Assessing the effects of sampling time on the accuracy of extracting water quality indicators through geospatial techniques: A case study of Rietvlei Dam

by

Student Name: Mokabolane Innocent Rabohale

Student No: 21015278

A dissertation submitted in partial fulfilment for the degree of Master of Science in Environmental Sciences, in the Department of Geography and Environmental Sciences,

Faculty of Science, Engineering and Agriculture

University of Venda
Thohoyandou, Limpopo
South Africa

Supervisor: Dr Oupa E. Malahlela

Co-Supervisor: Dr Lutendo Mugwedi

January 2024

ABSTRACT

Freshwater is scarce, and the Earth's ability to supply it is equally limited. Anthropogenic activities, urbanisation, and the effects of climate change degrade freshwater quality at a rapid rate. To design an effective water quality monitoring strategy that can respond to the rapid changes in surface water, regular sampling and analyses are essential. In developing countries, traditional methods of water sampling and laboratory analyses are adopted to generate water quality information. There is, however, a growing global adoption of optical remote satellite sensing for monitoring water quality, in conjunction with field surveys. One of the main challenges of optical remote sensing is sensitivity to cloud cover, which makes it difficult to synchronise with field surveys. This research addressed the question: "What is the effect of sampling time on the accuracy of extracting water quality indicators through remote sensing and geospatial techniques?", using a case study of Rietvlei Dam, South Africa. The research established the correlation between laboratory-analysed water quality and the Sentinel-2 (S2) data extracted from the months before, during and after the given field sampling date. This covered four sampling periods i.e., mid-winter, late winter, early spring, and mid-spring. Through multiple regression and accuracy analyses, the study determined the efficacy of the S2 remote sensing models for the chlorophyll-*a*, turbidity, coloured dissolved organic matter (CDOM) and total suspended solids (TSS) water quality indicators from the different field and satellite sampling periods. The S2 remote sensing models predicted the spatial distribution of the water quality indicators and high concentration area coverage of the dam in hectares (ha). S2 models were found to be best for early spring. Chlorophyll-*a* was extracted five days after field sampling at R^2= 0.79; for late winter CDOM was derived 44 days before the field sampling period at R^2 = 0.64; for late winter turbidity was optimally extracted 36 days after field sampling season at R^2= 0.54; and mid-winter TSS was extracted seven days before field sampling season at R^2=0.62. The turbidity and chlorophyll-*a* S2 models' spatial prediction of highly concentrated areas covered 36.14 and 28.92ha, respectively. In conclusion, the accuracies of mapping water quality indicators with remote sensing data and field observations differ based on an indicator measured and the time sampling difference between them. In general, the mapping accuracies decrease outside a seven-day sampling difference, whether before or after satellite image acquisition. This study contributes to the scientific understanding of remote sensing applications to the management of water quality in a given area, especially inland water quality.

Keywords: accuracy; regression models, remote sensing, sampling time, water quality

DECLARATION

I, **Mokabolane Innocent Rabohale [21015278]**, declare that this research dissertation is my original work and has not been submitted for any degree at any other university or institution. The dissertation does not contain other person's writing, unless specifically acknowledged and referenced accordingly.

Signed (Student):  Date: 16/02/2024

ACKNOWLEDGEMENTS

These acknowledgements are a tribute to all the individuals and organisations that made the academic journey worthwhile. I express my sincere gratitude to:

- Jehovah God, for the continuous strength, protection, and blessings throughout the journey of life.
- My supervisors, Dr OE Malahlela and Dr L Mugwedi. Words fail to express the amount of support, time and grooming poured in throughout the academic period.
- To the university staff, for the opportunities, space, and support.
- Tshwane municipality, for the permit and access to the water resources.
- To the REML laboratory team, for support in sampling and analyses.
- My dear parents, Phetole and Moyahabo Rabohale. Their continuous support, encouragement and love have kept me going.
- My siblings Mahlatse, Dineo and Lerato, for the emotional support and encouragement.
- To Mkhulu Geoffrey Bernard, for all the well-kept data.
- To Mkhulu Francis Girdlestone, for making sampling a safe and fun experience.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACRONYMS

Anova:	analysis of variance
BOA:	bottom of atmospheric
CDOM:	coloured dissolved organic matter
GIS:	geographic information system
GPS:	global positioning system
FD:	field date
Modis:	Moderate Resolution Imaging Spectroradiometer
NDVI:	normalised difference vegetation index
NTU:	nephelometric turbidity units
RS:	remote sensing
S2:	Sentinel-2
TSS:	total suspended solids
WQI:	water quality indicators

CHAPTER 1:
INTRODUCTION

1.1 BACKGROUND

Freshwater to sustain humans and habitats is limited. Anthropogenic activities such as urbanisation, and the effects of climate change exacerbate the declining quality of the available freshwater at a rapid rate (Khatri & Tyagi, 2015). Deterioration of surface water quality threatens the sustainability of social, ecological and economic developments globally. The declining surface water quality patterns vary notably within a country and from one country to another (Biswas & Tortajada, 2019). The rate at which surface water quality degrades prompts the need for continuous monitoring strategies with the ability to identify changes consistently, with confirmed validity (Namugize & Jewitt, 2018). Regular water sampling and analysis are components of effective water quality monitoring (DWS, 2019). Through these components, accurate information that meets the scientific basis for decision-making is generated (Guo, et al., 2020). Improved water quality information may lead to better governance strategies for freshwater (Ball, et al., 2022).

Traditional water sampling and laboratory analytical methods are well developed to generate water quality information that is accurate and meets the scientific basis for decision-making. However, poor water quality governance, increasing sampling costs, time-consuming analytical procedures, intensive labour requirements, and prolonged turnaround result in limited adaptation of the techniques, which results in the ineffective development of water management strategies (Namugize & Jewitt, 2018; Ball, et al., 2022). In short, traditional methods are ineffective in providing comprehensive temporal and spatial information of the water body (Gholizadeh, et al., 2016). To solve this problem, remote satellite sensing techniques are adopted to complement field sampling efforts (Gholizadeh, et al., 2016).

To advance the governance of water quality, geospatial techniques are increasingly adopted in water quality monitoring programmes. These techniques include remote satellite sensing, and geographic information systems (GIS), which are desirable because they provide wide spatial and temporal resolutions (Din & Zhan, 2017). Multispectral satellite sensors with free available data such as Moderate Resolution Imaging Spectroradiometer (Modis), Landsat, and Sentinel-2 are commonly used for inland water quality mapping (Xianghong, et al., 2015). Modis sensors inherently provide low-accuracy coarse spatial resolution images (250-1000 m) at high temporal resolution (1-2 days) and broad coverage. The Modis sensor's temporal resolution is ideal for regularly monitoring large land variations, but it is limited in extracting water quality of small inland water bodies (Huang, et al., 2018). On the other hand, the Landsat sensors offer spatial resolution images of 15-100m. While it is ideal for detecting most kinds of surface water bodies, its medium temporal resolution of 16 days hinders the desired outcome for frequent water quality extraction (Gholizadeh, et al., 2016). Sentinel-2 MSI sensors provide a high-quality image with a spatial resolution between 10 and 60m at a temporal resolution of five days (Huang,

et al., 2018). The Sentinel-2 MSI spectral, spatial and temporal capabilities are ideal for frequent (i.e., weekly) extraction of water quality indicators of inland waters (Potes, et al., 2018). The capabilities of Sentinel-2 sensors for monitoring water quality indicators such as chlorophyll-a, turbidity, suspended solids, and coloured dissolved organic matter (CDOM) are well established (Sent, et al., 2021). Remote sensing techniques can provide a comprehensive temporal and spatial sampling of water bodies. But their sampling capabilities are often negatively affected by dynamic atmospheric pollutants, alternating weather conditions, aerosols, and overall air quality (Joshi, et al., 2020). Further, even freely available satellite missions such as Sentinel-2 which orbit the Earth every five days and are faster at higher altitudes have limited images available for continuous water quality monitoring (ESA, 2016).

The water sampling time is usually defined as the time interval between successive collections of samples for the extraction of information about the water bodies (Rainwater & Thatcher, 1962). The frequency of sampling is often determined by the monitoring programme to respond to the needs of the water body. The quality of inland surface waters is highly dynamic and fluctuates according to direct and indirect pollution inflows. Extracting water quality information and governance of water resources remains a complex process that involves the assessment and inclusion of various environmental factors and stakeholders (Biswas & Tortajada, 2011; Ball, et al., 2022). Assessments of water quality often make use of traditional techniques of field sampling and laboratory measurements, which provide a limited temporal view. Consideration of many stakeholders is often required when the information is available. To establish the routine of continuous water quality monitoring, we need to understand the sampling time synergy between the traditional and remote satellite sensing sampling methods. This knowledge can assist in improving water quality monitoring programme strategies and governance.

1.2 PROBLEM STATEMENT

Extracting and mapping water quality indicators accurately at different sampling times is essential for identifying pollution sources and protecting surface freshwater. To design an effective management strategy that can respond to the rapid changes in surface water, regular sampling for extracting water quality information is essential. Developing countries rely heavily on traditional field sampling methods to extract water quality indicators. However, these methods are costly and affected by logistical problems, with a prolonged turnaround time, which results in ineffective monitoring programmes (Li, et al., 2021). The strength of remote satellite sensing is its ability to provide wider spatial and temporal information cost-effectively. However, remote satellite sensing sampling capabilities are affected by dynamic atmospheric pollutants, alternating weather conditions, aerosols, and overall air quality (Joshi, et al., 2020). Determining the synergy between the field sampling times on extracting water quality indicators of the two methods (satellite data and field data) can assist in shaping effective water quality monitoring programmes. Since there is limited information on the sampling time relations for the field and satellite methods (Gholizadeh, et al., 2016), this researcher had to identify, adapt, and enhance common statistical techniques to establish the process and models of identifying the relationship

between the two sampling methods. The accuracy and efficacy of these models require further evaluation to enable the development of a sampling framework. As indicated by Rwanga & Ndambuki (2017), developing a robust sampling framework by establishing the accuracy of extracting water indicators at different field sampling times to respond to rapid changes in surface water quality is crucial for effective decision-making. Hence the research addressed the question: "What is the effect of sampling time on the accuracy of extracting water quality indicators through remote sensing and geospatial techniques?"

1.3 RESEARCH AIM

The study is intended to assess the effects of field sampling time on the accuracy of extracting water quality indicators through remote sensing and geospatial techniques.

1.3.1 Research Objectives

a) To analyse the correlation between selected laboratory-analysed water quality indicators and remote sensing data at different sampling times.

b) To assess the efficacy of remote sensing models for estimating water quality indicators at different sampling times.

c) To establish the spatial map and area coverage of water quality indicators with S2 models.

1.3.2 Research Questions

a) What is the correlation between selected water quality indicators and remote sensing data at different sampling times, i.e. images acquired a month prior, during and after field sampling period?

b) What is the efficacy of remote sensing models for estimating water quality indicators at different sampling times?

c) How effective is the water quality indicator of the selected Sentinel-2 remote sensing models?

1.4 RELEVANCE OF RESEARCH

Sustainable living conditions and economic growth depend on our ability to respond quickly to protect available water resources. To protect the available water resources, the current rate of deteriorating water quality and quantity necessitates frequent pollution detection and control measures. This researcher sought to provide an optimal sampling window by coinciding with ideal field sampling conditions and minimising mapping errors generated during the sampling of remote satellite sensing data and field data of the water quality indicator extract process. This was to enhance sampling accuracy and frequency, which can result in water quality maps generated with minimised sampling error at a short turnaround time leading to effective implementation of water management strategies. The research is also intended

to improve confidence in the water quality product derived from optical remote sensing data, by establishing the thresholds beyond which mapping accuracy decreases. A sampling framework for water quality extraction through remote sensing was devised to assist with the overall water quality monitoring programme whilst contributing to the scientific body of research.

1.5 RESEARCH LIMITATIONS

The field sampling time to conduct the research was limited to two seasons (winter and spring), divided into four periods. These four periods were mid-winter, late winter, early spring, and mid-spring, being the months of June, July, August, and September respectively. The field data was held constant while satellite data was moved seasonally to estimate the fixed field data. The fixed field data was estimated by moving remote sensing data extracted the season before, during and after the field sampling season. Sampling for the field and the Sentinel-2 started in the mid-winter and ended at the end of mid-spring. Water sampling points were selected and limited due to the high cost of sampling and analysis. The selected sampling points were considered according to health and safety risk because there were wild animals in the sampling area. The study did not consider variations posed by sampling during other seasons such as summer and autumn, which could have produced differing results. During the study, a comprehensive number of models were generated, but only a limited number were considered.

1.6 RESEARCH CHAPTER SEQUENCE AND SUMMATION

Chapter 1: The first chapter introduces the research relating to the traditional and remote sensing sampling methods and their limitations in producing continuous water quality information of inland surface water. The chapter includes the overall research aim and objectives along with the relevance and limitations of the research.

Chapter 2: The chapter provides an analysis of the correlations between selected laboratory-analysed water quality indicators and remote-sensing data at different sampling times. The chapter displayed the summary of selected laboratory-analysed water quality indicator results from the four seasons. Thereafter, a short summary of the reflected date of Sentinel-2 images. The chapter provided correlation information of selected water quality indicators from field sampling times that were correlated with moving Sentinel- images extracted a season before, during and after field sampling times.

Chapter 3: This chapter delved into the models of the selected water quality indicators at different sampling times. Thereafter the best models were selected for further analysis.

Chapter 4: The chapter provided spatial maps of the best-performed water quality indicators estimated from the different sampling times along with the model spatial distribution. Thereafter, the chapter synthesised the research and provided a conclusion and recommendations for the entire research.

Chapter 5: This chapter synthesises and summarises the research.

CHAPTER 2

Analysing the correlations between selected laboratory-analysed water quality indicators and remote-sensing data at different field sampling times

Innocent Rabohale[1*], Oupa E. Malahlela[1], Lutendo Mugwedi[1]

[1] Department of Geography and Environmental Science, Engineering and Agriculture, University of Venda, Thohoyandou 0950, South Africa

[*] Corresponding author: innocentr@reml.co.za

ABSTRACT

Surface water sustains life on Earth, but it is limited and threatened. To ensure adequate water quantity, its quality needs preservation. Thus, continuous water quality information is required for protecting the limited available water resources. Water quality information is mainly generated through a traditional sampling for laboratory analysis technique, and to a lesser degree with satellite sampling techniques, especially in developing countries. This chapter established the synergy between traditional and Sentinel-2 (S2) satellite sampling at different field sampling times for selected water quality indicators (chlorophyll-*a*, turbidity, total suspended solids (TSS), and coloured dissolved organic matter (CDOM)). Field and S2 data were sampled in winter and spring from July to October 2022. Field data were analysed in the laboratory, then correlated with S2 data extracted a month before, at the same time as, and after the field sampling date. The results showed the potential synergy between the techniques' different sampling times, with the September field sampling producing strong correlated CDOM (r= 0.73) at ten (10) days' time lag and Chl-*a* (r=0.63) at five (5) days' time lag with remote sensing data sampled the same month as field date. The turbidity (r=0.71) can be produced within a month after field S2 satellite data at a time lag of 25 days. The TSS (r=0.72) August field data can be generated by same-month S2 satellite data at nine (9) days' time lag. The mean variation of correlated results at different sampling times for the water quality indicators was low at p= 0.73, implying there are other factors contributing to the relationships.

Keywords: Anova, correlation, remote sensing, sampling time, water quality

2.1 INTRODUCTION

Fresh water is required in adequate quantity and quality to sustain all aspects of life and for economic growth. Water quantity and quality are closely related, so the growing deterioration of water quality affects the quantity of water supply (Satinder, 2014). South Africa faces multifaceted challenges that affect water quality (DWS, 2017). Factors such as population growth, increasing pollution loads, climate change, and land use change contribute to surface water quality deterioration. Efforts to continuously monitor water quality are essential to respond proactively and reactively to protect and manage available freshwater resources. Water quality is usually defined as the biological, chemical, and physical properties intended for designated use, such as human consumption, and recreational or ecological use (Roy, 2019). All countries aim to ensure adequate supplies of safe water that meet intended use.

There is a consensus that water quality management in developing countries is inadequate, with some contributing factors including the lack of infrastructure, underinvestment, or poor management of resources (Nhapi, 2015). Traditional approaches which entail field sampling and laboratory analysis are the main methods used for monitoring water quality in developing countries. These approaches require extensive resources that are often costly, resulting in deficient or neglected water quality management in areas where financial resources are limited (Rivett, et al., 2013). Wherever water quality monitoring is considered possible, these approaches only provide point data with limited spatial and temporal information, limiting the ability to determine the extent of the water quality crisis (Van der Schyff & Garbutt, 2023). Extracting water quality data through laboratory techniques is preferable, because the methods are well established and have matured validation processes that meet the scientific basis for decision-making (Madrid & Zayas, 2007). However, the approaches are not economically sustainable for continuous water quality management.

Other techniques, such as remote satellite sensing platforms, are adopted to generate water quality information at various spatial and temporal resolutions (Gholizadeh, et al., 2016). For example, Elhag, et al., (2019) used remote sensing data of Sentinel-2 to estimate water quality indicators in arid environments and found high correlations between field and estimated mean values of chlorophyll (R^2 of 0.96), nitrate (R^2 of 0.94) and turbidity (R^2 of 0.94). The remote satellite sensing approaches generate wide spatial and temporal water quality information economically, but their capabilities are affected by atmospheric interference and cloud cover (Kapalanga, et al., 2021). Satellite techniques are gaining traction in water quality monitoring programmes, but validation processes of their models are still developing. Their ability to estimate accurate water quality information is still highly dependent on the quality of laboratory data used to train and validate the required models (Gholizadeh, et al., 2016).

The freely available satellite data from Sentinel-2 and Landsat satellites are commonly adopted for monitoring the water quality of surface water resources (Caballero, et al., 2022). The data of the two satellite sensors are widely compared in terms of extracting surface reflectance for mapping water quality indicators because of their similarities in spectral resolutions (Food, 2017). Sentinel-2

multispectral instrument provides a wider spatial resolution (10-60 m) at high temporal (5 days) resolutions than Landsat temporal (16 days) resolution (ESA, 2016). Sentinel-2 MSI temporal and spectral capabilities are ideal for regular monitoring and extraction of water quality indicators of surface waters (Elhag, et al., 2019). Considerable studies have successfully extracted water quality indicators using Sentinel-2 remote sensing data for inland surface waterbodies for indicators such as chlorophyll, turbidity, total suspended solids (TSS) and coloured dissolved organic matter (CDOM) (Pizani, et al., (2020); Yingxin, et al., (2021); Tham, et al., (2022)). High temporal satellite data are necessary for continuous water quality management.

The advantages of combined laboratory and remote satellite sensing techniques can improve the supply of continuous water quality information, the lack of which impedes decision-making in managing water resources (Qiaozhen, et al., 2020). Understanding the synergy between the sampling time of the techniques (i.e., field and satellite sampling) can contribute to the integration of the techniques in extracting continuous water quality data. This prompts the questions: What are the correlations between laboratory-extracted water quality indicators and multiple temporal satellite data extracted a month before, during and after field sampling dates? Also, is there a statistically significant difference between correlated data of laboratory-extracted water quality indicators and surface reflectance (Sentinel-2) data at different sampling times? In line with these questions, in this chapter the correlations between selected laboratory-analysed indicators and remotely sensed data at different field and satellite sampling times are explored.

2.2 METHODS

2.2.1 Study area

The research study area was Rietvlei Dam, which is in the city of Pretoria, Gauteng Province, South Africa. The area is within the southern region of the A21 catchment area. The dam is on the Sesmylspruit, a tributary of the Hennops River, at -25.87752 latitude and 28.27196 longitude, with a surface area of 1.89 kilometres (**Figure 2.1**). Apart from the Sesmylspruit River, the dam is fed by five fountains and five boreholes. About 20km upstream of Rietvlei Dam is Hartbeesfontein wastewater treatment works, to the east of Kempton Park. The treatment works' final effluent is released into the Swartspruit River. This river flows through agricultural land, Marais Lake, and wetlands before it joins the Sesymylspruit. The water quality information on the Rietvlei Dam is limited to the public (Roux & Oelofse, 2010).

The Rietvlei Dam was constructed between the years 1932 and 1934 to supply water to the residents of Pretoria (Rietvlei, 2006). The dam is within the Rietvlei Nature Reserve which covers 3 200 hectares. The dam wall is 32 metres high, with a capacity of 12 000 m^3 and a mean elevation above sea level of about 1 525 m, with the highest point at 1 542m and the lowest point at 1 473 m, at the dam outflow.

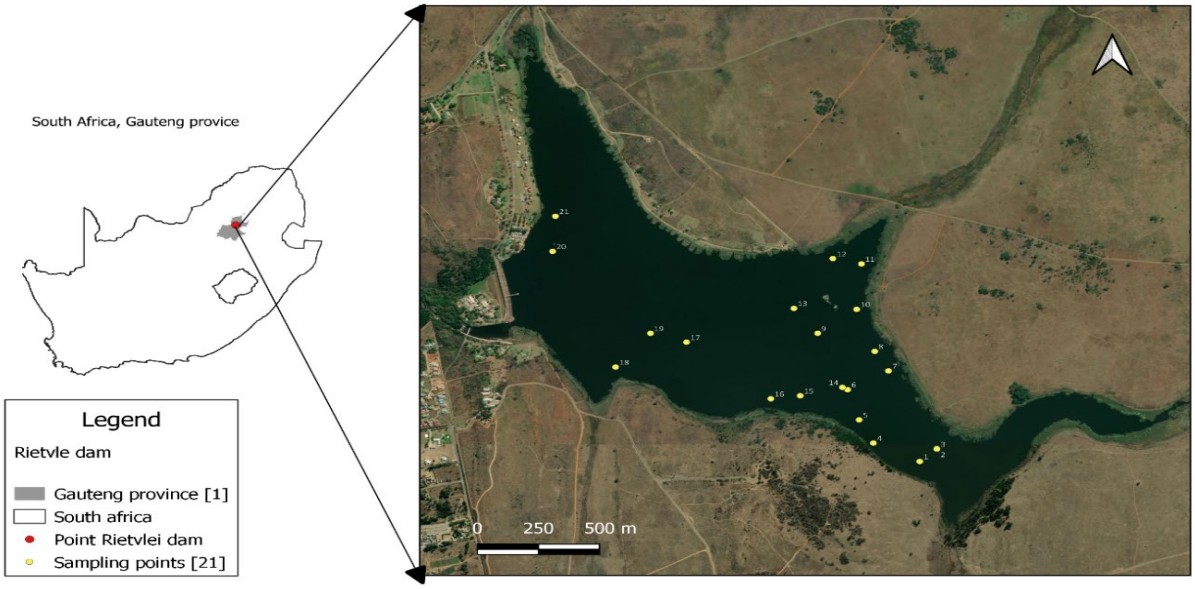


Figure 2.1: Rietvlei Dam and sampling points

2.2.2 Field data collection

A total of 21 points (n=21) were randomly sampled as grab water samples. The sampling points were selected, with a consideration of these factors: the depth of the dam, fishing line, the disparity of the dam water quality, and safety within a nature reserve due to the presence of wild animals (i.e., hippopotamus). The sampling points were randomly distributed and collected for four sampling seasons. The points were located by a handheld Garmin eTrex global positioning system (GPS).

The field sampling took place in two seasons, winter and spring. The two seasons were subdivided according to their months and categorised into early, mid, and late season. Sampling occurred from July to October 2022. July was categorised as mid-winter season, August as late winter season, September as early spring season, and October as mid-spring season. Field sampling took place once per season. Two-litre water samples were collected, preserved in a cooler box, and transported to the laboratory for immediate analysis. The sampling dates were classified with a unique number that started with the field data (FD) acronym, followed by a lower hyphen, and a corresponding date. The field sampling times are displayed in **Table 2.1** below.

The field sampling was intended to bi-weekly, starting in June 2022, to generate continuous monitoring information. However, due to poor weather conditions and high costs of sampling and laboratory analysis, sampling collection was limited to once per season on the dates listed in **Table 2.1**.

Table 2.1: Field sampling time

Season	Sampling dates	No. of Samples collected
Mid-winter	15 July 2022	21
Late winter	23 August 2022	21
Early spring	28 September 2022	21
Mid-spring	21 October 2022	21
	Cumulative total number of samples	**84**

2.2.3 Selected laboratory-analysed quality indicators

The water quality status of the dam can be classified with several indicators to determine its fitness for use for aquatic life and habitats. The selection of water quality indicators for the study area was to define the trophic status of the water resource while considering the capabilities of the Sentinel-2 satellite sensors (ESA, 2016). To define the trophic trends of the dam, i.e. turbidity (Tur), total suspended solids (TSS), colour dissolved organic matter (CDOM) and chlorophyll-*a* (Chl-*a*) water quality indicators were selected. The concentrations of the water quality indicators were measured according to the standard methods for the examination of water and wastewater (SMWW, 1998). The analytical principles and summarised methodologies were as follows.

(i) Turbidity estimation

Turbidity measures water clarity, and it is represented by the nephelometric unit (NTU). Turbidity meters estimate water clarity by measuring the amount of light scattered by the particles suspended in the water. A benchtop turbidity meter was used to estimate the turbidity of water samples from different sampling seasons.

(ii) Total suspended solids (TSS) estimation

Total suspended solids determine the concentration of suspended particles recorded in mg/L units. A gravimetric method was used to estimate the amount of TSS of samples. This was done through the filtration of known volumes of samples, drying the filters and captured solids, and then weighing the captured suspended solids in the sample. The calculation is the difference between the filter and the filtered samples over the volume of the sample.

