Journal of Digital Food, Energy & Water Systems [JD-FEWS]

JDFEWS 5 (1): 159-176, 2024
ISSN 2709-4529

Optimization of Reservoir Water Quality Parameters Retrieval and Treatment Using Remote Sensing and Artificial Neural Networks

Alice Nureen Adhiambo OMONDI
Moi University, Eldoret, Kenya.
nureenlol@gmail.com

Yashon OUMA
Moi University, Eldoret, Kenya.

Simon Njoroge MBURU
Moi Univerity, Eldoret, Kenya.

Cleophas Achisa MECHA
Moi University, Eldoret, Kenya.

Received: 24 May 2024
Review: 27 May 2024
Accepted: 27 June 2024
Published: 30 June 2024

Abstract— Inland water bodies are critical ecosystems that serve several functions, including providing fresh water, regulating climate and hydrological flows, and controlling pollution. Therefore, effective water resource monitoring and management is crucial in enhancing the sustainability of water supply systems. This study evaluated the possible use of satellite data to estimate water quality parameters (WQPs) in an inland water body. The study also used Artificial Neural Network (ANN) models and satellite data to determine the optimum coagulant dose for water treatment. Earth observations and machine learning methods have not been used extensively in developing countries, specifically in water quality monitoring and management. The study utilized empirical multivariate regression modeling (EMRM) of the spectral reflectances from satellite data to retrieve Chl-*a*, Total Suspended Solids (TSS), and Turbidity concentrations in an inland water body. Using Artificial Neural Networks-Multilayer Perceptron modeling, the extracted spectral reflectance values from the selected sampling points in the reservoir were used as model inputs to predict treated WQPs. A second Artificial Neural Networks-Multilayer Perceptron model was developed to predict the optimum coagulant dose required for raw water treatment. The R^2 values achieved with ANN model 1 were 0.81, 0.76, and 0.81, respectively for TSS, Turbidity, and Chl-*a*, and 0.99 for the optimum coagulant dose. The study concluded that spectral reflectance from medium-resolution satellite data products can be used to estimate WQPs from inland water bodies. Further, the ANN models demonstrate that extracted water quality data from satellite images can be used for water quality predictions and the optimization of water treatment plant operations.

Keywords— *Artificial Neural Network, Remote sensing, Water quality monitoring, water treatment.*

Journal of Digital Food, Energy & Water Systems [JD-FEWS]

JDFEWS 5 (1): 159-176, 2024
ISSN 2709-4529

1 Introduction

About two-thirds of the earth's surface is covered by water resources, yet only 3% of this is portable water (1). Furthermore, 25% of the global population still lacks access to sufficient quantities of good quality water (1). Inland water bodies are critical ecosystems that serve several functions, including providing fresh water, regulating climate and hydrological flows, pollution control, recreation, and irrigation (2). However, factors such as urbanization, climate change, environmental degradation, and unsustainable agricultural practices continue to contribute to deteriorating water quality (3). These changes negatively impact the sustainability of inland water resources by advancing algae proliferation, accelerating eutrophication, and increasing the concentration of turbidity and total suspended solids (TSS). Effective management and use of the existing water resources can be achieved by measuring water quality parameters (WQPs) thus facilitating effective management of the surface water resources.

Using satellite data for water quality monitoring (WQM) has proven useful in protecting water sources and improving the effectiveness of the water treatment processes (4). WQM and forecasting also predict pollution levels, allowing for advanced action plans to optimize water treatment plant (WTP) operations (4). Remote sensing techniques have continued to advance, thus finding wider applications in WQM (3, 5, 6). Remote sensing techniques can be used with artificial neural networks (ANNs), which can model non-linear geophysical transfer functions, including non-linear environmental processes (7). Combining remote sensing and ANN can also help to reduce the limitations of *in situ* sampling by using satellite data to predict WQPs at unsampled locations in a water supply reservoir. The concept can then be used to inform decisions on potable water treatment by using the satellite data and ANN to create a system to determine the optimum coagulant dose for portable water treatment.

The use of satellite imagery, specifically, Landsat images, has been used in extracting water quality data in inland water bodies in recent decades. Patra, Dubey (8) used in situ water quality data and Landsat Operational Land Imager (OLI) images to estimate the concentration of Chl-a. To correlate the in situ Chl-a concentrations with different OLI bands, a Pearson correlation analysis was done. The ratio between OLI5/OLI4 was the most optimal band combination for the estimation of Chl-a with an R2 value of 0.85 (8). Bonansea, Ledesma (9) in Cassafouth Reservoir in Cordoba, Argentina where Landsat-8 and Sentinel-2A images were used to estimate and map secchi disk transparency. The developed models achieved R^2 values between 0.6 and 0.94. The achieved R2 values showed the feasibility of using satellite data in estimating and mapping reservoir WQPs.

Pizani, Maillard (10) compared water quality data measured in situ from a hydroelectric reservoir in Brazil to data obtained from Landsat-8 OLI and Sentinel-2 Multispectral Instrument (MSI). Both sensors estimated the optically active WQPs with R^2 values greater than 0.6. The use of regression models, specifically empirical multivariate regression models (EMRM) in water quality modeling is demonstrated by (11) and (12). Meng, (11) used EMRM in Shanmei Reservoir in Fahzou City to estimate of turbidity, Chl-a, and algae density using Sentinel-2 MSI, and Landsat 8-9 OLI. Landsat-8 OLI gave better estimates than Sentinel-2 MSI with R^2 values of 0.70, 0.81, and 0.80, respectively, for Chl-a, algal density, and turbidity. Ouma, Noor and Herbert (12) also used EMRM in relating the performance of Landsat-8 OLI and Sentinel-2A/MSI in retrieving the concentration of turbidity, TSS, and Chl-a based on in situ WQPs from a water supply reservoir. Both studies showed the viability of the sensors and the regression models in retrieving water quality data from satellite images.

The use of ANN models to predict treated WQPs was tested at the Akron WTP in Ohio, USA. Four ANN models were developed and used to predict treated water turbidity and dissolved organic materials (13). The treated water turbidity was predicted with an R2 value of 0.91, suggesting that ANN models can be useful in predicting treated WQPs (13). The ANN modeling approach has also been tested in several studies. Haghiri, Daghighi and Moharramzadeh (14) tested an ANN-MLP model in a portable WTP in Ardabil province in Iran. The model had a reasonable accuracy with R^2 values greater than 0.8 in predicting treated WQPs, an R^2 value of 0.95, and a MSE of 0.12 in predicting the optimum coagulant dose. The use of ANN models for the prediction of the optimal coagulant dose for water treatment has also been tested in several studies. (15) developed ANN models for optimal coagulant dose forecasting in portable WTPs. Two models were developed depending on the quality of the raw water, and these models were tested on a large database of the selected WTPs. The models predicted the optimal coagulant dose with R^2 values above 0.8, and mean absolute errors (MAE) above 5.4g/m^3 (15). The study showed the possibility of adopting and using these models in WTPs to optimize the coagulant dosing practice. Kote and Wadkar (16) developed ANN models based on data from Maharashtra WTP, India to estimate the

Journal of Digital Food, Energy & Water Systems [JD-FEWS]

JDFEWS 5 (1): 159-176, 2024
ISSN 2709-4529

optimum coagulant and chlorine doses. The CFNN model was the most effective coagulant dose model with an R^2 value of 0.947while the most effective chlorine dose model was the RBFNN model with an R^2 value of 0.999 (16).

This study presents a process for optimizing portable water treatment processes using remote sensing and ANN through a completely predictive WTP model starting from the reservoir to the final point of the WTP. The concept has also been explored by (14) and (16), but the focus of these studies was limited to the water treatment works only. A completely predictive WTP model factors catchment processes that negatively impact water quality. As a result, the catchment processes largely influence the decisions made at the WTP, including the type and quantity of chemicals used. The study used satellite data and *in situ* water quality data to estimate turbidity, TSS, and Chl-a from Landsat-8 OLI using EMRM. The extracted WQPs from the Landsat images were then used in ANN modeling to predict treated WQPs for the Sosiani WTP. A second ANN model was then trained, validated, tested, and used to determine the optimum coagulant dose for the WTP.

