
- Model-2: Trained on Fold-1 + Fold-3 +…. Fold-10; Test on Fold-2; Record R^2 & RMSE

- Model-n: Trained on Fold-1 + Fold-2 + Fold-n +…. Fold-10; Test on Fold-n; Record R^2 & RMSE (n = a value between 4 and 9)
- Model-10: Trained on Fold-1 + Fold-3 +…. Fold-9; Test on Fold-10; Record R^2 & RMSE

All the 10 models were expected to be equivalent to the best-fit model and the R^2 and RMSE were used to evaluate this expectation. The procedure is summarized in Figure 2.

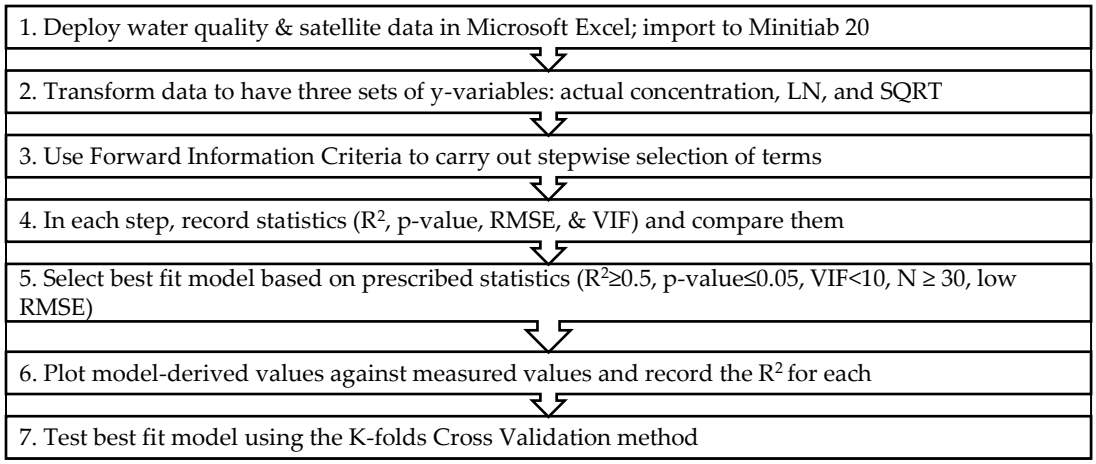

Figure 2 Flow chart for development of best-fit model

2.6 Best-fit model development and validation for single versus multiple reservoirs

This study compared the statistics of regression models with a single reservoir to those with multiple reservoirs in order to delineate the effects of multiple reservoirs on the performance of the three satellites for Chl-a detection. The same number of data points was used to calculate the regression statistics of each reservoir. The obtained statistics were then averaged to give the representative statistics for a single reservoir. The largest data size for Landsat-5 was 33 data points. Hence, all the reservoirs with at least 30 data points were selected and the sample size, N =30, used for the single

reservoir analysis. Landsat-8 had reservoirs with up to 48 data points allowing the sample size of 40 for the single reservoir analysis. The Sentinel-2 sample size, N= 30, was used for the single reservoir analysis. Random samples were selected without replacement until the required sample size (N = 30, 40, and 30 for Landsat-, Landsat-8, and Sentinel-2, respectively) was attained from each of the reservoir. All the data were selected for the analysis for a reservoir that had exactly the required sample size. Multiple regression models were developed using the stepwise FIC and the best-fit model was the one with the highest R^2, lowest RMSE, lowest VIF, and significant p-value. The average best-fit R^2 and RMSE were recorded as the representative statistics for a single reservoir.

The same sample sizes used in the single reservoir analysis were used for multiple reservoir analysis. An equal number of subsamples (n) were randomly drawn without replacement from the data for each of the ten study reservoirs (n = 3, 4, and 3, for Landsat-5, Landsat-8, and Sentinel-2) to give the total required sample size (N = 30, 40, and 30 for Landsat-5, Landsat-8, and Sentinel-2, respectively). Multiple regression models were then developed using the stepwise FIC and this was followed by selection of the best-fit model based on the R^2, RMSE, VIF, and p-value. This process was repeated multiple times and the average R^2 and RMSE were recorded as the representative statistics for multiple reservoirs. Figure 3 presents the steps followed in the single versus multiple reservoir analysis.

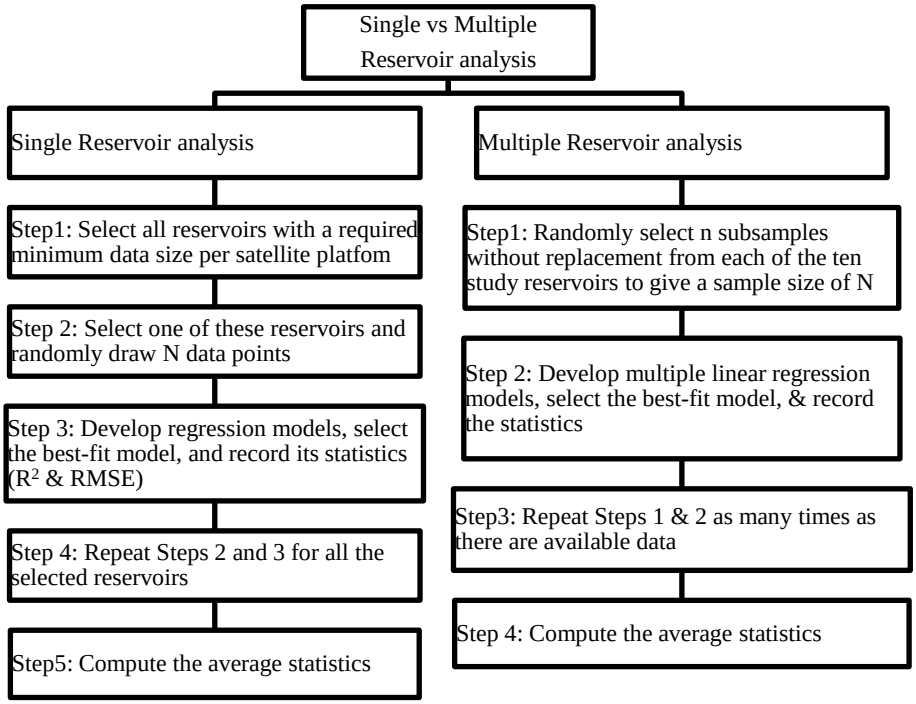

Figure 3 Flow chart showing the procedure for single and multiple reservoir analysis used for the study

CHAPTER III

RESULTS

3.1 Data size per time window

A total of 1,371 datapoints were available between 2006 and 2020 (14 years) of ground-based Chl-a data collected in the ten study reservoirs. Landsat-5 had 15 data points that were collected on the same day as the ground-based data, Landsat-8 had 24, and Sentinel-2 had 20 during this period. These were too few data points to pair with the Chl-a data to give reliable regression models (McCullough et al., 2012, Boucher et al., 2018). This observation confirmed the need to expand the time window beyond the same date of image acquisition and ground-based sampling to allow the use of enough data points**.** The number of data points and Chl-a range (µg/l) for each time window (0 to ± 8 days) for Landsat-5, Landsat-8, and Sentinel-2 is presented in Table 3.

Table 3 Number of data points associated with each of the increasing time window (0 to ± 8days) along with the Chl-a concentration range in (µg/l) for Landsat-5, Landsat-8, and Sentinel-2.

			Time window							
Platform	**Attribute**	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Landsat-5	Data size (N)	15	37	84	101	139	168	179	191	197
	Chl-a conc. range (µg/l)	1.32-34	1.32-34	1.32-48	1.32-48	1.32-48	1.32-58.5	1.32-58.5	1.32-58.5	1.32-58.5
Landsat-8	Data size (N)	24	73	110	181	235	304	329	408	432
	Chl-a conc. range (µg/l)	0.64-21.6	0.64-45.30	0.64-101	0.64-101	0.62-101	0.62-101	0.62-101	0.62-101	0.62-101
Sentinel-2	Data size (N)	20	49	67	124	195	237	259	291	333
	Chl-a conc. range (µg/l)	1.37-101	1.32-101	0.64-101	0.64-106	0.64-106	0.64-106	0.64-106	0.64-106	0.64-106

3.2 Best-fit model per satellite

Table 4 presents the statistical results (R^2, RMSE, 10-fold CV R^2, and 10-fold CV RMSE) of the best-fit regression model for each satellite platforms. Details of the regression models for all the time windows are shown in Appendix I.

Table 4 Statistics (R^2, RMSE, 10-fold CV R^2, and 10-fold CV RMSE) of the best-fit model per satellite sensor.

