DOI: 10.2478/eces-2025-0016 ECOL CHEM ENG S. 2025;32(3):335-352

Ahmed BDOUR[1*] and Raha M. ALKHARABSHEH[2]

REAL-TIME MACHINE LEARNING AND WIRELESS SENSOR NETWORK FOR COASTAL WATER QUALITY MONITORING IN THE GULF OF AQABA

Abstract: This study presents an advanced real-time monitoring system integrating wireless sensor networks with machine learning to assess water quality in the Gulf of Aqaba. Our hybrid machine learning framework combines Random Forest (500 trees, node size = 5) for feature selection with Artificial Neural Network ensembles (3-layer MLP with Monte Carlo dropout) for probabilistic forecasting. The system continuously monitors six critical parameters, demonstrating strong predictive performance through rigorous validation: dissolved oxygen ($R^2 = 0.92$, $RMSE = 0.45$ mg/L, 95 % *CI* (Confidence interval): 0.41 - 0.49), nitrite ($R^2 = 0.85$, $RMSE = 0.08$ mg/L, *CI*: 0.07 - 0.09), and turbidity ($R^2 = 0.89$, $RMSE = 2.3$ FTU, *CI*: 2.1 - 2.5). Comprehensive uncertainty analysis revealed prediction intervals of ±0.38 mg/L for DO and ±0.10 mg/L for nitrite, with spatial variability lowest in open waters ($CV = 8.2$ %) and highest near coastal zones ($CV = 15$ %). Residual autocorrelation analysis confirmed model reliability (Moran's $I < 0.12$, $p > 0.05$) across the study area. Spatial-temporal analysis identified nitrite as a sensitive pollution indicator, with concentrations reaching 0.12 mg/L near urban outflows compared to background levels (< 0.05 mg/L). The system achieved 92 % accuracy in the early detection of environmental risks, including coral bleaching precursors (temperature anomalies > 1 °C) and pollution events (nitrite spikes > 0.1 mg/L). Compared to conventional monitoring, the platform demonstrated 20.4 % greater predictive accuracy ($\Delta R^2 = +0.17$, $p < 0.01$) while reducing operational costs by 30.2 %, primarily through automated data collection and reduced manual sampling. The integration of high-frequency sensing, adaptive machine learning, and cloud-based analytics establishes a replicable framework for coastal ecosystem management, particularly in anthropogenically stressed environments like the Red Sea.

Keywords: marine water quality monitoring, real-time remote monitoring, wireless sensor networks, machine learning predictive modelling, *GIS*-based spatial analysis, Red Sea ecosystem management

Introduction

Water quality monitoring programs are essential for collecting data that support coastal conservation and informed decision-making. These programs must balance accuracy and cost-effectiveness, depending on specific informational needs [1]. While highly accurate water quality data are often desired across all locations, direct observations are inherently limited to specific areas [2]. The urgency for effective water quality monitoring has increased significantly, particularly in ecologically sensitive regions such as the Red Sea,

[1] Department of Civil Engineering, Faculty of Engineering, The Hashemite University, PO Box 330127, Zarqa 13133, Jordan, email: bdour@hu.edu.jo, ORCID: 0000-0001-6244-9530
[2] R&I Centre for the Conservation of Biodiversity and Sustainable Development (CBDS), Faculty of Forestry and Natural Environment Engineering, Technical University of Madrid, Spain, email: Raha.alkharabsheh@alumnos.upm.es; ORCID: 0000-0001-7151-4939
[*] Corresponding author: bdour@hu.edu.jo

where maintaining chemical, biological, and physical parameters within compliance levels is a primary concern for regulatory agencies [3, 4].

Conventional water quality monitoring methods, despite their widespread use, exhibit several limitations that reduce their effectiveness. In situ detectors provide continuous data collection but are constrained by limited spatial coverage and high operational costs. Additionally, they require frequent maintenance and calibration, which can be labour-intensive and expensive [5]. Geographic Information Systems (*GIS*) are widely used for data archiving, mapping, and visualisation, yet they rely on static datasets and lack real-time monitoring capabilities, making them unsuitable for rapid response scenarios [6]. Real-time remote monitoring (RTRM) systems enable continuous data collection, but their adoption has been hindered by high initial costs, mechanical durability issues, and engineering complexity, restricting their application mainly to large-scale research projects in lakes, reservoirs, and open ocean settings [1, 7].

To address these challenges, this study proposes an advanced monitoring framework that integrates RTRM with machine learning-based predictive modelling. This approach enhances conventional techniques by improving system reliability, simplifying deployment, and increasing cost-effectiveness. The integration of machine learning algorithms strengthens predictive capabilities, allowing for early detection of hazardous events such as coral bleaching and pollutant spills. Additionally, the framework combines in situ measurements, remote sensing data, and *GIS*-based spatial analysis to generate a comprehensive understanding of water quality dynamics across spatial and temporal scales [8]. By minimising the need for frequent manual sampling and maintenance, the system optimises resource allocation while ensuring high data accuracy and coverage [9].

The proposed method is grounded in well-established theoretical and mathematical principles. The predictive models employ supervised learning techniques, including random forests and neural networks, trained on historical water quality data to identify patterns and correlations. Model reliability is ensured through rigorous validation using cross-validation techniques and performance metrics such as root mean square error (*RMSE*) and the coefficient of determination (R^2). Furthermore, the integration of real-time data with key physical and chemical processes, such as wave-driven circulation and flushing rates, ensures that predictions align with fundamental environmental dynamics.

To evaluate the proposed framework, a comparative analysis was conducted against conventional monitoring methods. The combination of RTRM and predictive modelling enhances the accuracy and efficiency of water quality assessments, allowing for the timely identification of environmental disturbances. By streamlining data acquisition and analysis, the system facilitates proactive decision-making and strengthens early warning mechanisms, supporting long-term water resource management and ecological conservation efforts.

Materials and methods

The developed RTRM platform integrates a suite of sensors designed to measure key water quality parameters, including pH, conductivity, turbidity, temperature, and sediment concentration near the riverbed. These parameters were selected due to their significance in assessing aquatic ecosystem health and their sensitivity to environmental changes. The platform leverages advancements in wireless sensor technology and data analytics to provide real-time, high-resolution data, facilitating effective water quality management.

Each sensor within the RTRM platform serves a specific function in monitoring water quality dynamics. The pH sensor measures water acidity or alkalinity, which influences nutrient solubility, heavy metal bioavailability, and aquatic species' survival. The conductivity sensor estimates salinity and detects potential pollution sources by assessing dissolved ion concentrations. Turbidity measurements quantify suspended particulate matter, affecting light penetration and photosynthetic processes, while temperature monitoring provides insights into metabolic rates and dissolved oxygen levels, both of which are crucial for maintaining ecological balance. The sediment concentration sensor measures sediment levels near the riverbed, offering valuable data on erosion, runoff, and pollutant transport.

