Assessment and estimation of water quality using multi-linear regression.

C Lavanya[1*], *M* Nikitha[1], *N* Swetha[1], *K* Nikhitha[1], and *Laith* Hussein[2]

[1]Department of Civil Engineering, GRIET, Hyderabad, Telangana, India.
[2]College of Technical Engineering, The Islamic University, Najaf, Iraq.

Abstract. All living things require water in order to survive. Water that is drinkable for people should be safe and avoid of dangerous microorganisms. Therefore, water that has been purified should be made available to the general public before it can be utilized for human consumption. To ensure that the parameters are within allowable bounds, specific standards in the water must be investigated. Prior to use, the ground water must be evaluated and cleaned. The statistical method known as multiple linear regression makes use of two or more independent variables to forecast the value of a dependent variable. With the use of this method, we can ascertain the model's variation as well as the proportionate contributions of each independent variable to the overall variance. The assessment of water quality can be done by conducting the pH, turbidity, conductivity, acidity, alkalinity, chloride, iron, total solids, dissolved oxygen tests. Multiple samples are collected from various locations for accuracy of the test results. The aim of the study is to assess the water quality parameters through laboratory experiments and to relate the water parameters to water quality index using the multilinear regression.

1 Introduction

Water is the chief essential natural resource of around 71% on the earth's surface. Oceans holds nearly 96% of them and the remaining 4% of ground water. Most of the living organisms like plants, animals and human beings depend on water in day-to-day life and utilize water for different purposes like domestic use, irrigation purpose, industrial use, etc. Daily basis, the consumption of water increases due to increase of population rapidly and due to this water scarcity may arise. Due to the rapid industrialization and globalization and increase in population growth, it impacts the quality of water on the surface as well as in the sub-surface. Water quality can be defined by taking all the required parameters into consideration. Parameters such as physical, chemical and biological properties of water usually must be within the respective limits to its suitability for a required use. If any of the need qualities are not met, then appropriate steps should be made to enhance the water quality. Good quality water should be suitable for its intended purpose, such as irrigation, domicile usage, industrial use and hydropower generation.

* Corresponding author: lavanya.cc@gmail.com

Water quality unquestionably has an impact on ecosystem health, human safety and the availability of drinking water, all of which depend on water availability. There are so many water quality standards widely accepted around the world which are involved in their respective users. According to India, "Bureau of Indian standards" set some water quality protocols which can be used by entire country to lead a proper healthy life and to maintain good ecosystem. Water that is easily accessible and safe is essential for maintaining public health, regardless of whether it is used for drinking, household chores, cooking, etc. Better management of water resources, together with better sanitation and water delivery, may significantly reduce poverty and accelerate economic growth in nations. The UN General Assembly formally acknowledged the human right to clean water and sanitation in 2010. For home and personal use, everyone has the right to an adequate to supply of continuously flowing, safe, acceptable, physically accessible and reasonably priced water.

Water Quality Index (WQI) is a critical issue for environmental, health and water resource management. The physical, chemical and biological characteristics of water that may be used to forecast the water quality and help to determine the level of water purity are collectively referred to as Water Quality Index. Water Quality Index as developed by Brown et al. (1972) is shown in Table 1 below.

Table 1. Water Quality Index Developed by Brown et al.

Water Quality Index	Rating of Water Quality
0-25	Excellent Water Quality
25-50	Good Water Quality
50-75	Poor Water Quality
75-100	Very Poor Water Quality
Above 100	Unsuitable For Drinking

2 Literature Review

According to studies, human health greatly benefits from the assessment of drinking water quality based on aesthetic qualities and water quality indicators. Using the knowledge that families had about the water quality, aesthetic characteristics including colour, odour, and taste were assessed [1]. Developed multilinear regression models for predicting iron toxicity to aquatic organisms. The MLR models based on effective concentration 10% and 20% values were developed and performed reasonably well [2]. A single numerical expression which captures the combined influence of important parameters of water quality is feasible [3]. CCME WQI is an effective tool to evaluate water quality for water supply systems [4]. Groundwater quality is monitored through the collection of representative water samples and the examination of their physical and chemical properties at various sites [5]. The intelligent models could serve as reliable and useful tools in estimating the water quality index of the river [6].