Platform	Time window	Model (s)	R^2	RMSE	10-Folds CV	
					R^2	RMSE
Landsat-5	0*	Sqrt(Chl-a) = 5.957 - 13.93* d_1	0.55	1.04	0.44	1.10
Landsat-8	1	Sqrt(Chl-a)=1.840 + 48.78 d_{21} + 10.62 d_{23} + 21.57 d_{25}	0.60	0.88	0.56	0.94
Sentinel-2	8	Sqrt (Chl-a) = 6.075 - 2.670 *(Red/RE1) - 6.807 *d_6 + 18.80* d_{23}	0.83	1.52	0.80	1.55

Note: For Landsat-5, $d_1 = (\rho_G - \rho_B) / (\lambda_G - \lambda_B)$; For Landsat-8, $d_{21} = (d_B - d_{CA})/0.5* (\lambda_G - \lambda_{CA})$, $d_{23} = (d_R - d_G)/0.5* (\lambda_{NIR} - \lambda_G)$, $d_{25} = (d_{SWIR1} - d_{NIR})/0.5* (\lambda_{SWIR2} - \lambda_{NIR})$; For Sentinel-2, $d_6 = (\rho_{RE3} - \rho_{RE2}) / (\lambda_{RE2} - \lambda_{RE3})$, $d_{23} = (d_R - d_G)/0.5* (\lambda_{RE1} - \lambda_G)$ where ρ is the surface reflectance and λ is the center wavelength.

The best-fit model for each satellite fulfilled all these criteria: $R^2 \geq 0.5$, low RMSE values, VIF< 10, $p < 0.05$, Minimum data points (N) ≥ 30. For Landsat-5, all the criteria were fulfilled except for the requirement of $N \geq 30$. The R^2 decreased when the number of data points were increased for Landsat-5.

Figure 4 presents scatter plots showing relationships between the values of Chl-a (µg/L) predicted by the Landsat- 5 (L5), Landsat-8 (L8), and Sentinel-2 (S2) spectra and those measured in the study reservoirs. The R^2 values are displayed at the top of each graph along with the satellite platforms for developing the predictive model . The x-axis represents the measured values, and the y-axis represents the fitted values. All three satellites showed strong relationships between measured and fitted values of Chl-a.

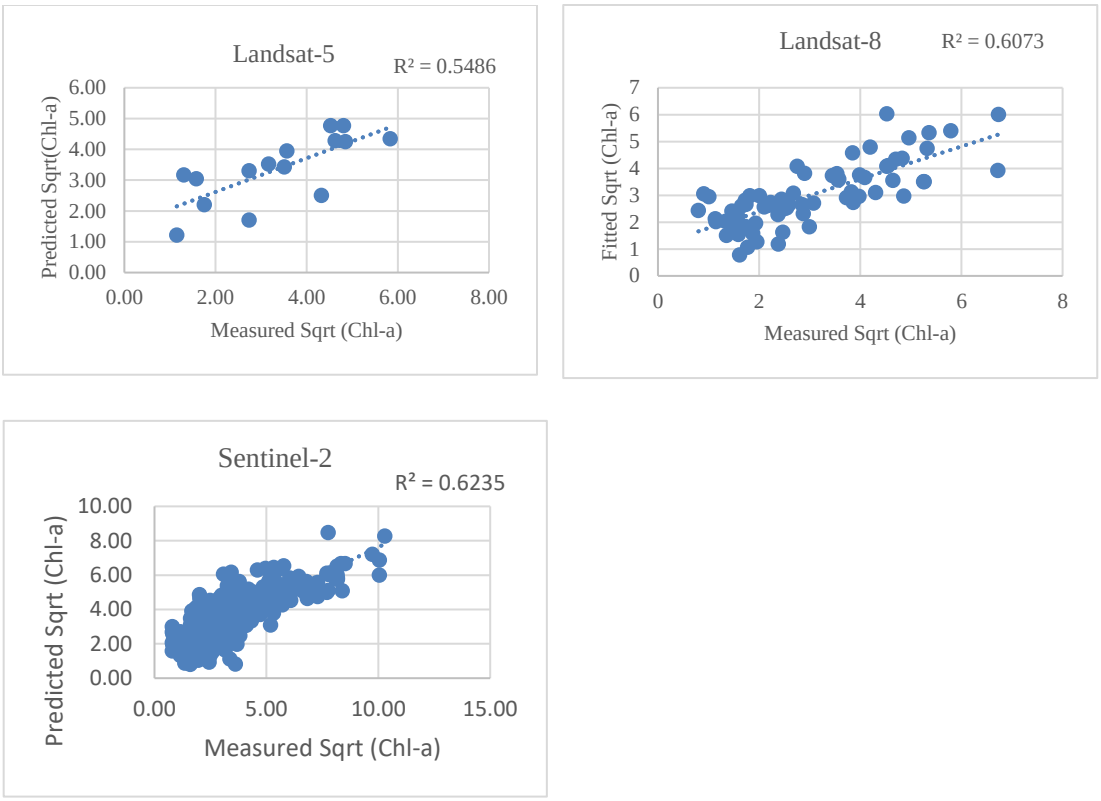

Figure 4 Scatter plots showing relationships between the model-derived concentrations and measured concentrations of Chl-a (µg/l) with Landsat-5 (L5), Landsat-8 (L8), and Senintel-2 (S2). The R^2 values are presented at the top of each graph with the satellite platform.

3.2.1 Comparison of different time windows (R^2)

Figure 5 shows the relationship between the coefficient of determination (R^2) and increasing time window for the three satellites. The blue curve represents the trend in R^2 for Landsat-5, the green and red represent those for Landsat-8 and Sentinel-2, respectively. The x-axis represents the number of sampling days before or after satellite image acquisition, and the y-axis presents the R^2.

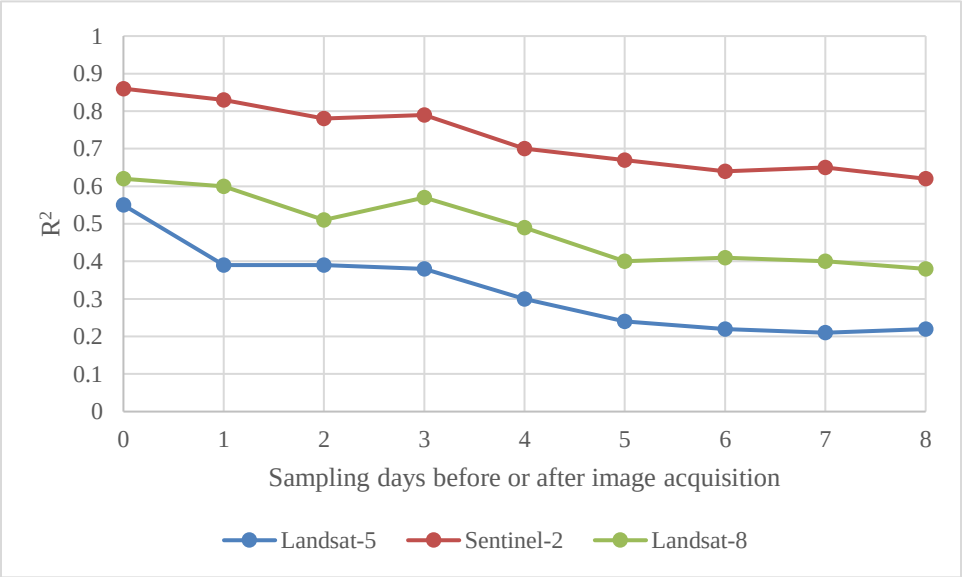

Figure 5 Graph of R^2 for the different time windows with Landsat-5, Landsat-8, and Sentinel-2

For all the three satellites, highest R^2 values were observed for the temporally coincident paired data. Generally, the R^2 values dropped as the time window increased. The strength of relation between the satellite spectra and ground-based data decreased as the temporal disparity increased. Sentinel-2 gave the highest R^2 values, followed by Landsat-8 and then Landsat-5 for all the time windows. Only Sentinel-2 gave $R^2 \geq 0.5$ in all nine time windows. Sentinel-2 provided stronger relationship between the satellite and ground-based data for each time window up to 8 days.

3.2.2 Comparison of different time windows (RMSE)

Figure 6 presents the RMSE values (µg/l) for the multiple regression analysis of Chl-a and the satellite spectra for Landsat-5, Landsat-8, and Sentinel-2 with increasing time window. The RMSE values obtained for all the three satellites were low (Chl-a <5 µg/l).

