Norges miljø- og biovitenskapelige universitet

Master's Thesis 2023 30 ECTS (STP)
Faculty of Science and Technology
(Fakultet for realfag og teknologi)

Water Quality Estimation Using High Resolution Satellite Imaging around Addis Ababa

Kvalitetestimering av vann ved bruk av høyoppløselig satellittbilder rundt Addis Ababa

Yonatan Bisrat Taye
Geomatics
(Geomatikk)

Preface

This is the acknowledgements.

Thanks to my supervisors for guidance and reassurance when I was feeling lost or stuck.

Thanks to the NORPART partners here at NMBU and Ethiopia for organising the field trip, improving this study with hands on experience, face to face correspondence with experts, and having the opportunity to travel worry free. Most of the worries where delt by Kotebe University of Education, as they took good care of me during my stay.

And thank you my dear editorial army for reading through these tough draft so late in the summer.

እግዚአብሔር ይመስገን

Takk for nå, Ås.

Abstract

There are many issues surrounding surface water quality in Ethiopia, and a lot can be traced back to Addis Ababa. Indiscriminate dumping, insufficient waste management infrastructure and the cost of monitoring are parts of the problem.

To protect the valuable water resources early detection and mitigation is key, and remote sensing could be a more sensible solution to these problems. In addition to cost savings of satellite based remote sensing, there are no physical boundaries, larger holistic samples and twenty years of research in this field. However, most of these techniques are based on low to medium resolution satellite images, which reduces smaller lakes and rivers. Others are based on expensive and complicated satellites, and this is a high bar of entry for low-end users. These low-end users are also the people who need better access to tools to manage their resources.

Therefore, this thesis set out to produce water quality models that are accessible for everyone with access to the Internet and a computer. By utilising the high resolution (<5 meters) satellite images from NICFI Basemap and conducting both a literature and field study, there was a hope to make viable models for different parameters.

The literature study reviewed many satellites, water quality parameters and the NICFI specifications in an effort to find matching spectral characteristics for all three. This would allow users to utilise Water Quality Models (WQM) and produce their own Parameter Distribution Maps (PDM) in their lakes of interest. However, there where many limitations when borrowing one WQM to another lake and instrument. The literature study concluded with four parameters that could be estimated: Total Suspended Matter, Chlorophyll-a, Turbidity and Coloured Dissolved Organic Matter. But the statistical significance was not calculated for these models.

A field study was conducted in Addis Ababa to gather samples to produce an original WQM using NICFI Basemap. There were not enough samples taken, so Assegide et al.'s field samples from Koka Reservoir were used. The Original WQM for chlorophyll-a had a R^2 of 0.72 with 25 samples, giving us a statistically significant result. The Novel PDM produced was accurate.

Recommendations for further research would be to use more field samples to get better models. Would also recommend that the Novel PDM should be tested on other lakes. However, this WQM should be appropriate for low-end users to monitor their essential water resources.

Sammendrag

Det er mange utfordringer rundt overflatevann i Etiopia, og mye kan spores tilbake til Addis Abeba. Ukritisk forsøpling, utilstrekkelig avfallshåndtering og høye kostnader ved overvåking er deler av problemet. For å beskytte de verdifulle vannressursene, så er tidlig gjenkjenning og skadebegrensning nøkkelen, og fjernmåling kan være en mer fornuftig løsning. I tillegg til kostnadsbesparelser ved satellittbasert fjernmåling, er det ingen fysiske grenser, helhetlige feltmålinger og tjue års forskning på dette feltet. Imidlertid er de fleste av disse teknikkene basert på satellittbilder med lav til middels oppløsning, noe som reduserer brukbarheten på innsjøer og elver. Andre er basert på dyre og kompliserte satellitter, og dette er en høy inngangsbar for «low-end» brukere. Disse brukerne er også personene som trenger bedre tilgang til verktøy for å administrere ressursene sine, på grunn av dårligere råd.

Derfor har denne oppgaven som mål å lage vannkvalitetsmodeller som er tilgjengelige for alle med tilgang til internett og en datamaskin. Ved å bruke de høyoppløselige (<5 meter) satellittbildene fra NICFI Basemap og gjennomføring av både en litteratur- og feltstudie, var det et håp om å lage levedyktige modeller for forskjellige parametere.

Litteraturstudien gjennomgikk mange satellitter, vannkvalitetsparametere og NICFI-spesifikasjonene i et forsøk på å finne samsvarende spektralegenskaper for alle tre. Brukere med vannkvalitetsmodellene (WQM) kan da produsere sine egne parameterdistribusjonskart (PDM) i sine sjøer av interesse. Imidlertid var det mange begrensninger når man låner en WQM til en annen innsjø og instrument. Litteraturstudien konkluderte med fire parametere som kunne estimeres: Totalt suspendert stoff, klorofyll-a, turbiditet og farget oppløst organisk stoff (CDOM). Men den statistiske signifikansen ble ikke beregnet for disse modellene.

Feltarbeid i Addis Ababa var gjennomført, men måtte låne andres feltprøver for å produsere en original WQM ved å bruke NICFI Basemap. Denne nye originale WQM for klorofyll-a hadde en R^2 0.72 med 25 prøver, noe som ga oss et statistisk signifikant resultat. Den nye PDM-en som ble produsert var nøyaktig.

Anbefalinger for videre forskning vil være å bruke flere feltprøver for å få bedre modeller med dette instrumentet . Imidlertid bør denne WQM være passende for lavende brukere for å overvåke deres essensielle vannressurser.

Table of Contents

1. INTRODUCTION

Water is required to sustain life, and is used domestically for hydration, food preparation and hygiene. Contaminated sources lead to a plethora of diseases, harming both the community and the environment itself. Even when disregarding anthropogenic pollution, freshwater bodies can be harmful for human use. Furthermore, the increased pressure on the world's water resources leads humanity to a vulnerable situation where conflicts can arise from scarcity. The UN estimates that 3.6 billion people live in areas that are potentially water-scarce one month a year (UNESCO, 2018). This number could grow to 5 billion by 2050. This development is evident in Ethiopia with the drought that came this spring and impacted 28 million people (WFP, 2023). In addition, the city of Addis Ababa has sprawled so uncontrolled that the sanitary infrastructure has been falling further behind. This affects the surface waters through indiscriminate dumping (Yohannes & Elias, 2017). The unproportionate growth of the city has reduced the amount of treated wastewater from 6% to 2% in Addis Ababa between 2005 and 2015. This means that less wastewater is being treated compared to the produced amount, and Little and Great Akaki River are the recipient of this waste as they are used as the city's illicit "dumping ground" (Assegide et al., 2022). In addition to the indiscriminate dumping, Assegide et al. points out a sextupling of fertiliser use between 1996 and 2015.

As outlined above, there are many issues surrounding surface water quality in Ethiopia, and a lot can be traced back to Addis Ababa. The interconnected lakes and their critical function contribute to harm people in Akaki Kality industrial zone (Yohannes & Elias, 2017). Yohannes and Elias documented diseases such as cough, diarrhoea, typhoid, typhus and other health issues. In addition, the rivers spread health risks through bioaccumulation in vegetables irrigated on polluted waters. These crops absorb heavy metals and toxins before being eaten by the population (Yohannes & Elias, 2017). Therefore, one must take care of this resource, to reduce harm and retain capacity. This means stopping pollution and maintaining the water bodies in an efficient and reliable way.

The sustainable management of resources is so important that world leaders at the United Nations have agreed on including several water-related issues in The Sustainable Development Goals (SDG), showing the significance of proper water conservation and management. The SDGs provide a global framework for sustainable development and guide nations to a better future (O'Connor et al., 2020).

A key element for water quality protection is monitoring the quality. This allows for early action to mitigate a crisis. Traditional water quality monitoring is tedious work with high costs in time and resources, and often done from stationary sampling points. This does not allow for large scale monitoring (Yang et al., 2022). While traditional monitoring methods are limited, remote sensing can infer information from mere images of surface water.

Remote sensing of the environment is not a new thing. Since the first imaging satellite was launched, images have been used to tell something about the Earth from space for over 50 years (Assegide et al., 2023). Over time, these images have been better in a multitude of ways. Spatial, with more detail being visible on the ground. Temporal, images being taken over time of the same area to see changes. Spectral, a new dimension of information available in the colours. It all culminates to the abundant number and configurations of sensors available today.

The advantages of satellite based remote sensing are plenty. The coverage is global with no restriction caused by terrain or political boundaries (O'Connor et al., 2020). In addition to the concrete information the images carry, more data can be inferred and complement conventional methods. In the Earth Observation compendium, produced by the European Space Agency (ESA), they conclude that satellite-based remote sensing can directly and indirectly contribute with 11 Sustainable Development Goals through 34 indicators. The Sustainable Development Goals (SDG) are goals ratified by world leaders at the United Nations (UN) to provide a global framework for sustainable development and guide nations to a better future. (O'Connor et al., 2020) Attaining these goals require a lot of work, and different methods to achieve them. Remote sensing does not solve SDG by handing out food, cleaning the air, or stopping conflicts. Instead, satellite remote sensing gives access to data from unreachable places, and with a higher frequency than traditional surveying methods. This helps the decision makers combat problems using the most accurate data obtainable.

"Clean water and sanitation" is the 6th SDG and states to "Ensure availability and sustainable management of water and sanitation for all." In the compendium O'Conner et al. evaluates remote sensing as a viable tool and methodology for this SDG. They also specify four targets and six indicators for the water SDG. This thesis will focus on the target 6.3: "By 2030, improve water quality by reducing pollution, eliminating dumping and minimizing release of hazardous chemicals and materials, halving the proportion of untreated wastewater and substantially increasing recycling and safe reuse globally." (O'Connor et al., 2020). Indicators 6.3.1 and .2 are about wastewater treatment and ambient water quality. By measuring the quality of wastewater in waterbodies, it is possible to detect pollution and estimate the quality of treatment. This monitoring allows for early treatment of water bodies, mitigating ecological harm to the environment and the people living there. Identifying illicit dumping sites would also help to secure these areas against future pollution.

Earlier studies show viable methods to estimate water quality from images. By comparing samples from water bodies with coinciding satellite images one can find correlation between the samples and image. This methodology is the empirical mode (Yang et al., 2022). There are many different water quality parameters and instruments used for capturing the images. They have an impact on both accuracy and applicability. However, there is a huge gap between the more powerful users and those without training or resources.

The methods available today require expertise, hardware, and field sampling materials. Furthermore, free sources, such as Sentinel and Landsat, have a lower spatial resolution, making rivers and smaller water bodies unmeasurable. For example, Brezonik et.al states that Landsat 5's 30-meter resolution can only estimate lakes that are at least 80 000m^2 (8ha). High resolution images are expensive but are crucial to monitor small changes and tackle pollution early and accurately.

Therefore, this study will use high resolution (<5 meter) satellite images to develop a methodology to estimate water quality in lakes around Addis Ababa with and without field samplings. The satellite images are freely available and are supplied by the Norway's International Climate and Forest Initiative (NICFI). Their goal is to reduce and reversing tropical forest loss, combating climate change, and facilitate sustainable development. (Planet, 2023) This aligns with the thesis' purpose. Furthermore, by

combining the images with the open-source program Quantum GIS (QGIS), the method developed could allow anyone with a computer and access to the internet to estimate water quality for any body of water in NICFI's database, no matter the physical, political, or financial restrictions. If this is successful, the methodology developed would build capacity and strengthen countries with less resources to work on their sustainability goals. With this background this study will attempt at answering these research questions:

1. Which water quality parameters are measurable using the NICFI satellite images?
2. To what degree can remotely sensed water quality models be extrapolated to different lakes?
3. How viable is NICFI for low-end users to monitor water quality?

2. BACKGROUND

2.1 Remote Sensing

The study of water quality is described by Gholizadeh et al. as “the process of determining characteristics of waterbodies and identifying the possible contamination sources that degrade the quality of water.” By using remote sensing techniques from satellite images, one can address both the water quality and the location of contaminants. The characteristics of water can be broken down to four factors. Physical (temperature, turbidity, salinity etc.), biological (bacteria, algae), chemical (oxygen, nitrogen, compounds etc.) and aesthetic (odour, taste, floating matter etc.)(Norsaliza & Ismail, 2010).

However, remote sensing cannot determine all these parameters, and different sensors have various technological limitations which reduces the detectability of certain parameters. Several studies have shown which types of parameters are more accurately measurable. Gholizadeh et al.’s review article has summarised them as mostly “optically active variables”, such as chlorophyl-a and turbidity. “Optically active” means that the parameter interacts with light and changes the energy spectrum of reflected solar radiation (Gholizadeh et al., 2016). To determine which parameters are detectable one must know two things: The parameter’s optical properties, as in “how it interacts with light”, and the properties of the measuring sensor.

The optical properties of different objects can be determined in laboratories using spectrometry to create the absorption spectrum for the object. This would gauge which wavelengths that are absorbed and reflected the most and works as the object’s “fingerprint” (Helseth, 2022). After identifying these peaks and troughs it is possible to determine which wavelength bands that should be used to identify the objects with the highest accuracy. Bands are widths of wavelengths that sensors are manufactured to measure, and different sensors have different number of bands, widths and sensitivity they can measure (Lied & Birkeland, 2022). Figure 1 shows the spectral property of chlorophyl-a and the band widths of Landsat 7 ETM. This visualisation shows how one can determine the best bands to use.

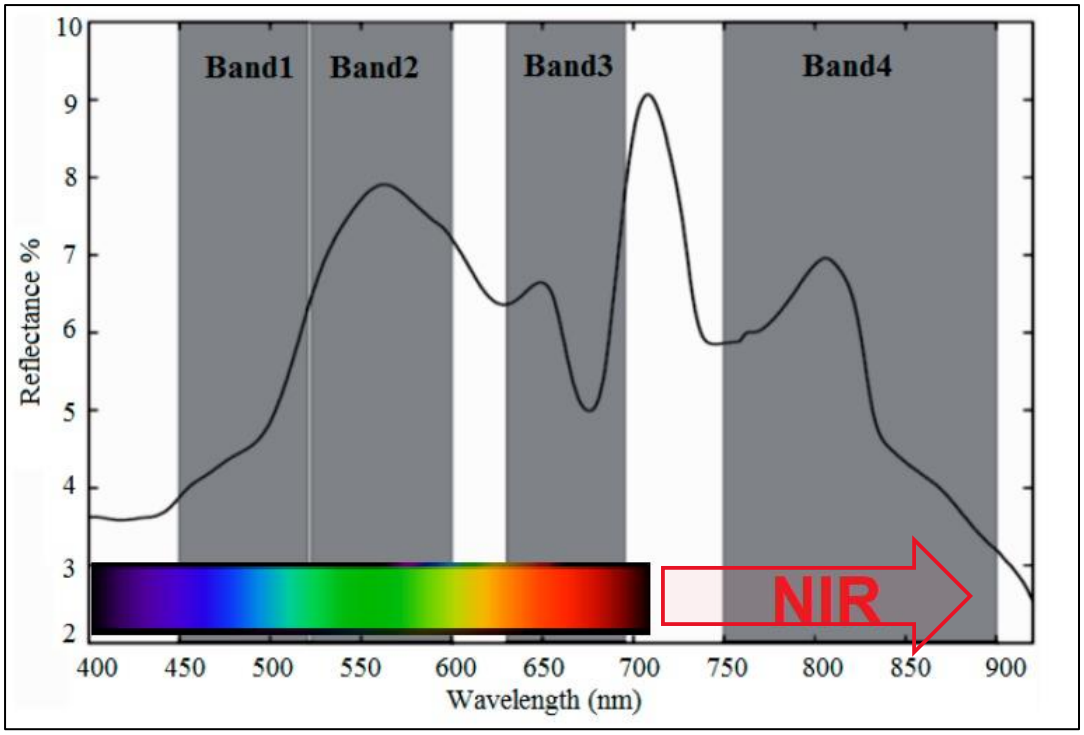


Figure 1 Example of a spectral property graph for Landsat Earth Thematic Mapper 7's bands (the bars), chlorophyll's reflectance (line graph) and approximate positioning of colours in terms of wavelength. The infrared spectrum starts after 700nm. Composite image from (Gholizadeh et al., 2016) and (Helmenstine, 2021). Original graph used under CC BY 4.0.

As seen Figure 1, bands can be thought of as the colours being measured. The greener an object, the higher is the measured radiation with the wavelengths the human eye perceive as green (Holtsmark & Angelo, 2021). Some bands in sensors are within the visible spectrum, and multispectral sensors can have bands outside wavelengths humans can perceive, such as the near infrared (NIR). Furthermore, the figure shows the amount of reflected radiation that hits a sensor. This is a ratio between the reflected light and the original radiation that lights up that surface or object being measured (NV5-Geospatial). Direct measurement of the light energy, radiance, can also be used for a sensor. For both however, one must correct the measurements for effects, such as atmospheric interference, and individual sensor calibrations. This is the basis to have an absolute measurement to compare observations between sensors.

In addition to using the bands, it is possible to get more information by examining the ratio between two or more bands. Ratios can also be combined with the difference between band signals. Therefore, studying the differences and ratios in peaks and troughs in the spectral signature can determine the object with higher certainty. Normalised Difference Vegetation Index (NDVI) is a common example of this. NDVI

makes an index to categorise quality and type of vegetation with values between -1 and 1 by using this formula: $\frac{NIR-Red}{NIR+Red}$ (Dick & Birkeland, 2023). Other indexes exist, and are relatively easy to use and common, but it only works as an index or guide instead of an estimation of quantity. Nevertheless, the band ratios can therefore be counted as additional bands, as they contain new spectral information in a similar way as the original bands. Gholizadeh et al says that using ratios "can reduce irradiance, atmospheric and air-water surface influences", and in turn improves the accuracy of models. Assegide et al found in 2023 concurring evidence of this when developing their Water Quality Model (WQM).

2.2 Sensor Review

To both develop and utilise a Water Quality Model (WQM), there are several strategies and factors to consider. The first factor is knowing the measuring sensor.

2.2.1 NICFI Database

The free high resolution satellite images are a mosaic product from Planet Labs' satellite constellation that stitches together "the global tropics" in a continuous area between 30°N and 30°S (Pooja Pandey et al., 2023). Figure 2 shows the countries that are included in this program. This is the NICFI Satellite Data Program, and the mosaic is called "NICFI Basemap".

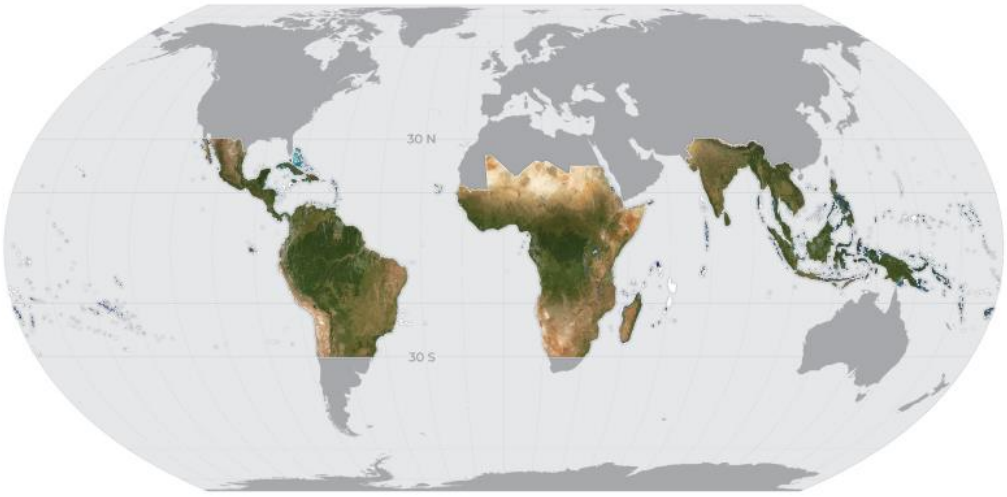


Figure 2: NICFI Basemap coverage overview. Gray area are outside the program. Imagery © 2023 Planet Labs Inc. All use subject to the Participant License Agreement

The Satellite Data program gives users free access to high resolution satellite imagery of the tropics to contribute to NICFI's goals of reducing and reversing tropical forest loss, combating climate change, and facilitate sustainable development.[1] The Basemap is easily accessible for non-commercial users and is of high technical quality. On the other hand, their *Participant License Agreement* explicitly state that the use of their product "cannot be for financial gain", but rather for sustainable resource management, document illegal land grabs, monitor environmental risk, or similar kinds of research that adheres to their purpose.[2]

The Basemap has 4 bands representing Blue (B1), Green (B2), Red (B3), and Near Infrared (B4). Near Infrared is shortened to NIR. Earliest images are from 2017, meaning shorter temporal resolution than Sentinel and Landsat. However, the spatial resolution of < 5 meters, is better than Landsat's 30 meters and Sentinel's 10 meters multispectral bands. The Basemap images are made out of smaller satellite pictures, called scenes, from various satellites and sensors. To avoid inconsistency between images, they are all harmonized and normalized to Landsat 8 SR data and Sentinel-2.[3]

2.2.2 Harmonisation and normalisation

Harmonisation and normalisation are techniques to correct and equalise different outputs from different sensors. The earliest images in the Basemap are from 2015 and were captured on Rapideye satellites. Later various Dove satellites were used, and harmonisation corrects the small differences between each satellite, but also to have comparable results to both Landsat 8 and Sentinel-2.[2]

Table 1 shows the differences in the satellite sensors and the original bands they are harmonised from. These sensors build up the Basemap, and the harmonisation allows for seamless use across the timeline of available images. Therefore, it could allow users to detect water quality parameters with the Basemap when the previous literature uses similar sensors as well. The harmonisation process has been described in depth by

[1] https://www.planet.com/nicfi/
[2] https://assets.planet.com/docs/Planet_ParticipantLicenseAgreement_NICFI.pdf
[3] https://support.planet.com/hc/en-us/articles/4407660177553-I-want-to-know-the-wavelength-information-for-the-NICFI-basemaps-
https://assets.planet.com/docs/scene_level_normalization_of_planet_dove_imagery.pdf

Planet Labs in their documentations.[4] Further reading is available in (Houborg & McCabe, 2018).