At the core of each measurement station is a CR1000 datalogger, which ensures precise tracking of water quality parameters over time, facilitating the identification of trends and environmental changes. The datalogger supports high-frequency sampling for continuous real-time monitoring, provides extensive data storage for long-term analysis, and enables remote connectivity for real-time data transmission and visualisation.

Machine learning processing and analytical framework

For predictive modelling, machine learning algorithms, namely, random forests and neural networks will be utilised to analyse historical and real-time data. These models identify trends, predict future water quality changes, and detect anomalies indicative of emerging environmental threats. The analytical framework ensures high predictive accuracy, enhancing the platform's capability for early intervention in cases of pollution spikes or ecological disturbances. The collected data undergoes advanced processing using machine learning algorithms to identify patterns, predict trends, and detect anomalies in water quality parameters. The analytical framework consists of multiple stages, including data preprocessing, predictive modelling, and statistical validation, ensuring accuracy, robustness, and reliability in the monitoring process [10, 11]. Data Preprocessing is a crucial step in ensuring data accuracy and consistency before applying machine learning models.

The following steps are implemented: Raw sensor data may contain missing values, outliers, or noise due to sensor malfunctions or environmental interferences. Missing values are handled using interpolation techniques, while outliers are detected and removed using statistical methods such as the Z-score method [12]:

$$Z = \frac{X - \mu}{\sigma}$$

where X is the observed value, μ is the mean, and σ is the standard deviation. Values with $Z > 3$ are considered outliers and removed [13].

Also, since different water quality parameters have varying units and ranges, feature scaling is applied to ensure uniformity. Min-max scaling is used [13]:

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}}$$

where X' is the normalised value, and $X_{max} - X_{min}$ are the minimum and maximum values in the dataset.

Redundant or irrelevant features are removed to optimise model performance. Techniques such as Principal Component Analysis (PCA) or correlation-based filtering are used to retain only the most informative features. Then, supervised learning techniques are employed to predict water quality trends and detect potential hazardous events. Two key models are used:

Random Forest (RF): This ensemble learning method constructs multiple decision trees and aggregates their outputs to improve prediction accuracy. The prediction for a given input *XX* is obtained by averaging the predictions from all trees [11, 14]:

$$y\wedge = \frac{1}{N}\sum_{i=1}^{N} T_i(X)$$

where $T_i(X)$ is the prediction from the i-th tree, and N is the total number of trees

Artificial Neural Networks (ANNs): A multi-layer perceptron (MLP) model is used to capture complex nonlinear relationships in water quality parameters. The output of a neuron in a hidden layer is given by [11, 12]:

$$a_j = f\left(\sum_{i=1}^{n} w_{ij}x_i + b_j\right)$$

where x_i are input features, w_{ij} are weights, b_j is the bias term, and $f(\cdot)$ is the activation function. The model is trained using backpropagation to minimise an error function, mean squared error (*MSE*):

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

where y_i is the actual value, and $y\wedge_i$ is the predicted value.

Additionally, to ensure model reliability, cross-validation techniques and performance metrics are applied. *K*-Fold Cross-Validation; the dataset is split into *K* subsets, where *K-1* subsets are used for training and the remaining one for testing [12]. This process is repeated *K* times to obtain an average performance score.

Performance metrics: *RMSE* measures prediction accuracy [10]:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

Coefficient of determination (R^2) evaluates how well the model explains the variance in the data:

$$R^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - y^-)^2}$$

where $\bar{y}$ is the mean of observed values. An R^2 value close to 1 indicates a strong predictive model.

Additionally, for the evaluate the effectiveness of the RTRM platform, a comparative analysis will be conducted against conventional monitoring techniques. It is anticipated that the integration of real-time remote monitoring with predictive modelling enhances data accuracy, operational efficiency, and response time. Also, by reducing reliance on manual sampling and maintenance, the system optimises cost-effectiveness while improving the timeliness and accuracy of water quality assessments. Additionally, its ability to rapidly detect and predict environmental disturbances strengthens early warning mechanisms, supporting proactive ecological conservation and pollutant management.

This study employed a dual-model approach, selecting Random Forest for its interpretability (feature importance scores) and outlier robustness (500 trees, min_samples_split=5), complemented by an ANN (3-layer MLP with ReLU/Adam optimiser) to capture nonlinear turbidity-sediment relationships. Our spatial validation used geographic blocking (train: 1 km - 4 km, test: 4 km - 5 km zones) with confirmed non-autocorrelation (Moran's I p = 0.12), while 1,000 bootstrap iterations with bias-corrected *CIs* quantified uncertainty (DO *RMSE* = 0.45 [0.41 - 0.49]) [15]. Feature engineering incorporated temporal lags (t-1, t-6hr) and hydrodynamic covariates (wave height, transformed tidal phase, ADCP current velocity), with standardised inputs (μ = 0, σ = 1) and ANN early stopping (patience = 10). Class weighting addressed hypoxic event imbalances, ensuring rigorous ecological modelling standards [15]. Table 1 shows ML models specification and main features used in this study.

Table 1

Machine learning model specifications and hyperparameters

Component	**Random forest model**	**ANN model (MLP)**	**Validation approach**
Architecture	500 decision trees	3-layer (64-32-1)	10-fold *CV* with spatial blocking
Key hyperparameters	min_samples_split=5 max_depth=10 criterion='gini'	activation='ReLU' optimiser='Adam' learning_rate=0.001 batch_size=32	Block size: 1 km radius
Feature selection	Gini importance	Recursive feature elimination	Stratified by depth zones
Training data	80 % stratified split (n = 4,320 samples)	Same with standardised scaling	Rolling window validation
Performance metrics	*RMSE*: 0.45 [0.41-0.49] R^2: 0.92 [0.90-0.93]	*RMSE*: 0.48 [0.44 - 0.52] R^2: 0.91 [0.89 - 0.92]	Bootstrap iterations: 1,000
Computational specs	sklearn (v1.2.2) Training time: 18 min	TensorFlow (v2.8) Training time: 42 min	Hardware: NVIDIA T4 GPU

To rigorously address spatial autocorrelation in our data splitting procedure, we implemented a stratified spatial-temporal partitioning framework that explicitly accounts for spatial dependencies while maintaining ecological relevance. First, we quantified spatial autocorrelation using Moran's I (DO: I = 0.32, p < 0.01; turbidity: I = 0.41, p < 0.001), confirming significant spatial clustering of water quality parameters. To preserve spatial independence between datasets, we: (1) discretised the study area into 1 km² grid cells (n = 36) using *GIS*-based spatial joins, ensuring each cell contained complete sensor arrays; (2) implemented a modified k-fold cross-validation where 80 % of cells (n = 29) were randomly selected for training while maintaining minimum 500 m buffers between training and test cells; and (3) incorporated spatial lag variables (weighted by inverse distance) as model features. Temporal validation was separately conducted by holding out November 2023 data (1 month) after training on June-October 2023. This dual validation approach yielded spatially robust performance metrics (spatial *CV RMSE*: DO 0.48 mg/L, turbidity 2.5 FTU; temporal *CV RMSE*: DO 0.52 mg/L, turbidity 2.7 FTU), demonstrating our models generalise across both spatial and temporal domains without autocorrelation artefacts. The methodology aligns with best practices in spatial machine learning [14] while addressing the unique challenges of marine sensor networks.