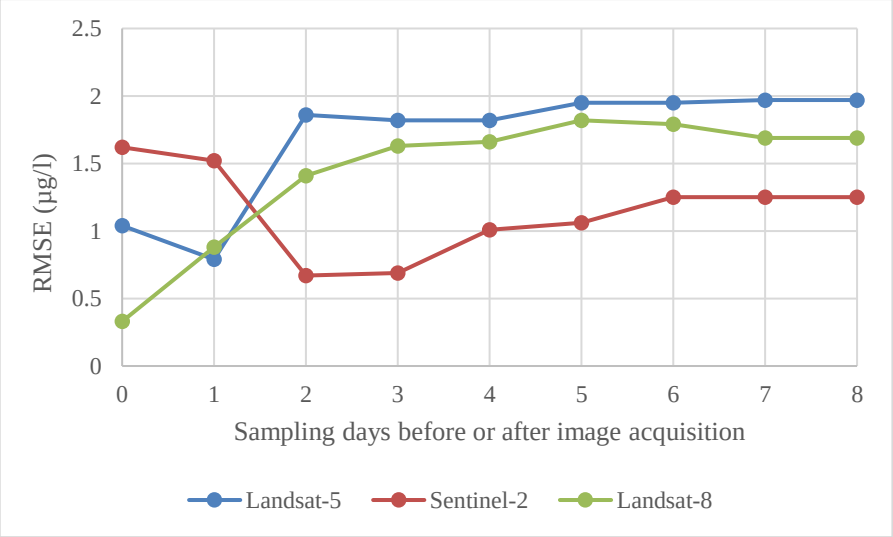

Figure 6 Graph of RMSE for the different time windows with Landsat-5, Landsat-8, and Sentinel-2.

Comparatively higher RMSE values were observed from ± 2 days to ±8 days for Landsat-5 as compared to other satellite platforms. The lowest RMSE value was observed for same day in the case of Landsat-8. The RMSE values increased after ± 1 day. After ± 3 days, the RMSE was constant till ±8 days. Sentinel-2 had the lowest RMSE values compared to the other two satellites with the increasing time window. The low RMSE values for Sentinel-2 throughout the time window from same day to ± 8 days. These results imply that Sentinel-2 regression models can better predict Chl-a when a temporal window of ± 8 days is considered.

3.2.3 Validation of the best-fit models per time window

Figure 7 shows the R^2 values obtained in the 10-folds cross-validation (CV) of the best-fit models for Landsat-5, Landsat-8, and Sentinel-2 with the increasing time window. For all three satellites, CV R^2 values dropped when the time window was increased from 0 to ± 8 days. Cross-validation results show that Sentinel-2 has the highest value (R^2=0.82) followed by Landsat-8 (R^2 = 0.46) and Landsat-5 (R^2 = 0.44).

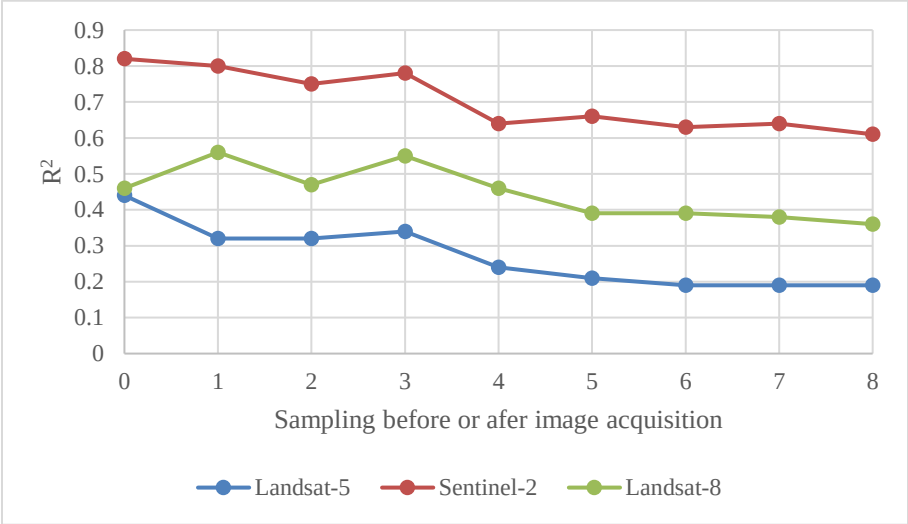

Figure 7 Graph of cross-validation R^2 versus time window for Landsat-5, Landsat-8, and Sentinel-2

For Landsat-5, only the temporally coincident days between satellite overpass and ground-data gave $R^2 \geq 0.5$ as compared to the other time windows. In the case of Landsat-8, results show that for ± 3 days of temporal disparity, Landsat-8 models can be used to predict Chl-a in reservoirs. Sentinel-2 models gave $R^2 \geq 0.5$ for all the time windows compared to Landsat-5 and Landsat-8. Based on the Sentinel-2 R^2 values, up to 8 days of time window is acceptable for pairing satellite and ground-based data.

Figure 8 presents the RMSE values obtained in the 10-folds CV of the best-fit models for Landsat-5, Landsat-8, and Sentinel-2 with the increasing time window. The RMSE values were low for all the sensors (RMSE< 5 µg/l).

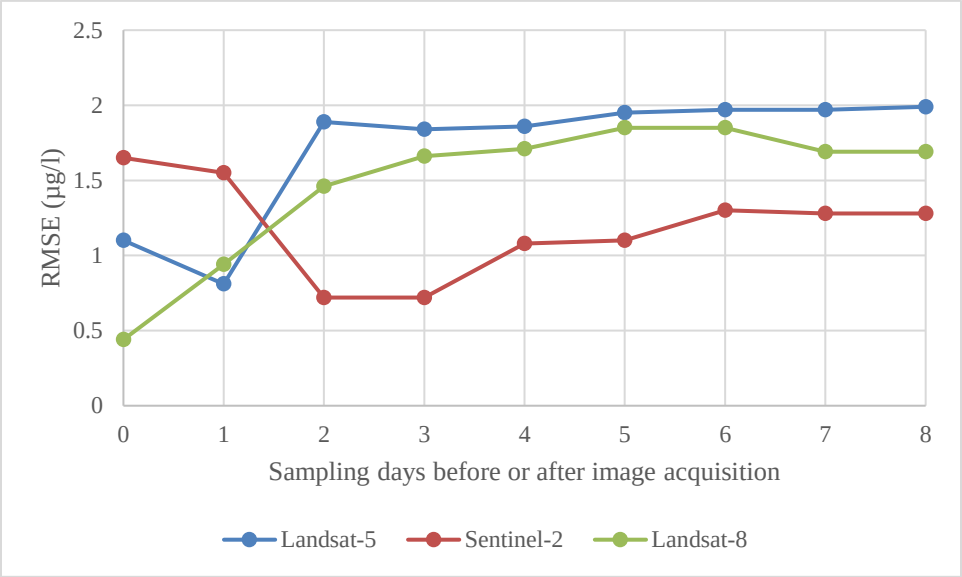

Figure 8 Graph of cross-validation RMSE versus time window for three satellite sensors

The RMSE values for Landsat-5 was lowest for ± 1 day of temporal disparity. The RMSE values increased from ± 2 days and remained constant after ± 5 days. For Landsat-8, lowest RMSE value was observed for temporally coincident day. The RMSE values followed an increasing trend after ± 1 day. The lowest RMSE values were observed for Sentinel-2 from ± 2 to ± 8 days compared to Landsat-5 and Landsat-8. Landsat-5's RMSE values were high as compared to Landsat-8 and Sentinel-2. The cross-validation RMSE results show that Sentinel-2 can be used to predict Chl-a when the temporal disparity between satellite overpass and ground-based sampling is ± 8 days.

3.3 Best-fit model development for single versus multiple reservoir

Table 5 presents the R^2, RMSE, average R^2, and average RMSE obtained for single reservoir using same number of data points (N=30, 40, and 30) and for Landsat-5, Landsat-8, and Sentinel-2, respectively for each reservoir. Table 6 presents the regression statistics (R^2, RMSE, average R^2 and average RMSE) obtained after repeating the regression analysis for three times using the same number of data points from multiple reservoirs as it was used for single reservoir for Landsat-5, Landsat-8, and Sentinel-2.` For this multiple regression analysis, spectral and ground-based Chl-a data obtained within a constant number of days ± 4 between satellite image acquisition and ground-based sampling were taken.

Table 5 Statistics RMSE, average R^2, average RMSE) for single reservoirs with their respective data points used for the regression analysis with constant number of days (Days = ± 4) between satellite image acquisition and ground-based sampling for Landsat-5, Landsat-8, and Sentinel-2.

