

Modelling the Water Quality of River Mara Using an Adaptive Neuro-Fuzzy Inference System (ANFIS)

Stephano M. Alphayo

`stephen.alphayo@yahoo.com`

Mbeya University of Science and Technology

Zacharia Katambara

Mbeya University of Science and Technology

Livingstone Swilla

Water Institute

Research Article

Keywords: ANFIS, Ecosystem Maintenance, Transboundary Resources, Water Pollution, Water Quality Indices

Posted Date: October 23rd, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11065815/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Modelling the Water Quality of River Mara Using an Adaptive Neuro-Fuzzy Inference System (ANFIS)

Abstract

Water is vital for human beings and the ecosystem. Thus, water quality (WQ) is extremely important. Seasonal variability and nonlinear interactions among WQ variables present major challenges for effective river WQ assessment and prediction. This study investigated seasonal variations in WQ and evaluated the performance of an Adaptive Neuro-Fuzzy Inference System (ANFIS) for modelling a River Pollution Index (RPI) in the Tanzanian section of the Mara River reach. The 16-year dataset from 2010-2026, in 12 sampling locations, was analysed in the dry and wet seasons. Total suspended solids (TSS), dissolved oxygen (DO), five-day biological oxygen demand (BOD₅) and Ammonia Nitrogen (NH₃-N) were used to characterise WQ conditions and derive the RPI. Descriptive statistics, skewness, kurtosis, the Shapiro–Wilk test, and Q–Q plots were used to assess seasonal and distributional characteristics. The results showed that TSS had comparable mean concentrations during the dry 51.94±17.17 and wet 52.55±19.38 seasons, whereas mean DO and BOD₅ increased from 3.83±0.92 to 4.09±0.96 and from 9.84±7.77 to 13.07±5.86, respectively. NH₃-N showed a relatively small increase from 1.79±0.89 during the dry season to 1.87±0.84 during the wet season. All variables showed statistically significant departures from normality ($p<0.05$). The ANFIS model demonstrated a very good performance under both seasonal conditions, with slightly better performance during the dry season ($r=0.897$, $R^2=0.805$) than during the wet season ($r=0.888$, $R^2=0.789$). The findings demonstrate that ANFIS can effectively capture the dominant nonlinear relationships between key WQ variables and RP conditions, providing a useful framework for long-term river WQ assessment.

Keywords ANFIS, Ecosystem Maintenance, Transboundary Resources, Water Pollution, Water Quality Indices.

Introduction

River systems provide essential ecosystem services and support domestic water supply, agriculture, livestock production, fisheries, biodiversity, and broader socioeconomic development (Yalew et al., 2021). However, river water quality is increasingly affected by interactions among natural processes and anthropogenic pressures, including land-use change, agricultural activities, sediment mobilisation, wastewater discharge, and catchment disturbance (Hussain & Karuppannan, 2023; Nkwanda et al., 2021). These pressures can alter the physical, chemical, and biological characteristics of river systems and create substantial spatial and temporal variability in water-quality conditions (Lubanga et al., 2021).

Seasonality is a major driver of river water-quality dynamics, particularly in tropical and savanna catchments where wet and dry periods are associated with pronounced changes in rainfall, runoff, sediment transport, nutrient mobilisation, dilution, and in-stream biogeochemical processes (Das & Biswas, 2023; Feng et al., 2023). Evidence from Das & Biswas (2023) who investigated the seasonal water quality parameters of the River Mara Bharali, Assam, showed that physicochemical parameters changed in the monsoon and winter season. Likewise, Al-Sudani (2021) Performed a study on the Tigris River to assess seasonal variation in water quality by using the Overall Index of Pollution (OIP) and identified the cause of impairment of water quality. Also, Younis et al. (2024) Studied the River Nile and identified the stations which were most affected by river pollution.

African river systems, including the Mara River Basin, have demonstrated substantial seasonal variation in water-quality variables, with parameters such as total suspended solids (TSS) often increasing during wet periods because of enhanced runoff and sediment transport (Kimario et al., 2025; Mwangi et al., 2020; Ngonyani et al., 2025). Similar studies in Tanzanian rivers have reported pronounced seasonal differences in dissolved oxygen (DO), biochemical oxygen demand (BOD), and suspended materials, highlighting the importance of considering hydrological seasonality in water-quality assessment (Alphayo et al., 2026; William et al., 2025).

Among the parameters commonly used to characterise river water quality, TSS, DO, five-day biological oxygen demand (BOD₅), and Ammonia Nitrogen (NH₃-N) provide complementary information on sediment loading, oxygen conditions, organic pollution, and nutrient enrichment (Marselina et al., 2022; Sepahvand et al., 2025). Their

concentrations may respond differently to seasonal hydrological conditions and catchment processes (Čerkasova et al., 2018; Swilla et al., 2024). Previous investigations within the Lake Victoria Basin have also documented seasonal variability in nutrient concentrations, including Ammonia Nitrogen, in the Mara River and other Tanzanian water bodies (Kiyemba et al., 2023; Nakkazi et al., 2024). These findings emphasise the need for long-term and seasonally resolved assessments to adequately characterise river water-quality dynamics.

The Mara River is a major transboundary river shared between Tanzania and Kenya and is of considerable ecological and socioeconomic importance (Mwangi et al., 2020). Previous research in the broader Mara River Basin has shown that water-quality patterns can vary with season, land use, and catchment characteristics, reinforcing the importance of examining seasonal controls on river pollution dynamics. Taking into consideration the study conducted by Mwangi et al. (2020) of Modelling the impact of land use change on water quality in the Upper Mara River Basin and a study conducted by Lubanga et al. (2021) of modelling the impact of land use change on water quality in the Upper Mara River Basin, Kenya.

Although individual water-quality variables provide valuable information, the interpretation of multiple parameters simultaneously can be challenging. Composite water-quality or pollution indices provide a practical means of integrating several indicators into a single measure that represents overall water-quality conditions (Al-Bahathy et al., 2023; Alphayo et al., 2026; Marselina et al., 2022; Shakerdonyavi et al., 2023). Such approaches can facilitate temporal and spatial comparisons and support communication of complex environmental information for water-resource management.

Recent work on the Tanzanian Mara River, like Kimario et al. (2025) who used NSFQI has similarly demonstrated the usefulness of index-based approaches for identifying spatial and temporal water-quality variation. However, the relationships between individual water-quality variables and an integrated pollution condition are often nonlinear and may vary under contrasting seasonal regimes (Anh et al., 2023; Kang et al., 2024). Conventional statistical approaches may therefore have limited ability to represent the complex interactions among environmental variables. Artificial intelligence approaches provide an alternative framework for modelling nonlinear environmental systems (Al-Adhaileh & Alsaade, 2021; Rajaei et al., 2020). Among these approaches, the Adaptive Neuro-Fuzzy Inference System (ANFIS) combines the adaptive learning capability of artificial neural networks with the rule-based reasoning of fuzzy inference systems. This structure allows ANFIS to represent nonlinear relationships between multiple predictor variables and a response variable. Previous applications like study done by Al-Adhaileh & Alsaade (2021) wrote that the ANFIS model showed accuracy during the testing phase, with a regression coefficient of 96.17% for predicting WQI. Similarly, Vun et al. (2024) assessed the Semyenih River, Malaysia, using ANN (Artificial Neural Network), and the prediction model showed an RMSE and a coefficient of correlation value of 0.20658 and 0.78803, and recommended ANN for future studies. The literature in river water-quality modelling has demonstrated the potential of ANFIS for representing complex relationships among water-quality variables and for predicting WQ parameters.

Therefore, this study investigated seasonal water-quality characteristics and evaluated the performance of an Adaptive Neuro-Fuzzy Inference System (ANFIS) for modelling river pollution conditions in the Tanzanian section of the Mara River. The study used a 16 years dataset spanning from 2010–2026, incorporating secondary water-quality records and primary observations collected from 12 sampling locations. Four key water-quality parameters TSS, DO, BOD₅, and NH₃-N, were used to characterise water-quality conditions and to derive a River Pollution Index (RPI). The novelty of this study lies in integrating a long-term seasonal water-quality dataset with a composite River Pollution Index and season-specific ANFIS modelling. By evaluating the nonlinear relationships between TSS, DO, BOD₅, NH₃-N, and integrated river pollution conditions separately under contrasting seasonal regimes, the study provides a data-driven framework for understanding seasonal water-quality variability and supporting long-term river-water-quality assessment and management in the Mara River Basin.

Materials and Methods

Study area

This study was conducted in the Mara River Basin, Tanzanian section spanning roughly from 1°20' S to 1°52' S latitude and 33°47' E to 35°30' E longitude, respectively. The Mara River is an important transboundary river shared by Tanzania and Kenya. The Mara River system supports ecological, domestic, agricultural, and other socioeconomic activities within its catchment (Mwangi et al., 2020). The river originates in Kenya's Mau Forest Complex, flows through the Maasai Mara National Reserve and Serengeti National Park, and empties into Lake Victoria as shown in Fig. 1. The figure illustrates the study area in the East African region (Fig. 1a), the shared Mara River Basin (Fig. 1b) showing Tanzania and Kenya, and the specific focus of this research, the Lower Mara River catchment in Tanzania (Fig. 1c).