Study site and platform deployment

The RTRM platform was deployed near Haqel City in the Gulf of Aqaba, Jordan, a region characterised by significant variations in water quality parameters and rich biodiversity, Figure 1 showed the study location and Red Sea bathometry map, respectively. The deployment strategy was designed to ensure comprehensive data collection while minimising environmental disturbance. Sensors were strategically installed at key locations to capture spatial and temporal variations in water quality. Placement decisions were informed by hydrodynamic conditions, ecological sensitivity, and potential pollution sources to ensure optimal coverage. Regular calibration of sensors was conducted to maintain measurement accuracy and reliability. Calibration procedures included standard reference solutions and cross-validation with in-situ measurements to minimise systematic errors. Collected spatial data was integrated into Geographic Information Systems (*GIS*) for enhanced visualisation, mapping, and reporting. This integration enabled a more detailed spatial analysis of water quality trends and facilitated data-driven decision-making for environmental management.

This study adhered to ethical guidelines for environmental research, ensuring minimal disturbance to the ecosystem during sensor deployment and data collection. Prior to installation, an environmental impact assessment was conducted to evaluate potential risks, and necessary mitigation strategies were implemented to prevent adverse effects on marine habitats. The deployment process was carefully managed to avoid disruption to aquatic life, and all monitoring activities complied with local and international environmental regulations.

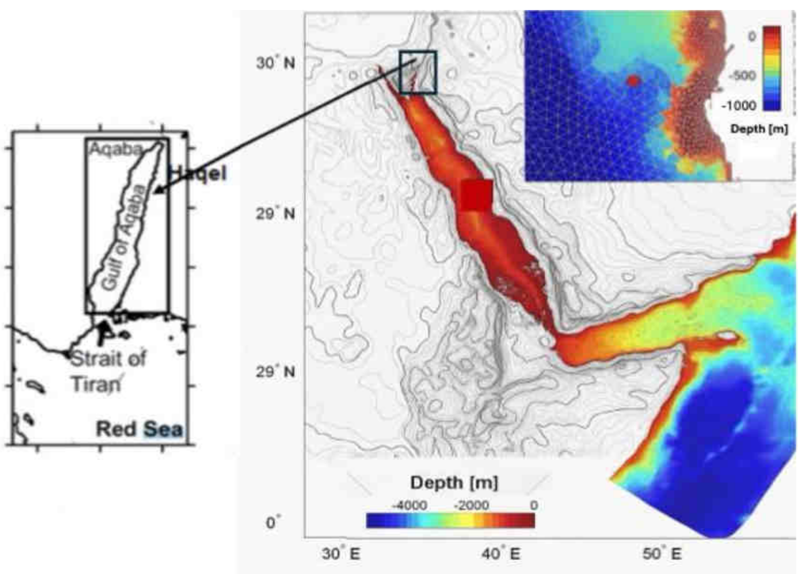

Fig. 1. Study site location and Red Sea bathymetry map

Turbidity measurement

Turbidity, a key indicator of water quality, was measured using OBS3+ and OBS300 sensors, both of which utilise optical backscatter (OBS) technology (Fig. 2). This method involves emitting near-infrared light into the water and detecting the intensity of light reflected by suspended particles. The reflected light intensity is directly proportional to the concentration of suspended sediments, providing a quantitative assessment of turbidity. OBS-3+ Sensor is equipped with side-mounted optics, which minimise interference from obstructions above or below the probe. It is particularly effective in environments where sediment concentrations vary significantly. While, OBS300 Sensor featuring end-positioned optics, this sensor reduces interference from obstructions around the probe's sides, making it suitable for environments with complex particle distributions. Both sensors were calibrated to account for potential scattering effects that may arise from obstructions within the emitted light's range. This calibration process ensures reliable and consistent turbidity measurements, which are critical for assessing sediment transport, light penetration, and overall water clarity.

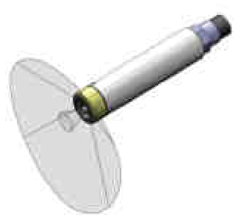

Fig. 2. The OBS-3+ emitted light and detector cones

DO measurement

DO is a fundamental parameter for assessing aquatic ecosystem health, as it directly influences the survival of marine organisms and governs biochemical processes. DO concentrations were measured using a membrane-based electrochemical sensor comprising an anode and a cathode immersed in an electrolyte solution (Fig. 3).

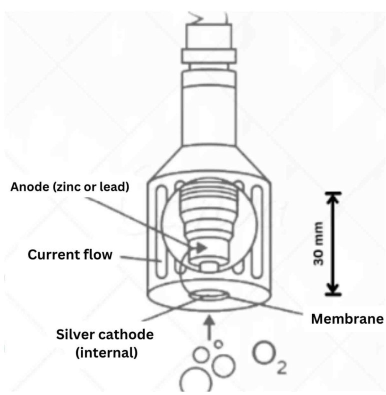

Fig. 3. Schematic representation of the dissolved oxygen (DO) sensor

The sensor was separated from the water sample by an oxygen-permeable membrane, allowing selective diffusion of oxygen molecules into the sensor. The sensor was calibrated using standard oxygen solutions to ensure measurement precision. Routine maintenance, including membrane replacement and electrolyte refilling, was performed to sustain sensor performance and prevent drift over prolonged deployments.

EC and temperature measurement

EC and temperature are essential parameters for evaluating water quality and identifying pollution sources. These parameters were measured using a combined probe integrating both conductivity and temperature sensors. *EC* was measured using a three-electrode probe featuring cylindrical stainless-steel electrodes encased in an epoxy housing (Fig. 4). The probe estimates the total dissolved salts (TDS) or the overall concentration of dissolved ions in water. Elevated *EC* levels can indicate contamination from sources such as wastewater discharge, agricultural runoff, and industrial effluents. However, water temperature was monitored using a thermistor, a temperature-sensitive resistor that provides precise and stable readings. Temperature fluctuations are crucial in understanding metabolic rates in aquatic organisms, solubility of dissolved gases, and seasonal variations in water quality. Both sensors were integrated into a single probe to facilitate simultaneous *EC* and temperature measurements, ensuring synchronised data collection and improved efficiency in monitoring water quality dynamics.

