WATER QUALITY PARAMETERS PREDICTION OF TIGRIS RIVER USING SENTINEL-2 DATA AND LASSO REGRESSION

Suhaib Saad[a], Adel Elshazly[b], Ahmad M. Senousi[b], Walid Darwish[b,*], Moustafa Baraka[c], Wael Ahmed[b]

[a] ENV. Engineer, Ministry of Environment, Karada, Baghdad 10062, Iraq
[b] Public Works Department, Faculty of Engineering, Cairo University, Giza 12613, Egypt
[c] Civil Engineering Program, German University in Cairo, Cairo, Egypt

KEY WORDS: Sentinel-2, Remote Sensing, Water Quality Index, LASSO, Tigris River

ABSTRACT:

Water-related issues have become a growing concern due to the impact of human activities and climate change unpredictability, and Iraq faces a significant challenge with freshwater scarcity and poor quality. Field measurements are expensive, while remote sensing is a cost-effective alternative. This study aimed to monitor the water quality of the Tigris River in Baghdad using Sentinel-2 satellite images. The study employed the least absolute shrinkage and selection operator (LASSO) and field data from 14 different stations during 2018 and 2019 to measure water quality parameters, including temperature (Temp), electrical conductivity (Cond), total dissolved solids (TDS), potential of hydrogen (pH), turbidity (Turb), Chlorophyll-a (Chl_a), Blue-Green Algae (BGA), and Dissolved Oxygen (DO). The incorporation of spectral indices significantly improved the models' effectiveness, with R^2 greater than 0.8, except for Cond, which was 0.73. Water Quality Index (WQI) based on Iraqi standards was estimated and showed that the river water quality was categorized as poor and very poor. This approach can enhance water resource management and decision-making in areas where traditional monitoring approaches may be challenging and costly.

1. INTRODUCTION

Water-related issues have garnered increased attention due to negative human activities and the unpredictability of climate change (Mahmoodi et al., 2021). Assessing the chemical, physical, and biological properties of water can help determine its quality, which is critical for human health and farming activities (Chadli et al., 2021). In Iraq, the scarcity and poor quality of freshwater is a major concern, with studies predicting insufficient good quality water to meet developmental needs in the near future (Al-Ansari et al., 2020). The water quality of Iraqi rivers is influenced by various factors, including climate change-related ones, which are deemed uncontrollable. However, factors like dams and irrigation projects are controllable but pose significant challenges, particularly due to their dependency on neighboring nations like Turkey and Iran. Therefore, international decisions have a substantial impact on the water quality of Iraqi rivers.

Field measurements are widely recognized as the primary method for monitoring water quality as they provide accurate information on the water quality of the rivers being monitored (Koparan et al., 2018). Sensors measure physical and chemical parameters of a body of water, and since water quality parameters tend to change rapidly, measurements are taken at the time of sampling. The most common approach to measuring field water quality is through multi-parameter water quality devices, which consist of a collection of probes that measure individual parameter to form the sonde of a multi-parameter water quality sensor. Dissolved oxygen (DO), temperature, pH, electrical conductivity (EC), turbidity, and depth are common probe configurations, and other factors like chlorophyll, oxidation-reduction potential (ORP), ammonia, ammonium, nitrate, and chloride can also be measured using probes. However, calibration of measured data is typically required in a laboratory, as the raw data may not be applicable but can assist in interpreting other water quality results. One drawback of field water quality measurements is their high cost. Sensor calibration, cleaning, and technical difficulties can all contribute to the cost of water quality sampling through these approaches. Therefore, new techniques such as remote sensing have been developed and are already used in various water resource management applications.

* Corresponding author

Remote sensing with free access (e.g., Landsat, Sentinel) can provide images with high spatial and spectral resolution covering large geographic areas. Although remote sensing does not directly measure water quality, various algorithms have been proposed to establish correlations between spectral information from image data and water quality parameters, resulting in mathematical equations that can predict water quality based on data obtained from remote sensors (Ewaid et al., 2018).

Retrieving water quality parameters from satellite image data is challenging due to the optical complex conditions that exist in lakes (Koparan et al., 2018). Additionally, shallow lakes may be subject to interference from the lake bottom. Furthermore, dynamic changes in water quality further complicate the retrieval of water quality data. Thus, remote sensing of water quality is limited to the retrieval of parameters such as water clarity, turbidity, water color, and the concentrations of optically active constituents such as algal pigments, suspended solids, and colored dissolved organic matter. The calibration and validation of models for the retrieval of water quality data from satellite images usually require collecting ground truth data and laboratory analysis of water samples (Cao et al., 2022).