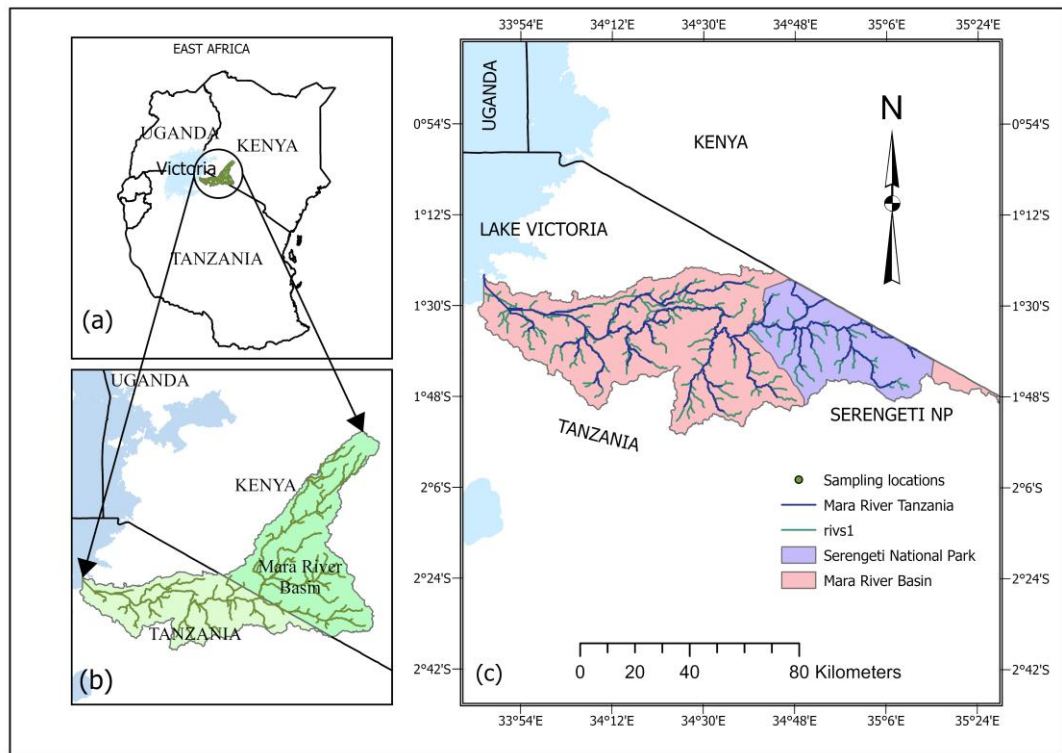


Fig. 1 The study area showing (a) East Africa Region map, (b) Mara River Basin, (c) Lower Mara River Basin

This lower catchment forms a vital ecological and hydrological zone that supports local biodiversity, wildlife habitats, tourism, agriculture, fisheries, and livestock keeping (Ngonyani et al., 2025). The region features a tropical savannah climate with a bimodal rainfall pattern, where long rains occur from March to May and short rains from October to December (Alphayo et al., 2026). Annual precipitation ranges between 800 mm and 1200 mm, and temperatures fluctuate between 20°C and 30°C based on elevation and season; these factors directly regulate river flow, sediment transport, and water quality. Topographically, the catchment consists of plains, river valleys, and wetlands (Kimario et al., 2025).

Land use and land cover (LULC)

The land cover is dominated by savannah grasslands that support wildlife and livestock, with agricultural activities concentrated near settlements and riverbanks, and wetlands distributed along the low-lying floodplains, as shown in Fig. 2. The land use/land cover (LULC) composition of the study area was classified into six major categories: forest, wetland, cropland, built-up areas, grassland, and water bodies. The LULC classification shows that grassland was 63.63% of the total study area, followed by cropland at 17.60%. Forest cover constituted 11.42%, while built-up areas and water bodies accounted for 2.88% and 2.59%, respectively.

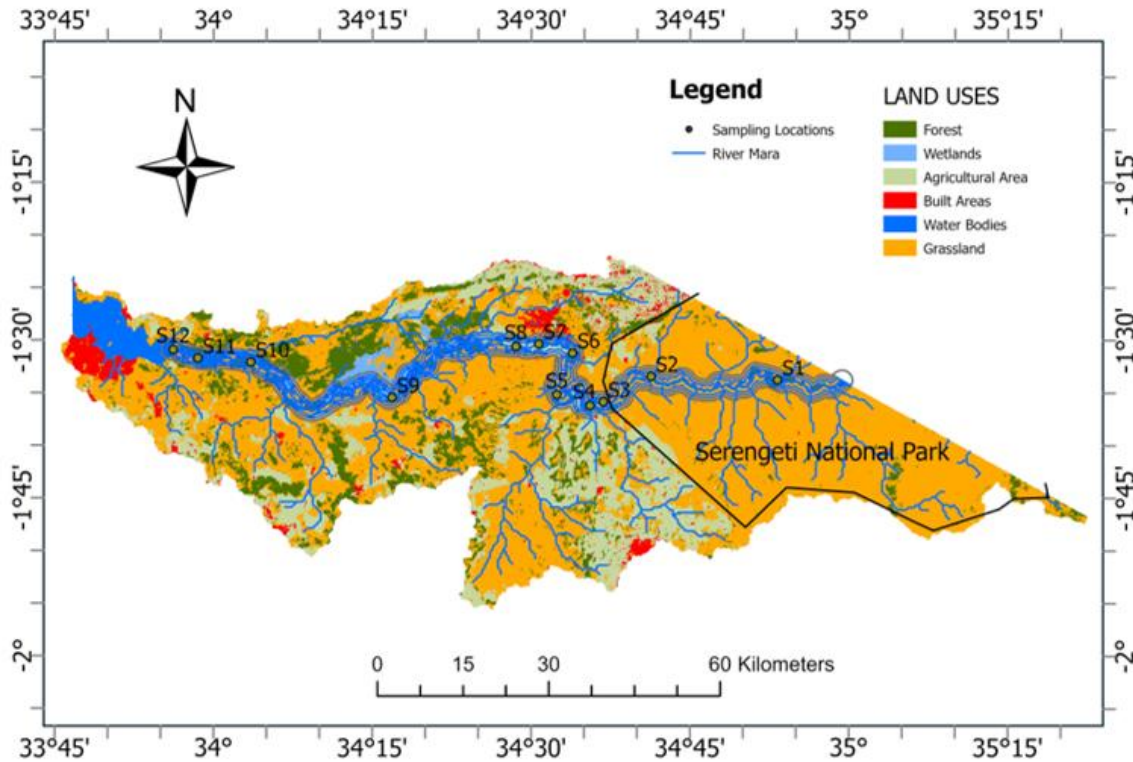


Fig. 2 Land use/land cover (LULC) classification of Mara River Catchment

Water-quality data collection and analysis

A total of twelve (12) sampling locations (Table 1) were included in the study to capture spatial variability in river water quality along the Tanzanian section of the Mara River. The study combined secondary and primary water-quality data covering 16 years (2010–2026). Secondary water-quality data for the period of 14 years (2010–2024) were obtained from historical data that were collected by the Mwanza and Musoma laboratories as part of their monitoring campaigns. Primary water-quality data were subsequently collected during 2025–2026 and analysed in the Musoma Water Quality Laboratory following standard methods recommended by APHA (American Public Health Association et al., 2017).

Table 1: River Mara Sampling Locations

ID	Easting (x)	Northing (y)	Sampling Locations Name
S1	709982	9827169	Kogatende
S2	687833	9827770	Sweto/Swetu
S3	679464	9823325	Tobora
S4	677144	9822694	Nyansurura Bridge
S5	671340	9824589	Somoche
S6	674110	9831897	Mrito
S7	668199	9833498	Mara Mine Before Discharge
S8	664157	9833035	Mara Mine After Discharge
S9	642496	9824117	Tigite/Magatini
S10	617749	9830319	Nyandoto/Sibora
S11	608481	9831006	Kirumi Bridge
S12	604192	9832541	Mara Mouth

Mwanza Zonal Water Quality Laboratory is accredited to ISO/IEC 17025:2017, with facility accreditation number TEST-5 0011, and Musoma Water Quality Laboratory is accredited to ISO/IEC 17025:2017 with facility accreditation number TEST-5 0055. Four water-quality parameters considered in this study are Total suspended solids (TSS), Dissolved oxygen (DO), Five-day biochemical oxygen demand (BOD₅) and Ammonia Nitrogen (NH₃-N). Water-quality sampling was conducted four times per year, comprising two sampling campaigns during the wet season and two during the dry season. This seasonal sampling design was adopted to capture variations in river water quality under contrasting hydrological conditions. The final dataset was classified into dry and wet season observations for subsequent descriptive, statistical, and modelling analyses.

Assessment of distributional characteristics and normality

The distributional characteristics of the water-quality variables were evaluated separately for the dry and wet seasons using skewness and kurtosis statistics (Dong, 2023). These measures provided an initial assessment of distribution symmetry and tail characteristics. Data normality was subsequently assessed using the Shapiro–Wilk test at a significance level of 5% ($\alpha = 0.05$). The null hypothesis assumed that the data were drawn from a normally distributed population. Normality was rejected when the associated p-value was less than 0.05. Normal Q–Q plots were additionally used to visually assess the agreement between the observed distributions and the theoretical normal distribution (Harbison & Simmons, 2024). Departures of the observed quantiles from the theoretical reference line were interpreted as evidence of deviations from normality. The statistical and graphical assessments were interpreted jointly because the relatively large sample size could increase the sensitivity of formal normality tests to minor deviations from a normal distribution.

River Pollution Index development

The four water-quality parameters, namely TSS, DO, BOD₅, and NH₃-N, were integrated to generate a composite River Pollution Index (RPI) representing the overall water-quality condition (Ati et al., 2022). The RPI uses four water-quality variables to assess river pollution. These variables are Dissolved Oxygen (DO), Biochemical Oxygen Demand (BOD₅), Total Suspended Solids (SS), and Ammonia Nitrogen (NH₃-N), which were converted to four-state quality sub-index scores (Table 2). The overall index obtained is subdivided into four pollution levels as non-polluted, lightly-polluted, moderately-polluted, and grossly-polluted (Yavuz et al., 2024).

Table 2: RPI for water quality parameters (Ati et al., 2022)

Parameters	Good	Slightly polluted	Moderate polluted	Gross polluted
DO (mg/L)	>6.5	4.6-6.5	2.0-4.5	<2.0
BOD ₅ (mg/L)	<3	3-4.9	5-15	>15
TSS (mg/L)	<20	20-49	50-100	>100
NH ₃ -N (mg/L)	<0.5	0.5-0.99	1.0-3.0	>3.0
Index Scores (Si)	1	3	6	10
Sub-index	<2	2.0-3.0	3.1-6.0	>6.0

The RPI was used as the target variable for the ANFIS model, while TSS, DO, BOD₅, and NH₃-N served as the model input variables. This approach enabled the model to capture the combined influence of multiple water-quality parameters on the overall river pollution condition.

ANFIS model development

An Adaptive Neuro-Fuzzy Inference System (ANFIS) was employed to model the nonlinear relationship between the selected water-quality variables and the River Pollution Index. The ANFIS model was developed using the Sugeno fuzzy inference structure, which combines fuzzy rule-based reasoning with the adaptive learning capabilities of artificial neural networks (Fig. 3.)

