GROUNDWATER QUALITY ASSESSMENT AND MODELING USING WATER QUALITY INDEX AND MATHEMATICAL MODELS: CASE STUDY OF BAM PROVINCE (BURKINA FASO)

Abstract:

In light of groundwater's fundamental role in providing drinking water to populations in the Sahel, its management and monitoring are vital for predicting and mitigating future crises. The goal of this study is to assess the groundwater quality in Bam province using the Water Quality Index (WQI), and to predict these indices through the application of Artificial Neural Networks (ANN) and traditional Multiple Linear Regression (MLR) techniques. In this context, a variety of physicochemical parameters such as Total Hardness (TH), pH, Electrical Conductivity (EC), calcium, magnesium, sodium, potassium, ammonium, bicarbonate, chloride, sulphate, nitrite, nitrate, phosphorus, and fluoride were collected from 154 boreholes, analysed, and used to calculate the WQIs. These parameters were employed as inputs, while the WQIs served as the output target for the models. The data were arranged in ascending order based on the indices, with 70% of the data reserved for training the models and 30% for testing.

The groundwater quality in the study area, characterized by its geological heterogeneity and discontinuous fractures affecting groundwater flow, is predominantly excellent, with 95.45% of the samples having WQIs below 50. The remaining 4.55% is split between good (3.90%) and poor (0.65%) quality, with WQIs ranging from 50–100 and 100–200, respectively. In terms of predictive modeling, the ANN method provided the most accurate results, with R^2 (coefficient of correlation) = 0.99, RMSE (Root Mean Scare Error) = 0.0037, and MAE (Mean Absolute Error) = 0.0032 for the training set, and R^2 = 0.96, RMSE = 4.46, and MAE = 3.27 for the testing set. By comparison, the traditional method showed lower accuracy with R^2 = 0.61, RMSE = 2.71, MAE = 2.02 for the training set, and R^2 = 0.93, RMSE = 7.97, MAE = 6.32 for the testing set. The slight decrease in model accuracy during the testing phase is attributed to the challenge of modeling strong indices with weaker ones, as well as the geological heterogeneity and discontinuities that complicate groundwater quality prediction. However, this does not affect the model's ability to predict extreme situations, such as water pollution events.

Key words: Bam province, Groundwater, Water Quality Index (WQI), Artificial Neural Networks (ANN), Multiple Linear Regression (MLR).

1. Introduction:

In regions with arid climates, the scarcity of drinking water frequently leads people to prioritize the search for water over evaluating its quality for human consumption or other purposes. Unfortunately, this tendency to use water without proper assessment can lead to health problems. It is important to remember that both the quantity and the quality of water are critical, regardless of its use [1]. Certainly, each parameter contributes to the determination of water quality, but none can independently define it. This makes it necessary to consider several physicochemical and/or bacteriological parameters when assessing water quality. Ideally, a wide range of parameters should be included to define the quality comprehensively. The quality of water is assessed based on the results of the analysis of its measured parameters, obtained through either field or laboratory testing. These findings are generally compared to established quality standards, which may include national (e.g., Türkiye, Canada, USA), regional (e.g., EU), or global (e.g., WHO) guidelines. A significant difficulty in this evaluation process arises when certain parameters of the same water sample comply with the standards, while others do not. This creates a challenge in accurately determining the water's quality, exposing the limitations inherent in comparison methods. The concept of a water quality index was created with the purpose of addressing this uncertainty, incorporating both physicochemical and/or bacteriological parameters in a comprehensive way to generate an index that is simple, accessible and understandable. The first technique for calculating the Water Quality Index (WQI) was introduced by [2], which took into account ten water quality parameters including sewage treatment, dissolved oxygen (DO), pH, coliforms, electroconductivity (EC), Carbon Chloroform Extract (CCE), alkalinity, chloride, temperature and obvious pollution. Subsequently, other methods for calculating the WQI have been developed, including the National Sanitation Foundation WQI [3], the Oregon WQI [4], and more recently, the Canadian WQI [5]. The WQI method is a fundamental approach that simplifies large datasets into a single value, reflecting the overall quality of water [2]. In recent years, the methods for calculating the WQI have significantly advanced, making it easier to assess and monitor water quality both spatially and temporally. The Canadian method for calculating the WQI differs from other approaches in that it treats all parameters equally in the index calculation. In contrast, other methods assign weights (ranging from 1 to 5) to each parameter based on its contribution or importance to water quality. However, the implementation of the Canadian method is challenging, as it requires at least four measurements for a minimum of four physicochemical parameters in order to calculate the index [5]. In one way or another, continuous monitoring of water resources is essential for defining a WQI, as it directly affects how we interpret their spatial and temporal fluctuations [6]. Aside from the indices used to assess water quality for human consumption, other researchers [7, 8, 9] have created indices that evaluate water quality for agricultural and plant-related uses. Through Hierarchical Cluster Analysis of WQIs, it has been shown that certain parameters do not consistently influence the index, regardless of their inclusion or exclusion, while others have a variable impact, either enhancing or diminishing water quality. The parameters falling into the latter category should be retained in the WQI calculation, whereas those in the former including pH, Zn, Cu and SO_4 may be excluded to enhance cost-efficiency [1]. So, by setting these parameters aside, we not only preserve the quality of the results but also reduce costs by not analysing them further.

In a country where agriculture is the main occupation for the majority of the population, and with the severe lack of surface water often exacerbated by climatic uncertainties, groundwater is an essential resource for addressing the water needs of the population. Agricultural, mining, and industrial activities can significantly affect the quality of water resources [10]. Therefore, it is essential to monitor water quality to ensure that its intended use yields the desired outcomes. Accurate forecasting of the WQI is vital for effective freshwater resource management, as it enables early detection of potential risks and facilitates timely interventions to prevent environmental disasters [11]. Water quality indices provide a reliable and efficient means of assessing the condition of water resources. Although the methodologies for calculating the WQI using various parameters are scientifically validated, these techniques are undeniably time-consuming.

Therefore, the development of mathematical models that bypass this calculation process would significantly enhance both time efficiency and effectiveness [12]. Such models are essential for ensuring the effective management of water resources and the safe use of water, without compromising public health or the environment. Artificial Neural Networks (ANN) are commonly employed either alone or in conjunction with techniques such as Multiple Linear Regression (MLR), Deep Neural Network (DNN), Adaptive Neuro Fuzzy Inference System (ANFIS) and other predictive methods to forecast WQI in groundwater [13, 14, 15, 16, 17] or surface water [18, 11, 19]. ANN, inspired by the structure and function of the human brain, are used to predict unknown data after a learning phase [20]. Through this technique, various layers are interconnected to model the relationship between inputs and outputs [21]. ANN offers a mathematical framework for quickly and accurately predicting water quality parameters, thereby saving time and effort [22].

Several studies have been conducted in Burkina Faso regarding water quality assessment and management. [23] assess groundwater quality in Manga and surrounding areas using physicochemical parameters and the Water Quality Index (WQI). This study indicate that most samples have excellent water quality, influenced by natural geochemical processes such as silicate and carbonate weathering, and ion exchange. [24] examine the impact of artisanal gold mining on water quality in Méguet, highlighting elevated levels of mercury, arsenic, and lead in both surface and groundwater. These contaminants, resulting from mining activities, pose significant health risks to the local population. The study emphasizes the need for improved environmental oversight and the adoption of sustainable practices to safeguard both public health and water resources. [25] investigate the quality of groundwater and spring water in the Pala locality (Bobo Dioulasso), influenced by Sotouba sandstone. Their findings indicate that water quality remains within WHO standards, with low mineralization and safe levels of heavy metals. The study underscores the importance of geological processes in water quality and recommends the establishment of a monitoring system to address potential pollution risks in the future. [26] assess wells water quality in Dédougou, finding contamination with nitrates, iron, ammonium, and faecal indicators above national standards. The study highlights the need for improved sanitation, better protection of wells, and enhanced public hygiene to prevent waterborne diseases in the region. [27] evaluate the groundwater quality in the basement aquifers of central Burkina Faso using hydrogeochemical analysis, Water Quality Index (WQI), and geostatistical methods. While the majority of borewells provide safe drinking water, high concentrations of arsenic and fluoride were observed in specific areas, influenced by both natural and anthropogenic factors. The study recommends targeted monitoring in recharge and discharge zones, particularly near Birimian Schist and granitic formations, to mitigate future contamination risks. [28] analyse arsenic concentrations in groundwater from Ouahigouya, finding levels that exceed WHO guidelines, particularly in areas associated with gold-bearing Birimian rocks. The study suggests a potential link between high arsenic levels and skin disorders, emphasizing the need for further research on arsenic contamination in the region. Finally, [29] evaluate the use of artificial neural networks (ANNs) to predict groundwater level fluctuations in the Gondo aquifer, comparing various models, including IDNN (Input Delay Neural Network), RNN (Recurrent Neural Network), and RBF (Radial Basis Function). The study concludes that the Elman-type RNN model outperforms the others, offering reliable forecasts and demonstrating its potential for groundwater management in semiarid regions with limited data.