Platform	**Days (±)**	**Single Reservoir**	**No. of data points**	**R^2**	**RMSE**	**Average R^2**	**Average RMSE (µg/l)**
Landsat-5	4	Lake Texoma	30	0.40	1.46	0.50	0.91
		Eufaula Lake	30	0.60	0.36		
Landsat-8	4	Kaw Lake	40	0.50	2.51	0.46	1.98
		Oologah Lake	40	0.42	1.44		
Sentinel-2	4	Grand Lake	30	0.75	0.50	0.81	0.44
		Oologah Lake	30	0.87	0.37		

Data points from Lake Texoma and Eufaula Lake were considered for single reservoir analysis using Landsat-5. The averaged R^2 and RMSE obtained for those two reservoirs were 0.50 and 0.91 µg/l. Kaw Lake and Oologah Lake were used for the single reservoir analysis for Landsat-8. The average R^2 and RMSE obtained were 0.46 and 1.98

µg/l. For Sentinel-2, data from Grand Lake and Oologah Lake were utilized. The average R^2 and RMSE observed were 0.81 and 0.44 µg/l. Sentinel-2 gave the highest average R^2 (R^2 = 0.81), followed by Landsat-5 (R^2 = 0.50) and Landsat-8 (R^2 = 0.46). Sentinel-2 provided the lowest average RMSE value (0.44 µg/l), followed by Landsat-5 (0.91 µg/l) and Landsat-8 (1.98 µg/l).

When carrying out the analysis for single reservoir in Minitab 20, particularly for Oologah lake using Landsat-8, very high VIF was observed (VIF >50). The multicollinearity observed between the independent variables was too high although R^2 > 0.5 was observed. The response variable Chl-a gave a R^2 ($R^2 \geq 0.5$) but VIF was greater than 80. The natural logarithm of Chl-a provided high to low VIF, however the R^2 associated with low VIF (VIF<10) was low and for high VIF (VIF > 400), the R^2 was high. The square root of Chl-a comparatively gave VIF around 10 and R^2 of 0.42, so that model was selected. The reason behind high multicollinearity between the independent variable is still not clear. This observation lowered the average R^2 for Landsat-8.

Table 6 Statistics (R^2, RMSE, average R^2 and average RMSE) obtained after repeating the regression analysis for three times using the same number of data points from multiple reservoirs with constant number of days (Days = ± 4) between satellite image acquisition and ground-based sampling for Landsat-5, Landsat-8, and Sentinel-2.`

Platform	**Days (±)**	**Regression analysis for multiple reservoirs**	**R^2**	**RMSE**	**Average R^2**	**Average RMSE (µg/l)**
Landsat-5	4	Run 1	0.38	1.14	0.34	1.62
		Run 2	0.41	1.90		
		Run 3	0.22	1.82		
Landsat-8	4	Run 1	0.65	2.10	0.62	1.81
		Run 2	0.62	1.42		
		Run 3	0.60	1.90		
Sentinel-2	4	Run 1	0.67	0.83	0.72	0.88
		Run 2	0.76	0.86		
		Run 3	0.72	0.94		

The average R^2 and RMSE values obtained after running the analysis for three times using data from multiple reservoirs were 0.34 and 1.6 µg/l for Landsat-5. For each run, Landsat-5 gave a low R^2 ($R^2 \leq 0.5$). Landsat-8 gave an average R^2 and RMSE values of 0.62 and 1.81 µg/l. Landsat-8 gave a $R^2 \geq 0.5$ for all the repeated analysis. Sentinel-2 gave an average R^2 and RMSE values of 0.72 and 0.88 µg/l. Each repeated analysis for Sentinel-2 using data from multiple reservoirs gave a $R^2 \geq 0.5$. The average R^2 for Sentinel-2 was the highest ($R^2 = 0.72$), followed by Landsat-8 ($R^2 = 0.62$) and Landsat-5 ($R^2 = 0.34$). The average RMSE values were observed to be lowest for Sentinel-2 (0.88 µg/l), followed by Landsat-5 (1.62 µg/l) and Landsat-8 (1.81 µg/l).

Landsat-5 exhibited poor performance when data from multiple reservoirs were used. The R^2 values dropped from (0.50-0.34). The performance of Landsat-8 was better when multiple reservoirs were used. An increase in R^2 value from 0.46 to 0.62 was observed for Landsat-8. Sentinel-2 showed good results in both the cases of single and

multiple reservoirs with high R^2 ($R^2 \geq 0.5$) values and low RMSE values. Sentinel-2 can predict Chl-a accurately regardless of the number of reservoirs being utilized.

CHAPTER IV

DISCUSSIONS

4.1 Time Window

Algal blooms have adverse impacts on the health and well-being of human, animal, and aquatic ecosystems. Chlorophyll-a present in major groups of algae is an indicator of algal blooms in inland water bodies (Topp et al., 2019). Those blooms are spatially and temporally dynamic in nature. Ground-based sampling alone may not be enough to monitor those blooms. Satellites remote sensing have been used to monitor algal blooms in water bodies. One of the challenges of satellite remote sensing is that it is difficult to capture the blooms due to the temporal resolution of satellites.

Studies have suggested specific number of dates that sampling can be carried out before or after satellite overpass to obtain images for developing empirical relationships between spectral and ground-based data. For this study we used a temporal gap of same day to eight days before or after satellite image acquisition to develop regression models for the estimation of Chl-a for each satellite sensors. Single bands, band ratios, and spectral derivatives were regressed against ground-based Chl-a data to obtain those regression models. The R^2 values obtained for the regression models for Landsat-5,

Landsat-8, and Sentinel-2 developed per time window followed a decreasing pattern. This result was consistent with the study carried out by Kloiber et al. (2001).

The strength of correlation between satellite and ground-based data was observed to be strong only for the temporally coincident date for Landsat-5. A decrease in the strength of models to accurately predict Chl-a was observed as the time window was expanded. Olmanson et al. (2007) carried out a study to develop a database for water clarity of Minnesota's lakes. They used a time window of up to ± 10 days when data points were sparse. They concluded that the larger number of ground observations with longer time window increased the R^2 values. In their study, when the number of data points were increased from 12 to 26 for ± 1 to ± 7 days' time window, the R^2 increased from 0.85 to 0.88. This result was opposite to what we achieved in our study. When the data points were increased with the increased time window from 0 to ± 8 days, the R^2 values decreased from 0.55-0.22. The RMSE values also increased from 1.04 – 1.97 µg/l with the increasing data points and expanded time window. The conclusion provided by Olmanson et al., 2007 might not be applicable under all scenarios as they carried out the study for water clarity and it usually does not exhibit large and rapid fluctuations. Landsat-5 exhibited poor performance as compared to Landsat-8 and Sentinel-2 in Chl-a sensing. Landsat-5 has a broad band spectrum and lacks wavelengths significant to the detection of Chl-a (Chipman et al., 2004). From the observation of our study and dataset, we conclude that Landsat-5 is not a good choice to use for Chl-a sensing in reservoirs when longer time window between satellite and ground-based data is considered.

The regression models developed for Landsat-8 provided strong relationship when up to ± 1 day of temporal disparity between satellite and ground-based data was

considered. Landsat-8 models did not provide statistically good results when the time window was increased beyond ± 1 day. Landsat-8 performed comparatively better than Landsat-5. Landsat-8 has two SWIR bands that contribute significantly to the Chl-a reflectance in the study reservoirs absent in Landsat-5. This result is consistent with the results obtained in the previous studies (Mansaray et al., 2021). The two SWIR bands present in Landsat-8 are less prone to be affected by turbidity (Kuhn et al., 2019). Similar to the case of Landsat-5, the strength of relationship decreased when the time window was increased. This finding was also opposite to the study carried out by Olmanson et al. (2007). The increase in data points did not improve the R^2 values for the regression models.

The Sentinel-2 regression models can reliably predict Chl-a up to ± 8 days of temporal disparity between satellite and ground-based data. The presence of specific bands sensitive to Chl-a estimation in Sentinel-2 gives it an advantage over Landsat-8 and Landsat-5. The Red-Edge band present in Sentinel-2 is significant as algae cause a peak reflectance near 700 nm and is less susceptible to absorption and scattering (Bramich et al., 2021). Wavelengths in the RE region have a strong linear relationship with Chl-a reflectance (Gitelson et al., 2009). The two SWIR bands also present in Sentinel-2 are important predictors of Chl-a in water bodies. The first derivatives of Red and RE bands also improve the estimation of Chl-a in inland waters (Han, 2005). Sentinel-2 is the best choice among Landsat-5 and Lansdat-8 when larger time window is considered. Ground-based data before or after 8 days of Sentinel-2 satellite overpass can be used when data points are sparse and there is a need to increase the time window between ground-based sampling and satellite image acquisition.

Overall, all three sensors can estimate Chl-a in the study reservoirs. Sentinel-2 outperformed Landsat-5 and Landsat-8 in the detection of Chl-a throughout the incremental time window. It demonstrates the potential applicability of Sentinel-2 in remote of Chl-a in reservoirs over a longer time window. Sentinel-2 can be effectively used in the management of algal blooms and make it easier for the lake managers and responsible authorities to take necessary strategic actions to reduce its adverse impacts on the environment.

