SMART MONITORING OF WATER SUPPLY TREATMENT PLANT PROCESSES USING REMOTE SENSING AND ARTIFICIAL NEURAL NETWORKS

BY

ALICE NUREEN ADHIAMBO OMONDI

A Thesis Submitted to the Department of Civil and Structural Engineering in Partial Fulfilment of the Requirements for the Award of the Degree of Master of Science in Water Engineering

Moi University

2023

DECLARATION

Declaration by the Candidate

I hereby declare that this thesis is my original work and has not been presented for a degree in any other university. No part of this thesis may be reproduced without the prior written permission of the author and/or Moi University.

Omondi Adhiambo N. Alice …… Date…….

(TEC/PGCS/02/15)

Declaration by the Supervisors

This thesis has been submitted for examination with our approval as university supervisors.

Prof. Yashon O. Ouma …… Date…….

Department of Civil and Structural Engineering,

Moi University, Eldoret, Kenya.

Dr. Njoroge S. Mburu …… Date…….

Department of Civil and Structural Engineering,

Moi University, Eldoret, Kenya.

Dr. Achisa C. Mecha …… Date…….

Department of Chemical and Process Engineering,

Moi University, Eldoret, Kenya.

DEDICATION

To my husband, my sons, my dear parents, and my sisters without whose unwavering love and support the completion of this work would not have been possible.

ACKNOWLEDGEMENT

I am grateful to the Almighty God for His guidance and sustenance throughout my studies. My heartfelt gratitude also goes to my supervisors Prof. Yashon O. Ouma, Dr. Njoroge S. Mburu, and Dr. Achisa Cleophas for their advice, support, and constructive suggestions throughout the study period. Special thanks to all staff in the Department of Civil and Structural Engineering, Moi University for their assistance throughout the study.

I am greatly indebted to the Building Capacity in Water Engineering for Addressing Sustainable Development Goals in East Africa (CAWESDEA) project led by Global Water Partnership Tanzania for the financial support which helped to facilitate part of the project. To the late Dr. Job Rotich Kosgey for introducing me to the CAWESDEA project and for the invaluable advice that helped enrich this research. Special thanks also goes to Eldoret Water and Sanitation Company (ELDOWAS) for the opportunity to do my research within their facility.

Lastly, I am grateful to my husband, sons, parents, and sisters for their constant reassurance and moral support.

ABSTRACT

The increasing global population continues to be a major threat to the world's ecological resources including lakes and rivers as people continue to clear the environment to create settlement spaces. Effective monitoring and management of a country's water resources is critical for sustainable water supply systems. Nonetheless, the conventional water quality monitoring (WQM) approach is laborious, time consuming, and costly. The main objective of this study was to evaluate the potential application of estimated water quality parameters (WQPs) from Landsat-8 OLI satellite data to determine the optimum coagulant dose for water treatment using Artificial Neural Network (ANN) models. The specific objectives of the study were: to estimate the concentration of Turbidity, total suspended solids (TSS), and Chlorophyll-*a* (Chl-*a)* from Landsat-8 OLI in correlation with *in situ* water quality data using empirical multivariate regression modelling (EMRM); to assess the spatial distribution and variability of the estimated and *in situ* WQPs for the selected period, and to use ANN modelling in predicting treated WQPs, and for determining the optimum coagulant dose for a water supply treatment plant. The study used satellite images and EMRM to estimate Chl-*a*, TSS, and Turbidity concentrations at different points in a water supply reservoir using same season data. Ordinary Kriging was used in the development of spatial maps showing the distribution and variability of the Landsat-predicted and *in situ* WQPs. The extracted spectral reflectance values from satellite images were then used as input for the first ANN model to predict treated WQPs, and in the second ANN model to predict the optimum coagulant dose required for water treatment. The results of the study show that, for all the samples, Turbidity, TSS, and Chl-*a* were estimated with R^2 values of 0.76, 0.81, and 0.81, respectively. The ANN model 1 for the prediction of treated WQPs had dependable accuracy with R^2 values of 0.99, 1.00, and 0.87 in predicting Turbidity, TSS, and Chl-*a*, respectively. Respectively, the *in situ* turbidity, TSS, and Chl-*a* for data collected in November 2020, December 2020 and January 2021 was (7.38 NTU, 7.08 NTU, and 8.62 NTU), (271.15 mg/L, 281.42 mg/L and 281.17 mg/L), and (37.17 mg/L, 50.86 mg/L and 44.75 mg/L) against a Landsat-estimated turbidity, TSS, and Chl-*a* of (7.44 NTU, 6.25 NTU, and 7.99 NTU), (268.17 mg/L, 279.89 mg/L 285.07 mg/L), (37.44 mg/L, 52.35 mg/L and 49.73 mg/L) for the specific months respectively. On the other hand, the ANN model 2 also had a high accuracy in predicting the optimum coagulant dose an R^2 value of 0.99. The actual coagulant dose for all the sampling days was 40 mg/l against the second ANN model's optimum coagulant dose of 39.95 mg/l for all the sampling days. Based on the results, the study concluded that satellite data products can be used for the retrieval of reservoir WQPs with reasonable accuracy. Furthermore, the ANN models also highlight the possibility of using extracted spectral reflectance values for water quality predictions, and optimizing water treatment plant operations. Since the study used same season data, it did not account for the temporal variability in WQPs. It is recommended that the model accuracies, and dependability be improved by using fairly extensive datasets collected at different seasons of the year. The concept will also increase the likelihood of using the models for water quality predictions in other reservoirs within, and outside the catchment.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACRONYMS AND ABBREVIATIONS

ANN	Artificial Neural Network
CFNN	Cascade Feedforward Neural Network
Chl-a	Chlorophyll-a
DNs	Digital Numbers
DOS	Dark Object Subtraction
EMRM	Empirical Multivariate Regression Modelling
FFBP	Feed Forward Back Propagation
FFNN	Feed Forward Neural Network
FLAASH	Fast Line-of-sight Atmospheric Analysis of Hypercubes
GRNN	Generalized Regression Neural Network
IDW	Inverse Distance Weighting
MLP	Multi-Layer Perceptron
MSI	Multispectral Instrument
NTU	Nephelometric Turbidity Unit
OLI	Operational Land Imager
RBFNN	Radial Basis Function Neural Network
Rrs	Remote Sensing Reflectance
ROI	Region of Interest
TIRS	Thermal Infrared Sensor
TM	Thematic Mapper
TOA	Top of Atmosphere
TSS	Total Suspended Solids
USGS	United States Geological Survey
WQM	Water Quality Monitoring

WQPs	Water Quality Parameters
WQIs	Water Quality Indices

CHAPTER 1: INTRODUCTION

1.1 Background Information

The increasing global population continues to be a major threat to the world's ecological resources including lakes and rivers as people continue to clear the environment to create settlement spaces (Black, 2016). Continuous watershed monitoring is crucial for the sustainable use of the world's freshwater resources (Emelko et al., 2011). Furthermore, there is also a need for water quality assessment at the source of the specific water treatment plants since this influences the type and cost of chemicals used to treat the water (Li et al., 2021).

Eldoret town in Kenya is one of the country's fastest growing towns with an estimated urban population of 475,716 based on the 2019 census (KNBS, 2019). The town relies on Chebara Water Treatment Works, Sosiani Water Treatment Works, Kapsoya Water Treatment Works, and Ellegerini Water Treatment works with a production capacity of 28,000 m^3/day, 14,959 m^3/day, and 7,500 m^3/day respectively (Mwaniki, 2018). According to Eldoret Water and Sanitation Company (ELDOWAS) the demand for water to Eldoret Municipality is estimated at 60,000 m^3/day, against a production of 36,400 m^3/day (Kimutai et al., 2018). According to Kibii et al (2021) mismanagement in the catchment is partly responsible for the huge disparity between demand and supply due to recent conversion of forested land into subsistence agriculture. This has led to flash floods, erosion and sedimentation which decrease the quality of surface water.

Water is a finite and vulnerable resource essential for sustaining life and development (Benson et al., 2020). The deteriorating water quality experienced in the catchment in recent years is mainly attributed to climate change, increasing population, and poor agricultural practices. According to a report by the World Health Organization (2017) an increase in turbidity necessitates higher disinfectant doses or more contact time for

the desired disinfection efficiency in drinking water treatment to be realized. Population growth has also increased the demand for water among diverse users with negative impacts on water quality in the water supply reservoir.

The conventional water quality monitoring (WQM) methodology involves *in situ* sampling followed by a physical, chemical, and biological analysis of samples in the laboratory (Gholizadeh et al., 2016). However, the approach is laborious, and cannot be used to effectively assess the spatial-temporal variations of the water quality indices (Gholizadeh et al., 2016). Moreover, dependence on traditional methods in addition to geographical restrictions, and inadequate regional spatial-temporal data also limits WQM and management of large water bodies (Duan et al., 2013).

Remote sensing techniques have continuously advanced over the decades and found wider applications in WQM (Papenfus et al., 2020). The techniques provide an avenue for WQM over larger temporal and spatial scales which help to understand and determine the water quality changes in line with the increasing water demands, high turbidity, and algal blooms (Papenfus et al., 2020). The method can also be used for water quality parameters (WQPs) estimations at different locations which are essential for monitoring and assessment of water quality at these locations. Artificial Neural Networks (ANNs) are useful in this regard for monitoring non-linear environmental processes since they are capable of modeling non-linear geophysical transfer functions (Chiappini et al., 2020). Remote sensing can be applied in combination with ANNs to overcome the challenges of *in situ* sampling and laboratory analysis by predicting WQPs at unsampled locations in the reservoir based on the extracted spectral reflectance values from remote sensing imaging. This concept was used to create a system for determining the optimum coagulant dose that is adaptive to real-time changes in raw water quality.

1.2 Problem Statement

The conventional method of WQM involves collection of samples followed by a physical, chemical, and biological analysis in the laboratory. The approach only gives an estimate of the parameters at specific points in time and space. In this regard, obtaining the spatial-temporal water quality variations for such extensive water bodies is virtually impossible using *in situ* sampling and analysis techniques (Gholizadeh et al., 2016).

Coagulation is an essential factor in portable water treatment that helps to maintain the turbidity of treated water within the acceptable range while helping to realize economic plant operation (Haghiri et al., 2018). The existing practice of determining coagulant dose based on the prevailing raw and settled water quality is not ideal since errors are bound to occur hours after dosing and these cannot be handled promptly by operators (Zhang & Stanley, 1999).

Previous researchers like Haghiri et al. (2018) and Kote and Wadkar (2019) have explored the possibility of developing a completely predictive water treatment plant model but the focus has been from the inlet of the water treatment works to the final point of treatment. However, there is need to incorporate information from the catchment since there are various catchment processes that contribute to deteriorating water quality. The WQM aspect can then be combined with predictive models for running the water treatment plants. Modelling helps in the prediction of treated WQPs based on process inputs mainly the optically active chlorophyll-*a* (Chl-*a*) and suspended particulate matter (SPM) that can be derived from remote sensing data. This helps to stabilize the treatment plant operation through a real time assessment of the optimal coagulant dosage.

1.3 Justification of the Study

The increasing demands for fresh water and environmental waste water discharge necessitates continuous surface WQM. Remote sensing is a fast, programmed, low-cost technology for monitoring aquatic environments (Garaba et al., 2015). Furthermore, remote sensing also allows for the identification of the optically active WQPs depending on their interaction with light and the ensuing energy change of the reflected incident radiation from the water surface (Garaba et al., 2015). Thus, remote sensing can avail a means for routine and accurate monitoring in line with sustainable water resource management.

The extracted WQPs from remote sensing can be used to inform ANN models for determining the optimal coagulant dose for water treatment. ANNs are one of the machine learning methods that can model the non-linear and multifaceted connections between WQPs since they can learn the temporal system dynamics with limited input data and to solve multifaceted problems efficiently (Ouma et al., 2020). The completely predictive water treatment plant model that begins from the catchment is crucial in helping to understand the efficiency of portable water treatment plants under adverse water quality conditions. The predictive model could also help to determine bottlenecks and ensure that the water treatment plant is operated at maximum efficiency. The machine learning methods will also reduce the sampling frequency while minimizing the costs associated with laboratory analysis.

1.4 Objectives

1.4.1 Main Objective

The purpose of this study was to evaluate the potential application of estimated WQPs from Landsat-8 OLI to determine the optimum coagulant dose for water treatment using Artificial Neural Network (ANN) models.

1.4.2 Specific Objectives

i. To evaluate the concentration of turbidity, TSS, and Chl-*a* from Landsat-8 OLI in correlation with *in situ* water quality data using EMRM.

ii. To determine the spatial distribution and variability of the Landsat-estimated and *in situ* WQPs for the selected period.

iii. To train, validate, test, and apply an ANN model to predict treated water quality parameters.

iv. To train, validate, test, and apply an ANN model to determine the optimum coagulant dose for a water supply treatment plant.

1.5 Research Questions

i. How do the extracted spectral reflectance values compare to the laboratory derived water quality values?

ii. How sufficient is the information gathered on the distribution of WQPs in understanding and deducing accurate water quality forecasts in the entire water body?

iii. What is the efficiency of ANN in predicting water quality parameters?

iv. Can ANN be applied to determine the optimum coagulant dose for a water supply treatment plant?

1.6 Scope and Limitations of the Study

The study was carried out at Two Rivers Dam in Uasin Gishu County. Sampling was done between November 2020 and January 2021 which is a dry period. Therefore, the models developed were calibrated and validated using same season data and they can only be applied for the Two River Dam system or in other reservoirs within the same geographical location for the specific season under investigation.

Sampling dates were prefixed based on the Landsat-8 OLI acquisition tool. Landsat-8 OLI has a 16-day repeat cycle and this means that the satellite acquires images of a specific area after every 16 days (USGS EROS Archive, 2018). The 16-day repeat cycle also limits intra-seasonal monitoring specifically in regions frequently covered by clouds since it could be difficult to relate the WQP characteristics with the optically active water constituents extracted from the satellite sensor (Amin et al., 2022). Two Rivers Dam is also highly silted and the increased water levels during the rainy season coupled with the lack of a motorized boat (a rowboat was used) made it difficult to collect samples during the rainy season when the water turbidities are expected to be higher.

CHAPTER 2: LITERATURE REVIEW

2.1 Introduction

This section presents a review of remote sensing techniques used in estimating WQPs in different water bodies. Water quality monitoring (WQM) of reservoirs is discussed in line with remote sensing techniques for extracting WQPs from satellite images, and discussions comparing extracted spectral reflectance values from satellite imagery and *in situ* WQPs. Literature on spatial distribution and variability of estimated and predicted WQPs is also presented. This section also highlights ANN techniques used in making water quality predictions and predicting the optimum coagulant dose for water treatment.

Inland water bodies are some of the valuable ecosystems that serve multiple functions including transport, hydropower generation, recreation, and providing irrigation water (Brooks et al., 2016). However, an increase in nutrients from urban development, forestry, and agriculture has greatly contributed to the increasing observed algal blooms in inland water bodies which are harmful to both human and animal life (Wurtsbaugh et al., 2019). Therefore, frequent monitoring of inland water bodies is thus a necessity. WQM using remote sensing has continued to gain popularity in WQM projects by reinforcing the abilities of decision makers and researchers to effectively monitor water bodies (Brooks et al., 2016). The approach is mainly used to measure the qualitative WQPs including colored dissolved organic matter, total suspended solids (TSS), turbidity, and chlorophyll-*a* (Chl-*a*) (Brooks et al., 2016).

The use of imaging to detect WQPs is based on the notion that water pollutants absorb and scatter the incident solar radiation thus enabling the correlation of water quality with the optical qualities of the water columns including color (Gholizadeh, 2016). Chl-*a* and suspended particulate matter (SPM) are optical WQPs that can be derived from

remote sensing data. The suspended particles are responsible for scattering light particles while Chl-*a* and dissolved or colored particulate matter is responsible for the absorption of light particles (Torbick et al., 2018). Algal blooms are directly related to the concentration of Chl-*a* in freshwaters which is essential for photosynthesis and it also influences the concentration of dissolved oxygen (Wernand et al., 2013). Turbidity is also an optical WQP that depends on the quantity of SPM in water. The more the SPM, the higher the reflectance value and consequently the turbidity of the water sample.

2.2 Water Quality Monitoring

WQM is critical in integrated watershed management in helping decision makers to understand, interpret, and use the information to support watershed management activities (Behmel et al., 2016). The WQM aspect entails sampling and analysis of water conditions and constituents including point pollutants such as metals, oils, and pesticides (Mustafa et al., 2021). Furthermore, WQM is also used to evaluate the concentration and effects of naturally occurring constituents such as nutrients, bacteria, and dissolved oxygen which can be affected by anthropogenic activities (Behmel et al., 2016). One of the main goals of WQM is to establish whether the water is meeting the core needs of its competing users. For instance, if the chemical concentration exceeds the required threshold, then, the water may not support beneficial uses such as fishing, and domestic uses for which it has been designated (Mustafa et al., 2021). Therefore, WQM is critical in comparing the concentration of the chemicals found in the streams, rivers, or reservoirs to the national standards to establish whether the water quality aligns with the standards required for different uses (Behmel et al., 2016). WQM is also critical in identifying the types, and sources of pollutants in order to create a framework for WQM in line with catchment preservation, and management practices (Behmel et

al., 2016). WQM can also help to assess the pollution trends over time, and the key WQPs to be monitored over time.

2.2.1 Water Quality Monitoring of Reservoirs

The point sampling approach for WQM in reservoirs is mainly limited by the variable spatial distribution of suspended sediment, dissolved material, and algae (Gholizadeh et al., 2016). Remote sensing overcomes this limitation by providing the spatial-temporal water quality distribution information (Gholizadeh et al., 2016). Thus, remote sensing can also be used to estimate the clarity or turbidity of the water. Even though there are several WQPs that can be monitored in a reservoir, the main constituents that affect water clarity are the colored segment of dissolved organic material, the algae cells, and non-algal particles that form the suspended matter (Mustafa, 2018). These optically active water constituents cause scattering and absorption of light through the water column thus affecting light penetration (Mustafa, 2018). Consequently, the optically active water constituents can be used to estimate other WQPs, and they also informed the decision to specifically focus on turbidity, TSS, and Chl-a.

According to Najafzadeh and Ghaemi (2021) water quality indices are crucial in describing the essential characteristics of water pollutants. Therefore, there is need for accurate predictions of water quality indices in order to gain insights on the patterns of pollutants in natural streams (Najafzadeh & Ghaemi, 2021). Furthermore, Najafzadeh et al. (2021) also notes that one of the most difficult issues in the in studies of water quality specifically, surface water resources, is getting an accurate estimate of WQIs. Even though there are numerous conventional methodologies for evaluating the WQIs, the limitations that exist among the traditional models have brought the need to employ Data-Driven Models in assessing the WQIs of natural streams.

Portable water is mainly required to be free from disease causing organisms and it should contain no chemicals that could adversely affect human health (WASREB, 2008). The water should also have low turbidity and it should contain no compounds that could impart an unpleasant taste or smell (WASREB, 2008). Generally, surface water sources must be adequately treated for them to meet these requirements.

2.2.2 Turbidity

Turbidity is an optical water characteristic that defines the relative liquid clarity by measuring light quantity either scattered or absorbed by the water column when light shines through it (Pier & Mateo, 2020). The particles also provide attachment sites for bacteria and other metals and this means that turbidity measurements are an indicator of potential water pollution (Poonam et al., 2013). Thus, high turbidity creates conducive environment for the regrowth of pathogens in water bodies and this can lead to an outbreak of waterborne diseases (Farrell et al., 2018). Turbidity is a significant parameter in portable water treatment since it facilitates the growth of pathogenic microorganisms (Zhang et al., 2021). The particulate matter can shelter pathogens from disinfection processes and also serve as food sources for the microorganisms (Farrell et al., 2018). For drinking water treatment plants, rapid turbidity fluctuations impact negatively on the operational performance of the plant due to a lag in process parameter adjustments (Zhang et al., 2021). Thus, safe drinking water can only be guaranteed if these risks are managed. The turbidity measurements also form the basis of the jar test experiments which are done to determine the coagulant dose needed to bring the turbidity level in water treatment to an acceptable level. The turbidity requirement for drinking water is less than 5 NTU (WASREB, 2008).

2.2.3 Total Suspended Solids

The total suspended solids (TSS) denote water particles with a size in excess of two microns (Lintern et al., 2018). A surge in TSS levels in surface waters result from different factors including erosion of river banks and streams which result in the suspension of soil and other particles across the water (Adjovu et al., 2023). Human activities like farming and fishing also contribute to a high TSS through dissolved pollutants like pathogens and heavy metals that attach to suspended particles in water or through the re-suspension of settled particles (Lintern et al., 2018). The algae also contribute to the TSS since the death of these organisms releases organic matter into the water thus adding on to the bulk of suspended matter while reducing the oxygen levels in the water (Adjovu et al., 2023).

A high TSS in water reduces the dissolved oxygen concentration and increases the water temperature (Rahmanian et al., 2015). The concept reduces the survival rate of organisms living in the water (Rahmanian et al., 2015). TSS also limits the penetration of sunlight in the water body which in turn halts the process of photosynthesis and reduces the survival rate of aquatic plants (Rahmanian et al., 2016). TSS also has detrimental effects on human health. For instance, the consumption of algae and bacteria may cause gastrointestinal issues while heavy metals may cause poisoning and even death (WASREB, 2008). TSS in treated water may cause corrosion and deterioration of pipes, plumbing, and fittings used for water supply (WASREB, 2008). As a result, a lot of money is wasted in replacement of plumbing and fixtures used for water supply and other appliances in the home. The guidelines for water quality and effluent monitoring state that drinking water should have no suspended matter (WASREB, 2008).

2.2.4 Chlorophyll-*a*

Chlorophyll-a (Chl-*a*) is the main photosynthetic pigment in algae and is largely used as a proxy algae biomass indicator to assess the trophic status of inland water bodies, and the extent and severity of algal blooms (Liu & Georgakakos, 2021). High levels of Chl-*a* in water mainly results from eutrophication which leads to excessive growth of algae (Liu & Georgakakos, 2021). The negative implications of Chl-*a* in drinking water manifests through higher levels of trihalomethane precursors, increased levels of cyanotoxins, taste and order problems, and increased levels of turbidity at the water source (Watson et al., 2015). The excess nutrients in the water bodies create conducive environment for cyanobacteria which are highly toxic and potentially harmful to both humans and animals to thrive (KDHE, 2011). The concentration of Chl-*a* provides a good assessment of the algal activities or primary production a water body (Knight, 2017). In essence, algal blooms are often linked to increased total nitrogen and/or total phosphorous in the water body (Knight, 2017). The increased frequency and severity of algal blooms increases the turbidity of the inland water body which in turn produces harmful toxins to both animals and humans while also lowering the levels of dissolved oxygen (Liu & Georgakakos, 2021).

2.3 Estimating the Concentration of WQPs from Landsat-8 OLI in Correlation with *in situ* Water Quality Data using EMRM

To correlate extracted spectral reflectance values from Landsat images and laboratory derived water quality data of Chl-*a* (Nalban Lake of East Kolkota Wetland, India), Patra et al. (2016) used Landsat Operational Land Imager (OLI) images and *in situ* water quality data to estimate the concentration of Chl-*a*. Pearson correlation analysis was then done between *in situ* Chl-*a* concentrations, and the possible OLI bands, and band ratios from the study points to determine the most suitable band ratio for estimating

Chl-*a* (Patra et al., 2016). The band ratio of OLI5/OLI4 was found to have the highest correlation coefficient with an R^2 value of 0.85 (Patra et al., 2016). The prediction model was then developed by applying regression analysis between the band ratio OLI5/OLI4 and Chl-a concentration of the study points (Patra et al., 2016). The reflectance ratios of the validation points were given as input on the prediction model and the model output was considered as predicted Chl-a values of the validation points (Patra et al, 2016). The validation was used to check the efficiency of the prediction model (Patra et al., 2016). A high correlation with an R^2 value of 0.78 was established between laboratory-derived and model-fitted Chl-*a* values (Patra et al., 2016).

González-Márquez et al. (2018) used Landsat-8 data together with empirical models to determine the spatial-temporal variations of WQPs including turbidity, TSS, and electrical conductivity in Playa Colorada Bay in Mexico. Multiple regression analysis was then done between *in situ* WQPs and the reflectance of the pixels corresponding to the sampling stations. The developed water quality models achieved R^2 values that ranged between 0.64 and 0.96 and this showed the feasibility of the Landsat-8 images in the estimation of the WQPs in the entire bay. The effectiveness of Landsat-8 in the assessment of reservoir water quality was also established by Bonansea et al. (2019) through the estimation and mapping of secchi disk transparency in Cassaffousth Reservoir in Cordoba, Argentina. Ground observations, and a dataset of four Landsat 8 and four Sentinel-2A images were used to create and validate models to estimate secchi disk transparency in the reservoir (Bonansea et al., 2019). The selected algorithms were used to obtain graphic representations of water clarity. Landsat-8 OLI estimated the secchi disk transparency with an R^2 value of 0.9 (Bonansea et al., 2019). The study showed the importance of Landsat-8 data in spatio-temporal water quality analysis

particularly in developing nations where conventional WQM programs are limited (Bonansea et al, 2019).

Pizani et al. (2020) paralleled the performance of Landsat-8 OLI and Sentinel-2 MultiSpectral Instrument (MSI) in developing multiple regression models for the optically active and optically non-active WQPs in a hydroelectric reservoir in Brazil. The physical and chemical WQPs were measured *in situ* using sensors, and also analyzed in laboratory (Pizani et al., 2020). The sampling date corresponded to the simultaneous overflight of Sentinel-2B, and Landsat-8 satellites which provided a means to perform a fair comparison of the two sensors (Pizani et al., 2020). Four optically active WQPs were considered: chlorophyll-a, Secchi disk depth, turbidity and temperature, and other six optically non-active parameters were also considered (Pizani et al., 2020). The multiple regression models used the spectral reflectance bands from both sensors (separately) as predictors, and the reflectance values were based on averaging kernels of 30 m and 90 m (Pizani et al., 2020). Both Landsat-8 OLI and Sentinel-2 MSI performed well in the estimation of optically active constituents specifically, turbidity, Chl-*a*, and secchi disk depth with R^2 values greater than 0.6.

Meng et al. (2022) used the EMRM algorithm to estimate the concentration of Chl-*a*, algae density, and turbidity using Landsat 8-9 OLI and Sentinel-2 MSI images in Shanmei Reservoir in Fazhou City. This study collected and sorted the water quality data measured at the site in 2020 to 2022, and Landsat 8-9 OLI and Sentinel-2 MSI images, simulated the chlorophyll-a (Chl-a) concentration, algae density, and turbidity using empirical multivariate regression (Meng et al., 2022). The researchers also explored the relationship between different WQPs using correlation analysis and principal component analysis (Meng et al., 2022). Landsat performed better that Sentinel-2 MSI with R^2 values of 0.70, 0.81, and 0.80 for Chl-*a*, algal density, and

turbidity respectively. Similarly, Ouma et al. (2020) related the performance of Landsat-8 OLI and Sentinel-2A/ MSI in the retrieval of optically active WQPs including turbidity, TSS, and Chl-*a* in reservoirs based on an EMRM algorithmic approach. In the development of the EMRM, ten of the sampling point data were used in the regression modelling in model calibration and the remaining five sampling points that were not used in model development were used in the validation of the model (Ouma et al., 2020). Both sensors estimated the reservoir WQPs with R^2 values greater than 0.7 and this showed the validity with which the developed algorithms can retrieve WQPS from satellite imagery.

2.4 Assessment of Spatial Distribution and Variability of Estimated and Predicted WQPs

Bresciani et al. (2019) used Landsat-7, Landsat-8 OLI, and Sentinel-2 data from 2013 to 2018 on two freshwater dammed reservoirs named Mulargia and Aposelemis situated in Sardinia (Italy) and Crete (Greek) respectively. Physically based models were used on 400 cloud-free images to retrieve Chl-*a* concentration, turbidity, and secchi disk depth. The derived estimates were effectively validated using the lower Pearson correlation value *r* which gave a value of 0.88 for Chl-*a*. The multi-temporal analysis results showed a reduction of sechi disc depth due to an increase in turbidity in Mulargia or an increase in Chl-*a* in Aposelemis. Thus, both lakes were assigned mesotrophic conditions based on the satellite-derived trophic state index.

In another study, Keith et al. (2018) examined the possibility of using Landsat-8 OLI imagery to estimate and map the spatial Chl-*a* distribution in Jordan Lake, North Carolina. Using the provisional Land Surface Reflectance product and *in situ* Chl-*a* concentrations from drinking water monitoring programs in North Carolina, and Rhode Island, the study compared five established approaches for estimating chl-

a concentrations using spectral data (Keith et al., 2018). The relationship between individual OLI bands reflectance values and the *in situ* concentration of Chl-*a* was used to isolate bands sensitive to Chl-*a* (Keith et al., 2018). Optical-band based equations were developed and the relationship between the ratio-based spectral index and the *in situ* Chl-*a* from various points derived (Keith et al., 2018). Spatial maps showing the distribution of Chl-*a* were also developed (Keith et al., 2018). From the study it was established that there was a substantial link between the spectral reflectance values from Landsat images and Chl-*a* concentration in the lake (Keith et al., 2018).

Silveira et al. (2020) used different spatial interpolation techniques including inverse distance weighting and random forest to develop spatial distribution maps for *in situ* and Landsat-estimated water quality variables in two study areas. Two case studies in distinct water bodies were performed, each using different spatial resolution data from Sentinel-2 spectral images and unmanned aerial vehicles together with laboratory analysis data (Silveira et al., 2020). In consonance with the methodology, supervised machine learning algorithms were trained to predict the concentration of TSS and Chl-*a* (Silveira et al., 2020). The predictions were evaluated separately in both study areas, where both TSS and Chl-*a* models achieved R^2 values above 0.8 (Silveira et al., 2020). The spatial maps developed for each of the WQPs demonstrated that the spatial distribution was consistent with the expected results from *in situ* measurements. Based on this, it was concluded that it is possible to generate spatial distribution water quality maps using interpolation techniques, and this could help in generating additional data points for improved training of machine learning algorithms for predicting WQPs (Silveira et al., 2020).