The objective of this study is to evaluate and model the WQI of groundwater in Bam province using ANN and MLR methods. This research is unique in that it focuses on issues that have not been previously explored in the region and attempts to model high WQI values based on a data collected from 154 boreholes. To achieve this, sixteen physicochemical parameters including

Total Hardness (TH), pH, Electrical Conductivity (EC), calcium (Ca), magnesium (Mg), sodium (Na), potassium (K), ammonium (NH_4), bicarbonate (HCO_3), chloride (Cl), sulphate (SO_4), nitrite (NO_2), nitrate (NO_3), phosphorus (PO_4), and fluoride (F) were incorporated for both WQI calculation and modeling, offering valuable insights into the groundwater characteristics of the study area.

2. MATERIALS AND METHODS

2.1 Geographical overview of the study area

Bam Province is located in the northwestern part of Burkina Faso, within the North-Central region. It lies between longitudes 1°22' and 1°55' West, and latitudes 12°59' and 13°55' North. It is bordered to the north by the Yatenga Province, to the east by the Sanmatenga Province, and to the south by the Passoré and Zondoma provinces. The province has a strategic position as part of the larger Sahelian zone, which influences its climate and socio-economic conditions.

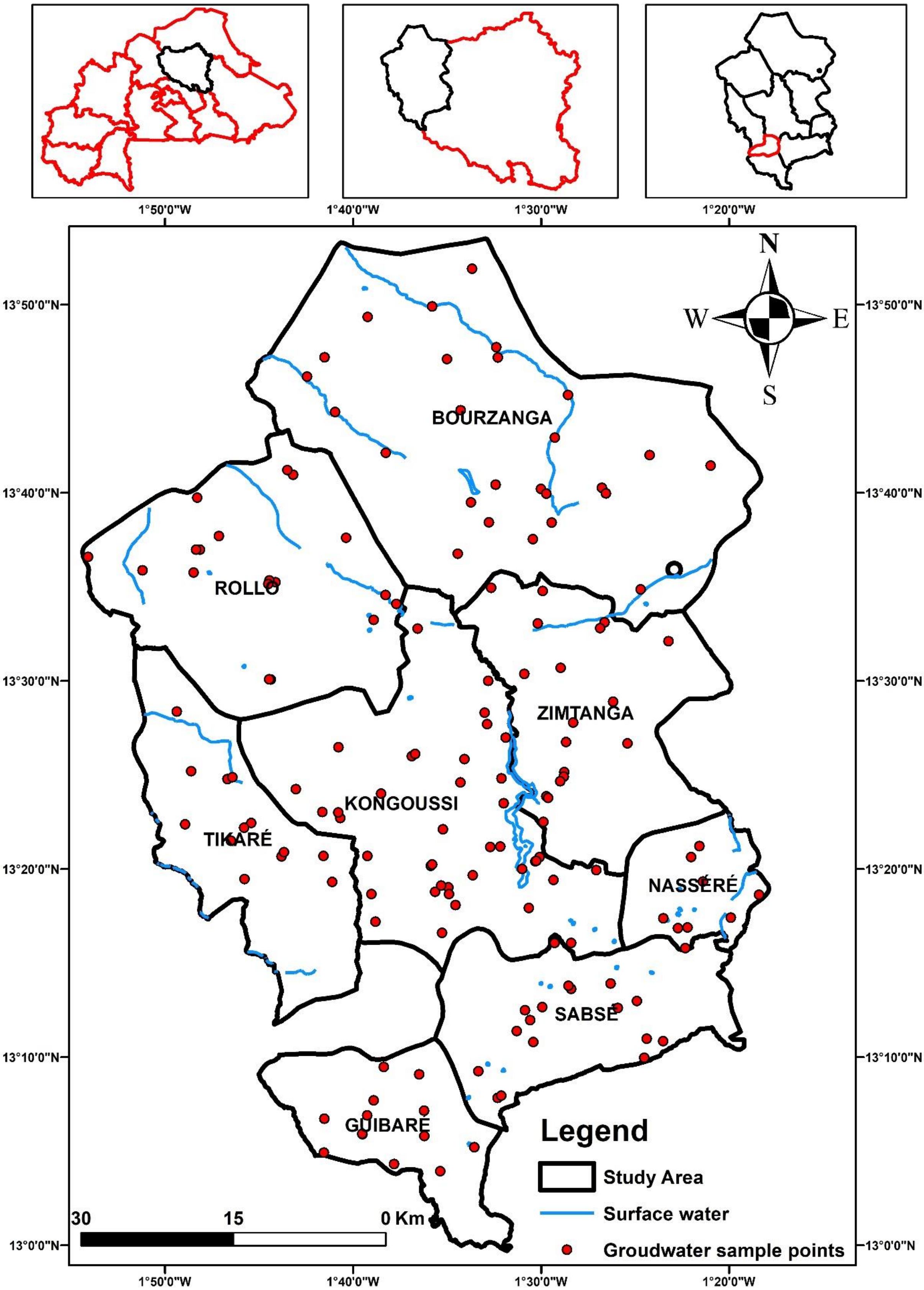

Fig. 1: Geographical localisation of study area and sampling points

The province is characterized by a prolonged dry season and a relatively short rainy season. The rainy season typically lasts three to four months, depending on the year, with annual precipitation ranging from 600 to 750 mm, which is relatively low compared to the more humid regions of the country. Influenced by the climate, the region is characterized by a vegetation type of the Sudanian-Sahelian zone. In the southern part, this consists of wooded savannah, while the northern areas feature tall grass savannah, gradually replacing the thorny steppes. The hydrography is characterized by the absence of permanent rivers, which results in often challenging water management during the dry season. Water reservoirs and wells are used to supply water during the dry period.

2.2 Geology and hydrogeology of the study area

The geology of the study area is closely linked to that of the West African Craton. As is the case for the majority of Burkina Faso (approximately 80%), this area is predominantly composed of igneous formations. More specifically, the northern part of the Bam province is dominated by metamorphic and anatectic complexes (migmatitic and anatectic gneisses), while the southern part is primarily occupied by volcano-sedimentary belts (basalt, arkosic sandstone, siltstone, claystone, pelite, quartzite, etc.). Within these two major geological units, granite and basalt intrusions are present. Structurally, the area exhibits faults oriented along a North-East/South-West direction, resulting from shear phenomena and, to a lesser extent, thrusting. Geological structures are significantly more numerous in the southern part of the study area than in the northern part. This disparity is likely to have a considerable impact on groundwater exploration in the region, where water tends to concentrate in faults and fractures. (Fig. 2).

The Bam province is endowed with groundwater resources that are crucial for supplying drinking water to local populations, particularly in rural areas. The recent model of basement aquifers, proposed by several authors [30, 31, 32], identifies three distinct "layers" based on hydrogeological properties, arranged from the bottom to the top: the top of the fractured basement consists of fractures of tectonic origin, usually sub-vertical, and functions as a significant drainage system at depth; the intermediate fractured zone is notably developed in granite formations, can extend to several tens of meters in thickness. It is characterized by primarily horizontal fractures; finally, the weathered upper layer corresponds to the weathered portion of the bedrock, which is typically more porous and permeable compared to the underlying layers. Water from the first two aquifers is generally of good quality compared to last one, but the resources can be limited and hard to access.