4.2 Single Versus Multiple Reservoirs

Ground-based data from single as well as multiple reservoirs were paired up with spectral data to develop predictive models to estimate Chl-a. The comparison of regression statistics that we obtained using single and multiple reservoirs showed varying results for the three satellites. We have observed that the spectral resolution of the satellite platforms plays a role in the estimation of Chl-a.

Landsat-5 showed good performance when data points from single reservoirs were used. This result was consistent with the study carried out by Tan et al. (2017) to estimate Chl-a in Erhai lake. When data points were used from multiple reservoirs, decrease in the R^2 value was observed. Poor performance was exhibited by Landsat-5 because of its broad band spectrum that lacks wavelengths significant to the detection of Chl-a (Chipman et al., 2004). Our result suggests that Landsat-5 cannot reliably predict Chl-a when data from multiple water bodies are used.

Landsat-8 provided strong relationship between ground-based and spectral data from multiple reservoirs were used to develop predictive models. The result was

consistent with a study carried out by Markogianni et al. (2020) in 50 natural/ artificial lakes of Greece. Landsat-8 also has spectral bands in the SWIR region that is important in the detection of Chl-a in water bodies. Landsat-8 had comparatively lower average R^2 for single reservoir. Oologah Lake, that was used for the single reservoir analysis of Landsat-8 exhibited high multicollinearity among the independent variables, that contributed to lower average R^2 for Landsat-8. The reason behind high multicollinearity among the independent variables is not clear and requires further analysis.

Sentinel-2 had high R^2 and low RMSE for both the single and multiple reservoirs. Sentinel-2 has RE bands absent in Landsat-5 and Landsat-8. SWIR bands not available in Landsat-5. The wavelengths in those regions play significant role in Chl-a detection (Bramich et al., 2005). In addition to that, RE and SWIR bands are less susceptible to absorption and scattering in turbid water bodies (Kuhn et al., 2019).

Other important factors that we have not considered in the study are the hydroclimatic, anthropogenic, and biophysical processes that takes place in those reservoirs. The ten study reservoirs range from eutrophic to hypereutrophic trophic status and have different watersheds. Even the zones of those study reservoirs (riverine, transition, and lacustrine) vary in terms of trophic state (OWRB, 2018). Li et al., 2017 concluded in their study that the concentration of Chl-a varies within the different zones of the reservoirs. It is usually higher in the riverine zone due to nutrient input from the watershed's runoff. The Chl-a concentration varies from site by site as well, because of anthropogenic activities such as intensive agricultural practices, industrialization, fishing, recreation, and irrigation in the surrounding water shed (Mamun et al., 2021).

We also observed varying range of Chl-a concentrations for different sample sites of the same reservoir in our study. The study reservoirs have different watersheds with different hydroclimatic and anthropogenic factors governing the water quality of the reservoirs. They undergo different biophysical conditions such as nutrient exchange rates, limnetic mixing, turbidity, change in lake water temperature, weather and wind conditions, biological oxygen demand (BOD) at different rates. During high wind conditions, the predicted Chl-a concentrations are usually low as they get mixed throughout the water column, providing unreliable estimations (Tan et al., 2017). These conditions of the reservoirs are important to understand to develop better predictive models using both the single as well as multiple reservoirs.

The result of this study suggests that the number of reservoirs do not matter as much as the time window and the spectral resolution of the satellite. However, further assessment and validation to delineate the effects of multiple reservoirs must consider the aforementioned hydroclimatic, anthropogenic, and biophysical conditions for the reservoirs to understand the effects of using multiple reservoirs on the satellite platforms.

CHAPTER V

CONCLUSIONS

Algal blooms are spatially and temporally dynamic in nature. These blooms are erratic and can form and dissipate in few hours or days. Chlorophyll-a is a photosynthetic pigment present in major group algae. It is used widely to monitor algal blooms in water bodies. Ground-based monitoring program alone cannot study the dynamic nature of the blooms. These blooms may form in areas where it is difficult to access (coves, difficult terrains). Ground-based sampling methods are limited to specific locations and times. Satellite remote sensing coupled with ground-based monitoring can fill the monitoring gaps of ground-based monitoring and improve the detection of Chl-a in water bodies. There are some challenges associated with satellite remote sensing. The lack of digital images due to cloud cover, spatial, temporal, and spectral resolution of satellite sensors to capture the bloom events might impede the efficacy of monitoring programs based on satellite remote sensing.

We used 14 years (2006-2020) of ground-based Chl-a and Landsat-5, Landsat-8, and Sentinel-2's spectral data to develop multiple regression models. These models utilized surface reflectance from bands, band ratios and combinations, and spectral derivatives. Our study determined the number of days we can go before or after the

satellite overpass and utilize the obtained ground-based data for developing regression models. The analysis was carried out for Landsat-5, Landsat-8, and Sentinel-2. We also compared the statistical results of Chl-a regression models using data from a single reservoir to those that were developed from multiple reservoirs. A specific time window of ±4 days between ground-based sampling and satellite image acquisition was used to carry out that analysis.

From the results of this study, we observed that Sentinel-2 models developed per time window performed better than Landsat-5 and Landsat-8. The spectral resolution of the satellite sensors used to develop the models played a role in explaining the variations observed in Chl-a estimation. This study revealed that an 8-day time window could be used to pair up satellite and ground-based data when Sentinel-2 is used and yield good results. The spectral bands present in Sentinel-2 contribute to the continuous prediction of Chl-a in reservoirs. This study also revealed that the number of reservoirs does not matter as much as the time window and the spectral resolution of the satellite being utilized. Landsat-8 and Sentinel-2 performed better than Landsat-5 when multiple reservoirs were used. Also, Landsat-8 and Sentinel-2 are open-source satellites, and the images are available free of cost. Lake managers and monitoring authorities can incorporate Landsat-8 and Sentinel-2 in their monitoring strategies along with ground-based sampling methods. This would allow the timely detection of algal blooms in water bodies, minimize the exposure risk to humans and animals, and avoid economical loss caused by algal blooms.

Future research should include hydroclimatic conditions, anthropogenic activities, biophysical conditions such as limnetic mixing, wind and weather conditions,

turbidity, BOD, lake water temperature, cumulative sunny days and precipitation events before bloom take place using available meteorological data (Tan et., 2017) for the reservoirs to fully understand the effects of multiple reservoirs. This study did not consider those conditions for the study reservoirs. Inclusion of those parameters will improve the efficacy of Chl-a predictive models using satellite remote sensing to monitor algal blooms. Study of those processes within the different zones of a reservoir (riverine, transition, and lacustrine zones) and its interaction with Chl-a could help to develop better predictive models for long term monitoring of algal blooms.

REFERENCES

Ansper, A., & Alikas, K. (2019). Retrieval of chlorophyll a from Sentinel-2 MSI data for the European Union water framework directive reporting purposes. *Remote Sensing*, *11*(1). https://doi.org/10.3390/rs11010064

Becker, B. L., Lusch, D. P., & Qi, J. (2005). Identifying optimal spectral bands from in situ measurements of Great Lakes coastal wetlands using second-derivative analysis. *Remote Sensing of Environment*, *97*(2), 238–248. https://doi.org/10.1016/j.rse.2005.04.020

Bonansea, M., Rodriguez, M. C., Pinotti, L., & Ferrero, S. (2015). Using multi-temporal Landsat imagery and linear mixed models for assessing water quality parameters in Río Tercero reservoir (Argentina). *Remote Sensing of Environment*, *158*, 28–41. https://doi.org/10.1016/j.rse.2014.10.032

Bonansea, M., Ledesma, M., Bazán, R., Ferral, A., German, A., O'Mill, P., … Pinotti, L. (2019). Evaluating the feasibility of using Sentinel-2 imagery for water clarity assessment in a reservoir. *Journal of South American Earth Sciences*, *95*(May), 102265. https://doi.org/10.1016/j.jsames.2019.102265

Boyer, T. A., Daniels, B., & Melstrom, R. T. (2015). *Algal Blooms in Oklahoma : Lake Warnings*.

Brezonik, P., Menken, K. D., Bauer, M., Brezonik, P., Menken, K. D., & Bauer, M. (2009). *Lake and Reservoir Management Landsat-based Remote Sensing of Lake Water Quality Characteristics , Including Chlorophyll and Colored Dissolved Organic Matter (CDOM) Landsat-based Remote Sensing of Lake Water Quality Characteristics , Including Chlorophy*. *2381*. https://doi.org/10.1080/07438140509354442

Buma, W. G., & Lee, S. Il. (2020). Evaluation of Sentinel-2 and Landsat 8 images for estimating Chlorophyll-a concentrations in Lake Chad, Africa. *Remote Sensing*, *12*(15). https://doi.org/10.3390/RS12152437

Chipman, J. W., Lillesand, T. M., Schmaltz, J. E., Leale, J. E., & Nordheim, M. J. (2004). Mapping lake water clarity with Landsat images in Wisconsin, U.S.A. *Canadian Journal of Remote Sensing*, *30*(1), 1–7. https://doi.org/10.5589/m03-047

Dale, B., Edwards, M., & Reid, P. C. (2006). Climate Change and Harmful Algal Blooms. *Ecology of Harmful Algae*, *189*(Hallegraeff 1993), 367–378. https://doi.org/10.1007/978-3-540-32210-8_28

Deutsch, E., & Alameddine, I. (2014). *Developing Landsat Based Algorithms To Augment In Situ Monitoring Of Freshwater Lakes And Reservoir*

Dodds, W. and Whiles, M. (2010) Freshwater Ecology: Concepts and Environmental Applications of Limnology. 2nd Edition, Elsevier, Amsterdam, 330-333.

