Journal of Engineering and Sustainable Development
http://jeasd.uomustansiriyah.edu.iq/index.php
https://doi.org/10.31272/jeasd.25.3.9

Vol. 25, No. 03, May 2021
ISSN 2520-0917

# MULTIVARIATE STATISTICAL TECHNIQUES OF WATER QUALITY IN TIGRIS RIVER WITHIN BAGHDAD CITY

**Noor Q. Ahmed[1]** ***Karim R. Gubashi[2]**

1) Environmental Engineering Department, College of Engineering, Mustansiriyah University, Baghdad, Iraq
2) Water Resources Engineering Department, College of Engineering, Mustansiriyah University, Baghdad, Iraq

***Received 25/9/2020*** ***Accepted in revised form 15/11/2020*** ***Published 1/5/2021***

**Abstract:** This study was conducted to assess the water quality of Tigris River within Baghdad city by using Weighted Arithmetic model. The studay area included five sites: Thiraa-Tigris ($S_1$), Al-Muthana bridge ($S_2$), Al-Shuhadaa bridge ($S_3$), Al-Doraa ($S_4$) and confluence point of the Diyala river ($S_5$). Ten water quality parameters were used in this study, Total Hardness (TH), Calcium (Ca), Hydrogen Ion concentration (pH), Chloride (Cl), Magnesium (Mg), Nitrate ($NO_3$), Sodium (Na), Boron (B), Turbidity (TUR) and Sulfate ($SO_4$). Statistical analysis such as correlation and regression coefficient using the statistical program SPSS was used to evaluate the results of the water quality index as well as to find a relationship between the water quality index and the variables. Several statistical methods, Factor Analysis (FA), Principal Component Analysis (PCA), Discriminant Analysis (DA), and Time Series Analysis assess parameters affecting water quality during the study period (August-December) 2019. The results showed Poor to Unsuitable quality index in Tigris River at Baghdad city except for Al-Muthna Bridge ($S_2$) was grade good quality index during the time. From the analysis, that the worst water quality index was found at confluence place of Diyala River ($S_5$) and grade unsuitable quality index.

**Keywords**: *statistical analysis, quality index; water quality, factor analysis*

## 1. Introduction

Much of the pollution loads occurring in the water result from the excessive use of water resources [12 and 16]. Because of the human and natural affects that lead to pollution occurring in water, and its effects on human health monitored, so a large number of variables causing pollution are measured [11 and 14]. Better, to understand water quality and environmental status, there are various statistical applications such as cluster analysis (CA), factor analysis (FA), and time series analysis. These statistical methods provide reliable management of water resources and give quick solutions to pollution problems [10, 13, 15, and 18]. One of the methods used to explain effects of surface water pollution is to changes in model water pollution in the river in terms of the index of quality. Changes of water quality factors used in such a model. Water quality concept defined as a set of testes for many parameters, formed in certain type, which indicates the possibility of its anthropic usage to meet certain [3]. Due to agricultural and human activities along the Tigris River there are have a significant negative impact on the river quality and its associated ecosystems. Consequently, water becomes seriously polluted, losses its clarity

**Corresponding Author*: karimgubashi@yahoo.com

Work of This Research is Licensed under CC BY

transparency, and self-purification rate decreases in year by year. Water contamination of the Tigris River is continuously increasing from upper to its lower reach. The water quality index gives about values ranging from zero to one hundred so the high values of the water quality index will indicate that the water quality is good and low values indicate that it is of poor quality. Therefore, a numerical index used as a management tool in water quality assessment [4]. Water quality index used to evaluate the water quality of Dokan Lake, Kurdistan region, by weighted arithmetic water quality index method depended on ten parameters (pH, DO, Turbidity, Conductivity, Hardness, Alkalinity, Na, BOD, $NO_3$ and $NO_2$). The results showed that the water quality in Dokan changed from good quality in the years 1978, 1979, 1980, 1999, 2000 and 2008 to poor quality in 2009 [1]. Multivariate statistical analysis used cluster analysis (CA), principal component analysis (PCA), factor analysis (FA) and discriminant analysis (DA), to assess water quality changes over time and distance and illustrate water information problems from the Songhua River [19]. Applied multivariate statistical techniques and used basic ingredient (PC) analysis, discriminant analysis (DA), and multiple linear regression analysis (MLRA) to find affected parameters of water quality through the year 2017-2018 in Tigris River. The result of MLRA presented that BOD5, Na+1, DO, and PO4-3 are the affected parameters for evaluating the COD value as a pointer of organic and nonorganic pollution [2]. The purpose of the present study was to assess the water quality of the Tigris River within Baghdad city using Weighted Arithmetic Indices and was to analyze the relation between parameters and water quality index using multivariate statistical software SPSS (version 24).

## 2. Methodology

The study presents the water quality indices of Tigris River within Baghdad city in terms of parameters such as pH, total hardness (TH), Calcium (Ca), Magnesium (Mg), Sulphate ($SO_4$), Boron (B), Chlorides (Cl), Nitrate ($NO_3$), Sodium (Na), and turbidity (TUR). Sampling points selected after a thorough field survey. Five sampling points taken starting from Thiraa Tigris ($S_1$), the first sampling point is the upstream of the river and the last is $S_5$, which is the confluence point of the Diyala River (see Fig.1). Samples collected from August to December 2019. Brief description of located sites sampling shown in Table 1. Depending on the locations in the stations shown in Table 1 take two samples for each month, from August to December 2019. The number of samples measured in the laboratory was 50 samples to measure the polluted variables. The measurements made in the laboratories of the Iraqi Water Resources Directorate, and the devices used for the tests shown in Table 2.

### 2.1 Water Quality Index, WQI

To compute WQI can be adopting the method of the Weighted Arithmetic Index developed by Tiwari and Mishra, 1985 [17]. For estimating the water quality in this research, firstly, the value of quality scale ($q_i$) for each parameter is calculated using the equation.

$$q_i = 100\left[\frac{V_i - V_o}{S_i - V_o}\right] \quad (1)$$

$q_i$ = quality rating scale for each parameter, $V_i$= estimated concentration of ith parameter in the analyzed water, $S_i$ = recommended standard value of i parameter, and $V_o$ = the ideal value of this parameter in the pure water. Then, the relative weight ($W_i$) was calculated by a value inversely proportional to the recommended standard ($S_i$ ) for the corresponding parameter using the following relation:

$$W_i = \frac{1}{S_i} \quad (2)$$

The overall WQI is calculated using the following equation:

$$WQI = \frac{\sum W_i q_i}{\sum W_i} \quad (3)$$

The grading of water quality obtained using the following types shown in Table 3.