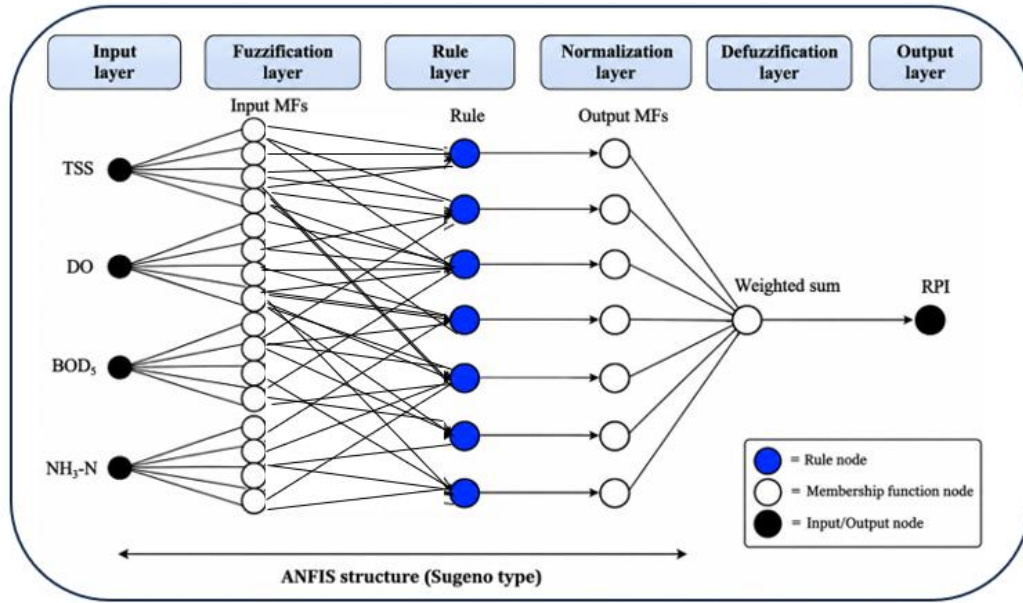


Fig. 3 ANFIS Sugeno-type structure used in this study

The general rule structure of a first-order TSK model can be expressed as follows (Elsabagh et al., 2024; Elsheikh et al., 2025; Khan et al., 2023; Machesa et al., 2023):

i) Rule Structure of First-Order TSK Model

A first-order TSK fuzzy system consists of fuzzy if-then rules, where the consequent part is a linear function of input variables. For a two-input system (x and y), the fuzzy rules take the form:

Rule i : If x is A_i and y is B_i then

$$f_i = p_i x + q_i y + r_i \dots \dots \dots (Eqn 2.1)$$

Where:

- A_i and B_i are fuzzy membership functions (MFs) for inputs x and y

- pi , qi , and ri Are the consequent parameters that will be optimised using the ANFIS learning algorithm.
- fi is the output of the rule.

ii) ANFIS Model Equations for First-Order TSK System

ANFIS has a five-layer architecture, and the mathematical representation for each layer is as follows:

Layer 1: Input Membership Function (Fuzzification Layer)

Each input is passed through fuzzy membership functions to determine its degree of belonging:

$$\mu_{Ai}(x) = \frac{1}{1 + (\frac{x - ci}{ai})^{2bi}} \quad \dots \dots \dots (Eqn 2.2)$$

$$\mu_{Bi}(y) = \frac{1}{1 + (\frac{y - di}{ei})^{2fi}} \quad \dots \dots \dots (Eqn 2.3)$$

Where ai, bi, ci, di, ei and fi are premise parameters that define the shape of the membership functions (typically Gaussian, Bell-shaped, or Triangular).

Layer 2: Rule Strength (Firing Strength Calculation)

Each node represents a fuzzy rule and applies a T-norm operator (e.g., multiplication) to calculate the firing strength of each rule:

$$wi = \mu_{Ai}(x) \times \mu_{Bi}(y) \quad \dots \dots \dots (Eqn 2.4)$$

Layer 3: Normalisation of Firing Strengths

Each rule's firing strength is normalised by dividing by the sum of all rule strengths:

$$\bar{wi} = \frac{wi}{\sum wi} \quad \dots \dots \dots (Eqn 2.5)$$

Where $\bar{wi}$ represents the normalised activation of each rule.

Layer 4: Defuzzification (Weighted Rule Output Computation)

Each rule generates an output, and the defuzzified output is calculated using a linear function:

$$\bar{wif_i} = \bar{wi}(pix + qiy + ri) \quad \dots \dots \dots (Eqn 2.6)$$

where pi, qi and ri are the consequent parameters optimised during training.

Layer 5: Overall Output (Aggregation Layer)

The final crisp output is computed as the weighted sum of all rule outputs:

$$f = \sum \bar{wif_i} = \sum \bar{wi}(pix + qiy + ri) \quad \dots \dots \dots (Eqn 2.7)$$

An adaptive neural network structure presented in Figure 2.1 was functionally similar to a fuzzy inference system. It was observed that the values of premise parameters and the overall output ' f ' were a linear combination of the consequent parameters (Shende & Sahoo, 2024).

The output ' f ' can be formulated as shown in Fig. 4 by the two passes in the hybrid learning algorithm of ANFIS.

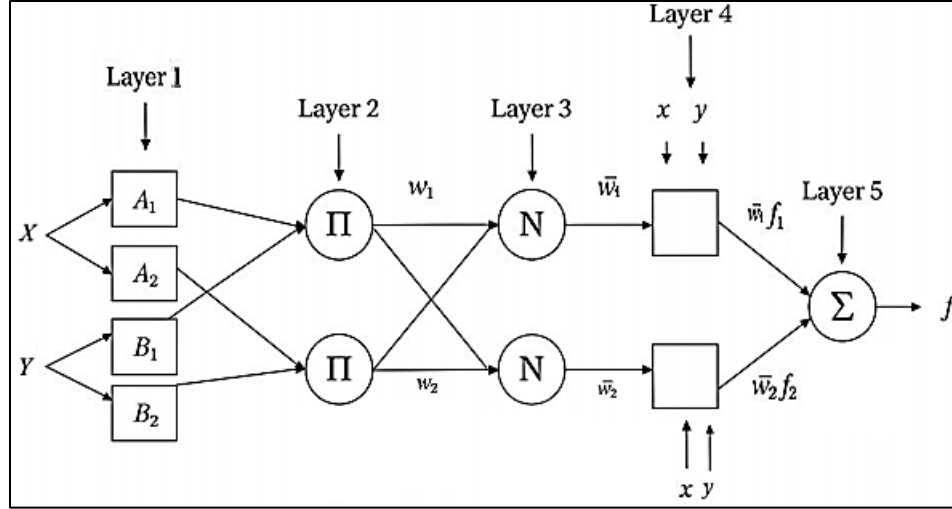


Fig. 4 A typical architecture of the ANFIS system (Shende & Sahoo, 2024).

Separate analyses were conducted for dry and wet season data to evaluate the ability of ANFIS to represent seasonal differences in river water-quality conditions. The seasonal approach allowed the model performance to be compared under contrasting hydrological conditions. The seasonal datasets were divided into training and testing subsets. A total of 70% of the observations were used for model training, while the remaining 30% were used for model testing (Khan et al., 2023). The training dataset was used to adjust the parameters of the ANFIS model and establish the relationship between the four water-quality parameters and the RPI. The independent testing dataset was subsequently used to evaluate the predictive performance of the developed model.

The predictive model was developed as a Sugeno-type Adaptive Neuro-Fuzzy Inference System (ANFIS) implemented in MATLAB R2024b through the ANFIS graphical user interface (GUI). The architecture consists of 4 input variables and 1 output variable (Fig. 3). The model learning process involved the adjustment of the fuzzy inference system parameters to reduce the difference between the observed and predicted RPI values. Initial fuzzy inference rules and input membership functions were automatically generated using subtractive clustering with a range of influence of 0.3, a squash factor of 1.25, an acceptance ratio of 0.5, and a rejection ratio of 0.15. This clustering process yielded a total of 28 fuzzy rules and a corresponding 28 Gaussian membership functions per input variable. The network parameters were optimised using the hybrid learning algorithm, combining the Least-Squares Estimator with Backpropagation Gradient Descent over 100 training epochs to minimise root-mean-square error.

The predictive performance of the ANFIS model was evaluated separately for the dry and wet seasons using the following statistical indicators; Pearson correlation coefficient (r), Coefficient of determination (R^2), Adjusted coefficient of determination (adjusted R^2) and Residual sum of squares (RSS). The Pearson correlation coefficient was used to evaluate the strength of association between observed and predicted RPI values. The coefficient of determination quantified the proportion of variability in the observed RPI explained by the ANFIS model. The adjusted R^2 was used as an additional measure of explanatory performance after accounting for model complexity. Higher values of r , R^2 , and adjusted R^2 , together with lower residual variation, were interpreted as indicators of improved model performance.

Model performance was further evaluated graphically using fitted curves and residual plots for each season. The fitted curves were used to visually compare observed and predicted RPI values and assess the ability of the ANFIS model to reproduce the dominant patterns and variability in river pollution conditions. Residual plots were examined to identify systematic prediction errors, trends, heterogeneity in residual variance, and potential outliers. A random distribution of residuals around the zero-reference line was considered indicative of limited systematic bias. In contrast, systematic patterns or increasing residual dispersion were interpreted as evidence of model limitations or unaccounted variability.

Results and Discussion

Descriptive statistics of water quality parameters

Descriptive statistics revealed distinct seasonal patterns and distributional characteristics across the water-quality parameters (Table 3). TSS showed comparable mean concentrations in the dry (51.94 ± 17.17) and wet (52.55 ± 19.38) seasons, although variability and positive skewness were greater in the wet season. DO increased modestly from 3.83 ± 0.92 in the dry season to 4.09 ± 0.96 in the wet season and exhibited near-symmetric distributions in both periods. In contrast, BOD₅ was substantially higher in the wet season (13.07 ± 5.86) than in the dry season (9.84 ± 7.77), while the dry-season distribution showed pronounced positive skewness and heavier tails. NH₃-N concentrations differed only marginally between seasons, increasing from 1.79 ± 0.89 in the dry season to 1.87 ± 0.84 in the wet season; however, both seasonal distributions were positively skewed and leptokurtic, indicating the influence of relatively high observations. These results indicate that seasonal differences were more evident in the variability and distributional characteristics of the water-quality parameters than in their mean concentrations, except for BOD₅. Table 3 summarised the descriptive statistics of the river.