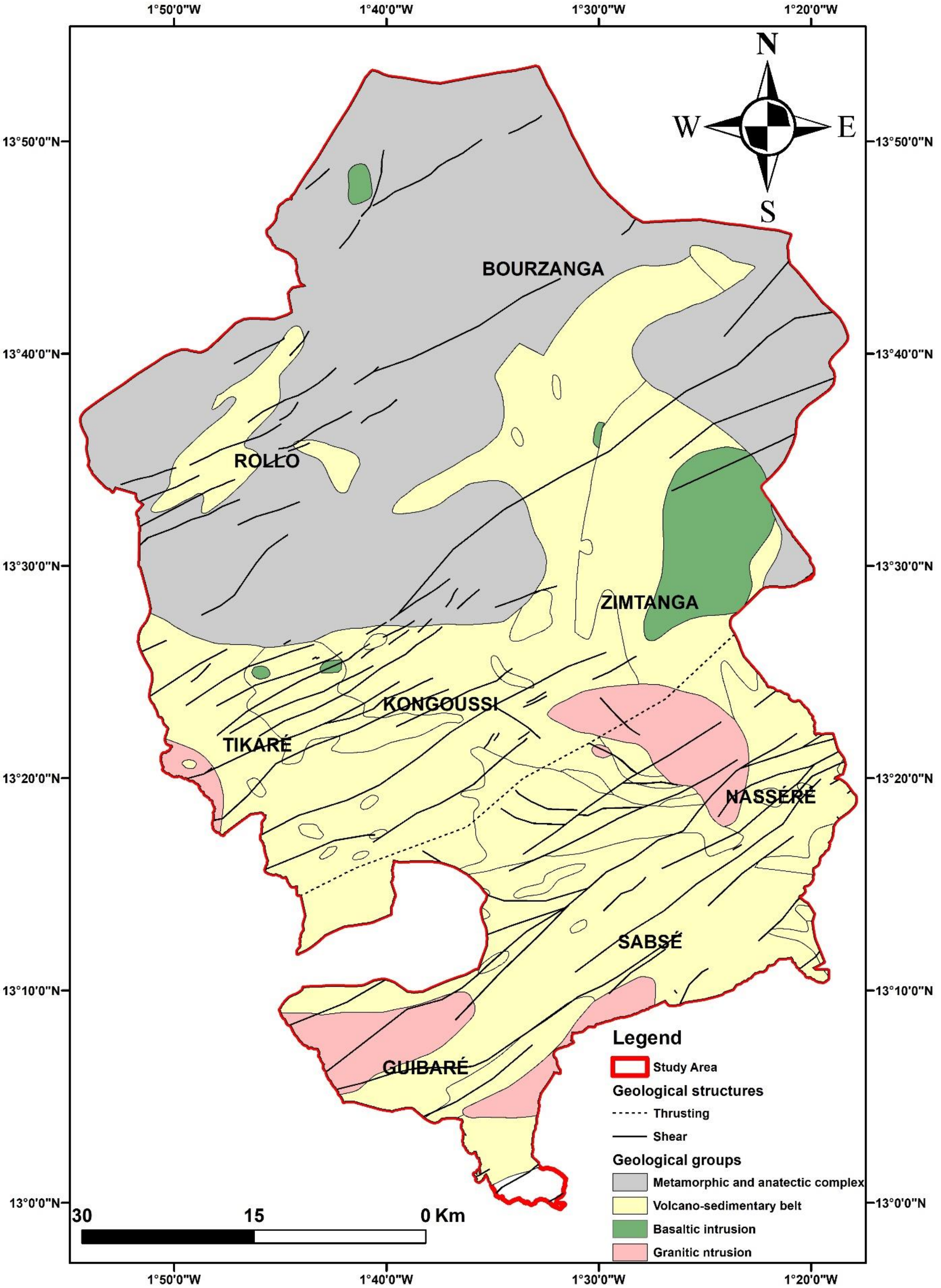

Fig. 2: Geographical localisation of study area and sampling points

2.3 Sampling and measurements

This study is based on the results of analyses conducted on 154 boreholes in 2013 by the Directorate General of Water Resources (DGRE) of Burkina Faso (Fig. 1). Out of a total of 154, the sampling points are distributed across the various communes of Bam Province as follows: Bourzanga (27 boreholes), Rollo (18 boreholes), Zimtanga (16 boreholes), Kongoussi (43 boreholes), Tikaré (13 boreholes), Guibaré (12 boreholes), Sabcé (17 boreholes), and Nasséré (8 boreholes).

The dataset includes both physical and chemical parameters. Physical parameters, such as pH and electrical conductivity (EC), were directly measured in the field using a multiparameter device. Chemical parameters were analysed by ion chromatography at the DGRE laboratory in Ouagadougou.

2.4 Methods

Before proceeding with the analysis of the data collected from the field and laboratory, particular attention was paid to verifying the quality of the obtained results. To this end, the Ionic Balance (IB) of the 154 analysed samples was determined using the following formula:

$$\mathbf{IB} = \frac{|\Sigma(\mathbf{Cations}) - \Sigma(\mathbf{Anions})| \times \mathbf{100}}{\Sigma(\mathbf{Cations}) + \Sigma(\mathbf{Anions})} \quad \text{Equation 1}$$

Where, each cation (Ca, Mg, Na, K, NH_4) and anion (HCO_3, Cl, SO_4, NO_2, NO_3, PO_4, F) parameter is in Meq/L.

Only after confirming the reliability of the results through this verification was the analytical approach, in accordance with the previously defined objectives, applied to the data. This method enabled the extraction of relevant information.

a. Water Quality Index calculation

The Weighted Arithmetic Water Quality Index (WAWQI) method was employed to determine the water quality index of the groundwater in the study area. This technique reduces multiple water quality parameter results to a single, easily interpretable value, facilitating straightforward evaluation. This advantage, along with the fact that a single analysis result from a sample point can be used, distinguishes it from the Canadian WQI method, which requires at least the results of four sampling campaigns [5]. As a result, the WAWQI method has gained worldwide recognition [33, 34].

In this work, the weighted arithmetic WQI values were determined based on the following procedure:

In the first step, the physicochemical parameters (pH, EC, Ca, Mg, Na, K, NH_4 , Fe, HCO_3 , Cl, SO_4 , NO_2 , NO_3 , PO_4 , and F), which are to be used in the calculation of the water quality index, are identified. The guideline values for each of these parameters, as per the World Health Organization (WHO) standards, are then determined (Table 1).

Table 1: WHO Drinking Water Standards, Weights, and Relative Weights used in the calculation of the groundwater quality index

Parameters	Units	WHO (2011)	Weight (wi)	Relative weight (Wi)
pH	-	6,5-8,5	4	0,08
EC at 20°C	(µS/cm)	2500	5	0,10
Total Hardness	(mg/L)	500	2	0,04
Ca	(mg/L)	75	2	0,04
Mg	(mg/L)	50	2	0,04
Na	(mg/L)	200	2	0,04
K	(mg/L)	12	2	0,04
Fe	(mg/L)	0,3	3	0,06
NH_4	(mg/L)	1,5	4	0,08
HCO_3	(mg/L)	120	1	0,02
Cl	(mg/L)	250	3	0,06
SO_4	(mg/L)	250	4	0,08
NO_2	(mg/L)	3	5	0,10
NO_3	(mg/L)	50	5	0,10
PO_4	(mg/L)	1	4	0,08
F	(mg/L)	1,5	4	0,08
Total	**-**	**-**	**52**	**1,00**

During the second step, a weight (wi) is assigned to each selected parameter based on its influence on health and its relative importance for drinking water quality, according to WHO standards. The weights range from 1 to 5, depending on the parameter's relevance in determining the overall water quality for consumption (Table 1).