According to Ouabo et al. (2020) suitable interpolation techniques can be applied on correctly sampled data in order to make inferences on the distribution and variability of

the WQPs at the unsampled locations. However, there is a need to have detailed information on the distribution of WQPs in the reservoir so as to make precise water quality predictions for a specific point in the reservoir (Ouabo et al., 2020). In this study, ordinary Kriging and inverse distance weighting were used for spatial analysis and surface mapping and ordinary Kriging performed better than inverse distance weighting method (Ouabo et al., 2020). Ouma et al. (2020) developed spatial distribution maps of both *in situ* and Landsat-8 OLI estimated WQPs including turbidity, TSS, and Chl-*a* for Chebara Dam in Kenya using ordinary Kriging. It was observed that the distribution trend for both Landsat-8 OLI and *in situ* measurements were closely correlated since most of the high and low concentration regions coincided with slight variations in concentration at some spatial locations (Ouma et al., 2020).

2.5 Remote Sensing and ANN Model Approach for Predicting Water Quality Parameters

ANN is a computing system that mimics the human brain in analyzing and processing information in order to solve complex problems (Mohamed, 2019). Artificial neural networks are made up of a set of simple elements; the artificial neurons are motivated by the biological nervous systems Gurney, 2018). The self-learning capabilities of the ANNs enable them to improve their performance as more data becomes available. According to Mohamed (2019) the performance accuracy of ANN models is dictated by the correctness of the data used and the model architecture selected. Additionally, modelling is an empiric approach and the model accuracy is influenced by the precision of the input data (Mohamed, 2019). The hidden layers in the ANNs are used to transform data and produce an output by applying weights to the layers (Jorgensen & Gromiec, 2016). Despite the complexity involved in developing ANN models, accurate correlations are guaranteed to a great extent.

Panda et al. (2004) applied ANNs to indirectly determine the concentration of Chl-*a* and SPM in Beaver Reservoir, Northwest Arkansas. An ANN-radial basis function was developed and spectral reflectance values from Landsat Thematic Mapper (TM) used as input with Chl-*a* or suspended matter concentration as output (Panda et al., 2004). Chl-*a* and suspended matter concentrations were predicted with a coefficient of determination R^2 of 0.55 and 0.90 respectively for the actual and predicted values respectively (Panda et al., 2004).

Ribeiro et al. (2008) studied the feasibility of WQM in large reservoirs using remote sensing and ANNs as an indirect method of estimating Chl-*a* concentrations in Tucurui Reservoir. An ANN, radial basis function model was developed to estimate Chl-*a* concentration. The spectral data from the satellite images was used as input to the neural network while Chl-*a* concentration was used as the output (Ribeiro et al., 2008). To train and validate the model the researchers used data from the years 1987, 1988, 1995, 1999, 2000 and 2004 (Ribeiro et al., 2008). The model estimated Chl-*a* concentration with a correlation coefficient of 0.92 implying that the model is useful for Chl-*a* predictions (Ribeiro et al., 2008).

Kennedy et al. (2015) developed four ANN models at the Akron Water Treatment Plant in Ohio, USA to predict treated water turbidity and dissolved organic material based on variable raw WQP values and coagulant doses. ANN models were built using operational data to predict each of the fluorescence components and turbidity after coagulation based on variable raw water quality and chemical doses (Kennedy et al., 2015). The final turbidity was predicted with a correlation coefficient of 0.91 (Kennedy et al., 2015). The results suggested that the models are useful in predicting treated water turbidity as a function of the raw water quality and coagulant dose (Kennedy et al., 2015).

Setshedi et al. (2021) used an ANN technique to develop a model for predicting WQPs using ANN-radial basis function (RBF) and ANN-Multi-Layer Perceptron (MLP) using data obtained from three district municipalities in South Africa. Two input combination models were trained, verified, and tested for their predictive performance ability, and their physicochemical prediction accuracy was compared by using each model's observed data with the predicted data (Setshedi et al., 2021). The MLP model achieved accuracies denoted by a mean square error of 39.07 and R^2 of 0.99 for the *in situ* and predicted water quality (Setshedi et al., 2021). Based on the accuracies achieved, it was recommended that the models be scaled up and used to make water quality predictions for waste water treatment plants (Setshedi et al., 2021).

2.6 Use of ANN Models to Determine Optimal Coagulant Dose for Water Treatment

Sengul and Gormez (2013) used ANN models to predict treated WQPs and optimal coagulant dose in using operation data from a water treatment plant in Istanbul, Turkey. ANN was used to estimate non-linear relationships between inputs and outputs in the dataset. Several architectures with one hidden layer and hidden neurons were built to obtain the right and proper ANN model. The multilayer ANN model showed that the model is efficient in predicting treated WQPs and the optimal coagulant does that is adaptable to real-time raw water quality changes. Furthermore, ANN models can also be used as a fast and efficient method of optimizing coagulant dosage.

Haghiri et al. (2018) developed an MLP-ANN with one hidden layer and used it to model jar test experiments for determining the optimal coagulant dose in drinking water treatment processes for a water treatment plant located in Ardabil Province in Iran. To evaluate the performance of the model, the mean squared error (MSE) and correlation coefficient (R^2) parameters were used (Haghiri et al., 2018). The model accomplished

reasonable accuracy in predicting the optimal coagulant dose with R^2 value of 0.95 and MSE of 0.12 mg/l (Haghiri et al., 2018). The concept established that MLP-ANN models have a reasonable accuracy in estimating WQPs and the optimal coagulant dose (Haghiri et al., 2018).

In order to overcome the limitations of the jar test experiment and to enable water treatment plant operators to save time and money, Baouab and Cherif (2018) developed ANN models for predicting optimal coagulant dose using data from three water treatment plants. ANN technique was applied in the study to large databases of three treatment plants with different processes in order to build models to predict the optimal dose of coagulant (Baouab & Cherif, 2018). Pre-modeling techniques, like data scaling and training database choice, were used to guarantee models with the lowest errors (Baouab & Cherif, 2018). Two models were then selected, with turbidity, conductivity, and pH as inputs for both raw, and treated water. The first model, L45-MOD, was specific to raw water with less than 45.5 NTU turbidity, or else the second model ATP-MOD would be adopted (Baouab & Cherif, 2018). The two models developed performed great when used on several databases with R^2 greater than 0.8 and a mean absolute error greater than $5.4 g/m^3$ (Baouab & Cherif, 2018). The study also highlighted the possibility of extrapolating and adopting such models for use in treatment plants that rely on coagulation (Baouab & Cherif, 2018).

Kote and Wadkar (2019) also developed models to estimate optimum coagulant and chlorine doses for Maharashtra Water Treatment Plant in India. Separate models for coagulant and chlorine doses were developed and the possibility of applying feed-forward neural network (FFNN), radial basis function neural network (RBFNN), cascade feed forward neural network (CFNN), and generalized regression neural network (GRNN) explored (Kote & Wakdar, 2019). For modeling, daily water quality

data of the last four years was collected from the plant laboratory of the water treatment plant in Maharashtra, India (Kote & Wakdar, 2019). To improve the performance of the models, the models were developed by varying input variables, hidden nodes, training functions, spread factor, and epochs (Kote & Wakdar, 2019). The best models were then selected based on the comparison of performance measures (Kote & Wakdar, 2019). The best performing coagulant dose model was the CFNN model with Bayesian regularization training function with an R^2 value of 0.947 (Kote & Wakdar, 2019). Similarly, the best performing chlorine dose model was the RBFNN model with an R^2 value of 0.999 (Kote & Wakdar, 2019).

2.7 Gaps in Knowledge to be Filled by the Study

This section mainly reviewed literature on estimation of WQPs from Landsat-8 images in correlation with *in situ* water quality data in addition to literature on spatial distribution and variability of estimated and *in situ* WQPs. The use of ANNs in water quality predictions and the estimation of optimum coagulant dose for water treatment was also highlighted.

Based on the literature review, the gap in knowledge stems from the need to have a completely predictive water treatment plant model that begins from the reservoir to the final point of treatment. Previous researchers like Haghiri et al. (2018) and Kote &Wadkar (2019) have explored this possibility with the focus being from the inlet of the water treatment works to the final point of treatment. The information from the catchment is what is used to inform the treatment plant processes in terms of the choice of chemicals mainly for the coagulation processes. Thus, having a model that starts from the reservoir may prove to be an effective tool to determine bottlenecks and make quick decisions to maximize the efficiency of treatment.

Previous researchers have also not explored the possibility of combining WQM with the predictive models for running the water treatment plants. The optimum coagulant dose to be used for treatment has mainly been determined based on the characteristics of the input water to the treatment plants. There are various processes in the catchment that contribute to deteriorating water quality. Thus, combining the monitoring aspect using remote sensing will help to give a better estimate of the optimum coagulant dose based on the changes in the catchment.

CHAPTER 3: MATERIALS AND METHODS

3.1 Area of Study

Two Rivers Dam Reservoir is situated in Uasin Gishu County, Kenya at a longitude of 35° 35' 14" and latitude of 0° 46' 88" (Figure 3.1).

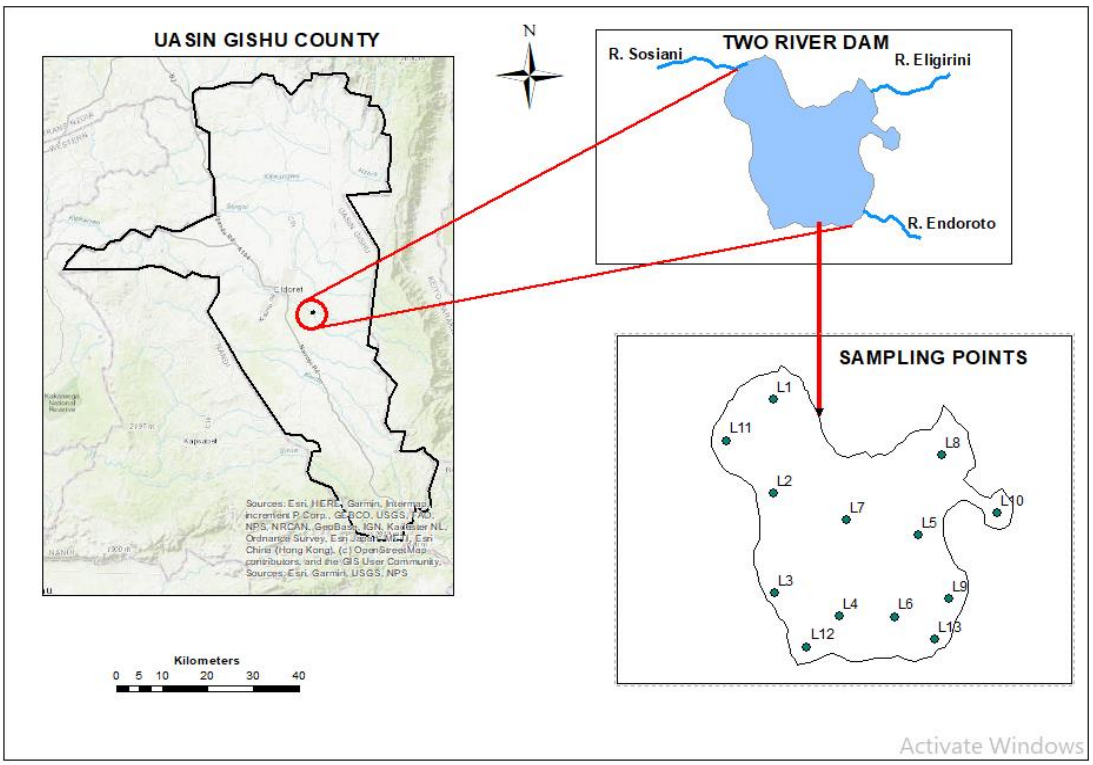


Figure 3.1: Two Rivers Dam Reservoir and selected sampling points

The water abstracted from the dam for treatment and subsequent distribution is 14,959 m^3/day. The water is treated at Sosiani Treatment Plant before it is supplied to Eldoret municipality and its environs. Increased agricultural activities and soil erosion have resulted in increased turbidity in the reservoir and an overall loss of water quality. There is an eminent presence of algal bloom especially during the dry season which necessitates pre-chlorination of raw water to oxidize the organic matter at the beginning of the treatment process. Generally, turbidity ranges between 50-200 NTU thus necessitating the use of poly aluminium chloride as a coagulant before the subsequent

processes of sedimentation, sand filtration, and chlorination. The presence of algal blooms is an indication of increased nutrient transport in the catchment.

3.1.1 Water Sampling and Testing

To monitor the water quality of Two Rivers Dam Reservoir, thirteen (13) stratified and coordinated sampling stations were set up at distances approximately 150 meters apart (Figure 3.2). The selection of the sampling points was influenced by the physical characteristics of the area with more points being located in areas where the influent streams (R. Endoroto and R. Elligirini) enter the dam, and regions close to the point where R. Sosiani exits the dam to capture the changing influent and runoff characteristics (Wilde et al., 2005). At the mid-section of the reservoir few points were selected since the flow is uniform, there is a stable bottom contour, and the water constituents are well mixed along the cross section (Wilde et al., 2005). More points were also located close to the edges where the water quality variability is highly likely to be high (Wilde et al., 2005). A mesh grid was drawn over the reservoir and based on a systematic sampling approach and the above guiding criteria, the 13 sampling points were selected.

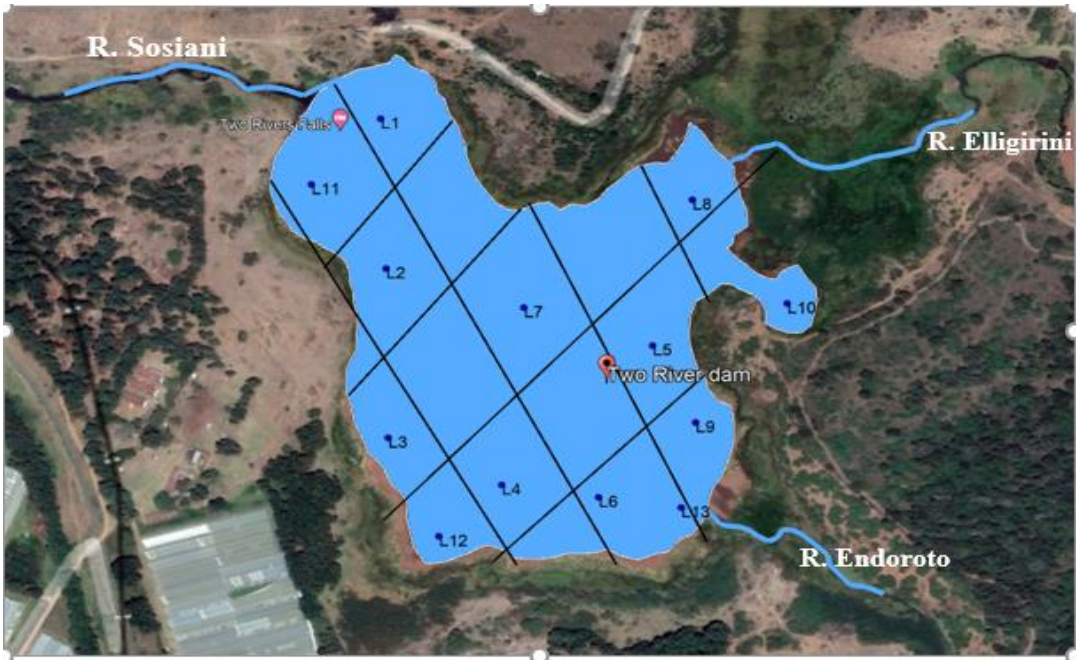


Figure 3.2: Grid system used for the selection of sampling points

Water samples (1500 ml) were then collected at 0.5 meter depths and the concentration of turbidity, TSS, and Chl-*a* determined following the standard laboratory protocols. The duplicate samples were collected at each of the 13 sampling points for each of the pre-determined sample collection dates. The duplicate samples for determining the concentration of turbidity, TSS, and Chl-*a* were collected using dark bottles at each of the 13 sampling points. The sample containers were completely filled and capped to prevent aeration, and each of these containers was then labeled and transported to the laboratory using a cooler box. The containers were then stored in a dark place prior to testing. Since Chl-*a* is a key requirement for photosynthesis, the water samples were collected in the photic zone which is close to the surface of the reservoir because, for stratified lakes or reservoirs, the epilimnion or the top layer mostly coincides with the photic zone of the water body (KDHE, 2011). Furthermore, for case I waters such as Two Rivers Dam the optical properties mainly detected by Landsat-8 OLI are mainly determined by phytoplankton and related colored dissolved organic matter, and detritus

degradation products (Gholizadeh et al., 2016). Thus, for meaningful inference between laboratory and satellite-derived WQPs (specifically, turbidity and TSS) to be made, sampling was done at 0.5 meter depth where the concentration of phytoplankton and related colored dissolved organic matter were likely to be highly concentrated.

Turbidity measurements were done in the lab using LaMotte turbidity meter (Model 2008 Code 1790, USA). The water sample was placed in a cylindrical vial to a volume of 10 ml as marked on the vial. A second vial was then filled to the 10 ml mark using distilled water which was the standard solution. The turbidity meter was then blanked using the standard solution after which the sample turbidity measurement was then taken.

TSS was determined by gravimetric method (APHA). Dry and clean filter papers were first weighed to obtain their masses. The sample solution was filtered using sintered glass lined with the 45µm filter paper by use of a filter pump. The filter papers were then dried in a desiccator and the new weights recorded. The concentration of TSS was then obtained as the difference between the weight of the clean, dry filter paper and the used filter paper divided by the sample volume.

$$TSS\ \left(\frac{mg}{L}\right) = \frac{(A-B)*1000}{Sample\ Volume(ml)} \quad (1)$$

Where:

A: Dry weight of residue and filter paper (mg)

B: Dry weight of the filter (mg)

Chl-*a* was determined by the spectrophotometric method where the optical density of the extracted Chl-*a* was measured at four wavelengths (750, 663, 645, and 630nm) and the resulting concentration determined based on the SCOR-UNESCO's equations (SCOR-UNESCO, 1966). The chlorophyll pigments from the water sample were

extracted using acetone by filtering the water sample through a 45µm filter. The filter was then placed in a test tube and 5 ml of acetone at 90% concentration added to the tube. The samples were then stored in a refrigerator for 24 hours after which 3 ml of the filtrate was drawn and placed in a spectrophotometer cell from which the transmittance was obtained and used to compute the chlorophyll-a concentration based on equation 2.

$$Chl - a = \frac{26.7*(664b-665a)*Ve}{Vs*L} \quad (2)$$

Where:

Ve: Volume of extract (L)

Vs: Volume of sample used (m^3)

L: Light path of the cuvette

A GPS receiver was used to locate the sampling points during sample collection thus enabling meaningful seasonal inferences to be made for the specific locations. The sampling locations were located using a GPS receiver.

The timing for *in situ* sampling coincided with the satellite overpass schedule, specifically, the satellite image acquisition dates to within ±1 day for best calibration results to be realized. In essence, the Landsat acquisition tool was used to view the paths scheduled for image acquisition on any day from which the sampling dates were selected.

The coagulant dose for water treatment for the specific sampling days was determined at the water treatment plants based on jar test experiments. Aluminum sulfate was the main coagulant used for the jar test experiments, and water treatment. The stock solution for the coagulant was made by mixing 20 grams of alum in 20 liters of water.

A constant dose of 40 liters of the stock solution was used in water treatment for each of the three days that the water samples were collected.

3.2 Retrieval of Water Quality Parameters from Landsat-8 OLI and Correlation Analysis

The Landsat-8 Operational Land Imager (OLI) and thermal infrared sensor (TIRS) is a push broom sensor with nine bands. That is, band 1-7 and band 9 at 30 m spatial resolution and band 8, which is a panchromatic band, at 15 m spatial resolution. It also has two Thermal Infrared bands with a spatial resolution of 100 m. The satellite has a 16-day repeat cycle and a swath width of 185 km (USGS EROS Archive, 2018).

The Landsat images used for the study were acquired between November 2020 and January 2021 in line with the satellite image acquisition schedules (Appendix A). The images were downloaded from United States Geological Survey (USGS) earth explorer (https://earthexplorer.usgs.gov/).

The procedures for the retrievals of WQPs from Landsat-8 OLI are outlined in the steps below.

a) **Determination of Radiance and Reflectance**

   The Fast Line-of-sight Atmospheric Analysis of Hypercubes (FLAASH) model was used to process the Landsat-8 OLI images where radiometric calibration was done to convert the digital numbers (DNs) to top of atmosphere (TOA) radiance (Appendix B). For a given wavelength, the TOA radiance is obtained from the following equation.

$$L_\lambda = G_\lambda * DN + O_\lambda \quad (3)$$

Equation (1) can also be written as:

$$L_\lambda = \left\{\frac{L_{max\lambda} - L_{min\lambda}}{DN_{max}}\right\} * DN + L_{min\lambda} \quad (4)$$

WhereL_λ is the TOA radiance given in units of ($Wm^{-2}Sr^{-1}\,\mu m^{-1}$),$G_\lambda$is the gain or multiplicative scaling factor ($Wm^{-2}Sr^{-1}\,\mu m^{-1}$)$/DN$ and $O_\lambda$is the offset or the rescaled bias ($Wm^{-2}Sr^{-1}\,\mu m^{-1}$), $L_{min\lambda}$is the spectral radiance scaled to DN_{min}(Wm^{-2}Sr^{-1}μm^{-1}),$L_{max\lambda}$ is the spectral radiance scaled to DN_{max} , DN_{max} is the maximum DN and DN is the quantized calibrated pixel values in DN specific to a sample point.

Atmospheric correction was then done to convert TOA radiance to TOA reflectance after which Dark Object Subtraction (DOS) was done to convert TOA reflectance to surface reflectance (Ouma et al. 2018). The extraction of the region of interest was then done and the resulting images used in the ensuing processing stages. The surface reflectance values were then divided by π to convert them into remote sensing reflectance (Moses et al. 2012) (Appendix C).

b) Correlation of Reflectance with Laboratory Water Quality Parameters

The 13 sampling points were then overlaid on each of the image's regions of interest (ROIs) and the spectral profile for each specific point used to determine the TOA reflectance for each of the bands from band 1-7. To reduce the errors in sample site locations and geometric corrections, and to correlate the reflectance and WQPs, an average spectral reflectance of 3×3 pixel neighborhood configuration was used (Reddy, 1997). It was appropriate to use the 3 x 3 pixel window for each of the sampling points in this study since most of the sampling was done close to the edges of the reservoir where water quality

variability is high (Reddy, 1997). Thus, the 3 x 3 window could include the shallow water near the banks.

c) WQPs Estimation Using Empirical Regression Modeling

The study used the Empirical Multivariate Regression Modeling (EMRM) approach for multivariate correlations between the reflectance from the different sensor bands and the WQPs measured *in situ* as described in Ouma et al. (2018). In this case, the retrieval objects were turbidity, TSS, and Chl-*a* and the different band combinations used for retrieval. The EMRM algorithm is useful in predicting WQPs from spectral features of satellite images. According to Aiman et al. (2014), standard multiple regression makes it possible to establish how well the dependent variables (WQPs) can be predicted by each independent variable (the spectral reflectance values). The main water quality retrieval algorithms used are single band, band ratios, linear band combinations, and mixed-band combinations (Meng et al., 2022). However, regression models with spectral ratio are more robust and dependable compared to regression models with single bands (Aiman et al., 2014).

Field data was first correlated with remote sensed data from which the equations for water quality retrieval were developed. Regression analysis was then carried out to determine the relationship between *in situ* and satellite-derived water quality data. Eight (8) points were used to calibrate the empirical multivariate regression model algorithms (EMRM) and five (5) points used for algorithm validation. Table 3.1 presents the remote sensing band combination(s) considered for the EMRM analysis of Landsat-8 OLI data in the estimation of water quality parameters.

Table 3.1: RS bands and band combinations for estimation of water quality parameters

Type of Band Combination	Band Combination
Single Band	B1, B2, B3, B4, B5, B6, B7
Band Ratio	B2/B1, B3/B1, B1/B3, B4/B3, B1/B4, B4/B1, B3/B2, B4/B2
Linear Combination	B1+B2, B1+B3, B1+B4, B2+B3, B2+B4, B3+B4, B1+B2+B3, B2+B3+B4, B3+B4+B1, B4+B1+B2
Mixed Combination	(B1/B3)+B1,(B1/B4)+B1,(B1/B4)+B2,(B1/B3)+B2,(B1/B3)+B3, (B4/B1)+B4

The regression analysis was done by comparing the predicted and lab-measured water quality values and the equations with the highest R^2 values chosen (Ouma et al., 2018). The regression equations were developed based on data from eight (8) sampling locations and equations with the highest Coefficient of Determination (R^2) values selected and then tested on the five sampling points not used in model development to establish their accuracy in deriving the WQPs. The accuracy of the WQPs derived from the sensor-based empirical models was determined using R^2, Bias, and Normalized Root Mean Square Error (NRMSE) error estimators shown in Equations 3-5 below.

1. Coefficient of Determination (R^2)

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}} \quad (5)$$

Where:

n - Number of pairs of values

∑xy - Sum of the products of x and y values

∑x -Sum of x values

∑y - Sum of y values

$\sum x^2$ - Sum of the squared x values

$\sum y^2$ - Sum of the squared y values

2. Bias

$$Bias = \frac{1}{n}\sum_{i=1}^{n}\lfloor X_i - Y_i \rfloor \quad (6)$$

Where:

X_i - Estimated water quality for the i^{th} sample point

Y_i - Lab measured value for the i^{th} sample point

n- Total number of sample points

3. Normalized Root Mean Square Error (NRMSE)

$$NRMSE = \frac{RMSE}{y_{imax} - y_{imin}} \quad (7)$$

Where:

RMSE= Root Mean Square Error

$$RMSE = \sqrt[2]{\frac{\sum_{i=1}^{n}(x_i - y_i)^2}{n}} \quad (8)$$

x_i - estimated water quality for the i^{th} sample point

y_i - lab measured value for the i^{th} sample point

n - Total number of sample points

y_{max} – maximum estimated water quality value

y_{min} – minimum estimated water quality value

3.3 Spatial Mapping of Water Quality Parameters

Spatial mapping of the WQPs is crucial in understanding how the different parameters are distributed in a reservoir and in predicting the concentration of a given WQP at a specific location (Adusei et al., 2021). The approach can be used to overcome the

limitations of point sampling particularly in large water bodies where sample collection and analysis at every point is virtually impossible (Adusei et al., 2021). Although a variety of deterministic and geostatistical interpolation methods can be used to estimate variables at unsampled locations, accuracies vary widely among methods (Ouma et al., 2018). While there are several types of Kriging used in geostatistical data interpolation such as Ordinary Kriging, Universal Kriging, and median polish, ordinary Kriging is the basis of geostatistics and gives the optimal data predictions (Ouma et al., 2012). The main assumption for ordinary Kriging is that the process is second-order stationary with normal distribution (Ouma et al., 2012). Murphy et al. (2010) related the performance of inverse distance weighting (IDW), and Kriging-based methods for spatial mapping of WQPs. The Kriging-based methods achieved higher accuracies compared to the IDW method (Murphy et al., 2010). Therefore, this study used Kriging to estimate the WQPs.

3.4 ANN Model Training, Validation, Testing, and Application

There are different architectures and models for ANN, namely, multi-layer perceptron (MLP), adaptive neuro fuzzy inference system (ANFIS), recurrent neural network (RNN), generalized regression neural network (GRNN), and radial basis function network (RBFN). These ANN models can be categorized into feedforward neural networks and recurrent neural networks. MLP neural networks, trained with a backpropagation learning algorithm, are the most popular feedforward neural networks (FFNNs) and have been widely used in hydrologic forecasting models (Da Silva et al., 2017; Singh et al., 2014; Sumi et al., 2012). The advantage of FNN is that with as few as a single hidden layer and arbitrary bounded and smooth activation functions, the system can approximate a continuous nonlinear function (Da Silva et al., 2017).

3.4.1 Selection of Network Architecture

The ANN architecture used in this study consisted of a MLP which is an example of a FFNN where the inputs are sent from one layer to the next by one-way weights (Haghiri, 2018). The choice to use feed forward back propagation (FFBP) MLPs was informed by the concept that MLPs have been used to successfully predict optimal coagulant doses (Baouab & Cherif, 2018; Haghiri, 2018). This study had few input data and the MLP-ANN network architecture was preferred in order to provide quick predictions with high accuracies. Even though MLP-ANN works well with large input data, the same accuracies can still be achieved even with small data sets (Jorgensen & Gromiec, 2016).

3.4.2 Number of Layers and Neurons

The ANN models used in this study had three layers. That is an input, hidden, and output layers. The optimum number of hidden layers and the optimum number of nodes was found by trial and error. According to Leon-Luque et al. (2016), any continuous function can be approximated using one hidden layer given sufficient degrees of freedom. One can also get a decent performance by setting the number of hidden layers to one or setting it to an average of the neurons in the input and output layers without iterations (Kahani et al., 2018). In this case, the ANN model 1 for predicting the treated WQPs and ANN model 2 for predicting the optimal coagulant dose each had one hidden layer. This means that both models were shallow learning neural networks that depended on the input data to create the prediction model. The ideal number of nodes in the hidden layer was determined through trial and error by either increasing or decreasing the initial neural number until the models reached the desired performance; the performance was evaluated after each iteration (Jalali et al., 2019). The iterations

stopped when the change in the coefficient of determination (R^2) was not considered significant anymore (Jalali et al., 2019).

3.4.3 Selection of Network Functions

The selection is governed by the characteristics of the available data and the learning type of the network (Hayou et al., 2018; Pratiwi et al., 2020). However, the linear and sigmoid functions are the most common activation functions (Hayou et al., 2018; Pratiwi et al., 2020).

3.4.4 Normalization

The normal distribution function was used to normalize the data before it was entered into the models. This helped to improve the system performance by preventing excessive fluctuations that could interfere with the learning function. According to Oostwal et al., (2021), scaling the input data linearly into 0 and 1 ensures that it is in agreement with the limits of the transfer function in the ANN output layer (sigmoidal).

That is:

$$Z = \frac{(x-\bar{x})}{s} \tag{9}$$

where x is the primary quantity, $\bar{x}$ is the average of the data, and s the standard deviation.