Optimised for deforestation

However, Planet Labs states in the Basemap Addendum that "the Basemap is optimised for deforestation detection via short-term differences in Normalized Difference Vegetation Index (NDVI), qualities prioritised by the NICFI technical advisors." They conclude that it "impacts other potential applications", so this "optimisation" must be considered when developing and evaluating the methodology.[3]

Wavelength information has been added in Table 1. Original band short tags are included in parentheses to make it easier to compare NICFI's bands with other images. The different numbers in the short tags between sensors comes from the difference in number of bands. The NICFI Basemap is spectrally constrained compared to the other products. Planet Labs' Super Dove PSB.SD has 8 bands, Landsat-8 has 11, and Sentinel-2 has 13 bands. However, NICFI's four bands are the standard "colours" in most multispectral sensors, meaning more existing models could be applicable. Even sensors on non-spaceborne platforms, such as planes and drones have Red, Green, Blue, and sometimes NIR too. This could allow for broader use of these techniques.

Table 1 Bands for different satellites used in NICFI database. All wavelengths are in µm and original band tags are included in parentheses.

NICFI Bands (short tag) NICFI	Rapideye	Super Dove PSB.SD	Landsat 8 OLI NICFI harmonised	Sentinel-2 NICFI harmonised
Blue – (B1)	0.440-0.510 (B1)	0.465-0.515 (B2)	0.45-0.51 (B2)	0.458-0.522 (B2)
Green – (B2)	0.520-0.590 (B2)	0.547-0.583 (B4)	0.53-0.59 (B3)	0.543-0.577 (B3)
Red – (B3)	0.630-0.685 (B3)	0.650-0.680 (B6)	0.64-0.67 (B4)	0.650-0.680 (B4)
Near Infrared NIR – (B4)	0.760-0.850 (B5)	0.845-0.885 (B8)	0.85-0.88 (B5)	0.785-0.899 (B8)

[4] https://assets.planet.com/docs/scene_level_normalization_of_planet_dove_imagery.pdf

Sources: [5]

2.2.3 Spectral characteristics

The United States Geological Survey (USGS) has a Spectral Characteristics Viewer that allows for comparing different sensors' characteristics.[6] This "spectral characteristic" is before any harmonisation is done, showing that the widths and peaks are different, but similar, just like Table 1 shows.

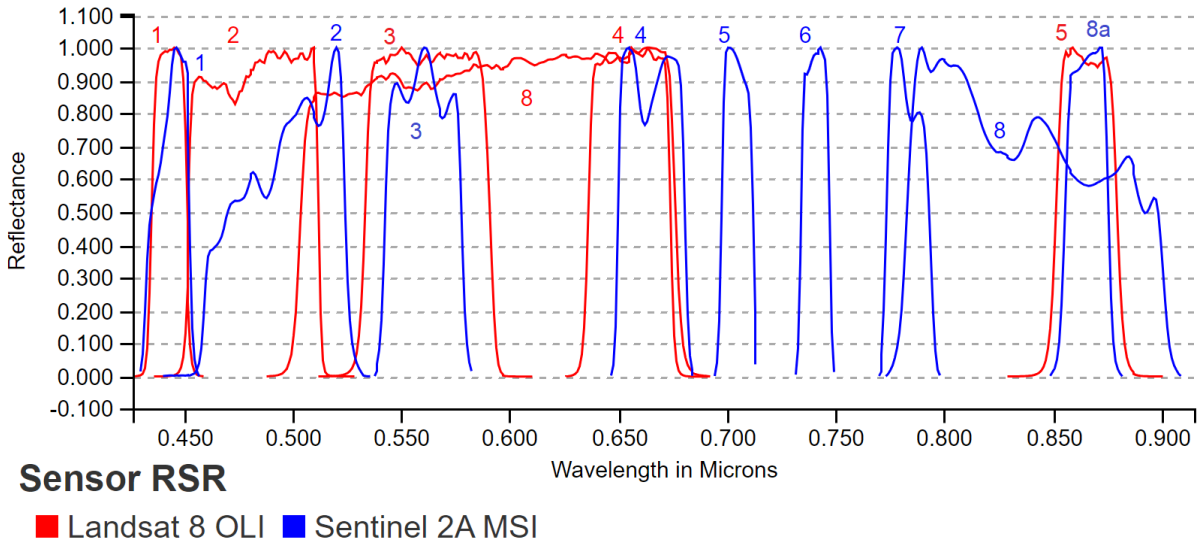


Figure 3: Comparing Landsat and Sentinel band widths using USGS Spectral Characteristics Viewer. Relevant band numbers are presented in Table 1

By using Landsat 8 as a baseline, just like Planet has[3], one can start researching detectable parameters.

However, there were some critical misunderstandings between radiance vs. reflectance that affected the thesis. This discussed in chapter 6.1 Procedural Knowledge.

[5] https://support.planet.com/hc/en-us/articles/4407660177553-I-want-to-know-the-wavelength-information-for-the-NICFI-basemaps- (Planet)
https://sentinel.esa.int/web/sentinel/user-guides/sentinel-2-msi/resolutions/spectral (ESA)
https://earth.esa.int/eogateway/missions/rapideye (ESA)
https://assets.planet.com/docs/Planet_Combined_Imagery_Product_Specs_letter_screen.pdf
[6] https://landsat.usgs.gov/spectral-characteristics-viewer

2.2.4 Increasing pixel window

To reduce inaccuracy and noise from the satellite images, one can use more pixels to estimate the water quality parameter (Brivio et al., 2001). By taking the average spectral value of the surrounding pixels instead of the single pixel containing the sample site, one can achieve a higher accuracy in the model. Several studies used this technique to their advantage, such as Assegide et al. 2023, Mancino et al. and Brivio et al. Brivio et al. states that these "average pixel windows" can be different sizes but if they are "too large" the "value may not be representative". Ekercin's case study in Turkey tested several "pixel window sizes" to find the optimal size correlating water quality with satellite imagery (Ekercin, 2007). The most optimal window for Ekercin was a 5x5 window. The satellite he used had a spectral resolution of 1m panchromatic and 4m for visual spectrum. This is very similar to NICFI except for the panchromatic band. When analysing Ekercin's optimal window results, the differences between the best and worst regression was 6% for SDD, 12% for chlorophyll-a, and 15% for TSS. Therefore, a noticeable amount could be gained by examining optimal pixel window. Nevertheless, the 3x3 window was the most accurate for NICFI Basemap when calculating regression for chlorophyll-a in Koka. This was tested against 5x5 pixel window, and the 3x3 was slightly better in the R^2 value.

2.2.5 Correlation

The extracted pixel, or window of pixels, are correlated with the water quality parameter to develop distribution models. (Ekercin, 2007) Multiple linear regression is one of them, and is the most relevant method for multispectral images, as this technique uses several independent variables to explain one dependant variable. For example, the four bands are independent while chlorophyll-a is dependant. This is also called empirical analysis where ground truths are correlated with another type of observation (Yang et al., 2022). This correlation is also called Pearson's r.[7]

[7] https://www.scribbr.com/statistics/pearson-correlation-coefficient/

2.3 Water Quality Parameter Review

A literature review was conducted to determine which parameters were possible to detect and estimate using the available sensor. As some parameters are more detectable to than others based on the different wavelengths, this must be considered.

This literature review was not conducted using any systematic methodology. Instead, research was done with database searches in NMBU's Oria and Google Scholar. Further reading was done in the sources of the sources. Yang et al and Gholizadeh et al.'s summary and review articles supplied the literature review with further research to investigate. After examining their findings, several Water Quality Models (WQM) were retrieved. The studies researched were conducted the last 20 years, and as recent as 2023. However, there are differences in their instruments, environment, and results. None of the papers had the exact same WQM, so several WQMs were tested. Some several times for the same parameter, such as chlorophyll-a. The literature review conclusion is presented in Table 2 below.

Table 2 Result from literature review. It is organised by the sources and includes details related to their results and methodology. All details can be found in their respective sources. Scale factor of 0.01 is used on non-ratioed addends. NICFI band names: B1=Blue, B2=Green, B3=Red, B4=Near Infrared | "n"=number of observations (water samples). Note that µg/L = mg/m3 for chl-a.

Source	Parameter	Model Equation	R^2	Type description
Miller 2004	Total Suspended Matter Concentration C_TSM (mg/L)	−1.91×1140.25 × (B3)	0.89 n=52	One band model using Red band. MODIS Terra with 250 m resolution used on the coast of the Northern Gulf of Mexico. Variating water quality dependant on river discharges.
Bilge 2003	Nitrate NO3-N (mg/L)	2.84 - 0.06×(B1) - 0.05×(B2) + 0.06×(B3) + 0.38×(B4)	0.86 n=14	4 band model with Landsat 4 TM with 30 m resolution used in Porsuk Dam reservoir, Turkey. Turbid water with nutrient loading.
	Suspended Sediment SS (mg/L)	6.50-0.73×(B1) -1.16×(B2) + 3.00×(B3) + 3.65×(B4)	0.92 n=14	
	Chlorophyll-a Chl-a (µg/L)	44.20 - 1.17×(B1) - 0.88×(B2) + 1.49×(B3) + 4.08×(B4)	0.87 n=14	
Kutser 2005	Coloured Dissolved Organic Matter CDOM (m^{-1})	5.13 × (B2/B3)^-2.67	0.73 n=14	One band ratio. EO-1's Advanced Land Imager with 30 m resolution used on boreal lakes in Finland and Sweden
Brezonic 2005	Secchi Disk Transparency ln(SDT) (m)	-2.663 - 0.03191×(B1) + 1.1030×((B1)/(B3))	0.91 n=39	Single band and one band ratio with natural logarithm. Landsat 5 TM with 30 m resolution used on different quality lakes in Minnesota, USA
		e^(-2.663 - 0.03191×(B1) + 1.1030×((B1)/(B3)))		
Allan 2007	Chlorophyll-a ln(chl-a) (µg/L)	14.141 - 5.0568×((B1)/B3))	0.91 n=16	One band ratio with natural logarithm Landsat 7 ETM+ with 30 m resolution used on degrading and loaded lakes in Northern New Zealand.
		e^(14.141 - 5.0568×((B1)/B3)))		
Keith 2018	Chlorophyll-a Chl-a (µg/L)	60.703 × $((B1)^{-1} - (B2)^{-1})$×(B4) + 10.386	0.86 n=20	One band ratio with three bands B4/(B1×B2)

Mancino 2009	Chlorophyll-a Chl-a (mg/m^3)	-47.515 + 9.516×(B3/B2) + 20.952×(B1/B2) - 873.0×(B2) + 34.889×(B2/B1)	0.72 n=45	Three band ratios plus one band model. Mesotrophic volcanic lakes in Southern Italy. Landsat 5 TM
Giardino 2001	Chlorophyll-a Chl-a (mg/m^3)	11.18×(B1) - 8.96×(B2) - 3.28	0.85 n=4	Two band model. Landsat 5 TM on lakes in Northern Italy 30 m resolution
Assegide 2023	Turbidity TU (NTU)	282.88 × (B3/B2) − 206.15	0.92 n=21	Band ratio model with red and green band. Sentinel-2 on Koka Reservoir in Ethiopia.
	Total Suspended Solid TSS (mg/L)	3938.9 × (B3) − 536.9	0.67 n=27	One band model with red band. Sentinel-2 on Koka Reservoir in Ethiopia.

More information can be found in the original papers that were summarised in the table above.

2.3.1 Water Quality Parameters

The first thing an observer might notice about a water body is the clarity, and it is defined as the turbidity. This the way water scatters and absorbs the light instead of going straight down, and potentially bounce back from the ground beneath. (Gholizadeh^) The turbidity is based on the amount of floating material in the water. This could be organic matter, sediments, particulates and more. The murkiness, as Gholizadeh puts it, is measured by the turbidity and Total Suspended Matter/Solid (TSM/TSS), and directly associated with Secchi Disk Depth/Transparency (SDD or SDT). Another parameter is Coloured Dissolved Organic Matters (CDOM). It makes water yellowish brown in high concentration and it affects aquatic ecology and the carbon dynamics. (Gholizadeh) CDOM consists of decomposed or decomposing material in the water that can originate from runoff from land or from the water itself.[8] However, many scales exist and can differ in method or type of result, and they affect other parameters as well, such as chlorophyll-a(Gholizadeh et al., 2016). Therefore, several of these "murky" parameters were included in the literature review.

Chlorophyll-a

Chlorophyll is a green pigment that is found in plants. It helps plants absorb the energy from light and drives photosynthesis process. There are different types of chlorophyll, and chlorophyll-a is the primary photosynthesis pigment, and is found in plants and algae, as well as the hazardous cyanobacteria.[9] The Great Norwegian Encyclopaedia explains that algae are the cornerstone for practically all life under the sea and an important primary producer in lakes. Entire ecosystems are built on them, however, when the abundance of them nears a critical line, for example during algal bloom, it can cause ecological death.[10] (https://snl.no/alger) When the algae amount increases from, for example, fertiliser run offs and other nutrient pollution, they block out light for other plant species in the water. If they use up all the nutrients in the water and start dying off at once, then the decomposition process will remove oxygen from the water, hurting aquatic life there. Furthermore, some types of algae are poisonous for animals, and with

[8] https://www.fondriest.com/environmental-measurements/parameters/water-quality/chromophoric-dissolved-organic-matter/
[9] https://snl.no/klorofyll
[10] https://snl.no/alger

high enough concentration can cause damage to people drinking or bathing in these waters.[11]

Algal blooms occur naturally from nutrients upwelling from below or run-off from the soil. However, this has been further aggravated by several factor. Two direct contributors are agricultural and industrial, and one indirect is land use changes. Agricultural fertilisation[11] has led to nutrients leaking from the farms to the water sources. Improper sewage and industry waste handling is rich of nutrients that contribute to bloom. These anthropogenic activities have changed the nutrient balance in waters leading to harsher blooms with an exacerbated effect on the water bodies. Koka Reservoir has also had occurrence of blue-green algal bloom, including overgrowth of the invasive water hyacinth.(Abhachire, 2014)

Algal bloom is therefore something that should be especially aware of in drinking sources. At the later stages of a bloom, it is easily detectable as a coloured hue in the water. The colour depends on the type of algae. This is also how early detection of algae is done. By measuring the chlorophyll-a in water samples it can estimate the amount of algae there is. (Estimating and mapping chlorophyll a concentration in Pensacola Bay Florida using Landsat ETM data.pdf) describe the spectral properties of chlorophyll-a as strong absorption between 400-500nm (blue) and at 680nm (red), and peak reflectance at 550nm (green) and 700nm (NIR). (Yang et al., 2022) have found that when the chl-a concentration increases, the reflectance in the green and red bands increases, while the reflectance in the blue band decreass. Furthermore, the peak position moved from about 680nm to around 715 nm with the increase in the peak amplitude. They also found evidence of the reflection peak at 700 nm was important for calculation of chlorophyll-a concentration in inland waters. Their conclution was therefore: selecting the optimal bands depends on Chlorophyl-a concentration. [bad paragraph structure]

This result is also similar to different review article on water quality parameter estimation by remote sensing. Gholizadeh et al. presents chlorophyll-a's spectral features as strong absorption between 450 – 475 nm (blue) and at 670 nm (red), and peak reflectance near 550 nm (green) and near 700 nm (NIR). [The reflectance peak

[11] https://snl.no/algeoppblomstring & https://snl.no/bl%C3%A5gr%C3%B8nnbakterier

near 700 nm and its ratio to the reflectance at 670 nm have been used to develop a variety of algorithms to retrieve chlorophyll-a in turbid water.

3. MATERIALS

For this thesis several equipment, methods, and locations were studied and experimented on to answer the research questions about water quality estimation key components for the result.

3.1 Case Study Areas

Five lakes around Addis Ababa were studied. These lakes are all part of the same water catchment and draining area, i.e. Awash River Basin, and are connected by Awash River and its smaller branches.

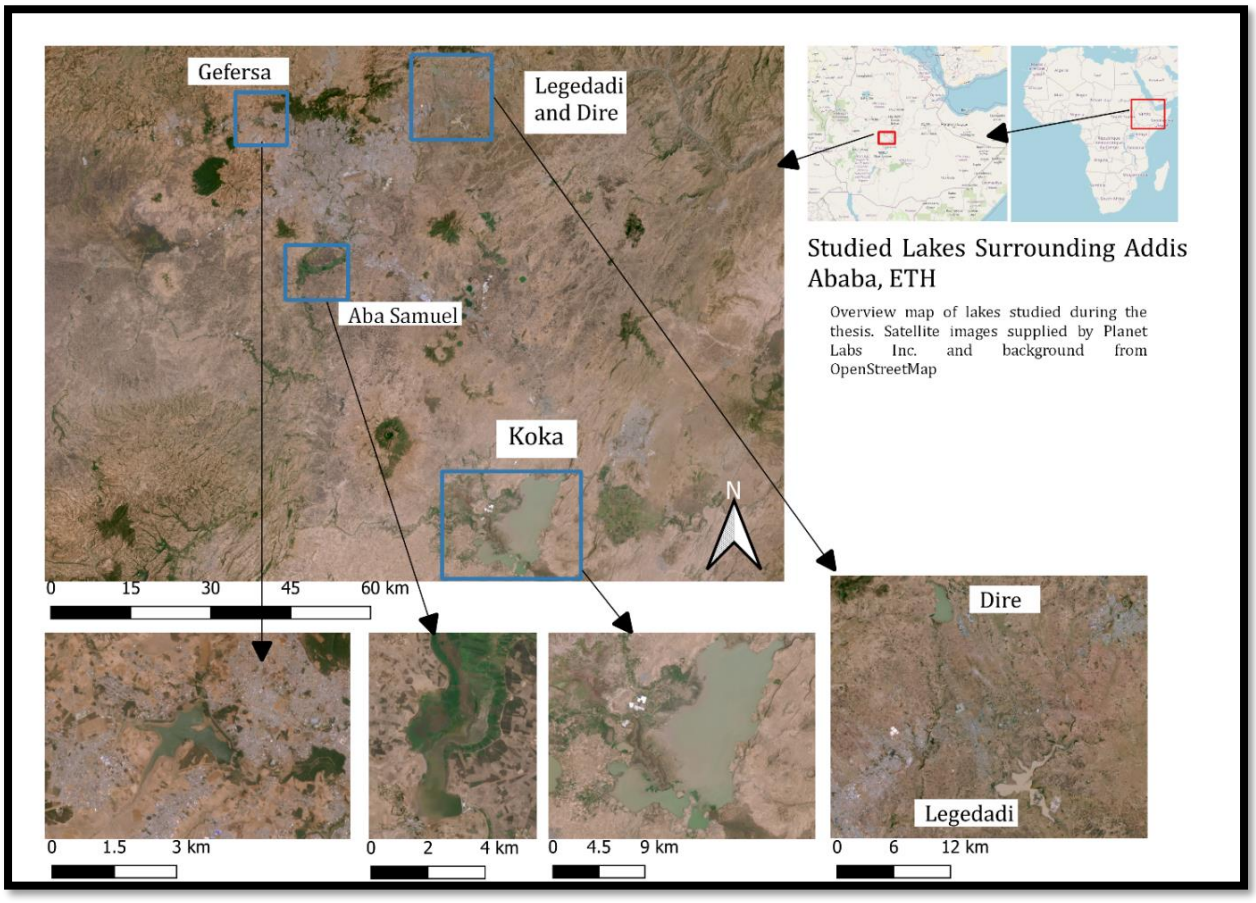


Figure 4 Overview map of all lakes in case study

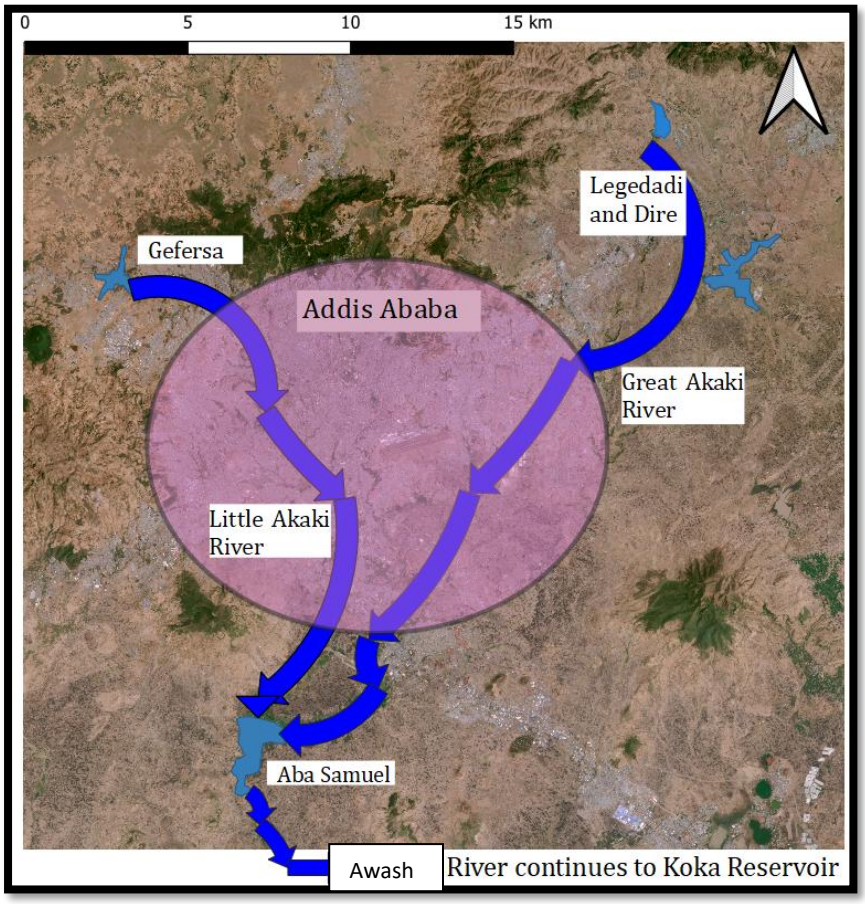


Figure 5 Simplified map of lakes and rivers around Addis Ababa. Inspire by Engida et al.'s figure.[12]

https://www.researchgate.net/publication/346321864_An_overview_of_water_pollution_status_in_Ethiopia_with_a_particular_emphasis_on_Akaki_RiverA_Review

Awash river and its branches are considered the most affected rivers of domestic, agricultural, industrial and recreational purposes in Ethiopia(Yohannes & Elias, 2017). Two of its branches, Little and Greater Akaki run through Addis Ababa, and are used as the city's "dumping ground" (Assegide et al., 2022). In addition to the indiscriminate dumping, Assegide et al. points out that the sextuples of fertiliser use between 1996 and 2015.