In the third step, the relative weight of each parameter was determined (Table 1) using Equation 2.

$$W_i = \frac{w_i}{\sum_{i=1}^{n}(w_i)} \qquad \text{Equation 2}$$

Wi represents the relative weight of each parameter, while wi refers to the assigned weight. The total number of parameters considered is denoted by n (5). In this context, Table 1 provides the standard values based on WHO guidelines, along with the assigned and relative weights for each parameter, which are used in calculating the WQI. A weight of 5 was assigned to NO_3 , NO_2 and EC [12, 35], 4 to pH, SO_4, F [27], NH_4, PO_4, and 3 to Fe [12] and Cl [27], 2 to TH, Ca, Mg, Na, and K [27], and 1 to HCO_3 [27; 35], according to their relative importance (Table 1).

In the fourth step, the quality-rating scale (Qi) for each parameter was determined using Equation 3.

$$Q_i = \frac{(C_i - V_i)}{(S_i - V_i)} \times 100 \quad \text{Equation 3}$$

Where **Ci** represents the estimated concentration of each parameter and **Si** denotes the recommended value according to WHO drinking water quality standards, the value of **Vi** is set to zero for all parameters, except for pH, which is set to 7 [35, 36].

In the fifth step, the sub-index value for each parameter is calculated using Equation 4.

$$SI_i = Q_i \times W_i \quad \text{Equation 4}$$

Finally, the WQI for each groundwater sampling point is determined using Equation 5.

$$WQI = \sum_{i=1}^{n} (SI_i) \quad \text{Equation 5}$$

The calculated WQI values were compared to the levels provided in Table 2 to determine the water quality class. Based on the results, spatial variations in the WQI can be easily observed. This allows for the quick identification of critical points, enabling the implementation of protective and/or remedial measures. When the WQI value is below 50, from 50 to 100, 100 to 200, 200 to 300, and above 300, the water is classified as excellent, good, poor, very poor, and unsuitable for consumption, respectively.

Table 2. Classification of water according to the WQI

WQI level	**Water class**
< 50	Excellent
50-100	Good
100-200	Poor
200-300	Very poor
> 300	Inadequate for drinking

The physicochemical parameters were used as inputs, and the calculated WQIs as outputs, to develop the models using MLR and ANN. Most authors [12, 16, 10], in the context of their studies, have divided the data into three groups (training, validation, and testing) with proportions such as 60-20-20 or 70-15-15. However, some authors [37, 38] develop their models using a 70-30 split, where 70% of the data is used for training and 30% for testing the model's performance. In this study, we adopted the latter approach to place greater emphasis on the training phase while ensuring that a sufficient portion of the data remained for evaluating the model's effectiveness.

Whether for the model based on MLR or ANN, it is the training group that was used to develop the models, while the other group, was used to test the performance of the developed models. The input parameters (pH, EC, Ca, Mg, Na, K, NH_4, Fe, HCO_3, Cl, SO_4, NO_2, NO_3, PO_4, and F) used for modeling the WQIs in both the MLR and ANN techniques were identical.

b. Multiple Linear Regression (MLR)

MLR is a statistical method used to assess the relationship between a dependent variable (output) and multiple independent variables (inputs). It helps to understand how changes in the independent variables affect the dependent variable. This method, an extension of simple linear regression, is widely applied in fields such as economics, finance, and social sciences for prediction and analysing relationships between variables [12, 10]. The model was developed using SPSS software (version 25). The general form of the MLR equation is as follows:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \ldots + \beta_n X_n \qquad \text{Equation 6}$$

Where Y is the dependent variable (output), X_1, X_2, X_3, ..., X_n are the independent variables (inputs), and β_0, β_1, β_2, ..., β_n are the coefficients that are adjusted to minimize errors.

c. Artificial Neural Networks (ANN)

As previously mentioned, MLR provides a clear mathematical formula for calculating the output parameter (WQI), whereas ANN does not, and is often referred to as a "black box" technique. The application of ANN modeling was performed using MATLAB R2016b. In this study, the Multi-Layer Perceptron (MLP) architecture was chosen, which includes three distinct layers: input, hidden, and output, along with the Feed Forward Back Propagation (FFBP) algorithm (Fig. 3) for modeling the WQI [12, 10]. This algorithm operates in two phases: forward propagation and back propagation. In forward propagation, data flows from the input layer to the output layer through the hidden layer, with weights assigned to connections between nodes. The final output is derived from the aggregation of node values and weights. In back propagation, weights are adjusted to reduce errors between the calculated and modelled WQI, a technique widely used for its error-reducing capabilities [10, 39, 40]. Neurons in the network have specific values and weights, and the number of nodes in the hidden layer is optimized through trial-and-error to achieve the best configuration [10, 41].

The tuning of hidden layer nodes is influenced by the number of input features, with the goal of maximizing the coefficient of determination (R^2) and minimizing root mean square error (RMSE) and mean absolute error (MAE). As the number of neurons strays from the optimal configuration, R^2 decreases, while RMSE and MAE increase, signalling the end of testing. Different activation functions, such as linear, sigmoid, and hyperbolic tangent, were evaluated to identify the best setup. Several constants were maintained during tuning, including the sigmoid activation function for input-hidden layers, linear activation for hidden-output layers, a learning coefficient of $\lambda = 0.50$, a momentum coefficient of $\alpha = 0.50$, a maximum of 10,000 iterations, and a single output neuron for the WQI [12, 10].