Systematic calculations of correlation coefficients between water parameters and regression analysis were found useful for rapid monitoring of water quality [7]. Multilinear regression and artificial intelligence-based models were presented for estimating the water quality index, emphasizing the reliability of intelligent models for this purpose [8]. Changes in water parameters due to seasonal variations, particularly in lake assessments, were studied [9]. Research focused on predicting dissolved oxygen in rivers using feedforward neural network (FNN) and multilinear regression (MLR) models, achieving optimal results by combining various water quality parameters [10]. Evaluation of the relationship between water pH and

 https://doi.org/10.1051/e3sconf/202452903007

physicochemical properties, controlling for heavy metals and bacteriological factors using logistic regression [11]. The absence of a universally accepted water quality index despite various efforts and ongoing research in this area was noted [12]. These points collectively contribute to understanding, monitoring, and managing water quality, highlighting the importance of multidisciplinary approaches and predictive models in water quality assessment and management [13].

3 Methodology

3.1 Collection of Water Samples from Various Locations

Drinking water samples have been collected three different locations and from different houses. Five samples have been collected from each location. In Hyderabad, the locations chosen were from

I. Balanagar
II. Bollaram
III. Khairtabad

3.2 Conducting Various Tests on the Collected Samples

Various lab experiments were conducted to determine the water quality parameters of the water samples that were collected from various locations. The various lab experiments conducted on the water samples are pH, Turbidity, Total Solids, Dissolved Oxygen and Chlorides as shown in Figure 1 to Figure 3.

Fig. 1. Determination of pH Value

Fig. 2. Determination of Dissolved Oxygen

Fig. 3. Estimation of Chlorides

3.3 Calculation of Water Quality Index (WQI)

The details for Calculating WQI Values is as follows.

$$WQI = \sum WnQn / \sum Wn$$

Where Wn = Unit weight factors for each parameter.

Qn = Sub Index

Calculation of WQI involves the following step by step procedure.

Step 1: Calculate Wn for each parameter by using the formula

$$Wn = K/Sn$$

Where $K = 1/ (1/S1+1/S2+\ldots+1/Sn) = 1/\sum 1/Sn$

Sn = Standard desirable value of the nth parameters on summation of all selected parameters unit weight factors, Wn = 1(unity).

Step 2: Calculate the Sun-Index (Qn) value by using the formula.

$$Qn = [(Vn - Vo)]/[(Sn-Vo)] *100$$

Where, Vn = mean concentration of the nth parameters

Sn = Standard desirable value of the nth parameters

Vo = Actual values of the parameters in pure water (generally Vo = 0), for most parameters except for PH .

Step 3: Combining step1 & step2, WQI is calculated as follows

$$WQI= \sum Wn\ Qn/\sum Wn$$

3.4 Multi-linear Regression Analysis

Multi-linear regression analysis is done by using excel. The multi linear regression equation is created from the summary output.

Formula for multi-linear regression equation:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_p x_{ip} + \epsilon$$

where, for i =n observations:

y_i = dependent variable

x_i = explanatory variables

β_0 = y-intercept (constant term)

β_p = slope coefficients for each explanatory variable

ϵ = the model's error term (also known as the residuals

4 Experimental Investigation and Results

4.1 Experimental Values

Various tests like pH, Acidity, Chlorides, Total solids and Dissolved oxygen were conducted for the water samples which were collected from different locations. Table 2 shows the results of the tests for the samples collected from Balanagar (location I). Similarly, Table 3 and Table 4 shows the results for the samples collected from Bollaram (location II) and Khairtabad (location III) respectively. The water quality index values provide an overall assessment of water quality, with lower values generally indicating better water quality. Based on the WQ, while certain parameters meet acceptable standards. Water Quality Index of all the locations are shown in Table 5. Based on WQI, it has been observed that the value between 26-50 indicating Grade B and it is good quality water for two samples whereas for three samples it is more than 100 indicating the sample belongs to grade E and it is unsuitable for drinking at location I. Similar observation is made for location III also. In location II, WQI for only sample it is more than 100 indicating unsuitable for drinking whereas for other four samples, WQI is having a good water quality.