Water quality assessment typically relies on a range of factors, encompassing both chemical and physical parameters. Mohammed and Bamarni (2019) conducted a study to assess the

https://doi.org/10.5194/isprs-annals-X-1-W1-2023-863-2023 |

water quality of the Duhok Dam for drinking and agricultural purposes. The study collected water samples from two locations on a monthly basis over a period of eight years and analyzed 14 water quality parameters. The results showed high concentrations of electrical conductivity, total dissolved solids, sulfate, and total hardness in all water samples, with some concentrations exceeding the guidelines prescribed by the WHO in 2008. The study also identified temporal variations in most parameters and concluded that the quality of water was influenced by both geological formations and anthropogenic activities around the dam.

Another research study aims to assess the water quality of the Tigris River using the water quality index method and GIS software was done by (Chabuk et al., 2020). The study collected data on 12 parameters from 14 stations along the river and used the weighted arithmetic method to compute the water quality index. The study also applied the interpolation method to produce prediction maps for the 12 parameters at 11 stations during the wet and dry seasons in 2016. Regression prediction was then applied to compare observed values with predicted values, and the results showed acceptable values for the determination coefficient (R^2). The study found that the state of water quality for the Tigris River was degraded downstream, particularly at station 8 in Aziziyah during both the wet and dry seasons, and at Qurnah in the south of Iraq. The study highlights the importance of considering the entire length of the river in order to gain comprehensive knowledge about contamination, analyze the problem, and find appropriate solutions.

Adopting Landsat imagery, a research study investigates the seasonal changes that occurred on the Tigris River in 2018 (Allawai and Ahmed, 2019). The study focuses on analyzing the differences in reflectance values at different locations along the river, as well as seasonal differences and changes in algal amounts. The results show distinct reflectance differences among the downstream, midstream, and upstream areas, and reflectance values are significantly associated with the seasonal factor. The modeling of chlorophyll-a and Secchi disk depth indicates a decrease in water clarity over time, but chlorophyll-a amounts have decreased. The study suggests that the decreasing water clarity is due to factors other than chlorophyll-a.

Another study carried by Ewaid et al. (2018) was aimed to analyze monthly water quality data sets of ten stations on the Tigris River within Baghdad for the year 2016. The water quality index (WQI) was calculated using 11 parameters and used as the dependent parameter in stepwise multiple linear regression (MLR) analysis to develop a water quality model (WQM) for the river. The study found significant differences in WQI values among months and stations and developed a WQM consisting of five parameters. The study shows that the use of WQI as the dependent parameter input improves the prediction of MLR model as a tool to understand and simplify water quality variation.

Water quality modeling methods are generally categorized into three groups: statistical techniques (e.g., Salarijazi and Ghorbani, 2019), machine learning (e.g., Ighalo et al., 2021), and integrated methods (e.g., Lu and Ma, 2020). Sakizadeh (2015) employed Linear discriminant and naive Bayesian classification methods to estimate water quality in Karaj River, Iran, with results demonstrating that the linear discriminant outperformed the naive Bayesian classification. In another study, Salarijazi and Ghorbani (2019) utilized linear and non-linear regression models as water quality prediction methods, with the models proving to be efficient in predicting electrical conductivity in a case study of Zarringol and Ramian River, Iran.

Furthermore, Avila et al. (2018) compared the performance of several statistical models, including regression tree, Markov chain, and Bayesian network, among others, for freshwater quality prediction. Their results indicated that the Bayesian network was the most effective in handling missing data and outliers, allowing for continuous updating in real-time. In addition, Babbar and Chaubey (2022) developed six multiple regression models for water quality prediction in the St. Joseph River basin, USA, with their experiments demonstrating that the ridge regressor was the best predictor for nonpoint water quality parameters. In another study, Zhang et al. (2021) employed a temporal LASSO regression model to forecast suspended sediment concentrations in coastal oceans. Their findings showed that the model was concise and practical, effectively shrinking the interrelated parameters into representative ones while achieving one-hour ahead forecasting with higher accuracy compared to other data-driven methods. Finally, Alnahit et al. (2022) used three parameter selection methods such as stepwise regression, Least Absolute Shrinkage and Selection Operator, and genetic algorithm for water quality prediction based on Random Forest and Boosted regression tree models. Both models obtained good results with the selected parameters by the models.