(iii) Colour-dissolved organic matter (CDOM) estimation

The colour (or chromophoric) dissolved organic matter is part of the organic matter in water that absorbs strongly in the blue and ultraviolet (UV) spectrum. The concentration of CDOM affects the submerged aquatic vegetation. The estimations of CDOM were made using a UV-VIS spectrophotometer, with a CDOM absorption coefficient of 440nm.

(iv) Chlorophyll-a estimation

Chlorophyll-*a* is an important indicator in trophic assessments of water bodies (Aranha, et al., 2022). In this study, chlorophyll-*a* samples were prepared by filtering them through 0.45µm pore filters. They

were thereafter kept frozen at -20 °C. Chlorophyll-*a* was extracted in 80% acetone by centrifugation, with the concentration estimation conducted using a spectrophotometer.

2.2.4 Satellite data acquisition

Satellite data from Sentinel-2 (S2) 2A/B twin mission was used in this study. The Sentinel-2 (S2) is a multispectral, wide-swath platform used for monitoring water quality, land surface and vegetation. The ESA sentinel-2 user handbook specifies the satellite temporal, spectral, radiometric, and spatial information as noted in **Table 2.2** (ESA, 2023). The S2 data is pled with multi-spectral band,: four bands (i.e., visible and near infrared) at 10m, six bands (i.e., red-edge and shortwave) at 20m and three bands (i.e., atmospheric) at 60m spatial resolution at a field view of 290km.

Table 2.2: Sentinel-2 specification

Resolutions	Spectral (bands)	Temporal (repeat cycle)	Spatial resolution (m)
Specification	13 spectral ranging From 443nm to 2190nm	5	10, 20, 60

The five days' temporal resolution was deemed suitable to afford weekly sampling frequency for water quality monitoring. Satellite imageries are not necessarily suitable for water quality assessment, due to cloud cover and atmospheric interferences (Said & Khan, 2021). Hence, only S2 images with cloud cover of less than 10% and atmospheric interference were acquired for water quality assessment. **Table 2.3** lists S2 images that were available for the research during the sampling period (i.e., June to October) and acquisition dates. A cumulative total of 16 L2A images were sampled at different sampling seasons from the Copernicus Open Access Hub.

For ease of image management, images were furnished with labels starting with the S2 (Sentinel-2) acronym, followed by a hyphen, and acquisition date or month as noted in **Table 2.3**. The lists of Sentinel-2 images that were acquired covered the period June to October 2022.

The poor weather conditions observed during the mid-winter (i.e., July) and mid-spring (i.e., September) seasons affected the number of images available and suitable for study. Nonetheless, favourable weather conditions were observed during the later winter period (i.e., August) and early spring (i.e., September) seasons resulting in more images available and suitable for the study. See **Table 2.3** available images acquired and processed during the study.

Table 2.3: Satellite image acquisition dates for the analysis

Month	S2-Sampling season	S2-Acquisition date	Comments
July	Mid-winter	05 July 10 July 15 July ~~20 July~~ ~~25 July~~ ~~30 July~~	Acquired Acquired Acquired Cloud cover Cloud cover Cloud cover
August	Late-winter	04 August 09 August 14 August ~~19 August~~ 24 August 29 August	Acquired Acquired Acquired Cloud cover Acquired Acquired
September	Early-spring	03 September 08 September ~~13 September~~ 18 September 23 September 28 September	Acquired Acquired Cloud cover Acquired Acquired Acquired
October	Mid-spring	03 October ~~08 October~~ ~~13 October~~ ~~18 October~~ 23 October 28 October	Acquired Cloud cover Cloud cover Cloud cover Acquired Acquired
Total acquired images. The dashed-out dates = the images were not suitable for water quality assessment due to cloud cover.		**16**	

2.2.5 Image pre-processing

The sixteen (16) acquired images were pre-processed using the Quantum Geographic Information System (QGIS) version 3.24.1-Tisler software. The bottom of atmospheric (BOA) images were corrected from atmospheric effects using the Sen2Cor tool (https://sentinel.esa.int/). The Sen2Cor tool is a processor in SNAP software that converts Sentinel-2 products from Level 1C to Level 2A for atmospheric and cirrus correction (ESA, 2016). The images acquired were all atmospheric corrected Level 2A images.

The 10m spectral bands are effective in the estimation of water quality indicators in small inland water bodies. However, four bands are limited in estimating some water quality indicators. Hence, the six red-edge to shortwave bands with 20m spatial resolution were resampled to a lower spatial resolution of 10m for better application of water quality assessment. The three 60m spatial resolution bands (atmospheric) were excluded from the study because of their calibration limitations, which could result

in vertical banding images (Pizani, et al., 2020). The ten bands of each Sentinel-2 image acquired at different sampling times were pre-processed.

The BOA spectral data were converted to reflectance data with a semi-automation classification plugin (SCP) tool of the QGIS (Congedo, 2021). The SCP tool was used for any further image pre-processing routines such as correcting irradiance and for image sub-setting. The points surface reflectance data that corresponded to the field sampling points as identified with the GPS were extracted, and adopted for correlation assessment.

2.2.6 Correlation analysis

Correlation analysis is a statistical assessment technique to quantify the relationship between the variables and thereafter determine the strength of the relation (Schober, et al., 2018). The correlation assessment between laboratory-extracted water quality data and the reflectance data was done using MS Excel. Four sets of water quality data extracted at the laboratory were correlated with bands reflectance data extracted from images from the season before, during and after field sampling dates. The decision for the relationship strong linear relationship between the two variables suggested a greater potential to quantify the exact and reliable water quality indicators, whilst the weak relationship shows the opposite (Schober, et al., 2018).

Pearson's correlation coefficient, as represented by the symbol r, was used to determine the linear correlation relationship between sets of laboratory data and S2 reflectance data sampled the season before, same and after the four of the field sampling dates. Equation 1 (eq. 1) shows the Pearson's correlation coefficient formula.

$$r = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=}^{n}(x_i-\bar{x})^2 \Sigma_{i=1}^{n}(y_i-\bar{y})^2}} \quad \text{..... eq. (1)}$$

where: r = the sample correlation coefficient

x_i = the value of the independent variable

$\bar{x}$ = the mean value of the independent variable

y_i = the value of the dependent variable

$\bar{y}$ =the mean value of the independent variable

n= the number of paired data points in the sample

The correlation coefficient ranges between -1 and +1, with the positive and negative sign indicating the direction of the relationship between the independent (laboratory data) and dependent (S2 reflectance data) variables. In this study, the strength of the relationship between the variables was classified as weak (when the range was 0.3 to 0.49), moderate (when the range was 0.5 to 0.69), strong (when the range was 0.7 to 1) and no correlation relationship when it was between 0 and 0.29. When a weak to

non-linear relationship existed between the variables, no further assessment to explore other non-linear relationships was done during the study (Schober, et al., 2018).

In order to establish a stronger correlation relationship between the laboratory-extracted data and the reflectance data from the S2 images, 10 bands were selected. The S2 images were extracted a season before, during and after the field sampling date. A study by Bande, et al., (2018), adopted the all-band correlation assessment strategy to improve the chance of best correlation for chlorophyll-*a* and images from Sentinel-2 and Landsat satellite data. Hence, a similar strategy of adopting 10 bands was employed.

2.2.7 Analysis of variance

To establish any statistical difference of maximum mean correlated data extracted a season before, during and after field sampling date for the four water quality indicators, a single-way analysis of variance (Anova) technique was employed. The test was performed with the Ms Excel. The Anova analysis was used to assess the overall significant difference between the correlated means of water quality indicators extracted a season before, during and after the field sampling date.

2.3 RESULTS

2.3.1 Descriptive statistics results

2.3.1.1 Summary and seasonal trends of laboratory extracted water quality results

A summary of laboratory-extracted water quality results for four field sampling times is presented in **Table 2.4**. The four-field sampling dates were during the mid-winter, late winter, early spring, and mid-spring seasons. The concentrations for the indicators are reported as ranges, and the means with related standard deviations. The indicators were presented as Tur for turbidity, TSS for total suspended solids, CDOM for coloured dissolved organic matter, and Chl-*a* for chlorophyll-*a*. The accumulative water quality means at four different field sampling time are displayed in **Figure 2.2**.

Table 2.4: Summary statistics data for water quality indicators from different sampling dates

Field sampling season & date	Statistics	Tur (NTU)	TSS (mg/L)	CDOM (1/m)	Chl-*a* (ug/L)
Mid-winter FD_15July	Range	1.6 - 4.3	4.0 -14.0	3.22 - 4.14	0.57 - 70.99
	Mean ±SD	2.91 ± 0.81	9.38 ± 2.80	3.59 ± 0.25	15.90 ± 15.74
Late winter FD_23August	Range	0.8 - 2.5	4 – 8	2.30 - 3.91	4.43 - 66.83
	Mean ±SD	1.53±0.43	4.90±1.41	3.07±0.47	30.09±18.50
Early spring FD_28September	Range	1.2 - 3.9	2 – 31	1.61 - 2.53	5.14 -79.85
	Mean ±SD	2.27±0.71	8.14±7.34	2.12±0.29	39.55±16.81
Mid-spring FD_21October	Range	1.2 - 3.8	5 – 35	2.76 - 5.06	80.13 - 583.68
	Mean ±SD	2.31±0.71	13.43±8.33	3.77±0.66	191.53±121.84

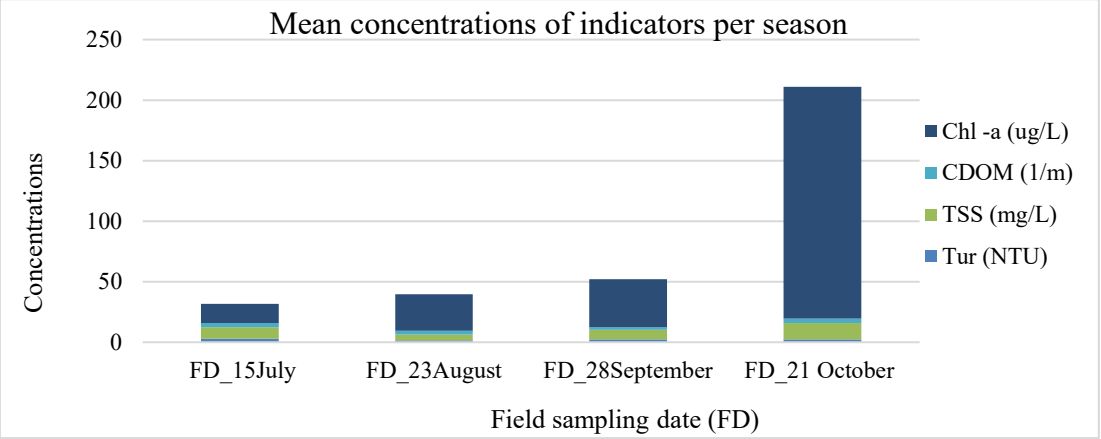


Figure 2.2: Trends of cumulative mean concentrations of water quality indicators for four seasons

On average, the cumulative mean concentrations of the four water quality indicators showed an increasing trend from mid-winter to mid-spring. The highest turbidity concentration of 4.3 NTU was seen during the mid-winter season, and the lowest concentration of 0.8 NTU was during detected during the late winter season. The TSS's highest concentration of 35mg/L was measured during the mid-spring season, contrary to its lowest concentration of 2mg/L observed during early spring. The CDOM's highest concentration of 5.06 1/m was detected during the mid-spring season, while its lowest concentration of 1.61 1/m was observed in the early spring. Notably, the Chl-*a* concentration progressive trend was apparent, with the highest concentration of 583.68µg/L during the mid-spring season, different from the initial lowest concentration of 0.57µg/L. The overall percentage increase of chlorophyll-*a* indicator showed an average increase of over 1000% from the first measured concentration in the mid-winter to mid-spring season. Meanwhile, the other indicators showed an overall increase of less than 50% from the initial measured concentration.

2.3.2 Correlation results

Twenty-one (21) field points of laboratory-extracted data were correlated with related point Sentinel-2 (S2) reflectance data of 10 bands. The points correlation results were extracted for the month before, during and after the field sampling date for the water quality indicators. The means for points correlated results of bands were determined. Thereafter, the minimum and maximum mean of monthly images bands were recorded for each water quality indicator at different sampling times. The monthly minimum and maximum mean correlation results extracted a month before, during and after field sampling date were summarised, and are provided in **Table 2.5**. The field and satellite sampling period covered four months, July to October. The field data was correlated with two months' image reflectance data.

It was observed that the strongest maximum mean correlation for turbidity water quality indicator was r=-0.72, between the September field sampling and October satellite sampling times. The TSS water quality indicator displayed the stronger correlation of r=0.71 between the August field and satellite sampling times. The CDOM water quality indicator's highest correlation of r=0.73 was between the

September field and satellite sampling time. Finally, the Chl-*a* indicator's highest observed correlation of r=0.62 was between the September field and satellite sampling times. The strongest correlations for the TSS, CDOM and Chl-*a* were observed between same months sampling times of the two techniques. The minimum mean correlation observed at different sampling times ranged from -0.09 to 0.

Table 2.5: Minimum and maximum mean correlation results at different sampling times

Field sampling date	Sentinel-2 Sampling month	Correlation of water quality indicators							
		Turbidity		TSS		CDOM		Chl-*a*	
		Min	Max	Min	Max	Min	Max	Min	Max
FD_15Jul	S2_Jul	0.01	0.59	0.07	0.6	-0.01	0.4	-0.02	-0.49
	S2_Aug	0.01	0.55	0.00	0.5	0.00	0.51	0.01	0.51
FD_23Aug	S2_Jul	0.01	-0.47	0.01	0.32	-0.07	-0.62	0.01	-0.34
	S2_Aug	0.01	0.59	0.01	0.71	0.01	0.49	0.01	-0.42
	S2_Sep	0.01	0.30	0.00	0.4	0.02	-0.51	0.01	0.43
FD_28Sep	S2_Aug	0.01	0.71	0.00	0.64	0.01	0.67	0.01	0.61
	S2_Sep	0.05	0.64	0.01	-0.34	0.03	0.73	0.00	0.62
	S2_Oct	0.01	-0.72	0.01	-0.43	0.04	0.42	-0.01	0.58
FD_21Oct	S2_Sep	0.00	0.36	0.00	-0.45	0.01	0.5	-0.05	-0.42
	S2_Oct	0.00	-0.27	-0.09	0.61	0.03	0.53	-0.04	-0.36

2.3.3 Maximum and mean correlation results at different sampling times

The maximum mean correlation results for water quality indicators extracted the month before, during and after the field sampling date at different sampling periods are presented below.

2.3.3.1 Correlation between July laboratory-extracted data and Sentinel-2 (S2) data acquired during the month of, and the month after, the field sampling date (FD_15Jul)

The maximum mean correlation results for water quality indicators from July field sampling time (FD_15Jul) and Sentinel-2 (S2) satellite data from the same month as field sampling (S2_Jul) and a month after (S2_Aug) the sampling field date are depicted in **Figure 2.3**.

The highest strong correlation for turbidity and TSS water quality indicators of r=0.59 and r = 0.60 were observed between the same (July) month field and S2 sampling times. While a moderately strong correlation for Chl-*a* and CDOM water quality indicators were the same at r = 0.51, with S2_Aug data extracted a month after field sampling date (FD_15Jul). The TSS water quality indicator correlation results between the July field and satellite sampling month outperformed the correlation of other indicators during extracted the same month and the month after field sampling date (FD_15Jul).

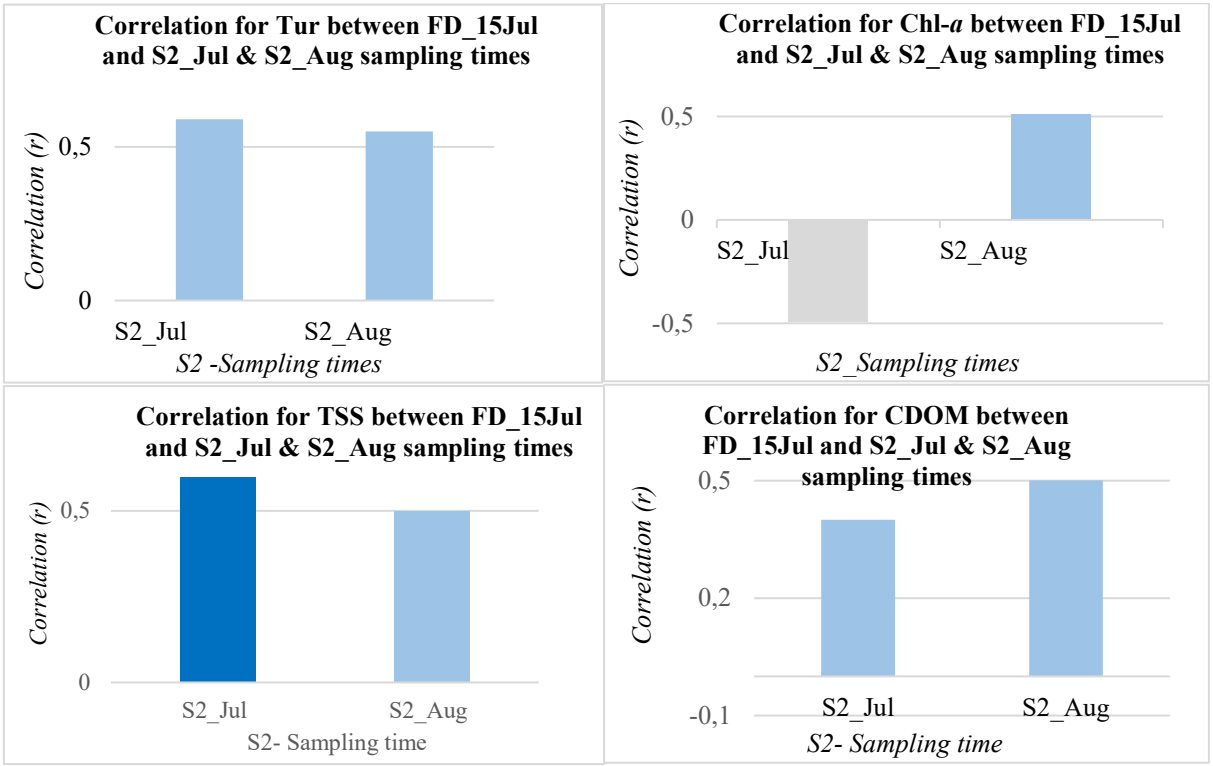


Figure 2.3: Correlation results between FD_July and S2_Jul & S2_Aug sampling times

2.3.3.2 Correlation between August laboratory-extracted data and Sentinel-2 (S2) data acquired a month before, during, and after the field sampling dates (i.e., FD_23Aug).

The maximum mean correlation of water quality indicators between the August laboratory-extracted data and the Sentinel (S2) data acquired a month before (S2_Jul), same (S2_Aug) and after (S2_Sep) field sampling date (FD_23Aug) were as depicted in **Figure 2.4**.

The turbidity (Tur) water quality indicator displayed the highest correlation of a strong value at r=0.69 between the field and satellite (S2) during the same sampling month of August. The similar sampling times difference produced the highest strong correlation for TSS water quality indicator at r=0.71. The CDOM water quality indicator produced the strongest negative correlation (-0.62) between the August field sampling time and the satellite data extracted in July, which was a month before field sampling date. The Chl-*a* highest maximum correlation r=0.43 was seen between August field sampling month and the September satellite extracted data, which is after the field sampling date (FD_23Aug). The TSS water quality indicator displayed the highest maximum mean correlation during the August field sampling and satellite data.

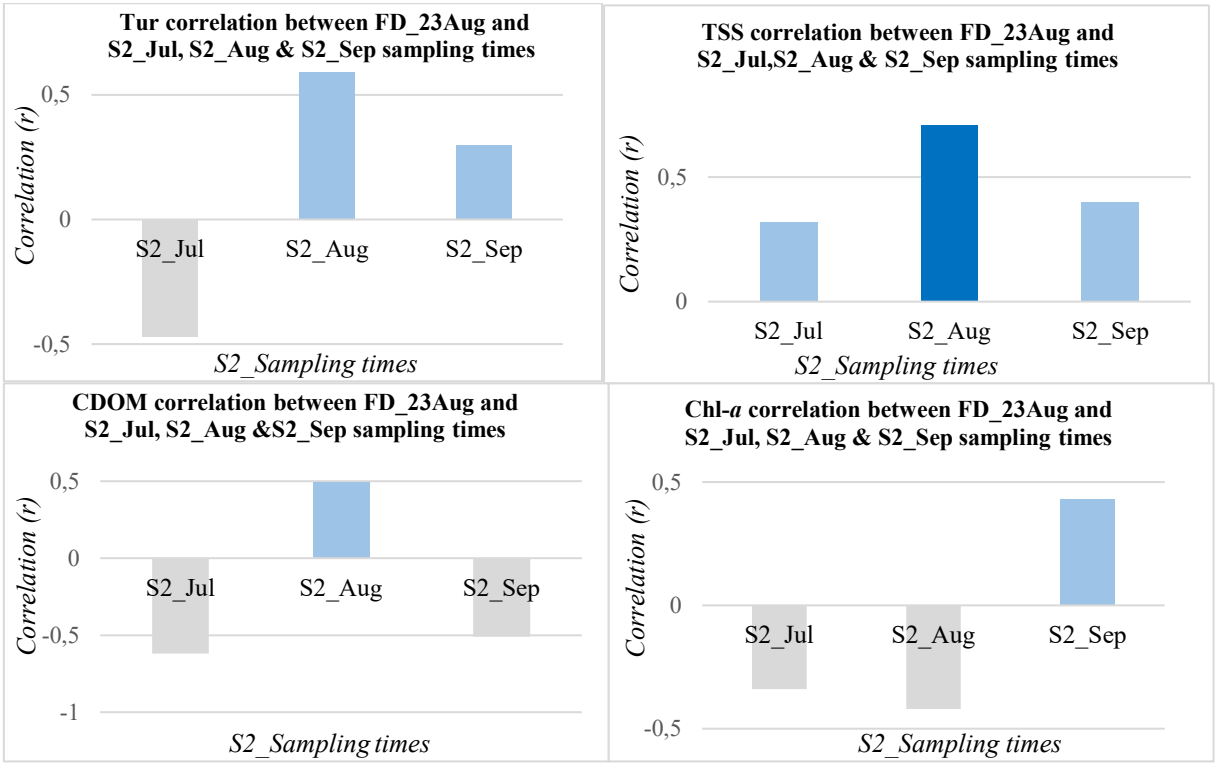


Figure 2.4: Correlation results between FD_August and S2_July, S2_Aug and S2_Sep sampling times

2.3.3.3 Correlation between September laboratory-extracted data and Sentinel-2(S2) data acquired the month before, during and after the field sampling date (FD_28Sep).

The maximum mean correlation results of the water quality indicators between the August field sampling month and satellite (S2) data acquired a month before (S2_Aug), during (S2_Sep) and after (S2_Oct) the sampling field date (FD_28Sep) were as depicted in **Figure 2.5**.

The turbidity water quality indicator had its strongest correlation of r=0.71 between September field sampling and the S2 satellite data extracted a month prior (i.e. August) field date. The TSS displayed the highest correlation of r=0.64 at a similar sampling times difference as that of turbidity water quality indicator. The CDOM highest correlation data was observed at r=0.73, between the September field work and S2 satellite data extracted the same month. The chlorophyll-*a* water quality indicator had the highest correlation of r=0.62, between the September field sampling and S2 satellite date acquired the same month. The CDOM water quality correlation between the September field and satellite sampling times outperformed other water quality indicators during the same sampling period.

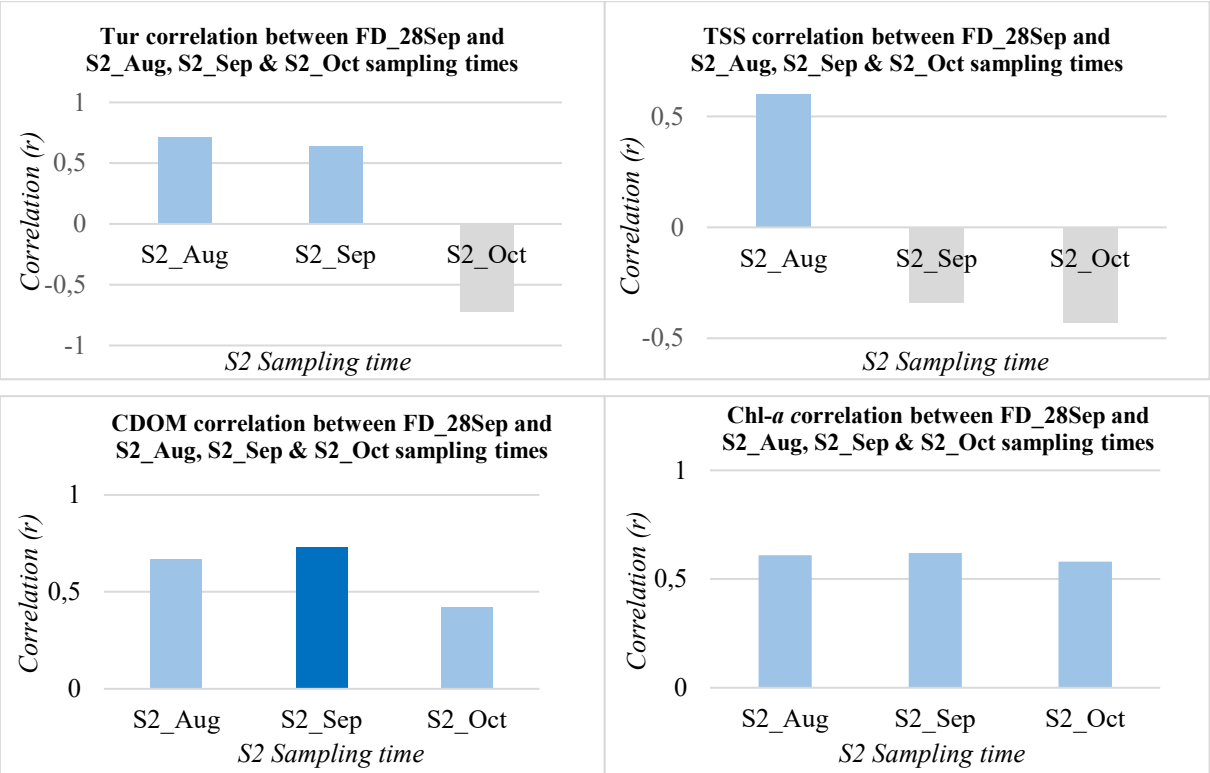


Figure 2.5: Correlation results between FD_September and S2_Aug;S2_Sep and S2_Oct sampling times

2.3.3.4 Correlation for the October laboratory-extracted data and Sentinel-2 (S2) data acquired a month before, during and after the field sampling date (FD_21Oct).