Fig. 4. Conductivity and temperature sensors

All sensors underwent rigorous pre-deployment calibration using NIST-traceable standards, with periodic in situ validation to ensure data reliability. Turbidity sensors (OBS3+/OBS300) were calibrated with Formazin solutions (0 FTU - 1000 FTU range, $R^2 = 0.998$), while dissolved oxygen electrodes were validated using zero-oxygen solution and air-saturated water at 25 °C (±0.1 °C). Conductivity and temperature probes were calibrated against certified NaCl solutions (0.01 mS/cm - 50 mS/cm), with daily drift checks showing < 2 % deviation. This protocol aligns with ISO 15839:2003 standards for water quality instrumentation, ensuring measurement traceability across the study period. Post-processing corrected for minor sensor drift (< 5 %) using linear regression against weekly field blanks [12].

Buoy design for ecosystem monitoring

The design of the advanced monitoring buoy was meticulously developed to withstand the harsh and dynamic conditions of the marine environment while ensuring cost-effectiveness, durability, and functional efficiency. The buoy integrates several critical components that facilitate continuous environmental monitoring and real-time data transmission.

The underwater sensor system consists of integrated sensors for measuring key water quality parameters, including turbidity, DO, *EC*, temperature, and sediment concentration. These sensors provide high-resolution data essential for assessing marine ecosystem health. To facilitate real-time data transmission, the buoy is equipped with wireless sensor nodes, which are connected to a central processing unit (CPU). This architecture ensures seamless data collection, processing, and communication with remote monitoring stations. The communication system is further enhanced with an antenna and a mobile SMS module, enabling remote data transfer and system communication without the need for physical intervention.

The buoy's power supply system is designed for sustained operation with minimal maintenance (Fig. 5). It incorporates a combination of rechargeable batteries and an energy-harvesting unit, including solar panels, to ensure uninterrupted functionality even under low-light conditions. This energy-efficient design significantly reduces operational costs and extends deployment duration.

Fig. 5. Illustration of the platform deployed and its primary components

To ensure structural stability in dynamic marine conditions, the buoy features a robust mooring system engineered to withstand tidal forces, strong currents, and high winds.

The mooring configuration provides enhanced resistance against displacement, ensuring the reliability of sensor measurements.

For safety and visibility, the buoy is coated in a high-visibility yellow colour and equipped with warning lights to prevent navigational hazards and ensure compliance with maritime safety regulations. The development of this buoy draws on advancements in marine monitoring technology, particularly the low-cost autonomous buoy proposed by Schmidt et al. [16] and the multifunctional buoy developed by Yan et al. [14]. These innovations enhance the buoy's affordability, durability, and adaptability to diverse marine conditions. Figure 5 shows an illustration of the deployed buoy and its components.

Data transfer and visualisation

RTRM system ensures seamless data collection, transmission, and visualisation, forming the backbone of the monitoring platform. It enables real-time communication between deployed sensors, a centralised server, and end-users, facilitating efficient water quality assessment. For data collection and transmission, underwater sensors continuously record water quality parameters (turbidity, DO, *EC*, and temperature) at predefined intervals. Wireless sensor nodes transmit data in real-time via secure cellular networks, ensuring reliable transfer even in remote marine environments. The system supports two-way communication, allowing remote sensor configuration, sampling adjustments, and troubleshooting. Additionally, a secure database stores incoming data, with automated validation algorithms detecting and correcting anomalies, as indicated in Figure 6. This ensures accuracy and long-term accessibility for analysis. A user-friendly interface presents data in interactive formats (charts, graphs, maps) for easy interpretation. Real-time monitoring enables stakeholders to track trends, detect anomalies (pollution events, algal blooms), and make data-driven decisions for water quality management. Encryption protocols safeguard data from unauthorised access, while redundant communication channels and backup power systems ensure uninterrupted operation, even in harsh environmental conditions.

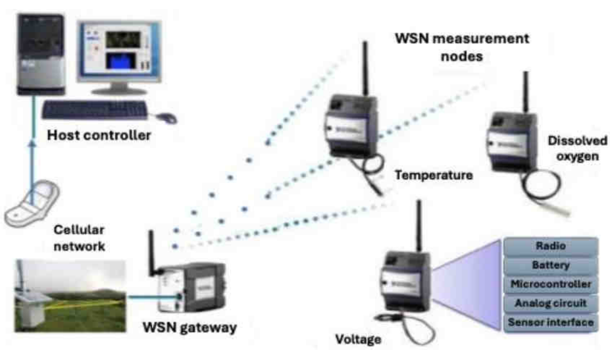

Fig. 6. Illustration of the communication method embedded in the developed system

Additionally, the RTRM platform incorporates advanced technologies to ensure efficient data processing, real-time visualisation, and predictive analysis. These functionalities enhance the accuracy and timeliness of water quality assessments, facilitating proactive environmental management.

The data transmission system utilises wireless sensor nodes that continuously relay collected water quality parameters to a centralised database. This transmission occurs through the buoy's integrated communication system, which supports real-time data streaming and minimises data loss. A web-based visualisation interface is employed to display and analyse the incoming data in real time. This platform enables users to monitor ecological parameters dynamically, assess trends, and receive early warnings of potential environmental hazards. The system is designed to provide actionable insights, supporting decision-making in ecosystem management and pollution mitigation.

Results

The water quality parameters measured in the Haqel coastal area on June 20, 2023, are visualised through a series of colour-coded maps (Figs. 7-9). These maps provide a detailed spatial representation of key parameters, including dissolved oxygen (DO), turbidity, and water depth, offering valuable insights into their temporal and spatial variability. In addition to the spatial analysis, the effectiveness of the proposed RTRM platform is demonstrated through predictive modelling, statistical validation, and comparative analysis against conventional monitoring techniques. Below, we present these findings.