Table 2. Observational Values at Balanagar

Parameters	Sample 1	Sample 2	Sample 3	Sample 4	Sample 5	Acceptable Limits (As per IS 10500)
pH	6.62	5.86	6.79	6.28	6.92	6.5-8.5
Acidity	11.01	13.33	15.71	12.23	14.16	<200 mg/L as $CaCO_3$
Chlorides	175.48	203.84	194.975	191.43	212.7	<250 mg/L as Cl
Total solids	1000	4000	3000	2000	3000	<500 mg/L
Dissolved oxygen	7.98	8.07	0	8.27	0	<5 mg/L

Table 3. Observational Values at Bollaram

Parameters	Sample 1	Sample 2	Sample 3	Sample 4	Sample 5	Acceptable Limits (As per IS 10500)
pH	7.29	8.35	6.97	5.94	6.67	6.5-8.5
Acidity	35.53	21.74	31.58	24.36	30.11	<200 mg/L as $CaCO_3$
Chlorides	336.78	3.55	443.13	336.78	868.53	<250 mg/L as Cl
Total solids	2000	1000	4000	4000	3000	<500 mg/L
Dissolved oxygen	0	9.85	0	0	0	<5 mg/L

Table 4. Observational Values at Khairtabad

Parameters	Sample 1	Sample 2	Sample 3	Sample 4	Sample 5	Acceptable Limits (As per IS 10500)
pH	6.84	7.21	6.75	6.85	6.93	6.5-8.5
Acidity	88	72	66	82	68	<200 mg/L as $CaCO_3$
Chlorides	186.11	191.43	21.27	173.705	47.85	<250 mg/L as Cl
Total solids	0	0	2000	1000	2000	<500 mg/L
Dissolved oxygen	0	7.78	0	8.17	7.88	<5 mg/L

Table 5. Water Quality Index of All the Locations

Sample No.	Balanagar (Location I)	Bollaram (Location II)	Khairatabad (Location III)
1	127.14	35.04	30.37
2	128.84	156.42	126.53
3	33.3	36.6	31.46
4	129.62	31.68	130.95
5	33.92	36.2	128.26

4.2. Regression Plots

Graphs are plotted between water quality index and water parameters. The linear regression equation is found from the graphs as shown in Figure 4 to Figure 8. The below figures provide information about relations of pH, acidity, chloride levels, total solids concentration, and dissolved oxygen content with the Water Quality Index. Understanding these relationships between the water parameters affecting water quality and guiding efforts to calculate, maintain or improve water safety standards.

