Water Practice & Technology

IWA PUBLISHING

Water Practice & Technology Vol 18 No 2, 428 doi: 10.2166/wpt.2023.010

Estimation and mapping of water quality parameters using satellite images: a case study of Two Rivers Dam, Kenya

Alice Nureen Omondi [a,*], Yashon Ouma[a], Job Rotich Kosgei[a], Victor Kongo[b], Ednah Jelagat Kemboi[a], Simon Mburu Njoroge[a], Achisa Cleophas Mecha[c] and Emmanuel Chessum Kipkorir[a]

[a] Department of Civil and Structural Engineering, Moi University, P.O. Box 3900-30100, Eldoret, Kenya
[b] Global Water Partnership (SA)/Tanzania Water Partnership, P.O. Box 32334, Dar es Salaam, Tanzania
[c] Department of Chemical and Processing Engineering, Moi University, P.O. Box 3900-30100, Eldoret, Kenya
*Corresponding author. E-mail: nureenlol@gmail.com

ANO, 0000-0002-9544-3757

ABSTRACT

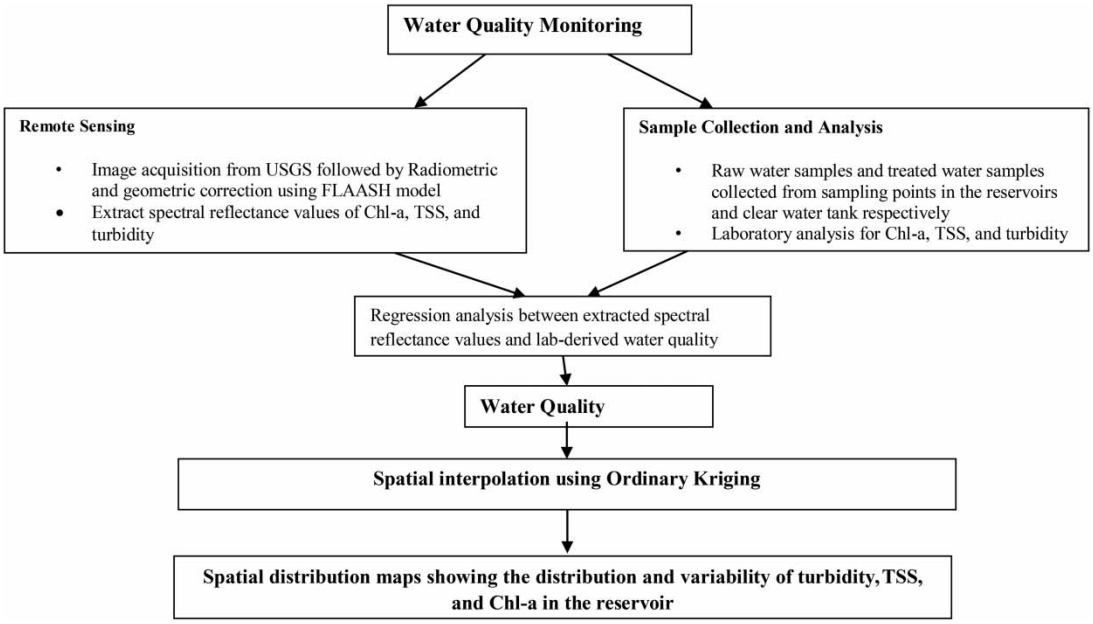
The continuous water quality monitoring (WQM) of watersheds and the existing water supplies is a crucial step in realizing sustainable water development and management. However, the conventional approaches are time-consuming, labor intensive, and do not give spatial–temporal variations of the water quality indices. The advancements in remote sensing techniques have enabled WQM over larger temporal and spatial scales. This study used satellite images and an empirical multivariate regression model (EMRM) to estimate chlorophyll-a (Chl-*a*), total suspended solids (TSS), and turbidity. Furthermore, ordinary Kriging was applied to generate spatial maps showing the distribution of water quality parameters (WQPs). For all the samples, turbidity was estimated with an R^2 and Pearson correlation coefficient (r) of 0.763 and 0.818, respectively while TSS estimation gave respective R^2 and r values of 0.809 and 0.721. Chl-*a* was estimated with accuracies of R^2 and r of 0.803 and 0.731, respectively. Based on the results, this study concluded that WQPs provide a spatial–temporal view of the water quality in time and space that can be retrieved from satellite data products with reasonable accuracy.

Key words: Chlorophyll-*a*, empirical multivariate regression modeling, Landsat-8, TSS, turbidity

HIGHLIGHTS

- Remote sensing could avail a cost-effective option for the continuous monitoring of watersheds and water resources.
- Satellite-derived data could inform water quality monitoring decisions.
- Ordinary Kriging enabled the development of water quality spatial distribution maps for the water supply reservoir.
- An empirical multivariate regression modeling (EMRM) approach is used for the development of model coefficients.

GRAPHICAL ABSTRACT



1. INTRODUCTION

The world's population is fast growing and this keeps on pushing more people to settle in fragile ecosystems mainly the shores of lakes and rivers, and coastal regions (Bar-Massada *et al.* 2014). The increasing anthropogenic activities in these areas are a major threat to most of the world's freshwater resources (Vörösmarty *et al.* 2010) and could create irreversible negative impacts, especially in a changing climate. Continuous watershed monitoring is a crucial concept that will help realize sustainable water supplies. According to Najafzadeh & Niazmardi (2021), the quality of surface water plays a key role in the sustainability of ecological systems. Measuring water quality parameters (WQPs) is of high importance in the management of surface water resources. Furthermore, the treatment and supply of water to meet the needs of the various end users require an understanding of the real-time quality of water at the source which influences the choice of chemicals and quantities used in the treatment process.

Eldoret is a fast-growing town with an estimated urban population of 475,716 in 2019 and a growth rate of 3.82% per annum. According to the Eldoret Water and Sanitation (ELDOWAS) Company, the demand for water in Eldoret Municipality is estimated at 60,000 m^3/day, against 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 the recent conversion of forested land into subsistence agriculture. This has led to flash floods, erosion, and sedimentation which decrease the quality of surface water. In addition, competing users and uses have contributed to a substantial increase in freshwater requirements. Furthermore, factors such as climate change, population growth, and inadequate conservation practices in the catchment negatively impact the water quality as exhibited through increased turbidity and algal blooms (Ontumbi *et al.* 2015; Barasa & Perera 2018).

Monitoring surface water quality is crucial mainly in the context of increasing freshwater demands and wastewater discharged to the environment (Chen & Han 2018). The traditional approach of water quality monitoring (WQM) entails the collection of samples in the field followed by a water quality analysis in the laboratory. However, the approach is time-consuming, labor intensive, and it does not give the spatial-temporal variations of the water quality indices (WQIs). Furthermore, reliance on conventional methods also limits the possibility of monitoring, forecasting, and managing entire water bodies due to the large extent of the water surface, lack of spatial-temporal data on a regional scale, and geographical limitations (Gholizadeh *et al.* 2016).

According to Najafzadeh *et al.* (2021), WQIs are crucial in describing the essential characteristics of water pollutants and this creates the need for accurate predictions of WQIs in order to gain insights into the patterns of pollutants in natural streams. Furthermore, Najafzadeh *et al.* (2021) also note that one of the most difficult

issues in the 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 (DDMs) in assessing the WQIs of natural streams. The WQM challenges can also be overcome by using satellite images which avail a smart, rapid, and low-cost WQM tool.

