

UNIVERSITY OF THE WITWATERSRAND



UNIVERSITY OF THE  
WITWATERSRAND,  
JOHANNESBURG

FACULTY OF SCIENCE

# **Using Multispectral Remote Sensing for Mapping and Monitoring Water Quality at the Vaal Dam**

Akum Ernest Tanjeck (759483)

**School of Geography, Archaeology and Environmental Studies**

**A research report submitted to the Faculty of Science, University of the Witwatersrand  
in partial fulfilment of the requirements of the degree of Masters of Science  
(Geographical Information Science & Remote Sensing)**

**Supervisor:** Dr. Elhadi Adam

**Co-Supervisors:** Prof. Jasper Knight, Dr Mohamed Abd Elbasit

**Johannesburg, 2019**

## DECLARATION

I declare that this Theses “**Application of Multispectral Remote Sensing for Mapping and Monitoring Water Quality at the Vaal Dam**” is my own, unaided work. It is being submitted for the Degree of Masters of Science in Geographical Information Systems & Remote Sensing at the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination at any other University.

----- Signature of candidate

----- Day of -----20----- in .....

## Abstract

This project aimed to test the use of two multispectral sensors in estimating and mapping chlorophyll-a and turbidity concentrations, at the Vaal Dam, South Africa. Landsat 8 and Sentinel-2 satellite data were acquired on the 04/10/2016 and 13/11/2016 respectively. Image processing was carried out using atmospheric correction by applying the FLAASH model on the calibrated radiance to obtain atmospherically corrected Landsat 8 images. With Sentinel-2 data, the atmospheric correction was performed using Sen2cor from the Sentinel toolbox to obtain a geometrically corrected Sentinel-2 multispectral image. Band reflectance values were extracted from the two remotely sensed data, and laboratory measurements for chlorophyll-a and turbidity concentrations were obtained from samples collected from 23 sampling points at the dam on the 26-28 October 2016. The remotely sensed data were then cross-validated with the field data in mapping and predicting chlorophyll-a and turbidity concentrations. Two regression models were used in this study; multiple stepwise linear regression, and random forest regression models, to predict chlorophyll-a and turbidity concentrations from the remotely sensed data. The performance and accuracy of these regression models were evaluated by correlating the predicted against the observed chlorophyll-a and turbidity concentrations from the satellite data. Random forest regression gave a higher performance and accuracy than stepwise multiple regression based on their R squared and RMSE values. In mapping chlorophyll-a and turbidity concentrations from both remotely sensed datasets, stepwise linear regression analysis was used to derive estimates for chlorophyll-a and turbidity concentrations. The high estimate values were multiplied with their corresponding bands and added up with intercepts (the expected mean values for chlorophyll-a and turbidity) from coefficients derived from the regression analysis. An equation was then developed using the raster calculator in ArcGIS to map chlorophyll-a and turbidity concentrations from the remotely sensed data for the entire dam. The test of the new generation multispectral sensors Landsat 8 and Sentinel-2 using random forest and multiple stepwise regression models in mapping and predicting chlorophyll-a and turbidity concentrations at the dam was a success. The random forest model gave a better performance than multi stepwise regression. In some cases, the performance of the models was poor as a result of the field data collection date not coinciding with the date of satellite data collection.

## **Acknowledgements**

I would like to give an acknowledgement to the Agricultural Research Council-Institute for Soil, Climate, and Water (ARC- ISCW), the Japanese team that took part in the research, and also helped in analysing the laboratory data. I am also thankful for my supervisor and Dr Elhadi Adam and Prof. Jasper Knight for giving me guidance, help and corrections for my project. Furthermore, I will also like to thank Dr Mohamed Abd Elbasit of ARC as my co-supervisor in this project as well as the University of Wits for providing me with the good academic atmosphere for my Master's program. Lastly, I will also like to acknowledge IRG - Japan Society for the Promotion of Science / NRF Research Cooperation Programme project number 101569 in funding this project.

## Table of contents

DECLARATION .....	ii
Abstract.....	iii
Table of contents .....	iv
List of Figures .....	vi
List of Tables .....	vii
CHAPTER 1: Introduction .....	1
1.1 Background .....	1
1.2 Problem statement .....	3
1.3 Aims and objectives .....	4
CHAPTER 2: Literature review.....	5
2.1 Water quality and sources of pollution .....	5
2.2 The impacts of water pollution.....	5
2.3 Monitoring water quality.....	6
2.4 Remote sensing for water quality.....	6
2.4.1 Remote sensing for water quality using Landsat 8 and Sentinel-2 sensor .....	7
2.4.2 Atmospheric interference of satellite sensors.....	8
2.4.3 Spectral signatures of water and the impact of pollution .....	9
2.4.4 Remote sensing data for water quality.....	9
2.5 Comparison between field and satellite remote sensing in monitoring water quality .....	10
2.6 Application of remote sensing in monitoring water quality parameters .....	10
2.6.1 Suspended matter.....	11
2.6.2 Phytoplankton.....	11
2.6.3 Turbidity.....	12
2.6.4 Dissolved organic matter .....	13
2.7 Challenges and opportunities of remote sensing for water quality management.....	13
2.8 Comparison and integration of Landsat 8/OLI and Sentinel-2A/MSI data for mapping and monitoring water quality at the Vaal Dam .....	14
2.9 Regression methods used for water quality .....	15
CHAPTER 3: Materials and methods.....	17
3.1 Study area .....	17
3.2 Methodology.....	18
3.2.1 Remote sensing data acquisition and pre-processing .....	18
3.2.2 Field measurements.....	21

3.2.3. Regression modelling.....	23
3.2.4 Accuracy assessment of models .....	25
CHAPTER 4: Results .....	26
4.1. Descriptive statistics .....	26
4.2 Comparison of in situ measured and predicted chlorophyll-a and turbidity values .....	26
4.2.1 Predicting chlorophyll-a using different regression models from remotely sensed data ...	31
4.2.2 Predicting turbidity using different regression models from remotely sensed data.....	31
4.3 Accuracy assessment .....	31
4.4 Mapping spatial patterns for chlorophyll-a and turbidity at the dam .....	32
CHAPTER 5: Discussion.....	36
5.1 Findings obtained from the random forest and stepwise multiple linear regression models ...	36
5.2 Statistical analysis of chlorophyll-a and turbidity from field measurements .....	38
5.3 Spatial patterns for chlorophyll-a and turbidity concentrations at the dam.....	38
5.4 Relationship between band reflectance and predicted chlorophyll-a and turbidity values .....	39
Conclusions .....	40
References.....	42

## List of Figures

Figure 1: Map of the Vaal dam showing the study area .....	18
Figure 2: Map showing sampling points 1-23 at the dam.....	22
Figure 3: Graph showing predicted vs observed chlorophyll-a concentrations with Landsat 8 in evaluating random forest regression performance.....	27
Figure 4: Graph showing predicted vs observed chlorophyll-a concentrations with Landsat 8 data in evaluating stepwise multiple linear regression performance .....	27
Figure 5: Graph showing predicted vs observed turbidity concentrations with Landsat 8 data in evaluating random forest regression performance.....	28
Figure 6: Graph showing predicted vs observed turbidity concentrations with Landsat 8 data in evaluating stepwise multiple linear regression performance.....	28
Figure 7: Graph showing predicted vs observed chlorophyll-a concentrations with Sentinel-2 data in evaluating stepwise multiple linear regression performance .....	29
Figure 8: Graph showing predicted vs observed chlorophyll-a concentrations with Sentinel-2 data in evaluating random forests regression performance.....	29
Figure 9: Graph showing predicted vs observed turbidity concentrations with Sentinel-2 data in evaluating random forest regression performance.....	30
Figure 10: Graph showing predicted vs observed turbidity concentrations with Sentinel-2 data in evaluating stepwise multiple linear regression performance.....	30
Figure 11: Spatial pattern for chlorophyll-a concentrations estimated from Landsat 8 OLI data on 4 October 2016 at the dam (blue on the legend represent high and red to orange represent low) .....	33
Figure 12: Spatial pattern for turbidity concentrations estimated from Landsat 8 OLI data on 4 October 2016 at the dam.....	33
Figure 13: Spatial patterns for chlorophyll-a levels estimated from Sentinel-2A MSI data on 13 November 2016 at the dam.....	34
Figure 14: Spatial pattern for turbidity concentrations estimated from Sentinel-2A MSI data on 13 November 2016 at the dam.....	34

## List of Tables

Table 1: Relationship between chlorophyll-a (Ug/L) concentrations and trophic condition of lakes and reservoirs (Adapted from Boyd et al., 1997) .....	12
Table 2: The spectral bands of Landsat 8 sensor and its corresponding wavelengths (from Roy et al., 2014) ( <a href="https://landsat.gsfc.nasa.gov/landsat-8/landsat-8-overview/">https://landsat.gsfc.nasa.gov/landsat-8/landsat-8-overview/</a> ).....	19
Table 3: Sentinel-2 sensor with its spectral properties ( <a href="https://www.satimagingcorp.com/satellite-sensors/other-satellite-sensors/sentinel-2a/">https://www.satimagingcorp.com/satellite-sensors/other-satellite-sensors/sentinel-2a/</a> ).....	20
Table 4: The sensors used for this study and a comparison of their spectral properties.....	20
Table 5: Water quality parameters from each sampling point at the dam .....	22
Table 6: Statistical results for chlorophyll-a and turbidity concentrations from field data .....	26
Table 7: Performance of random forest regression and stepwise linear regression models for prediction of chlorophyll-a concentrations using reflectance values from satellite data.....	31
Table 8: Performance of random forest regression and stepwise linear regression models for prediction of turbidity levels using band reflectance values from satellite data.....	31
Table 9: Equations derived from estimates of chlorophyll-a and turbidity with their corresponding bands and intercepts used in mapping chlorophyll-a and turbidity concentrations using Landsat 8 and Sentinel-2 imagery.....	32

## CHAPTER 1: Introduction

### 1.1 Background

The Vaal Dam is one of the most significant dams in South Africa which provides fresh water to the Witwatersrand region (DWAF, 2008). Most developed and developing countries today are faced with the challenges of water pollution (Azizullah *et al.*, 2011). The water quality of this dam has been affected by different water quality parameters especially chlorophyll-a and turbidity (DWAF, 2008). Freshwater from the dam provides different uses like irrigation, domestic, industrial, and recreational uses which are all being affected by chlorophyll-a and turbidity concentrations. Each of these uses has different defined standards which support that use. Chlorophyll-a and turbidity concentrations at the dam which are the main water variables for this study are caused by eutrophication from agricultural lands that use chemical fertilisers, sewage from septic tanks from residents and acid mine drainage (Cech, 2010).

Surface water systems represent the major source for fresh water supply for economic and domestic uses in South Africa and around the world (World Health Organization, 2011). Estimating and mapping water quality variables is relevant in water management and planning, and therefore water quality information is needed for understanding the current conditions of natural water bodies and future trends in water quality (Pimentel *et al.*, 2004). The pollutants that can change the reflectance properties of water can be identified and quantified using hyperspectral and multispectral remote sensing techniques (Ritchie *et al.*, 2000; Schiebe *et al.*, 2000). Therefore, remote sensing techniques can be used in detecting and mapping variables for water quality in assisting decision-making organisations for future planning of water resource allocation, as well as agricultural and industrial impacts (Brivio *et al.*, 2001). This is because, with remote sensing techniques, water quality parameters for the entire water body can be estimated and mapped whereas with in situ measurement techniques, water variables can only be measured at specific points where water samples are collected, and the results may not be true for other areas. The advantage of remote sensing is that it can give a general view with continuous monitoring of the entire water body (Dube *et al.*, 2015). The deterioration of water quality is a major concern regarding issues related to public health, disease vectors, sanitation (DWAF, 2008), grey water recycling, the potable water security, and economic productivity of agriculture and industry (Dekker *et al.*, 1995). Therefore, treating, managing and monitoring water quality in surface waters can be of substantial financial cost to a country, especially with the presence of ongoing climate change

and increased water demands (Ritchie and Schiebe, 2000; Ritchie, 2000). Furthermore, the deterioration of water quality affects entire ecosystems, the food chain, and environmental quality within the catchment. Therefore, systematic detection and mapping of water quality variables are critical for managing and improving such water resources (Ritchie *et al.*, 2003).

The classification of water pollution sources includes point sources (e.g. pollution from industrial waste) and non-point sources (e.g. pollution from floods and runoff) (Chawira *et al.*, 2013; Dube *et al.*, 2015). Runoff from urban environments can contribute organic and inorganic substances particulates into freshwater sources (Ngabe *et al.*, 2000; Davis *et al.*, 2001; Huang *et al.*, 2004; Bian & Zhu, 2009) particularly when there is an increase in infrastructure and urbanisation. Runoff that contributes different pollutants to freshwater bodies can severely damage aquatic ecosystems by inhibiting the diffusion of dissolved oxygen for aerobic respiration (Marsalek *et al.*, 1999; Gobel *et al.*, 2007; Wei *et al.*, 2015). Also, freshwater bodies have been faced with pollution crises due to an increase in water demand from agricultural and industrial sectors (Jofre *et al.*, 2008).

Water quality is strongly influenced by the amount of inflowing water into the system which changes with catchment rainfall patterns, land use and geoengineering. For example, droughts reduce the amount of water inflow in a water body. This can cause an increase in pollutant concentration as the volume of water becomes smaller. This will have impacts on aquatic life, water usability and total resource volume. On the other hand, floods can increase water turbidity and nutrient fluxes due to erosion and high runoff and may increase water pollution by sewage overflow and acid mine drainage which may cause loss of biodiversity in the aquatic ecosystem as well as health problems to the public (Chawira *et al.*, 2013).

Detecting and mapping water variables for water quality assessment can improve water management and conservation as well as more general environmental quality (Ritchie *et al.*, 1996; Nellis *et al.*, 1998; Östlund *et al.*, 2001). Water quality mapping can be facilitated by predictive and modelling tools that can simulate the effect of climate and land use change on water quality (Park *et al.*, 2010). Furthermore, assessing water quality in water bodies can also be used to develop a comprehensive tool for government and health supervision agencies to make informed decisions on water management (Wang *et al.*, 2004; Coskun *et al.*, 2008).

In the past, water quality assessment was based on the estimation of water variables using field data obtained from specific field locations, which give measurements of water quality parameters only at that particular point. However, this type of approach is expensive and

time-consuming if being used to assess the water quality status for large water bodies. It also does not give reliable information on the spatial distribution of different water quality parameters for the entire water body (De Carlo *et al.*, 2007). Despite these limitations, some fixed devices/apparatus have been used to monitor water quality continuously (Schofield *et al.*, 2002; Glasgow *et al.*, 2004; Lee *et al.*, 2005; De Carlo *et al.*, 2007).

Remote sensing methods have been used extensively in detecting and mapping different physical variables of water bodies (e.g. Lavery and Pattiaratchi, 1993; Chen *et al.*, 2007; Gitelson *et al.*, 2007). Satellite sensors with a high spatial resolution like Sentinel-2, Worldview-2, MODIS, IKONOS, and SPOT (Satellite Pour l'Observation de la Terre) are easily obtained. With their high spectral and spatial resolutions, they are useful in detecting and mapping water quality variables across large areas ( Ekercin, 2007; Blondeau-Patissier *et al.*, 2014).

In this project, chlorophyll-a and turbidity concentrations at Vaal dam, Free State, South Africa, were mapped using spectral reflectance from remote sensing datasets. Chlorophyll-a and turbidity concentrations are affected by pollutants such as nitrogen and phosphorus from chemical fertilisers used in agriculture, carried into the dam by agricultural runoff. This causes excessive growth of algal blooms which produce an abundance of chlorophyll-a pigments. Moreover, suspended particulates from within the catchment and adjacent settlements are also washed into the dam. These pollutants also affect chlorophyll-a and turbidity concentrations. Hence, this project will be significant for water managers to be able to take precautions against pollution in the dam which is a significant source of freshwater to the Witwatersrand metropolitan area.

## **1.2 Problem statement**

The Vaal dam experiences pollution with biological, chemical and physical substances washed in from agricultural farmland, mines and settlements in the catchment (DWAF, 2008). This makes the dam potentially affected by the growth of algal blooms due to the presence of ion concentrations like nitrates ( $\text{NO}_3$ ) and phosphates ( $\text{PO}_4$ ) brought in during the inflow of sediments and biological substances from sewage (Bai *et al.*, 2009). The Vaal area has experienced a growth in urban development, agriculture and mining activities in the past decades. These activities have severely affected the dam's ecosystems and water quality (DWAF, 2008). There is, therefore, a need to measure and map chlorophyll-a and turbidity concentrations as the main parameters affecting water quality at the dam.

Water is a scarce resource in South Africa, where agricultural sectors, industrial and domestic sectors suffer from the variable quality and quantity of freshwater supplies. Water quality is an issue at the Vaal dam which is one of the most important sources of freshwater to urban areas within Gauteng. Therefore, there is the need to use remote sensing application for water quality assessment at the dam using remotely sensed data obtained from Landsat 8 Operational Land Imager (OLI) and Sentinel-2 together with field measurements for easy and continuous assessment of the water quality variables that is chlorophyll-a and turbidity concentrations (Dekker *et al.*, 2001; Salama *et al.*, 2009). The use of satellite data and field measurements of water quality parameters may prevent any pollution disaster that may arise at the dam (Salama *et al.*, 2009). This will save cost and time for water quality assessment.

### **1.3 Aims and objectives**

This project aims to test the use of two multispectral sensors in estimating and mapping chlorophyll-a, and turbidity concentrations, which are the primary water quality parameters affecting the Vaal Dam. The specific objectives are:

- 1) To assess the feasibility of integrating Landsat 8 and Sentinel-2 data with in situ measurements in estimating and mapping chlorophyll-a concentrations and turbidity levels at the Vaal Dam.
- 2) To use random forests regression and multiple stepwise linear regression models in predicting chlorophyll-a and turbidity concentrations by cross-validating measurements of spectral reflectance for chlorophyll-a and turbidity from satellite data with that from field measurements.
- 3) Using both random forest and multiple stepwise regression methods to execute both correlation and regression analysis to investigate and study the relationship between in situ measurements and remotely sensed data of the study area.

