

https://dspace.org/

# Water quality assessment at Vaalkop Dam in North-West Province using multispectral remote sensing data

**Mthimkulu, Lindy Ndzumbululo**

**2022**

University of the Witwatersrand, Johannesburg

https://hdl.handle.net/10539/37007

Mthimkulu, Lindy Ndzumbululo. (2022). Water quality assessment at Vaalkop Dam in North-West Province using multispectral remote sensing data [Master's dissertation, University of the Witwatersrand, Johannesburg]. WireDSpace. https://hdl.handle.net/10539/37007

Downloaded from a DSpace repository. https://dspace.org/

# UNIVERSITY OF WITWATERSRAND



UNIVERSITY OF THE
WITWATERSRAND,
JOHANNESBURG

## FACULTY OF SCIENCE

# Water Quality Assessment at Vaalkop Dam in North-West Province using Multispectral Remote Sensing Data

Lindy Ndzumbululo Mthimkulu (378855)

**Animal, Plant and Environmental Sciences**

**A research report submitted to the Faculty of Sciences, the University of Witwatersrand in partial fulfilment of the degree of Master of Science**

**(**Environmental Sciences**)**

**Supervisor:** Prof. Elhadi Adam

**Johannesburg, October 2022**

# Declaration

I declare that the work on this dissertation, "Water quality assessment at Vaalkop Dam in North-West Province using multispectral remote sensing data," is completely my own. This has been submitted for the University of the Witwatersrand's Master of Environmental Sciences. It has never been submitted for any other university's degree or qualification.

Mthimkulu ---------- Signature of the candidate

---------- Day of ---04/10---------2022---------- in……

# Acknowledgments

I would like to thank my employer, Magalies Water, for granting me the opportunity to study and invest in my education.

My appreciation goes to Prof. Elhadi Adam for assisting and supervising this study. I want to also convey my appreciation to Norval Lippiatt and his wife for providing me with their boat and spending the full day sampling the dam with me.

My sincere gratitude to my husband (Leabua Mthimkulu), who assisted me to gather 108 samples in 8 hours. I further owe the time I spent working on this dissertation to my children Lesedi and Mohau; Mohau was only two months old when I collected the data.

# Table of Contents

**CHAPTER 1** ..... 1
**INTRODUCTION** ..... 1
**1.1 General Introduction** ..... 1
**1.3 Aim and Objectives** ..... 4
**1.3.1 Aim** ..... 4
**1.3.2 Objectives** ..... 4
**CHAPTER 2** ..... 5
**LITERATURE REVIEW** ..... 5
**2.1 Factors affecting water quality at the source** ..... 5
**2.1.1 Natural Factors** ..... 5
**2.1.2 Human Factors** ..... 5
**2.2 Water Quality Monitoring** ..... 6
**2.4 The use of Multispectral Data for Monitoring Water Quality** ..... 7
**2.5 Remote sensing's challenges and benefits for water quality** ..... 8
**2.6 Regression models for Water Quality Assessment** ..... 8
**CHAPTER 3** ..... 10
**MATERIALS AND METHODS** ..... 10
**3.1 Study Area** ..... 10
**3.2 Methodology** ..... 11
**3.1.1 Sampling Data Acquisition** ..... 11
**3.1.2 Remote Sensing Data Acquisition and Pre-processing** ..... 12
**3.1.3 Pre-processing and atmospheric correction for Sentinel-2 data** ..... 13
**3.1.4 Landsat-8 Data pre-processing and atmospheric correction** ..... 14
**3.1.5 Description of water quality parameters** ..... 16
**3.1.6 Regression Modeling** ..... 17
**3.1.7 Accuracy Assessment** ..... 18
**CHAPTER 4** ..... 19
**RESEARCH FINDINGS** ..... 19
**4.1 Descriptive analysis of the ground measured Chlorophyll-a and Turbidity** ..... 19
**4.2 Spatial pattern of the ground measured Chlorophyll-a and Turbidity** ..... 20
**4.3 Chlorophyll-a Results** ..... 20
**4.3.1 Correlation results** ..... 20

**4.3.2 Support vector regression results** ..... 22
**4.4 Turbidity Results** ..... 24
**4.4.1 Correlation Results** ..... 24
**4.4.2 Support vector regression results** ..... 26
**CHAPTER 5** ..... 28
**DISCUSSION** ..... 28
**CHAPTER 6** ..... 30
**CONCLUSION** ..... 30
**6.1 Conclusion** ..... 30
**6.2 Recommendations and Limitations** ..... 30
**REFERENCES** ..... 31

# List of Figures

**Figure 1:** Shows the location of the study areas on the map of the Bojanala district in the North-West Province ..... 11
**Figure 2:** Map showing sampling points 1-27 at the dam ..... 12
**Figure 3:** Sentinel-2 image ..... 13
**Figure 4:** Landsat-8 8 OLI image ..... 15
**Figure 5**: Density distribution of ground measured chlorophyll-a (on left) and turbidity (on right) ..... 19
**Figure 6:** Box and whisker plot showing the variability of the ground measured chlorophyll-a and turbidity respectively ..... 19
**Figure 7**: Spatial pattern of the ground measured Chlorophyll-a and Turbidity ..... 20
**Figure 8**: Chlorophyll-a correlation results with A. Sentinel-2 was acquired on 16th October 2020 and B. Landsat-was 8 8 acquired on 22nd October 2020 respectively ..... 21
**Figure 9**: Chlorophyll-a correlation results with (A) Sentinel-2 acquired on 25th November 2020 and (B) Landsat-8 8 acquired on 07th November 2020 respectively ..... 21
**Figure 10:** Landsat-8 8 plots of showing relationships between actual and predicted chlorophyll-a for (a) training set (n=20), (b) validation set (n=8) ..... 23
**Figure 11:** Sentinel-2 plots of showing relationships between actual and predicted chlorophyll-a for (a) training set (n=20), (b) validation set (n=8) ..... 24
**Figure 12:** Turbidity correlation results with (A) Sentinel-2 acquired on 16th October 2020 and (B) Landsat-8 8 acquired on 22nd October 2020 respectively ..... 24
**Figure 13** Turbidity correlation results with A. Sentinel-2 was acquired on 25th November 2020 and B. Landsat-was 8 8 acquired on 07th November 2020 respeon ctively ..... 25
**Figure 14**: Sentinel-2 plots showing relationships between actual and predicted turbidity ..... 26
**Figure 15**: Landsat-8 8 plots showing relationships between actual and predicted turbidity ..... 26

# List of Tables

**Table 1** Specifications for Sentinel-2 remote sensor ..... 13
**Table 2** Spectral properties for Sentinel -2 vs Landsat-8 8 ..... 15
**Table 3** Specifications for Landsat-8 8 remote sensor ..... 15
**Table 4** Chlorophyll-a Landsat-8 8 acquired on 22nd October 2020 and Sentinel-2 ..... 22
**Table 5** Chlorophyll-a Landsat-8 8 and Sentinel-2 correlation with Sentinel-2 and Landsat-8 8 ..... 22
**Table 6** SVR chlorophyll-a results ..... 23
**Table 7** Turbidity Landsat-8 8 acquired on 22nd October 2020 and Sentinel-2 acquired on 16th October 2020 correlation ..... 25
**Table 8** Turbidity Landsat-8 8 acquired on 07th November 2020 and Sentinel-2 correlation acquired on 25th November 2020 ..... 25
**Table 9** SVR turbidity results in ..... 26

# List of Abbreviations

| | |
|---|---|
| DN | Digital Number |
| ENVI | Environment for Visualising Images |
| ESA | European Space Agency |
| ETM | Enhanced Thematic Mapper |
| FLAASH | Fast Line-of-sight Atmospheric Analysis of Hypercubes |
| GEMI | Global Environmental Monitoring Index |
| GOCI | Geostationary Ocean Colour Imager |
| MERIS | Medium Resolution Imaging Spectroradiometer |
| MIP | Modular Inversion and Processing System |
| MODIS | Moderate Resolution Imaging Spectroradiometer |
| MSI | Multspectral Instruments |
| MSS | Multispectral Scanner |
| NASA | National Aeronautics and Space Administration |
| NIR | Near InfraRed |
| NTU | Nephelometric Turbidity Unit |
| OLI | Operational Land Imager |
| RF | Random Forests |
| RFR | Random Forests Regression |
| RMSE | Root Mean Square Error |
| RSR | Remote sensing reflectance |
| SPOT | Satellite Pour I'Observation de la Terre |
| SVM | Support Vector Machine |
| SVR | Support Vector Regression |
| SWIR | ShortWave InfraRed |
| TIRS | Thermal Infrared Sensor |
| TM | Thematic Mapper |
| TOA | Top of the Atmosphere |
| TSS | Total Suspended Solids |
| Ug/l | Micrograms per litre |

# CHAPTER 1
# INTRODUCTION

## 1.1 General Introduction

Water is a life-sustaining resource for all organisms. As much as nature share the value of water with human being it cannot supersede how human have become dependent on water for drinking, sanitation, farming, mining, and other industrial purposes such as power generation. Water is perceived as a vital component in the fight against poverty, the foundation of any successfully growing and developed country is also measured on water and sanitation provision (Basson *et al.,* 1997). The water law in South Africa stresses that essential human and ecological necessities must be given (DWAF, 1998) and that the misuse of water from all perspectives should be feasible (New, 2002).

South Africa is one of the nations in the globe receiving less than 500 mm of mean annual rainfall (Dallas & Rivers-Moore, 2014). According to the Department of Water and Sanitation Master plan 2017/18 report: water security is a critical challenge directly facing South Africa which affects the social welfare and economic growth of the country. Urbanisation, population growth, degradation of wetlands, decrease in rainfall, and water losses are all estimated to increase the current water scarcity. The degree of water availability and other dependencies, such as the availability of potable water and sewerage system and contamination of the water supply, affect the country's economic development (Meissner *et al.,*2018). When the interest in water use increases due to pressure in human population and improvement in economic activities, the ecosystem of water bodies will keep on decaying except if they are overseen in a maintainable manner. The outstanding expansion in the entire population, the continued refinement of its necessities and the application for the support of the current way of life, and the cycle of industrialisation have not exactly come about in tremendously expanded pressing factor and exhaustion of water asset, however, they have likewise caused the current day era of enormous amounts of waste (Strydom *et al.,* 2009).

The physicochemical properties, as well as biological components such as algae, are used to determine its quality, physical condition (e.g., temperature), and microbiological composition (Matthews *et al.,* 2015). Poor water quality is likely to cause waterborne diseases which can endanger human health and safety, reduced visibility with negative economic impacts that cause the cost of treatment to increase (Matthews *et al.,* 2015). Both the quality and quantity of water are required to meet essential needs and to be shared amongst industry, agricultural and ecosystem regimes (Matthews *et al.,* 2015). Nitrates such as (fertilizers, septic systems, animal feedlots, industrial waste, and food processing waste) are a major factor in the degradation of water quality (Shabalala *et al.,* 2013). Disintegrated natural carbon and pathogen-inducing chemicals, as well as salts and warm pollution from the oceans, are expected to be compounded by rising temperatures and longer periods of low flow (Bates *et al.,* 2008). Freshwater habitats in the country are also being affected by climate change, which has direct effect on the quality of water (Dallas & Rivers-Moore, 2014). A change in the quality of water and its temperature influences the solvency of oxygen and the different gases, toxicity, reaction rates of chemical and microbial activity (Dallas

& River-Moore, 2014). A decrease in the concentration of dissolved oxygen, especially in respect of joined impacts of high temperatures and low streams is averse to aquatic organisms including amphibians (Dallas & Day, 2004). Water quality changes with time and seasons as variables such as turbidity, nutrients, and phosphorus respond to more rainfall events with sediments washed in from the catchment (Oberholster & Ashton, 2008). According to Dallas and Rivers-Moore, the freshwater in South Africa is currently enhanced from moderate to high eutrophic levels. The effects in further changes in nutrient loads and nutrient cycles may bring about increased algal development, changes in eutrophic conditions, increased frequencies of cyanotoxins that influence human well-being contrarily. In South Africa, most eutrophic waterways and reservoirs have predominant phytoplankton genera, cyanobacteria *Microcystis sp.* and *Anabaena sp* (Dallas & River-Moore, 2014).

The freshwater quality of the available sources has deteriorated because of rising pollution by mining, agriculture, afforestation, urbanization, and industry such as power generation (Oberholster & Ashton, 2008). Monitoring water quality is critical to detect any current or potential problems, especially considering the rising pollution levels in our water resources. Monitoring water quality can also reveal any long-term patterns or changes in water bodies (Wang & Yang, 2019) while weak monitoring systems pose a high risk to decision making and planning due to a lack of data and information, government, and businesses such as water boards are required to meet a range of water quality goals (DWS 2017/18). Therefore, new monitoring technologies and improving data management need to be adopted. However, biological, and chemical analyses of water quality utilising in-situ water samples and monitoring are expensive, consume time, as well as labour-intensive, and carry the risk of one slipping off into the reservoirs accidentally (Dube *et al.,* 2015). Due to the constraints of traditional in-situ sampling methods, satellite image processes enable the utilisation of remote-sensing data for vast and long-term water quality assessment (Wang & Yang, 2019). Remote-sensing techniques have progressed from the limitations of the multispectral sensors to hyperspectral sensors, achieving maximum regression and the production of detailed water quality mapping in a less time and labor (Wang & Yang, 2019). Satellite imagery in water allows for the determination of water bodies' physico-chemical (temperature, turbidity, color, salinity, pH, and dissolved oxygen) and biological (bacteria and algae) features (Dube *et al.,* 2015). Although it is costly to use remote-sensing methods for observing huge bodies with high temporal coverage at a reasonable level of accuracy (Chawla *et al.,* 2020). Recent research in many parts of the world involving water quality have made use of data from remote sensing, Pizani *et al.,* 2020 compared the performance of Sentinel-2 MSI and Landsat-8 OLI sensors to produce multiple regression models of water quality parameters (chlorophyll-a, Secchi disk depth, turbidity, and temperature) in a hydroelectric reservoir in Brazil. The results showed that optical active parameters yielded strong regression models from both the Sentinel-2 and Landsat-8 sensors, all with $R^2 > 0.75$. Another study was conducted in Lake Chad to test suitability of Sentinel-2 Multispectral Imager's (MSI) data for mapping different lake water quality parameters ( chlorophyll- a (Chl a), water color, colored dissolved organic matter (CDOM) and dissolved organic carbon (DOC) .They were able to show that there is good correlation between band ratio algorithms calculated from Sentinel-2 MSI data and lake water parameters like Chl a ($R^2 = 0.83$), CDOM ($R^2 = 0.72$) and DOC ($R^2 = 0.92$) concentrations as well as water color ($R^2 = 0.52$) (Buma *et al.*, 2020)

## 1.2 Problem Statement

The quality of water must be monitored and assessed on a regular basis to sustain the water resource management. The Vaalkop Dam in the Province of North-West is crucial for balancing and conserving water for urban and rural usage, farm management and aquaculture (DWS, 2012, Swanelpoel *et al.,* 2017). Hartbeespoort dam is bounded by mines and agricultural activities which have given rise to the increase of mining towns (Davis, 2017). Due to the expansion of the population which led to the demand for water for commercial farming, the amount of sewage entering the dam has caused excessive contamination of water (Ashton *et al.,* 1985). The cause of water pollution such as nutrients influx from the rivers, led to the development of algal blooms and an increase rate of eutrophication was identified in the Hartbeespoort dam (Harding, 2008). Through a framework of canals within the lower waterway, the Hartbeespoort dam can moreover bolster the Vaalkop dam. However, the water abstraction from the Vaalkop Nature Reserve dam consists of tremendous concentrations of cyanobacteria, cyanotoxins, heavy metals, and infestation of water bodies with hyacinth (Swanelpoel *et al.,* 2017). The presence of emerging contaminants from upstream Wastewater Treatment Plants further adds to poor water quality and fertilizers from commercial farming through eutrophication and development of algal blooms (Swanelpoel *et al.,* 2017). The relationship between the dams is that the entrance of chemically enriched water through the canal from the Roodekopjes dam which receives water from Hartbeespoort Dam adds to the causes of reduced water quality in the Vaalkop Dam (Swanelpoel *et al.,* 2017).