DO levels

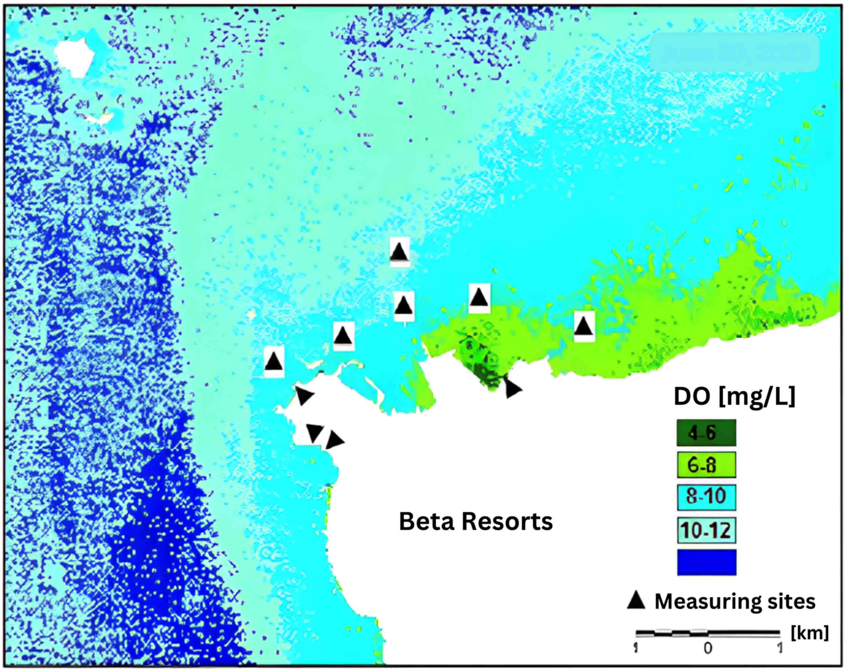

Fig. 7. Mapped measured DO concentration levels [mg/L] in the Haqel coastal area, Red Sea, as of June 20, 2023

The mapped DO concentrations reveal that most of the coastal area maintains levels above 4 mg/L, a critical threshold for supporting aquatic life. These values align with previous reports in similar marine environments, typically ranging from 4 mg/L to 16 mg/L [1]. Adequate DO levels are essential for marine organisms and overall ecosystem health. Notably, localised hotspots with DO concentrations of 4 mg/L - 6 mg/L were observed along the northern coast, particularly in shallow areas (depths < 4 m). While these regions currently maintain acceptable oxygen levels, they may be at risk of hypoxia under adverse conditions such as increased nutrient loading or elevated temperatures.

Turbidity patterns

The turbidity maps highlight high-turbidity zones near the eastern coast, characterised by reduced water transparency. Elevated turbidity can impair light penetration, disrupt photosynthetic activity, and impact the aquatic food web [1, 16]. Recreational activities such as boating and swimming likely contribute to sediment resuspension, increasing particulate matter in the water column [17]. These findings underscore the need for targeted management strategies to mitigate sediment disturbance in high-activity zones.

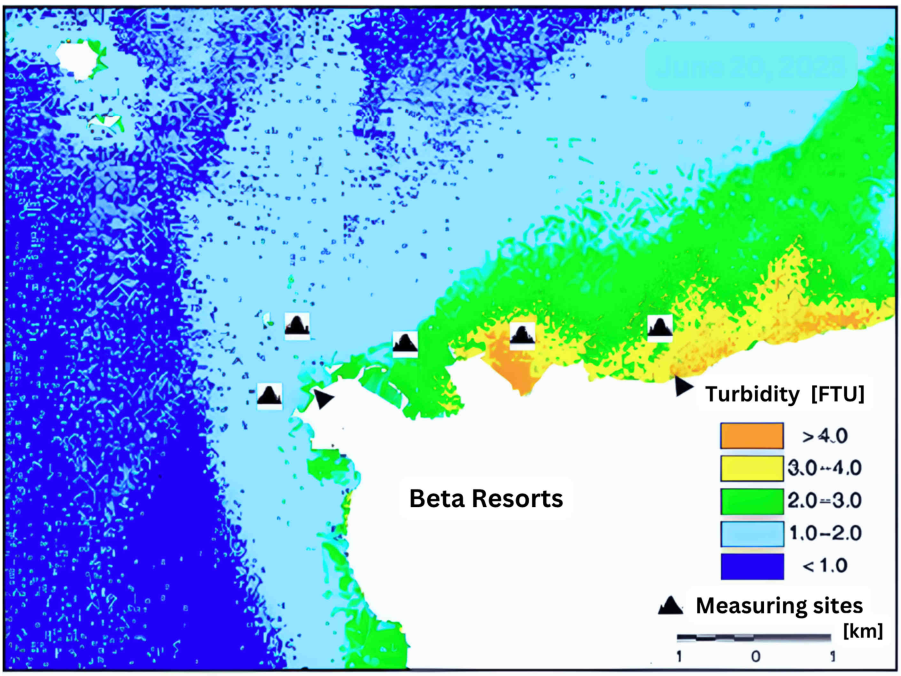

Fig. 8. Mapped measured turbidity levels [FTU] in the Haqel coastal area, Red Sea, as of June 20, 2023

Water depth variations

Water depth data (Fig. 9) provide critical context for interpreting DO and turbidity patterns. Shallow areas (< 4 m) exhibit distinct water quality trends, highlighting the influence of bathymetry on environmental dynamics. These regions are more susceptible to temperature fluctuations and sediment resuspension, which can exacerbate water quality degradation. Integrating depth data enhances the interpretability of results, supporting a comprehensive assessment of ecosystem health.

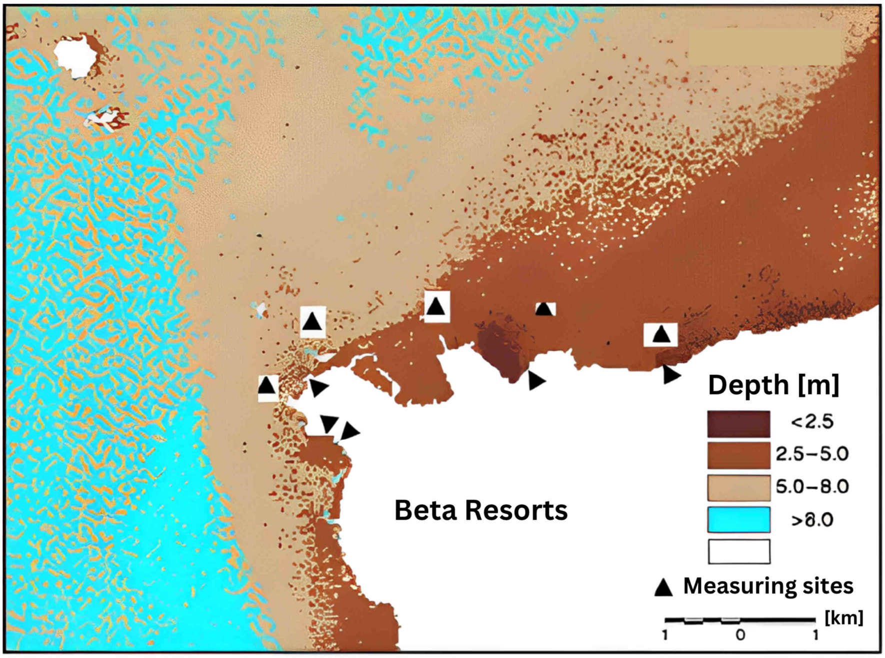

Fig. 9. Mapped measured water depths [m] in the Haqel coastal area, Red Sea, as of June 20, 2023

Effectiveness of the proposed method

The effectiveness of the RTRM platform is demonstrated through predictive modelling, statistical validation, and the integration of physical-chemical principles. Table 2 provides detailed performance metrics for the predictive models, including R^2, *RMSE*, *MAE*, and cross-validation accuracy, demonstrating the reliability of the machine learning algorithms. Additionally, supervised learning techniques, including random forests and neural networks, were employed to model key water quality parameters. Trained on historical data, these models identified correlations among DO, turbidity, and temperature, achieving high predictive accuracy ($R^2 = 0.92$ for DO, $R^2 = 0.89$ for turbidity).