## **CHAPTER 2: Literature review**

### **2.1 Water quality and sources of pollution**

Understanding water quality parameters and detecting different water pollutants from agriculture, acid mine drainage, sewage discharge, industries and municipal wastes, which cause eutrophication, are essential in managing water resources and the conservation of water ecosystems and the environment. This is because clean water provides the foundation for prosperous communities (Dekker *et al.*, 1995). Water pollution is seen as a serious global issue and especially in South Africa, which requires ongoing evaluation of water resource policies at all levels. Moreover, the world is heading towards a water crisis, caused by climate changes which impact on lakes, dams and rivers. Advanced remote sensing methods provide opportunities for understanding a variety of changes on earth, and can facilitate the monitoring of water bodies and the water quality parameters associated with these (DWAF, 2008). Freshwater bodies have been affected by severe pollution as a result of increased water demand by industries, agriculture, urban areas and by population increase. Therefore, there is need to perform water quality monitoring in overcoming the issues that affect water quality in freshwater bodies, to enable the government to create policies that will help manage these resources such that there are no health effects to the public or ecosystems.

### **2.2 The impacts of water pollution**

Surface water bodies are the primary water source for South Africa's domestic, agricultural, and industrial water requirements. Therefore, maintaining water quality in surface water bodies is of major concern. These water bodies are highly sensitive to pollutants from both point and non-point sources (Mayo *et al.*, 1995; Zhang *et al.*, 2002; Hadjimitsis, 2010; Kneese *et al.*, 2013). Point source pollutants can be traced to a single source whereas non-point source pollutants are diffused across the landscape with the movement of water. Non-point sources are usually associated with climate, landscape processes such as soil erosion, land use management, and human activities within the catchment. Deterioration of water quality by high sediment loads, dissolved biogeochemical and organic and inorganic constituents is a significant issue for public health, disease vectors, sanitation, grey water recycling, the economic productivity of agriculture and industry, and potable water security (Rinta-Kanto *et al.*, 2009). Thus, treating, managing and monitoring of water quality is very important but can cost large amounts of money, especially in light of ongoing climate change and increased water demands (Zhang *et al.*, 2002). Moreover, the deterioration of water quality may affect a catchment's entire ecosystem, the food chain and overall environmental quality. The

motivation of this research was to develop a methodology that can capture and evaluate changes in water quality at a synoptic scale that is relevant to existing water resource management frameworks in South Africa.

Increased nitrogen and phosphorus loads in surface water can cause severe health issues to the public, such as blue baby syndrome (methemoglobinemia) (Knobeloch *et al.*, 2000) and cancer affecting vital body organs (Falconer *et al.*, 1999; Chorus and Fastner, 2001). Furthermore, the presence of nutrients at higher concentrations is a threat to water treatment plants that get their source of tap water from groundwater, lakes and coastal waters. Nitrogen and phosphorus at high levels may affect water bodies through increased chlorophyll-a pigments due to the growth of algal blooms (Alleman, 2013). This has led to the need for constant monitoring of water quantity and quality at a regional scale concerning their nutrient concentrations.

### **2.3 Monitoring water quality**

Good water quality standards can be achieved by monitoring changes in water quality in different water bodies (Ritchie and Cooper, 2001). Monitoring water quality is fundamental because it gives a better understanding of the state of the system and the level to which water quality may be affected by land use, land cover change, climate change etc. This will enable effective and efficient management by knowing what the drivers of water quality changes are.

### **2.4 Remote sensing for water quality**

The impacts of land use on site-scale water quality measurements have been considered in several studies. However, the effects of land use on distributed water quality have not been considered because water quality data have been usually measured at only a few points (Wang *et al.*, 2006; Chen *et al.*, 2007). These measuring points are usually located in populated areas, not across whole mixed (urban/rural) catchments. In this project, remote sensing-based water quality data were used to develop an algorithm to measure reflectance for chlorophyll-a and turbidity along with other water pollution constituents.

The study of water quality parameters such as turbidity and chlorophyll-a over space and time can be achieved through the use of remotely sensed data. In this study, the detection of these variables will be based on experimental models that will link field values of these parameters with reflectance values from remotely sensed data, according to different sectors of the electromagnetic spectrum (Cox *et al.*, 1998; Mueller, 2000; Giardino *et al.*, 2001; Wang *et al.*, 2006; Chen *et al.*, 2007b; Penafior *et al.*, 2007; Kabbara *et al.*, 2008).

Better methods for assessing water quality in different water bodies can be achieved with the use of remote sensing applications although, in developing countries, this approach has not been exploited (e.g. Chawira *et al.*, 2013; Kibena *et al.*, 2014; Majozi *et al.*, 2014). Besides the use of remote sensing applications for water quality, other factors affect the uptake of this method, including technical know-how, insufficient finance, inaccessibility and availability of the correct remotely sensed data for water quality studies (Majozi *et al.*, 2014).

Living components like cyanobacteria and algae which can be estimated from satellite remote sensing contain chlorophyll-a pigments (Chawira *et al.*, 2013; Kibena *et al.*, 2014; Qin *et al.*, 2014; Majozi *et al.*, 2014; Rinaldi *et al.*, 2013; Dube *et al.*, 2015). The chlorophyll-a present in the chloroplast of algae cells reflects green light but absorbs red and blue light which makes the algae exhibit a green colour (Dlamini *et al.*, 2016). With remote sensing applications on water quality, water can be classified into two categories; clean and polluted waters (Morel, 1977). Clean water refers to water source that does not pose substantial threat to humans. Polluted water is that which has been degraded by living and non-living elements such as the production of chlorophyll-a from algae and contamination of the water body with organic and inorganic pollutants. They affect chlorophyll-a and turbidity concentrations and total dissolved solutes in freshwater bodies (Gin, 2003). The complex interaction of both living and non-living substances in hard waters hinders the success of remote sensing applications for water quality (Novo *et al.*, 1989; Quibell, 1991; Lodhi *et al.*, 1997).

Different water parameters exhibit different spectral attributes which can be quantitative or qualitative, and the variables that cause a change in spectral signatures are chlorophyll-a, turbidity, and coloured dissolved organic matter. Deriving a relationship for the reflectance of satellite data and optical properties like backscattering coefficients and total absorption can be achieved with remote sensing (Gordon *et al.*, 1994).

#### **2.4.1 Remote sensing for water quality using Landsat 8 and Sentinel-2 sensor**

Landsat 8 Operational Land Imager (OLI) launched in 2013 as well as the multispectral imager (MSI) Sentinel-2 have high spatial and spectral resolution and have potential for detecting and mapping surface water quality variables (Pahlevan *et al.*, 2014; Franz *et al.*, 2015). The first of the Copernicus high resolution optical satellites, Sentinel-2, was launched in June 2015, and provides free data and better capabilities over the OLI with higher pixel resolution (that is 10 and 20 m per pixel compared to 30 m per pixel of OLI) in the visible

and near-infrared bands, and additional red-edge bands which enable the sensor to absorb chlorophyll-a of very small concentrations. The advantages of using high spatial resolution sensors (10-60 m) for monitoring coastal and inland waters is in their ability to map and monitor water quality. This has also allowed for resolving human processes and their effects from space such as quarrying (Barnes *et al.*, 2015; Vanhellemont and Ruddick, 2015) and offshore construction (Vanhellemont and Ruddick, 2014). Turbid rivers have been analysed in more detail as a result of the high spatial resolution and atmospheric correction using shortwave-infrared bands (Giardino *et al.*, 2014; Ruddick *et al.*, 2006), and retrieval of various optically active constituents (Concha and Schott, 2016). In some recent studies, Landsat 8 imagery has been used in analysing turbid tidal waves (Pacheco *et al.*, 2015), and also damage of coral reefs found associated with island building (Barnes and Hu, 2016).

Sentinel-2 imagery comes with 10 m, 20 m and 60 m spatial resolution, meaning small water bodies can be studied, and its data are acquired in 13 spectral bands. The radiometric resolution of the sensor is 12-bit. Since the multispectral image was not designed for aquatic remote sensing, its usefulness for water applications has been only for separating water bodies from land. Mapping water bodies from space using MSI data allows the improvement of mapping of global inland water bodies from current 14.25 m resolution to 10 m.

The red-edge band of aerial corrected images of Sentinel-2A (S2A/MSI) is of importance for monitoring water quality, as used by some European countries under the Water Framework Directive (WFD, 2000/60/EC and amendments). These waters present challenges for both in situ and remote sensing monitoring systems. The red-edge band shows good performance on atmospherically corrected images (S2A)/MSI and hence the possibility to map and quantify chlorophyll-a concentration in coastal and inland waters (Vanhellemont and Ruddick, 2016; Toming *et al.*, 2016).

#### **2.4.2 Atmospheric interference of satellite sensors**

The Sentinel-2 sensor records radiation leaving the water surface and transmitting through the atmosphere. This radiation then undergoes backscattering of atmospheric constituents with air molecules and dust particles and absorption and emission of radiation by greenhouse gases and water vapour (Salama *et al.*, 2012).

In remote sensing, cloud cover has severe impacts as the visible and near-infrared (NIR) radiation cannot penetrate cloud cover, and this makes the sensor unable to register the reflectance values of water. That is why cloud-free images should be acquired. Other

atmospheric effects can be corrected for satellite data and different atmospheric correction models that can improve water reflectance quality (Gordon, 1997; Ruddick *et al.*, 2000; Chomko *et al.*, 2003).

### **2.4.3 Spectral signatures of water and the impact of pollution**

A spectral signature is the variation of reflectance or emittance of a material concerning electromagnetic wavelengths. The spectral signature of water and spectral bands of Landsat 8 and Sentinel-2 are characterised by high absorption at visible and infrared wavelengths. Water bodies with high chlorophyll-a and turbidity concentrations have higher reflectance in the visible region than clear water. Thus, the reflectance patterns can be used to detect the presence of chlorophyll-a and turbidity concentrations in water bodies (Chawira *et al.*, 2013). From previous studies on water quality, some multispectral sensors like Landsat 5 and Moderate Resolution Imaging Spectro-radiometer (MODIS) were able to measure chlorophyll-a in different water bodies with the help of their spectral bands (Gilerson *et al.*, 2010; Olmanson *et al.*, 2013; Shuchman *et al.*, 2013). In most studies carried out on water quality, some variables were not easily measured due to the low spatial and spectral resolution of the sensors used. There are other methods through which water bodies have been effectively assessed for water quality, such as aircraft techniques in obtaining images with better spatial resolution. Despite the use of this method, there are some factors like high financial cost and a lower resolution involved in the acquisition of images. With Sentinel-2 imagery, the higher spatial and spectral resolution and 5 to 10 days revisit time mean that regular surveillance of some water quality parameters can take place.

### **2.4.4 Remote sensing data for water quality**

Water variables and their variability can be measured over time from remotely sensed data obtained with high spatial resolution sensors. The interaction of light, water and water variables on water quality can be understood using remotely sensed data (Richardson, 1996; Park and Ruddick, 2010).

Water quality in different sectors of any given water body can be assessed using narrow wavebands of the electromagnetic spectrum of multispectral sensors (Nieuwoudt and Backeberg, 2011; Soticha *et al.*, 2014). Some satellite sensors do not easily measure water parameters that occur at very low concentrations due to their spectral bands exhibiting a low resolution (Gower *et al.*, 2005). The type of any water variable measured can be determined by either the nature of the water body being studied or the sensor used in the study (Liu *et al.*,

2003). The amount of algae in any water body is related to the concentration of nutrients like nitrogen and phosphorus and can be estimated through chlorophyll-a levels (Letelier and Abbott, 1996; Vermote and Vermeulen, 1999; Wu et al., 2009). This has been done in previous studies with the help of multispectral sensors like MODIS and Landsat 5 and 7 (Verstraete et al., 1999; Chawira et al., 2013; Ndungu, 2014). Different regression models have been used in selecting bands or band ratios for estimating any water parameter (Carpenter and Carpenter, 1983; Lathrop and Lillesand, 1986; Ritchie *et al.*, 2003).

## **2.5 Comparison between field and satellite remote sensing in monitoring water quality**

The traditional method used in assessing water quality is based on the analysis of field measurements only. This method gives high accuracy in measurements of water variables at points where the samples were obtained. Despite the advantage of this method, there are limitations including the time and expense involved and that it does not give a general view of water quality status of the entire water body (Brivio *et al.*, 2001). Water quality assessment can also be achieved with remote sensing techniques by correlating variables obtained from field data with band reflectance from remotely sensed data (Schiebe *et al.*, 1992; Dekker and Peters, 1993; Zilioli and Brivio, 1997; Fraser, 1998; Kloiber *et al.*, 2000, 2002; Giardino *et al.*, 2001).

## **2.6 Application of remote sensing in monitoring water quality parameters**

Water quality involves the determination of chemical, physical and biological components of water bodies and identifying any possible source of pollution or contamination that might cause degradation of water quality. Water quality indicators are categorized as (i) Biological: bacteria, algae, (ii) Physical: temperature, turbidity and clarity, colour, salinity, suspended solids, dissolved solids, (iii) Chemical: pH, dissolved oxygen, biological oxygen demand, nutrients (including nitrogen and phosphorus), organic and inorganic compounds (including toxicants), and (iv) Aesthetic: odour, taints, colour, and floating matter. Previous studies on water quality have shown the importance of using remote sensing as a monitoring tool (Ritchie and Cooper, 1996; Schalles *et al.*, 1998; Dekker *et al.*, 2002).

The spectral reflectance obtained from the remote sensing data can give information about the band or wavelengths which are suitable for that particular water quality parameter. An empirical equation can be developed to relate spectral signature to each water quality parameter. A brief description of some water quality parameters has been seen in past water quality papers (Ritchie *et al.*, 2003).

### 2.6.1 Suspended matter

Suspended matter comprises organic and inorganic matter. It plays an essential role in water quality management and is related to the fluxes of heavy metals and micro-pollutants. Previous studies have shown the relationship between suspended sediment and spectral reflectance. Suspended sediments increase the radiance from surface water in visible and near-infrared bands (Ritchie and Cooper, 2001). In situ and laboratory measurements have also shown that the surface water radiance is affected by the type of sediments, texture, colour, sensor view and sun angle, as well as water depth (Usali and Ismail, 2010). Remote sensing techniques can be used to estimate and map the concentrations of suspended matter in inland waters, with the aid of both spatial and temporal information. Remote sensing studies for suspended matter have been carried out using several satellite platforms such as Landsat, Sentinel-2, MODIS and Worldview-2 (Miller and McKee, 2004). These studies have shown a good relationship between suspended matter and radiance or reflectance from either single band or a combination of some bands in satellite or airborne platforms. Results from in situ studies have shown that wavelengths between 700 and 800 nm are the most useful for determining the suspended matter in surface water (Ritchie *et al.*, 1987). The absorption of radiance by suspended sediment is generally much smaller than that of chlorophyll, but the scattering is much higher. Regarding optical wavelength, remote sensing of terrestrial environments is more extensive compared to remote sensing of water bodies, which is restricted to a relatively narrow range. Therefore, the range of 400 to 850 nm is specifically used for water quality parameters. Investigations from several remote sensing studies have found that thematic mapper (TM4), a Landsat product, has a good relationship with total suspended matter (Dekker *et al.*, 2002).

### 2.6.2 Phytoplankton

In many freshwater bodies, eutrophication is quantified regarding the concentration of chlorophyll contained in the algal plankton cells. Chlorophyll-a is one of the photosynthetic pigments that contribute to the colour of water. Many studies use remote sensing for mapping chlorophyll-a, which is an indicator of algal concentration and a key parameter in water quality assessment (Schalles *et al.*, 1998). Most inland waters (lakes, rivers, reservoirs or dams) usually have a high range of chlorophyll-a concentrations as a result of eutrophication. Knowledge of the amount of phytoplankton is vital in understanding the extent of eutrophication in these water bodies. Considerable uncertainty in the detection of chlorophyll during cyanobacterial blooms is a result of the short period involved in the biological process,

which becomes difficult for satellite sensors to detect as their temporal resolution are longer. Despite this shortcoming, success in the quantitative mapping of chlorophyll-a and algal blooms has been achieved using remote sensing especially with high temporal resolutions sensors (Dekker, 1993). Multispectral sensors like IKONOS and MODIS (Siegel and Gerth, 2000) and synthetic aperture radars (Svejkovsky and Shandley 2001) have also been used to map the extent of cyanobacterial blooms. Remote sensing has also been used to measure chlorophyll-a concentration and patterns (Ritchie and Charles, 1996). Most of these studies are based on empirical relationships between radiance in narrow bands or band ratios and chlorophyll-a concentrations. Measurements from aircraft, satellite sensors like Landsat, SPOT, Sentinel-2, MODIS and Worldview-2 use a variety of algorithms and wavelengths to map chlorophyll-a of the oceans, estuaries and fresh water. Suitable techniques for estimating chlorophyll-a pigments in water bodies are band rationing and regression (Dekker *et al.*, 2002).

Since this present study was based on chlorophyll-a and turbidity estimation and mapping, Table 1 shows chlorophyll-a concentrations and trophic conditions of water bodies, to determine the amount of chlorophyll-a levels which can be used to explain the trophic state of chlorophyll-a in any given water body.

Table 1: Relationship between chlorophyll-a (ug/L) concentrations and trophic condition of lakes and reservoirs (Adapted from Boyd et al., 1997)

Minimum (Ug/L)	Maximum (Ug/L)	Trophic state
≤ 2	≤ 5	Oligotrophic, water quality status is pleasing with minimal concentrations of phytoplankton.
2-5	6-17	Mesotrophic, there is the presence of some algae, aesthetic nature of water reduced and some oxygen depletion.
6-17	17-45	Mesotrophic, reasonable amount of algae growth, poor water colour, oxygen depletion.
>20	>45	Eutrophic, a large quantity of phytoplankton production, severe oxygen depletion in bottom waters, and reduction in other uses.

### 2.6.3 Turbidity

Water turbidity is an expression of the optical properties of water, which causes light to be scattered and absorbed rather than transmitted in straight lines. It is therefore commonly regarded as the opposite of clarity. Turbidity measurements have often been used in

calculating fluvial suspended sediment concentrations (Wass *et al.*, 1997). The best correlation of turbidity from in situ measurements and reflectance from satellite images is in the red band of the electromagnetic spectrum (Lathrop and Lillesand, 1986). Other sensors such as IKONOS demonstrate the usability of high-resolution satellite data for mapping turbidity in freshwater bodies (Hellweger *et al.*, 2007) and its higher resolution is a potential solution to the limitations of spatial scale. Liversedge, (2007) has also conducted turbidity studies in marginal ice lakes at the Bering Glacier, Alaska, where simple and multiple linear regression analyses were performed using different Landsat 7 ETM-F bands to determine the best predictors of turbidity. The final algorithm utilised in the study was Landsat 7 ETM+ Band 3 (red portion of the electromagnetic spectrum) and Band 4 (a near-infrared part of the electromagnetic spectrum) to predict turbidity levels. In this study, turbidity maps created using the algorithm mentioned above can be used to help determine inter- and intra-annual sediment dynamics. The turbidity maps can also show where floods discharge sediments into the water body (Miller *et al.*, 2004).

