

UNIVERSITY OF THE WITWATERSRAND

JOHANNESBURG



UNIVERSITY OF THE
WITWATERSRAND,
JOHANNESBURG

**A comparison of Landsat 8 and Sentinel 2 in mapping chlorophyll-a
and turbidity in the Vaal Dam, and Lakes Shinji and Nakaumi of
Japan**

by

Prosper Bande (924631)

A research report submitted to the Faculty of Science, University of the
Witwatersrand, Johannesburg, in partial fulfilment of the requirements for the degree
of Master of Science (Geographical Information Systems and Remote Sensing) at
the School of Geography, Archaeology & Environmental Studies

Supervisor: Dr Elhadi Adam (Wits University)

Declaration

I, Prosper Bande, declare that this research report is my own unassisted work. It is being submitted as partial fulfilment for the degree Master of Science in Geographical Information Systems and Remote Sensing to the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination at any other University.

**Prosper
Bande** Digitally signed by
Prosper Bande
Date: 2020.06.12
13:32:08 +02'00'

Signature of candidate

..... Day of.....**2020**.....at.....**Braamfontein**.....Johannesburg

Abstract

The Vaal Dam of South Africa and Lakes Shinji and Nakaumi of Japan play an important part in human activities such as agriculture, industry and domestic utilities. Two major problems confronting both the Vaal Dam and Lakes Shinji and Nakaumi are eutrophication and high levels of turbidity. Remote sensing technologies provide an effective means of monitoring the quality of water bodies. Landsat 8, Sentinel 2 and field hyperspectral sensors were investigated using support vector regression to find the bands and band ratios that best predict turbidity and chlorophyll-a in Lakes Shinji and Nakaumi (LSN), and the Vaal Dam. For Vaal Dam chlorophyll-a, the best bands/band ratios were found to be the global environmental monitoring index (GEMI) for Landsat 8, B2/B1 for Sentinel 2 and Rrs687/Rrs675 for the field hyperspectral remote sensing reflectance (Rrs) data. For Lakes Shinji and Nakaumi, the best bands/band ratios for chlorophyll-a were found to be B2/B3 for Landsat 8 B7/B6 for Sentinel 2 and Rrs881/Rrs882 for the field hyperspectral Rrs data. For retrieving turbidity, the best bands/band ratios for Vaal Dam were found to be B4/B1 for Landsat 8, B3/B1 for Sentinel 2 and Rrs819/Rrs794 for the hyperspectral Rrs data. For LSN, the best bands/band ratios for turbidity were found to be B9/B8 for Landsat 8, B11/B1 for Sentinel 2 and Rrs580/Rrs575 for the hyperspectral data. Support vector regression was found to perform well even with relatively small samples for the Vaal Dam achieving R^2 values of 0.92 for Hyperspectral Rrs data, 0.91 for Sentinel 2 and 0.79 Landsat 8 in predicting chlorophyll-a, and 0.68 for hyperspectral Rrs data, 0.89 for Sentinel 2 and 0.70 for Landsat 8 in predicting turbidity. Lakes Shinji and Nakaumi did not achieve support vector regression results as good as those of the Vaal Dam due to small sample size. For chlorophyll-a, R^2 values of 0.75 for the field hyperspectral Rrs data, 0.70 for Sentinel 2 and 0.13 for Landsat 8 were achieved. Turbidity R^2 values were 0.68 for the field hyperspectral data, 0.39 for Sentinel 2 and 0.44 for Landsat 8. This work finds that Landsat 8 and Sentinel 2 can be effectively used to monitor water quality in small water bodies. As a general deduction from the results, Sentinel 2's performance is better than that of Landsat 8 in predicting both chlorophyll-a and turbidity and is comparable to the field hyperspectral Rrs data. However, it is acknowledged that Landsat 8 has more historical archives useful for time series

analyses. The two satellites should be used in a complementary manner for greater probabilities of obtaining cloud-free images.

Keywords: *Chlorophyll-a, Turbidity, Sentinel 2, Landsat 8, Hyperspectral Rrs, Support Vector Regression.*

Acknowledgements

Firstly, I would like to thank God Almighty through Jesus Christ His Son who gives me strength and whose grace is sufficient for me every single day.

I would also like to thank Dr Elhadi Adam for his guidance and supervision throughout this work. Thank you also to the Agricultural Research Council for the role they played in the collection of the field data used in this research.

Last but certainly not least, I would like to thank my husband and children for their unwavering support and encouragement throughout this work.

Dedication

This research is dedicated to my family, my husband Tapiwa, and my daughters Tanatswa and Anotida who sacrificed a lot, encouraged and motivated me throughout my masters degree, and made sure I finished this research report.

Table of Contents

Declaration	i
Abstract	ii
Acknowledgements	iv
Dedication	v
List of Figures	viii
List of tables	ix
List of equations	ix
List of Abbreviations	x
Chapter one	1
1. Introduction	1
1.1 Background	1
1.2 Problem statement	3
1.3 Aim and Objectives	5
1.3.1 Aim	5
1.3.2 Objectives	5
1.4 Research report outline	6
Chapter two	8
2. Literature Review	8
2.1 Hyperspectral remote sensing for water quality	8
2.2 Multispectral remote sensing for water quality	11
Chapter three	15
3. Methodology	15
3.1 Study area	15
3.2 Field data collection	17
3.2.1 Field spectral measurements	17

3.2.2 Water quality sampling	18
3.3 Remote sensing data acquisition and pre-processing	19
3.3.1 Landsat 8 data.....	21
3.3.2 Sentinel 2 data	21
3.4 Statistical analysis	22
Chapter four.....	25
4. Results.....	25
4.1 Descriptive statistics	25
4.2 Chlorophyll-a results.....	28
4.2.1 Correlation results	28
4.2.2 Support vector regression results	31
4.3 Turbidity results.....	35
4.3.1 Correlation results	35
4.3.2 Support vector results	38
Chapter five	42
5. Discussion and conclusions	42
5.1 Discussion.....	42
5.2 Conclusion	45
5.3 Limitations and recommendations	45
References	46

List of Figures

Figure 1 Outline of the research report showing the sequential workflow that was followed ..	7
Figure 2 Vaal Dam, South Africa	16
Figure 3 Lake Shinji and Lake Nakaumi of Japan.....	17
Figure 4 Chlorophyll-a summary statistics for (a) Vaal Dam and (b) LSN Lakes. The mean, standard deviation and total number of samples are reported.....	26
Figure 5 Turbidity summary statistics for (a) Vaal Dam and (b) LSN Lakes. The mean, standard deviation and total number of samples are reported.....	27
Figure 6 Correlation plots showing correlation coefficients between chlorophyll-a and field hyperspectral Rrs band ratios obtained from all possible two-band combinations using bands located between 400 and 900nm for (a) Vaal Dam and (b) LSN Lakes	29
Figure 7 Field hyperspectral Rrs data plots showing relationships between actual and predicted chlorophyll-a for (a) Vaal Dam training set (n=13), (b) Vaal Dam validation set (n=4), (c) LSN training set (n=7) and (d) LSN validation set (n=3). The R^2 values are reported for each model.	33
Figure 8 Landsat 8 plots showing relationships between actual and predicted chlorophyll-a for (a) Vaal Dam training set (n=15), (b) Vaal Dam validation set (n=7), (c) LSN training set (n=7) and (d) LSN validation set (n=3). The R^2 values are reported for each model.	34
Figure 9 Sentinel 2 plots of showing relationships between actual and predicted chlorophyll-a for (a) Vaal Dam training set (n=15), (b) Vaal Dam validation set (n=7), (c) LSN training set (n=7) and (d) LSN validation set (n=3). The R^2 values are reported for each model.....	35
Figure 10 Correlation plots showing correlation coefficients between turbidity and field hyperspectral Rrs band ratios obtained from all possible two-band combinations using bands located between 400 and 900nm for (a) Vaal Dam and (b) LSN Lakes	36
Figure 11 Field hyperspectral Rrs data plots showing relationships between actual and predicted turbidity for (a) Vaal Dam training set (n=13), (b) Vaal Dam validation set (n=4), (c) LSN training set (n=7) and (d) LSN validation set (n=3). The R^2 values are reported for each model.	39
Figure 12 Landsat 8 plots showing relationships between actual and predicted turbidity for (a) Vaal Dam training set (n=16), (b) Vaal Dam validation set (n=7), (c) LSN training set (n=7) and (d) LSN validation set (n=3). The R^2 values are reported for each model.	40

Figure 13 Sentinel 2 plots showing relationships between actual and predicted turbidity for (a) Vaal Dam training set (n=16), (b) Vaal Dam validation set (n=7), (c) LSN training set (n=7) and (d) LSN validation set (n=3). The R^2 values are reported for each model.41

List of tables

Table 1 Spectral bands and resolutions of Landsat 8 and Sentinel 2 sensors (Korhonen et al., 2017).....	20
Table 2 Chlorophyll-a field hyperspectral Rrs data band correlation coefficients (R)	30
Table 3 Chlorophyll-a Landsat 8 and Sentinel 2 band correlation coefficients (R).....	30
Table 4 SVR chlorophyll-a results.....	32
Table 5 Turbidity field hyperspectral Rrs data correlation coefficients (R)	37
Table 6 Turbidity Landsat 8 and Sentinel 2 correlation coefficients (R)	37
Table 7 SVR Turbidity results	38

List of equations

Equation 1 - Support vector regression equation.....	23
Equation 2 – Root mean square error	24

List of Abbreviations

ENVI – Environment for Visualising Images
ESA – European Space Agency
ETM – Enhanced Thematic Mapper
FLAASH – Fast Line-of-sight Atmospheric Analysis of Hypercubes
GEMI – Global Environmental Monitoring Index
GOCI – Geostationary Ocean Colour Imager
LSN – Lakes Shinji and Nakaumi
ETM – Enhanced Thematic Mapper
MERIS – Medium Resolution Imaging Spectroradiometer
MIP – Modular Inversion and Processing System
MODIS – Moderate Resolution Imaging Spectroradiometer
MSI – MultiSpectral Instruments
MSS – Multispectral Scanner
NASA – National Aeronautics and Space Administration
NIR – Near InfraRed
OLI – Operational Land Imager
RF – Random Forest
RFR – Random Forest Regression
RMSE – Root Mean Square Error
Rrs – Remote sensing reflectance
SPOT – Satellite Pour l’Observation de la Terre
SVM – Support Vector Machine
SVR – Support Vector Regression
SWIR – ShortWave InfraRed
TM – Thematic Mapper
TSS – Total Suspended Solids

Chapter one

1. Introduction

1.1 Background

Lakes and reservoirs such as dams are critical resources which have an essential role in ecosystems. Pollution, as a result of human activities, has resulted in the decline of water quality in the majority of these ecosystems globally (Bonansea et al., 2019). According to Mushtaq and Lala (2017, p274), “more than half of the world’s populations live adjacent to water bodies and carry out activities that increase aquatic stressors such as anthropogenic eutrophication and algal blooms”. To meet the demands of agriculture, industry, conservational and ecosystem uses, as well as basic human needs, water of sufficient quantity and quality is required (Matthews and Bernard, 2015). It is therefore essential that the quality of water in these lakes and reservoirs be constantly monitored to ensure clean and safe water for all its uses.

Water quality indicators, such as chlorophyll-a concentration, total suspended solids, Secchi disk depth, temperature, pH, turbidity, salinity, total phosphorus and dissolved organic carbon are often used to assess water quality (Haji Gholizadeh et al., 2016). The most commonly used parameter to assess the quality of water in inland water bodies is chlorophyll-a, due to its role as the primary photosynthetic pigment of all plant life, with its concentration indicating phytoplankton biomass and eutrophication (Wang et al., 2017). Turbidity is a measurement of the ability of light to penetrate the depth of a water body and consequently provides quantitative information as to the state of the quality of the water (Wu et al., 2014). Water bodies are broadly divided into case one waters whose optical properties are primarily a result of the effects of phytoplankton and other organic materials, and case two waters whose optical properties are primarily determined by non-organic materials such as mineral particles (Gholizadeh et al., 2016).

Water quality indicators are normally determined by collecting water samples and analysing them in the laboratory and taking in situ measurements in the field (Haji Gholizadeh et al., 2016). The disadvantage of using these traditional methods is that they are laborious, time consuming and expensive (Abdelmalik, 2018). Investigating

water quality trends and variations resulting from spatial and temporal effects using the traditional methods is difficult, if not impossible to achieve (Gholizadeh et al., 2016). Accessibility of some inland water bodies due to factors such as surrounding topography, remoteness and expanse of the water body can also hinder traditional methods of water quality monitoring (Gholizadeh et al., 2016). “Remote sensing tools provide spatial and temporal views of surface water quality parameters that are not readily available from traditional measurements, thus making it possible to monitor the landscape effectively and efficiently, identifying and quantifying water quality parameters and problems” (Ritchie et al., 2003, p695). As Hansen et al., (2017, p1) outlines, “the general process of developing an empirical remote sensing model for water quality parameters typically involves downloading and processing of remote sensing imagery, collecting coincident (or near-coincident) field measurements of the specific parameter being investigated, and using regression or other statistical modelling techniques to develop a relationship between the field-measured concentrations and remotely sensed reflectance from the corresponding pixel or group of pixels”.

The remote sensing of water quality for inland water bodies was previously hampered by a lack of appropriate sensors (Palmer et al., 2015). The launch of the new multispectral sensors Landsat 8 and Sentinel-2 has opened unprecedented opportunities for long-term high-frequency monitoring of inland water bodies (Mandanici and Bitelli, 2016). The imagery from these sensors is freely available at the respective websites making it cheaper to use compared to the commercial sensor images that are available at a high cost.

There are however some challenges faced in using remote sensing data for estimation of inland water quality parameters. In water quality remote sensing, the time at which the satellite image is taken should be a match or a close match to the time of field sample data collection (Bonansea et al., 2019). The time difference between the date of the satellite overpass and the date of field data collection is known as the time window and this is determined by natural variations of the process being measured such as algal blooms (Bonansea et al., 2019). When using historical records, it is difficult to find exact matches between satellite and field data, which often results in close matches being used instead (Hansen et al., 2017). This shows remote sensing is not entirely independent of traditional methods of water quality monitoring (Masocha

et al., 2018). Moreover, only cloud free images are useful for remote sensing therefore weather conditions such as cloud cover affect field data collection missions on planned dates (Masocha et al., 2018). According to Moses and Miller (2019, p1), “atmospheric correction, though successful for open-ocean waters, is still a challenge for coastal, estuarine, and inland waters because the widely variable and complex optical conditions encountered in these waters invalidate some basic assumptions in typical atmospheric correction models”. For example, the assumption that SWIR reflectance is from aerosols in aquatic correction routines results in it being subtracted from the rest of the spectra during the step of aerosol removal (Moses and Miller, 2019). However, it has widely been documented that SWIR signals over inland water pixels are due to strong adjacency effects most of the time and not aerosols (Kuhn et al., 2019). The depth to which light is able to penetrate water also limits remote sensing data, which means that the estimates of algal populations or other water quality parameter measurements are confined to the layers close to the water surface (Hansen et al., 2017). However, there are varying levels of algal populations with increase in depth, and near surface measurements may not be representative of the entire water body. Spaceborne hyperspectral or airborne data is expensive, as well as the required equipment for in situ field hyperspectral measurement, and this is also a major hinderance to using remote sensing methods for monitoring water quality (Gholizadeh et al., 2016).

“South Africa is a semi-arid country with inefficient management of water that results in overall deterioration in its quality” (Plessis et al., 2015, p648). As outlined by Malahlela et al. (2018), South Africa is currently faced with a myriad of water quality issues which all centre around eutrophication. Japan, however, generally has good water quality for human consumption (Kataoka, 2011). However, its coastal areas and inland water bodies close to human activity often receive more pollutants resulting in poor compliance with the set standards for human consumption (Kataoka, 2011). Lake Shinji and Lake Nakaumi are experiencing eutrophication and consequently a decline in the quality of water with some algal bloom occurrences typical of water bodies near urban areas (Sakuno et al., 2007).

1.2 Problem statement

The Vaal Dam and Lakes Shinji and Nakaumi are small inland water bodies that need constant monitoring because of the importance of their water for human activities.

These include domestic, industrial and agricultural uses. “The Vaal River catchment is part of the Vaal River system, which is deemed to be the ‘workhorse’ of South Africa as it is located within the economic heart of the country” (Plessis et al., 2015, p661). There are conflicting uses of water in the Vaal dam with water being used for domestic and industrial uses on one hand and the same water body being used for disposal of industrial and domestic effluent. The latter has resulted in high levels of dissolved nitrates and phosphates which cause algal blooms (Cloot and Roux, 1997). Malahlela et al. (2018) argues that although there is adequate public knowledge of the effects of eutrophication in South Africa, the knowledge on spatial and temporal distribution of chlorophyll-a within the Vaal Dam is lacking. This information is needed for reservoirs to be managed more efficiently. Moreover, recent years have seen a further increase in the significance of water quality monitoring in the Vaal Dam because of problems such as uneven rainfall distribution due to climate change, acidic mine drainage and increased demand of water caused by rapid urbanization (Sakuno et al., 2018). Because of the importance of these water bodies for human activities, their water quality must be constantly monitored. Traditional methods alone are not enough to provide the constant monitoring required. As outlined by Masocha et al. (2017), field measurements are essential, but they have the disadvantages of being expensive, time consuming and inefficient. This is especially significant in African countries where resources for water quality monitoring are often scarce. Even in countries like Japan where adequate resources are available, traditional methods, being labour intensive and time-consuming mean that it is not feasible to provide a simultaneous water quality database on a regional scale (Gholizadeh et al., 2016). Moreover, spatial and temporal variations in water quality cannot be identified using traditional methods (Gholizadeh et al., 2016). Remote sensing, therefore, provides an affordable and efficient means of constantly monitoring water quality.