The quality of water utilised for residential use and other purposes, and the security of water assets, especially in a water-scarce nation as South Africa, is fundamentally everyone's concern and ought to be overseen concurring to scientific standards (DWAF,1998). The national government has executed controls and approaches to provide secure water to all, but a few water supplier entities a municipality getting bulk water from State-owned entities have not fundamentally caught up with the national rules. The small compliance rate is mostly due to certain reasons such as skill shortages, under-resourcing, need of understanding of required measures, need of intercession to address issue zones, lacking administration, and impediments on accounts, resources, and monetary responsibility (Du Plessis, 2006).

The water quality monitoring of the dams has for longest time been conducted at fixed points i.e., dam wall, and these using the in-situ water sampling method, which is costly, time-consuming, and very labour-intensive (Dube *et al.,* 2015). Consequently, there is now a requirement to assess the usability of methodologies that allow for routine water quality assessment at a lower cost and bring value to water resource management decision-making. Water quality monitoring techniques based on remote sensing are seen as cost-effective ways to monitor water quality since they enable for continuous monitoring of small to large areas in a reasonable amount of time (Somvanshi *et al.,* 2012).

## 1.3 Aim and Objectives

### 1.3.1 Aim

This research report aimed to assess the feasibility of integrating remote sensing and in-situ measurements in the water quality assessment in the Vaalkop dam. To compare Landsat-8, Sentinel-2, and in-situ data in mapping the quality of water in Vaalkop Dam.

### 1.3.2 Objectives

- To test the effectiveness of Sentinel-2 and Landsat-8 data and Support Vector Machine in mapping water turbidity and chlorophyll-a in the study area.

- To compare the performance of Sentinel-2 and Landsat-8 bands in detecting and mapping water turbidity and chlorophyll-a.

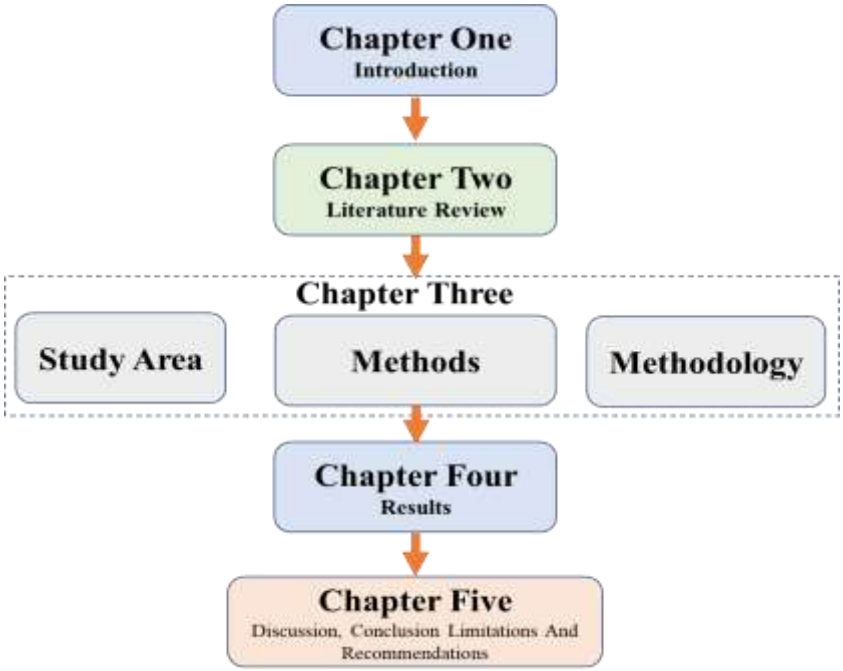
## 1.4 Outline of the research report

The report was organized in to five chapters:

- Chapter 2: The literature review which focused on multispectral remote sensing water.

- Chapter 3: This report described the study area, how data was collected and processed, and how forecast models were developed for turbidity and chlorophyll-a.
- Chapter 4: The results of the research in a comparative manner.

- Chapter 5: Focused on discussion, conclusion, the research's limitations, and recommendations




**Outline of the research report showing the sequential workflow that was followed**

# CHAPTER 2
# LITERATURE REVIEW

## 2.1 Factors affecting water quality at the source

Water quality is a term that refers to the chemical, physical, and biological characteristics of water in terms of its suitability for a specific purpose (Bartram, 1996). Most studies have shown that water quality is impacted from agriculture, waste treatment, and industrial discharges, or naturally (Dutta et al., 2005; Ogleni & Topal, 2011). The magnitude of the effects of human actions on freshwater varies by region. Changes in water quality have an impact on both people and ecosystems (Wagner et al., 2013). Water contamination and inefficient utilisation of freshwater would cripple any form of life and any development endeavours. For instance, inadequate water quality results in a variety of financial costs, including corrupt practices of environmental administrations, wellness expenses, implications on financial activities which including agribusiness, mechanical generation, and leisure, increased water cost of treatment, and decreased land values, and many others (UNESCO, 2012). The Vaalkop dam receives water from the Hex, Elands, and Crocodile Rivers. The following sections discuss the variables that impact quality of the water at the source.

### 2.1.1 Natural Factors

Climate change, algal bloom, high nutrient loading, siltation, and hyacinth infestation of water bodies are all natural elements affecting the quality of water at the source, dam, or river. There are secondary problems such as eutrophic water from the decomposition of aquatic plants hence depleting the water dissolved oxygen (Ashton *et al.,* 1985). Eutrophication is a result of natural and human interference, natural processes of eutrophication are controlled by the topsoil composition, local geology, and natural environmental factors of the catchment (Walmsley, 2003). Climatic condition among the Crocodile (West) Marico water management area is temperate and semi in the east to dry in the west. Temperature variations in the area are minimal meaning that Hex River does not experience serious variations in water temperature except additional heat or cold is added by anthropogenic (Du Plessis, 2006).

### 2.1.2 Human Factors

A growing threat to water resources and biodiversity has emerged from human-induced river pollution. As water quality declines, this may have adverse effect on human health and negative impact on the economy, particularly in agriculture and industries. Disease-causing microbes can be found in tainted drinking water that can cause waterborne illnesses such as cholera and other gastrointestinal issues (Ashton *et al.,* 2008). Even though the waterways have a self-decontamination limit, this limit has been altered caused by human activity in the stream catchment, prompting its eradication of such an important biological system (Smith, 2003). Surface water is generally exposable to contamination on account of its effective openness for the removal of wastewaters. The anthropogenic impacts also increase the exploitation of the water resources caused by agricultural, mining, industrial, and urban activities (Samarghadi *et al.,* 2007). The human factors that affect water quality are agricultural activities, heavy metals from the

mining industry, and wastewater treatment work (Ashton et al., 1985. As a result of agricultural fertilizer runoff, vegetation and periodical algal blooms emerge, degrading the oxygenation regime and resulting in deteriorated water quality (Strydom *et al.,* 2009). Mineral waste flows into surface waters near mining operations pose a significant danger of contamination (UNEP, 2010). Underground stopes and explosive residues have been shown to have adverse effects of varying intensity on water surfaces due to many platinum deposits as there are several mines around Vaalkop Dam catchment mined by different mining companies (Hamann, 2003). The tremendous concentration of lead and cadmium was detected in the Crocodile River which feeds both dams with water (DWS, 2018).

## 2.2 Water Quality Monitoring

Freshwater is quickly becoming the world's most pressing natural concern. The world's population is currently expanding, and fresh water supplies are limited while consumption for the resources rises (Hinrichsen &Tacio, 2002). As the demand for the resources continues to increase it tends to be overused and becomes difficult to monitor, leading to depletion and pollution of surface and groundwater (Hinrichsen &Tacio, 2002). All kinds of life and development can be adversely affected by water pollution, which has an indirect effect on economic costs, health-related costs, property values, and the cost of water treatment (Kapalanga, 2016). Therefore, the continuous monitoring of the quality of water is essential to the preservation of freshwater resources (Kannel *et al.,* 2007).

The focal goal of monitoring water resources in South Africa is to ensure the adequate and effective use of water to assist the proper and sustainable social and economic advancement (Du Plessis, 2006). To manage water requires that the accessible water be managed appropriately, designated reasonably and that the nature of accessible water is secured. Polluted water must be cleaned and ensure the quality of water resources should not be compromised (Barnard, 1999). Physical and chemical characteristics have been used to assess the quality of freshwater systems using a variety of approaches around the world (Adokole *et al.,* 2003). Most researchers have recently added techniques for assessing water quality particularly in the use of remote sensing (Hansen et al., 2015).

## 2.3 Remote Sensing of Water Quality

Based on remotely sensed reflectance, remote sensing can be used to determine the physical, chemical, and biological aspects of water bodies (Dube *et al.,* 2015). Broadband reflectance correlations with water characteristics including chlorophyll-a concentrations and total suspended sediments have been employed in the 1980s to evaluate the inland water quality via satellite remote sensing (Mancino *et al.,* 2009). The use of satellite images to assess water quality in Africa appears to be under-researched (Dube *et al.,* 2015). Continuous water quality monitoring is made possible using remote sensing data (Ritchie *et al.,* 2003). There has been some success in measuring chlorophyll-a using the remote sensing ratio of near-infrared (705nm) reflectance to red (670nm) reflectance. Other investigations have identified a substantial positive association between suspended sediments and spectral radiance (Masocha *et al.,* 2017). Remote sensing data was

utilized to develop an algorithm for assessing reflectance of chlorophyll-a, turbidity, and other water pollution elements in this study.

### 2.4 The use of Multispectral Data for Monitoring Water Quality

Landsat-8 multispectral sensor (USGS) is overseen by the United States Geographical Survey (USGS). Landsat information congruity mission (LDCM) completed data collection, which started with the launch of Landsat 1 in 1972 and provided a more thorough explanation of the sensor's features (Roy *et al.*, 2014). A Landsat-8 instrument called Operational Land Imager (OLI) was launched in 2013 with a resolution of thirty meters (15 meters for panchromatic) and a revisiting time of 16 days. It tries to keep the Landsat earth observation program's data stable (Mandanici & bitchli, 2016). Using a 185-kilometer sweep area, Landsat-8 operates as a push-broom satellite sensor. From the near-invisible to the shortwave infrared, there are nine bands on the operational land imager (OLI). One band is consistent with previous Landsat sensors, but the other two are brand new. Infrared cirrus band and deep blue seaside/airborne band are the two. Landsat-8 also has a thermal infrared sensor (TIRS). Temperature measurements are made using infrared bands 10 and 11. For the first time ever because of Landsat's addition of two thermal bands, a wide range of thermal applications have opened. For each red, blue, green, and wavelength, and the blue-green ratio, OLI RRS values were comparable with in situ values for the OLI sensor, demonstrating OLI's suitability for surface chlorophyll-a monitoring in nearshore coastal waters of Santa Monica Bay. Between the OLI and in situ Rrs values, an $R^2$ value of 0.78 was found, indicating a strong correlation between the two data sets (Trinh *et al.*, 2017).

Europe's Space Agency (ESA) runs Sentinel-2 (ESA). As part of the Copernicus program, earth observation satellites were to be used to monitor agriculture and forestry (Drusch *et al.*, 2012Sentinel-2A and Sentinel-2B were launched into orbit in 2015 and 2017, respectively. The sensors can only have five-days intervals between visits. The spectral resolutions of Landsat-8 and Sentinel-2 vary greatly. The spectral resolutions of Landsat-8 and Sentinel-2 vary greatly. It also scans the Earth at the same time as Landsat-8 so that historical comparisons can be made. Landsat-8 records information using only eight bands, but Sentinel-2 has thirteen. Although these bands have the same names as their Landsat-8 counterparts, they do not include thermal infrared bands. Using two identical multispectral sensors (MS) on two separate satellites (Sentinels 2A and 2B), the Sentinel-2 mission can collect data in thirteen bands with varying geographical resolutions (between 10m and 60m). It is designed to keep the SPOT missions running smoothly (Mandanici & bitchli, 2016). The ability to map and analyse water quality is one of the advantages of using high spatial resolution sensors (10-60 m) for purposes of monitoring coastal and inland waters (Vanhellemont & Ruddick, 2015). Sentinel-2 MSI data were utilized to map a variety of lake water quality metrics. Based on Sentinel-2 Level-1C and Level-2A imagery, in situ chlorophyll, colored dissolved organic matter (CDOM), and dissolved organic carbon (DOC) data were gathered in nine lakes (Toming *et al.*, 2016). For example, Chl a ($R^2 = 0.83$), CDOM (0.72), and DOC (0.92), as well as paintwork (0.52), were shown to be strongly linked to band ratio algorithms estimated from Sentinel-2 MSI findings ($R^2 = 0.92$).

When using Landsat 7 and 8 and Sentinel-2, there are several ways to determine a chlorophyll-a concentration of water bodies. Methods such as these use a ratio of the red and infrared bands, as

well as a ratio of the green and blue bands (Allan *et al.,* 2007; Turner, 2010). As a result of the study of Vaal dam's chlorophyll-a concentrations and water turbidity, Landsat-8 and Sentinel-2 have proven to be especially useful for mapping chlorophyll-a concentrations, as high R-squared values were obtained. Sentinel-2 and Landsat-8 are therefore excellent tools for monitoring water quality (Bande *et al.,* 2018). With an accuracy of 0.98 discovered for this study using Landsat-8 over Olushanga Dam, Landsat's estimation ability is clearly outstanding (Kapalanga, 2015). Landsat-8 bands 2 and 5 were also found to be correlated with suspended sediment concentrations through regression and algorithms were developed for the calculation of these bands. More ocean and coastal water can be monitored with these two sensors because of their improved radiometric quality.

### 2.5 Remote sensing's challenges and benefits for water quality

In the past, government agencies have been unable to collect sufficient samples for long-term water quality monitoring because it was too cumbersome and expensive (Schaeffer et al., 2013). Whether you have an ambitious sampling plan, you may only be able to sample a small portion of a water body's total area. In many cases, measurements are taken in a small area, so the results may not be accurate. One-day samples cannot accurately represent the water quality over a week, month, or season. Since the sampling is too small to accurately measure changes in water quality, it is used instead (Schaeffer et al., 2013). Using satellite remote sensing, one can assess the long-term viability of watershed habitats and their services under current and future land-use practices. It is best to monitor the environment on a large scale to ensure that management activities are efficient. Satellite technology makes it possible to monitor water quality on a local and global level.

Using direct solar radiation, which is absorbed or scattered by the water column's constituents, it is possible to calculate water quality parameters by satellite. Water-leaving radiance is reflected into the atmosphere by the radiation. To calculate reflectance, one divides the amount of reflected radiation (W m-2) by the amount of direct sunlight absorbed by the sea surface. To derive the ocean color signature, remote sensing reflectance is used to incorporate the spectral absorption and backscattering properties of all materials present in the water column (Coble et al., 2004). These physical descriptions are inaccessible to managers and policymakers who lack specialized knowledge.