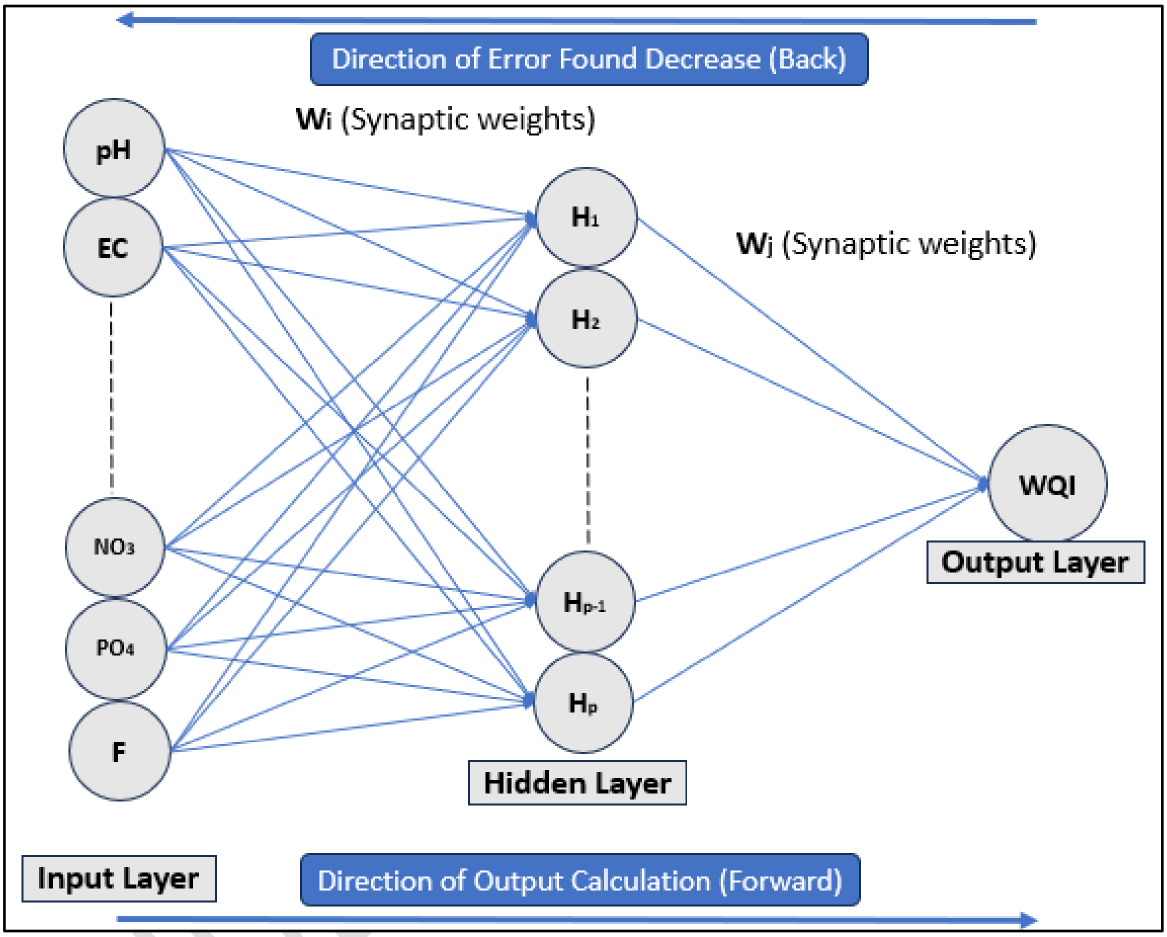

Fig. 3: Feed Forward Back Propagation Neural Network Representation

W_i and W_j represent the weights associated with the connections between the input-hidden and hidden-output layer neurons, respectively. Each input layer neuron (pH, EC, Ca, Mg, Na, K, NH_4 , Fe, HCO_3 , Cl, SO_4 , NO_2 , NO_3 , PO_4 , and F) has "p" weights, corresponding to the number of hidden layer nodes. In this study, each hidden layer neuron has a single weight, as there is only one output layer node (Fig. 3). To minimize the differences between the maximum and minimum measured data and enhance ANN efficiency, input and output layer data were normalized between 0.1 and 0.9, as shown in Equation 7 [12, 10].

$$X_{ni} = \frac{0.8(X_i - X_{min})}{(X_{max} - X_{min})} + 0.2 \quad \text{Equation 7}$$

Here, X_{ni} represents the normalized value for data point i, while X_{min} and X_{max} refer to the minimum and maximum values of the entire dataset. To obtain X_i from the model, the denormalization process is applied by using X_{min} and X_{max}, transforming the results based on the formula above.

d. Evaluation of modeling results

The performance of predictions from the MLR and ANN models was evaluated using three metrics: the coefficient of determination (R^2), Root Mean Square Error (RMSE), and Mean Absolute Error (MAE). These parameters were calculated for the training, and test datasets, as outlined in Equations 8, 9, and 10.

$$R^2 = \left[\frac{\left[\sum_{i=1}^{n}(Y_{\text{Measured } i} - \bar{Y}_{\text{Measured}})(Y_{\text{Modelled } i} - \bar{Y}_{\text{Modelled}})\right]^2}{\left[\sum_{i=1}^{n}(Y_{\text{Measured } i} - \bar{Y}_{\text{Measured}})^2\right]\left[\sum_{i=1}^{n}(Y_{\text{Modelled } i} - \bar{Y}_{\text{Modelled}})^2\right]}\right] \qquad \text{Equation 8}$$

$$\text{RMSE} = \sqrt{\frac{1}{n}\left(\sum_{i=1}^{n}(Y_{\text{Measured } i} - Y_{\text{Modelled } i})^2\right)} \qquad \text{Equation 9}$$

$$\text{MAE} = \frac{1}{n}\left(\sum_{i=1}^{n}|Y_{\text{Measured } i} - Y_{\text{Modelled } i}|\right) \qquad \text{Equation 10}$$

Where, $Y_{\text{Measured } i}$ represents the ith calculated WQI based on measured physicochemical parameters (from both field and laboratory data), $\bar{Y}_{\text{Measured}}$ is the average of the calculated WQIs, $Y_{\text{Modelled } i}$ denotes the ith modelled WQI, and $\bar{Y}_{\text{Modelled}}$ is the average of modelled WQIs.

R^2 indicates the accuracy of the model, with a value close to 1 signifying a better fit between the measured and modelled values. RMSE and MAE measure the deviation between observed and predicted values, with values close to 0 indicating higher model accuracy. These metrics were used to compare the performance of the two models, MLR and ANN, across different datasets and tests. For MLR, the optimal outcome is easily identified based on the input parameters (pH, EC, Ca, Mg, Na, K, NH_4 , Fe, HCO_3 , Cl, SO_4 , NO_2 , NO_3 , PO_4 , and F), while for ANN, several tests were conducted to identify the best model.

3. Results and Discussion

3.1 Water quality parameters

The calculated IB are situated between 0.06 and 7.33. additionally, only 5.19 % of the all samples have IB value superior to 5 which denote the quality of samples analysis devices.

Table 3 presents the descriptive statistical analysis of the physicochemical parameters of surface water in the study area. The pH values of the water samples range from 4.70 to 9.60, with a mean value of 6.89, indicating that the waters are generally slightly acidic to slightly basic. Electrical conductivity (EC) exhibits significant variability, with minimum, maximum, and mean values of 32.00 µS/cm, 1059.00 µS/cm, and 405.88 µS/cm, respectively. Total hardness (TH) values vary from 2.44 to 52.59, with an average of 20.68, suggesting that the waters are predominantly soft. The calcium concentrations range from 1.24 mg/L to 184.80 mg/L, with a mean of 35.35 mg/L. Magnesium concentrations range from 0.00 mg/L to 80.63 mg/L, with an average of 29.09 mg/L. Sodium concentrations fluctuate between 0.10 mg/L and 94.80 mg/L, with a mean of 7.08 mg/L, while potassium concentrations range from 0.00 mg/L to 9.60 mg/L, averaging 1.91 mg/L. Total iron concentrations range from 0.00 mg/L to 6.40 mg/L, with an average of 0.25 mg/L. Ammonium levels vary from 0.00 mg/L to 2.00 mg/L, with a mean concentration of 0.11 mg/L. Bicarbonate concentrations range from 45.14 mg/L to 668.56 mg/L, with an average of 267.30 mg/L. Chlorine concentrations fluctuate between 0.33 mg/L and 5.91 mg/L, with a mean of 1.06 mg/L. Sulphate concentrations range from 2.00 mg/L to 100.00 mg/L, with an average of 6.74 mg/L. Nitrite concentrations range from 0.00 mg/L to 0.18 mg/L, with an average of 0.02 mg/L. Nitrate concentrations vary from 0.44 mg/L to 33.88 mg/L, with a mean of 4.93 mg/L, while phosphate levels range from 0.25 mg/L to 2.05 mg/L, with an average of 0.78 mg/L. Finally, fluoride concentrations fluctuate between 0.02 mg/L and 2.58 mg/L, with a mean of 0.26 mg/L.

With the exception of bicarbonate and EC, which have relatively high standard deviations (126.49 and 223.34, respectively), the other parameters have rather low values for this statistical parameter, ranging between 0.03 and 25.15. Despite its significant variation (32-1059 µS/cm), EC remains within the recommended range by the WHO, indicating good water quality. This parameter reflects the mineralization of the water, its richness in ions and ability to conduct electricity. This high variability can be explained by the petrographic diversity in the study area. The majority of the chemical elements in groundwater come from ion exchanges between water and rock [10] or sea water intrusion. This value is more important in groundwater compared to surface water due to their important contact and time with rocks. As for bicarbonate, the process is much more complex. Indeed, rainwater rich in carbonic acid (CO_2) or a geological environment rich in limestone can promote high concentrations.