The advancements in computer science and remote sensing techniques have made remote sensing find wider applications in WQM (Usali & Ismail 2010; Gholizadeh *et al.* 2016). The use of remote sensing techniques and satellite images allows for continuous WQM over larger spatial and temporal scales thus improving the water management practices for vast geographical areas (Japitana & Burce 2019). The concept could also be extended to determine the impacts of anthropogenic activities in different catchments, pollution management, and watershed management. The development of WQM systems that incorporate remote sensing increases the efficiency with which individuals respond to emergency ecological challenges such as point and non-point pollution, algal blooms, and floods. Real-time measurements also enable data to be analyzed rapidly and effectively while limiting errors that come with sample collection and laboratory analysis.

Ouma *et al.* (2018) note that empirical models leverage bivariate and/or multiple regressions between data acquired from sensors and WQPs measured *in situ* by correlating the sensor radiance values and their band combinations with WQPs collected and measured based on the sensor overpass schedule. For instance, empirical multivariate regression modeling (EMRM) simulations are done to determine the multivariate correlations between the reflectance from sensor bands and WQPs measured *in situ*. Furthermore, Najafzadeh *et al.* (2018) also highlight the need for evolutionary computing-based formulations including the application of equations extracted from gene expressive programming, and evolutionary polynomial regression (EPR) in the prediction of WQPs. This is in line with the increasing recommendation over the last decade to use artificial intelligence models in the prediction of WQPs.

Landsat-8 Operational Land Imager (OLI) is one of the intelligent tools that avails a simple, automated, fast, inexpensive, and noninvasive technology for operational and productive aquatic environmental monitoring (Garaba *et al.* 2015). Furthermore, the assessment of Landsat-8 imagery also allows for the identification of the optically active water constituents based on their interaction with light and the subsequent energy change of the incident radiation reflected from the water body (Garaba *et al.* 2015). Numerous algorithms have been developed based on Landsat data for the retrieval of WQP values from remotely sensed imagery. Thus, Landsat-8 OLI is one of the tools that could be used to enable accurate and routine monitoring of water bodies in line with the need for sustainable management of water resources.

According to Ouabo *et al.* (2020) suitable interpolation techniques can be applied to 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. A study by Murphy *et al.* (2010) compared the performance of ordinary Kriging, inverse distance weighting (IDW), and universal Kriging for spatial interpolation of WQPs. The Kriging-based methods gave better estimates compared to the IDW method with an accuracy of more than 10%.

This study presents a smart WQM approach for the assessment and estimation of WQPs (turbidity, total suspended solids (TSS), and chlorophyll-*a* (Chl-a)) in Two Rivers Dam, Uasin Gishu County in Kenya. The study used satellite images acquired from Landsat-8 OLI. Ordinary Kriging was applied to estimate the WQPs at the unsampled locations in the reservoir from which the spatial distribution maps were developed.

2. MATERIALS AND METHODS

2.1. Study area

Two Rivers Dam is located in Uasin Gishu County, Kenya at a longitude of 35° 35′ 14″ and latitude of 0° 46′ 88″ as shown in Figure 1. The dam is one of three dams that serve Eldoret town and its surroundings and is situated at the confluence of River Endoroto and River Ellegerini and which are both headwaters of River Sosiani. The other sources are Elligrini and Chebara Dams.

The data used in this study were collected from the 13 sampling points, L1–L13 (Figure 1). A high concentration of points was near the edges because of the highly variable water quality characteristics at these locations.

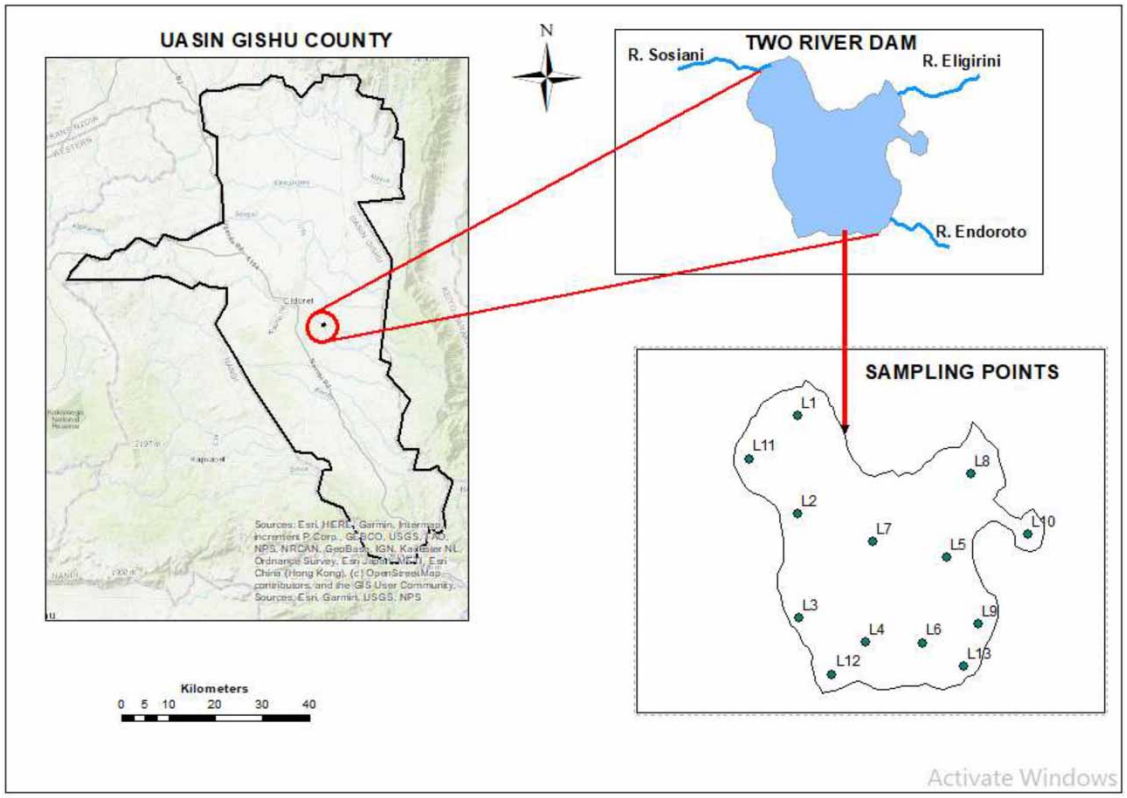


Figure 1 | Two Rivers Dam reservoir and the selected sampling points.

2.2. Landsat image acquisition and processing

The study used Landsat-8 OLI images acquired between 25 November 2020 and 28 January 2021. The selection of *in situ* sampling dates coincided with the satellite overpass schedule to a tolerance of ± 1 day.

2.3. Determination of radiance and reflectance

Image processing was done using the Fast Line-of-sight Atmospheric Analysis of Hypercubes (FLAASH) model where the digital numbers were converted to top-of-atmosphere (TOA) radiance through radiometric calibration. The TOA radiance was then converted to TOA reflectance through atmospheric correction and the surface reflectance was obtained after Dark Object Subtraction (DOS) (Ouma *et al.* 2018). After the DOS, the region of interest (ROI) for each of the images was then extracted and used in the subsequent processing steps. The Two Rivers Dam shapefile was then used to extract the ROI from the processed images and the ROI used for the subsequent processing.