Remote sensing satellite sensors are increasingly being used for water quality monitoring because of their efficiency and effectiveness even in inaccessible and complex water bodies. As Toming et al. (2016) outline, ocean colour sensors such as Moderate Resolution Imaging Spectroradiometer (MODIS), have frequent revisit time and radiometric resolution of 12-bit which is sufficient for dark objects like waterbodies. They fail, however, when it comes to small inland water bodies because of their coarse spatial resolution (300-1000m) which cannot capture the information from the small

lakes. The introduction of the new multispectral sensors, Sentinel 2 and Landsat 8 makes available free imagery of medium resolution capable of obtaining water quality data from small inland water bodies. Although the primary purpose of these sensors is for land applications, they offer unforeseen opportunities for the remote sensing of small inland water bodies (Dörnhöfer et al., 2016). Hyperspectral sensors have high spatial resolution and are good for remote sensing of inland water, but the high cost of the images leaves researchers resorting to hyperspectral field data more and more (Chen et al., 2016). There is, therefore, a need to explore the use of hyperspectral field data, Sentinel 2 and Landsat 8 bands in retrieving the water quality parameters of Vaal Dam and Lakes Shinji and Nakaumi (LSN) to effectively monitor the quality of these water bodies. As outlined by Murugan et al. (2016), in optical remote sensing, the additive effect of chlorophyll-a and other components such as minerals give the radiance from a water body. Water bodies have varying amounts of minerals and other constituents. As a consequence, spectral bands suitable for estimation of chlorophyll-a or other water quality parameters also vary with differing locations of the water bodies (Murugan et al., 2016). It is therefore necessary to also determine if the sensors perform the same in water bodies located in different geographical locations.

Because of the cost involved in acquiring in situ field measurements, sample sizes are often small and cannot be successfully used in conventional regression models. As pointed out by Hafeez et al. (2019), support vector regression, a machine learning method, has demonstrated strong predictability even when training samples are limited. There is, therefore, a need to explore the use of support vector regression for the Vaal Dam and Lakes Shinji and Nakaumi, given the limitation of sample size that is often experienced.

1.3 Aim and Objectives

1.3.1 Aim

To compare Landsat 8, Sentinel 2 and field hyperspectral measurement data in mapping the quality of water in the Vaal Dam, and Lakes Shinji and Nakaumi.

1.3.2 Objectives

1. To determine Landsat 8, Sentinel 2 and field hyperspectral Rrs bands which best retrieve turbidity and chlorophyll-a in the Vaal Dam and in Lakes Shinji and Nakaumi.

2. To investigate the performance of Support Vector regression in predicting chlorophyll-a and turbidity using different remote sensing data.

1.4 Research report outline

Figure 1 shows the outline and workflow of the research report. The report is organised in five chapters:

- Chapter two is the literature review which focuses on hyperspectral and multispectral remote sensing of water.
- Chapter three focuses on the description of the two geographical study areas, how data was collected and processed, and the development of support vector regression prediction models for chlorophyll-a and turbidity.
- Chapter four summaries the results of the research for the two geographical study sites in a comparative manner.
- Chapter five is divided into the discussion, conclusion and finally, limitations and recommendations of the research.

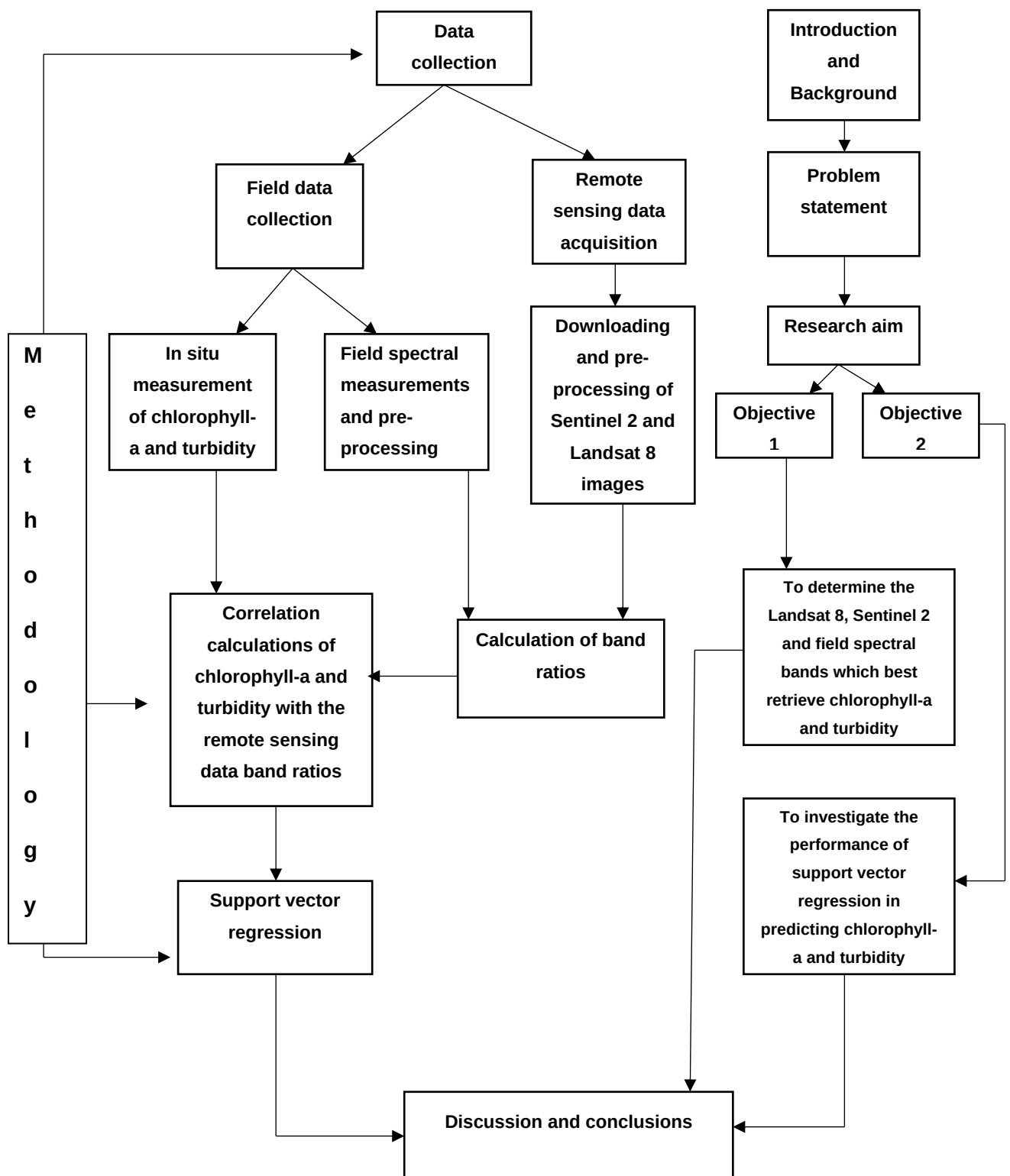


Figure 1 Outline of the research report showing the sequential workflow that was followed

Chapter two

2. Literature Review

Anthropogenic pressures on lake and reservoir ecosystems have resulted in threats such as eutrophication and sediment loading (Kuhn et al., 2019). Chlorophyll-a is a fundamental water quality parameter which can be used to assess eutrophication and detect algal blooms. Low and high chlorophyll-a levels indicate good and poor water quality respectively (Coskun et al., 2008). “Turbidity is mainly caused by the concentration of suspended matter in the water column and causes light to be scattered and absorbed rather than transmitted in straight lines through the sample” (Saylam et al., 2017, p195). As with chlorophyll-a, low and high levels of turbidity indicate good and poor water quality respectively. The environmental threats both chlorophyll-a and turbidity pose are such that only remote sensing methods coupled with advances in computer technology can enable timeous exposure of environmental threats such as algal blooms and eutrophication as well as providing improved real-time water quality monitoring (Matthews, 2011). However, many obstacles such as the radiometric, spatial, and spectral resolution of sensors used, are an impediment to the achievement of sufficient accuracy for the remote sensing of inland water bodies (Kutser et al., 2016). Choice of an appropriate sensor depends largely on the spatiotemporal variability and the size of the water body (Hansen et al., 2017). Both multispectral and hyperspectral sensors have been used successfully in the remote sensing of inland water quality (Haji Gholizadeh et al., 2016).

2.1 Hyperspectral remote sensing for water quality

Hyperspectral remote sensing techniques have the ability to retrieve spectral data of the earth’s surface. Data retrieval is almost continuous using a large number of very narrow bands (typically 10nm) throughout the visible and infrared regions of electromagnetic spectrum with no wavelength omissions (Chen et al., 2016). “Consequently, each pixel in a hyperspectral image contains a spectrum representing the light absorbing and/or scattering properties of the spatial region represented by that pixel, and the resulting spectrum acts like a signature, which can be used to classify or estimate composition of the material it represents” (Burger and Gowen, 2011, p13). In outlining the origins and history of hyperspectral remote sensing, Pu (2017) introduces four development stages: firstly hyperspectral remote sensing

preparation before the 1980s, secondly first generation of airborne hyperspectral sensors in the early 1980s, thirdly second generation of airborne hyperspectral sensors in the late 1980s and lastly satellite hyperspectral sensors in the late 1990s. Hyperspectral satellite sensors have only recently been moving beyond the testing phase owing to instrument design limitations such as signal to noise ratio, the large quantities of data that must be transmitted and stored, and difficulties with atmospheric correction (Pu 2017). Hyperion imaging spectrometer, launched in 2001, was the first experimental spaceborne hyperspectral sensor to orbit the earth and has been used for the remote sensing of inland and near-coastal waters to retrieve detailed water quality parameter information (Matthews, 2011). Hyperspectral remote sensing has since flourished in the defence and commercial sectors because of faster computers that have been developed that can better handle the large amount of data from new and improved spectrometers (Borengasser et al., 2007).

Airborne sensors and hand-held spectrometers are used mostly for hyperspectral data acquisition (Adam et al., 2010. "A hand-held spectrometer is an optical instrument used for measuring the spectrum emanating from a target in one or more fixed wavelengths in the laboratory and the field" (Adam et al., 2010, p286). Although field spectrometers have mostly been used as calibration and validation devices of satellite measurements, they are increasingly being used on their own (Kutser et al., 2016). Field hyperspectral remote sensing is sensitive to the time of day data is collected with the ideal time period being between 10:00 a.m. and 2:00 p.m., and this must be considered when planning data collection missions (Borengasser et al., 2007). This is important in providing for maximum illumination and a favourable angle of incidence, with the aim being to minimise light spillage, shadows and tonal variation (Borengasser et al., 2007).

Sakuno et al. (2018) tested 2 band and 3 band models to find correlations with chlorophyll-and turbidity for Lakes Shinji and Nakaumi and Vaal Dam. They found that for chlorophyll-a, there was very good correlation ($R^2 = 0.7$, $RMSE = 2 \text{ mgm}^{-3}$) using a two-band model and three-band model, with remote sensing reflectance (R_{rs}) values of $R_{rs}(687)/R_{rs}(672)$ and $[R_{rs}^{-1}(687) \times R_{rs}^{-1}(672)] \times R_{rs}(832)$. For turbidity they found a good correlation ($R^2 = 0.7$, $RMSE = 260 \text{ NTU}$) using two band and three band models, with $R_{rs}(763)/R_{rs}(821)$ and $R_{rs}(810) \times [R_{rs}(730) + R_{rs}(770)]/2$ (Sakuno et al 2018). However, they used simulated Sentinel 2 data and did not use the actual satellite data.

They also used a low number of samples with a total of 10 for Lakes Shinji and Nakaumi and a total of 28 for Vaal Dam. They did not make corrections for sun and glint however, due to equipment and observation time. This could have impacted the credibility of their final spectral measurements.

Yang and Anderson (2016) estimated chlorophyll-a concentrations in different water bodies using a band ratio index that they developed. They collected hyperspectral data from various water bodies in the coastal St. Jones River Watershed, Delaware-USA, from March 2006 to May 2007 using an Ocean Optics Spectroradiometer (300–900 nm), and also collected chlorophyll-a and total suspended solids field samples for laboratory analysis. Their finding was that the new band ratio R716/R667 which they developed had a good correlation of 0.82 with in situ chlorophyll-a and that it was not sensitive to total suspended solids in different water bodies and across varied seasons (Yang and Anderson, 2016). The sampling period for the study spanned more than a year from March 2006 to May 2007. The validation dataset was however, collected in a time period that did not span all the seasons as the training dataset. The Root Mean Square Error (RMSE) for the validation dataset showed as a result, that their index did not provide accurate predictions when chlorophyll-a concentrations were lower than 15mg/m³. To avoid this, the original dataset spanning all seasons could have been split into a training and validation dataset.

In their research, Kišević et al. (2016) investigated the use of hyperspectral data estimating chlorophyll-a in River Sava, Croatia. They found the best bands to be R745/R418 with an R squared value of 0.72. The sample size of only 14 was however very small. The accuracy of the predictions could have been improved with a larger number of samples.

Although very useful in the remote sensing of water, only a few studies have used hyperspectral data for monitoring water quality especially in Africa due to high cost of both acquiring and processing the data (Adam et al., 2010). “The high dimensional complexity of hyperspectral data can be problematic in terms of image processing algorithms, an excessive demand for sufficient field samples, high cost, and overfitting when using multivariate statistical techniques” (Adam and Mutanga, 2009, p613). In order to address these challenges, multispectral sensors are increasingly being used for the remote sensing of inland water bodies (Malahlela et al., 2018).

2.2 Multispectral remote sensing for water quality

As outlined by Haji Gholizadeh et al. (2016, p3146), “Spaceborne sensors are passive sensors, which typically cover larger geographic areas compared to airborne sensors and have relatively coarse spatial and spectral resolutions”. The majority of spaceborne sensors are multispectral sensors, and compared to hyperspectral sensors, they measure relatively few, broad spectral bands (Haji Gholizadeh et al., 2016). “Moderate to very high spatial resolution satellite remote sensing systems typically include multispectral bands, designed primarily for surface classification applications, and a higher spatial but lower spectral resolution panchromatic band designed for positioning and mensuration applications” (Zhang and Roy, 2016, p1). Remote sensing techniques for monitoring quality of water have been developed starting in the early 1970s (Ritchie et al., 2003). The earliest used sensor for monitoring water quality is the Multispectral Scanner (MSS) with the Thematic Mapper (TM) becoming the most used in water quality remote sensing in the 1990s (Wang and Yang, 2019). The late twentieth century saw the emergence of multispectral sensors such as MODIS, Landsat-7 ETM (Enhanced TM) and SPOT which have greatly enhanced the data resource series available for the remote estimation of water quality parameters (Wang and Yang, 2019).

As outlined by Hansen et al. (2017), the characteristics of a sensor can greatly influence the use of the resulting model and dataset of a particular investigation. The choice of a sensor is affected by factors such as spatial resolution, spectral resolution, temporal resolution, spatial variability within a water body, length of the historic record and available resources for data processing (Hansen et al., 2017).

Due to a lack of sensors with suitable spatial resolution, studies of inland water quality based on remote sensing were mostly done using airborne data or confined to large water bodies where ocean colour sensors like MERIS or MODIS have been used (Dörnhöfer and Oppelt, 2016). These ocean colour sensors have coarse spatial resolution which is not capable of extracting water quality data from small inland water bodies. Commercial multispectral sensors such as Worldview-2 are suitable for remote sensing of small inland water bodies but the high cost of the images restricts access for many remote sensing data users. The Landsat program has been monitoring the earth for more than forty years, providing the most reliable historical data (Concha and Schott, 2016). Landsat sensors have been used in the remote sensing of water to

assess water quality indicators such as chlorophyll-a concentration, turbidity, Secchi disc depth and TSS, and largely to show how inland water quality parameters and spectral reflectance are related (Mushtaq and Lala, 2017). The previous Landsat series (Landsat 1–7) satellites however, had good spatial resolution (30–79 m) but the radiometric resolution (6–8 bits) was low and could only be used to a limited extent for mapping inland water quality parameters (Toming et al., 2016). The latest generation of Landsat 8 as well as the new medium resolution multispectral sensor Sentinel 2 satellites with a radiometric resolution of 12 bits, have opened opportunities for advanced high-frequency monitoring applications in small inland water bodies (Bresciani et al., 2018).

However, a major disadvantage of Landsat 8 is its repeat cycle of 16 days which poses problems in areas with frequent cloud cover due to a reduction in the number of cloud free images that can be used as well as instances where intra-seasonal monitoring is required (Gholizadeh et al., 2016). As pointed out by Kutser et al. (2016), many processes such as algal blooms can have a life span that is less than the 16 days and fail to be picked up using Landsat 8 even when consecutive cloud-free images are available. The primary purpose of the Sentinel 2 mission is to allow frequent monitoring of worldwide land use or landcover change at all spatial levels (Pahlevan et al., 2017). However, the similarities of Sentinel 2 to Landsat 8 mean that Sentinel 2 has also opened opportunities for mapping water quality parameters enabling the provision of high-quality remote sensing products over coastal and inland waters (Pahlevan et al., 2017). Sentinel 2 has the advantage of a higher temporal resolution of 5 days for the combined Sentinel 2A and 2B and combining Landsat 8 and Sentinel 2A/B data, a multispectral global coverage with a pixel size of 10–30m is available approximately every 3 days (Bresciani et al., 2018).

Yang and Anderson (2016) used Landsat 8 to map chlorophyll-a in Jordan Lake, a freshwater lake in different seasons using individual bands and band ratios. Their finding was that chlorophyll-a concentration could be estimated using a different spectral ratio index they developed in different seasons with good accuracy. The most useful band ratios for retrieving chlorophyll-a in their study were found to be B5/B3 and B3/B1. However, their study was confined to summer and autumn only neglecting winter and spring. Their findings can therefore not be applied throughout the year.