### 2.6 Regression models for Water Quality Assessment

Statistical regression methods consist of two techniques - parametric methods, which show a direct relationship between water quality measures including chlorophyll-a and band ratios/indices, and non-parametric regression, which describes regression models generated from training data (Mao *et al.,* 2019). Machine-learning regression methods are nonparametric non-linear statistical methods. Using regression analysis, researchers have been able to show connections between sensor spectral reflectance and field water quality metrics (Mushtaq & Lala, 2017). Choosing a regression method and independent variables that provide an $R^2$ value between 0 and 1 is the key to performing an efficient regression study (Mushtaq & Lala, 2001). To deal with the problem of nonlinearity in water quality results, some studies have investigated the prediction of water quality

parameters using machine learning models (Bande *et al.,* 2018). Multispectral data at medium or fine resolution can be used by more complex algorithms to monitor water quality at a smaller scale, according to (Dube *et al.* 2015). Only a few studies have utilized remote sensing data to estimate coastal or ocean water quality using machine learning algorithms such as random forest and support vector regression (Kim *et al.,* 2014). The ability of SVR to overcome issues of high dimensional data and limited samples, makes it have more benefits than other machine learning approaches (Banadkooki *et al.,* 2020). Water quality field measurements are tedious, labour-intensive, and expensive, so samples are generally small. Consequently, regression methods with high precision with limited samples are preferred

# CHAPTER 3
# MATERIALS AND METHODS

## 3.1 Study Area

The Vaalkop Dam is located fifty-five kilometers from Brits in the Province of North-West, within Bojanala Platinum District Municipality. The study between the Hex River and Elands River in the Bushveld Basin Aquatic Ecoregion which is also part of the Crocodile basin (Swanelpoel *et al.,* 2017). The Bojanala Platinum district municipality has the most platinum group metal reserves in the country which makes up 80% of the world's reserves (Steyn, 2006). Vaalkop Dam is within the nature reserve and has a surface area of about 110 ha and 11.4 M deepness at the dam wall **(**Swanelpoel *et al.,* 2017**).** When Vaalkop Dam was built in 1972 the main purpose was to serve municipal water supply through Magalies Water treatment works, industrial uses, and irrigation, it was then expanded in 2018 to meet supply water demand for the mining operations within the area (Swanelpoel *et al.,* 2017). The dam is an appealing site for boating, fishing, bird watching, and angling. Along the Eastern side of the dam, there is a private residential area developing into a small estate called Bushwillows game farm, 70% of the property owners used the area as holiday houses and 30% are retired couples. The North-West Parks Bird Sanctuary, an 800-hectare section of the reserve, also lies along this leg of the dam. The district's land area focuses on commercial–land farming, commercially irrigated, and subsistence dry-land farming. Along the Vaalkop dam the mixes-crop farming such as maize, sunflowers, and lucern is dominant and herds of cattle that belongs to the village dwellers (Ellison, 1990). The area received an average of 300 to 700 millimeters of rain per year. In the summer, temperatures range from 22 to 34 degrees Celsius, and in the winter, dry, sunny days are interspersed with frigid nights. A day's temperature might fluctuate from 2 to 20 degrees Celsius on average in the winter (May to July). According to a species diversity report (Swanelpoel *et al.,* 2017**),** the Hex River part of the dam has a concerning poor water quality. The inflow of synthetically enhanced water through the canal from Roodekopjes Dam additionally adds to poor water quality in the Vaalkop Dam.

The Bojanala Platinum District, according to the 2016 community census, is marked by significant population expansion, with a present population of approximately 1.7 million people (StatsSA, 2016). The Vaalkop Dam is crucial for balancing and storing water for the basin's population, as well as supporting farmers and nature reserves. The dam also provides water to the municipality of Thabazimbi and mines in Limpopo. Assessing the quality of the dam's water, as a result, is critical to ensuring that it does not further deteriorate and to minimising the adverse effect on

human, animal, and environmental health. Farmers that rely on the dam water will be unable to produce and supply essential food products to the market.

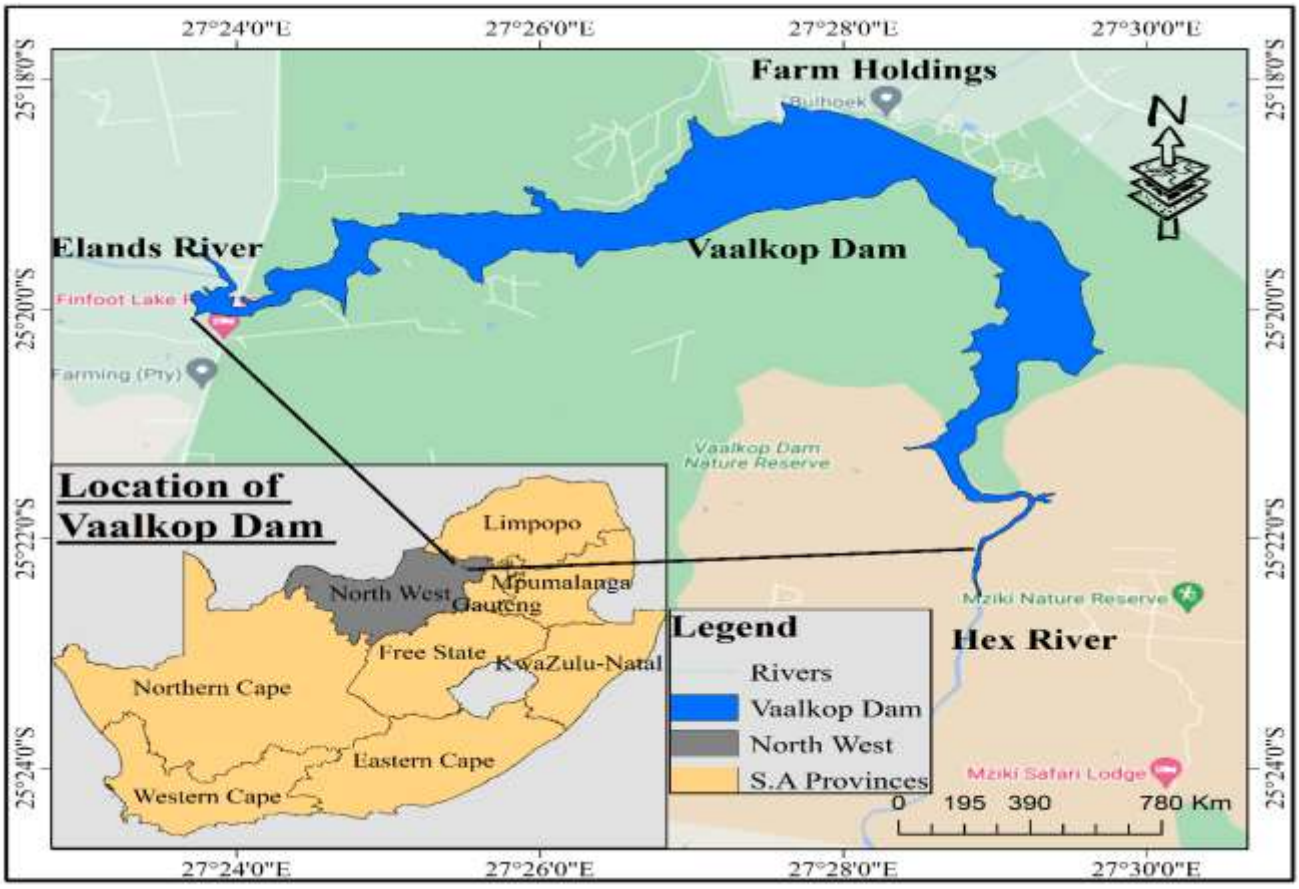



**Figure 1:** The research area is shown on a map of the Bojanala district in the North-West Province (Generated using ArcGIS 10.6.1 version).

### 3.2 Methodology

#### 3.1.1 Sampling Data Acquisition

On the 28th of October 2020, just after the dry season, 108 samples were collected in 27 different locations at Vaalkop Dam, and a Global positioning system was utilised to capture locations from each sample site. Samples were gathered 300 mm below the water's surface using a 5-liter bucket tied with a rope which were then placed in high-density sterile bottles and refrigerated to 3°C to 5°C. Ground information was collected from 27 sample sites in various parts of the Vaalkop Dam. The sample areas in the dam were carefully selected to consider the effects of the major rivers flowing into the dam from the surrounding areas. Locations have been chosen based on farming abstraction locations and water hyacinth population increase. Sample locations in the canal were at the dam's entry point. An ISO-certified Magalies Water laboratory determined physicochemical

characteristics for the purpose of investigating possible relationships between dam water quality and water characteristics. Physio-chemical water quality parameters such as chlorophyll-a, turbidity, colour, hardness, electrical conductivity, pH, E-coli, mercury, nitrate, lead, and ammonia were tested in these samples. However, only chlorophyll-a and turbidity were of interest in this research due to constraints budget. Figure 2 shows sampling point locations.

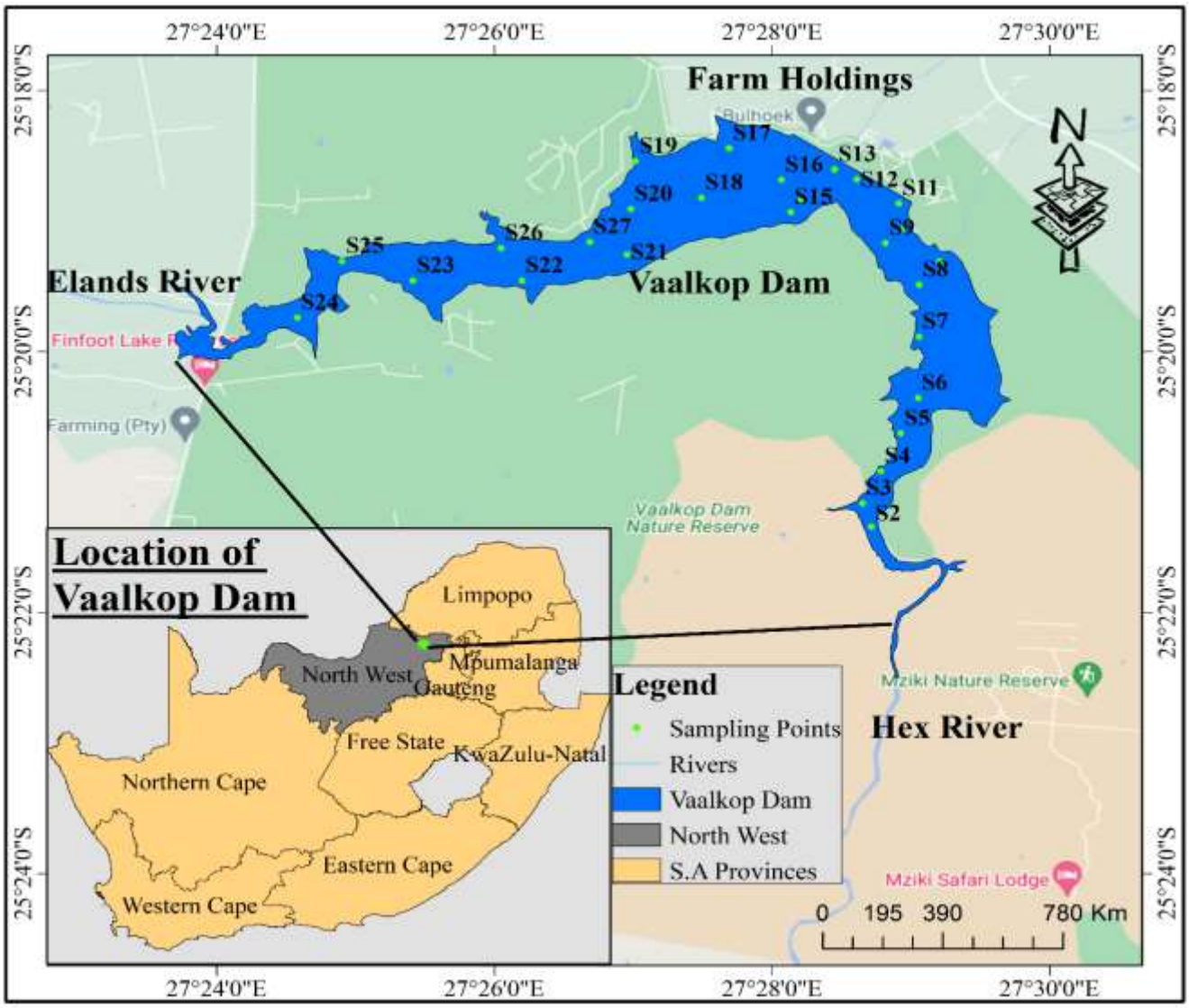



**Figure 2:** A map of the dam's sample points 1–27 is shown.

### 3.1.2 Remote Sensing Data Acquisition and Pre-processing

Regarding its synoptic coverage and high spatial resolution, remote sensing is perfectly suited to fulfil the standards of water quality monitoring (Ghodizadah et al., 2016). Data from the Vaalkop Dam's water quality was collected using Landsat-8 and Sentinel-2, two multi-temporal remote sensors. Landsat-8 and Sentinel-2 multi-temporal images were taken on October 16th, 2020

(sampling month). Sentinel-2 and Landsat-8 were used to assess and map the dam's chlorophyll-a and turbidity concentration trends

### 3.1.3 Pre-processing and atmospheric correction for Sentinel-2 data

Imagery from Sentinel-2's study area 171077 was obtained from the Open Access Hub – Copernicus website at https://scihub.copernicus.eu on October 16th, 2020. Using Sen2cor v.02.08.00 of the Sentinel-2 Toolbox (S2TBX) together with the Sentinel Application Platform (SNAP) v8.0, a bottom-of-the-atmosphere reflectance image was generated, which was utilised for additional visualisation and interpretation of band values (see Figure 3 and Table 1 below). visualisation and interpretation of band values (see Figure 3 and Table 1 below).