The Tigris River in Baghdad city is monitored for water quality using sensors that measure various water quality parameters. However, the calibration of these sensors is not effective due to the high cost of maintenance. Therefore, there is a need to develop cost-effective methods based on remote sensing to more accurately measure water quality and provide timely maps for efficient water resource planning. The aim of this study is to provide low-cost monitoring of river water quality without relying on field measurements. The objectives of this study include: **(1)** Acquisition and preparation of necessary image data and water quality samples along the Tigris River in Baghdad, **(2)** Statistical analysis of spectral relationships between water quality parameters and satellite image data, **(3)** Development of regression models to predict water quality parameters and validation of predictions with field measurements, and **(4)** Development of a water quality index (WQI) for the Tigris River in Baghdad based on arithmetic methods and Iraqi water quality standards. This study incorporates dissolved oxygen data, which is available within the field data and has a certain concentration within the determinants.

The rest of the paper is structured as follows: Section 2 describes the study area and datasets, Section 3 explains the methodology of water quality monitoring, Section 4 presents the results and discussion, and Section 5 presents the conclusions of this research.

2. STUDY AREA AND DATASETS

The Tigris River (as shown in Figure 1), located in Baghdad, Iraq, is one of the largest rivers in the Middle East, spanning 110 km through the city and dividing it into two parts. The river is a major source of water supply for the city, with many water purification plants located along its course. This paper focuses on assessing the water quality of the Tigris River by analyzing the values of eight water quality parameters over two years (2018 and 2019) at 14 sampling stations located along the river. The 14 sampling stations selected for this study were Karkh, Taje, 7-Abkar, Salameat, Karama, Wathba, Senak, Melea, Jaderea,

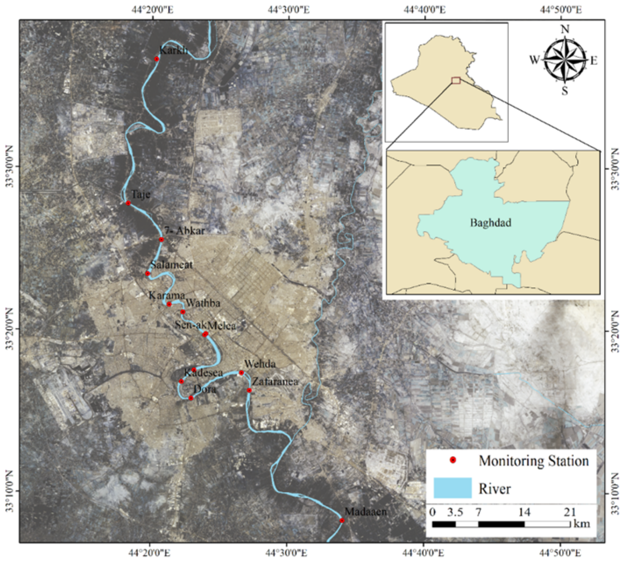


Figure 1. Map of Tigris River crossing Baghdad with monitoring station in red circles.

Kadesea, Dora, Wehda, Zafaranea, and Madaaen, as shown in Table 1. The water quality parameters measured were temperature (Temp), electrical conductivity (Cond), total dissolved solids (TDS), pH, turbidity (Turb), Chlorophyll-a (Chl_a), Blue-Green Algae (BGA), and Dissolved Oxygen (DO).

Sentinel-2B satellite images for 2018 and 2019 were used to ensure the data was accurate and geometrically correct. The images were processed to be in BOA and projected using Universal Transverse Mercator (UTM) with a World Geodetic System 84 (WGS84) datum. The images were atmospherically calibrated using the QUick Atmospheric Correction (QUAC) module in Exelis Visual Information Solutions software. Finally, the image bands were co-registered with other data to remove any geometric errors

3. METHODOLOGY

The methodology adopted in this research for monitoring water parameters is illustrated in Figure 2. The primary datasets used to evaluate water quality at the study sites are field-based water quality sampling and remote sensing satellite images (Sentinel-2B). Each phase of the proposed methodology will be discussed in detail.

3.1 Field Data Measurements

The spatial variations of water quality were studied at fourteen different sites across the Tigris River in Baghdad. The temporal variations were analyzed from January to December for 2018 and 2019. The seasonal variations were assessed using the two major seasons of Iraqi climate i.e., summer (May-October) and winter (November-April). 70% of the filed data were conducted for calibration of the predicted model, while the remain were implemented for evaluation.