2 Materials and methods

2.1 Study site

Two Rivers Dam Reservoir is situated in Uasin Gishu County, Kenya, at a longitude of 35° 35' 14" and latitude of 0° 46' 88", as shown in Figure 1 (3). The dam supplies the Sosiani Water Treatment Plant, where the water is treated and distributed to Eldoret town and its environs.

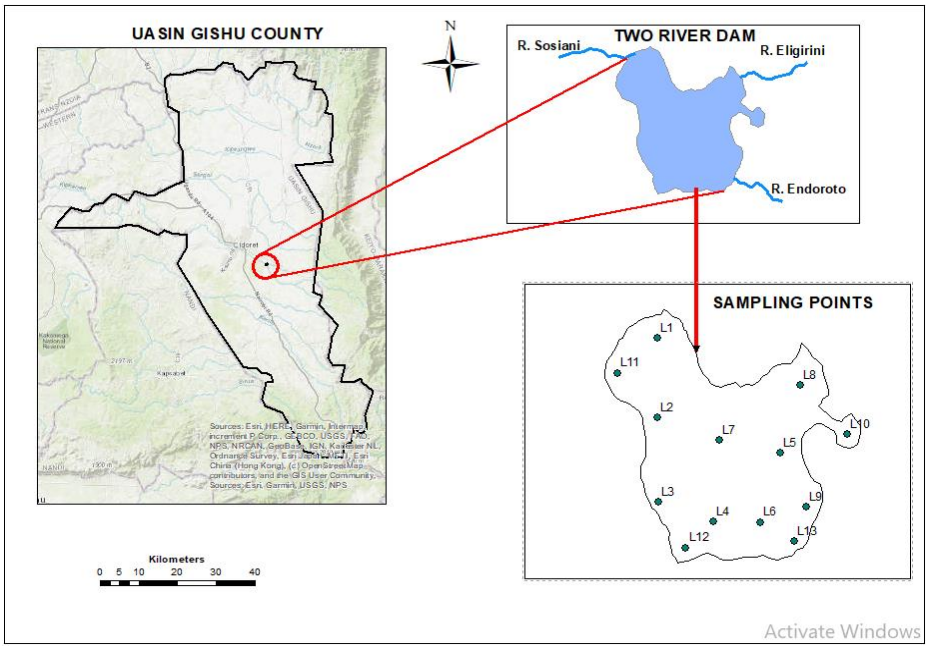

Fig. 1. Two Rivers Dam and the sampling points

2.2 Sample site selection

Thirteen in situ water sampling points were set up in the dam (3). Several sampling points were located where River Elligirini and River Endoroto enter the dam and, in the region, where River Sosiani exits the dam. More sampling points were also located at the edges where there is a high likelihood of water quality variability (3, 17). A GPS receiver was used to locate the sample site.

JDFEWS 5 (1): 159-176, 2024
ISSN 2709-4529

2.3 Sampling and laboratory water quality testing

The satellite overpass schedule determined the in-situ sample collection dates. Duplicate 1500ml samples were collected at 0.5-meter depths at each sampling point, and the turbidity, TSS, and Chl-a concentration was determined (3). The coagulant dose for water treatment was determined using jar test experiments. The stock solution for the coagulant (Aluminum sulfate) was 1 g/L of alum (18).

2.4 Determination of radiance and reflectance

The Landsat image acquisition schedule was used to obtain the Landsat images used for the study. Image acquisition was done between November 2020 and January 2021. The satellite images were processed using the Fast Line-of-sight Atmospheric Analysis of Hypercubes (FLAASH) model, as explained in (19).

2.5 Correlation of satellite data with laboratory water quality data

The 13 sampling points were then overlaid on the images' regions of interest (ROIs). For each of the sampling points, the spectral profile for each of the OLI bands (1-7) was then used to determine the top of atmosphere (TOA) reflectance. To minimize errors in locating the sampling points and to improve the EMRM analysis, an average spectral reflectance of 3 × 3 pixel neighborhood configuration was used (Reddy, 1997).

2.6 Empirical multivariate regression modeling

EMRM was used to correlate satellite and *in situ* WQPs (19). Different band combinations were used for the retrieval of turbidity, TSS, and Chl-a from satellite images (table 1) (3, 11). The water quality retrieval algorithms were developed by correlating *in situ* water quality data with remotely sensed data, forming the basis for the regression analysis. The calibration of the EMRM was done using eight sampling points, and model validation was performed using five points (3).

Table 1. Remote sensing band combinations (3)

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

2.7 ANN model training, validation, testing, and application

ANN models can be used to evaluate complex, non-linear relationships to establish specific data characteristics and predict future events (20-22). ANNs, specifically the ANN multilayer perceptron (MLP), have also been used to make water quality predictions (20).

Selection of network architecture, layers, and neurons The ANN MLP network architecture was used (14, 15). The MLP-ANN network architecture gives accurate predictions even with limited datasets, as in his study (23). Two ANN models, each with three layers, were developed to predict treated WQPs, and the optimal coagulant dose as described in (18).

a) Model 1-Prediction of treated water quality parameters A schematic of the model to predict treated WQPs is shown in Figure 2.

Journal of Digital Food, Energy & Water Systems [JD-FEWS]

JDFEWS 5 (1): 159-176, 2024
ISSN 2709-4529

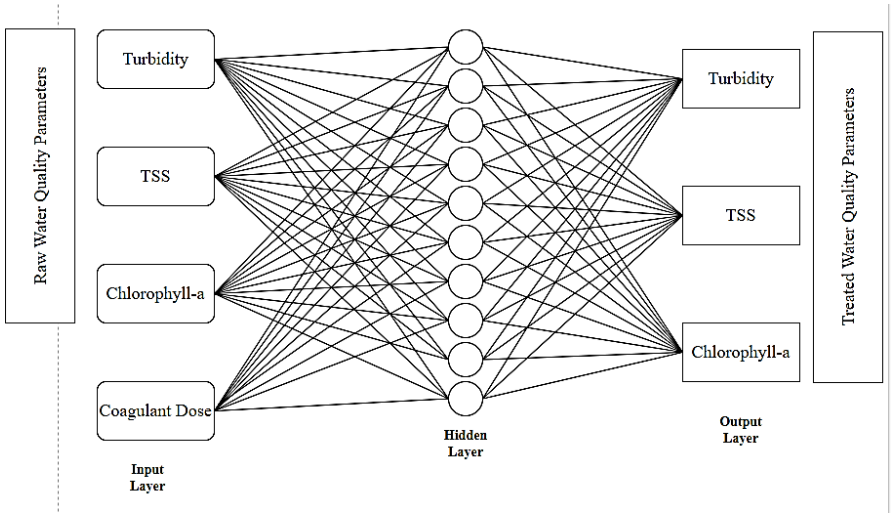

Fig. 2. ANN model 1 for predicting treated water quality parameters

An ANN model was developed using the MATLAB software. The process inputs were the raw WQPs, and the treated WQPs were the process outputs being modeled. The process control parameter was the alum dose. The inputs for the ANN model 1 were the normalized spectral reflectance values from satellite imagery for each specific sampling day, while the targets were the treated WQPs as shown in the figure.

b) ANN model 2-Prediction of optimal coagulant dose for water treatment The developed ANN model 2 was used to predict the optimum coagulant dose for water treatment. A schematic summarizing the process is shown in Figure 3.