Table 3. Descriptive statistics of the physicochemical parameters and WQI

Parameters	Maximum	Minimum	Average	Std. Dev.	Number of Non-Compliant Points	WHO (2011)
pH	9,60	4,70	6,89	0,60	36	**6,5-8,5**

EC	1059,00	32,00	405,88	**223,34**	0	**2500**
TH	52,59	2,44	20,68	10,52	0	**500**
Ca	184,80	1,24	35,35	25,15	10	**75**
Mg	80,63	0,00	29,09	17,78	20	**50**
Na	94,80	0,10	7,08	10,75	0	**200**
K	9,60	0,00	1,91	1,82	0	**12**
Fe	6,40	0,00	0,25	0,68	17	**0,3**
NH_4	2,00	0,00	0,11	0,22	0	**1,5**
HCO_3	668,56	45,14	267,30	**126,49**	135	**120**
Cl	5,91	0,33	1,06	0,67	0	**250**
SO_4	100,00	2,00	6,74	10,65	0	**250**
NO_2	0,18	0,00	0,02	0,03	0	**3**
NO_3	33,88	0,44	4,93	4,96	0	**50**
PO_4	2,05	0,25	0,78	0,30	25	**1**
F	2,58	0,02	0,26	0,32	0	**1,5**
WQI	148.82	11.30	28.80	17.50	-	-

Std. dev. Means Standard deviation

Among the sixteen parameters, EC, TH, Na, K, NH_4, Cl, SO_4, NO_2, NO_3, and F parameters comply with the standards (WHO) while pH, Ca, Mg, Fe, HCO_3 and PO4 don't comply with in some boreholes. The vast majority of physicochemical parameters (10 out of 16) for the 154 groundwater samples are in compliance with the standards, indicating that the groundwater quality in the area is generally quite good (Table 3). Among the six non-compliant parameters, with the exception of bicarbonate which exceeds the standard 135 times, the other parameters exceed the limit between 10 and 35 times. This result suggests a generally good state of groundwater quality in the Bam province. However, one may question the overall quality of water when some parameters are compliant and others are not. Therefore, the Water Quality Index was calculated for each sample to assess the overall quality of each sample, taking all parameters into account.

3.2 Water Quality Index (WQI)

The descriptive statistics of the WQI results are summarized in Table 3. The maximum WQI value recorded was 148.82 (observed at Kouka village, Konguoussi commune), while the minimum was 11.30 (observed at Kalagré-Mossi village, Zimtenga commune). The mean WQI value was 28.80, with a standard deviation of 17.50. According to the WQI classification system used, values below 50 are classified as excellent, those between 50 and 100 as good quality, values ranging from 100 to 200 as poor quality, those between 200 and 300 as very poor quality, and values exceeding 300 are considered inadequate for drinking (Table 2). The standard deviation suggests that the WQI values do not exhibit significant variation, indicating that the majority of the values are concentrated around the mean. This result is generally favourable in terms of water quality, as the mean WQI falls within the excellent category.

Out of the total of 154 boreholes, 147 exhibit water quality indices classified as excellent, 6 as good, and 1 as poor. No samples fall under the categories of very poor or inadequate for consumption. Although a single water point is classified as poor, it can be concluded that the overall water quality in the study area is very good. The non-compliant parameters observed in some boreholes do not significantly affect the water quality. In other words, despite some water points having non-compliant parameters, the overall water quality remains generally satisfactory. Other studies conducted in the same region [23, 25, 27] or in different parts of the world [42, 13, 16] have yielded similar results regarding the high quality of groundwater. The groundwater in this area benefits from natural protection due to its position, often being shielded from direct surface contamination. However, this should not overshadow the urgent need for conservation efforts, as it is technically very challenging and extremely costly to remediate polluted groundwater once contamination has occurred. Additionally, artisanal mining activities have led to the contamination of both surface and groundwater [24], and wells have been contaminated by nitrates, iron, ammonium, and faecal matter due to inadequate sanitation and lack of well protection [26] in neighboring localities.

3.3 Groundwater Quality Index modelling

In previous studies, parameters have been classified according to various criteria, such as the measurement years [10], a gradient from upstream to downstream [12], or even randomized [37, 38, 43]. In contrast, as outlined in the methodology, once the WQI values were computed, the results were ranked in ascending order based on these indices. This approach enabled the evaluation of the ability of lower indices to predict higher indices, while also considering extreme pollution scenarios where high indices could be observed. In these cases, the model must be able to accurately capture the dynamics and potentially nonlinear relationships between the lower and higher indices, ensuring robust predictions even under severe pollution conditions. Both MLR and ANN were employed to train the models, using 16 physicochemical parameters as input and the WQI as the output, based on 70% of the available data. After training, the model performance was tested using the remaining 30% of the data.

The MLR technique was applied to derive the following formula for calculating the WQI:

$$WQI = 27.947 - 2.722pH - 0.006(EC) - 0.034(Ca) - 0.017(Mg) - 0.027(Na) + 0.288(K) + 16.018(Fe) + 4.55(NH4) + 0.034(HCO3) + 0.326(Cl) + 0.062(SO4) + 24.317(NO2) + 0.207(NO3) + 6.899(PO4) + 4.534(F) + 0(TH) \quad \text{Equation 11}$$

This formula was subsequently applied to both the training and testing datasets. Finally, the results obtained using this formula were compared with those calculated without the application of the MLR technique. The performance of both the MLR and ANN models for the WQI is presented in Table 4.

Table 4. Performance of MLR and ANN models of WQI

Models		**MLR**	**ANN**
Training set	R^2	0.61	0.999
	RMSE	2.71	0.0037
	MAE	2.02	0.0032
Test set	R^2	0.93	0.96
	RMSE	7.97	4.46
	MAE	6.32	3.27

In the case of the ANN method, the modeling process was carried out in accordance with the conditions specified in the methodology section. In contrast to the MLR approach, the ANN method offers greater flexibility for experimentation, allowing for adjustments to various parameters, such as the number of neurons in the hidden layer, the number of inputs and outputs, learning rate, activation functions, and others. For this study, the only parameter adjusted was the number of neurons in the hidden layer to optimize the model. The optimal model was selected based on several evaluation criteria, including the correlation coefficient (R^2) and error metrics (RMSE, MAE). The best-performing model was obtained after 7 iterations, with the number of neurons in the hidden layer set to 23. The results of the ANN model for the Water Quality Index (WQI) are presented in Table 4.

The model based on ANN has shown particularly promising results, especially when considering the correlation coefficients. During the training phase, the ANN model achieves a coefficient of determination (R^2) of 0.999, indicating a highly precise fit between the calculated WQI and the modelled WQI. In comparison, the model based on Multiple Linear Regression (MLR) also demonstrates the ability to model the WQI, but with less favourable evaluation coefficients (R^2, RMSE, MAE), as shown in Fig. 4.

In the testing phase, a similar trend is observed, with the ANN model still outperforming the MLR model. However, an important nuance arises: although the ANN model remains superior, the MLR model achieves a better R^2 value during testing than during training, as depicted in Fig. 4. This suggests that while both techniques are capable of modeling the WQI, the learning process differs between the two methods, and the ANN technique proves to be more robust in terms of fitting the data during training and testing.