#### **2.6.4 Dissolved organic matter**

Dissolved organic matter affects water colour. Dissolved organic matter refers to all soluble natural substances (that can pass a 0.45  $\mu\text{m}$  filter) which contribute to the absorption of light at specific wavelengths. Coloured Dissolved Organic Matter (CDOM) mostly consists of naturally occurring constituents which are water soluble, biogenic, heterogeneous organic substances that are yellow to brown (Aiken *et al.*, 1996). CDOM is the fraction of dissolved organic matter that absorbs light in both ultraviolet and visible ranges. Dissolved organic matter also affects the volume of reflectance in a spectrum, and almost exclusively at shorter wavelengths (Bukata *et al.*, 1991). CDOM absorbs visible light especially below  $\sim 500 \mu\text{m}$ , and there is an exponential increase in its absorbance with decreased wavelengths. According to Strömbeck, (2001), the absorbance of red light can be significant at high CDOM concentrations. Also, an increase in the level of CDOM absorption can affect primary productivity and ecosystem structure in freshwater bodies by reducing the amount and quality of photosynthetically-active radiation to phytoplankton (Del Castillo *et al.*, 1999).

### **2.7 Challenges and opportunities of remote sensing for water quality management**

Most water bodies exhibit complex optical properties which do not result from phytoplankton, and this makes the measurements of these water parameters difficult using remote sensing applications. This is because satellite sensors with low spatial and spectral

properties cannot detect such water parameters (Morel and Prieur, 1977). Furthermore, satellite sensors sometimes record spectral signatures from the floor of water bodies if they are shallow. This increases the backscattered signal values recorded by the sensor. Thus, models used in detecting water variables in large water bodies cannot be used in small water bodies (O'Reilly *et al.*, 1998).

In estimating water quality variables, there is a need to understand backscattered signals from the bottom of the water body and the variables present in it. It is also essential to know how spectral signatures measured by the sensor are affected by backscattering coefficients of the water column. Moreover, the surroundings of water bodies sometimes may affect the measurements of spectral signatures by sensors, thus giving wrong results of water variable measurements (O'Reilly *et al.*, 2000). Accurate measurements of water variables from remote sensing, therefore, determine the performance of the regression model used, especially when it coincides with field measurements. Water variables like chlorophyll-a and turbidity can easily be estimated with suitable regression models which can pick out relevant spectral signatures from different inherent optical properties (IOPs) in water bodies (Olmanson *et al.*, 2013).

Atmospheric correction models are significant in eliminating atmospheric effects like aerosols, haze and sky glint which affect remotely sensed data obtained from water surfaces (Wang, 2004). The correction can be done by calibrating the remotely sensed data through the conversion of digital numbers (DN) into radiance values (Hadjimitsis, 1999).

## **2.8 Comparison and integration of Landsat 8/OLI and Sentinel-2A/MSI data for mapping and monitoring water quality at the Vaal Dam**

The biospheres of aquatic environments have been monitored by earth observation sensors which provide a general overview of key water quality variables at the high temporal frequency. The multi-spectral instrument of Sentinel-2 has 13 spectral bands: four of them at 10 m, six at 20 m, and three bands at 60 m spatial resolution. The twin satellites of Sentinel-2 offer data with a 5-day revisit time. Thus, biochemical processes over short time frames can be monitored by Sentinel-2 satellite, whereas Landsat 8 with a revisit time of 16 days cannot do this. In turbid coastal waters, Landsat 8/OLI can mainly be used for mapping of suspended sediments and turbidity. The Multispectral Imager on Sentinel-2 has a similar spatial resolution (10-60 m), but an increased spectral and temporal coverage and additional red-edge bands at 20 m spatial resolution, which detects a chlorophyll-a absorption peak at

around 665 nm (Toming *et al.*, 2016). Furthermore, the launch of Landsat-8 in February 2013 ensures the continuous stream of satellite data which is essential for monitoring and is comparable to other satellite sensors regarding spatial and spectral coverage of land cover (Irons *et al.*, 2012).

Satellite sensors for Landsat 8 and Sentinel-2 operate at different spectral, spatial and temporal resolutions (Roy *et al.*, 2014; Hagolle *et al.*, 2015). Thus, this may be seen as a problem when different band placements and sensitivities for specific wavelengths give different results for optically active constituents (OAC) concentrations for the same water body. Also, the improvement of the width of spectral bands for both Landsat 8 and Sentinel-2 leads to stronger correlations of OACs of satellite data and field measurements, and additional bands in Sentinel-2 allow for a greater range of OACs to be retrieved (Dekker, 1993). Due to the high spectral resolution of both sensors, they will be able to differentiate different OACs in water bodies and as well as likely derive a high accuracy of chlorophyll-a concentrations and turbidity when correlated with field measurements (Dekker, 1993), and this has been shown to improve the success of water quality assessment (Aurin and Dierssen, 2012). Sentinel-2 provides good spectral and temporal coverage of water bodies, while Landsat 8 provides a good spatial coverage, and hence the need for integrating both satellites with field data to allow for useful estimation and mapping of chlorophyll-a and turbidity concentrations of an entire water body.

## **2.9 Regression methods used for water quality**

Regression methods have been widely used for water quality monitoring and mapping in lakes and dams. Random forest and multiple stepwise regression analysis are most often used to demonstrate relationships between the spectral reflectance of satellite data and water quality parameters collected in the field (Khattab & Merkel 2014). The main reason for using regression analysis for water quality is to select an appropriate regression method and independent variables that result in a high  $R^2$  value. An  $R^2$  value approaching 1 indicates a good correlation between the predicted and measured data (Wang *et al.* 2006). In most water quality studies, regression analysis is performed using in situ estimated values of water quality parameters and mean reflectance values of remotely sensed data (Dewidar & Khedr 2005; Somvanshi *et al.*, 2012). Thus, in this study, the two regression models mentioned above were used in predicting chlorophyll-a and turbidity concentrations by cross-validating measurements of spectral reflectance for chlorophyll-a and turbidity from satellite data with that from field measurements. Furthermore, both regression models were used to execute

both correlation and regression analysis to investigate and study the relationship between in situ measurements and remotely sensed data of the study area based on their R<sup>2</sup> values (Wang et al. 2006).

## CHAPTER 3: Materials and methods

### 3.1 Study area

The Vaal Dam is located some 56 km south of Johannesburg, close to Vereeniging (Figure 1). The Vaal dam is one of the major dams in South Africa, and it is of great significance to the public and its surrounding environments. It was constructed in the 1930s by Rand Water and the Department of Irrigation. The Witwatersrand region obtains water supplies from the dam, and it can hold 2575 million m<sup>3</sup> of water from inflow from an external source (DWAF, 2008). The dam has a surface area of 321 km<sup>2</sup> with a maximum depth of 22.5 m. The dam is seen to be the most valuable dam as it is the leading supplier of fresh water to the most densely populated region in South Africa (DWAF, 2003). Due to increased demand for water supply, there has been the construction of the highland water project in Lesotho to increase the volume of water at the dam. This will enable the dam to meet its high demands for water supply within Gauteng province (DWAF, 2008).

There are three leading causes of water pollutants identified in the Vaal dam, namely agricultural practices, mining activities, and sewage from settlements within the catchment (Braune and Rogers, 1987). Contaminants include pesticides and fertilisers in agriculture, waste from acid mine drainage, and sewage from toilets and septic tanks. These are then deposited into the dam and can cause an increase in the concentration of ions (nitrates and phosphates), suspended matter which increases turbidity, and promotes excessive growth of algae which increases the chlorophyll content. Rand Water has monitored water quality status at the Vaal Dam which is a voluntary programme, that monitors water quality for the end consumer which has been running since March 1998 (DWAF, 2008). The monitoring program involves the measurements of water quality parameters such as Dissolved Oxygen (DO), chlorophyll-a pigments, electrical conductivity (EC), salinity, temperature, turbidity, and ion concentrations including phosphates (PO<sub>4</sub>), nitrates (NO<sub>3</sub>) and pH. These measurements are obtained through lab analysis of water samples collected at different field sampling locations. In this project, a total of 23 sampling points were used to estimate and map chlorophyll-a and turbidity concentrations to locate areas that are not polluted, slightly polluted and heavily polluted.

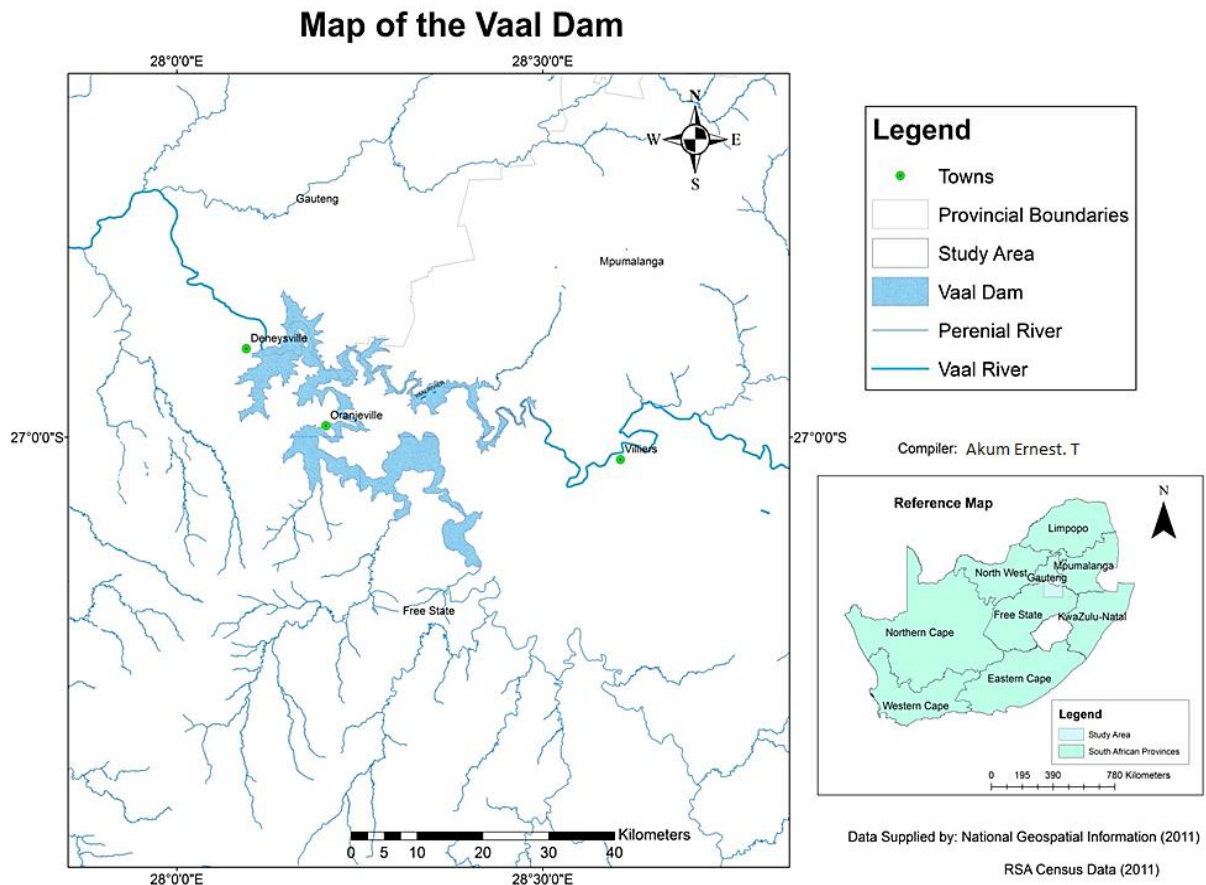


Figure 1: Map of the Vaal dam showing the study area

## 3.2 Methodology

### 3.2.1 Remote sensing data acquisition and pre-processing

In this project, Landsat-8 and Sentinel-2 images were used to estimate and map the spatial patterns of chlorophyll-a and turbidity concentrations at the dam.

#### 3.2.1.1 Landsat 8 Data acquisition and pre-processing

Landsat 8 OLI operational land images (OLI) were downloaded from <https://earthexplorer.usgs.gov> where the satellite overpassed the dam on 04/10/2016. Images obtained on or close to the dates of field data collection were covered with clouds and could not be used for this study. The only cloud-free image was acquired on 04/10/2016 which had a 22 days' time difference with the field data. Landsat 8 consists of the operational land imager (OLI) and thermal infrared sensors (TIRS). The OLI sensor uses nine spectral bands in the visible near-infrared (VNIR) and shortwave infrared (SWIR) regions to sense the entire

earth. There is also an addition of two new deep blue spectral bands which are bands 1 and nine that detect aerosols, clouds and haze. Radiometric atmospheric corrections were performed on the satellite imagery to eliminate aerosol, water vapour and clouds, and digital numbers (DN) were converted to top of atmosphere (TOA) reflectance using radiometric calibration with effective bandwidths for each band. Atmospheric correction was then carried out on the radiance image using the FLAASH model to derive an aerial corrected image from which reflectance values were then extracted using the ArcGIS software (Roy *et al.*, 2014). The spectral bands of Landsat 8 and their corresponding wavelengths can be seen in Table 2.

Table 2: The spectral bands of Landsat 8 sensor and its corresponding wavelengths (from Roy *et al.*, 2014) (<https://landsat.gsfc.nasa.gov/landsat-8/landsat-8-overview/>)

Band number	Sensor type	Wavelength ( $\mu\text{m}$ )	Band centre (nm)	Spatial resolution (m)
Band 1- coastal aerosol	OLI	0.43- 0.45	443	30
Band 2- Blue	OLI	0.45- 0.51	482	30
Band 3- Green	OLI	0.53- 0.59	562	30
Band 4- Red	OLI	0.64- 0.67	655	30
Band 5- Near infrared (NIR)	OLI	0.85- 0.88	865	30
Band 6- SWIR 1	OLI	1.57- 1.65	1610	30
Band 7- SWIR 2	OLI	2.11- 2.29	2200	30
Band 8- Panchromatic	OLI	0.50- 0.68	590	15
Band 9- Cirrus	OLI	1.36- 1.38	1372	30
Band 10- Thermal infrared (TIRS)	TIRS 1	10.60- 11.19	10800	100
Band 11- Thermal infrared (TIRS)	TIRS 2	11.50- 12.51	12000	100

### 3.2.1.2 Sentinel-2 Data pre-processing and atmospheric correction

Sentinel-2 Level-1C (L1C) multispectral imagery (MSI) data were acquired from Sentinel's Scientific Data Hub (<https://scihub.copernicus.eu/>). Sentinel-2 images acquisition date was 13/11/2016. The Sentinel-2 Level-1C product comprises of 100 km<sup>2</sup> tiles (ortho-images in UTM/WGS84 projection). The TOA reflectance for all parameters was transformed into radiance by applying radiometric measurements per pixel of the remotely sensed data. The Sentinel-2 Level-1C image was then resampled with the aid of Ground Sampling Distance (GSD) of 10 m, 20 m and 60 m resolution of the different spectral bands. The image used was geometrically corrected to WGS84, UTM, zone 35S at 10 m and 20 m resolution.

Atmospheric correction for the Sentinel-2 Level-1C image was then performed using Sentinel-2 Toolbox (S2TBX) of the Sentinel Application Platform (SNAP). For each sampling location at the dam, 3 m x 3 m cloud-free pixels were extracted, and the mean values of these pixels were used for analysis. The Sen2cor atmospheric correction model was then finally applied on the Sentinel-2 Level-1C image to obtain Sentinel-2 Level 2A (L2A) Bottom of Atmosphere (BOA) reflectance image. Table 3 shows the spectral properties of the Sentinel-2 MSI sensor used in this study.

Table 3: Sentinel-2 sensor with its spectral properties (<https://www.satimagingcorp.com/satellite-sensors/other-satellite-sensors/sentinel-2a/>)

Band number	Central wavelength (µm)	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	30	60
11 SWIR	1.610	90	20
12 SWIR	2.190	180	20

Table 4: The sensors used for this study and a comparison of their spectral properties

Satellite	Landsat 8	Sentinel-2
Satellite sensor systems	OLI/TIRS	MSI
Spatial resolution (m)	30	10, 20, 60
Number of spectral bands	11	12
Revisit cycle (days)	16	5
Swath width (km)	185	290
Launch date	Feb 2015	June 2015
Years in orbit/minimum design life (years)	2/5	0/7
Spectral resolution (µm)	0.85-0.88	2.3 nm spectral resolution in the 400–1000 nm range

### 3.2.2 Field measurements

The Vaal dam was sampled by the Agricultural Research Council (ARC) and a water quality assessment team from Japan in October 2016. The field measurements for water quality parameters at the Vaal dam was sampled for three days, and the locations where the samples were collected are shown in Figure 2. In situ data were obtained on 26-28 October 2016 using a Hydrolab MS5 – Multiparameter Mini Sonde sensor. This device can measure up to 10 parameters simultaneously (<http://www.ott.com/en-us/products/water-quality-2/hydrolab-ms5multiparameter-mini-sonde-57/>).

Field measurements were collected from 23 sampling locations selected as random points in a polygon to represent the Vaal dam, Vaal River input and Wilger River input, which the two main rivers feeding the dam. At every sampling location, measurements were taken from the surface. A hand-held Trimble GPS receiver simultaneously determined sample positions. The field data were collected covering the 23 representative sites in the different parts of the Vaal Dam. Water samples for analysis were collected in cleaned polyethene bottles rinsed with distilled water. At each sample site, a 100 ml of water sample was collected in containers rinsed twice with the water to be sampled before filling, and the samples were then taken to the lab for analysis.

Different water quality parameters were analysed from these samples (chlorophyll-a, turbidity, dissolved oxygen, electrical conductivity, pH, temperature, Reduction-oxidation (ORP) and salinity) although only chlorophyll-a and turbidity are of concern in this study. Values for chlorophyll-a and turbidity concentrations analysed from the water samples collected in the polyethene bottles from the 23 sampling sites were used in the prediction of chlorophyll-a and turbidity concentrations from the remotely sensed data. Table 5 shows field measurements for the different water quality parameters that were analysed from the 23 sampling points at the dam.