2.4. Laboratory WQP determination

A total of 78 water samples were collected for the entire sampling duration and the standard laboratory protocols were applied in determining the concentration of turbidity, TSS, and Chl-*a*. For each of the three sampling dates, two water samples were collected at each of the 13 points. Each sample and its replicate were then tested for turbidity, TSS, and Chl-*a*, and the average value was recorded. Turbidity measurements were done in the laboratory using the Hanna Portable turbidity meter (model HI98703) and TSS was determined by the gravimetric method (APHA 1975). 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 630 nm) and the resulting concentration was determined based on the SCOR-UNESCO's equations (SCOR-UNESCO 1966). 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.

2.5. Correlation of spectral reflectance with WQPs

The correlation analysis entailed overlaying the sampling points on the ROI extracted from the processed satellite images for each of the sample collection days. An average spectral reflectance of 3 × 3 pixel neighborhood configuration as proposed by Reddy (1997) was used in order to reduce errors in locating the sampling sites, correlate the reflectance and the WQPs as well as address the high water quality variability since the sampling points were close to the edges of the reservoir. Thus, the 3 × 3 window could include the shallow water near the banks. To convert the surface reflectance values to remote sensing reflectance (Rrs), the surface reflectance values were divided by π (Moses *et al.* 2015).

2.6. WQPs estimation using empirical regression modelling

The EMRM approach was used to correlate remote sensing reflectance and the WQPs measured *in situ* as described in Ouma *et al.* (2018). The band combinations considered for EWRM analysis were single band, band ratio, linear, and mixed combination.

The predicted and laboratory-measured water quality values were compared based on the EWRM algorithm and the equations with the highest R^2 values were selected (Wang *et al.* 2006). Eight sampling points were used in the development of the model. The appropriate regression equation was selected based on the value of the coefficient of correlation (R) which was used as a measure of accuracy for the derived equations and five points were used for model validation. The accuracy of the regression results was then determined using the coefficient of determination, Pearson correlation coefficient, mean absolute error or bias, and normalized root mean square error (NRMSE) estimators.

3. RESULTS AND DISCUSSION

3.1. Comparisons between spectral reflectance values and *in situ* WQPs

3.1.1. Turbidity

The average *in situ* turbidity for the entire sampling period varied between 4 and 17 NTU with an average of 7.69 NTU. The reservoir turbidity was generally low since sampling was done during the dry season. 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 been reduced as a result of plain sedimentation which refers to the quiescent settling of water in a reservoir for extended durations without the aid of chemicals especially 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. The reflectance from the blue, green, and red bands yielded the highest correlation coefficient between *in situ* and Landsat-derived water quality values as shown in Table 1. Ouma *et al.* (2020) also demonstrated that turbidity could be estimated using remote sensing by utilizing the green, blue, and red bands of Landsat-8 OLI. Similar results were also obtained by Lotfi *et al.* (2019) with the highest correlation obtained between the reflectance values of red and blue bands and *in situ* turbidity. The significance of the red and blue band in the estimation of turbidity is also emphasized in a study by Kalele (2019) where the best-performing model

Table 1 | Regression equations for turbidity, Chl-*a*, and TSS estimation (B2 = blue, B3 = green and B4 = red)

		Regression equation	Band Combination	R^2	R	nRMSE	Bias
25/11/2020	Turbidity	$y = -1{,}169x^2 + 3{,}694x - 2{,}908$	(B1/B4) + B2	0.797	0.720	0.257	0.084
	TSS	$y = -5{,}340\ln(x) + 2{,}754$	B3/B2	0.788	−0.808	0.704	−9.981
	Chl-*a*	$y = 7{,}820x^2 - 20{,}734x + 13{,}767$	B1/B3	0.802	0.854	0.227	3.092
11/12/2020	Turbidity	$y = 68{,}165x^2 - 15{,}713x + 908.2$	B3 + B4 + B1	0.757	0.886	0.631	−1.572
	TSS	$y = 635.9e^{53.65x}$	B4	0.853	0.723	0.376	2.660
	Chl-*a*	$y = 7{,}556x^2 - 6{,}881x + 1{,}591$	(B4/B1) +B4	0.682	0.648	0.525	10.966
28/01/2021	Turbidity	$y = 29.02\ln(x) + 117.7$	B4	0.688	0.620	0.364	−0.402
	TSS	$y = -6{,}131x^2 + 25{,}640x - 23{,}721$	(B1/B4) + B1	0.757	0.700	0.497	15.862
	Chl-*a*	$y = -9{,}145x^2 + 9{,}639x - 2{,}444$	B4/B1	0.926	0.691	0.676	16.234

was a combination of the reflectance values of the red and blue bands with R and RMSE values of 0.841 and 0.828, respectively. Furthermore, the model validation dataset resulted in R and RMSE values of 0.832 and 0.430, respectively. The results from this study show that turbidity for inland waters can be estimated using remote sensing reflectance values from the visible bands of the satellite imagery.

3.1.2. Total suspended solids

The average *in situ* TSS for the entire sampling period varied between 247 and 321 mg/L with an average of 277.91 mg/L. The highest concentration of TSS was recorded at the points where River Endoroto and River Ellegerini entered the reservoir since the inflow from the two rivers agitated and suspended the settled sediments from the bottom of the reservoir. Based on the EWRM algorithmic approach shown in Table 1, TSS was best estimated from the coastal aerosol, blue, green, and red bands.

The results in Table 1 relating B3 (green) and B2 (blue) gave an R^2 value of 0.788. This can be compared with the results of Jaelani *et al.* (2016) where the logarithmic regression algorithm based on the band ratio of the remote sensing reflectance of B2 (blue) to B3 (green) gave an R^2 value of 0.79. The concept was also established in a study by Ouma *et al.* (2020) where the liner regression model from the band ratio between the B3 (green) and B2 (blue) resulted in an R^2 value of 0.9249. From this study, the single B4 (red) regression model yielded the highest R^2 value of 0.853. This can be compared to the study by Yanti *et al.* (2016) where it was established that the estimation and mapping of TSS concentration can be done using a single B4 (red) band. In the study, the single band linear regression model relating *in situ* TSS to remote sensing reflectance (Rrs) of B4 (red) gave an R^2 value of 0.5431. 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. The results from this and similar studies show that TSS can be estimated with relatively high accuracy from the visible bands of satellite imagery. 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 the TSS concentration of inland waters.

3.1.3. Chlorophyll-*a*

During the entire sampling period, the average *in situ* Chl-*a* ranged between 23.58 and 83.15 mg/L with an average of 46.51 mg/L. From Table 1, the highest concentration of Chl-*a* was recorded in the same regions where high values of TSS were observed. 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 thus leading to the observed high concentration of both TSS and Chl-*a* in the same regions.

The Rrs based on coastal aerosol, green, and red bands gave the best estimate of Chl-*a*. Similar results were also obtained by Jaelani *et al.* (2016) where Chl-*a* concentration retrieval algorithms based on band ratios involving B1, B2, B3, and B4 gave a high determination coefficient ($R^2 > 0.5$). The study by Watanabe *et al.* (2015) where the Rrs spectra computed from *in situ* data showed high absorption at the blue (B2) and red (B4) spectral regions with the reflection peak being at the green (B3) region and the beginning of the near-infrared region close to the end of the red edge (B4) also confirms that B2, B3, and B4 can be used to estimate Chl-*a* from Landsat OLI images. 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 acknowledge that if only the near-infrared and blue bands are used for Chl-*a* retrieval, then the correlation is not ideal. Like most of the cited studies, the current study showed that satellite data such as that from Landsat-8 OLI coastal aerosol, blue, green, and red can

be used for estimating and monitoring the seasonal variations of Chl-*a* in reservoirs and this could help detect, in advance, the occurrence of possible algal blooms.