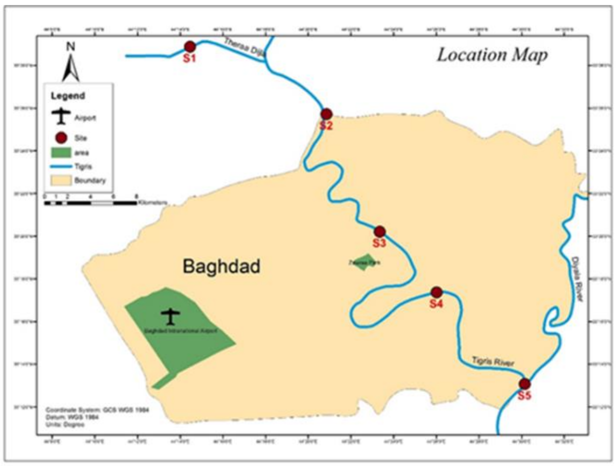


**Figure 1.** Map showing sampling sites

**Table 1.** Locations of sites sampling in Tigris River

| No. of site | Location | Geographic location: Latitudes / North | | Meridians / east | |
|---|---|---|---|---|---|
| | | Grade | Minute | Grade | Minute |
| $S_1$ | Thiraa-Tigris | 33 | 28.887 | 44 | 14.457 |
| $S_2$ | Al-Muthana bridge | 33 | 25.728 | 44 | 20.842 |
| $S_3$ | Al-Shuhadaa bridge | 33 | 20.217 | 44 | 23.313 |
| $S_4$ | Al-Doraa | 33 | 17.369 | 44 | 25.997 |
| $S_5$ | The confluence point of Diyala river | 33 | 13.073 | 44 | 30.154 |

**Table 2.** Laboratory measurements

| Parameter | Method of device |
|---|---|
| Chloride (Cl) | Titration by $AgNO_3$ |
| Total Hardness(TH) | Titration by EDTA |
| Calcium (Ca) | Titration by EDTA |
| Magnesium (Mg) | Titration by EDTA |
| Boron (B) | Spectrophotometer |
| Nitrate ($NO_3$) | Spectrophotometer |
| Sodium(Na) | Flame photometer |
| Sulfate($SO_4$) | Titration by EDT |
| pH | pH meter |
| Turbidity | Turbidity meter |

**Table 3.** Water Quality Rating as per Weight Arithmetic Water Quality Index Method

| *WQI Value* | *Rating of Water quality* | *Grading* |
|---|---|---|
| 0-25 | Excellent Water quality | A |
| 26-50 | Good Water quality | B |
| 51-75 | Poor water quality | C |
| 76-100 | Very Poor water quality | D |
| Above 100 | Unsuitable | E |

## 2.2. Statistical Analysis

To verify the water quality index, WQI a statistical analysis applied using the Social Sciences Statistical Package 24 SPSS. This Package used to evaluate Factor Analysis (FA), Principle Component Analysis (PCA), Discriminant Analysis (DA), and Time- series Analysis considered in this study for the water quality modeling. The methodology of statistical analysis is as follows.

*2.2.1. Factor Analysis (FA)*

Factor analysis is a multivariate statistical method used in Software SPSS to analyze data or correlation matrices, or matrix variations of variables ($x_1, x_2, \dots\dots x_n$ ) and their product. The goal is to clarify the relationships between these variables, and it results in several of new or assumed variables called factors ($f_1, f_2, \dots. f_n$ ). It aims to analyze a set of correlation coefficients between several variables and reduce them to a smaller number of factors, that is, it helps to understand the structure of the correlation matrix or common variance through a smaller number of factors. In mathematical forms, may be written as [20].

$$x_1 = \lambda_{11} f_1 + \lambda_{12} f_2 \dots\dots + \lambda_{1n} f_n + u_1$$

$$x_2 = \lambda_{21} f_1 + \lambda_{22} f_2 \dots\dots + \lambda_{2n} f_n + u_2 \quad (4)$$

$$x_n = \lambda_{n1} f_1 + \lambda_{n2} f_2 \dots\dots + \lambda_{nn} f_n + u_n$$

There are several methods of estimation, the most common of which is analyzing the main factors and analyzing the maximum probability factors to find the factors for the variables so that the relationships of the observed variables arise from their relationship with the factors.

*2.2.2. Principal Component Analysis (PCA)*

This analysis in Software SPSS aims to reduce data or simplify the structure of the phenomenon, that is, to simplify the presentation and representation of the phenomenon under study whenever possible without compromising important information in the hope that this will facilitate the understanding of the phenomenon [8]. The analysis of the basic components is to find the distinctive roots and distinct vectors of the correlation matrix of Eq. 4, which are represented by variables ($\lambda_{1,}\ \lambda_{2,\ \dots}\ \lambda_n$), which are called the main components of the matrix [7]. Then find the correlation coefficient between these main components and then allow them to be combined with each other, after which the distinctive roots are determined that correspond to those major components. In this method, the original variables are converted into new variables (axes) called the basic compounds and represent linear groups of the original variables, knowing that the FA analysis accompanies the PCA analysis. The primary scope of the FA is to convert the big data into a smaller number of amounts and there is another decrease in the variables groups so that they form the least influential variables during the rotation of the axis and called, in this case, PCA.

*2.2.3. Cluster Analysis (CA)*

The aim of this analysis in Software SPSS is to discover a specific pattern that organizes variables and divides them into groups whose components have common properties. Clustering analysis usually begins by forming groups based on the distance ($d_{ij}$ , for individuals i and j) between the variables, the spacing, or the convergence of the variables, as in the following equation (Euclidean distance) [7].

$$d_{ij} = \left[\sum_{l=1}^{n} \left(x_{il} - x_{jl}\right)^2\right]^{\frac{1}{2}} \quad (5)$$

The method of connecting variables together in the form of groups is called clustering algorithms. The idea is to link variables that are similar to each other in separate groups. Then these variables are re-mixed and redistributed again based on group averages as a second step. At the end of this step, group averages are calculated. This process repeated until the variables settle into specific groups. This method called k-means cluster analysis. At each stage, more and more individuals are linked together to form larger and larger clusters of increasingly dissimilar elements. The series of steps in this type of clustering conveniently summarized in a tree-like diagram known as a dendrogram.

