Journal of Mines, Metals and Fuels, *73(11): 3599-3605; 2025. DOI: 10.18311/jmmf/2025/50089*

Print ISSN : 0022-2755

Journal of Mines, Metals and Fuels

Contents available at: *www.informaticsjournals.co.in/index.php/jmmf*

A Case Study on Modeling Water Quality Index of River using Multiple Linear Regression

Rohini S. More* **and Sameer S. Shastri**

Department of Civil Engineering, Sinhgad College of Engineering, Pune - 411041, Maharashtra, India; morerohini345@gmail.com

Abstract

Water pollution assessment remains a complex and critical task. To facilitate comprehensive evaluation, the Water Quality Index (WQI) serves as a reliable metric to quantify overall water quality in terms of type and extent. In this study, the WAWQI method was employed to calculate the WQI based on key parameters including pH, DO, COD, BOD, TS, TSS, TDS, turbidity, and hardness at the Omkareshwar Sampling Station, Pune district, Maharashtra, India, over a 4-year period (2019-2022). The calculated WQI values ranged between 55 and 57, classifying the river water quality as "poor". To explore interrelationships among water quality parameters and their predictive utility, Pearson correlation and Multiple Linear Regression (MLR) analyses were conducted. A strong correlation was observed among the studied parameters, indicating potential for model-based water quality assessment. The Regression analysis for the year 2019 yielded four MLR models: MLR1 (R^2 = 0.9799), MLR2 (R^2 = 0.9828), MLR3 (R^2 = 0.9837), and MLR4 (R^2 = 0.9953). These models were subsequently applied to estimate WQI for the year 2022. The predictive model explained 74% of the variance in WQI (R^2 = 0.74, $p < 0.05$), confirming the applicability of historical water quality data in forecasting future trends. The study highlights the utility of correlation and regression approaches for water quality monitoring and supports the development of data-driven WQI models for effective environmental management.

Major Findings: *The study assessed water quality of the Mutha River (2019-2022) using the WAWQI method, with WQI values (55-57) indicating poor water quality. Significant correlations among parameters were identified, and multiple linear regression models (R^2 up to 0.9953) were developed to predict WQI, explaining 74% of the variance for 2022.*

Keywords: *Correlation, Multiple Linear Regression, River Mutha, Statistical Analysis, Water Quality Index*

1.0 Introduction

Rivers are essential for the growth of society and the economy[1]. When pollutants are released into the environment, they can cause a variety of pollution types, including water pollution[2]. Human society's development along the riverbank is the principle cause of river pollution. Major rivers receive millions of liters of wastewater from homes, farms, and industries[3]. Water quality suffers as a result, and aquatic life and people are harmed[4]. Knowing the water's quality before drinking it is crucial since drinking water that contains contaminants can lead to several water-borne illnesses. According to researchers[5] monitoring water quality is essential for assessing potential environmental risks and determining the extent of contamination. Due to the river's significance, monitoring river water quality continuously is required to assess contamination levels. Additionally, humans, animals, and aquatic plants can all suffer from severe river pollution. The first step is to properly evaluate the river's water quality before implementing any treatment system[6]. Numerous techniques have been used over the

*Author for correspondence

years to predict the quality of water. One such method is statistical analysis, which goes hand in hand with water body monitoring. The statistical regression and correlation analysis are a time saving and cost-effective technique suggested by researchers[7]. From a previously complicated set of facts[8,9], these methods offer the benefit of a straightforward and impartial interpretation. The main objectives of this research are to: (1) Assess the water quality of the Mutha rver using WQI. (2) To identify the primary factors influencing WQI, employ Multiple Linear Regression (MLR). (3) Predicting the new WQI using a Multiple Linear Regression (MLR) model.

2.0 Methodology

Water quality evaluation is the overall process of examining the water's physical, biological, and chemical characteristics[10] to conduct the assessment, several parameters were tested, including pH, TS, TDS, TSS, DO, BOD, COD, turbidity, and hardness using standard methods as prescribed by APHA (2017) for 4 years from 2019 to 2022. A total of 62 water samples for 4 consecutive years were examined for the parameters in the laboratory of Sinhgad College of Engineering, Department of Civil Engineering, Pune, Maharashtra.