12 https://www.researchgate.net/publication/346321864_An_overview_of_water_pollution_status_in_Ethiopia_with_a_particular_emphasis_on_Akaki_RiverA_Review

As stated above, there are many issues surrounding surface water quality in Ethiopia, and a lot comes from Addis Ababa.

Awash River Basin, is populated by 18 million people, and they are directly affected by or affecting the rivers and lakes (Assegide et al., 2022). The Upper basin starts north of the city and are collected in the drinking water sources of Addis Ababa, Gefersa and Legedadi Reservoirs. Later this water flows through the city in Little and Great Akaki River, respectively, and ends up in Aba Samuel Lake. All of these are man-made reservoirs, however, Aba Samuel is only used as irrigation water and the dam holding it back is for electricity and silt filtration

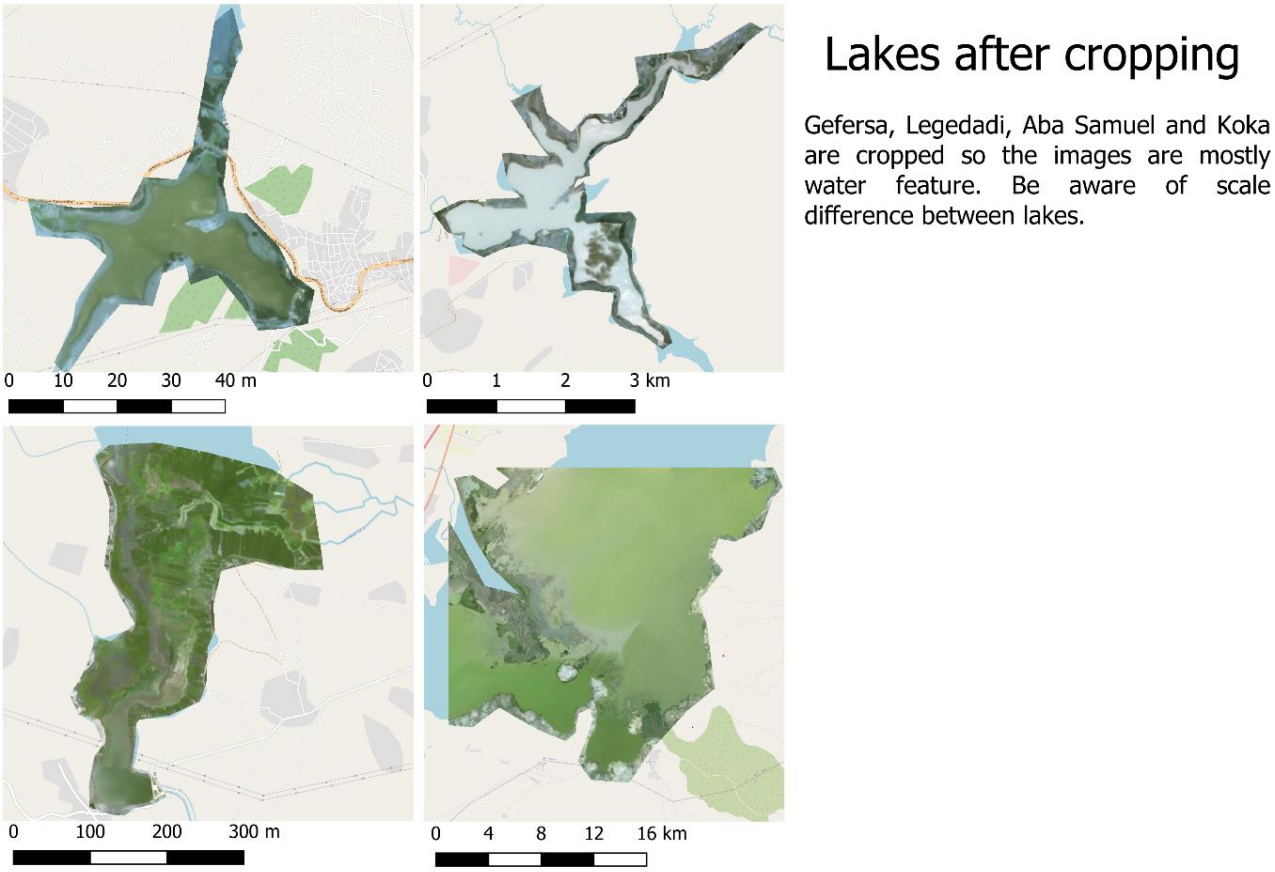


Figure 6 Overview satellite images of lakes in March 2023 except for Koka in February 2022.

Table 3 Location and type information for case study lakes around Addis Ababa.

Visited lakes	Location (Distance)	Type	Uses
Legedadi (+Dire) Reservoir	9.15° N 38.93° E (10km East)	Freshwater reservoir	Dam built in 1971 for drinking water. 44 million m^3 (Dire 19 million m^3)
Aba Samuel Reservoir	8.79° N 38.71° E (9km South)	Irrigation water and hydroelectric	In situ testing Nitrate test strip Laboratory testing
Gefersa Reservoir	9.06° N 38.64° E (5km West)	Freshwater reservoir	Main dam built in 1938 for drinking water storage. 9 million m^3

3.1.1 Legedadi and Dire Reservoir

Legedadi Reservoir is a manmade lake built to supply water to Addis Ababa in 1967. It is located 9.06°N and 38.98°E East of Addis Ababa. The water is piped to Legedadi Water Treatment Plant before being sent to consumers. Wastewater from the treatment plant and excess flow is sent down the Great Akaki River.

Dire Reservoir is a separate freshwater lake 9 km north of Legedadi Reservoir. It is located at 9.16°N and 38.93°E. The water is like Legedadi, and it is transported to the same treatment plant though pipes. It acts as an additional water source for Addis Ababa.[13]

3.1.2 Aba Samuel Reservoir

Aba Samuel changes throughout the season. From a width and length of 1 km by 9 km after the rain season, down to <10m narrow river in the dry season (Ingwani et al., 2010). Located south of Addis Ababa, at 8.79° N 38.71° E, it catches both Akaki rivers and sends them though Aba Samuel Dam, that is used for sedimentation and

[13] Personal correspondence with Zeleke Teferi from Addis Ababa Water and Sewage Authority.

hydroelectricity. Highly polluted water and algal and hyacinth growth is visible from space in the rainy season.

3.1.3 Gefersa Reservoir

Situated west of Addis Ababa is Gefersa Reservoir. More specifically at 9.06° N 38.64 and is a drinking water source for the capital. The reservoir is held back by the Gefersa Dam, and spills down Little Akaki River through Addis Ababa. As a drinking source the reservoir has had adequate results from a water quality study from 2020 (Mekuria et al., 2020). Though still some trace elements of heavy metals where found, but it was not as bad as the quality of Little Akaki River flowing through Addis Ababa.

3.1.4 Koka Reservoir

The Koka reservoir is a lake that covers an area between 90 km^2 and 152km^2 after its dry and wet seasons in June and October, respectfully. It is located at 8.43° N and 39.18° E at 1588 meters above sea level at the dam outlet (Assegide et al., 2023).

Table 3 shows the laboratory results from Assegide et al.'s field samples in Koka Reservoir between the 25th and 26th of February 2022.

Table 4 Laboratory results for water quality in Koka Reservoir Febrary 2022 by Assegide et al. 2023. Last values are the average for each parameter.

id	Chlorophyl-a µg/L	Turbidity NTU	Total Suspended Solid	id	Chlorophyl-a	Turbidity	Total Suspended Solid
1	3.475	38	218	**15**	19.112	36	246
2	18.243	38	286	**16**	16.062	-	197
3	12.162	-	222	**17**	20.849	40	247
4	23.456	44	288	**18**	19.112	52	212
5	21.718	52	228	**19**	17.375	46	402
6	16.506	46	308	**20**	18.687	40	226
7	21.718	64	210	**21**	19.112	52	223
8	17.031	100	338	**22**	14.768	52	827
9	18.849	48	860	**23**	17.012	48	235
10	10.425	42	192	**24**	52.718	-	318
11	105.98	-	514	**25**	77.375	44	606
12	49.517	34	436	**26**	396.14	148	317
13	15.212	-	247	**27**	-	72	227
14	17.819	-	226	**$\bar{X}$**	40	54	328

The sample sites coordinates were determined using a hand held "Garmin 60 s GPS receiver" (Assegide et al., 2023) that has an accuracy up to 3-5 meters, [which is typical for handheld devices] (https://static.garmin.com/pumac/GPS60_OwnersManual.pdf).

The coordinates were made available by the author[14] in UTM 37 N Adidan (ETRS: 20137), a projection with a ground accuracy up to 6 metres in Ethiopia (https://epsg.io/20137). The specific method of sampling for these parameters are all described in great detail by Assegide et al. 2023.

The Sentinel-2 sample sites reflectance is presented in figurex. The figure is a modified version of Assegide et al.'s 2023 graph so only relevant bands are presented. It is uncertain what happened to their 27th sample site .

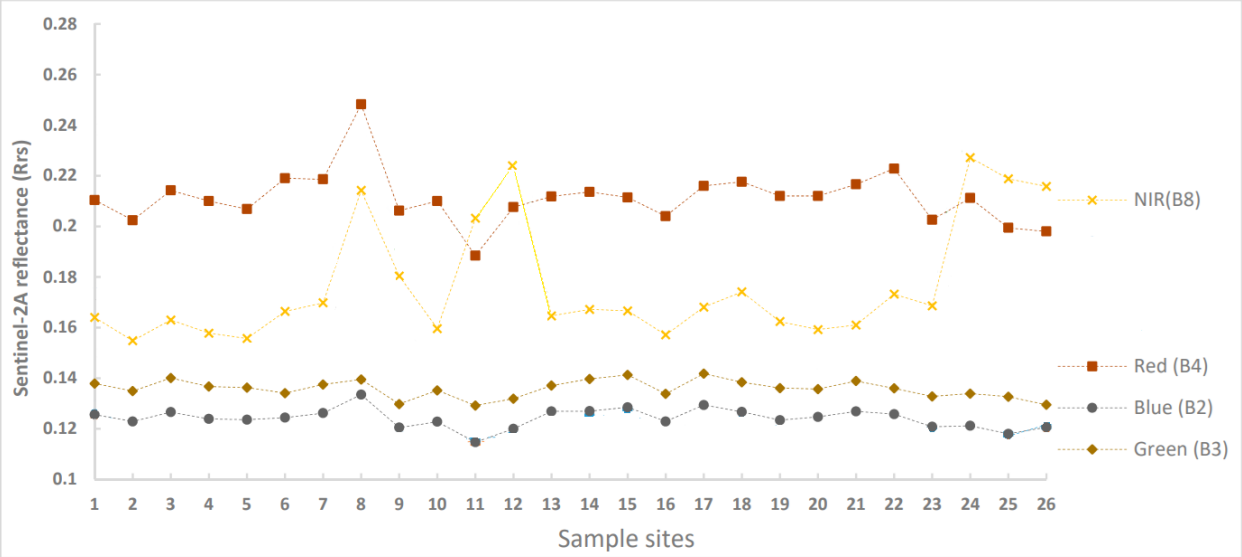


Figure 7 Reflectance on Koka sample sites with Sentinel-2 by Assegide et al. 2023. Modified graph where only relevant bands are included.

Assegide: Red shows the highest reflectance (characteristically) with exceptions to sample sites 11, 12, 24, 25 and 26 where NIR jumps up and takes the lead . These sites are also where NIR seams to deviate from a steady state. Sample site 8 has the highest total reflectance, while there is a dip at site 11 except for NIR band. When comparing with NICFI, these trends are slightly off.

3.1.5 Meteorological conditions

Weather data acquired from the Ethiopian Metrological Institute during field study in Addis Ababa. The data was ordered for a symbolic cost of 100 Ethiopian Birr (aprox 20

[14] Personally met Endaweke Assegide at Water & Land Resource Centre – Addis Ababa University and received his data.

NOK/2USD). The data was comprehensive with daily observations from two weather stations. These stations, “Bole” and “Obs”, are in Addis Ababa. One by Bole Airport and second by the Ethiopian Metrological Institute’s offices by Black Lion Hospital (Tikur Anbessa). These weather stations are between 13-80km away from the lakes. The institute insisted during acquisition of this data, that the distance between the weather station and Aba Samuel Reservoir was no issue. However, later this was not taking into account the 80 km to Koka Reservoir. Additionally, any gap in the graph below is due to missing data in the metrological database.

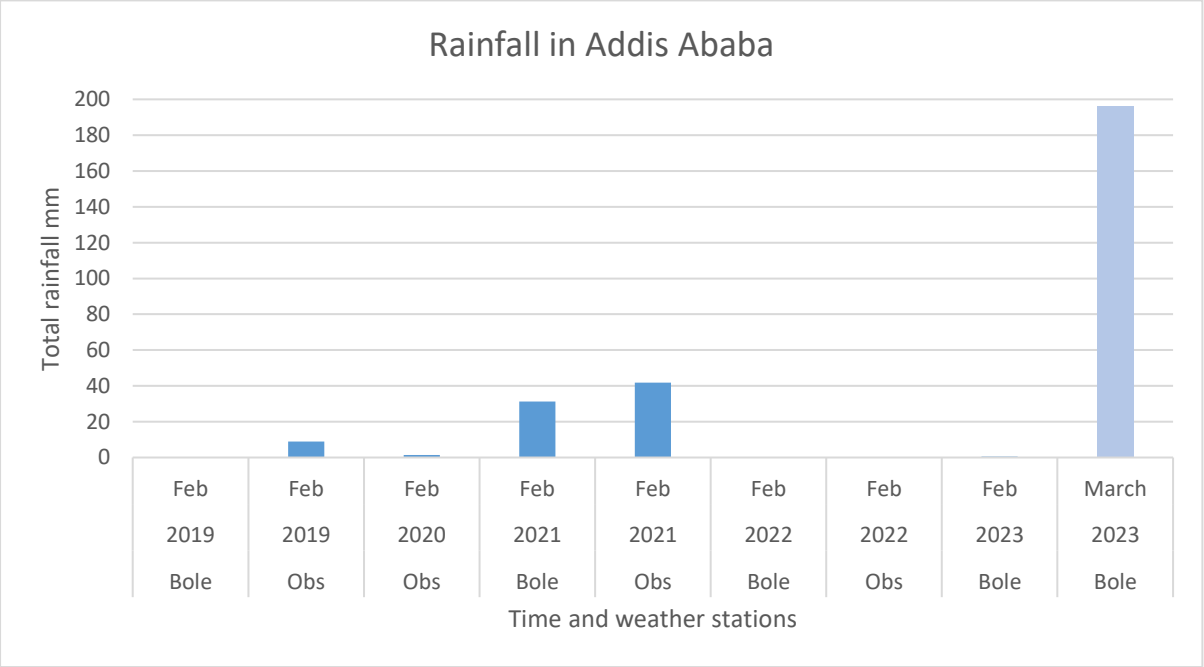


Figure 8 Rainfall per February and measuring station in Addis Ababa, Ethiopia. March 2023 was added to show the difference between February and March. Also the time weather during field trip to Addis Ababa.

Figure of rainfall by month and measuring station. The March 2023 was added to show the difference between February and March. Rainfall per day in March 2023 was also acquired.

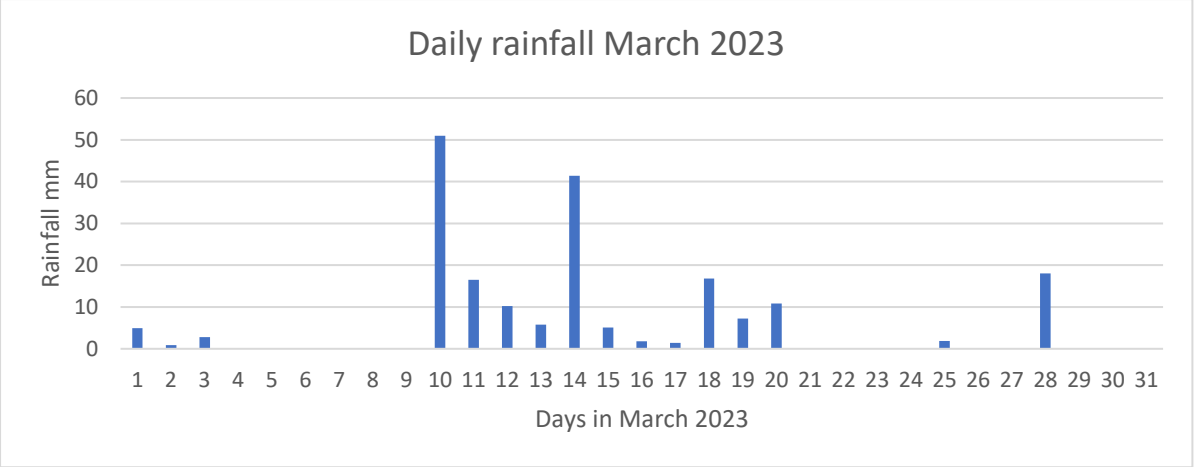


Figure 9 Daily rainfall during March 2023. Field trip to Addis Ababa was conducted during this time frame.

3.2 Satellite Images

3.2.1 Satellite data

The satellite images were from Norway's International Climate and Forest Initiative (NICFI) Satellite Data Program. Their high-resolution images were taken from the Rapideye and Dove nano satellites produced by Planet Labs PBC. These satellites are 10x10x30cm and weigh about 4 kg (Houborg & McCabe, 2018). The images are available for the tropical regions of the planet with monthly updated Basemaps from September 2020 to August 2023.[15] Before September 2020 there are biannual Basemaps back to December 2015. As mentioned earlier, there are four bands available with wavelengths presented in Table 1. The colours are Blue (B1), Green (B2), Red (B3) and NIR (B4).

The Basemaps are normalised so that the product of combined sensors look alike and are harmonised with Sentinel-2 and Landsat 8. This culminates to the "Normalized Analytical Mosaic" with 4 bands and unsigned 16-bit data depth for the reflectance values. The mosaics are composed of "PlanetScope Ortho Scenes" meaning that the images are geometric, radiometric, and atmospherically corrected. This results to a spatial resolution of 4.77m per pixel for their surface reflectance product packed in a

[15] https://assets.planet.com/docs/NICFI_General_FAQs.pdf

GEOTIFF file. The reflectance values are scaled by 10 000 to reduce quantisation errors (rounding off errors). The Digital Numbers (DNs) are values between 0-10 000, but the bit depth allows for values up to 65535. To get the reflectance one must divide the DN with 10 000.[16]

3.2.2 Temporal image information

Throughout a month, Planet Labs' satellites record all the images that will be part of that specific monthly mosaic. Since the mosaics are compiled after the month of capturing, it takes up to two weeks into the next month before the Basemap is available. In this time Planet Labs' "Basemap production algorithm" works on producing the best mosaic by removing issues like glare and clouds, as well as calibration and corrections, such as sensor specific and atmospheric. This leads to a mosaic of many different images from different days, but the result is clearer and of higher quality, though they do not guarantee 100% cloudless images.

However, for an empirical analysis, the images have to be captured as close to the field sample date as possible, so that the condition for correlation is "true". Too many of days off, and there could be shifts in the water body. The images would then not be capturing the same "lake" the sample was taken from.

The three basemap tiles that cover lake Koka were captured on different days and satellites throughout February 2022. This is similar for the other lake images as well.

3.3 Image processing

Image processing and raster analysis was done using Quantum GIS (QGIS) and its built in tools and plug ins. When the pixel data was extracted from the satellite images, Excel was used for statistical analysis and empirical model building.

[16] https://assets.planet.com/docs/Planet_Combined_Imagery_Product_Specs_letter_screen.pdf

3.4 Water Quality Assessment Equipment

During the field study, water quality equipment was used to determine quality and quantity of select parameters based on available equipment and expertise.

3.4.1 Nitrate strip

Nitrate was measured with MQuant's colorimetric test strips. These test strips measure nitrate in the form as NO3 and NO3-N, with concentrations between 0 - 500 mg/L for NO3 and 0 – 113 mg/L for NO3-N.[17]

3.4.2 In situ tester TRACE2O Hydrolite HT102-EC

In addition to the test strips, in situ water testers were used to determine several parameters. pH, Electric conductivity, Total Dissolved Solids and temperature was measured using Trace2O's Hydrolite HT102-EC and HT101-pH. These work by taking water samples and dipping a digital sensor to measure the different parameters. Information on how this works is described in the manual.[18]

3.4.3 Laboratory testing of field water samples

Laboratory tests were conducted using colorimetric method by a photometer (DR7100). These tests gave results for 8 parameters: Phosphate, Phosphorus, ammonia (NH3), ammonium (NH4), nitrate (N-NO3 and NO3), nitrite (NO2) and nitrogen (N). These test were conducted by the Ministry of Water and Energy by Yirgalem Esuneh Endalew.