The water quality indicators correlation results between the October field sampling and satellite (S2) data sampled a month before and month after field sampling date (FD_21Oct) were as displayed in **Figure 2.6**.

The TSS and CDOM water quality indicators displayed the highest maximum mean correlation of r=0.61 and r=0.53 respectively, between the same month of October field and S2 satellite sampling times. While the turbidity water quality indicator displayed a weak positive correlation of r=0.36 between the October field sampling and September S2 satellite sampling time. The chlorophyll-*a* water quality indicator had a negative correlation between the laboratory-extracted data and the satellite data acquired the month before (S2_Sep) and the same month (S2_Oct) as field sampling date (FD_21Oct). The chlorophyll-*a* highest correlation of r=0.42 was observed between the same month of October field and S2 satellite sampling times. The correlation results of TSS water quality indicator between the October field and S2 satellite sampling times surpassed other indicators during the same sampling period.

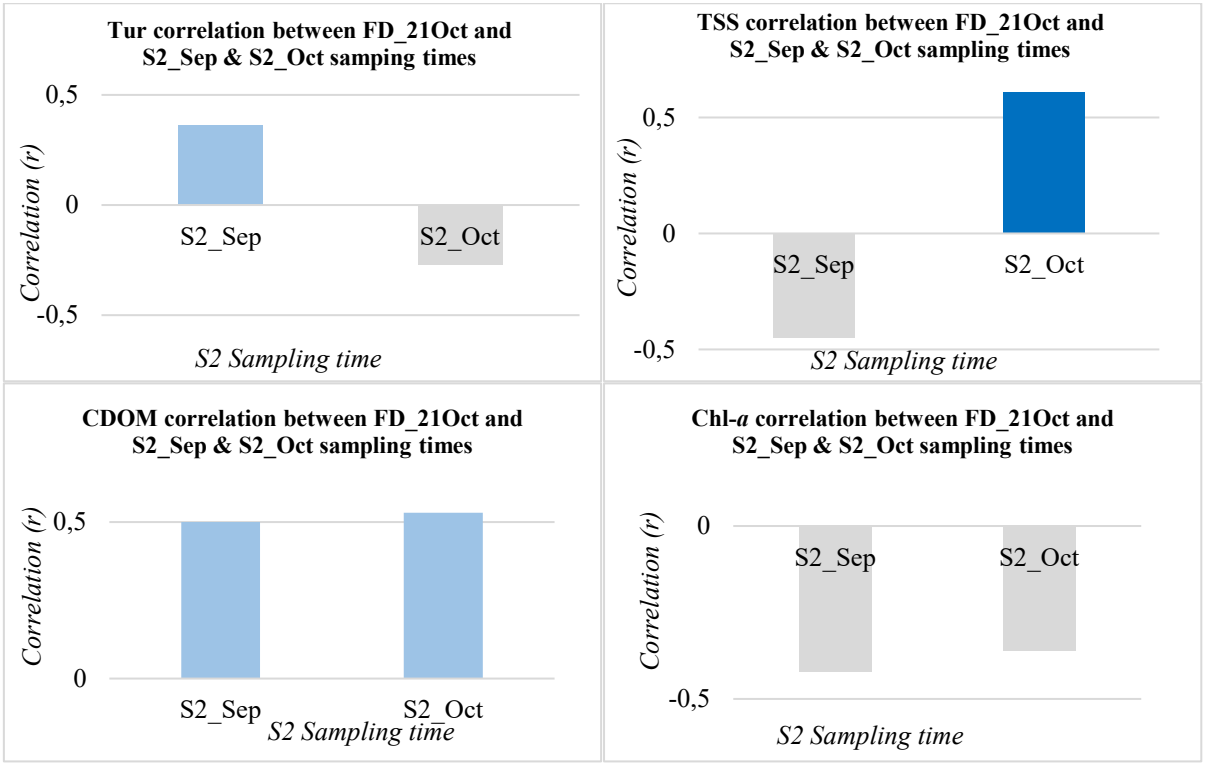


Figure 2.6: Correlation results between FD_October field and S2_Sep and S2_Oct sampling times

2.3.4 Analysis of variance (Anova) results

A one-way Anova technique was used to determine whether the maximum mean correlations of water quality indicators from different sampling times periods were the same. Water quality indicators were turbidity, TSS, CDOM and Chl-*a*. There was no statistically significant difference between groups as determined by the one-way Anova (F(0.4334) =2.866; P-value= 0.73). The results were as depicted in **Table 2.6** below. Maximum mean correlations for the water quality indicators were independent of the sampling times difference between the field and satellite samplings.

Table 2.6: The Anova result for different sampling times and one-way p-value

Anova: Single Factor
SUMMARY

Groups	*Count*	*Sum*	*Average*	*Variance*
Turbidity	10	2.28	0.228	0.269373
TSS	10	2.56	0.256	0.222871
CDOM	10	3.12	0.312	0.224484
Chl-*a*	10	0.72	0.072	0.25824

Anova

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between groups	0.31692	3	0.10564	0.433409	0.730373	2.866266
Within groups	8.77472	36	0.243742			
Total	9.09164	39				

2.4 DISCUSSION

2.4.1 Estimations of water quality indicators at different sampling periods

Correlation assessment described the strength of the linear relationship between the laboratory-extracted water quality data and the Sentinel-2 (S2) data extracted a month before, during and after the field sampling date, at different sampling periods. The strength of the relationship was classified as weak (0.3 to 0.49), moderate (0.5 to 0.69), strong (0,7 to 1) or no correlation (0 to 0.29). The moderate to strong statistical relationship were potential good estimators of the field water quality data for the sampling period.

The mid-winter field sampling occurred on the 15th of July, and the results showed the potential to extracted turbidity and TSS water quality indicators by all Sentinel-2 (S2) data acquired the month during (S2_July) and after (S2_August) field sampling date at moderate association. These physical indicators are essential for monitoring the incoming sunlight, which affect the growth of algae and plankton (Gholizadeh, et al., 2016). The July CDOM and Chl-*a* field data had a moderate statistical relationship only with S2_August data, at a time lag of 20 and 35 days, respectively. Consequently, the synergy between the field and S2 remote sensing techniques to extract mid-winter field (FD_15July) water quality indicators is through estimating the turbidity and TSS with any S2_July and S2_August data. The CDOM and Chl-*a* indicators can be estimated with S2_August data only.

The late-winter (FD_23August) laboratory-extracted data had an inverse moderate statistical relationship for CDOM indicator for sampling difference of all S2 data from a month prior (S2_July) until 11 days' time lag of the following (S2_September) month from field date. The TSS and turbidity indicators displayed a strong and moderate statistical relationship respectively, with sampling time difference during the month of field sampling at a time lag of nine days from field date. To optimise the water quality extraction for the August TSS and turbidity laboratory-extracted data, cooperation between the same month sampling difference with S2 remote sensing data can be adopted. The CDOM indicator can be associated with the previous and following month sampling difference S2 remote sensing data. The Chl-*a* indicator had little possibility of being extracted by linear correlation during the correlated sampling times. Other empirical approaches can be employed to estimate the chlorophyll-*a* (Lins, et al., 2017).

While the early-spring (FD_28September) turbidity and Chl-*a* laboratory-extracted data can be complemented with S2 data extracted a month before (S2_August), during (S2_September) and after (S2_October) sampling. The CDOM indicator was estimated with S2 data from the month before and during the field sampling month. The TSS indicator was estimated by S2 data from the month before the field date at a time lag of 30 days. The strongest correlation for turbidity indicator had a time lag of 25 days: the CDOM at a time lag of 10 days and the Chl-*a* at a time lag of five days sampling times difference between sampling techniques. This was contrary to the sampling times difference window of about three days for generating highly correlated measurements (Bramich, et al., 2021). The strong

correlations observed with S2 data from the months before and during was partly due to the improved illumination and improved weather conditions (Wang & Menenti, 2021).

Meanwhile, the mid-spring (FD_21October) laboratory-extracted data had the poorest correlation with S2 data from the month before (S2_Septemeber) and during (S2_October) the sampling. The TSS indicators had the potential to be complemented by S2 data from during the month sampling times difference. The CDOM indicator could be estimated by the S2_September and S2_October data. The mid-spring month had limited S2 remote sensing data, and poor weather conditions (Dubovik, et al., 2021).

The synergy between field and S2 satellite sampling times for extracting selected water quality indicator allows for water monitoring programmes to alter between the traditional and S2 satellite techniques. Thus, reliance on extraction of water quality with traditional methods only, or coincidental and near coincidental data only, is reduced (Omondi, et al., 2023; Kapalanga, et al., 2021; Govedarica & Jakovljević, 2019).

Caution is required whenever non-coincidental data is adopted, since the null hypothesis at 5% level of significant was accepted, indicating that there was not enough evidence to conclude that the water quality correlations were not equal at different sampling times. The Anova results showed that the strongly correlated water quality indicators were independent of the time difference between the field and satellite data. Thus, other factors that influence the difference in the correlations results require consideration. The following indicators are assessed for model development turbidity (r= -0.72), TSS (r=0.71), CDOM (r=0.73) and Chl-*a* (r=0.62) at different sampling periods. The common approach for developing water quality models requires field and satellite data that is acquired at the same sampling times (Aranha, et al., 2022; Congalton, 2001; Yadav, et al., 2019; Bande, et al., 2018). This provides an opportunity for continuous water quality extraction of waterbodies at different sampling times.

2.4.2 Spectral information of strongly correlated water quality indicators

The spectral bads that contributed to the highest correlation highlighted their potential to be used to estimate and map water quality indicators. Thus, the Sentinel-2 spectral bands that contributed to the strong correlation for CDOM (0.73) and Chl-*a* (0.62) were a green band (B3) and a red-edge band (B6) respectively. The retrieval of the CDOM water quality indicator with the B3 band can be supported by other studies that used a combination of Sentinel-2 bands including the green band (Toming, et al., 2016; Nadil et al., 2021). The red-edge band (B6) contributed towards a strong correlation for Chl-*a*. The Sentinel red-edge band has been used for vegetation classification and wide extraction of chlorophyll-*a* model (Bramich, et al., 2021). The bands that contributed to the turbidity and TSS high correlation were the red-edge band (B5) and the near infrared band (B8) respectively. According to the review by Gholizadeh, et al., (2016), the single band when selected appropriately can provide a robust estimation algorithm.

2.5 CONCLUSION

This chapter established the minimum and maximum mean correlations for water quality indicators at different field and satellite sampling times. There was a strong correlation between laboratory-analysed and satellite data extracted at different sampling times. The single spectral band that contributed most to the highest strong correlation per water quality indicator was considered. The following water quality indicators showed correlation, and their field and satellite sampling time differences are as follows:

- TSS indicator correlation (0.71) was in the same month of field and S2 sampling time at a time lag of nine days.
- These indicators were highly correlated for September field sampling (early spring season):
    - The CDOM correlation (0.73) was between the same month (September) of field and S2 satellite sampling time at time lag of 10 days.
    - Chlorophyll-*a* correlation (0.62) was between same month field and S2 satellite sampling time at time lag of five days.
    - The turbidity correlation (-0.72) with S2 extracted the following month of field sampling date at a time lag of 25 days.

Ascertaining the strong correlation of water quality indicators at different sampling times is a step towards developing the continuous monitoring programme using both techniques. The findings in this chapter opened opportunities to further analyse models suitable for estimating surface water quality. The water quality indicators with moderate to strong maximum mean correlation were selected for further analyses in the next chapter (Chapter 3), to develop a model for estimating the surface water quality at different sampling times.

2.6 REFERENCES

Adusei, Y. Y., Quaye-Ballard, J., Adjaottor, A. A. & Mensah, A. A., 2021. Spatial prediction and mapping of water quality of Owabi reservoir from satellite imageries and machine learning models. *The Egyptian Journal of Remote Sensing and Space Science*, 24(3), pp. 825-833.

Aranha, T., Martinez, J. M. S. E. & Barros, M. U., 2022. Remote analysis of the chlorophyll-*a* concentration using Sentinel-2 MSI image in a semiarid environment in Northeastern Brazil. *Water*.

Barsi, J. et al., 2018. Sentinel MSI and Landsat-8 OLI radiometric cross comparison over desert sites. *European Journal of Remote Sensing*, 51(1).

Biorestita, F. et al., 2018. The use of Sentinel-2 imagery for total suspended solids (TSS) estimation in Porong River, SiDoarjo. *Journal of Geodesi dan Geomatika*, 1(1), pp. 1-6.

Bramich, J., Bolch, C. & Fischer, A., 2021. Improved red-edge chlorophyll-a detection for Sentinel 2. *Ecological Indicators*.

Bramich, J., Bolch, C. J. & Fischer, A., 2022. Improved red-edge chlorophyll-a detection for Sentinel 2. *Ecological Indicators*, 120 (2021).

Caballero, I. et al., 2022. Use of the Sentinel-2 and Landsat-8 satellites for water quality monitoring: An early warning tool in the Mar Menor coastal lagoon. *Remote Sensing*.

Dubovik, O. et al., 2021. Grand challenges in satellite remote sensing. *Frontiers in Remote Sensing*, Volume 2.

DWAF, 2002. National Eutrophication Monitoring Programme. *South African National Water Quality Monitoring Programmes Series*.

DWAF, 2004. National Water Resources Strategy.

DWS, 2017. *Water Quality Management Policies and Strategies for South Africa*, Pretoria, South Africa: s.n.

Elhag, M., et al., 2019. Assessment of water quality parameters using temporal remote sensing spectral reflectance in arid environments, Saudi Arabia. *Water*, pp. 1-14.

European Space Agency (E. S. A.), 2016. Sen2Cor Configuration and User Manual.

European Space Agency (E. S. A.), 2023. *The European Space Agency*. [Online] Available at: https://sentinels.copernicus.eu/documents/247904/685211/Sentinel-2_User_Handbook.pdf/8869acdf-fd84-43ec-ae8c-3e80a436a16c?t=1438278087000 [Accessed 14 August 2023].

Food, N., 2017. Comparing Sentiel-2A and Landsat 7 and 8 using surface reflectance over Australia. *Remote Sensing*, 9(7).

Food, N., 2017. Comparing Sentinel 2A and Landsat 7 and 8 using surface reflectance over Australia. *Remote Sensing*, 9(7).

Gholizadeh, M. H., Melesse, A. M. & Reddi, L., 2016. A comprehensive review on water quality parameters estimation using remote sensing techniques. *Sensors*, Volume 16.

Govedarica, M. & Jakovljević, G., 2019. *Monitoring spatial and temporal variation of water quality parameters using time series of open multispectral data*. Paphos, Cyprus, s.n.

Kapalanga, T., Hoko, Z., Gumindoga, W. & Chikwiramakomo, L., 2021. Remote-sensing-based algorithms for water quality monitoring in Olushandja Dam, North-Central Namibia. *Water Supply*, 21(5), pp. 1878-1894.

Malahlela, O., 2016. Inland waterbody mapping: towards improving discrimination and extraction of inland surface water features. *International Journal of Remote Sensing*, 37(19), pp. 4574-4589.

Mamuye, T. E. & Mutiu, K. A., 2021. Overview of water quality modeling. *Cogent Engineering*, 8(1).

Nadil, H., Mohammed, A. & Sohib, A., 2021. *Application of Sentinel 2 to evaluate colored dissolved organic matter algorithms for inlands water bodies in Jordan*. Jordan, s.n.

Ncube, E., Voyi, K. & Preez, H., 2012. Implementing a protocol for selection and prioritisation of organic contaminants in the drinking water value chain: case study of Rand Water, South Africa. *Water SA*, 38(4), pp. 487-504.

Nguyen, T. et al., 2020. Total suspended solid distribution in Hau River using Sentinel-2A satellite imagery. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, pp. 91-97.

Nhapi, I., 2015. Challenges for water supply and sanitation in developing countries: Case studies from Zimbabwe. *Understanding and Managing Urban Water in Transition*, Volume 15.

Omondi, A. N. et al., 2023. Estimation and mapping of water quality parameters using satellite images: a case study of Two Rivers Dam, Kenya. *Water Practice & Technology*, pp. 428-443.

Onyutha, C., 2020. From R-squared to coefficient of model accuracy for assessing “goodness-of-fits”. *Geoscientific Model Development*.

Pizani, F., Maillard, P., Ferreira, A. & Amorim, C., 2020. Estimation of water quality in a reservoir from Sentinel-2 MSI and Landsat-8 OLI sensors. *V-3-2020. 401-408. 10.5194/isprs-annals-V-3-2020-401-2020*.

Potes, M. et al., 2018. Use of Sentinel 2 – MSI for water quality monitoring at Alqueva reservoir, Portugal. *International Association of Hydrological Sciences*, pp. 73-79.

Qiaozhen, G. et al., 2020. an integrated study on change detection and environment evaluation of surface water. *Applied Water Science*, pp. 1-15.

Rivett, U., Champanis, M. & Wilson-Jones, T., 2013. Monitoring drinking water quality in South Africa: Designing information systems for local needs. *Water SA*.

Roux, S. & Oelofse, S., 2010. *The rising costs of both sewage treatment and the production of potable water associated with increasing levels of pollutions in a portion of the Crocodile-West Marico water management area (Case study)*. Pretoria, CSIR.

Roy, R., 2019. An Introduction to water quality analysis. *International Research Journal of Engineering and Technology*, 6(1).

Sagan, V. et al., 2020. Monitoring inland water quality using remote sensing: potential and limitations of spectral indices, bio-optical simulations, machine learning, and cloud computing. *Earth-Science Reviews*, Volume 205.

Said, S. & Khan, A. S., 2021. Remote sensing-based water quality index estimation using data-driven approaches: A case study of the Kali River in Uttar Pradesh, India. *Environment, Development and Sustainability*, Volume 23, pp. 18252-18277.

SANews, 2019. *93% of South Africans have access to water services*. [Online] Available at: https://www.sanews.gov.za/south-africa/93-south-africans-have-access-water-services#:~:text=Ninety%20three%20percent%20of%20the%20South%20African%20population,the%20population%20have%20access%20to%20basic%20sanitation%20services. [Accessed 05 10 2023].

Satinder, A., 2014. *Comprehensive water quality and purification*. Elsevier.

Shi, X. et al., 2022. Retrieval of chlorophyll-*a* concentrations using Sentinel-2 MSI imagery in Lake Chagan based on assessments with machine learning models. *Remote Sensing*, 14(19).

Smith, G., 2018. Step away from stepwise. *Journal of Big Data*, Volume 5.

SMWW, 1998. *Standard Methods for the Examination of Water and Wastewater*, s.l.: s.n.

Toming, K. et al., 2016. First experiences in mapping lake water quality parameters with Sentinel-2 MSI imagery. *Remote Sensing*.

Tran, M. et al., 2023. Band ratios combination for estimating chlorophyll-*a* from Sentinel-2 and Sentinel-3 in coastal waters. *Remote Sensing*, 15(6).

Uddin, G., Nash, S. & Olbert, A., 2021. A review of water quality index models and their use for assessing surface water quality. *Ecological indicators*.

USGS, 2019. *USGS science for a changing world*. [Online] Available at: https://www.usgs.gov/special-topics/water-science-school/science/how-much-water-there-earth. [Accessed 05 10 2023].

Van der Schyff, A. & Garbutt, M., 2023. Water quality monitoring systems for sub-Saharan Africa: Towards an effectiveness framework. International Conference on Information Systems and Advanced Technologies.

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

Wang, Z. & Menenti, M., 2021. Challenges and opportunities in Lidar remote sensing. *Remote Sensing*.

Weather&Climate, 2023. Weather-and-Climate. [Online]
Available at: https://weather-and-climate.com/average-monthly-precipitation-Rainfall-inches,centurion-gauteng-za,South-Africa. [Accessed 02 August 2023].

World Health Organization, 2004. Guidelines for drinking water quality recommendations, Volume 1.

Yang, H. et al., 2022. A review of remote sensing for water quality retrieval: progress and challenges. *Remote Sensing*, 14(1770).

CHAPTER 3

Assessing the efficacy of remote sensing models for estimating water quality indicators at different sampling times: a case study in Gauteng, South Africa

Innocent Rabohale[1*], Oupa E. Malahlela[1], Lutendo Mugwedi[1]

[1] Department of Geography and Environmental Science, Engineering and Agriculture, University of Venda, Thohoyandou 0950, South Africa

[*] Corresponding author: Innocentr@reml.co.za

ABSTRACT

Growing deficiencies of freshwater necessitate continuous generation of water quality information to protect water resources from pollution and other factors. The use of traditional and remote satellite sensing technique to generate water quality information is widely accepted. However, traditional and remote sensing data may be acquired at different sampling times. The uncertainties introduced thereby can hamper effective decision-making. The extent to which a time lag may exist between field survey and remote satellite sensing data, and its effect on the accuracy of modelling, has received little attention. Therefore, this study is intended to assess the efficacy of remote sensing models for estimating water quality indicators at different sampling times. The multi-stepwise regression analyses of laboratory-extracted data and Sentinel-2 (S2) satellite data acquired a month before, during and after sampling date were adopted to develop models. The S2 models' efficacy was determined by R^2, MAE and RMSE assessment. The results showed that S2 models produced at different sampling times of the field survey and satellite sampling, can efficiently predict the Chl-*a* (R^2 =0.79, RMSE =64.99 µg/mL and MAE =38.35) and TSS (at R^2=0.62; RMSE =19.62mg/L and MAE=14.16) at shorter sampling times difference window of a week. The CDOM (R^2=0.64, RMSE =10.31 1/cm and MAE=10.30) and turbidity (R^2=0.56, RMSE=0.37 NTU and MAE=0.35) have a wider sampling time difference window of over a month. The optimising of water quality extraction may be achieved through continuous sampling, including both field survey and satellite data techniques.

Keywords: accuracy, correlation, efficacy, remote sensing, models, water quality, validation

3.1 INTRODUCTION

Developing countries face a significant challenge in protecting water quality while trying to improve water supply (Uddin, et al., 2021). Water quality monitoring is undertaken to inform policymakers, environmental managers and communities of the status and emerging threats to their water bodies. Water quality status provides guidelines for the usefulness of the water to meet human consumption, recreational use, habitat and species diversity support and contribution to economic growth. Efforts to assess the water quality status to meet the different uses are commonly done by traditional techniques of in-situ sampling and laboratory analysis (Gholizadeh, et al., 2016). These techniques have proven to provide accurate water quality information. However, they are affected by logistical limitations and high costs (Ncube, et al., 2012), thus limiting their effectiveness in continuous water quality monitoring. To improve the decision-making capabilities of water resource custodians, managers, and policymakers need continuous water quality information. This in turn will align with the global efforts under Sustainable Development Goal Six, which aims to ensure the availability and sustainable management of water and sanitation for all.

There is a consensus among some researchers on the integration of traditional and remote sensing techniques to generate continuous water quality data to improve water resources management programmes (Muhoyi, et al., 2022; Bugnot, et al., 2018). However, remote satellite sensing is highly affected by atmospheric interference and poor weather conditions, limiting the data extraction for water quality determination (Gholizadeh, et al., 2016). Nonetheless, remote-sensing data application in water quality applications has gained traction and numerous methods to pursue high-accuracy estimations are being developed (Wang & Yang, 2019). To retrieve the water quality indicators through remote satellite sensing, three modes are commonly adopted i.e., empirical, semi-empirical and analytical, to establish the relationship between the laboratory-analysed data and satellite reflectance data. In application, the empirical mode is simple, and it is easy to refine the selection of more precise channels to increase the accuracy of water quality indicators when compared to the other modes (Yang, et al., 2022). The empirical mode is adopted because of the complexity introduced by the inherent optical properties of the water bodies, and continuous changes in water quality indicators concentration in seasons (Yang, et al., 2022; Pizani, et al., 2020).

Researchers in many parts of the world mostly use freely available satellite data from Sentinel-2 and Landsat satellites to monitor water quality for surface water bodies (Pizani, et al., 2020; Bande, et al., 2018; Du, et al., 2016; Gholizadeh, et al., 2016). The two satellite instruments are widely compared in terms of extracting surface reflectance for mapping water quality indicators because of their similarities in spectral resolutions (Food, 2017). Sentinel-2 multispectral instrument provides a wider spatial resolution (10-60m) at high temporal (5 days) resolutions than Landsat temporal (16 days) resolution (ESA, 2016). Sentinel-2 MSI temporal and spectral capabilities are ideal for regular monitoring and extraction of water quality indicators of surface waters (Potes, et al., 2018). Many studies have successfully extracted water quality indicators using Sentinel-2 remote sensing data for inland surface

waterbodies for indicators such as chlorophyll-a, turbidity, total suspended solids (TSS) and CDOM (Potes, et al., 2018; Pizani, et al., 2020; Aranha, et al., 2022). The complications of separating spectral properties of water quality indicators are due to the inherent optical properties of the water body (Yang, et al., 2022).