3.2. Model validation using estimated and *in situ* water quality measurements

The validation of the developed regression algorithms was done using data from five sampling stations (L1, L2, L4, L5, and L9 in Figure 1). Table 2 presents calibration and validation results.

Table 2 | Validation results for *in situ* and estimated water quality measurements

	Water quality parameter	Estimation method	Sample (*n*)	Min.	Max.	Med.	Avg.	SD	CV (%)	SE
25/11/2020	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
		Landsat-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 DDM and DDMs have been frequently used to assess the water quality index (WQI) for natural streams (Najafzadeh *et al.* 2021). The results from all the sampling dates show that TSS had the highest variation in concentration followed by Chl-a and turbidity. Based on the standard deviation (SD), coefficient of variance (CV), and standard error (SE) metrics, the satellite imagery tended to mostly underestimate the concentration of the WQPs but with a very small margin as seen through the low coefficient of variation values. The accuracies obtained based on the EMRM can be compared to the results obtained by Najafzadeh *et al.* (2021) who used four well-known DDMs including 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 was controlled through techniques like Forward Selection (FS), and Gamma Test and the FS-M5 MT gave the best estimate of the WQI. Even though in this study the number of input variables was 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 2. This shows that the EMRM approach is a reliable DDM for the estimation of WQPs from Landsat-8 OLI. Najafzadeh & Niazmardi (2021) also developed a Multiple Kernel-Support Vector Regression (MKSVR) algorithm to estimate chemical oxygen demand (COD) and biological oxygen demand (BOD) of Karun River, Iran using different WQPs as input variables. The MKSVR model performed best in estimating BOD with a correlation coefficient R of 0.8 and RMSE of 4.76 mg/l. In comparison with this study, the MKSVR model is more accurate since the EWRM algorithm gave lower correlation coefficient values as shown in Table 1. Even so, the relatively high accuracies could also be attributed to the high number of input parameters considered all of which affect water quality. The idea of using more input parameters to increase the accuracy of WQPs predictions is also supported by the findings from the study by Najafzadeh *et al.* (2018) where nine input parameters (specifically, Ca^{2+}, Na^{+}, Mg^{2+}, NO^{2-}, NO_3^-, PO_4^{3-}, EC, PH, and turbidity) were used to estimate BOD, COD, and dissolved oxygen using evolutionary computing-based formulations. All three models tested achieved high-performance accuracies as indicated by correlation coefficient (R) of 0.86, 0.76, and 0.84 for GEP, MT, and EPR, respectively.

3.3. Graphical analysis of *in situ* and Landsat-estimated validated results

A graphical analysis of *in situ* and Landsat-estimated validation results for turbidity, TSS, and Chl-*a* for the respective data collection dates are presented in Figures 2–4. The error bars represent the coefficient of variation values between Landsat-estimated and *in situ* results. The results from Figure 3 show that some Landsat-predicted measurements coincided with the actual measured values mainly for TSS where the coefficient of variation for the Landsat-estimated measurements was 11%. However, there were slight variations for some points where Landsat OLI either overestimated or underestimated the predicted values mostly with a small error margin since the coefficient of variation was less than 30% for almost all the points.

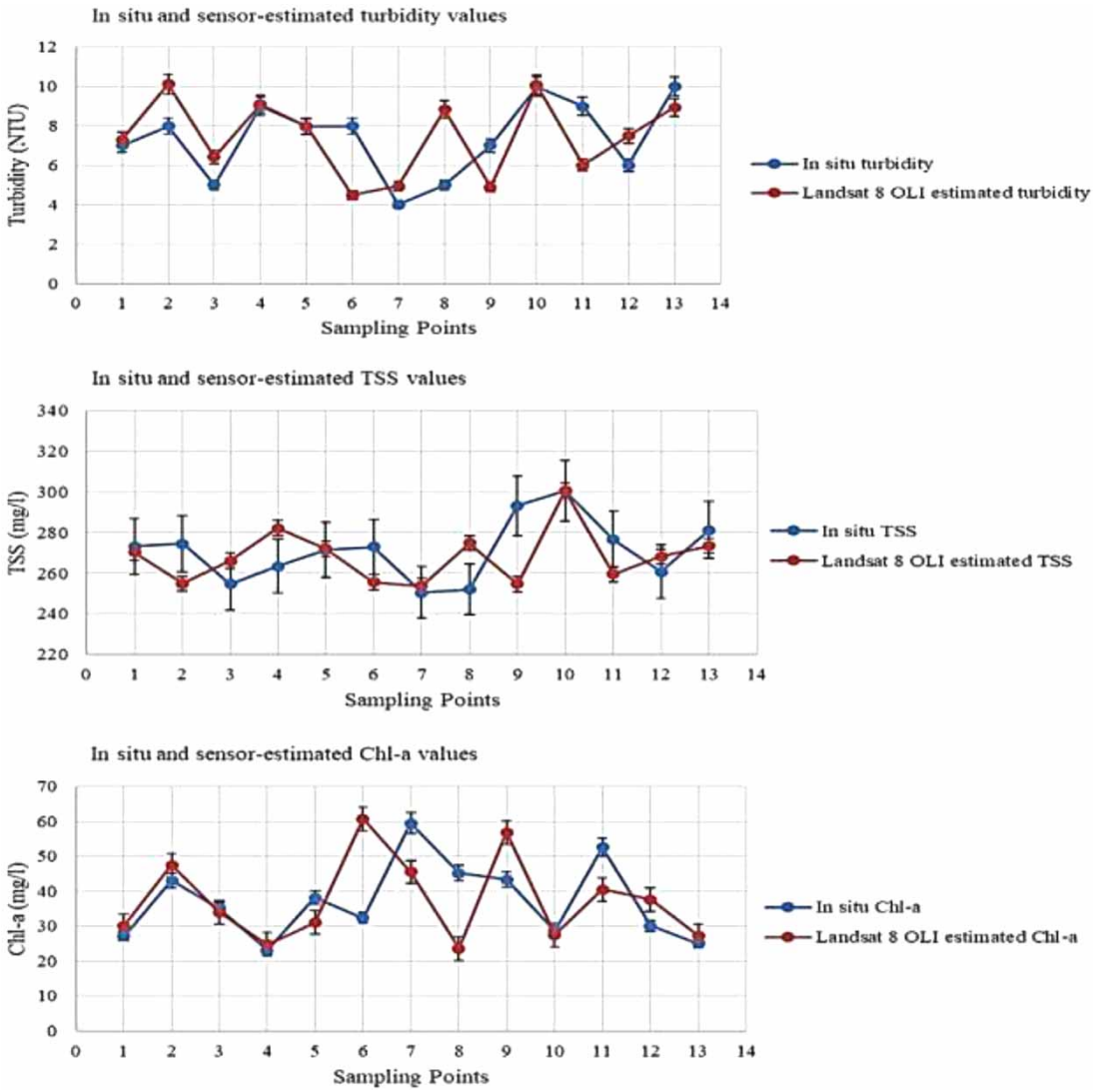


Figure 2 | Turbidity, TSS, and Chl-*a* variations for data collected on 25/11/2020.