### 2.2.4. Discriminant Analysis (DA)

It is one of the multivariate analysis methods in Software SPSS. In this method, the variables analyzed in a correlation way, taking into account the interrelationships between these variables, as it seeks to create a statistical model that depicts the interrelationship between the various variables. The discriminatory analysis allows estimation of the membership of a variable within a group based on certain variables. The purpose of a Discriminant Analysis is to classify the elements into categories within the information consisting of the predicted N group. Between the many available methods, the most popular and simplest technique is linear discriminant analysis (LDA). This analysis is used Fisher's Linear Discriminant Function. Fisher's proposition was to seek a linear transformation of the variables. Fisher suggestion linear formulation of a variable [7] as

$$z = a_1x_1 + a_2x_2 + \cdots + a_nx_n \quad (6)$$

Where z = discriminate function, a = discriminant coefficients and x = independent variables. Fisher showed that gives the coefficients for such a transformation.

$$a = s^{-1}(\bar{x}_1 - \bar{x}_2) \quad (7)$$

Where $a^T = [a_1, a_2 \ldots\ldots, a_n]$ and $\bar{x}_1$ , $\bar{x}_2$ are the group mean vectors of sample means in the two groups and s is the pooled within-groups covariance matrix, calculated from the separate within-group covariance matrices $s_1$ $and$ $s_2$ [7] as

$$s = \frac{1}{n_1+n_2-2}[(n_1 - 1)s_1 + (n_2 - 1)s_2] \quad (8)$$

Where $n_1$ and $n_2$ are the sample sizes of each group. If an individual has a discriminant score closer to $\bar{z}_1$ than to $\bar{z}_2$ , assignment is to group 1, otherwise it is to group 2. The classification rule formalized by defining a cut-off value $\bar{z}_c$ given by [7].

$$\bar{z}_c = \frac{\bar{z}_1+\bar{z}_2}{2} \quad (9)$$

Now, assuming that $\bar{z}_1$ is the larger of the two means, the classification rule of an individual with discriminant score $z_i$ is

Assign individual to group 1 if $z_i - z_c > 0$

Assign individual to group 2 if $z_i - z_c \leq 0$

Assignment techniques when there are more than two groups, for example, canonical variants analysis used.

### 2.2.5. Forecasting using Time Series

A time series in Software SPSS is a collection of readings obtained from the measurement of one variable during a regular time [6]. The primary purpose of time series modeling is to generate future readings from old checks that are in the form of a time series. The time series includes a modeling expert who assesses or defines the best fit of ARIMA (Autoregressive, Integrated, Moving Average) model or an exponential model from the series of approved variables and thus can dispense with the mathematical solution in which the trial and error. A time series is a series of data points that measured successively. For example readings for the variables pH, Ca, TH, and so on mathematically defined as a set of directions $x_t = 0, 1, 2, \ldots n$ when t represents elapsed time [5 and 6]. The measurements made during an event arranged in a time series in an appropriate chronological order. The time series containing records of one variable called univariate. However, if the records are considered to be more than one variable, then they are set as multivariate.

## 3. Results and Discussions

### 3.1 Results of Weighted Arithmetic Index Method

Water quality index for Thiraa-Tigris, $S_1$, Al-Muthana bridge, $S_2$, Al-Shuhadaa bridge, $S_3$, Al-Doraa, $S_4$, and the meeting place in the Diyala River, S5 were calculated and found ranged from 44.1 to 125.95 indicate the results revealed from good to unsuitable as shown in Fig. 2. Station $S_2$ has grade Good in October and November while station $S_5$ at all times the quality of water is Unsuitable because that is where water dumped from the Rustumiya treatment plant and the confluence of Diyala River. Other stations have quality of water Poor to Very poor during all of the times. Fig. 3 shows the variation of water quality index for stations with time. The figure presented that there is only Good quality in station $S_2$ for a period of two months (October and November). As for the rest of the time, the quality ranges between Poor and Unsuitable. The figure also shows that station $S_5$ has a quality that is Unsuitable for all study time.

### 3.2 Results of Statistical Analysis

To verify the water quality index, ten parameters studied. Then with the help of the multivariate analysis of water quality data sets for the Tigris River achieved through FA, PCA, DA, and time series analysis. FA and PCA applied to data sets (10 variables). An Eigenvalue presents a value of the sign for the factor, which with the highest Eigenvalue is the majority significant. The eigenvalue of 1.0 or greater is studied important [8]. From the results, the first two eigenvalues was greater than one. Fig. 4 shows the scree plot of the analysis. The first two components extracted and the other component eliminated. This means that most of the total variance of the first two factors as shown in Table 4 assumed the original data. Table 5 presents the distribution of total variance induced by the first two factors after rotated factor loadings. It is clear from Table 4, that 78.181% of the total variance of the measured water quality data used and the remaining 8 components only explain 21.819 %. Also, Table 5 reveals that factor loading is classified as strong corresponding to absolute loading values of greater than 0.75 ( strong : 0.75 , moderate: 0.75-0.5 and weak:0.5-0.3, developed by Liu et al. [9]. Table 4 presented the first factor presented 49.695% of the total variance and was grade strong positive loading SO4, Ca, TH, Cl, and Mg, while it show moderate positive loading with both B and Na. The second factor explained 28.487% of the total variance and grade strong positive loading Tur. and NO3, moderate positive loading Na, and B, while Mg and TH are classified weakly loading factor and weakly negative loading pH. Table 5 also shows the parameters, SO4, Ca, TH, and Cl have maximum loadings in the first factor and these parameters indicate that there is a great pollution in the river of these parameters. In the second factor, there are Tur and NO3 have the maximum loadings indicate to other pollution in the river. Using multiple linear regressions between water quality index and the two factors, that results from factor analysis in SPSS program as shown in Tables 6 and 7. From Tables, we can easily find the equation model is

$$WQI = 85.198 + 9.728\, F_1 + 18.114\, F_2 \qquad (10)$$

Table 6 shows that model fits 88% accuracy. To make true findings from regression, the residuals of the regression should follow a normal distribution. The residuals are simply the error values, or the differences between the measured value of the dependent parameter and the calculated value. The variables of factors $F_1$ *and* $F_2$ in equation 10 are evaluated by using nonlinear multiple regression.