Table 3: Descriptive Statistics for Mara River Water Quality Parameters

Parameter	Season	N	Minimum	Maximum	Mean±SD	Skewness	Kurtosis
TSS	Dry	312	15.06	102.83	51.94±17.17	0.39524	-0.22186
	Wet	300	19.24	113.56	52.55±19.38	0.95527	0.32491
DO	Dry	312	1.59	5.91	3.83±0.92	0.00677	-0.59038
	Wet	300	1.48	6.25	4.09±0.96	-0.22595	-0.55901
BOD ₅	Dry	312	0.06	38.46	9.84±7.77	1.14551	1.03977
	Wet	300	1.06	28.69	13.07±5.86	0.40801	-0.0744
NH ₃ -N	Dry	312	0.265	4.57	1.79±0.89	1.18142	1.0637
	Wet	300	0.364	5.276	1.87±0.84	1.1061	1.32031

Box plots in Fig. 5 of the water-quality parameters revealed distinct differences in their distributions and variability between the dry and wet seasons. TSS (Fig. 5a) exhibited a relatively wider spread during the wet season, indicating greater variability, with higher upper values and evidence of right-skewness compared with the dry season. DO (Fig. 5b) showed comparatively compact distributions in both seasons, with a modest upward shift in the wet-season central values, indicating generally higher dissolved oxygen concentrations. BOD₅ (Fig. 5c) displayed substantial variability, particularly during the dry season, with a wider spread and potential high-value observations consistent with its positive skewness. In contrast, wet-season BOD₅ values were more concentrated around the central distribution despite having a higher mean. NO₃ showed positive skewness in both seasons, with the upper tails extending towards higher concentrations, although its overall spread was relatively similar between seasons.

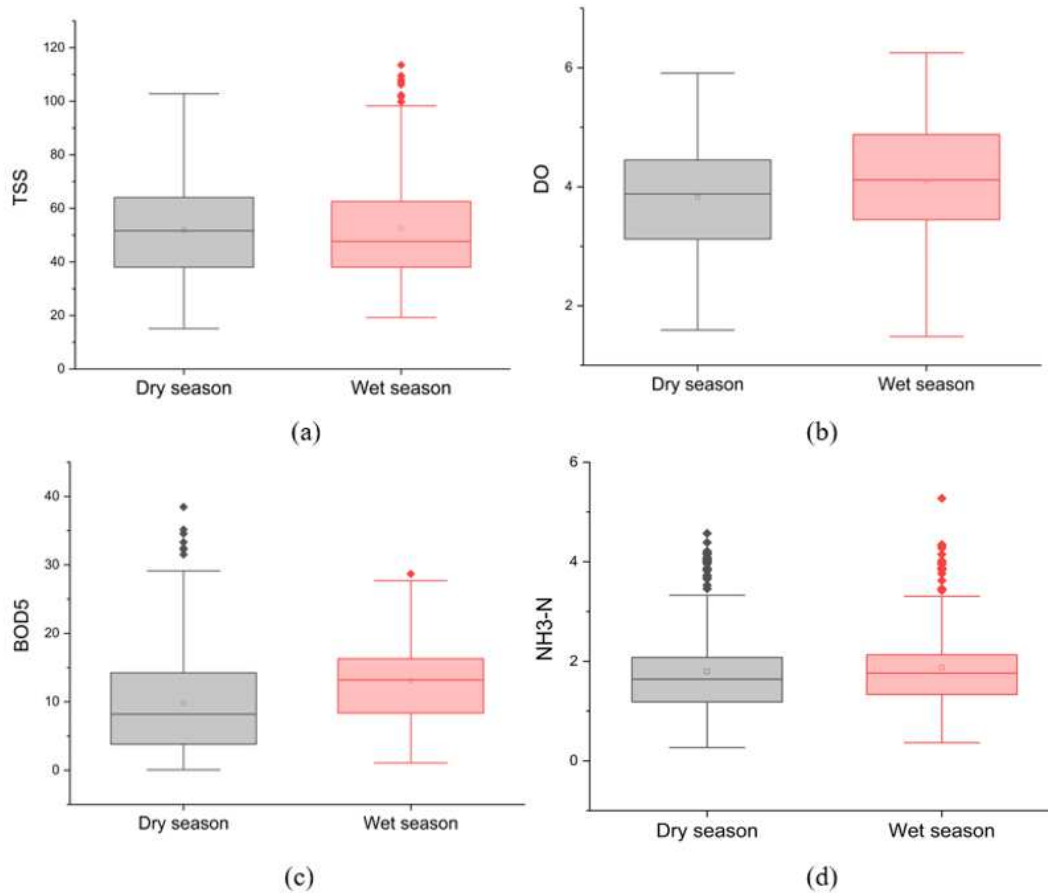


Fig. 5 Seasonal box plots showing the distribution of (a) TSS, (b) DO, (c) BOD₅, and (d) NH₃-N concentrations during the dry and wet seasons.

Data normality and distributional characteristics

The Shapiro–Wilk test indicated statistically significant departures from normality for all water-quality parameters in both seasons (Table 4). TSS deviated significantly from normality during both the dry season ($W = 0.982$, $p = 0.0006$) and wet season ($W = 0.922$, $p < 0.0001$). Similarly, DO showed statistically significant, although comparatively smaller, departures from normality in the dry ($W = 0.991$, $p = 0.041$) and wet ($W = 0.987$, $p = 0.008$) seasons. More pronounced departures were observed for BOD₅ and NH₃-N, particularly during the dry season, with W values of 0.903 and 0.895, respectively (both $p < 0.0001$).

Table 4: Shapiro–Wilk test showing significant departures from normality

Parameter	Season	n	Shapiro–Wilk W	p-value	Decision ($\alpha = 0.05$)
TSS	Dry season	312	0.98202	0.0000596	Reject normality
	Wet season	300	0.92204	<0.0001	Reject normality
DO	Dry season	312	0.9905	0.04102	Reject normality
	Wet season	300	0.98697	0.00829	Reject normality
BOD5	Dry season	312	0.90308	<0.0001	Reject normality
	Wet season	300	0.9808	0.000472	Reject normality
NO3	Dry season	312	0.89508	<0.0001	Reject normality
	Wet season	300	0.92005	<0.0001	Reject normality

The normal Q–Q plots showed varying degrees of departure from the theoretical normal distribution among the four water-quality parameters and between seasons (Fig. 6&7). TSS exhibited noticeable deviations from the reference line, particularly at the lower and upper tails. The deviation was more pronounced during the wet season (Fig. 6b), where the points showed greater curvature and departure at the extremes, indicating a positively skewed, non-normal distribution. The dry-season TSS (Fig. 6a) distribution was comparatively closer to the reference line but still showed deviations at the lower and upper quantiles.

For DO (Fig. 6 c&d), the Q–Q plots showed the closest agreement with the normal reference line among the investigated parameters. In both the dry and wet seasons, most observations were distributed relatively close to the reference line, particularly across the central portion of the distribution. However, deviations at the lower and upper tails were evident, indicating that the DO data were only approximately, rather than strictly, normally distributed. This observation is consistent with the relatively small skewness values obtained for DO.

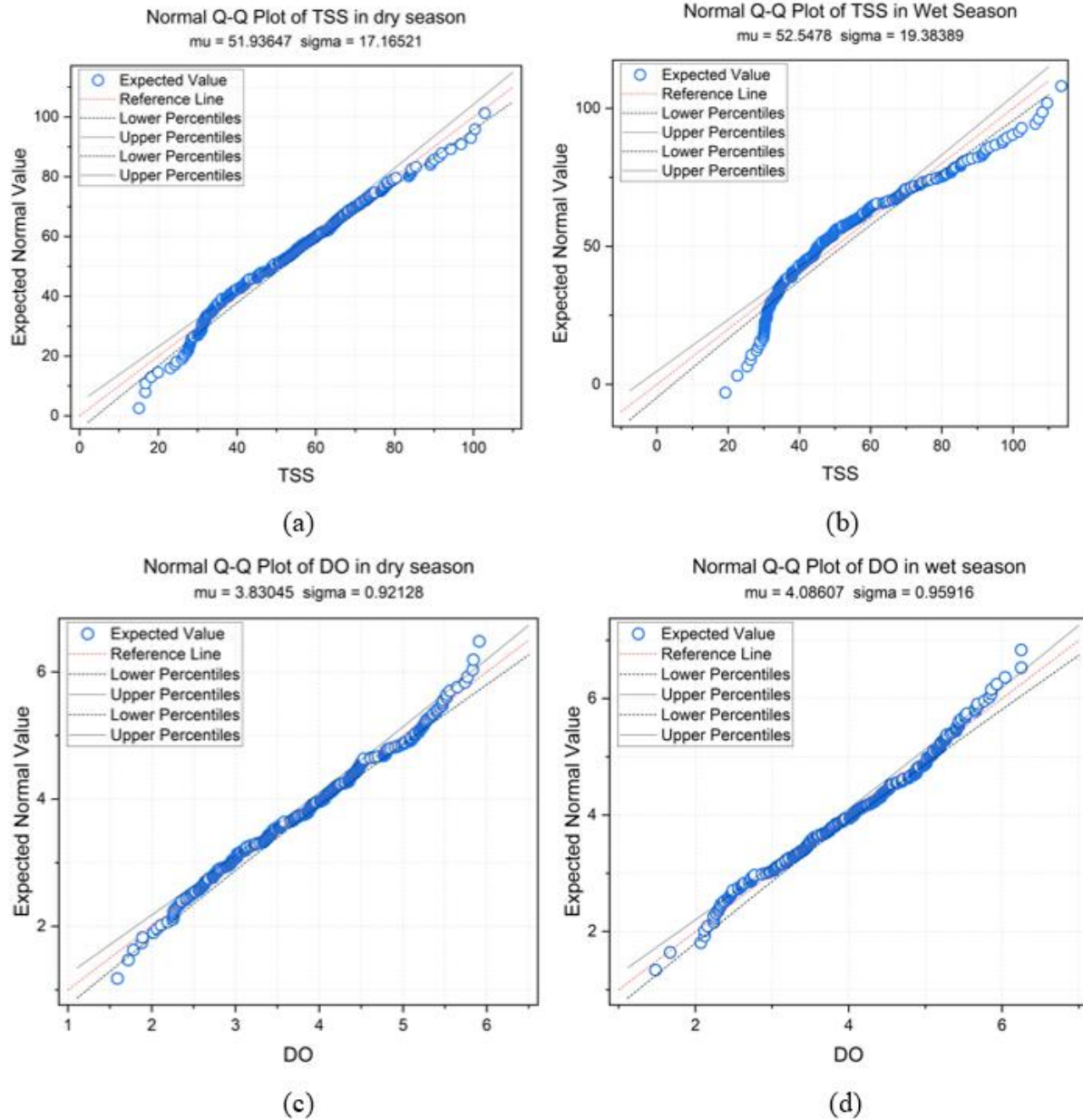


Fig. 6 The Quantile-Quantile (Q-Q) plots for TSS (a&b) and DO (c&d) in dry and wet season respectively

The Q-Q plots for BOD₅ (Fig. 7 e&f) revealed more substantial deviations from normality, particularly during the dry season. The pronounced curvature of the points, especially at the lower quantiles and upper tail, indicates asymmetry and the presence of extreme observations, consistent with the relatively high positive skewness and kurtosis. During the wet season, the BOD₅ observations showed comparatively closer alignment with the reference line across the central distribution; nevertheless, departures at the distributional tails remained evident, confirming that the data did not fully satisfy the normality assumption.