In Figure 3, the actual and predicted values followed a similar trend line with most points coinciding with 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 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%.

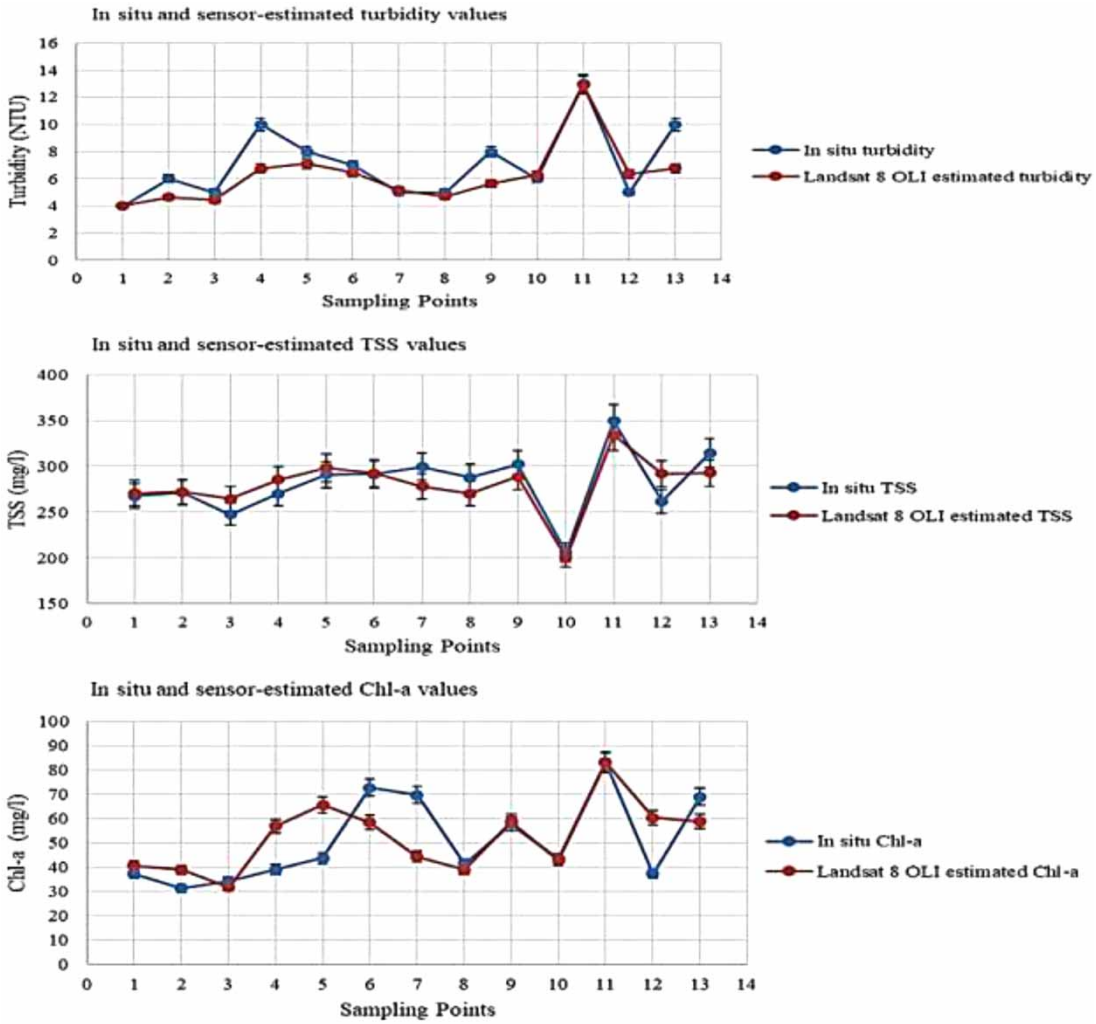


Figure 3 | Turbidity, TSS, and Chl-*a* variations for data collected on 11/12/2020.

The model results can be compared with the results from the study by Nafsin & 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.

Overall, Landsat-8 OLI performed well in the prediction of WQPs with reasonable variation based on the SD, CV, and SE metrics. Generally, the effectiveness of atmospheric correction plays a great role in dictating the accuracy of water quality modeling (Bonansea *et al.* 2019). It is clear that satellite images like those obtained from Landsat-8 OLI can be used as a cost-effective and high-frequency tool for monitoring the water quality of inland waters if adequate radiometric and atmospheric corrections are done.

3.4. Spatial distribution and variability of *in situ* and estimated WQPs

It is important to determine the spatial distribution of the WQPs in order to visualize the variation in water quality for the entire reservoir from the sampled locations. The spatial maps for the distribution and variability of the observed and estimated WQPs were developed using ordinary Kriging to enable further model performance analysis.

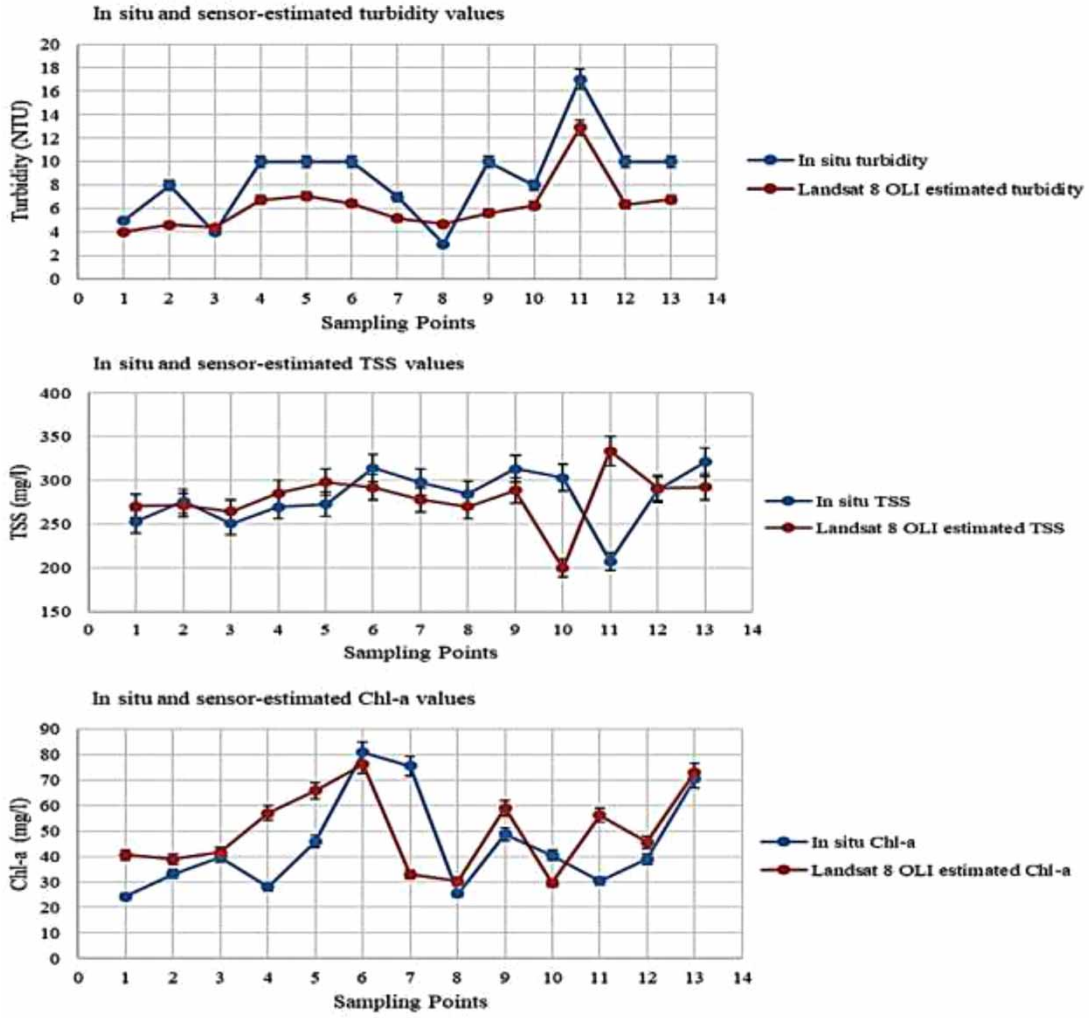


Figure 4 | Turbidity, TSS, and Chl-*a* variations for data collected on 28/01/2021.