The results calculated in Tables 8, 9, 10, and 11. The equations of factors $F_1$ *and* $F_2$ found in Tables 8 - 10. The statistical results show that the equation of the first factor fits with $R^2 = 0.989$ (Table 8 and the second factor fits with $R^2 = 0.973$ (Table 11).

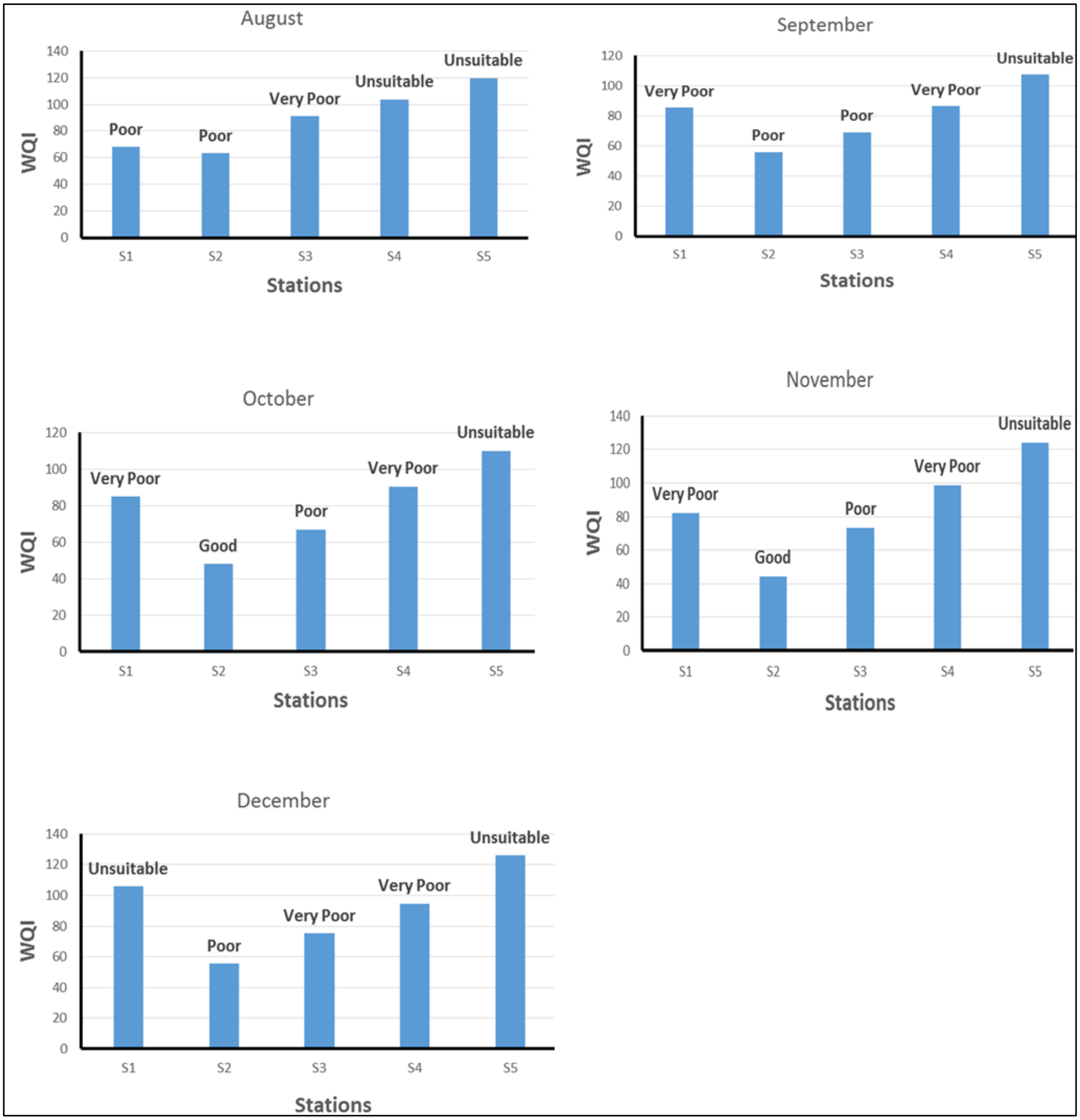


**Figure 2.** Water quality index within Baghdad city during 2019

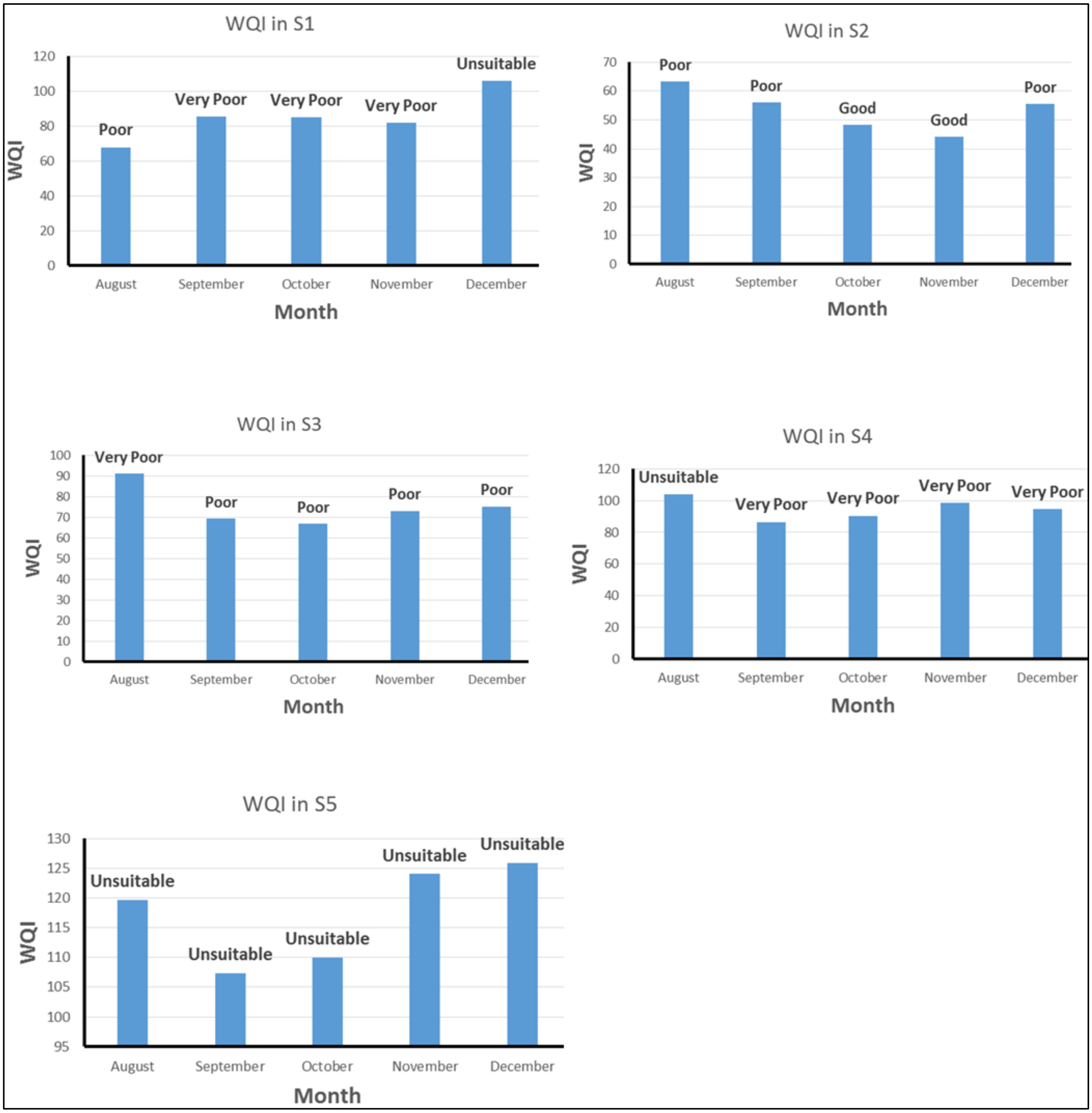


**Figure 3.** Variation of water quality index with time for stations within Baghdad city during 2019

$$F_1 = -0.536\, TH^{0.21} + 0.011\, Ca + 0.008\, Cl - 3.626\, Mg^{-0.482} - 0.95\, B + 0.004\, SO4 \quad (11)$$

$$F_2 = 0.073\, Tur + 0.389\, NO3 + 0.001\, Na - 0.356\, pH \quad (12)$$

The results of the statistical analysis time series are shown in Table 12. Goodness fit measures (RMSE, MAPE, MaxAPE, MAE, MaxAE, and Normalized BIC) revealed reliable and stable values of the developed model with $R^2$=0.98. Tables 13 and 14 presented three predictors, $SO_4$, B, and Turbidity have more significant in forecasting the water quality index values. In order to make valid inferences from regression, the residuals of regression should follow a normal distribution. The residuals are simply the error terms, or the difference between the observed value of the dependent variable and the predicated value as shown in Fig. 5. Fig. 6 shows the comparison between observed values of water quality index and that forecasted by the model. The results of this study presented the time series modeling is capable of identifying and forecasting water quality index values. Fig. 7 presented the results of the Discriminant analysis. All five sampling stations on the river classified into three statistically significant groups: DA1 ($S_2$, $S_3$, and $S_4$), DA2 ($S_5$) and DA3 ($S_1$). These discriminants of sampling sites indicated that each discriminant had a water quality of its have, which was different from the other discriminants. The DA results presented that the model was useful in showing the best category of surface water in an entire area and important for water resources of management. Thus, it can be said that for rough calculations of water quality, one station gave the sample in each discriminant is enough to evaluate the water quality of the entire station. Fig.8 shows similar to the previous results, but the type of water quality was used as a dependent variable in DA analysis and it was found that water quality is not appropriate (unsuitable grade) and very poor in the river. These results are the same as that of the Weighted Arithmetic Index Method.