[17] https://www.fishersci.com/content/dam/fishersci/en_US/documents/programs/scientific/brochures-and-catalogs/brochures/emd-millipore-mquant-mcolorphast-brochure.pdf
[18] https://www.manualslib.com/manual/2448209/Trace2o-Aquasafe-Wsl25-Plus.html?page=76#manual https://www.trace2o.com/_files/ugd/f8d288_a0018fc347bc46b3b13cd62e59bcdb8f.pdf

4. METHODS

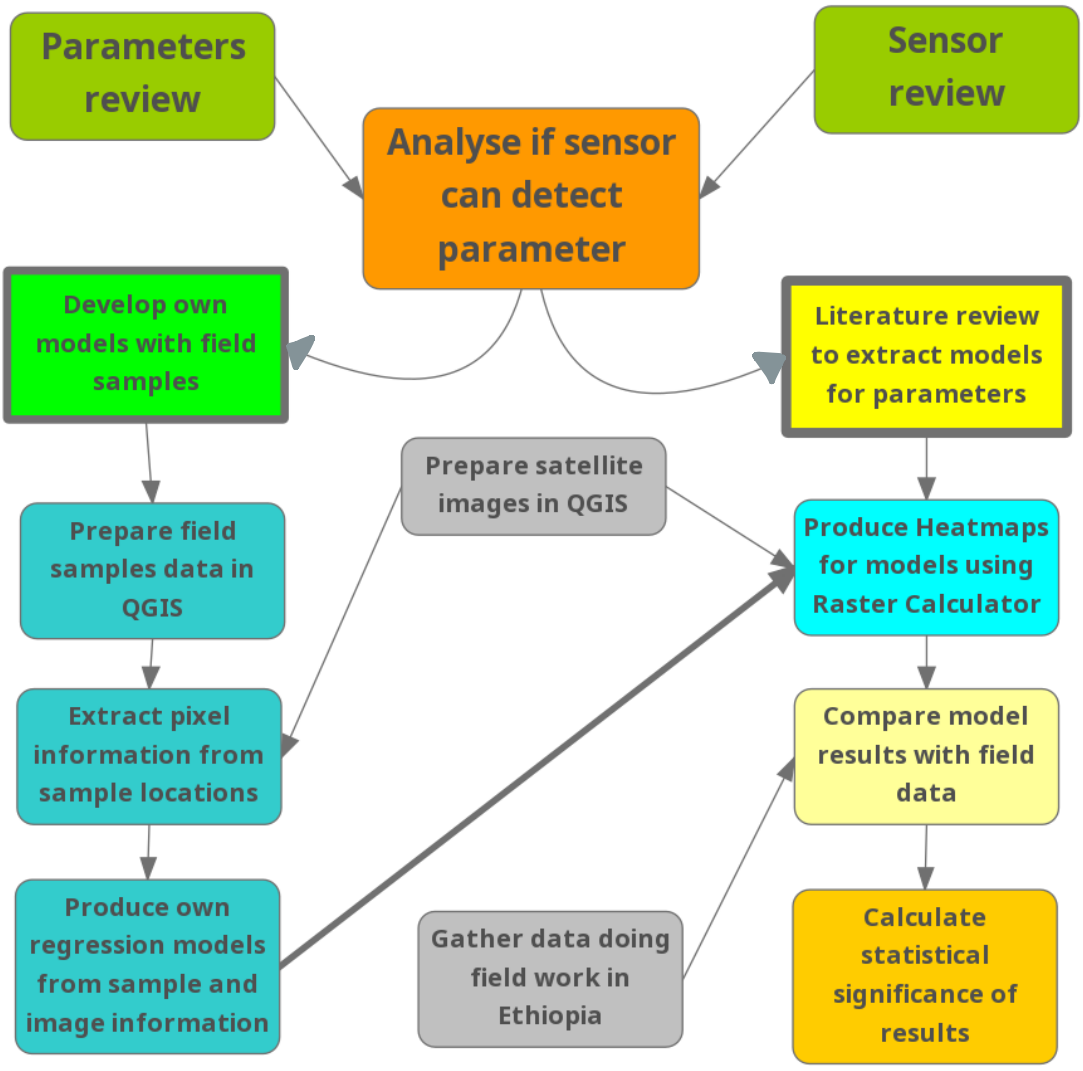


Figure 10 Flowchart of Methodology. Made using mindmup.com

Figure above is a simplified flowchart of the methodology in this thesis. Two key direction are presented by the two arrows that come out of "Analyse if sensor can detect parameter". One methodology was to develop an Original Water Quality Model and subsequent Novel Parameter Quality Maps. The other methodology was to extract the existing WQMs to use them on the case study lakes and instruments.

4.1 Field work around Addis Ababa

Surface water samples were taken from the lakes surrounding Addis Ababa during the field trip in March 2023. The various lakes were visited and tested in slightly different ways.

The table below gives an overview of the tests conducted. In addition to the water quality tests the areas were visually inspected/surveyed .

Table 5 Overview of field work done around Addis Ababa. Includes approximate distance from Addis Ababa city boarders.

Visited lakes	Location (Distance)	Date	Type of water quality tests
Legedadi (+Dire) Reservoir	9.15° N 38.93° E (10km East)	11/03/2023	Nitrate Test
Aba Samuel Reservoir	8.79° N 38.71° E (9km South)	14/03/2023	In situ testing Nitrate test strip Laboratory testing
Gefersa Reservoir	9.06° N 38.64° E (5km West)	21/03/2023	Laboratory testing

4.1.1 Legedadi and Dire Reservoir field work

Legedadi Reservoir and the connected Dire Dam was visited 11th of March 2023. At Dire Reservoir a nitrate test was conducted using MQuant's Nitrate Test. The test strip was dipped on the surface of the water for one second. After the dip, the strip was airdried for one minute before reading the results.

Nitrate test was not conducted at Legedadi Reservoir, but the water at Dire is pumped to the Legedadi Water Treatment Plant together with the lake water from Legedadi itself, and the experts consider them similar enough.

Nitrate test taken at 12:58 at 9.1494679° and 38.9315299°.

The weather was overcast.

Visual inspection of the reservoir was also conducted.

4.1.2 Aba Samuel Reservoir field work

During the visit, two areas were used as testing sites.

Figure 11 Sample sites at Aba Samuel Reservoir. Yellow = Site AS-A and Red = Site AS-B

The tests conducted on the sites were similar except for an issue at site Aba Samuel A (AS-A). Due to deep mud, there were issues with walking the remaining 100m to the shore. Therefore, a local worker for the nearby groundwater pump station, Dega Gidimga, assisted by walked the remaining distance to gather water sample. Therefore, no visual inspection was done on the shoreline.

Sampling at site Aba Samuel A (AS-A) was conducted 11:55 at 8.7979697° and 38.7121640°. The second sampling was conducted on site Aba Samuel B (AS-B) at 12:56 at 8.7886237° and 38.7056267°. Both were on the 14th of March 2023. The water was gathered in an ordinary plastic drinking water bottle of 0.5 litre[how to say in a concise way??]. The bottle was filled from the shore and rinsed three times before being filled up as a sample. After filling and retrieving the bottle, in situ testing with TRACE2O Hydrolite HT102-EC and HT102-pH was conducted on the sample water. The set

instructions for the device were followed by expert Yirgalem . The measured parameters were pH, Electric conductivity, Total Dissolved Solids, and temperature.

Additionally, MQuant test strips were used to check the nitrate in the sample. At site AS-B this was done by pouring some water into the bottlecap and dipping the nitrate strip there. This was to avoid contaminating the water sample with the paper and its indicating chemicals.

The bottle samples were sent to the Ministry of Water and Energy with Yirgalem to conduct the laboratory testing. There was about four hours between the first sample was taken and arriving at the Ministry. The laboratory testing was colorimetric method by a photometer (DR7100). The parameters tested were phosphate, phosphorus, ammonium, ammonia, nitrate nitrogen, nitrate, nitrite, and nitrogen.

The weather during the field work was overcast with some cloud breaks later in the day, but no direct sunlight.

Visual inspection was also conducted at AS-B.

4.1.3 Gefersa Reservoir field work

Gefersa Reservoir was visited 21st of March 2023 guided by Solomon from Addis Ababa Water and Sewage Authority (AAWSA). Two water samples were collected from the shore using 0.6 litre store brought water bottles. The bottles were rinsed three times in the lake before being filled.

First sampling was Gefersa A (GA) conducted 12:50 at 9.0629466° and 38.6414824°. The second sampling Gefersa B (GB) was conducted 13:08 at 9.0646032° and 38.6420233°.

At sample site GB there were trees close to the shore above the test site. This could have an impact on the satellite image. Samples were taken around 13:00 and were placed in a fridge at 16:02. The next day the samples were given to the Ministry of Water and Energy (MoWE) with about 4 hours between coming out of the fridge until arrival at their offices. The laboratory test conducted was colorimetric method by photometer (DR7100) again.

The weather during the field work was slightly overcast with breaks in the clouds.

4.2 Production of Parameter Distribution Maps for Literature Study Water Quality Parameters

The literature study produced several models that are applicable on NICFI's database. They are presented in Table 2 in the Theory chapter. The methodology to go from literature model to a map in QGIS can be explained using the model presented by Miller et al.. They produced a model to estimate Total Suspended Matter Concentration (C_TSM). The model is a linear regression using one band from the MODIS multispectral sensor. As this sensor is different from NICFI's, one must find the wavelength of this band and match it up to NICFI's available bands. In this example, Miller et al. used MODIS' first band with a wavelength width between 620-670nm. This band would correspond with NICFI's harmonised band 3 – Red 640-670nm. There is some difference [as mentioned earlier], but it is the closest fit, as both are considered the "red band" of the sensors. In addition, the band is multiplied by the radiometric scale factor of 0.01 so that the DN's are now radiance. This is only done to bands that are not part of a ratio, as the ratio process would not be affected by both dividend and divisor being multiplied by the same number.

Now that the bands and formulas were identified, and corrected in some cases, QGIS' Raster Calculator was used to create the model with the NICFI image. This generated a heatmap for C_TSM, however the initial map looked wrong as the MINMAX values have not been adjusted yet. Since the image is whole and contains land features surrounding the lake, the model would misinterpret this information and attempt to make an estimation of a water quality parameter. This could have been solved by clipping the water body out of the image and work strictly on that, however, tweaking the MINMAX of the displayed heat map by visually finding the highest and lowest values in the water body was just as accurate, though more time consuming and manual. [move last part to discussion].

To conclude the example, the result of this process was a better visual representation of the C_TSM for the water body. This example methodology was applied to all water quality parameters found and tested in the literature study. There are differences

between bands and ratios, and they are highlighted in the same table. More information about the individual methodologies can be found in their respective sources.

4.3 Produce Original Water Quality Model on Koka

Creating a new regression based model requires joining the in situ samples data with the images. For this thesis the samples are from Endaweke Assegide that was used in his paper *Spatiotemporal Dynamics of Water Quality Indicators in Koka Reservoir, Ethiopia*.

4.3.1 Preprocessing

Importing image

First step was to import the images for Koka. Koka is so large that it takes three basemap tiles, so three images will be used. Therefore, it required repeating any further processes for each tile when the images are used.[improve]

Importing sample sites

To extract the DN of the image from the sample site, one must add the site coordinates into the QGIS project. Before that, the coordinate systems of the image and sites must be concurring. The samples were supplied in the local Ethiopian projection ETRS: 20137 UTM 37N Adidan while the images are in EPSG: 3857 WGS84/Pseudo-Mercator. Therefore, it was decided to reproject the 27 coordinates from Adidan to Pseudo-Mercator using the Python package PROJ.

```
from pyproj import Transformer, transform
transformer = Transformer.from_crs(20137, 3857)
new_coord = transformer.transform(Easting, Northing)
```

image of the code used in Python. Easting and Northing are the sample coordinates in a list separated by X and Y direction .

The transformation output was a tuple containing the new X and Y coordinates in Pseudo-Mercator. The unit was given in meters [explain why], [remove].

The transformed samples site location information was plotted to a Comma-Separated Values file (CSV-file). This CSV file was added to the QGIS project and the “Create Points Layer From Table” tool was used to produce vector layer containing the samples .

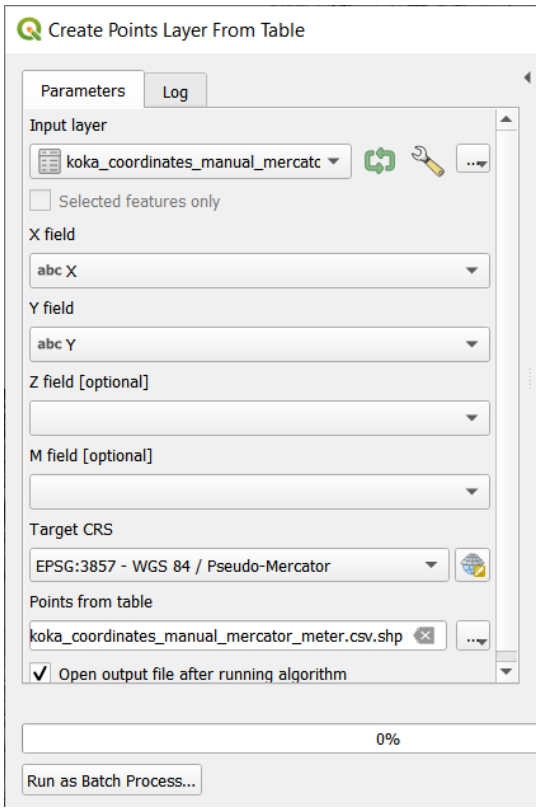


Figure 12 Screenshot from "Create Points Layer From Table" tool in QGIS. Settings used are visible

Input layer was the CSV file with coordinates and site numbers. Figure 12 shows the settings used. The X field is West/East and Y field is North/South. “Target CRS” is the coordinate system the output points will have. The result was a vector layer with points signifying the location of a sample site. These points had only fields containing site number and location information. The next step was to change the site from being a point vector to a polygon containing the 3x3 pixel window with the sample site in the centre pixel.

Buffering from point to polygon

Several buffering methods were attempted using QGIS tools, and the resulting method was manually creating polygons in the QGIS project. The first step was creating a new shapefile layer with one field containing the ID, which had to concur with the sample site ID. The geometry type was polygon with the spatial reference set to Pseudo-

Mercator. Second step was manually creating a 3x3 polygon surrounding the sample site pixel. Each polygon ID must have the exact same site ID, as these polygons will act as the sample sites from this point on.

Now with the polygons created, it was possible to use QGIS' "Zone Statistics" to extract DN values from the raster below the polygons, specifically the average value of these 9 pixels in the 3x3 polygon. However, "Zonal Statistics" can only input one raster layer, one band from that raster, and one polygon layer. Meanwhile, the lake has three raster tiles with four bands where the sample sites are scattered. Therefore, a model was built in QGIS Model Builder to automate the process.

Extract pixel values around the points

The figure below explains the process. This model extracted the values in the three by three polygon surrounding the sample site. The rasters are named after their cardinal direction from each other, so the NE raster is the image that is to the top right, and so on.

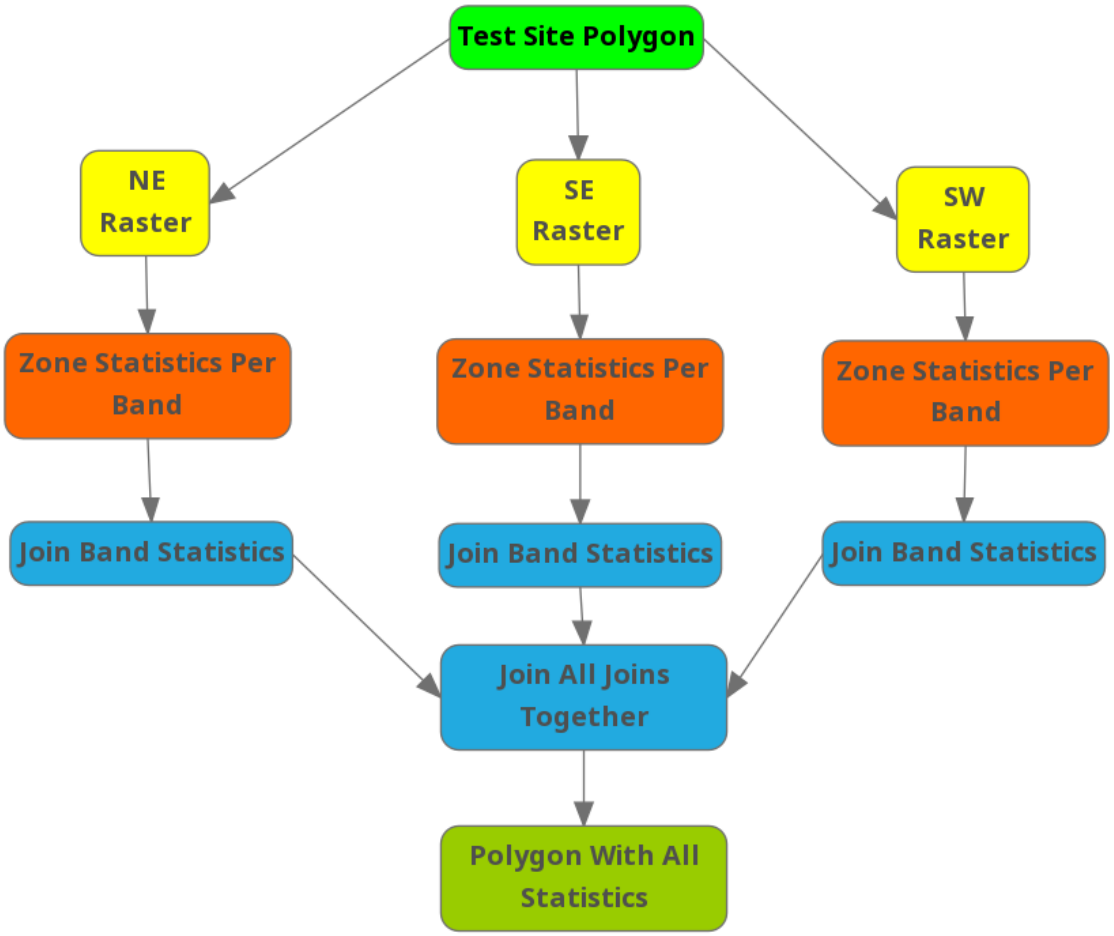


Figure 13 Simplified flow diagram of QGIS Model Builder for joining raster information with sample sites. Made using mindmup.com

Inputs for this model were the three rasters and the 3x3 polygon shapefile, shown as the yellow and lime-green boxes in the figure. The model took one raster and one of its bands and implemented "Zone statistics" for the area on the raster covered by the polygon. Since "Zone Statistics" only applies to one band at a time, so for each raster, the statistics must be calculated four times. The statistic of interest is the mean of the Digital Number in the band.

The mean value of each band in each raster was joined together using "Join attributes by field value", as the first layer of the model creates four vector layers with the mean values. This joining tool combines only two objects at a time, so this was repeated several times in the model until all the means are combined into one vector layer. This process is seen in the second and subsequent layers in Figure 13. The result was an object containing all the mean values of all bands in each raster.

The resulting table was messy, as the joins included everything in each joining. This meant that "Zonal Statistics" were attempted on raster bands without containing the specific polygons. There were 36 fields, and only 4 of them contain the mean value, and one field for identifying the sample site. However, it was easy to manually extract the data, as unrelated fields were either repeating sample site IDs or blank. This automation model is not optimised, but the relevant results were easily extracted, and the processing time is negligible compared to manually processing the statistics and joining tools. Furthermore, this model can also take in any polygon vector layer and three rasters to produce an output. The shapes or locations do not matter.

4.3.2 Water Quality Regression Model in Excel

Water Quality Parameters Model Development in Excel

To develop the models for the selected water quality parameters the extracted DN values from QGIS were combined with the water quality sample data from Koka in Excel based on sample location/ID. This is added to appendix A-5. Each sample site has now four columns for DNs and one column for each measured parameter, Chlorophyll-a (chl-a), Turbidity (TU) and Total Suspended Solid (TSS). However, some sites are missing observations.

Empirical analysis was used to create a model to predict water quality parameters from remotely sensed images. The built in Data Analysis Tool from Excel was implemented to calculate the regression model. Inputs for this tool were field observed water quality parameters as the "Y Range" and the DN for each band were the "X Range". This regression method is repeated for each water quality parameter.

However, some of the sample observations are missing for chlorophyll-a and turbidity. These parameters were rearranged so that the blank data was not included in the regression model. Furthermore, the authors of the Koka article omitted some observations in their model development for chlorophyll-a, as they determined them as outlier. Omitting outliers identified by Assegide et al. 2023 was done in this step of the methodology.

Correlation

Correlation was calculated using the Data Analysis Tool in Excel. DNs and sample data are used as input and the correlation is calculated for all variables.

Band ratios as new dependent variable

Band ratios were created by taking one band's DN and dividing it with a second band's DN. The quotient was a new variable that was used for model development, specifically as part of the independent variables. These band ratios were used in the same manner as the studied models from Mansino, Allen and Kutser. Meaning they were either combined with together or used individually for the regression model development. Different combinations were tested for their R^2 values.

4.4 Produce Novel Parameter Displacement Map

The multivariate equation created by the regression model was used to make maps of the water quality parameter on Koka. This was the same method as described earlier with the models extracted from literature, meaning QGIS Raster Calculator was used to create the maps.

5. RESULTS

5.1 Field work around Addis Ababa, Ethiopia

Results from the field work sampling and visual inspection

The table below are all the test results from the field work conducted.[19]

Table 6 Results from in situ water quality testing March 2023. µS = microSiemens = $\mu\Omega^{-1}$ |

Parameter [unit]	AS-A	AS-B	Dire Dam
pH	7.4	7.51	-
Electric conductivity (EC) [µSiemens]	495	694	-
Total Dissolved Solids (TDS) [mg/L]	248	348	-
Temperature [°C]	23.7	25.1	-
Nitrate NO3 (Test strip) [mg/L]	25	50	10
Nitrate NO3-N (Test strip) [mg/L]	5.6	11	2.3

Table 7 Laboratory results from field samples March 2023. Results supplied by Yirgalem Esuneh Endalew from Ministry of Water and Energy. All values are in mg/L.

Sample Site	Phosphate	P	NH3	NH4	N-NO3	NO3	NO2	N
AS-A	37.9	12.5	0.52	0.55	20	88	1.32	0.4
AS-B	19	6.3	0.44	0.47	25	112	4.6	1.4
GA	5.2	1.7	0.18	0.19	19	82	0.056	0.017
GB	5.3	1.7	0.27	0.29	21	92	0.049	0.015

Images and observations are presented in the sections below.