3.4.1. Turbidity distribution

The measured and estimated turbidity distribution and variation for the selected dates are shown in Figure 5. The *in situ* measured and Landsat-estimated turbidity distributions specifically for dates 11/12/2020 and 28/01/2021 were closely correlated. Notably, the highest concentration points were at (L10 and L13 for data collected on 25/11/2020 and L11 for data collected on both 11/12/2020 and 28/01/2021). However, from the statistical results, the regression models for all 3 days estimated the turbidity in the reservoir with an accuracy of more than 0.7. This means that the discrepancy noted for the measured and estimated turbidity on 25 November 2020 through a shift in the highest concentration point from point L11 in the laboratory-derived spatial map to point L2 in Landsat-predicted map could be attributed to the Landsat's overestimation and underestimation of predicted turbidity at points L1 and L11, respectively. 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.

3.4.2. TSS distribution

The spatial maps for TSS distribution in the dam show a high correlation between the laboratory measurements and estimations from Landsat OLI, as illustrated in Figure 6. The highest TSS concentration for both laboratory and in situ measurements were recorded at sampling points L10 (L11 and L13) and (L4, L9, and L13) on 25 November 2020, 11 December 2020, and 28 January 2021, respectively. This means that for each of the data collection days, the peak locations of high TSS concentrations are in near-perfect coincidence between measured

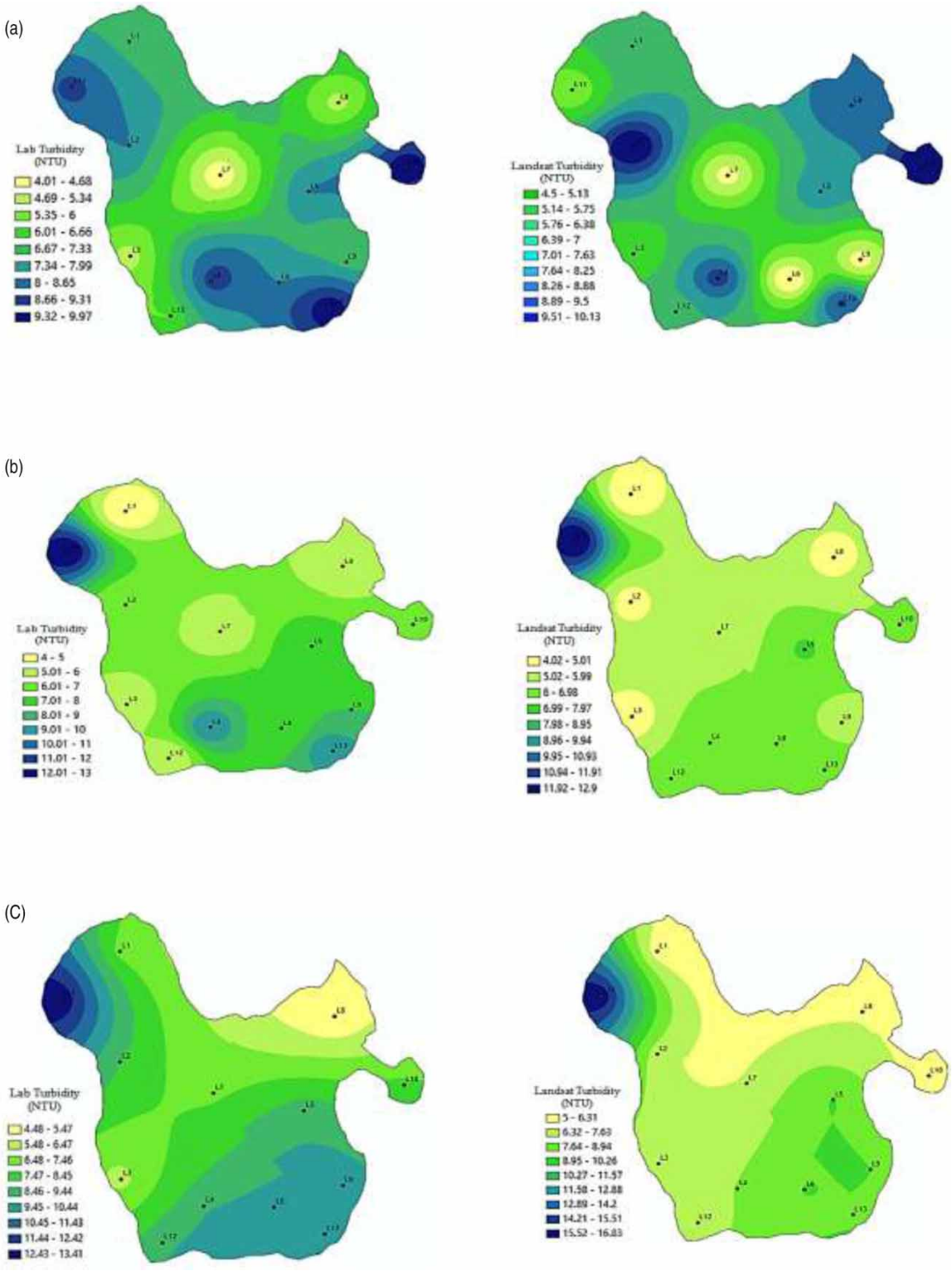


Figure 5 | Measured and estimated turbidity distribution and variation from 25/11/2020 to 28/01/2021. (a) Measured and estimated turbidity measurements on 25 November 2020. (b) Measured and estimated turbidity measurements on 11 December 2020. (c) Measured and estimated turbidity on 28 January 2021.

values and satellite-derived estimates. Notably, almost all the places that have a high TSS concentration also have a 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