Masocha et al. (2018) tested the utility of Landsat 8 in mapping water quality in two subtropical lakes. The water quality parameters investigated were chlorophyll-a and turbidity. Least squares regression was used for the statistical analysis of their work and their findings were that bands B4 and B5/B4 were the best fit for chlorophyll-a. For turbidity the band ratio B4/B2 was the best fit. However, they acknowledge that the algorithms they used cannot be used in different seasons and in other lakes and recommend further testing of algorithms.

Kutser et al. (2016) researched the importance of Sentinel 2 in the remote sensing of black lakes which are lakes with a high amount of coloured dissolved organic matter. In these water bodies, remote sensing sensors can only detect the only water-leaving signal at two peaks near 710 nm and 810 nm. The height of the 810 peak was found to be in good correlation with chlorophyll-a. Sentinel-2 band 5 (705 nm) was found to be capable of detecting the peak and hence mapping chlorophyll-a. Landsat 8 could not detect the peaks (Kutser et al., 2016). However, they could not validate their atmospheric correction methods and their in-situ database was limited. Liu et al. (2017) used Sentinel 2 to map suspended particulate matter in Poyang Lake, China and found that Sentinel 2 bands B4 and B7 have the best fit for retrieving suspended particulate matter.

Yadav et al. (2019) compared Landsat 8 and Sentinel 2 in mapping chlorophyll-a in Lake Biwa, a freshwater lake, and the coastal water of Wakasa Bay in Japan. Both satellite sensors achieved better accuracy for the coastal water compared to the inland water. They also found that there was a slight underestimation of chlorophyll-a by Landsat 8 with increase in chlorophyll-a concentrations. Sentinel 2 results were found to be relatively better than Landsat 8 results for the freshwater lake (Yadav et al., 2019). The in-situ data sample was small however, and their study could have benefited from a larger dataset.

The research of Bonansea et al. (2019) investigated the capabilities of Landsat 8 and Sentinel 2 in estimating Secchi disc depth in Cassaffousth Reservoir (Córdoba, Argentina). Both sensors were found to be suitable for mapping inland water quality. However, they found that although Sentinel 2 has higher spatial resolution, there was no evidence of it achieving results better than Landsat in assessing Secchi disc depth in medium sized water bodies (Bonansea et al., 2019). Although they achieved good

results, their findings are specific for the lake they investigated and cannot to applied to other lakes due to the optical complexity of freshwater.

Chapter three

3. Methodology

The water quality parameters investigated in this research are chlorophyll-a and turbidity.

3.1 Study area

The study areas are Vaal Dam in South Africa located at 26°56'49"S and 28°14'30"E (Figure 2), and Lake Shinji located at 35°26'N, 132°57'E and Lake Nakaumi located at 35°32'N, 133°09'E in Japan (Figure 3). The Vaal Dam is a water body that is located at the boundary between three provinces in South Africa namely Gauteng, Mpumalanga and the Free State (Malahlela et al., 2018). The largest portion of Vaal Dam catchment (60%) is located in the Free State Province with the remainder being located in the Gauteng and Limpopo provinces (Bertasso, 2008). The construction of the Vaal dam was done in 1938 with a dam wall height of 54.2 m which was later elevated to 63.5 m in 1985 due to increased volumes of water (Malahlela et al., 2018). The Vaal Dam, with a surface area of approximately 32107 hectares and a capacity of 2536 million cubic metres, supplies water to Gauteng province, the industrial hub of South Africa (Bertasso, 2008). As outlined by Birch (1983), the geology of the Vaal Dam comprises mostly of rocks of the Karoo supergroup which when eroded result in predominantly illitic sediments with small quantities of smectite and kaolinite sediments. The Vaal Dam has a sub-tropical dry savannah climate with mean annual evaporation of 1 300 mm which far exceeds the annual rainfall of approximately 600 mm (Jury, 2016).

Lakes Shinji and Nakaumi are slightly salty lakes situated at the border between Japan's Shimane and Tottori prefectures and have high concentrations of suspended solids and phytoplankton (Sakuno et al., 2018). The two lakes are connected by the Ohashi River which is 7.3-km long. The origin of the freshwater of Lake Shinji is the Hii River which accounts for 90% of Lake Shinji fresh water (Sakuno et al., 2018). There is eutrophication and formation of a layer of water with little oxygen at the bottom of the lakes in summer due to increased pollution from organic substances and nutrients (Nakata et al., 2000). There are Benthic filter-feeders in both lakes, and these remove phytoplankton and detritus materials from the water thus preventing eutrophication

(Nakata et al., 2000). These Benthic feeders need to be protected through the constant monitoring of the water quality for plant biomass levels to remain low. The average transparency for both lakes is about 1m with frequent occurrences of cyanobacteria (blue-green algae) in Lake Shinji and red tide blooms in Lake Nakaumi (Sakuno et al., 2018). The geology of Lakes Shinji and Nakaumi subwatersheds consists of base rock of granite, granite porphyry, and granodiorite (Matsunaga et al., 1999). The land use around the lakes is dominated by agricultural rice paddies with Matsue City situated east of Lake Shinji (Yamamuro and Kanai, 2005). The agricultural and urban activities are a source of nutrient loading causing eutrophication in the lakes (Yamamuro and Kanai, 2005).

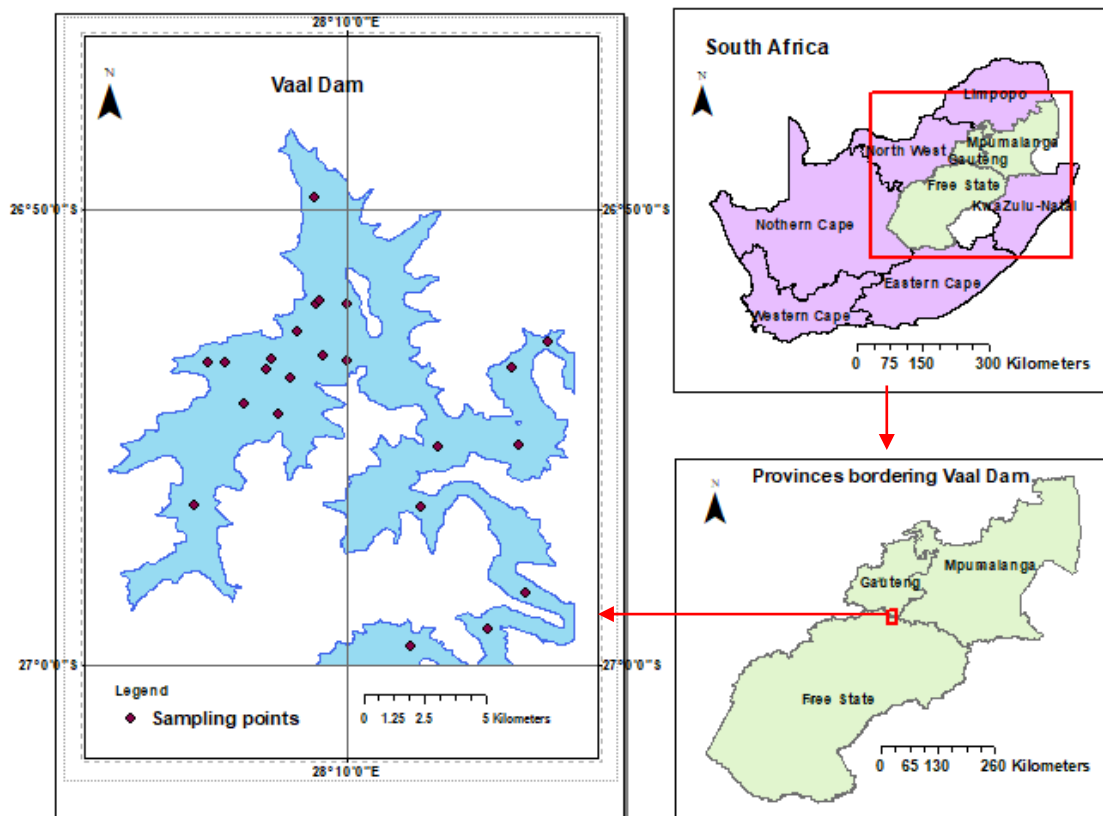


Figure 2 Vaal Dam, South Africa

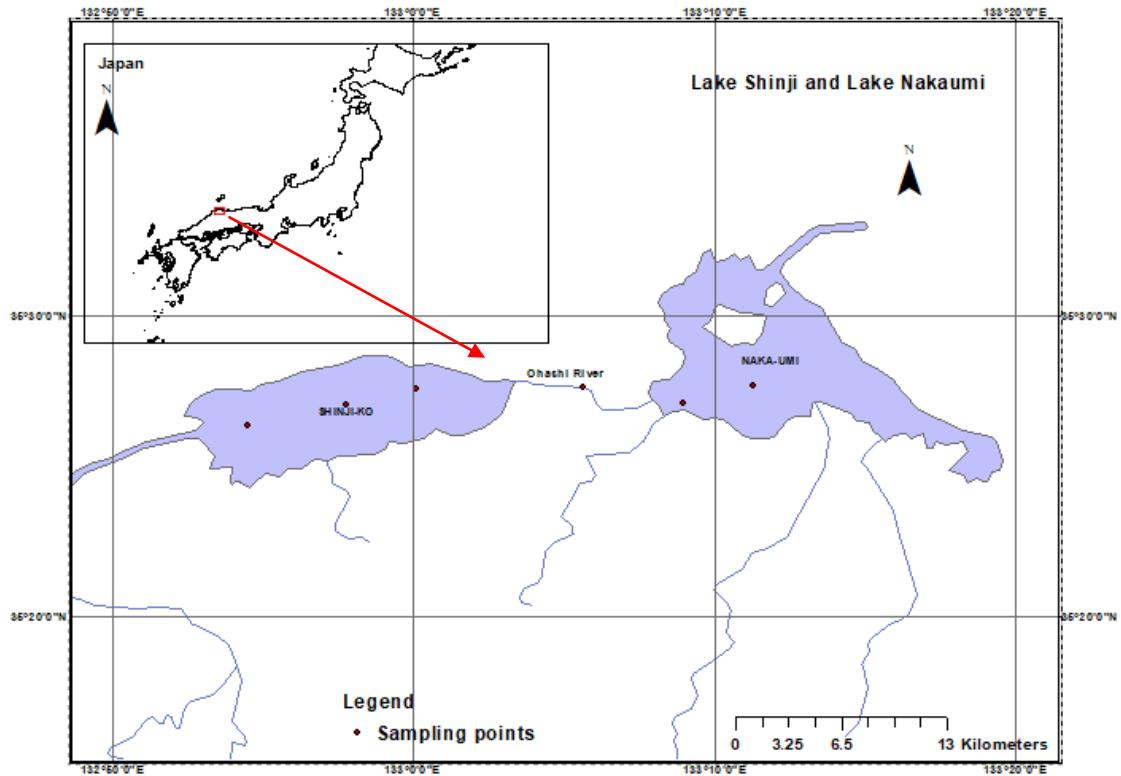


Figure 3 Lake Shinji and Lake Nakaumi of Japan

Vaal Dam and the LSN are in different parts of the world and are physically far removed from each other. However, as Sakuno et al. (2018) point out, the historical background of the lakes and the problems they are currently facing are similar and researchers in the two countries deemed it necessary to collaborate efforts for the improvement of water quality monitoring techniques in the two countries.

3.2 Field data collection

The collection of field data was done in both South Africa and Japan. Different field measurements were done including field spectral measurements and in situ measurements of water quality parameters namely chlorophyll-a and turbidity.

3.2.1 Field spectral measurements

An MS 720 portable spectral radiometer hand-held sensor produced by EKO Ltd. (Guangzhou, China) with a spectral range of 350–1050 nm and sampling interval of 3.3 nm was used to measure water reflectance (Sakuno et al., 2018). It measures spectral irradiance in the visible to near-infrared range. Measurements using the wavelengths from the 350 to 1050 nm portion of the spectrum only were collected due

to their low noise-to-signal ratio at both ultraviolet and near-infrared wavelengths (Rostom et al., 2017).

Spectral irradiance was measured immediately above the surface of the water. Measurements of solar radiation reflected by a ZB 6010 white board were also done three times and the remote sensing reflectance (R_{rs}) was then determined by calculating the respective average reflectance and dividing it by π (Sakuno et al., 2018). A major problem encountered in interpreting the data collected above the water surfaces is removal of sun and sky glint (Tan et al., 2015). In the measurement of field spectra in this study, auxiliary observations such as sky radiance were not done due to equipment and observation time, therefore removal of glint from the sun and sky was consequently not done (Sakuno et al., 2018).

Because of the large volume of data stored and analysed in hyperspectral remote sensing, it is necessary to reduce the dimension of the data and to compress it to enable effective data handling (Burger and Gowen, 2011). In hyperspectral imaging, dimension reduction is usually done through band selection to reduce the number of spectral dimensions in the hypercube (Burger and Gowen, 2011). As is done with band ratios and indices, the resulting smaller dataset after dimension reduction must contain most if not all the information in the original large dataset (Holloway and Mengersen, 2018). In this study, 2-band ratios were calculated as a means of dimension reduction. The band ratio calculations and correlations of the band ratios with chlorophyll-a and turbidity were done in R statistical software using the packages 'corrplot', 'gdata', 'gtools' and 'Hmisc'.

3.2.2 Water quality sampling

Water quality in situ field measurements and field hyperspectral measurements were done on the 26th of October 2016 at 23 sampling points on the Vaal dam. Stratified random sampling was done with a bias towards representation of all the land uses around the dam. These land uses include agricultural activities, mining, and urban build-up (Plessis et al., 2015). 23 points were randomly generated across the Vaal Dam area and the points were then uploaded into a GPS to navigate to the locations using a boat. For the LSN in situ water quality measurements and hyperspectral field measurements were done on the 12th of July 2016 and on the 12th of September 2016. Fewer points, 10 in total were collected for the LSN using the same random sampling

method used for Vaal Dam. Chlorophyll-a and Turbidity were measured in the field at each sampling point using Hydrolab Data Sonde 5x instruments.

The hyperspectral individual bands and 2 band ratios were tested for correlation with turbidity and chlorophyll-a using R statistical software to accommodate the large number of band ratios. The top 10 bands and band ratios were used in SVR models in R statistical software. Sentinel 2 and Landsat 8 correlations with chlorophyll-a and Landsat 8 were done using Microsoft excel. The top 10 bands and band ratios were also used in the SVR models using R statistical software.

3.3 Remote sensing data acquisition and pre-processing

Sentinel-2A and 2B are recent multi-spectral imaging missions with wide swath and high resolution launched by the European Space Agency (ESA) in June 2015 and December 2017, respectively (Poddar et al., 2019). The aim of the Sentinel 2 twin satellite mission is to provide enhanced continuity to the SPOT missions and to complement the Landsat programme (Bonansea et al., 2019). The twin satellites carry on board MultiSpectral Instruments (MSI), which measure the reflected solar spectral radiances in 13 spectral bands ranging from the visible to the shortwave infrared (SWIR) bands (Pahlevan et al., 2017). Landsat 8 was launched on the 11th of February 2013 as a joint venture between the United States Geologic Survey (USGS) and National Aeronautics and Space Administration (NASA) and carries onboard the Operational Land Imager (OLI) (Kuhn et al., 2019). The spatial resolution of Sentinel 2 bands is 10–60 m with a 290 km swath width (Van der Werff and Van der Meer, 2016). Landsat 8 has 11 bands with a spatial resolution of 30m (15 for panchromatic) and a swath width of 185km (Mandanici and Bitelli, 2016). Both sensors, Landsat 8 and Sentinel 2, have a radiometric resolution of 12 bit (Toming et al., 2016). Table 1 summarises the properties of Sentinel 2 and Landsat 8 bands.

The revisit time of Sentinel 2A and 2B combined is 5 days whilst that of Landsat 8 is 16 days (Van der Werff and Van der Meer, 2016). As Pahlevan et al. (2017) outline, compared to Landsat 8 OLI, Sentinel 2 MSI offers more spectral measurements within the near-infrared (NIR). These additional spectral bands have an advantage over coastal or inland waters which are optically complex and will provide opportunities for improved retrievals of total suspended solids and chlorophyll-a in hypertrophic,

coloured dissolved organic matter rich or extremely turbid waters (Pahlevan et al., 2017).

Table 1 Spectral bands and resolutions of Landsat 8 and Sentinel 2 sensors (Korhonen et al., 2017)

Landsat 8			Sentinel 2		
Band	Wavelength range (nm)	Resolution	Band	Wavelength range (nm)	Resolution
1 Coastal aerosol	433-453	30	1 Coastal aerosol	433-453	60
2 Blue (B)	450-515	30	2 Blue	458-523	10
3 Green (G)	525-600	30	3 Green	543-578	10
4 Red (R)	630-680	30	4 Red	650-680	10
			5 Red edge 1 (RE1)	698-713	20
			6 Red edge 2 (RE2)	733-748	20
			7 Red edge 3 (RE3)	773-793	20
5 Near infrared	845-885	30	8 Near infrared (NIR)	785-900	10
			8a Near infrared narrow (NIRn)	855-875	20
			9 Water vapour	935-955	60
9 Shortwave infrared/ cirrus	1360-1390	30	10 Shortwave infrared/ Cirrus	1360-1390	60
6 Shortwave infrared 1 (SWIR1)	1560-1660	30	11 Shortwave infrared 1 (SWIR1)	1656-1655	20
Shortwave infrared 2 (SWIR2)	2100-2300	30	12 Shortwave infrared 2 (SWIR20)	2100-2280	20
Panchromatic	500-680	15			

Satellite remote sensing radiation, unlike taking in situ measurements with a spectroradiometer, travels through the atmosphere for long distances on its way to the earth, during which the radiation intensity and direction are transformed by atmospheric effects such as absorption and scattering from different molecules (Murugan et al., 2016). This travel necessitates radiometric and atmospheric correction of the satellite data in order to remove the atmospheric effects enabling the development of credible relationships between remote sensing data and water quality variables (Bonansea et al., 2019). Atmospheric correction over water, unlike over land, requires greater accuracy because more than seventy percent of the top of the atmosphere signal over water is known to be from atmospheric effects and sun and sky glint from the water surface (Kuhn et al., 2019). Atmospheric correction for Landsat 8 and Sentinel 2 water applications is still a subject of much research interest with no single robust atmospheric correction algorithm for the processing of large datasets yet to be developed (Vanhellemont, 2019).