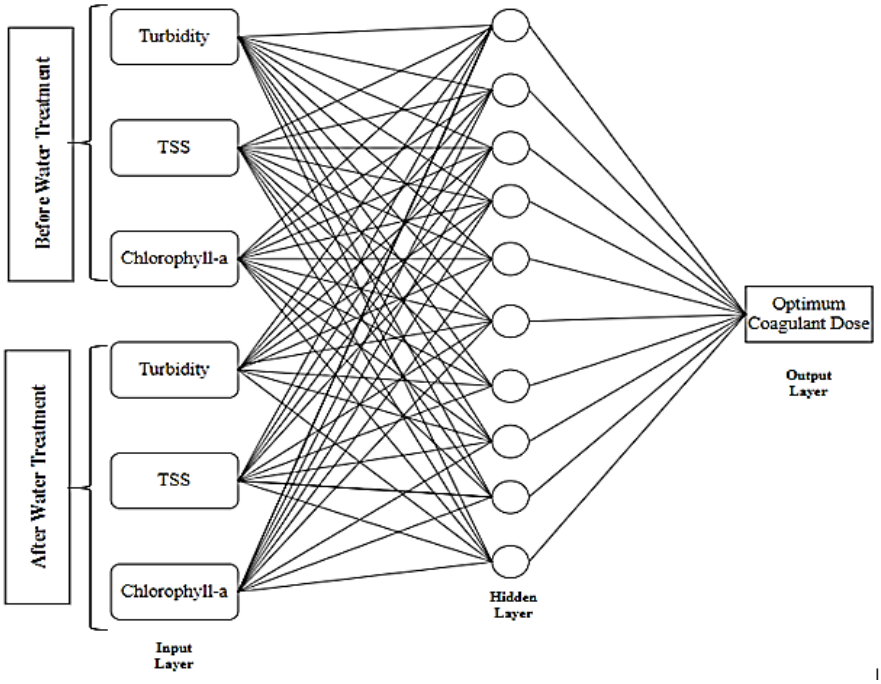

Fig. 3. ANN model 2 for predicting optimal coagulant dose

Normalization Prior to entering the data into the developed ANN models, the data was normalized using the normal distribution function shown in equation 1. As a result, the models' performance was improved (24).

That is:

$$(x+a)^n = \sum_{k=0}^{n} \binom{n}{k} x^k a^{n-k} \quad (1)$$

JDFEWS 5 (1): 159-176, 2024
ISSN 2709-4529

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

Data division and pre-processing The default MATLAB data division was applied to split the data set for training, testing, and validation for both models that were developed.

Model performance and evaluation Model performance was evaluated using root mean squared error (RMSE), mean absolute error (MAE) or bias, and the coefficient of determination (R^2) as shown in equation 2, 3, and 4.

Coefficient of determination (R^2)

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

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

Bias

$$Bias = \frac{1}{n}\sum_{i=1}^{n}[X_i - Y_i] \quad (3)$$

Where:

X_i - Estimated WQP value for the i^{th} sampling location
Y_i – Lab measured WQP value for the i^{th} sampling location
n- Total number of sampling locations

Normalized root mean squared error

$$NRMSE = \frac{RMSE}{y_{imax} - y_{imin}} \quad (4)$$

Where:

RMSE= Root Mean Square Error

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

a_i - estimated WQP value at the i^{th} sampling location
b_i – *in situ* WQP value for the i^{th} sampling location
n - Total number of sample locations
y_{max} – maximum estimated WQP value
y_{min} – minimum estimated WQP value

3 Results and discussion

3.1 Retrieval of turbidity from satellite images

The average in situ turbidity was 7.69 NTU. The turbidity in Two Rivers Dam reservoir was low since water samples were collected during the dry season between November and January. Therefore, the sediment inflows from rainwater discharge into the reservoir was very low. The low sediment loads could also be attributed to plain

JDFEWS 5 (1): 159-176, 2024
ISSN 2709-4529

sedimentation (25). Table 2 presents the Landsat band combinations and the associated regression equations through which turbidity values were derived from the remote sensing reflectance (Rrs).

Table 2. Regression equations and band combinations for turbidity estimation (18)

	Regression Equation	**Band Combination**	**R^2**	**nRMSE (NTU)**	**MAE (NTU)**
25/11/2020	**$y=-1169x^2+3694x2908$**	(B1/B4)+B2	0.80	0.26	0.08
	$y= -996.8x^2+2853x-2031$	B1/B4	0.68	0.31	-0.79
	$y = -1552x^2+2831x-1280$	B4/B2	0.73	0.29	0.55
11/12/2020	**$y=68165x^215713x+908.2$**	B3+B4+B1	0.76	0.63	-1.57
	$y=67304x^2-11390x+484.4$	B2+B3+B4	0.74	0.57	-1.61
	$y = 87184x^2-8803+226.4$	B2+B4	0.73	1.06	-1.34
28/01/2021	$y = 29.02In(x)+117.7$	B4	0.69	0.36	-0.40
	$y = 712.1x^2-704.5x+175.2$	B3/B1	0.74	0.36	-1.80
	$y = 130x^2-493.6x+471.4$	(B1/B3)+B2	0.74	0.36	-1.80

The visible Band 2, Band 3, and Band 4 gave the highest correlation coefficient between *in situ* and satellite-derived turbidity, as shown in Table 2. For each sampling date, the polynomial regression models achieved R^2 values of 0.80, 0.76, and 0.74, respectively. Similar results were also obtained from the study by Lotfi, Ahmadi Nadoushan, and Abolhasani (26) and Ouma, Noor, and Herbert (12) where turbidity was best estimated using the visible Landsat-8 OLI bands. Kalele (27) also demonstrated the viability of the visible red and blue bands in the estimation of turbidity, which was achieved with R^2 and RMSE values of 0.83 and 0.43, respectively. In a study by Hossain, Mathias, and Blanton (28) Landsat-8 OLI data was used in correlation with near real-time *in situ* water quality data in estimating the turbidity of the selected points in Tennessee River in the United States. The Red band was more effective in estimating turbidity with an R^2 value of 0.95, even though the suitability of using single-band water quality retrieval algorithms for the estimation of turbidity is demonstrated by Hossain, Mathias, and Blanton (28); the findings of this study similar to what was established by Kalele (27), Lotfi, Ahmadi Nadoushan and Abolhasani (26), and Ouma, Noor and Herbert (12) high levels of success can also be achieved in using multiple bands in the regression models for turbidity estimation from satellite images. Furthermore, in this study, some of the sampling points considered are in the shallow regions of the reservoir. Therefore, the adjacency of these points to the land surface could erroneously increase the reflectance from the near-infrared bands (28, 29). However, the effectiveness of the visible bands in such scenarios is still high since the visible bands are largely unaffected by the location of some sampling points in the shallow regions of the reservoir (28, 29). Thus, the visible bands are more effective in shallow or narrow fluvial environments in estimating optical water properties from satellite imagery.

3.2 Retrieval of turbidity from satellite images

The average in situ TSS for the entire sampling period was 277.91 mg/L. Table 3 presents the regression equations and band combinations that retrieve TSS from satellite imagery. TSS concentration was notably high at the points where River Ellegerini and River Endoroto entered the reservoir, causing sediment resuspension.

JDFEWS 5 (1): 159-176, 2024
ISSN 2709-4529

Table 3. Retrieval of total suspended solids from satellite images (18)

	Regression Equation	**Band Combi-nation**	**R^2**	**nRMSE (mg/l)**	**Bias (mg/l)**
25/11/2020	**y=5340In(x)+2754**	B3/B2	0.79	0.70	-9.98
	y = 7930x – 9457	(B1/B4) +B2	0.80	0.49	0.24
	$y=17955x^2$–47361x +33603	(B1/B3) + B2	0.81	0.73	-7.84
11/12/2020	**$y = 635.9e^{53.65x}$**	B4	0.85	0.38	2.66
	$y=4640x^2$-23267x+31152	(B1/B4) + B2	0.80	0.29	4.20
	$y=77928x^2$-64825x+15537	(B4/B1) + B4	0.82	0.29	3.51
28/01/2021	**$y=-6131x^2+25640x– 23721$**	(B1/B4) + B1	0.76	0.50	15.86
	$y=-6139x^2$+25265x– 22910	(B1/B4) + B2	0.76	0.50	15.70
	$y=-79430x^2$+83798x– 18978	(B4/B1) + B4	0.77	0.54	16.55

The optimal bands for TSS estimation in this reservoir were the visible B1, B2, B3, and B4. TSS was best estimated using the exponential, logarithmic, and polynomial regression models with R^2 values of 0.85, 0.81, and 0.77, as shown in Table 3. Studies have shown that high concentrations of TSS are very sensitive when detected using B4, while low concentrations of TSS are detected at B3 (30). Furthermore, the combination of B2 and B3 with B4 gives higher accuracy in extracting water quality data from satellite images, as shown in the study by Adawiah and. Similarly, in this study, the regression models incorporating B4 achieved the highest R^2 values, which denotes the sensitivity of the red band (B4) to the presence of TSS in inland water bodies. Yanti, Susilo, and Wicaksono (31) also showed the effectiveness of the single band 4 in their study, where TSS estimation using B4 and a linear regression model yielded an R^2 value of 0.543. In this study, the single-band regression model for B4 based on the logarithmic regression model achieved a high accuracy in estimating the concentration of TSS with R^2 value of 0.85. However, Yanti, Susilo and Wicaksono (31) demonstrated that the single band red cannot be used in isolation to effectively estimate TSS in inland water bodies. However, this study demonstrated the effectiveness of the single red band in TSS estimation and mapping.