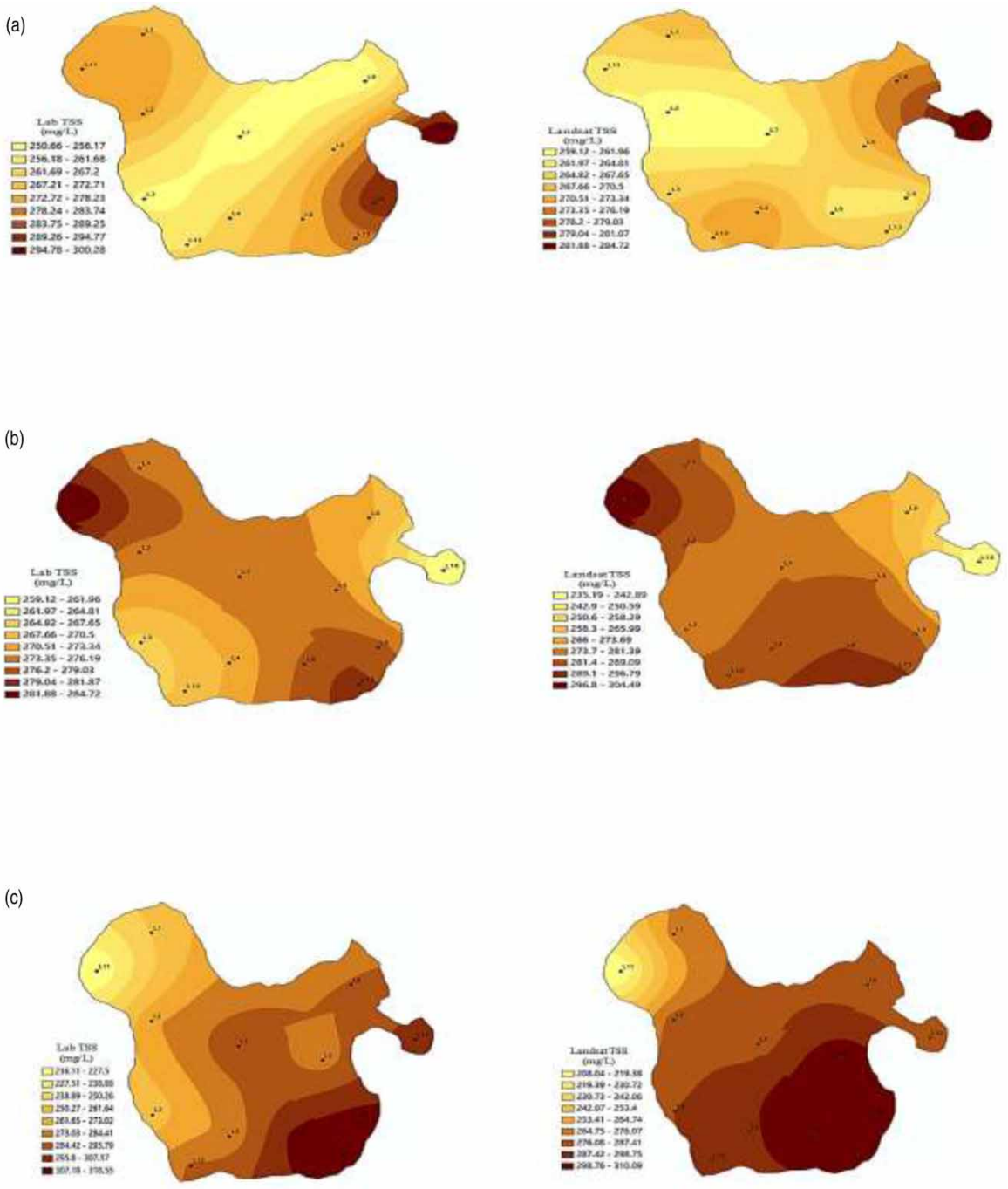


Figure 6 | Measured and estimated TSS distribution and variations from 25/11/2020 to 28/01/2021. (a) Measured and estimated TSS on 25 November 2020. (b) Measured and estimated TSS on 12 December 2020. (c) Measured and estimated TSS on 28 January 2021.

turbidity. The spatial interpolation results confirm that Landsat performed best in the estimation of TSS concentration. Thus, the sensor can satisfactorily be used for estimating and mapping TSS distribution in reservoirs.

3.4.3. Chl-*a* distribution

Figure 7 shows measured and estimated Chl-*a* distribution and variation for the sampling period. For data collected on date 25/11/2020, the lowest Chl-*a* concentration was recorded at points L1, L4, and L10 for both laboratory and *in situ* measurements. On the other hand, the highest Chl-*a* concentration was recorded at

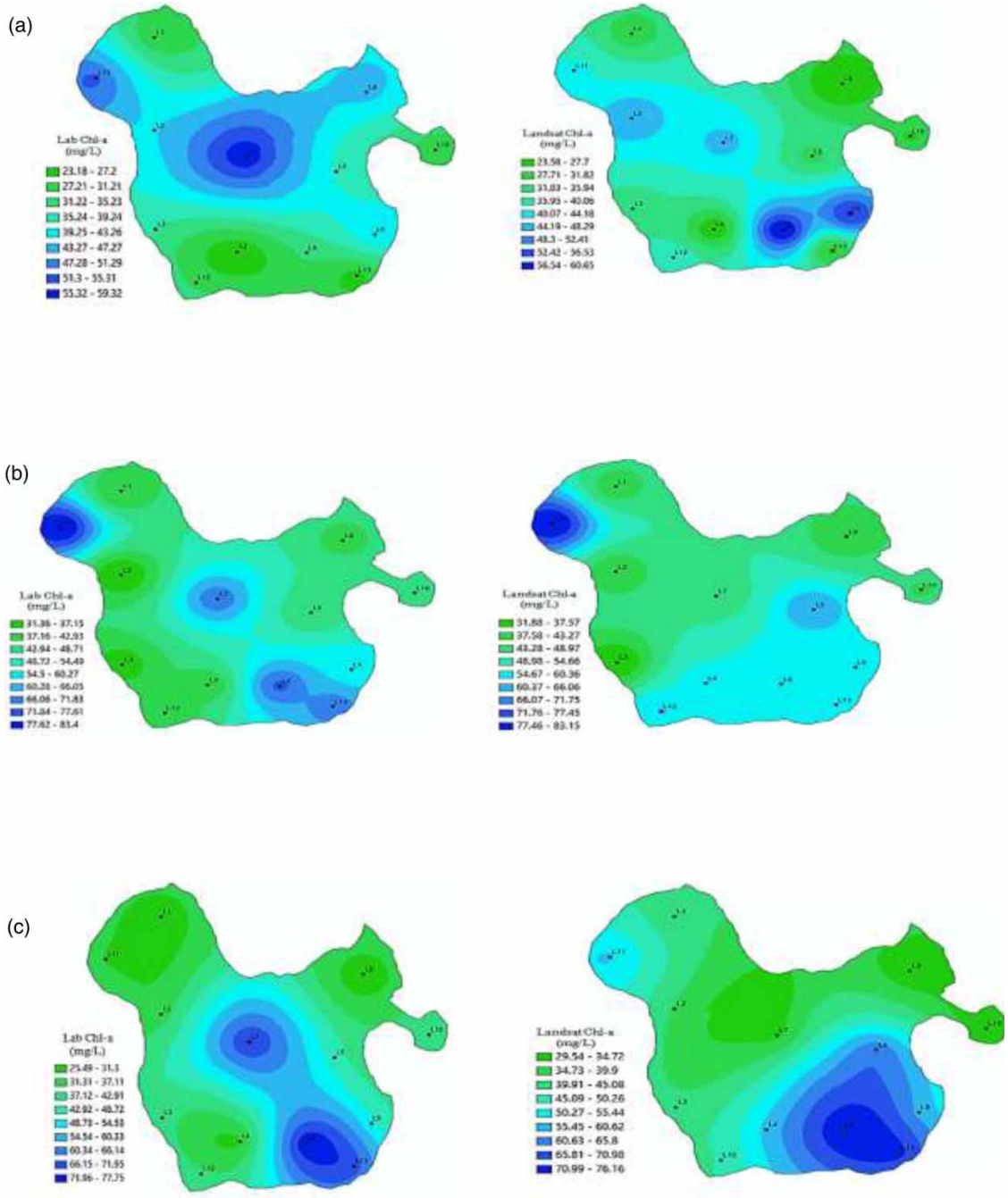


Figure 7 | Measured and estimated Chl-*a* distribution and variations from 25/11/2020 to 28/01/2021. (a) Measured and estimated Chl-*a* on 25 November 2020. (b) Measured and estimated Chl-*a* on 12 December 2020. (c) Measured and estimated Chl-*a* on 28 January 2021.