In practice, in-situ and satellite remote sampling times vary, unless intentionally designed. To use the methods simultaneously for continuous water quality monitoring purposes, a comprehensive understanding of the synergy of both methods at different sampling times is needed. The research's primary objective was to assess the efficacy of remote sensing models for estimating water quality indicators at different sampling times. Different sampling times referred to the moving remote sensing sampling time to estimate the fixed in-situ sampling data of a specific sampling date. The three secondary research objectives were: firstly, to develop the remote sensing models using Sentinel-2 data to estimate the water quality indicators (WQI). The Sentinel-2 data were extracted a season before, the same season and the season after in-situ sampling time. Secondly, the research identified and selected the best-performing models established in the season before, in the same season, and after in-situ sampling time. Thirdly, make recommendations for the identification of the best possible matches between field sampling and satellite data for specific water quality indicators. Through these objectives, the efficacy of remote sensing models for estimating water quality indicators at different sampling times was successfully established. Throughout the data processing, in-situ water quality indicators data were held constant, while different remote satellite sensing data from different seasons were estimated the water quality data.

3.2 METHODS

3.2.1 Study area

The research study area was Rietvlei Dam, which is in Centurion, Gauteng Province, South Africa. The area is within the southern region of the A21 catchment area. The dam is within the Sesymylspruit, a tributary of the Hennops River, at -25.87752 latitude and 28.27196 longitude, with a surface area of 1.89 kilometres (**Figure 3.1**). The dam is fed by the Sesmylspruit River, five fountains and five boreholes. About 20km upstream of Rietvlei Dam, the Hartbeesfontein wastewater treatment works is located, to the east of Kempton Park. The treatment works final effluent is released into the Swartspruit River. This river flows through agricultural land, Marais Lake, and wetlands before it joins the Sesymylspruit. The water quality information on the Rietvlei Dam is limited to the public (Roux & Oelofse, 2010). The dam is within the Rietvlei Nature Reserve which covers 3 200 hectares. The dam wall is 32 meters high with a capacity of 12 000 m^3 and a mean elevation above sea level of about 1 525 m, with the highest point at 1 542m and the lowest point at 1 473m, at the dam outflow.

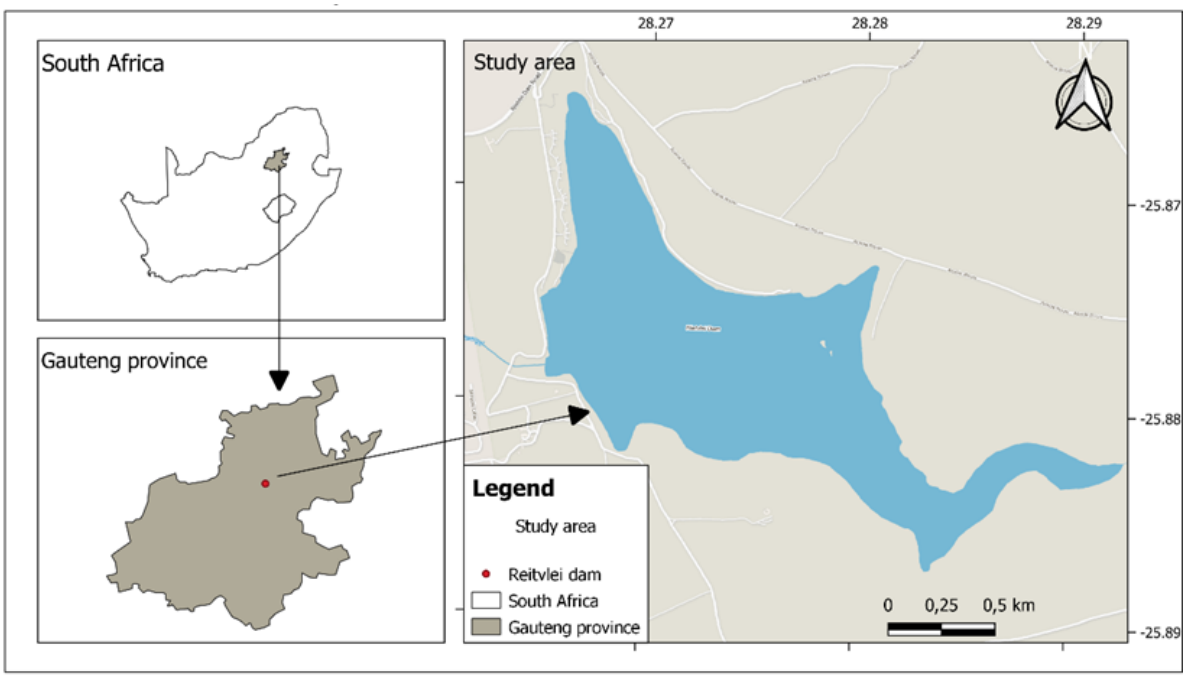


Figure 3.1: Rietvlei dam and sampling points

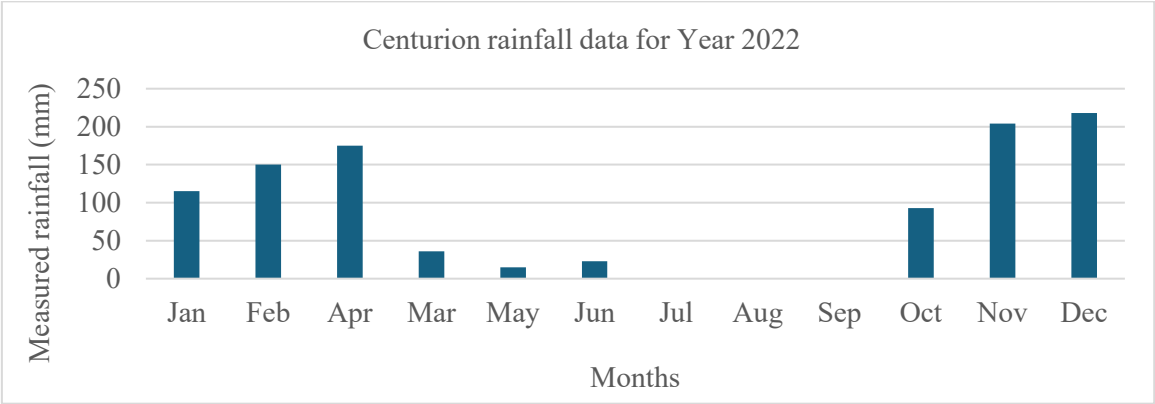


Figure 3.2: Centurion rainfall data

Centurion usually has dry winters, which normally fall in the months of May, June, July, August, and September. The lowest temperature was mainly observed in June and July. The temperature started to increase in August 2022.

As depicted in **Figure 3.2**, in the year 2022 July, August and September were the driest months with 0mm of precipitation. A considerable amount of precipitation fell in November and December, more than 200mm. The total annual precipitation in the year 2022 was 1029mm, which was significantly above the average annual precipitation of 630mm (Weather&Climate, 2023).

3.2.2 Field sampling

A desktop random sampling plan was designed using GIS software, and 40 points (n =40) were pre-allocated to guide the collection of water samples. However, only 21 points (n=21) were randomly

sampled. At each point, two litres of grab water samples were collected within the Rietvlei Dam. The samples were preserved in a cooler box and transported to the laboratory for water quality analyses.

The selection of field sampling points was guided by several factors such as the depth of the dam, fishing line, the dam's water quality disparity, and health and safety issues, since the dam is within a nature reserve that contains wild animals (i.e., hippopotamuses). Hence, this resulted in a total of sampling 21 points randomly distributed and sampled. Field sampling was divided into four months. To locate the same sampling points each season, a handheld Garmin eTrex global positioning system (GPS) was used.

The monthly sampling was guided by principles provided by the national eutrophication monitoring programme while considering the associated sampling and analytical costs (DWAF, 2002). Initially, sampling dates were pre-selected to be every two weeks to adhere to the minimum requirement of the trophic monitoring programme (DWAF, 2002). However, due to poor climatic conditions and the costs of sampling and laboratory analysis, field sampling took place once per month.

The sampling months ran from July 2022 to October 2022. The seasons were classified as follows: July was classified as mid-winter, August as late winter, September as early spring, and October as mid-spring. Each field sampling date was furnished with a unique number. Each unique number had the prefix letter FD as an abbreviation for field date, followed by a lower hyphen (_), and thereafter the sampling date. The field sampling times are displayed in **Table 3.1** below.

Table 3.1: Field sampling time

Field date unique number	Field sampling Season	Field sampling dates	Sample no
FD_15July	Mid-winter	15 July 2022	21
FD_23August	Late winter	23 August 2022	21
FD_28September	Early spring	28 September 2022	21
FD_21October	Mid-spring	21 October 2022	21
		Cumulative total number of samples	**84**

3.2.3 Acquisition of Sentinel-2 (S2) images

The S2 images were acquired to address the aim of the research. The S2 is based on a twin satellite that provides a medium to high spatial resolution of 10m, 20m, and 60m images, with a temporal resolution of five days pled with multi-spectral information of 13 bands covering visible-infrared red and short-wave infrared at a field view of 290km.

The S2's five days of temporal resolution allow for weekly sampling frequency. This amounts to an accumulative amount of six images per month. The five days' temporal resolution is ideal for the weekly sampling frequency/monitoring of water quality indicators. However, images are not readily suitable for water quality assessment, because of cloud cover and atmospheric interference (Said & Khan, 2021). Satellite images with less than 1% of cloud cover and cleared atmospheric interference were adopted for water quality assessment. **Table 3.2** lists S2 images and the quality of those that were available

during the research period (i.e., June to October) and acquisition dates. Out of 24 images, a total of 17 were acquired and processed. The acquired S2 images did not coincide with the in-situ sampling date.

Satellite acquisition dates were also categorised into sampling season, which was defined according to the month in which the images were acquired i.e., July as the mid-winter season; August as the late winter season; September as the early spring season and October as the mid-winter season. For ease of data management, images were given labels with the prefix S2 which stood for Sentinel-2, followed by the hyphen (_), and thereafter the acquired date. As displayed in **Table 3.2**.

In line with the climatic conditions of the study, poor weather conditions were observed during the mid-winter and mid-spring seasons where fewer images (i.e., three for both seasons) were acquired. Overall, clear favourable weather conditions for satellite image acquisition were seen during the late winter and early spring seasons where most images (i.e., nine and five images respectively) were obtained.

3.2.4 Pre-processing of Sentinel-2 (S2) images

The S2 satellite images consist of 13 spectral bands (443–2190nm), which are more sensitive to optically active water quality indicators. At 10m spatial resolution there are four visible and near-infrared bands, at 20m spatial resolution there are six red-edge/shortwave bands, and at 60m there are three atmospheric correction bands. The visible and near-infrared (VNIR) spectral bands are sensitive in the estimation of chlorophyll-*a*, suspended solids, and turbidity water quality indicators of inland small water bodies (Shi, et al., 2022). The short-wave infrared (SWIR) bands have been successfully used to detect water quality indicators such as total suspended solids, dissolved organic matter and coloured dissolved organic matter (Nguyen, et al., 2020). **Table 3.3** lists all the spectral bands adopted in the study. Ten spectral bands were selected and pre-processed using the QGIS software.

Table 3.2: Sentinel-2 (S2) image acquisition dates and quality

Sampling season	Month	Adopted labels	Sampling date (2022)	Observation
Mid-winter	July	S2_186	S2_ 05 Jul	Low sunlight
		S2_191	S2_10 Jul	Low sunlight
		S2_196	S2_15 Jul	Low sunlight
		Not adopted	20 Jul	Clouds
		Not adopted	25 Jul	Clouds
		Not adopted	30 Jul	Clouds
Late winter	August	S2_216	S2_04 Aug	low sunlight
		S2_221	S2_09 Aug	High sunlight
		S2_226	S2_14 Aug	High sunlight
		Not adopted	19 Aug	Clouds
		S2_236	S2_24 Aug	Low sunlight
		S2_241	29 Aug	Low sunlight
Early spring	September	S2_246	S2_03 Sep	High sunlight
		S2_251	S2_08 Sep	High sunlight
		Not adopted	13 Sep	Clouds
		S2_261	S2_18 Sep	High sunlight
		S2_266	S2_23 Sep	High sunlight
		S2_271	S2_28 Sep	High sunlight
Mid-spring	October	S2_276	S2_03 Oct	High sunlight
		Not adopted	08 Oct	Clouds
		Not adopted	13 Oct	Clouds
		Not adopted	18 Oct	Clouds
		S2_296	S2_23 Oct	High sunlight
		S2_301	S2_28 Oct	High sunlight

Table 3.3: Sentinel-2 spectral bands selected and processed in the study (ESA, 2016)

Band	Description	Spatial resolution (m)	Central wavelength (nm)	Adopted in the study (Yes/No)
B1	Coastal aerosol	60	443	No
B2	Blue	10	490	Yes
B3	Green	10	560	Yes
B4	Red	10	665	Yes
B5	Vegetation edge	20	705	Yes
B6	Vegetation edge	20	740	Yes
B7	Vegetation edge	20	783	Yes
B8	NIR	10	842	Yes
B8A	Vegetation edge	20	865	Yes
B9	Water vapour	60	945	No
B10	SWIR-cirrus	60	1375	No
B11	SWIR	20	1610	Yes
B12	SWIR	20	2190	Yes

3.2.5 Multiple regression models

Multiple linear regressions were conducted with RStudio (R i386 4.1.3) to determine the relationship between selected laboratory-analysed water quality indicators and remote sensing data from different

sampling times. Different sampling times refer to the moving remote sensing sampling time to estimate the fixed in-situ sampling data from a sampling date. The remote sensing sampling times were of images sampled from a month before, during, and after the in-situ sampling date. As noted by Schober, et al., (2018), the coefficient of determination (R^2) provides an estimation of the proportions of variance explained by the model. The coefficient of determination (R^2) variable determined the performance of the statistical model to predict the desired outcome. A bi-directional stepwise method uses the combination of forward selection and backward elimination to develop a desired model. The step-by-step process iteratively prunes a list of poor contributors/variables, while keeping the most useful variables in the model (Smith, 2018). This empirical approach developed remote sensing models to estimate water quality indicators with satellite reflectance data extracted a month before, during and after the in-situ sampling time.

The coefficient of determination value ranges between zero (0) and one (1). The model (estimator) was not suitable or valid to estimate water quality whenever the value was R^2 was below 0.4. However, a model with a coefficient of determination close to 1 was considered valid and a good estimator. The model follows Equation 1 below.

Multiple linear regression model equation

$$Y_i = \beta_0 + \Sigma_{i=1}^{n} (\beta_i x_i) + \varepsilon \quad \text{eq.(1)}$$

Y_i = dependent variable
β_0 = constant/intercept
β_i = regression coefficient
ε = random error

The performance of a liner model for water quality indicators at different sampling times first selection criteria was the coefficient of determination (R^2), which determined the estimation of the proportion of variance justified by the model. The second criterion assessed the highest significance value with the lowest p-value. The significance level of 95% ($p<0.05$) was the first option, whenever the models generated presented a lower significance value within a sampling period a significance level of 90% ($p<0.1$) was accepted. Estimators that met the criteria were assessed for their efficacy.

3.2.6 Validation metrics

The accuracy of the prediction of the derived models helped in determining their efficacy. Hence, validation was essential for ensuring that the remote sensing models were accurate and reliable. The best estimator remote sensing models were subjected to a validation process. These validation metrics were adopted: Pearson's correlation (R^2) root mean square error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE). The metrics were used to evaluate the models' performance to

estimate water quality indicators extracted before, during and after the field sampling date. The RMSE, MAE and MAPE are defined as:

$$\textbf{MAE}= \sum_{i=1}^{n} \frac{|Pi-Qi|}{n} \quad \text{eq. (2)}$$

$$\textbf{MAPE} = \sum_{i=1}^{n} \frac{|Pi-Qi|}{n} * 100 \quad \text{eq. (3)}$$

$$\textbf{RMSE} = \sqrt{\sum_{i=0}^{n} \frac{(Pi-Qi)^2}{n}} \quad \text{eq. (4)}$$

P_i is the satellite pixel value of the estimated/ to be validated

Q_i laboratory data

n the number of samples

3.3 RESULTS

Remote sensing models for laboratory-analysed water quality indicators (WQI) were presented according to their field sampling date, and the related satellite data extracted a month before, during, and after the in-situ sampling date.

3.3.1 Selected remote sensing regression models at different sampling times

Sentinel-2 (S2) linear regression models and their performance for estimating the FD_15July, FD_23August, FD_28September, and FD_21October field water quality indicators are presented in Appendix 1 and Appendix 2, respectively. The results only reported best-performed water quality remote sensing models of each indicator are displayed in **Table 3.4** and **Figure 3.3** below.

The best performed remote sensed model (R^2 =0.64) to estimate CDOM water quality indicator was observed between the FD_23 August field sampling time and S2 data extracted before the field survey date at the time lag of 44 days. The best turbidity water quality model (R^2=0.56) was between the FD_23August field survey and S2 data acquired the following month at a 36 days' time lag. The TSS best model (R^2=0.62) was extracted for FD_21October field survey and S2 data from a month before field date at seven days' time lag. These models had a potential to predict the water quality indicators at different sampling times between the field survey and satellite remote techniques successfully. However, the coefficient of determination is not adequate in providing model performance, because of its limitations in providing models residual information (Onyutha, 2020).

Table 3.4: Selected Sentinel-2 (S2) remote sensing models

Water quality indicator	Field sampling date (FD)	Sentinel-2 (S2) acquisition date	Time-lag (days)	Regression model
CDOM	FD_23Aug	S2_10 July	44	-0.9898 + (154.5726 * B2) + (-60.1605 * B3) + (-162.0896 * B4) + (08.4425 * B12)
Turbidity	FD_23Aug	S2_28 Sep	36	-6.399 + (150.284 * B4) + (274.522 * B6) + (-573.285 * B7) + (290.755 * B11) + (-64.415 * B12)
Chl-*a*	FD_27Sep	S2_18 Sep	5	36.82 + (-9035.51 * B4) + (-13894.56 * B5) + (16435.66 * B6) + (8007.64 * B12)
TSS	FD_21Oct	S2_03 Sep	7	5453 + (-100809 * B2) + (157610 * B3) + (-151696 * B5) + (-87892 * B6) + (212177 * B7) + (98223 * B8) + (-179430 * B8a)

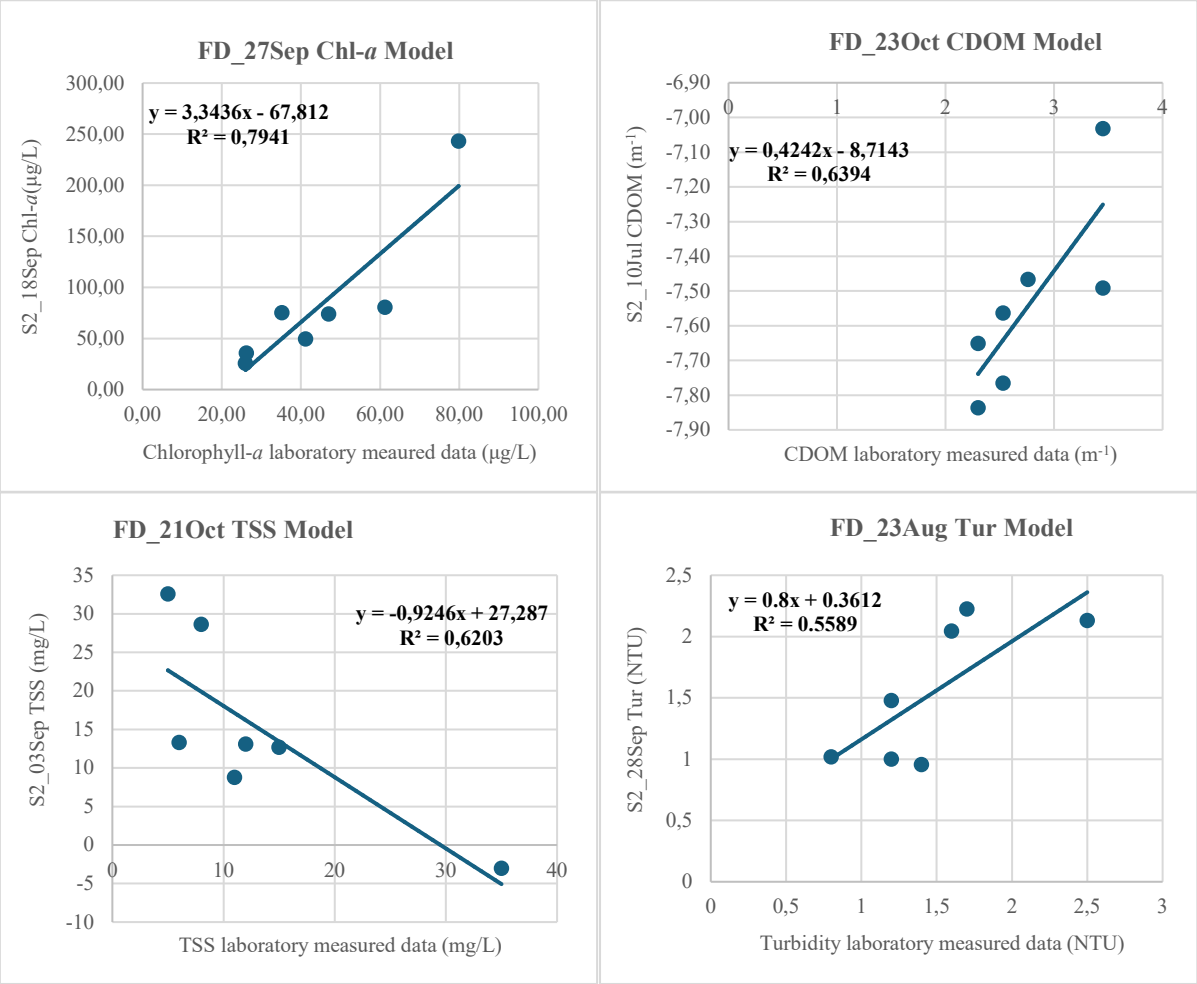


Figure 3.3: Predicted vs. actual using the remote sensing models

3.3.2 Water quality indicators remote sensing models accuracy performance

The accuracy performance provides the efficacy of the Sentinel-2 (S2) remote sensing models for estimating water quality indicators at the different sampling time with specified time lags as listed in **Table 3.5**. About 30% of the data was used for S2 remote sensing model accuracy assessment.

The R-squared as indicated in scatter plots in **Figure 3.3** displayed how well the S2 models predicted the laboratory-extracted indicators at different sampling times. The selected laboratory-extracted indicators explain over 50% of the variation in the S2 remote sensing models acquired at different sampling times. The models are the best fit to estimate the water quality indicator at different sampling times. The R^2 of chlorophyll-*a* indicator was the highest at 0.79 at five-day sampling times difference. The chlorophyll-*a* mean absolute error was 38.35 and the RMSE of 64.99 µg/L. However, The MAPE provided a limited information on the ability of S2 remote sensing models to predict field water quality data.

Table 3.5: The accuracy results of remote sensing models for respective sampling times

Water quality indicators	Time-lag (days)	Model accuracy performance			
		R^2	RMSE	MAE	MAPE
CDOM	44	0.64	10.31 1/cm	10.30	381.53
Turbidity	36	0.56	0.37 NTU	0.35	-6.53
Chlorophyll-*a*	5	0.79	64.99 µg/L	38.35	-66.22
TSS	7	0.62	19.62 mg/L	14.16	-1.99

3.4 DISCUSSION

To determine the efficacy of Sentinel-2 (S2) remote sensing models, the models were initially developed at different sampling times between the traditional and satellite techniques. Thereafter, model performance water established. The water quality models are discussed first and then their efficacy is revealed.

3.4.1 Remote sensing models for water quality indicators

The remote satellite sensing models were for predicting water quality indicators sampled during the July, August, September and October field months. The laboratory and satellite data were acquired at non-coincidental sampling times. The best selected model optimal band combinations for the four water quality indicators:

(a) Chlorophyll-*a* S2 model was generated with field and satellite data with five days sampling time difference within the same month. The Chl-*a* model consisted of four bands (red, vegetation edges at varying wavelength, and SWIR) at wavelength of about 665nm, 705nm, 740nm and 2190nm. The chlorophyll-*a* fluorescence was over the red and NIR domain of the electromagnetic spectrum (Tran, et al., 2023). The Chl-*a* model has the potential to estimate the 21September field Chl-*a* indicator, which absorbed most of the sun radiation at red visible near infra-red band (B4 and B5) and reflected the radiation at visible near infra-red band (B6) and short infra-red band (B12). This could have been as a result of high agal biomass which coincides with the blooming phase usually observed during the early spring season where high

photosynthesis takes place. The chlorophyll-*a* spectral coverage of the near-infrared is preferred because it is less affected by the CDOM (Bramich, et al., 2022).

(b) The determination of 23August field CDOM with the S2 remote sensing model developed from July satellite data at 44 days' sampling times difference, helped in assessing the physical and chemical conditions of the waterbody. The CODM model was generated by a combination of four bands in the visible spectrum and SWIR. The coloured component of the CDOM has strong ultraviolet and short-wavelength visible light-absorption properties. The S2_10July model strongly absorbed the light at the red band, followed by the green band. The insignificant changes in the CDOM concentration throughout the study period indicated that other factors contribute toward the indicator levels limitations.