To examine possible correlations in the water quality data and remote sensing spectral data, a linear correlation was used. The correlation with calibrated spectral bands as well as other extracted data such as spectral indices was investigated. The evaluation was based on visual interpretation as well as standard quantitative approaches such as the coefficient of determination (R^2).

3.2 Spectral Indices

Band ratios, which include spectral indices, improve the spectral contrast between distinct targets and thus help to improve water quality prediction models. Nine spectral indices calculated from Sentinel-2B datasets were used in this study: Difference Vegetation Index (DVI), Green Normalized Difference Vegetation Index (GDVI), Modified Normalized Difference Water Index (MNDWI), Modified Soil Adjusted Vegetation Index (MSAVI), Normalized Difference Vegetation Index (NDVI), Normalized Difference Water Index (NDWI), Normalized Ratio Vegetation Index (NRVI), Specific Leaf Area Vegetation Index (SLAVI), and Simple Vegetation Ratio (SVR).

3.3 LASSO Regression

Introducing a penalty item in linear regression can be an effective way to reduce a model's variance, especially when dealing with high-dimensional predictors (Zhang et al., 2021). The LASSO regression (see Equation 1) is a type of linear regression that incorporates a penalty item ($L1$) to force some coefficient estimations to exactly equal 0 with a sufficiently large tuning parameter. As a result, LASSO can automatically select the most important independent parameters by shrinking the less important predictors to 0. In the context of a dataset (x_i, y_i), where x_i is a p-dimensional vector, LASSO can be an effective tool for feature selection. Automatic feature selection is performed. However, it is critical to investigate the effect of feature selection on water quality prediction accuracy. The number of features (k) included in regression models to predict Tigris River water quality was investigated in this study. A range of 2-22 k was analyzed for the Sentinel-2B dataset.

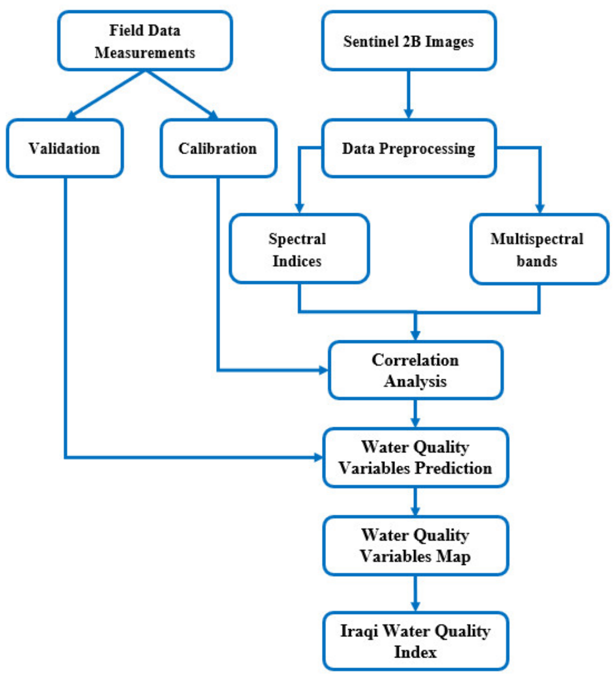


Figure 2. Schematic diagram for the adopted methodology for monitoring water quality of TIGRIS River.

$$\Big(\sum_{i=1}^{n}\big(y_i - \sum_{j=1}^{p}(x_{ij})\big)\Big)^2 + \alpha \sum_{j=1}^{p}|\beta_j| \tag{1}$$

Where n indicates The total number of training examples in the dataset, x_{ij} represents the hypothetical function for prediction, y_i represents the weight for i^{th} feature, α indicates the regularization strength,and β_j represents the weight of j^{th} feature

3.4 Water Quality Mapping

The Inverse Distance Weighted (IDW) interpolation (Liu et al., 2021) was used to construct the water quality map for each parameter in the Tigris River during the summer and winter seasons.

$$Z_0 = \frac{\sum_{i=1}^{n} \frac{Z_i}{X_i^r}}{\sum_{i=1}^{n} \frac{1}{X_i^r}} \tag{2}$$

where Z_0 indicates the estimated value of point zero, Z_i indicates the value of known point i, X_i represents distance between i and zero point, r indicates the specified power >1, and n indicates the number of known points used in estimation.