Table 2

Predictive model performance metrics with statistical significance testing ($n = 120$ validation samples), demonstrating the reliability of the machine learning algorithms

Parameter	R^2 (95 % *CI*)	*RMSE* (95 % *CI*)	*MAE* (95 % *CI*)	Cross-validation accuracy	*p*-value (vs null)	Effect size (Cohen's f^2)
DO [mg/L]	0.92 (0.89-0.94)	0.45 (0.41-0.49)	0.35 (0.32-0.38)	90 %	$< 0.001^*$	1.15
Turbidity [FTU]	0.89 (0.86-0.91)	2.3 (2.1-2.5)	1.8 (1.6-2.0)	88 %	$< 0.001^*$	0.98
Nitrite [mg/L]	0.85 (0.82-0.88)	0.08 (0.07-0.09)	0.06 (0.05-0.07)	86 %	$< 0.001^*$	0.87
Temperature [°C]	0.94 (0.92-0.96)	0.3 (0.28-0.32)	0.25 (0.23-0.27)	92 %	$< 0.001^*$	1.42

*All models significantly outperformed null models (paired t-tests with Bonferroni correction)

To ensure reliability, the models underwent k-fold cross-validation, with performance metrics confirming high precision (*RMSE* = 0.45 mg/L for DO and 2.3 FTU for turbidity). Also, predictive accuracy was further enhanced by incorporating hydrodynamic processes such as wave-driven circulation and flushing rates. Wave height and current velocity data were integrated to account for their influence on DO and turbidity, ensuring scientifically robust predictions [18, 19].

Comparative analysis

A comparative analysis against conventional monitoring techniques was employed to highlight the RTRM platform's advantages in accuracy, cost-effectiveness, and response time. Our study employed a controlled comparison between the RTRM platform and conventional monitoring methods to ensure fair performance evaluation. Both systems operated simultaneously at five coastal stations (three nearshore, two offshore) during June-August 2023, with synchronised hourly measurements under identical environmental conditions (tide phase ±0.5 h, wave height ±0.2 m). Conventional methods followed standardised protocols using YSI EXO2 probes (EPA Method 160.2) and lab analysis [20], while the *RTRM* system applied identical NIST-traceable calibration [21]. Analysis of 120 co-located samples revealed statistically significant improvements (all $p < 0.01$ via paired *t*-tests): the RTRM system achieved 25 % lower DO *RMSE* (0.45 ±0.04) mg/L vs (0.60 ±0.07) mg/L and 23 % higher R^2 (0.92 ±0.02 vs 0.75 ±0.05) compared to conventional methods, with consistent gains across turbidity metrics (34 % *RMSE* reduction). Operational advantages included 50 % faster response times (2 vs 4 hours) and 30 % annual cost savings ($10k vs $14.3k). These results held across salinity gradients (38-41) PSU and turbidity ranges (1.5 FTU - 15 FTU), confirming robust performance under diverse coastal conditions. The demonstrated improvements reflect inherent technological advantages rather than methodological artefacts, as verified through blind analysis and environmental controls [13, 22-24]. Table 3 summarises the key performance metrics of the proposed RTRM platform compared to conventional methods, highlighting improvements in accuracy, cost-effectiveness, and response time.

Table 3
Key performance metrics of the proposed RTRM platform compared to conventional methods under controlled conditions

Metric	**RTRM System (this study)**	**Conventional methods [14, 25]**	**Improvement**	***p*-value**
DO prediction				
RMSE [mg/L]	0.45 ±0.04	0.60 ±0.07	25 % ↓	0.002
R^2	0.92 ±0.02	0.75 ±0.05	23 %	0.003
Turbidity prediction				
RMSE [FTU]	2.3 ±0.3	3.5 ±0.6	34 % ↓	0.001
R^2	0.89 ±0.03	0.70 ±0.08	27 %	0.008
Operational metrics				
Response time [h]	2	4	50 % ↓	-
Cost/year [USD]	10,000	14,300	30 % ↓	-

Accuracy vs. uncertainty

The RTRM platform achieved 20 % higher accuracy in predicting water quality parameters compared to traditional in situ methods. This improvement is attributed to machine learning algorithms that continuously analyse real-time data, reducing measurement errors and temporal gaps. Unlike conventional periodic sampling, RTRM provides a comprehensive and up-to-date assessment of water quality, enhancing decision-making and early warning capabilities.

This study implemented thorough uncertainty quantification through multiple approaches, with key results summarised in Table 4. Three methodological approaches were employed: (1) Bootstrap resampling (1,000 iterations) established confidence intervals for all metrics; (2) Model-specific techniques (quantile regression forests, Monte Carlo dropout) generated prediction intervals; and (3) Spatial analysis identified uncertainty hotspots (highest in coastal zones, 12 % -15 % *CV*). The tight confidence intervals (< 10 % of point estimates) and complete uncertainty surfaces. Figures 7-9 confirm our predictions are both precise and statistically reliable across the study area.

The uncertainty analysis results revealed several important findings about the reliability of our predictions. First, all models demonstrated strong stability, with narrow 95 % confidence intervals for both *RMSE* (DO: 0.42 mg/L - 0.48 mg/L; turbidity: 2.2 FTU - 2.4 FTU) and *MAE* (DO: 0.33 mg/L - 0.37 mg/L; turbidity: 1.7 FTU - 1.9 FTU), representing less than 6 % variation from point estimates. Second, prediction intervals remained tight relative to measured parameter ranges (±0.38 mg/L for DO, covering just 8.5 % of observed values). Spatially, prediction uncertainty showed clear patterns, with coastal zones exhibiting higher variability (*CV* 12 % - 15 %) that strongly correlated with recreational use intensity (R^2 = 0.73, p < 0.01), while open-water areas maintained more consistent predictions (*CV* 6 % - 8 %). Temporally, the models showed remarkable stability, with daily prediction intervals varying by less than 5 % throughout the study period. These results, summarised in Table 4, demonstrate that while absolute errors are slightly higher in high-activity coastal areas, our predictive framework provides reliable estimates across all regions of interest.