3.0 Experimental Setup

The Mutha river flows through the center of Pune, which is located at 18°31' N and 73°51' E[7]. River Mutha flows through Pune city and is considered a lifeline of Pune city, Maharashtra, India[10,11]. Due to rapid increase of industrialisation, urbanisation, etc, the health of river Mutha is deteriorating, which has direct and indirect effects on living species. Human activities are the principal source of pollution in the rivers. Untreated household trash is dumped into the river[12]. 14 untreated sewage systems and natural rainwater outfalls contribute to the Mutha river's increased pollutant load[13] Pune produces 744 MLD from sewage and other sources, according to Pune Municipal Corporation. A total of 744 MLD of trash, of which 177 MLD are untreated, are entering the river[14]. Pune is the 2nd biggest city in Maharashtra and the 8th most visited metropolis in India, has a population of 7.4 million, which is served by a total of 9 sewage treatment plants. In the Sahyadri hills of the Mulshi taluka of the Pune district, the Mutha river rises close to the village of Vegare. Upstream of the Khadakwasla dam, the river is joined by 2 of its tributaries, Ambi and Moshi. One sampling station was selected along the stretch of the river Mutha for sample collection in the study region to assess the water quality index for the consecutive four years, *i.e.,* from 2019 to 2022 as shown in Figure 1. Study area river Mutha with sampling station Omkareshwar temple with latitude 18°41' N and 73°78' E is as shown in Figure 2 and Figure 3.



Figure 1. Sample location for collected samples.

4.0 Results and Discussion

4.1 Statistical Analysis

The water quality indicators minimum values, maximum values, mean, and standard deviation values were ascertained using statistical analysis. Table 1 provides a summary of the findings. The water was slightly acidic to neutral, with a mean of 6.76 and a low standard deviation (σ = 0.49). The DO (Dissolved Oxygen) levels varied from 0.8 mg/l to 5.6 mg/l, with a mean of 2.78 mg/l. Considering that most aquatic organisms need DO levels above 5 mg/l to survive, the low DO values in some samples (*e.g.*, 0.8 mg/l) suggest possible stress for aquatic life. BOD levels varied from 54 to 256 (mg/l), with an average value of



Figure 2. Actual site of Omkareshwar sampling station in the study area.

120.45 mg/l, and the range of COD levels from 120 to 800 (mg/l), with an average of 318.45 mg/l. Significant organic pollution, most likely from home or industrial wastewater discharge, is indicated by these high values. TS, TDS, and TSS: With a mean of 443.12 mg/l, the TS values varied from 280 mg/l to 720 mg/l. Variations in dissolved and suspended solids, as indicated by the TDS and TSS values, could have an impact on treatment

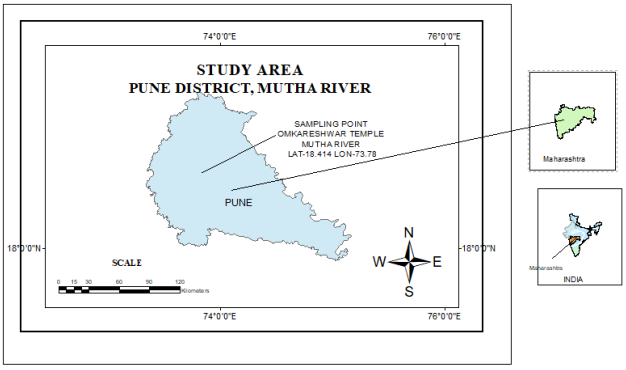


Figure 3. Omkareshwar sampling station in the study area.

Table 1. Statistical analysis of parameters

Parameter	Mean	Minimum	Maximum	Standard Deviation
pH	6.76	6	8	0.49
DO	2.78	0.8	5.6	1.12
COD	318.45	120	800	161.23
BOD	120.45	54	256	55.67
TS	443.12	280	720	120.34
TDS	252.34	155	460	85.67
TSS	193.45	35	320	78.23
Hardness	116.78	100	137	8.45
Turbidity	11.23	6.4	24	3.89

procedures and water clarity. The mean hardness value was 116.78 mg/l, with a range of 100 mg/l to 137 mg/l. This suggests fairly hard water, which is usually suitable for the majority of applications. The mean turbidity value was 11.23 NTU, with a range of 6.4 NTU to 24 NTU. Elevated levels of turbidity indicate the existence of suspended particles, which may affect the quality of the water and the effectiveness of treatment.

4.2 Water Quality Index

A "water quality index" is a collection of all the biological, chemical and physical properties of water that can be used to predict its quality and assist in determining its degree of purity[15-17]. A composite indicator known as the Water Quality Index (WQI) compiles information on the overall quality of water based on a number of factors, including pH, dissolved oxygen and other contaminants. Table 2 below displays the water quality rating suggested by Brown.