Gholizadeh, M. H., Melesse, A. M., & Reddi, L. (2016). A comprehensive review on water quality parameters estimation using remote sensing techniques. *Sensors (Switzerland)*, *16*(8). https://doi.org/10.3390/s16081298

Gitelson, A. A., Dall'Olmo, G., Moses, W., Rundquist, D. C., Barrow, T., Fisher, T. R., … Holz, J. (2008). A simple semi-analytical model for remote estimation of chlorophyll-a in turbid waters: Validation. *Remote Sensing of Environment*, *112*(9), 3582–3593. https://doi.org/10.1016/j.rse.2008.04.015

Graham, J. L., Ziegler, A. C., Loving, B. L., & Loftin, K. (2012). Fate and Transport of Cyanobacteria and Associated Toxins and Taste-and-Odor Compounds from Upstream Reservoir Releases in the Kansas River, Kansas, September and October 2011. In *US Department of the Interior, US Geological Survey, 2012*. (Vol. 5129). Retrieved from papers://aff512a5-579d-44ca-9504-d5ff894b570f/Paper/p4365

Han, L., & Rundquist, D. C. (1997). Comparison of NIR/RED ratio and first derivative of reflectance in estimating algal-chlorophyll concentration: A case study in a turbid reservoir. *Remote Sensing of Environment, 62*(3), 253-261. doi:10.1016/S0034-4257(97)00106-5

Han, L. (2005). Estimating chlorophyll-a concentration using first-derivative spectra in coastal water. *International Journal of Remote Sensing*, *26*(23), 5235–5244. https://doi.org/10.1080/01431160500219133

Ho, J. C., Michalak, A. M., & Pahlevan, N. (2019). Widespread global increase in intense lake phytoplankton blooms since the 1980s. *Nature*, *574*(7780), 667–670. https://doi.org/10.1038/s41586-019-1648-7

Keith, D., Rover, J., Green, J., Zalewsky, B., Charpentier, M., Thursby, G., & Bishop, J. (2018). Monitoring algal blooms in drinking water reservoirs using the landsat-8 operational land imager. *International Journal of Remote Sensing*, *39*(9), 2818–2846. https://doi.org/10.1080/01431161.2018.1430912

Kloiber, S. M., Brezonik, P. L., Olmanson, L. G., & Bauer, M. E. (2002). A procedure for regional lake water clarity assessment using Landsat multispectral data. *Remote*

Sensing of Environment, *82*(1), 38–47. https://doi.org/10.1016/S0034-4257(02)00022-6

Kuhn, C., Matos, A. De, Ward, N., Loken, L., Oliveira, H., Kampel, M., … Butman, D. (2019). Remote Sensing of Environment Performance of Landsat-8 and Sentinel-2 surface reflectance products for river remote sensing retrievals of chlorophyll- a and turbidity. *Remote Sensing of Environment*, *224*(February), 104–118. https://doi.org/10.1016/j.rse.2019.01.023

Kupssinskü, L. S., Guimarães, T. T., De Souza, E. M., Zanotta, D. C., Veronez, M. R., Gonzaga, L., & Mauad, F. F. (2020). A method for chlorophyll-a and suspended solids prediction through remote sensing and machine learning. *Sensors (Switzerland)*, *20*(7). https://doi.org/10.3390/s20072125

Liu, L. W., & Wang, Y. M. (2019). Modelling reservoir turbidity using Landsat 8 satellite imagery by gene expression programming. *Water (Switzerland)*, *11*(7). https://doi.org/10.3390/w11071479

Mamun, M., Ferdous, J., & An, K.-G. (2021). Empirical Estimation of Nutrient, Organic Matter and Algal Chlorophyll in a Drinking Water Reservoir Using Landsat 5 TM Data. *Remote Sensing*, *13*(12), 2256. https://doi.org/10.3390/rs13122256

Mansaray, A. S., Dzialowski, A. R., Martin, M. E., Wagner, K. L., Gholizadeh, H., & Stoodley, S. H. (2021). Comparing planetscope to landsat-8 and sentinel-2 for sensing water quality in reservoirs in agricultural watersheds. *Remote Sensing*, *13*(9), 1–20. https://doi.org/10.3390/rs13091847

Mason, A. M., & Triplett, J. R. (2016). Controlling environmental crisis messages in uncontrollable media environments: The 2011 case of blue-green algae on Grand Lake O' the Cherokees, OK. *Advances in Natural and Technological Hazards Research*, *45*, 189–204. https://doi.org/10.1007/978-3-319-20161-0_12

McCullough, I. M., Loftin, C. S., & Sader, S. A. (2012). Combining lake and watershed characteristics with Landsat TM data for remote estimation of regional lake clarity. *Remote Sensing of Environment*, *123*, 109–115. https://doi.org/10.1016/j.rse.2012.03.006

Moore, S. K., Trainer, V. L., Mantua, N. J., Parker, M. S., Laws, E. A., Backer, L. C., & Fleming, L. E. (2008). Impacts of climate variability and future climate change on harmful algal blooms and human health. *Environmental Health: A Global Access Science Source*, *7*(SUPPL. 2), 1–12. https://doi.org/10.1186/1476-069X-7-S2

Olmanson, L. G., Bauer, M. E., & Brezonik, P. L. (2008). A 20-year Landsat water clarity census of Minnesota's 10,000 lakes. *Remote Sensing of Environment*, *112*(11), 4086–4097. https://doi.org/10.1016/j.rse.2007.12.013

Ouma, Y. O., Noor, K., & Herbert, K. (2020). Modelling Reservoir Chlorophyll- a, TSS, and Turbidity Using Sentinel-2A MSI and Landsat-8 OLI Satellite Sensors with Empirical Multivariate Regression. *Journal of Sensors*, *2020*. https://doi.org/10.1155/2020/8858408

OWRB (2010). *Development of a Probability Survey Design for Lakes & Reservoirs Final Report Approved August 31 , 2010*.

OWRB. (2018, May 17). *Monitoring and Assessment*. (Oklahoma Water Resources Board) Retrieved from http://www.owrb.ok.gov/quality/monitoring/monitoring.php

OWRB. (2018, March 14). *Water Facts*. Retrieved from The Oklahoma Water Resources Board: https://www.owrb.ok.gov/util/waterfact.php

Pirasteh, S., Mollaee, S., Fatholahi, S. N., & Li, J. (2020). Estimation of Phytoplankton Chlorophyll-a Concentrations in the Western Basin of Lake Erie Using Sentinel-2 and Sentinel-3 Data. *Canadian Journal of Remote Sensing*, *46*(5), 585–602. https://doi.org/10.1080/07038992.2020.1823825

Song, K., Fang, C., Jacinthe, P., Wen, Z., Liu, G., Xu, X., … Lyu, L. (2021). *Climatic versus Anthropogenic Controls of Decadal Trends (1983 − 2017) in Algal Blooms in Lakes and Reservoirs across China*.https://doi.org/10.1021/acs.est.0c06480

Sun, D., Hu, C., Qiu, Z., & Shi, K. (2015). Estimating phycocyanin pigment concentration in productive inland waters using Landsat measurements: A case study in Lake Dianchi. *Optics Express*, *23*(3), 3055. https://doi.org/10.1364/oe.23.003055

Tan, W., Liu, P., Liu, Y., Yang, S., & Feng, S. (2017). *A 30-Year Assessment of Phytoplankton Blooms in Erhai Lake Using Landsat Imagery : 1987 to 2016*. https://doi.org/10.3390/rs9121265

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

Torbick, N., & Corbiere, M. (2015). *A Multiscale Mapping Assessment of Lake Champlain Cyanobacterial Harmful Algal Blooms*. 11560–11578. https://doi.org/10.3390/ijerph120911560

Urbanski, J., Wochna, A., Bubak, I., Grzybowski, W., Lukawska-Matuszewska, K., Łącka, M., … Zajączkowski, M. (2016). Application of Landsat 8 imagery to regional-scale assessment of lake water quality. *International Journal of Applied Earth Observation and Geoinformation*, *51*, 28–36. https://doi.org/10.1016/j.jag.2016.04.004