Legedadi and Dire Reservoir

[19] Laboratory results: Result supplied by Yirgalem Esuneh Endalew

Figure 14 Photographs from Dire Reservoir. Sample Site is in left picture.

Water is slightly brown and turbid. The rocky area smells like algae, however when moving on the bridge over the reservoir there is no smell. Explained as algae in the rocks forming over time while the water has been treated against algae. The algae is therefore only in the rocks and not the water.

Aba Samuel Test Site A (AS-A)

Figure 15 Photographs overlooking Aba Samuel Reservoir close to Sample Site A.

Too muddy to walk to the shore line. Abundance of trash left behind as the lake shrinks throughout the season. Plastic containers and shoes were easily spottable on the field.

Aba Samuel Test site B (AS-B)

Figure 16 Photographs from Aba Samuel Reservoir Sample Site B.

Gefersa Reservoir

Gefersa Test Site A (GA)

Figure 17 Photographs from Gefersa Reservoir Sample Site A (GA). Colour of water and transparency is accurate compared to personal observation.

The site area appeared as a place where birds gather, as several were observed flying and swimming away when approached. The area was littered with bird feathers and waste.

Gefersa Test Site B (GB)

Figure 18 Photographs from Gefersa Reservoir Sample Site B (GB).

Very similar water as GA site. [ADD more] Appeared less green, but could have been the direction of the Sun.

5.2 Parameter Distribution Maps of Literature Water Quality Models

Nine Water Quality Models (WQMs) where found and calculated for. The Parameter Distribution Maps (PDMs) are presented below using QGIS's Print Layout.

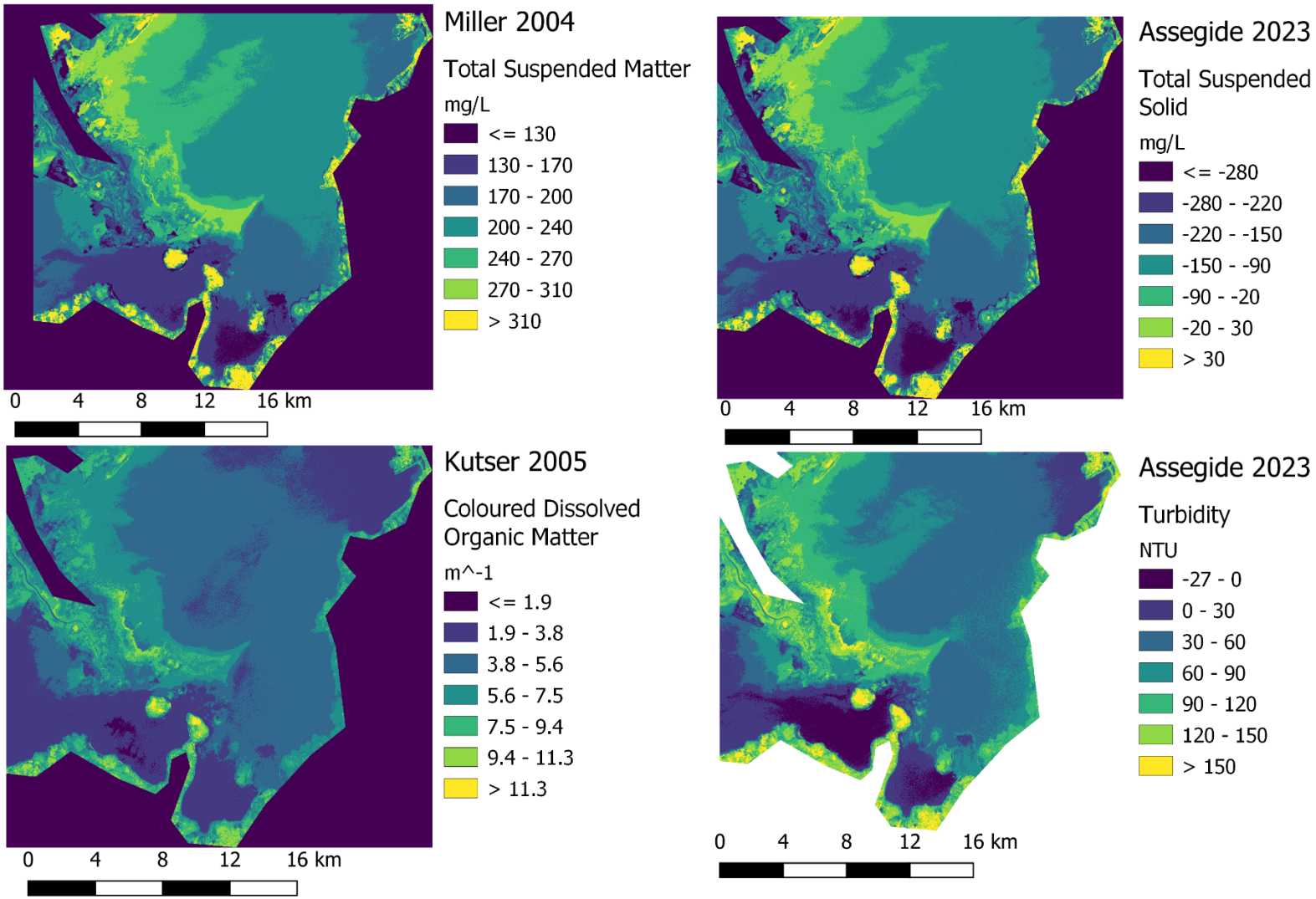


Figure 19 Parameter Distribution Map of Koka Reservoir South East Tile for February 2022. Miller 2004 TSM (forced positive), Assegide 2023 TSS, Kutser 2005 CDOM, and Assegide 2023 TU. Dsaa

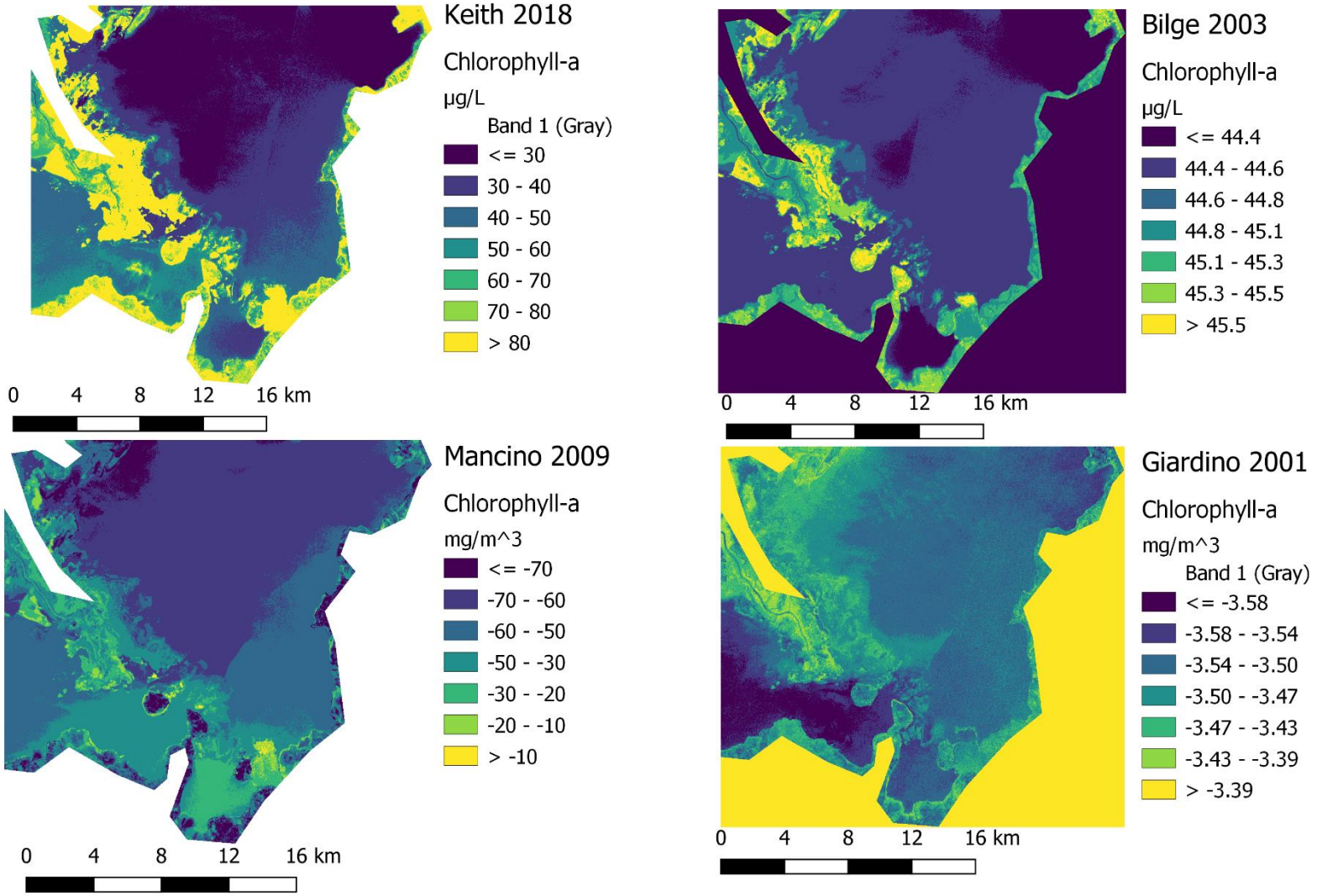


Figure 20 Map of Chlorophyll-a distribution based on literature formula on Koka Reservoir. Keith 2018, Bilge 2003, Mancino 2009 and Giardino 2001.

Figure19

Miller 2004 Total Suspended Matter Concentration C_TSM (mg/L)

The Parameter Distribution Map (PDM) for TSM in Koka Reservoir shows values around 100 to 310 mg/L in the water. The south point has the lowest concentration while the highest is in the centre of the map, moving along the coast in a north east direction. The bright yellow spots in the middle are land masses, so should be ignored.

Assegide 2023 Total Suspended Solid

The top right map in Figure 19 shows the TSS spread being -300 to 20 mg/L in the reservoir. The map is very similar to Miller 2004, but differences can be seen in the centre of the image and up a bit, as there is different spread of the third brightest green [move disc?]. Similar distribution as Miller 2004, with lowest values in south point, though all the values in the water are negative except at the same high concentration point as Miller 2004.

Kutser 2005 Coloured Dissolved Organic Matter CDOM (m^{-1}) [update ?]

The spread of CDOM is between 1.5 and 8 m^{-1} in the water. Lowest value is found south southwest of the lake. [improved image with better spread update?]. Difficult to see where the shore ends and the water starts.

Assegide 2023 Turbidity

The turbidity values in the water are between -27 and 120 NTU according to the PDM. Just like the other maps in Figure 19 the distribution is similar, as in lowest values at the south and south southwest, and higher values in the centre of the map.

Figure 20 contains four chlorophyl-a Parameter Distribution Maps from different WQM.

Figure 20

Keith 2018 Chlorophyll-a Chl-a (µg/L)

The top right PDM shows Keith 2018's distribution. The concentration of chlorophyl-a in the water is between 20 and 120 µg/L. The lowest values are from the centre and north, with the highest values being in the south southwest region. The small yellow area under the big central one is in the water, and not a land feature.

Bilge 2003 Chlorophyll-a Chl-a (µg/L)

Bilge 2003's PDM is different from Keith 2018. The lowest values are in both the southern and the northeastern parts. Highest values are still placed close to the peak at Keith, in the south southwest, The minimal and maximal values are between 44.3 and 44.8 µg/L.

Mancino 2009 Chlorophyll-a Chl-a (mg/m^3) [UPDATE ?]

[move to discussion Mancino 2009 comes with (mg/m^3) which can be converted to µg/L since] The PDM from Mancino 2009 have values which are all negative in the water, from -80 to -35 (mg/m^3). Highest, or least negative, values in the south and south southwest, and peak negativity in the north west.

Giardino 2001 Chlorophyll-a Chl-a (mg/m^3)

Similar to Bilge 2003, Giardino 2001's PDW for chlorophyll-a has a narrow width of observations. The values are also all negative and between -3.58 and -3.39 mg/m^3. The lowest values are from south southwest to the west, and the highest are hugging the coast in the centre and moving north.

5.3 Original Water Quality Model on Koka

5.3.1 Pre-processing results

Imported images in QGIS

Koka Reservoir is so large that it required three Basemap tiles. The first pre-processing results were the three raster tiles. Figurex shows them with the sample sites as well, and the whole transformed coordinate list is in appendix X. The visualisation is in RGB though some colours are odd in addition to the visible seam line between the tiles, as the tiles. This is due to the automatically balanced tiles based on their individual "colour" distribution. Only a visual issue.

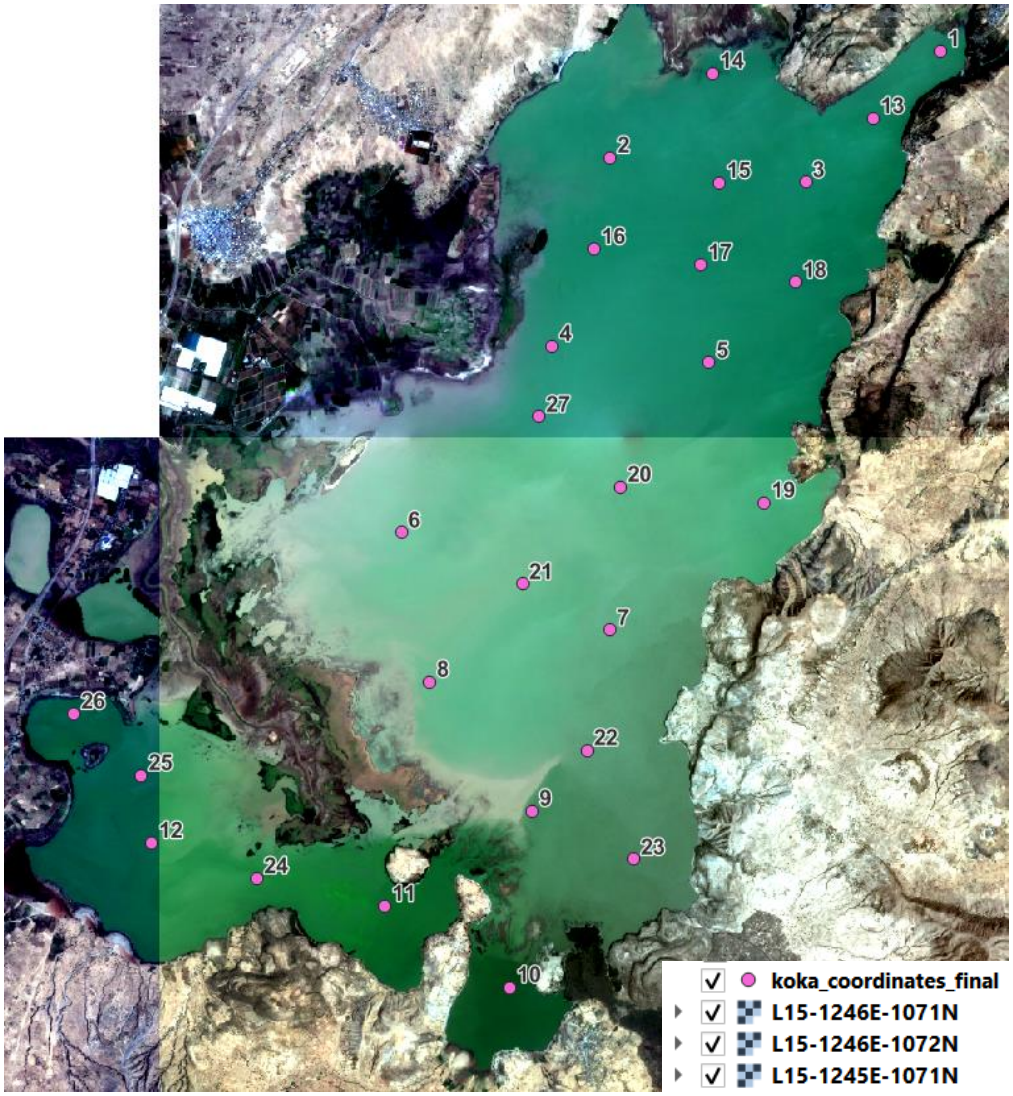


Figure 21 Imported Basemap of Koka Reservoir February 2022 with the 27 sample sites.

Importing sample sites

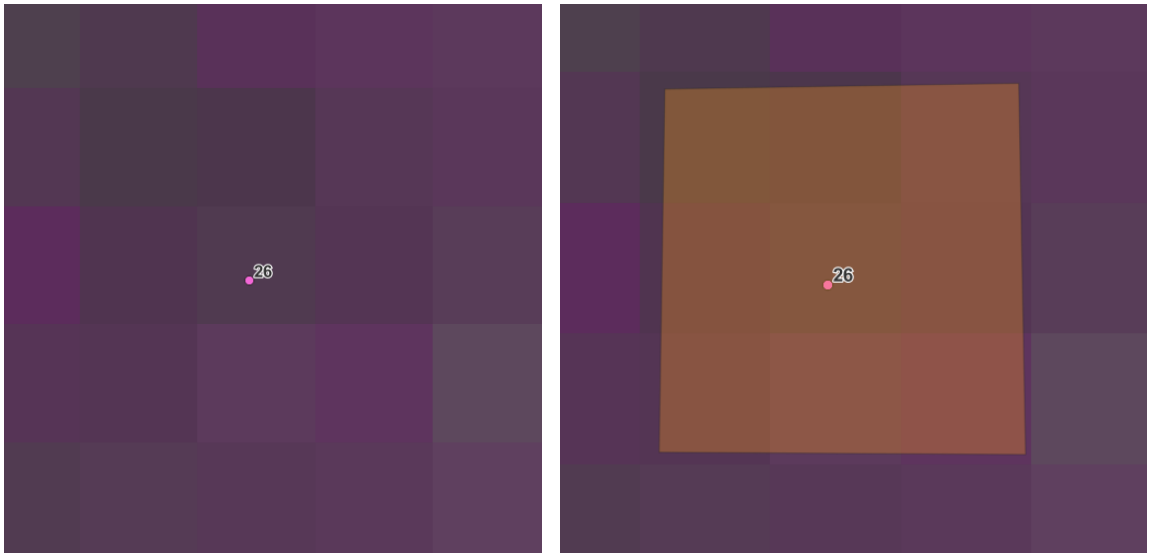
The 27 sample sites spread on the lake visualised in QGIS with the basemap tiles in the background. The numbers are the site IDs.

(Merging the layers together would lead to interpolation and change of the pixel values. Though these changes are miniscule, it would add an unnecessary uncertainty to the data, as this would affect the data values. move).

The sample sites are spread throughout the lake.

Buffering from point to polygon

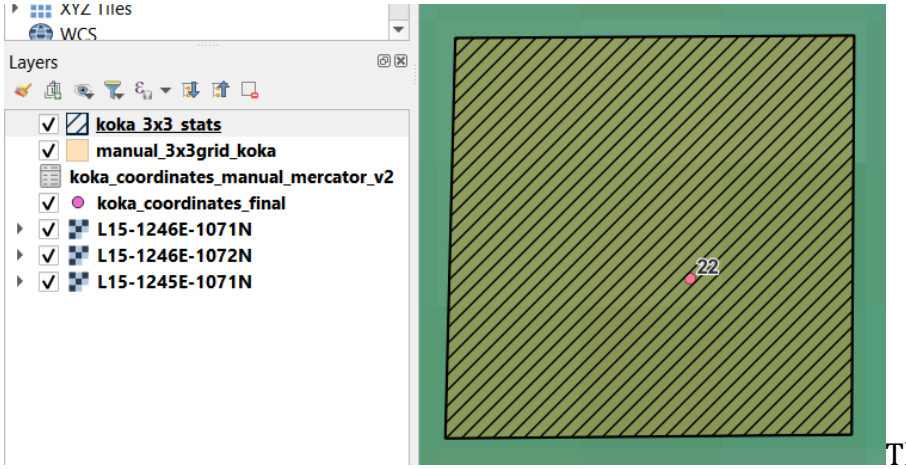
Visualisation of point sample 26



before and after adding polygon. The transparent layer is the polygon covering the point sample in the centre of the three-by-three pixels of the raster. The colours of the raster was changed to visualise the different pixel easier.

Extract pixel values around the points

The resulting polygon shapefile after running the "QGIS Model". The polygons in QGIS were the same but their fields are different as they now contain all the Zone statistics.



The shaded area is the new polygon. It is the same as the buffered polygon.

The average three-by-three DN for each sample site was then extracted to Excel for further processing and analysing. As the DNs are only scaled spectral reflectance, the reflectance was calculated and is presented in the figure below to compare with Assegide (2023) et al.'s results.