3.4.5 Data Division and Pre-Processing

Two ANN models were calibrated, validated, and used. The default MATLAB data division was applied in splitting the data set for the sampling days. That is 70% for model training, 15% for model validation, and the remaining 15% for model testing. The first model was then applied to predict treated WQPs and the second model used to determine the optimum coagulant dose.

a) **Training phase**: Training algorithms were selected to facilitate ANN learning by changing the weights of the nodes. The main objective of the training was to reduce the mean squared error (MSE) which ultimately resulted in the successful construction of the model (Ahmed et al., 2018). Training also allowed the model to see and learn from the data.

b) **Validation phase**: Model validation after training was then done. The prediction of the accuracy and the trained model's performance were evaluated. The network's ability to generalize was also be measured by checking whether the variance between the predicted and the actual values of the outputs had been minimized (Ahmed et al., 2018).

c) **Testing phase**: This was done to evaluate the learning progress made by the network thus optimizing the structure of the ANN. It also helped to establish when to cease training and to ensure that the model gave an accurate prediction on any data that is outside the training set (Chen et al., 2020; Dawood et al, 2021). This also helped the model to predict instead of giving values based on the memorized input pattern (Ahmed et al., 2018).

d) **Application**: The first model was then used to predict treated WQPs actively controlled by coagulant dose (that is turbidity, Chl-*a*, and total suspended solids) and the second model was used to predict the optimum coagulant dose based on the input WQPs.

3.4.6 Model Performance Evaluation Functions

The performance of both models was evaluated to establish the cumulative values between the target output values and created outputs by the network. Different

statistical methods including mean absolute error (MAE) or bias, the coefficient of determination (R^2), and root mean square error (RMSE) were used.

3.5 Model 1- Prediction of Treated Water Quality Parameters

The model for predicting treated WQPs was first prepared. A schematic of the model inputs and outputs is shown in Figure 3.3.

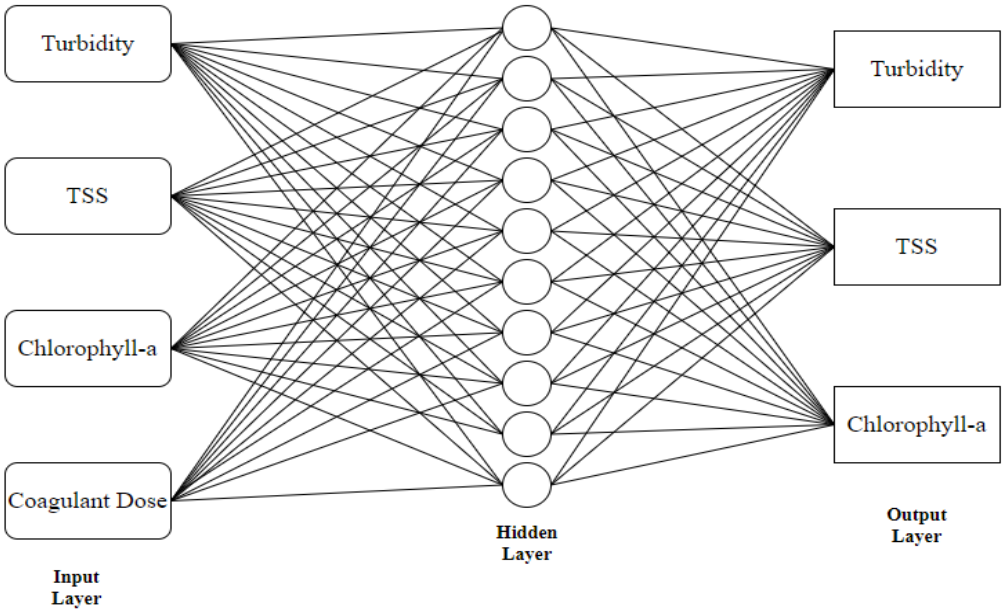


Figure 3.3: Schematic diagram of the input and output layers of ANN model 1.

In this case, the raw WQPs were the process inputs, the alum dose was the process control parameter, and the process outputs being modelled were the treated water quality parameters. From the EWRM, the regression equations with the highest R^2 values were used for estimating the values of turbidity, TSS, and Chl-*a* from the Landsat-8 OLI band combinations. The normalized values of estimated Landsat-8 OLI measurements together with the coagulant doses for each of the specific sampling days were the inputs for training, validating, and testing the ANN model while the corresponding treated WQPs were the targets as shown in Figure 3.4. The input data was normalized and an ANN-MLP with a single hidden layer and 10 neurons developed

and used. The output of this model was the desired treated WQPs (turbidity, Chl-*a*, and TSS). The model development was done using MATLAB software.

3.6 ANN Model 2- Prediction of Optimum Coagulant Dose for Water Treatment

A second model was used to anticipate the amount of coagulant necessary depending on the values of the raw WQPs and the desired treated WQPs values. An MLP-ANN with one hidden layer and 10 neurons was developed and used. The input data consisted of normalized values of: raw water turbidity, TSS, and Chl-*a* (estimated from Landsat-8 OLI) and treated water turbidity, Chl-*a*, and TSS (based on laboratory experiments) and the model output was the optimal coagulant dose (Figure 3.4).

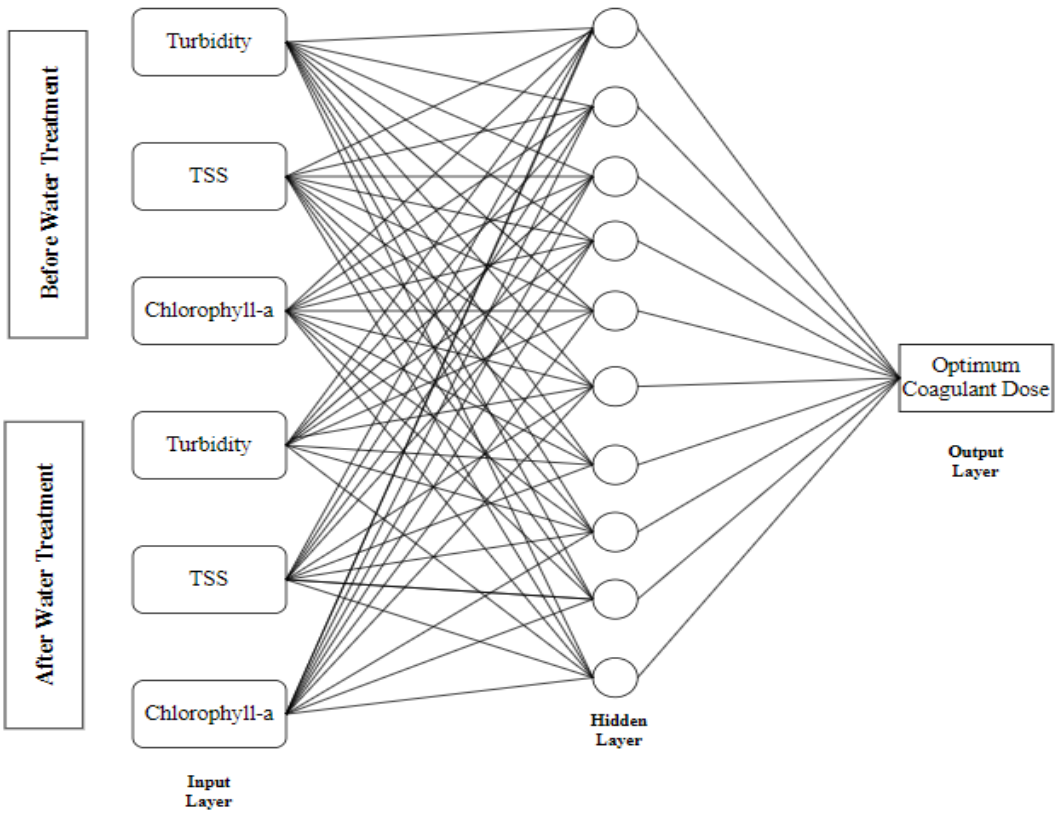


Figure 3.4: Schematic diagram of the input and output layers of ANN model 2.

Similar to case model 1, regression equations with the highest R^2 values were also used for estimating the values of turbidity, TSS, and Chl-*a* from the Landsat-8 OLI band combinations. The normalized values of Landsat-estimated raw water turbidity, TSS, and Chl-*a* and treated water turbidity, Chl-*a*, and TSS (based on laboratory experiment) for each of the specific sampling days were the inputs for training, validating, and testing the ANN model while the corresponding optimal coagulant dose was the output.

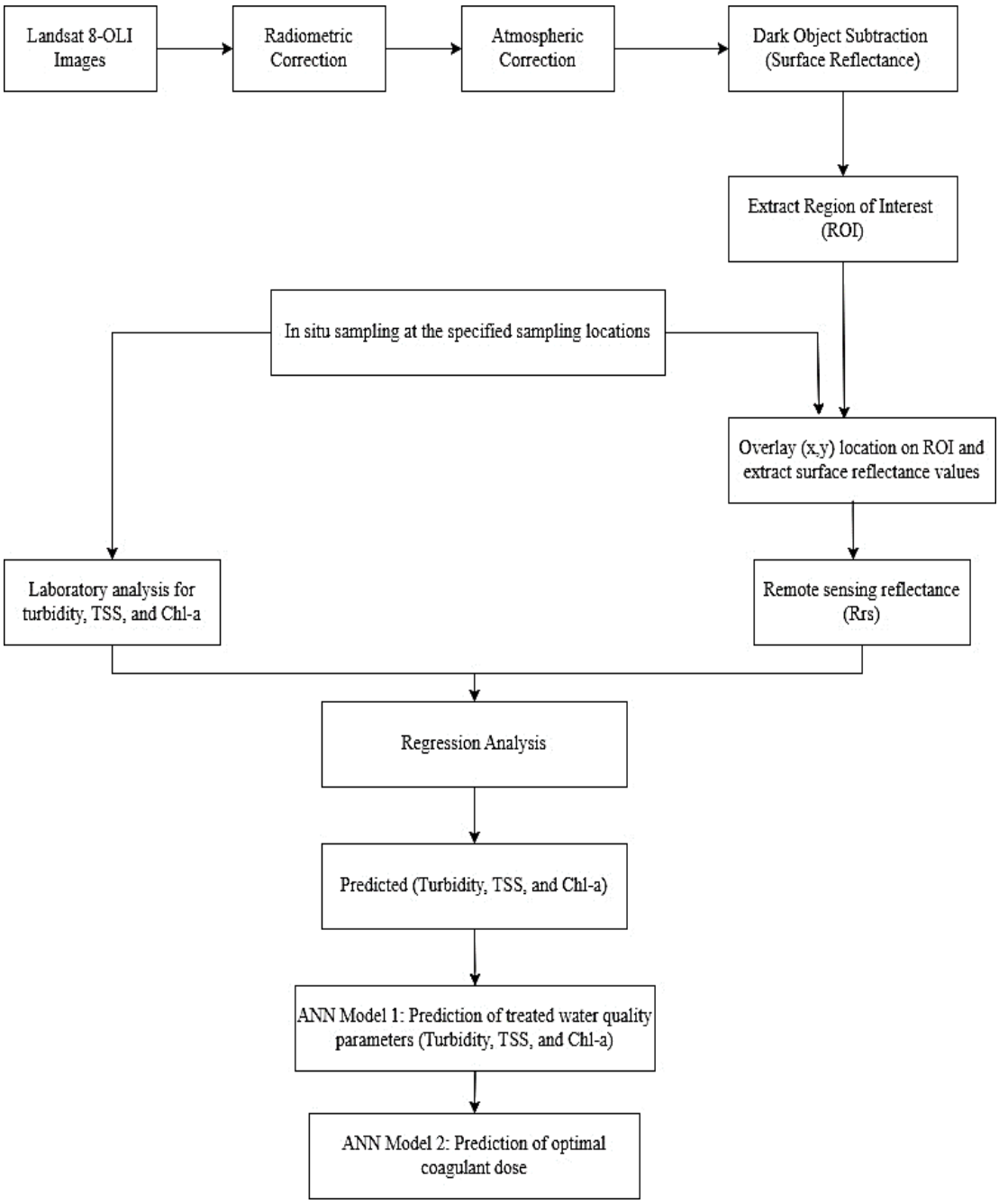


Figure 3.5: Flow diagram of the summarized methodology.

CHAPTER 4: RESULTS AND DISCUSSION

4.1 Estimation of turbidity, TSS, and Chl-a in Correlation with in situ Water Quality Data using EMRM

4.1.1 Estimation of Turbidity

The turbidity values based on the lab measurements ranged between 4-10 NTU, 4-13 NTU and 3-17 NTU for data collected in November 2020, December 2020, and January 2021, respectively as shown in Table 4.1. The estimated turbidity measurements ranged between 5-10 NTU, 4-13 NTU and 5-17 NTU for data collected in November 2020, December 2020 and January 2021, respectively as shown in Table 4.1.

Table 4.1: *In situ* and estimated turbidity for the specific sampling days

Sampling Station	*In situ* Turbidity (NTU)			Estimated Turbidity (NTU)		
	25/11/2020	11/12/2020	28/01/2021	25/11/2020	11/12/2020	28/01/2021
1	7.0	4.0	5.0	7.3	4.0	5.4
2	8.0	6.0	8.0	10.1	4.6	6.8
3	5.0	5.0	4.0	6.4	4.4	6.6
4	9.0	10.0	10.0	9.1	6.8	7.7
5	8.0	8.0	10.0	8.0	7.1	9.0
6	8.0	7.0	10.0	4.5	6.5	9.0
7	4.0	5.0	7.0	4.9	5.2	6.4
8	5.0	5.0	3.0	8.8	4.7	6.2
9	7.0	8.0	10.0	4.9	5.6	9.0
10	10.0	6.0	8.0	10.1	6.3	5.4
11	9.0	13.0	17.0	6.0	12.9	16.9
12	6.0	5.0	10.0	7.5	6.4	6.9
13	10.0	10.0	10.0	9.0	6.8	8.7

Generally, the turbidity in the reservoir was observed to be low mainly because of low sediment inflows throughout the sampling duration and plain sedimentation taking place in the reservoir. The sampling was done between November and January which is a dry period meaning there was no sediment inflow from rainwater discharge into the

reservoir. Sediment loads could have also reduced as a result of plain sedimentation which refers to the quiescent settling of water in a lake, basin, or reservoir for extended durations without the aid of chemicals particularly when the water source is polluted or highly turbid (Mehdinejad et al., 2012). The concept is more like natural water treatment that results in the settlement of suspended solids, removal of color, hardness reduction, breakdown of organic chemicals, and unfavorable conditions that lead to the death of pathogens (Mehdinejad et al., 2012). The band combinations and ratios from the Landsat data that gave the best estimate of the turbidity based on the remote sensing reflectance (Rrs), the regression equations and associated errors are shown in Table 4.2.

Table 4.2: Regression equations and associated errors for turbidity estimation

	Regression Equation	**Band Combination**	**R^2**	**nRMSE (NTU)**	**Bias (NTU)**
25/11/2020	**$y = -1169x^2+3694x-2908$**	(B1/B4)+B2	0.797	0.257	0.084
	$y = -996.8x^2+2853x-2031$	B1/B4	0.680	0.308	-0.792
	$y = -1552x^2+2831x-1280$	B4/B2	0.729	0.286	0.554
11/12/2020	**$y = 68165x^2-15713x+908.2$**	B3+B4+B1	0.757	0.631	-1.572
	$y= 67304x^2-11390x+484.4$	B2+B3+B4	0.736	0.571	-1.608
	$y = 87184x^2-8803+226.4$	B2+B4	0.726	1.060	-1.342
28/01/2021	$y = 29.02In(x)+117.7$	B4	0.688	0.364	-0.402
	$y = 712.1x^2-704.5x+175.2$	B3/B1	0.736	0.355	-1.796
	$y = 130x^2-493.6x+471.4$	(B1/B3)+B2	0.736	0.360	-1.80

The reflectance from the visible bands based on the polynomial regression models highlighted in Table 4.2 yielded the highest correlation coefficient between the laboratory and satellite-derived data with R^2 values of 0.797, 0.757, 0.736 for the

specific data collection dates. The findings in this study for turbidity regressions are also supported by the results of Ouma et al. (2020) which showed the viability of using visible Landsat-8 OLI bands for turbidity estimations. The results from Lotfi et al. (2019) also highlighted the usefulness of the visible Landsat-8 OLI bands in estimating *in situ* turbidity. The significance of the red and blue band in estimation of turbidity is also emphasized in a study by Kalele (2019) where the model that gave the best results combined the reflectance values of the red and blue bands and it achieved R and RMSE values of 0.841 and 0.828 respectively. The model validation dataset gave R^2 and RMSE values of 0.832 and 0.430 respectively. Furthermore, the nRMSE and Bias values shown in Table 4.2 were lower than the smallest and the average *in situ* turbidity values and this is an indication of less variation between laboratory and satellite-derived values. Thus, the model is suitable for estimating turbidity in inland waters.

4.1.2 Estimation of Total Suspended Solids

The TSS values based on the lab measurements ranged between 250.6 – 300.4 mg/L, 247.6 -314.1 mg/L and 250.8-321.3 mg/L for data collected on 25th November 2020, 11thDecember 2020 and 28th January 2021 respectively as shown in Table 4.3. The estimated TSS measurements ranged between 253.63 – 300.67 mg/L, 200.02 -333.88 mg/L and 207.85-308.58 mg/L for data collected on 25th November 2020, 11thDecember 2020 and 28th January 2021 respectively as shown in Table 4.3.

Table 4.3: *In situ* and estimated TSS for the specific sampling days

Sampling Station	*In Situ* TSS (mg/l)			Estimated TSS (mg/l)		
	25/11/2020	11/12/2020	28/01/2021	25/11/2020	11/12/2020	28/01/2021
1	273.1	267.2	253.0	270.3	270.7	269.5
2	274.5	271.4	276.2	254.9	271.9	282.3
3	254.7	247.6	250.8	266.1	264.7	287.5
4	263.4	270.2	269.7	282.1	285.6	296.4
5	271.4	290.8	273.2	272.0	298.1	308.3
6	273.0	291.3	314.2	255.6	292.3	308.0
7	250.6	299.2	298.0	253.8	278.1	282.0
8	252.1	287.6	284.8	274.8	270.1	280.0
9	293.2	302.2	313.6	254.7	288.7	308.6
10	300.4	205.8	303.2	300.7	200.0	278.9
11	276.5	349.4	207.6	259.5	333.9	207.9
12	260.8	261.6	289.6	268.2	291.7	290.2
13	281.2	314.1	321.3	273.5	292.8	306.4

The band combinations and ratios that gave the best estimate of TSS, the regression equations and associated errors are shown in Table 4.4. TSS concentration was notably high at areas where River Ellegerini and River Endoroto flowed into the reservoir because the inflow from these two rivers caused sediment resuspension.

Table 4.4: Regression equations and associated errors for TSS estimation

	Regression Equation	**Band Combination**	**R^2**	**nRMSE (mg/l)**	**Bias (mg/l)**
25/11/2020	**y = -5340In(x) + 2754**	B3/B2	0.788	0.704	-9.981
	y = 7930x – 9457	(B1/B4)+B2	0.804	0.493	0.242
	y = 17955x^2 – 47361x +33603	(B1/B3) + B2	0.808	0.729	-7.835
11/12/2020	**y = 635.9$e^{53.65x}$**	B4	0.853	0.376	2.660
	y = 4640x^2-23267x+31152	(B1/B4) + B2	0.801	0.293	4.201
	y = 77928x^2-64825x+15537	(B4/B1) + B4	0.818	0.292	3.505
28/01/2021	**y = -6131x^2+ 25640x – 23721**	(B1/B4) + B1	0.757	0.497	15.862
	y = -6139x^2 +25265x – 22910	(B1/B4) + B2	0.757	0.496	15.704
	y = -79430x^2 + 83798x – 18978	(B4/B1) + B4	0.766	0.543	16.546

Table 4.4 shows that the logarithmic, exponential, and polynomial regression models gave the best estimate of TSS with R^2 values of 0.808, 0.853, and 0.766 respectively for the specific data collection dates. The visible bands (coastal aerosol, blue, green, and red) proved to be the optimal bands for TSS estimation in this reservoir.

The results from the logarithmic model in Table 4.4 based on the band ratio between B3 (green) and B2 (blue) gave an R^2 value of 0.788. Similar results were obtained by Jaelani et al. (2016) in which the logarithmic model developed from the Rrs band ratio of B2 (blue) and B3 (green) resulted in an R^2 value of 0.79. The viability if the single B4 (red) for TSS estimation from Landsat imagery is also demonstrated by Yanti et al. (2016) where TSS estimation using B4 and a linear regression model yielded an R^2 value of 0.543. Ouma et al. (2020) also achieved an R^2 value of 0.9249 based on B3 (green) and B2 (blue) ratio using a linear regression model. In comparison with the

other studies, Yanti et al. (2016) concluded that the red band alone is not that informative in the retrieval of TSS. However, in this study, the red band alone was quite informative in the retrieval of TSS from the reservoir. However, just like Yanti et al. (2016) suggested, combining the red band and other visible bands proved to be quite effective in the estimation and mapping of the WQPs in this reservoir. Generally, the nRMSE and Bias errors for this study were also lower compared to the TSS values measured *in situ* and this proves that the method is sufficient for estimating TSS concentration of inland waters.

4.1.3 Estimation of Chlorophyll-*a*

The Chl-*a* values based on the lab measurements ranged between 23.08 – 59.42 mg/L, 31.36 -83.40 mg/L and 24.22-80.86 mg/L for data collected on 25th November 2020, 11thDecember 2020 and 28th January 2021, respectively as shown in Table 4.5. The estimated Chl-*a* measurements ranged between 23.58 – 60.67 mg/L, 31.87 -83.15 mg/L and 29.52-76.22 mg/L for data collected on 25th November 2020, 11thDecember 2020 and 28th January 2021, respectively as shown in Table 4.5.

Table 4.5: *In situ* and estimated Chl-*a* for the specific sampling days

Sampling Station	*In Situ* Chl-*a* (mg/l)			EMRM Estimated Chl-*a* (mg/l)		
	25/11/2020	**11/12/2020**	**28/01/2021**	**25/11/2020**	**11/12/2020**	**28/01/2021**
1	27.16	37.44	24.22	30.11	40.65	40.65
2	43.01	31.36	33.16	47.49	38.96	38.96
3	35.14	34.39	39.78	33.97	31.87	41.55
4	23.08	39.11	28.07	24.8	57.02	57.02
5	38.2	43.8	45.86	31.05	65.73	65.73
6	32.35	72.8	80.86	60.67	58.57	76.22
7	59.42	69.75	75.42	45.49	44.46	32.91
8	45.2	41.41	25.4	23.58	38.96	30.3
9	43.34	57.84	48.79	56.8	58.91	58.91
10	28.65	43.4	40.39	27.53	42.86	29.52
11	52.52	83.4	30.4	40.42	83.15	56.20
12	30.04	37.38	38.86	37.63	60.50	45.56
13	25.09	69.08	70.51	27.19	58.90	72.96

The band combinations and ratios that gave the best estimate of Chl-*a*, the regression equations and associated errors are shown in Table 4.6.

Table 4.6: Regression equations and associated errors for Chl-*a* estimation

	Regression Equation	Band Combination	R^2	nRMSE (mg/l)	Bias (mg/l)
25/11/2020	$y = 21293x^2 – 32118x + 12136$	B3/B1	0.799	0.238	3.674
	$y = 7820x^2 – 20734x + 13767$	B1/B3	0.802	0.227	3.092
	$y = 16746x^2 – 53631x + 42964$	(B1/B4) + B1	0.807	0.406	-3.696
11/12/2020	$y = 8797x^2 – 7585x + 1658$	B4/B1	0.68	0.493	10.344
	$y = 7556x^2 – 6881x + 1591$	(B4/B1) +B4	0.682	0.525	10.966
28/01/2021	$y = -8121x^2 + 8983x – 2389$	(B4/B1) + B4	0.920	0.716	16.886
	$y = -9145x^2 + 9639x – 2444$	B4/B1	0.926	0.676	16.234

Farming is one of the key economic activities in the catchment. Therefore, the notably high Chl-*a* concentrations could be attributed to the inflow of fertilizer leachate into the reservoir. The highest concentration of Chl-*a* specifically in December 2020 and January 2021 was noted in the same locations where TSS concentration was high. The inflow of water from River Endoroto and River Ellegerini increases the concentration of particulate matter and nutrients at these points. This is because farming is the leading economic activity around the reservoir and this means that the observed Chl-*a* concentrations can be linked to diffuse pollution by fertilizer leachate from the nearby farms, specifically, an influx of total phosphorous and total nitrogen which are the main variables that contribute to nutrient enrichment. Consequently, the concentration of Chl-*a* which is the response variable increases. The problem is further worsened by the rainy season which facilitates significant nutrient runoff followed by a dry season, which provides perfect conditions for algae incubation (KDHE, 2011). Furthermore, the increased concentration of particulate matter provides attachment sites for the algae and this enables the algal bloom concentration to be propagated (KDHE, 2011). Thus, this led to the observed high concentration of both TSS and Chl-*a* in the same regions.

The remote sensing reflectance (Rrs) based on second order polynomial regression models for band ratios and mixed band combinations gave the best estimate of Chl-*a*. The visible bands B1, B3, and B4 (coastal aerosol, green, and red) gave the highest R^2 values of 0.802, 0.682, 0.926 for the specific data collection dates between the laboratory and satellite-derived Chl-a values (Table 4.6). Jaelani et al. (2016) also obtained comparable results using band ratios (B1, B2, B3, B4) in developing retrieval algorithms for Chl-*a* with the accuracy denoted by $R^2>0.5$. As opposed to the study by Lai et al. (2021) which states that the best band combination for the retrieval of Chl-a is that which includes the blue and near infrared bands, the near infrared band is not

that informative in the retrieval of Chl-*a* in this reservoir. Regardless, Lai et al. (2021) also acknowledges that if only the near infrared and blue bands are used for Chl-*a* retrieval, then the correlation is not ideal. Therefore, the performance evaluation confirmed that the most appropriate algorithms for the retrieval of Chl-*a* in the reservoir using Landsat-8 OLI are those developed from band combinations of B2, B3, and B4 (Table 4.6).

4.1.4 Validation of Predicted Water Quality Parameters with *in situ* Measurements

Five (5) sampling stations were used to validate the regression equations developed. The validation results are presented in Table 4.7 with inclusion of data from the sampling locations used in model calibration.

Table 4.7: Validation results for predicted and *in situ* water quality measurements

25/11/2020	**Water Quality Parameter**	**Data Source**	**Sample (n)**	**Min.**	**Max.**	**Med.**	**Avg.**	**SD**	**CV (%)**	**SE**
	Turbidity	*In situ*	13	4.00	10.00	8.00	7.38	1.94	26.25	0.54
		Landsat -8 OLI	13	4.50	10.13	7.49	7.44	1.96	26.31	0.54
	TSS	*In situ*	13	250.6	300.4	273.00	271.15	15.04	5.55	4.17
		Landsat -8 OLI	13	253.75	300.67	268.23	268.17	13.37	4.99	3.71
	Chl-*a*	*In situ*	13	23.08	59.42	35.14	37.17	11.04	29.71	3.06
		Landsat -8 OLI	13	23.58	60.67	33.97	37.44	12.08	32.26	3.35
11/12/2020	Turbidity	*In situ*	13	4.00	13.00	6.00	7.08	2.63	37.15	0.73
		Landsat -8 OLI	13	4.02	12.90	6.27	6.25	2.24	35.92	0.62
	TSS	*In situ*	13	205.80	349.40	287.60	281.42	34.57	12.29	9.59
		Landsat -8 OLI	13	200.02	333.88	285.62	279.89	29.82	10.66	8.27
	Chl-*a*	*In situ*	13	31.36	83.40	43.40	50.86	17.38	34.17	4.82
		Landsat -8 OLI	13	31.87	83.15	57.02	52.35	14.16	27.06	3.93
28/01/2021	Turbidity	*In situ*	13	3.00	17.00	10.00	8.62	3.55	41.18	0.98
		Landsat -8 OLI	13	5.38	16.86	6.89	7.99	2.97	37.22	0.82
	TSS	*In situ*	13	207.60	321.30	284.80	281.17	31.56	11.23	8.75
		Landsat -8 OLI	13	207.85	308.58	287.54	285.07	26.64	9.35	7.39
	Chl-*a*	*In situ*	13	24.22	80.86	39.78	44.75	19.19	42.88	5.32
		andsat-8 OLI	13	29.52	76.22	45.56	49.73	15.89	31.94	4.41

The EMRM algorithm used for the prediction of WQPs is a data-driven model and data driven models have been frequently used to assess the water quality index (WQI) for natural streams (Najafzadeh et al., 2021). Based on the results shown in Table 4.7 the variation in concentration was highest for TSS then Chl-a, and lastly Turbidity. The standard deviation (SD), coefficient of variation (CV), and standard error (SE) metrics in Table 4.7 show that Landsat-8 OLI mostly underestimated the concentration of the WQPs nonetheless with a small margin as denoted by the small difference in the CV values achieved for both *in situ* and Landsat-predicted WQPs. For instance, the coefficient of variation for *in situ* turbidity and Landsat-estimated turbidity are 26.25% and 26.31% respectively, which is notably a low difference. This means that both Landsat and laboratory measurements can reliably give accurate results.

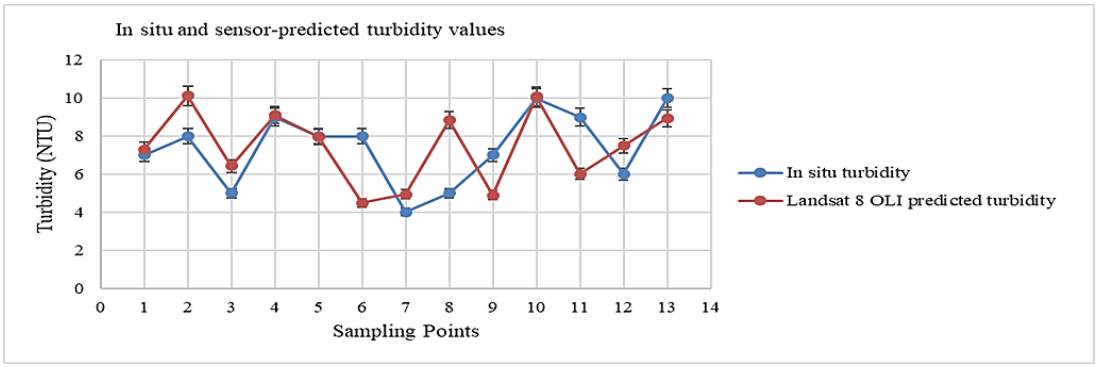
The accuracies obtained based on the EMRM model can be compared to the results obtained by (Najafzadeh et al., 2021) who used four well-known DDMs including Evolutionary Polynomial Regression (EPR), M5 Model Tree (MT), Gene-Expression Programming (GEP), and Multivariate Adaptive Regression Spline (MARS) for the prediction of the WQI in Karun River, Iran. The number of DDMs feeding-input variables were controlled through techniques like Forward Selection (FS), and Gamma Test and the FS-M5 MT gave the best estimate of the WQI.