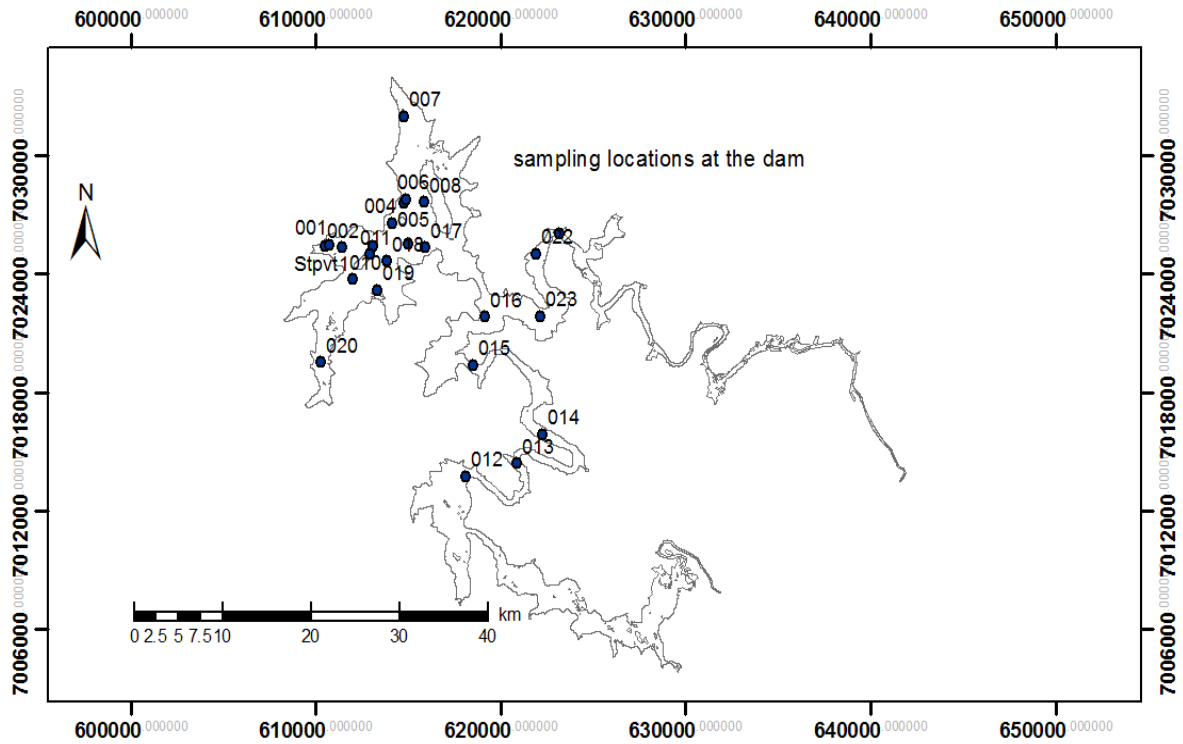


Figure 2: Map showing sampling points 1-23 at the dam

Table 5: Water quality parameters from each sampling point at the dam as seen in Figure 2

Sampling Site #	Temp (°C)	EC (us/cm)	DO (mg/l)	Chll-a (Ug/l)	Turb (NTU)	pH	ORP (mV)	Salin (ppt)
1	19.5	125.6	7.5	3.2	768	7.6	28.1	0.04
2	19.5	125.7	7.6	2.9	1085	7.6	100.2	0.03
3	19.7	125.3	7.6	3.2	700.5	7.5	94.5	0.04
4	20.1	124.2	7.6	2.9	714.8	7.6	76.0	0.03
5	20.7	123.4	7.7	2.9	773.5	7.6	105.1	0.04
6	21.1	123.1	7.7	3.5	867.1	7.6	83.6	0.04
7	20.3	119.9	7.7	3.3	696.5	7.7	97.8	0.03
8	21.1	117.5	7.8	4.1	797.6	7.7	117.2	0.05
9	20.8	119.4	7.7	3	733.1	7.6	104.6	0.03
10	20.2	122	7.6	4	723.6	7.7	158.2	0.05
11	19.6	120.8	7.6	3.5	777.1	7.6	110.5	0.04
12	21.7	87.5	7.4	6.2	1550.9	8.1	92.7	0.07
13	22.8	88.9	7.6	6.9	1506	7.7	117.5	0.08
14	23.8	88.8	7.5	6	1621.3	7.5	149.4	0.07
15	24	90.7	7.8	6.5	1371.9	7.8	141.8	0.08
16	24.2	108.1	8.7	12.4	1036.1	8.2	119.7	0.15
17	23.3	116.6	7.8	3.1	747.1	8.1	104	0.04
18	24.2	119.2	8.4	4	822.9	8	104	0.05
19	22.8	120.3	7.9	4.3	752.3	8	77.4	0.05
20	24.5	124.4	9	13.3	1328	8.6	96.1	0.14
21	23.4	164.3	8.4	13.3	750	8.4	64.5	0.16
22	24	138.8	8.6	12.	913.3	8.4	99.9	0.16
23	23.5	117	7.8	11.8	943.6	8.2	103.0	0.14

### **3.2.3. Regression modelling**

In this study, two regression models (stepwise multiple linear regression, and random forest regression) were used to predict and map chlorophyll-a and turbidity concentrations (Rawlings, 1988; Neter et al., 1989). This was done by cross-validating measurements of spectral reflectance for chlorophyll-a and turbidity from satellite data with that from field measurements. The justification for using random forest and multiple stepwise linear regression models was because they have been used in other water quality studies (Agirre-Basurko et al., 2006; Kovdjenko et al., 2010; Ferraro and Giordani, 2012) and they gave good performance and high accuracy when comparing remotely sensed and field data. They are also found to be more robust and reliable, meaning they can handle many input variables at the same time. This implies that the regression models allow each independent variables (spectral values) of the remote sensing data to predict the dependent variables (chlorophyll-a and turbidity concentrations) (Vincent et al., 2004).

The in situ chlorophyll-a and turbidity values from the 23 sampling points across the dam and reflectance values from both remote sensing data were divided into 70% training and 30% testing datasets. Both regression methods gave seven predictive models for chlorophyll-a and turbidity concentrations for both satellite data with in situ data, which were then used to create linear regression plots in Excel. These 7 points were used as a percentage of the 23 points for testing of the models' performance and accuracy for both satellite data (Grimm et al., 2008).

#### **3.2.3.1. Random forest model**

Random forest is a machine learning algorithm which aggregates numerous decision trees to obtain a consensus prediction of the response categories (Breiman, 2001). When using the random forest regression model, two main parameters must be put in place; the number of trees to be used in the sample (Ntree) and the number of variables to be taken for every try (Mtry). Multiple bootstrap samples from the training data create multiple regression trees which were recursively partitioned according to a given random subset of predictor variables, and a predetermined number of decision trees were developed. With each new tree, the sample data subset was randomly selected, and with each new split, the subsets of predictor variables were randomly selected. Each tree was structured with a random subset of band predictors to determine the best bands for predicting chlorophyll-a and turbidity concentrations (Breiman, 2001). Results obtained from all predictive bands were averaged to

get overall prediction accuracy, and the random forest reduced bias and variance to obtain a higher accuracy (Grimm et al., 2008). The number of observations used in building the model was 23, and the number of trees used was 500 while the variables tried at each split was 3. The data was set as ‘in bag’ data (total number of samples that will be drawn back from the entire sample size) and the excluded data were set as ‘out-of-bag’ data (OOB) (Breiman, 1996). For every regression tree, the model calculated mean square error which is the difference in prediction of using OOB and in bag data (Palmer et al., 2007). The model then predicted chlorophyll-a and turbidity concentrations through cross-validation of remote sensed and field data, by dividing the datasets into training and testing datasets. The in bag samples were used to do the training and the OOB to do the accuracy assessment. The training dataset was then used to evaluate the accuracy of the model, which gives an optimistic estimate of the performance of the model based on the R-squared values obtained for both chlorophyll-a and turbidity.

### ***3.2.3.2 Multiple stepwise linear models***

Multiple stepwise linear regression is a method of regressing multiple variables while simultaneously removing the weakest correlated variable (Coelho-Barros et al., 2008). This method was used in studying the relationship between the dependent variables chlorophyll-a and turbidity concentrations and the independent variable of reflectance values from remote sensing data (Coelho-Barros et al., 2008). The formula of multiple stepwise linear regression is given by:

$$Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \dots + \beta_k X_{kt} + \epsilon_t, t = 1, 2, \dots, N$$

Where Y is the dependent variable, X<sub>1</sub>, X<sub>2</sub>, ..., X<sub>k</sub> are the independent or explanatory variables, t index the number of observations in the sample N, and the term  $\epsilon$  is a random error term.

The stepwise multiple linear regression model was used to estimate chlorophyll-a and turbidity concentrations from remotely sensed data using single band input and non-reciprocal reflectance band ratios. That is using single reflectance values from each band width and at the same time removing the weakest reflectance band values to create a new channel of data (Eregno, 2013). The regression model was then calibrated and cross-validated using coincident pairs from the satellite images and field data. The best single band and band

ratio models were then applied to the entire satellite scene to estimate chlorophyll and turbidity concentrations for the whole dam.

### 3.2.4 Accuracy assessment of models

Accuracy assessment is a measurement of the total number of predictions from band reflectance of satellite data when correlated with the observed chlorophyll-a and turbidity from field measurements. This was done by obtaining R-squared and root mean square error (RMSE) values from both Landsat 8 and Sentinel-2 data through correlation with field measurements (Hung and Lin, 2006). The R-squared values obtained were then used to evaluate the performance and accuracy of the regression models. Predicted chlorophyll-a and turbidity concentrations were then used to map spatial patterns across the dam.

The RMSE is a measure of the average level of error. Its values range from 0 to  $\infty$ . RMSE with smaller values shows a good fit of a model. RMSE is good for measuring how accurate a model predicts a variable from the observed values (Khattab and Merkel, 2014, El Saadi et al., 2014, Yüzügüllü and Aksoy, 2011, Coskun et al., 2008). The RMSE can be obtained using the formula

$$RMSE = \sqrt{\frac{\sum_{i=1}^n [\log(x_i^{\text{estimated}}) - \log(x_i^{\text{measured}})]^2}{n-2}}$$

Where  $x_i^{\text{measured}}$  represents measured field chlorophyll-a concentrations and turbidity levels at the validation sites and  $x_i^{\text{estimated}}$  is the model estimated chlorophyll-a concentrations and turbidity levels from satellite band reflectance and n is the number of validation points.

An R-squared value (R<sup>2</sup>) tells if a model has a good fit which helps in understanding the proximity of data points with a regression line, and ranges from 0 to 1. The higher the R<sup>2</sup> value, the better the model fits the data (Coskun et al., 2008).

An average y value associated with a given x value can be predicted using a regression line. RMSE values obtained were used to indicate if there was a small or wide variation between predicted and measured values. This was to check if the use of remote sensing application is conducive for estimating and mapping chlorophyll-a and turbidity concentrations for water quality assessment (El Saadi et al., 2014).

## CHAPTER 4: Results

### 4.1. Descriptive statistics

Table 6 shows descriptive statistics for measured chlorophyll-a and turbidity concentrations in the field.

Table 6: Statistical results for chlorophyll-a and turbidity concentrations from field data

Statistics	Chlorophyll-a (Ug/L)	Turbidity (NTU)
Maximum	13.386	1621.31
Minimum	2.9210	696.59
Mean	5.90	955.71
Standard deviation	3.80	302.17
Sample size	23	23

### 4.2 Comparison of in situ measured and predicted chlorophyll-a and turbidity values

Graphs comparing in situ measured and predicted chlorophyll-a and turbidity concentrations are presented in Figures 3 to 10. The data points show a good fit along the regression line. A poor fit may reflect the limited number of sampling points used for satellite data prediction for the entire dam. Furthermore, remote sensing data not coinciding with field data collection dates as a result of satellite data that overpassed the dam on field data collection dates were affected by cloud cover rendering it not fit to be used for this study. Below are Graphs 3- 10 showing the comparison of in situ chlorophyll-a and turbidity concentrations predicted from band values from satellite data for both sensors through statistical analysis using random forest and multiple linear regression methods.

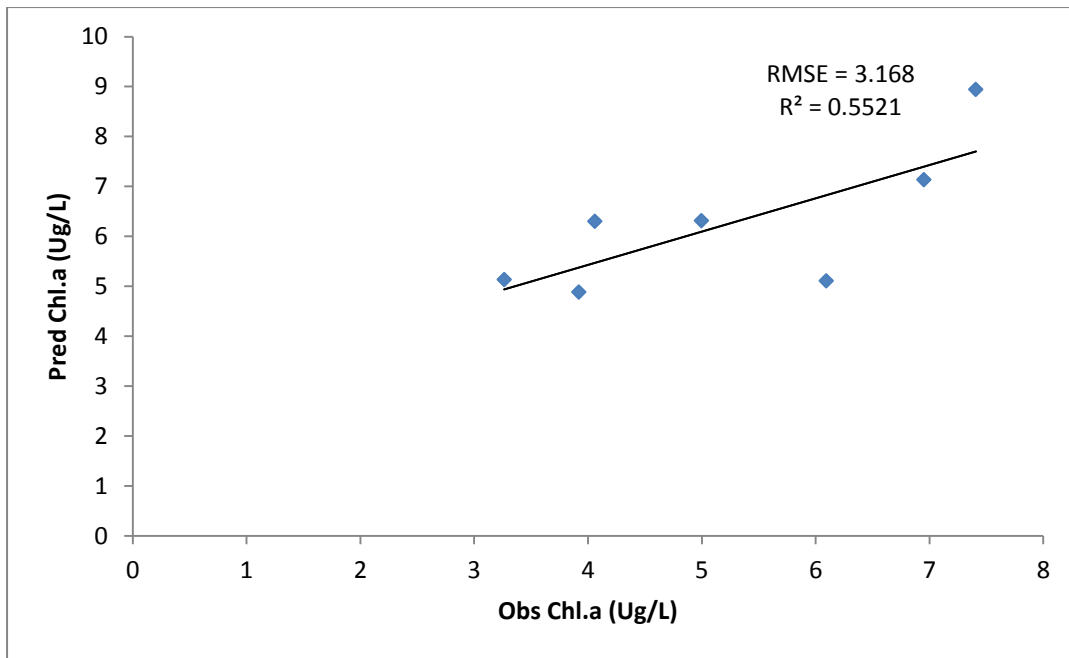


Figure 3: Graph showing predicted vs observed chlorophyll-a concentrations with Landsat 8 in evaluating random forest regression performance

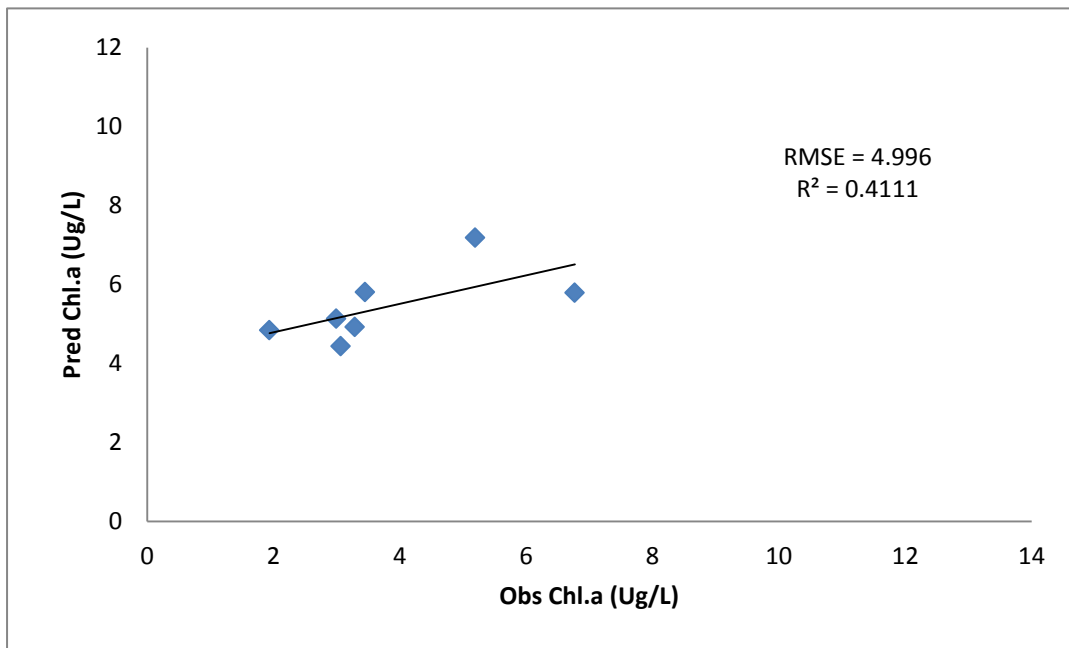


Figure 4: Graph showing predicted vs observed chlorophyll-a concentrations with Landsat 8 data in evaluating stepwise multiple linear regression performance

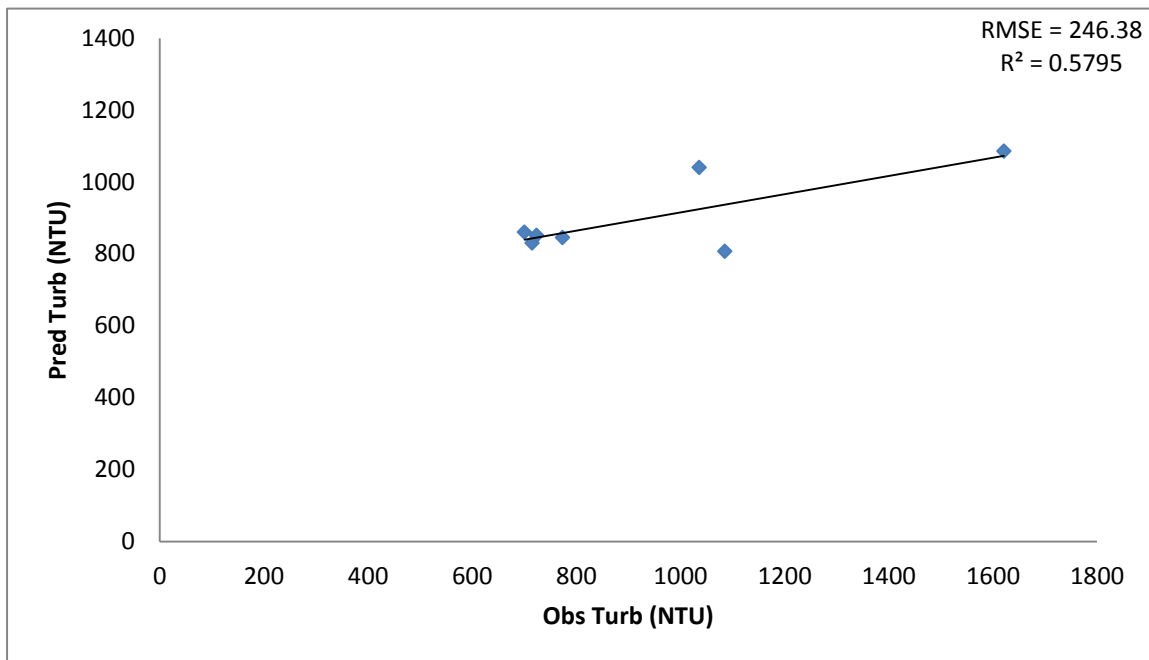


Figure 5: Graph showing predicted vs observed turbidity concentrations with Landsat 8 data in evaluating random forest regression performance