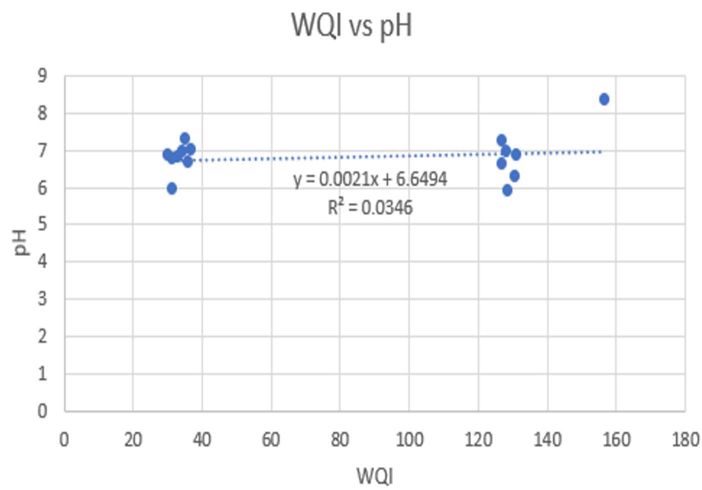

Fig. 4. Graph between WQI and pH

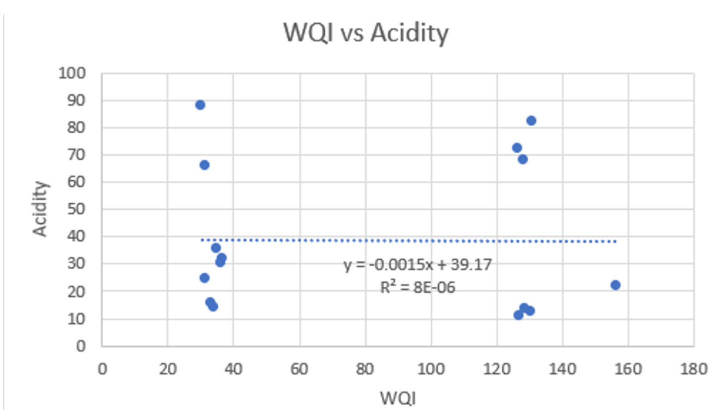

Fig. 5. Graph between WQI and Acidity

 https://doi.org/10.1051/e3sconf/202452903007

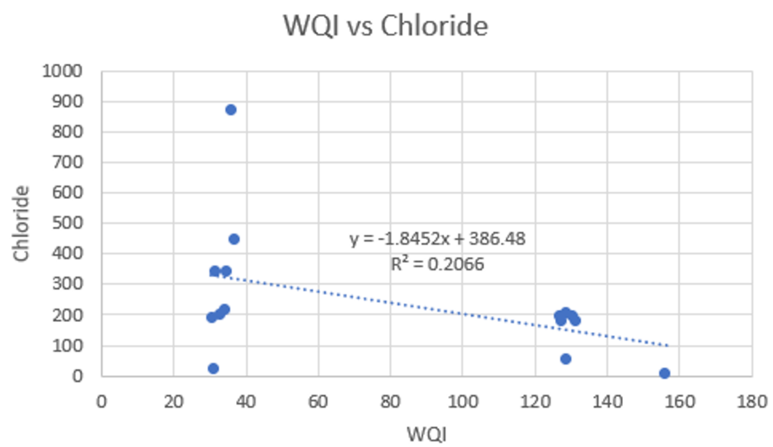

Fig. 6. Graph between WQI and Chloride

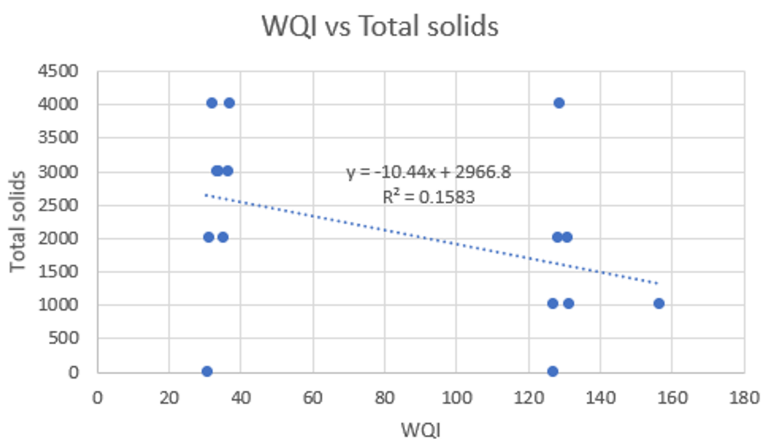

Fig. 7. Graph between WQI and Total Solids

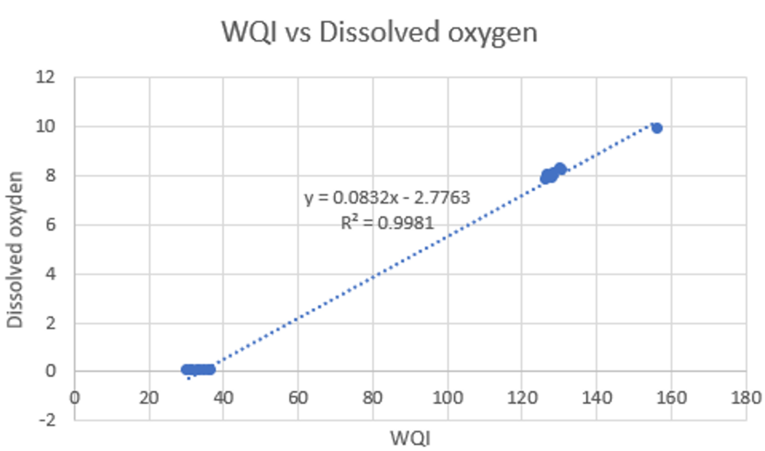

Fig. 8. Graph between WQI and Dissolved Oxygen

4.3. Multi-Linear Regression Analysis

The purpose of this analysis is to develop an equation by multiple linear regression analysis, that relates various water quality parameters to different Water Quality Index of water samples collected. The dataset used in this analysis has results of various water quality tests conducted on collected water samples. The water parameters tested include pH, Dissolved Oxygen, Acidity, Chlorides and Total Solids. A multiple linear regression model is generated to establish relationships between the water parameters and Water Quality Index as shown in Figure 9 and Figure 10.

The following are the equations obtained from the multi linear regression analysis based on the test results.

The linear equation between pH and Water Quality Index is

$$Y = 0.0021x + 6.6494.$$

The linear equation between Acidity and Water Quality Index is

$$Y = -0.0015x + 39.17$$

The linear equation between Chloride and Water Quality Index is

$$Y = -1.8452x + 386.48$$

The linear equation between Total Solids and Water Quality Index is

$$Y = -10.44x + 2966.8.$$

The linear equation between Dissolved Oxygen and Water Quality Index is

$$Y = 0.0832x - 2.7763$$

The multi-linear regression equation between Water Quality Index and the water parameters obtained is