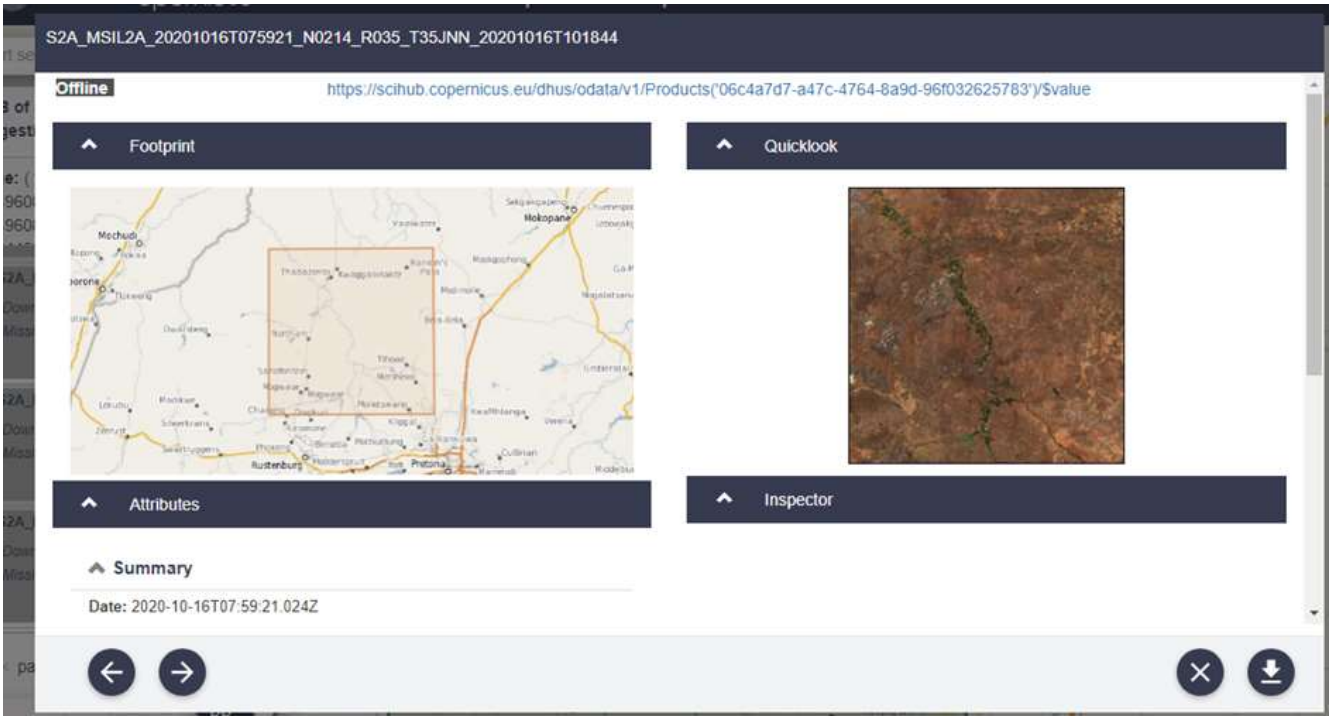



**Figure 3:** Sentinel-2 image

**Table 1** For Sentinel-2's remote sensor specifications (adapted from Roy *et al.*, 2014 and Drusch *et al.*, 2012)

| Band number | Central wavelength (um) | Bandwidth (nm) | Spatial resolution (m) |
|---|---|---|---|
| 1. Coastal aerosol | 0.443 | 20 | 60 |
| 2. Blue | 0.490 | 65 | 10 |
| 3. Green | 0.560 | 35 | 10 |
| 4. Red | 0.665 | 30 | 10 |

| | | | |
|---|---|---|---|
| 5. Vegetation Red Edge | 0.705 | 15 | 20 |
| 6. Vegetation Red Edge | 0.740 | 15 | 20 |
| 7. Vegetation Red Edge | 0.783 | 20 | 20 |
| 8a NIR | 0.842 | 115 | 10 |
| 8b Vegetation Red Edge | 0.865 | 20 | 20 |
| 9. Water Vapour | 0.945 | 20 | 60 |
| 10. SWIR-Cirrus | 1.375 | 20 | 60 |
| 11. SWIR | 1.610 | 90 | 20 |
| 12. SWIR | 2.190 | 180 | 20 |

(Adapted from Roy et al., 2014 and Drusch *et al.,* 2012)

### 3.1.4 Landsat-8 Data pre-processing and atmospheric correction

On October 22, 2020, Landsat-8 OLI operational land imagery was retrieved from https://earthexplorer.usgs.gov (Table 3). Landsat-8 is made from the operational land imager (OLI) and thermal infrared sensors (TIRS). The OLI sensor detects the entire earth using nine spectral information in the visible near-infrared (VNIR) and shortwave infrared (SWIR) areas. Radiometric atmospheric corrections were performed to satellite data to eliminate aerosol, water vapor, and clouds, and digital numbers (DN) have been translated to top of atmosphere (TOA) reflectance through radiometric calibration for accurate bandwidths for every band. Two new deep blue spectral bands detect aerosols, clouds, and haze, namely bands 1 and 9. Using FLAASH model in ENVI software Landsat-8 image was atmospherically corrected on the radiance image to convert it to reflectance values, Top of Atmosphere reflectance. After the conversion to radiance images, ArcGIS version 10.8 was used to extract respective band values from both resultant reflectance of Landsat-8 OLI image and Sentinel-2C image for each study area site point (Roy *et al.,* 2014) see Figure 4 and Table 2.

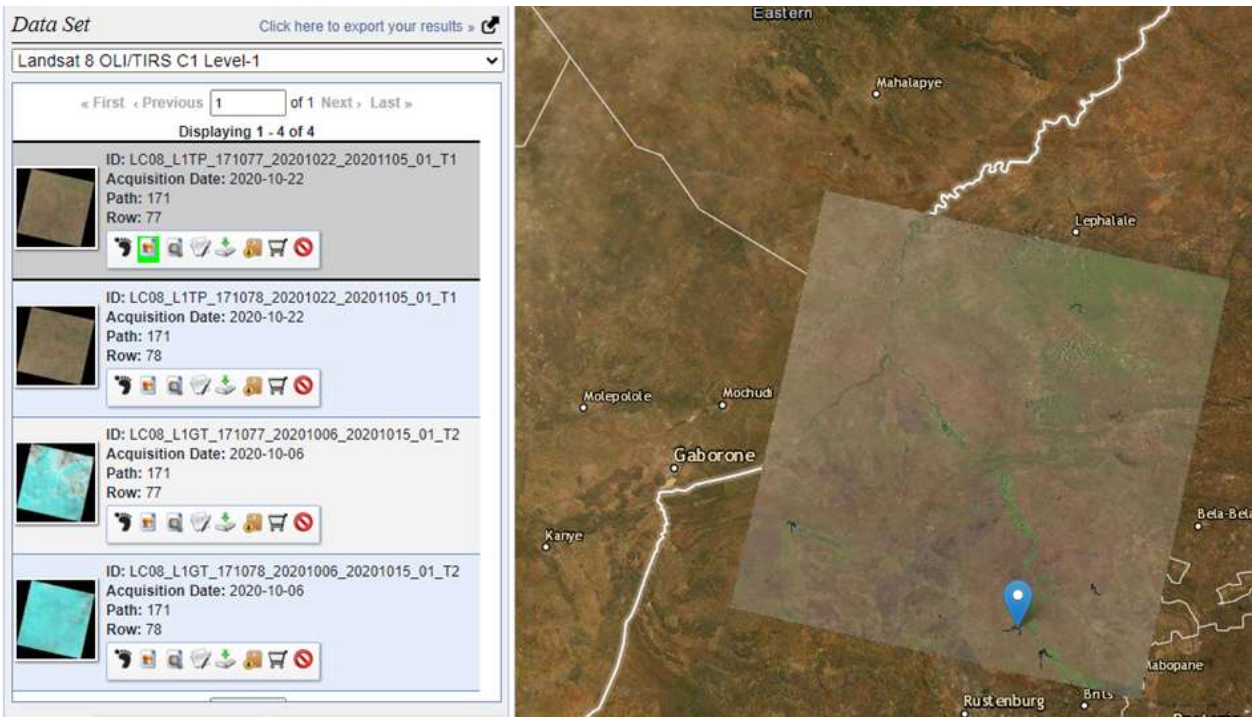



**Figure 4:** Landsat-8 OLI image (https://scihub.copernicus.eu/dhus/odata/v1)

**Table 2**: Spectral properties for Sentinel -2 vs Landsat-8

| Comparison specification | Sentinel-2 | Landsat 8 |
|---|---|---|
| Satellite sensor system | MSI | OLI/TIRS |
| Spatial resolution range | 10-60m | 15-100m |
| Resolution in radiometric | 12 bits per pixel(wide) | 12 bits per pixel (wide) |
| Resolution of time (Temporal) | 5 days | 16 days |
| Altitude of orbit | 786 km | 705km |

**Table 3**: Specifications for Landsat-8 remote sensor *(Adapted from Roy et al., 2014 and Drusch et al., 2012)*

| Band number | Sensor type | Wavelength (µm) | Spatial resolution (m) |
|---|---|---|---|
| Band 1-Coastal aerosol | OLI | 0.443 | 30 |
| Band 2 -Blue | OLI | 0.482 | 30 |
| Band 3-Green | OLI | 0.562 | 30 |
| Band 4-Red | OLI | 0.665 | 30 |
| Band 5-Near Infrared (NIR) | OLI | 0.865 | 30 |
| Band 6-SWIR 1 | OLI | 1.610 | 30 |
| Band 7-SWIR 2 | OLI | 2.200 | 30 |
| Band 8-Panchromatic | OLI | 0.590 | 15 |
| Band 9-Cirrus | OLI | 1.375 | 30 |
| Band 10-Thermal Infrared (TIRS) | TIRS 1 | 10.900 | 100 |
| Band 11-Thermal Infrared (TIRS) | TIRS 2 | 12.000 | 100 |

*(Adapted from Roy et al., 2014 and Drusch et al., 2012)*

### 3.1.5 Description of water quality parameters

The physical, chemical, and biological composition of watercourses can be determined, as well as any possible pollution or contamination sources that could contribute to a decrease in water quality.

#### 3.2.5.1 Turbidity

Too much turbidity is likely to cause Dissolved Oxygen (DO) stratification and temperature increase in waterways (Tessema *et al.,* 2014). High Total suspended solids levels in surface water absorb heat from the sun, raising the temperature of water and decreasing dissolved oxygen levels (Iqbal et al., 2010). Turbidity measurements are mostly used to calculate fluvial suspended sediment concentrations (Wass *et al.,* 1997). The red band of the electromagnetic spectrum has a significant impact on turbidity measured in situ and reflectance measured from satellite images (Lathrop & Lillesand, 1986). Liversedge conducted simpler and various linear regression analyses using different Landsat 7 ETM-F bands to discover the best predictors of turbidity in marginal ice lakes at the Bering Glacier in Alaska (2007). Turbidity levels were predicted using the Landsat 7 ETM+ Band 3 (red portion of the electromagnetic spectrum) and Band 4 (near-infrared portion of the electromagnetic spectrum) algorithms. The algorithm was utilised for this study to create turbidity maps to aid in determining inter- and intra-annual sediment dynamics. Turbidity maps can also show where sediments are being discharged during floods.

#### 3.2.5.2 Chlorophyll-a

The quantity of dissolved oxygen in the water is heavily influenced by algae in the water. During the day, algae take in carbon dioxide and release oxygen through photosynthesis, therefore the quantity of dissolved oxygen in the water increases (Shock *et al.,* 2003). The occurrence of chlorophyll-a in the water surface is associated with the presence of algal bloom, which also

illustrates the amount of eutrophication in the water surface (Li & Li, 2004). The amount of chlorophyll in the algal plankton cells in various freshwater bodies is used to determine eutrophication. Chlorophyll-a is a photosynthetic pigment that contributes to the color of the water. Many studies have employed remote sensing to map chlorophyll-a, a fundamental measure in water quality evaluation and an indication of algal population (Schalles *et al.,* 2008). Eutrophication causes a wide variety of chlorophyll-a concentrations in most inland waters (lakes, rivers, reservoirs, and dams). Understanding the amount of phytoplankton in these water bodies is critical to determining the extent of eutrophication. During cyanobacterial blooms, satellite sensors can have trouble in detecting chlorophyll because of the short cycle involved in the biological activity, which gets more difficult to detect as the temporal resolution increases. Although this drawback, remote sensing has been effective in quantifying chlorophyll-a and algal blooms, especially with the use of satellites. To map chlorophyll-a in the seas, estuaries, and freshwater, scientists use a range of algorithms and wavelengths from aircraft and satellite sensors including Landsat, Sentinel-2, and MODIS. Band rationing and regression are effective methods for estimating chlorophyll-a pigments in water bodies (Dekker *et al.,*2002).

### 3.1.6 Regression Modelling

SVMs are divided into two types: SVMs for classification and SVMs for regression (Gupta *et al.,* 2015). Support vector regression (SVR) is a probability model that uses the structural risk minimisation principle to minimise a generalisation error bound (Li *et al.,* 2016 To predict an independent variable, SVR aims to develop ideal hyperplanes which isolate groups of input data with equal feature attributes (Kim *et al.,* 2014). SVR has many advantages, including the ability to solve learning problems with limited sample sizes, nonlinearity, or high dimensions (Wang *et al.,* 2011). The SVR equation is as follows:

$$f_{(x)} = wTx + b\ (w, x \in Rd) \qquad \textit{Equation 1}$$

x ∈ Rd refers to any non-linear function that takes input data into a higher dimensional space (Xiao *et al.,* 2018). The model's unknown parameters are w, a regular to the hyper-plane weight vector, and b, the hyper-plane bias (Rodriguez-Galiano *et al.,* 2015). (Xiao *et al.,* 2018).) The penalisation factor (C), the ε (epsilon) insensitive region, and the kernel parameter σ (sigma) is the three hyper-parameters that must be tuned (Wang et al., 2011). Kernel functions that specify the form, magnitude, and precision of the SVR model are commonly polynomial, linear, sigmoid, Gaussian, radial basis functions and spectral angle (Kim *et al.,* 2014). Provided the trial-and-error method involved in determining the right parameters, the correct ones are harder to achieve, making it difficult to distinguish between the best and worst models (Wang *et al.,* 2011). "The linear kernel is a special case of the radial basis function, and the radial basis function can better accommodate nonlinear input-output relationships" (Li *et al.,* 2016, p73). Furthermore, the radial basis function kernel poses fewer computational challenges than the polynomial kernel, which contains more hyperparameters than the radial basis function kernel (Li *et al.,* 2016). One of the most used kernel functions is the radial basis function.

SVR were chosen because they had already been applied in other water quality projects and provided excellent accuracy and performance in comparing remote sensing data and field datasets.

Wang et al. (2011) are among the first to apply support vector regression to the remote sensing of inland water quality. Permanganate index, ammonia-nitrogen, and chemical oxygen demand were all calculated using SVR. With small samples, the results showed that SVR provided better predictions than the traditional multiple regression process. (Zhou *et al.,* 2016). Kim et al. (2014) compared SVR, RFR, and Cubist regression to measure chlorophyll-a and suspension particle matter in coastal using the Geostationary Ocean Colour Imager. SVR surpassed the other two machine learning algorithms, according to Kim et al. (2014), with $R^2$ values of 0.91 for chlorophyll-a and 0.98 for particulate suspended matter. Ruescas et al. (2018) used simulated reflectance data from Sentinel-2 and Sentinel-3 to compare five regression models, including RFR and SVR, for retrieving colored dissolved organic matter. Although all the regression models worked well, the SVR model outperformed the others.

In this study, the Support Vector Machine was used to determine the regression model that best retrieves each of the water quality parameters. Vector regression was performed using the R packages "caret" and "e1070" (Forkuor *et al.,* 2017). The regression models were built using bands and band ratios from field hyperspectral and satellite data that had a close relationship to chlorophyll-a and turbidity. A training data set (70 percent) and a test (validation) dataset were created from the data (30 percent).

### 3.1.7 Accuracy Assessment

The R-squared and root mean square error (RMSE) values were acquired by matching Sentinel-2 and Landsat-8 data with field measurements (Hung & Lin, 2006). The acquired R-squared values were then utilised to evaluate the accuracy and efficiency of the regression models. The expected chlorophyll-a and turbidity concentration patterns were then mapped across the dam. The average degree of error is represented by the RMSE. It has a range of values ranging from 0 to 1. The RMSE indicates a strong model match when it is smaller. The RMSE is a helpful metric for measuring how well a model predicts a variable given observed values (Khattab & Merkel, 2014; Coskun et al., 2008). The following is the equation used to calculate the RMSE:

$$\text{RMSE} = \sqrt{(\sum WQpred, i - WQmeas, i\ N\ i{=}1\ N)} \qquad \text{Equation 2}$$

WQpred i, is the anticipated chlorophyll-a or turbidity, and WQmeas i, is the measured chlorophyll-a or turbidity in Equation 2, where N represents the size of samples, WQpred is the predicted chlorophyll-a or turbidity, and WQmeas is the measured chlorophyll-a or turbidity (Sakuno *et al.,* 2018). The R-squared value ($R^2$), which ranges from 0 to 1, indicates whether a model is well-fitting and assists in determining how close datasets are to a linear regression. The $R^2$ number denotes how well the model predicts the outcomes (Coskun *et al.,* 2008).