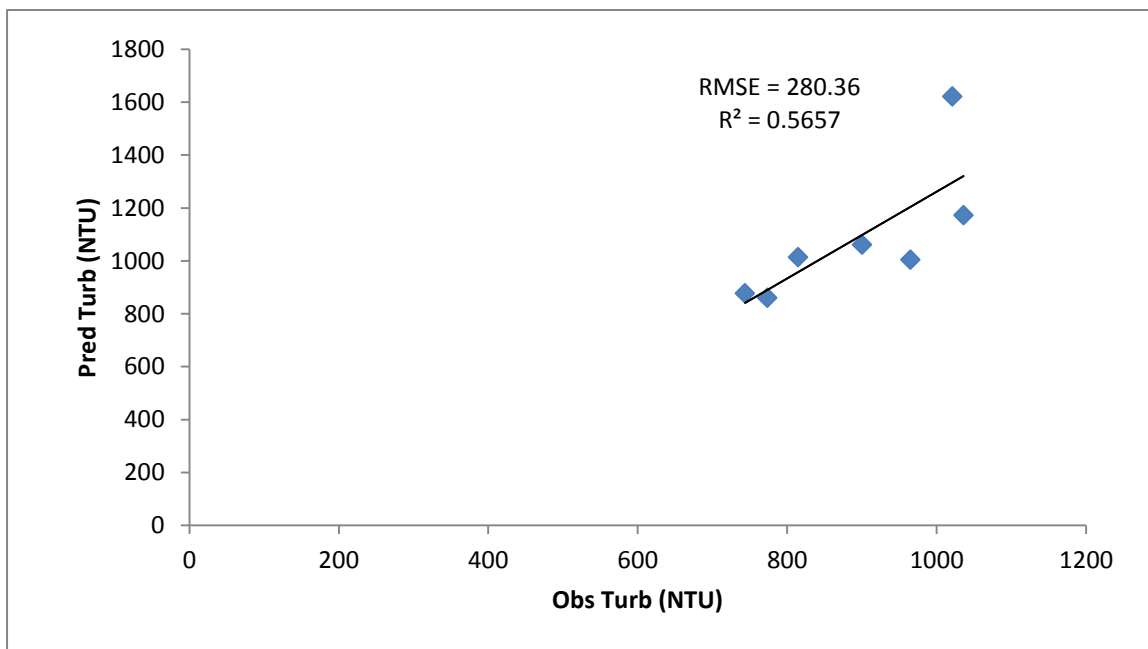


Figure 6: Graph showing predicted vs observed turbidity concentrations with Landsat 8 data in evaluating stepwise multiple linear regression performance

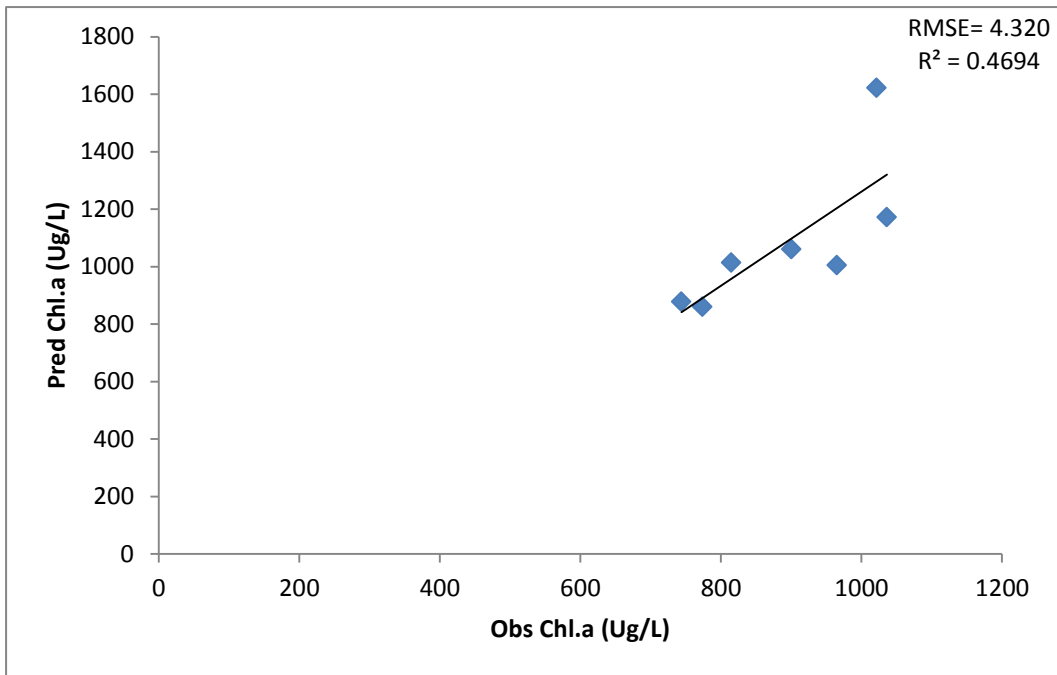


Figure 7: Graph showing predicted vs observed chlorophyll-a concentrations with Sentinel-2 data in evaluating stepwise multiple linear regression performance

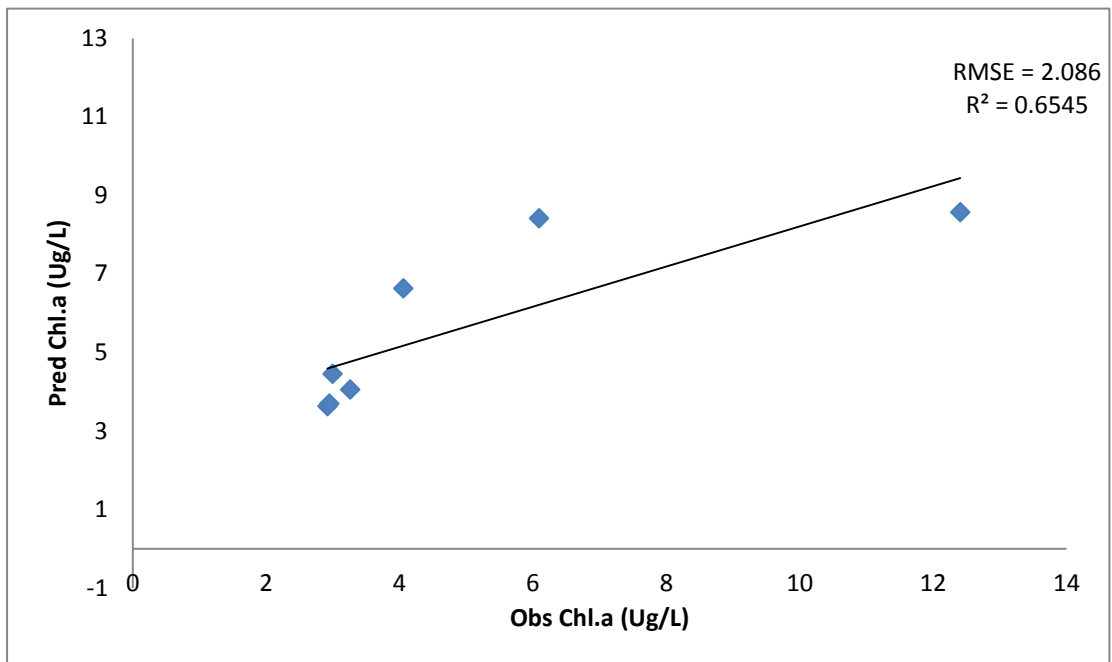


Figure 8: Graph showing predicted vs observed chlorophyll-a concentrations with Sentinel-2 data in evaluating random forests regression performance

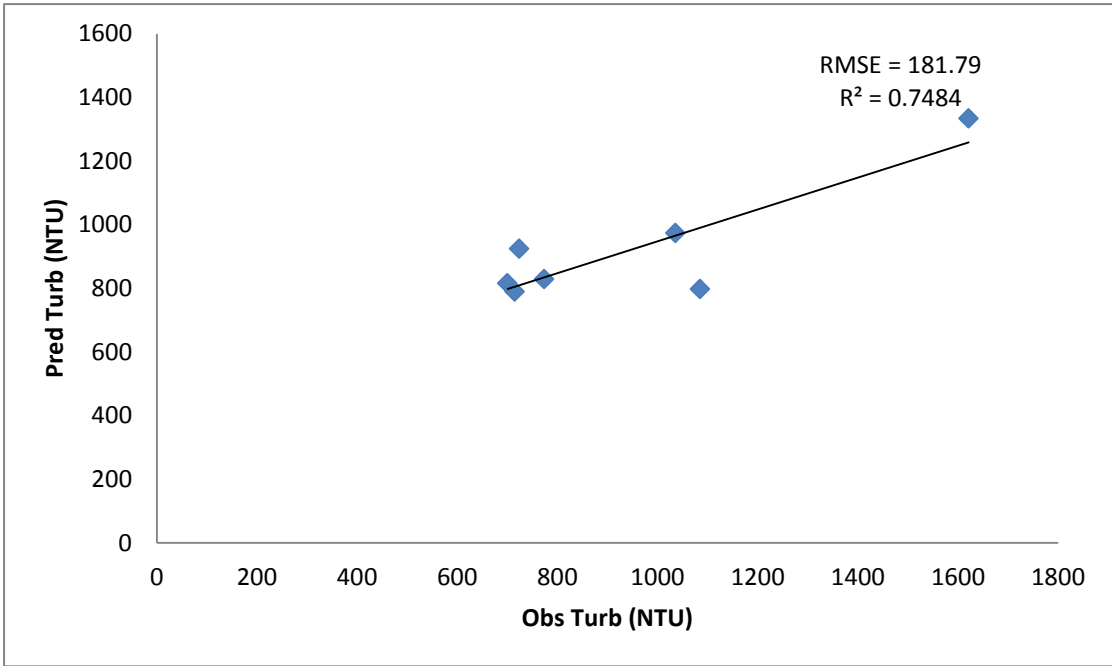


Figure 9: Graph showing predicted vs observed turbidity concentrations with Sentinel-2 data in evaluating random forest regression performance

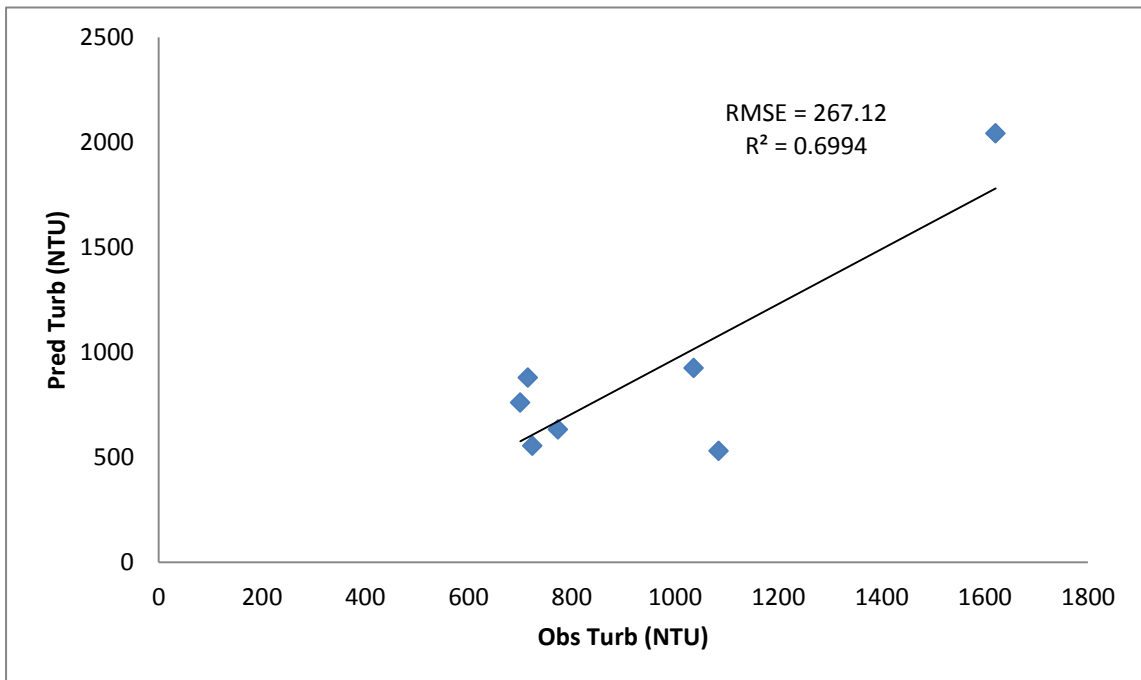


Figure 10: Graph showing predicted vs observed turbidity concentrations with Sentinel-2 data in evaluating stepwise multiple linear regression performance

### 4.2.1 Predicting chlorophyll-a using different regression models from remotely sensed data

Random forest and stepwise linear regression models were used to compare in situ measured and predicted chlorophyll-a from Landsat 8 data. The R-squared values and RMSE are presented in Table 7.

Table 7: Performance of random forest regression and stepwise linear regression models for prediction of chlorophyll-a concentrations using reflectance values from satellite data

Remote sensing dataset	Random forest regression			Stepwise regression		
	R square	RMSE	p-value	R square	RMSE	p-value
Sentinel-2	0.6545	2.086	0.02755	0.4694	4.320	0.08113
Landsat-8	0.5521	3.168	0.2085	0.4111	4.996	0.0007525

### 4.2.2 Predicting turbidity using different regression models from remotely sensed data

Random forest and stepwise linear regression model were also used in comparing field measured and predicted turbidity with Landsat 8 data to see the performance of the models. Results are presented in Table 8.

Table 8: Performance of random forest regression and stepwise linear regression models for prediction of turbidity levels using band reflectance values from satellite data

Remote sensing dataset	Random forest regression			Stepwise regression		
	R square	RMSE	p-value	R square	RMSE	p-value
Sentinel-2	0.7484	181.79	0.01192	0.6994	267.12	0.0002417
Landsat-8	0.5795	246.38	0.04683	0.5657	280.36	0.1019

## 4.3 Accuracy assessment

The prediction of chlorophyll-a and turbidity concentrations from field data using satellite data showed the performance and accuracy assessment of random forest and stepwise linear regression models. The random forest regression model showed a higher performance and accuracy compared to stepwise multiple linear regression based on their R-squared and RMSE values. Thus the random forests model showed a high correlation between predicted and observed chlorophyll-a and turbidity concentrations over the stepwise multiple linear regression model.

#### 4.4 Mapping spatial patterns for chlorophyll-a and turbidity at the dam

The best performing model (random forest regression) was used to derive equations used in mapping chlorophyll-a and turbidity concentrations at the dam. The empirical approach was able to quantify chlorophyll-a and turbidity concentrations by using bands with high coefficients derived from both Landsat 8 and Sentinel-2 images. This was done by deriving estimates for chlorophyll-a and turbidity concentrations. The highest estimated values were selected and multiplied with their corresponding bands from the two remotely sensed data and their respective intercepts (the expected mean values for chlorophyll-a and turbidity concentrations) from coefficients derived from the regression model used. An equation was then formed using the raster calculator in ArcGIS to map spatial patterns of chlorophyll-a and turbidity concentrations for the entire dam using the remotely sensed data shown in Table 9.

In mapping, the spatial patterns for chlorophyll-a and turbidity concentrations at the dam, the equations in Table 9 were applied to the Landsat 8 OLI and Sentinel-2A MSI image to derive a map showing the spatial patterns for chlorophyll-a and turbidity concentrations. Mapped patterns exhibit a good correlation with field measured chlorophyll-a and turbidity. This can be seen in Figures 11 to 14. From the spatial patterns, it can be observed that there are spatial variations in chlorophyll-a and turbidity concentrations across the entire dam.