$$Y=4.21(X1) +0.007(X2) + 0.004(X3) +0.001(X4) + 12.17(X5) + (-0.007)$$

Where y = dependent variable i.e. Water Quality Index

X = independent variables

X1 = pH

X2 = acidity

X3 = chloride

X4 = total solids

X5 = dissolved oxygen

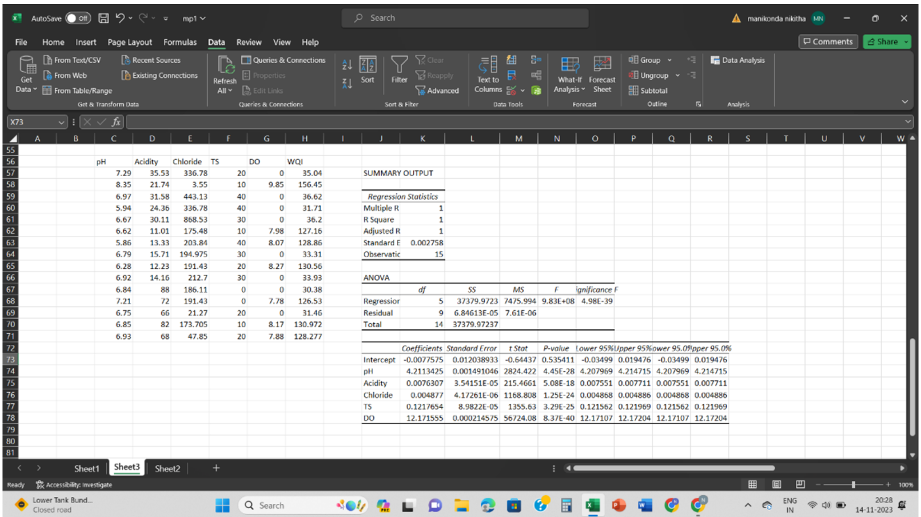

Fig. 9. Multi-Linear Regression Values

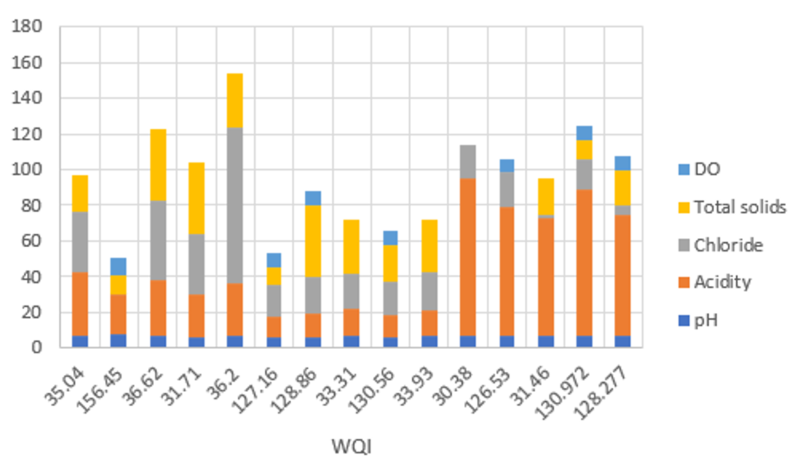

Fig. 10. Multi-Linear Regression Plot

5 Conclusions

The assessment of Water Quality Index (WQI) involves analysis of various parameters including pH, Acidity, Chloride, Total Solids and Dissolved Oxygen. Through the linear equations provided for each parameter and a multi-linear regression equation, the WQI can be determined based on the measured values of these parameters. This methodology allows for a comprehensive evaluation of water quality, enabling identification of samples suitable or unsuitable for drinking and irrigation purposes. By computing the WQI for each sample, decisions can be made regarding water usage, ensuring environmental and public health considerations taking into account. Multi-linear regression analysis serves as a valuable tool for understanding the relationship between various water parameters and the overall WQI. By incorporating multiple independent variables such as pH, Acidity, Chloride, Total Solids and Dissolved Oxygen, the multi-linear regression equation provides a comprehensive model for predicting the WQI of a given sample.

References

1. Addisie, M. B. (2022).” Evaluating drinking water quality using water quality parameters and esthetic attributes.” Air, Soil and Water Research, 15, 11786221221075005, pp. 1-8.
2. Brix, K. V., Tear, L., DeForest, D. K., Adams, W. J. (2023).” Development of multiple linear regression models for predicting chronic iron toxicity to aquatic organisms.” Environmental Toxicology and Chemistry, pp.1371-1385.
3. Brown, R. M., McClelland, N. I., Deininger, R. A., Tozer, R. G. (1970).” A water quality index-do we dare.” Water and sewage work, pp.117,110.
4. Damo, R., Icka, P. (2013).” Evaluation of Water Quality Index for Drinking Water.” Polish Journal of Environmental Studies, 22(4),pp.1045-1051.
5. Dohare, D., Deshpande, S., Kotiya, A. (2014).” Analysis of ground water quality parameters: A Review.” Research Journal of Engineering Sciences ISSN, 2278, 9472,3(5), pp 26-31.
6. Ewaid, S. H., Abed, S. A., Kadhum, S. A. (2018).” Predicting the Tigris River water quality within Baghdad, Iraq by using water quality index and regression analysis.” Environmental Technology Innovation, 11, pp.390-398.
7. Gaya, M. S., Abba, S. I., Abdu, A. M., Tukur, A. I., Saleh, M. A., Esmaili, P., Wahab,N. A. (2020).” Estimation of water quality index using artificial intelligence approaches and multi-linear regression.” Int. J. Artif. Intell. ISSN, 2252, 8938,9(1), pp.126-134.
8. Gorde, S. P., Jadhav, M. V. (2013).” Assessment of water quality parameters: a review.” JEng Res Appl, 3(6), 2029-2035. pp.31
9. Heydari, M. M., Abbasi, A., Rohani, S. M., Hosseini, S. M. A. (2013).” Correlation study and regression analysis of drinking water quality in Kashan City, Iran.” Walailak Journal of Science and Technology (WJST), 10(3), pp.315-324.
10. Joarder, M. A. M., Raihan, F., Alam, J. B., Hasanuzzaman, S. (2008).” Regression analysis of ground water quality data of Sunamganj District, Bangladesh.”, pp.291-296.
11. Kadam, A. K., Wagh, V. M., Muley, A. A., Umrikar, B. N., Sankhua, R. N. (2019).” 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, 5, pp.951-962.
12. Kumar, N., Sinha, D. K. (2010).” Drinking water quality management through correlation studies among various physico-chemical parameters: A case study.” International journal of environmental sciences, 1(2), pp.253-259.
13. Ouma, Y. O., Okuku, C. O., Njau, E. N. (2020).” Use of artificial neural networks and multiple linear regression model for the prediction of dissolved oxygen in rivers: Case study of hydrographic basin of River Nyando, Kenya.” Complexity, 2020, pp.1-23.