3.3 Retrieval of chlorophyll-*a* from satellite images

The average in situ Chl-a for the entire sampling period was 46.51 mg/l. One of the main economic activities in the catchment is farming, meaning that the inflow of fertilizer leachate into the catchment could contribute to the high Chl-a concentration. Nutrient flows and algae proliferation in the reservoir increases due to rainy seasons, which facilitate the inflow of nutrients into the reservoir. Furthermore, the dry season that ensues after the rainy season also provides conducive conditions for the incubation of algae (32). Table 4 summarizes the regression equations and the band combinations used for the retrieval of Chl-a concentration.

JDFEWS 5 (1): 159-176, 2024
ISSN 2709-4529

Table 4. Regression equations and band combinations 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	0.80	0.24	3.67
	$y = 7820x^2 – 20734x + 13767$	B1/B3	0.80	0.23	3.09
	$y = 16746x^2 – 53631x + 42964$	(B1/B4) + B1	0.81	0.41	-3.70
11/12/2020	$y = 8797x^2 – 7585x + 1658$	B4/B1	0.68	0.49	10.34
	$y = 7556x^2 – 6881x + 1591$	(B4/B1) +B4	0.68	0.53	10.97
28/01/2021	$y = -8121x^2 + 8983x – 2389$	(B4/B1) + B4	0.92	0.72	16.89
	$y = -9145x^2 + 9639x – 2444$	B4/B1	0.93	0.68	16.23

Chl-*a* was best estimated using the polynomial regression models specifically from the B1, B3, and B4 with R^2 values of 0.80, 0.68, 0.93 for the specific sampling dates (18). Similarly, Jaelani, Limehuwey (33) also obtained comparable results using water quality retrieval algorithms developed using the visible bands, and the models achieved R^2 values greater than 0.5. However, Lai, Zhang (34) showed that combining the visible B2 and the near infrared bands could be the optimal band combination for Chl-*a* data retrieval. On the contrary, in this study, the near-infrared bands did not effectively retrieve Chl-*a* data from this reservoir. The findings were confirmed by Lai, Zhang (34), who found that the combination of near-infrared and blue Landsat-8 OLi bands was not so effective in the retrieval of Chl-*a* from inland water bodies.

3.3 Validation of *in situ* and predicted water quality parameters

The developed regression equations were validated using data from the five (5) sampling stations not used for model development. The validation results are presented in Table 5 (3).

Table 5. Validation results for in situ and estimated water quality parameters (3)

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

JDFEWS 5 (1): 159-176, 2024
ISSN 2709-4529

		Land-sat-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
		Land-sat-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
		Land-sat-8 OLI	13	31.87	83.15	57.02	52.35	14.16	27.06	3.93
28/01/2021	Tur-bidity	*In situ*	13	3.00	17.00	10.00	8.62	3.55	41.18	0.98
		Land-sat-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
		Land-sat-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
		Land-sat-8 OLI	13	29.52	76.22	45.56	49.73	15.89	31.94	4.41

Figure 4(a-c), Figure 5 (a-c), and Figure 6 (a-c) present the laboratory-measured and Landsat-predicted results for turbidity, TSS, and Chl-*a*.

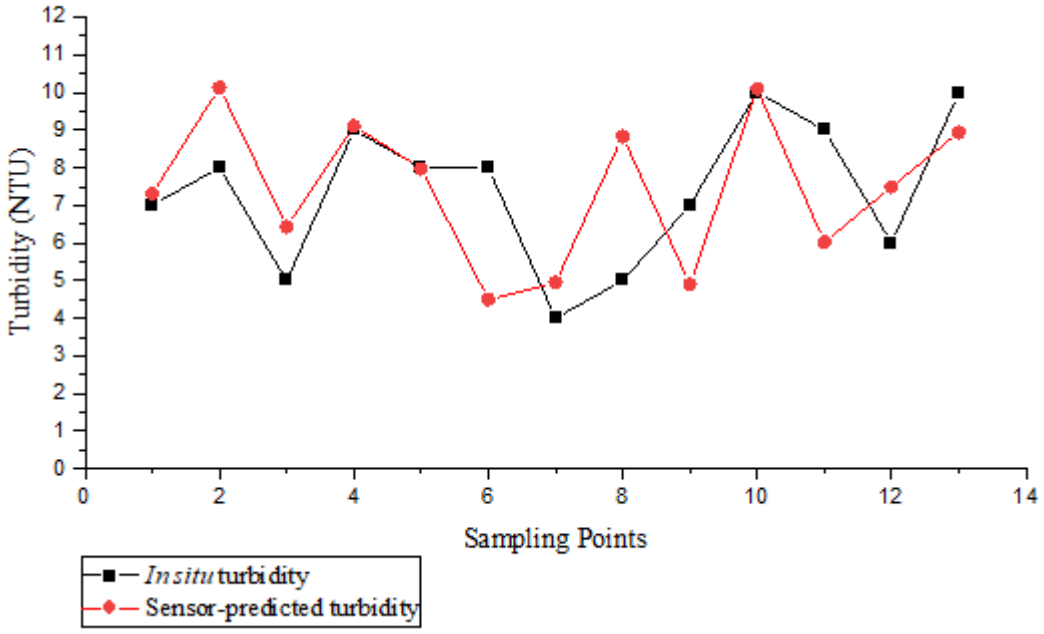

Fig. 4a: *In situ* and sensor-predicted turbidity variations for 25/11/2020.

JDFEWS 5 (1): 159-176, 2024
ISSN 2709-4529

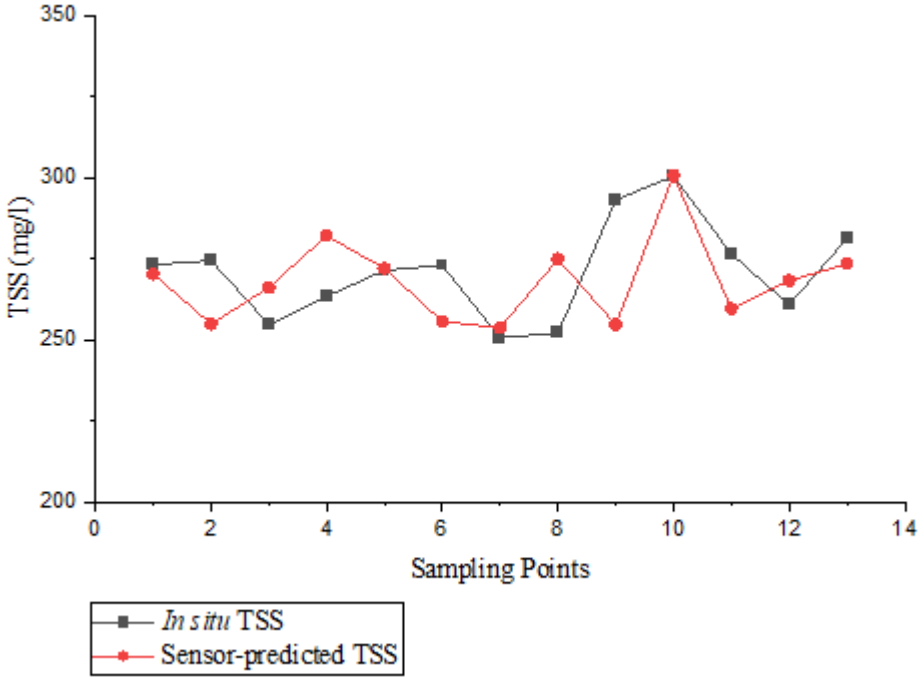

Fig. 4b: *In situ* and sensor-predicted TSS variations for 25/11/2020.