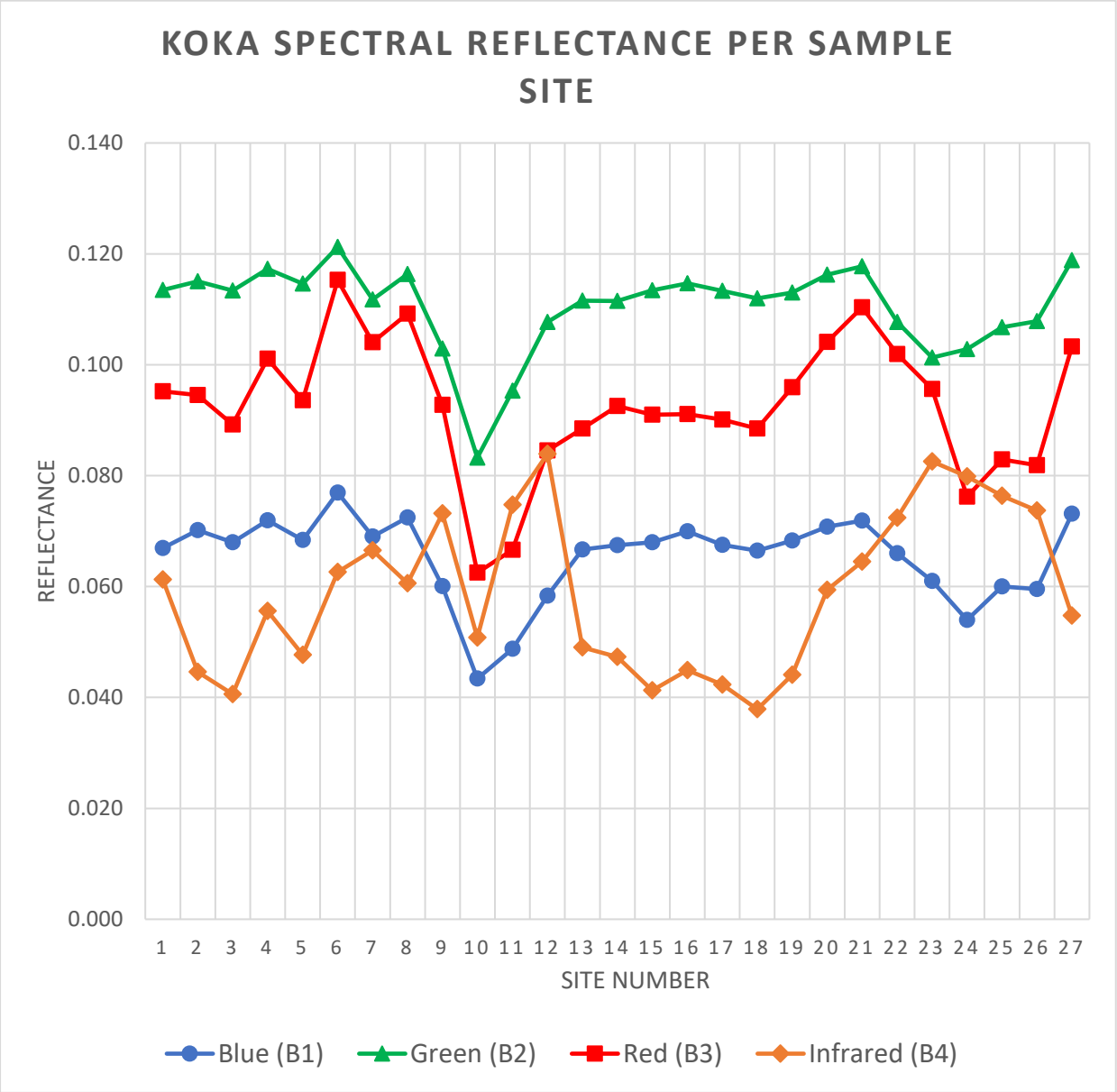


Figure 22 Koka spectral reflectance per sample site February 2022

The highest reflectance band for NICFI was Green (B1) across all sample sites with Red right below. The lowest band was NIR with exceptions at sites 9 to 12 and 22 to 26. Lowest total reflectance was at site 10. All bands have a dip, with Green and Red losing the most. For the RGB bands there was a V-shape trend from sample site 8 down to 10, and back up again at site 13.

Average for each band Blue: 0.0654 | Green: 0.110 | Red: 0.0927 | Infrared: 0.0590

Temporal displacement between image and sample date

The three basemap tiles that cover lake Koka were captured on different days and satellites throughout February.

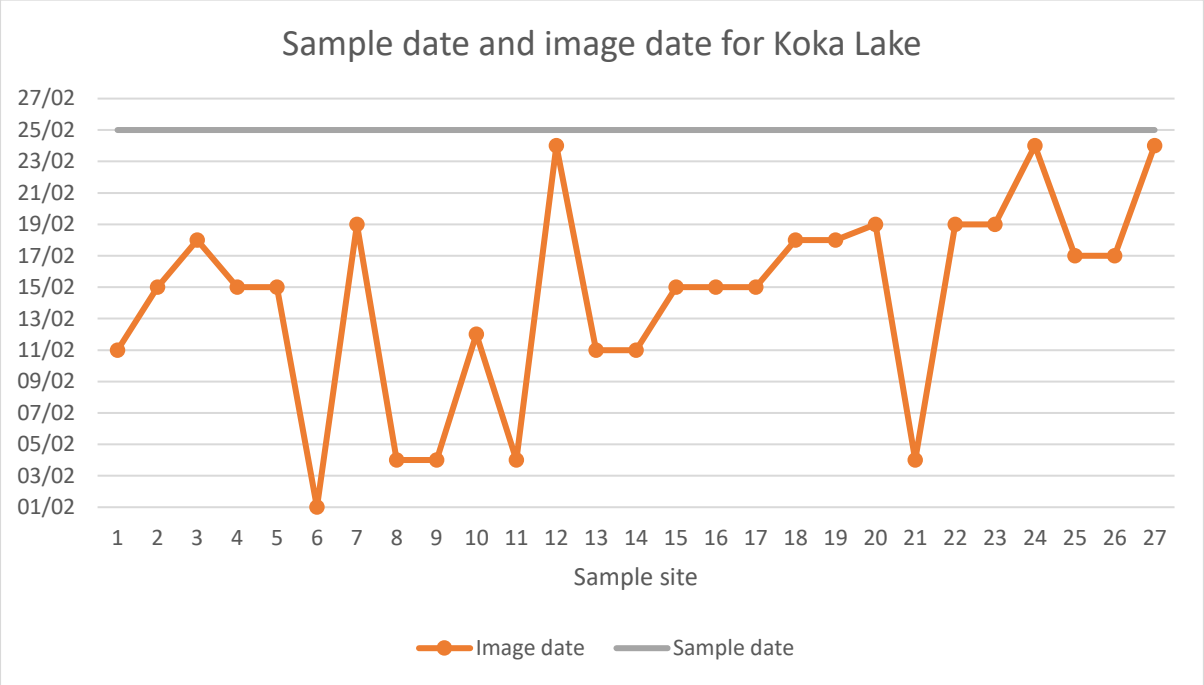


Figure 23 Temporal difference between sample date and satellite image date for Koka Lake Febrary 2022.

The average number of days off between the NICFI satellite images and the sample date on Koka by Assegide et al. was 10.6 days.

5.3.2 Regression Models from Excel

The results from empirical method.

Turbidity

Correlation

Table 8 Turbidity Correlation

	Blue	***Green***	***Red***	***Infrared***	***Turbidity***
Blue	1				
Green	0.96	1			
Red	0.90	0.80	1		
Infrared	-0.27	-0.24	0.053	1	

Turbidity	0.0025	0.051	0.025	0.21	1

SUMMARY OUTPUT Turbidity

Regression Statistics	
Multiple R	0.276389981
R Square	0.076391422
Adjusted R Square	-0.154510723
Standard Error	27.98174721
Observations	21

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	4	1036.158693	259.0396732	0.330839	0.853168118
Residual	16	12527.65083	782.9781769		
Total	20	13563.80952			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	-51.47696116	145.7139845	-0.353273993	0.728495	-360.376809	257.422887	-360.376809	257.4228867
Blue	-0.088663464	0.911950783	-0.097223958	0.923756	-2.021912762	1.84458583	-2.02191276	1.844585833
Green	0.147050582	0.46216073	0.318180607	0.754463	-0.832686399	1.12678756	-0.8326864	1.126787562
Red	-0.03029626	0.286083166	-0.105900186	0.916978	-0.63676548	0.57617296	-0.63676548	0.576172959
Infrared	0.049840415	0.09292383	0.536357734	0.59909	-0.147149304	0.24683013	-0.1471493	0.246830134

R^2 = 0.0764

Very low. Not applicable at all

Band ratio testing:

Band combination	**Types of combination**	**R^2**	**Source**
B3/B2	1 ratio	0.00015	Koka paper
B3/B2 + B1/B3 + B1/B2	3 ratios	0.0677	Self
B4/B3	1 ratio	0.0381	Self

It made things worse..

remove

Total Suspended Solid

Table 9 Total Suspended Solid Correlation

	Blue	***Gree n***	***Red***	***Infrared***	***Total Suspended Solids***
Blue	1				

Green	0.96	1			
Red	0.90	0.81	1		
Infrared	-0.42	-0.36	-0.091	1	
Total Suspended Solids	-0.26	-0.27	-0.074	0.51	1

SUMMARY OUTPUT Total Suspended Sediment

Regression Statistics	
Multiple R	0.527304282
R Square	0.278049806
Adjusted R Square	0.146786134
Standard Error	165.2388993
Observations	27

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	4	231346.3358	57836.58396	2.118254	0.112706843
Residual	22	600685.6642	27303.89383		
Total	26	832032			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	565.9638661	773.9564134	0.731260645	0.472338	-1039.1235	2171.051228	-1039.123495	2171.051228
Blue	1.434173068	4.264431545	0.336310491	0.739823	-7.40971666	10.2780628	-7.409716664	10.2780628
Green	-1.298732218	2.242490352	-0.579147294	0.568373	-5.94937256	3.351908127	-5.949372564	3.351908127
Red	-0.150508555	1.334111712	-0.112815556	0.9112	-2.9172869	2.616269795	-2.917286904	2.616269795
Infrared	0.674745919	0.498702644	1.35300249	0.189796	-0.35950006	1.708991902	-0.359500063	1.708991902

R^2 = 0.278

Results from removing the highest observations (sample 9 and 22)

Regression Statistics	
Multiple R	0.64
R^2	0.41
Adjusted R Square	0.29
Standard Error	87.11
Observations	25

Increased the R^2 to 0.41.

cannot use this if I cannot justify the omitted observations.

Excel trendline idk

Chlorophyll-a

Correlation:

Table 10 Chlorophyll-a Correlation

	Blue	***Green***	***Red***	***Infrared***	***Chlorophyl-a***
Blue	1				
Green	0.96	1			
Red	0.90	0.81	1		
Infrared	-0.41	-0.36	-0.082	1	
Chlorophyl-a	-0.30	-0.16	-0.31	0.34	1

Chlorophyll-a regression result without sample 1 and 26, as they have been omitted in the source material as well. **R^2= 0.72**

Testing without nr. 1 and 26

SUMMARY OUTPUT

Regression Statistics	
Multiple F	0.849452
R Square	0.721568
Adjusted I	0.662951
Standard E	13.17007
Observati	24

ANOVA

	df	*SS*	*MS*	*F*	*gnificance F*
Regressio	4	8540.594	2135.149	12.30982	4.17E-05
Residual	19	3295.566	173.4508		
Total	23	11836.16			

	Coefficients	*andard Err*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*ower 95.0%*	*pper 95.0%*
Intercept	-123.49	64.14805	-1.92508	0.069322	-257.754	10.77299	-257.754	10.77299
X Variable	-0.08896	0.34281	-0.25949	0.798048	-0.80647	0.628554	-0.80647	0.628554
X Variable	0.293198	0.182103	1.610069	0.123872	-0.08795	0.674344	-0.08795	0.674344
X Variable	-0.19439	0.107656	-1.8057	0.086836	-0.41972	0.030932	-0.41972	0.030932
X Variable	0.112203	0.040426	2.775523	0.012049	0.027591	0.196816	0.027591	0.196816

Other attempts (from June meeting document progressM_june) are added in the appendix .

Total correlation

Total correlation with all inputs with sample sites containing all parameters tested. N=20 as some parameters are not tested on all locations.

Table 11 Correlation between all parameters for Koka samples and bands. Greener means stronger correlated.

	Blue	***Green***	***Red***	***Infrared***	***Chl-a***	***Turbidity***	***TSS***
Blue	1.00						
Green	0.96	1.00					
Red	0.90	0.79	1.00				
Infrared	-0.26	-0.23	0.07	1.00			
Chl-a	-0.24	-0.09	-0.30	0.29	1.00		

Turbidity	-0.03	0.02	0.00	0.23	0.82	1.00	
TTS	-0.17	-0.18	0.00	0.46	0.02	-0.01	1.00

Strongest correlation is between chlorophyll-a and turbidity. This adds up, since the increased green pigments would reduce visibility in water. The "greener" the more turbid. The second strongest correlation is between Total Suspended Solid and Infrared. This also adds up, since deep clear waters absorb infrared radiation, so when the water is murkier and filled with suspended solids, more infrared radiation is reflecting off the water.

5.3.3 Original Water Quality Model for chl-a

The developed WQM for chlorophyll-a was based on samples from Koka Reservoir and regression analysis on NICFI's Basemap images from February 2022.

Original Water Quality Model for chlorophyll-a:

$$-123.4904225 - 0.088955872 \times (B1) + 0.293198207 \times (B2) - 0.194393362 \times (B3) + 0.112203264 \times (B3)$$

5.4 Novel Parameter Distribution Maps

Chlorophyll-a

By using the field samples from Assegide et al. 2023 and regression statisitcs on the NICFI satellite images the Original Water Quality Model (WQM) for chlorophyll-a (chl-a) was computed. The Novel Parameter Distribution Maps (PDM) were created using this WQM and they are presented below in Figure 24. These maps show the distribution of chl-a at Gefersa Reservoir, Legedadi Reservoir, and Aba Samuel Reservoir for March 2023 and for Koka Reservoir Febrary 2022.

The distribution between the bodies of water differ. Koka has a peak consentration of 200 µg/L chl-a, and goes as low as -20 chl-a. Highest consentration in the south southwest and west of the map. Low values at the centre moving north.

At Aba Samuel the range is between -30 to 40 µg/L in the water. Higest values by the south westmoste shoreline.

Legedadi has some negative values, but in the water that is not covered clouds, the range is 7 to 100 µg/L. The dark blue spots are cloud shaddows in the centre, and have values as low as -30 g/L.

The distribution at Gefersa is spatially more even, however the western shoreline has the higest constration, while the remaining shoreline is generally higher. There is an exception at the easternmost shoreline where the consentration droppes to the lowest values. The range at Gefersa in the water is -12 to 35 µg/L.

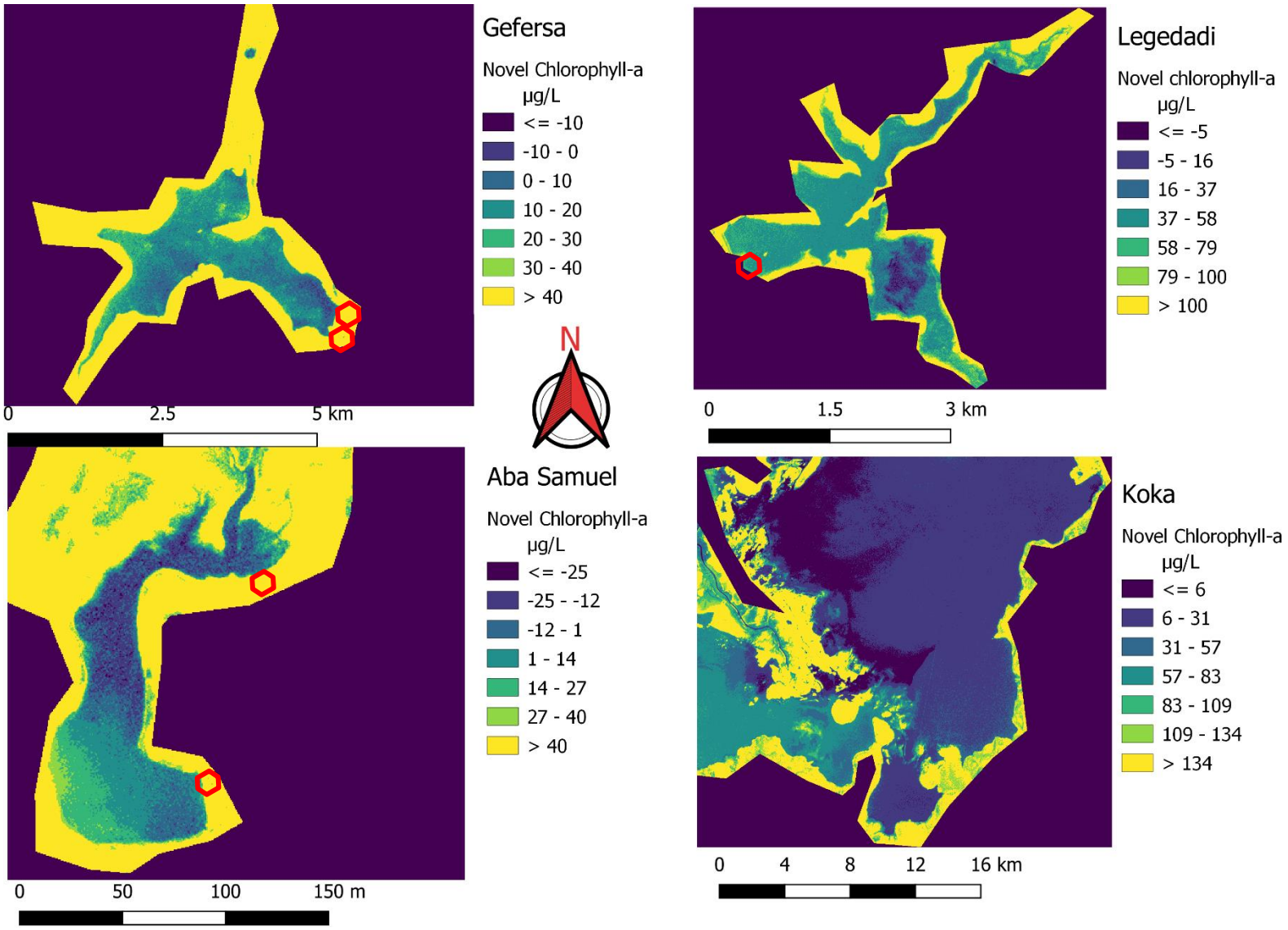


Figure 24 Novel PDM for Chlorophyll-a on Gefersa, Legedadi, Aba Samuel and Koka Reservoirs. Red hexagons are visited sample sites.

6. DISCUSSION

6.1 Procedural knowledge

Throughout the researching and experimenting for this thesis, there was procedural learning, so by reading, attempting and failing, more information was gained. This meant that important discoveries came later, and it had an impact on the structure and results of the methodology. The procedural knowledge has been taken into account, but due to time restriction some factors might be unadjusted for.

6.1.1 Field Challenges

Planning an expedition to Addis Ababa was complicated and resulted with some changes to the plans. The assistance received from Kotebe University of Education (KUE), the NORPART partners, supervisors, and the experts in Addis Ababa was excellent, but there were shortcomings. The field trip was delayed by a month, causing a major delay for the thesis progress. When finally in Addis Ababa, insufficient water quality testing equipment was brought there. It was mistakenly assumed that KUE would and could supply equipment for the field sampling, but this was not possible. According to the Chief of Water Chemical Engineering at KUE[20], they had either not the relevant equipment or it was currently not working. Issues such as lack of reagents for the chemical tests or broken equipment with no means to repair them. Costs were a big issue in both this thesis and water management systems in developing countries as a whole. KUE was surprised that this research would be dependent on the scientific resources and equipment of a developing nation. Unfortunately, there were problems with borrowing scientific equipment and reagents from NMBU's labs, as they would be used in lectures and courses locally instead.

Fortunately, KUE organised the relevant documentation and administration so that the research could be assisted by both Addis Ababa Water and Sewage Authority (AAWSA) and the Ministry of Water and Energy (MoWE). AAWSA administrates the drinking

[20] Tesfaye _____

water and sewages of Addis Ababa. They[21] granted access and were guides during the field trips to Legedadi, Dire, and Gefersa Reservoir. MoWE is responsible for preparing national water policy, strategy, and action plans. They have jurisdiction on Aba Samuel Dam, so they were the guides to Aba Samuel Reservoir. In addition, the guide from MoWE took samples and provided in situ and laboratory results for this thesis.[22]

Zeleke(https://urbanagetaskforce.net/addisababa/ scroll^)

Make a list of challenges:

6.1.2 Lack of boat

Unfortunately, there was not enough planning done to be able to conduct a large enough amount of sample taking. Research that involves the spatial distribution of a parameter requires the samples to be spatially distributed. However, neither AAWSA[23] nor MoWE[24] had any boats available to use in these lakes, either due to being far away or broken. These boats would have been used to get samples from different locations instead of only by the walkable portions of the shore. Furthermore, samples taken by the shore would be negatively affected by edge effect, as shore pixels would be mixed with the water. In addition, there is a difference between the shallows and the deeper waters. Therefore, the number, distribution, and location of the samples might have a negative effect on the results of the sample based experiments .

6.1.3 Misunderstanding of reflectance

Until early July, the research had been wrongfully based on the theory that the NICFI Basemap mosaic was a radiance product. The digital numbers (DNs) were between 0-10000, giving some certainty that the DNs had to be something else than reflectance. Further misinterpretation of Planet's and NICFI's product documentation and specifications solidified this mistake. This led to a lot of time and energy being wasted

[21] Head of Department at AAWSA Zeleke Tefari was guid and expert to Legedadi and Dire. Environmentalist from AAWSA, Solomon Tadesse, was the guide to Gefersa.
[22] Water quality expert and Water Quality Directorate at MoWE Yirgalem Esuneh Endalew guided trip to Aba Samuel, gave insight and prepared the in situ and laboratory tests and results for this thesis.
[23] Confirmed during meeting with Head of Department Zeleke Teferi in AAWSA and saw the broken boats myself at both Gefersa and Dire.
[24] Claimed their nearest boat was at Bahir Dar (500km away) by the Ambassador Asfaw Dingamo (State Minister of Water Supply and Sanitation)

on model testing with the wrong data fundament. Most of the models produced had unattainable results, but some had accurate results, compared to available data. A lot of research had to be done to find the correct scaling factor, and KSAT[25] supplied the missing information. Therefore, this mistake was identified and the methodology was repeated. Without this procedural knowledge, these mistakes would have been persisting in the whole thesis, and be a large systematic mistake.

6.2 Sources of Error

As there are several layers and steps required to get this data and results, each step could contain several sources of error and reduced the statistical accuracy.

6.2.1 Positional accuracy

The NICFI Basemap has a specific positional accuracy, which determines the correctness of coordinates between the image pixel and real-world location. The product specification stats 10m Root Mean Square Error (RMSE), and this is lower than the < 5m per pixel resolution (4.77 m). This was solved by using a bigger pixel window of 3x3 to extract the image information. The total positional accuracy would then be "swallowed up" by extracting from a bigger area. This "window" was also to account for GNSS inaccuracies when measuring in the field by Assegide et al.. Their handheld GNSS unit had 3-5m accuracy, while the mobile phone GNSS unit used at Gefersa, Legedadi and Aba Samuel had 4m accuracy.