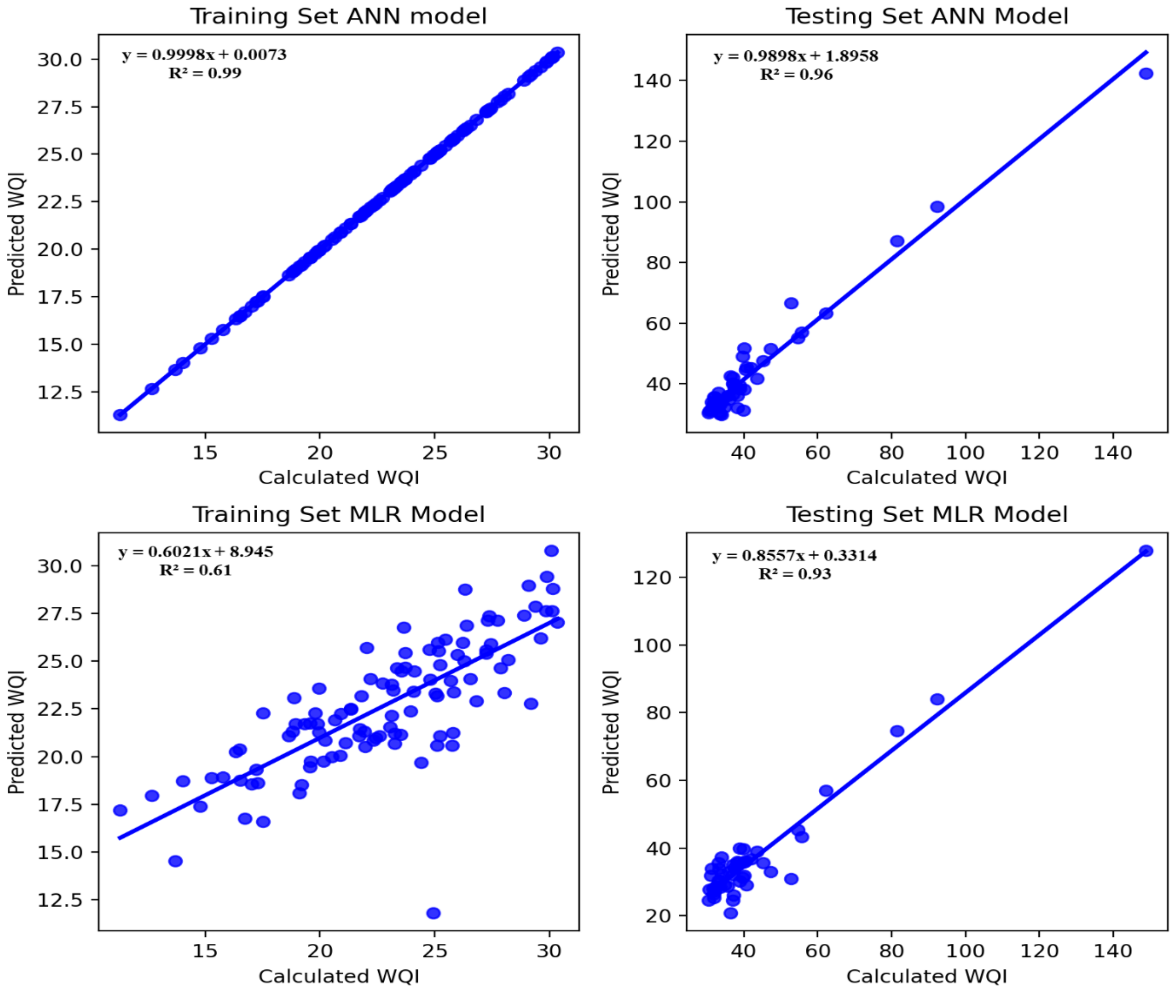

Fig. 4: Relationship between calculated and predicted WQI with ANN and MLR technics in training and testing sets

The results from the learning phase are overall more satisfactory than those obtained during the testing phase. This observation can be attributed to the challenge of modeling strong indices using weak indices. During the learning phase, the model is exposed to data that do not fully represent the reality of the field, particularly in terms of weak indices, which it struggles to capture.

The two techniques have been used by other authors to model the WQI of surface waters [12], where the data were ordered from upstream to downstream, and of groundwater [37, 38, 43], where the data were randomized, with lower error rates compared to this study. This highlights the difficulty of predicting high values based on lower ones. Moreover, the geological context of the study area, characterized by a wide variety of geological formations and discontinuous fractures through which groundwater circulates, also contributes to this disparity. This geological configuration leads to heterogeneity in water quality, due to the presence of multiple, poorly interconnected water flow pathways. Such limited interconnection restricts the relationships between different water points, thus complicating the modeling of the water quality index. However, this difficulty should not divert the research objectives in the field of water pollution, which remain the necessity for models to effectively predict extreme pollution events.

To enhance model performance despite this complex geological context and the data configuration, it would be beneficial to conduct sampling at various times of the year, particularly during the wet and dry seasons. Incorporating additional data from these periods could help better capture seasonal variations in water quality, thereby enabling the construction of a more robust and accurate model.

4. Conclusion

The groundwater quality in the study area is predominantly excellent, with 97% of the wells showing very good quality, and only one out of 154 wells having a poor WQI. While most of the physicochemical parameters analysed in this study remain within acceptable limits, certain parameters, including pH, phosphate, magnesium, iron, and occasionally calcium, exceeded the recommended limits set by the World Health Organization (WHO). Notably, bicarbonate levels surpassed the WHO standards 135 times, indicating a frequent deviation from the acceptable range. However, these exceedances have not significantly affected the overall good quality of the groundwater.

The modeling approaches employed in this study are capable of developing a WQI model based on low-quality indices and subsequently predicting higher WQI values. When comparing the results of the two techniques, it is evident that ANN outperform traditional MLR, exhibiting lower error rates (RMSE and MAE) and stronger correlations (R^2).

Despite the overall good quality of groundwater, the proliferation of small-scale artisanal gold mining, involving uncontrolled mercury and cyanide use, along with agricultural practices that rely on fertilizers and pesticides, and the insecurity linked to the use of Improvised Explosive Devices (IEDs) containing high levels of nitrates, highlight the importance of increased monitoring of water resources. Special attention should be paid to parameters such as pH, phosphate, magnesium, iron, calcium, and bicarbonate, which have exceeded the recommended standards, as well as to heavy metals and other pollutants. This targeted monitoring will be crucial in ensuring the continued protection of water quality in the Bam province.

Disclaimer (Artificial intelligence)

Option 1: Author(s) hereby declare that NO generative AI technologies such as Large Language Models (ChatGPT, COPILOT, etc.) and text-to-image generators have been used during the writing or editing of this manuscript.

References:

[1] Konare M., Gültekin F., and Temizel E. H. Using Canadian Water Quality Index method to evaluate the spatio-variation of water quality and the impacts of quality parameters: a case study of Amasya's surface water (Northern Turkey). Environmental monitoring and assessment. 2023; 195(1), 188.

[2] Horton R. K. An index number system for rating water quality. Journal of Water Pollution Control Federation. 1965; 37(3), 300–306.

[3] Brown R. M., McClelland N. I., Deininger R. A., and Tozer R. Z. A water quality index-do we dare? Water and Sewage Works. 1970; 117, 339–343.

[4] Dunnette D. A. A geographically variable water quality index used in Oregon. Water Pollution Control Federation. 1979; 53–61 ISO 690.

[5] CCME. Canadian water quality guidelines for the protection of aquatic life: Canadian Water Quality Index 1.0 Technical Report. Canadian Environmental Quality Guidelines Winnipeg Manitoba. 2001.

[6] Kazi T. G., Arain M. B., Jamali M. K., Jalbani N., Afridi H. I., Sarfraz R. A., Baig J. A., and Shah A. Q. Assessment of water quality of polluted lake using multivariate statistical techniques: A case study. Ecotoxicology and Environmental Safety. 2009; 72(20), 301–309. Available: https://doi.org/10. 1016/j.ecoenv.2008.02.024.

[7] Anyango, G. W., Bhowmick, G. D., and Bhattacharya, N. S. A Critical Review of Irrigation Water Quality Index and Water Quality Management Practices in Micro-Irrigation for Efficient Policy Making. Desalination and Water Treatment. 2024; 100304.

[8] Kandil, N. M., Rayan, R., and Sadek, M. A. Development of Irrigation Water Quality Index Using Artificial Neural Network. Journal of Scientific Research in Science. 2023; 40(1), 86-101.

[9] Majeed, S. A. A. D., Mohammed, O. I., and Hassan, W. H. Determining irrigation water quality index for evaluation groundwater quality of green-belt zone, Karbala, Iraq. In 4th International Congress on Civil Engineering, Architecture and Urban Development. 2016; (pp. 1-12).

[10] Konare M. Amasya İli Yüzey ve Yeraltısuyu Kalite Parametrelerinin Yapay Sinir Ağları Yaklaşımı ile Belirlenmesi ve Değerlendirilmesi (Doctoral Dissertation). Karadeniz Technical University. 2021.

[11] Zamili H., Bakan G., Zubaidi S. L., and Alowsi M. A. Water quality index forecast using artificial neural network techniques optimized with different metaheuristic algorithms. Modelling Earth Systems and Environment. 2023; 9(4), 4323-4333.