3.3.1 Landsat 8 data

Landsat 8 images for the lakes were downloaded from the United States Geological Survey (USGS) website (<https://earthexplorer.usgs.gov>). The downloaded images should have been taken at dates that were as close as possible to the field data collection dates and the images should have been free of clouds. The Vaal Dam image was taken on the 4th of October 2016 whilst that for LSN was taken on the 11th of August 2016. There was a time gap of 22 days and 31 days with the field data acquisition for the Vaal Dam and LSN lakes respectively. This large time gap is not desirable as many bio geophysical processes can occur during this time resulting in a mismatch for the image data and the in-situ data. Radiometric and atmospheric correction for the level 1C Landsat 8 images were done in ENVI software using Fast Line-of-sight Atmospheric Analysis of Hypercubes (FLAASH) method, resulting in level 2C bottom of the atmosphere reflectance images. The corrected images were transferred to ArcGIS software as TIFF images where all possible band ratios and band indices were calculated in the ArcGIS image analysis window. The raster values for the bands, band ratios and band indices were extracted to the in-situ data points for chlorophyll-a and turbidity, and the tables produced exported to excel from ArcGIS ArcMap. In excel, the correlations of band ratios with chlorophyll-a and turbidity were calculated and the top ten band ratios identified.

3.3.2 Sentinel 2 data

Sentinel 2 Level 1C images for the LSN lakes and Vaal Dam were downloaded from the USGS website (<https://earthexplorer.usgs.gov>). The images should also have been taken at dates that were as close as possible to the dates of the field data collection. The Vaal Dam image was taken on the 4th of October 2016 whilst that of the LSN was taken on the 28th of July 2016. There was a time gap between field data acquisition of 22 days for Vaal Dam and 31 days for the LSN lakes. This large time gap is not ideal as biological events such as algal blooms and influx of sediments due to rainfall can occur during this time.

The Sentinel 2 level 1C images are top of the atmosphere reflectance. A Sen2cor module available from the European Space Agency (ESA) platform was used for pre-processing of the images using the Windows command prompt to produce level 2C bottom of the atmosphere reflectance (Kuhn et al., 2019). The corrected images were also transferred to ArcGIS ArcMap software where all the possible band ratios and

band indices were calculated. As with the Landsat 8 data, the raster values of the bands, band ratios and band indices were extracted to the field data points and the tables exported to excel. The correlations with chlorophyll-a and turbidity were calculated and the best correlations used for the regression analysis.

3.4 Statistical analysis

Regression analysis is the method usually used to show relationships between the field water quality parameters and the spectral reflectance from sensors (Mushtaq and Lala, 2017). In regression analysis, the key is to select an appropriate regression method and independent variables that result in a high R^2 value from a range of zero being the minimum and 1 being the maximum (Mushtaq and Lala, 2017). Statistical regression methods are broadly divided into parametric methods which usually consist of a direct relationship between water quality parameters such as chlorophyll-a and band ratios/indices, and non-parametric regression methods which define regression models learnt from the training dataset (Mao et al., 2019). Non-parametric models can be linear or non-linear, and machine learning regression methods are non-parametric non-linear statistical methods (Mao et al., 2019).

In recent years, a number of studies have investigated the prediction of water quality parameters using machine learning models in order to deal with the problem of nonlinearity in water quality data (Li et al., 2019). Dube et al. (2015) suggest that there is a necessity for the development of more complex algorithms that utilise medium to fine-resolution multispectral data for the monitoring of water quality at smaller scales. As Kim et al. (2014, p159) point out, “although machine learning approaches such as random forest (RF) and support vector regression (SVR) have advantages such as no assumption in data distribution in predicting various measures, there are few studies that have applied such approaches in coastal or ocean water quality estimation from remote sensing data”. There are even fewer studies that have also applied these techniques in the remote sensing of inland water quality. SVR offers more advantages over other machine learning methods, because of its ability to resolve issues of high dimensional data and small samples (Abobakr Yahya et al., 2019). The time consuming, laborious and costly process of water quality field measurements usually mean that samples are limited. Regression methods that have high accuracy with small samples are therefore preferred.

Support vector machine (SVMs) were first introduced by Vapnik in the late 1960s (Rodriguez-Galiano et al., 2015). SVMs are structured in two categories, SVMs for the classification and SVMs for the regression (Gupta et al., 2015). Support vector regression (SVR) is a prediction model based on the structural risk minimization principle and aims at minimizing a bound on the generalization error (Li et al., 2016). SVR predicts an independent variable by trying to find optimum hyperplanes that separate groups of input features with similar feature characteristics (Kim et al., 2014). Advantages of SVR include its ability to solve learning problems with properties of small sample size, nonlinearity or high dimension (Wang et al., 2011). The SVR equation can be defined as

$$f(x) = w^T x + b \quad (w, x \in R^d) \quad \text{Equation 1}$$

$x \in R^d$ is any non-linear function that maps the input data into the high-dimensional feature space (Xiao et al., 2018). The unknown parameters of the model are w , a weight vector which is normal to the hyper-plane, and b , the hyper-plane bias (Rodriguez-Galiano et al., 2015), (Xiao et al., 2018). There are three hyper-parameters that must be tuned, and these are the penalization factor (C), the ε (epsilon) insensitive zone, and the kernel parameter σ (sigma) (Wang et al., 2011). Commonly used kernel functions which determine the type, complexity and precision of the SVR model include linear, polynomial, Gaussian, sigmoid, spectral angle, and radial basis functions (Kim et al., 2014). Given the trial and error process involved in finding the best parameters, the optimal ones are not easy to find and therefore it is difficult to distinctively obtain the best model (Wang et al., 2011). “The linear kernel is a special case of radial basis function, and radial basis function can better handle cases when the relationship between inputs and outputs is nonlinear” (Li et al., 2016, p73). Moreover, the radial basis function kernel involves fewer numerical difficulties than the polynomial kernel, which has more hyperparameters than the radial basis function (Li et al., 2016). The radial basis function is the most popular of the kernel functions and it is the one that was used in this study.

Wang et al. (2011) were among the first researchers to use support vector regression in the remote sensing of water quality of inland waters. They used SVR to retrieve permanganate index, ammonia-nitrogen and chemical oxygen demand. The results showed that SVR produced predictions that are better than those of the normal

multiple regression method, with limited samples (Wang et al., 2016). Kim et al. (2014) compared SVR, RFR and Cubist regression to estimate chlorophyll-a and suspended particulate matter in coastal waters using Geostationary Ocean Colour Imager. They found that SVR performed better than the other two machine learning approaches, with R^2 values of 0.91 for chlorophyll-a and 0.98 for particulate suspended matter (Kim et al., 2014). Ruescas et al. (2018) compared five regression models including RFR and SVR in retrieving coloured dissolved organic matter applied to Sentinel-2 and Sentinel-3 simulated reflectance data. Although all the regression models performed well, the SVR model produced better results than most of the models.

Support vector regression was used for all the study sites in this research to determine the regression model that best retrieves each of the water quality parameters. The R “caret” and “e1070” packages were used for support vector regression (Forkuor et al., 2017). Bands and band ratios with good correlation to chlorophyll-a and turbidity that were obtained from the field hyperspectral and satellite data were used for the regression models. The data was divided into a training data set (70%) and a test (validation) dataset (30%). For Vaal Dam this resulted in a dataset of 16 samples for the training set and 7 samples for the test set. LSN had a training set of 7 samples and a validation set of 3 samples. To assess model performance, the coefficient of determination (R^2) to account for goodness-of-fit was used, and the root mean square error (RMSE) to gauge accuracy (Wang et al., 2016). The equation used in the calculation of the RMSE is shown below:

$$RMSE = \sqrt{\left(\frac{\sum_{i=1}^N WQ_{pred,i} - WQ_{meas,i}}{N} \right)^2} \quad \text{Equation 2}$$

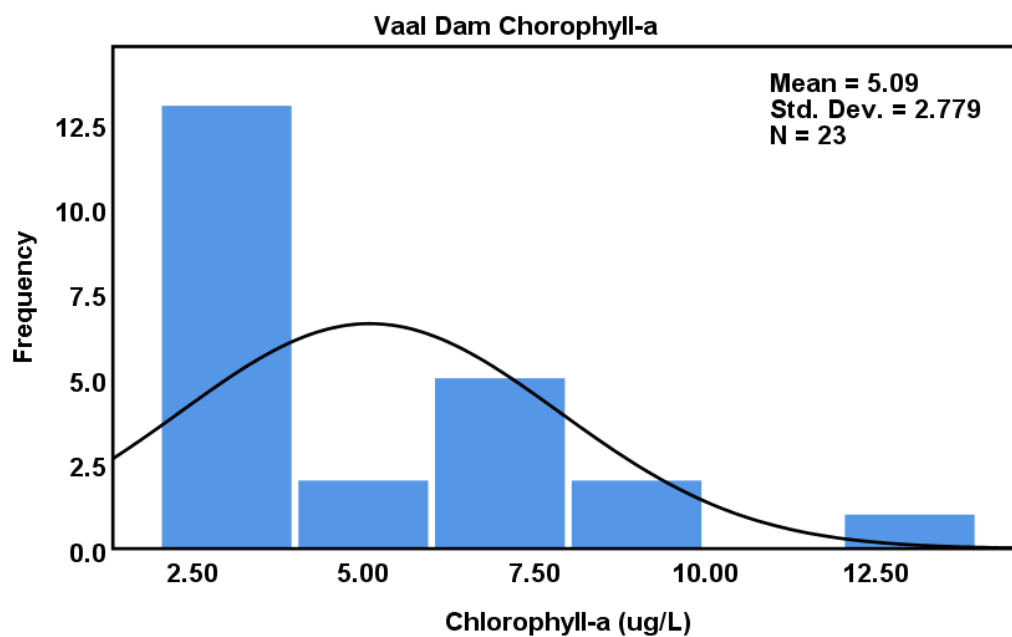
where N is the number of samples, $WQ_{pred,i}$ is the predicted chlorophyll-a or turbidity, and $WQ_{meas,i}$ is the measured chlorophyll-a or turbidity (Sakuno et al., 2018).

Chapter four

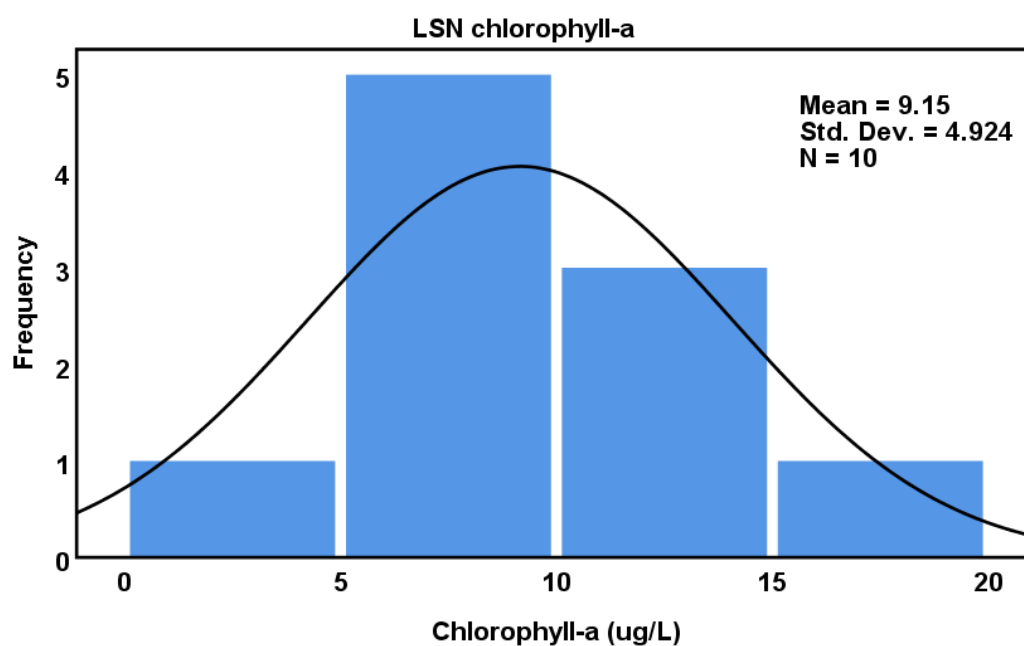
4. Results

4.1 Descriptive statistics

Descriptive statistics for chlorophyll-a and turbidity in the Vaal Dam and Lakes Shinji and Nakaumi are shown in Figure 4 and Figure 5. The mean and standard deviation of the chlorophyll-a plots show that there is more chlorophyll-a in the Vaal Dam with mean of 9.15 ug/L compared to the mean of 5.09 ug/L for the LSN lakes. The Vaal Dam chlorophyll-a data had an extreme outlier which was removed. The turbidity for Vaal Dam was significantly much higher than the Japanese lakes as shown by the mean of 380.42 NTU compared to the mean of 2.44 NTU for the LSN lakes. The turbidity levels in the Vaal Dam are therefore more than a hundred times the turbidity levels in the LSN lakes.

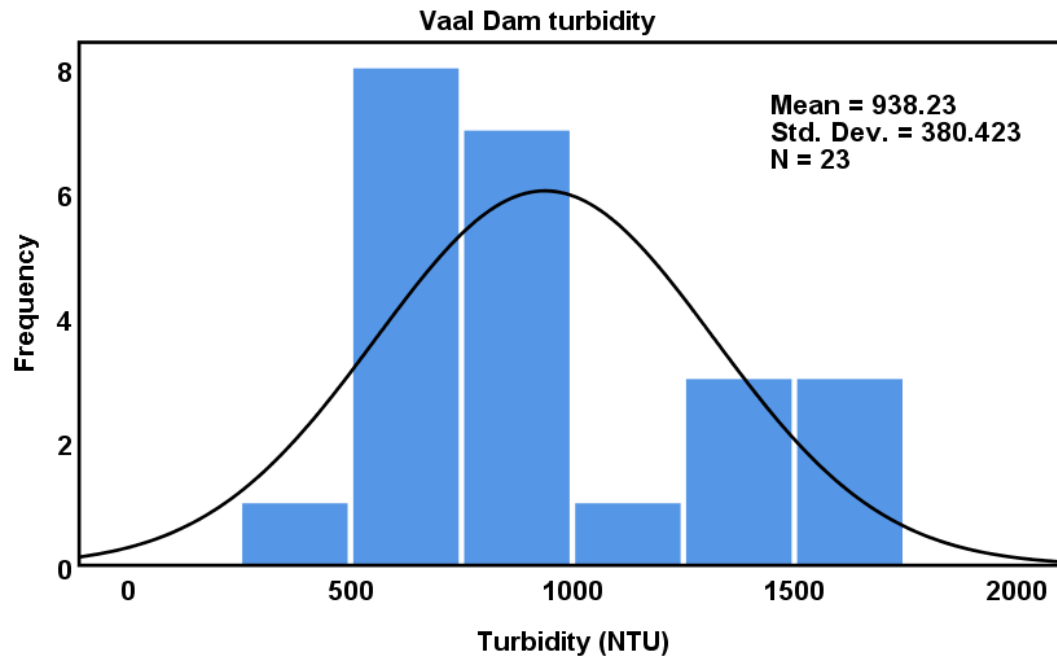


(a)

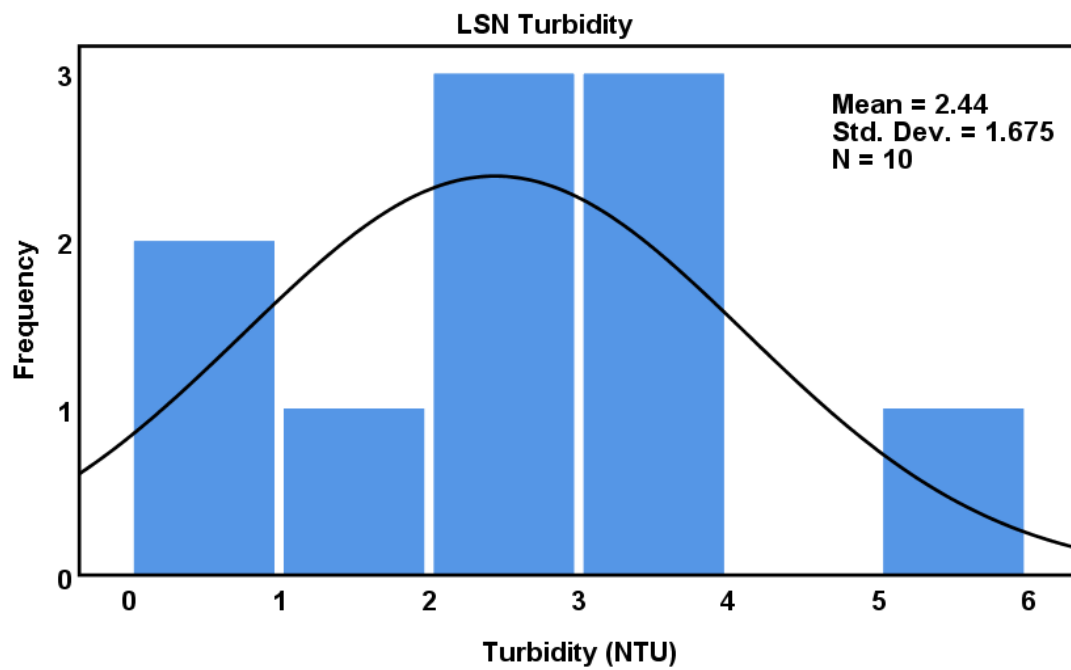


(b)