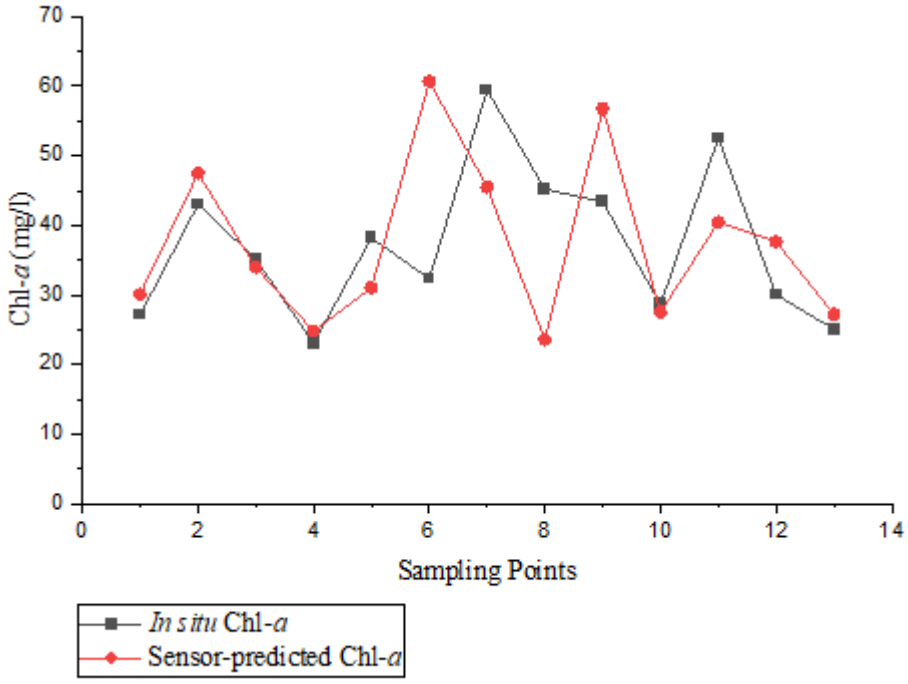

Fig. 4c: *In situ* and sensor-predicted Chl-*a* variations for 25/11/2020.

JDFEWS 5 (1): 159-176, 2024
ISSN 2709-4529

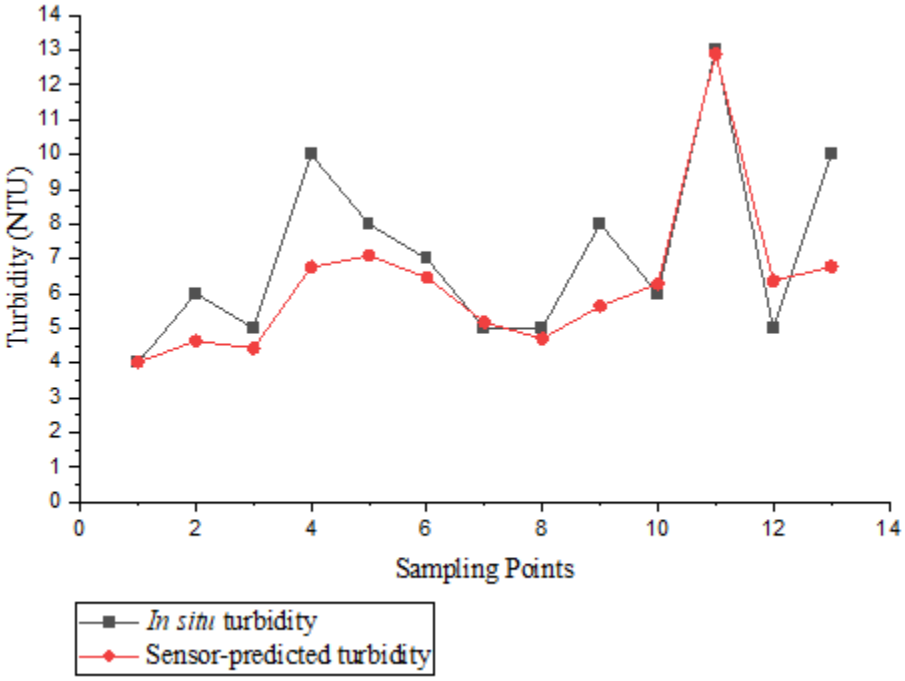

Fig. 5a: *In situ* and sensor-predicted turbidity variations for 11/12/2020.

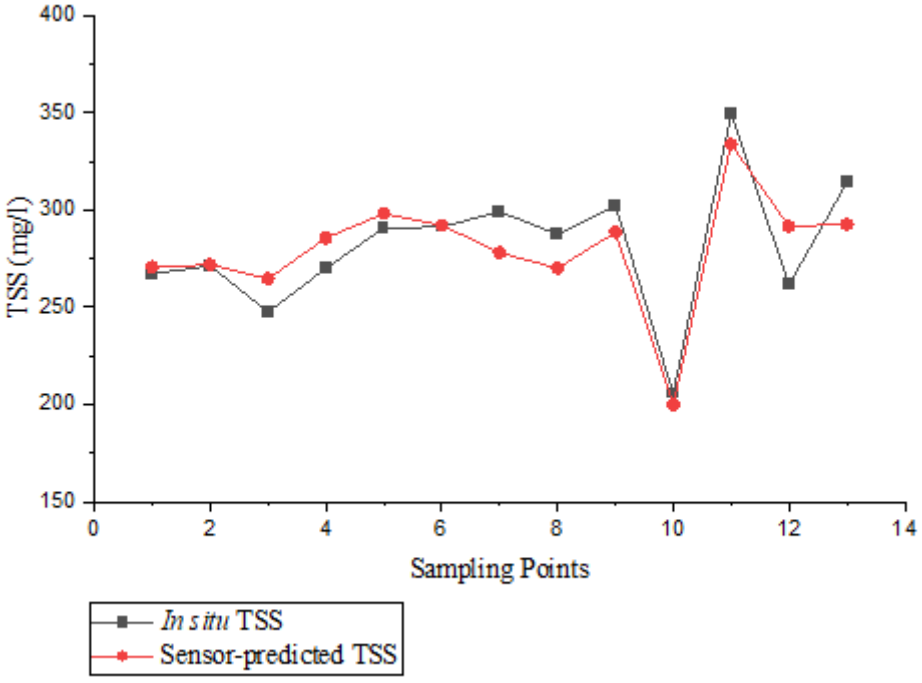

Fig. 5b: *In situ* and sensor-predicted TSS variations for 11/12/2020.