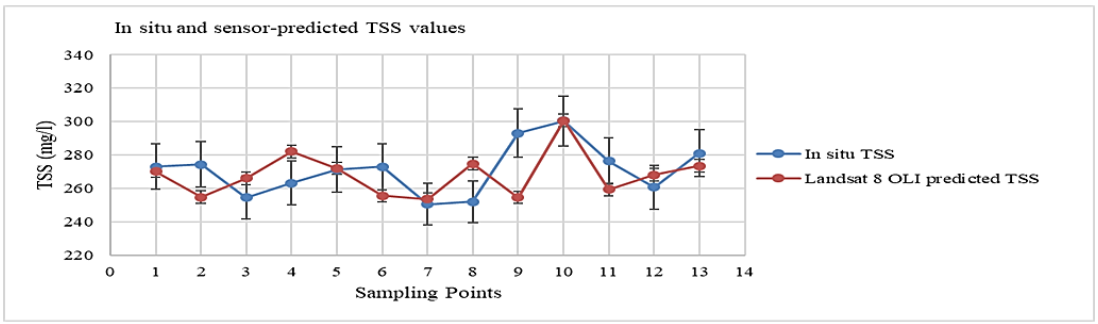
Figure 4.1, Figure 4.2, and Figure 4.3 present the laboratory-measured and Landsat-predicted results for turbidity, TSS, and Chl-*a*. Figure 4.2 shows that some Landsat-predicted values matched the *in situ* measurements. For instance, Landsat-predicted and *in situ* TSS measurements mostly coincided and the CV between the values was 11%. Additionally, the overestimated or underestimated Landsat-predicted values in most cases had a small margin of error. Even though in this study the number of input variables were not controlled because of their low numbers (only three inputs

specifically turbidity, TSS, and Chl-a), great accuracies were still achieved as shown through the validation results in Table 4.7. This shows that the EMRM approach is a reliable DDM for the estimation of WQPs from Landsat-8 OLI.

In Figure 4.2, the actual and predicted values followed a similar trend line with most points coinciding for actual and predicted turbidity and TSS values. For Chl-*a*, notable variations were at points (4,5, and 12) and points (6 and 7) where Landsat-8 values were overestimated and underestimated respectively by a significant margin. Figure 4.3 shows that there was a significant variation of the actual and predicted values specifically for turbidity and Chl-*a*. Landsat underestimated the turbidity values while Chl-*a* values were overestimated. However, there was a slight variation between Landsat-predicted and *in situ* TSS since Landsat estimated the TSS values with a coefficient of variation of less than 10%.

The model results can be compared with the results from the study by Nafsin and Li (2022) which investigated the effectiveness of four stand-alone machine learning (ML) algorithms and six novel hybrid algorithms in predicting the 5-day BOD of Buriganga River, Bangladesh. The Random Forest-Support Vector Machine (RF-SVM), Artificial Neural Network-Support Vector Machine (ANN-SVM), and Gradient Boosting Machine-Support Vector Machine (GBM-SVM) achieved high prediction accuracies of 91%, 89.6%, and 88.8% respectively. This means that the ML algorithms, just like the EWRM algorithm, can also be used to improve the accuracy of water quality parameter predictions from satellite imagery. The high prediction accuracies could significantly reduce the coefficient of variation between *in situ* and Landsat-predicted WQPs.





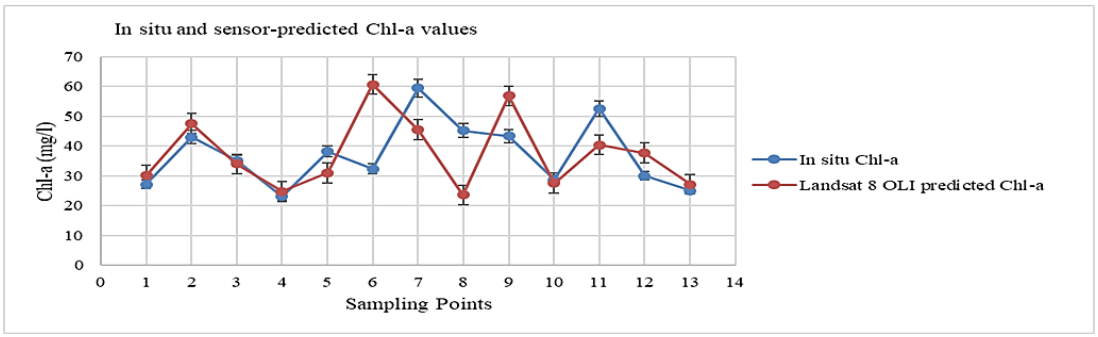


Figure 4.1: Variations between prediction and validation results for turbidity, TSS, and Chl-*a* (25/11/2020).

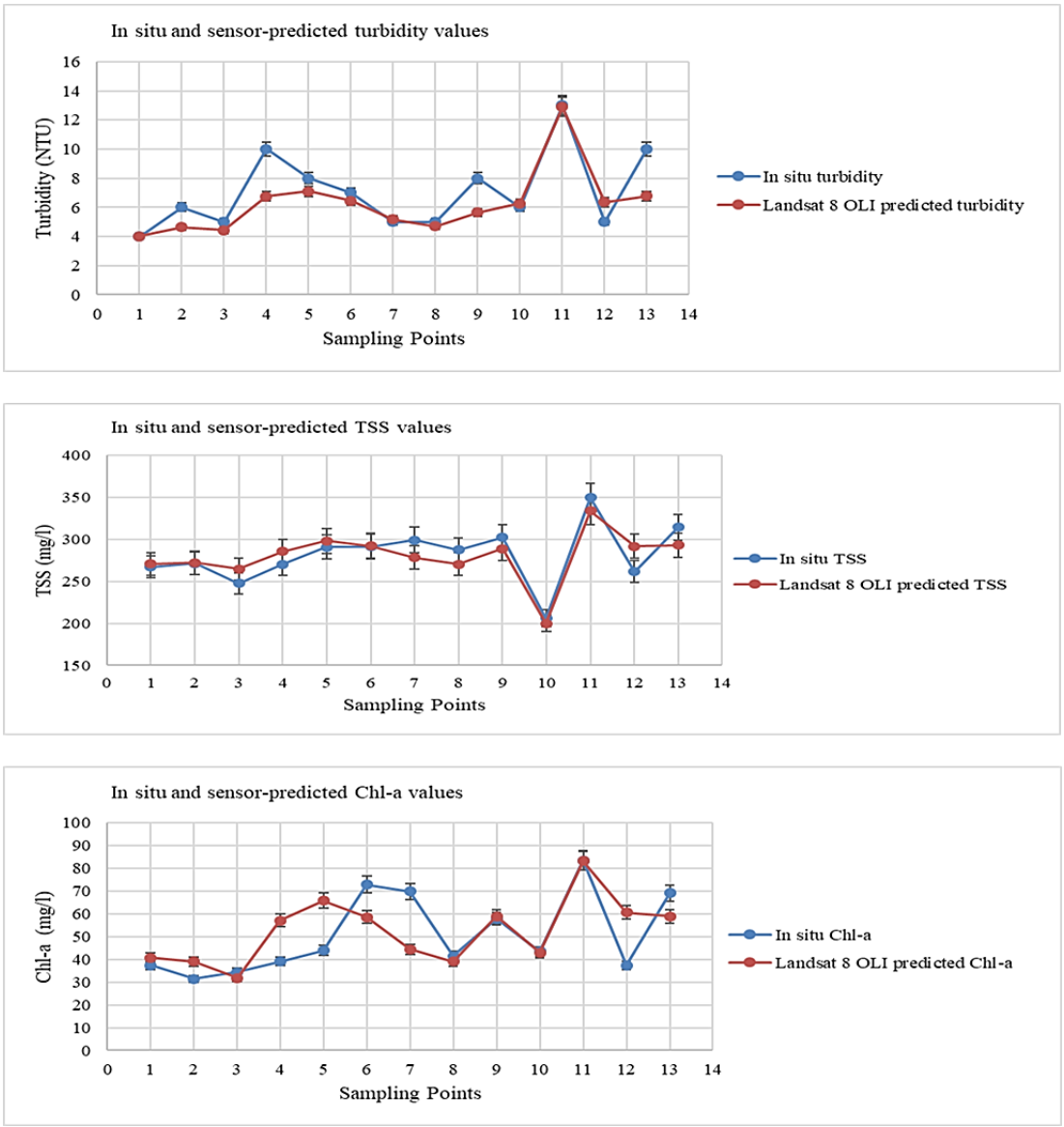


Figure 4.2: Variations between prediction and validation results for turbidity, TSS, and Chl-*a* (11/12/2020).

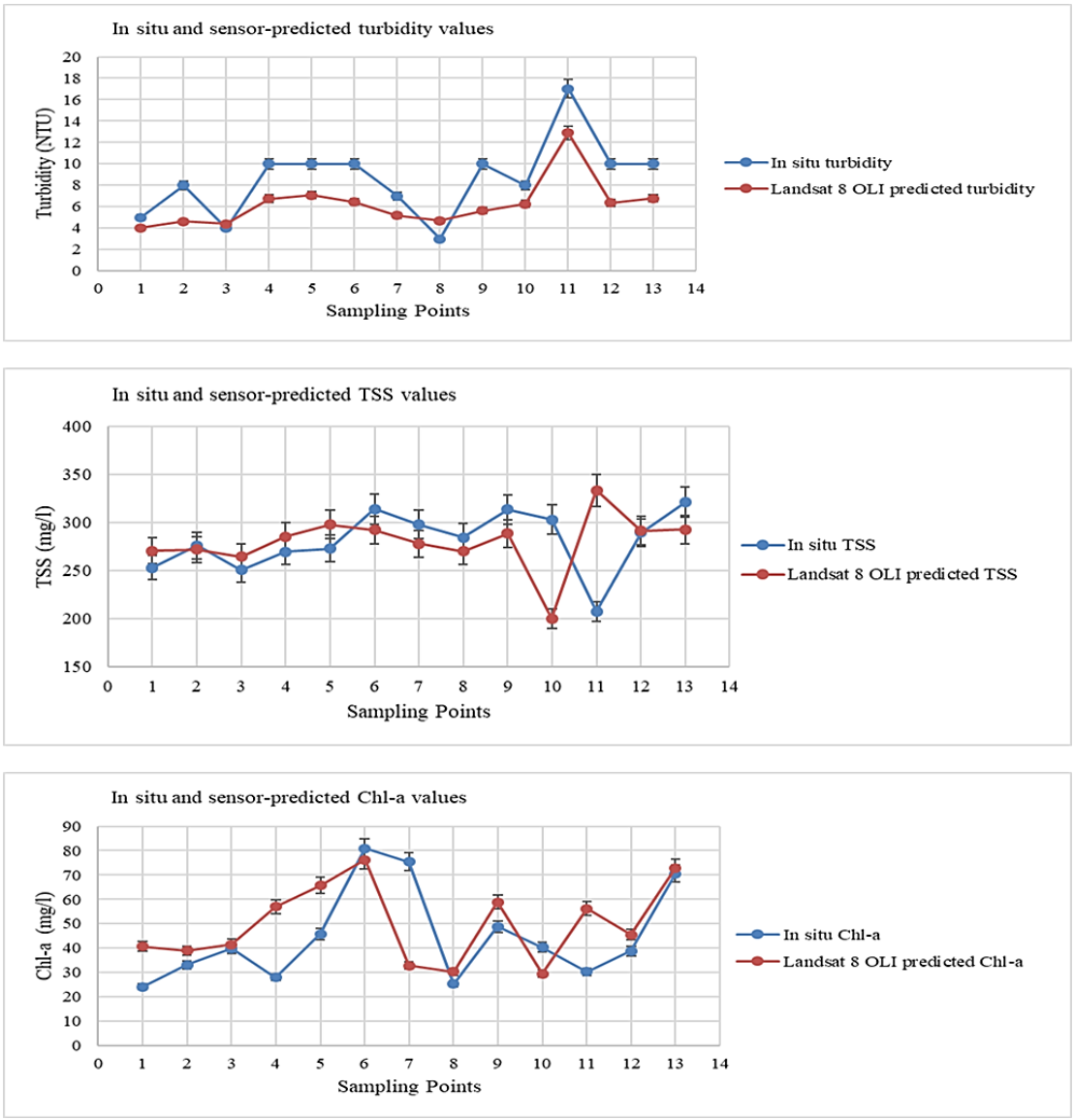


Figure 4.3: Variations between prediction and validation results for turbidity, TSS, and Chl-*a* (28/01/2021).

Generally, Landsat-8 OLI did great in the estimation of WQPs with realistic variations between the predicted and *in situ* measurements based on the metrics in Table 4.7. Even though the lowest variation was achieved in TSS estimation, the accuracy of WQM is highly dependent on the efficacy of atmospheric correction (Bonansea et al., 2019). Overall, Landsat-8 OLI avails a reliable and affordable technique for WQM, the effectiveness of which is dictated by the adequacy of the atmospheric and radiometric corrections.

The use of Landsat 8 OLI to estimate the influent water quality characteristics can be adapted to help maintain a stable performance of the water treatment plant. There is a significant relationship between the raw WQPs and the quantity of flocculants and coagulants used in water treatment. In practice, the influence WQP estimations are estimated based on the conventional sampling and laboratory analysis method. However, changing weather patterns such as rainfall events may cause large variations in the influent water quality characteristics, and this could reduce the efficiency of the water treatment plants. Therefore, using Landsat data can help anticipate and effectively plan for, and monitor the water treatment plant processes with reference to the types of chemicals to be used.

4.2 Spatial Distribution and Variability of Estimated and Predicted Water Quality Parameters

Ordinary Kriging was used in the development of spatial maps showing the distribution and variability of the Landsat-predicted and *in situ* WQPs. The spatial distribution maps gave a visual analysis of the spatial distribution of the WQPs in the reservoir based on data collected from the specific sampling points.

4.2.1 Spatial Distribution and Variability of Turbidity

Figure 4.4 shows that for dates 11/12/2020 and 28/01/2021 the spatial turbidity distributions for Landsat-estimated and laboratory measurements were closely related. The regression models developed based on the data collected for the entire sampling period estimated the reservoir turbidity with more than 70% accuracy. Thus, even though there are some discrepancies highlighted in the spatial distribution maps which indicate a mismatch between Landsat-predicted and *in situ* measurements at some specific sapling points, the differences could be attributed to overestimation or underestimations by Landsat-8 OLI. The difference is also evident in Figure 4.1 and

Figure 4.4 which show lower coincidence between the laboratory and validated results at these specific points. Overall, the Landsat-8 OLI estimations in this study are similar to the observations reported by Ouma et al. (2020) in that estimated results are closely correlated with *in situ* turbidity both in spatial location and aerial distribution.

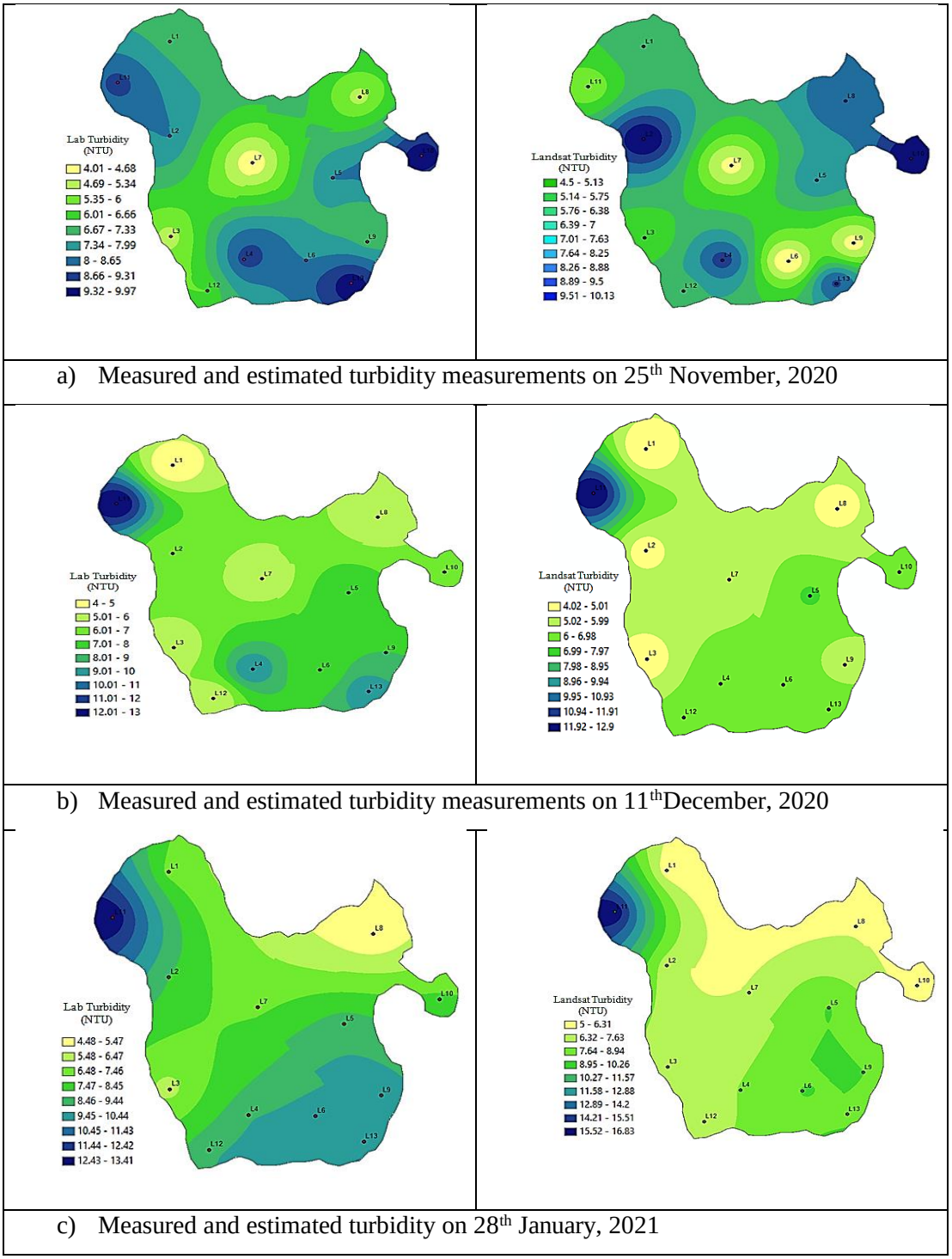
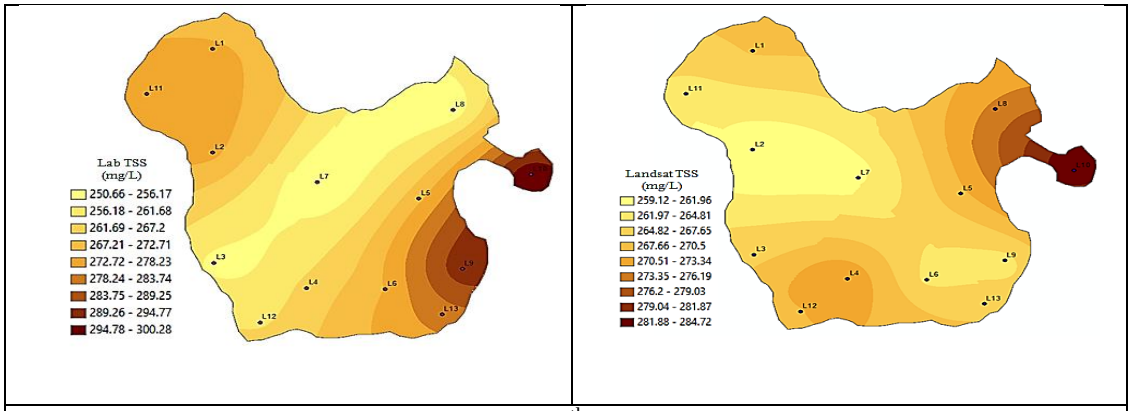


Figure 4.4: Spatial distribution of measured and estimated turbidity.

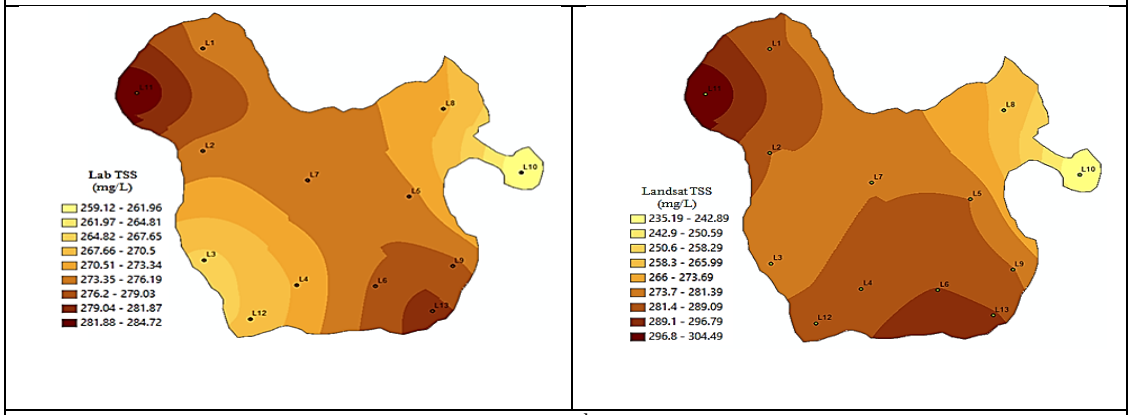
4.2.2 Spatial Distribution and Variability of TSS

The TSS spatial distribution maps show that the *in situ* and Landsat-estimated measurements are highly correlated, Figure 4.5. For both Landsat-8 OLI and *in situ* measurements, high TSS values were registered at sampling locations L10, (L11 & L13), and (L4, L9, &L13) in November 2020, December 2020, and January 2021 respectively. This shows that for the entire sampling period, the points with the highest TSS for both *in situ* and Landsat-estimated concentrations coincided.

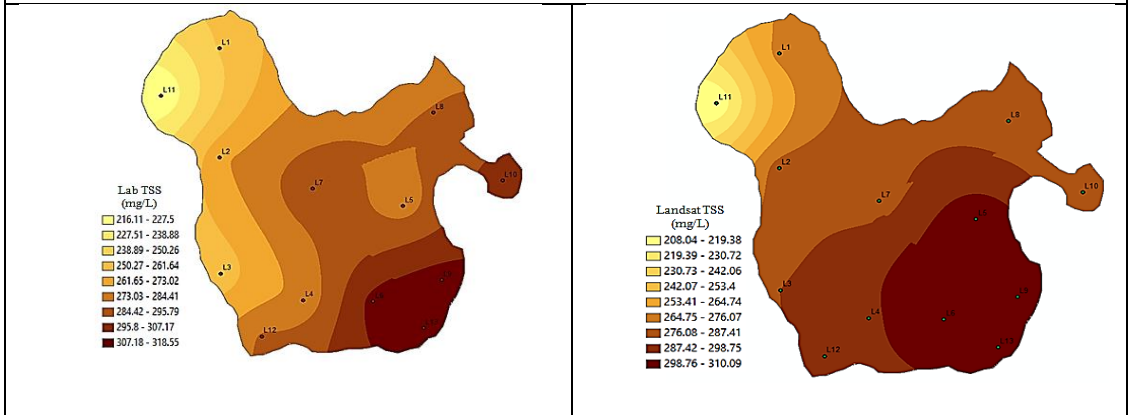
Notably, almost all the places that have a high TSS concentration also have high concentration of turbidity. This could be explained based on the findings by Lotfi et al. (2019) which confirmed that the total amount of suspended solids is a determining factor that contributes to an increase in turbidity. The spatial distribution and variability maps show that Landsat-8 OLI achieved great results in estimating TSS as indicated in Figure 4.1, Figure 4.2, and Figure 4.3 and in the statistical results in Table 4.4 and Table 4.7. Therefore, Landsat-8 OLI is useful for estimation and mapping reservoir TSS.



a) Measured and estimated TSS on 25th November, 2020



b) Measured and estimated TSS on 12thDecember, 2020



c) Measured and estimated TSS on 28th January, 2021

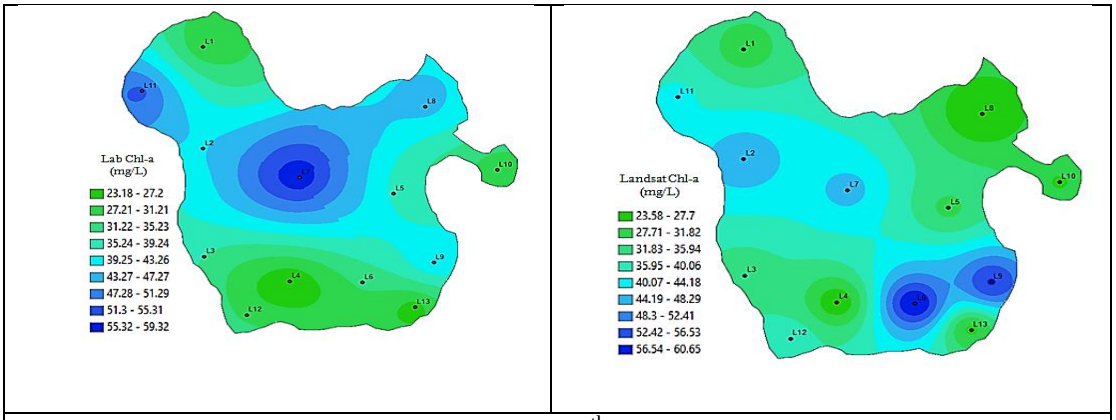
Figure 4.5: Spatial distribution of measured and estimated TSS.

4.2.3 Spatial Distribution and Variability of Chlorophyll-*a*

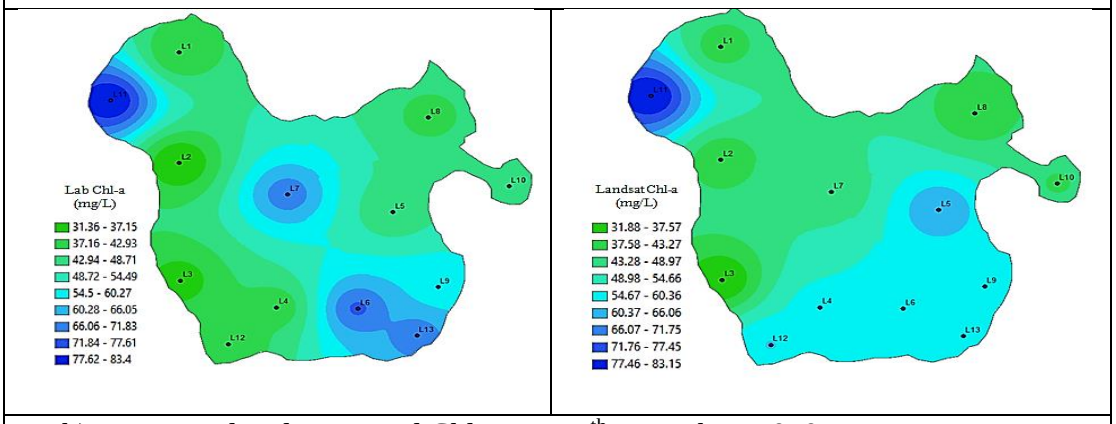
The spatial maps for Chl-*a* distribution for both the lab and *in-situ* measurements were closely correlated. For instance, points (L11 and L13) and point L8 for data collected in December and January respectively registered the highest Chl-*a* values, Figure 4.6.

However, in November, points L1, L4, and L10 had the lowest *in situ* and Landsat-estimated Chl-*a* concentration. The regions where Chl-*a* concentrations are high are in close proximity to areas where River Endoroto joins the reservoir. The river introduces point and non-point pollutants from the adjoining catchment areas where the main economic activities are manufacturing, and agriculture.

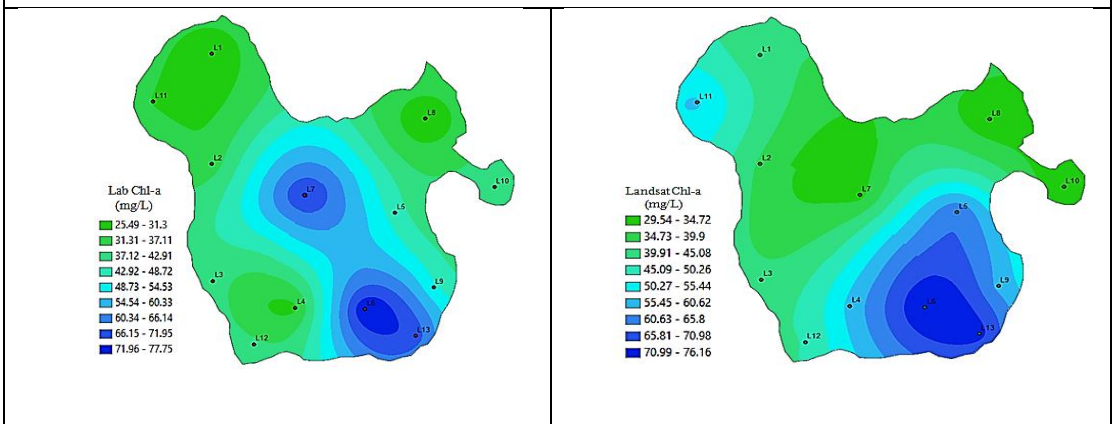
Sampling was done during the dry season and this means that there was decreased nutrient transport into the reservoir. However, Poddar et al. (2019) notes that phytoplankton blooms are expected during the dry season due to the favorable water temperatures and the existence of pre-deposited nutrients carried by the rivers during the rainy season. Consequently, high Chl-*a* is observed at these points close to where the rivers enter the reservoir. The EMRM approach used in this study and the resultant models could be used to predict and map turbidity, Chl-*a*, and TSS in Two Rivers Dam reservoir and other reservoirs in the catchment. The near-real time monitoring approach allowed for the analysis of the selected physical WQPs based on laboratory measurements and Rrs values from Landsat-8 OLI. The models are only applicable to Two Rivers Dam and reservoirs in the region. However, the model constants should be tailored based on the hydrological characteristics of the reservoirs and the periodic changes in hydrological and climatic conditions (Ouma et al., 2020).



a) Measured and estimated Chl-*a* on 25th November, 2020



b) Measured and estimated Chl-*a* on 12th December, 2020



c) Measured and estimated Chl-*a* on 28th January, 2021

Figure 4.6: Spatial distribution of measured and estimated Chl-*a*.