Figure 4 Chlorophyll-a summary statistics for (a) Vaal Dam and (b) LSN Lakes. The mean, standard deviation and total number of samples are reported



(a)



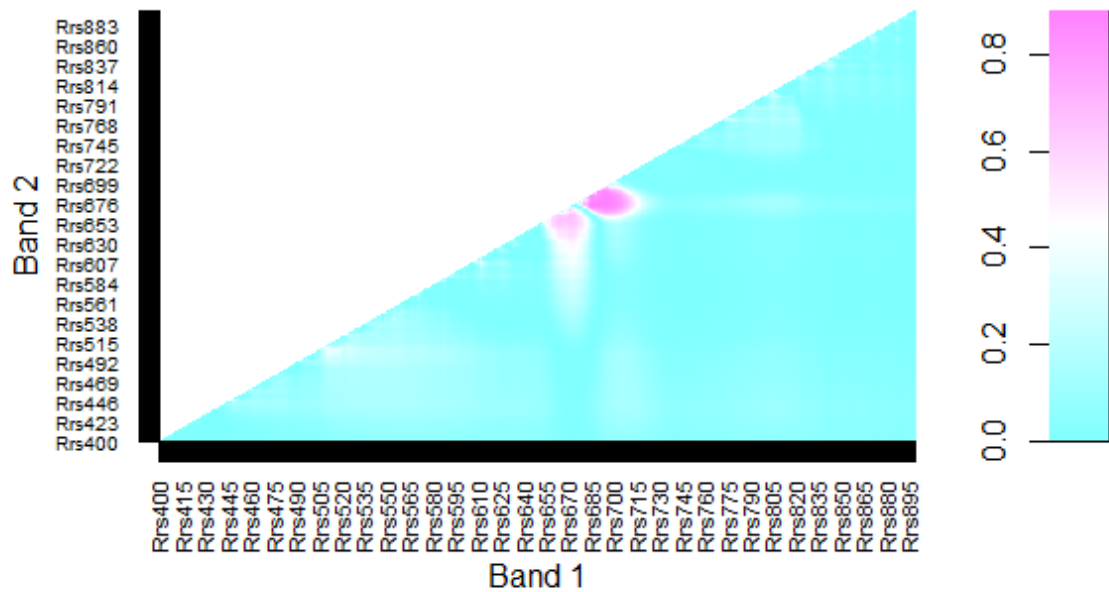
(b)

Figure 5 Turbidity summary statistics for (a) Vaal Dam and (b) LSN Lakes. The mean, standard deviation and total number of samples are reported

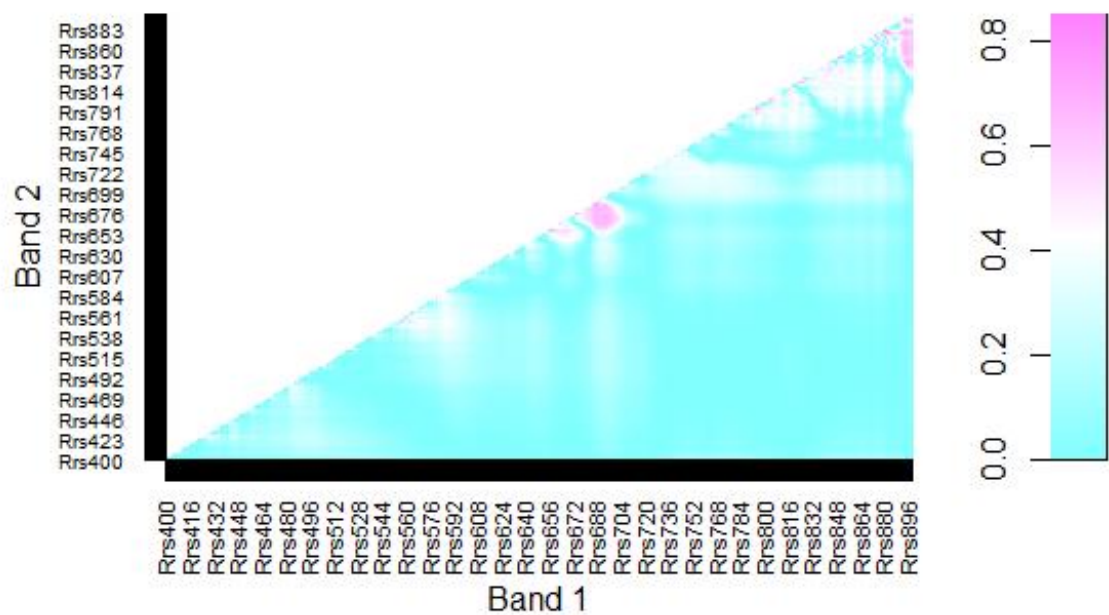
4.2 Chlorophyll-a results

4.2.1 Correlation results

The correlation plots are shown in Figure 6, and Table 2 shows the top 10 bands correlated with chlorophyll-a. The general outlook of the plots shows that the band ratios that are best correlated to chlorophyll-a are in similar regions of the electromagnetic spectrum (Rrs640 – Rrs715) for both Vaal Dam and LSN lakes. The bands best correlated to chlorophyll-a were found to be in the red region of the electromagnetic spectrum. However, the LSN lakes have an additional correlation region utilising bands in within Rrs800 – Rrs900. This concurs with the findings of Kutser et al. (2016) who found chlorophyll-a to have a peak around Rrs810. The Vaal Dam chlorophyll-a band ratio Rrs687/Rrs675 had the highest correlation of 0.8929. This concurs with the findings of (Sakuno et al., 2018) who found the best band ratio for chlorophyll to be Rrs(687)/Rrs(672). The best band ratio for the LSN lakes is Rrs881/Rrs882. The top 10 best correlated band ratios were used in the support vector regression. Support vector regression can only work with two variables at a time therefore the band ratio of the top 10 achieving the highest R^2 value was adopted as the SVR regression model algorithm for the water quality parameter being investigated.



(a)



(b)

Figure 6 Correlation plots showing correlation coefficients between chlorophyll-a and field hyperspectral Rrs band ratios obtained from all possible two-band combinations using bands located between 400 and 900nm for (a) Vaal Dam and (b) LSN Lakes

Table 2 Chlorophyll-a field hyperspectral Rrs data band correlation coefficients (R)

Vaal Dam chlorophyll-a		LSN chlorophyll-a	
Band ratio	R	Band ratio	R
Rrs687Rrs675	0.892919243	Rrs881Rrs882	0.852724184
Rrs697Rrs683	0.892645392	Rrs882Rrs881	0.852554766
Rrs691Rrs675	0.891975402	Rrs879Rrs883	0.829767584
Rrs694Rrs675	0.891554585	Rrs883Rrs879	0.826373666
Rrs693Rrs675	0.891472479	Rrs852Rrs850	0.818382912
Rrs696Rrs675	0.891089159	Rrs850Rrs852	0.81733694
Rrs697Rrs675	0.890816202	Rrs879Rrs884	0.813666268
Rrs695Rrs675	0.890691851	Rrs884Rrs879	0.807423375
Rrs696Rrs683	0.890156257	Rrs878Rrs884	0.807129142
Rrs697Rrs679	0.889635565	Rrs881Rrs883	0.806200595

Table 3 Chlorophyll-a Landsat 8 and Sentinel 2 band correlation coefficients (R)

Landsat 8 chlorophyll-a				Sentinel 2 chlorophyll-a			
Vaal Dam		LSN		Vaal Dam		LSN	
Band/Band ratio	R	Band/Band ratio	R	Band/Band ratio	R	Band/Band ratio	R
GEMI	0.570579	B2/B3	0.228627	B2/B1	0.82416	B7/B6	0.69776
MSAVI	0.532526	B3/B2	0.22163	B1/B2	0.81687	B1/B2	0.69721
B4/B5	0.53229	B5/B2	0.1712	B1	0.81048	B4	0.694576
B2/B5	0.50668	B5/B4	0.17042	B3/B4	0.68764	B6/B7	0.691845
B7/B1	0.506308	NDVI	0.16741	B4/B3	0.685425	B2/B1	0.687139
B1/B5	0.50531	B4/B2	0.16731	B1/B4	0.68352	B5/B3	0.65556
OSAVI	0.504202	B7/B2	0.16605	B4/B1	0.664088	B5	0.634927
B7/B2	0.50308	B5/B1	0.16596	B2	0.6569	B4/B6	0.63482
SAVI	0.502724	B6/B2	0.16593	VARI	0.63037	B4/B5	0.63268
NDVI	0.502038	B6/B1	0.16256	B10/B1	0.579667	B7	0.630015

Table 3 shows the best 10 Landsat 8 and Sentinel 2 bands and band ratios best correlated to chlorophyll-a for Vaal Dam and the LSN lakes. The bands and band ratios best correlated to measured chlorophyll-a were used in the SVR models. For Vaal Dam these are the Global Environmental Monitoring Index (GEMI) which utilises the red and near-infrared bands for Landsat 8, and B2/B1 for Sentinel 2. For the LSN lakes these are B2/B3 for Landsat 8 and B7/B6 for Sentinel 2. For chlorophyll-a, the best correlated band ratios utilised the coastal aerosol, blue, green, red and NIR bands. The LSN Landsat correlation values were very low achieving a value of only 0.23 with

band ratio B2/B3. This could be due to the large time window between the image acquisition date and the field data collection date.

4.2.2 Support vector regression results

Table 4 shows the SVR results for chlorophyll-a. For Vaal Dam, field hyperspectral Rrs data achieved an R^2 value of 0.92 for chlorophyll-a, whilst Sentinel-2 achieved an R^2 of 0.91 and 0.74 for Landsat 8. The RMSE value for Sentinel-2 (0.55 ug/L) was the lowest compared to that of Landsat 8 (0.86ug/L) and field hyperspectral Rrs data (1.09ug/L) which indicates better accuracy achieved with Sentinel-2. The Sentinel 2 results are comparable with the field hyperspectral Rrs data in predicting chlorophyll-a. The LSN SVR chlorophyll-a training models had R^2 values of 0.75 for the field hyperspectral data Rrs, 0.70 for Sentinel 2 and 0.13 for Landsat 8. Although the Sentinel 2 R^2 value was high, the validation dataset R squared value was low. The RMSE values were high with 2.72ug/L for the field hyperspectral Rrs data, 2.96ug/L for Sentinel 2, 5.05ug/L for Landsat 8. The poor result for Landsat 8 can be attributed to its lower spatial resolution (30m). Although SVR is known to perform well even when sample size is small, this particular dataset had too few samples in the training set (7 samples) for tuning of the SVR models to be done. Tuning of the model can only be done with training samples of at least 10. Another contributing factor to the poor results could be the differences in the field data collection dates. Although field data was collected on two dates, the samples were too few for two images to be used. The Sentinel 2 images selected that were cloud free did not coincide with either of the field data collection dates. A lot of changes can occur in the water body in the time difference between data collection and image acquisition and the chlorophyll-a and turbidity values could have significantly changed.

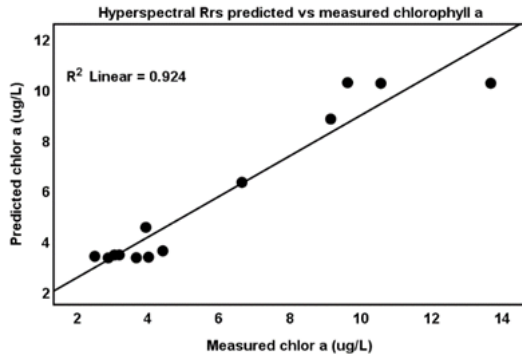
Although Sentinel-2 shows an advantage over Landsat 8 in predicting chlorophyll-a in the Vaal Dam due to the higher resolution of some bands and the presence of extra bands in the red and near-infrared region, both models have good R squared values and can be successfully used in the remote sensing of water quality. Comparing Sentinel-2 and Landsat 8 using support vector regression for LSN lakes also shows that Sentinel 2 has a distinct advantage over Landsat 8 in retrieving chlorophyll-a. The better Sentinel 2 result can be attributed to the higher spatial resolution of coastal aerosol, blue and green bands (10m) used in the chlorophyll-a band ratios. For both

Vaal Dam and LSN lakes, the field hyperspectral Rrs data performed well as expected, due to its higher resolution.

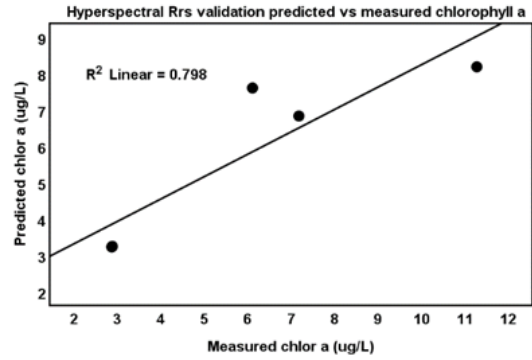
Table 4 SVR chlorophyll-a results

	Vaal Dam				LSN Lakes			
	Training set (70%)		Validation set (30%)		Training set (70%)		Validation set (30%)	
Sensor	R ²	RMSE (ug/L)	R ²	RMSE (ug/L)	R ²	RMSE (ug/L)	R ²	RMSE (ug/L)
Field spectral Rrs data	0.92	1.09	0.75	1.56	0.75	2.72	0.63	1.16
Landsat 8	0.79	0.86	0.74	0.02	0.13	5.05	0.71	1.02
Sentinel 2	0.91	0.55	0.95	0.55	0.70	2.96	0.30	1.59

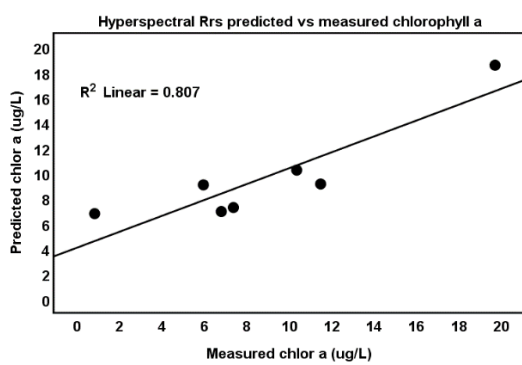
The SVR regression prediction results were plotted against the measured results. The plots are shown in Figures 7, 8 and 9. For Vaal Dam chlorophyll-a, Sentinel-2 and field hyperspectral Rrs results were very good with an R² value of 0.91 and 0.92 respectively, whilst Landsat 8 had an R² value of 0.55. The validation results were good for all of them with Sentinel achieving an R² value of 0.97, Landsat 8 0.85, and field hyperspectral Rrs data 0.80. For the LSN lakes chlorophyll-a, the predictions plots yielded R² values of 0.81 for the field hyperspectral Rrs data, 0.82 for Sentinel 2 and 0.16 for Landsat 8. The LSN Sentinel 2 and field hyperspectral Rrs data regression results show that support vector regression achieves good results even when the samples size is small. This agrees with the findings of Wang et al. (2016), who found that SVR can achieve predictions that are better than the traditional statistical regression methods, with limited samples. However, small sample size did not allow the SVR model to be tuned. The validation results were good for all of them with R² values of 0.89 for field hyperspectral Rrs data, 0.96 for Sentinel 2 and 0.92 for Landsat 8. The results of Sentinel 2 are comparable with the results of the field hyperspectral Rrs data. Comparing Sentinel-2 and Landsat 8 using support vector regression shows that Sentinel 2 has a distinct advantage over Landsat 8 for predicting chlorophyll-a for both study sites. Although Landsat 8 achieved good prediction results for the Vaal Dam, the prediction results for the LSN were poor and show that Landsat 8 cannot be used to predict chlorophyll-a using the current dataset.



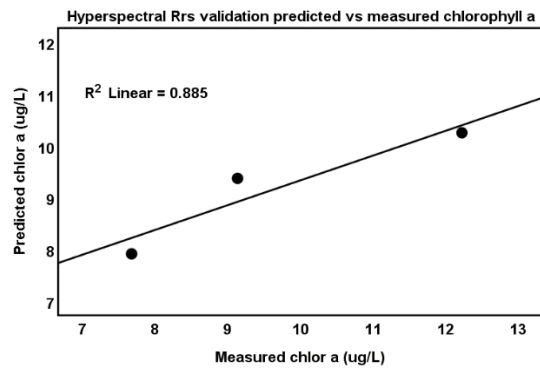
(a) Vaal Dam training set



(b) Vaal dam validation set

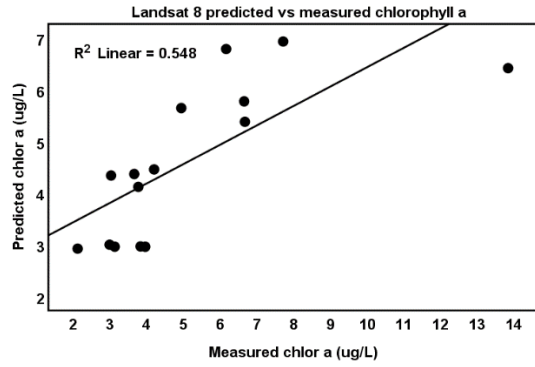


(c) LSN training set

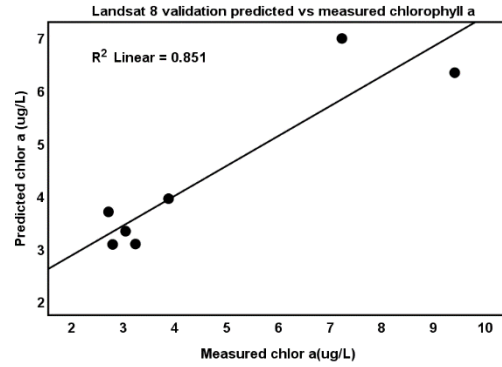


(d) LSN validation set

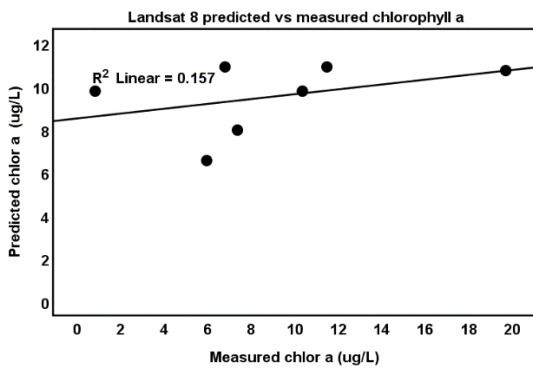
Figure 7 Field hyperspectral Rrs data plots showing relationships between actual and predicted chlorophyll-a for (a) Vaal Dam training set ($n=13$), (b) Vaal Dam validation set ($n=4$), (c) LSN training set ($n=7$) and (d) LSN validation set ($n=3$). The R^2 values are reported for each model.