# CHAPTER 4
# RESEARCH FINDINGS

## 4.1 Descriptive analysis of the ground measured Chlorophyll-a and Turbidity

Figure 5 below shows derived density distribution plots of ground measured Turbidity and Chlorophyll-a values. Measured Chlorophyll-a values of Vaalkop Dam peak around 70-80 µg/L while those of turbidity peak around 30-35 (NTU) values. Interpretation of the median and quartiles positions using the box and whisker plot in Figure 6 below suggest that chlorophyll-a distribution is negatively skewed with a great variation between the sample points comparing to turbidity values which have small variation with a normally the distribution of values measured

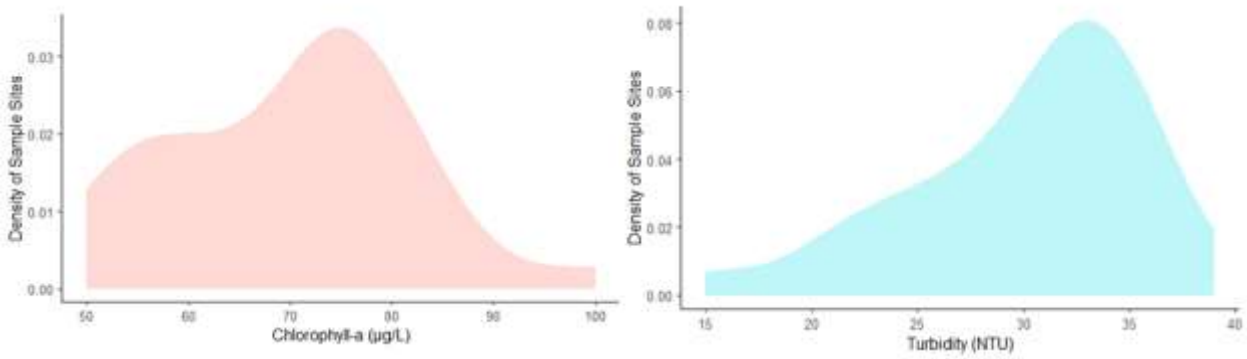



**Figure 5**: Density pattern of chlorophyll-a observed in the ground (on the left) and turbidity (on the right)

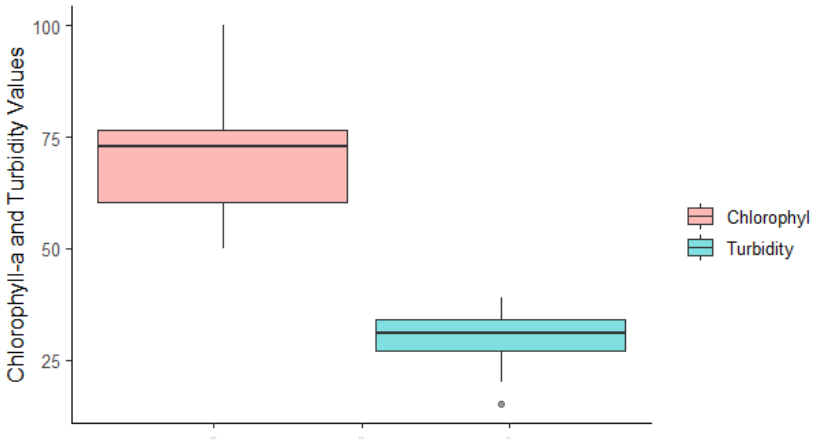



**Figure 6:** The variability of ground-measured chlorophyll-a and turbidity is shown in a box and whisker plot.

## 4.2 Spatial pattern of the ground measured Chlorophyll-a and Turbidity

This figure (7) illustrates the spatial pattern for parameters of water quality in the Vaalkop Dam as measured by ground turbidity and Chlorophyll-a. The dam's innermost section (Crocodile canal) is the most contaminated, with pollution decreasing as it approaches the Elands River canal. The higher concentration of chlorophyll-a and turbidity on the Crocodile canal might have been caused by the pollution by the farming activities which inputs of organic and inorganic sediments along the Crocodile River canal which also carries water from Hartbeespoort dam.

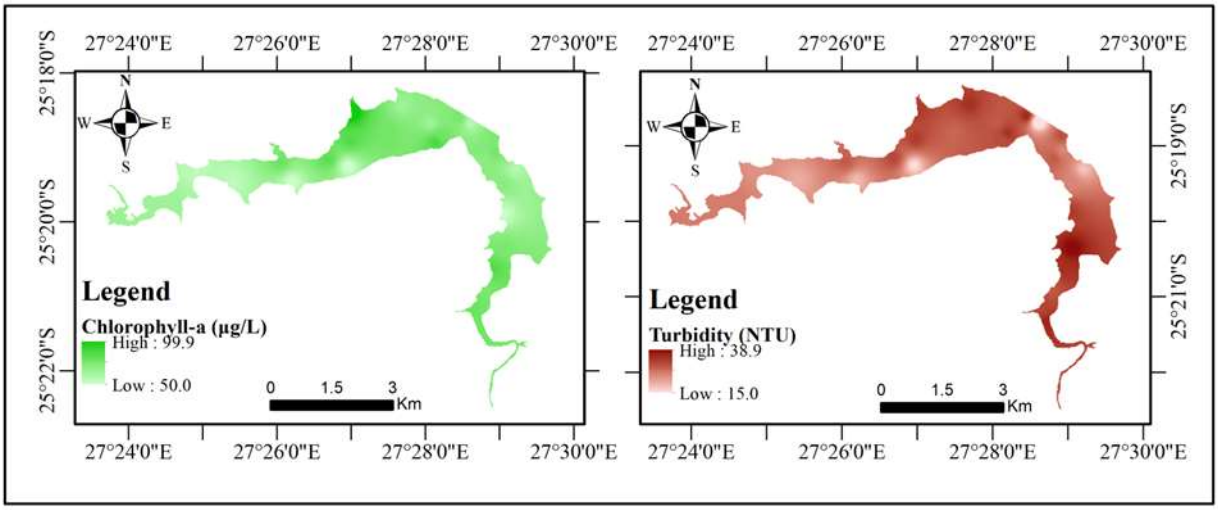



**Figure 7**: Spatial pattern of the ground measured Chlorophyll-a and Turbidity

## 4.3 Chlorophyll-a Results

### 4.3.1 Correlation results

For Landsat-8 and Sentinel-2, different band ratios were created, which were then correlated with Chlorophyll-a to determine the best band combinations that could be used for support vector machine predictions. The correlation plot in Figure 8 below illustrates the relationship between Sentinel-2 and Landsat-8 reflectance values acquired within one month of ground measurements and ground measured chlorophyll-a values. Bands that were found to have the most correlation coefficient value with Chlorophyll-a were Landsat-8 bands with Band 7 Short Wave Infrared (SWIR2) and Band 6 Short Wave Infrared (SWIR1) being the best-correlated bands. Sentinel-2 bands have shown a weak correlation with Chlorophyll-a values.

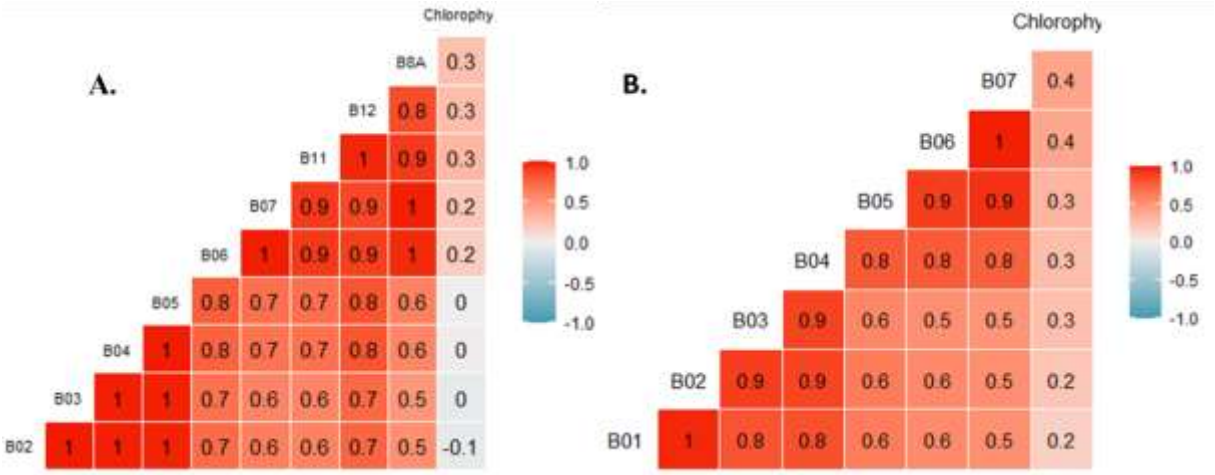



**Figure 8**: Chlorophyll-a correlation results with A. Sentinel-2 acquired on 16th October 2020 and B. Landsat-8 acquired on 22nd October 2020 respectively.

Although there is difference significance in correlation values but the pattern of the band correlation values for images acquired within a month of ground measurement is more like that of images acquired a month later after ground measurements (Figure 9). Images acquired a month after ground measurements have the lowest correlation coefficient. The changes that occur over a month are likely to be the main cause for the difference in correlation coefficients observed between images of these difference dates.

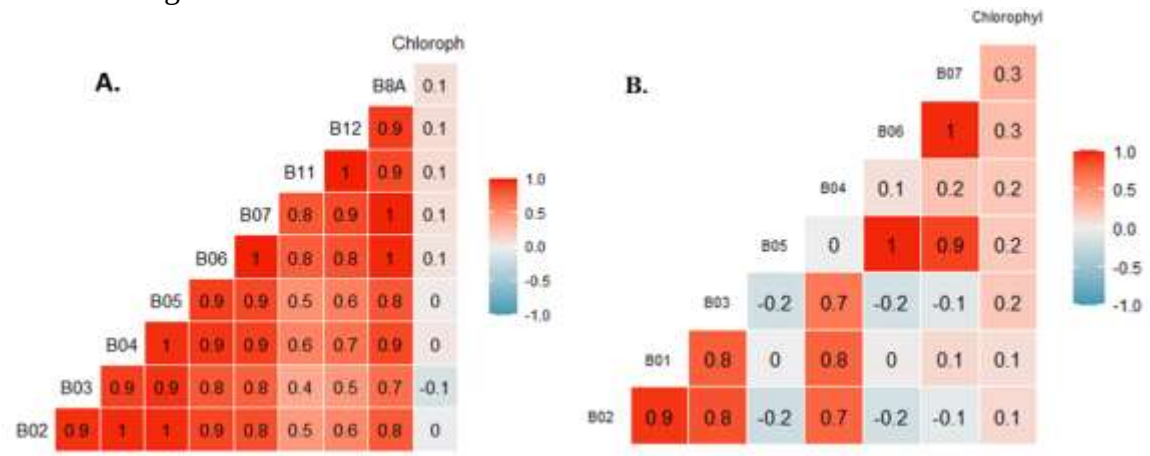



**Figure 9**: Chlorophyll-a correlation results with (A) Sentinel-2 acquired on 25th November 2020 and (B) Landsat-8 acquired on 07th November 2020 respectively

Table 5 below summarises the correlation analysis results of the best band ratio combinations. Just as in single-band correlation results in Figure 10 above, Landsat-8 band ratios correlated the most with chlorophyll-a than Sentinel-2 band ratios. Global Environmental Monitoring Index (GEMI) utilizes the red and near-infrared bands for Landsat-8 (Yang *et al.*, 2017). Landsat-8 bands 4, 5, and 2 did show the best correlation which agrees with results by (Bande, 2020) whereas Sentinel-2 band 2, 4, and 6 ratios did show the most correlation value.

**Table 4** Chlorophyll-a Landsat-8 acquired on 22nd October 2020 and Sentinel-2 *(Acquired on 16th October 2020 correlation)*

| Landsat 8 Chlorophyll-a | | Sentinel 2 Chlorophyll-a | |
|---|---|---|---|
| Band/Band ratio | Correlation coefficient | Band/Band ratio | Correlation coefficient |
| B04/ B05 | 0.553 | B02/ B04 | 0.496 |
| B03/ B07 | 0.541 | B06/B01 | 0.462 |
| B04/ B06 | 0.522 | B03/ B11 | 0.431 |
| B02/B04 | 0.518 | B05/ B11 | 0.402 |
| GEMI | 0.504 | B06/ B07 | 0.401 |
| B02/B07 | 0.502 | B04/ B12 | 0.392 |

Furthermore, the band combinations observed in table 4 for images acquired a month after measurements on the ground showed a weak correlation with the values derived from the ground measurements from chlorophyll-a (Table 6).

**Table 5** Chlorophyll-a Landsat-8 and Sentinel-2 Chlorophyll-a correlation

| Landsat 8 Chlorophyll-a | | Sentinel 2 Chlorophyll-a | |
|---|---|---|---|
| **Band/Band ratio** | **Correlation coefficient** | **Band/Band ratio** | **Correlation coefficient** |
| B04/ B05 | 0.391 | B02/ B04 | 0.263 |
| B03/ B07 | 0.385 | B06/B01 | 0.241 |
| B04/ B06 | 0.344 | B03/ B11 | 0.208 |
| B02/B04 | 0.294 | B05/ B11 | 0.187 |
| GEMI | 0.256 | B06/ B07 | 0.175 |
| B02/B07 | 0.201 | B04/ B12 | 0.167 |

*Sentinel-2 acquired on 25th November 2020 and B. Landsat-8 acquired on 07th November 2020 respectively*

These results exhibit that images acquired within a month of ground measurements they correlate the best with ground measured chlorophyll-a values thus they were used for support vector machine regression.

### 4.3.2 Support vector regression results

Table 6 below shows the SVR prediction results for chlorophyll-a. Sentinel-2 achieved the highest $R^2$ value of 0.70 while Landsat-8 achieved $R^2$ of 0.58 for chlorophyll-a predictions. Sentinel-2 had the lowest RMSE value of 7.43ug/L when compared to Landsat-8 forecasts, which had an RMSE of 7.12ug/L, showing that Sentinel-2 had greater prediction accuracy than Landsat-8. The higher precision for Sentinel-2 in prediction over Landsat-8 can be caused by the higher resolution (20m) of the product comparing to 30m resolution of Landsat-8 products. An extra red bands and near-infrared bands also added an advantage to Sentinel-2 products predictions.

**Table 6:** SVR chlorophyll-a results

| RS Imagery | Training Set (70%) | | Validating Set (30%) | |
|---|---|---|---|---|
| | $R^2$ | **RMSE (**ug/L) | $R^2$ | **RMSE (**ug/L) |
| **Sentinel-2** | 0.70 | 6.28 | 0.681 | 7.43 |
| **Landsat-8** | 0.58 | 8.36 | 0.594 | 7.12 |

The link between band ratio combinations as independent factors and ground measured chlorophyll-a as dependent variables was presented in figure 5 and 6 below, with the blue line indicating the SVR linear relation between these two variables. From these plots, the validation results for Sentinel-2 were better with $R^2$ of 0.681 than validation for Landsat-8 which produces an $R^2$ of 0.594 which is again down to improved spatial and spectral resolution of Sentinel-2 comparing Landsat-8 products. These results are congruent with Wang et al. (2016), whose findings explained that SVR performs the best in prediction with a small sample size than convolutional statistical regression models (Figure 10 & Figure 11). The scattered results in Figure 11 could be a consequence of the other suspended solids, which can influence R in a relatively large selection of wavelengths (Barbieux *et al.*, 2018).