(c) The August field turbidity indicator had potential to be estimated by S2_September models extracted 36 days after field sampling date. The turbidity S2 model was because of five band combination consisting of the red, vegetation edge at varying wavelength and SWIR at the different wavelength. The turbidity water quality indicator mainly scatters and absorbs light (Gholizadeh, et al., 2016). The turbidity of water is dependent on the presence of suspended solids. The turbidity remote sensing model absorbed light largely at the red-edge band (B7) and to a lesser extent at the short infrared band (B12). According to Caballero, et al., (2018), the use of red-edge and NIR regions of the spectrum allows turbidity to be detected in shallow waters. The August field turbidity was better estimated by the September S2 satellite data, since August marked the end of dry season of increased evaporation, as noted with reduced water level seen at 2m. The September satellite images had a clearer illumination to view and estimate the turbidity concentration of the water body.

(d) The October TSS field data was modelled by S2 September data extracted seven days before field sampling. The TSS model band combination included seven bands, two from the visible spectrum, vegetation edges in different wavelength, and NIR. The TSS remote sensing model produced a linear band combination reflected at the green band (B3), red-edge band (B7) and NIR (B8). The model had the transmission bands as the blue band (B2), red-edge (B6) and red-edge (B8a). The early spring TSS remote sensing model was robust enough to retrieve the mid-spring water which was seven days away. The models for retrieving the TSS water quality indicator are usually tweaked to suit the optical properties of the water body (Biorestita, et al., 2018).

The liner remote sensing models band combinations are site-specific but can produce robust estimates of water quality indicators when selected appropriately for the different sampling time between the tradition and satellite remote techniques (Gholizadeh, et al., 2016).

3.4.2 Remote sensing models' efficacy at different sampling times

A total of 12 remote sensing models were developed. However, only the four best performing models according to the performance criteria were selected per water quality indicator for efficacy assessment.

The CDOM remote sensing model for estimating FD_23 August field data was generated with S2 satellite data extracted 44 days before filed sampling date, at the efficacy of R^2 0.64, RMSE of 10.31 1/m and MAE of 10.30. The August turbidity field data was predicted by September S2 remote sensing data extracted 36 days after field sampling date with the efficacy of R^2 0.56, RMSE of 0.37 NTU and MAE of 0.35. Further, the September chlorophyll-*a* field data was predicted by S2 remote sensing model extracted five days before field sampling date at the R^2 0.79, RMSE of 64.99 µg/L and MAE of 38.35. Lastly, the October TSS field data was estimated by S2 remote sensing model acquired seven days after field sampling date at R^2 of 0.62, RMSE of 19.62 mg/L and MAE of 14.16. The efficacies of the S2 remote sensing models performed for indicators permit for the estimation of the field indicators with other S2 satellite data at relatively similar sampling times difference, with the ability to account for inaccuracies.

The planned acquisition schedule for the S2 affords the opportunity to combine the field survey and remote satellite sensing technique to extract water quality indictors at high efficacy. The changes of environmental factors are always expected. However, the established model error assessments for different sampling times can manage the potential inaccuracies. The error assessments for extracting water quality at different sampling time require further development as the techniques are adopted in water quality monitoring programme (Wang & Yang, 2019).

3.5 CONCLUSION AND RECOMMENDATIONS

This research has demonstrated the efficacy of remote satellite sensing models for estimating water quality indicators at different sampling times. Therefore, water quality monitoring programmes can adopt field sampling and remote satellite sensing techniques simultaneously to generate continuous water quality information when field sampling is hindered by costs and logistical limitations, or planned remote satellite sensing data extraction is affected by cloud cover and atmospheric interferences. The sampling framework consisting of both traditional field survey and Sentinel-2 (S2) remote sensing technique for different sampling times can be adopted. The sampling framework should incorporate the indicators' model sampling time difference sensitivity of the techniques to optimise the monitoring programme:

- The CDOM and turbidity remote sensing model can be applied for a wider field sampling difference widow of over a month. The remote sensing data extracted a month before and after field sampling can be incorporated in extracting the CDOM and turbidity, respectively.
- The Chl-*a* and TSS S2 models are appropriate for shorter sampling time difference widow between the techniques. The Chl-*a* model best performed with same-month data.
- The accuracy determination of S2 models opens opportunities to estimate same field data with suitable S2 data extracted at difference times, with known accounted errors.

3.6 LIMITATIONS

The research had the following limitations:

- Sentinel-2 (S2) remote sensing models generated were site-specific; their adaptability in other water resources needs to be tested.
- The sampling times were limited to the mid-winter and mid-spring seasons.
- Sampling times were limited to the available satellite and field sampling period.
- Other factors than atmospheric and cloud cover that might influence the accuracy of models at different sampling times were held constant.

The linear regression models ascertained, allow for water quality monitoring programmes to adopt both techniques simultaneously, since S2 remote sensing models sampling time differences and accuracy are known. The S2 remote sensing models are essential for predicting the spatial distribution of the water quality of waterbody to design an effective mitigation strategy for a continuous monitoring programme. The next chapter (Chapter 4) provides spatial distribution of water quality indicators and water quality maps along with the area coverage provided by the models.

3.7 REFERENCES

Adusei, Y. Y., Quaye-Ballard, J., Adjaottor, A. A. & Mensah, A. A., 2021. Spatial prediction and mapping of water quality of Owabi reservoir from satellite imageries and machine learning models. *The Egyptian Journal of Remote Sensing and Space Science*, 24(3), pp. 825-833.

Aranha, T., Martinez, J. M. S. E. & Barros, M. U., 2022. Remote analysis of the chlorophyll-*a* concentration using Sentinel-2 MSI image in a semiarid environment in Northeastern Brazil. *Water*.

Barsi, J. et al., 2018. Sentinel MSI and Landsat-8 OLI radiometric cross comparison over desert sites. *European Journal of Remote Sensing*, 51(1).

Biorestita, F. et al., 2018. The use of Sentinel-2 imagery for total suspended solids (TSS) estimation in Porong River, SiDoarjo. *Journal of Geodesi dan Geomatika*, 1(1), pp. 1-6.

Bramich, J., Bolch, C. J. & Fischer, A., 2022. Improved red-edge chlorophyll-*a* detection for Sentinel 2. *Ecological Indicators*, 120(2021).

Caballero, I. et al., 2022. Use of the Sentinel-2 and Landsat-8 satellites for water quality monitoring: An early warning tool in the Mar Menor coastal lagoon. *Remote Sensing*.

Dubovik, O. et al., 2021. Grand challenges in satellite remote sensing. *Frontiers in Remote Sensing*, Volume 2.

DWAF, 2002. National Eutrophication Monitoring Programme. *South African National Water Quality Monitoring Programmes Series*.

DWAF, 2004. National Water Resources Strategy.

DWS, 2017. *Water Quality Management Policies and Strategies for South Africa*, Pretoria, South Africa: s.n.

Elhag, M. et al., 2019. Assessment of water quality parameters using temporal remote sensing spectral reflectance in arid environments, Saudi Arabia. *Water*, pp. 1-14.

ESA, 2016. Sen2Cor Configuration and User Manual.

ESA, 2023. *The European Space Agency*. [Online]
Available at: https://sentinels.copernicus.eu/documents/247904/685211/Sentinel-2_User_Handbook.pdf/8869acdf-fd84-43ec-ae8c-3e80a436a16c?t=1438278087000.
[Accessed 14 August 2023].

Food, N., 2017. Comparing Sentinel-2A and Landsat 7 and 8 using surface reflectance over Australia. *Remote Sensing*, 9(7).

Gholizadeh, M. H., Melesse, A. M. & Reddi, L., 2016. A comprehensive review on water quality parameters estimation using remote sensing techniques. *Sensors*, Volume 16.

Govedarica, M. & Jakovljević, G., 2019. *Monitoring spatial and temporal variation of water quality parameters using time series of open multispectral data*. Paphos, Cyprus, s.n.

Kapalanga, T., Hoko, Z., Gumindoga, W. & Chikwiramakomo, L., 2021. Remote-sensing-based algorithms for water quality monitoring in Olushandja Dam, north-central Namibia. *Water Supply*, 21(5), pp. 1878-1894.

Malahlela, O., 2016. Inland waterbody mapping: towards improving discrimination and extraction of inland surface water features. *International Journal of Remote Sensing*, 37(19), pp. 4574-4589.

Mamuye, T. E. & Mutiu, K. A., 2021. Overview of water quality modeling. *Cogent Engineering*, 8(1).

Nadil, H., Mohammed, A. & Sohib, A., 2021. *Application of Sentinel 2 to evaluate colored dissolved organic matter algorithms for inlands water bodies in Jordan*. Jordan, s.n.

Ncube, E., Voyi, K. & Preez, H., 2012. Implementing a protocol for selection and prioritisation of organic contaminants in the drinking water value chain: case study of Rand Water, South Africa. *Water SA*, 38(4), pp. 487-504.

Nguyen, T. et al., 2020. Total suspended solid distribution in Hau River using Sentinel-2A satellite imagery. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, pp. 91-97.

Nhapi, I., 2015. Challenges for water supply and sanitation in developing countries: Case studies from Zimbabwe. *Understanding and Managing Urban Water in Transition*, Volume 15.

Omondi, A. N. et al., 2023. Estimation and mapping of water quality parameters using satellite images: A case study of Two Rivers Dam, Kenya. *Water Practice & Technology*, pp. 428-443.

Onyutha, C., 2020. From R-squared to coefficient of model accuracy for assessing “goodness-of-fits”. *Geoscientific Model Development*.

Pizani, F., Maillard, P., Ferreira, A. & Amorim, C., 2020. Estimation of water quality in a reservoir from Sentinel-2 MSI and Landsat-8 OLI sensors. *V-3-2020. 401-408. 10.5194/isprs-annals-V-3-2020-401-2020.*

Potes, M. et al., 2018. Use of Sentinel 2 – MSI for water quality monitoring at Alqueva reservoir, Portugal. *International Association of Hydrological Sciences*, pp. 73-79.

Qiaozhen, G. et al., 2020. An integrated study on change detection and environment evaluation of surface water. *Applied Water Science*, pp. 1-15.

Rivett, U., Champanis, M. & Wilson-Jones, T., 2013. Monitoring drinking water quality in South Africa: designing information systems for local needs. *Water SA*.

Roux, S. & Oelofse, S., 2010. *The rising costs of both sewage treatment and the production of potable water associated with increasing levels of pollutions in a portion of the Crocodile-West Marico water management area (Case study)*. Pretoria, CSIR.

Roy, R., 2019. An introduction to water quality analysis. *International Research Journal of Engineering and Technology (IRJET)*, 6(1).

Sagan, V. et al., 2020. Monitoring inland water quality using remote sensing: Potential and limitations of spectral indices, bio-optical simulations, machine learning, and cloud computing. *Earth-Science Reviews*, Volume 205.

Said, S. & Khan, A. S., 2021. Remote sensing-based water quality index estimation using data-driven approaches: a case study of the Kali River in Uttar Pradesh, India. *Environment, Development and Sustainability*, Volume 23, pp. 18252-18277.

SANews, 2019. *93% of South Africans have access to water services*. [Online]
Available at: https://www.sanews.gov.za/south-africa/93-south-africans-have-access-water-services#:~:text=Ninety%20three%20percent%20of%20the%20South%20African%20population,the%20population%20have%20access%20to%20basic%20sanitation%20services.
[Accessed 05 10 2023].

Satinder, A., 2014. *Comprehensive water quality and purification*. Elsevier.

Shi, X. et al., 2022. Retrieval of chlorophyll-*a* concentrations using Sentinel-2 MSI imagery in Lake Chagan based on assessments with machine learning models. *Remote Sensing*, 14(19).

Smith, G., 2018. Step away from stepwise. *Journal of Big Data*, Volume 5.

SMWW, 1998. *Standard methods for the examination of water and wastewater*, s.l.: s.n.

Toming, K. et al., 2016. First Experiences in Mapping Lake Water Quality Parameters with Sentinel-2 MSI Imagery. *Remote Sensing*.

Tran, M. et al., 2023. Band ratios combination for estimating chlorophyll-*a* from Sentinel-2 and Sentinel-3 in coastal waters. *Remote Sensing*, 15(6).

Uddin, G., Nash, S. & Olbert, A., 2021. A review of water quality index models and their use for assessing surface water quality. *Ecological indicators*.

USGS, 2019. *USGS Science for a changing world*. [Online]
Available at: https://www.usgs.gov/special-topics/water-science-school/science/how-much-water-there-earth. [Accessed 05 10 2023].

Van der Schyff, A. & Garbutt, M., 2023. Water quality monitoring systems for sub-Saharan Africa: Towards an effectiveness framework. International Conference on Information Systems and Advanced Technologies.

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

Wang, Z. & Menenti, M., 2021. Challenges and opportunities in Lidar remote sensing. *Remote sensing*.

Weather&Climate, 2023. *Weather-and-Climate*. [Online]
Available at: https://weather-and-climate.com/average-monthly-precipitation-Rainfall-inches,centurion-gauteng-za,South-Africa. [Accessed 02 August 2023].

World Health Organization, 2004. Guidelines for drinking water quality recommendations. Volume1.

Yang, H. et al., 2022. A review of remote sensing for water quality retrieval: progress and challenges. *Remote sensing*, 14(1770).

CHAPTER 4

Spatial mapping of water quality indicators from Sentinel-2: A case of Rietvlei Dam, South Africa

Innocent Rabohale[1*], Oupa Malahlela[1], Lutendo Mugwedi[1]

[1] Department of Geography and Environmental Science, Engineering and Agriculture, University of Venda, Thohoyandou 0950, South Africa

[*] Corresponding author: Innocentr@reml.co.za

ABSTRACT

Determining spatial patterns of water quality indicators is the key to water management programmes. Spatial mapping provides information essential for visualising and identifying contaminated areas requiring attention. Advancements in remote sensing technique has made it possible to model and visualise water quality concentration of waterbodies. The integration of the field survey and satellite technique to extract water quality at different sampling times at great accuracy has increase the potential to optimise water monitoring programme. The visualisation of water quality developed from models generated at different sampling times aids in the developing of mitigation and protection strategies of water resources. The main aim of this study was to create the spatial mapping of the four water quality indicators (i.e., TSS, CDOM, turbidity, chlorophyll-*a*) from Sentinel-2 (S2) imagery. The two objectives of the study were to map the surface concentration of water quality indicators, using Sentinel-2 models. Thereafter, to determine the area cover of elevated concentration of the water quality indicators in hectares. The study revealed the surface spatial distribution of the four water quality indicators. The Chl-*a* S2 model predicted the September field data, with the elevated concentration covering 28.9 hectares. The turbidity model successfully estimated the August field data, with the elevated concentration covering 36.14 hectares surface area. The TSS modelled the October elevated concentrations area of 1.45 hectares. Therefore, operational adaptations of determining elevated concentrations of water quality indicators to catch extremities of indicator concentration is necessary to combat pollutions and protect the resources. The models generated from data with different sampling times can be incorporated in water monitoring programmes to determine spatial distribution of water resources.

Keywords: accuracy, models, monitoring, sampling time, spatial map, validation, water quality

4.1 INTRODUCTION

Seventy-one percent of the Earth's surface is covered by water (USGS, 2019). However, freshwater resources such as rivers, streams, lakes and groundwater, which contain water for human consumption and other natural communities, are a mere 3% of the total (Adusei, et al., 2021). Freshwater is a scarce resource, threatened by myriad factors such as anthropogenic factors, population growth, economic development, polluted runoff, and climate change (World Health Organization, 2004). In South Africa, the distribution of water and the monitoring of surface water quality is uneven, due to a lack of access to funds and infrastructure (SANews, 2019). The continuous monitoring of surface freshwater water quality indicators ensures that water quality data are available on time for effective decision-making, and to prompt effective management strategies.

Sampling is an essential step in water quality monitoring management programmes. Developing countries depend on traditional techniques to gather water quality data. Traditional sampling and laboratory analysis methods to monitor water quality to identify pollutants and protect freshwater resources are time-consuming and expensive (DWAF, 2004). Advancements in remote satellite sensing technology and model development have made it possible to improve water quality monitoring programmes (Sagan, et al., 2020). Remote satellite sensing has been proven to be efficient in monitoring water quality indicators (Barsi, et al., 2018; Bramich, et al., 2022). Satellite observations complement the field sampling to provide continuous measurements of water quality data. However, sampling satellite data is often affected by cloudy weather, climate change, and atmospheric interference (Dubovik, et al., 2021). The synergy between satellite observation and field sampling has been established in Chapter 3. The freely available data from Sentinel-2 and Landsat missions are widely used in the continuous assessment of water quality assessment. Remote satellite sensing techniques quantify surface water, leaving reflectance stored in sensor image (ESA, 2023). The Sentinel mission has a higher temporal resolution of five days, which is better than that of the Landsat mission (Barsi, et al., 2018). The optical water quality indicators such as chlorophyll-*a*, turbidity, total suspended solids (TSS), and coloured dissolved organic matter (CDOM) have been successfully extracted with remote satellite sensing techniques (Barsi, et al., 2018; Biorestita, et al., 2018; Aranha, et al., 2022; Bramich, et al., 2022).

Developing remote satellite sensing models is commonly done by statistical techniques of correlating the water point surface leaving reflectance on images and related in-situ water quality concentrations (Mamuye & Mutiu, 2021). The remote satellite sensing models are an integral part of predicting the temporal and spatial distribution of surface water quality indicators. Researchers usually acquire satellite data and field data at the same time to improve the accuracy and performance of models for better temporal and spatial predictions (Elhag, et al., 2019; Aranha, et al., 2022). In Chapter 3 of this study it was apparent that remote sensing models developed with non-coincidental in-situ data extracted from the laboratory and satellite data from different sampling times had high efficacy of R^2 of 0.79, 0.64, 0.56 and 0.62 for chlorophyll-*a*, CDOM, turbidity and TSS water quality indicators, respectively. The

sampling difference between the field and the satellite sampling times of 44 days before August field sampling times for chlorophyll-*a*; 36 days after August field sampling time for CDOM; five days after early spring field sampling times for turbidity; and seven days before October field sampling time for TSS allows for continued assessment of surface water quality data.

The best-performed remote satellite sensing models were further used to estimate the surface quality concentration of the dam. Thus, these remote satellite sensing models were used to draw spatial maps for the four water quality indicators. Thereafter, the area coverage of the remote satellite sensing models for the water quality indicators were compared at different sampling times. Hence, the aim of this chapter is to establish the spatial mapping of water quality indicators for Sentinel-2: a case of Rietvlei Dam, South Africa. The aim is addressed through two objectives. Firstly, the map the surface concentration of water quality indicators using the Sentinel-2 models generated at the different sampling times between the fields survey and satellite techniques. Thereafter, determine the actual area in hectares of the elevated concentration of water quality indicators within the water body.

4.2 STUDY AREA AND METHODOLOGY

The study area of the research was Rietvlei Dam, which is in Gauteng Province, South Africa. The area is within the southern region of the A21 catchment area. The dam is on the Sesymylspruit, a tributary of the Hennops River, at -25.87752 latitude and 28.27196 longitudes with a surface area of 1.89 kilometres (**Figure 4.1**). The dam is fed by the Sesmylspruit River, five fountains, and five boreholes. About 20km upstream of Rietvlei Dam, Hartbeesfontein wastewater treatment works is located to the east of Kempton Park. The treatment works final effluent is released into the Swartspruit River. This river flows through agricultural land, Marais Lake, and wetlands before it joins the Sesymylspruit. The water quality information on the Rietvlei Dam is limited to the public (Roux & Oelofse, 2010). The dam is within the Rietvlei Nature Reserve which covers 3 200 hectares. The dam wall is 32 meters high with a capacity of 12 000 m3 and a mean elevation above sea level of about 1 525 m, with the highest point at 1 542m and the lowest point at 1 473m, at the dam outflow.

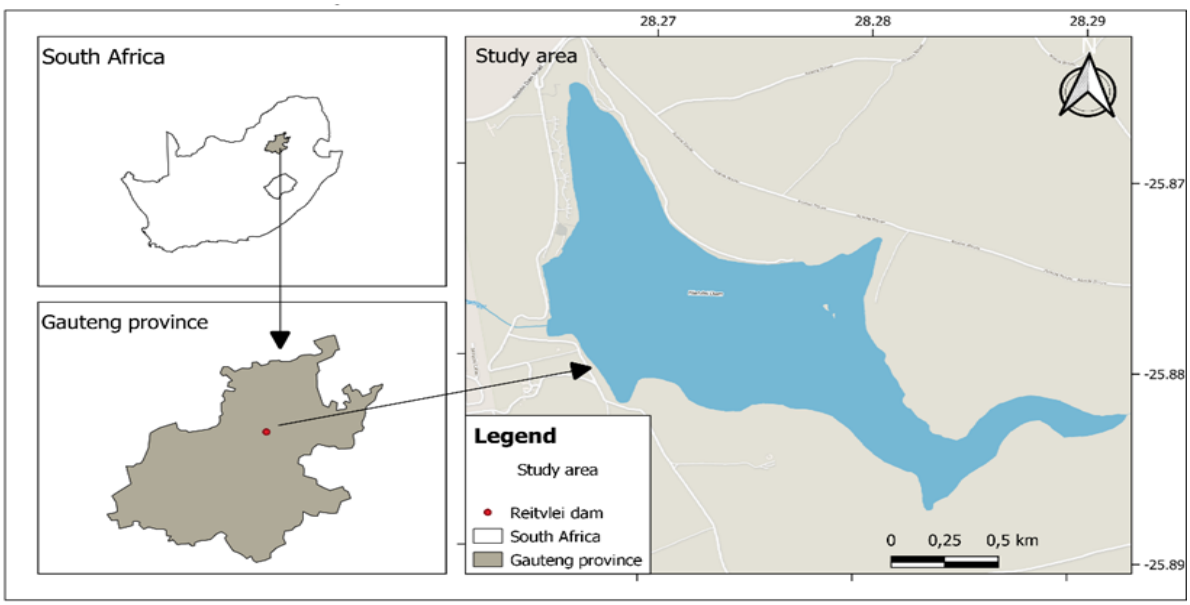


Figure 4.1: Rietvlei dam and sampling points

The linear regression between the laboratory-analysed water quality data and reflectance data of Sentinel-2 bands was conducted to generate remote sensing linear models. In Chapter 3, the four best-performed Sentinel-2 remote sensing models of each water quality indicator were extracted and selected. The linear models are listed in **Table 4.1**. The models were generated with data from different field and satellite sampling times (non-coincidental) and thereafter subjected to efficacy assessment.

Table 4.1: Remote sensing linear models for water quality indicators at different sampling seasons

Water quality indicator	Field sampling date (FD)	Sentinel-2 (S2) acquisition date	Time lag (days)	Regression model
CDOM	FD_23Aug	S2_10 July	44	-0.9898 + (154.5726 * B2) + (-60.1605 * B3) + (-162.0896 * B4) + (08.4425 * B12)
Turbidity	FD_23Aug	S2_28 Sep	36	-6.399 + (150.284 * B4) + (274.522 * B6) + (-573.285 * B7) + (290.755 * B11) + (-64.415 * B12)
Chl-*a*	FD_27Sep	S2_18 Sep	5	36.82 + (-9035.51 * B4) + (-13894.56 * B5) + (16435.66 * B6) + (8007.64 * B12)
TSS	FD_21Oct	S2_03 Sep	7	5453 + (-100809 * B2) + (157610 * B3) + (-151696 * B5) + (-87892 * B6) + (212177 * B7) + (98223 * B8) + (-179430 * B8a)

The linear models along with Sentinel-2 images were used to predict the 10m spatial resolution maps of water quality indicators of the dam. The open-source Quantum Geographic Information System (QGIS) software was used to generate the spatial maps and to estimate the coverage area of the remote sensing models of the water quality indicators extracted at different sampling times.

4.2.1 Image processing

The acquired Sentinel-2 Level-2 images were radiometrically corrected and with atmospherically corrected surface reflectance. These images were suitable for water quality assessment. These products were adopted for water quality indicator measurements. The Sentinel-2 remote sensing model bands were adopted for the prediction of the spatial water quality indicator surface concentration.

4.2.2 Water body extraction

To overcome the difficulty of discriminating the water body from other land features, such as vegetation and rocks, a simple water index (SWI) Formula (1) was adopted for extracting the inland surface body (Malahlela, 2016).

$$SWI = 1/\sqrt{(Blue\ band - SWIR\ band)} \quad \text{eq.(1)}$$

4.2.3 Area coverage of elevated concentration as depicted by water quality model

The Sentinel-2 remote sensing models were further assessed to determine the high concentrations of water quality indicators. The South African National Standard (SANS) 241 Drinking water specification was adopted to set the minimum operational requirement for the turbidity (NTU) water quality indicator. Standards such as those used to determine the distribution of the water quality indicators of the dam.

4.3 RESULTS

The results are laid out in two sections. The first section displays the spatial distribution of water quality indicators extracted from Sentinel-2 models generated at different sampling times from the field sampling dates. The second section shows the sentinel-2 remote sensing model surface area coverage.

4.3.1 Surface spatial distribution of water quality indicators at different sampling times

The surface spatial distribution of the water quality indicators derived with Sentinel-2 remote sensing models extracted from different field sampling times are presented in **Figures 4.2 to 4.5**.

The spatial distribution of the CDOM remote sensing model was generated 44 days before the field sampling period. The model surface spatial distribution with the variance contributions of blue, green, red, and shortwave infrared bands with the efficacy of $R^2 = 0.64$. Low values of CDOM measured at values closer to 0 m-1 were generally observed from the laboratory-measured data for the mid-winter period, as represented by **Figure 4.2**.