Similarly, NH₃-N (Fig. 7. g&h) showed systematic deviations from the reference line in both seasons. The curvature observed in the lower and upper portions of the Q-Q plots indicate positive skewness and departures from a normal distribution. The deviations were evident in both dry- and wet-season datasets, suggesting that elevated nitrate

observations contributed to extended upper tails. This pattern agrees with the positive skewness and kurtosis values and the significant Shapiro–Wilk test results.

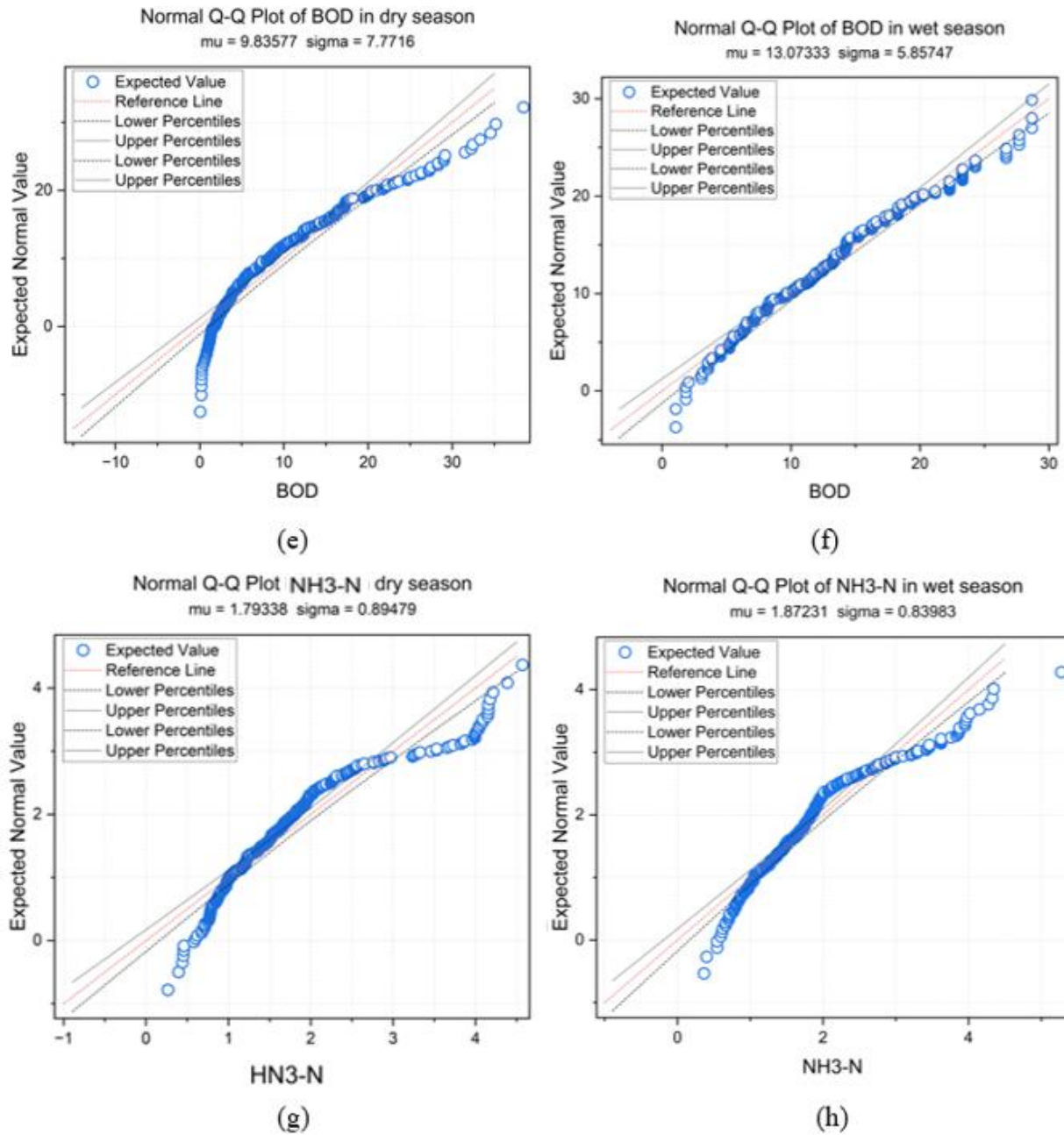


Fig. 7 The Quantile-Quantile (Q–Q) plots for BOD₅ (e&f) and NH₃-N (g&h) in dry and wet seasons, respectively

Relationship between water-quality parameters and RPI

Figure 8 illustrates the spatial variation in RPI values across the sampling locations during the dry and wet seasons. The boxplots indicate differences in the distribution and variability of RPI among locations, reflecting spatial heterogeneity in water-quality conditions. Seasonal differences in RPI are also evident, suggesting that hydrological conditions influence pollution levels and their distribution along the sampling locations.

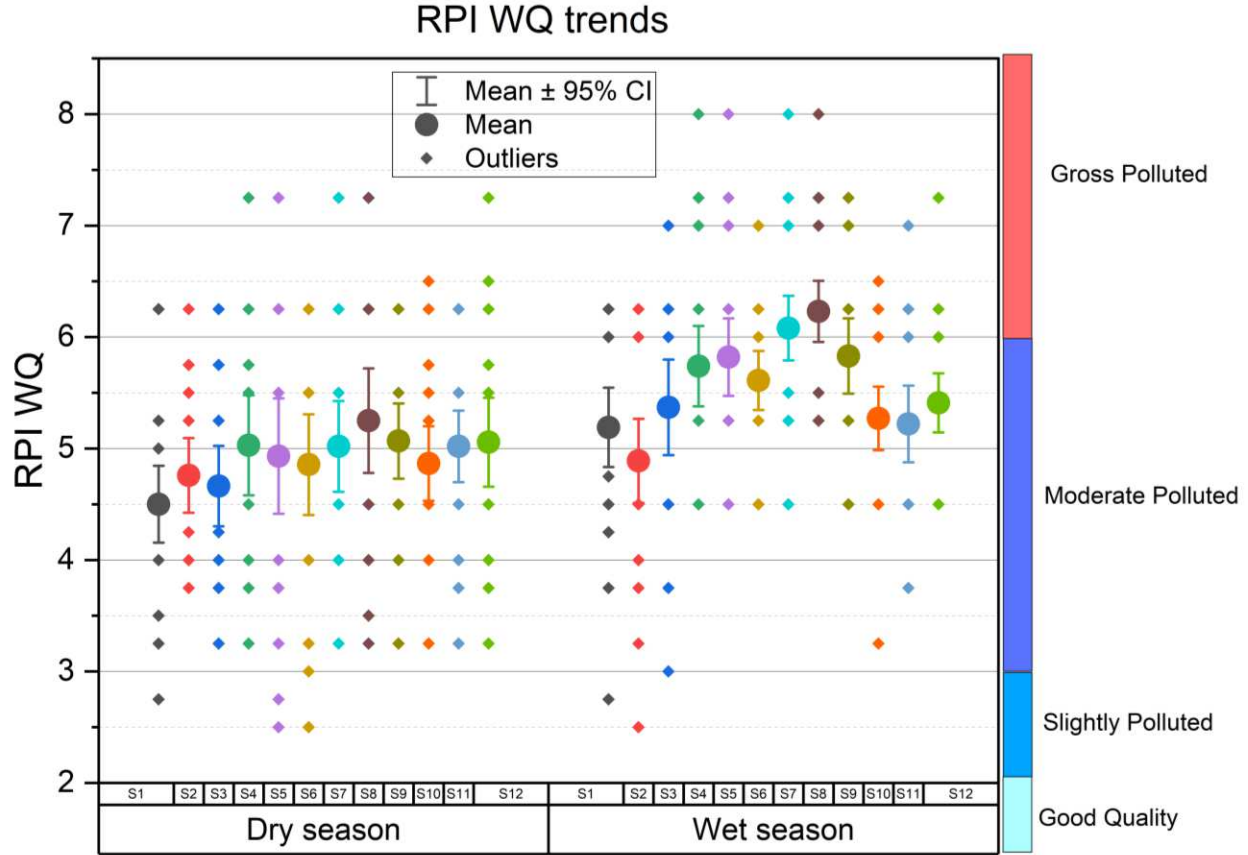


Fig. 8 Spatial variation in RPI values in the sampling locations during the dry and wet seasons

Spearman's rank correlation analysis was conducted to evaluate the relationships between individual water-quality parameters and the River Pollution Index (RPI) during the dry and wet seasons (Table 5). The results revealed substantial seasonal differences in the association between water-quality parameters and RPI. During the dry season, BOD exhibited the strongest positive correlation with RPI ($r_s = 0.802$, $p < 0.0001$), indicating that organic pollution was strongly associated with increasing RPI values. $\text{NH}_3\text{-N}$ also showed a significant positive correlation with RPI ($r_s = 0.376$, $p < 0.0001$), although the magnitude of this relationship was moderate. In contrast, DO was significantly negatively correlated with RPI ($r_s = -0.196$, $p = 0.00051$), suggesting that increasing pollution levels were associated with reduced dissolved oxygen concentrations. TSS showed no significant relationship with RPI during the dry season ($r_s = 0.024$, $p = 0.672$).