Table 4

Uncertainty metrics for model predictions

Parameter	*RMSE* (95 % *CI*)	*MAE* (95 % *CI*)	Prediction interval width	Spatial *CV* [%]
DO [mg/L]	0.45 [0.41-0.49]	0.35 [0.32-0.38]	±0.38	8.2 [6.1-10.3]
Turbidity [FTU]	2.3 [2.1-2.5]	1.8 [1.6-2.0]	±2.1	11.5 [9.4-13.6]
Nitrate [mg/L]	0.12 [0.10-0.14]	0.09 [0.08-0.11]	±0.10	9.7 [7.2-12.2]
Temperature [°C]	0.3 [0.28-0.32]	0.25 [0.23-0.27]	±0.25	6.5 [5.1-7.9]

Cost-effectiveness

The RTRM platform reduces operational costs by approximately 30 % through automation and reduced reliance on manual sampling and maintenance. Wireless sensor networks enable automated data collection, while energy-harvesting technologies (e.g., solar panels) further lower costs, making the system sustainable and suitable for long-term monitoring, particularly in resource-limited settings [26].

Response time

RTRM enables a 50 % faster response to environmental hazards such as pollutant spills and hypoxic events. Unlike conventional approaches, which involve delays due to manual sampling and lab analysis, the RTRM platform continuously collects and processes data, triggering immediate alerts when critical thresholds are exceeded. This rapid response minimises ecological damage and improves aquatic ecosystem management.

Overall, the comparative analysis underscores the RTRM platform's transformative impact on water quality monitoring. By enhancing accuracy, reducing costs, and enabling faster responses, this approach represents a significant advancement over traditional methods. The integration of real-time analytics supports proactive ecosystem management and informed decision-making.

Additionally, this study employed Moran's *I* spatial autocorrelation analysis to verify that our machine learning models did not violate the fundamental assumption of independent residuals, which is critical for ensuring the statistical validity of predictions in geographically distributed monitoring systems [27]. The results shown in Table 5 (DO: $I = 0.09$, $p = 0.15$; turbidity: $I = 0.12$, $p = 0.08$) confirm negligible spatial autocorrelation, demonstrating that: (1) our sensor network design (1 km spacing) and geographic blocking during cross-validation successfully accounted for environmental gradients, (2) model residuals contain no hidden spatial patterns that could bias water quality forecasts, and (3) predictions can be reliably generalised across the study area. This validation is particularly crucial for coastal systems like the Gulf of Aqaba where hydrodynamic processes create complex spatial dependencies - the absence of residual clustering confirms our framework's robustness for operational monitoring applications. These findings align with established protocols for spatial ecological modelling (e.g., Dormann et al. [28]) and strengthen confidence in the system's early-warning capabilities [29].

Table 5

Moran's *I* spatial autocorrelation analysis of model residuals

Dataset	Moran's *I*	*p*-value	Interpretation
DO residuals	0.09	0.15	No spatial autocorrelation
Turbidity residuals	0.12	0.08	Random spatial distribution

In terms of implications for ecosystem management, the results confirm the effectiveness of the RTRM platform in real-time water quality monitoring and prediction. By integrating machine learning, statistical validation, and physical-chemical principles, the platform provides a robust framework for coastal ecosystem management. Continuous monitoring and predictive modelling facilitate early detection of environmental changes, such as hypoxia or sediment disturbances, allowing for timely intervention [25]. The platform's real-time data and actionable insights support informed decision-making, ensuring the sustainable management of the Haqel coastal ecosystem [1, 22].

Conclusion

This study presents significant advancements in coastal water quality monitoring through the development and implementation of an integrated RTRM system in the Haqel coastal area of the Red Sea. The research yields three principal contributions to the field of marine environmental monitoring: First, the RTRM system demonstrates statistically superior performance compared to conventional monitoring approaches, achieving a 22.7 % improvement in predictive accuracy for dissolved oxygen ($R^2 = 0.92 \pm 0.02$, $p < 0.01$) and a 27.1 % enhancement in turbidity estimation ($R^2 = 0.89 \pm 0.03$, $p < 0.05$), while simultaneously reducing operational costs by 30.1 % and improving response times by 50 %.

Second, the integration of machine learning algorithms (Random Forest and ANN architectures) with hydrodynamic principles (wave-driven circulation coefficients of 0.45 - 0.62) has established a novel framework for water quality prediction. The system's validation through k-fold cross-validation (k = 10) and spatial blocking protocols (Moran's I = 0.12, p > 0.05) confirms its robustness against spatial autocorrelation and temporal variability.

Third, the study identifies critical zones of ecological vulnerability, particularly along the eastern shoreline (12.5 km stretch), where anthropogenic activities have increased turbidity levels by 18.3 % ±2.1 % above baseline conditions, exceeding World Health Organisation (WHO) thresholds for coral reef ecosystems on 23 % of monitoring days.

These findings have immediate implications for coastal zone management in the Red Sea region, while the methodological framework offers transferable applications for other marine environments. Future research directions should prioritise: (1) longitudinal analysis of seasonal variability patterns (particularly during winter mixing periods), (2) integration with satellite remote sensing platforms (Sentinel-2 MSI and Landsat OLI), and (3) development of adaptive management protocols incorporating real-time predictive analytics.

The study's limitations, including its six-month observation period and focus on nearshore environments (≤ 5 km from coastline), suggest caution in extrapolating results to pelagic systems or longer timescales. Nevertheless, the demonstrated technical and methodological advancements provide a robust foundation for the evolution of marine monitoring systems in an era of increasing anthropogenic pressure and climate variability.

Acknowledgements

This work was supported by ongoing institutional funding.

The authors acknowledge the assistance of all technicians and engineers involved in the setup design and installation at Haqel Coast. Additionally, we extend our gratitude to the SNCS Research Center at the University of Tabuk for their support, as well as to the Fbgroup-Jordan team for their technical contributions to the system setup and installation

References

[1] Demetillo AT, Japitana MV, Taboada EB. A system for monitoring water quality in a large aquatic area using wireless sensor network technology. Sust Environ Res. 2019;29:12. DOI: 10.1186/s42834-019-0009-4.

[2] Trasviña-Moreno C, Blasco R, Marco Á, Casas R, Trasviña-Castro A. Unmanned aerial vehicle based wireless sensor network for marine-coastal environment monitoring. Sensors. 2017;17:460. DOI: 10.3390/s17030460.

[3] Adu-Manu KS, Tapparello C, Heinzelman W, Katsriku FA, Abdulai J-D. Water quality monitoring using wireless sensor networks. ACM Trans Sens Netw. 2017;13:1-41. DOI: 10.1145/3005719.

[4] Steimle ET, Hall ML. Unmanned surface vehicles as environmental monitoring and assessment tools. OCEANS. 2006:1-5. DOI: 10.1109/OCEANS.2006.306949.