4.3 ANN Model 1 – Prediction of Treated Water Quality Parameters

ANN model 1 results related to treated water turbidity, TSS, and Chl-*a* are shown in Figure 4.7, with the related R^2 and Mean Squared Error (MSE) shown in Table 4.8.

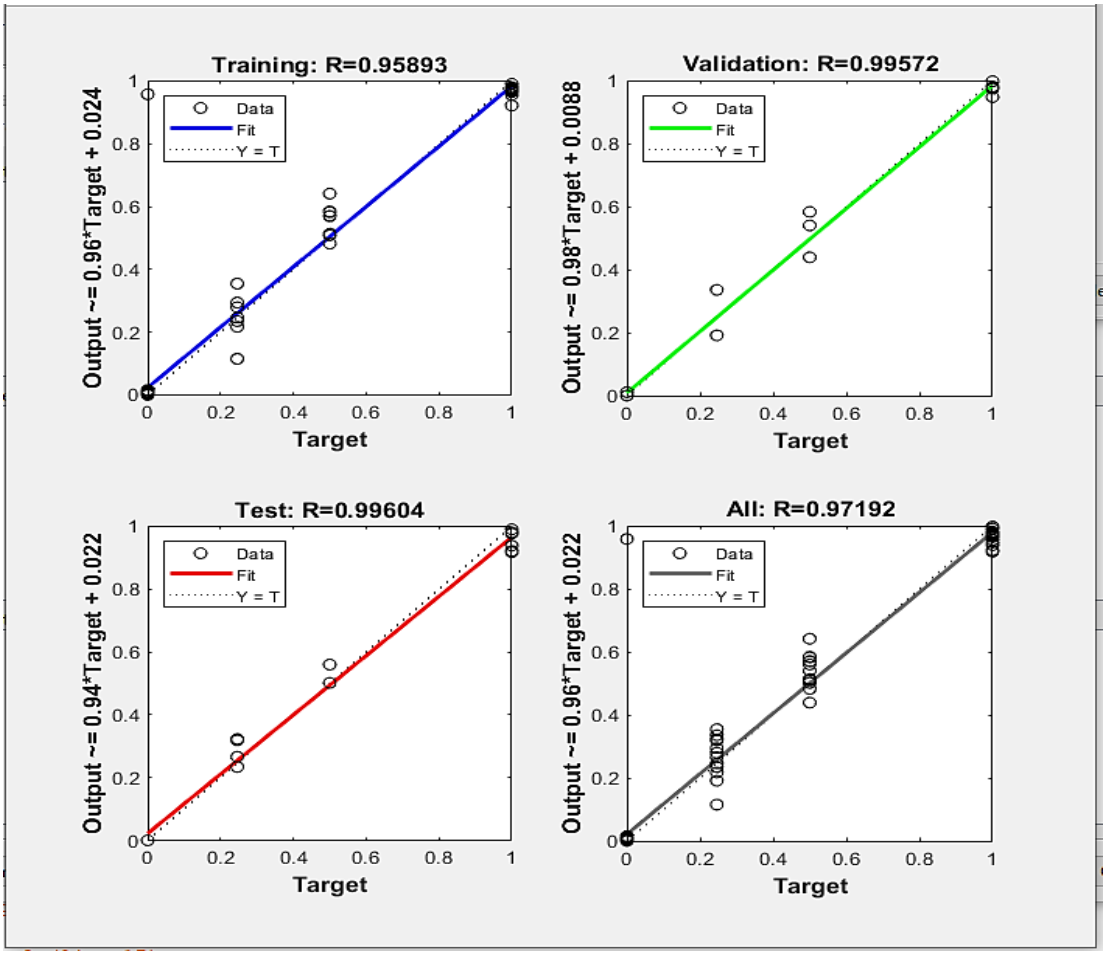


Figure 4.7: ANN Model 1 predictions of treated water quality parameters.

Table 4.8: ANN Model 1 results

Model 1 Output	**R^2**	**MSE**
Turbidity	0.9933	0.0026 NTU
TSS	1	0 mg/l
Chl-*a*	0.8665	0.0252 mg/l

The model has dependable accuracy in predicting the treated WQPs with R^2 values of 0.99, 1.0, and 0.87 for turbidity, TSS, and Chl-*a* respectively (Table 4.8). The MSE for the predictions are also low. That is errors of 0.0026, 0, and 0.0252 for turbidity, TSS, and Chl-*a* respectively. An ANN-MLP model used by Haghiri et al. (2018) to predict treated WQPs based on the raw water turbidity, pH, alkalinity, and alum dose. The

ANN model also attained reasonable accuracy in predicting the treated WQPs with R^2 values of 0.85 to 0.94. The MSE values are low with an average of 0.11NTU, 0.01pH, and 26.31mg/l for turbidity, pH, and alkalinity respectively. Setshedi et al. (2021) also developed an ANN-MLP model to predict WQPs in three rivers located in three municipalities in the Eastern Cape Province of South Africa. The accuracy of the model was tested by comparing the observed results to the predicted values. Just like in this scenario, the model exhibited a great understanding of the data sets and it also gave a good predictive accuracy with an R^2 of 0.99 and MSE of 39.07 (Setshedi et al., 2021).

Kennedy et al. (2015) developed four ANN models to predict treated WQPs, specifically dissolved organic matter and turbidity at the Akron Water Treatment Plant in Ohio, USA. Even though the models were limited to turbidity and dissolved organic matter, the results highlight the validity of applying ANN models in the prediction of treated WQPs as a function of the raw WQPs and the coagulant dose. Similarly, Seo et al. (2016) also developed an ANN model for forecasting eight WQPs (total nitrogen, total phosphorous, electric conductivity, Chl-*a*, dissolved oxygen, temperature, pH, and turbidity) downstream of Cheongpyeong Dam in Seoul city. The average R^2 values for seven of these parameters ranged between 0.881 to 0.998 with RMSE values ranging between 0.001 and 0. 360. The R^2 value for turbidity was slightly lower at 0.638 with RMSE of 3.208.

Unlike in the current study where the model was applied to a water treatment plant, the predictive capacity of the models by Setshedi et al. (2021), Kennedy et al. (2015), and Seo et al. (2016) was tested on dams and rivers. However, the accuracy with which the models predicted the WQPs shows the opportunities for using ANN as a predictive tool and the possibility of scaling up these models and using them for water quality predictions in water treatment plants.

4.4 ANN Model 2 – Prediction of Optimum Coagulant Dose

Results for optimal coagulant dose obtained by using model 2 are shown in Figure 4.8.

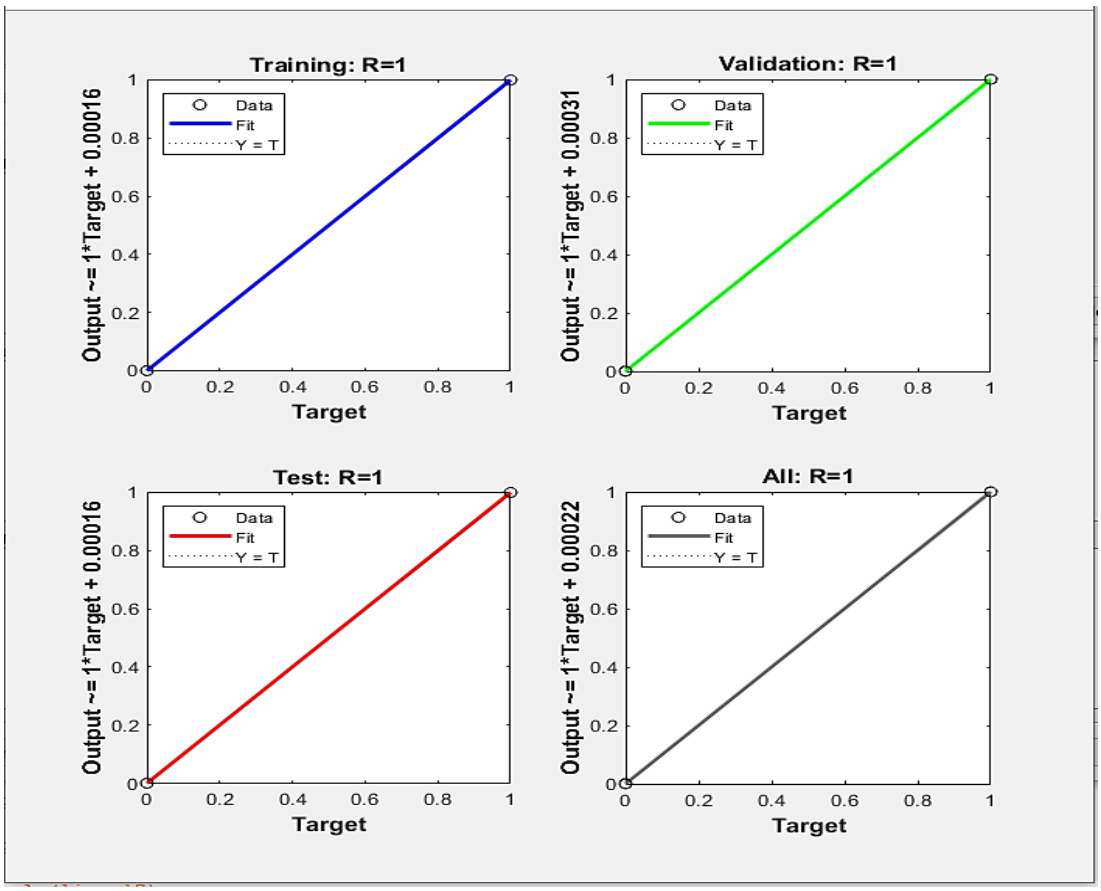


Figure 4.8: ANN Model 2 predictions of optimal coagulant dose.

The model also has a high accuracy in predicting the optimum coagulant dose required for treatment with an R^2 of 0.9987 and a MSE of 0 mg/l. A study by Kote and Wadkar (2019) used ANN to model chlorine and coagulant dose in a water treatment plant based on jar test experiments. The predicted coagulant doses were highly related to the actual doses with R^2 of 0.81 and MSE of 5.47g/m^3. The accuracy of the predictions indicates that ANN is capable of precisely modelling the coagulation process. This is also the case for the study done by Haghiri et al. (2018) which aimed at forecasting the optimum coagulant dose by using ANN to model jar test experiments. Just like in this study, Haghiri et al. (2018) developed an ANN-MLP with a single hidden layer to predict the

optimum coagulant dose for the water treatment process and the model also attained a reasonable accuracy with R^2 of 0.93 and MSE of 0.37. Sengul and Gormez (2013) also developed an ANN-MLP to determine the optimum coagulant dose in drinking water treatment. The model also achieved reasonable accuracies with R^2 value of 0.8 and RMSE of 12.51. Sengul and Gormez (2013) also worked with limited data just like in this study and they noted that the performance of the network is affected by the quality and completeness of the data. Regardless, the accuracies realized show the potential of ANN models in coagulant forecasting with the core focus being increasing the size of the dataset in order to improve the system's performance. The value of large datasets in increasing ANN model accuracies was demonstrated by Baouab and Cherif (2018) who used data from large databases of three water treatment plants in developing ANN models to predict the optimal coagulant dose. The large datasets resulted in high accuracies and it also showed the possibility of model transfer for use in similar water treatment plants.

Jayaweera and Aziz (2018) used a different ANN modelling approach by developing an extreme learning machine (ELM) single layer feed forward neural network, ELM-radial basis neural network, and an ANN-MLP neural network for prediction of the optimum coagulant dosage. All the models developed performed well with R^2 exceeding 0.97. This study only used the MLP neural network modelling approach but the accuracies obtained are highly comparable. Even though the ELM-radial basis neural network gave better results than the ANN-MLP model used in this study, the prediction accuracy obtained with the ANN-MLP model is still highly accurate. Thus, the ANN-MLP can be used to efficiently determine the coagulant dose required for treatment.

4.5 Summary

Models always carry a certain degree of error and this could limit the possibility of developing a completely predictive water treatment plant model. However, there are endless possibilities for using ANNs in water treatment by creating a system that runs in parallel with the main treatment plant processes. The results from this study highlights the usefulness of ANN models in simulating the processes that occur in jar test experiments depending on the training data range. Additionally, the values obtained are also within acceptable limits and this demonstrates the accuracy with which ANN models can be used to predict WQPs and the optimal coagulant dose. Therefore, the process models developed in this study are useful in understanding the coagulation process based on different factors that affect raw water quality. The process models can also be combined with operators' experience to enhance coagulant dosing practices at water treatment plants. An analysis of the existing errors indicate that the ANN models performed even when model development was done with limited data. Generally, the model accuracy could be increased if the modelling process leveraged on more data since ANN modelling is a parametric method. The ANN models for use in water treatment plants can be combined with periodic water quality data from Landsat to help improve the efficiency of the treatment process, and to optimize the coagulant dosing practices.

The study explored the possibility of a predictive water treatment plant model with the focus being from the inlet of the water treatment works to the final point of treatment. The knowledge gained is useful in informing decisions to develop a predictive water treatment plant model that starts from the reservoir. This may prove to be an effective tool to determine bottlenecks and make quick decisions to maximize the efficiency of treatment. The modelling aspect can also be combined with predictive models for

running the water treatment plants. Thus, combining the monitoring aspect using remote sensing will help to give a better estimate of the optimum coagulant dose based on the changes in the catchment.

CHAPTER 5: CONCLUSION AND RECOMMENDATIONS

5.1 Conclusion

The study aimed at evaluating the possibility of applying a smart monitoring approach using remote sensing and ANN to enhance the control of water treatment plant processes.

From the study results, the following conclusions were drawn:

1. A comparison of the extracted spectral reflectance values from Landsat-8 OLI satellite data and laboratory-derived water quality data of Chl-*a*, turbidity, and TSS established that Landsat-8 OLI availed an effective approach for reservoir water quality retrieval. For the retrieval of turbidity, the second-order polynomial regression models for dates 25/11/2020, 12/12/2020, and 28/01/2021 respectively gave the best results with R^2 values of 0.797, 0.757, 0.736. TSS was best estimated by logarithmic, exponential, and polynomial regression models with R^2 values of 0.808, 0.853, and 0.766 for the specific data collection dates. For the estimation of Chl-*a*, the Rrs values based on second order polynomial regression models performed best with R^2 values of 0.802, 0.682, 0.926 for the specific data collection dates. Furthermore, remote sensing could enable continuous WQM and/or management.
2. The spatial distribution maps developed to highlight the distribution and variability of the estimated and predicted WQPs for the selected period were comparable to a great extent as evidenced by the almost similar values obtained at most of the sampling points. The regression models developed based on the data collected for the entire sampling period estimated the reservoir turbidity with more than 70% accuracy. This means that the sampling data obtained at

specific locations either through sampling or from satellite images can be used to interpret and infer water quality predictions in the entire water body.

3. The ANN model developed for predicting treated WQPs based on the extracted spectral reflectance values from Landsat-8 OLI highlighted some of the possibilities for using ANNs in water treatment. The first ANN model developed to predict treated water quality parameters had a dependable accuracy with R^2 values of 0.99, 1.0, and 0.87 for turbidity, TSS, and Chl-*a* respectively. The MSE for the predictions were 0.0026 NTU, 0 mg/l (meaning that the model has no error in predicting TSS), and 0.0252 mg/l for turbidity, TSS, and Chl-*a* respectively. The ANN model developed showed the possibility of using extracted spectral reflectance values from satellite images for water quality predictions and optimizing water treatment plant operations.
4. The ANN model developed to determine the optimum coagulant dose required for water treatment had a high accuracy with R^2 of 0.9987 and a MSE of 0 mg/l. This showed that even with limited data, the MLP-ANN can be used for optimizing the coagulant dosing estimations in water treatment plants. Furthermore, the accuracies achieved also highlight the possibility of developing ANN that are adaptive to changes in raw water quality in water treatment plants, even though the monitoring and predictive capacity of ANNs in combination with Landsat 8 and 9 data is presently limited to 8 day intervals.

5.2 Recommendations

1. To increase the efficacy and dependability of Landsat-8 OLI in WQPs retrieval, model coefficients should be developed using *in situ* data from a significant number of sampling locations for varied seasons. Furthermore, combining Landsat-8 and Landsat-9 will also reduce the revisit time for data collection to

8 days. This means that the model constants can be improved even further since the combination of these two satellites could allow for inter-seasonal monitoring.

2. The empirical models developed are only applicable to Two Rivers Dam system and other reservoirs in the region. However, the development of model transfer functions should be done in line with the hydrological characteristics of the reservoirs and the seasonal changes in hydrological and climatic conditions to enable the developed models to be used for water quality predictions in other reservoirs within the catchment.
3. Given that the MLP-ANN model is a parametric method, then the accuracy of the model depends on the quality and quantity of the collected data. Therefore, more studies should be done with fairly large data sets collected at different seasons of the year in order to test the accuracy of the model with highly variable input data.
4. Models carry a certain margin of error and this makes it quite difficult to develop a completely predictive water treatment plant model. However, efforts should still be invested in developing completely predictive treatment plant models to run in parallel with the main treatment plant operations. This will help to understand the water treatment plant performance under different conditions thus helping to determine bottlenecks and influencing decision-making processes. Furthermore, the concept can also be used to maximize water treatment plants' efficiency.

REFERENCES

Adjovu, G. E., Stephen, H., James, D., & Ahmad, S. (2023). Measurement of Total Dissolved Solids and Total Suspended Solids in Water Systems: A Review of the Issues, Conventional, and Remote Sensing Techniques. *Remote Sensing*, *15*(14), 3534.

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), 825-833.

Ahmed, A. N., Othman, F. B., Afan, H. A., Ibrahim, R. K., Fai, C. M., Hossain, M. S., ... & Elshafie, A. (2019). Machine learning methods for better water quality prediction. *Journal of Hydrology*, *578*, 124084.

Amin, E., Belda, S., Pipia, L., Szantoi, Z., El Baroudy, A., Moreno, J., & Verrelst, J. (2022). Multi-Season Phenology Mapping of Nile Delta Croplands Using Time Series of Sentinel-2 and Landsat 8 Green LAI. *Remote sensing*, *14*(8), 1812.

APHA (American Public Health Association). (1975). *Standard Methods for the Examination of Water and Wastewater.* 14th Edition. American Public Health Association. Washington, D.C.

Baxter, C. W., Stanley, S. J., & Zhang, Q. (1999). Development of a full-scale artificial neural network model for the removal of natural organic matter by enhanced coagulation. *Journal of Water Supply: Research and Technology—AQUA*, *48*(4), 129-136.

Baouab, M. H., & Cherif, S. (2018). Prediction of the optimal dose of coagulant for various potable water treatment processes through artificial neural network. *Journal of Hydroinformatics*, *20*(6), 1215-1226.

Behmel, S., Damour, M., Ludwig, R., & Rodriguez, M. J. (2016). Water quality monitoring strategies—A review and future perspectives. *Science of the Total Environment*, *571*, 1312-1329.

Benson, D., Gain, A. K., & Giupponi, C. (2020). Moving beyond water centricity? Conceptualizing integrated water resources management for implementing sustainable development goals. *Sustainability science*, *15*(2), 671-681.

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

Black, M. (2016). *The atlas of water: mapping the World's most critical resource*. Univ of California Press.

Bratby, J. (2016). *Coagulation and flocculation in water and wastewater treatment*. IWA publishing.

Bresciani, M., Giardino, C., Stroppiana, D., Dessena, M. A., Buscarinu, P., Cabras, L., ...&Tzimas, A. (2019). Monitoring water quality in two dammed reservoirs from multispectral satellite data. *European Journal of Remote Sensing*, 1-10.

Brooks, B. W., Lazorchak, J. M., Howard, M. D., Johnson, M. V. V., Morton, S. L., Perkins, D. A., ... & Steevens, J. A. (2016). Are harmful algal blooms becoming the greatest inland water quality threat to public health and aquatic ecosystems?. *Environmental toxicology and chemistry*, *35*(1), 6-13.

Chen, Y., Song, L., Liu, Y., Yang, L., & Li, D. (2020). A review of the artificial neural network models for water quality prediction. *Applied Sciences*, *10*(17), 5776.

Chiappini, F. A., Teglia, C. M., Forno, Á. G., & Goicoechea, H. C. (2020). Modelling of bioprocess non-linear fluorescence data for at-line prediction of etanercept based on artificial neural networks optimized by response surface methodology. *Talanta*, *210*, 120664.

Dawood, T., Elwakil, E., Novoa, H. M., & Delgado, J. F. G. (2021). Toward urban sustainability and clean potable water: Prediction of water quality via artificial neural networks. *Journal of Cleaner Production*, *291*, 125266.

Da Silva, I. N., Hernane Spatti, D., Andrade Flauzino, R., Liboni, L. H. B., dos Reis Alves, S. F., da Silva, I. N., ... & dos Reis Alves, S. F. (2017). *Artificial neural network architectures and training processes* (pp. 21-28). Springer International Publishing.

Duan, W., Takara, K., He, B., Luo, P., Nover, D., &Yamashiki, Y. (2013). Spatial and temporal trends in estimates of nutrient and suspended sediment loads in the Ishikari River, Japan, 1985 to 2010. *Science of the Total Environment*, *461*, 499-508.

Emelko, M. B., Silins, U., Bladon, K. D., & Stone, M. (2011). Implications of land disturbance on drinking water treatability in a changing climate: Demonstrating the need for "source water supply and protection" strategies. *Water research*, *45*(2), 461-472.

Farrell, C., Hassard, F., Jefferson, B., Leziart, T., Nocker, A., & Jarvis, P. (2018). Turbidity composition and the relationship with microbial attachment and UV inactivation efficacy. *Science of the total environment*, *624*, 638-647.

Garaba, S., Friedrichs, A., Voß, D., & Zielinski, O. (2015). Classifying natural waters with the Forel-UleColour index system: results, applications, correlations and crowdsourcing. *International journal of environmental research and public health*, *12*(12), 16096-16109.

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

González-Márquez, L. C., Torres-Bejarano, F. M., Rodríguez-Cuevas, C., Torregroza-Espinosa, A. C., & Sandoval-Romero, J. A. (2018). Estimation of water quality parameters using Landsat 8 images: application to Playa Colorada Bay, Sinaloa, Mexico. *Applied Geomatics*, *10*(2), 147-158.

Gurney, K. (2018). *An introduction to neural networks*. CRC press.

Haghiri, S., Daghighi, A., &Moharramzadeh, S. (2018). Optimum coagulant forecasting by modeling jar test experiments using ANNs. *Drinking Water Engineering and Science*, *11*(1), 1-8.

Hamed, M. M., Khalafallah, M. G., &Hassanien, E. A. (2004). Prediction of wastewater treatment plant performance using artificial neural networks. *Environmental Modelling& Software*, *19*(10), 919-928.

Hayou, S., Doucet, A., & Rousseau, J. (2018). On the selection of initialization and activation function for deep neural networks. *arXiv preprint arXiv:1805.08266*.

Jaelani, L. M., Limehuwey, R., Kurniadin, N., Pamungkas, A., Koenhardono, E. S., & Sulisetyono, A. (2016). Estimation of Total Suspended Sediment and Chlorophyll-A Concentration from Landsat 8-Oli: The Effect of Atmospher and Retrieval Algorithm. IPTEK The Journal for Technology and Science, 27(1).

Jalali, S. M. J., Ahmadian, S., Kebria, P. M., Khosravi, A., Lim, C. P., & Nahavandi, S. (2019). Evolving artificial neural networks using butterfly optimization algorithm for data classification. In *Neural Information Processing: 26th International Conference, ICONIP 2019, Sydney, NSW, Australia, December 12–15, 2019, Proceedings, Part I 26* (pp. 596-607). Springer International Publishing.

Jayaweera, C. D., & Aziz, N. (2018, November). Development and comparison of extreme learning machine and multi-layer perceptron neural network models for predicting optimum coagulant dosage for water treatment. In *Journal of Physics: Conference Series* (Vol. 1123, No. 1, p. 012032). IOP Publishing.

Jorgensen, S. E., & Gromiec, M. J. (Eds.). (2016). *Developments in environmental modelling*. Elsevier.

Kahani, M., Ahmadi, M. H., Tatar, A., & Sadeghzadeh, M. (2018). Development of multilayer perceptron artificial neural network (MLP-ANN) and least square support vector machine (LSSVM) models to predict Nusselt number and pressure drop of TiO2/water nanofluid flows through non-straight pathways. *Numerical Heat Transfer, Part A: Applications*, *74*(4), 1190-1206.

Kalele, A. S. (2019). *Estimation and Mapping of Turbidity in the Lower Charles River Using Landsat 8 OLI Satellite Imagery* (Doctoral dissertation, Northeastern University).

KDHE (Kansas Department of Health and Environment). (2011). Water Quality Standards White Paper: Chlorophyll-a Criteria for Public Water Supply Lakes or Reservoirs.

Keith, D., Rover, J., Green, J., Zalewsky, B., Charpentier, M., Thursby, G., & Bishop, J. (2018). Monitoring algal blooms in drinking water reservoirs using the Landsat-8 Operational Land Imager. *International journal of remote sensing*, *39*(9), 2818-2846.

Kennedy, M. J., Gandomi, A. H., & Miller, C. M. (2015). Coagulation modeling using artificial neural networks to predict both turbidity and DOM-PARAFAC component removal. *Journal of Environmental Chemical Engineering*, *3*(4), 2829-2838.

Kenya National Bureau of Statistics. (2019). 2019 Kenya population and housing census results.

Kibii, J. K., Kipkorir, E. C., & Kosgei, J. R. (2021). Application of Soil and Water Assessment Tool (SWAT) to evaluate the impact of land use and climate variability on the Kaptagat catchment river discharge. *Sustainability*, *13*(4), 1802.

Kimutai, J. C., Evans, W. O., & Ekai, P. (2018). Assessment of water shortages and coping measures at household level in the informal settlements of Eldoret Municipality, Uasin Gishu County, Kenya. *IOSR Journal of Environmental Science, Toxicology and Food Technology (IOSR-JESTFT)*, *12*(3), 57-71.

Knight, G. L. (2017). Kansas Lake and Wetland Monitoring Program 2017 Annual Report.

Kote, A. S., &Wadkar, D. V. (2019). Modeling of Chlorine and Coagulant Dose in a Water Treatment Plant by Artificial Neural Networks. *Engineering, Technology & Applied Science Research*, *9*(3), 4176-4181.

Lai, Y., Zhang, J., Song, Y., & Gong, Z. (2021). Retrieval and Evaluation of Chlorophyll-a Concentration in Reservoirs with Main Water Supply Function in Beijing, China, Based on Landsat Satellite Images. *International Journal of Environmental Research and Public Health*, *18*(9), 4419.

León-Luque, A. J., Barajas, C. L., & Peña-Guzmán, C. A. (2016). Determination of the optimal dosage of Aluminum Sulfate in the coagulation-flocculation process using an Artificial Neural Network. *International Journal of Environmental Science and Development*, *7*(5), 346-350.

Li, L., Rong, S., Wang, R., & Yu, S. (2021). Recent advances in artificial intelligence and machine learning for nonlinear relationship analysis and process control in drinking water treatment: A review. *Chemical Engineering Journal*, *405*, 126673.

Lintern, A., Webb, J. A., Ryu, D., Liu, S., Waters, D., Leahy, P., ... & Western, A. W. (2018). What are the key catchment characteristics affecting spatial differences in riverine water quality?. *Water Resources Research*, *54*(10), 7252-7272.

Liu, X., & Georgakakos, A. P. (2021). Chlorophyll a estimation in lakes using multi-parameter sonde data. *Water Research*, *205*, 117661.

Lotfi, G., Ahmadi Nadoushan, M., & Abolhasani, M. (2019). The feasibility of using Landsat OLI images for water turbidity estimation in Gandoman wetland, Iran. *Journal of Radar and Optical Remote Sensing*, *2*(2), 49-62.

Mehdinejad, M. H., Bina, B., & Hadian, S. (2012). The survey of removal of suspended solids from river at flooding period by plain sedimentation process. *Advances in Environmental Biology*, *6*(1), 358-361.

Meng, H., Zhang, J., & Zheng, Z. (2022). Retrieving Inland Reservoir Water Quality Parameters Using Landsat 8-9 OLI and Sentinel-2 MSI Sensors with Empirical Multivariate Regression. *International Journal of Environmental Research and Public Health*, *19*(13), 7725.

Michalopoulos, S. (2019). *Can space technologies improve drinking water quality?*. [online] euractiv.com. Available at: https://www.euractiv.com/section/agriculture-food/special_report/can-space-technologies-improve-drinking-water-quality/ [Accessed 16 May 2019].

Meng, H., Zhang, J., & Zheng, Z. (2022). Retrieving Inland Reservoir Water Quality Parameters Using Landsat 8-9 OLI and Sentinel-2 MSI Sensors with Empirical Multivariate Regression. *International Journal of Environmental Research and Public Health*, *19*(13), 7725.

Mohamed, Z. E. (2019). Using the artificial neural networks for prediction and validating solar radiation. *Journal of the Egyptian Mathematical Society*, *27*(1), 1-13.

Mustafa, S. M. T. Quantification of model uncertainties in groundwater drought simulations under climate change.

Mustafa, H. M., Mustapha, A., Hayder, G., & Salisu, A. (2021, January). Applications of iot and artificial intelligence in water quality monitoring and prediction: A review. In *2021 6th international conference on inventive computation technologies (ICICT)* (pp. 968-975). IEEE.

Mwaniki, S. (2018). World Bank Funds More Projects. *Africa Water & Sanitation & Hygiene*, (6). Retrieved from https://afriwater.org/articles/346-world-bank-funds-more-projects

Murphy, R. R., Curriero, F. C., & Ball, W. P. (2010). Comparison of spatial interpolation methods for water quality evaluation in the Chesapeake Bay. *Journal of Environmental Engineering*, *136*(2), 160-171.

Nafsin, N., & Li, J. (2022). Prediction of 5-day biochemical oxygen demand in the Buriganga River of Bangladesh using novel hybrid machine learning algorithms. *Water Environment Research*, *94*(5), e10718.