3.5 Iraqi Water Quality Index

The method of weighted arithmetic was used to calculate the WQI. This study used an overall of eight parameters, of which four (Turb, pH, TDS, DO, and Cond.) were used for the calculation of WQI. These parameters have been used in many previous studies of water quality assessment for Iraqi rivers (Chabuk et al., 2020). The WQI for each selected station in the Tigris River was calculated using the following equations (Chabuk et al., 2020):

$$Q_i = \big(\frac{N_i - N_0}{ST_i - N_0}\big) \times 100 \tag{3}$$

$$W_i = \frac{1}{ST_i} \tag{4}$$

$$WQI = \frac{\sum Q_i \times W_i}{\sum W_i} \tag{5}$$

Where Q_i is the sub-index of the i_th parameter, W_i is the inverse weight of the standard value ST_i of the i_th parameter, ST_i is the standard value of the i_th parameter, N_i is the measured concentration value for the i_th parameter, and N_0 is the ideal value for each parameter in water that has zero value, excluding the pH which is equal to 7.

WQI is a quick and easy way to evaluate water quality using a set of parameters. However, it offers a water quality value that can be difficult to interpret in real-world situations. As a result, using the WQI value to classify water quality for distinct category classes is critical. The water quality rating (WQR), which is a clear linguistic rating for a WQI value (Alsaqqar et al., 2015), is shown in Table 1.

Value of WQI	Water Quality Rating (WQR)
< 50	Excellent
50–100	Good
100–200	Poor
200–300	Very poor
300–400	Polluted
> 400	Vary polluted

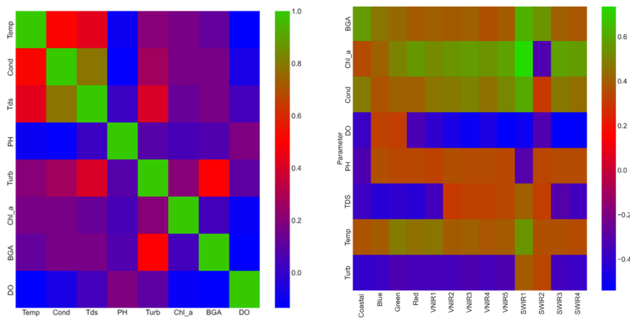
Table 1. Water quality rating classification based on WQI value.

4. RESULTS ANALYSIS AND DISCUSSION

Fourteen water samples were collected along the Tigris River in Baghdad and analyzed for different water quality parameters. The descriptive statistics were measured for (Temp., Cond., TDS, pH, Turb., Chl_a, BGA, and DO) over 2018 and 2019. For the 2018 dataset, the concentrations ranged from 10 c° to 16.5 c° with an average of 11.8 c°, 125 µs/cm to 1742 µs/cm with an average of 1058.26 µs/cm, 445 g/l to 942 g/l with an average of 673.24 g/l, 6.7 to 8.9 with an average of 8, 1.45 NTU to 23.5 NTU with an average of 6.83 NTU, 0.1 mg/l to 4.3 mg/l with an average of 1.7 mg/l, 21 mg/ml to 1378 mg/ml with an average of 280.24 mg/ml, 9.2 mg/l to 12.2 mg/l with an average of 10.66 mg/l for Temp., Cond., TDS, pH, Turb., Chl_a, BGA, and DO, respectively.

In 2019 dataset, the concentrations ranged from 10.2 c° to 13.5 c° with an average of 11.59 c°, 125 µs/cm 1742 µs/cm with an average of 1063.27 µs/cm, 0.445 g/l to 0.945 g/l with an average of 0.67 g/l, 6.7 to 8.9 with an average of 8.01, 1.45 NTU to 138 NTU with an average of 7.79 NTU, 0.1 mg/l to 4.5 mg/l with an average of 1.75 mg/l, 21 mg/ml to 1368 mg/ml with an average of 285.95 mg/ml, 9.2 mg/l to 12.2 mg/l with an average of 10.71 mg/l for Temp., Cond., TDS, pH, Turb., Chl_a, BGA, and DO, respectively. Temp. and TDS had lower average values in 2019 compared to the previous year (2018). The other parameters all had higher values. While lower TDS means higher quality of water, the water of higher Cond. indicates more chemicals dissolved in the water. In addition, higher turbidity levels are caused by solid particles suspended in the water. High Chl_a can cause algae to grow or bloom. Higher DO makes the water taste better and higher BGA indicates lower quality of water and may cause diarrhea, nausea or vomiting.