Table 2. Water quality rating

Water Quality Index	Water Quality Rating
0 to 25	Excellent
26 to 50	Good
51 to 75	Poor
76 to 100	Very Poor
Above 100	Unsuitable For Drinking

The weighted arithmetic index approach by Horton, as outlined by Cude, C is used to calculate the index of water quality.

5 steps were followed to calculate WQI for consecutive four years from 2019 to 2022

1. Select water quality parameters
2. Determine the Weight (Wn) for each parameter based on its proportional significance in determining the quality of drinking water.

$$W_n = \frac{W_i}{\sum_i^n W_i} = 1 \quad (1)$$

3. Compute the Quality rating scale (Q_n) for measured parameters using Table 3

$$Q_n = \left(\frac{V_{Standard} - V_{Ideal}}{V_n - V_{Ideal}} \right) \times 100 \quad (2)$$

4. Calculate the Sub-Index (SI)

$$SIn = Wn \times Qn \quad (3)$$

5. Determine the Water Quality Index (WQI)

$$WQI = \sum SIn \quad (4)$$

The average values during the study were 7.08, 56.04, 55.89 and 55.22 for the 4 consecutive years shown in Table 4.

5.0 Correlation and Regression

To depict realistic water scenarios, the mathematical models used to determine water quality require 2 parameters: The strong, moderate, and weak correlation between the parameters. By attempting to ascertain the

Table 3. Ideal and standard values of the water parameters

Description	Ideal Values (V_{ideal})	Standard Values ($V_{Standard}$)
pH	7	8.5
Hardness	75	300
Turbidity	1	5
COD	0	250
BOD	0	30
DO	5	7
TS	0	500
TDS	0	500
TSS	0	100

Table 4. Status of river water quality

Year	Average WQI by WAWQI	Status
2019	57.08	Poor water quality
2020	56.04	Poor water quality
2021	55.89	Poor water quality
2022	55.22	Poor water quality

type of link between the variables, this analysis offers a forecasting or prediction mechanism. The likelihood that the variables x and y will have a linear relationship is indicated whether the coefficient of correlation is nearer +1 or –1. To determine the connection between the factors of water quality examined in the water samples that were gathered, Pearson's correlation was used. The correlation between WQI and individual water quality parameters helps identify the important factors affecting water quality shown in Table 5. WQI and pH are very negatively correlated (r = -0.98987). According to this, the WQI tends to drastically drop when pH rises and vice versa. In this dataset, pH is probably a major determinant of WQI. DO has a moderate negative correlation with WQI (r = -0.25581). This indicates that higher DO levels are associated with better water quality (lower WQI). COD and BOD have strong negative correlations with WQI (COD: r = -0.47124, BOD: r = -0.7465), meaning higher COD and BOD levels are associated with poorer water quality. WQI and TS have a substantial negative connection (r = -0.6995), indicating that lower water quality is associated with higher total solids. TDS and WQI have a moderately

Table 5. The correlation matrix for the year 2019

	pH	DO	COD	BOD	TS	TDS	TSS	Hardne-ss	Turbidity	WQI
pH	1									
DO	0.24647	1								
COD	0.437783	0.627536	1							
BOD	0.746215	0.353551	0.793239	1						
TS	0.671916	0.015919	0.626034	0.843952	1					
TDS	0.534375	-0.32936	0.295891	0.401712	0.629903	1				
TSS	0.162451	0.400273	0.38602	0.516999	0.434198	-0.42613	1			
Hardne-ss	0.431878	0.965227	0.725415	0.536042	0.169751	-0.15706	0.379534	1		
Turbidi-ty	0.085308	-0.08477	0.165235	0.449799	0.552335	-0.17406	0.845294	-0.08643	1	
WQI	-0.98987	-0.25581	-0.47124	-0.7465	-0.6995	-0.57635	-0.14579	-0.44423	-0.0592	1

negative connection (r = -0.57635), suggesting that lower water quality is linked to higher TDS levels. TSS appears to have little direct influence on WQI in this dataset, as evidenced by the poor negative correlation between TSS and WQI (r = -0.1457). WQI and hardness have a negative association (r = -0.444), suggesting that hardness has little effect on water quality. Turbidity and WQI have a very weakly negative correlation (r = -0.059), indicating that there is hardly any direct association between the two. There are moderately favourable association between total solids and dissolved solids (r = 0.629903), hardness and DO (r =0.965227), turbidity and suspended solids (0.845294), total solids and BOD (0.843952), *etc*. The parameters that have the biggest effects on WQI in this dataset are pH, BOD, and TS. Strong relationships between PH, BOD, TS, and TDS imply that these factors are connected and could affect water quality as a whole.