Table 9: Equations derived from estimates of chlorophyll-a and turbidity with their corresponding bands and intercepts used in mapping chlorophyll-a and turbidity concentrations using Landsat 8 and Sentinel-2 imagery

<b>The sensor used: Landsat 8</b>	
<b>Water quality parameter</b>	<b>Formula derived</b>
Chlorophyll-a (µg/L)	$-121.017+(42.915*\text{band1})+(21.362*\text{band2})+(12.498*\text{band5})+(106.086*\text{band6})$
Turbidity (NTU)	$9629.7292+(788.2332*\text{band1})+(0.9489*\text{band3})+(2343.6296*\text{band4})+(6155.1291*\text{band6})+(2227.8601*\text{band7})$
<b>The sensor used: Sentinel-2</b>	
<b>Water quality parameter</b>	<b>Formula derived</b>
Chlorophyll-a (µg/L)	$57.666708+(0.012515*\text{band3})+(0.037702*\text{band5})+(0.005658*\text{band6})+(0.064291*\text{band8})+(0.13529*\text{band12})$
Turbidity (NTU)	$-2878.86024+(3.51175*\text{band4})+(0.10694*\text{band5})+(3.98848*\text{band6})+(2.99105*\text{band12})$

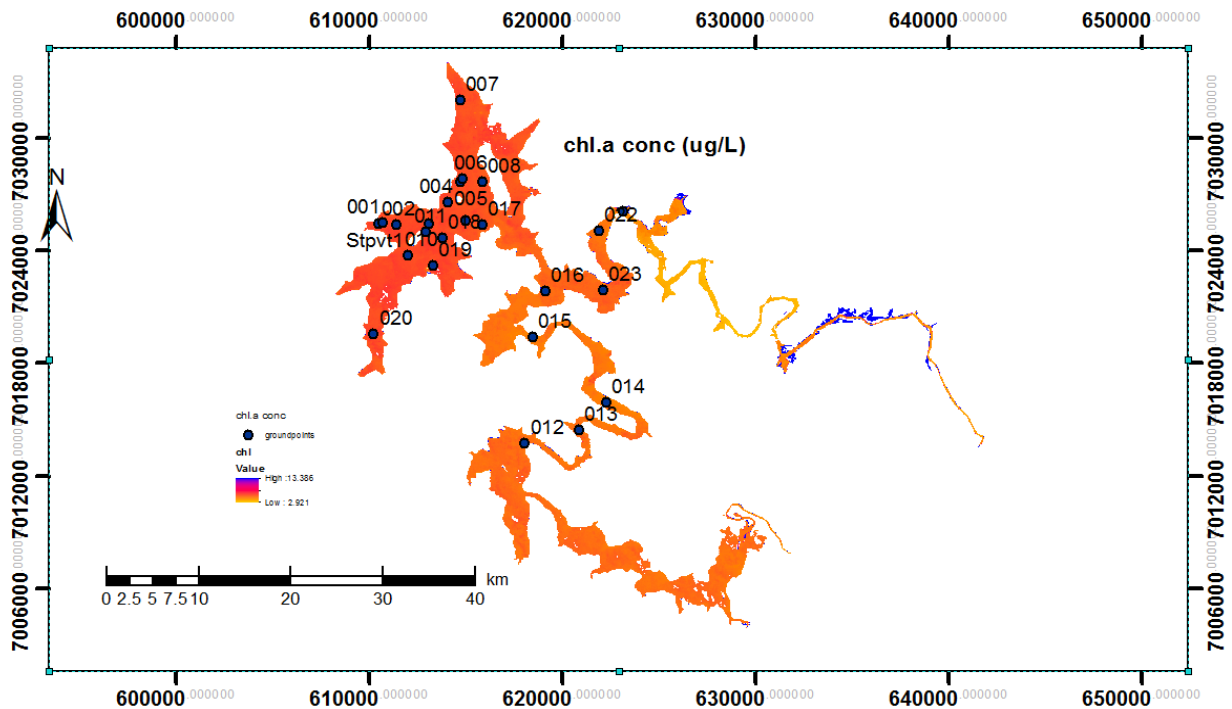


Figure 11: Spatial pattern for chlorophyll-a concentrations estimated from Landsat 8 OLI data on 4 October 2016 at the dam (blue on the legend (13.386) represent high and red to orange (2.921) represent low)

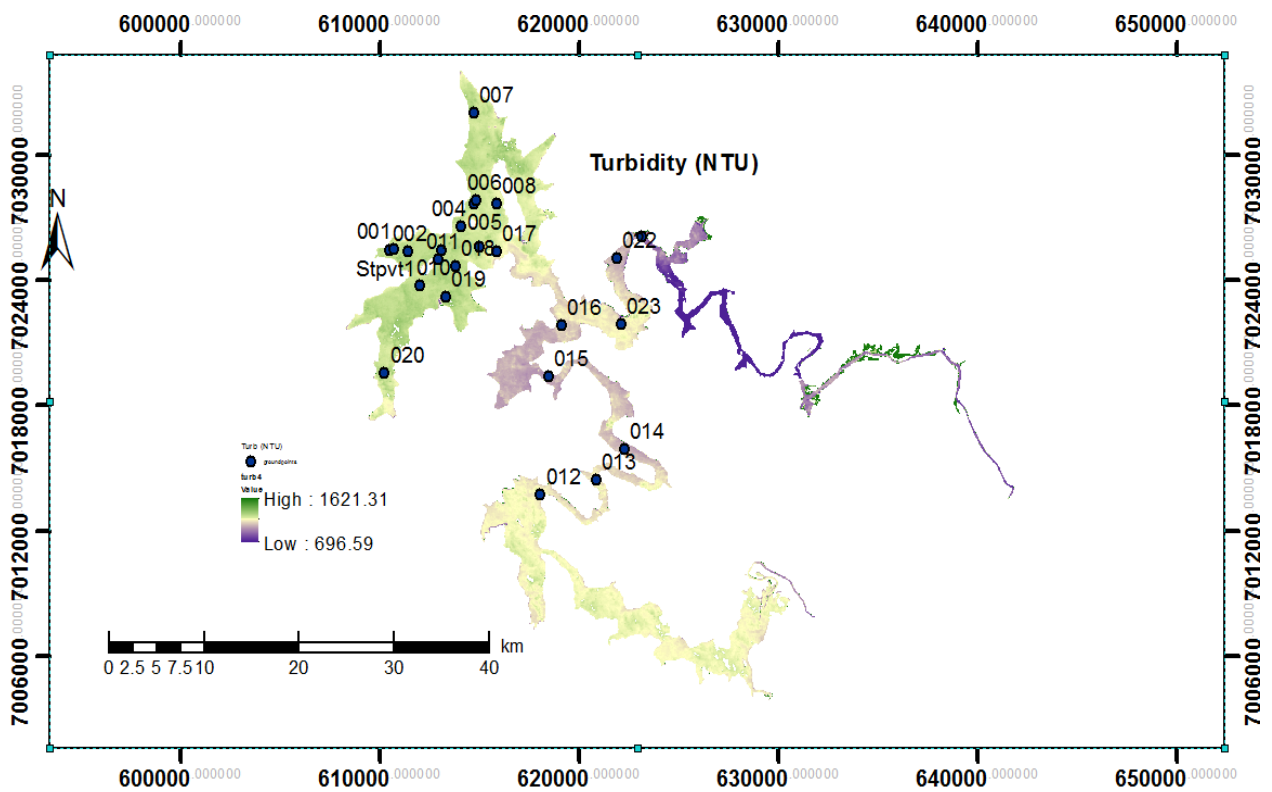


Figure 12: Spatial pattern for turbidity concentrations estimated from Landsat 8 OLI data on 4 October 2016 at the dam

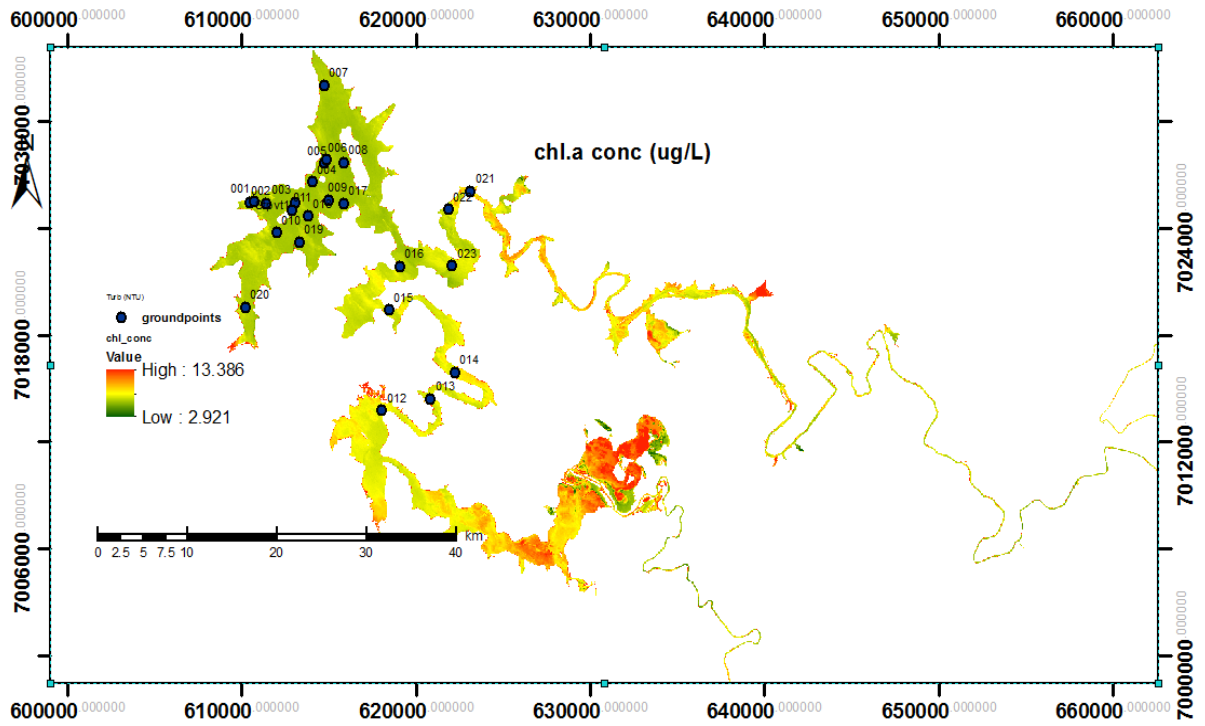


Figure 13: Spatial patterns for chlorophyll-a levels estimated from Sentinel-2A MSI data on 13 November 2016 at the dam.

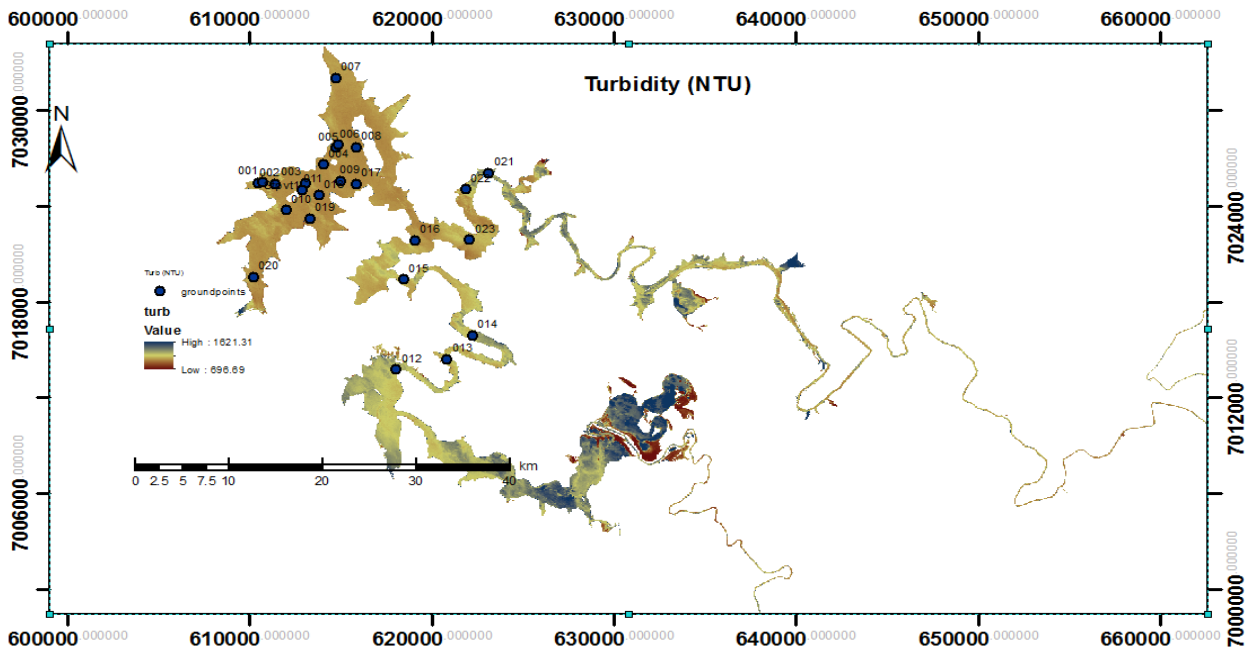


Figure 14: Spatial pattern for turbidity concentrations estimated from Sentinel-2A MSI data on 13 November 2016 at the dam

The spatial patterns of chlorophyll-a & turbidity concentrations obtained from satellite data with field measurements show higher values mostly at the shorelines of the dam and further away from the reservoir where few sampling points were taken. The low values were mainly at the inner section of the dam especially the area of the map in figures 11- 14 where many sampling points were obtained. Thus, the spatial pattern for chlorophyll-a & turbidity concentrations increases outwards at the dam. This might have been as a result of eutrophication along the shorelines of the dam where lots of organic and inorganic sediments from farmlands and built-up areas are being deposited whereas only a small proportion of the sediments are carried into the dam by wind movements and wave currents.

Generally, chlorophyll-a and turbidity concentrations at the Vaal dam were relatively low to average according to the trophic condition of freshwater bodies adapted from (Boyd *et al.*, 1997) of Table 1. In the mapping of chlorophyll-a and turbidity concentrations from Landsat 8 and Sentinel-2 data, different colour ramps were used to show spatial patterns for chlorophyll-a and turbidity concentrations at the dam as shown on Figure 11 to 14.

## CHAPTER 5: Discussion

According to (Lu, 2006), some water quality studies have been successful using different multispectral and hyperspectral sensors to estimate and map chlorophyll-a and turbidity concentrations. In this study, predicting and mapping chlorophyll-a and turbidity concentrations was a success though not as compared to other studies on water quality. This might have been as a result of remote sensing data not coinciding with field data and the water quality might have changed over relatively short periods due to change in wind direction or river inputs. This has also been seen in other studies where by chlorophyll-a concentrations in leaves were being predicted from field measurements using satellite data (Adam et al., 2010). Therefore, in order to improve the success in predicting and mapping chlorophyll-a and turbidity concentrations using remote sensing applications, the remotely sensed data and field data collection dates must coincide as so as to improve the prediction of chlorophyll-a and turbidity concentrations from field measurements using satellite data (Thenkabail *et al.*, 2004; Lu, 2006).

### 5.1 Findings obtained from the random forest and stepwise multiple linear regression models

From past water quality studies, random forest and stepwise multiple linear regression have been used in predicting water quality variables from field data using satellite data (Pal, 2005; Gislason *et al.*, 2006; Lawrence *et al.*, 2006; Chan and Paelinckx, 2008; Stumpf and Kerle, 2011; Adam *et al.*, 2012). In this study, the use of random forest and stepwise multiple linear regression models on Landsat 8 and Sentinel-2 data showed their suitability through their performance and accuracy in predicting chlorophyll-a and turbidity concentrations from field data. The results obtained from the study showed that random forest displayed a higher performance and accuracy compared to stepwise multiple linear regression (Figures 3-10). This might be because the random forest model was able to handle many input variables without variable deletion, generates an internal unbiased estimate of the generalisation error as the forest building progresses, and estimates missing data. These gave the random forest an advantage over stepwise multiple linear regression with its higher R-squared values and lower RMSE values (Breiman, 2001). In other studies, the random forest has been frequently used to yield high R-squared values and performance as well as in predicting chlorophyll-a in vegetation and some water bodies (Stumpf and Kerle, 2011; Adam *et al.*, 2012).

Since the RMSE measures how much error there is between the observed chlorophyll-a and turbidity concentrations from field measurements and predicted chlorophyll-a and turbidity concentration from satellite data, the random forest model was a better predictor compared to the stepwise linear model as seen from its lower RMSE values which indicate a better fit. This implies that most chlorophyll-a and turbidity concentrations observed from field measurements were predicted from the satellite data (Hollister et al., 2016). The R-squared values obtained from the random forest regression model indicated a good fit thus the model showed a positive correlation between the observed and predicted chlorophyll-a and turbidity concentrations over stepwise linear model. This implies that the data points were near the regression line, meaning most of the chlorophyll-a and turbidity concentrations in this case observed from field data were predicted from the remotely sensed data (Figures 5, 8, 9, 10). In some cases, the models did not show a strong positive correlation between the observed and predicted chlorophyll-a and turbidity concentrations (Figures 3, 4, 6, 7). Thus, the correlation between the observed and predicted values, in this case, had a moderate positive relationship although the overall prediction was a success.

Furthermore, the relatively low RMSE values obtained from both models indicated that there was a small variation in error for observed and predicted chlorophyll-a and turbidity concentrations from field measurements using the band reflectance in the remotely sensed data. Based on the overall performances and accuracy of the models, random forest regression gave a better performance and accuracy than stepwise multiple regression, which has also been seen in other water quality studies (Pal, 2005; Gislason *et al.*, 2006; Lawrence *et al.*, 2006; Chan and Paelinckx, 2008; Stumpf and Kerle, 2011; Adam *et al.*, 2012). One of reasons while this study didn't give a very much high performance with the models as seen in other studies might be because the remote sensing data acquisition date not coinciding with the field data collection date may have affected the regression model results in some cases and may be the water quality at the dam might have change over this period of the RS data acquisition with the field sampling validation/ground truthing dates. Moreover, the number of sampling points used for this study may also have caused the poor performance in some instances, where there was a wide variation in the values of chlorophyll-a and turbidity concentrations obtained from field data.

Recommendations to improve the performance of these models in this study are that the field data collection coincides with remote sensing data acquisition date, and more field sampling

points should be used. However, in some cases, water quality may not change drastically over long periods if there are no human activities such as agriculture or flood events which may affect water quality through eutrophication which are the main processes that really affect water quality. But river inputs and wind direction change over relatively short periods but will not affect water quality of the dam significantly.

## **5.2 Statistical analysis of chlorophyll-a and turbidity from field measurements**

The descriptive statistics shown in Table 6 were obtained from field measurements collected from 23 sampling points across the dam. Comparing the mean values for chlorophyll-a concentrations and turbidity, it was found that chlorophyll-a has a mean value of 5.9  $\mu\text{g/L}$  and 955.71 NTU for turbidity. The range between the maximum and minimum values for both chlorophyll-a and turbidity concentrations is large. This means that water quality is slightly polluted at one part, moderately polluted in some other areas and highly polluted in other areas like the shoreline of the dam. The minimum and maximum values for chlorophyll-a and turbidity concentrations indicate that they are not evenly distributed and there is a high variation in chlorophyll-a and turbidity concentrations across the dam. From the trophic state index adopted from Boyd (1997) in Table 1, concerning the range of chlorophyll-a and turbidity measurements obtained from the dam, the different trophic states of the dam can be mapped. This indicates the presence of some algae turbidity, the aesthetic nature of the water reduced, and some oxygen depletion. High turbidity levels exist towards the shorelines of the dam (Figure 2) as a result of deposition of nutrients from agricultural areas and sewage from septic tanks, causing eutrophication in which chlorophyll-a pigments can colour the water (Boyd, 1997). This knowledge will help the management authorities with management strategies for water treatment and conservation.