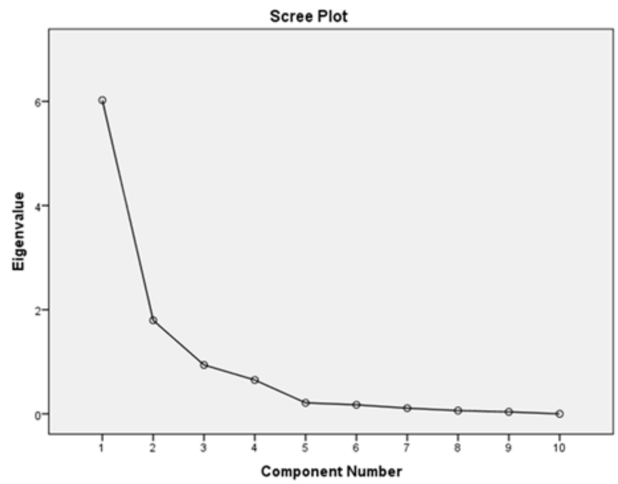


**Figure 4.** Scree plot for observed water quality

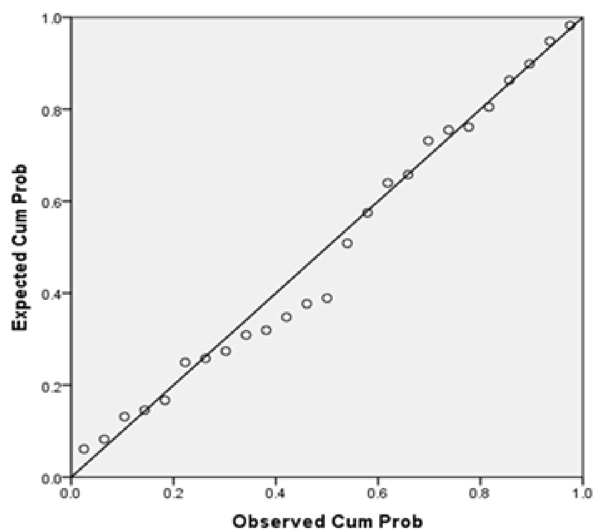


**Figure 5.** Shows normal P-P plot of regression standardized residual dependent variable

**Table 4.** Total variance explained

| Component | Initial Eigenvalues | | | Extraction sums of squared loadings | | | Rotation sums of squared loadings | | |
|---|---|---|---|---|---|---|---|---|---|
| | Total | % of Variance | Cumulative % | Total | % of Variance | Cumulative % | Total | % of Variance | Cumulative % |
| 1 | 6.023 | 60.226 | 60.226 | 6.023 | 60.226 | 60.226 | 4.969 | 49.695 | 49.695 |
| 2 | 1.796 | 17.955 | 78.181 | 1.796 | 17.955 | 78.181 | 2.849 | 28.487 | 78.181 |
| 3 | 0.938 | 9.381 | 87.562 | | | | | | |
| 4 | 0.65 | 6.499 | 94.061 | | | | | | |
| 5 | 0.211 | 2.112 | 96.173 | | | | | | |
| 6 | 0.174 | 1.736 | 97.909 | | | | | | |
| 7 | 0.109 | 1.09 | 99.00 | | | | | | |
| 8 | 0.062 | 0.621 | 99.62 | | | | | | |
| 9 | 0.038 | 0.38 | 100.0 | | | | | | |
| 10 | 1.232E-6 | 1.232E-6 | 100.0 | | | | | | |

Extraction Method: Principal Component Analysis

**Table 5.** Varimax rotated factor matrix for the whole data sets[a]

| | Component 1 | Component 2 |
|---|---|---|
| Sulfate | .949 | |
| Calcium | .946 | |
| Total hardness | .936 | .311 |
| Chloride | .924 | |
| Magnesium | .794 | .491 |
| Boron | .619 | .555 |
| Turbidity | | .873 |
| Nitrate | | .824 |
| Sodium | .583 | .677 |
| Hydrogen ion con. | | -.435 |

Extraction Method: Principal Component Analysis.

Rotation Method: Varimax with Kaiser Normalization.[a]

a. Rotation converged in 3 iterations.

**Table 6.** Model summary[b]

| Model | R | R Square | Adjusted R Square | Std. Error of the Estimate | Change Statistics | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | R Square Change | F Change | df1 | df2 | Sig. F Change |
| 1 | .884[a] | .781 | .761 | 11.366 | .781 | 39.271 | 2 | 22 | .000 |

a. Predictors: (Constant), REGR factor score 2 for analysis 1, REGR factor score 1 for analysis 1

b. Dependent Variable: Water quality index

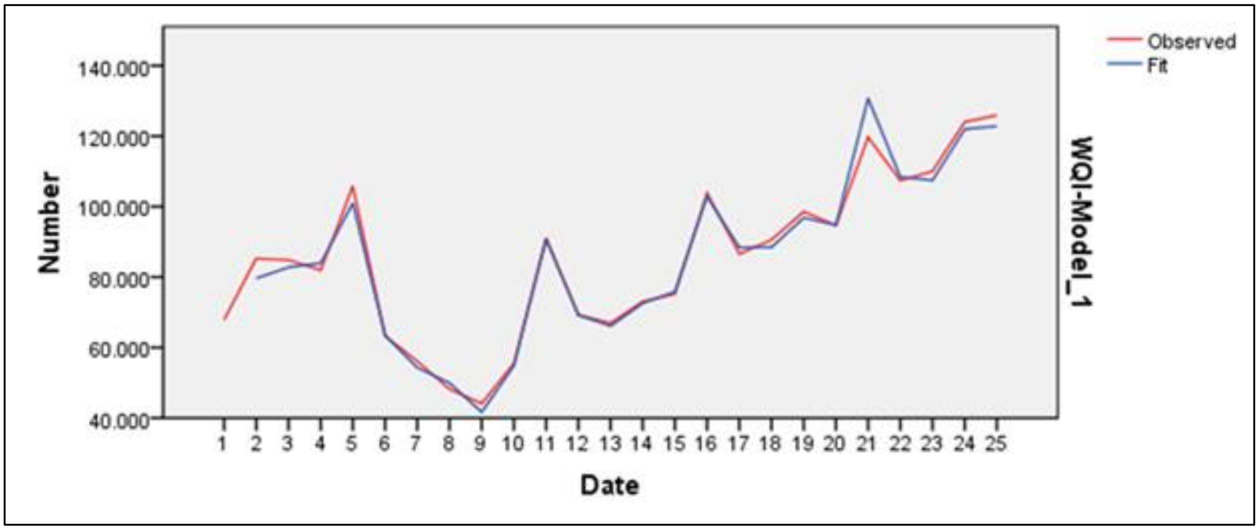


**Figure 6**. Time series plot of WQI

**Table 7**. Coefficients of regression model[a]

| Model | | Unstandardized Coefficients | | Standardized Coefficients | t | Sig. | 95.0% Confidence Interval for B | |
|---|---|---|---|---|---|---|---|---|
| | | B | Std. Error | Beta | | | Lower Bound | Upper Bound |
| 1 | (Constant) | 85.198 | 2.273 | | 37.480 | .000 | 80.484 | 89.912 |
| | REGR factor score 1 for analysis 1 | 9.728 | 2.320 | .418 | 4.193 | .000 | 4.917 | 14.540 |
| | REGR factor score 2 for analysis 1 | 18.114 | 2.320 | .779 | 7.808 | .000 | 13.303 | 22.926 |

a. Dependent Variable: Water quality index

**Table 8.** Parameter Estimates

| Parameter | Estimate | Std. Error | 95% Confidence Interval Lower Bound | Upper Bound |
|---|---|---|---|---|
| a1 | -.536 | 6.044 | -13.289 | 12.216 |
| b1 | .011 | .012 | -.015 | .036 |
| c1 | .008 | .002 | .003 | .013 |
| d1 | -3.626 | 17.571 | -40.696 | 33.445 |
| e1 | -.950 | .559 | -2.130 | .229 |
| f1 | .004 | .000 | .003 | .004 |
| a11 | .210 | 1.536 | -3.030 | 3.451 |
| d11 | -.482 | 3.358 | -7.567 | 6.603 |

* NonLinear Regression.

```
COMPUTE PRED_=a1*(TH) ** a11+b1*(Ca)+c1*(Cl)
+d1*(Mg) ** d11+e1*(B)+f1*(SO4).
 FAC1_1
   /PRED PRED_
  /SAVE PRED
  /CRITERIA ITER 100 STEPLIMIT 2 ISTEP 1E+20.
```