During the wet season, BOD remained the parameter most strongly associated with RPI ($r_s = 0.614$, $p < 0.0001$), although the magnitude of the correlation was lower than that observed during the dry season. TSS exhibited a moderate positive correlation with RPI ($r_s = 0.385$, $p < 0.0001$), whereas $\text{NH}_3\text{-N}$ showed a weaker but statistically significant positive relationship ($r_s = 0.255$, $p < 0.0001$). DO remained negatively correlated with RPI ($r_s = -0.153$, $p = 0.008$), although the relationship was relatively weak.

Table 5: Spearman's rank correlation of RPI with respect to individual water quality parameters

Parameter	Dry season Spearman r_s	p-value	Wet season Spearman r_s	p-value
TSS	0.25	<0.0001	0.492	<0.0001
DO	-0.1889	0.00079	-0.175	0.00234
BOD	0.73245	<0.0001	0.624	<0.0001
$\text{NH}_3\text{-N}$	0.37933	<0.0001	0.2399	<0.0001

Seasonal performance of the ANFIS model

The ANFIS model demonstrated good predictive performance in both the dry and wet seasons, although its performance was slightly stronger during the dry season; see Table 6. For the dry-season dataset ($n = 312$), the model achieved a Pearson correlation coefficient of $r = 0.897$ and an R^2 of 0.805, indicating a strong association between the observed and predicted values and suggesting that approximately 80.5% of the variability in the modelled water-quality parameter was explained by the ANFIS model. The adjusted R^2 (0.804) was nearly identical to the R^2 , indicating that the explanatory performance remained stable after adjustment for model complexity. The residual sum of squares (RSS) was 52.8, reflecting the magnitude of unexplained variation in the dry-season predictions.

Table 6: ANFIS predictive performance in both the dry and wet seasons

Description	ANFIS in Dry season	ANFIS in Wet season
Number of Points	312	300
Degrees of Freedom	310	298
Residual Sum of Squares	52.79839	65.80349
Pearson's r	0.8972	0.88834
R-Square (COD)	0.80497	0.78914
Adj. R-Square	0.80435	0.78844

During the wet season ($n = 300$), the ANFIS model also demonstrated strong performance, with a Pearson correlation coefficient of $r = 0.888$ and an R^2 of 0.789. Thus, the model explained approximately 78.9% of the variability in the observed data. The adjusted R^2 of 0.788 was similarly close to the corresponding R^2 value, suggesting stable model performance. However, the wet-season RSS (65.8) was higher than that obtained during the dry season, indicating greater residual variation in the model predictions. The results indicate that the ANFIS model captured a substantial proportion of the seasonal variability under both hydrological conditions, with marginally better predictive performance during the dry season.

The fitted curve (Fig. 9) for the dry season showed good agreement between the observed and ANFIS-predicted values, indicating that the model successfully reproduced the dominant variation in the response variable. The predicted values generally followed the temporal pattern of the observations, although deviations were evident for some individual observations and extreme values. This visual agreement was consistent with the quantitative performance metrics, with the model explaining 80.5% of the observed variability ($R^2 = 0.805$).

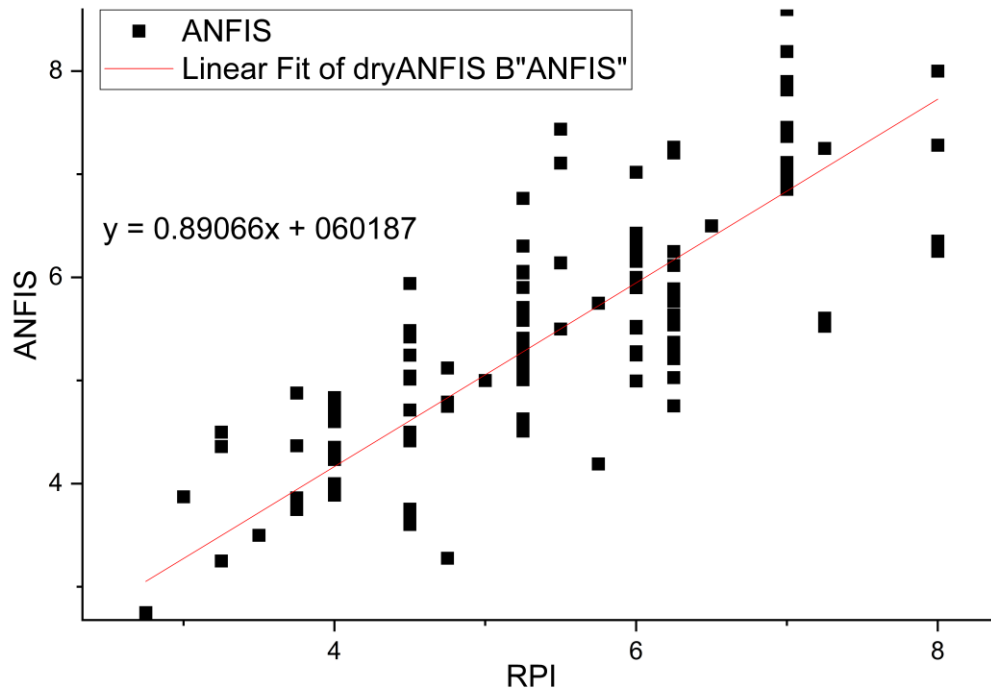


Fig. 9 ANFIS fitted curve for the dry season

The dry-season residuals plot (Fig. 10) was generally distributed around the zero-reference line without an apparent systematic pattern, indicating that the ANFIS model did not exhibit substantial systematic overprediction or underprediction. The relatively consistent dispersion of the residuals further suggests that the model captured the principal structure of the dry-season data.

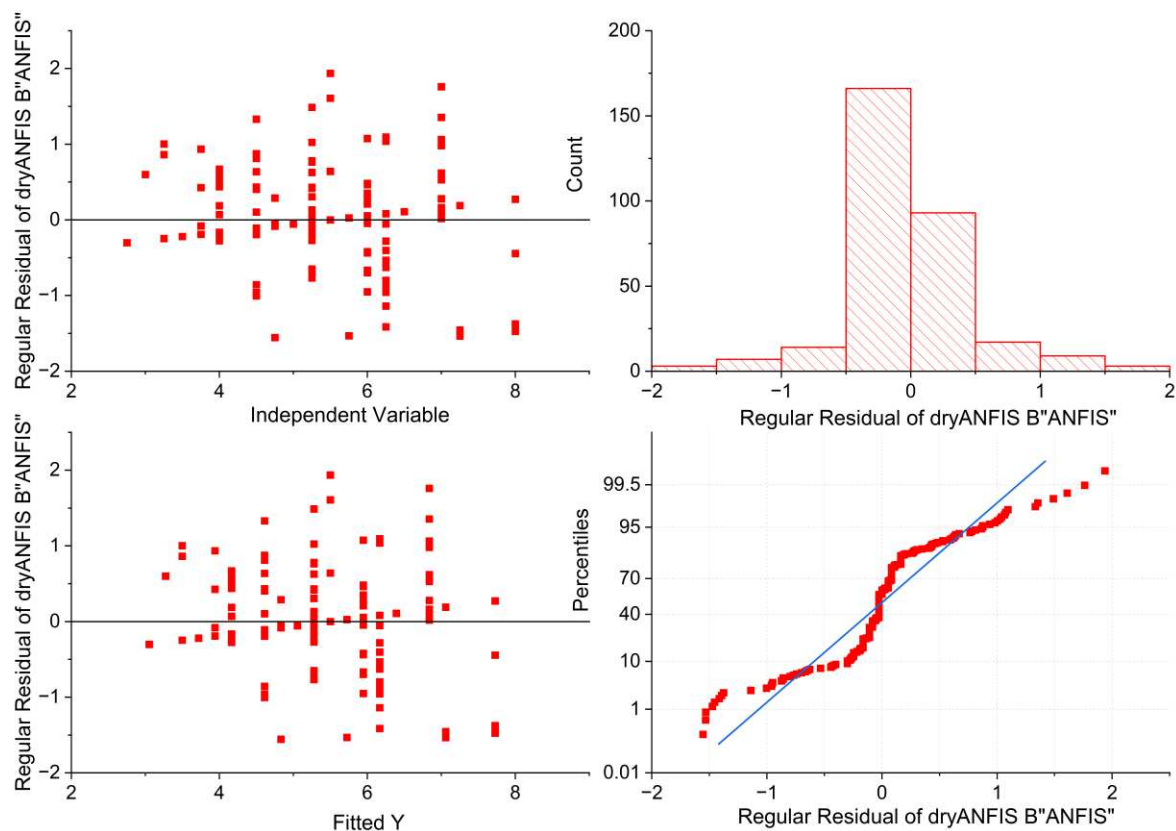


Fig. 10 ANFIS dry-season residuals plot

During the wet season, the fitted curve (Fig. 11) also showed substantial agreement between observed and predicted values; however, the deviations between the two series were comparatively greater than those observed during the dry season. The ANFIS model captured the general pattern of variation but showed reduced accuracy for some fluctuations or extreme observations. This observation is consistent with the slightly lower coefficient of determination ($R^2 = 0.789$), indicating that approximately 78.9% of the observed variability was explained by the model.