points (L11 and L13) and point L8 for data collected on dates 11/12/2020 and 28/01/2020, respectively. The areas with a high concentration of Chl-*a* are close to the regions where River Endoroto enters the reservoir. This could be explained by the river draining adjoining areas of the catchment where agriculture, tourism, and manufacturing are the main economic activities that introduce several point and non-point pollutants. Furthermore, sampling was carried out during the dry season and this means that there was decreased nutrient transport into the dam. However, Poddar *et al.* (2019) note that phytoplankton blooms are expected during the dry season due to the conducive water temperatures and the presence 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 algorithmic models developed in this study could be used for the prediction and mapping of turbidity, Chl-*a*, and TSS in Two Rivers Dam reservoir.

4. CONCLUSION

The study evaluated the performance of Landsat-8 OLI in predicting the turbidity, TSS, and Chl-*a* for an inland water reservoir based on *in situ* measurements at specified sampling points in the reservoir. The results revealed that the mean values of laboratory-measured turbidity, TSS, and Chl-*a* were 7.69 NTU, 277.9 mg/L, and 46.51 mg/L, respectively and this was highly comparable to Landsat-8 estimated values of 7.22 NTU, 277.71 mg/L, and 46.51 mg/L, respectively. For all the samples, turbidity was estimated using a polynomial regression model with both R^2 and Pearson correlation coefficient (r) greater than 75%. TSS was best estimated by exponential and polynomial regression models with respective mean R^2 and r of 0.809 and 0.721. Chl-*a* was best estimated using polynomial regression models with mean R^2 and r of 0.803 and 0.731, respectively.

From the study, it has been shown that satellite images including Landsat-8 OLI images avail a cheap and cost-effective tool for reservoir WQM and management. The remote sensing approach ensures continuous water quality assessment and/or management and increases spatial-temporal reservoir monitoring. However, in order to improve the effectiveness and reliability of Landsat-8 OLI in water quality parameter retrieval, it is recommended that the development of the model coefficients be based on data collected on a larger extent of the reservoir for the different seasons in a year. In line with this, model transfer functions should also be developed to enable the algorithms to be used for water quality predictions in other reservoirs within the same location. To apply these models to other localities, the model coefficients must be revised in line with the reservoirs' hydrological characteristics and the seasonal variations in the specific climatic and hydrological conditions.

ACKNOWLEDGEMENTS

The work reported here was undertaken as part of the Building Capacity in Water Engineering for Addressing Sustainable Development Goals in East Africa (CAWESDEA) project which is part of the IDRC funded program on Strengthening Engineering Ecosystems in sub-Saharan Africa. CAWESDEA Project is led by Global Water Partnership Tanzania in collaboration with Makerere University (Uganda), Moi University (Kenya) and University of Dar es Salaam (Tanzania). We acknowledge the support from the Eldoret Water and Sanitation Company (ELDOWAS) for hosting the research reported herein.

DATA AVAILABILITY STATEMENT

Data cannot be made publicly available; readers should contact the corresponding author for details.

CONFLICT OF INTEREST

The authors declare there is no conflict.

REFERENCES

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

Barasa, B. N. & Perera, E. D. P. 2018 Analysis of land use change impacts on flash flood occurrences in the Sosiani River basin Kenya. *International Journal of River Basin Management* **16**(2), 179–188.

Bar-Massada, A., Radeloff, V. C. & Stewart, S. I. 2014 Biotic and abiotic effects of human settlements in the wildland–urban interface. *Bioscience* **64**(5), 429–437.

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.

Chen, Y. & Han, D. 2018 Water quality monitoring in smart city: A pilot project. *Automation in Construction* **89**, 307–316.

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.

Jaelani, L. M., Limehuwey, R., Kurniadin, N. & Pamungkas, A. 2016 *Estimation of TSS and Chl-A Concentration From Landsat 8-OLI: The Effect of Atmosphere and Retrieval Algorithm*.

Japitana, M. V. & Burce, M. E. C. 2019 A satellite-based remote sensing technique for surface water quality estimation. *Engineering, Technology & Applied Science Research* **9**(2), 3965–3970.

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*.

Kenya National Bureau of Statistics 2019 *The 2019 Kenya Population and Housing Census: Population by County and Sub-County*. Kenya National Bureau of Statistics. Nairobi, Kenya.

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.

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.

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.

Moses, W. J., Bowles, J. H. & Corson, M. R. 2015 Expected improvements in the quantitative remote sensing of optically complex waters with the use of an optically fast hyperspectral spectrometer – a modeling study. *Sensors* **15**(3), 6152–6173.

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. & Niazmardi, S. 2021 A novel multiple-kernel support vector regression algorithm for estimation of water quality parameters. *Natural Resources Research* **30**(5), 3761–3775.

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.

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.

Ontumbi, G., Obando, J. & Ondieki, C. 2015 The influence of agricultural activities on the water quality of the river sosiani in uasin gishu county, Kenya. *International Journal of Research in Agricultural Sciences* **2**(1), 2348–3997.

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), 1–19.

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**, 1–32.

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**, 1–21.

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.

Reddy, M. A. 1997 A detailed statistical study on selection of optimum IRS LISS pixel configuration for development of water quality models. *International Journal of Remote Sensing* **18**(12), 2559–2570.

SCOR-UNESCO, W 1966 Determination of photosynthetic pigments. In: Determination of photosynthetic pigments in sea-water. UNESCO, Ed.; UNESCO: Paris, France, pp. 9–18.

Usali, N. & Ismail, M. H. 2010 Use of remote sensing and GIS in monitoring water quality. *Journal of Sustainable Development* **3**(3), 228.

Vörösmarty, C. J., McIntyre, P. B., Gessner, M. O., Dudgeon, D., Prusevich, A., Green, P., Glidden, S., Bunn, S. E., Sullivan, C. A., Liermann, C. R. & Davies, P. M. 2010 Global threats to human water security and river biodiversity. *Nature* **467**(7315), 555–561.

Wang, F., Han, L., Kung, H. T. & Van Arsdale, R. B. 2006 Applications of Landsat-5 TM imagery in assessing and mapping water quality in Reelfoot Lake, Tennessee. *International Journal of Remote Sensing* **27**(23), 5269–5283.

Watanabe, F. S. Y., Alcântara, E., Rodrigues, T. W. P., Imai, N. N., Barbosa, C. C. F. & Rotta, L. H. D. S. 2015 Estimation of chlorophyll-a concentration and the trophic state of the Barra Bonita hydroelectric reservoir using OLI/Landsat-8 images. *International Journal of Environmental Research and Public Health* **12**(9), 10391–10417.

Yanti, A., Susilo, B. & Wicaksono, P. 2016 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. IOP Publishing, p. 012028.

First received 11 December 2021; accepted in revised form 18 January 2023. Available online 31 January 2023