The correlation between concentrations of water quality parameters to each other was calculated, as shown in Figure 3-a. The correlation matrix for water quality parameters. Strong correlation is indicated by green, weak correlation by blue, and moderate correlation by in-between colors. The majority of the parameters are just weakly to moderately associated with one another. The dataset revealed no strong correlations. TDS and Cond. have a stronger moderate correlation than any other pair of parameters. Cond. and TDS are moderately associated with water temperature. Turb. was also found to have a moderate relationship with TDS and BGA. The correlation between optical water quality measurements (such as Turb.) and non-optical parameters can let satellite sensors measure the latter indirectly. Light signals from the surface of aquatic bodies are measured by satellite sensors. As a result, satellite sensors are assured to measure the optical parameters that modify the color of the water. However, non-optical parameters have an issue since their alterations in water do not affect the color of the water and hence cannot be directly monitored by satellite sensors. The relationship between optical and non-optical parameter concentrations is therefore crucial. Strong to moderate correlation may greatly aid in the development of models for non-optical parameter water quality prediction based on remote sensing data.



(a) (b)

Figure 3. The water quality parameters correlation: (a) between different parameters, and (b) between parameters and spectral bands.

Water quality parameters concentrations were substantially connected with Sentinel-2 bands such as Coastal, Blue, and SWIR1 as shown in Figure 3-b. The spectral bands of Sentinel-2 were intended to aid in building models for predicting water quality. Due to its capacity to enhance spectral contrast between distinct targets, spectral indices may also be useful in predicting concentrations of water quality parameters.

LASSO regression models were developed for water quality prediction using Sentinel-2B spectral bands and spectral indices derived from these satellite data. Models for water quality parameters like Temp, Cond, TDS, pH, Turb., Chl a, BGA, and DO were created. Coastal, Blue, Green, Red, VNIR1, VNIR2, VNIR3, VNIR4, VNIR5, SWIR1, SWIR2, SWIR3, and SWIR4 are among the thirteen bands on Sentinel-2B. Similar to the Landsat 8 data, nine spectral indices were added to the Sentinel-2B spectral data for water quality modeling: DVI, GNDVI, MNDWI, MSAVI, NDVI, NDWI, NRVI, SLAVI, and SVR. The models were created to correlate with measurements of water quality parameters.

The performance of the developed models is presented in Table 2. The results include the optimal k values used for each water quality parameter. The models trained on 70% of the samples and validated on the remaining 30%. The results show that the models were significantly estimated the water quality measurements using the Sentinel-2B data and its derived spectral indices.The most accurate models were Trub (RMSE = 1.04 NTU, $R^2 = 0.96$), DO (RMSE=0.18 mg/l, $R^2 = 0.91$), and Temp (0.16 c^o, $R^2 = 0.90$) respectively. The models of Cond and pH showed the lowest performance with RMSE and R^2 of 93.03 µs/cm and 0.07, 0.73, and 0.81, respectively. The other models showed moderately accurate performance with R^2 ranging from 0.86 for BGA and Chl_a to 0.89 for TDS. Water parameters prediction thematic maps estimated for the whole area of Tigris River are shown in Figure 4.

This study used an arithmetic method to calculate a water quality index (WQI) based on Iraqi drinking water quality standards and five parameters: Turb, pH, TDS, Cond, and DO. The WQI readings for the 2018 summer range from 157.60 in the Poor category (July) to 135.86 in the Poor category (August). In summer 2019, WQI ranged from 209.19 (Very Poor) to 139.10 (Poor). In winter of 2018 and 2019, WQI was estimated to be Very Poor (223.08 and 227.92, respectively) in April and Poor in other months.

parameters	k	RMSE	R^2
BGA	22	94.42 mg/ml	0.86
Chl_a	18	0.24 mg/l	0.86
Cond	20	93.03 µs/cm	0.73
DO	22	0.18 mg/l	0.91
pH	21	0.07	0.81
TDS	19	39.36 g/l	0.89
Temp	21	0.16 c^o	0.90
Turb	19	1.04 NTU	0.96

Table 2. Statistics and accuracy measures of the proposed models for water quality prediction of the Tigris River in Baghdad.