Watanabe, F., Alcântara, E., Imai, N., Rodrigues, T., & Bernardo, N. (2018). Estimation of chlorophyll-a concentration from optimizing a semi-analytical algorithm in productive inland waters. *Remote Sensing*, *10*(2), 1–18. https://doi.org/10.3390/rs10020227

Watanabe, F., Alcântara, E., Bernardo, N., de Andrade, C., Gomes, A. C., do Carmo, A., … Rotta, L. H. (2019). Mapping the chlorophyll-a horizontal gradient in a cascading reservoirs system using MSI Sentinel-2A images. *Advances in Space Research*, *64*(3), 581–590. https://doi.org/10.1016/j.asr.2019.04.035

Yacobi, Y. Z., Giltelson, A., & Mayo, M. (1995). Remote sensing of chlorophyll in Lake Kinneret using highspectral-resolution radiometer and Landsat TM: spectral features of reflectance and algorithm development. *P1ankton Research*, *17*(11), 2155-2173. Retrieved from http://digitalcommons.unl.edu/natrespapers/

Yan, Y., Bao, Z., & Shao, J. (2018). Phycocyanin concentration retrieval in inland waters: A comparative review of the remote sensing techniques and algorithms. *Journal of Great Lakes Research*, *44*(4), 748–755. https://doi.org/10.1016/j.jglr.2018.05.004

Yip, H. D., Johansson, J., & Hudson, J. J. (2015). A 29-year assessment of the water clarity and chlorophyll-a concentration of a large reservoir: Investigating spatial and temporal changes using Landsat imagery. *Journal of Great Lakes Research*, *41*, 34–44. https://doi.org/10.1016/j.jglr.2014.11.022

APPENDICES

Appendix I. Best-fit models obtained for the three satellite sensors with their statistics

Table 7 Predictive models for three satellites for time window with their statistics (R^2, RMSE, 10 fold CV $R^2$and RMSE) for Landsat-5

Platform	D	Data points	Model (s)	R^2	RMSE	10-Folds CV	
						R^2	RMSE
Landsat-5 TM	0	15	Sqrt(Chl-a) = 5.957 - 13.93* d1	0.55	1.04	0.44	1.10
Landsat-5 TM	1	37	Sqrt(Chl-a) = -1.14 + 3.54* (G/R) + 5.96 NIR1	0.39	0.79	0.32	0.81
Landsat-5 TM	2	84	Ln(Chl-a) = 4.936 - 1.051*(G/B) - 0.791*(R/NIR1)	0.39	1.86	0.32	1.89
Landsat-5 TM	3	101	Ln(Chl-a) = 3.705 - 0.736 * (R/NIR1) - 1.103* d1	0.38	1.82	0.34	1.84
Landsat-5 TM	4	139	Ln(Chl-a) = 3.224 - 0.5633* (R/NIR1) - 1.185 *d2 + 8.82* d5	0.30	1.82	0.24	1.86
Landsat-5 TM	5	168	Ln(Chl-a) = 3.269 - 0.418* (R/NIR1) + 1.275 d3 + 15.08 * d5	0.24	1.95	0.21	1.95
Landsat-5 TM	6	179	Ln(Chl-a) =3.425 - 0.497 * (R/NIR1) - 0.828*d1 + 7.97*d5	0.22	1.95	0.19	1.97
Landsat-5 TM	7	191	Ln(Chl-a) = 3.716 - 7.24 * NIR1 - 0.6473*(R/NIR1)	0.21	1.97	0.19	1.97
Landsat-5 TM	8	197	Ln(Chl-a) =3.732 - 0.000754 * B4 - 0.6544 *(R/NIR1)	0.22	1.97	0.19	1.99

Table 8 Predictive models for three satellites for time window with their statistics (R^2, RMSE, 10 fold CV $R^2$and RMSE) for Landsat-8

Platform	D	Data Points	Model (s)	R^2	RMSE	10-Folds CV	
						R^2	RMSE
Sentinel-2	0	20	Ln (Chl-a) =2.618 + 4.415* d4 + 3.57 * d8A	0.86	1.62	0.82	1.65
Sentinel-2	1	49	Ln (Chl-a) = 2.620 - 0.599 *(NIR-SWIR1/NIR+SWIR1) + 4.835* d4	0.83	1.52	0.80	1.55
Sentinel-2	2	67	Sqrt (Chl-a) = 2.695 + 10.245 *d4 + 9.44 *d7 – 11.72 * d22	0.78	0.67	0.75	0.72
Sentinel-2	3	124	Sqrt (Chl-a) = 2.802 + 9.908 *d4 + 10.06 *d7 – 11.59* d22	0.79	0.69	0.78	0.72
Sentinel-2	4	195	Sqrt (Chl-a) = 5.766 - 3.473 *(Red/RE1) + 0.653*(G/B) - 6.567 *d6 + 17.33 *d23	0.70	1.01	0.64	1.08
Sentinel-2	5	237	Sqrt (Chl-a) =5.802 - 2.148 *(NIR-Red/ NIR+Red) - 3.437*(Red/RE1) - 5.576* d6 + 21.12* d23	0.67	1.06	0.66	1.1
Sentinel-2	6	259	Sqrt (Chl-a) =6.404 - 2.844 * (Red/RE1) + 77.13* d11 + 18.91 * d23	0.64	1.25	0.63	1.30
Sentinel-2	7	291	Sqrt (Chl-a) = 6.857 - 3.214 *(Red/RE1) + 77.97 *d11 + 18.62 * d23	0.65	1.25	0.64	1.28
Sentinel-2	8	333	Sqrt (Chl-a) = 6.075 - 2.670 *(Red/RE1) - 6.807 *d6 + 18.80* d23	0.62	1.25	0.61	1.28

Table 9 Predictive models for three satellites for time window with their statistics (R^2, RMSE, 10 fold CV R^2 and RMSE) for Senitnel-2

Platform	D	Data Points	Model (s)	R^2	RMSE	10-Folds CV	
						R^2	RMSE
Landsat-8	0	24	Sqrt (Chl-a)= 5.011- 11.61*d1	0.62	0.35	0.46	0.44
Landsat-8	1	73	Sqrt(Chl-a)=1.840 + 48.78 d21 + 10.62 d23 + 21.57 d25	0.6	0.88	0.56	0.94
Landsat-8	2	110	Sqrt(Chl-a) = 6.382 - 3.945 * (NIR/red) + 60.73 *d21 - 10.91 *d24	0.51	1.41	0.47	1.46
Landsat-8	3	181	Sqrt(Chl-a) =6.066 - 3.658 * (NIR/red) + 65.17 *d21 - 10.784 d24	0.57	1.63	0.55	1.66
Landsat-8	4	235	Sqrt(Chl-a) =5.876 - 3.480 * (NIR/red) + 55.47 *d21 - 9.871* d24	0.49	1.66	0.46	1.71
Landsat-8	5	304	Sqrt(Chl-a)= 4.782 - 2.160 * (NIR/red) + 53.32* d21 - 7.778 *d24	0.40	1.82	0.39	1.85
Landsat-8	6	329	Sqrt(Chl-a) = 2.883 + 1.678 * ((G-SWIR)/(G+SWIR))^2 + 12.756 * d4 + 52.84 * d21	0.41	1.79	0.39	1.85
Landsat-8	7	408	Sqrt(Chl-a) =3.850 + 69.87 *d6 + 55.47* d21 - 6.508 * d24	0.40	1.69	0.38	1.69
Landsat-8	8	432	Sqrt(Chl-a)=3.742 + 62.74 * d6 + 54.95 *d21 - 6.239 * d24	0.38	1.69	0.36	1.69

Appendix II. List of reservoirs with their date of image acquisition and ground-based sampling for the three satellites

Table 10 Difference in days between Landsat-8 image acquisition and in situ sampling dates for the 10 Oklahoma reservoirs

S.N	Name of Reservoir	Date of image acquisition	Date of in situ sampling	(±)Lag days
1.	Robert S. Kerr Reservoir	11/18/2015	11/18/2015	0
		9/17/2016	9/13/2016	-4
2.	Grand Lake	4/24/2015	4/28/2015	+4
		8/6/2018	8/7/2018	+1
3.	Lake Carl Blackwell	4/25/2013	4/24/2013	-1
		12/18/2015	12/21/2015	+3
		9/24/2016	9/27/2016	+3
		2/9/2018	2/13/2018	+4
		6/1/2018	6/5/2018	+4
4.	Lake McMurtry	4/12/2014	4/7/2014	-5
		10/1/2016	10/3/2016	+2
		4/4/2017	4/4/2017	0
		7/9/2017	7/5/2017	-4
		1/4/2019	1/2/2019	-2
		6/29/2019	7/2/2019	-3
5.	Lake Texoma	10/13/2017	10/16/2017	+3
		8/31/2013	8/26/2013	-5
6.	Oologah lake	3/11/2014	3/10/2014	-1
		8/27/2014	8/25/2014	-2
		11/11/2016	11/14/2016	+3
		5/31/2017	5/31/2017	0
		8/19/2017	8/15/2017	-4
		1/29/2019	1/30/2019	+1
		4/26/2019	4/29/2019	+3
		7/24/2019	7/29/2019	+5
		7/31/2019	7/29/2019	-2
7.	Altus Reservoir	9/4/2015	9/8/2015	+4
		10/17/2019	10/14/2019	-3
		8/21/2016	8/22/2016	+1
8.	Waurika Lake	4/14/2018	4/10/2018	-4
		11/29/2014	12/2/2014	+3