The surface spatial distribution of turbidity water quality indicator is displayed in **Figure 4.3**. A high concentration of light-absorbing particles was seen around the edge of the dam. The edge of the dam was shallow during the winter season field sampling times. The predicted surface spatial distribution of turbidity water quality indicator was mainly due to the variance contribution of red, red-edge, and

shortwave infrared with the efficacy of R^2=0.56. The CDOM and turbidity water quality indicators were generated from the late winter field water samples, and their relationship was consistent. The slightly elevated turbidity water quality indicator showed an inverse relationship to the light absorption coefficient of CDOM. Granted, the remote sensing models were extracted with reflectance data from the season before and the season after the field sampling date, for CDOM and turbidity respectively. The light-limiting by turbidity effect on photosynthesis by reducing the light that passes through the to reach the bottom of the dam.

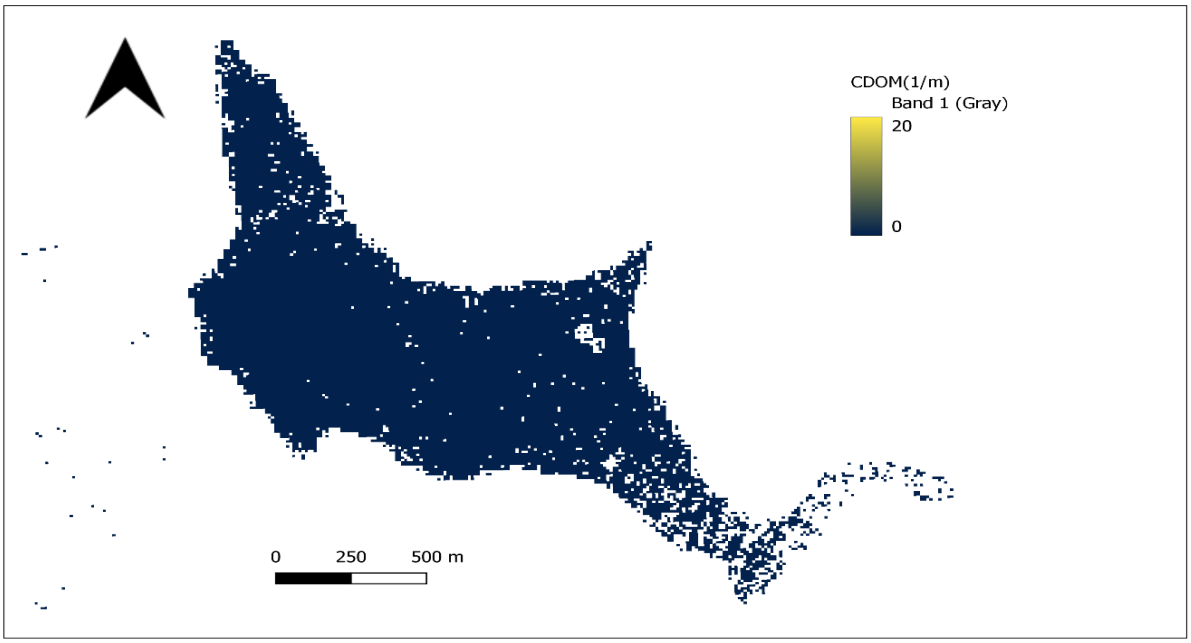


Figure 4.2: August CDOM concentration derived with S2 data extracted 44 days before field sampling date

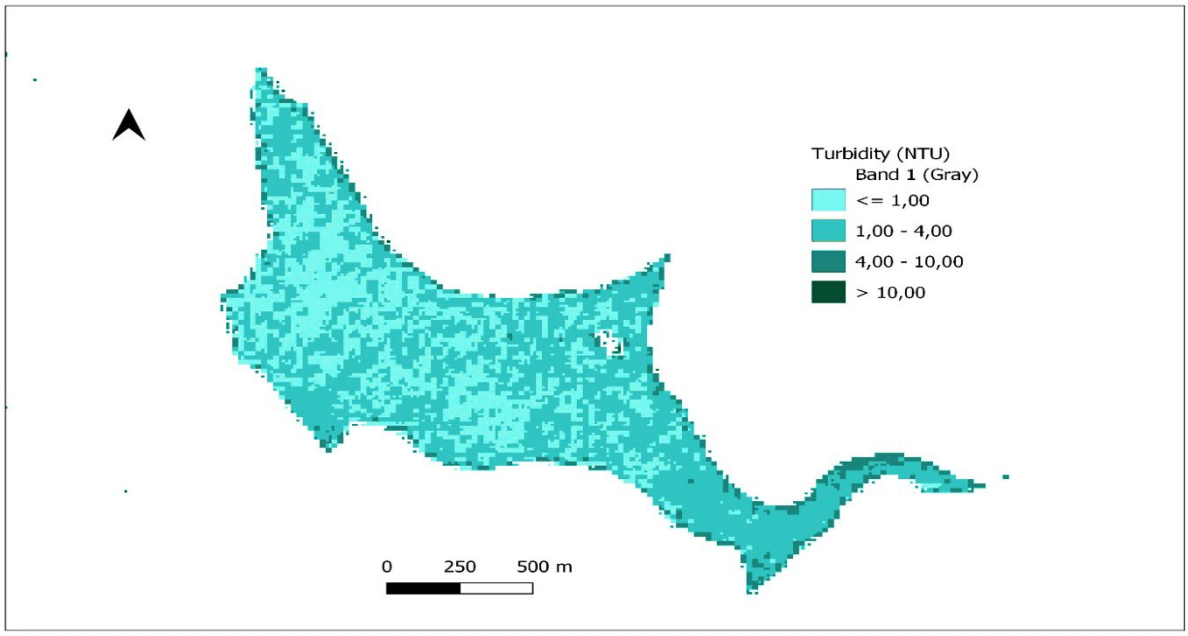


Figure 4.3: August turbidity concentration derived with S2 data extracted 36 days after field sampling date

The chlorophyll-*a* remote sensing model surface spatial distribution is represented by **Figure 4.4**. The model shows the early spring water quality concentrations. The concentration levels were mostly high at the edge of the dam during the early spring. There were patches of slightly elevated concentrations (25-50 µg/L) within the dam and more towards the outflow (northern part) of the dam. The variance contribution of the chlorophyll-*a* model was red, red-edge, and shortwave infrared bands with a correlation coefficient of R^2=0.79.

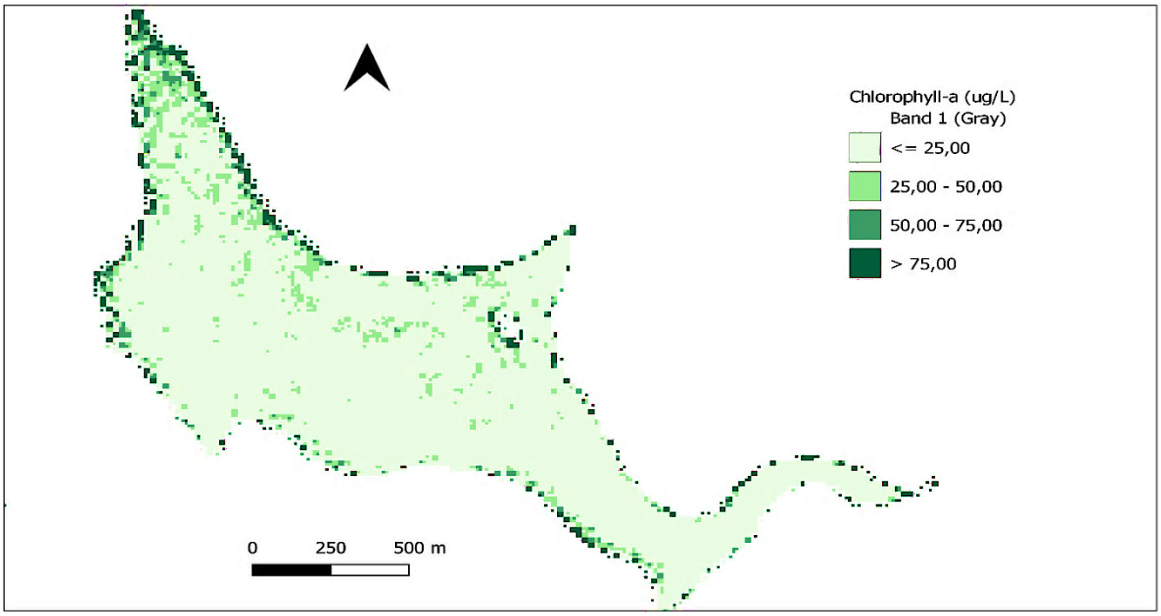


Figure 4.4: September chlorophyll-*a* concentration derived with S2 data extracted five days after field sampling date

The surface spatial distribution of the mid-winter TSS water quality indicator, as predicted by the early-spring Sentinel-2 remote sensing model, shows patches of elevated concentrations in the dam. The elevated concentrations are mainly between 500 and 750 mg/L, as depicted in **Figure 4.5**. The mid-spring TSS surface distribution was modelled by late winter Sentinel-2 reflectance data, which was just seven days before the field sampling date.

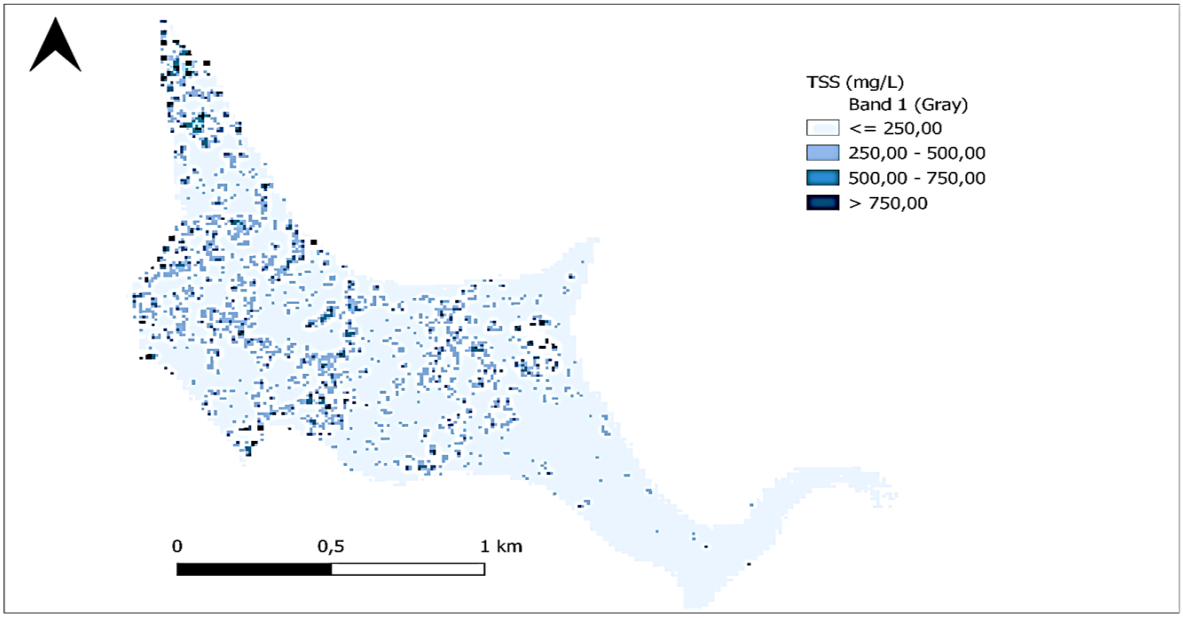


Figure 4.5: October TSS concentration derived with S-2 data extracted seven days before field sampling date

4.3.2 Surface spatial distribution area coverage of the water quality indicators

The Sentinel-2 remote sensing model for the four water quality indicators extracted at different sampling times showed a varied spatial distribution coverage, similar to other quantitative maps compiled through remote satellite sensing techniques (Kapalanga, et al., 2021; Aranha, et al., 2022). The concentration levels of the water quality indicators as extracted through laboratory contributed to the yield of spatial distribution response of the remote sensing models. The area coverage of the high concentrations of the water quality indicator as covered by the Sentinel-2 remote sensing models. The histogram of the water quality indicators spatial distribution shows the high spatial coverage of the turbidity and chlorophyll-*a*, as depicted in **Figure 4.6**. Other researchers determine the trends of the surface water quality distributions with validated models (Jin, et al., 2023; Kowe, et al., 2023).

The Sentinel-2 remote sensing model for CDOM had no high concentration area coverage as depicted in the figure below. Nonetheless, the turbidity remote sensing model spatial distribution coverage of elevated concentration (turbidity $\geq$ 1 NTU) covered about 36 hectares of surface area. While chlorophyll-*a* remote model for estimating early spring concentrations spatial distribution area for high concentrations (Chl-*a* $\geq$ 25 µg/L) covered a surface area of about 28 hectares.

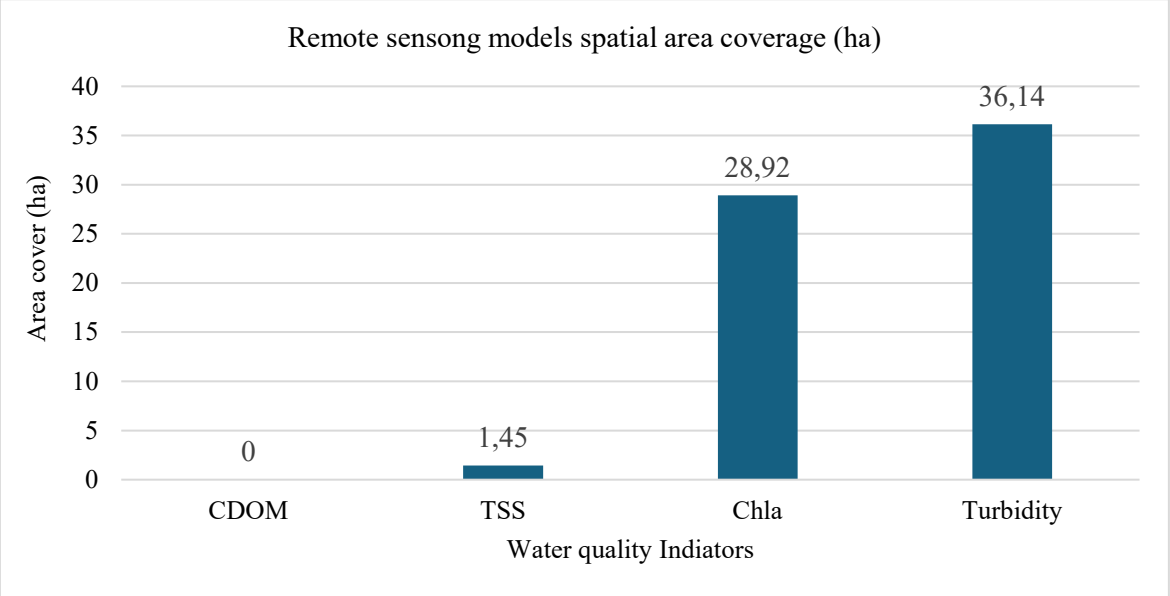


Figure 4.6: Spatial distribution of the high concentration of the water quality indicator at different sampling times

4.4 DISCUSSION

4.4.1 Surface spatial distribution of water quality indicators at different sampling times

The spatial distribution of the water quality indicators is essential to visualise the variation in the water quality for the entire dam, which aids in devising remediation efforts (Ramadas & Samantaray, 2017). The spatial maps of the four best-performed S2 remote sensing models extracted at the different sampling times from field sampling showed a high accuracy. The laboratory extracted CDOM water quality concentration was consistently low-concentrations during the late winter season. According to the study by Zhang, et al., (2021), the CDOM concentration is often elevated in summer and autumn periods, which were seasons excluded from this study. Nonetheless, the extracted Sentinel-2 remote sensing model sensitivity aligned with the field observation of overall low CDOM measurements.

The spatial distribution of the chlorophyll-*a* and turbidity concentration retrieved from satellite data confirmed the parches patterns of the concentration observed as field measurements. The agreement between field observation and satellite measurements is consistent with findings by other researchers (Adusei, et al., 2021; Kapalanga, et al., 2021). The chlorophyll-*a* concentration observed during early spring (September) qualified the degree of trophy of the Rietvlei dam, which ranged between the eutrophic and hypereutrophic at the edge. The dam is experiencing rising pollution because of upstream industrial and human activities, which lead to the growing eutrophic concerns (Roux & Oelofse, 2010). The reliability and accuracy of evaluated surface spatial distribution of chlorophyll-*a* with remote sensing model extracted from data of about ±2 to 12 days sampling times difference has been attained by other researchers (Aranha, et al., 2022; Li, et al., 2021).

The August CDOM and turbidity indicators were predicted by Sentinel-2 remote sensing model with over a month sampling time difference from field sampling date. The September chlorophyll-*a* was estimated with the remote sensing models of a week sampling time difference not more than five days. The October TSS indicator was estimated with S2 remote sensing of a week sampling time difference. As noted in the review by Llodrà-Llabrés, et al., (2023), the performance of the indicators is contributed by the water resource and the season. The field sampling from the two seasons (i.e., winter and spring) considered in this study, indicated that Sentinel-2 models with sampling time difference can satisfactorily estimate and map water quality indicators.

4.4.2 Surface spatial distribution area coverage of the water quality indicators

The surface spatial distribution of water quality indicators is significant because water quality indicators are highly dynamic. The temporal and spatial distribution of water course are driven by the concentrations of the water indicator at the time of assessment (Llodrà-Llabrés, et al., 2023). The water quality indicators extracted by early spring (i.e., September) Sentinel-2 remote sensing models displayed the surface area coverage for elevated concentrations distinctly at different concentration levels. As stated by World Health Organization (2004), the desirable threshold for turbidity for manging operational risks is > 1 NTU, but for the aesthetic risk is 5 NTU. The patches or unevenly distributed turbidity concentration on the surface could be afforded by the turning turbines in the dam. While the chlorophyll-*a* concentration levels indicated the trophic level (25µg/mL) of the dam, consistent with the national classification of the dam (DWAF, 2002). Accurate maps are essential to effectively manage the water resource. Although the spatial distribution coverage focused on the elevated concentrations of the water quality indicators, the ability for the Sentinel-2 remote sensing models to distinguish concentration level is of great relevance for continuous water quality assessment. The chlorophyll-*a* and turbidity remote sensing models covered an extensive spatial large area for elevated concentrations. A large spatial distribution coverage area of chlorophyll-*a* and turbidity can assist in the predictions of future change scenarios in water quality monitoring using sampling times difference data.

4.5 CONCLUSION AND LIMITATIONS

This study evaluated the spatial distribution of the best-performed remote sensing models extracted at different sampling times for field sampling. Sentinel-2 remote sensing models were successful in providing the spatial distribution of the water quality indicators. The remote sensing models estimated the area coverage of high concentrations of water quality indicators, and show that:

- The August turbidity of >1 NTU concentration covered an area of 38.14ha with the Sentinel-2 (S2) model generated with September data at sampling time difference of 36 days.
- The September Chl-*a* eutrophic levels of >25µg/L coverage was 28.92ha with a S2 model October S2 satellite data at five days sampling times difference window.
- The October TSS remote at the elevated concentration of >250mg/L area coverage was 1.45ha with the S2 remote sensing model with 5-days sampling times difference window.

- Less can be said about the area coverage of the August CDOM water quality indicator because of its overall low concentration during the late winter field sampling period. It had an inverse relationship with the turbidity water quality indicator. Nonetheless, the S2 remote sensing model was generated from 44-days sampling time difference. The model can be applied for the longer sampling time difference windows.

This research finding makes it feasible to adopt both methods to manage the water quality of surface water resources. When there are limited funding and logistical issues hinder field sampling, the remote sensing models can be used to fill the gaps for effective decision-making.

4.6 LIMITATIONS

- The generated models were generated for the eutrophic water body, thus their applicability in other water systems is not known.
- The limited satellite images due to atmospheric and weather conditions limited the amount of data to generate remote sensing models.
- The cost associated with field sampling and laboratory analyses limited the continuous field data for generating and testing remote sensing indicators.

4.7 REFERENCES

Adusei, Y. Y., Quaye-Ballard, J., Adjaottor, A. A. & Mensah, A. A., 2021. Spatial prediction and mapping of water quality of Owabi reservoir from satellite imageries and machine learning models. *The Egyptian Journal of Remote Sensing and Space Science*, 24(3), pp. 825-833.

Aranha, T., Martinez, J. M. S. E. & Barros, M. U., 2022. Remote analysis of the chlorophyll-*a* concentration using Sentinel-2 MSI image in a semiarid environment in northeastern Brazil. *Water*.

Barsi, J. et al., 2018. Sentinel MSI and Landsat-8 OLI radiometric cross comparison over desert sites. *European Journal of Remote Sensing*, 51(1).

Biorestita, F. et al., 2018. The use of Sentinel-2 imagery for total suspended solids (TSS) estimation in Porong River, SiDoarjo. *Journal of Geodesi dan Geomatika*, 1(1), pp. 1-6.

Bramich, J., Bolch, C. & Fischer, A., 2021. Improved red-edge chlorophyll-*a* detection for Sentinel 2. *Ecological Indicators*.

Bramich, J., Bolch, C. J. & Fischer, A., 2022. Improved red-edge chlorophyll-*a* detection for Sentinel 2. *Ecological Indicators*, 120(2021).

Caballero, I. et al., 2022. Use of the Sentinel-2 and Landsat-8 satellites for water quality monitoring: An early warning tool in the Mar Menor coastal lagoon. *Remote Sensing*.

Dubovik, O. et al., 2021. Grand challenges in satellite remote sensing. *Frontiers in Remote Sensing*, Volume 2.

DWAF, 2002. National Eutrophication Monitoring Programme. *South African National Water Quality Monitoring Programmes Series*.

DWAF, 2004. National Water Resources Strategy.

DWS, 2017. *Water Quality Management Policies and Strategies for South Africa*, Pretoria, South Africa: s.n.

Elhag, M. et al., 2019. Assessment of water quality parameters using temporal remote sensing spectral reflectance in arid environments, Saudi Arabia. *Water*, pp. 1-14.

ESA, 2016. Sen2Cor Configuration and User Manual.

ESA, 2023. The European Space Agency. [Online]
Available at: https://sentinels.copernicus.eu/documents/247904/685211/Sentinel-2_User_Handbook.pdf/8869acdf-fd84-43ec-ae8c-3e80a436a16c?t=1438278087000
[Accessed 14 August 2023].

Food, N., 2017. Comparing Sentiel-2A and Landsat 7 and 8 using surface reflectance over Australia. *Remote Sensing*, 9(7).

Gholizadeh, M. H., Melesse, A. M. & Reddi, L., 2016. A comprehensive review on water quality parameters estimation using remote sensing techniques. *Sensors*, Volume 16.

Govedarica, M. & Jakovljević, G., 2019. *Monitoring spatial and temporal variation of water quality parameters using time series of open multispectral data*. Paphos, Cyprus, s.n.

Kapalanga, T., Hoko, Z., Gumindoga, W. & Chikwiramakomo, L., 2021. Remote-sensing-based algorithms for water quality monitoring in Olushandja Dam, north-central Namibia. *Water Supply*, 21(5), pp. 1878-1894.

Malahlela, O., 2016. Inland waterbody mapping: Towards improving discrimination and extraction of inland surface water features. *International Journal of Remote Sensing*, 37(19), pp. 4574-4589.

Mamuye, T. E. & Mutiu, K. A., 2021. Overview of water quality modeling. *Cogent Engineering*, 8(1).

Nadil, H., Mohammed, A. & Sohib, A., 2021. *Application of Sentinel 2 to evaluate colored dissolved organic matter algorithms for inlands water bodies in Jordan*. Jordan, s.n.

Ncube, E., Voyi, K. & Preez, H., 2012. Implementing a protocol for selection and prioritisation of organic contaminants in the drinking water value chain: Case study of Rand Water, South Africa. *Water SA*, 38(4), pp. 487-504.

Nguyen, T. et al., 2020. Total suspended solid distribution in Hau River using Sentinel-2A satellite imagery. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, pp. 91-97.

Nhapi, I., 2015. Challenges for water supply and sanitation in developing countries: Case studies from Zimbabwe. *Understanding and Managing Urban Water in Transition*, Volume 15.

Omondi, A. N. et al., 2023. Estimation and mapping of water quality parameters using satellite images: A case study of Two Rivers Dam, Kenya. *Water Practice & Technology*, pp. 428-443.

Onyutha, C., 2020. From R-squared to coefficient of model accuracy for assessing “goodness-of-fits”. *Geoscientific Model Development*.

Pizani, F., Maillard, P., Ferreira, A. & Amorim, C., 2020. Estimation of water quality in a reservoir from Sentinel-2 MSI and Landsat-8 OLI sensors. *V-3-2020. 401-408. 10.5194/isprs-annals-V-3-2020-401-2020*.

Potes, M. et al., 2018. Use of Sentinel 2 – MSI for water quality monitoring at Alqueva reservoir, Portugal. *International Association of Hydrological Sciences*, pp. 73-79.

Qiaozhen, G. et al., 2020. An integrated study on change detection and environment evaluation of surface water. *Applied Water Science*, pp. 1-15.

Rivett, U., Champanis, M. & Wilson-Jones, T., 2013. Monitoring drinking water quality in South Africa: Designing information systems for local needs. *Water SA*.

Roux, S. & Oelofse, S., 2010. *The rising costs of both sewage treatment and the production of potable water associated with increasing levels of pollutions in a portion of the Crocodile-West Marico water management area (Case study)*. Pretoria, CSIR.