5. CONCLUSIONS

The study provided low-cost monitoring of river water quality without relying on field measurements, by establishing correlations between spectral information from image data and water quality parameters. LASSO regression models were developed for the prediction of water quality parameters using Sentinel-2B spectral bands and spectral indices derived from these satellite data. LASSO was conducted on field data from 14 different stations during 2018 and 2019 to measure water quality parameters, including temperature (Temp), electrical conductivity (Cond), total dissolved solids (TDS), pH, turbidity (Turb), Chlorophyll-a (Chl_a), Blue-Green Algae (BGA), and Dissolved Oxygen (DO). The results of the study showed that the LASSO regression models developed significantly estimated the water quality measurements.

The most accurate models were Trub, DO, and Temp, while the models of Cond and pH showed the lowest performance. The other models showed moderately accurate performance, with R2 ranging from 0.86 for BGA and Chl_a to 0.89 for TDS. Four out of the estimated eight parameters (Turb, pH, TDS, DO, and Cond.) were used for the calculation of the Water Quality Index for each selected station, and the water quality rating (WQR) was categorized as poor and very poor. In future work, it would be valuable to investigate how water quality is influenced when considering the modifiable areal unit problem (MAUP) and modifiable Temporal unit problem (MTUP)..

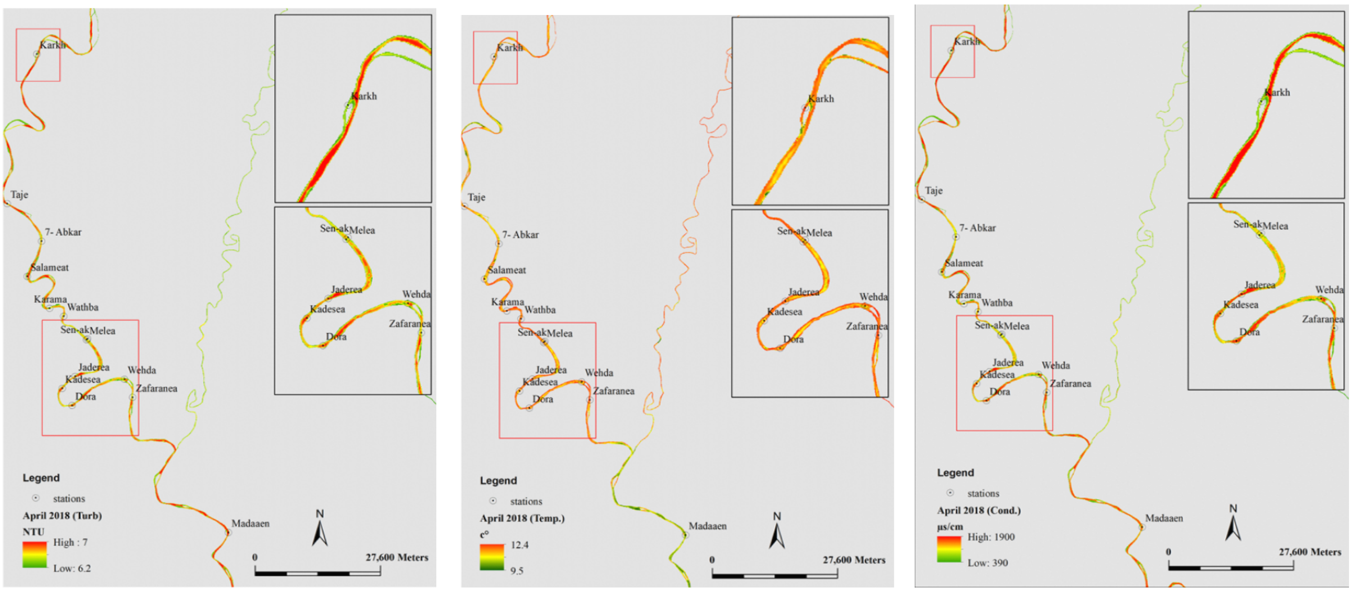
References

Al-Ansari, N. N., Abbas, N., Laue, J., Knutsson, S., 2020. Water Scarcity: Problems and Possible solutions. *Journal of Earth Sciences and Geotechnical Engineering*, 243-312. http://www.scienpress.com/Upload/GEO/Vol%2011$_{27}$*.pdf*.

Allawai, M. F., Ahmed, B. A., 2019. Using GIS and Remote Sensing Techniques to Study Water Quality Changes and Spectral Analysis of Tigris River within Mosul City, North of Iraq. *Iraqi Journal of Science*, 60, 2300-2307. https://ijs.uobaghdad.edu.iq/index.php/eijs/article/view/1222.