S.N	Name of Reservoir	Date of image acquisition	Date of in situ sampling	(±)Lag days
9.	Eufaula Lake	8/14/2015	8/10/2015	-4
		1/23/2017	1/18/2017	-5
		7/9/2017	7/10/2017	+1
		10/12/2019	10/7/2019	-5
		1/7/2020	1/7/2020	0
10.	Kaw Lake	1/16/2015	1/13/2015	-3
		4/25/2013	4/22/2013	-3
		4/22/2015	4/21/2015	-1
		7/4/2015	7/6/2015	-2
		6/1/2018	6/4/2018	+3
		8/4/2018	8/7/2018	+3
		11/4/2019	11/5/2019	+1
		2/8/2020	2/10/2020	+2
		9/26/2020	9/21/2020	-5

Table 11 Difference in days between Sentinel-2 image acquisition and ground-based sampling for 10 reservoirs of Oklahoma

S.N	Name of reservoir	Date of image acquisition	Date of in situ sampling	(±)Lag days
1.	Robert S. Kerr Reservoir	2/21/2020	2/24/2020	-3
2.	Grand Lake	10/19/2019	10/22/2019	-3
		3/7/2020	3/4/2020	+3
		8/24/2020	8/25/2020	-1
		1/27/2018	1/23/2018	+4
		10/29/2017	10/24/2017	+5
		4/12/2018	4/17/2018	-5
		8/19/2020	8/25/2020	-6
		10/14/2019	10/22/2019	-8
3.	Lake Carl Blackwell	11/16/2017	11/14/2017	+2
		8/3/2018	8/6/2018	-3
		2/9/2018	2/13/2018	-4
		6/9/2018	6/5/2018	+4
4.	Lake McMurtry	1/5/2019	1/2/2019	+3
		3/31/2019	4/1/2019	-1
		7/9/2019	7/2/2019	-7
		6/24/2019	7/2/2019	-8
5.	Lake Texoma	7/6/2018	7/2/2018	+4
		10/22/2017	10/16/2017	+6
		10/9/2017	10/16/2017	-7
		1/15/2018	1/8/2018	+7
		4/10/2018	4/2/2018	+8
6.	Oologah Lake	1/30/2019	1/30/2019	0
		4/27/2019	4/29/2019	-2
		10/27/2018	10/29/2018	-2
		7/31/2019	7/29/2019	-2
		2/16/2017	2/13/2017	+3
		1/27/2019	1/30/2019	-3
		7/26/2019	7/29/2019	-3
		8/23/2017	8/15/2017	+8
		7/21/2019	7/29/2019	-8

S.N	Name of reservoir	Date of image acquisition	Date of in situ sampling	(±)Lag days
7.	Altus reservoir	10/15/2019	10/14/2019	+1
		1/18/2020	1/21/2020	-3
		10/10/2019	10/14/2019	-4
		10/20/2019	10/14/2019	+6
		1/13/2020	1/21/2020	-8
8.	Waurika Lake	4/10/2018	4/10/2018	0
		8/17/2020	8/17/2020	0
		1/20/2018	1/17/2018	+3
		9/27/2019	9/30/2019	-3
		10/27/2017	10/31/2017	-4
		4/15/2018	4/10/2018	+5
		8/22/2020	8/17/2020	+5
		7/4/2018	7/10/2018	-6
		10/7/2019	9/30/2019	+7
		1/20/2020	1/13/2020	+7
		1/5/2020	1/13/2020	-8
9.	Eufaula Lake	8/9/2020	8/10/2020	-1
		8/14/2020	8/10/2020	+4
		7/6/2017	7/10/2017	-4
10.	Kaw Lake	9/21/2020	9/21/2020	0
		11/6/2019	11/5/2019	+1
		2/9/2018	2/12/2018	-3
		5/30/2018	6/4/2018	-4
		8/3/2018	8/7/2018	-4
		11/1/2019	11/5/2019	-4
		6/9/2018	6/4/2018	+5
		9/16/2020	9/21/2020	-5

Table 12 Difference in days between Sentinel-2 image acquisition and ground-based sampling for 10 reservoirs of Oklahoma

S.N	Name of reservoir	Date of image acquisition	Date of in situ sampling	(±)Lag days
1.	Robert S. Kerr Reservoir	2/21/2020	2/24/2020	-3
2.	Grand Lake	10/19/2019	10/22/2019	-3
		3/7/2020	3/4/2020	+3
		8/24/2020	8/25/2020	-1
		1/27/2018	1/23/2018	+4
		10/29/2017	10/24/2017	+5
		4/12/2018	4/17/2018	-5
		8/19/2020	8/25/2020	-6
		10/14/2019	10/22/2019	-8
3.	Lake Carl Blackwell	11/16/2017	11/14/2017	+2
		8/3/2018	8/6/2018	-3
		2/9/2018	2/13/2018	-4
		6/9/2018	6/5/2018	+4
4.	Lake McMurtry	1/5/2019	1/2/2019	+3
		3/31/2019	4/1/2019	-1
		7/9/2019	7/2/2019	-7
		6/24/2019	7/2/2019	-8
5.	Lake Texoma	7/6/2018	7/2/2018	+4
		10/22/2017	10/16/2017	+6
		10/9/2017	10/16/2017	-7
		1/15/2018	1/8/2018	+7
		4/10/2018	4/2/2018	+8

S.N	Name of reservoir	Date of image acquisition	Date of in situ sampling	(±)Lag days
6.	Oologah Lake	1/30/2019	1/30/2019	0
		4/27/2019	4/29/2019	-2
		10/27/2018	10/29/2018	-2
		7/31/2019	7/29/2019	-2
		2/16/2017	2/13/2017	+3
		1/27/2019	1/30/2019	-3
		7/26/2019	7/29/2019	-3
		8/23/2017	8/15/2017	+8
		7/21/2019	7/29/2019	-8
7.	Altus reservoir	10/15/2019	10/14/2019	+1
		1/18/2020	1/21/2020	-3
		10/10/2019	10/14/2019	-4
		10/20/2019	10/14/2019	+6
		1/13/2020	1/21/2020	-8
8.	Waurika Lake	4/10/2018	4/10/2018	0
		8/17/2020	8/17/2020	0
		1/20/2018	1/17/2018	+3
		9/27/2019	9/30/2019	-3
		10/27/2017	10/31/2017	-4
		4/15/2018	4/10/2018	+5
		8/22/2020	8/17/2020	+5
		7/4/2018	7/10/2018	-6
		10/7/2019	9/30/2019	+7
		1/20/2020	1/13/2020	+7
		1/5/2020	1/13/2020	-8

S.N	Name of reservoir	Date of image acquisition	Date of in situ sampling	(±)Lag days
9.	Eufaula Lake	8/9/2020	8/10/2020	-1
		8/14/2020	8/10/2020	+4
		7/6/2017	7/10/2017	-4
10.	Kaw Lake	9/21/2020	9/21/2020	0
		11/6/2019	11/5/2019	+1
		2/9/2018	2/12/2018	-3
		5/30/2018	6/4/2018	-4
		8/3/2018	8/7/2018	-4
		11/1/2019	11/5/2019	-4
		6/9/2018	6/4/2018	+5
		9/16/2020	9/21/2020	-5

VITA

Priya Kayastha

Candidate for the Degree of

Master of Science

Thesis: MATCHING SATELLITE AND GROUND-BASED DATA FOR CHLOROPHYLL-A SENSING IN RESERVOIRS

Major Field: Environmental Science

Biographical:

Education:

Completed the requirements for the Master of Science in Environmental Science at Oklahoma State University, Stillwater, Oklahoma in July, 2021.

Completed the requirements for the Bachelor of Science in Environmental Science, Dhulikhel, Kavre, Nepal, 2017.

Experience:

- Graduate Research Assistant
  Oklahoma State University — August 2019-July 2021
- Environmental Consultant
  Raj Bandhu Engineering Services
  Nepal — January 2019- July 2019
- Intern
  The Small Earth Nepal (SEN)
  Nepal — February 2018-July 2018