Najafzadeh, M., Homaei, F., & Farhadi, H. (2021). Reliability assessment of water quality index based on guidelines of national sanitation foundation in natural streams: integration of remote sensing and data-driven models. *Artificial Intelligence Review*, *54*(6), 4619-4651.

Najafzadeh, M., Ghaemi, A., & Emamgholizadeh, S. (2018). Prediction of water quality parameters using evolutionary computing-based formulations. *International Journal of Environmental Science and Technology*, *16*(10), 6377-6396.

Oostwal, E., Straat, M., & Biehl, M. (2021). Hidden unit specialization in layered neural networks: ReLU vs. sigmoidal activation. *Physica A: Statistical Mechanics and its Applications*, *564*, 125517.

Ouabo, R. E., Sangodoyin, A. Y., & Ogundiran, M. B. (2020). Assessment of ordinary Kriging and inverse distance weighting methods for modeling chromium and cadmium soil pollution in E-waste sites in Douala, Cameroon. *Journal of Health and Pollution*, *10*(26).

Ouma, Y. O., Noor, K., & Herbert, K. (2020). Modelling reservoir chlorophyll-a, TSS, and turbidity using Sentinel-2A MSI and Landsat-8 OLI satellite sensors with empirical multivariate regression. *Journal of Sensors*, *2020*.

Ouma, Y. O., Owiti, T., Kipkorir, E., Kibiiy, J., & Tateishi, R. (2012). Multitemporal comparative analysis of TRMM-3B42 satellite-estimated rainfall with surface gauge data at basin scales: daily, decadal and monthly evaluations. *International Journal of Remote Sensing*, *33*(24), 7662-7684.

Ouma, Y. O., Okuku, C. O., & Njau, E. N. (2020). Use of artificial neural networks and multiple linear regression model for the prediction of dissolved oxygen in rivers: case study of hydrographic basin of River Nyando, Kenya. *Complexity*, *2020*.

Ouma, Y. O., Waga, J., Okech, M., Lavisa, O., & Mbuthia, D. (2018). Estimation of reservoir bio-optical water quality parameters using smartphone sensor apps and Landsat ETM+: review and comparative experimental results. *Journal of Sensors*, *2018*.

Papenfus, M., Schaeffer, B., Pollard, A. I., & Loftin, K. (2020). Exploring the potential value of satellite remote sensing to monitor chlorophyll-a for US lakes and reservoirs. *Environmental Monitoring and Assessment*, *192*(12), 808.

Patra, P., Dubey, S. K., Trivedi, R. K., Sahu, S. K., & Rout, S. K. (2016). Estimation of chlorophyll-a concentration and trophic states for an inland lake from Landsat 8 OLI data: a case of Nalban Lake of East Kolkata Wetland, India.

Pier, B., & Mateo, S. (2020). Continuous Water-Quality and Suspended-Sediment Transport Monitoring in the San Francisco Bay, California, Water Years 2016–17.

Pizani, F. M., Maillard, P., Ferreira, A. F., & de Amorim, C. C. (2020). Estimation of Water Quality in A Reservoir from Sentinel-2 MSI And Landsat-8 OLI Sensors. *ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences*, *5*(3).

Poonam, T., Tanushree, B., & Sukalyan, C. (2013). Water quality indices-important tools for water quality assessment: a review. *International Journal of Advances in chemistry*, *1*(1), 15-28.

Poddar, S., Chacko, N., & Swain, D. (2019). Estimation of Chlorophyll-a in northern coastal Bay of Bengal using Landsat-8 OLI and Sentinel-2 MSI sensors. *Frontiers in Marine Science*, *6*, 598.

Pratiwi, H., Windarto, A. P., Susliansyah, S., Aria, R. R., Susilowati, S., Rahayu, L. K., ... & Rahadjeng, I. R. (2020, February). Sigmoid activation function in selecting the best model of artificial neural networks. In *Journal of Physics: Conference Series* (Vol. 1471, No. 1, p. 012010). IOP Publishing.

Rahmanian, N., Ali, S. H. B., Homayoonfard, M., Ali, N. J., Rehan, M., Sadef, Y., & Nizami, A. S. (2015). Analysis of physiochemical parameters to evaluate the drinking water quality in the State of Perak, Malaysia. *Journal of Chemistry*, *2015*.

Ribeiro, H. M. C., da Costa Almeida, A., Rocha, B. R. P. D., &Krusche, A. V. (2008). Water quality monitoring in large reservoirs using remote sensing and neural networks. *IEEE Latin America Transactions*, *6*(5), 419-423.

SCOR-UNESCO, W. (1966). Determination of photosynthetic pigments. *Determination of Photosynthetic Pigments in Sea-water*, 9-18.

Sengul, A. B., & Gormez, Z. (2013). Prediction of optimal coagulant dosage in drinking water treatment by artificial neural network. In *1st EWAS-MED International Conference: Improving Efficiency of Water Systems in a Changing Natural and Financial Environment*.

Seo, I. W., Yun, S. H., & Choi, S. Y. (2016). Forecasting water quality parameters by ANN model using pre-processing technique at the downstream of Cheongpyeong Dam. *Procedia Engineering*, *154*, 1110-1115.

Setshedi, K. J., Mutingwende, N., & Ngqwala, N. P. (2021). The use of artificial neural networks to predict the physicochemical characteristics of water quality in three district municipalities, Eastern Cape Province, South Africa. *International Journal of Environmental Research and Public Health*, *18*(10), 5248.

Silveira Kupssinskü, L., Thomassim Guimarães, T., Menezes de Souza, E., C. Zanotta, D., Roberto Veronez, M., Gonzaga Jr, L., & Mauad, F. F. (2020). A method for chlorophyll-a and suspended solids prediction through remote sensing and machine learning. *Sensors*, *20*(7), 2125.

Singh, S. K., Jain, S. K., & Bárdossy, A. (2014). Training of artificial neural networks using information-rich data. *Hydrology*, *1*(1), 40-62.

Sumi, S. M., Zaman, M. F., & Hirose, H. (2012). A rainfall forecasting method using machine learning models and its application to the Fukuoka city case. *International Journal of Applied Mathematics and Computer Science*, *22*(4), 841-854.

Torbick, N., Ziniti, B., Stommel, E., Linder, E., Andrew, A., Caller, T., ...& Shi, X. (2018). Assessing cyanobacterial harmful algal blooms as risk factors for amyotrophic lateral sclerosis. *Neurotoxicity research*, *33*(1), 199-212.

U.S. Geological Survey, 2018, Lakes and reservoirs—Guidelines for study design and sampling: U.S. Geological Survey Techniques and Methods, book 9, chap. A10, 48 p., https://doi.org/10.3133/tm9a10.[Supersedes USGS Techniques of Water-Resources Investigations, book 9, chap. A10, version 1.0]

USGS EROS Archive - Landsat Archives - Landsat 8 OLI (Operational Land Imager) and TIRS (Thermal Infrared Sensor) Level-1 Data Products | U.S. Geological Survey. Usgs.gov. (2018). Retrieved 18 July 2022, from https://www.usgs.gov/centers/eros/science/usgs-eros-archive-landsat-archives-landsat-8-oli-operational-land-imager-and

Valentin, N. and Denœux, T. (2001).A neural network-based software sensor for coagulation control in a water treatment plant. *Intelligent Data Analysis*, 5(1), pp.23-39.

WASREB (Water Services Regulatory Board). (2008). Guidelines on drinking water quality and effluent monitoring.

Watson, S. B., Whitton, B. A., Higgins, S. N., Paerl, H. W., Brooks, B. W., & Wehr, J. D. (2015). Harmful algal blooms. In *Freshwater Algae of North America* (pp. 873-920). Academic Press.

Wernand, M. R., van der Woerd, H. J., &Gieskes, W. W. (2013). Trends in ocean colour and chlorophyll concentration from 1889 to 2000, worldwide. *PLOS one*, *8*(6), e63766.

Wilde, F. D., Radtke, D. B., Gibs, J., & Iwatsubo, R. T. (2005). National field manual for the collection of water-quality data. *US Geological Survey Techniques in Water-Resources Investigations, Book*, *9*.

World Health Organization. (2017). *Water Quality and Health Review of Turbidity: Information for regulators and water suppliers*. World Health Organization. Retrieved from https://apps.who.int/iris/bitstream/handle/10665/254631/WHO-FWC-WSH-17.01-eng.pdf

Wurtsbaugh, W. A., Paerl, H. W., & Dodds, W. K. (2019). Nutrients, eutrophication and harmful algal blooms along the freshwater to marine continuum. *Wiley Interdisciplinary Reviews: Water*, *6*(5), e1373.

Xie, H., Luo, X., Xu, X., Pan, H., & Tong, X. (2016). Automated subpixel surface water mapping from heterogeneous urban environments using Landsat 8 OLI imagery. *Remote sensing*, *8*(7), 584.

Xiong, L., Deng, R., Li, J., Liu, X., Qin, Y., Liang, Y., & Liu, Y. (2018).Subpixel surface water extraction (SSWE) using Landsat 8 OLI data. *Water*, *10*(5), 653.

Yanti, A., Susilo, B., & Wicaksono, P. (2016, November). The aplication of Landsat 8 OLI for total suspended solid (TSS) mapping in Gajahmungkur reservoir Wonogiri regency 2016. In IOP Conference Series: Earth and Environmental Science (Vol. 47, No. 1, p. 012028). IOP Publishing.

Zhang Q, Stanley SJ (1999) Real-time water treatment process control with artificial neural networks. J Environ Eng 125(2):153–160.

Zhang, Y., Yao, X., Wu, Q., Huang, Y., Zhou, Z., Yang, J., & Liu, X. (2021). Turbidity prediction of lake-type raw water using random forest model based on meteorological data: A case study of Tai lake, China. *Journal of Environmental Management*, *290*, 112657.

APPENDICES

Appendix A: Raw Landsat-8 images used for the study

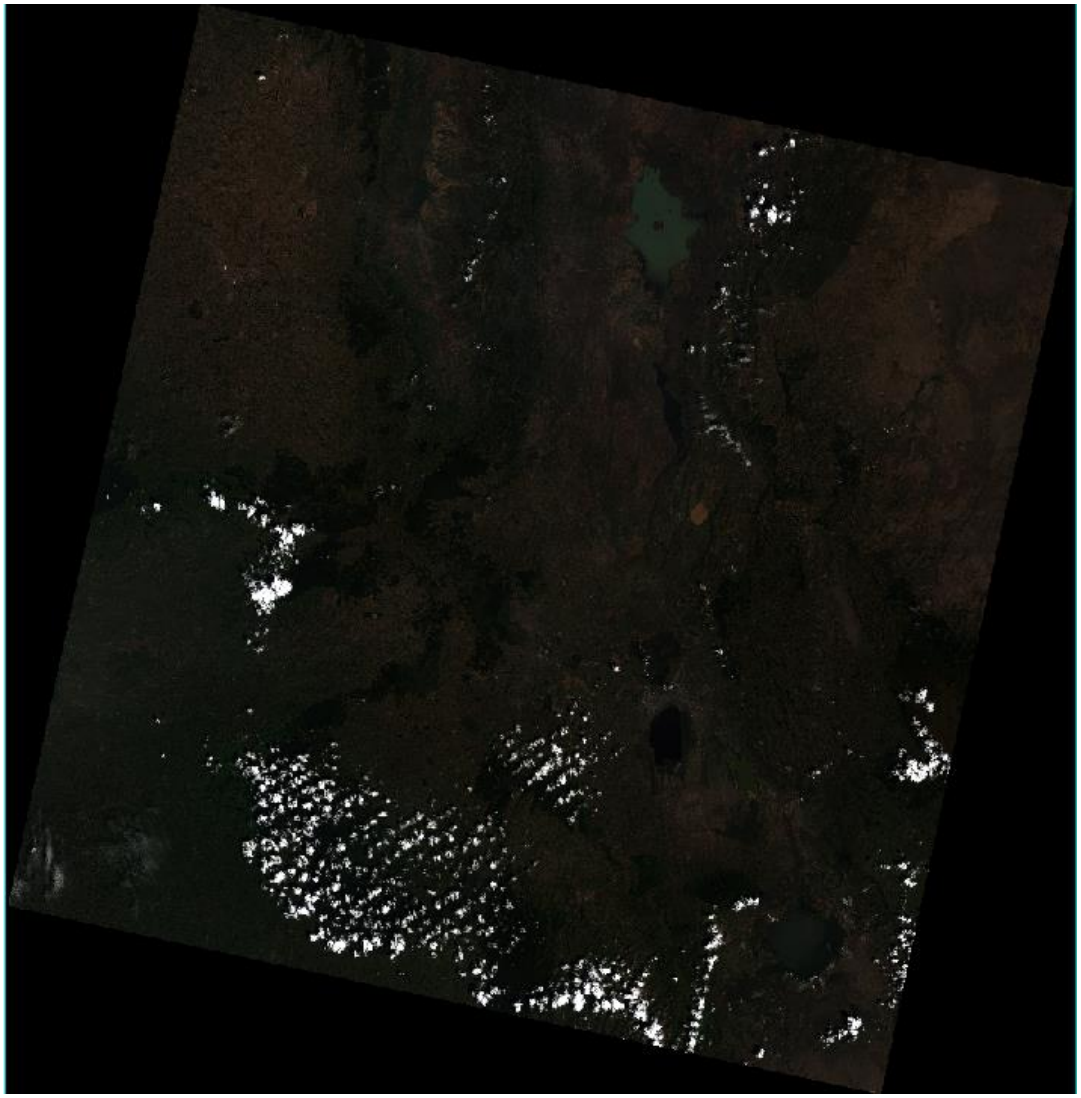


Image acquired on 25/11/12/2020

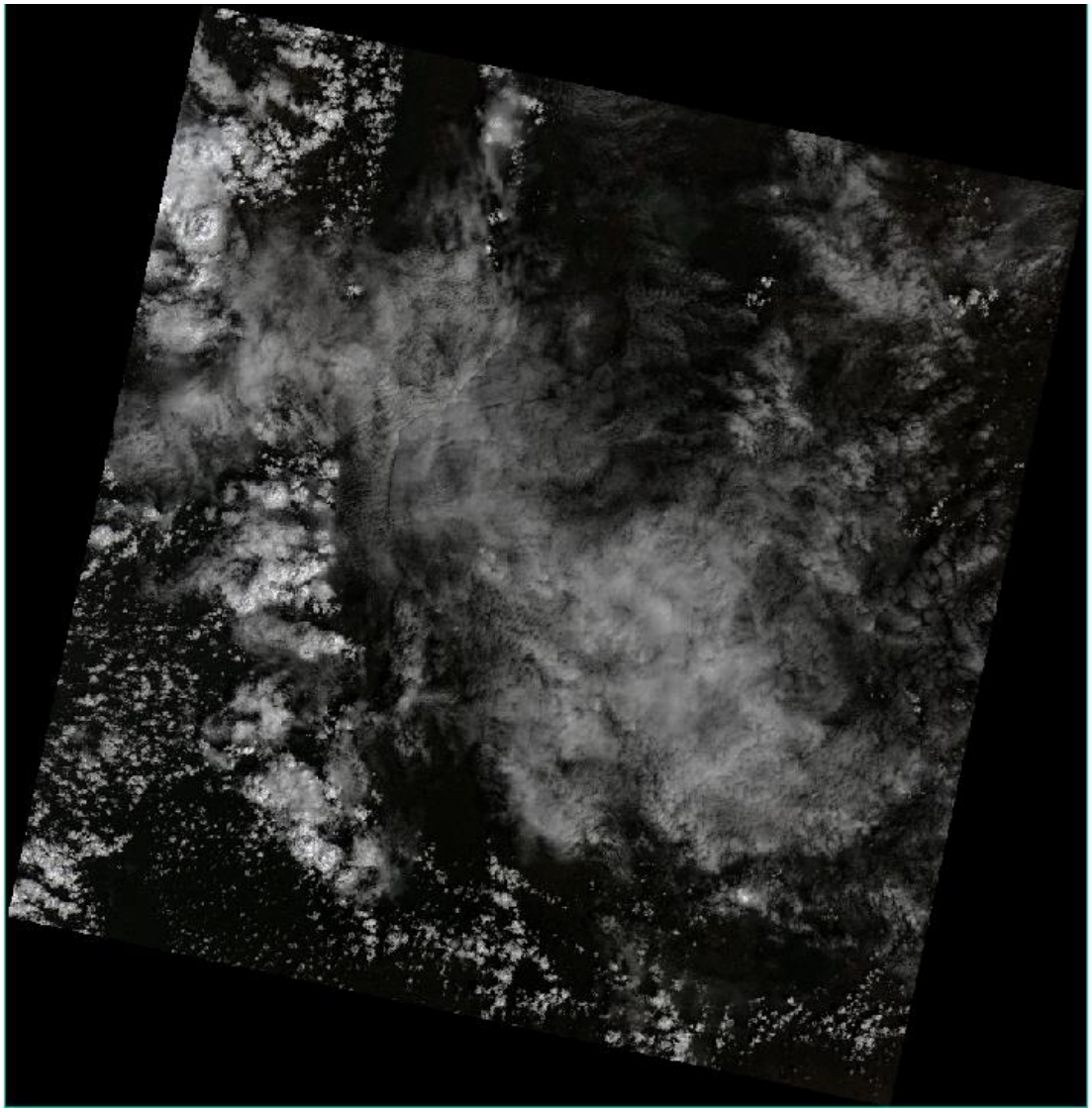


Image acquired on 11/12/2020

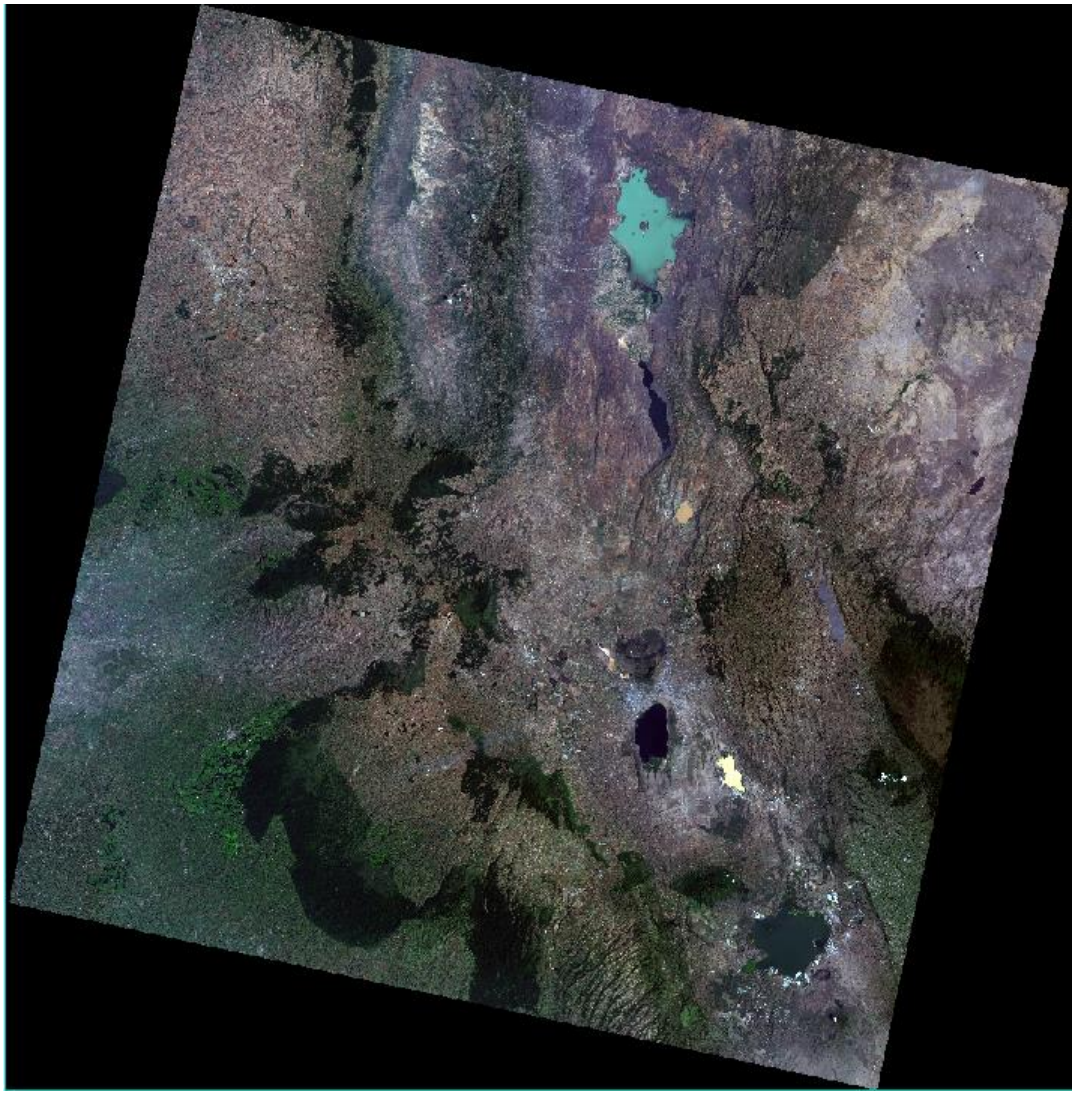


Image acquired on 28/01/2021

Appendix B: Surface Reflectance and Top of Atmosphere (TOA) Reflectance values

The TOA reflectance values are obtained from radiometric calibration in the FLAASH model which converts the digital numbers obtained from satellite images to TOA reflectance values.

TOA Reflectance (25/11/2020)							
Sampling Point	**Band 1**	**Band 2**	**Band 3**	**Band 4**	**Band 5**	**Band 6**	**Band 7**
L1	0.6546	0.4974	0.5049	0.4772	0.5288	0.4091	0.3495
L2	0.6192	0.4663	0.4876	0.4352	0.5034	0.4129	0.3439
L3	0.6462	0.4928	0.5013	0.473	0.5288	0.4144	0.3552
L4	0.6821	0.5286	0.5197	0.494	0.5467	0.4288	0.3694
L5	0.6716	0.5145	0.5188	0.4893	0.5418	0.4375	0.3818
L6	0.6741	0.5142	0.5364	0.5007	0.5466	0.4559	0.3972
L7	0.5966	0.4477	0.4688	0.4354	0.4947	0.3991	0.3404
L8	0.5618	0.4204	0.4224	0.398	0.4723	0.3695	0.3204
L9	0.6525	0.4945	0.5177	0.4815	0.5308	0.4438	0.3871
L10	0.6402	0.4975	0.4747	0.4538	0.5244	0.4124	0.3643
L11	0.6228	0.476	0.4869	0.4551	0.5115	0.4168	0.3526
L12	0.6817	0.5259	0.5313	0.4997	0.5508	0.4398	0.3788
L13	0.6455	0.4945	0.4951	0.4646	0.5226	0.4212	0.3616
TOA Reflectance (11/12/2020)							
Sampling Point	**Band 1**	**Band 2**	**Band 3**	**Band 4**	**Band 5**	**Band 6**	**Band 7**
L1	0.1786	0.0826	0.1124	0.085	0.0556	0.0386	0.0249
L2	0.1803	0.0843	0.1133	0.0854	0.0321	0.0169	0.0119
L3	0.1804	0.0849	0.114	0.0835	0.0343	0.0188	0.0132
L4	0.1813	0.0854	0.1158	0.0893	0.0317	0.0145	0.0102
L5	0.1805	0.0852	0.1165	0.0904	0.0315	0.0147	0.0099
L6	0.1805	0.085	0.1157	0.0893	0.0314	0.0141	0.0095

L7	0.1798	0.0845	0.115	0.0864	0.0296	0.0138	0.0097
L8	0.1788	0.0844	0.1158	0.0847	0.0316	0.016	0.0109
L9	0.1788	0.0841	0.1155	0.0885	0.0385	0.0252	0.0156
L10	0.1749	0.0776	0.0985	0.0671	0.0846	0.0763	0.0422
L11	0.189	0.0932	0.1145	0.0971	0.1669	0.1645	0.1085
L12	0.1797	0.0855	0.1163	0.0892	0.0336	0.0155	0.0108
L13	0.1806	0.0852	0.1163	0.0894	0.0331	0.0154	0.0105
TOA Reflectance (28/01/2021)							
Sampling Point	**Band 1**	**Band 2**	**Band 3**	**Band 4**	**Band 5**	**Band 6**	**Band 7**
L1	0.1511	0.0459	0.0838	0.0655	0.0537	0.0273	0.0183
L2	0.1546	0.05	0.0881	0.0687	0.0274	0.0152	0.0117
L3	0.1516	0.048	0.0874	0.0682	0.0284	0.0151	0.0109
L4	0.1543	0.0494	0.0899	0.071	0.0245	0.0106	0.007
L5	0.1542	0.0502	0.093	0.0742	0.0245	0.0093	0.0071
L6	0.1544	0.0506	0.0931	0.0742	0.0242	0.01	0.0068
L7	0.1529	0.0483	0.088	0.0679	0.0242	0.011	0.0078
L8	0.1524	0.0475	0.0874	0.0674	0.0257	0.0116	0.0077
L9	0.1537	0.0498	0.0931	0.0741	0.0402	0.0182	0.0116
L10	0.1485	0.0445	0.0867	0.0656	0.1094	0.0567	0.0338
L11	0.1592	0.0557	0.1025	0.0973	0.1603	0.1739	0.1259
L12	0.1524	0.0484	0.0881	0.069	0.0258	0.0115	0.0078
L13	0.1539	0.0499	0.0928	0.0734	0.0265	0.0113	0.0073

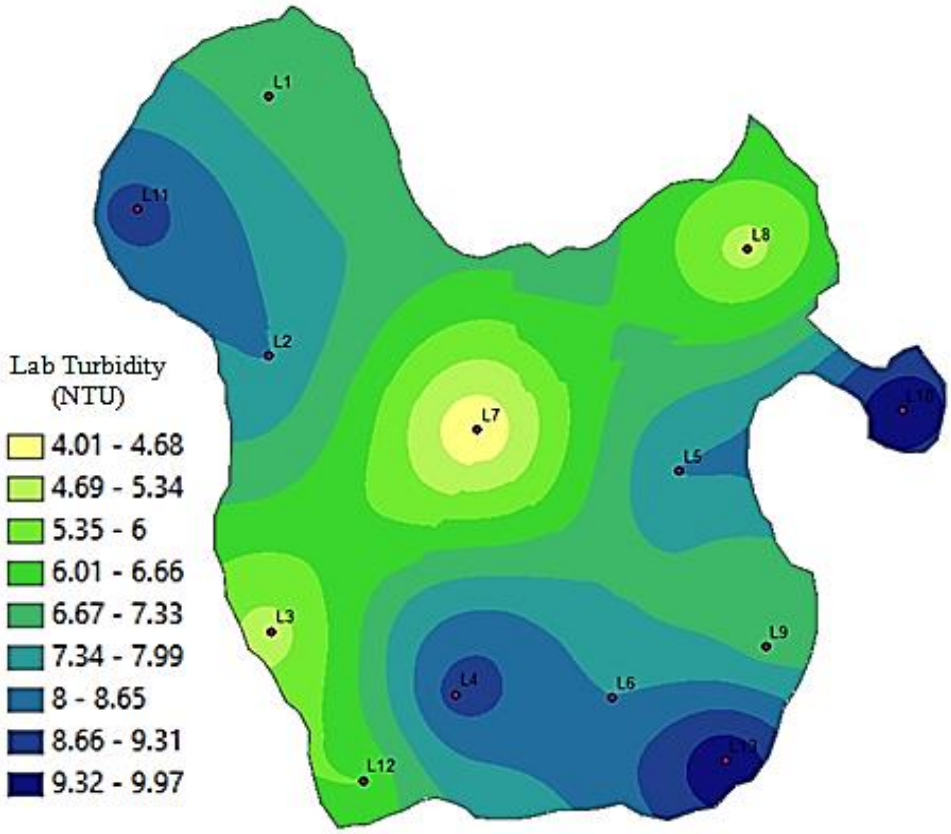
Appendix C: Remote sensing reflectance (Rrs)

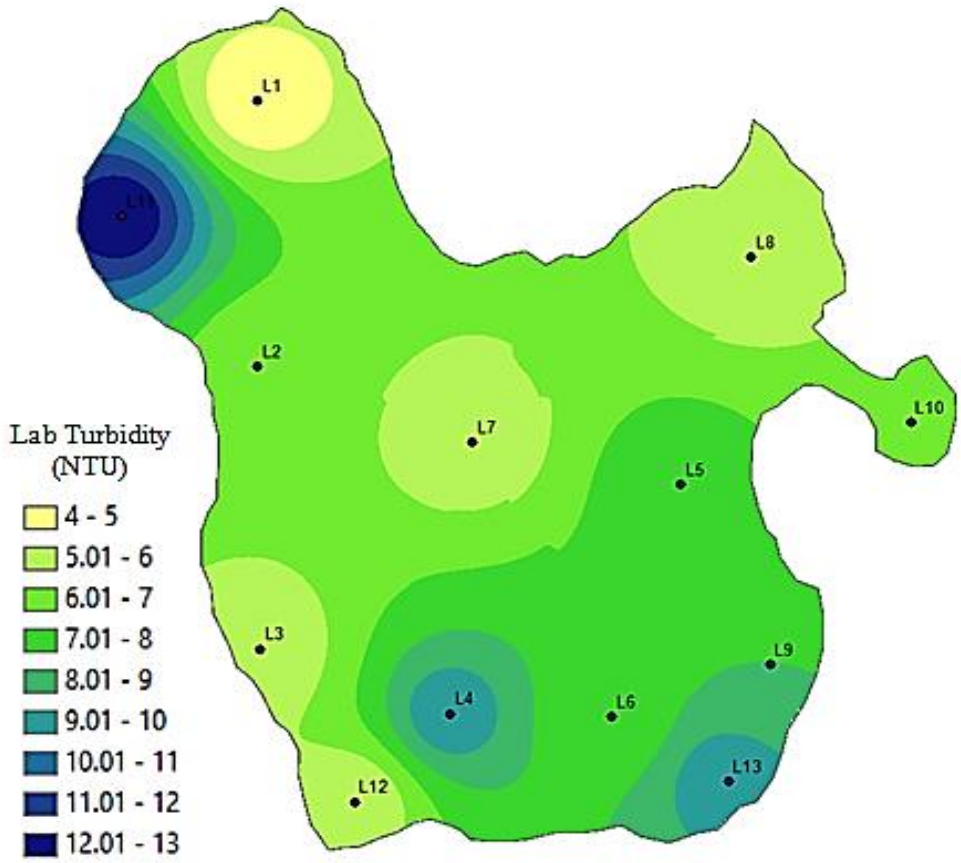
The Rrs is obtained after dark object subtraction which converts TOA reflectance to surface reflectance. The surface reflectance values are then divided by π to obtain Rrs.