JDFEWS 5 (1): 159-176, 2024
ISSN 2709-4529

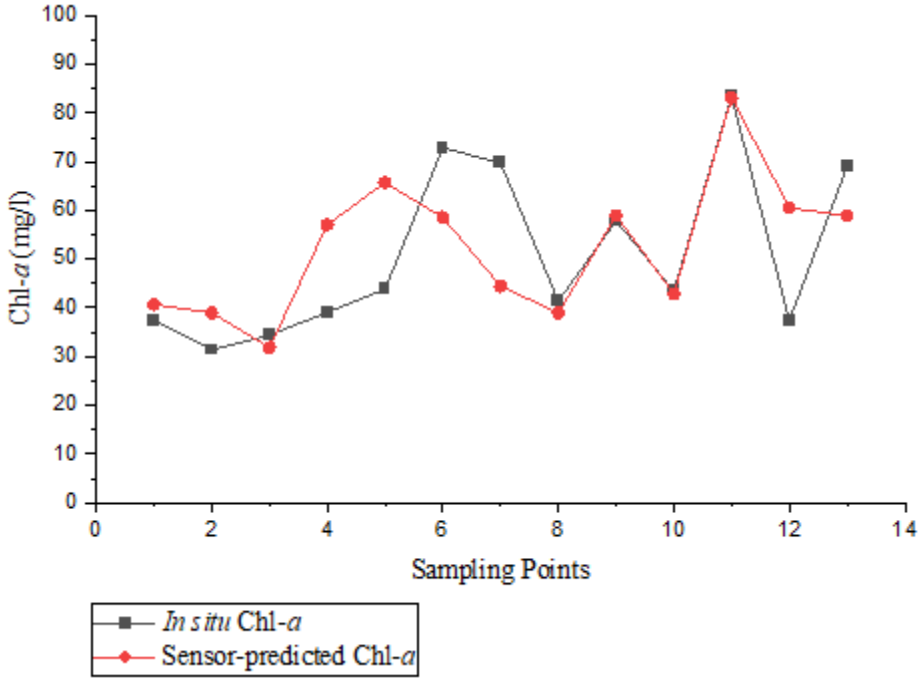

Fig. 5c: *In situ* and sensor-predicted Chl-a variations for 11/12/2020.

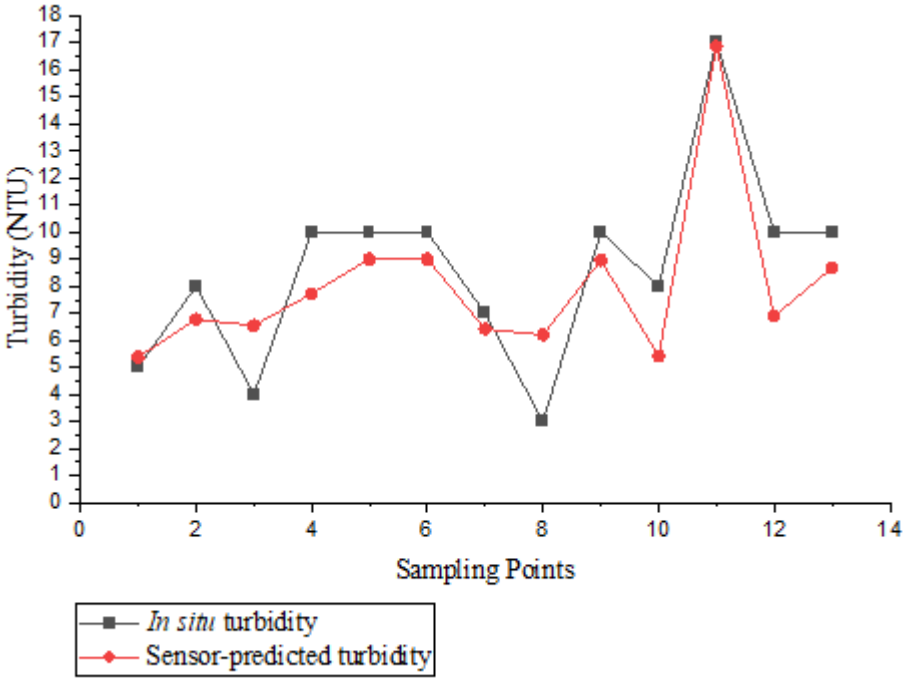

Fig. 6a: *In situ* and sensor-predicted turbidity variations for 28/01/2021.

JDFEWS 5 (1): 159-176, 2024
ISSN 2709-4529

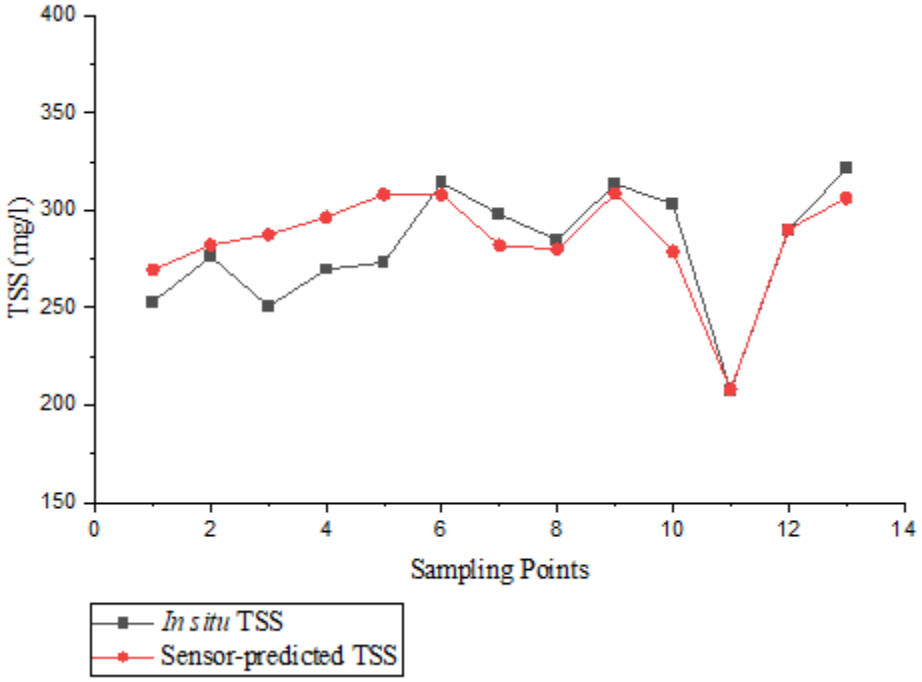

Fig. 6b: *In situ* and sensor-predicted TSS variations for 28/01/2021.

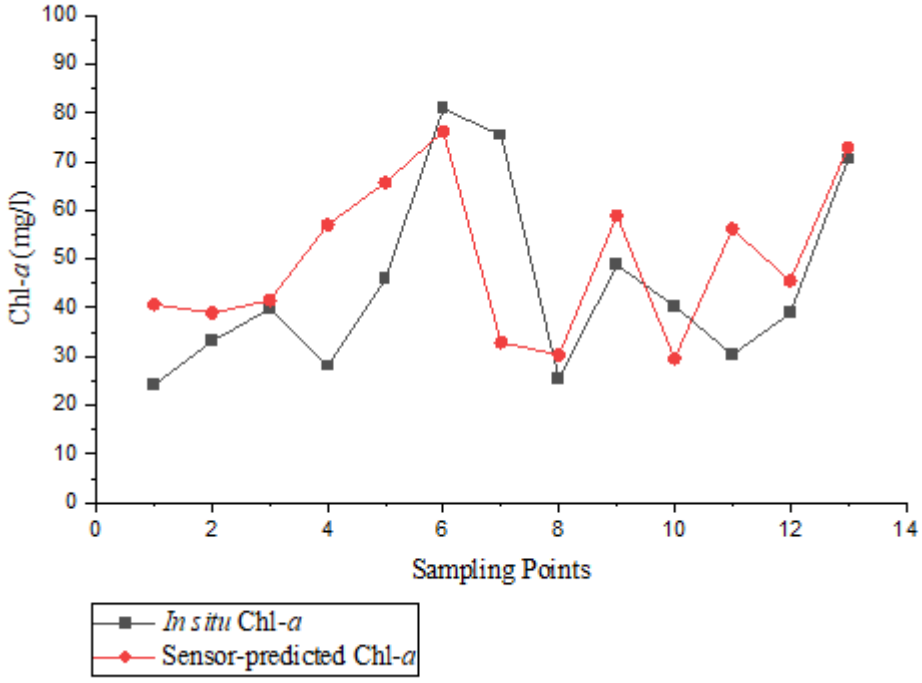

Fig. 6c: *In situ* and sensor-predicted Chl-*a* variations for 28/01/2021.

The use of satellite data to estimate WQPs from the water supply reservoir can be useful in improving the efficiency of the WTPs' operations. This is particularly important because of the changing weather patterns, which largely contribute to the variability of water quality characteristics in the catchment. Therefore, the satellite data can be used to anticipate and plan for WTP processes, mainly, the chemicals used.