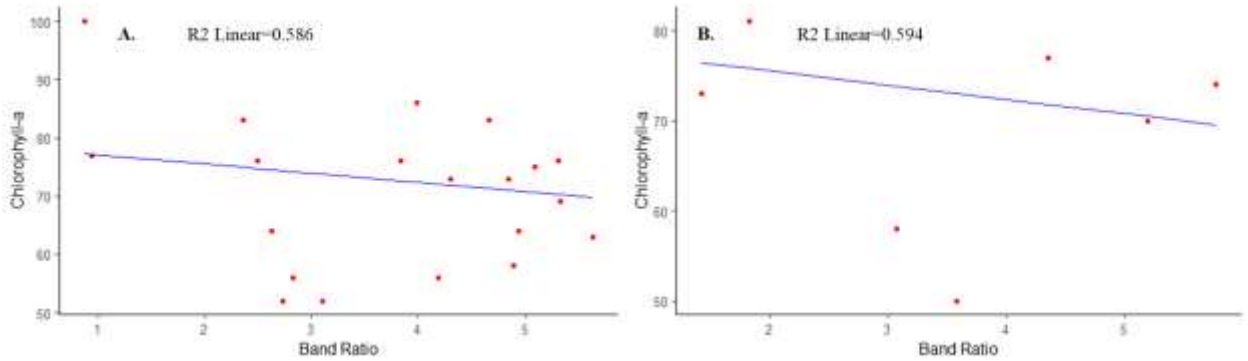



**Figure 10:** Landsat-8 plots of showing relationships between actual and predicted chlorophyll-a for (a) training set (n=20), (b) validation set (n=8)

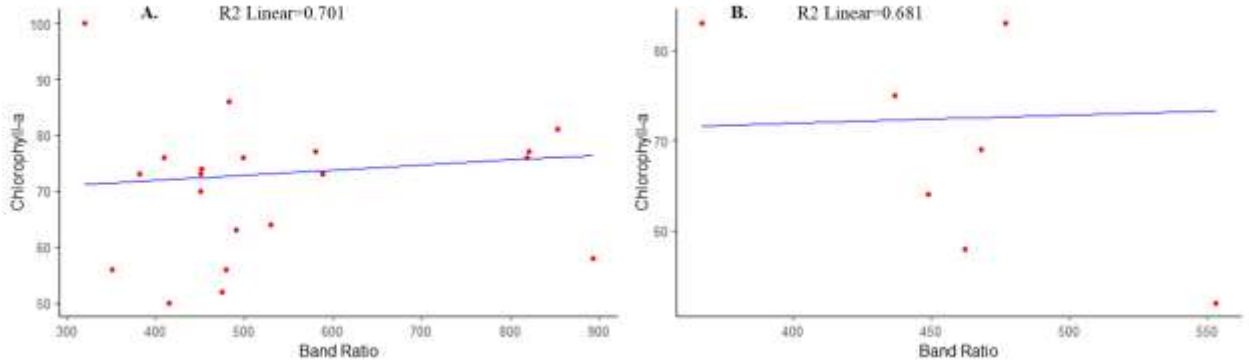



**Figure 11:** Sentinel-2 plots of showing relationships between actual and predicted chlorophyll-a for (a) training set (n=20), (b) validation set (n=8)

## 4.4 Turbidity Results

### 4.4.1 Correlation Results

Just as in the case of chlorophyll-a, bands and band ratios were correlated with turbidity to find the best-correlated band combinations to be used in support vector machine predictions. Figure 4 and 5 below summarizes the results of turbidity correlations with Sentinel-2 and Landsat-8 bands respectively acquired in different dates of study period. Results have shown that Landsat-8 band 03 (Green band) and band 04 (Red band) and Sentinel-2 band 11 (SWIR band) have the best correlation values with ground measured turbidity values (Figure 12 and Figure 13).

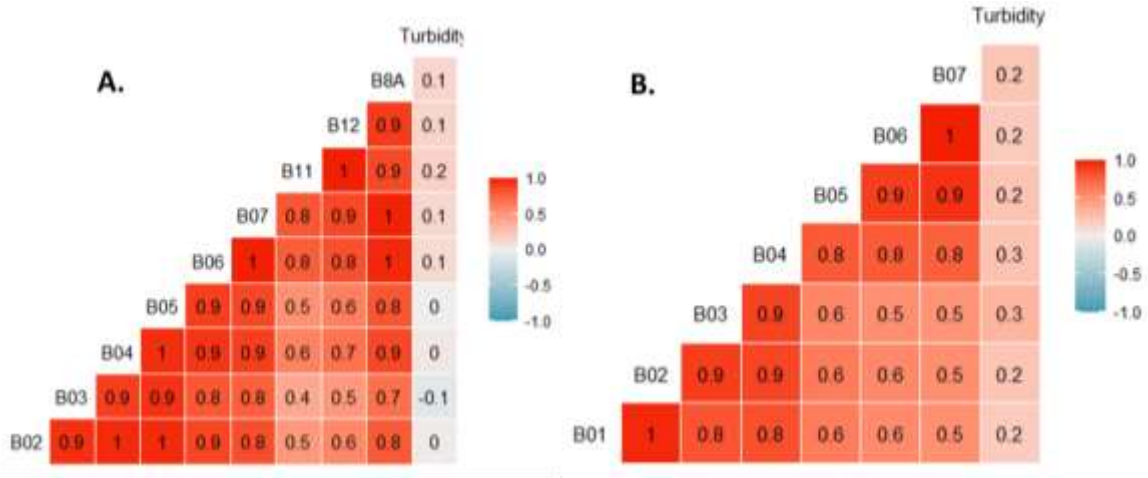



**Figure 12:** Turbidity correlation results with (A) Sentinel-2 acquired on 16th October 2020 and (B) Landsat-8 acquired on 22nd October 2020 respectively

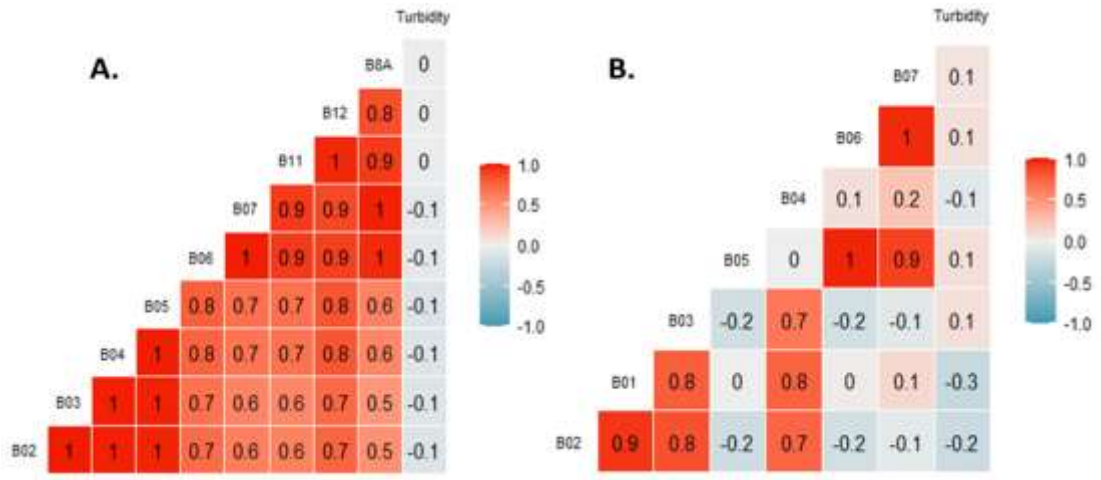



**Figure 13** Turbidity correlation results with A. Sentinel-2 acquired on 25th November 2020 and B. Landsat-8 acquired on 07th November 2020 respectively

Results summarised on table 3 below shows that Landsat-8 bands 01, 04, 03 and 07 ratios are the best band combinations to be carried and used in support vector machine regression predictions whilst for the case of Sentinel-2 bands 03, 02 and 01 ratios show the best correlation coefficient values. Moreover, Table 7 and Table 8 below summarises the correlation results for the turbidity and Landsat-8 and Sentinel-2 images acquired a month later after ground measurements of water quality data, which is still having the lowest correlation due to temporal changes over month period difference.

**Table 7** Turbidity Landsat-8 acquired on 22nd October 2020 and Sentinel-2 acquired on 16th October 2020 correlation

| **Landsat-8 Turbidity** | | **Sentinel-2 Turbidity** | |
|---|---|---|---|
| **Band/Band ratio** | **Correlation coefficient** | **Band/Band ratio** | **Correlation coefficient** |
| B01/ B04 | 0.482 | B03/ B02 | 0.458 |
| B02/ B07 | 0.474 | B02/B01 | 0.443 |
| B03 B04 | 0.431 | B04/ B02 | 0.416 |
| B02/B03 | 0.412 | B03/ B04 | 0.408 |
| B02/B05 | 0.402 | B05/ B12 | 0.394 |

**Table 8** Turbidity Landsat-8 acquired on 07th November 2020 and Sentinel-2 correlation acquired on 25th November 2020

| **Landsat-8 Turbidity** | | **Sentinel-2 Turbidity** | |
|---|---|---|---|
| **Band/Band ratio** | **Correlation coefficient** | **Band/Band ratio** | **Correlation coefficient** |
| B01/ B04 | 0.184 | B03/ B02 | 0.271 |
| B02/ B07 | 0.172 | B02/B01 | 0.229 |
| B03/ B04 | 0.149 | B04/ B02 | 0.204 |
| B02/ B03 | 0.109 | B03/ B04 | 0.145 |
| B02/ B05 | 0.106 | B05/ B12 | 0.104 |

### 4.4.2 Support vector regression results

SVR predictions for turbidity may be seen in Table 9 of this report. The $R^2$ values of 0.593 and 0.581 for Sentinel-2 and Landsat-8 training set predictions are comparable. The RMSE results for Sentinel-2 and Landsat-8 were 8.28NTU and 8.36NTU respectively.

**Table 9** SVR turbidity results

| | **Training Set (70%)** | | **Validating Set (30%)** | |
|---|---|---|---|---|
| | $R^2$ | **RMSE (NTU)** | $R^2$ | **RMSE (NTU)** |
| **Sentinel-2** | 0.593 | 8.28 | 0.525 | 7.32 |
| **Landsat-8** | 0.581 | 8.36 | 0.601 | 7.43 |

On the other hand, the validation results on figure plots figure 9 and figure 10 below show that Landsat-8 validation results have better accuracy than that of Sentinel-2. The findings of the study (Bande, 2020) that validates the SVR model using Landsat-8 outperforms that using Sentinel-2 data are consistent with the results presented here.

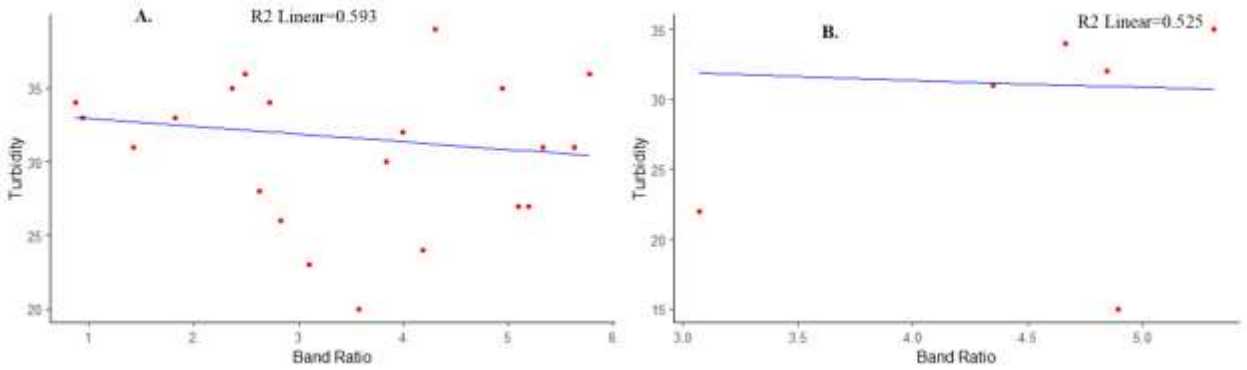



**Figure 14**: Sentinel-2 plots illustrate relationships between actual and predicted turbidity for (a) training set (n=20), (b) validation set (n=8)

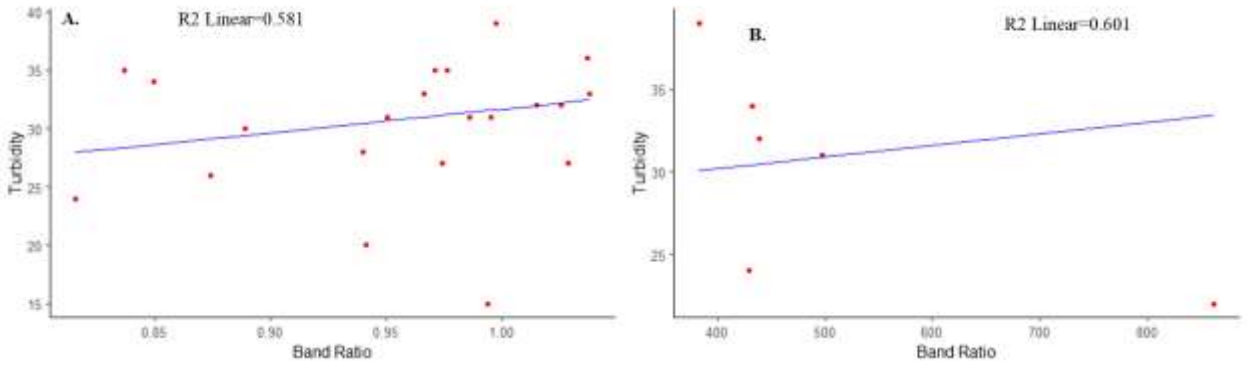



**Figure 15**: Landsat-8 plots illustrate relationships between actual and predicted turbidity for (a) training set (n=20), (b) validation set (n=8)

Figures 14 and 15 illustrate the predicted vs measured turbidity. The turbidity SVR $R^2$ values for Sentinel-2 and Landsat-8 were extremely low with the values of 0.59 and 0.58, respectively. The results shows that the two sensors used, Sentinel-2 is the best sensor for predicting turbidity in the Vaalkop Dam.

# CHAPTER 5
# DISCUSSION

Using a machine learning algorithm called support vector machine, this research project examined the capability of Sentinel-2 and Landsat-8 imagery in predicting turbidity and chlorophyll-a in the Vaalkop dam. Both images are useful considerably well in predicting the two water quality indicators, which is consistent with the findings of Bande's (2020) study, which revealed that Sentinel-2 and Landsat-8 images can be used to predict turbidity and chlorophyll-a, respectively. Considering the comparing the efficiency of the two images, Sentinel-2 images performed better in predicting chlorophyll-a than Landsat-8 image. The reason for the improved accuracy of Sentinel-2 prediction is due its 10m spatial resolution which is higher compared to 30m resolution of Landsat-8 which makes it most efficient to capture the variations on the small area of Vaalkop dam (Zhou *et al.,* 2020). Turbidity predictions utilizing Landsat-8 and Sentinel-2 imagery were not good enough comparing with the regression coefficients ($R^2$) of chlorophyll-a on the validation results. (Kallel *et al.,* 2018)

The low $R^2$ in the validation of prediction model was due to different in dates for data acquisitions in which ground measurements of water quality parameters was done on 28th October 2020 whilst the downloaded cloud free Sentinel-2 and Landsat-8 images were retrieved on 12th October 2020 and 16th October 2020 respectively (Radoux *et al.,* 2016). To achieve better analysis, our study aimed at finding the cloud free satellite images. For this reason, used images for prediction had 12- and 16-days' time different with the date of ground measurement for Landsat-8 and Sentinel-2 individually. Studies on the subject have indicated that the maximum reasonable time different between ground measured and satellite image acquisition to be ± 8 days for water environmental parameters (Nguyen *et al.,* 2020) and the prediction using data with small time different between ground measured the levels for turbidity and chlorophyll-a as well as the satellite image acquisition, the better the validation results (Oucher *et al.,* 2018 ; Wang *et al.,* 2021). For that case, factors such as changes on weather and hydrological parameters over a time difference have been the influence on the prediction accuracy.