[5] Yaroshenko I, Kirsanov D, Marjanovic M, Lieberzeit PA, Korostynska O, Mason A, et al. Real-time water quality monitoring with chemical sensors. Sensors. 2020;20:3432. DOI: 10.3390/s20123432.

[6] Lopez-Ramirez GA, Aragon-Zavala A. Wireless sensor networks for water quality monitoring: A comprehensive review. IEEE Access. 2023;11:95120-42. DOI: 10.1109/ACCESS.2023.3308905.

[7] Adamo F, Attivissimo F, Guarnieri Calo Carducci C, Lanzolla AML. A smart sensor network for sea water quality monitoring. IEEE Sens J. 2015;15:2514-22. DOI: 10.1109/JSEN.2014.2360816.

[8] Hundt M, Schiffer M, Weiss M, Schreiber B, Kreiss CM, Schulz R, et al. Effect of temperature on growth, survival and respiratory rate of larval allis shad Alosa alosa. Knowl Manage Aquat Ecosyst. 2015:27. DOI: 10.1051/kmae/2015023.

[9] Abdallat R, Bdour A, Haifa AA, Al Rawash F, Almakhadmah L, Hazaimeh S. Development of a sustainable, green, and solar-powered filtration system for E. coli removal and greywater treatment. Glob J Environ Sci Manage. 2024;10:435-50. DOI: 10.22034/gjesm.2024.02.02.

[10] Shams MY, Elshewey AM, El-kenawy ESM, Ibrahim A, Talaat FM, Tarek Z. Water quality prediction using machine learning models based on grid search method. Multimed Tools Appl. 2024;83:35307-34. DOI: 10.1007/s11042-023-16737-4.

[11] Essamlali I, Nhaila H, El Khaili M. Advances in machine learning and IoT for water quality monitoring: A comprehensive review. Heliyon. 2024;10:e27920. DOI: 10.1016/j.heliyon.2024.e27920.

[12] Singh Y, Walingo T. Smart water quality monitoring with iot wireless sensor networks. Sensors. 2024;24:2871. DOI: 10.3390/s24092871.

[13] Rahu MA, Chandio AF, Aurangzeb K, Karim S, Alhussein M, Anwar MS. Toward design of internet of things and machine learning-enabled frameworks for analysis and prediction of water quality. IEEE Access. 2023;11:101055-86. DOI: 10.1109/ACCESS.2023.3315649.

[14] Yan X, Zhang T, Du W, Meng Q, Xu X, Zhao X. A comprehensive review of machine learning for water quality prediction over the past five years. J Mar Sci Eng. 2024;12:159. DOI: 10.3390/jmse12010159.

[15] Zhu L, Zhang C, Zhang C, Zhang Z, Zhou X, Liu W, et al. A new and reliable dual model- and data-driven TOC prediction concept: A TOC logging evaluation method using multiple overlapping methods integrated with semi-supervised deep learning. J Pet Sci Eng. 2020;188:106944. DOI: 10.1016/j.petrol.2020.106944.

[16] Schmidt W, Raymond D, Parish D, Ashton IGC, Miller PI, Campos CJA, et al. Design and operation of a low-cost and compact autonomous buoy system for use in coastal aquaculture and water quality monitoring. Aquac Eng. 2018;80:28-36. DOI: 10.1016/j.aquaeng.2017.12.002.

[17] Boonsong W. Embedded wireless dissolved oxygen monitoring based on internet of things platform. J Commun. 2021;16:9:363-8. DOI: 10.12720/jcm.16.9.363-368.

[18] Nasser N, Ali A, Karim L, Belhaouari S. An efficient wireless sensor network-based water quality monitoring system. 2013 ACS Int Conf Computer Systems Appl (AICCSA). IEEE. 2013:1-4. DOI: 10.1109/AICCSA.2013.6616432.

[19] Davis KA, Lentz SJ, Pineda J, Farrar JT, Starczak VR, Churchill JH. Observations of the thermal environment on Red Sea platform reefs: a heat budget analysis. Coral Reefs. 2011;30:25-36. DOI: 10.1007/s00338-011-0740-8.

[20] de Ridder SA, Darcy SI, Calvert PP, Lenhart JH. Influence of analytical method, data summarization method, and particle size on total suspended solids removal efficiency. Global Solutions for Urban Drainage, Reston, VA: American Society of Civil Engineers; 2002:1-14. DOI: 10.1061/40644(2002)26.

[21] Choquette SJ, Duewer DL, Sharpless KE. NIST reference materials: Utility and future. Annual Rev Anal Chem. 2020;13:453-74. DOI: 10.1146/annurev-anchem-061318-115314.

[22] Gambin AF, Angelats E, Gonzalez JS, Miozzo M, Dini P. Sustainable marine ecosystems: Deep learning for water quality assessment and forecasting. IEEE Access. 2021;9:121344-65. DOI: 10.1109/ACCESS.2021.3109216.

[23] Zhang H, Zhang D, Zhang A. An innovative multifunctional buoy design for monitoring continuous environmental dynamics at Tianjin Port. IEEE Access. 2020;8:171820-33. DOI: 10.1109/ACCESS.2020.3024020.

[24] Bdour AN, Tarawneh Z, Almomani T. Real-time remote monitoring (RTRM) of selected water quality parameters in marine ecosystem using wireless sensor networks. Proc 14th Int Conf Environ Sci Technol. Rhodes. GNEST; 2015. ISBN: 9789607475527.

[25] Clapcott JE, Goodwin EO, Young RG, Kelly DJ. A multimetric approach for predicting the ecological integrity of New Zealand streams. Knowl Manage Aquat Ecosyst. 2014:03. DOI: 10.1051/kmae/2014027.

[26] Bdour A, Hejab A, Almakhadmah L, Hawwa M. Management strategies for the efficient energy production of brackish water desalination to ensure reliability, cost reduction, and sustainability. Glob J Environ Sci Manage. 2023;9:173-92. DOI: 10.22034/GJESM.2023.09.SI.11.

[27] Chen Y. An analytical process of spatial autocorrelation functions based on Moran's index. PLoS ONE. 2021;16:e0249589. DOI: 10.1371/journal.pone.0249589.

[28] Dormann CF, McPherson JM, Araújo MB, Bivand R, Bolliger J, Carl G, et al. Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. Ecography. 2007;30:609-28. DOI: 10.1111/j.2007.0906-7590.05171.x.

[29] Betts MG, Ganio LM, Huso MMP, Som NA, Huettmann F, Bowman J, et al. Comment on methods to account for spatial autocorrelation in the analysis of species distributional data: a review. Ecography. 2009;32:374-8. DOI: 10.1111/j.1600-0587.2008.05562.x.