Journal of Digital Food, Energy & Water Systems [JD-FEWS]

JDFEWS 5 (1): 159-176, 2024
ISSN 2709-4529

3.5 ANN model 1: Prediction of treated water quality parameters

ANN model 1 results are shown in Figure 7, and Table 6.

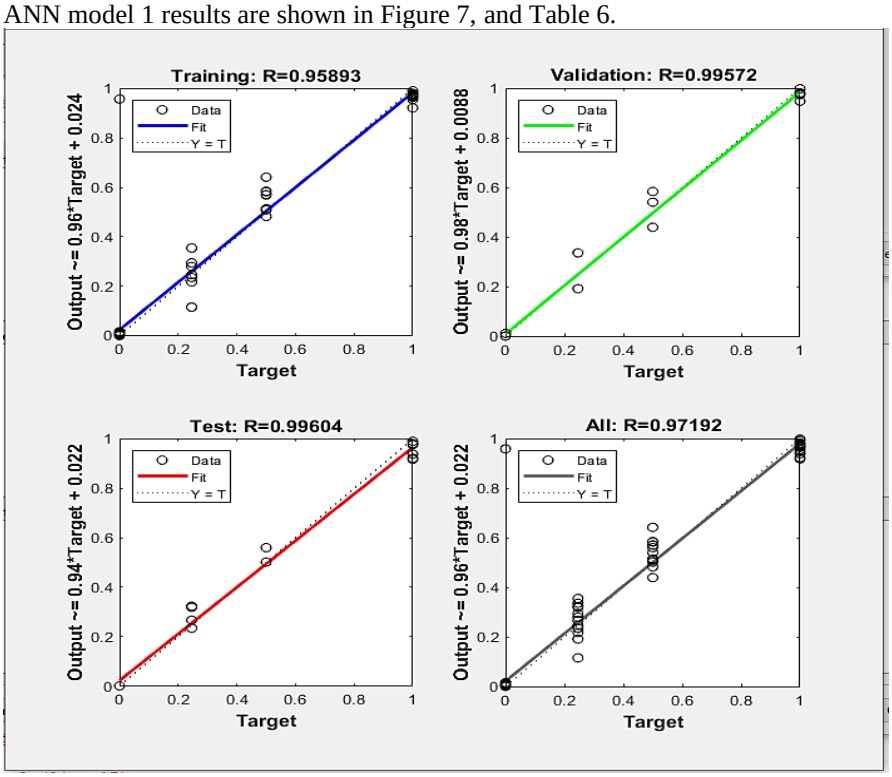

Fig. 7. ANN model 1 predictions of treated water quality parameters

In comparing the in situ and satellite-derived water quality data, the ANN 1 model predictions achieved R^2 values of 0.87, 0.99, and 1.0 for Chl-*a*, turbidity, and TSS, respectively (Table 6). The MSE values from the model were 0.0252, 0, and 0.0026 for Chl-*a*, TSS, and turbidity, respectively. Similar ANN-MLP models were used by Haghiri, Daghighi, and Moharramzadeh (14) and Setshedi, Mutingwende, and Ngqwala (35) to predict treated WQPs. The models learned and understood the data, as evidenced by the predictive accuracies, and achieved R^2 values greater than 0.8 (35). Similar studies by Kennedy, Gandomi, and Miller (13) and Seo, Yun, and Choi (36) also developed an ANN model to predict treated WQPs. The studies also showed that ANN models can be used to predict treated WQPs as a function of the coagulant dose.

The ANN model in this study was used in a WTP, unlike the models in the studies by Setshedi, Mutingwende, and Ngqwala (35), Kennedy, Gandomi and Miller (13), and Seo, Yun, and Choi (36), which were tested on dams and rivers. However, these models were accurate in predicting the WQPs, and this showed the opportunity to extrapolate these models for water quality predictions in a portable WTP.

JDFEWS 5 (1): 159-176, 2024
ISSN 2709-4529

3.6 ANN Model 2: Prediction of Optimum Coagulant Dose

ANN model 2 results are shown in Figure 8.

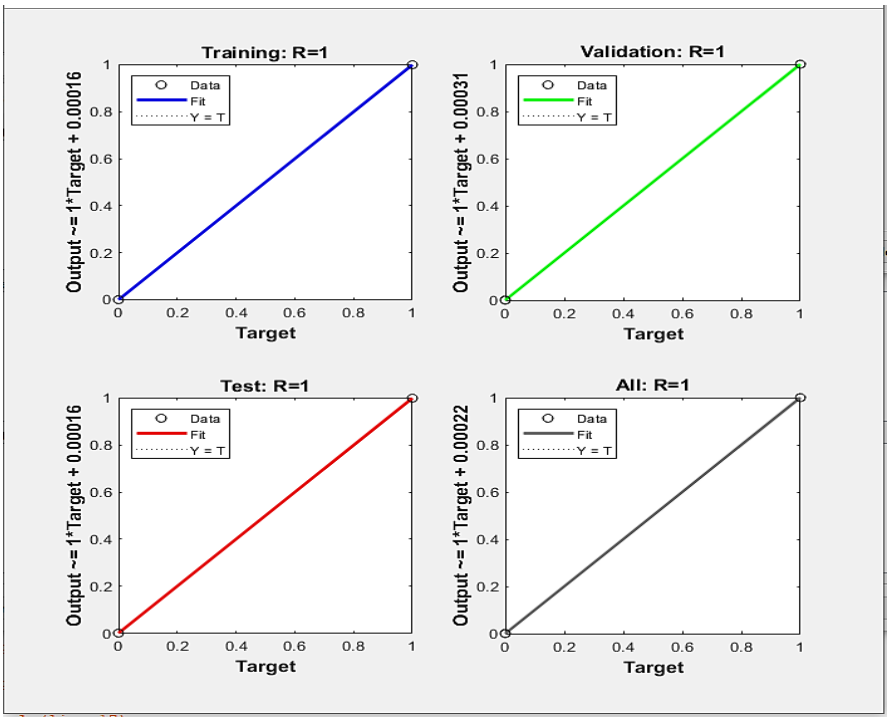

Figure 8: ANN model 2 predictions of optimum coagulant dose

The model predicted the optimum coagulant dose with an R^2 of 0.9987 and an MSE of 0 mg/l. (16) developed an ANN model to mimic jar test experiments in determining the optimum coagulant dose. Compared to the actual coagulant dose for water treatment, the predicted coagulant dose was comparable with R^2 values of 0.81 and MSE of 5.47g/m^3. Haghiri, Daghighi and Moharramzadeh (14) also obtained similar results, who also modeled jar test experiments, and the model predictions achieved an R^2 value of 0.93 and an MSE value of 0.37. (37) also developed an ANN-MLP model with limited data, similar to this study, for optimum coagulant dose prediction in a WTP. The model achieved an R^2 value of 0.8 and an RMSE value of 12.51. The accuracy of the ANN models shows the potential of ANN modeling for coagulant dose forecasting, and the model accuracies can be increased by increasing the dataset. A similar study used data from large databases to predict the optimal coagulant dose (15). The accuracies achieved by these models was relatively high compared to that which can be achieved using small datasets. Furthermore, in using relatively large datasets, the study also showed the possibility of model transfer for use in similar WTPs.

4 Conclusions

The results from this study showed the potential of using Landsat 8-OLI for water quality predictions and estimation of the optimal coagulant dose for water treatment. The visible Band 2, Band 3, and Band 4 gave the highest correlation coefficient between *in situ* and satellite-derived turbidity with R^2 values greater than 0.7. The visible B1, B2, B3, and B4 were ideal for TSS estimation, with R2 values above 0.7 when comparing the *in situ* and satellite-derived water quality data. B1, B3, and B4 were ideal for Chl-a estimation with R2 values above 0.68 when comparing the satellite-derived and in situ water quality data. In comparing the in situ and satellite-derived water quality data, the ANN 1 model predictions achieved R^2 values of 0.87, 0.99, and 1.0 for Chl-*a*, turbidity, and TSS, respectively. The MSE values from the model were 0.0252, 0, and 0.0026 for Chl-*a*, TSS, and turbidity, respectively. ANN model 2 predicted the optimum coagulant dose with an R^2 of 0.9987 and an MSE of 0 mg/l.