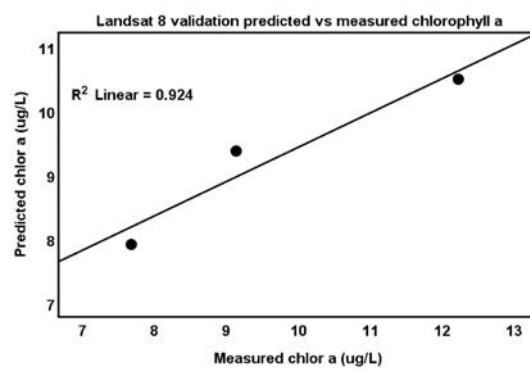
(a) Vaal Dam training set



(b) Vaal Dam validation set

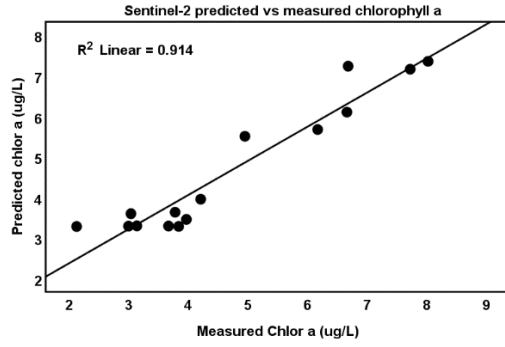


(c) LSN training set

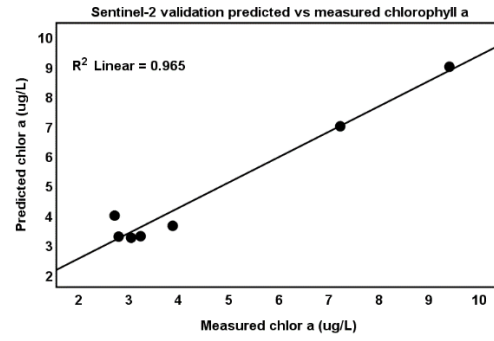


(d) LSN validation set

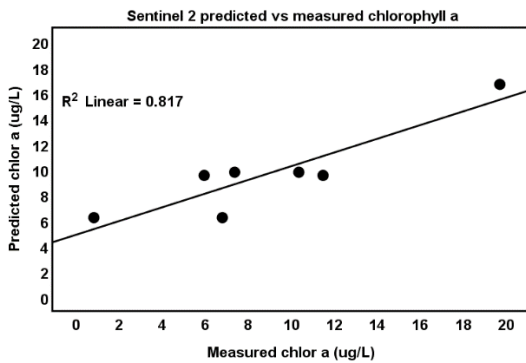
Figure 8 Landsat 8 plots showing relationships between actual and predicted chlorophyll-a for (a) Vaal Dam training set ($n=15$), (b) Vaal Dam validation set ($n=7$), (c) LSN training set ($n=7$) and (d) LSN validation set ($n=3$). The R^2 values are reported for each model.



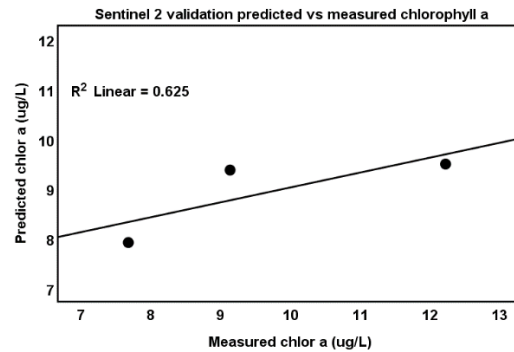
(a) Vaal Dam training set



(b) Vaal Dam validation set



(c) LSN training set



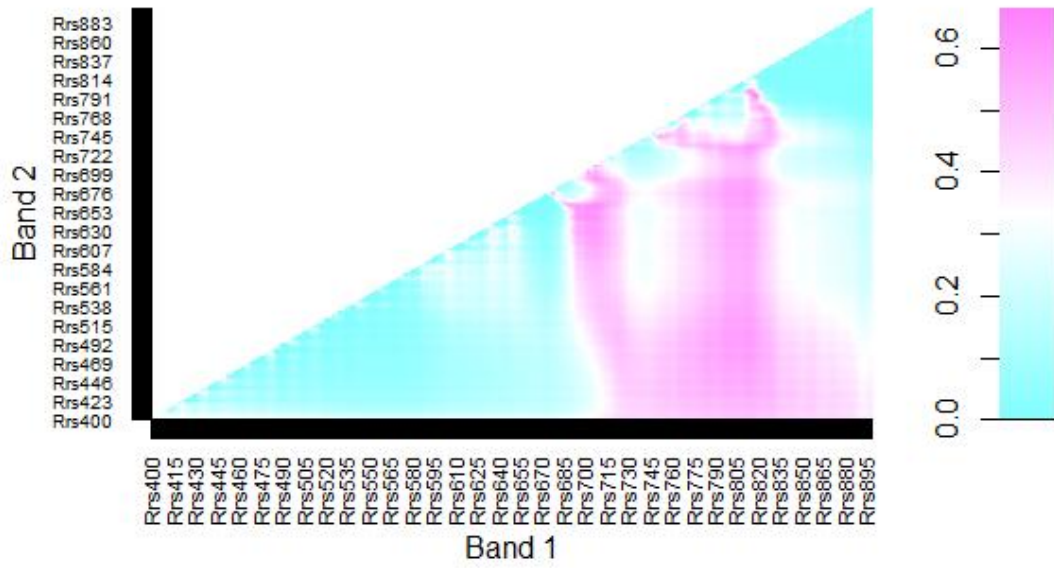
(d) LSN validation set

Figure 9 Sentinel 2 plots of showing relationships between actual and predicted chlorophyll-a for (a) Vaal Dam training set ($n=15$), (b) Vaal Dam validation set ($n=7$), (c) LSN training set ($n=7$) and (d) LSN validation set ($n=3$). The R^2 values are reported for each model.

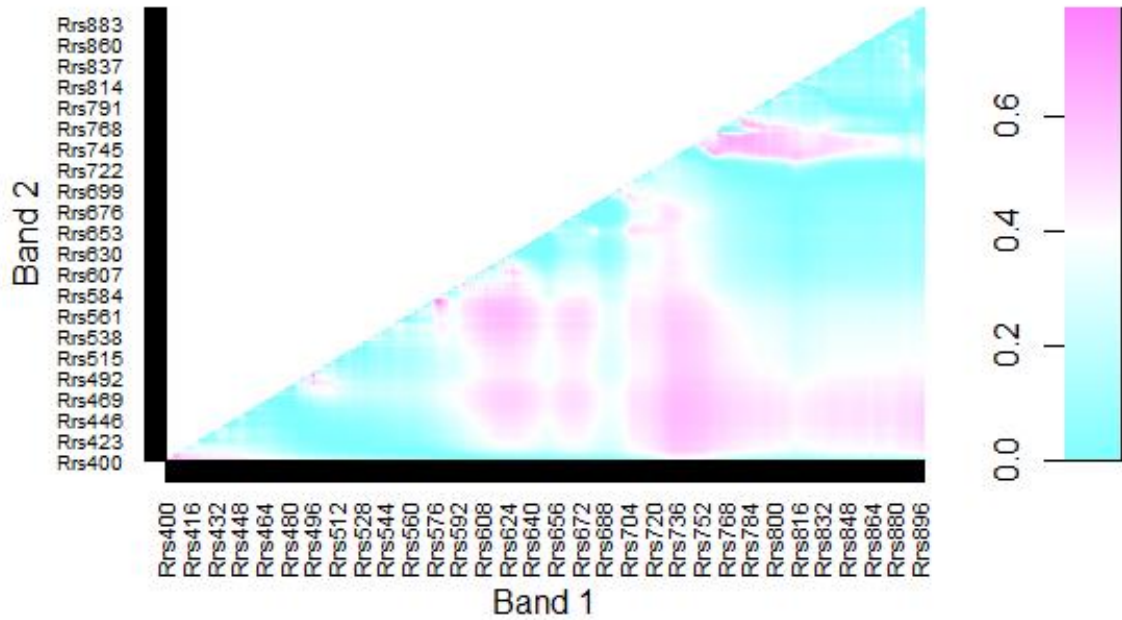
4.3 Turbidity results

4.3.1 Correlation results

Figure 10 shows the field hyperspectral Rrs data correlation plots for Vaal Dam and the LSN. Table 4 shows the top 10 band ratios with the highest correlation rates. These bands are within the near-infrared region of the electromagnetic spectrum. The best band ratio correlated to turbidity for Vaal Dam was found to be Rrs881/Rrs882 with a correlation coefficient of 0.6726. For the LSN the best band ratio was found to be Rrs580/Rrs575 with a correlation coefficient of 0.7950. The general outlook of the correlation plots shows that there are some band ratios that are common to both the LSN and Vaal Dam, but band ratios correlated to turbidity are specific to the site.



(a)



(b)

Figure 10 Correlation plots showing correlation coefficients between turbidity and field hyperspectral Rrs band ratios obtained from all possible two-band combinations using bands located between 400 and 900nm for (a) Vaal Dam and (b) LSN Lakes

Table 5 Turbidity field hyperspectral Rrs data correlation coefficients (R)

Vaal Dam Turbidity		LSN Turbidity	
Band ratio	R	Band ratio	R
Rrs819/Rrs794	0.672652618	Rrs580/Rrs575	0.795090946
Rrs816/Rrs489	0.672070819	Rrs575/Rrs580	0.791325861
Rrs816/Rrs490	0.671549498	Rrs580/Rrs577	0.782442184
Rrs816/Rrs488	0.670859669	Rrs577/Rrs580	0.780534856
Rrs816/Rrs491	0.670777967	Rrs580/Rrs576	0.765611107
Rrs815/Rrs489	0.670731148	Rrs576/Rrs580	0.763040384
Rrs816/Rrs492	0.670457369	Rrs579/Rrs575	0.760396671
Rrs815/Rrs490	0.670213928	Rrs582/Rrs575	0.75924166
Rrs816/Rrs493	0.669663019	Rrs575/Rrs579	0.758180908
Rrs815/Rrs488	0.669490782	Rrs579/Rrs576	0.754473845

The highest band ratio for Vaal Dam was found to be Rrs819/Rrs794 with 0.6726. This band ratio is also in the same spectral range as the band ratio found by Sakuno et al. (2018). For the LSN, the best band ratio was found to be Rrs580/Rrs575 with a correlation coefficient of 0.7950. The band ratios best correlated to turbidity were found to be those utilising the NIR, green and blue regions of the electromagnetic spectrum.

Table 6 Turbidity Landsat 8 and Sentinel 2 correlation coefficients (R)

Landsat 8 Turbidity (NTU)				Sentinel 2 Turbidity (NTU)			
Vaal Dam		LSN Lakes		Vaal Dam		LSN Lakes	
Band/ Band ratio	R	Band/ Band ratio	R	Band/ Band ratio	R	Band/ Band ratio	R
B4/B1	0.590997	B9/B8	0.69815	B3/B1	0.68382	B11/B1	0.724339
B4/B2	0.567733	B9/B8A	0.69815	B1/B3	0.673	B11/B2	0.723833
B2/B4	0.53537	B8/B9	0.694722	B3/B2	0.658068	B10/B5	0.721069
B1/B4	0.50532	B8A/B9	0.694722	B2/B3	0.65005	B11/B3	0.71942
B3/B1	0.484005	B9/B7	0.62411	B4/B2	0.58153	B10/B6	0.714155
B5/B1	0.483816	B7/B9	0.617332	B4/B1	0.575226	B10/B4	0.711319
B7/B6	0.480805	B9/B6	0.6153	B2/B4	0.565	B1/B11	0.70706
B3/B2	0.480535	B6/B9	0.607238	B5/B1	0.562118	B11/B4	0.705758
B5/B2	0.479998	B12/B1	0.6034	B5/B2	0.552135	B10/B7	0.692504
B6/B7	0.46706	B12/B1	0.60317	B1/B4	0.55023	B5/B10	0.692337

Table 9 shows the top ten Sentinel 2 and Landsat 8 bands and band ratios best correlated to turbidity for the Vaal Dam and LSN lakes. For turbidity, the top ten correlated bands utilised the coastal aerosol, blue, green, red and near infrared bands. The best Landsat 8 band ratio for Vaal Dam was found to be B4/B1 whilst that of the LSN was found to be B3/B1. For Sentinel 2, the best band ratio for Vaal Dam was

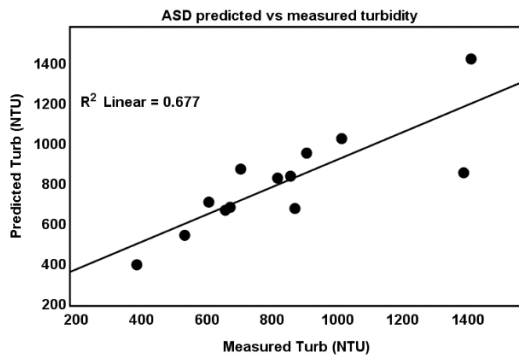
found to be B9/B8 and B11/B1 for the LSN lakes. These correlation results show that it is not the same band ratios that are correlated to turbidity in different water bodies.

4.3.2 Support vector results

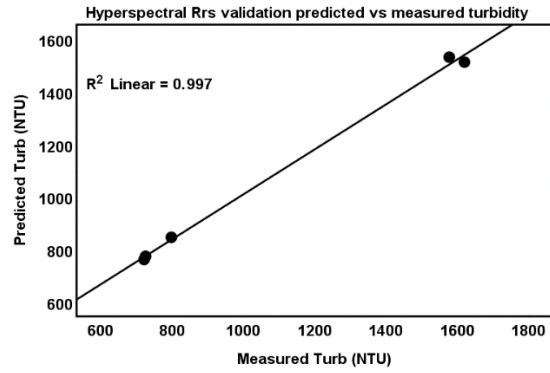
Table 7 SVR Turbidity results

Sensor	Vaal Dam				LSN Lakes			
	Training set (70%)		Validation set (30%)		Training set (70%)		Validation set (30%)	
	R ²	RMSE (NTU)	R ²	RMSE (NTU)	R ²	RMSE (NTU)	R ²	RMSE (NTU)
Field spectral Rrs data	0.68	169.5	0.98	62.9	0.68	0.73	0.99	0.18
Landsat 8	0.70	196.6	0.34	324.3	0.44	0.97	0.99	0.21
Sentinel 2	0.89	118.2	0.27	342.6	0.39	1.54	0.99	0.18

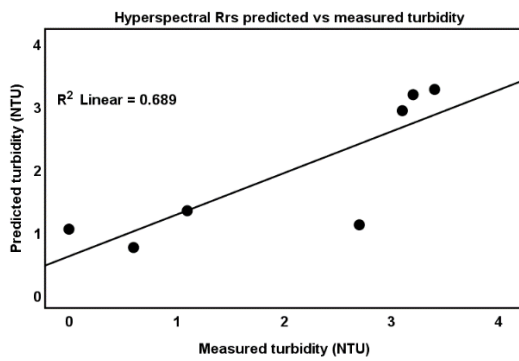
Table 7 shows the SVR turbidity results. For Vaal Dam, Sentinel-2 achieved the best result with an R squared value of 0.89 compared to the lower values achieved with Landsat 8 of 0.70 and 0.68 for the field hyperspectral Rrs data. Sentinel 2 had the lowest RMSE of 118.2 NTU whilst Landsat 8 had the highest RMSE of 196.6 NTU. The Landsat 8 B4/B2 band ratio is consistent with the finding of Masocha et al (2018) who also used the same band ratio for predicting turbidity. The Vaal Dam turbidity validation models for Landsat 8 and Sentinel 2 produced low R squared values. However, the hyperspectral data turbidity validation R squared value was high (0.98). For the LSN lakes, only the field hyperspectral Rrs data achieved a good R² value of 0.68. Landsat 8 and Sentinel 2 achieved an R² value of 0.44 and 0.39 respectively. Field hyperspectral Rrs data achieved the lowest RMSE of 0.73. Landsat 8 and Sentinel 2 had RMSE values of 0.97 and 1.09 respectively. All the validation datasets achieved high R² values of 0.99. These results show that for predicting turbidity, field hyperspectral Rrs data performs better than the multispectral sensors Landsat 8 and Sentinel 2. Apart from the higher resolution of the hyperspectral data, its performance was better because the data was collected on the same day as the in-situ turbidity measurements, whereas the Landsat 8 and Sentinel 2 satellite overpass and in-situ measurement match ups were poor.