## **5.3 Spatial patterns for chlorophyll-a and turbidity concentrations at the dam**

The spatial patterns for chlorophyll-a and turbidity concentrations at the dam were mapped by applying the equations of the intercept and coefficients of high band performance in Table 9, from the analysis performed using the random forest regression model to the atmospherically corrected Landsat 8 and Sentinel-2 image data. Results show a low concentration of chlorophyll-a pigments and turbidity in some parts, moderate chlorophyll-a and turbidity concentrations in other regions, and a high concentration of chlorophyll-a and

turbidity in other areas, especially at the shorelines. Generally, there were no very high concentrations of chlorophyll-a and turbidity in the entire dam, indicating that algae production was low at the time of field data collection. R-squared values for chlorophyll-a and turbidity were much higher with Sentinel-2 compared to Landsat 8 data. This shows that there was a stronger correlation between band reflectance values for Sentinel-2 with field data than Landsat 8. This may be because the Sentinel-2 sensor has a higher temporal and spectral resolution which enables it to detect chlorophyll-a and turbidity at very low concentrations though Landsat had a good performance for turbidity with random forest but not as compared to Sentinel-2. The additional red edge bands of the Sentinel-2 sensor absorbed most of the chlorophyll-a and turbidity at a wavelength peak of 0.705 to 0.783  $\mu\text{m}$ .

#### **5.4 Relationship between band reflectance and predicted chlorophyll-a and turbidity values**

The additional red edge band reflectance for Sentinel-2 data gave an advantage in the prediction of more chlorophyll-a and turbidity concentrations compared to Landsat 8 data. Although the correlation was satisfactory, it can be compared with results obtained from other similar studies in Lake Victoria, Omerli dam, and Lake Nablun, which show similar estimated chlorophyll-a and turbidity concentrations from Landsat 8 data and field measurements (Irons et al., 2012).

Sentinel-2 data was acquired on 13/11/2016 while Landsat 8 was acquired on 04/10/2016 and field measurements were taken on 26-28/10/2016. Thus there was a gap in time between the field measurements and remote sensing data acquisition. Therefore there might have been a more significant change in water quality for Landsat 8 with field data compared to Sentinel-2. This may also explain why Sentinel-2 data gave a much better correlation with field measurements compared to Landsat 8. In this study, it is seen that the reliability in estimating and mapping chlorophyll-a and turbidity is better achieved with multispectral remote sensing application, especially with Sentinel-2A MSI.

## Conclusions

This study has shown that the high spatial resolution of the Landsat 8 sensor and the high spectral and temporal resolution of the Sentinel-2 sensor provided data suited for the estimation and mapping of chlorophyll-a and turbidity concentrations at Vaal dam. With the utilisation of simple processing methods and readily available software, the acquisition, processing methods of Landsat 8 and Sentinel-2 data were possible.

The R-squared results obtained showed some good performance in some instances and poor performance in some other cases especially with stepwise multiple linear regression compared to a random forest, although these regression models have demonstrated high performance and accuracy in predicting chlorophyll-a and turbidity concentrations in other studies. The reason for the low performances and accuracy based on their R-squared values in some instances could be a result of the difference in data acquisition date for the remote sensing and field data. This was because data acquisition of remotely sensed data on the dates of field data collection or dates closed to the field data collection dates were affected by clouds and could not be used for the study. The only remotely sensed data available were the ones acquired on the dates mentioned above in the methodology section of remote sensing data acquisition and pre-processing. These are some of the challenges that may occur in water quality projects.

Furthermore, the poor performance of the regression models might also have been due to the few sample points used in the study. Some recommendations in addressing these issues affecting the performance of the regression models were suggested. Firstly, ensuring that acquisition dates of remotely sensed data should coincide with field data acquisition dates to avoid change in water quality which might affect the results from the statistical analysis and that more sample points across a wider range of concentrations (high to low range) should be used to have better predictive variables for chlorophyll-a and turbidity concentrations when building the predictive models.

Obtaining field measurements from the 23 water sampling points gave accurate information but only at these points. Therefore, there was the need to use remotely sensed data from the multispectral sensors Landsat 8 and Sentinel-2. Their spectral and spatial properties enable water quality to be monitored over shorter and longer periods in the visible and near-infrared bands, which enhances the prediction and mapping of chlorophyll-a and turbidity

concentrations at the dam. The study shows that remotely sensed data and field measured chlorophyll-a and turbidity at different sampling points could be related through random forest regression and stepwise linear regression analysis.

The growth in agricultural and industrial sectors within the catchment is deteriorating water quality in the dam and measures should be taken to address this. This study concluded that the use of satellite data in estimating and mapping chlorophyll-a and turbidity from field measurements was a success. Random forest regression modelling showed much better performance and accuracy in prediction compared to stepwise multiple regression.

The chlorophyll-a and turbidity estimated and mapped from the remotely sensed data in this study showed similarity to results obtained from field data. The R-square values indicated that Sentinel-2 was a much better sensor in predicting and mapping chlorophyll-a and turbidity concentrations compared to Landsat 8 as a result of its overall better performance with the regression models compared to Landsat 8. These differences were due to sensor spectral properties. The water quality maps produced in this study showed the spatial patterns of modelled chlorophyll-a and turbidity concentrations, which give a clear understanding of water quality status and the implementation of good management policies for water quality of the dam. Mapping of turbidity concentrations has also provided an insight into the hydrologic routing of the Vaal dam flood system by showing areas where suspended sediments have been discharged into the dam, which will help to pinpoint key sources and assist in water management.

## References

- Adam, E., Mutanga, O. and Rugege, D. 2010. Multispectral and hyperspectral remote sensing for identification and mapping of wetland vegetation: a review. *Wetlands Ecology and Management*, 18, 281-296.
- Adam, E.M., Mutanga, O., Rugege, D. and Ismail, R. 2012. Discriminating the papyrus vegetation (*Cyperus papyrus L.*) and its co-existent species using random forest and hyperspectral data resampled to HYMAP. *International Journal of Remote Sensing*, 33, 552-569.
- Agirre-Basurko E., Ibarra-Berastegi G. and Madariaga, I. (2006) Regression and multilayer perceptron-based models to forecast hourly O<sub>3</sub> and NO<sub>2</sub> levels in the Bilbao area. *Environmental Modelling & Software*, 21, 430-446.
- Aiken, J., Moore, G.F., Trees, C.C., Hooker, S.B. and Clark, D.K. 1996. The SeaWiFS CZCS-type pigment algorithm. *Oceanographic Literature Review*, 33, 315-316.
- Alleman, T.L., Fouts, L. and Chupka, G., 2013. *Quality parameters and chemical analysis for biodiesel produced in the United States in 2011* (No. NREL/TP-5400-57662). National Renewable Energy Lab.(NREL), Golden, CO (United States).
- Azizullah A, Khattak MNK, Richter P, Hder DP (2011) Water pollution in Pakistan and its impact on public health. *Rev Environ Int* 37:479–497.
- Aurin, D.A. and Dierssen, H.M., 2012. Advantages and limitations of ocean color remote sensing in CDOM-dominated, mineral-rich coastal and estuarine waters. *Remote Sensing of Environment*, 125, pp.181-197.
- Bai, J., Cui, B., Xu, X., Ding, Q. and Gao, H., 2009. Heavy metal contamination in riverine soils upstream and downstream of a hydroelectric dam on the Lancang River, China. *Environmental Engineering Science*, 26(5), pp.941-946.
- Barnes, B.B. and Hu, C., 2015. Cross-sensor continuity of satellite-derived water clarity in the Gulf of Mexico: Insights into temporal aliasing and implications for long-term water clarity assessment. *IEEE Transactions on Geoscience and Remote Sensing*, 53(4), pp.1761-1772.

- Barnes, B.B. and Hu, C., 2016. Island building in the South China Sea: detection of turbidity plumes and artificial islands using Landsat and MODIS data. *Scientific reports*, 6, p.33194.
- Blondeau-Patissier, D., Gower, J.F., Dekker, A.G., Phinn, S.R. and Brando, V.E., 2014. A review of ocean colour remote sensing methods and statistical techniques for the detection, mapping and analysis of phytoplankton blooms in coastal and open oceans. *Progress in oceanography*, 123, pp.123-144.
- Bian, B. and Zhu, W., 2009. Particle size distribution and pollutants in road-deposited sediments in different areas of Zhenjiang, China. *Environmental geochemistry and health*, 31(4), pp.511-520.
- Boyd, C.E., Diana, J.S., Szyper, J.P., Batterson, T.R. and Piedrahita, R.H. 1997. Water quality in ponds. In: *Dynamics of Pond Aquaculture*, eds. H.S. Egna and C.E. Boyd, pp. 53–71.
- Braune, E. and Rogers, K.H. 1987. *Vaal River catchment: problems and research needs*. National Scientific Programmes Unit: CSIR.
- Breiman, L. 1996. Bagging predictors. *Machine Learning*, 24, 123-140.
- Breiman, L. 2001. Random forests. *Machine Learning*, 45, 5-32.
- Brivio, P.A., Giardino, C. and Zilioli, E., 2001. Validation of satellite data for quality assurance in lake monitoring applications. *Science of the total environment*, 268(1-3), pp.3-18.
- Bukata, R.P., Jerome, J.H., Kondratyev, K.Y. and Pozdnyakov, D.V. 1991. Satellite monitoring of optically-active components of inland waters: an essential input to regional climate change impact studies. *Journal of Great Lakes Research*, 17, 470-478.
- Carpenter, D.J. and
- Carpenter, S.M. 1983. Modeling inland water quality using Landsat data. *Remote Sensing of Environment*, 13, 345-352.
- Cech, T.V., 2010. *Principles of water resources: history, development, management, and policy*. John Wiley & Sons.

- Chan, J.C.W. and Paelinckx, D. 2008. Evaluation of Random Forest and Adaboost tree-based ensemble classification and spectral band selection for ecotope mapping using airborne hyperspectral imagery. *Remote Sensing of Environment*, 112, 2999-3011.
- Chawira, M., Dube, T. and Gumindoga, W. 2013. Remote sensing based water quality monitoring in Chivero and Manyame lakes of Zimbabwe. *Physics and Chemistry of the Earth, Parts A/B/C*, 66, 38-44.
- Chen, Z., Hu, C. and Muller-Karger, F. 2007. Monitoring turbidity in Tampa Bay using MODIS/Aqua 250-m imagery. *Remote Sensing of Environment*, 109, 207-220.
- Chomko, R.M., Gordon, H.R., Maritorena, S. and Siegel, D.A. 2003. Simultaneous retrieval of oceanic and atmospheric parameters for ocean color imagery by spectral optimization: a validation. *Remote Sensing of Environment*, 84, 208-220.
- Chorus, I. and Fastner, J., 2001. Recreational exposure to cyanotoxins. *Cyanotoxins-occurrence, causes, consequences*, pp.190-199.
- Coelho-Barros, E.A., Simoes, P.A., Achcar, J.A., Martinez, E.Z. and Shimano, A.C. 2008. Methods of Estimation in Multiple Linear Regression: Application to Clinical Data. *Revista Colombiana De Estadística*, 31, 111-129.
- Concha, J.A. and Schott, J.R., 2016. Retrieval of color producing agents in Case 2 waters using Landsat 8. *Remote Sensing of Environment*, 185, pp.95-107.
- 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, 275-285.
- Cox Jr, R.M., Forsythe, R.D., Vaughan, G.E. and Olmsted, L.L., 1998. Assessing water quality in Catawba River reservoirs using Landsat thematic mapper satellite data. *Lake and Reservoir Management*, 14(4), pp.405-416.
- Davis, A.P., Shokouhian, M. and Ni, S., 2001. Loading estimates of lead, copper, cadmium, and zinc in urban runoff from specific sources. *Chemosphere*, 44(5), pp.997-1009.

- De Carlo, E.H., Hoover, D.J., Young, C.W., Hoover, R.S. and Mackenzie, F.T. 2007. Impact of storm runoff from tropical watersheds on coastal water quality and productivity. *Applied Geochemistry*, 22, 1777-1797.
- Dekker, A.G. and Hoogenboom, H. 1997. *Operational tools for remote sensing of water quality: A prototype toolkit*. Netherlands Remote Sensing Board (BCRS), Programme Bureau, Rijkswaterstaat Survey Department.
- Dekker, A.G. and Peters, S.W.M. 1993. The use of the Thematic Mapper for the analysis of eutrophic lakes: a case study in the Netherlands. *International Journal of Remote Sensing*, 14, 799-821.
- Dekker, L. and Ritsema, C. (1995). Fingerlike Wetting Patterns in Two Water-Repellent Loam Soils. *Journal of Environment Quality*, 24(2), p.324.
- Dekker, A.G., Vos, R.J. and Peters, S.W.M. 2001. Comparison of remote sensing data, model results and in situ data for total suspended matter (TSM) in the southern Frisian lakes. *Science of the Total Environment*, 268, 197-214.
- 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, 15-35.
- Del Castillo, C.E., Coble, P.G., Morell, J.M., López, J.M. and Corredor, J.E. 1999. Analysis of the optical properties of the Orinoco River plume by absorption and fluorescence spectroscopy. *Marine Chemistry*, 66, 35-51.
- Dewidar, K. and Khedr, A.A., 2005. Remote sensing of water quality for Burullus Lake Egypt. *Geocarto international*, 20(3), pp.43-49.
- Dlamini, S., Nhapi, I., Gumindoga, W., Nhwatiwa, T. and Dube, T. 2016. Assessing the feasibility of integrating remote sensing and in-situ measurements in monitoring water quality status of Lake Chivero, Zimbabwe. *Physics and Chemistry of the Earth, Parts A/B/C*, 93, 2-11.

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, 1-7.

DWAF Department of Water Affairs and Forestry (DWAF).(2008) *Strategic framework for water for sustainable growth and development*. Department of Water Affairs and Forestry, Pretoria, South Africa. (2008).

DWAF, 2008 Department of Water Affairs and Forestry (DWAF), 2008. *Integrated Water Quality Management Plan for the Vaal River System: Water Quality Management Strategy*. DWAF Report No. P RSA C000/00/2305/7. Department of Water Affairs, Pretoria.

DWAF 2003. Trophic status report: *trophic status of impoundments*. South African Department of Water Affairs and Forestry, 1, Pretoria.

Ekercin, S., 2007. Water quality retrievals from high resolution IKONOS multispectral imagery: A case study in Istanbul, Turkey. *Water, Air, and Soil Pollution*, 183(1-4), pp.239-251.

El Saadi, A.M., Yousry, M.M. and Jahin, H.S., 2014. Statistical estimation of Rosetta branch water quality using multi-spectral data. *Water Science*, 28(1), pp.18-30.

Eregno, F.E. 2013. *Multiple linear regression models for estimating microbial load in a drinking water source case from the Glomma river, Norway*. Masters thesis, Norwegian University of Life Sciences, Ås.

Falconer, I.R., 1999. An overview of problems caused by toxic blue-green algae (cyanobacteria) in drinking and recreational water. *Environmental Toxicology*, 14(1), pp.5-12.

Franz, B.A., Bailey, S.W., Kuring, N. and Werdell, P.J., 2015. Ocean colour measurements with the Operational Land Imager on Landsat-8: implementation and evaluation in SeaDAS. *Journal of Applied Remote Sensing*, 9(1), p.096070.

Ferraro, M.B. and Giordani, P. 2012. A multiple linear regression model for imprecise information. *Metrika*, 75, 1049-1068.

- Fraser, R.N. 1998. Multispectral remote sensing of turbidity among Nebraska Sand Hills lakes. *International Journal of Remote Sensing*, 19, 3011-3016.
- Giardino, C., Pepe, M., Brivio, P.A., Ghezzi, P. and Zilioli, E. 2001. Detecting chlorophyll, Secchi disk depth and surface temperature in a sub-alpine lake using Landsat imagery. *Science of the Total Environment*, 268, 19-29.
- Giardino, C., Bresciani, M., Cazzaniga, I., Schenk, K., Rieger, P., Braga, F., Matta, E. and Brando, V., 2014. Evaluation of multi-resolution satellite sensors for assessing water quality and bottom depth of Lake Garda. *Sensors*, 14(12), pp.24116-24131.
- 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, 24109-24125.
- Gin, K.H., Koh, S.T. and Lin, I.I., 2003. Spectral irradiance profiles of suspended marine clay for the estimation of suspended sediment concentration in tropical waters. *International Journal of Remote Sensing*, 24(16), pp.3235-3245.
- Gislason, P.O., Benediktsson, J.A. and Sveinsson, J.R. 2006. Random forests for land cover classification. *Pattern Recognition Letters*, 27, 294-300.
- Gitelson, A.A., Schalles, J.F. and Hladik, C.M. 2007. Remote chlorophyll-a retrieval in turbid, productive estuaries: Chesapeake Bay case study. *Remote Sensing of Environment*, 109, 464-472.
- Glasgow, H.B., Burkholder, J.M., Reed, R.E., Lewitus, A.J. and Kleinman, J.E., 2004. Real-time remote monitoring of water quality: a review of current applications, and advancements in sensor, telemetry, and computing technologies. *Journal of Experimental Marine Biology and Ecology*, 300(1-2), pp.409-448.
- Göbel, A., McArdell, C.S., Joss, A., Siegrist, H. and Giger, W., 2007. Fate of sulfonamides, macrolides, and trimethoprim in different wastewater treatment technologies. *Science of the Total Environment*, 372(2-3), pp.361-371.

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.

Gordon, H.R. 1997. Atmospheric correction of ocean color imagery in the Earth Observation System era. *Journal of Geophysical Research*, 102, 17081-17106.

Gower, J., King, S., Borstad, G. and Brown, L. 2005. Detection of intense plankton blooms using the 709 nm band of the MERIS imaging spectrometer. *International Journal of Remote Sensing*, 26, 2005-2012.

Grimm, R., Behrens, T., Märker, M. and Elsenbeer, H. 2008. Soil organic carbon concentrations and stocks on Barro Colorado Island—Digital soil mapping using Random Forests analysis. *Geoderma*, 146, 102-113.

Hadjimitsis, D.G., Hadjimitsis, M.G., Toullos, L. and Clayton, C., 2010. Use of space technology for assisting water quality assessment and monitoring of inland water bodies. *Physics and Chemistry of the Earth, Parts A/B/C*, 35(1-2), pp.115-120.

Hadjimitsis, D.G., Clayton, C.R.I. and Hope, V.S., 2004. An assessment of the effectiveness of atmospheric correction algorithms through the remote sensing of some reservoirs. *International Journal of Remote Sensing*, 25(18), pp.3651-3674.

Hagolle, O., Huc, M., Villa Pascual, D. and Dedieu, G., 2015. A multi-temporal and multi-spectral method to estimate aerosol optical thickness over land, for the atmospheric correction of FormoSat-2, LandSat, VEN $\mu$ S and Sentinel-2 images. *Remote Sensing*, 7(3), pp.2668-2691.