In this study, predictions of turbidity and chlorophyll-a was done based on ground measured data acquired in October which is in mid-spring season thus the predictions were limited on one season only. Studies have shown that predictive models developed using data from one season, may not be useful in other weather seasons (Oucher *et al.,* 2018). According to the researchers, seasonal fluctuations in hydro meteorological variables need seasonal adjustment of the predictive model (Boucher *et al.,* 2018) However, selective band choices based on the research geographic area bodies of water, as well as combining remote sensing data surface temperature of smaller satellites band, may limit the effects of seasonal changes. Several algorithms can be used for atmospheric correction of Sentinel-2 images for example Sen2Cor, ACOLITE and MIP (Modular Inversion and Processing System can used for atmospheric correction both in land and water parameter studies, and the choice of atmospheric correction influence the regression results (Dörnhöfer *et al.,*2016).

General results suggest that the developed Chlorophyll-a prediction model performs better using Sentinel-2 comparing to Landsat-8 images and the resolution difference between the two images

account for the difference in prediction accuracy. Furthermore, despite having sample size enough to tune the predictive model, other factors such as time different, atmospheric distortions, and the choice of atmospheric correction algorithm influenced the validation accuracy of the model.

# CHAPTER 6
# CONCLUSSION

## 6.1 Conclusion

The findings of this study reveal that a support vector machine could be utilized to monitor water quality (turbidity and chlorophyll-a) for a Vaalkop dam in South Africa using Sentinel-2 and Landsat-8 data. For chlorophyll-a predictions, Sentinel-2 bands B02, B04 and B06 and Landsat-8 bands B03, B04 and B07 performed better whilst Sentinel-2 bands B01, B03 and B12 and Landsat-8 bands B01, B05 and B07 performed better for the turbidity predictions. Although support vector machine regression was found to perform relatively well with the small sample sizes, there is a need to increase sample sizes to improve predictions.

## 6.2 Recommendations and Limitations

The fact that field measurements were only gathered for one season limited comparability and application to subsequent seasons in this study. Further studies are recommended to expand the scheme and use data from all four seasons and comparisons of the performance of algorithms for different seasons. Additionally, while this study only used a support vector machine method for estimating, other machine learning techniques such as Artificial Neural Networks (ANN) and Random Forests (RF) should be evaluated in comparison. Other atmospheric correction methods, other than Sen2Cor, should be studied and evaluated, according to this study (Geman *et al.*, 1992).

# REFERENCES

Abdi, H., 2010. Partial least squares regression and projection on latent structure regression (PLS Regression). *Wiley interdisciplinary reviews: computational statistics*, *2*(1), pp.97-106.

Adakole, J.A., Mbah, C.E. and Dalla, M.A., 2003. Physicochemical limnology of Lake Kubanni, Zaria-Nigeria.

Adam, E., Mutanga, O., 2009. Spectral discrimination of papyrus vegetation (Cyperus papyrus L.) in swamp wetlands using field spectrometry. ISPRS J. Photogrammetry. Remote Sens. 64, 612–620. https://doi.org/10.1016/j.isprsjprs.2009.04.004

Allan, M.G., Hicks, B.J. and Brabyn, L., 2007. Remote sensing of water quality in the Rotorua lakes.

Ashton, P.J., Chutter, P.M., Cochrane, K.L., De Moor, F.C., Hely-Hutchinson, J.R., Jarvis, A.C., Robarts, R.D., Scott, W.E., Thornton, J.A., Twinch, A.J. and Zohary, T., 1985. Limnology of Hartbeespoort dam. National Scientific Programmes Unit: CSIR.

Ashton, P.J., Hardwick, D., and Breen, C.M., 2008. Changes in water availability and demand within South Africa's shared river basins as determinants of regional social-ecological resilience. *Burns, MJ & Weaver, AvB*, pp.279-310.

Babin, M., Morel, A. and Gentili, B., 1996. Remote sensing of sea surface sun-induced chlorophyll fluorescence: consequences of natural variations in the optical characteristics of phytoplankton and the quantum yield of chlorophyll a fluorescence. International Journal of Remote Sensing, 17(12), pp.2417-2448.

Banadkooki, F.B., Ehteram, M., Panahi, F., Sammen, S.S., Othman, F.B. and Ahmed, E.S., 2020. Estimation of total dissolved solids (TDS) using new hybrid machine learning models. Journal of Hydrology, 587, p.124989.Bande, P., Adam, E., Elbasit, M.A.A. and Adelabu, S., 2018 comparing Landsat-8 and sentinel-2 in mapping water quality at Vaal dam.

Barbieux, M., Uitz, J., Bricaud, A., Organelli, E., Poteau, A., Schmechtig, C., Gentili, B., Obolensky, G., Leymarie, E., Penkerc'h, C. and d'Ortenzio, F., 2018. Assessing the variability in the relationship between the particulate backscattering coefficient and the chlorophyll a concentration from a global biogeochemical-Argo database. Journal of Geophysical Research: Oceans, 123(2), pp.1229-1250.

Barnard, D., 1999. Environmental Law for All: A Practical Guide for the Business Community, the Planning Professions, Environmentalists, and Lawyers. Impact Books.

Barrett, D.C., and Frazier, A.E., 2016. An automated method for monitoring water quality using Landsat imagery. Water, 8(6), p.257.

Bartram, J. and Ballance, R. eds., 1996. Water quality monitoring: a practical guide to the design and implementation of freshwater quality studies and monitoring programmes. CRC Press.

Basson, M.S., Van Niekerk, P.H. and Van Rooyen, J.A., 1997. Overview of water resources availability and utilisation in South Africa. Department of Water Affairs and Forestry.

Bates BC, Kundzewicz ZW, Wu S, Palutikof JP. Climate change and water. Technical paper of the Intergovernmental Panel on Climate Change. Geneva: IPCC Secretariat; 2008

Bellacicco, M., Volpe, G., Briggs, N., Brando, V., Pitarch, J., Landolfi, A., Colella, S., Marullo, S. and Santoleri, R., 2018. Global distribution of non-algal particles from ocean color data and implications for phytoplankton biomass detection. Geophysical Research Letters, 45(15), pp.7672-7682.

Binding, Caren E., John H. Jerome, Robert P. Bukata, and William G. Booty. "Spectral absorption properties of dissolved and particulate matter in Lake Erie." Remote sensing of environment 112, no. 4 (2008): 1702-1711.

Bonansea, M., Ledesma, M., Rodriguez, C. and Pinotti, L., 2019. Using new remote sensing satellites for assessing water quality in a reservoir. Hydrological sciences journal, 64(1), pp.34-44.

Boucher, J., Weathers, K.C., Norouzi, H. and Steele, B., 2018. Assessing the effectiveness of Landsat-8 chlorophyll a retrieval algorithm for regional freshwater monitoring. Ecological applications, 28(4), pp.1044-1054.

Buma, W.G. and Lee, S.I., 2020. Evaluation of sentinel-2 and Landsat-8 images for estimating chlorophyll-a concentrations in lake Chad, Africa. Remote Sensing, 12(15), p.2437.

Chawla, I., Karthikeyan, L., and Mishra, A.K., 2020. A Review of Remote Sensing Applications for Water Security: Quantity, Quality, and Extremes. Journal of Hydrology, p.124826.

Coble, P., Hu, C., Gould Jr, R.W., Chang, G. and Wood, A.M., 2004. *Colored Dissolved Organic Matter in the Coastal Ocean: An Optical Tool for Coastal Zone Environmental Assessment & Management*. naval research lab stennis space center ms oceanography div.

Coskun, H.G., Tanik, A., Alganci, U. and Cigizoglu, H.K., 2008. Determination of environmental quality of a drinking water reservoir by remote sensing, GIS, and regression analysis. *Water, air, and soil pollution*, *194*(1), pp.275-285.

Dallas HF, Day JA. The effect of water quality variables on aquatic ecosystems: A review. Water Research Commission technical report 224/04. Pretoria: Water Research Commission; 2004.

Dallas, H.F. and Rivers-Moore, N., 2014. Ecological consequences of global climate change for freshwater ecosystems in South Africa. *South African Journal of Science*, *110*(5-6), pp.01-11.

Davis, A., 2017. Hydrogeological Characteristics of Hartbeespoort Dam (Doctoral dissertation).

Dekker, A.G., Vos, R.J., and Peters, S.W.M., 2002. Analytical algorithms for lake water TSM estimation for retrospective analyses of TM and SPOT sensor data. *International journal of remote sensing*, *23*(1), pp.15-35.

Department of Water Affairs and Forestry, 1998a. Quality of Domestic Water Supplies.Volume1 Assessment Guide.WRC.TT101/98.2nd ed.

Dörnhöfer, K., Göritz, A., Gege, P., Pflug, B. and Oppelt, N., 2016. Water constituents and water depth retrieval from Sentinel-2A—The first evaluation in an oligotrophic lake. *Remote Sensing*, *8*(11), p.941.

Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., Hoersch, B., Isola, C., Laberinti, P., Martimort, P. and Meygret, A., 2012. Sentinel-2: ESA's optical high-resolution mission for GMES operational services. Remote sensing of Environment, 120, pp.25-36.

Du Plessis, J., 2006. The assessment of the water quality of the Hex River Catchment-North -West Province (Doctoral dissertation, University of Johannesburg).

Dube, T., Mutanga, O., Seutloali, K., Adelabu, S. and Shoko, C., 2015. Water quality monitoring in sub-Saharan African lakes: a review of remote sensing applications. *African Journal of Aquatic Science*, *40*(1), pp.1-7.

Dutta, S., Kole, R.K., Ghosh, S., Nath, D. and Vass, K.K., 2005. Impact assessment of lead on water quality of River Ganga in West Bengal, India. Bulletin of Environmental Contamination & Toxicology, 75(5).

DWA. 2012. Crocodile West River Reconciliation Strategy 2012. Department of Water Affairs Republic of South Africa Report number WMA 03/A31/00/6110/4 December 2012. https://www.dwa.gov.za/Projects/crocodilemaintenance/Documents/MOCWS%20Reports/CWRS%20Strategy%202012%20FINAL.pdf Date of access: December 2012.

DWAF (Department of Water Affairs and Forestry) 1998: Resource Directed Measures for Water Resource Protection: Integrated Report. Institute for Water Quality Studies Report No N/0000/00/_/REH0299. Unofficial first draft. Pretoria: Institute for Water Quality StudiesEllison, G.T.H., 1990. A note on the small mammal fauna of Vaalkop Dam Nature Reserve. *Koedoe*, *33*(1), pp.114-116.

Forkuor, G., Hounkpatin, O.K., Welp, G. and Thiel, M., 2017. High resolution mapping of soil properties using remote sensing variables in south-western Burkina Faso: a comparison of machine learning and multiple linear regression models. *PloS one*, *12*(1), p.e0170478.

Geman, S., Bienenstock, E. and Doursat, R., 1992. Neural networks and the bias/variance dilemma. Neural computation, 4(1), pp.1-58.

Gholizadeh, M.H., Melesse, A.M. and Reddi, L., 2016. A comprehensive review on water quality parameters estimation using remote sensing techniques. Sensors, 16(8), p.1298.

Gilerson, A.A., Gitelson, A.A., Zhou, J., Gurlin, D., Moses, W., Ioannou, I. and Ahmed, S.A., 2010. Algorithms for remote estimation of chlorophyll-a in coastal and inland waters using red and near infrared bands. Optics express, 18(23), pp.24109-24125.

Gitelson, A.A., Dall'Olmo, G., Moses, W., Rundquist, D.C., Barrow, T., Fisher, T.R., Gurlin, D. and Holz, J., 2008. A simple semi-analytical model for remote estimation of chlorophyll-a in turbid waters: Validation. Remote Sensing of Environment, 112(9), pp.3582-3593.

Gons, H.J., Rijkeboer, M. and Ruddick, K.G., 2002. A chlorophyll-retrieval algorithm for satellite imagery (Medium Resolution Imaging Spectrometer) of inland and coastal waters. Journal of Plankton Research, 24(9), pp.947-951.

Gordon, H.R. and Morel, A.Y., 2012. Remote assessment of ocean color for interpretation of satellite visible imagery: A review.

Gordon, H.R. and Wang, M., 1994. Retrieval of water-leaving radiance and aerosol optical thickness over the oceans with SeaWiFS: a preliminary algorithm. Applied optics, 33(3), pp.443-452.

Gupta, D.K., Prasad, R., Kumar, P., Mishra, V.N., Vishwakarma, A.K. and Srivastava, P.K., 2015. Support vector regression for retrieval of soil moisture using bistatic scatterometer data at X-band. *International Journal of Geological and Environmental Engineering*, *9*(10), pp.1201-1204.

Hafeez, S., Wong, M.S., Ho, H.C., Nazeer, M., Nichol, J., Abbas, S., Tang, D., Lee, K.H. and Pun, L., 2019. Comparison of machine learning algorithms for retrieval of water quality indicators in case-II waters: a case study of Hong Kong. Remote sensing, 11(6), p.617.

Hamann, R., 2003. Mining companies' role in sustainable development: The 'why' and' how' of corporate social responsibility from a business perspective. Development Southern Africa, 20(2), pp.237-254.

Hansen, C.H., Williams, G.P., Adjei, Z., Barlow, A., Nelson, E.J. and Miller, A.W., 2015. Reservoir water quality monitoring using remote sensing with seasonal models: a case study of five central-Utah reservoirs. Lake and Reservoir Management, 31(3), pp.225-240.

Harding, W.R., 2008. The determination of annual phosphorus loading limits for South African dams. Pretoria: Water Research Commission.

Hinrichsen, D. and Tacio, H., 2002. The coming freshwater crisis is already here. The linkages between population and water. Washington, DC: Woodrow Wilson International Center for Scholars, pp.1-26.

Hornik, K., Meyer, D. & Karatzoglou, A. 2006. Support vector machines in R. Journal of statistical software, 15, 1-28.

http://cs2016.statssa.gov.za

https://www.cogta.gov.za/ddm/wpcontent/uploads/2020/08/DistrictProfile_BOJANALA01072020.pdf

https://www.gcis.gov.za/content/resource-centre/sa-info/officialguide/2017-18
https://www.fse.org.za/index.php/water-issues/item/606-summary-of-water-related-challenges-in-south-africa-2018

Hung, J., Lin, C.H., Wang, J.D. and Chan, C.C., 2006. Exhaled carbon monoxide level as an indicator of cigarette consumption in a workplace cessation program in Taiwan. *Journal of the Formosan Medical Association*, *105*(3), pp.210-213.

Iqbal, J., Mumtaz, M.W., Mukhtar, H., Iqbal, T., Mahmood, S.H.A.H.I.D. and Razaq, A., 2010. Particle size distribution analysis and physico-chemical characterization of Chenab River water at Marala Headworks. Pak. J. Bot, 42(2), pp.1153-1161.