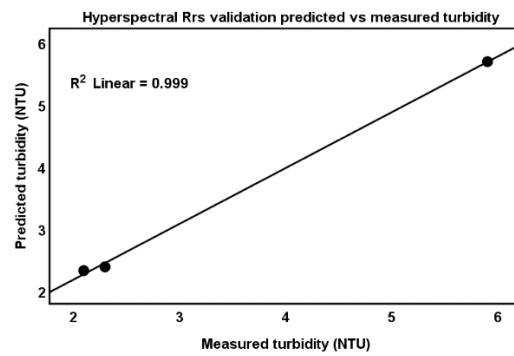
(a) Vaal Dam training set



(b) Vaal Dam validation set

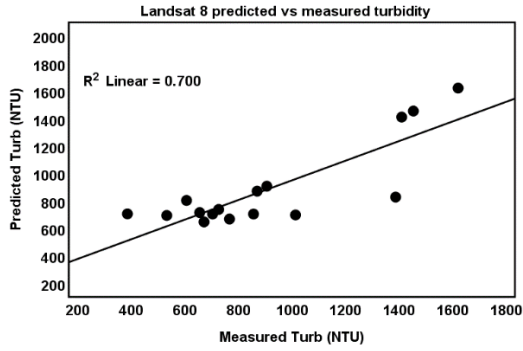


(c) LSN training set

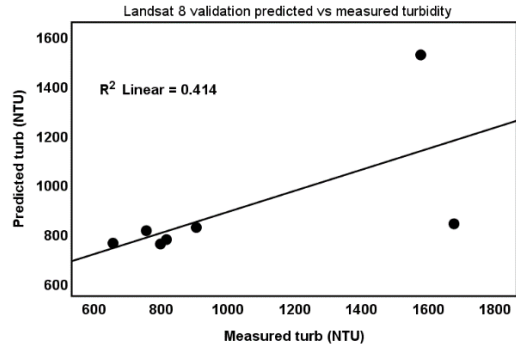


(d) LSN validation set

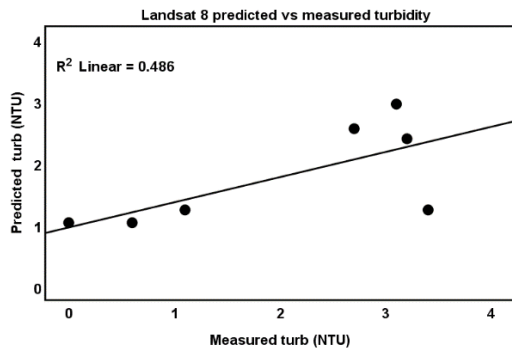
Figure 11 Field hyperspectral Rrs data plots showing relationships between actual and predicted turbidity for (a) Vaal Dam training set ($n=13$), (b) Vaal Dam validation set ($n=4$), (c) LSN training set ($n=7$) and (d) LSN validation set ($n=3$). The R^2 values are reported for each model.



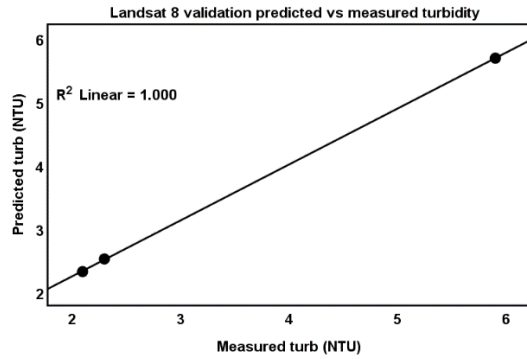
(a) Vaal Dam training set



(b) Vaal Dam validation set

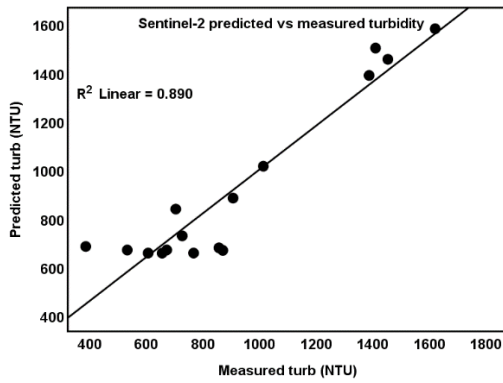


(c) LSN training set

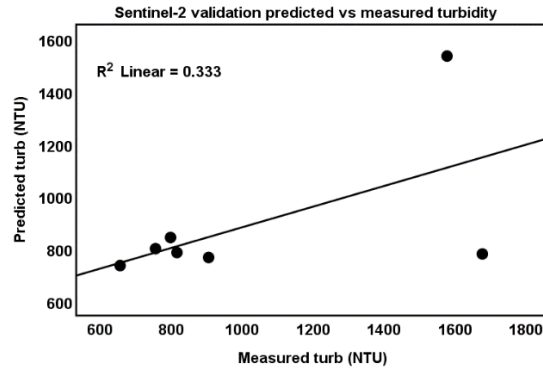


(d) LSN validation set

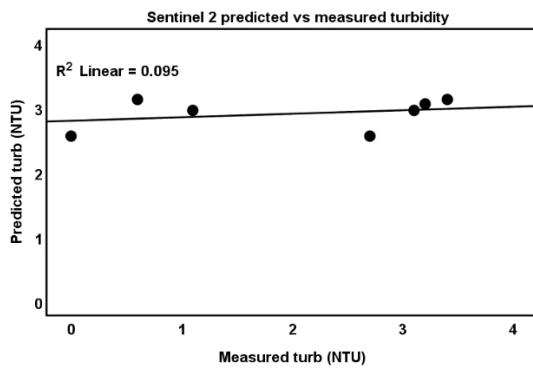
Figure 12 Landsat 8 plots showing relationships between actual and predicted turbidity for (a) Vaal Dam training set ($n=16$), (b) Vaal Dam validation set ($n=7$), (c) LSN training set ($n=7$) and (d) LSN validation set ($n=3$). The R^2 values are reported for each model.



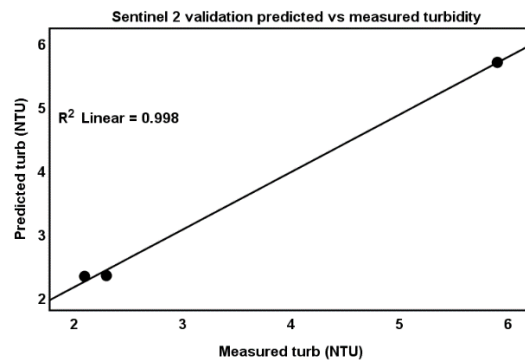
(a) Vaal Dam training set



(b) Vaal Dam validation set



(c) LSN training set



(d) LSN validation set

Figure 13 Sentinel 2 plots showing relationships between actual and predicted turbidity for (a) Vaal Dam training set ($n=16$), (b) Vaal Dam validation set ($n=7$), (c) LSN training set ($n=7$) and (d) LSN validation set ($n=3$). The R^2 values are reported for each model.

Figures 11, 12 and 13 show the predicted vs measured turbidity plots. For Vaal Dam, the Sentinel 2 algorithm achieved the best result with an R^2 value of 0.89 compared to the 0.68 and 0.70 for the field hyperspectral Rrs data and Landsat 8 respectively. However, the validation results for both Sentinel 2 and Landsat 8 were low. The field hyperspectral Rrs data validation result was good with an R^2 value 0.997. The LSN field hyperspectral Rrs data R^2 value is similar to that of Vaal Dam with 0.69. However, the LSN turbidity SVR R^2 values for Sentinel 2 and Landsat 8 were low with values of 0.095 and 0.49 respectively. All the validation R^2 values were high. These results show that of the three sensors used, the best sensor for predicting turbidity in both the Vaal Dam and the LSN lakes is the field hyperspectral Rrs data which achieved high R^2 values for both the training and validation data sets.

Chapter five

5. Discussion and conclusions

5.1 Discussion

The research aimed to use support vector regression in the comparison of Landsat 8, Sentinel 2 and field hyperspectral Rrs in predicting chlorophyll-a and turbidity in the Vaal Dam and Lakes Shinji and Nakaumi. The findings of this research concur with the findings of Hafeez et al. (2019) and Kim et al. (2014) who both found SVR to perform better than other machine learning methods such as random forest regression. In comparing Landsat 8 and Sentinel 2, the finding of Kutser et al. (2016) was that Sentinel 2 was able to detect the peak of chlorophyll-a at Rrs710, but Landsat 8 could not. This is due to the higher spatial resolution of Sentinel 2 and the extra bands in the red region of the electromagnetic spectrum which are missing in Landsat 8. This explains the advantage of Sentinel 2 that was observed in this research in the prediction of chlorophyll-a for both study sites. Yadav et al. (2019) also compared Sentinel 2 and Landsat 8 and their finding was also that Sentinel 2 performed better than Landsat 8, concurring with the findings of this research.

The lower SVR turbidity validation R^2 values for Sentinel 2 and Landsat 8 could be a result of the differences in the dates for field data collection and the taking of satellite images. Images used in remote sensing should be cloud free and this limits the number of available images that are useful for data analysis. This is one of the major drawbacks of using remote sensing data. According to Nguyen et al. (2020, p4), “several studies have discussed the time gap between in-situ measurements and the satellite overpass, indicating that a maximum of ± 8 days is reasonable in case of the stationary condition of the water environment”. The time gap in this investigation is 22 days for Vaal Dam and 31 days for the LSN lakes. The meteorological conditions of both the Vaal Dam and LSN lakes areas could not be ascertained during this time therefore there is an uncertainty in the validity of the satellite and in-situ data matchups. Boucher et al. (2018), through their investigation found that the smaller the time window between the measured chlorophyll-a concentration and the satellite image acquisition date, the better the performance of the algorithm. However, the challenge remains of matching in-situ data measurements and satellite overpass due to factors such as weather conditions. Missions to carry out in-situ data collection are

usually planned well in advance of the actual day, and it may not be possible to reschedule them in the event of adverse weather conditions such as cloud cover. In such cases, the nearest cloud free image is used as was done in this study. Although the matchups were not perfect, the prediction results as indicated by the R^2 values achieved were very good for chlorophyll-a at both study sites. However, the matchups need to be improved for Sentinel 2 and Landsat 8 for the successful prediction of turbidity.

Matchups of satellite overpass and in-situ data can be improved by using multiple sensors instead of relying on one sensor. The use of multiple sensors enables increased coverage with varying repeat cycles and extents, increasing the chances of obtaining band combinations which are best suited for specific water quality conditions (Hansen et al., 2017). As Bonansea et al. (2019) point out, using both Landsat 8 and Sentinel-2 imagery enables the acquisition of at least one cloud free image every month giving data users such as environmental managers a chance to maximise on the use of remote sensing for water quality. Moreover, although Sentinel 2 outperforms Landsat 8, Sentinel 2 is a relatively new satellite with no historical archives needed for many time series analyses. The two sensors, Sentinel 2 and Landsat 8, can therefore be used in a complementary manner as well as being used individually.

Field hyperspectral Rrs data showed consistently good prediction and validation results because there was no time gap between the hyperspectral field measurements and the in-situ chlorophyll-a and turbidity measurements. The high dimensionality of the data, however, presents problems in data handling. Although the dimension reduction was done through the calculation and use of 2 band ratios, there was still a lot of data to work with. Other dimension reduction methods that can reduce the amount of data further should be explored. Field hyperspectral Rrs data also has the disadvantages of being labour intensive (because of the excessive demand for sufficient field samples) and costly and can therefore not be used for the constant monitoring of water quality (Adam and Mutanga, 2009).

Page et al. (2019) argue that satellite and field measurements can never exactly match because of differences in spatial coverage and errors in field measurement, satellite atmospheric correction, and water quality model development. An element of human error is possible in field measurements especially when the equipment is handled by

more than one person and used on different days. The same atmospheric correction methods for Landsat 8 and Sentinel 2 were used for Vaal Dam and the LSN lakes. However, the atmospheric conditions are not the same over Japan and South Africa because the sites are in different parts of the world and the atmospheric correction methods used in this research are primarily designed for land. The research can be improved by testing different atmospheric correction methods for the two study sites, and, using the relatively new atmospheric methods that have been designed for coastal and inland water bodies such as ACOLITE. Dörnhöfer and Oppelt (2016) point out that atmospheric correction is made trickier by increased reflectance from close landmasses or clouds which affect the reflectance of water bodies, as well as challenging specification of aerosols. Dörnhöfer et al. (2016) tested three atmospheric correction methods for Sentinel 2 images, Sen2Cor, ACOLITE and MIP (Modular Inversion and Processing System) and found MIP to perform better than both Sen2Cor and ACOLITE. Both ACOLITE and MIP are specifically designed for water bodies and can be used for both Landsat 8 and Sentinel 2, whereas Sen2Cor (for Sentinel images) has been developed specifically for land and does not have any correction for the effects of water surfaces (Dörnhöfer et al., 2016).

The in-situ measurements were only done in spring and summer time seasons. However, Boucher et al. (2018) discovered that there was seasonal variation with the algorithm output. Because the algorithms used in this study were developed using data confined to mainly one season, it is possible that these algorithms may not be useful across all seasons.

The overall outlook of the results for the two sites shows that the results were better for the Vaal Dam compared to the LSN sites. This is due to the extremely small number of samples for the LSN lakes that did not allow tuning of the SVR models to obtain the best parameters for the algorithms. Although the Vaal Dam sample size is also relatively small, it was large enough to allow tuning of the SVR model to obtain the best parameters for the model. This is important for the management of water resources especially in African countries like South Africa where resources for in-situ data measurements are scarce, and do not allow for the purchase of high-resolution remote sensing images from commercial satellites or the extensive use of field spectrometers.

5.2 Conclusion

This study has shown that the new multispectral sensors Landsat 8 and Sentinel 2, and hyperspectral field data can be successfully used in monitoring the water quality parameters chlorophyll-a and turbidity for the Vaal Dam and Lakes Shinji and Nakaumi. The most important band ratios for extracting chlorophyll-a were found to be those that utilise the bands B1, B2, B4 and B5. The best bands for turbidity were found to be those that utilise bands B1, B2 B3 and B4. Sentinel 2 was found to perform better than Landsat 8 for both chlorophyll-a and turbidity in both the Vaal Dam and Lakes Shinji and Nakaumi. Both multispectral sensors are comparable to the hyperspectral data, in some cases even producing better than the field spectral data results for the Vaal Dam. Support vector regression was found to perform well even with relatively small sample sizes. However, there is a need to increase sample size to improve the results for Lakes Shinji and Nakaumi of Japan in future work. Lakes Shinji and Nakaumi and the Vaal Dam have different optical characteristics and algorithms specific to their respective optical characteristics should be used in predicting water quality parameters.

5.3 Limitations and recommendations

The major limitation of the study was a lack of adequate field data samples especially for the LSN lakes which had a combined total of only 10 samples. This research recommends that future research should utilise more field data samples in comparing Landsat 8 and Sentinel 2. Other atmospheric correction methods should also be explored for the study sites as these can affect the performance of the algorithms. Only one machine learning regression method was investigated in this research. This study also recommends that other machine learning methods such as random forest regression and artificial neural networks be investigated and compared with the SVR to find the best regression methods for retrieving chlorophyll a and turbidity. Only 2 band models and single bands were used in this study. Future research can utilise more complex algorithms to predict chlorophyll-a and turbidity. This research recommends the use of the multispectral sensors Sentinel 2 and Landsat 8, together with support vector regression for the remote sensing monitoring of small inland water bodies.

References

- Abdelmalik, K.W., 2018. Role of statistical remote sensing for Inland water quality parameters prediction. *Egypt. J. Remote Sens. Space Sci.* 21, 193–200. <https://doi.org/10.1016/j.ejrs.2016.12.002>
- Abobakr Yahya, A.S., Ahmed, A.N., Binti Othman, F., Ibrahim, R.K., Afan, H.A., El-Shafie, A., Fai, C.M., Hossain, M.S., Ehteram, M., Elshafie, A., 2019. Water Quality Prediction Model Based Support Vector Machine Model for Ungauged River Catchment under Dual Scenarios. *Water* 11, 1231. <https://doi.org/10.3390/w11061231>
- Adam, E., Mutanga, O., 2009. Spectral discrimination of papyrus vegetation (*Cyperus papyrus* L.) in swamp wetlands using field spectrometry. *ISPRS J. Photogramm. Remote Sens.* 64, 612–620. <https://doi.org/10.1016/j.isprsjprs.2009.04.004>
- Adam, E., Mutanga, O., Rugege, D., 2010. Multispectral and hyperspectral remote sensing for identification and mapping of wetland vegetation: a review. *Wetl. Ecol. Manag.* 18, 281–296. <https://doi.org/10.1007/s11273-009-9169-z>
- Bertasso, A., 2008. Ecological parameters of selected helminth species in *Labeobarbus aeneus* and *Labeobarbus kimberleyensis* in the Vaal Dam and an evaluation of their influence on indicators of environmental health.
- Bonansea, M., Ledesma, M., Rodriguez, C., Pinotti, L., 2019. Using new remote sensing satellites for assessing water quality in a reservoir. *Hydrol. Sci. J.* 64, 34–44. <https://doi.org/10.1080/02626667.2018.1552001>
- Borengasser, M., Hungate, W.S., Watkins, R., 2007. *Hyperspectral Remote Sensing: Principles and Applications*. CRC Press.
- Boucher, J., Weathers, K.C., Norouzi, H., Steele, B., 2018. Assessing the effectiveness of Landsat 8 chlorophyll a retrieval algorithms for regional freshwater monitoring. *Ecol. Appl.* 28, 1044–1054. <https://doi.org/10.1002/eap.1708>
- Bresciani, M., Cazzaniga, I., Austoni, M., Sforzi, T., Buzzi, F., Morabito, G., Giardino, C., 2018. Mapping phytoplankton blooms in deep subalpine lakes from Sentinel-2A and Landsat-8. *Hydrobiologia* 824, 197–214. <https://doi.org/10.1007/s10750-017-3462-2>
- Burger, J., Gowen, A., 2011. Data handling in hyperspectral image analysis. *Chemom. Intell. Lab. Syst., Analytical Platforms for Providing and Handling Massive Chemical Data* 108, 13–22. <https://doi.org/10.1016/j.chemolab.2011.04.001>
- Chen, F., Xiao, D., Li, Z., 2016. Developing water quality retrieval models with in situ hyperspectral data in Poyang Lake, China. *Geo-Spat. Inf. Sci.* 19, 255–266. <https://doi.org/10.1080/10095020.2016.1258201>
- Cloot, A., Roux, G.L., 1997. Modelling algal blooms in the middle vaal river: A site specific approach. *Water Res.* 31, 271–279. [https://doi.org/10.1016/S0043-1354\(96\)00248-5](https://doi.org/10.1016/S0043-1354(96)00248-5)
- Concha, J.A., Schott, J.R., 2016. Retrieval of color producing agents in Case 2 waters using Landsat 8. *Remote Sens. Environ., Landsat 8 Science Results* 185, 95–107. <https://doi.org/10.1016/j.rse.2016.03.018>
- Coskun, H.G., Tanik, A., Alganci, U., Cigizoglu, H.K., 2008. Determination of Environmental Quality of a Drinking Water Reservoir by Remote Sensing, GIS and Regression Analysis. *Water. Air. Soil Pollut.* 194, 275–285. <https://doi.org/10.1007/s11270-008-9716-x>