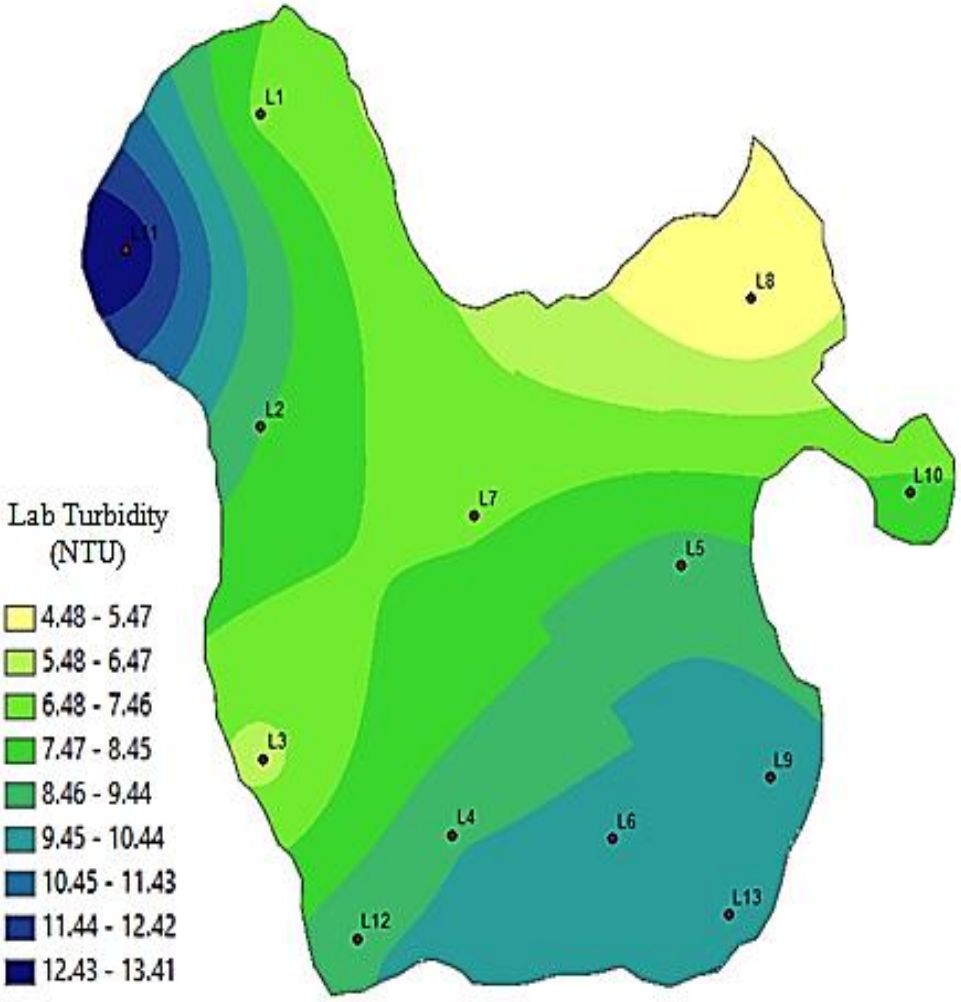
Remote Sensing Reflectance (Rrs) (25/11/2020)							
Sampling Point	**Band 1**	**Band 2**	**Band 3**	**Band 4**	**Band 5**	**Band 6**	**Band 7**
L1	0.2084	0.1583	0.1607	0.1519	0.1683	0.1302	0.1112
L2	0.1971	0.1484	0.1552	0.1385	0.1602	0.1314	0.1095
L3	0.2057	0.1569	0.1596	0.1506	0.1683	0.1319	0.1131
L4	0.2171	0.1683	0.1654	0.1572	0.1740	0.1365	0.1176
L5	0.2138	0.1638	0.1651	0.1557	0.1725	0.1393	0.1215
L6	0.2146	0.1637	0.1707	0.1594	0.1740	0.1451	0.1264
L7	0.1899	0.1425	0.1492	0.1386	0.1575	0.1270	0.1084
L8	0.1788	0.1338	0.1345	0.1267	0.1503	0.1176	0.1020
L9	0.2077	0.1574	0.1648	0.1533	0.1690	0.1413	0.1232
L10	0.2038	0.1584	0.1511	0.1444	0.1669	0.1313	0.1160
L11	0.1982	0.1515	0.1550	0.1449	0.1628	0.1327	0.1122
L12	0.2170	0.1674	0.1691	0.1591	0.1753	0.1400	0.1206
L13	0.2055	0.1574	0.1576	0.1479	0.1663	0.1341	0.1151
Remote Sensing Reflectance (11/12/2020)							
Sampling Point	**Band 1**	**Band 2**	**Band 3**	**Band 4**	**Band 5**	**Band 6**	**Band 7**
L1	0.0569	0.0263	0.0358	0.0271	0.0177	0.0123	0.0079
L2	0.0574	0.0268	0.0361	0.0272	0.0102	0.0054	0.0038
L3	0.0574	0.0270	0.0363	0.0266	0.0109	0.0060	0.0042
L4	0.0577	0.0272	0.0369	0.0284	0.0101	0.0046	0.0032
L5	0.0575	0.0271	0.0371	0.0288	0.0100	0.0047	0.0032
L6	0.0575	0.0271	0.0368	0.0284	0.0100	0.0045	0.0030
L7	0.0572	0.0269	0.0366	0.0275	0.0094	0.0044	0.0031
L8	0.0569	0.0269	0.0369	0.0270	0.0101	0.0051	0.0035
L9	0.0569	0.0268	0.0368	0.0282	0.0123	0.0080	0.0050
L10	0.0557	0.0247	0.0314	0.0214	0.0269	0.0243	0.0134

L11	0.0602	0.0297	0.0364	0.0309	0.0531	0.0524	0.0345
L12	0.0572	0.0272	0.0370	0.0284	0.0107	0.0049	0.0034
L13	0.0575	0.0271	0.0370	0.0285	0.0105	0.0049	0.0033
Surface Reflectance (28/01/2021)							
Sampling Point	**Band 1**	**Band 2**	**Band 3**	**Band 4**	**Band 5**	**Band 6**	**Band 7**
L1	0.0481	0.0146	0.0267	0.0208	0.0171	0.0087	0.0058
L2	0.0492	0.0159	0.0280	0.0219	0.0087	0.0048	0.0037
L3	0.0483	0.0153	0.0278	0.0217	0.0090	0.0048	0.0035
L4	0.0491	0.0157	0.0286	0.0226	0.0078	0.0034	0.0022
L5	0.0491	0.0160	0.0296	0.0236	0.0078	0.0030	0.0023
L6	0.0491	0.0161	0.0296	0.0236	0.0077	0.0032	0.0022
L7	0.0487	0.0154	0.0280	0.0216	0.0077	0.0035	0.0025
L8	0.0485	0.0151	0.0278	0.0215	0.0082	0.0037	0.0025
L9	0.0489	0.0159	0.0296	0.0236	0.0128	0.0058	0.0037
L10	0.0473	0.0142	0.0276	0.0209	0.0348	0.0180	0.0108
L11	0.0507	0.0177	0.0326	0.0310	0.0510	0.0554	0.0401
L12	0.0485	0.0154	0.0280	0.0220	0.0082	0.0037	0.0025
L13	0.0490	0.0159	0.0295	0.0234	0.0084	0.0036	0.0023

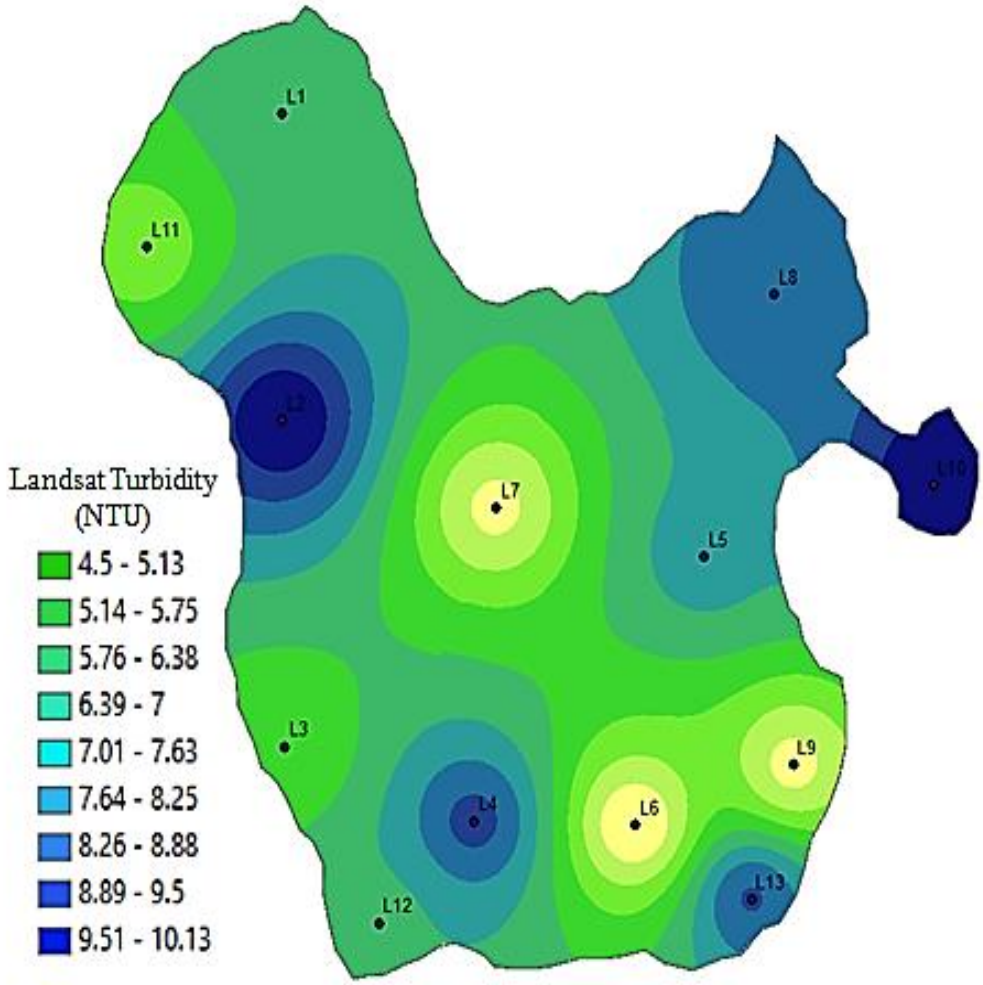
Appendix D: Spatial distribution and variability of *in situ* turbidity on 25/11/2020, 11/12/2020, and 28/01/2021 respectively.

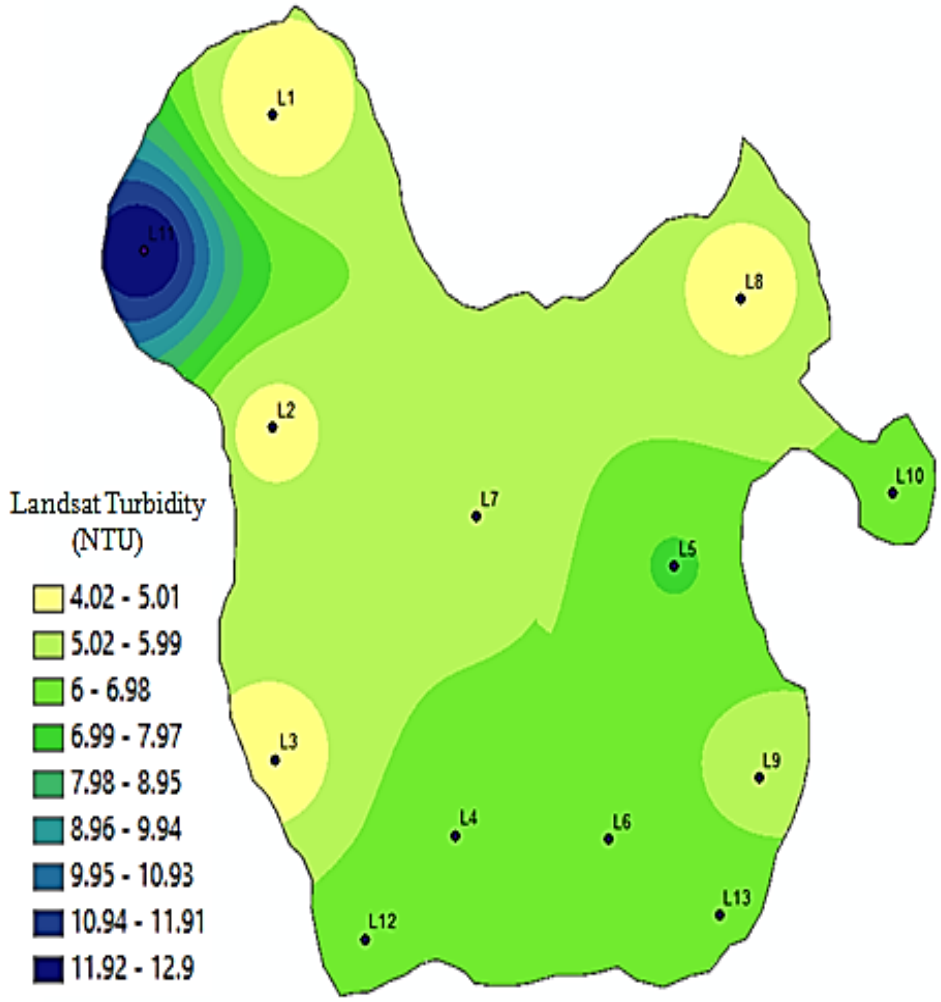


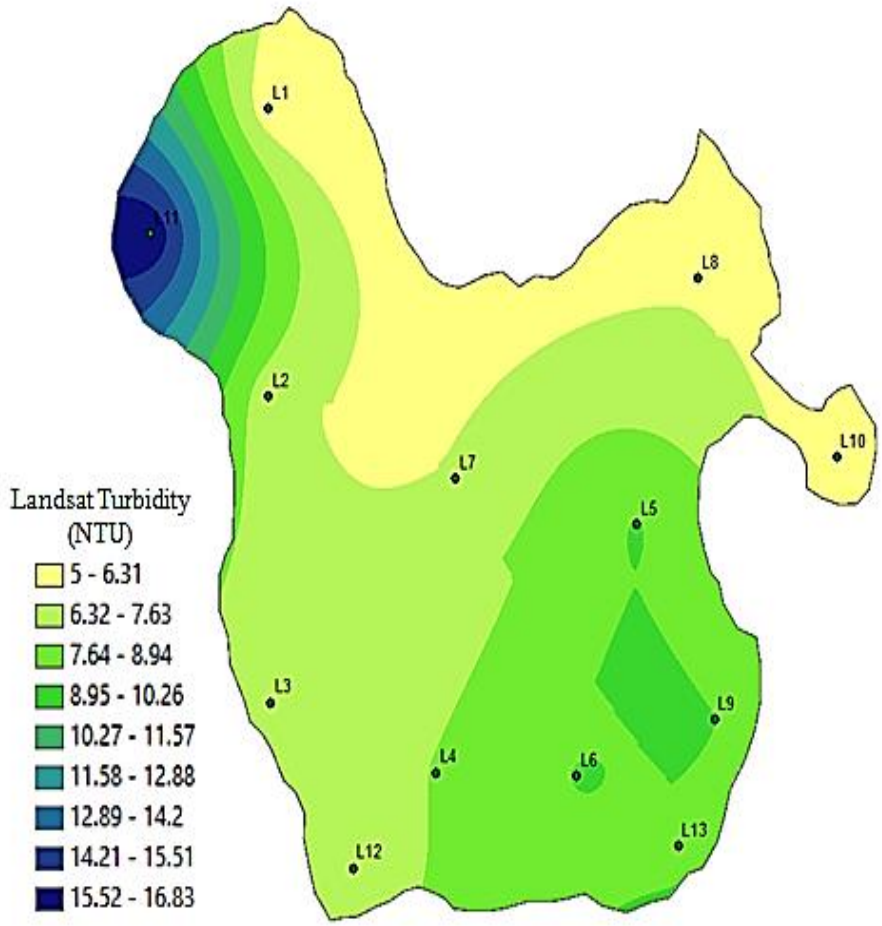

L1
L11
L8
L2
L7
L10
L5
L3
L9
L4
L6
L13
L12
Lab Turbidity (NTU)
4 - 5
5.01 - 6
6.01 - 7
7.01 - 8
8.01 - 9
9.01 - 10
10.01 - 11
11.01 - 12
12.01 - 13


L1
L11
L8
L2
L10
L7
L5
L3
L9
L4
L6
L13
L12
Lab Turbidity
(NTU)
4.48 - 5.47
5.48 - 6.47
6.48 - 7.46
7.47 - 8.45
8.46 - 9.44
9.45 - 10.44
10.45 - 11.43
11.44 - 12.42
12.43 - 13.41

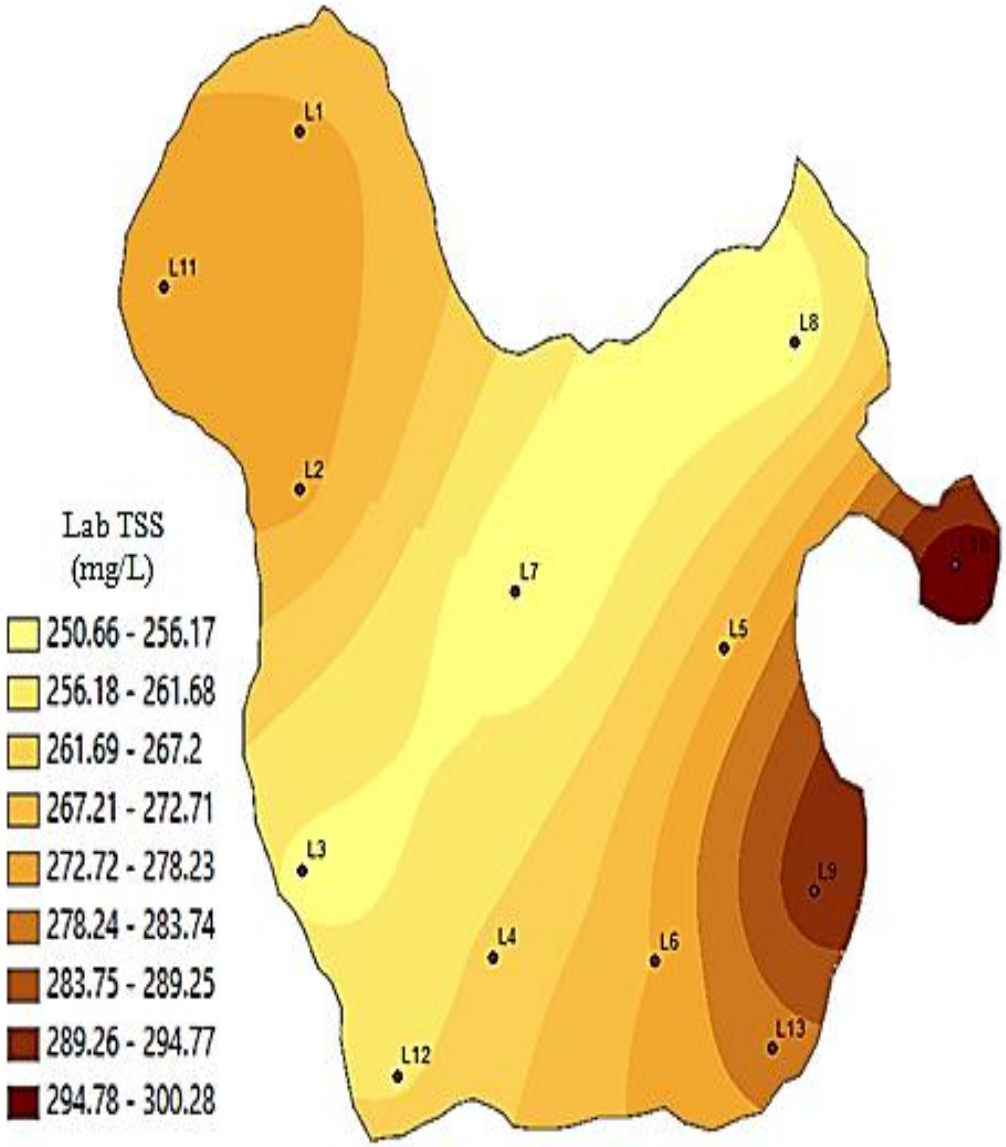
Appendix E: Spatial distribution and variability of Landsat-estimated turbidity on 25/11/2020, 11/12/2020, and 28/01/2021 respectively.

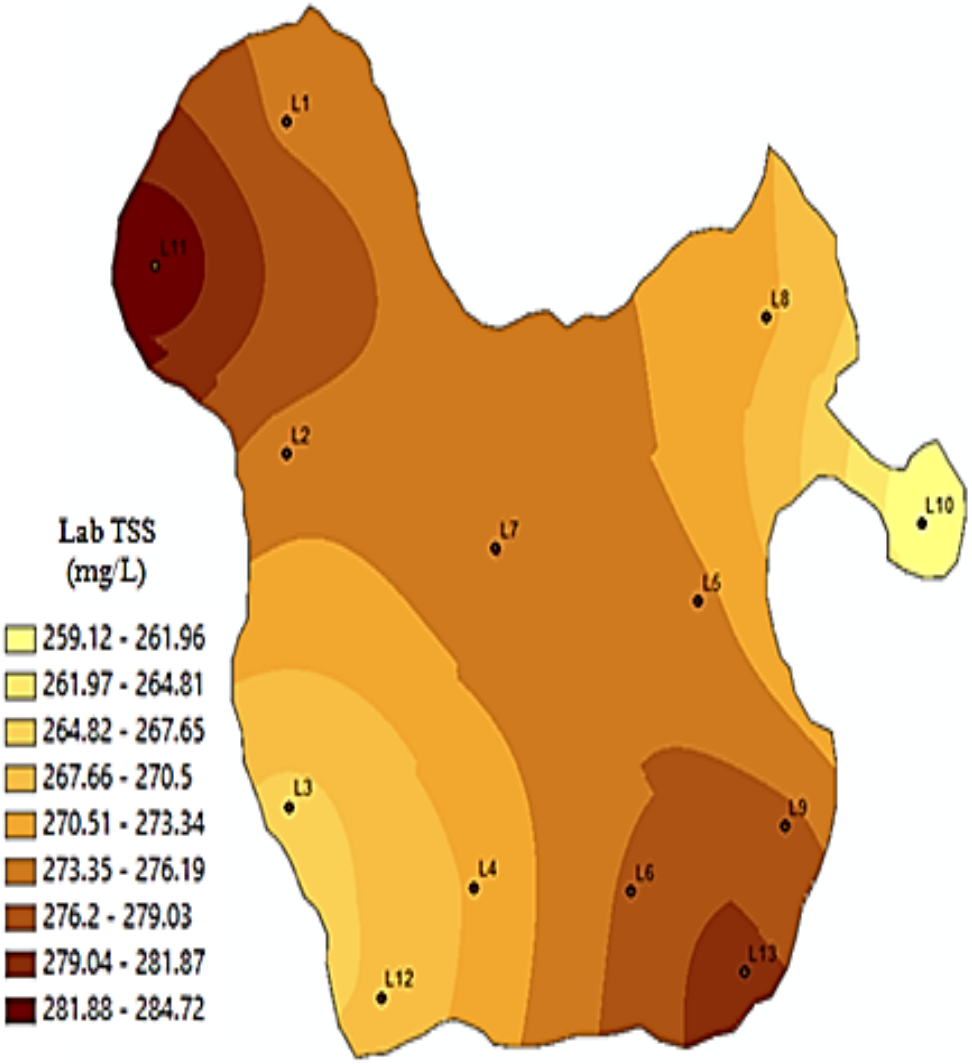


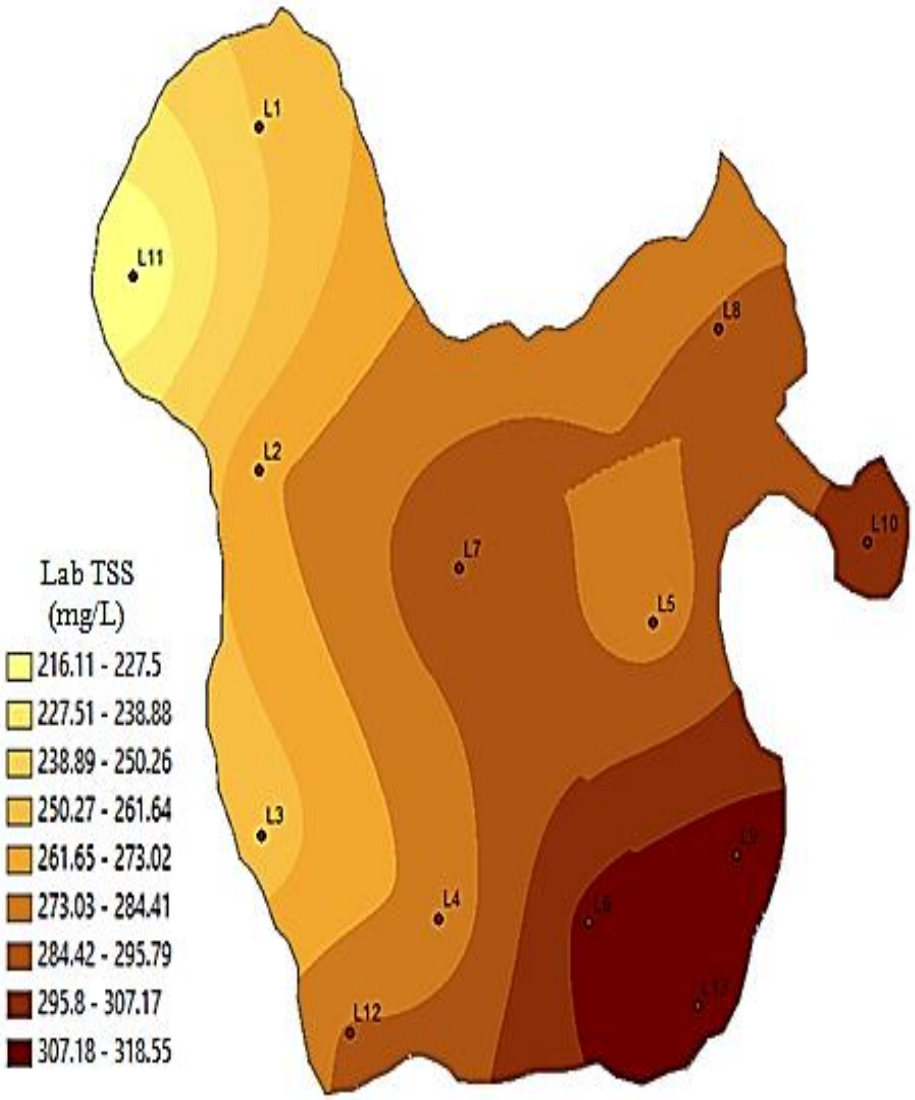

Landsat Turbidity
(NTU)
4.02 - 5.01
5.02 - 5.99
6 - 6.98
6.99 - 7.97
7.98 - 8.95
8.96 - 9.94
9.95 - 10.93
10.94 - 11.91
11.92 - 12.9
L1
L11
L8
L2
L10
L7
L5
L3
L9
L4
L6
L13
L12


L1
L11
L8
L2
L10
L7
L5
L3
L9
L4
L6
L13
L12
Landsat Turbidity (NTU)
5 - 6.31
6.32 - 7.63
7.64 - 8.94
8.95 - 10.26
10.27 - 11.57
11.58 - 12.88
12.89 - 14.2
14.21 - 15.51
15.52 - 16.83

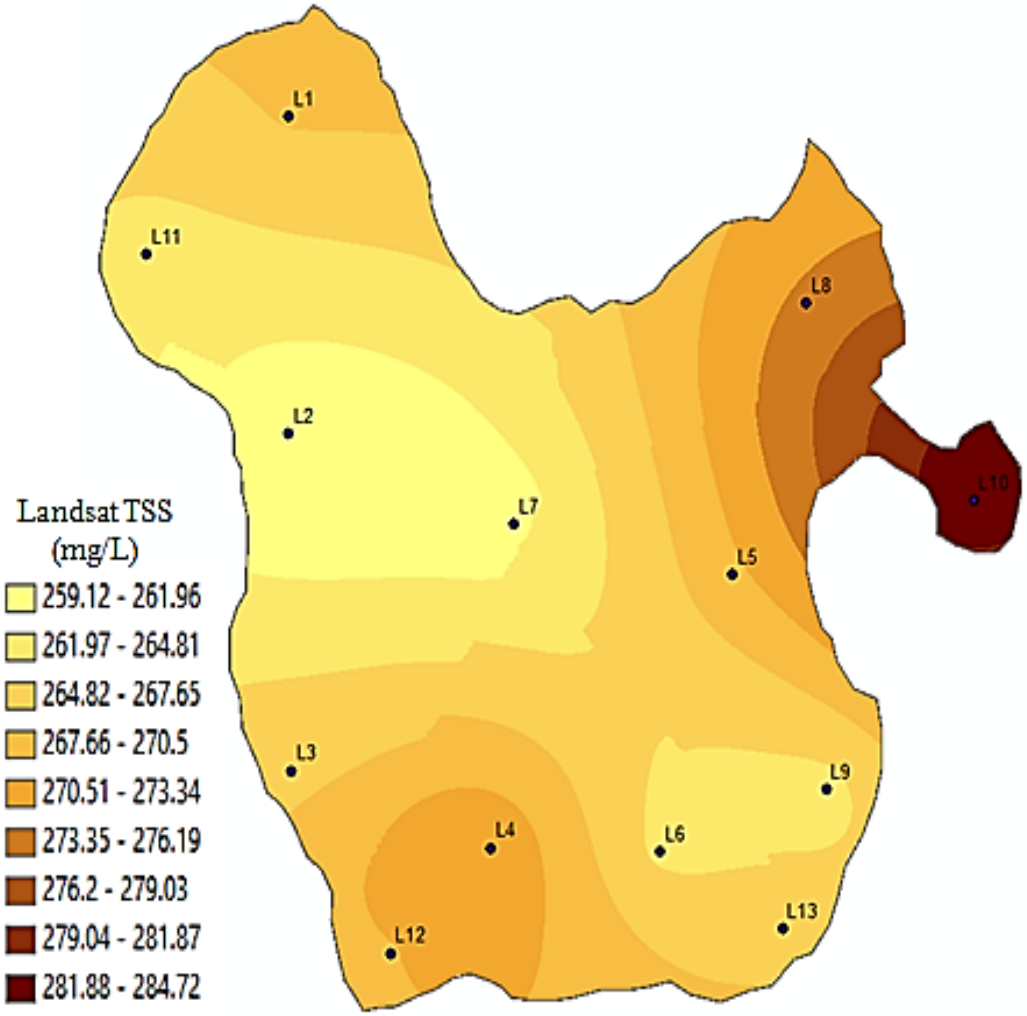
Appendix F: Spatial distribution and variability of *in situ* TSS on 25/11/2020, 11/12/2020, and 28/01/2021 respectively.