Hellweger, F.L., Miller, W. and Oshodi, K.S., 2007. Mapping turbidity in the Charles River, Boston using a high-resolution satellite. *Environmental monitoring and assessment*, 132(1-3), pp.311-320.

Hollister, J.W., Milstead, W.B. and Kreakie, B.J., 2016. Modeling lake trophic state: a random forest approach. *Ecosphere*, 7(3).

Huang, X., Pedersen, T., Fischer, M., White, R. and Young, T.M., 2004. Herbicide runoff along highways. 1. Field observations. *Environmental science & technology*, 38(12), pp.3263-3271.

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, 210-213.

- Irons, J.R., Dwyer, J.L. and Barsi, J.A. 2012. The next Landsat satellite: The Landsat data continuity mission. *Remote Sensing of Environment*, 122, 11-21.
- Jofre, J., Blanch, A.R. and Lucena, F., 2009. Water-borne infectious disease outbreaks associated with water scarcity and rainfall events. *In Water scarcity in the mediterranean* (pp. 147-159). Springer, Berlin, Heidelberg.
- Kabbara, N., Benkhelil, J., Awad, M. and Barale, V., 2008. Monitoring water quality in the coastal area of Tripoli (Lebanon) using high-resolution satellite data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 63(5), pp.488-495.
- 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.
- Kibena, J., Nhapi, I. and Gumindoga, W., 2014. Assessing the relationship between water quality parameters and changes in landuse patterns in the Upper Manyame River, Zimbabwe. *Physics and Chemistry of the Earth, Parts A/B/C*, 67, pp.153-163.
- Kloiber, S.M., Anderle, T.H., Brezonik, P.L., Olmanson, L., Bauer, M.E. and Brown, D.A., 2000. Trophic state assessment of lakes in the Twin Cities(Minnesota, USA) region by satellite imagery. *Advances in limnology. Stuttgart*, (55), pp.137-151.
- Kloiber, S.M., Brezonik, P.L., Olmanson, L.G. and Bauer, M.E., 2002. A procedure for regional lake water clarity assessment using Landsat multispectral data. *Remote sensing of Environment*, 82(1), pp.38-47.
- Kneese, A.V. and Bower, B.T., 2013. *Managing water quality: economics, technology, institutions*. Rff Press.
- Knobeloch, L., Salna, B., Hogan, A., Postle, J. and Anderson, H., 2000. Blue babies and nitrate-contaminated well water. *Environmental health perspectives*, 108(7), p.675.
- Kovdienko, N.A., Polishchuk, P.G., Muratov, E.N., Artemenko, A.G., Kuz'min, V.E., Gorb, L., Hill, F. and Leszczynski, J. 2010. Application of Random Forest and Multiple Linear Regression Techniques to QSPR Prediction of an Aqueous Solubility for Military Compounds. *Molecular Informatics*, 29, 394-406.
- Lathrop, R.G. 1986. Use of Thematic Mapper data to assess water quality in Green Bay and central Lake Michigan. *Photogrammetric Engineering and Remote Sensing*, 52, 671-680.
- Lavery, P., Pattiaratchi, C., Wyllie, A. and Hick, P., 1993. Water quality monitoring in estuarine waters using the Landsat Thematic Mapper. *Remote sensing of environment*, 46(3), pp.268-280.

- Lawrence, R.L., Wood, S.D. and Sheley, R.L. 2006. Mapping invasive plants using hyperspectral imagery and Breiman Cutler classifications (Random Forest). *Remote Sensing of Environment*, 100, 356-362.
- Lee, E.J. and Schwab, K.J., 2005. Deficiencies in drinking water distribution systems in developing countries. *Journal of water and health*, 3(2), pp.109-127.
- Letelier, R.M. and Abbott, M.R., 1996. An analysis of chlorophyll fluorescence algorithms for the Moderate Resolution Imaging Spectrometer (MODIS). *Remote Sensing of Environment*, 58(2), pp.215-223.
- Lodhi, M.A., Rundquist, D.C., Han, L. and Kuzila, M.S., 1997. THE POTENTIAL FOR REMOTE SENSING OF LOESS SOILS SUSPENDED IN SURFACE WATERS 1. *JAWRA Journal of the American Water Resources Association*, 33(1), pp.111-117.
- Liu, Y., Islam, M.A. and Gao, J., 2003. Quantification of shallow water quality parameters by means of remote sensing. *Progress in physical geography*, 27(1), pp.24-43.
- Liversedge, L., 2007. *Turbidity mapping and prediction in ice marginal lakes at the Bering Glacier system, Alaska* (Doctoral dissertation).
- Lu, D. 2006. The potential and challenge of remote sensing-based biomass estimation. *International Journal of Remote Sensing*, 27, 1297-1328.
- Majozzi, N.P., Salama, M.S., Bernard, S., Harper, D.M. and Habte, M.G., 2014. Remote sensing of euphotic depth in shallow tropical inland waters of Lake Naivasha using MERIS data. *Remote sensing of environment*, 148, pp.178-189.
- Marsalek, J., Rochfort, Q., Brownlee, B., Mayer, T. and Servos, M., 1999. An exploratory study of urban runoff toxicity. *Water science and technology*, 39(12), pp.33-39.
- Mayo, M., Gitelson, A., Yacobi, Y.Z. and Ben-Avraham, Z., 1995. Chlorophyll distribution in lake kinneret determined from Landsat Thematic Mapper data. *Remote Sensing*, 16(1), pp.175-182.
- Miller, R.L. and McKee, B.A. 2004. Using MODIS Terra 250 m imagery to map concentrations of total suspended matter in coastal waters. *Remote Sensing of Environment*, 93, 259-266.
- Morel, A. and Prieur, L. 1977. Analysis of variations in ocean colour. *Limnology and Oceanography*, 22, 709-722.
- Mueller, J.L., Davis, C., Arnone, R., Frouin, R., Carder, K., Lee, Z.P., Steward, R.G., Hooker, S., Mobley, C.D. and McLean, S., 2000. Above-water radiance and remote sensing reflectance measurements and analysis protocols. *Ocean Optics protocols for satellite ocean color sensor validation Revision*, 2, pp.98-107.

- Neter, J., Wasserman, W. and Kutner, M.H., 1989. Applied linear regression models.
- Ndungu, J.N. 2014. *Assessing water quality in Lake Naivasha*. PhD thesis, University of Twente.
- Nellis, M.D., Harrington Jr, J.A. and Wu, J., 1998. Remote sensing of temporal and spatial variations in pool size, suspended sediment, turbidity, and Secchi depth in Tuttle Creek Reservoir, Kansas: 1993. *Geomorphology*, 21(3-4), pp.281-293.
- Ngabe, B., Bidleman, T.F. and Scott, G.I., 2000. Polycyclic aromatic hydrocarbons in storm runoff from urban and coastal South Carolina. *Science of the total environment*, 255(1-3), pp.1-9.
- Nieuwoudt, W. and Backeberg, G. (2011). A review of the modelling of water values in different use sectors in South Africa. *Water SA*, 37(5).
- Novo, E.M.M., Hansom, J.D. and Curran, P.J., 1989. The effect of sediment type on the relationship between reflectance and suspended sediment concentration. *Remote Sensing*, 10(7), pp.1283-1289.
- Olmanson, L.G., Brezonik, P.L. and Bauer, M.E. 2013. Airborne hyperspectral remote sensing to assess spatial distribution of water quality characteristics in large rivers: The Mississippi River and its tributaries in Minnesota. *Remote Sensing of Environment*, 130, 254-265.
- O'Reilly, J.E. 2000. Ocean color chlorophyll a algorithms for SeaWiFS, OC2, and OC4: Version 4. *SeaWiFS Postlaunch Calibration and Validation Analyses*, 3, 9-23.
- 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, 24937-24953.
- Östlund, C., Flink, P., Strömbeck, N., Pierson, D. and Lindell, T., 2001. Mapping of the water quality of Lake Erken, Sweden, from imaging spectrometry and Landsat Thematic Mapper. *Science of the Total Environment*, 268(1-3), pp.139-154.
- Pacheco, A., Horta, J., Loureiro, C. and Ferreira, Ó., 2015. Retrieval of nearshore bathymetry from Landsat 8 images: a tool for coastal monitoring in shallow waters. *Remote Sensing of Environment*, 159, pp.102-116.
- Pahlevan, N., Lee, Z., Wei, J., Schaaf, C.B., Schott, J.R. and Berk, A., 2014. On-orbit radiometric characterization of OLI (Landsat-8) for applications in aquatic remote sensing. *Remote Sensing of Environment*, 154, pp.272-284.
- Pal, M. 2005. Random forest classifier for remote sensing classification. *International Journal of Remote Sensing*, 26, 217-222.

- Palmer, D.S., O'Boyle, N.M., Glen, R.C. and Mitchell, J.B. 2007. Random forest models to predict aqueous solubility. *Journal of Chemical Information and Modelling*, 47, 150-158.
- Park, J.H., Duan, L., Kim, B., Mitchell, M.J. and Shibata, H., 2010. Potential effects of climate change and variability on watershed biogeochemical processes and water quality in Northeast Asia. *Environment International*, 36(2), pp.212-225.
- Park, Y.J., Ruddick, K. and Lacroix, G., 2010. Detection of algal blooms in European waters based on satellite chlorophyll data from MERIS and MODIS. *International Journal of Remote Sensing*, 31(24), pp.6567-6583.
- Peñaflor, E.L., Villanoy, C.L., Liu, C.T. and David, L.T., 2007. Detection of monsoonal phytoplankton blooms in Luzon Strait with MODIS data. *Remote sensing of environment*, 109(4), pp.443-450.
- Pimentel, D., Berger, B., Filiberto, D., Newton, M., Wolfe, B., Karabinakis, E., Clark, S., Poon, E., Abbett, E. and Nandagopal, S., 2004. Water resources: agricultural and environmental issues. *BioScience*, 54(10), pp.909-918.
- Qin, B., Li, W., Zhu, G., Zhang, Y., Wu, T. and Gao, G., 2015. Cyanobacterial bloom management through integrated monitoring and forecasting in large shallow eutrophic Lake Taihu (China). *Journal of hazardous materials*, 287, pp.356-363
- Quibell, G., 1991. The effect of suspended sediment on reflectance from freshwater algae. *International Journal of Remote Sensing*, 12(1), pp.177-182.
- Rawlings, J.O. 1988. *Applied Regression Analysis: A Research Tool*. Pacific Grove: Wadsworth and Brooks/Cole Advanced Books and Software.
- Richardson, L.L., 1996. Remote sensing of algal bloom dynamics. *BioScience*, 46(7), pp.492-501.
- Rinaldi, M., Surian, N., Comiti, F. and Bussettini, M., 2013. A method for the assessment and analysis of the hydromorphological condition of Italian streams: The Morphological Quality Index (MQI). *Geomorphology*, 180, pp.96-108.
- Rinta-Kanto, J.M., Saxton, M.A., DeBruyn, J.M., Smith, J.L., Marvin, C.H., Krieger, K.A., Saylor, G.S., Boyer, G.L. and Wilhelm, S.W., 2009. The diversity and distribution of toxigenic *Microcystis* spp. in present day and archived pelagic and sediment samples from Lake Erie. *Harmful Algae*, 8(3), pp.385-394.
- Ritchie, J.C. and Cooper, C.M. 2001. Remote sensing techniques for determining water quality: Applications to TMDLs. In: *TMDL Science Issues Conference, Water Environment Federation, Alexandria, VA*, pp. 367-374.
- Ritchie, J.C. and Schiebe, F.R. 2000. Water quality. In: *Remote Sensing in Hydrology and Water Management* (pp. 287-303). Springer Berlin Heidelberg.

- Ritchie, J.C., Cooper, C.M. and Yongqing, J. 1987. Using Landsat multispectral scanner data to estimate suspended sediments in Moon Lake, Mississippi. *Remote Sensing of Environment*, 23, 65-81.
- Ritchie, J.C. and Rango, A., 1996. Remote sensing applications to hydrology: introduction. *Hydrological Sciences Journal*, 41(4), pp.429-431.
- Ritchie, J.C., Zimba, P.V. and Everitt, J.H. 2003. Remote sensing techniques to assess water quality. *Photogrammetric Engineering & Remote Sensing*, 69, 695-704.
- 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, 154-172.
- Ruddick, K.G., Ovidio, F. and Rijkeboer, M. 2000. Atmospheric correction of SeaWiFS imagery for turbid coastal and inland waters. *Applied Optics*, 39, 897-912.
- Ruddick, K.G., De Cauwer, V., Park, Y.J. and Moore, G., 2006. Seaborne measurements of near infrared water-leaving reflectance: The similarity spectrum for turbid waters. *Limnology and Oceanography*, 51(2), pp.1167-1179.
- Salama, M.S., Dekker, A., Su, Z., Mannaerts, C.M. and Verhoef, W., 2009. Deriving inherent optical properties and associated inversion-uncertainties in the Dutch Lakes. *Hydrology and Earth System Sciences*, 13(7), pp.1113-1121.
- Salama, M.S., Radwan, M. and van der Velde, R., 2012. A hydro-optical model for deriving water quality variables from satellite images (HydroSat): A case study of the Nile River demonstrating the future Sentinel-2 capabilities. *Physics and Chemistry of the Earth, Parts A/B/C*, 50, pp.224-232.
- 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 an eutrophic lake. *Journal of Phycology*, 34, 383-390.
- Schiebe, F.R., Harrington Jr, J.A. and Ritchie, J.C., 1992. Remote sensing of suspended sediments: the Lake Chicot, Arkansas project. *International Journal of Remote Sensing*, 13(8), pp.1487-1509.
- Schofield, O., Bergmann, T., Bissett, P., Grassle, J.F., Haidvogel, D.B., Kohut, J., Moline, M. and Glenn, S.M., 2002. The long-term ecosystem observatory: An integrated coastal observatory. *IEEE Journal of Oceanic Engineering*, 27(2), pp.146-154.
- Shuchman, R.A., Leshkevich, G., Sayers, M.J., Johengen, T.H., Brooks, C.N. and Pozdnyakov, D. 2013. An algorithm to retrieve chlorophyll, dissolved organic carbon, and suspended minerals from Great Lakes satellite data. *Journal of Great Lakes Research*, 39, 14-33.

- Siegel, H. and Gerth, M. 2000. Remote-sensing studies of the exceptional summer of 1997 in the Baltic Sea: The warmest August of the century, the Oder flood, and phytoplankton blooms. *Elsevier Oceanography Series*, 63, 239-255.
- 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), p.62.
- Soticha, K., Jareeya, Y., Sudjit, K. and Prapat, P. (2014). Assessing Water Quality of Rural Water Supply in Thailand. *Journal of Clean Energy Technologies*, pp.226-228.
- Strömbeck, N., 2001. *Water quality and optical properties of Swedish lakes and coastal waters in relation to remote sensing* (Doctoral dissertation, Acta Universitatis Upsaliensis).
- Stumpf, A. and Kerle, N. 2011. Object-oriented mapping of landslides using Random Forests. *Remote Sensing of Environment*, 115, 2564-2577.
- Svejkovsky, J. and Shandley, J. 2001. Detection of offshore plankton blooms with AVHRR and SAR imagery. *International Journal of Remote Sensing*, 22, 471-485.
- Thenkabail, P.S., Stucky, N., Griscom, B.W., Ashton, M.S., Diels, J., Van Der Meer, B. and Enclona, E. 2004. Biomass estimations and carbon stock calculations in the oil palm plantations of African derived savannas using IKONOS data. *International Journal of Remote Sensing*, 25, 5447-5472.
- 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, 640, doi:10.3390/rs8080640.
- Usali, N. and Ismail, M.H. 2010. Use of remote sensing and GIS in monitoring water quality. *Journal of Sustainable Development*, 3, 228, doi:10.5539/jsd.v3n3p228.
- Vanhellemont, Q. and Ruddick, K., 2016, May. Acolite for Sentinel-2: Aquatic applications of MSI imagery. *In Proceedings of the ESA Living Planet Symposium, Pragur, Czech Republic* (pp. 9-13).
- 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.
- Vanhellemont, Q. and Ruddick, K., 2014. Turbid wakes associated with offshore wind turbines observed with Landsat 8. *Remote Sensing of Environment*, 145, pp.105-115.
- Vermote, E.F. and Vermeulen, A., 1999. Atmospheric correction algorithm: spectral reflectances (MOD09). *ATBD version*, 4, pp.1-107.

- Verstraete, M.M., Pinty, B. and Curran, P.J., 1999. MERIS potential for land applications. *International Journal of Remote Sensing*, 20(9), pp.1747-1756.
- Vincent, R.K., Qin, X., McKay, R.M.L., Miner, J., Czajkowski, K., Savino, J. and Bridgeman, T. 2004. Phycocyanin detection from LANDSAT TM data for mapping cyanobacterial blooms in Lake Erie. *Remote Sensing of Environment*, 89, 381-392.
- Wang, F., Han, L., Kung, H.T. and Van Arsdale, R.B., 2006. Applications of Landsat-5 TM imagery in assessing and mapping water quality in Reelfoot Lake, Tennessee. *International Journal of Remote Sensing*, 27(23), pp.5269-5283.
- 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, 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, 263-283.
- Wei, K. and Yang, J., 2015. Oxidative damage induced by copper and beta-cypermethrin in gill of the freshwater crayfish *Procambarus clarkii*. *Ecotoxicology and environmental safety*, 113, pp.446-453.
- World Health Organization, 2011. Evaluating household water treatment options: Health-based targets and microbiological performance specifications.
- Wu, M., Zhang, W., Wang, X. and Luo, D. 2009. Application of MODIS satellite data in monitoring water quality parameters of Chaohu Lake in China. *Environmental Monitoring and Assessment*, 148, 255-264.
- Yüzügüllü, O. and Aksoy, A., 2011. Determination of Secchi Disc depths in Lake Eymir using remotely sensed data. *Procedia-Social and Behavioral Sciences*, 19, pp.586-592.
- Zhang, J. and Liu, C.L., 2002. Riverine composition and estuarine geochemistry of particulate metals in China—weathering features, anthropogenic impact and chemical fluxes. *Estuarine, coastal and shelf science*, 54(6), pp.1051-1070.
- Zilioli, E. and Brivio, P.A. 1997. The satellite derived optical information for the comparative assessment of lacustrine water quality. *Science of the Total Environment*, 196, 229-245.