Alnahit, A. O., Mishra, A. K., Khan, A. A., 2022. Stream water quality prediction using boosted regression tree and random forest models. *Stochastic Environmental Research and Risk Assessment*, 36, 2661-2680.

Alsaqqar, A. S., Hashim, A. M., Mahdi, A. A., 2015. Water Quality Index Assessment using GIS Case study: Tigris River in Baghdad City. *2515— International Journal of Current Engineering and Technology*, 5, 2515-2520. http://inpressco.com/water-quality-index-assessment-using-gis-case-study-tigris-river-in-baghdad-city/.



(a) (b) (c)

Figure 4. Selected water Quality parameters maps for the study area.

Avila, R., Horn, B., Moriarty, E., Hodson, R., Moltchanova, E., 2018. Evaluating statistical model performance in water quality prediction. *Journal of Environmental Management*, 206, 910-919.

Babbar, R., Chaubey, I., 2022. Multiple regression analysis for predicting few water quality parameters at unmonitored sub-watershed outlets in the St. Joseph River basin, USA. *Geocarto International*, 37, 8697-8723.

Cao, Q., Yu, G., Sun, S., Dou, Y., Li, H., Qiao, Z., 2022. Monitoring water quality of the haihe river based on ground-based hyperspectral remote sensing. *Water (Switzerland)*, 14.

Chabuk, A., Al-Madhlom, Q., Al-Maliki, A., Al-Ansari, N., Hussain, H. M., Laue, J., 2020. Water quality assessment along Tigris River (Iraq) using water quality index (WQI) and GIS software. *Arabian Journal of Geosciences*, 13.

Chadli, K., Boufala, . M., Responsible, O., Merkel, B. J., Boufala, M. . H., 2021. Assessment of water quality using Moroccan WQI and multivariate statistics in the Sebou watershed (Morocco) Nitrate-nitrogen NH 4 + Ion ammonium PO 4 3. *Arabian Journal of Geosciences*, 14, 27. https://doi.org/10.1007/s12517-020-06296-5.

Ewaid, S. H., Abed, S. A., Kadhum, S. A., 2018. Predicting the Tigris River water quality within Baghdad, Iraq by using water quality index and regression analysis. *Environmental Technology Innovation*, 11, 390-398. https://linkinghub.elsevier.com/retrieve/pii/S2352186418301548.

Ighalo, J. O., Adeniyi, A. G., Marques, G., 2021. Artificial intelligence for surface water quality monitoring and assessment: a systematic literature analysis. *Modeling Earth Systems and Environment*, 7, 669-681.

Koparan, C., Koc, A. B., Privette, C. V., Sawyer, C. B., 2018. In situ water quality measurements using an unmanned aerial vehicle (UAV) system. *Water (Switzerland)*, 10.

Liu, Z., Zhang, Z., Zhou, C., Ming, W., Du, Z., 2021. An Adaptive Inverse-Distance Weighting Interpolation Method Considering Spatial Differentiation in 3D Geological Modeling. *Geosciences*, 11, 51. https://www.mdpi.com/2076-3263/11/2/51.

Lu, H., Ma, X., 2020. Hybrid decision tree-based machine learning models for short-term water quality prediction. *Chemosphere*, 249.

Mahmoodi, N., Osati, K., Salajegheh, A., Saravi, M. M., 2021. Trend in river water quality: Tracking the overall impacts of climate change and human activities on water quality in the Dez River Basin. *Journal of Water and Health*, 19, 159-173.

Mohammed, N. I., Bamarni, K. A., 2019. Water quality monitoring of Duhok Dam (Kurdistan Region of Iraq). *ZANCO JOURNAL OF PURE AND APPLIED SCIENCES*, 31. https://zancojournals.su.edu.krd/index.php/JPAS/article/view/2410.

Sakizadeh, M., 2015. Assessment the performance of classification methods in water quality studies, A case study in Karaj River. *Environmental Monitoring and Assessment*, 187, 573. http://link.springer.com/10.1007/s10661-015-4761-6.

Salarijazi, M., Ghorbani, K., 2019. Improvement of the simple regression model for river' EC estimation. *Arabian Journal of Geosciences*, 12, 235. http://link.springer.com/10.1007/s12517-019-4392-2.

Zhang, S., Wu, J., Jia, Y., Wang, Y. G., Zhang, Y., Duan, Q., 2021. A temporal LASSO regression model for the emergency forecasting of the suspended sediment concentrations in coastal oceans: Accuracy and interpretability. *Engineering Applications of Artificial Intelligence*, 100.