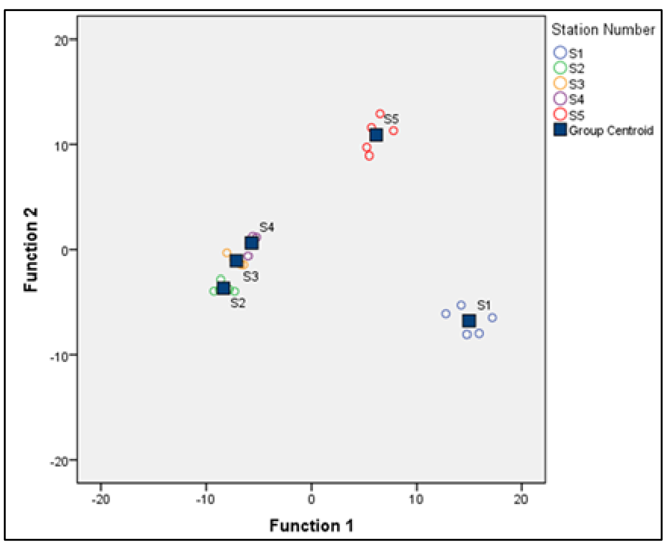


**Figure 7.** Canonical discriminant functions for stations groups

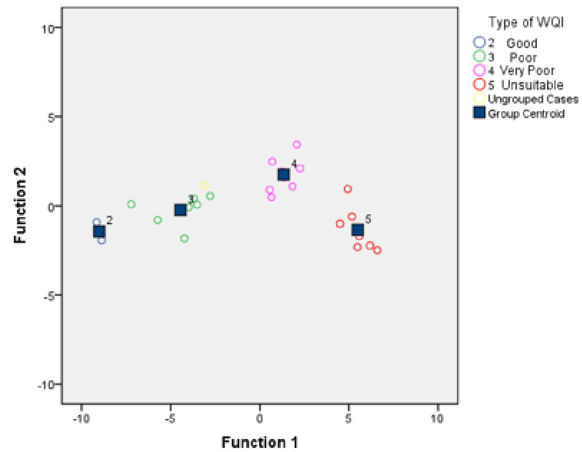


**Figure. 8** Canonical discriminant functions of WQI in Tigris River within Baghdad City during 2019

**Table 9.** ANOVA[a]

| Source | Sum of Squares | df | Mean Squares |
|---|---|---|---|
| Regression | 23.728 | 8 | 2.966 |
| Residual | .272 | 17 | .016 |
| Uncorrected Total | 24.000 | 25 | |
| Corrected Total | 24.000 | 24 | |

Dependent variable: REGR factor score 1 for analysis 1[a]

a. R squared = 1 - (Residual Sum of Squares) / (Corrected Sum of Squares) = .989.

**Table 10.** Parameter Estimates

| Parameter | Estimate | Std. Error | 95% Confidence Interval Lower Bound | Upper Bound |
|---|---|---|---|---|
| a1 | .073 | .00 | .057 | .088 |
| b1 | .389 | .033 | .321 | .458 |
| c1 | .001 | .002 | -.002 | .004 |
| d1 | -.356 | .014 | -.386 | -.326 |

* NonLinear Regression.

```
COMPUTE PRED_=a1*TUR+b1 * NO3+c1 * Na+d1 *
PH.
```

**Table 11**. ANOVA[a]

| Source | Sum of Squares | df | Mean Squares |
|---|---|---|---|
| Regression | 23.346 | 4 | 5.837 |
| Residual | .654 | 21 | .031 |
| Uncorrected Total | 24.000 | 25 | |
| Corrected Total | 24.000 | 24 | |

Dependent variable: REGR factor score 2 for analysis 1[a]

a. R squared = 1 - (Residual Sum of Squares) / (Corrected Sum of Squares) = .973.

**Table 12.** Model Fit

| Fit Statistic | Mean | SE | Minimum | Maximum | Percentile | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | 5 | 10 | 25 | 50 | 75 | 90 | 95 |
| Stationary R-squared | .976 | . | .976 | .976 | .976 | .976 | .976 | .976 | .976 | .976 | .976 |
| R-squared | .982 | . | .982 | .982 | .982 | .982 | .982 | .982 | .982 | .982 | .982 |
| RMSE | 3.328 | . | 3.328 | 3.328 | 3.328 | 3.328 | 3.328 | 3.328 | 3.328 | 3.328 | 3.328 |
| MAPE | 2.396 | . | 2.396 | 2.396 | 2.396 | 2.396 | 2.396 | 2.396 | 2.396 | 2.396 | 2.396 |
| MaxAPE | 9.205 | . | 9.205 | 9.205 | 9.205 | 9.205 | 9.205 | 9.205 | 9.205 | 9.205 | 9.205 |
| MAE | 2.110 | . | 2.110 | 2.110 | 2.110 | 2.110 | 2.110 | 2.110 | 2.110 | 2.110 | 2.110 |
| MaxAE | 11.020 | . | 11.020 | 11.020 | 11.020 | 11.020 | 11.020 | 11.020 | 11.020 | 11.020 | 11.020 |
| Normalized BIC | 2.802 | . | 2.802 | 2.802 | 2.802 | 2.802 | 2.802 | 2.802 | 2.802 | 2.802 | 2.802 |

**Time Series Modeler**

```
    /AUXILIARY  CILEVEL=95 MAXACFLAGS=24
  /MISSING USERMISSING=EXCLUDE
  /MODEL DEPENDENT=WQI INDEPENDENT=TH
Ca Mg Na SO4 NO3 Cl B PH TUR
    PREFIX='Model'
  /EXPERTMODELER TYPE=[ARIMA EXSMOOTH]
```

**Table 13.** Model Statistics

| Model | Number of Predictors | Model Fit statistics | Ljung-Box Q(18) | | | Number of Outliers |
|---|---|---|---|---|---|---|
| | | Stationary R-squared | Statistics | DF | Sig. | |
| Water quality index-Model_1 | 3 | .976 | 10.253 | 18 | .923 | 0 |

**Table 14**. ARIMA Model Parameters

| | | | | Estimate | SE | T | Sig. |
|---|---|---|---|---|---|---|---|
| Water quality index-Model_1 | Water quality index | Natural Logarithm | Difference | 1 | | | |
| | Sulfate | Natural Logarithm | Numerator Lag 0 | .115 | .030 | 3.850 | .001 |
| | | | Difference | 1 | | | |
| | Boron | Natural Logarithm | Numerator Lag 0 | .449 | .021 | 21.793 | .000 |
| | | | Difference | 1 | | | |
| | Turbidity | Natural Logarithm | Numerator Lag 0 | .361 | .086 | 4.171 | .000 |
| | | | Difference | 1 | | | |