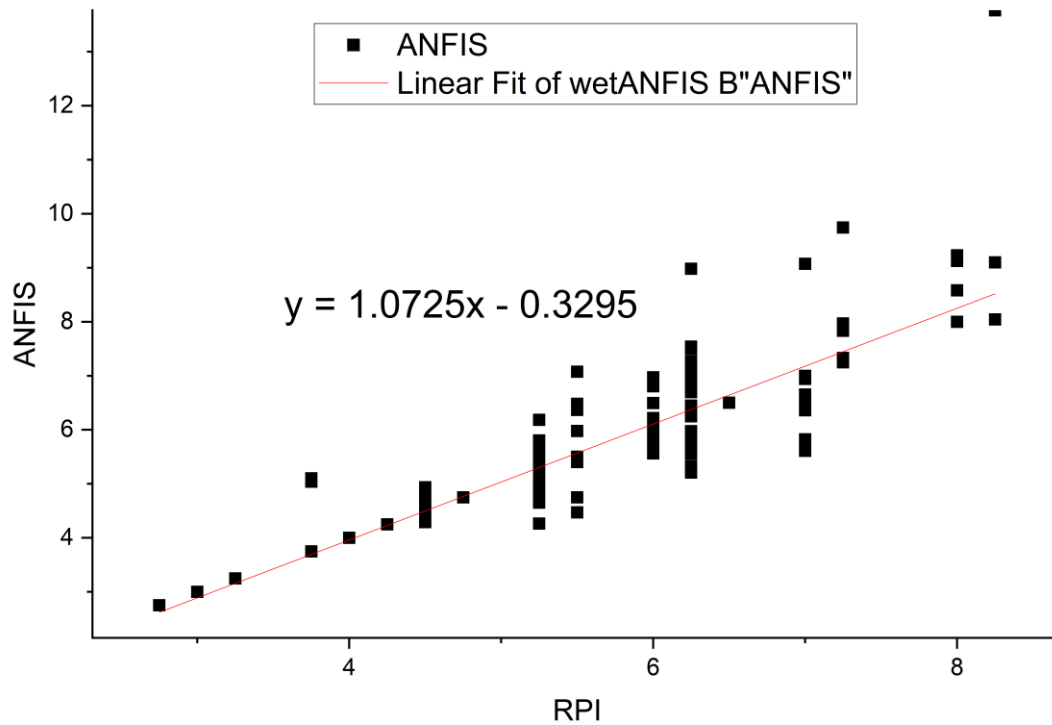


Fig. 11 ANFIS wet-season residual plot

The wet-season residual plot (Fig. 12) showed a wider dispersion around the zero-reference line compared with the dry season, indicating greater unexplained variability in the predictions. Nevertheless, the absence of a strong systematic pattern would suggest that the prediction errors were predominantly random rather than indicative of substantial model bias.

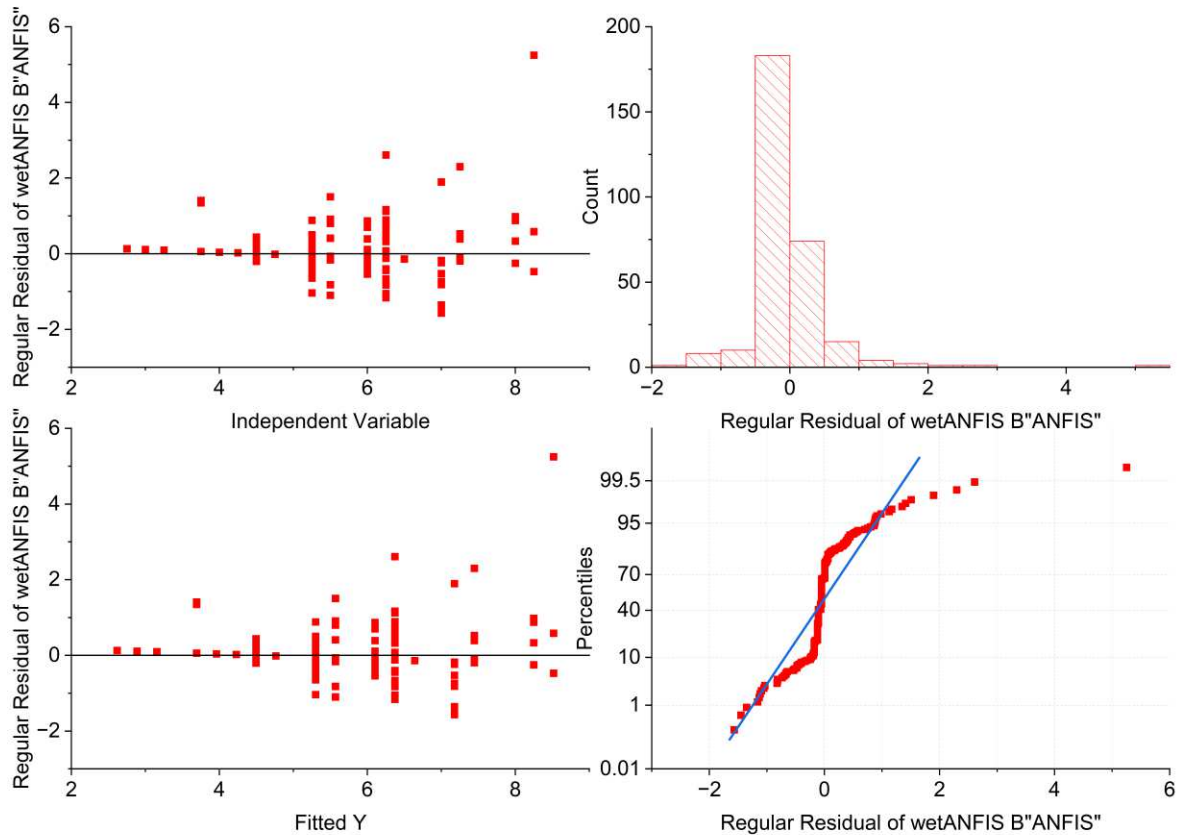


Fig. 12 ANFIS wet-season residuals plot

Discussion

The results demonstrate clear seasonal variability in the river water-quality characteristics and confirm the capability of the ANFIS model to reproduce the observed response under both dry- and wet-season conditions. Descriptive analysis showed that TSS concentrations were broadly comparable between seasons, whereas DO and BOD₅ were higher during the wet season. In contrast, NH₃-N exhibited only a modest seasonal increase. Seasonal differences were also reflected in data variability and distributional characteristics, with several parameters showing positive skewness and evidence of occasional elevated observations.

The distributional analysis further demonstrated that the water-quality variables did not strictly conform to a normal distribution. The Shapiro–Wilk test rejected the assumption of normality for all parameters in both seasons, although the magnitude of departure varied among variables. The Q–Q plots and skewness statistics indicated that DO was comparatively closer to a normal distribution, whereas BOD₅, particularly during the dry season, and NH₃-N in both seasons showed more pronounced departures from normality. These results highlight the heterogeneous and potentially nonlinear characteristics of river water-quality data and provide a rationale for applying flexible nonlinear modelling approaches such as ANFIS.

The ANFIS model achieved good predictive performance under both seasonal conditions. During the dry season, the model produced a Pearson correlation coefficient of 0.897 and an R² of 0.805, indicating that approximately 80.5% of the observed variability was explained by the model. Performance remained similarly strong during the wet season, with $r = 0.888$ and $R^2 = 0.789$, corresponding to 78.9% of the explained variability. The close agreement between R² and adjusted R² in both seasons indicates stable explanatory performance. However, the slightly higher explanatory power and lower residual sum of squares during the dry season suggest that the model achieved marginally better agreement with the observed data under dry-season conditions.

Collectively, these findings support the potential of ANFIS as a robust nonlinear modelling framework for seasonal river water-quality prediction, while also indicating that a proportion of the observed variability remains unexplained. Future work could improve predictive performance by incorporating additional hydrological, meteorological, or catchment-related predictors and by evaluating model generalisation using independent validation data and complementary error metrics.

Study limitations

Despite the useful insights provided by the long-term dataset and the application of season-specific ANFIS modelling, several limitations should be acknowledged. First, the study relied on four water-quality parameters (TSS, DO, BOD₅, and NH₃-N, depending on the final parameter set), which may not fully represent all physical, chemical, biological, and microbiological processes influencing river pollution. Additional variables, including streamflow, rainfall, temperature, pH, electrical conductivity, turbidity and microbial indicators, could improve understanding of the factors governing seasonal water-quality variability and potentially enhance model performance. Second, the long-term dataset combined secondary records with more recent primary measurements; therefore, differences in sampling procedures, analytical instruments, detection limits, and laboratory quality-control practices over time may have introduced additional uncertainty. Third, sampling was conducted four times per year, which may not capture short-term pollution events and rapid hydrological changes, particularly during intense rainfall and runoff periods. Finally, although the ANFIS model demonstrated good performance using a 70% and 30% training and testing partition, its generalizability should be further evaluated using independent external datasets and, where possible, additional validation approaches. Therefore, future studies should incorporate higher-frequency monitoring, a broader set of environmental predictors, and independent spatiotemporal validation to strengthen the robustness and transferability of the proposed modelling framework.

Conclusion

This study investigated the seasonal characteristics of river water quality and evaluated the performance of an Adaptive Neuro-Fuzzy Inference System (ANFIS) for modelling the observed water-quality response under dry and wet season conditions. The descriptive analysis demonstrated seasonal differences in the magnitude, variability, and distributional characteristics of TSS, DO, BOD₅ and NH₃-N. While TSS and NH₃-N showed relatively small differences in mean concentrations between seasons, DO and BOD₅ increased during the wet season. The distributional analysis further showed that the water-quality variables did not strictly satisfy the normality assumption, with the Shapiro–Wilk test rejecting normality for all parameters in both seasons. The degree of departure from normality varied among parameters, with more pronounced skewness observed for BOD₅ and NH₃-N.

Despite these distributional characteristics and the inherent variability of river water-quality data, the ANFIS model demonstrated good performance in both seasons. The model achieved slightly better results during the dry season ($r = 0.897$, $R^2 = 0.805$) than during the wet season ($r = 0.888$, $R^2 = 0.789$). These results indicate that ANFIS was able to capture a substantial proportion of the observed variability under both seasonal conditions, although a moderate proportion of the variability remained unexplained. The close agreement between R^2 and adjusted R^2 further supports the stability of the model performance.

The findings demonstrate the potential of ANFIS as an effective nonlinear modelling approach for river water-quality analysis under contrasting seasonal conditions. The slightly reduced performance during the wet season suggests greater unexplained variability during this period, highlighting the potential value of incorporating additional hydrological, meteorological, and catchment-related predictors in future modelling efforts. Further evaluation using independent validation datasets and complementary performance indicators would strengthen the assessment of model generalizability. Nevertheless, the present results indicate that season-specific ANFIS modelling provides a useful framework for characterising and predicting river water-quality variability and may support more effective seasonal water-quality monitoring and management strategies.