JDFEWS 5 (1): 159-176, 2024
ISSN 2709-4529

Thus, the ANN models can improve the efficiency of WTP operations by reducing the need for jar test experiments, especially with highly variable raw water quality characteristics mainly brought about by climate change and anthropogenic activities.

Furthermore, the model accuracies attained also demonstrate the opportunities for ANN modeling in portable WTPs by developing and using ANN models that are adaptive to raw water quality changes. The use of ANN models in combination with satellite data is currently restricted by Landsat 8-OLI eight-day repeat cycle. Furthermore, models carry a certain margin of error, which challenges developing a completely predictive WTP model. However, the ANN models can be developed to run parallel with the main WTP operations.

5 Acknowledgment

We want to acknowledge the Building Capacity in Water Engineering for Addressing Sustainable Development Goals in East Africa (CAWESDEA) project, an initiative of the Global Water Partnership Tanzania, to fund the research work reported herein. We would also like to acknowledge the Eldoret Water and Sanitation Company for hosting the research project. We are greatly indebted to the Renewable Energy, Nanomaterials, and Water (RENWA) research group, under Dr. Achisa's guidance, for the research mentorship.

6 References

[1] Adjovu GE, Stephen H, James D, Ahmad S. Overview of the Application of Remote Sensing in Effective Monitoring of Water Quality Parameters. Remote Sensing. 2023;15(7):1938.

[2] Brooks B. W, Lazorchak J. M, Howard M. D, Johnson M-VV, Morton S. L, Perkins D. A, et al. In some places, in some cases and at some times, harmful algal blooms are the greatest threat to inland water quality. Environmental toxicology and chemistry. 2017;36(5):1125.

[3] Omondi A. N, Ouma Y, Kosgei J. R, Kongo V, Kemboi E. J, Njoroge S. M, et al. Estimation and mapping of water quality parameters using satellite images: a case study of Two Rivers Dam, Kenya. Water Practice and Technology. 2023;18(2):428-43.

[4] Imen S, Chang N-B, Yang Y. J, Golchubian A. Developing a model-based drinking water decision support system featuring remote sensing and fast learning techniques. IEEE systems journal. 2016;12(2):1358-68.

[5] Gholizadeh M. H, Melesse A.M, Reddi L. A comprehensive review on water quality parameters estimation using remote sensing techniques. Sensors. 2016;16(8):1298.

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

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

[8] Patra P, Dubey S.K, Trivedi R.K, Sahu S.K, Rout S.K. 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. 2016.

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

[10] Pizani F.M, Maillard P, Ferreira A.F, de Amorim C.C. Estimation of water quality in a reservoir from Sentinel-2 MSI and Landsat-8 OLI sensors. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences. 2020;3:401-8.

[11] Meng H, Zhang J, Zheng Z. 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. 2022;19(13):7725.

[12] Ouma Y.O, Noor K, Herbert K. 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;2020:1-21.

[13] Kennedy M.J, Gandomi A.H, Miller C.M. Coagulation modeling using artificial neural networks to predict both turbidity and DOM-PARAFAC component removal. Journal of Environmental Chemical Engineering. 2015;3(4):2829-38.

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

JDFEWS 5 (1): 159-176, 2024
ISSN 2709-4529

[15] Baouab M.H, Cherif S. Prediction of the optimal dose of coagulant for various potable water treatment processes through artificial neural network. Journal of Hydroinformatics. 2018;20(6):1215-26.
[16] Kote A.S, Wadkar D.V. Modeling of chlorine and coagulant dose in a water treatment plant by artificial neural networks. Engineering, Technology & Applied Science Research. 2019;9(3):4176-81.
[17] Wilde F.D, Radtke D.B, Gibs J, Iwatsubo R.T. Chapter A1. Preparations for water sampling, USGS. 2005.
[18] Omondi A.N.A, Ouma Y, Mburu N.S, Achisa C.M. Smart Monitoring of Water Supply Treatment Plant Processes using Remote Sensing and Artificial Neural Networks: Moi University; 2023.
[19] Ouma Y.O, Waga J, Okech M, Lavisa O, Mbuthia D. Estimation of reservoir bio-optical water quality parameters using smartphone sensor apps and Landsat ETM+: review and comparative experimental results. Journal of Sensors. 2018;2018.
[20] Da Silva I.N, Hernane Spatti D, Andrade Flauzino R, Liboni L.H.B, Dos Reis Alves S.F, Da Silva I.N, et al. Artificial neural network architectures and training processes: Springer; 2017.
[21] Singh SK, Jain SK, Bárdossy A. Training of artificial neural networks using information-rich data. Hydrology. 2014;1(1):40-62.
[22] Sumi SM, Zaman MF, Hirose H. A rainfall forecasting method using machine learning models and its application to the Fukuoka city case. International Journal of Applied Mathematics and Computer Science. 2012;22(4):841-54.
[23] Jorgensen S, Gromiec M. Developments in environmental modelling: Elsevier; 2016.
[24] Oostwal E, Straat M, Biehl M. Hidden unit specialization in layered neural networks: ReLU vs. sigmoidal activation. Physica A: Statistical Mechanics and its Applications. 2021;564:125517.
[25] Mehdinejad MH, Bina B, Hadian S. The survey of removal of suspended solids from river at flooding period by plain sedimentation process. Advances in Environmental Biology. 2012;6(1):358-61.
[26] Lotfi G, Ahmadi Nadoushan M, Abolhasani M. The feasibility of using Landsat OLI images for water turbidity estimation in Gandoman wetland, Iran. Journal of Radar and Optical Remote Sensing. 2019;2(3.4):49-62.
[27] Kalele A.S. Estimation and mapping of turbidity in the lower Charles River Using Landsat 8 OLI satellite imagery: Northeastern University; 2019.
[28] Hossain A.A, Mathias C, Blanton R. Remote sensing of turbidity in the Tennessee River using Landsat 8 satellite. Remote Sensing. 2021;13(18):3785.
[29] Lobo F.L, Costa M.P, Novo E.M. Time-series analysis of Landsat-MSS/TM/OLI images over Amazonian waters impacted by gold mining activities. Remote Sensing of Environment. 2015;157:170-84.
[30] Adawiah S, Setiawan K, Parwati E, Faristyawan R, editors. Development of empirical model of Total Suspended Solid (TSS) by using Landsat 8 on the Coast of Bekasi Regency. IOP Conference Series: Earth and Environmental Science; 2021: IOP Publishing.
[31] Yanti A, Susilo B, Wicaksono P, editors. The aplication of Landsat 8 OLI for total suspended solid (TSS) mapping in Gajahmungkur reservoir Wonogiri regency 2016. IOP Conference Series: Earth and Environmental Science; 2016: IOP Publishing.
[32] KDHE. Water Quality Standards White Paper: Chlorophyll-a Criteria for Public Water Supply Lakes or Reservoirs. Bureau of Water, KDHE Topeka, Kan.; 2011.
[33] Jaelani L.M, Limehuwey R, Kurniadin N, Pamungkas A, Koenhardono E.S, Sulisetyono A. 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. 2016;27(1).
[34] Lai Y, Zhang J, Song Y, Gong Z. 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. 2021;18(9):4419.
[35] Setshedi K.J, Mutingwende N, Ngqwala N.P. 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. 2021;18(10):5248.
[36] Seo I, Yun S.H, Choi S.Y. Forecasting water quality parameters by ANN model using pre-processing technique at the downstream of Cheongpyeong Dam. Procedia Engineering. 2016;154:1110-5.
[37] Sengul A, Gormez Z, editors. Prediction of optimal coagulant dosage in drinking water treatment by artificial neural network. 1st EWAS-MED International Conference: Improving Efficiency of Water Systems in a Changing Natural and Financial Environment; 2013.