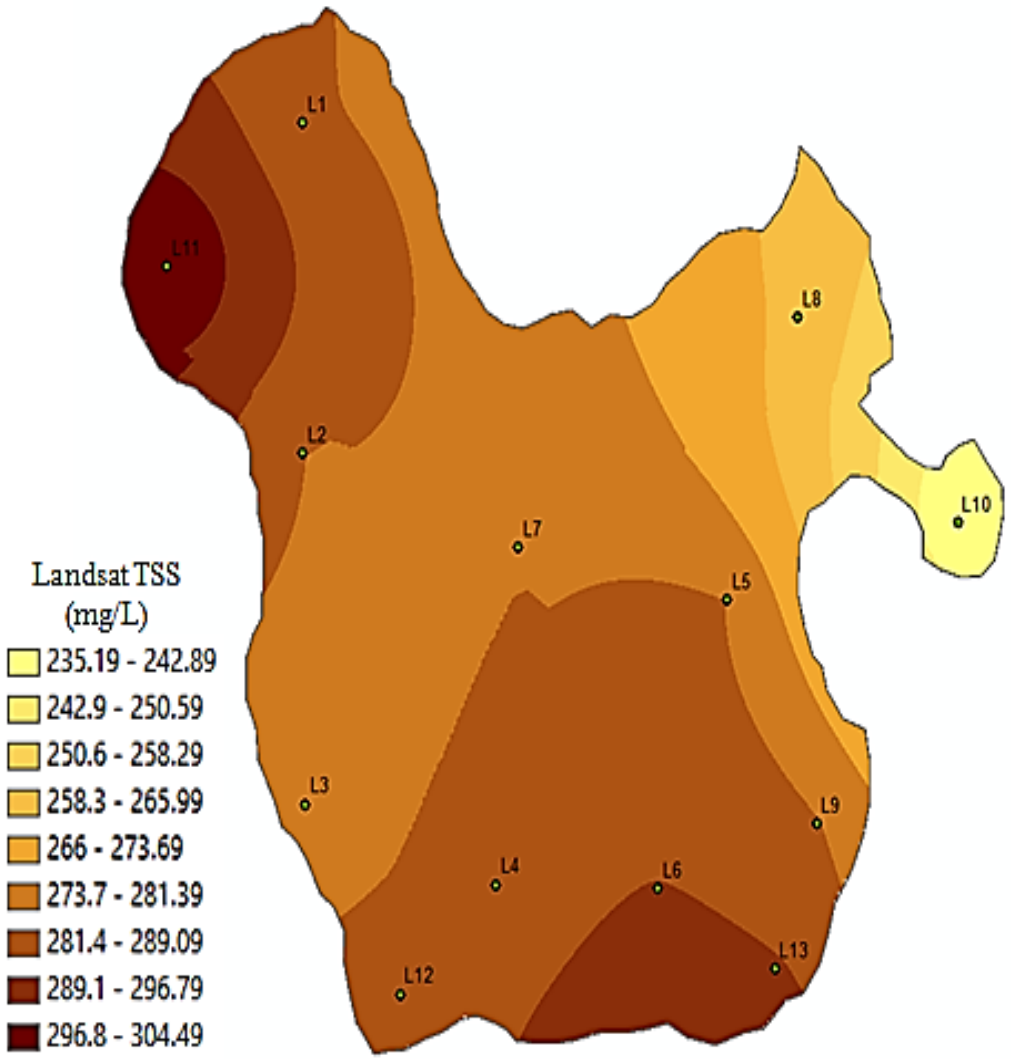


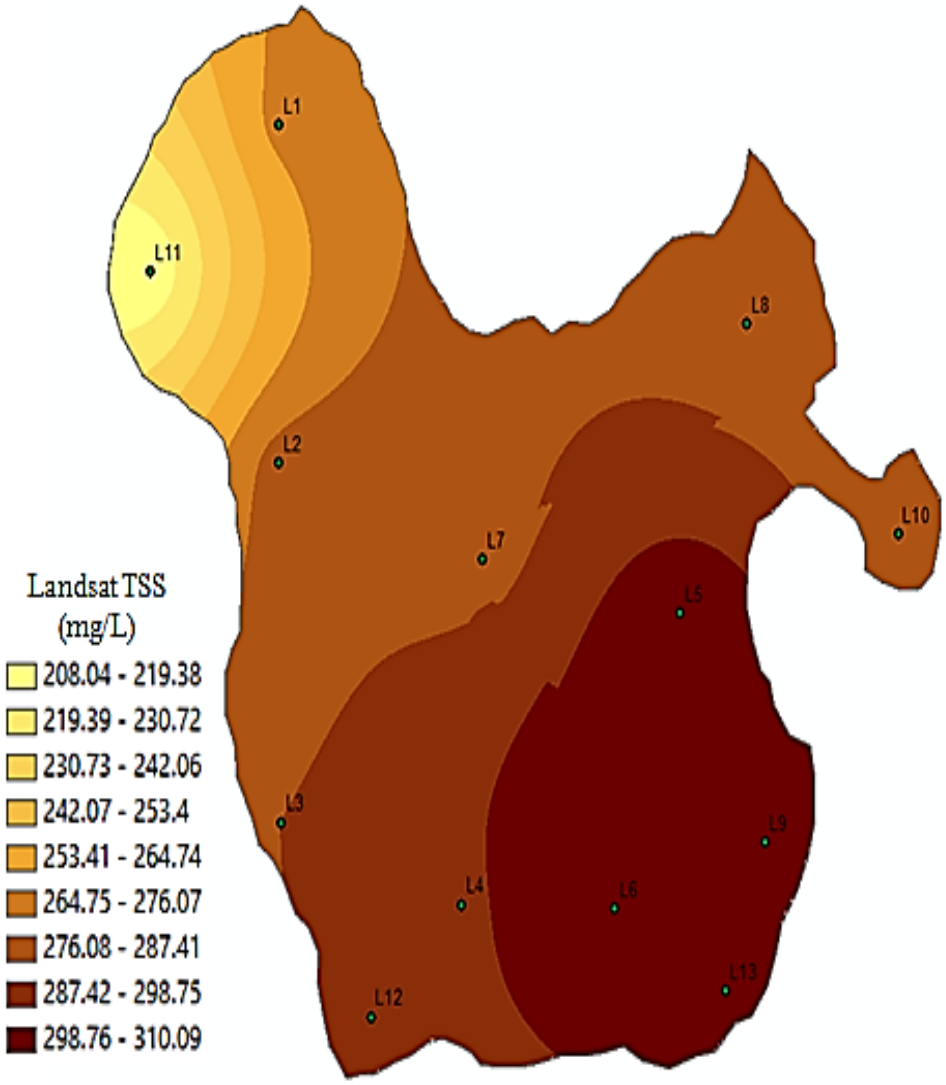

L1
L11
L8
L2
L10
L7
L5
L3
L9
L4
L6
L13
L12
Lab TSS
(mg/L)
259.12 - 261.96
261.97 - 264.81
264.82 - 267.65
267.66 - 270.5
270.51 - 273.34
273.35 - 276.19
276.2 - 279.03
279.04 - 281.87
281.88 - 284.72


L1
L11
L8
L2
L10
L7
L5
Lab TSS
(mg/L)
216.11 - 227.5
227.51 - 238.88
238.89 - 250.26
250.27 - 261.64
261.65 - 273.02
273.03 - 284.41
284.42 - 295.79
295.8 - 307.17
307.18 - 318.55
L3
L9
L4
L6
L12

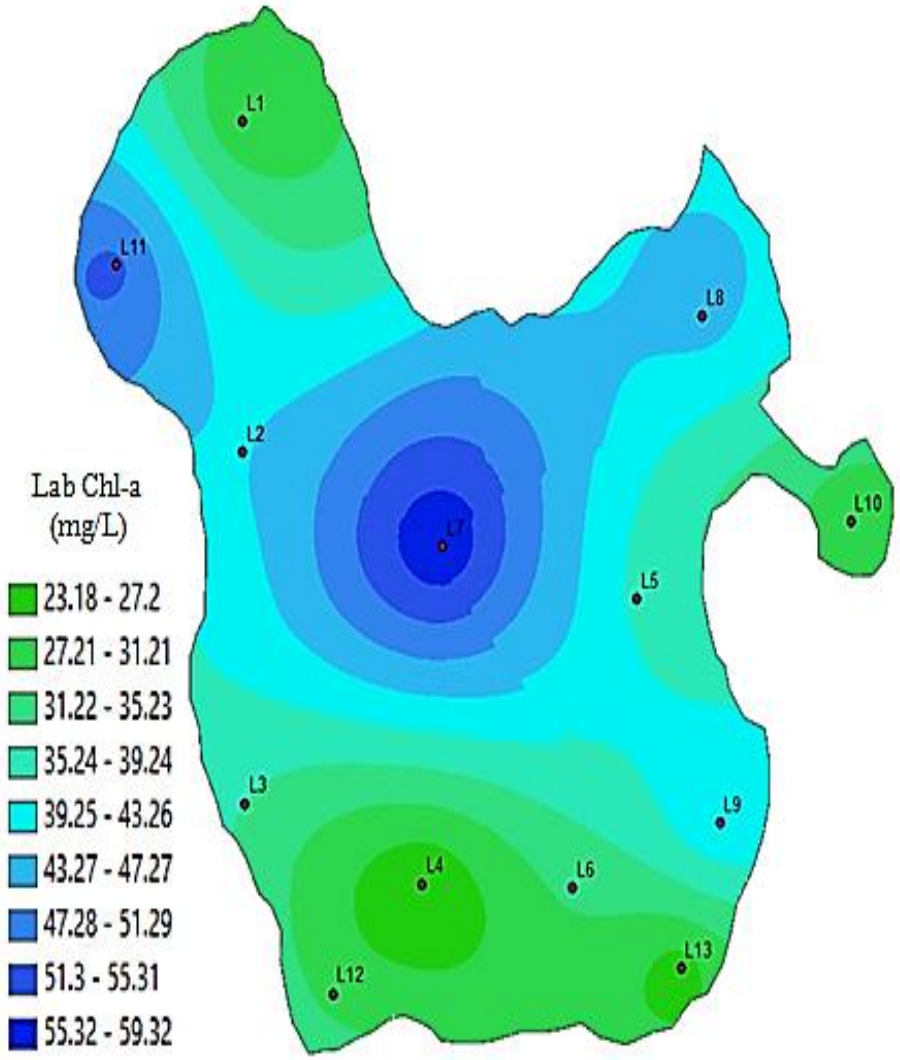
Appendix G: Spatial distribution and variability of Landsat-estimated TSS on 25/11/2020, 11/12/2020, and 28/01/2021 respectively.

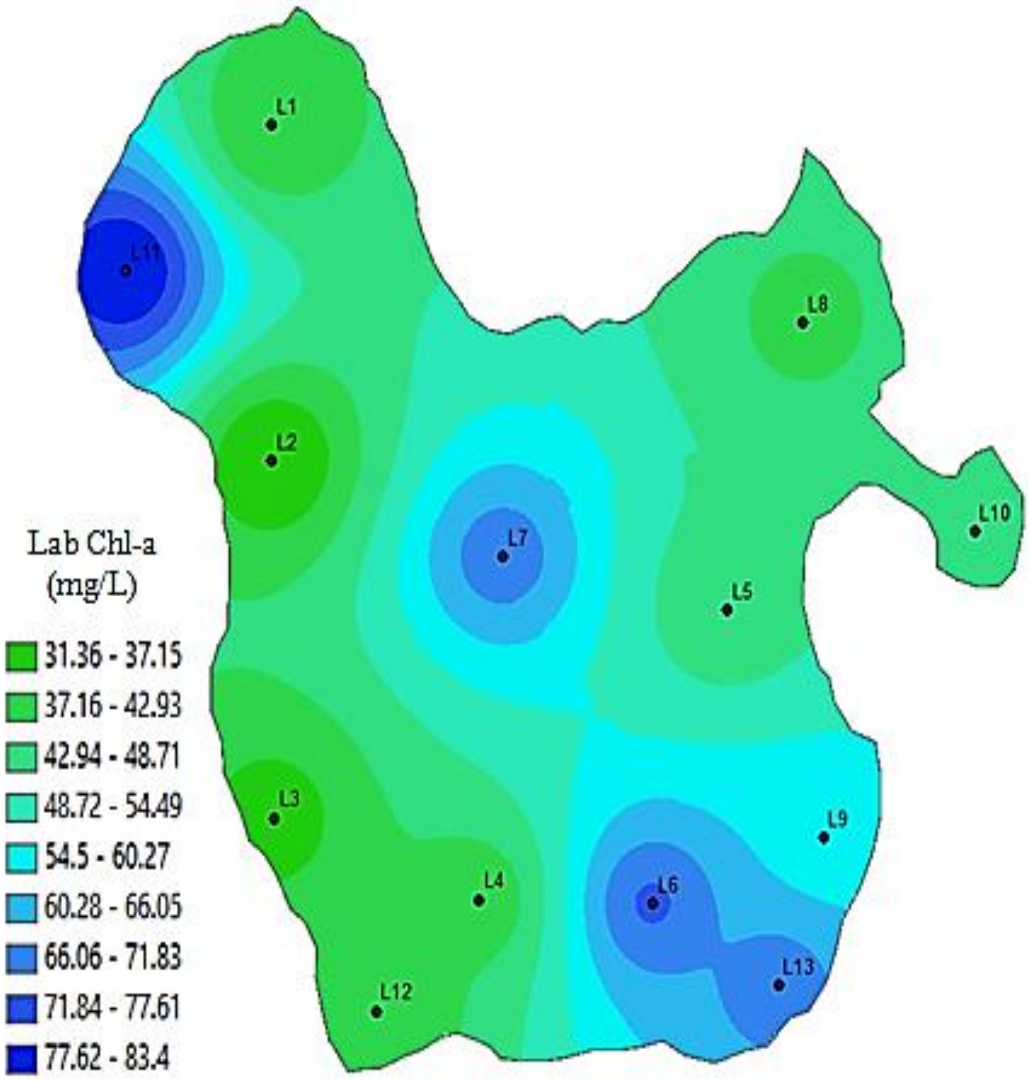


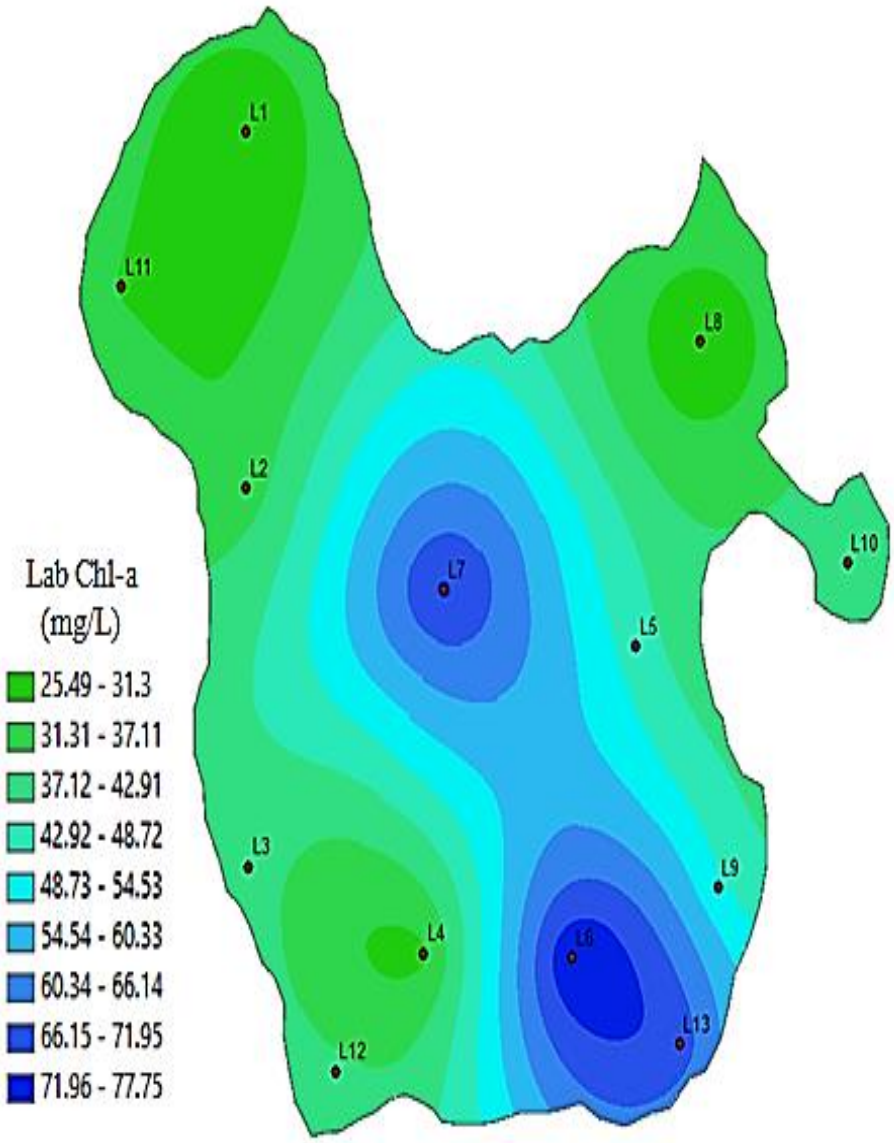

L1
L11
L8
L2
L10
L7
L5
L3
L9
L4
L6
L13
L12
Landsat TSS
(mg/L)
235.19 - 242.89
242.9 - 250.59
250.6 - 258.29
258.3 - 265.99
266 - 273.69
273.7 - 281.39
281.4 - 289.09
289.1 - 296.79
296.8 - 304.49


L1
L11
L8
L2
L10
L7
L5
L3
L9
L4
L6
L13
L12
Landsat TSS
(mg/L)
208.04 - 219.38
219.39 - 230.72
230.73 - 242.06
242.07 - 253.4
253.41 - 264.74
264.75 - 276.07
276.08 - 287.41
287.42 - 298.75
298.76 - 310.09

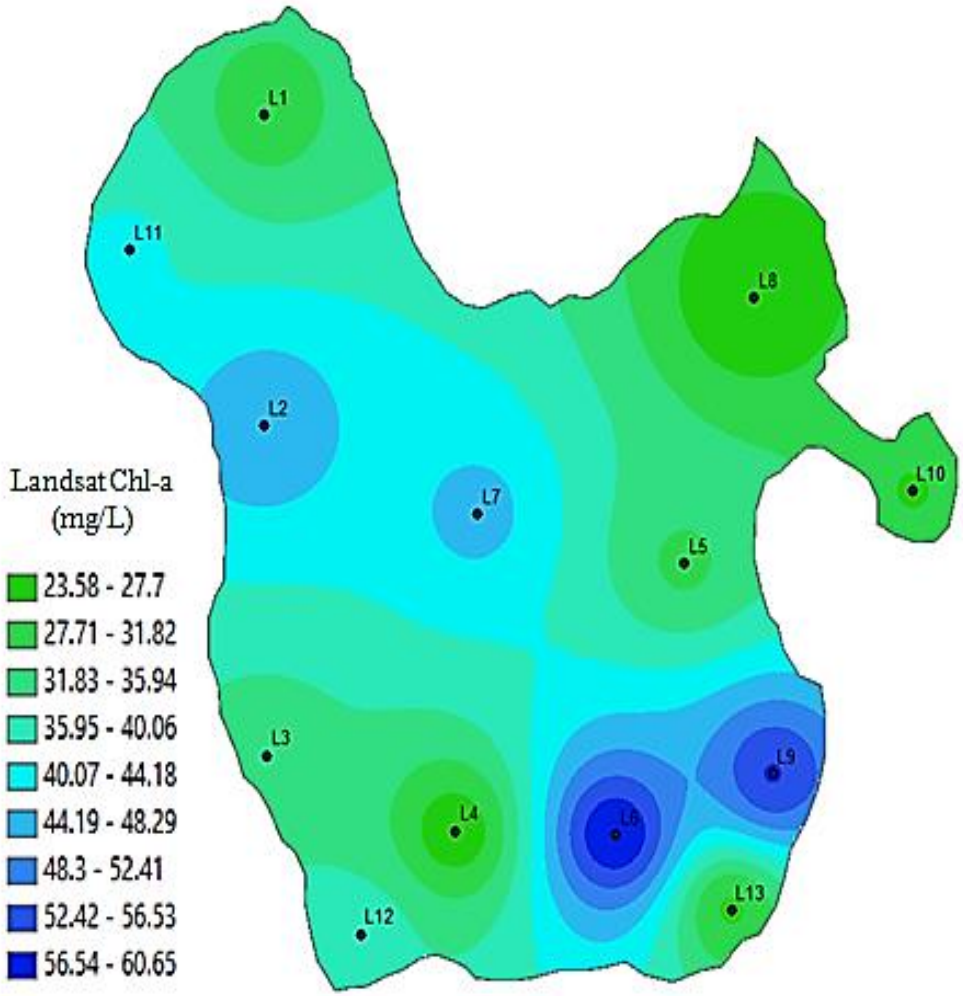
Appendix H: Spatial distribution and variability of *in situ* Chl-*a* on 25/11/2020, 11/12/2020, and 28/01/2021 respectively.

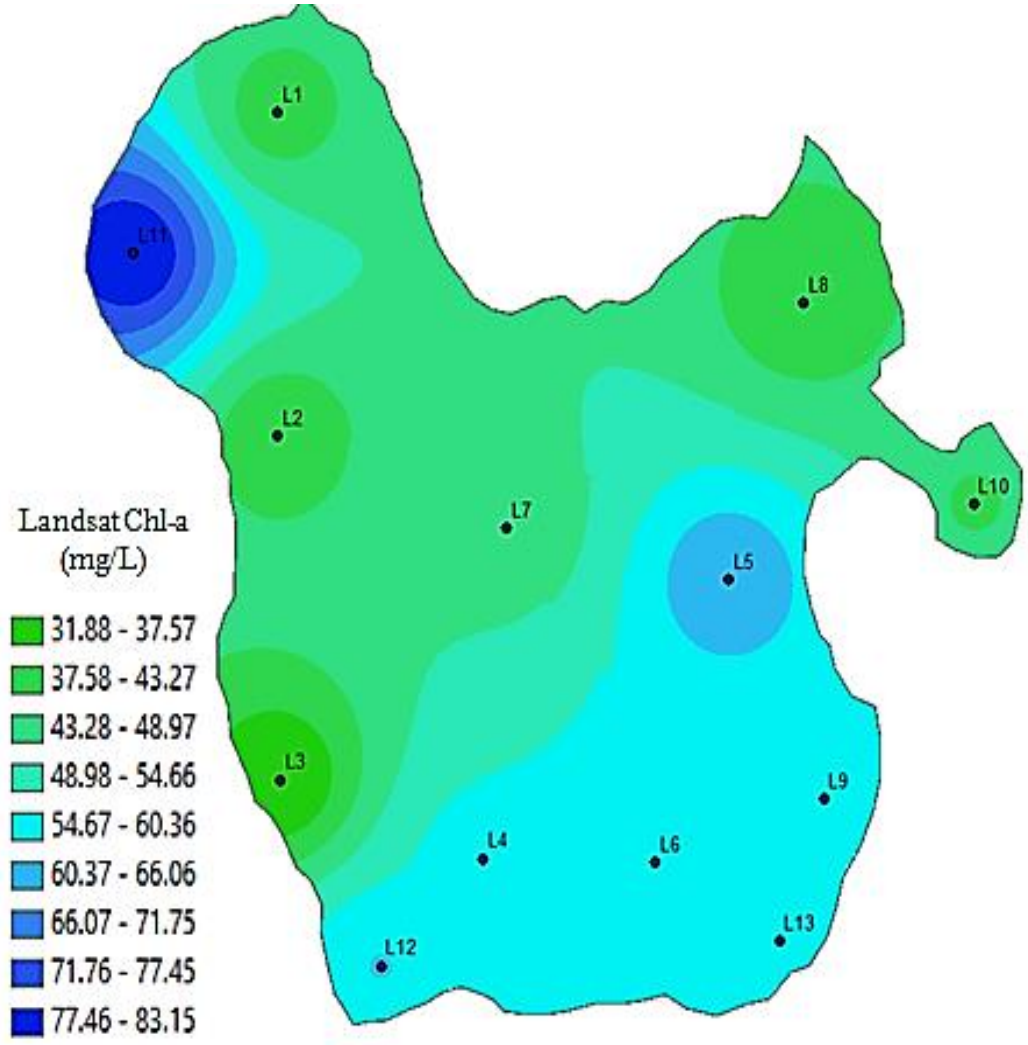


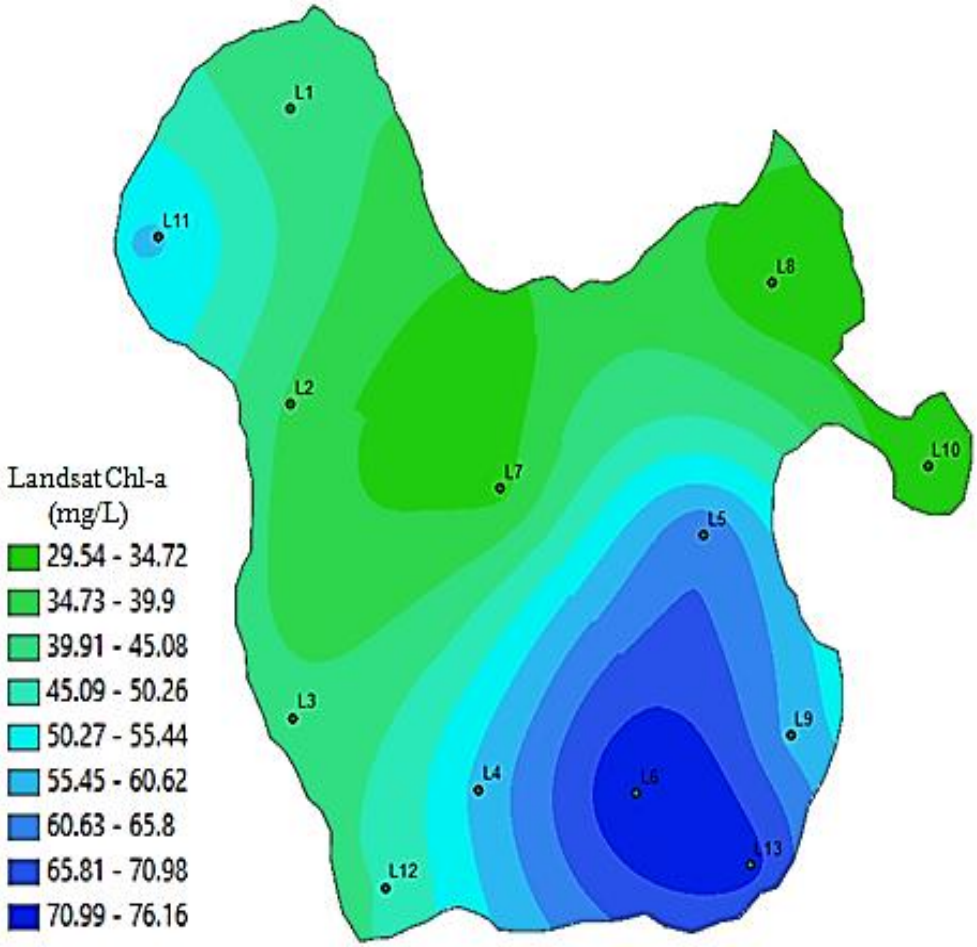

Lab Chl-a
(mg/L)
31.36 - 37.15
37.16 - 42.93
42.94 - 48.71
48.72 - 54.49
54.5 - 60.27
60.28 - 66.05
66.06 - 71.83
71.84 - 77.61
77.62 - 83.4
L1
L11
L8
L2
L10
L7
L5
L3
L9
L4
L6
L13
L12


L1
L11
L8
L2
L10
L7
L5
L3
L9
L4
L6
L13
L12
Lab Chl-a
(mg/L)
25.49 - 31.3
31.31 - 37.11
37.12 - 42.91
42.92 - 48.72
48.73 - 54.53
54.54 - 60.33
60.34 - 66.14
66.15 - 71.95
71.96 - 77.75

Appendix I: Spatial distribution and variability of Landsat-predicted Chl-a on 25/11/2020, 11/12/2020, and 28/01/2021 respectively.




L1
L11
L8
L2
L10
L7
L5
L3
L9
L4
L6
L13
L12
Landsat Chl-a
(mg/L)
31.88 - 37.57
37.58 - 43.27
43.28 - 48.97
48.98 - 54.66
54.67 - 60.36
60.37 - 66.06
66.07 - 71.75
71.76 - 77.45
77.46 - 83.15


L1
L11
L8
L2
L10
L7
L5
L3
L9
L4
L6
L13
L12
Landsat Chl-a
(mg/L)
29.54 - 34.72
34.73 - 39.9
39.91 - 45.08
45.09 - 50.26
50.27 - 55.44
55.45 - 60.62
60.63 - 65.8
65.81 - 70.98
70.99 - 76.16

Appendix J: Antiplagiarism Report from CERMESA

SR145

THESIS WRITING COURSE

PLAGIARISM AWARENESS CERTIFICATE

This certificate is awarded to

OMONDI ADHIAMBO N. ALICE

TEC/PGCS/02/15

In recognition for passing the University's plagiarism

Awareness test with a similarity index of 02% and

Striving to maintain academic integrity.

Awarded by:

Prof. John Changách, CERM-ESA Project Leader

05th /12/2022