Roy, R., 2019. An Introduction to water quality analysis. *International Research Journal of Engineering and Technology*, 6(1).

Sagan, V. et al., 2020. Monitoring inland water quality using remote sensing: Potential and limitations of spectral indices, bio-optical simulations, machine learning, and cloud computing. *Earth-Science Reviews*, Volume 205.

Said, S. & Khan, A. S., 2021. Remote sensing-based water quality index estimation using data-driven approaches: a case study of the Kali River in Uttar Pradesh, India. *Environment, Development and Sustainability*, Volume 23, pp. 18252-18277.

SANews, 2019. *93% of South Africans have access to water services*. [Online]
Available at: https://www.sanews.gov.za/south-africa/93-south-africans-have-access-water-services#:~:text=Ninety%20three%20percent%20of%20the%20South%20African%20population,the%20population%20have%20access%20to%20basic%20sanitation%20services.
[Accessed 05 10 2023].

Satinder, A., 2014. *Comprehensive water quality and purification*. Elsevier.

Shi, X. et al., 2022. Retrieval of chlorophyll-*a* concentrations using Sentinel-2 MSI imagery in Lake Chagan based on assessments with machine learning models. *Remote Sensing*, 14(19).

Smith, G., 2018. Step away from stepwise. *Journal of Big Data*, Volume 5.

SMWW, 1998. *Standard Methods for the Examination of Water and Wastewater*, s.l.: s.n.

Toming, K. et al., 2016. First experiences in mapping lake water quality parameters with Sentinel-2 MSI Imagery. *Remote Sensing*.

Tran, M. et al., 2023. Band ratios combination for estimating chlorophyll-*a* from Sentinel-2 and Sentinel-3 in coastal waters. *Remote Sensing*, 15(6).

Uddin, G., Nash, S. & Olbert, A., 2021. A review of water quality index models and their use for assessing surface water quality. *Ecological Indicators*.

USGS, W. S. S., 2019. USGS Science for a changing world. [Online]
Available at: https://www.usgs.gov/special-topics/water-science-school/science/how-much-water-there-earth. [Accessed 05 10 2023].

Van der Schyff, A. & Garbutt, M., 2023. Water quality monitoring systems for sub-Saharan Africa: Towards an effectiveness framework. International Conference on Information Systems and Advanced Technologies.

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

Wang, Z. & Menenti, M., 2021. Challenges and opportunities in Lidar remote sensing. *Remote Sensing*.

Weather&Climate, 2023. Weather-and-Climate. [Online]
Available at: https://weather-and-climate.com/average-monthly-precipitation-Rainfall-inches,centurion-gauteng-za,South-Africa. [Accessed 02 August 2023].

World Health Organization, 2004. Guidelines for drinking water quality recommendations. Volume1.

Yang, H. et al., 2022. A review of remote sensing for water quality retrieval: progress and challenges. *Remote Sensing*, 14(1770).

CHAPTER 5:
RESEARCH SYNTHESIS, SUMMARY AND CONCLUSION

5.1 RESEARCH SYNTHESIS AND SUMMARY

Continuous water sampling and analysis affects efficient water quality monitoring programmes, which contribute towards water quality and quantity protection (DWS, 2019). The growing adoption of optical remote satellite sensing for monitoring water quality in conjunction with field surveys globally, presents an opportunity for continuous water quality assessment and mapping. To adopt both remote satellite sensing and field survey simultaneously for continuous water quality mapping a sampling time synergy needed to be understood. Thus, this study assessed the effects of sampling time on the accuracy of extracting water quality indicators through remote sensing and geospatial techniques. The three objectives addressed the aim as introduced in Chapter . Each objective established unique findings also presenting opportunities for further research.

The first objective as addressed in Chapter 2 had a question: "What was the correlation between the laboratory-extracted data and Sentinel-2 satellite data acquired a month before, the during and after field sampling times?" The finding indicated correlations between the remote sensing data and field survey data: The strong correlation of four selected laboratory-extracted data were observed for the following sampling times difference:

- The September indicator had a strong correlation (r=-0.72) with October S2 satellite with a month sampling time difference.
- The September CDOM indicator was highly correlated (r=0.73) with S2 satellite data from extracted in the same month (sampling time difference).
- The September Chl-*a* moderately correlated (r=0.63) with S2 satellite data from the same month (sampling time difference).
- The October TSS strongly correlated (r=0.71) with the S2 satellite data a month before (sampling time difference).

The second objective addressed in Chapter 3 had a question: "What was the efficacy of the Sentinel-2 remote sensing models for estimating water quality indicators at different sampling times?"

The water quality indicators with moderate to strong correlations generated in Chapter 2 were adopted to develop remote sensing models thereafter determine their efficacy. A total of 12 remote sensing models were generated. Only the best-performed remote sensing models per water indicator were selected. Their efficacy and their sampling times difference results were as follows:

- Turbidity efficacy observed at R2 = 0.56; RMSE = 0.37 NTU, MAE =0.35; 36 days sampling time variation.

- CDOM efficacy was R^2=0.64; RMSE=10.311/m, MAE=10.30; 44 days sampling times difference.
- TSS efficacy was R^2 = 0.62; RMSE = 19.62mg/L, MAE =14.16; seven days sampling time difference.
- Chl-*a* efficacy was R^2 = 0.79; RMSE = 64.99 µg/L, MAE =38.35; five days sampling time difference.

Lastly, the third objective was addressed by answering the question: "What was the spatial mapping of the water quality indicators from Sentinel-2?" The water quality indicators with efficacy results generated in Chapter 3 were implemented to generate the map and estimate area coverage of high concentrations of the indicator. It was observed that:

- The turbidity covered an area of 38.14ha for concentration above 1 NTU.
- The Chl-*a* remote sensing area coverage of 28.92ha for concentrations above 25µg/L.
- The TSS remote sensing model area coverage was 1.45m for concentrations above 250mg/L
- Less could be mentioned about the area coverage of the CDOM water quality indicator because of its overall low concentration during the late winter field sampling period.

In conclusion, the effect of sampling time on the accuracy of extracting water quality indicators at different sampling times through remote sensing and geospatial techniques showed that the accuracy of extracting water quality indicators depended on the sampling times difference of the satellite images and field sampling period. The synergy between remote satellite sensing and field survey sampling methods was established, and it was observed that water quality indicators respond differently to sampling time differences. The TSS and chlorophyll-*a* water quality indicators were sensitive to sampling times difference, so a shorter sampling time difference of about week was more effective. The CDOM and turbidity water quality indicators could withstand a longer sampling time difference of over a month. Therefore, to implement continuous water quality assessment using both remote sensing and field survey sampling, the sampling time difference and the selected water quality indicators are to be considered. This study contributes to scientific understanding of the remote sensing applications in water quality management for the inland waterbodies. It is recommended that further research be considered for different sampling periods, different waterbodies, more water quality indicators, and more models be generated for the adoption of common water quality indices.

REFERENCES

Biswas, A., & Tortajada, C. (2019). Water quality management: A global neglected issue. *International journal of water resources development*, 35(6), 913-916.

Din, E. S., & Zhan, Y. (2017). Improving the accuracy of extracting surface water quality levels (SWQLs) using remote sensing and artificial neural network: A case study in the Saint John River, Canada. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 42(4).

ESA, (n.d.). *The European Space Agency.* Retrieved August 14, 2023, from https://www.esa.int/Applications/Observing_the_Earth/Copernicus/Sentinel-2/Satellite_constellation#:~:text=The%20Sentinel-2%20mission%20is%20based%20on%20a%20constellation,and%20this%20is%20even%20faster%20at%20higher%20latitudes

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

Huang, C., Chen, Y., Zhang, S., & Wu, J. (2018). Detecting, extracting, and monitoring surface water from space using optical sensors: A review. *Reviews of Geophysics*, 56, 333-360. doi:10.1029/2018RG000598

Joshi, A., Agrawal, S., & Chauhan, P. (2020). Reduction of atmospheric effects in satellite images during the COVID-19 induced lockdown. *Journal of the Indian Society of Remote Sensing*, 48, 1613–1625.

Khatri, N., & Tyagi, S. (2015). Influence of natural and anthropogenic factors on surface and groundwater quality in rural and urban areas. *Frontiers in Life Science*, 8(1), 23-39. doi:10.1080/21553769.2014.933716

Li, X., Ding, J., & Ilya, N. (2021). Machine learning method for quick identification of water quality index (WQI) based on Sentinel-2 MSI data: Ebinur Lake case study. *Water Supply*, 21(3).

Namugize, J., & Jewitt, G. (2018). Sensitivity analysis for water quality monitoring frequency in the application of a water quality index for the uMngeni River and its tributaries, KwaZulu-Natal, South Africa. *Water SA*, 44. doi:10.4314/wsa.v44i4.01

Potes, M., Rodrigues, G., Penha, A. M., Novais, M. H., Costa, M. J., Salgado, R., & Morais, M. M. (2018). Use of Sentinel-2 – MSI for water quality monitoring at Alqueva reservoir, Portugal. *International Association of Hydrological Sciences*, 73-79.

Rwanga, S., & Ndambuki, J. (2017). Accuracy assessment of land use/land cover classification using remote sensing and GIS. *International Journal of Geosciences, 8*, 611-622.

Sent, G., Biguino, B., Favareto, L., Cruz, J., Sá, C., Dogliotti, A., . . . Brito, A. (2021). Deriving water quality parameters using Sentinel-2 imagery: A case study in the Sado Estuary, Portugal. *Remote Sensing*, 13(1043). doi:10.3390/rs13051043

Xianghong, C., Min, F., Hao, J., Jia, S., & Bei, J. (2015). Downscaling MODIS surface reflectance to improve water body extraction. *Advances in Meteorology*. doi:10.1155/2015/424291

APPENDICES

APPENDIX 1

Table 3.1: Remote sensing regression models for water quality indicators at different sampling seasons

Field sampling season	Remote sensing model ID	Sampling time difference	Model description
Mid-winter FD_15Jul	S2_196 Tur (same season)	0	-15.31 + (343.99 * B3) + (-866.04 * B4) + (1389.23 * B5) + (-1043.97 * B6) + (-921.25 * B7) + (992.56 * B8) + (200.52 * B8a) + (-503.55 * B11) + (501.07 * B12)
	S2_191Chl-*a* (same season)	0	1426.7 + (-5007.9 * B2) + (-9737.8 * B4) + (12456.3 * B5) + (-8262.1 * B6) + (-8866.2 * B7) + (6932.2 * B12)
	S2_216CDOM (season after)	20	15.557 + (185.193 * B2) + (195.650 * B3) + (-242.007 * B5) + (-105.048 * B6) + (10.677 * B7) + (-279.340 * B8) + (74.272 * B8a) + (189.263 * B11) + (-249.785 * B12)
Late winter FD_23Aug	S2_191CDOM (season before)	44	-0.9898 + (154.5726 * B2) + (-60.1605 * B3) + (-162.0896 * B4) + (08.4425 * B12)
	S2_221CDOM (same season)	14	-11.51 + (-194.15 * B2) + (235.63 * B3) + (-362.54 * B6) + (659.75 * B7) + (94.53 * B8) + (-708.97 * B8a) + (-148.26 * B11) + (547.02 * B12)
	S2_261Chl-*a* (season after)	26	75.37 + (8887.23 * B2) + (-15093.58 * B3) + (10010.49 * B4) + (-0040.45 * B6) + (17296.59 * B7) + (17022.87 * B8) + (11826.88 * B11) + (1865.89 * B12)
	S2_271Tur (season after)	36	-6.399 + (150.284 * B4) + (274.522 * B6) + (-573.285 * B7) + (290.755 * B11) + (-64.415 * B12).
Early spring FD_28Sep	S2_236 Chl-*a* (season before)	45	116.8 + (-4791.7 * B3) + (2819.8 * B5) + (-7047.6 * B8) + (-3743.7 * B11) + (12460.5 * B12)
	S2_266 Tur (same season)	5	-23.285 + (526.094 * B3) + (-217.889 * B4) + (-165.786 * B5) + (-234.396 * B7) + (-160.776 * B8a) + (430.236 * B11)
	S2_296 Tur (season after)	25	-1.705 + (-104.581 * B3) + (-175.548 * B4) + (161.040 * B5) + (-207.977 * B7) + (380.023 * B12)
	S2_261 Chl-*a* (season after)	5	36.82 + (-9035.51 * B4) + (-13894.56 * B5) + (16435.66 * B6) + (8007.64 * B12)
Mid-spring FD_21Oct	S2_246 TSS (season before)	7	5453 + (-100809 * B2) + (157610 * B3) + (-151696 * B5) + (-87892 * B6) + (212177 * B7) + (98223 * B8) + (-179430 * B8a)

Detailed multiple regression results from different field sampling times correlated with the related satellite image extracted a season before, same and after in-situ sampling time.

Table 4.1(a): Stepwise regression models data for estimating Mid-winter FD_15Jul water quality indicators.

		Indicators	**Turbidity**		**TSS**		**CDOM**		**Chla**	
Field date	Image season	Images dates	r^2	P-value	r^2	P-value	r^2	P-value	r^2	P-value
FD_15Jul	Mid-winter	S2_186	0.9584	0.06927	0.9257	0.01845	0.7461	0.0659	0.6536	0.0334
		S2_191	0.5873	0.1441	**0.9762**	**0.006619**	**0.9631**	**0.05843**	**0.8032**	**0.03007**
		S2_196	**0.9291**	**0.05258**	0.9072	0.0855	0.2661	0.1824	0.5572	0.6376
	Late winter	**S2_216**	**0.8707**	**0.007786**	0.7455	0.4279	**0.9484**	**0.02918**	0.1861	0.1236
		S2_221	0.8986	0.01236	0.8033	0.07401	0.669	0.2601	0.8661	0.07022
		S2_226	0.5362	0.1076	0.5331	0.2138	0.3352	0.233	0.6192	0.04911
		S2_236	0.8586	0.07913	**0.8658**	**0.07062**	0.9076	0.009511	**0.9588**	**0.06824**
		S2_241	0.8504	0.036	0.6069	0.2294	0.5868	0.2615	0.5277	0.3649

Table 4.2(a): Stepwise regression models data for estimating Mid-winter FD_23 Aug water quality indicators.

		Indicator	**Turbidity**		**TSS**		**CDOM**		**Chl-*a***	
Field date	Image season	Image dates	r^2	p-Value	r^2	p-Value	r^2	p-Value	r^2	p-Value
FD_23Aug	Mid-winter	S2_186	0.7212	0.03736	0.4907	0.07027	0.7419	0.009789	0.6027	0.3825
		S2_191	0.8705	0.02432	**0.62**	**0.01771**	**0.7582**	**0.007418**	**0.9205**	**0.006219**
		S2_196	**0.9765**	**0.03074**	0.6498	0.08306	0.6441	0.1753	0.609	0.2262
	Late winter	S2_216	0.2931	0.1484	0.6398	0.4758	0.7848	0.03975	0.9497	0.0901
		S2_221	**0.9315**	**0.004065**	0.6966	0.00601	**0.9474**	**0.008167**	0.8433	0.004451
		S2_226	0.8899	0.001154	**0.8676**	**0.002343**	0.831	0.005923	0.782	0.01529
		S2_236	0.7272	0.08174	0.7271	0.6573	0.9014	0.01145	0.8449	0.3774
		S2_241	0.4809	0.2953	0.5453	0.1967	0.8202	0.4431	**0.8667**	**0.02631**
	Early spring	S2_246	0.838	0.04447	0.8962	0.01319	0.8408	0.01532	0.3395	0.391
		S2_251	0.8418	0.0418	0.8516	0.1905	0.8585	0.03096	0.3893	0.4653
		S2_261	0.912	0.07766	0.2575	0.5654	**0.8944**	**0.003991**	**0.965**	**0.003066**
		S2_266	0.8457	0.09549	0.697	0.2131	0.7932	0.1752	0.5491	0.09635
		S2_271	**0.8155**	**0.008227**	**0.9825**	**0.02007**	0.8796	0.1345	0.8669	0.002394

Table 4.3 (a): Stepwise regression models data for estimating early spring FD_28Sep water quality indicators.

		Indicator	**Turbidity**		**TSS**		**CDOM**		**Chl-*a***	
Field date	Image season	Images dates	r^2	P-value	r^2	P-value	r^2	P-value	r^2	P-value
FD_28Sep	Late winter	S2_216	0.7238	0.08478	0.9062	0.08706	0.9702	0.01023	0.7201	0.0882
		S2_221	0.9349	0.00352	0.8961	0.1042	0.8187	0.2619	0.8567	0.03203
		S2_226	**0.9913**	**9.873e-05**	0.5654	0.08337	0.7977	0.003457	0.3875	0.06745
		S2_236	0.8658	0.008793	**0.7773**	**0.04419**	**0.9257**	**0.001231**	**0.8111**	**0.008988**
		S2_241	0.9608	0.0008099	0.2522	0.2022	0.8214	0.02214	0.7258	0.01263
	Early spring	S2_246	0.9382	0.003029	0.6876	0.00692	0.8459	0.09525	0.6758	0.02543
		S2_251	0.8162	0.008122	0.9131	0.002083	**0.8601**	**0.01008**	0.4027	0.1454
		S2_261	0.8267	0.001773	0.3801	0.07207	0.5886	0.02584	**0.6906**	**0.02095**
		S2_266	**0.9367**	**0.0007134**	0.661	0.07427	0.3544	0.2055	0.7115	0.1903
		S2_271	0.8784	0.1367	**0.935**	**0.0001488**	0.8462	0.03874	0.4112	0.05432
	Mid-spring	**S2_276**	0.7848	0.09311	**0.8453**	**0.09599**	0.8193	0.05936	**0.6062**	**0.05629**
		S2_296	**0.8178**	**0.007858**	0.7144	0.1858	0.191	0.1181	0.5471	0.1943
		S2_301	0.7934	0.08397	0.6842	0.1253	**0.7137**	**0.04109**	0.5935	0.5686

Table 4.4 (a): Stepwise regression models data for estimating mid-spring FD_21Oct water quality indicators.

		Indicator	**Turbidity**		**TSS**		**CDOM**		**Chl-*a***	
Field date	Image season	Images dates	r^2	P-value	r^2	P-value	r^2	P-value	r^2	P-value
FD_21Oct	Early spring	**S2_246**	**0.8418**	**0.04181**	**0.9473**	**0.001908**	0.9211	0.02124	0.8735	**0.06197**
		S2_251	0.5306	0.9152	0.9426	0.03556	0.8407	0.01533	0.3733	0.9837
		S2_261	0.2875	0.1549	0.7823	0.04122	0.7707	0.2147	0.7565	0.1267
		S2_266	0.8341	0.1114	0.9215	0.02099	0.8285	0.05174	0.5106	0.3969
		S2_271	0.4747	0.7749	0.7332	0.1581	0.6906	0.05406	0.8292	0.1185
	Mid-spring	S2_276	0.8748	0.1437	0.1815	0.1288	0.7322	0.1595	0.5931	0.2512
		S2_296	**0.9737**	**0.0362**	0.7243	0.4752	0.782	0.3448	0.4972	0.9372
		S2_301	0.6917	0.2217	**0.7999**	**0.164**	**0.9856**	**0.015**	**0.7992**	**0.07808**

APPENDIX 2

Detailed validation results for all the best-performing models from all the related field data and Sentinel-2 data.

7.1 FD_15July models validation results

Table 4.2: Validation metrics results for remote satellite sensing models estimating FD_196 data.

Field date	Image season	Accuracy metrics	Turbidity (NTU)	TSS (mg/L)	CDOM (1/m)	Chl-*a* (µg/L)
FD_15Jul	Mid-winter	**R^2**	**0.32**	0.25	0.0007	**0.37**
		RMSE	1.24	3.95	0.471	124.26
		MAE	1.16	3.15	0.37	121.67
		MAPE	-52.03	9.93	2.56	3149.75
	Late winter	**R^2**	0.05	0.001	**0.43**	0.20
		RMSE	1.99	8.55	0.50	5685.65
		MAE	1.88	6.90	0.46	5685.55
		MAPE	-68.93	-29.59	-3.15	204852.72

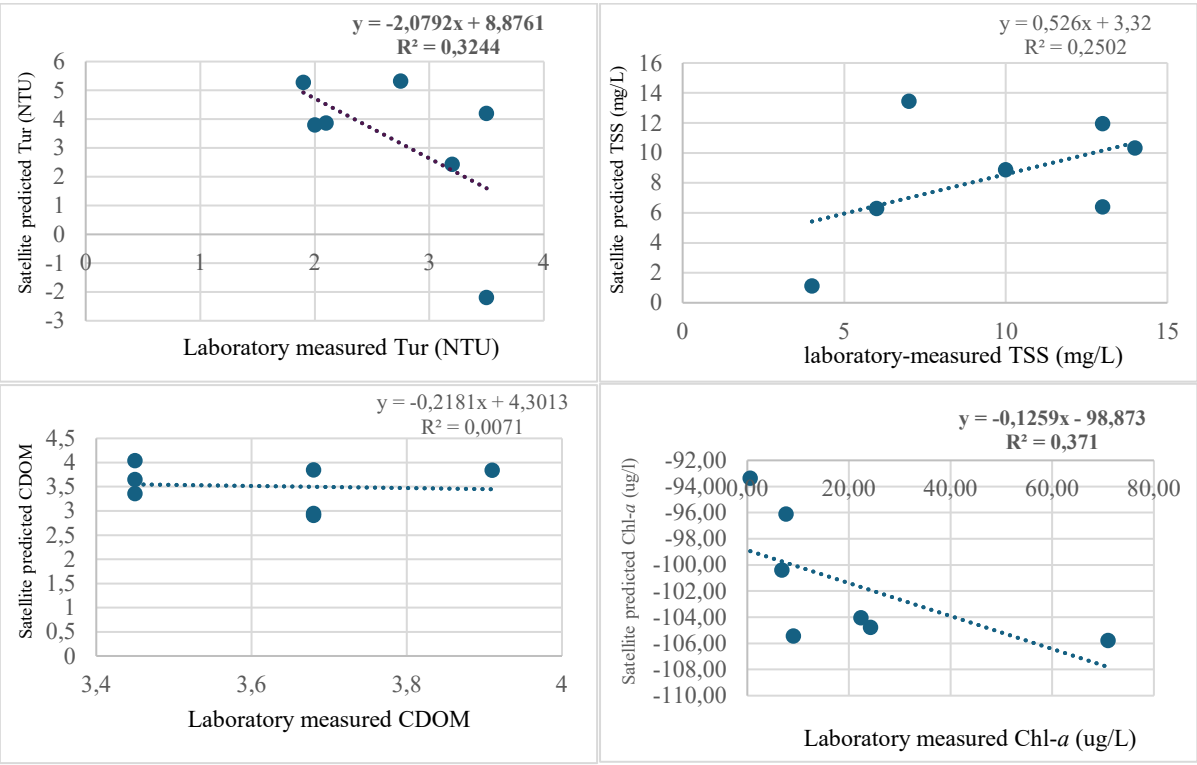
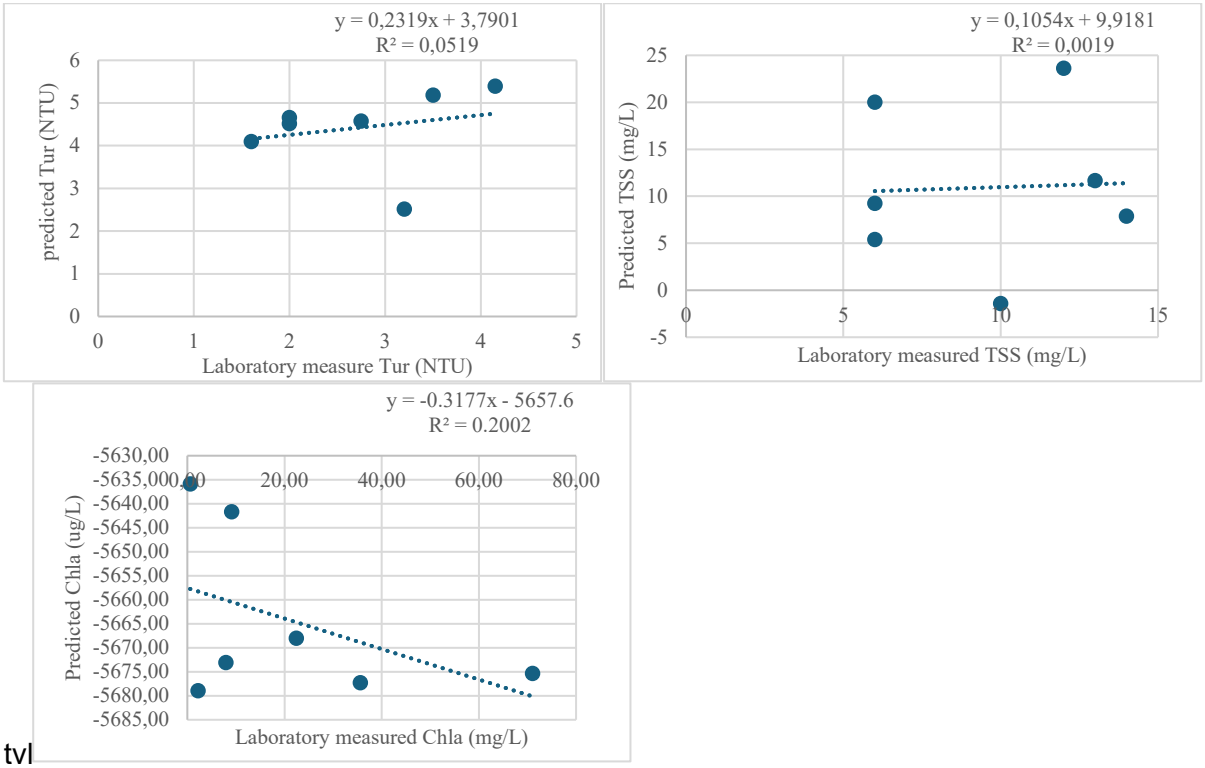