References

- Al-Adhaileh, M. H., & Alsaade, F. W. (2021). Modelling and prediction of water quality by using artificial intelligence. *Sustainability (Switzerland)*, 13(8). <https://doi.org/10.3390/su13084259>
- Al-Bahathy, I. A. A., Al-Janabi, Z. Z., Al-Ani, R. R., & Maktoof, A. A. (2023). Application of the Water Quality and Water Pollution Indexes for Assessing Changes in Water Quality of the Tigris River in the South Part of Iraq. *Ecological Engineering and Environmental Technology*, 24(5). <https://doi.org/10.12912/27197050/165901>
- Alphayo, S. M., Katambara, Z., & Swilla, L. (2026). Assessment of Mara River water quality in Tanzania under data constraints using the Canadian Council of Ministers of the Environment Water Quality Index (CCME-WQI). *H2Open Journal*, 9(3), 100012. <https://doi.org/10.1016/J.HTOPEN.2026.100012>
- Al-Sudani, I. M. (2021). Water quality assessment of Tigris River using overall index of pollution (OIP). *Baghdad Science Journal*, 18. [https://doi.org/10.21123/bsj.2021.18.2\(Suppl.\).1095](https://doi.org/10.21123/bsj.2021.18.2(Suppl.).1095)
- American Public Health Association, American Water Works Association, & Water Environment Federation. (2017). *Standard Methods for the Examination of Water and Wastewater* (23rd ed.). American Public Health Association, Washington DC.
- Anh, N. T., Can, L. D., Nhan, N. T., Schmalz, B., & Luu, T. Le. (2023). Influences of key factors on river water quality in urban and rural areas: A review. *Case Studies in Chemical and Environmental Engineering*, 8. <https://doi.org/10.1016/j.cscee.2023.100424>
- Ati, E. M., Abbas, R. F., Ajmi, R. N., & Zeki, H. F. (2022). WATER QUALITY ASSESSMENT IN THE AL-MUSAYYIB RIVER/EUPHRATES SYSTEM USING THE RIVER POLLUTION INDEX (RPI). *European Chemical Bulletin*, 11(5). <https://doi.org/10.31838/ecb/2022.11.05.008>
- Čerkasova, N., Umgieser, G., & Ertürk, A. (2018). Development of a hydrology and water quality model for a large transboundary river watershed to investigate the impacts of climate change – A SWAT application. *Ecological Engineering*, 124. <https://doi.org/10.1016/j.ecoleng.2018.09.025>
- Das, A., & Biswas, S. P. (2023). Studies on Seasonal Variation of Water Quality Parameters of River Mara Bharali in Sonitpur District of Assam. *Agricultural Science Digest*, 43(3). <https://doi.org/10.18805/ag.D-5706>
- Dong, Y. (2023). Descriptive Statistics and Its Applications. *Highlights in Science, Engineering and Technology*, 47. <https://doi.org/10.54097/hset.v47i.8159>
- Elsabagh, M. A., Emam, O. E., Gafar, M. G., & Medhat, T. (2024). Handling uncertainty issue in software defect prediction utilizing a hybrid of ANFIS and turbulent flow of water optimization algorithm. *Neural Computing and Applications*, 36(9). <https://doi.org/10.1007/s00521-023-09315-0>
- Elsheikh, A. H., Egiza, M., Diab, M. R., Nassar, M., Alhosary, M., Nassar, S., Rozza, M., Faisal, N., & Essa, F. A. (2025). Modelling and optimization of an inverted pyramid solar still using ANFIS-PSO: predictive analysis of water production, energy, and exergy efficiency. *Separation and Purification Technology*, 377. <https://doi.org/10.1016/j.seppur.2025.134492>
- Feng, S., Zhang, Z., Zhao, S., Guo, X., Zhu, W., & Das, P. (2023). Time lag effect of vegetation response to seasonal precipitation in the Mara River Basin. *Ecological Processes*, 12(1). <https://doi.org/10.1186/s13717-023-00461-w>
- Harbison, L., & Simmons, K. (2024). Fundamentals of Descriptive Statistics. *Journal of Dental Hygiene*, 98(5).

552 Hussain, S., & Karuppannan, S. (2023). Land use/land cover changes and their impact on land surface temperature
 553 using remote sensing technique in district Khanewal, Punjab Pakistan. *Geology, Ecology, and Landscapes*, 7(1).
 554 <https://doi.org/10.1080/24749508.2021.1923272>

555 Kang, T., Yu, N., Shin, M., Na, K., Lim, K. J., & Kim, J. (2024). Design of Non-Structural Practices for Sustainable
 556 Water Quality Improvement in an Urban River: A Case Study of South Korea. *Sustainability (Switzerland)* ,
 557 16(6). <https://doi.org/10.3390/su16062298>

558 Khan, M., Khan, A. U., Khan, J., Khan, S., Haleem, K., & Khan, F. A. (2023). Streamflow forecasting for the Hunza
 559 river basin using ANN, RNN, and ANFIS models. *Water Practice and Technology*, 18(5).
 560 <https://doi.org/10.2166/wpt.2023.060>

561 Kimario, I. C., Mkonda, M. Y., & Materu, S. F. (2025). The use of water quality indices in evaluating water quality
 562 trends from 2003 to 2023 in the Lower Mara River, Tanzania. *International Journal of River Basin Management*.
 563 <https://doi.org/10.1080/15715124.2025.2515244>

564 Kiyemba, H., Barasa, B., Asaba, J., Makoba Gudoyi, P., & Akello, G. (2023). Water Hyacinth's Extent and Its
 565 Implication on Water Quality in Lake Victoria, Uganda. *Scientific World Journal*, 2023.
 566 <https://doi.org/10.1155/2023/4947272>

567 Lubanga, H. L., Manyala, J. O., Sitati, A., Yegon, M. J., & Masese, F. O. (2021). Spatial variability in water quality
 568 and macroinvertebrate assemblages across a disturbance gradient in the Mara River Basin, Kenya. *Ecohydrology*
 569 *and Hydrobiology*, 21(4). <https://doi.org/10.1016/j.ecohyd.2021.03.001>

570 Machesa, M. G. K., Tartibu, L. K., & Okwu, M. O. (2023). Performance analysis of stirling engine using computational
 571 intelligence techniques (ANN & Fuzzy Mamdani Model) and hybrid algorithms (ANN-PSO & ANFIS). *Neural*
 572 *Computing and Applications*, 35(2). <https://doi.org/10.1007/s00521-022-07385-0>

573 Marselina, M., Wibowo, F., & Mushfiroh, A. (2022). Water quality index assessment methods for surface water: A
 574 case study of the Citarum River in Indonesia. *Heliyon*, 8(7). <https://doi.org/10.1016/j.heliyon.2022.e09848>

575 Mwangi, H. M. , Julich, S. , Patil, S. D. , McDonald, M. A. , & Feger, K. H. (2020). Modelling the impact of land use
 576 change on water quality in the Upper Mara River Basin, Kenya. . *Science of the Total Environment*, 728(138040).

577 Nakkazi, M. T., Nkwasa, A., Martínez, A. B., & van Griensven, A. (2024). Linking land use and precipitation changes
 578 to water quality changes in Lake Victoria using earth observation data. *Environmental Monitoring and*
 579 *Assessment*, 196(11). <https://doi.org/10.1007/s10661-024-13261-2>

580 Ngonyani, H., Gambo, B., Mourad, K. A., Rilwanu, T. Y., & Berndtsson, R. (2025). Indigenous Knowledge in
 581 Watershed Management: The Case of the Mara River Basin, Tanzania. *Water (Switzerland)*, 17(1).
 582 <https://doi.org/10.3390/w17010063>

583 Nkwanda, I. S., Feyisa, G. L., Zewge, F., & Makwinja, R. (2021). Impact of land-use/land-cover dynamics on water
 584 quality in the Upper Lilongwe River basin, Malawi. *International Journal of Energy and Water Resources*, 5(2).
 585 <https://doi.org/10.1007/s42108-021-00125-5>

586 Rajaei, T., Khani, S., & Ravansalar, M. (2020). Artificial intelligence-based single and hybrid models for prediction
 587 of water quality in rivers: A review. In *Chemometrics and Intelligent Laboratory Systems* (Vol. 200).
 588 <https://doi.org/10.1016/j.chemolab.2020.103978>

589 Sepahvand, M., Taghizade Firozjaee, T., Abdi, J., & Latifian, B. (2025). Assessment of water quality of the Kashkan
 590 River using water quality indicators. *Environmental Resources Research*, 13(1).
 591 <https://doi.org/10.22069/ijerr.2025.23069.1462>

- Shakerdonyavi, A., Cheloyan, P., Moridi, A. A., & Khalili, R. (2023). Assessment of Spatio-Temporal Changes in Water Quality of Babolroud River Using IRWQIsc, NSFWQI and CCME Water Quality Index. *Journal of Water and Irrigation Management*, 13(2).
- Shende, A. D., & Sahoo, M. M. (2024). Real time simulation of groundwater quality index using adaptive neuro-fuzzy inference. *Spatial Information Research*, 32(1). <https://doi.org/10.1007/s41324-023-00528-y>
- Swilla, L., Katambara, Z., & Lingwanda, M. (2024). Calibration and verification of a hydrological SWMM model for the ungauged Kinyerezi River catchment in Dar es Salaam, Tanzania. *Modeling Earth Systems and Environment*, 10(2). <https://doi.org/10.1007/s40808-023-01929-6>
- Vun, K. J., Arumugasamy, S. K., Azari, M., & Yenn, T. F. (2024). Water Level Forecasting Using Artificial Neural Network (Ann): A Case Study of Semenyih River, Malaysia. *Process Integration and Optimization for Sustainability*, 8(1). <https://doi.org/10.1007/s41660-023-00366-3>
- William, M., Katambara, Z., & Shegwando, O. (2025). Spatial assessment of water quality in a hierarchically structured river system using stream order classification and multivariate statistical techniques: A case study from Tunduma, Tanzania. *Asian Journal of Water, Environment and Pollution*, (2). <https://doi.org/10.36922/AJWEP025320248>
- Yalew, S. G., Kwakkal, J., & Doorn, N. (2021). Distributive Justice and Sustainability Goals in Transboundary Rivers: Case of the Nile Basin. *Frontiers in Environmental Science*, 8. <https://doi.org/10.3389/fenvs.2020.590954>
- Yavuz, V. S., Kartal, V., & Sambito, M. (2024). Comparative Analysis of Water Quality in Major Rivers of Türkiye Using Hydrochemical and Pollution Indices. *Water (Switzerland)*, 16(18). <https://doi.org/10.3390/w16182676>
- Younis, M. L., Rizk, E. S. T. E., Elewa, S. E., M. Abo-Elfotouh, O., & Mola, H. R. A. (2024). Seasonal evaluation of heavy metals and zooplankton distribution and their co-relationship in the Rosetta branch area of the Nile Delta in Egypt. *Applied Water Science*, 14(4). <https://doi.org/10.1007/s13201-024-02121-6>