[12] Konare M., Traore A., Hassan, A. M., Dembele A., Cissoko F., Sissoko D. Assessment and Modeling of Water Quality in the Niger River: A Comprehensive Study Using Water Quality Indices and Intelligent Techniques. International Journal of Environment and Climate Change. 2024; 14 (10):571-88.
Available: https://doi.org/10.9734/ijecc/2024/v14i104509.
[13] Gani, A., Singh, M., Pathak, S., and Hussain, A. Groundwater quality index development using the ANN model of Delhi Metropolitan City, India. Environmental Science and Pollution Research. 2023; 1-16.
[14] Taşan, S. Estimation of groundwater quality using an integration of water quality index, artificial intelligence methods and GIS: Case study, Central Mediterranean Region of Turkey. Applied Water Science. 2023; 13(1), 15.
[15] Al-Adhaileh, M. H., Aldhyani, T. H., Alsaade, F. W., Al-Yaari, M., and Albaggar, A. K., A. Groundwater quality: The application of artificial intelligence. Journal of Environmental and Public Health. 2022; 2022(1), 8425798.
[16] Kulisz, M., Kujawska, J., Przysucha, B., and Cel, W. Forecasting water quality index in groundwater using artificial neural network. Energies. 2021; 14(18), 5875.
[17] Kadam, A. K., Wagh, V. M., Muley, A. A., Umrikar, B. N., and Sankhua, R. N. Prediction of water quality index using artificial neural network and multiple linear regression modelling approach in Shivganga River basin, India. Modeling Earth Systems and Environment. 2019; 5, 951-962.
[18] Banda, T. D., and Kumarasamy, M. Artificial Neural Network (ANN)-Based Water Quality Index (WQI) for Assessing Spatiotemporal Trends in Surface Water Quality—A Case Study of South African River Basins. Water. 2024; 16(11), 1485.
[19] Xia, L., Han, Q., Shang, L., Wang, Y., Li, X., Zhang, J., and Liu, L. Quality assessment and prediction of municipal drinking water using water quality index and artificial neural network: A case study of Wuhan, central China, from 2013 to 2019. Science of the Total Environment. 2022; 844, 157096.
[20] Samson M, Swaminathan G, Venkat KN. Assessing groundwater quality for potability using a fuzzy logic and GIS – A case study of Tiruchirappalli city – India. Computer Modelling and New Technologies. 2010; 14(2):58–68.
[21] Chang FJ, Kao LS, Kuo YM, Liu CW. Artificial neural networks for estimating regional arsenic concentrations in a blackfoot disease area in Taiwan. J Hydrol. 2010; 388:65–76.

[22] Heidarzadeh N. A practical low-cost model for prediction of the groundwater quality using artificial neural networks, Journal of Water Supply: Research and TechnologyAQUA. 2017; 66 (2):86–95.
[23] Millogo C., Sako A., Sagnon B., and Nakolendousse S. Hydrochemical and Spatial Assessment of Groundwater Quality from Fractured Basement Aquifers in the South-Central of Burkina Faso, West Africa. West Africa. 2024.
Available: https://ssrn.com/abstract=4950052 or http://dx.doi.org/10.2139/ssrn.4950052.
[24] Ouedraogo I., Lankoande S., Konate Y., Sawadogo B., Kagambega N., and Lompo M. Assessing the Impact of Gold Mining on the Quality of Water Resources in the Commune of Meguet, Burkina Faso. Journal of Water Resource and Protection. 2024; 16(04), 281-292.
[25] Millogo C., Bakouan C., Sako A., and Nakolendoussé S. Hydrogeochemical Characterization and Multivariate Analysis of Groundwater in Pala, Burkina Faso: Implications for Sustainable Water Management. International Research Journal of Pure and Applied Chemistry. 2024; 25 (5):17-34.
Available: https://doi.org/10.9734/irjpac/2024/v25i5872.
[26] Karambiri I., Dabire A. M., Zoungrana B., Meda N. S. D., and Ouedraogo B. Évaluation de la qualité bactériologique et physico-chimique des eaux de puits de la ville de Dédougou, Burkina Faso. Afrique SCIENCE. 2023; 23(2), 1-13.
[27] Sako, A., and Kafando, S. Hydrogeochemical and spatial assessment of groundwater quality from basement aquifers in the Central Plateau Region of Burkina Faso, West Africa. Environmental Earth Sciences. 2021; 80(9), 358.
[28] Smedley P. L., Knudsen J., and Maiga D. Arsenic in groundwater from mineralised Proterozoic basement rocks of Burkina Faso. Applied geochemistry. 2007; 22(5), 1074-1092.
[29] Coulibaly P., Anctil F., Aravena R., and Bobee B Artificial neural network modeling of water table depth fluctuations. Water Resour Res. 2001; 37(4):885–896.
[30] Taylor R. G., and Howard K. W. F. A tectono-geomorphic model of the hydrogeology of deeply weathered crystalline rock: Evidence from Uganda. Hydrogeology Journal. 2000; vol. 8, pp. 279– 294, 2000.
[31] Lachassagne P., and Berkowitz R. J. Aquifères de socle. Nouveaux concepts. Application à la prospection et la gestion de la ressource en eau. BRGM. 2005; no. 2, pp. 32–37.
[32] Ouandaogo-Yameogo, S., Blavoux, B., Nikiema, J., and Savadogo, A. N. Caractérisation du fonctionnement des aquifères de socle dans la région de Ouagadougou à partir d’une étude de la qualité chimique des eaux. Revue des sciences de l’eau. 2013; 26(3), 173-191.

[33] Maraşlıoğlu F, Öbekcan H. Assessing the water quality and eutrophication of the Tersakan stream (Samsun-Amasya, Turkey) using water quality index (WQI) method. ICOCEE-Cappadocia Turkey. 2017.

[34] Fathi E., Zamani-Ahmadmahmoodi R., and Zare-Bidaki R. Water quality evaluation using water quality index and multivariate methods, Beheshtabad River Iran. Applied Water Science. 2018; 8, 210.

[35] Hassan A. M., and Fırat Ersoy A. Statistical assessment of seasonal variation of groundwater quality in Çarşamba coastal plain, Samsun (Turkey). Environ Monit Assess. 2022; 194 135.

[36] Chowdhur RM, Muntasir SY, Hossain MM. Water quality index of water bodies along Faridpur-Barisal road in Bangladesh. Glob Eng Tech Rev. 2012; 2(3) 1-8.

[37] Lee J. M., Ko K. S., and Yoo K. A machine learning-based approach to predict groundwater nitrate susceptibility using field measurements and hydrogeological variables in the Nonsan Stream Watershed, South Korea. Applied Water Science. 2023; 13(12), 242.

[38] Messier K. P., Wheeler D. C., Flory A. R., Jones R. R., Patel D., Nolan B. T, Ward M. H. Modeling groundwater nitrate exposure in private wells of North Carolina for the Agricultural Health Study. Sci Total Environ. 2019; 655:512–519. Available: https://doi.org/10.1016/j.scitotenv.2018.11.022.

[39] Ismael, M., Mokhtar, A., Farooq, M., & Lü, X. Assessing drinking water quality based on physical, chemical and microbial parameters in the Red Sea State, Sudan using a combination of water quality index and artificial neural network model. Groundwater for Sustainable Development. 2021; 14, 100612.

[40] Hameed, Mohammed, et al. "Application of artificial intelligence (AI) techniques in water quality index prediction: a case study in tropical region, Malaysia." Neural Computing and Applications. 2017; 28: 893-905.

[41] Kheradpisheh Z, Talebi A, Rafati L, Ghaneian MT, Ehrampoush MH. Groundwater quality assessment using artificial neural network: A case study of Bahabad plain, Yazd, Iran. Desert. 2015; 20(1): 65-71.

[42] Ram, A., Tiwari, S. K., Pandey, H. K., Chaurasia, A. K., Singh, S., and Singh, Y. V. Groundwater quality assessment using water quality index (WQI) under GIS framework. Applied Water Science. 2021; 11, 1-20.

[43] Yoo K., Shukla S. K., Ahn J. J., Oh K., Park J. Decision tree-based data mining and rule induction for identifying hydrogeological parameters that influence groundwater pollution sensitivity. J Clean Prod. 2016; 122:277–286.

Available: https://doi.org/10.1016/j.jclepro.2016.01.075.