In addition, at Aba Samuel site A, the location information is inaccurate since it was not possible to cross the mud all the way to the shore. Therefore, using the "window" would give comfort in parts of the inaccuracy.

[25] E-mail correspondence with Senior Project Manager, Charlotte Bishop, at Kongsberg Satellite Services

6.2.2 MQuant: Temperature and expiration

Other practical issues were noticed during the field testing in Aba Samuel. Firstly, the nitrate strips expired on the 31st of March 2021. This could influence the results as they were almost two years overdue. Secondly, the nitrate strips are supposed to be stored in a dry and cold space between +2°C and +8°C, in accordance to their instructions. However, since the retrieval of the nitrate strips on Thursday the 2nd of March 2023, they were not refrigerated at all. This meant they were out of refrigeration for the entire travel, from Norway to Ethiopia, until the second day at the field work at Aba Samuel on the 14th of March. These issues were noticed when discolouration on all the remaining unused test strips was observed at Aba Samuel. They appeared as they had already started reacting at parts of the reaction zone, as shown in the figure below. The dirt on the top corner was from the ground as it fell during handling, however, the discolouration was already there and unaffected by the drop. At this point it was decided not to use them further after this site. This meant that there were no nitrate strip tests conducted at Gefersa. However, the results are not removed from Aba Samuel nor Dire.

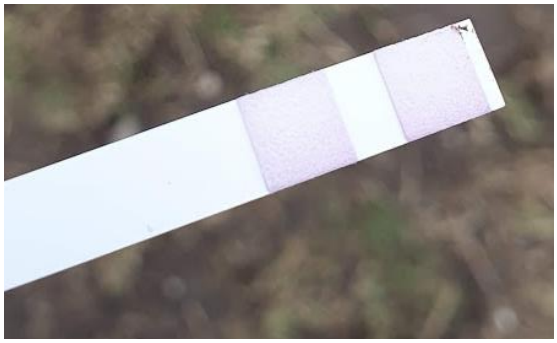


Figure 25 Nitrate strip that has reacted due to improper storing. Dirt in top corner can be ignored. All remaining samples look like this.

Personal correspondence with Doctor Melesse Eshetu Moges explains that expired nitrate strips can at strong concentrations show results within the range when compared to a Hack Lange test. However, these results might not qualify as “strong concentrations” as they are on the lower spectrum of the indicator’s scale. Due to this, we can assume that the first tests were to some extend less accurate measurements of the nitrate levels at that test site. On the other side, there were no discolouration issues on the unused test strips during the testing at Dire Dam on the 11th of March. This could mean that the MQuant test strip results could be used as indications rather than a

definite result. Meaning the nitrate strips are still comparable to the literature values and the laboratory measurements.

6.2.3 Sample size

This field study had a total of 5 sample sites. When counting sites with at least a two tests run on them, the number of samples reduces to four and three, as the same tests are not done for all sites. Five is too small of a sample size to do any proper statistical analysis with significant results. Therefore, this study will use these values as indicators, or approximate values, that gives a realistic expectation for the Parameter Distribution Maps (PDM) produced. This includes both the existing literature Water Quality Models experimented on, as well the Original WQM produced on Assegide et al's 2023 paper on Koka Reservoir. This could help validate or invalidate resulting PDMs.

6.3 Analyse field and laboratory results

The results from the laboratory tests seem plausible. For most of the parameters, Aba Samuel had consistently higher values compared to Gefersa, meaning it was more polluted. The exception was for NO3 and N-NO3 at Gefersa site B (GB), which were slightly higher.

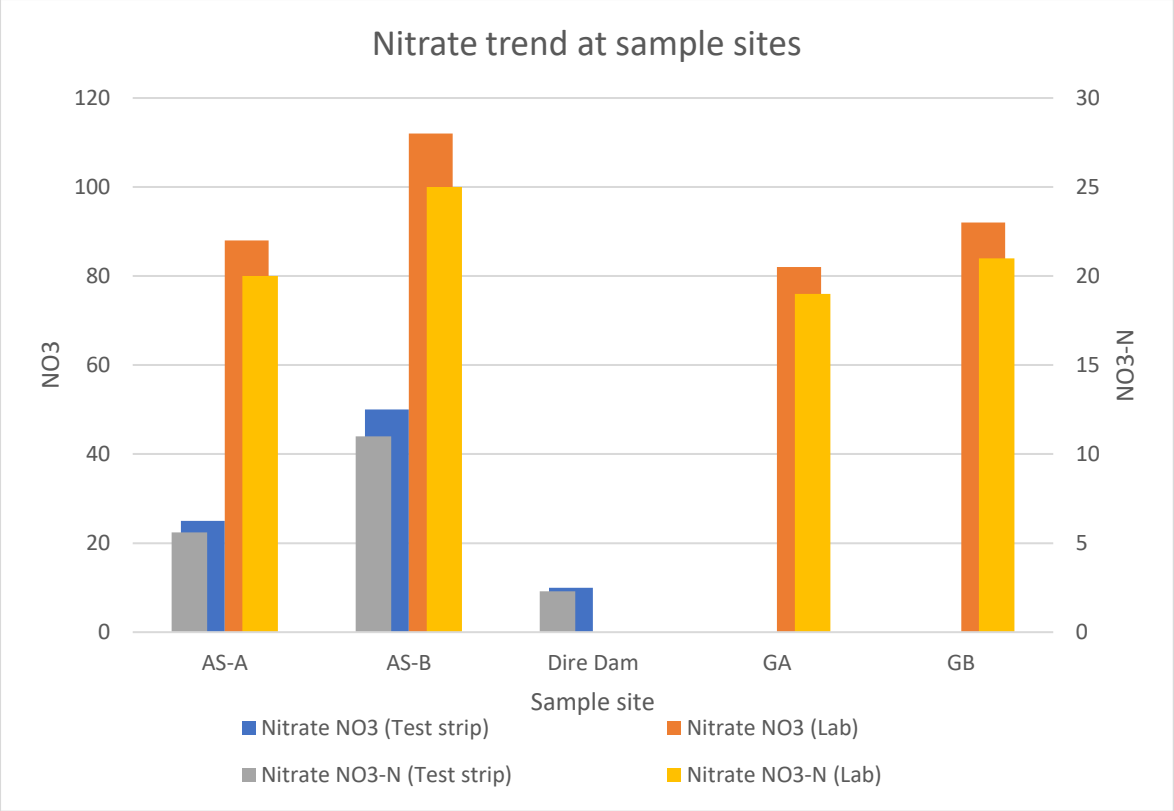


Figure 26 Nitrate results from field work at Aba Samuel, Dire, and Gefersa Reservoir

The nitrate results were expected to be higher at Aba Samuel, as it is dirtier(explain/point at literature). Even though there are differences in the accuracy [ADD accuracy description?maybe] between the laboratory and MQuant nitrate test strip, there was an observable trend in the results. There was a higher concentration of nitrate on Aba Samuel site B (AS-B) than site A (AS-A). CHECK EXPLAIN NITRATE IN THEORY . Site GB had also higher concentrated than GA. WHY MORE IN GA

The laboratory test and the test strip agreed on this finding, however, the laboratory test for NO3 showed only a 27% increase while the test strips showed a doubling, evident in **Error! Reference source not found.**, and this accounts for NO3-N as well.

The discrepancy between test strip and laboratory could be explained by the coarseness of the displayable results for the test strips. For NO3 the measurable results jump exponentially and can only be 0, 10, 25, 50, 100, 250 or 500mg/L.

Furthermore, the nitrate test strip results were higher at AS-B than AS-A, which was consistent with the laboratory results, even though the numerical values were different.

Thus, the test strips do appear to work to some extent and could be used to show general trends in the water.

Based on this, and the test strip results at Dire Reservoir, one could reasonably accept the lower levels of NO3 and NO3-N presented in Table 6 and the bar chart above. Since Dire Reservoir is part of Addis Ababa's drinking water source, this is ideal. However, Gefersa is also a drinking water source, but has very similar (<10% different at AS-A and <25% between AS-B and GB) nitrate concentrations as Aba Samuel Reservoir, which was the exception mentioned earlier.

On the other hand, Gefersa has significantly lower concentrations of the other parameters tested in the laboratory. Table 7 shows differences in results between a hundredth and 40% between Aba Samuel and Gefersa.

6.4 Discussion of literature PDMs

The results from the Parameter Distribution Maps (PDMs) from the literature study are presented in Table 12 below. This table categories the PDMs based on the resulting parameter values and whether they are good. Here "goodness" is based on realistic and probable values. A PDM has "Failed" if the values are physically impossible, like negative concentration, or has improbable values, such as no variation. Other Water Quality Models (WQM) from the literature study are added here without including them in the result section, due to their level of "failedness", such as the remaining Bilge 2003 models for Suspended Sediments and nitrate.

Table 12 Verdict on the Parameter Distribution Maps from literature study. [any modification discription]

Verdict	Reason	Models
Failed	All negative or extreme values	Miller TSM [normal]<(-240) -(-13) >, Allan Chl-a [nat log] <34000-89000>, Mancino Chl-a <-80-1.6>, Giardino Chl-a <(-3.69) - (-3.32)>, Assegide TTS <(-300) - 0>
Failed	No variation	**Every** Bilge (NO3,SS and Chl-a) <+-5%>, Allan Chl-a [non log] <10.35-11.65>, Brezonik SDT [non log] <(-2.664) - (-2.665)>, Brezonik SDT [nat log] <0.069+-0.0003>
Good	Realistic seeming values	Miller TSM [forced positive] <13-240>, Keith Chl-a <25-150>, Assegide TU <-27-150>, Kutser CDOM <1-8 >

Bilge et al. presented three WQMs that could be used with the NICFI satellite data. The bands are representing the same colours, though with slight width differences. However, all the PDM results were objectionable with little variation in the water. In addition to the PDMs, this result can be verified when investigating the model formula. For all three parameters, the formula uses the reflectance multiplied with coefficients to calculate the concentration. These coefficients are small numbers with values between zero and four, meaning the product of this multiplication is always a small value, since the reflectance from the NICFI bands are never over 0.12, as seen in Figure 22. This made

the constants in the formula by Bilge et al. the base value of the results, no matter the spectral input from NICFI database. Both the PDM and table above shows this trend.

The “small weight” the bands carry in this WQM is not unique for Bilge 2003. Brezonic 2005 WQMs have the same small result variation, due to the small coefficients multiplied with the small reflectance value. The ratio of two bands is also a small value causing the little variation in the results. The constant is still the key factor. Even when considering the natural logarithm in the formula, [remove calculating every result by placing it over Euler’s number], the result is still with little variation. frasing. With Euler, the values are even narrower for the Secchi Dick Transparency.

Allan 2007 does also use ratio and natural logarithm, and the results are still in the failed category for this study.

Allan: even though the WQM uses ratio of two bands, and with and without natural logarithm, it was not possible to get reasonable values using NICFI basemap. TERM? Allan had higher blue values than red, and it is opposite for NICFI. The background information was too different.

For all the PDMs that only contain negative values, these can only be defined as failed models. Concentrations are only positive values, and even when considering the uncertainty of regression, where one could represent negative values as zero, there is no appropriate way to do so when every value is negative. However, the forced positive Miller 2004 WQM model gave an interesting result with positive values. The argument for this test was in case there was any typos by the original authors making every result negative. However, further reading did not find any typos, so this forced positivity could not be a statistically viable methodology. It could be entire random that this “better result” came from switching the +- sign. Meaning this result has to be thrown away.

The other WQMs in the “failed due to negativity” category, Mancino 2009, Giardino 2001, and Assegide 2023 TTS, all share the same issue. Every model has a large negative constant that the results are not capable of overcoming, or a large negative coefficient keeping every result negative. As mentioned earlier, these concentration results are impossible.

The remaining WQMs are Kutser 2005 CDOM, Keith 2018 chl-a, and Assegide 2023 TU. These had positive values with a spread that is within a realistic window for their parameter. However, there is still a need to confirm their accuracy with the literature.

The turbidity can be directly compared to Assegide et al. 2023. The satellite images the PDM is based on are from the same lake and time period. Assegide et al.'s turbidity lab results were similar to the PDM. The laboratory values were between 38-148 NTU, and average was 52, while the PDM had 25-150 NTU in the water.

[MOVED WEAVE BETTER Turbidity: Very much in negative area. South is very dark/sub zero. When this is thought off as zero, it is not agreeing with RGB image of dark green water. The turbidity should be very high in that area southern area of the lake. Shows that TU is hard to measure in green waters. Water values from -27 to 120 (ish). The numerical values are very similar to observations made by Endaweke. The laboratory values are between 38-148

6.5 Discussion of Original WQM

The results from the Original Water Quality Model developed with multi variate regression was inconsistent between the different parameters. From an R^2 of 0.0764 for turbidity to the R^2 of 0.72 for chlorophyll. Due to the small sample size (between 21 and 27) the results could be inflated. However, for this study, these results are meaningful.

Turbidity's R^2 of 0.0764 means that the NICFI dataset is not a suitable tool for estimating turbidity in waterbodies of this nature (trophic state/ clarity). However, the literature has shown it is possible using the same bands. Assegide 's regression model achieved R^2 between 0.85 to 0.9156 using different band ratio that are all available on the NICI dataset. Due to the low results of the 4 bands, band ratios were also tested, as seen in tablex. The models developed had had R^2 that are even lower than the 4 band regression model. This could mean there is a systematic error from the statistical approach, or that radiance/reflectance issue is here.

add more here

Observing the correlation between turbidity and the DN this low performance could be predicted, as the highest correlation is only with NIR at 0.21, while the remaining are below 0.05.

The lower number of observations for this parameter is not explained in the source material, and this could have an impact on accuracy

- Discuss the differences between radians and reflectance

- Discuss the differences between instruments, band width and such

The reflectance recorded by the NICFI is similar to Assegide et al.'s. This is evident when comparing figurex and Assegide et al.'s results in the appendix.

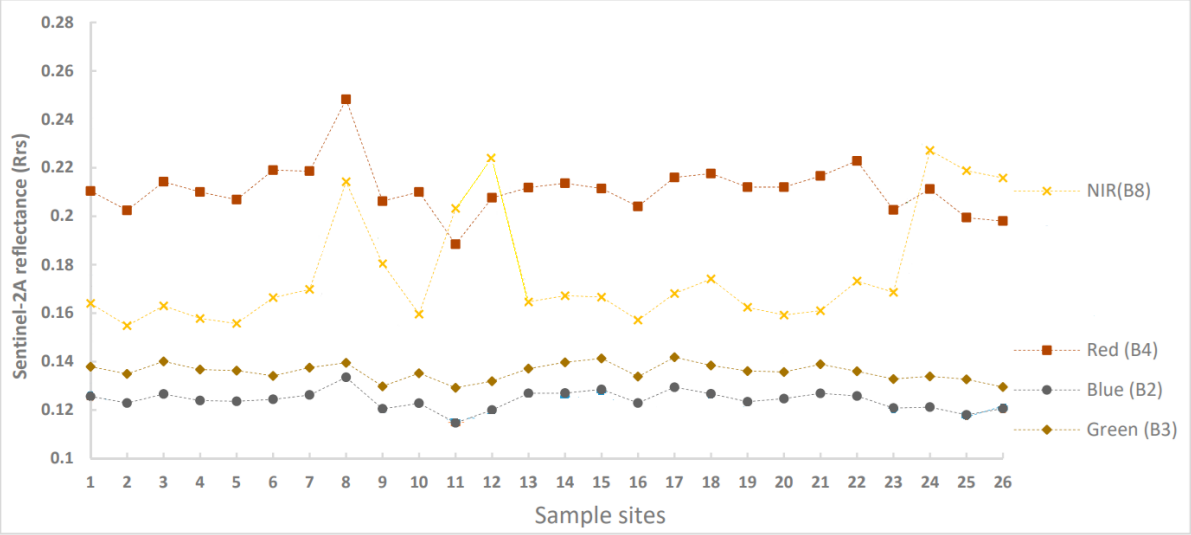
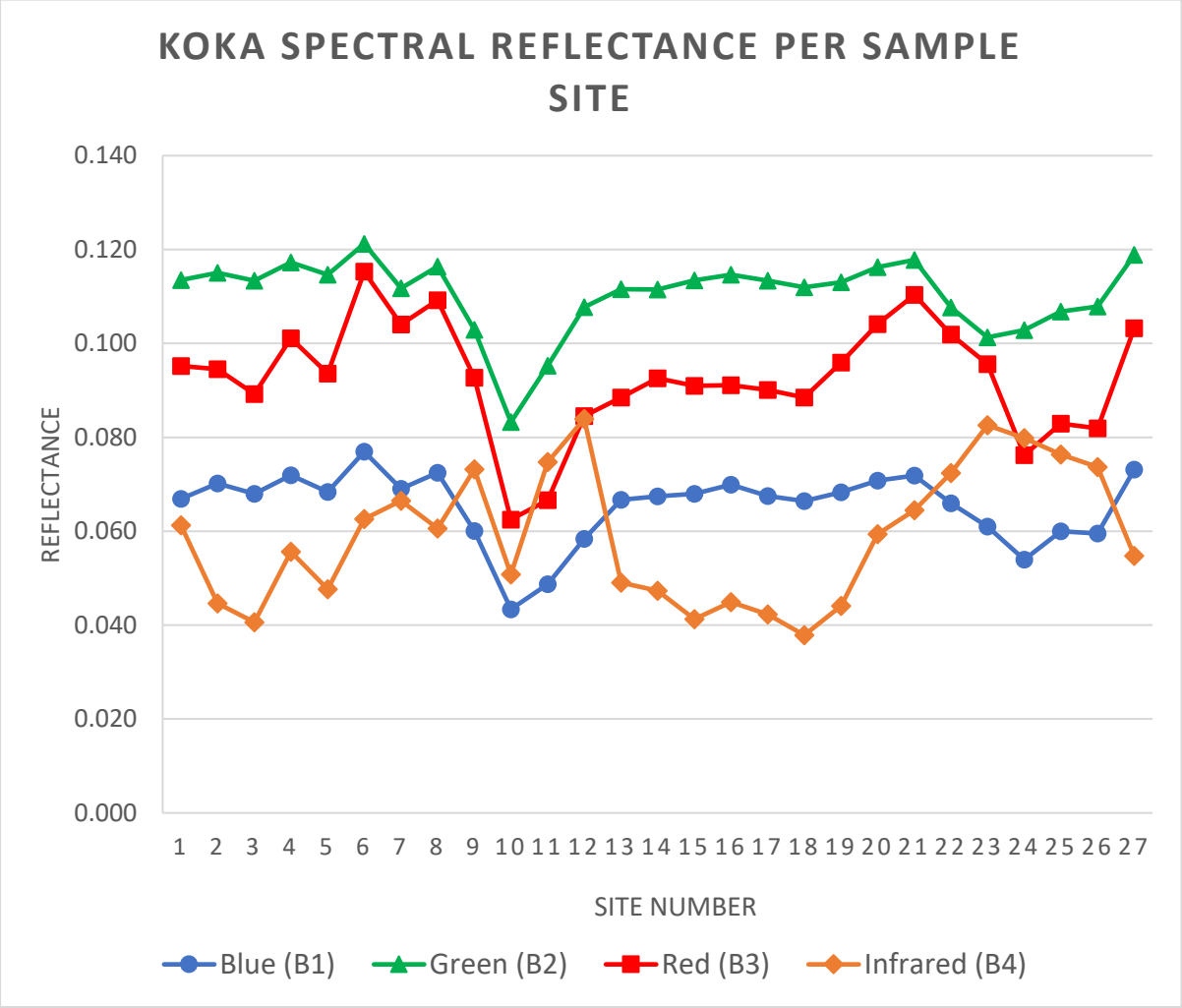


Figure 27 Reflectance on Koka sample sites with Sentinel-2 by Assegide et al. 2023. Modified graph where only relevant bands are included.



Firstly, the reflectance values are higher for Assegide et al. than NICFI. NICFI's results were as low as 0.04 and max out at 0.12 while Assegide et al. had reflectance from 0.11

up to 0.25. This accounts only for the relevant bands, and shows a huge discrepancy in reflectance. However, the reflectance trends between the sample sites are more important.

[MOVED to results(descriptive) NICFI: The highest reflectance band for NICFI was Green (B1) across all sample sites with Red right below. The weakest band was NIR with exceptions at site 9-12 and 22-26 . Total lowest reflectance was site 10. All bands have a dip, with Green and Red losing the most. For the RGB bands there was a V-shape trend from sample site 8 down to 10, and back up again at site 12 /13.]

But with

start explaining difference:

The different values were expected due to the different equipment.

- Regression issues and statistics
    - Few samples (n=27)
    - More?
    - 3x3 benefit

- Difference between satellite image date and sample date.
    - Koka
    - Other lakes

Parts of the inaccuracy in the models could be attributed to the difference between the in situ sample date in Koka by Assegide et al. and the NICFI Basemap images. 10.6 days is the average difference between the variables. This was illustrated in figurex (Sample date and image date Koka lake) in the materials section. As mentioned earlier, waterbodies are complex structures that change rapidly. sience point at theory . The discrepancy could reduce the accuracy dramatically, however, the weather and climate at Koka could reduce its impact.

Main changes to water bodies are (check theory) precipitation and river flow . Since these samples were taken in the end of February, the amount of rainfall would be low (Assegide et al.). Metrological data presented in figure (rainfall in Addis Ababa) agrees as well. However, the weather stations were not directly close to the sample site, but they are upstream to the Akaki/Awash River in the same Awash basin. Seeing less rain could mean that there has not been a big amount of flow down to the Koka Reservoir. This would support the argument that the reservoir was not drastically different between these days. Nevertheless, the number of days different should be kept in mind.