Kallel, A., Ksibi, M., Dhia, H.B. and Khélifi, N., 2017. Recent Advances in Environmental Science from the Euro-Mediterranean and Surrounding Regions Proceedings of Euro-Mediterranean Conference for Environmental Integration (EMCEI-1), Tunisia 2017. In Conference proceedings EMCEI (p. 167).

Kannel, P.R., Lee, S., Lee, Y.S., Kanel, S.R. and Khan, S.P., 2007. Application of water quality indices and dissolved oxygen as indicators for river water classification and urban impact assessment. Environmental monitoring and assessment, 132(1-3), pp.93-110.

Kapalanga, T.S., 2016. Assessment and development of remote sensing-based algorithms for water quality monitoring in Olushandja Dam, North-Central Namibia.

Khattab, M.F. and Merkel, B.J., 2014. Application of Landsat-5 and Landsat-7 images data for water quality mapping in Mosul Dam Lake, Northern Iraq. *Arabian Journal of Geosciences*, *7*(9), pp.3557-3573.

Kim, Y.H., Im, J., Ha, H.K., Choi, J.K. and Ha, S., 2014. Machine learning approaches to coastal water quality monitoring using GOCI satellite data. *GIScience & Remote Sensing*, *51*(2), pp.158-174.

Kühn, A.L., Venter, S.N., Van Ginkel, C., Vermaak, E. and Zingitwa, L., 2000, May. Identification of areas with faecally polluted surface water sources in South Africa. In Conference paper presented at WISA.

Kutser, T., Paavel, B., Verpoorter, C., Ligi, M., Soomets, T., Toming, K. and Casal, G., 2016. Remote sensing of black lakes and using 810 nm reflectance peak for retrieving water quality parameters of optically complex waters. Remote Sensing, 8(6), p.497.

Lathrop, R.G. and Lillesand, T.M., 1986. Use of Thematic Mapper data to assess water quality in Green Bay and central Lake Michigan. *Photogrammetric Engineering & Remote Sensing*, *52*(5), pp.671-680.

Li, B., Yang, G., Wan, R., Dai, X. and Zhang, Y., 2016. Comparison of random forests and other statistical methods for the prediction of lake water level: a case study of the Poyang Lake in China. *Hydrology Research*, *47*(S1), pp.69-83.

Li, R. and Li, J., 2004. Satellite remote sensing technology for lake water clarity monitoring: an overview. Environmental Informatics Archives, 2, pp.893-901.

Lim, J. and Choi, M., 2015. Assessment of water quality based on Landsat-8 operational land imager associated with human activities in Korea. Environmental monitoring and assessment, 187(6), p.384.

Liversedge, L., 2007. Turbidity mapping and prediction in ice marginal lakes at the Bering Glacier System, Alaska (Doctoral dissertation).

Mancino, G., Nolè, A., Urbano, V., Amato, M. and Ferrara, A., 2009. Assessing water quality by remote sensing in small lakes: the case study of Monticchio lakes in southern Italy. *iForest-Biogeosciences and Forestry*, 2(4), p.154.

Mandanici, E. and Bitelli, G., 2016. Preliminary comparison of sentinel-2 and Landsat-8 imagery for combined use. *Remote Sensing*, *8*(12), p.1014.

Mao, H., Meng, J., Ji, F., Zhang, Q. and Fang, H., 2019. Comparison of machine learning regression algorithms for cotton leaf area index retrieval using Sentinel-2 spectral bands. *Applied Sciences*, *9*(7), p.1459.

Masocha, M., Dube, T., Nhiwatiwa, T. and Choruma, D., 2018. The testing utility of Landsat-8 for remote assessment of water quality in two subtropical African reservoirs with contrasting trophic states. Geocarto International, 33(7), pp.667-680.

Masocha, M., Murwira, A., Magadza, C.H., Hirji, R. and Dube, T., 2017. Remote sensing of surface water quality in relation to catchment condition in Zimbabwe. *Physics and Chemistry of the Earth, Parts A/B/C*, *100*, pp.13-18.

Matthews, M.W., Bernard, S. and Winter, K., 2010. Remote sensing of cyanobacteria-dominant algal blooms and water quality parameters in Zeekoevlei, a small hypertrophic lake, using MERIS. *Remote Sensing of Environment*, *114*(9), pp.2070-2087.

Meissner, R., Steyn, M., Moyo, E., Shadung, J., Masangane, W., Nohayi, N. and Jacobs-Mata, I., 2018. South African local government perceptions of the state of water security. Environmental Science & Policy, 87, pp.112-127.

Moses, W.J., Gitelson, A.A., Berdnikov, S. and Povazhnyy, V., 2009. Estimation of chlorophyll-a concentration in case II waters using MODIS and MERIS data—successes and challenges. Environmental research letters, 4(4), p.045005.

Mountrakis, G., Im, J. and Ogole, C., 2011. Support vector machines in remote sensing: A review. ISPRS Journal of Photogrammetry and Remote Sensing, 66(3), pp.247-259.

Mushtaq, F. and Nee Lala, M.G., 2017. Remote estimation of water quality parameters of Himalayan Lake (Kashmir) using Landsat-8 OLI imagery. *Geocarto international*, *32*(3), pp.274-285.

New, M., 2002. Climate change and water resources in the southwestern Cape, South Africa. South African Journal of Science, 98(7), pp.369-376.

Nguyen, H.Q., Ha, N.T. and Pham, T.L., 2020. Inland harmful cyanobacterial bloom prediction in the eutrophic Tri and Reservoir using satellite band ratio and machine learning approaches. Environmental Science and Pollution Research, 27(9), pp.9135-9151.

Obaid, A.A., Ali, K.A., Abiye, T.A. and Adam, E.M., 2021. Assessing the utility of using current generation high-resolution satellites (Sentinel-2 and Landsat-8) to monitor large water supply dam in South Africa. Remote Sensing Applications: Society and Environment, 22, p.100521.

Oberholster, P.J. and Ashton, P.J., 2008. State of the nation report: An overview of the current status of water quality and eutrophication in South African rivers and reservoirs. *Parliamentary Grant Deliverable. Pretoria: Council for Scientific and Industrial Research (CSIR)*.

Ogleni, N. and Topal, B., 2011. Water quality assessment of the Mudurnu River, Turkey, using biotic indices. Water resources management, 25(10), p.2487.

O'Reilly, J.E., Maritorena, S., Mitchell, B.G., Siegel, D.A., Carder, K.L., Garver, S.A., Kahru, M. and McClain, C., 1998. Ocean color chlorophyll algorithms for SeaWiFS. Journal of Geophysical Research: Oceans, 103(C11), pp.24937-24953.

Pizani, F.M., Maillard, P., Ferreira, A.F. and de Amorim, C.C., 2020. Estimation of water quality in a reservoir from Sentinel-2 msi and Landsat-8 oli sensors. ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences, 5(3).

Radoux, J., Chomé, G., Jacques, D.C., Waldner, F., Bellemans, N., Matton, N., Lamarche, C., d'Andrimont, R. and Defourny, P., 2016. Sentinel-2's potential for sub-pixel landscape feature detection. *Remote Sensing*, *8*(6), p.488.

Ritchie, J.C., Zimba, P.V. and Everitt, J.H., 2003. Remote sensing techniques to assess water quality. Photogrammetric Engineering & Remote Sensing, 69(6), pp.695-704.

Robarts, R.D., Ashton, P.J., Thornton, J.A., Taussig, H.J. and Sephton, L.M., 1982. Overturn in a hypertrophic, warm, monomictic impoundment (Hartbeespoort Dam, South Africa). *Hydrobiologia*, *97*(3), pp.209-224.

Rodriguez-Galiano, V., Sanchez-Castillo, M., Chica-Olmo, M. and Chica-Rivas, M.J.O.G.R., 2015. Machine learning predictive models for mineral prospectivity: An evaluation of neural networks, random forest, regression trees and support vector machines. *Ore Geology Reviews*, *71*, pp.804-818.

Roy, D.P., Wulder, M.A., Loveland, T.R., Woodcock, C.E., Allen, R.G., Anderson, M.C., Helder, D., Irons, J.R., Johnson, D.M., Kennedy, R. and Scambos, T.A., 2014. Landsat-8: Science and product vision for terrestrial global change research. Remote sensing of Environment, 145, pp.154-172.

Ruescas, A.B., Hieronymi, M., Mateo-Garcia, G., Koponen, S., Kallio, K. and Camps-Valls, G., 2018. Machine learning regression approaches for colored dissolved organic matter (CDOM) retrieval with S2-MSI and S3-OLCI simulated data. *Remote Sensing*, *10*(5), p.786.

Sakuno, Y., Yajima, H., Yoshioka, Y., Sugahara, S., Abd Elbasit, M.A., Adam, E. and Chirima, J.G., 2018. Evaluation of unified algorithms for remote sensing of chlorophyll-a and turbidity in lake shinji and lake nakaumi of Japan and the vaal dam reservoir of South Africa under eutrophic and ultra-turbid conditions. *Water*, *10*(5), p.618.

Samarghandi, M.R., Nouri, J., Mesdaghinia, A.R., Mahvi, A.H., Nasseri, S. and Vaezi, F., 2007. Efficiency removal of phenol, lead and cadmium by means of UV/TiO 2/H 2 O 2 processes. International Journal of Environmental Science & Technology, 4(1), pp.19-25.

Schaeffer, B.A., Schaeffer, K.G., Keith, D., Lunetta, R.S., Conmy, R. and Gould, R.W., 2013. Barriers to adopting satellite remote sensing for water quality management. *International Journal of Remote Sensing*, *34*(21), pp.7534-7544.

Schalles, J.F., Gitelson, A.A., Yacobi, Y.Z. and Kroenke, A.E., 1998. Estimation of chlorophyll a from time series measurements of high spectral resolution reflectance in a eutrophic lake. *Journal of Phycology*, *34*(2), pp.383-390.

Shabalala, A.N., Combrinck, L. and McCrindle, R., 2013. Effect of farming activities on seasonal variation of water quality of Bonsma Dam, KwaZulu-Natal. South African Journal of Science, 109(7-8), pp.01-07.

Shock, C.C., Pratt, K. and Station, M.E., 2003, March. Phosphorus effects on surface water quality and phosphorus TMDL development. In Western nutrient management conference (Vol. 5, No. 21, p. 1).

Simis, S.G., Ruiz-Verdú, A., Domínguez-Gómez, J.A., Peña-Martinez, R., Peters, S.W. and Gons, H.J., 2007. Influence of phytoplankton pigment composition on remote sensing of cyanobacterial biomass. *Remote Sensing of Environment*, *106*(4), pp.414-427.

Smith, V.H., 2003. Eutrophication of freshwater and coastal marine ecosystems a global problem. Environmental Science and Pollution Research, 10(2), pp.126-139.

Somvanshi, S., Kunwar, P., Singh, N.B., Shukla, S.P. and Pathak, V., 2012. Integrated remote sensing and GIS approach for water quality analysis of Gomti river, Uttar Pradesh. International Journal of Environmental Sciences, 3(1), pp.62-74.

Steyn, S., 2006. *The management of aerial particulate pollution: the case of platinum industry smelters in the Rustenburg region of Northwest Province, South Africa* (Doctoral dissertation, University of Pretoria).

Strydom, H.A., King, N.D., Fuggle, R.F. and Rabie, M.A. eds., 2009. Environmental Management in South Africa. Juta and Company Ltd.

Sun, D., Li, Y. and Wang, Q., 2009. A unified model for remotely estimating chlorophyll-a in Lake Taihu, China, based on SVM and in situ hyperspectral data. IEEE Transactions on Geoscience and Remote Sensing, 47(8), pp.2957-2965.

Swanepoel, A., Du Preez, H.H. and Cloete, N., 2017. The occurrence and removal of algae (including cyanobacteria) and their related organic compounds from source water in Vaalkop Dam with conventional and advanced drinking water treatment processes. *Water SA*, *43*(1), pp.67-80.

Tessema, A., Mohammed, A., Birhanu, T. and Negu, T., 2014. Assessment of physico-chemical water quality of Bira dam, Bati Wereda, Amhara region, Ethiopia. Journal of Aquaculture Research and Development, 5(6).

Toming, K., Kutser, T., Laas, A., Sepp, M., Paavel, B. and Nõges, T., 2016. First experiences in mapping lake water quality parameters with Sentinel-2 MSI imagery. *Remote Sensing*, *8*(8), p.640.

Trinh, R.C., Fichot, C.G., Gierach, M.M., Holt, B., Malakar, N.K., Hulley, G. and Smith, J., 2017. Application of Landsat-8 for monitoring impacts of wastewater discharge on coastal water quality. Frontiers in Marine Science, 4, p.329.

Turner, Dan. "June 2010." Remote Sensing of Chlorophyll-a Concentrations to Support the Deschutes Basin Lake and Reservoirs TMDLs. Department of Environmental Quality. Portland, OR (2010).

UNESCO, T., 2012. Managing water under uncertainty and risk. The United Nations world water development report, 4.

United Nations Environment Programme. Division of Early Warning, Assessment, African Ministers' Council on Water, African Union. Commission, United States. Department of State and European Union, 2010. Africa water atlas (Vol. 1). UNEP/Earth print.

Vanhellemont, Q. and Ruddick, K., 2015. Advantages of high-quality SWIR bands for ocean colour processing: Examples from Landsat-8. *Remote Sensing of Environment*, *161*, pp.89-106.

Wagner, P.D., Kumar, S. and Schneider, K., 2013. An assessment of land-use changes impacts on the water resources of the Mula and Mutha Rivers catchment upstream of Pune, India. Hydrology & Earth System Sciences Discussions, 10(2).

Wang, M., Yao, Y., Shen, Q., Gao, H., Li, J., Zhang, F. and Wu, Q., 2021. Time-Series Analysis of Surface-Water Quality in Xiong'an New Area, 2016–2019. Journal of the Indian Society of Remote Sensing, 49(4), pp.857-872.

Wang, X. and Yang, W., 2019. Water quality monitoring and evaluation using remote-sensing techniques in China: A systematic review. Ecosystem Health and Sustainability, 5(1), pp.47-56.

Wang, X., Fu, L. and He, C., 2011. Applying support vector regression to water quality modelling by remote sensing data. *International journal of remote sensing*, *32*(23), pp.8615-8627.

Wang, Y., Xia, H., Fu, J. and Sheng, G., 2004. Water quality change in reservoirs of Shenzhen, China: detection using LANDSAT/TM data. *Science of the Total Environment*, *328*(1-3), pp.195-206.

Wass, P.D., Marks, S.D., Finch, J.W., Leeks, G.J.X.L. and Ingram, J.K., 1997. Monitoring and preliminary interpretation of in-river turbidity and remote sensed imagery for suspended sediment transport studies in the Humber catchment. *Science of the Total Environment*, *194*, pp.263-283.

Xiao, X., Zhang, T., Zhong, X., Shao, W. and Li, X., 2018. Support vector regression snow-depth retrieval algorithm using passive microwave remote sensing data. *Remote sensing of environment*, *210*, pp.48-64.

Yadav, S., Yamashiki, Y., Susaki, J., Yamashita, Y. and Ishikawa, K., 2019. Chlorophyll estimation of lake water and coastal water using landsat-8 and sentinel-2a satellite. International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences.

Zhou, X., Chen, J., Rakstad, T.E., Ploughe, M. and Tang, P., 2020. Water chlorophyll estimation in an urban canal system with high-resolution remote sensing data. IEEE Geoscience and Remote Sensing Letters, 18(11), pp.1876-1880.

Zhou, X., Zhu, X., Dong, Z. and Guo, W., 2016. Estimation of biomass in wheat using random forest regression algorithm and remote sensing data. *The Crop Journal*, *4*(3), pp.212-219.