- Dörnhöfer, K., Göritz, A., Gege, P., Pflug, B., Oppelt, N., 2016. Water Constituents and Water Depth Retrieval from Sentinel-2A—A First Evaluation in an Oligotrophic Lake. *Remote Sens.* 8, 941. <https://doi.org/10.3390/rs8110941>
- Dörnhöfer, K., Oppelt, N., 2016. Remote sensing for lake research and monitoring – Recent advances. *Ecol. Indic.* 64, 105–122. <https://doi.org/10.1016/j.ecolind.2015.12.009>
- Dube, T., Mutanga, O., Seutloali, K., Adelabu, S., Shoko, C., 2015. Water quality monitoring in sub-Saharan African lakes: a review of remote sensing applications. *Afr. J. Aquat. Sci.* 40, 1-7–7.
- EarthExplorer - Home [WWW Document], n.d. URL <https://earthexplorer.usgs.gov/> (accessed 1.17.20).
- Forkuor, G., Hounkpatin, O.K.L., Welp, G., Thiel, M., 2017. High Resolution Mapping of Soil Properties Using Remote Sensing Variables in South-Western Burkina Faso: A Comparison of Machine Learning and Multiple Linear Regression Models. *PLOS ONE* 12, e0170478. <https://doi.org/10.1371/journal.pone.0170478>
- Gholizadeh, M.H., Melesse, A.M., Reddi, L., 2016. A Comprehensive Review on Water Quality Parameters Estimation Using Remote Sensing Techniques. *Sensors* 16, 1298. <https://doi.org/10.3390/s16081298>
- Gupta, D.K., Prasad, R., Kumar, P., Mishra, V.N., Vishwakarma, A.K., Srivastava, P.K., 2015. Support Vector Regression for Retrieval of Soil Moisture Using Bistatic Scatterometer Data at X-Band. *Int. J. Geol. Environ. Eng.* 9, 1201–1204.
- Hafeez, S., Wong, M.S., Ho, H.C., Nazeer, M., Nichol, J., Abbas, S., Tang, D., Lee, K.H., Pun, L., 2019. Comparison of Machine Learning Algorithms for Retrieval of Water Quality Indicators in Case-II Waters: A Case Study of Hong Kong. *Remote Sens.* 11, 617. <https://doi.org/10.3390/rs11060617>
- Haji Gholizadeh, M., Melesse, A.M., Reddi, L., 2016. Spaceborne and airborne sensors in water quality assessment.
- Hansen, C.H., Burian, S.J., Dennison, P.E., Williams, G.P., 2017. Spatiotemporal Variability of Lake Water Quality in the Context of Remote Sensing Models. *Remote Sens.* 9, 409. <https://doi.org/10.3390/rs9050409>
- Holloway, J., Mengersen, K., 2018. Statistical Machine Learning Methods and Remote Sensing for Sustainable Development Goals: A Review. *Remote Sens.* 10, 1365. <https://doi.org/10.3390/rs10091365>
- Kataoka, Y., 2011. Water Quality Management in Japan: Recent Developments and Challenges for Integration. *Environ. Policy Gov.* 21, 338–350. <https://doi.org/10.1002/eet.591>
- Kim, Y.H., Im, J., Ha, H.K., Choi, J.-K., Ha, S., 2014. Machine learning approaches to coastal water quality monitoring using GOCI satellite data. *GIScience Remote Sens.* 51, 158–174. <https://doi.org/10.1080/15481603.2014.900983>
- Kišević, M., Morović, M., Andričević, R., 2016. The use of hyperspectral data for evaluation of water quality parameters in the River Sava. *Fresenius Environ. Bull.* 25, 4814–4822.
- Korhonen, L., Hadi, Packalen, P., Rautiainen, M., 2017. Comparison of Sentinel-2 and Landsat 8 in the estimation of boreal forest canopy cover and leaf area index. *Remote Sens. Environ.* 195, 259–274. <https://doi.org/10.1016/j.rse.2017.03.021>
- Kuhn, C., de Matos Valerio, A., Ward, N., Loken, L., Sawakuchi, H.O., Kampel, M., Richey, J., Stadler, P., Crawford, J., Striegl, R., Vermote, E., Pahlevan, N.,

- Butman, D., 2019. Performance of Landsat-8 and Sentinel-2 surface reflectance products for river remote sensing retrievals of chlorophyll-a and turbidity. *Remote Sens. Environ.* 224, 104–118. <https://doi.org/10.1016/j.rse.2019.01.023>
- Kutser, T., Paavel, B., Verpoorter, C., Ligi, M., Soomets, T., Toming, K., Casal, G., 2016. Remote Sensing of Black Lakes and Using 810 nm Reflectance Peak for Retrieving Water Quality Parameters of Optically Complex Waters. *Remote Sens.* 8, 497. <https://doi.org/10.3390/rs8060497>
- Li, B., Yang, G., Wan, R., Dai, X., Zhang, Y., 2016. Comparison of random forests and other statistical methods for the prediction of lake water level: a case study of the Poyang Lake in China. *Hydrol. Res.* 47, 69–83. <https://doi.org/10.2166/nh.2016.264>
- Li, L., Jiang, P., Xu, H., Lin, G., Guo, D., Wu, H., 2019. Water quality prediction based on recurrent neural network and improved evidence theory: a case study of Qiantang River, China. *Environ. Sci. Pollut. Res.* 26, 19879–19896. <https://doi.org/10.1007/s11356-019-05116-y>
- Liu, H., Li, Q., Shi, T., Hu, S., Wu, G., Zhou, Q., 2017. Application of Sentinel 2 MSI Images to Retrieve Suspended Particulate Matter Concentrations in Poyang Lake. *Remote Sens.* 9, 761. <https://doi.org/10.3390/rs9070761>
- Malahlela, O.E., Oliphant, T., Tsoeleng, L.T., Mhangara, P., 2018. Mapping chlorophyll-a concentrations in a cyanobacteria- and algae-impacted Vaal Dam using Landsat 8 OLI data. *South Afr. J. Sci.* 114, 1–9. <https://doi.org/10.17159/sajs.2018/4841>
- Mandanici, E., Bitelli, G., 2016. Preliminary Comparison of Sentinel-2 and Landsat 8 Imagery for a Combined Use. *Remote Sens.* 8, 1014. <https://doi.org/10.3390/rs8121014>
- Mao, H., Meng, J., Ji, F., Zhang, Q., Fang, H., 2019. Comparison of Machine Learning Regression Algorithms for Cotton Leaf Area Index Retrieval Using Sentinel-2 Spectral Bands. *Appl. Sci.* 9, 1459. <https://doi.org/10.3390/app9071459>
- Masocha, M., Dube, T., Nhiwatiwa, T., Choruma, D., 2018. Testing utility of Landsat 8 for remote assessment of water quality in two subtropical African reservoirs with contrasting trophic states. *Geocarto Int.* 33, 667–680. <https://doi.org/10.1080/10106049.2017.1289561>
- Matthews, M.W., 2011. A current review of empirical procedures of remote sensing in inland and near-coastal transitional waters. *Int. J. Remote Sens.* 32, 6855–6899. <https://doi.org/10.1080/01431161.2010.512947>
- Matthews, M.W., Bernard, S., 2015. Eutrophication and cyanobacteria in South Africa's standing water bodies: A view from space. *South Afr. J. Sci.* 111, 1–8. <https://doi.org/10.17159/sajs.2015/20140193>
- Moses, W.J., Miller, W.D., 2019. Editorial for the Special Issue “Remote Sensing of Water Quality.” *Remote Sens.* 11, 2178. <https://doi.org/10.3390/rs11182178>
- Murugan, P., Sivakumar, R., Pandiyan, R., Annadurai, M., 2016. Comparison of in-Situ Hyperspectral and Landsat ETM+ Data for Chlorophyll-a Mapping in Case-II Water (Krishnarajapuram Lake, Bangalore). *J. Indian Soc. Remote Sens.* 44, 949–957. <https://doi.org/10.1007/s12524-015-0531-8>
- Mushtaq, F., Lala, M.G.N., 2017. Remote estimation of water quality parameters of Himalayan lake (Kashmir) using Landsat 8 OLI imagery. *Geocarto Int.* 32, 274–285. <https://doi.org/10.1080/10106049.2016.1140818>

- Nakata, K., Horiguchi, F., Yamamuro, M., 2000. Model study of Lakes Shinji and Nakaumi — a coupled coastal lagoon system. *J. Mar. Syst.* 26, 145–169. [https://doi.org/10.1016/S0924-7963\(00\)00051-8](https://doi.org/10.1016/S0924-7963(00)00051-8)
- Nguyen, H.-Q., Ha, N.-T., Pham, T.-L., 2020. Inland harmful cyanobacterial bloom prediction in the eutrophic Tri An Reservoir using satellite band ratio and machine learning approaches. *Environ. Sci. Pollut. Res.* <https://doi.org/10.1007/s11356-019-07519-3>
- Page, B.P., Olmanson, L.G., Mishra, D.R., 2019. A harmonized image processing workflow using Sentinel-2/MSI and Landsat-8/OLI for mapping water clarity in optically variable lake systems. *Remote Sens. Environ.* 231, 111284. <https://doi.org/10.1016/j.rse.2019.111284>
- Pahlevan, N., Sarkar, S., Franz, B.A., Balasubramanian, S.V., He, J., 2017. Sentinel-2 MultiSpectral Instrument (MSI) data processing for aquatic science applications: Demonstrations and validations. *Remote Sens. Environ.* 201, 47–56. <https://doi.org/10.1016/j.rse.2017.08.033>
- Palmer, S.C.J., Kutser, T., Hunter, P.D., 2015. Remote sensing of inland waters: Challenges, progress and future directions. *Remote Sens. Environ., Special Issue: Remote Sensing of Inland Waters* 157, 1–8. <https://doi.org/10.1016/j.rse.2014.09.021>
- Plessis, A. du, Harmse, T., Ahmed, F., 2015. Predicting water quality associated with land cover change in the Grootdraai Dam catchment, South Africa. *Water Int.* 40, 647–663. <https://doi.org/10.1080/02508060.2015.1067752>
- Poddar, S., Chacko, N., Swain, D., 2019. Estimation of Chlorophyll-a in Northern Coastal Bay of Bengal Using Landsat-8 OLI and Sentinel-2 MSI Sensors. *Front. Mar. Sci.* 6. <https://doi.org/10.3389/fmars.2019.00598>
- Pu, R., 2017. *Hyperspectral Remote Sensing: Fundamentals and Practices*. CRC Press.
- Ritchie, J.C., Zimba, P.V., Everitt, J.H., 2003. *Remote Sensing Techniques to Assess Water Quality* [WWW Document]. <https://doi.org/info:doi/10.14358/PERS.69.6.695>
- Rodriguez-Galiano, V., Sanchez-Castillo, M., Chica-Olmo, M., Chica-Rivas, M., 2015. Machine learning predictive models for mineral prospectivity: An evaluation of neural networks, random forest, regression trees and support vector machines. *Ore Geol. Rev.* 71, 804–818. <https://doi.org/10.1016/j.oregeorev.2015.01.001>
- Rostom, N.G., Shalaby, A.A., Issa, Y.M., Afifi, A.A., 2017. Evaluation of Mariut Lake water quality using Hyperspectral Remote Sensing and laboratory works. *Egypt. J. Remote Sens. Space Sci., WATER SCARCITY AND HARVESTING* 20, S39–S48. <https://doi.org/10.1016/j.ejrs.2016.11.002>
- Ruescas, A.B., Hieronymi, M., Mateo-Garcia, G., Koponen, S., Kallio, K., Camps-Valls, G., 2018. Machine Learning Regression Approaches for Colored Dissolved Organic Matter (CDOM) Retrieval with S2-MSI and S3-OLCI Simulated Data. *Remote Sens.* 10, 786. <https://doi.org/10.3390/rs10050786>
- Sakuno, Y., Kozu, T., Tsuzuki, Y., Matsunaga, T., 2007. Monitoring of water pollution and aquatic plants in the coastal lagoon environments using ALOS data.
- Sakuno, Y., Yajima, H., Yoshioka, Y., Sugahara, S., Abd Elbasit, M.A.M., Adam, E., Chirima, J.G., 2018. Evaluation of Unified Algorithms for Remote Sensing of Chlorophyll-a and Turbidity in Lake Shinji and Lake Nakaumi of Japan and the Vaal Dam Reservoir of South Africa under Eutrophic and Ultra-Turbid Conditions. *Water* 10, 618. <https://doi.org/10.3390/w10050618>

- Saylam, K., Brown, R.A., Hupp, J.R., 2017. Assessment of depth and turbidity with airborne Lidar bathymetry and multiband satellite imagery in shallow water bodies of the Alaskan North Slope. *Int. J. Appl. Earth Obs. Geoinformation* 58, 191–200. <https://doi.org/10.1016/j.jag.2017.02.012>
- Tan, J., Cherkauer, K.A., Chaubey, I., 2015. Using hyperspectral data to quantify water-quality parameters in the Wabash River and its tributaries, Indiana. *Int. J. Remote Sens.* 36, 5466–5484. <https://doi.org/10.1080/01431161.2015.1101654>
- Toming, K., Kutser, T., Laas, A., Sepp, M., Paavel, B., Nõges, T., 2016. First Experiences in Mapping Lake Water Quality Parameters with Sentinel-2 MSI Imagery. *Remote Sens.* 8, 640. <https://doi.org/10.3390/rs8080640>
- Van der Werff, H., Van der Meer, F., 2016. Sentinel-2A MSI and Landsat 8 OLI Provide Data Continuity for Geological Remote Sensing. *Remote Sens.* 8, 883. <https://doi.org/10.3390/rs8110883>
- Vanhellemont, Q., 2019. Adaptation of the dark spectrum fitting atmospheric correction for aquatic applications of the Landsat and Sentinel-2 archives. *Remote Sens. Environ.* 225, 175–192. <https://doi.org/10.1016/j.rse.2019.03.010>
- Wang, L., Zhou, X., Zhu, X., Dong, Z., Guo, W., 2016. Estimation of biomass in wheat using random forest regression algorithm and remote sensing data. *Crop J.* 4, 212–219. <https://doi.org/10.1016/j.cj.2016.01.008>
- Wang, X., Fu, L., He, C., 2011. Applying support vector regression to water quality modelling by remote sensing data. *Int. J. Remote Sens.* 32, 8615–8627. <https://doi.org/10.1080/01431161.2010.543183>
- Wang, X., Yang, W., 2019. Water quality monitoring and evaluation using remote sensing techniques in China: a systematic review. *Ecosyst. Health Sustain.* 5, 47–56. <https://doi.org/10.1080/20964129.2019.1571443>
- Wang, Z., Kawamura, K., Sakuno, Y., Fan, X., Gong, Z., Lim, J., 2017. Retrieval of Chlorophyll-a and Total Suspended Solids Using Iterative Stepwise Elimination Partial Least Squares (ISE-PLS) Regression Based on Field Hyperspectral Measurements in Irrigation Ponds in Higashihiroshima, Japan. *Remote Sens.* 9, 264. <https://doi.org/10.3390/rs9030264>
- Wu, J.-L., Ho, C.-R., Huang, C.-C., Srivastav, A.L., Tzeng, J.-H., Lin, Y.-T., 2014. Hyperspectral Sensing for Turbid Water Quality Monitoring in Freshwater Rivers: Empirical Relationship between Reflectance and Turbidity and Total Solids. *Sensors* 14, 22670–22688. <https://doi.org/10.3390/s141222670>
- Xiao, X., Zhang, T., Zhong, X., Shao, W., Li, X., 2018. Support vector regression snow-depth retrieval algorithm using passive microwave remote sensing data. *Remote Sens. Environ.* 210, 48–64. <https://doi.org/10.1016/j.rse.2018.03.008>
- Yadav, S., Yamashiki, Y., Susaki, J., Yamashita, Y., Ishikawa, K., 2019. Chlorophyll Estimation Of Lake Water and Coastal Water Using Landsat-8 and Sentinel-2A Satellite, in: *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. Presented at the ISPRS Technical Commission III WG III/2, 10 Joint Workshop “Multidisciplinary Remote Sensing for Environmental Monitoring” (Volume XLII-3/W7) - 12–14 March 2019, Kyoto, Japan, Copernicus GmbH, pp. 77–82.* <https://doi.org/10.5194/isprs-archives-XLII-3-W7-77-2019>
- Yang, Z., Anderson, Y.Z., 2016. Estimating Chlorophyll-A Concentration in a Freshwater Lake Using Landsat 8 Imagery.

Zhang, H.K., Roy, D.P., 2016. Computationally Inexpensive Landsat 8 Operational Land Imager (OLI) Pansharpening. *Remote Sens.* 8, 180.
<https://doi.org/10.3390/rs8030180>