## 4. Conclusions

Based on the overall results analysis of this study, water quality index by using Weighted Arithmetic Index method of Tigris River in all of the sampling sites within Baghdad City found poor to unsuitable grade during August, September and December 2019. This method gives good water quality index for station $S_2$ during only two months October and November 2019. Station $S_5$ (confluence of Diyala river) always has an unsuitable water quality index for the entire study period. From statistical analysis results, Factor analysis extracted two components that most of the total variance of the measured data that cumulative 78.2% of the total calculated water quality data used. The first factor showed 49.695% of the total variance and was grade strong positive loading $SO_4$, Ca, TH, Cl., Mg, and B and was moderate positive loading Na. The second factor explained 28.487% of the total variance and was grade strong positive loading Tur. and NO3, were grade moderate positive loading Na, and B, while Mg and TH are classified weakly loading factor and weakly negative loading PH. Discriminant analysis, DA revealed that five sampling stations on the river within Baghdad City grouped into three statistically significant groups. The first one includes $S_2$, $S_3$, and $S_4$, the second group includes $S_5$ and the third group includes $S_1$. One sample of these three groups may be satisfactory to find the water quality of the entire area of Baghdad City. The type of water quality was used as dependent variable in DA analysis and it was found that water quality is not appropriate (unsuitable grade) and very poor in the Tigris River within Baghdad City.


## Acknowledgments

The authors would like to thank Mustansiriah University and staff of the Water Resources Engineering Department- College of Engineering for their support to conduct this research**.**


## Conflict of interest

The publication of this article causes no conflict of interest.

## 5. References


1. Abdul Hameed M Jawad, A., Haider S, A., & Bahram K, M. (2010). Application of water quality index for assessment of Dokan lake ecosystem, Kurdistan region, Iraq. *journal of water resource and protection, 2010.*

2. Abed, S. A., Ewaid, S. H., & Al-Ansari, N. (2019, September). Evaluation of Water quality in the Tigris River within Baghdad, Iraq using Multivariate Statistical Techniques. In *Journal of Physics: Conference Series* (Vol. 1294, No. 7, p. 072025). IOP Publishing.

3. Alberto, W. D., del Pilar, D. M., Valeria, A. M., Fabiana, P. S., Cecilia, H. A., & de los Ángeles, B. M. (2001). Pattern Recognition Techniques for the Evaluation of Spatial and Temporal Variations in Water Quality. A Case Study:: Suquía River Basin (Córdoba–Argentina). *Water research*, *35*(12), 2881-2894.

4. Avannavar, Santosh M., and S. Shrihari (2008), Evaluation of water quality index for drinking purposes for river Netravathi, Mangalore, South India, Environmental Monitoring and Assessment , pp. 279-290.

5. Cochrane, J. H. (2005). Time series for macroeconomics and finance. *Manuscript, University of Chicago*, 1-136.

6. Hipel, K. W., & McLeod, A. I. (1994). *Time series modelling of water resources and environmental systems*. Elsevier.

7. Johnson, R. A., & Wichern, D. W. (2002). *Applied multivariate statistical analysis* (Vol. 5, No. 8). Upper Saddle River, NJ: Prentice hall.

8. Kim, J. O., Ahtola, O., Spector, P. E., & Mueller, C. W. (1978). *Introduction to factor analysis: What it is and how to do it* (No. 13). Sage.

9. Liu,C.W., Lin, K.H. and Kuo, Y. M. (2003), Application of factor analysis in the assessment of groundwater quality in a Blackfoot disease area in Taiwan. Science of the total environment, 313,77-89.

10. Reghunath, R., Murthy, T. S., & Raghavan, B. R. (2002). The utility of multivariate statistical techniques in hydrogeochemical studies: an example from Karnataka, India. *Water research*, *36*(10), 2437-2442.

11. Saffran, K., Cash, K., & Hallard, K. (2001). Canadian water quality guidelines for the protection of aquatic life, CCME water quality Index 1, 0, User's manual. *Excerpt from Publication, 1299.*

12. Sargaonkar, A., & Deshpande, V. (2003). Development of an overall index of pollution for surface water based on a general classification scheme in Indian context. *Environmental monitoring and assessment*, *89*(1), 43-67.

13. Shrestha, S., & Kazama, F. (2007). Assessment of surface water quality using multivariate statistical techniques: A case study of the Fuji river basin, Japan. *Environmental Modelling & Software*, *22*(4), 464-475.

14. Simeonov, V., Einax, J. W., Stanimirova, I., & Kraft, J. (2002). Environmetric modeling and interpretation of river water monitoring data. *Analytical and bioanalytical chemistry*, *374*(5), 898-905.

15. Simeonova, P., Simeonov, V., & Andreev, G. (2003). Water quality study of the Struma river basin, Bulgaria (1989–1998). *Open Chemistry*, *1*(2), 121-136.

16. Singh, K. P., Malik, A., & Sinha, S. (2005). Water quality assessment and apportionment of pollution sources of Gomti river (India) using multivariate statistical techniques—a case study. *Analytica Chimica Acta*, *538*(1-2), 355-374.

17. Tiwari, T. N., & Mishra, M. A. (1985). A preliminary assignment of water quality index of major Indian rivers. *Indian J Environ Prot*, *5*(4), 276-279.

18. Wunderlin, D. A., Diaz, M. P., Ame, M. V., Pesce, S. F., Hued, A. C., & Bistoni, M. A. (2001). Pattern recognition techniques for the evaluation of spatial and temporal variations in water quality. A case study: Suquia river basin (Cordoba, Argentina). Water Research, 35, 2881–2894.

19. Yang, L. I., Linyu, X. U., & Shun, L. I. (2009). Water quality analysis of the Songhua River Basin using multivariate techniques. *Journal of water resource and protection, 2009.*
20. Yong, A. G., & Pearce, S. (2013). A beginner's guide to factor analysis: Focusing on exploratory factor analysis. *Tutorials in quantitative methods for psychology*, *9*(2), 79-94.