Figurex

from discussion, but moved here

Back from methods to discussion [figure out best text]^

The average number of days off between the NICFI satellite images and the sample date on Koka by Assegide et al. was 10.6 days. As shown in figurex (Sample date and image date Koka lake) in the materials section, there were temporal gaps between 1 and 24 days, which is could be a big source of error. Assegide et al. had the benefit of Sentinel-2's high temporal resolution, and opportunity to download single images, to have images up to one day off of the field sampling. Meanwhile, the NICFI basemap gives only access to finished mosaics of several satellite images with different time and dates to gain the highest quality image without clouds or shadows. This means there will be difficulties to get images on specific dates, as seen in this method.

Discuss own model used in QGIS results

Chlorophyll-a: The values are better. Negative values in dark blue area in the water. Not the best, but the 72% accuracy seems fine. Compared to the other methods, its fine.

Discuss more

7. CONCLUSION & RECCOMENDATIONS

Water quality assessment is tedious work. From manual trekking and testing lakes, to documenting water born illnesses in a population, all to determine the quality of a water source. The latter is an example of damage assessment instead of management, and this is something the United Nations 6th Sustainable Development Goal is about . By monitoring, treating and stopping pollution, health, environment and society is better [Refrase]. This thesis attempted to create a methodology to conduct cheaper and broader water quality monitoring. Differently from other literature, the end goal was to find a method that low-end users could easily use without fancy euqpment and experticce. By utilising free software (QGIS) and high resolution satellite images (NICFI) it could have been a vialbe method anyone with a computer and internet to monitor surface water quality for the Earth's Tropic regions .

Based on the literature study, the NICFI Basemap can assess several water quality parameters. By comparing sensor types and spectral characteristics there were few limitations on what could be measured. If there was more research time, the number would probably increase. The seven water quality parameters presented in Table 2 are all calculatable, however, the accuracy, and reliability are more important factors.

Each Water Quality Model (WQM) in the aforementioned table had their original R^2 presented. Their coefficient of determination, i.e. the models goodness of fit, indicated that they are accurate models. However, when extracting the WQM to be used in this study, the R^2 stops being valid, as the model is no longer comparing their proven correlated field data with their satellite images. Instead, new images from different lakes and sensors were forced into the WQM.

The production of Parameter Distribution Maps (PDM) from the literature WQM found only three WQMs that could produce realistic values for the lakes around Addis Ababa. This included attempts at manipulating the WQMs in unscientific manners, such as Miller 2004 WQM for Total Suspended Matter being forced positive.

Table 13 Water Quality Models with realistic values for lakes in Addis Ababa

WQM	Parameter (unit) <spread>	Formula

Miller 2004 (positive)	Total Suspended Matter TSM (mg/L) <13-240>	1.91×1140.25 × (B3)
Keith 2018	Chlorophyll-a (µg/L) <25-150	60.703 × ((B1)$^{-1}$ – (B2)$^{-1}$)×(B4) + 10.386
Assegide 2023	Turbidity TU (NTU) <-27-120>	282.88 × (B3/B2) − 206.15
Kutser 2005	Coloured Dissolved Organic Matter CDOM (m^{-1}) <1-8 >	5.13 × (B2/B3)^-2.67

Turbidity is great since the laboratory values were between 38-148 NTU, while the PDM had 25-150 NTU in the water.

By comparing literature values and field samples with the PDMs, it was possible to give a rough estimate of the accuracy and reliability of these WQMs on NICFI images. However, the field study in Addis Ababa did not yield enough field samples to compute the accuracy with statistical significance.

In addition to extract existing WQMs, an original WQM was produced using field samples from Assegide et al. 2023 on Koka Reservoir. The 27 sample sites were the basis for the three WQMs produced.

565.9638661 + 1.434173068 × (B1) - 1.298732218 × (B2) - 0.150508555 × (B3) + 0.674745919 × (B4)

WQM	**Goodness of fit**	**Formula**
Turbidity	R^2 = 0.0764	-51.47696116 - 0.088663464×(B1) + 0.147050582 ×(B2) - 0.03029626 ×(B3) + 0.049840415×(B4)
Total Suspended Solid	R^2 = 0.278	565.9638661 + 1.434173068 × (B1) - 1.298732218 × (B2) - 0.150508555 × (B3) + 0.674745919 × (B4)

Chlorophyll-a	$R^2 = 0.72$	-123.4904225 - 0.088955872×(B1) + 0.293198207×(B2) - 0.194393362×(B3) + 0.112203264×(B3)

Turbidity $R^2 = 0.0764$

Total Suspended Solid

$R^2 = 0.278$

Chlorophyll-a

Start working on a conclution

The results of these experiments differ a lot. The tested methodologies extrapolated from other studies from around the world show unrealistic results, and in some cases impossible values. Some of the issues could be attributed to the differences between our equipment and lakes tested compared to those from the studies.

- Many differences between literature experiments and mine
    - Equipment
        - Band width
        - Other, quality, age,
        - Quality of fieldwork
    - Their Field
        - Water type
        - Climate
    - Their processing
        - Atmospheric correction!
        - Other processes
- What worked
    - Formulas that were good
        - Why did work?

- What did not work
    - Formulas that did not
        - Why not

Recommend having bigger field samples. A future study that could build on these findings should be a comparative study. By using historic water quality samples and the corresponding satellite images in archives, it should be possible to increase the sample size. Furthermore, seasonal and regional WQM could be built if there are enough sample. By using other satellite images than NICFI, it would be possible to attain temporally better images, as the Basemap is bound by the nature of mosaic products and the necessary processing required to have a high quality and spatial resolution available on a free platform .

The best-case scenario for a low-end user with access to NICFI is when there already exists a lot of location based field samples. This would allow for a decent estimation and distribution for specific water quality parameters. However, the lack of temporal control might affect the user if there is high turbulence in the water, or heavy rainfall that changes the whole water composition in a short time. This would mean that stakeholders in areas with high flood risk would struggle to get good information from this method, even thgouh they statistically are lowest-end users with the most too loose. too sad

From discussion “

The turbidity can be directly compared to Assegide et al. 2023. The satellite images the PDM is based on are from the same lake and time period. Assegide et al.’s turbidity lab results were similar to the PDM. The laboratory values were between 38-148 NTU, and average was 52, while the PDM had 25-150 NTU in the water.

“ extract conclution

Answer these

1. Which water quality parameters are measurable using the NICFI satellite images?
2. To what degree can remotely sensed water quality models be extrapolated to different lakes?

3. How accurate can could a NICFI water quality model be without field samples? UNANSWERED
4. How viable is NICFI for low-end users to monitor water quality?

REFERENCES

Abhachire, L. W. (2014). Studies on Hydrobiological Features of Koka Reservoir and

Awash River in Ethiopia *Studies on Hydrobiological Features of Koka Reservoir and*

Awash River in Ethiopia 158-162. https://www.fisheriesjournal.com/archives/2014/vol1issue3/PartC/C24.pdf

Assegide, E., Alamirew, T., Dile, Y. T., Bayabil, H., Tessema, B., & Zeleke, G. (2022). A Synthesis of Surface Water Quality in Awash Basin, Ethiopia. *Frontiers in water*, *4*. https://doi.org/10.3389/frwa.2022.782124

Assegide, E., Shiferaw, H., Tibebe, D., Peppa, M. V., Walsh, C. L., Alamirew, T., & Zeleke, G. (2023). Spatiotemporal Dynamics of Water Quality Indicators in Koka Reservoir, Ethiopia. *Remote sensing (Basel, Switzerland)*, *15*(4), 1155. https://doi.org/10.3390/rs15041155

Brivio, P. A., Giardino, C., & Zilioli, E. (2001). Determination of chlorophyll concentration changes in Lake Garda using an image-based radiative transfer code for Landsat TM images. *International journal of remote sensing*, *22*(2-3), 487-502. https://doi.org/10.1080/014311601450059

Dick, Ø. B., & Birkeland, R. (2023). *Vegetasjonindeks* [Encyclopedia]. https://snl.no/vegetasjonsindeks

Ekercin, S. (2007). Water Quality Retrievals from High Resolution Ikonos Multispectral Imagery: A Case Study in Istanbul, Turkey. *Water, air, and soil pollution*, *183*(1-4), 239-251. https://doi.org/10.1007/s11270-007-9373-5

Gholizadeh, M. H., Melesse, A. M., & Reddi, L. (2016). A Comprehensive Review on Water Quality Parameters Estimation Using Remote Sensing Techniques. *Sensors (Basel)*, *16*(8), 1298. https://doi.org/10.3390/s16081298

Helmenstine, A. (2021). *Visible Light Spectrum Wavelengths and Colors*. Science Notes. Retrieved 10/07 from https://sciencenotes.org/visible-light-spectrum-wavelengths-and-colors/

Helseth, L. E. (2022). *Spektrometer* https://snl.no/spektrometer

Holtsmark, T., & Angelo, K. (2021). *Farger* https://snl.no/farge

Houborg, R., & McCabe, M. F. (2018). A Cubesat enabled Spatio-Temporal Enhancement Method (CESTEM) utilizing Planet, Landsat and MODIS data. *Remote Sensing of Environment*, *209*, 211-226. https://doi.org/https://doi.org/10.1016/j.rse.2018.02.067

Ingwani, E., Gumbo, T., & Gondo, T. (2010). The general information about the impact of water hyacinth on Aba Samuel Dam, Addis Ababa, Ethiopia: Implications for ecohydrologists. *Ecohydrology & Hydrobiology*, *10*(2), 341-345. https://doi.org/https://doi.org/10.2478/v10104-011-0014-7

Lied, F., & Birkeland, R. (2022). *Fjernmåling*. Store Norske Leksikon. Retrieved 10/07 from https://snl.no/fjernm%C3%A5ling

Mekuria, D. M., Kassegne, A. B., & Asfaw, S. L. (2020). Little Akaki River sediment enrichment with heavy metals, pollution load and potential ecological risks in downstream, Central Ethiopia. *Environmental Systems Research*, *9*(1), 23. https://doi.org/10.1186/s40068-020-00188-z

Norsaliza, U., & Ismail, M. H. (2010). Use of Remote Sensing and GIS in Monitoring Water Quality. *Journal of Sustainable Development*, *3*. https://doi.org/10.5539/jsd.v3n3p228

NV5-Geospatial. *Digital Number, Radiance, and Reflectance*. NV5 Geospatial. Retrieved 17/7 from https://www.nv5geospatialsoftware.com/Support/Maintenance-Detail/ArtMID/13350/ArticleID/16278/Digital-Number-Radiance-and-Reflectance

O'Connor, B., Moul, K., Pollini, B., de Lamo, X., & Simonson, W. (2020). *Earth Observation for SDG Compendium of Earth Observation contributions to the SDG Targets and Indicators* [Compendium](ESA Contract No. 4000123494/18/I-NB). https://eo4society.esa.int/wp-content/uploads/2021/01/EO_Compendium-for-SDGs.pdf

Planet. (2023). *Norway's International Climate and Forests Initiative Satellite Data Program*. Planet Labs. Retrieved 10/07 from https://www.planet.com/nicfi/

Pooja Pandey, Joe Kington, Anudeep Kanwar, Robert Simmon, & Abraham, L. (2023). *PLANET BASEMAPS FOR NICFI DATA PROGRAM ADDENDUM TO BASEMAPS PRODUCT SPECIFICATION*. Planet Labs. Retrieved 17/07 from https://assets.planet.com/docs/NICFI_Basemap_Spec_Addendum.pdf

UNESCO. (2018). *The United Nations World Water Development Report 2018: Nature-Based Solutions for Water.*: UNESCO Retrieved from https://unesdoc.unesco.org/ark:/48223/pf0000261424

WFP. (2023). *WFP Ethiopia Drought Response Situation Report #7 (January – March 2023)*. World Food Program Retrieved from https://reliefweb.int/report/ethiopia/wfp-ethiopia-drought-response-situation-report-7-january-march-2023

Yang, H., Kong, J., Hu, H., Du, Y., Gao, M., & Chen, F. (2022). A Review of Remote Sensing for Water Quality Retrieval: Progress and Challenges. *Remote sensing (Basel, Switzerland)*, *14*(8), 1770. https://doi.org/10.3390/rs14081770

Yohannes, H., & Elias, E. (2017). Contamination of Rivers and Water Reservoirs in and Around Addis Ababa City and Actions to Combat It. *Environment Pollution and Climate Change*, *01*. https://doi.org/10.4172/2753-458X.1000116

Appendix

Add appendix here

A-1 Raster Calculator input

Novel PDM for Chlorophyll-a Raster calculator:

-123.4904225 - 0.088955872 "L15-1246E-1071N@1" + 0.293198207 "L15-1246E-1071N@2" - 0.194393362 "L15-1246E-1071N@3" + 0.112203264 "L15-1246E-1071N@4"

The inputs to produce the PDMs in QGIS Raster Calculator and the value spread:

miller tsm

-1.91 1140.25 ("aba2@3"/10000)

1.91 1140.25 ("se2@3"/10000) (try positive)

1.91 1140.25 ("aba2@3"/10000)

130-240

bilge no3

aba:

2.84 - 0.06 ("aba2@1" 0.0001) - 0.05 ("aba2@2" 0.0001) + 0.06 ("aba2@3" 0.0001) + 0.38 ("aba2@4" 0.0001)

2.861-2.894

gef:

2.84 - 0.06 ("gef2@1" 0.0001) - 0.05 ("gef2@2" 0.0001) + 0.06 ("gef2@3" 0.0001) + 0.38 ("gef2@4" 0.0001)

2.861-2.894

chl:

44.20 - 1.17 ("se2@1" 0.00001) - 0.88 ("se2@2" 0.00001) + 1.49 ("se2@3" 0.00001) + 4.08 ("se2" 0.00001)

44.20+-2

koka CHL

44.20 - 1.17 ("se2@1"/10000) - 0.88 ("se2@2"/10000) + 1.49 ("se2@3"/10000) + 4.08 ("se2@4"/10000)

44-46 (ABSOLUT MINMAX)

SS koka

6.50 - 0.73 ("se2@1" 0.0001)-1.16 ("se2@2" 0.0001)+3.00 ("se2@3" 0.0001) + 3.65 ("se2@4" 0.0001)

6.5-8.3

ALL bilge will be barly any value over constant

CHL:

44.20-1.17 ("se2@1"/10000)-0.88 ("se2@2"/10000)+1.49 ("se2@3"/10000)+4.08 ("se2@4"/10000)

+-minimal

Keith chl:

60.703 (("se2@1"^-1) - ("se2@2"^-1)) "se2@4" + 10.386

25-150 chl-

Kutser CDOM

if("se2@2" > 0, 5.13 (("se2@2" / "se2@3") ^ -2.67), 0)

1-8 CDOM

Mancino chl

-47.515 + 9.516 ("se2@3"/"se2@2") + 20.952 ("se2@1"/"se2@2") - 873.0 ("se2@2"/10000) + 34.889 ("se2@2"/"se2@1")

-80-1.6 chl

Giardino chl

11.18 ("se2@1"/10000)-8.96 ("se2@2"/10000)-3.28

-3.69-(-3.2) SMOLL

Assegide TTS

3938.9 ("se2@3"/10000)-536.9

-300-0

Assegide TU Koka

282.88 ("1246-1071@3" / "1246-1071@2") - 206.15

-27-15

natural logarithm

Allan non log chl

14.141-5.0568 ("se2@1"/"se2@3")

10.35-11.65

Allan with log

2.718281828459045^(14.141-5.0568 ("se2@1"/"se2@3"))

34000-89000

Brezonik

non lin

-2.663-0.03191 ("aba2@1" 0.0001) + 1.1030 ("aba2@1" 0.0001/"aba2@3" 0.0001)

-2.664 and -2.665

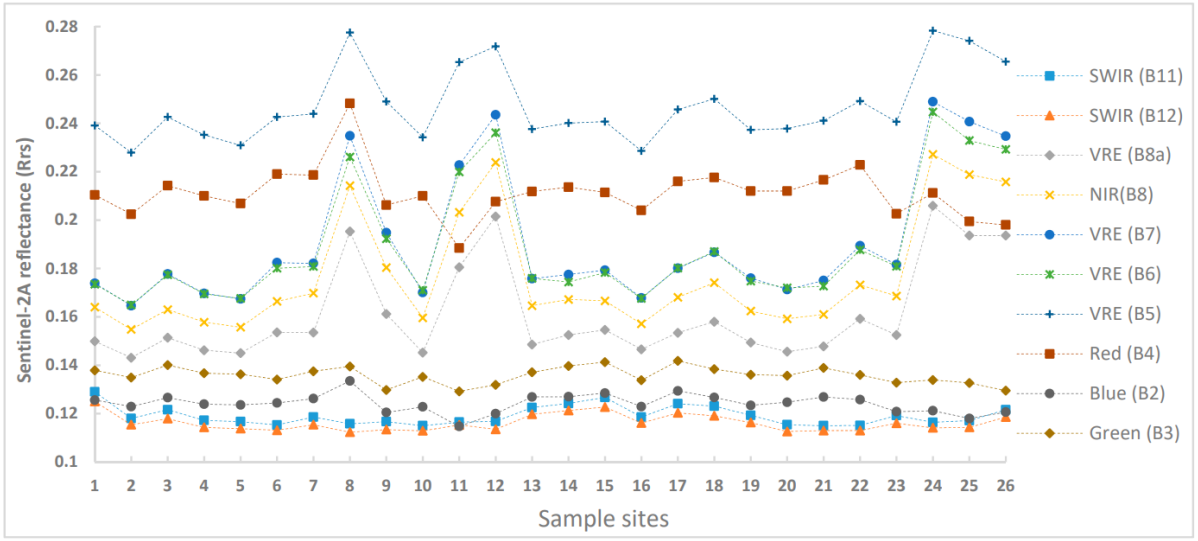
with nat lin

2.718281828459045^(-2.663-0.03191 ("aba2@1" 0.0001) + 1.1030 ("aba2@1" 0.0001/"aba2@3" 0.0001))

0.069 +-0.0003 SMOLL

A-2 Something results

Assegide et al. 2023 reflectance variation across the sample sites:



Original without removing bands unrelated to NICFI Basemap.

A-3 Residuals for Original WQM from Koka

RESIDUAL OUTPUT

Observatio	Predicted Y	Residuals	Standard Residuals
1	17.700426	0.542574	0.045327095
2	20.604895	-8.44289	-0.705326665
3	22.211305	1.244695	0.103982921
4	23.341958	-1.62396	-0.135666847
5	9.6887649	6.817235	0.569517662
6	15.062779	6.655221	0.555982903
7	8.9129314	8.118069	0.678190406
8	26.617774	-7.76877	-0.649009994
9	17.474771	-7.04977	-0.58894393
10	66.902089	39.07791	3.264602117
11	70.296391	-20.7794	-1.735928149
12	27.179598	-11.9676	-0.999783352
13	16.558815	1.260185	0.105276899
14	18.051185	1.060815	0.088621423
15	23.799761	-7.73776	-0.646419201
16	21.085371	-0.23637	-0.019746621
17	16.126871	2.985129	0.249380265
18	10.003881	7.371119	0.615789615
19	18.638455	0.048545	0.004055468
20	15.735208	3.376792	0.282100132
21	16.697316	-1.92932	-0.161176689
22	26.036778	-9.02478	-0.753937705
23	71.416191	-18.6982	-1.562062876
24	60.674487	16.70051	1.395175121

A-4 Nitrate Test Strip Photos

The MQuant Nitrate test strips indicate their results by comparing the reacted reagent squares with the scale supplied on the container. Results are presented in Chapter 5, but images are kept here to validate observation.

Dire Dam

Nitrate strips result from Dire Dam by Legedadi:

The result of the test appears to be 10mg/L for NO3- and 2.3 mg/L for NO3-N.

Aba Samuel

Result of Nitrate Strip for first site on Aba Samuel. No image from the sample site as it was too muddy to walk over to.

Aba Samuel Site A:

25mg/L NO3 and 5.6 mg/L NO3-N

Aba Samuel Site B

50mg/L NO3 and 11mg/L NO3-N but slightly close to the 25/5.6 value as well

A-5 Koka Samples and NICFI DNs

id	Blue	Green	Red	NIR	Chlorophyl-a	Turbidity	Total Suspended Solid
1	669.1	1135.0	952.1	612.8	3.475	38	218
2	701.6	1150.3	945.2	446.2	18.243	38	286
3	679.7	1133.8	892.1	406.0	12.162	-	222
4	719.4	1172.8	1011.1	556.1	23.456	44	288
5	683.9	1146.3	935.9	476.8	21.718	52	228
6	769.4	1212.3	1152.8	626.2	16.506	46	308
7	690.2	1117.4	1040.8	665.2	21.718	64	210
8	724.4	1163.7	1092.2	605.9	17.031	100	338
9	600.4	1028.9	927.3	731.9	18.849	48	860
10	433.9	832.4	625.0	507.9	10.425	42	192
11	487.6	952.9	666.1	747.4	105.98	-	514
12	583.6	1077.2	845.2	839.2	49.517	34	436
13	667.0	1115.4	885.1	490.3	15.212	-	247
14	674.7	1115.1	925.7	472.9	17.819	-	226
15	679.6	1134.1	909.8	412.9	19.112	36	246
16	699.3	1146.7	911.0	449.1	16.062	-	197
17	675.2	1133.4	901.1	423.2	20.849	40	247
18	664.6	1119.4	884.7	378.7	19.112	52	212
19	683.1	1130.0	959.4	440.8	17.375	46	402
20	708.1	1162.6	1041.0	593.8	18.687	40	226
21	718.7	1177.7	1103.3	644.8	19.112	52	223
22	659.9	1077.0	1019.1	723.9	14.768	52	827
23	609.8	1012.9	956.1	825.8	17.012	48	235
24	539.7	1028.1	762.0	798.6	52.718	-	318
25	599.9	1067.7	829.2	763.7	77.375	44	606
26	595.3	1078.7	818.9	737.0	396.14	148	317

27	731.4	1188.6	1032.7	547.3	-	72	227

Norges miljø- og biovitenskapelige universitet
Noregs miljø- og biovitskapelege universitet
Norwegian University of Life Sciences

Postboks 5003
NO-1432 Ås
Norway