Figure 4.1(a): August remote sensing model validation performance estimating FD_15Jul in-situ data



tvl

Figure 4.2(b): August remote sensing model validation performance estimating FD_15Jul in-situ data

7.2 FD_23 August models validation results

4.3.2 Model validation result

Field date	Image season	Accuracy metrics	Turbidity model	TSS model	CDOM model	Chl-*a* model
FD_235	Mid-winter	**R^2**	0.19	0.09	**0.64**	0.12
		RMSE	1.71	2.39	10.31	111.08
		MAE	1.13	1.87	10.30	78.39
		MAPE	-47.50	-43.40	381.53	-465.38
	Late winter	**R^2**	0.01	0.15	**0.46**	0.28
		RMSE	1.11	1.59	9.79	2297.34
		MAE	0.78	1.49	9.09	1707.65
		MAPE	-47.45	12.42	292.15	-8469.16
	Early spring	**R^2**	**0.56**	0.18	0.034	**0.58**
		RMSE	0.37	4.42	33.97	53.60
		MAE	0.35	3.36	23.28	37.21
		MAPE	-6.53	-0.46	701.55	129.21

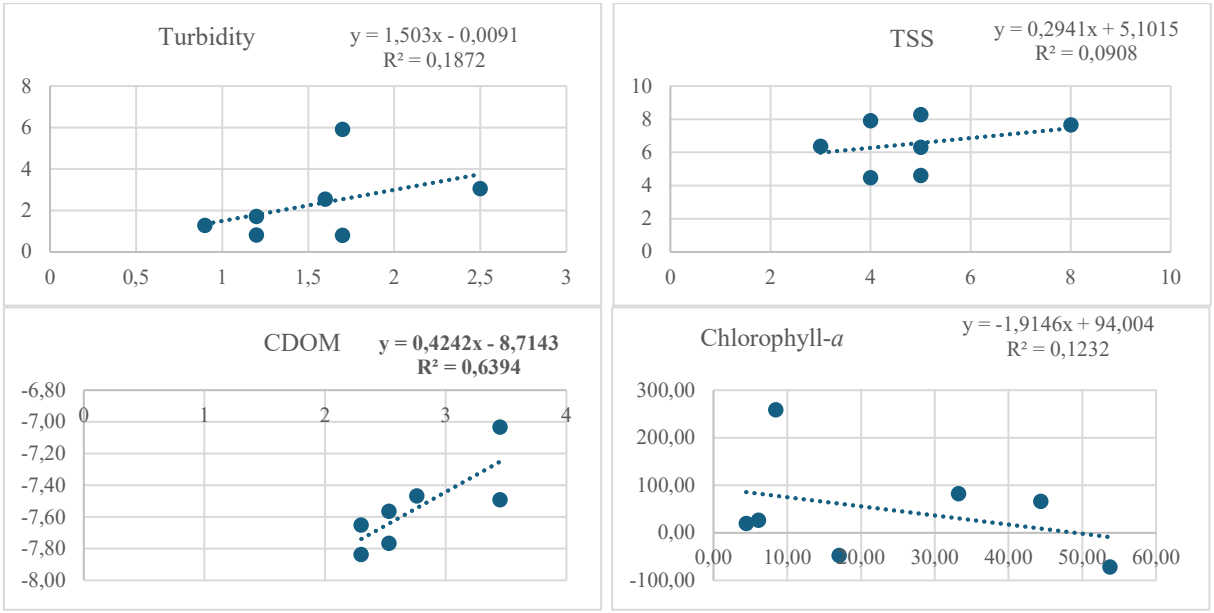


Figure 4.3(a): July remote sensing model validation performance estimating FD_23Aug in-situ data

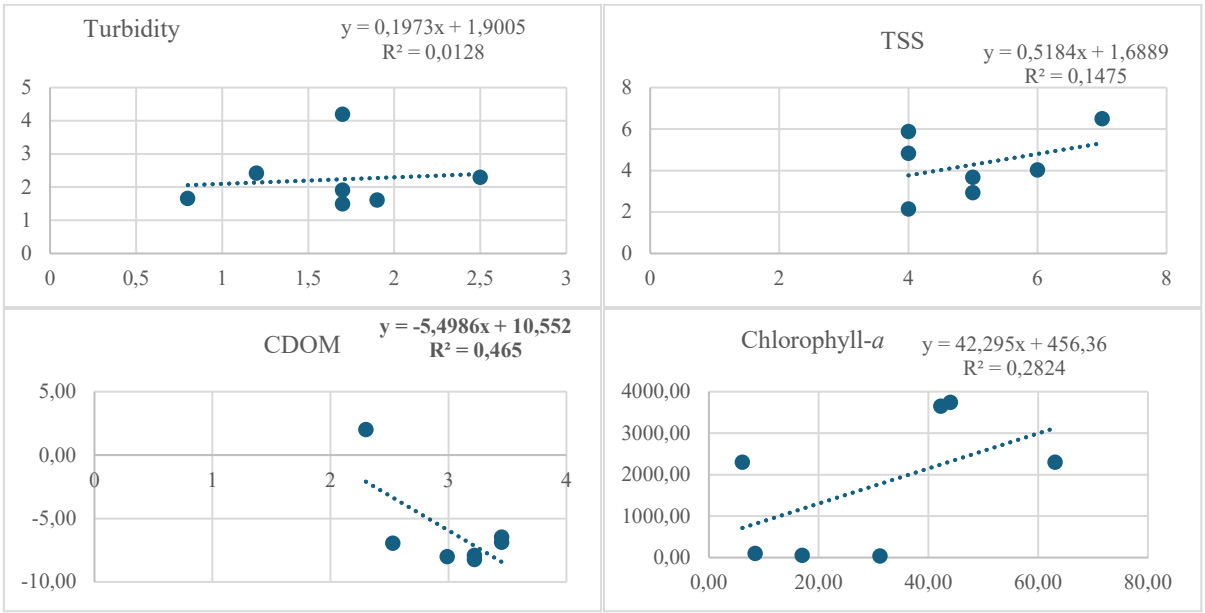


Figure 4.3(b): August remote sensing model validation performance estimating FD_23Aug in-situ data

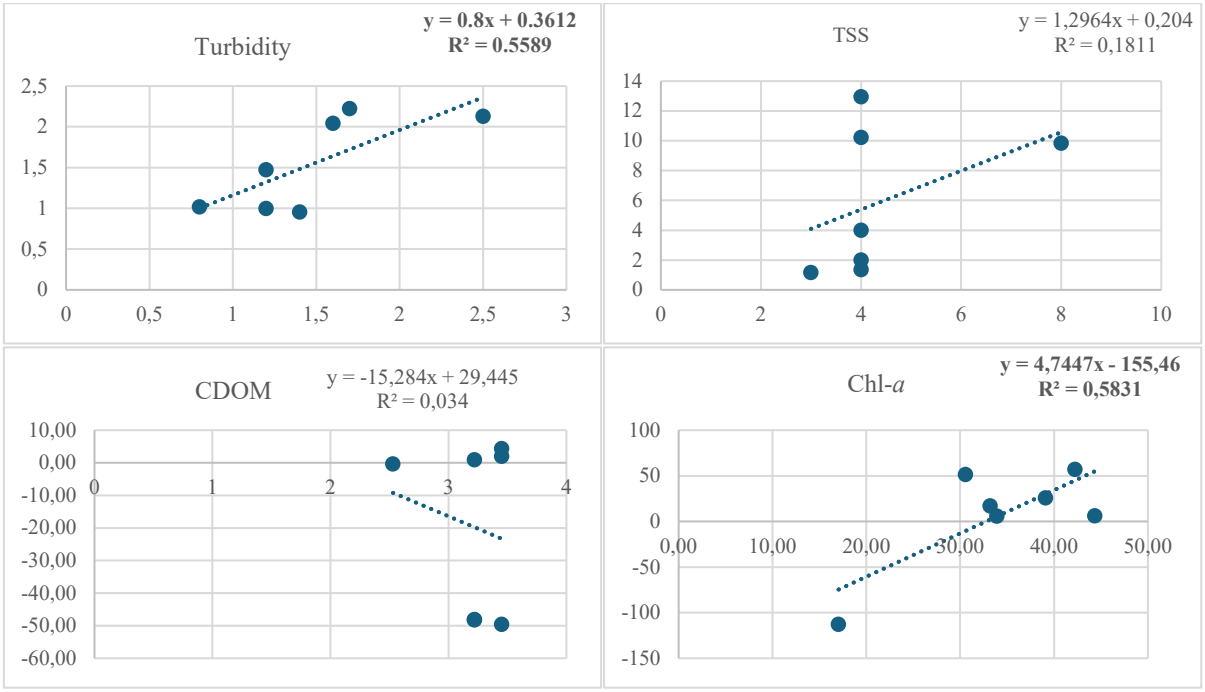


Figure 4.3(c): September remote sensing model validation performance estimating FD_23Aug in-situ data

7.2 FD_28September models validation results

Table 4.4: Summary of validation results of best-performing models estimating FD_28 Sep indicators

Field date	Image season	Accuracy metrics	Turbidity model	TSS model	CDOM model	Chl-*a* model
FD_271	Late-winter	**R^2**	0.16	0.006	0.03	**0.54**
		RMSE	2.21	26.01	0.68	43.35
		MAE	1.86	17.16	0.48	27.72
		MAPE	-16.99	-68.72	6.12	14.53
	Early spring	**R^2**	**0.46**	0.27	0.29	0.23
		RMSE	2.27	24.02	0.35	28.59
		MAE	2.12	20.69	0.26	22.33
		MAPE	89.91	211.22	-9.23	-82.43
	Mid spring	**R^2**	**0.48**	0.059	0.12	**0.79**
		RMSE	0.84	23.33	0.40	64.99
		MAE	0.67	15.42	0.28	38.35
		MAPE	3.02	66.10	-10.05	-66.22

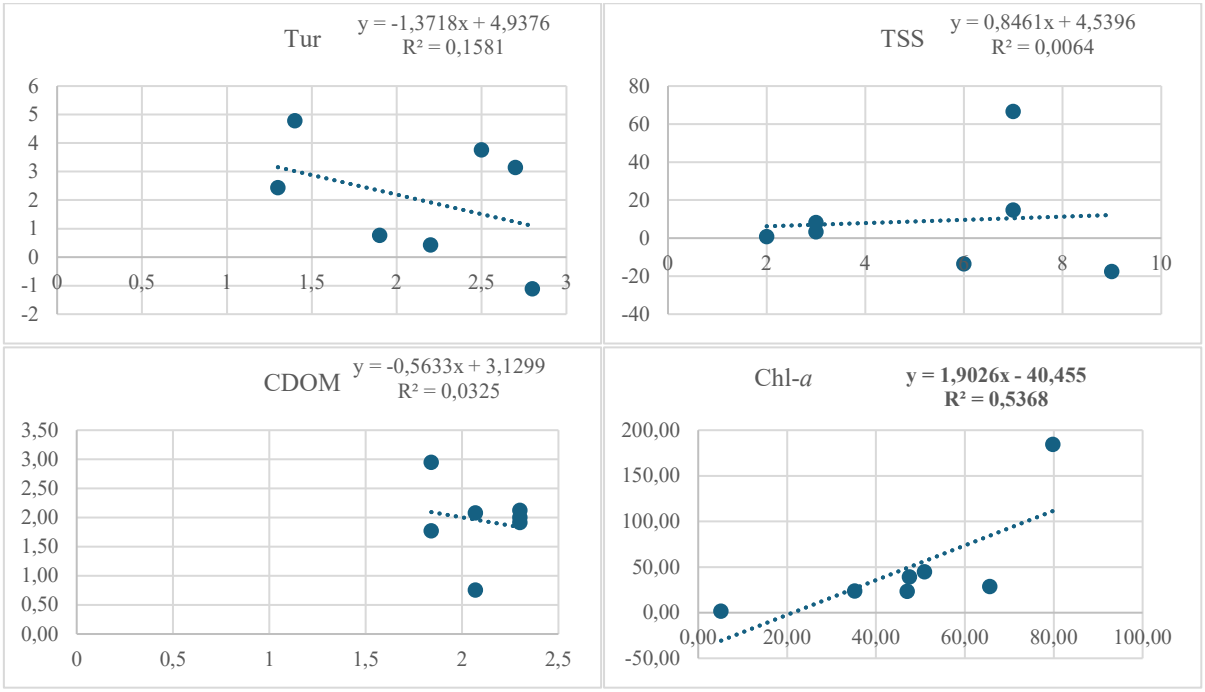


Figure 4.4(a): August remote sensing model validation performance estimating FD_28Sep in-situ data

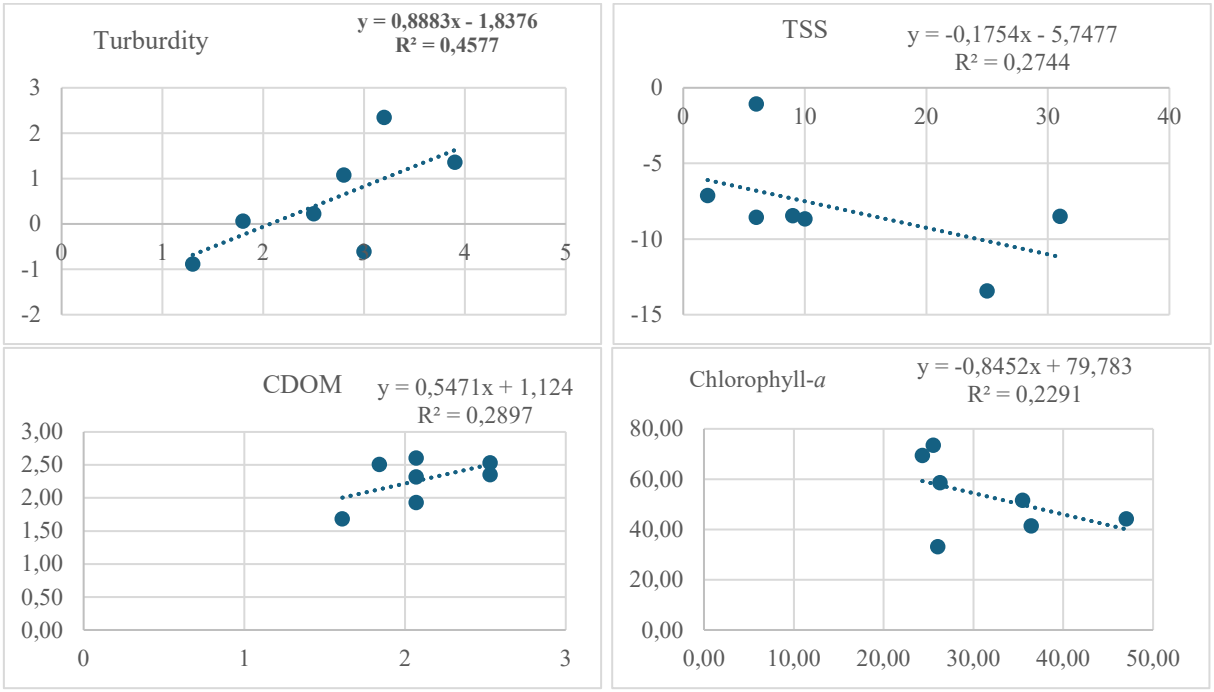


Figure 4.4(b): September S2 remote sensing model validation performance estimating FD_28Sep in-situ data

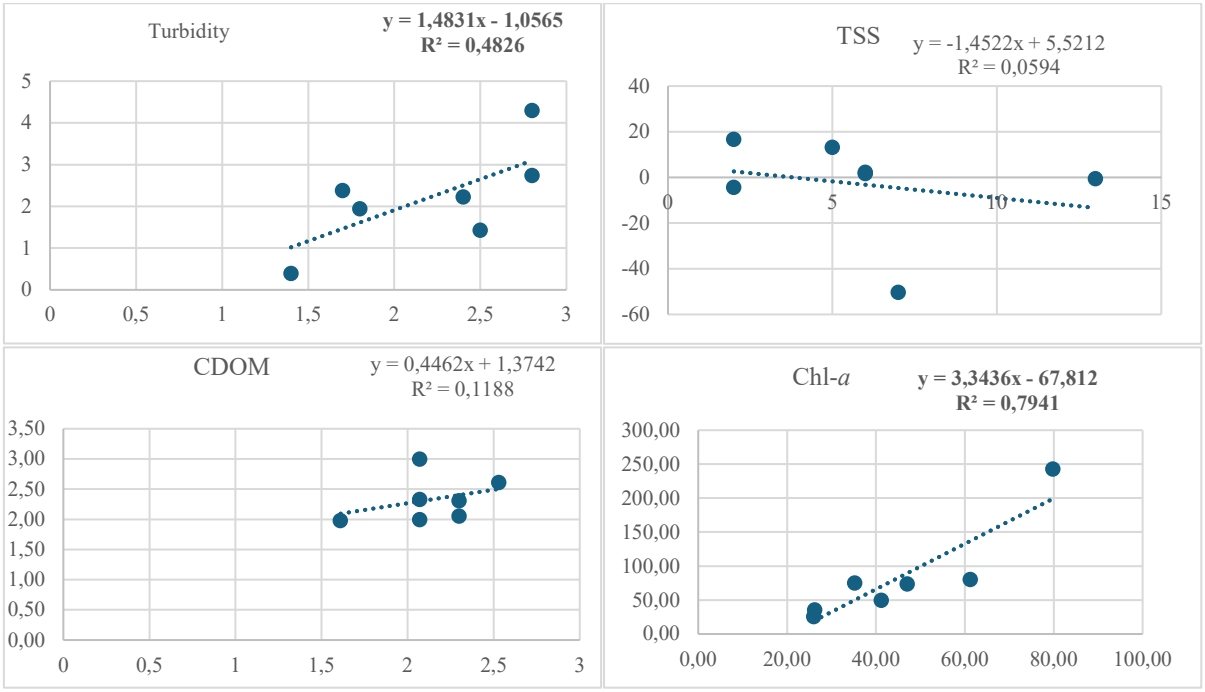


Figure 4.4(c): Mid-spring remote sensing model validation performance estimating FD_271 in-situ data

7.2 FD_21 October models validation results

Table 4.5: Best-performing validations results.

Field date	Image season	Accuracy metrics	Turbidity (NTU)	TSS (mg/L)	CDOM (1/m)	Chl-*a* (µg/L)
FD_294	Early spring	**R^2**	0.01	**0.62**	0.12	0.003
		RMSE	1.71	19.62	1.28	664.66
		MAE	1.09	14.16	1.01	631.40
		MAPE	-0.58	-1.99	-2.18	405.21
	Mid spring	**R^2**	0.00	0.02	0.17	0.07
		RMSE	111.34	27.70	2.33	958.90
		MAE	111.22	20.49	1.72	567.42
		MAPE	-4871.90	-223.56	1.79	-432.21

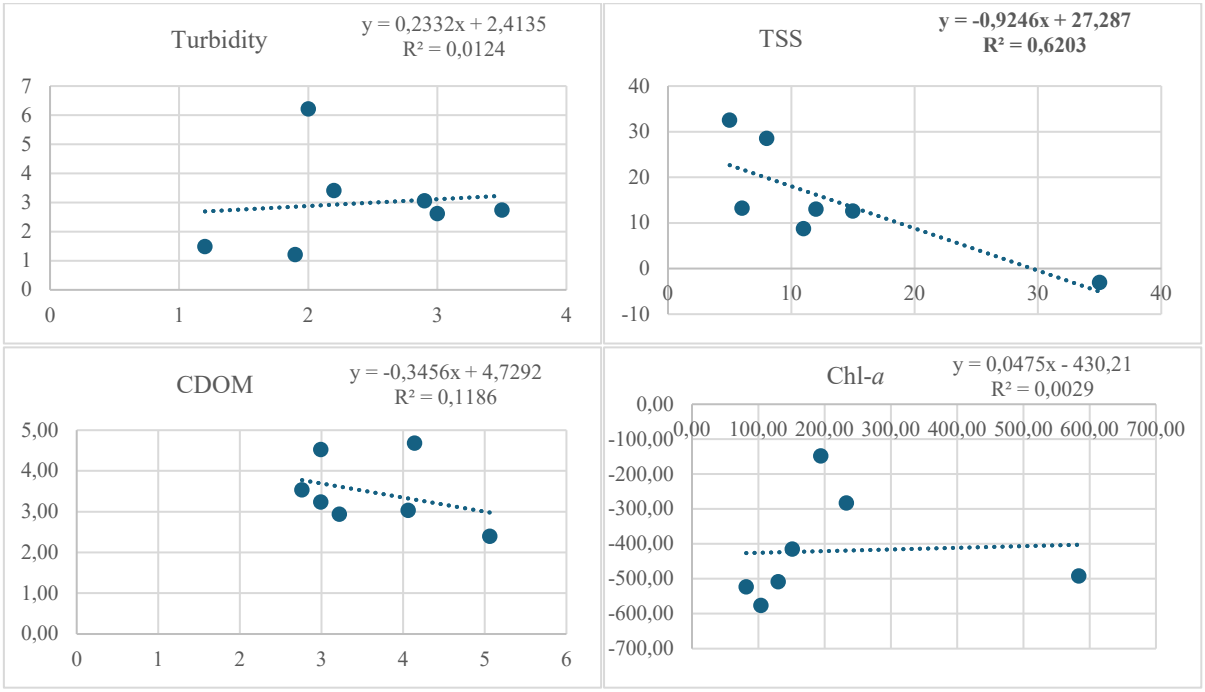


Figure 4.5(a): September remote sensing model validation performance estimating FD_21Oct in-situ data

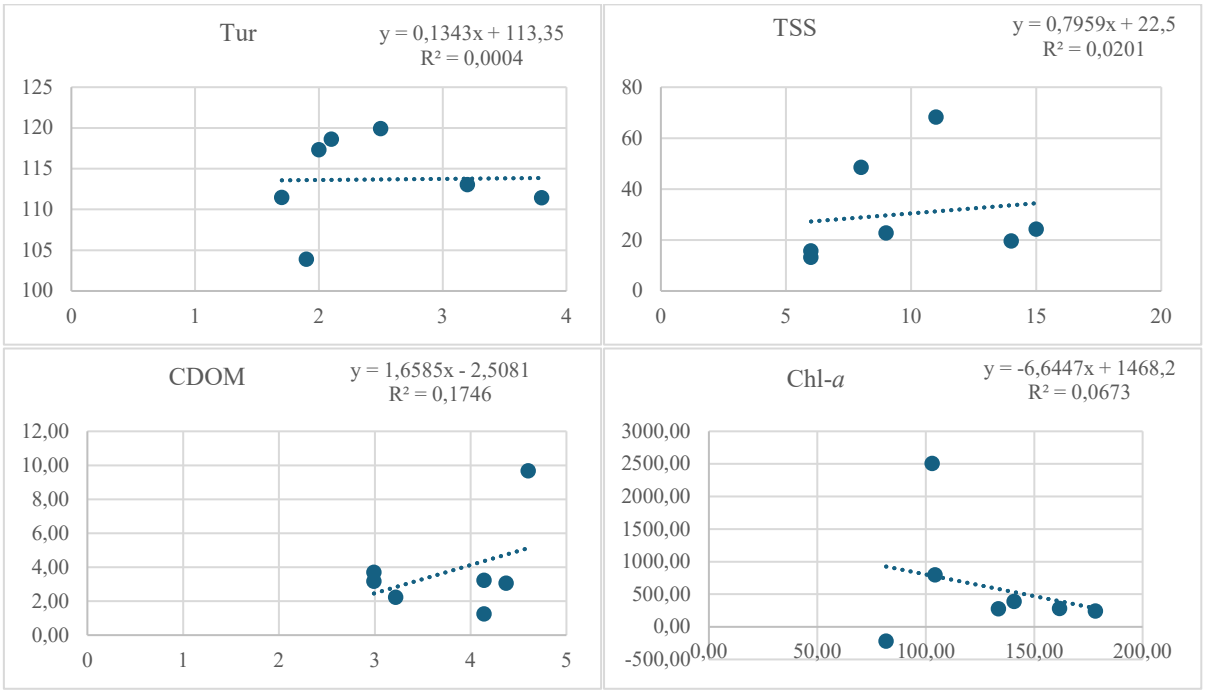


Figure 4.5(b): October S2 remote sensing model validation performance estimating FD_21Oct in-situ data

APPENDIX 3

CERTIFICATE OF EDITING

EDITOR'S LETTER

To whom it may concern

Dear Sir/Madam

This is to validate that I, **Zitha Innocent** have proofread and edited a dissertation for a degree of Master of Science in Environmental Sciences, in the Department of Geography and Environmental Sciences in the Faculty of Science, Engineering and Agriculture at the University of Venda, entitled: **Assessing the effects of sampling time on the accuracy of extracting water quality indicators through geospatial techniques: A case study of Rietvlei Dam**

By

Mokabolane Innocent Rabohale

I have further suggested several amendments that the student has undertaken to effect before the dissertation is finally submitted: spelling, grammar, structure, and coherence. This dissertation was inspected meticulously for consistency and correctness for register usage and citations. Should there be any inquiry, please do not hesitate to contact me.

Best Regards

21/02/2024

Contact No: 0715430998/ 015 962 8922
Email: Innocent.Zitha@univen.ac.za

BA (Communication & English), BA (Hons) English, MA (English), Lecturer (English) in the Science Foundation Programme, University of Venda