A statistical method for predicting the connection between one dependent (Y) and independent variables is called Multiple Linear Regression (MLR). By adding more factors, it improves the predictive accuracy of basic linear regression. The regression variables are more informative and fit better when the regression coefficient is higher. The general MLR model is expressed as follows:

$$Y = \beta + \alpha_1 x_1 + \alpha_2 x_2 + ... \alpha_i x_i \quad (5)$$

Where
Y dependent variable
β Intercept
α_1, α_2, α_i Coefficient of the independent variable
x_1, x_2, x_i independent variable

The regression relationship and the regression coefficient (R^2), of linear regression were used. Multiple Linear Regression (MLR) was performed using Microsoft Excel's Data Analysis Toolpak to assess the connection between water quality indicators and Water Quality Index (WQI)[18-22]. The independent variables included PH, DO, BOD, COD, TS, TSS, TSS, turbidity and hardness. The significance levels were set at $p< 0.05$. Four MLR models were developed using water quality parameters according to their contribution and statistical importance to the Water Quality Index (WQI). These parameters were selected through preliminary correlation analysis. MLR equations were developed with various input parameters shown in Table 6.

Poor water quality was indicated by the weighted arithmetic method's calculation of the average original WQI range for 2019-2022, which was 55-57. In 2022, the average predicted WQI values derived from MLR Equations 1, 2, 3, and 4 were 44.675 (R^2=0.74, $p<0.05$), 47.896 (R^2=0.52, $p<0.05$), 47.03 (R^2=0.65, $p<0.05$), and 56.677 (R^2=0.74, $p<0.05$), in that order. The predicted WQI (56.67) from MLR Equation 4 was closer to the original WQI (55.22) than from MLR Equation 1 (44.675), even though both models showed the same explanatory power (R2 = 0.74). The accuracy of the regression models was evaluated using the Root Mean Square Error (RMSE). With 2 parameters, MLR[1] produced an RMSE of 10.54, which is a comparatively high departure from the reported WQI. In contrast, MLR[4], which incorporated all parameters, produced an RMSE of 1.46, demonstrating significantly better predictive performance (R^2 = 0.74) shown in Figure 4. The additional components in MLR Equation 4

Table 6. MLR models with equation

Model	R Square	Input Model Parameters	MLR Equation
MLR[1]	0.9799	PH,BOD	NEW WQI=141.3247-(13.012*PH)-(0.00458*BOD)
MLR[2]	0.9828	PH,BOD,TS	NEW WQI=141.313-(12.8855*PH)+(0.015392*BOD)-(0.00472*TS)
MLR[3]	0.9837	PH,BOD,TS,TDS	NEW WQI=139.8796-(12.6244*PH)+(0.007026*BOD)-(0.00273*TS)-(0.00212*TDS)
MLR[4]	0.9953	All parameters	NEW WQI=21.44113-(10.1126*PH)-(3.1507*DO)+(0.031593*COD)-(0.04183*BOD)-(0.95617*TS)+(0.940297*TDS)+(0.931183*TSS)+(0.926856*Hardness)+(0.926865*Turbidity)

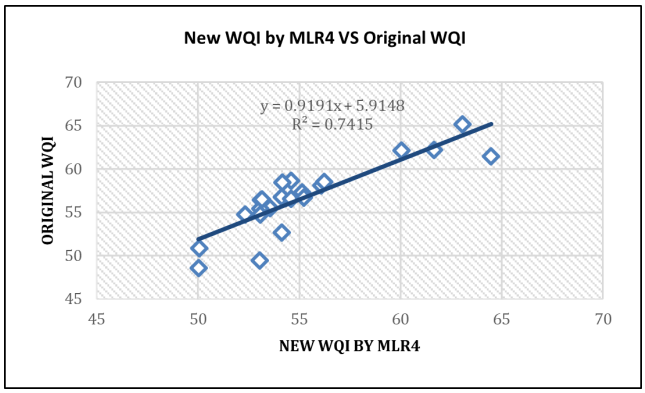


Figure 4. Linear regression of predicted WQI *vs* original WQI.

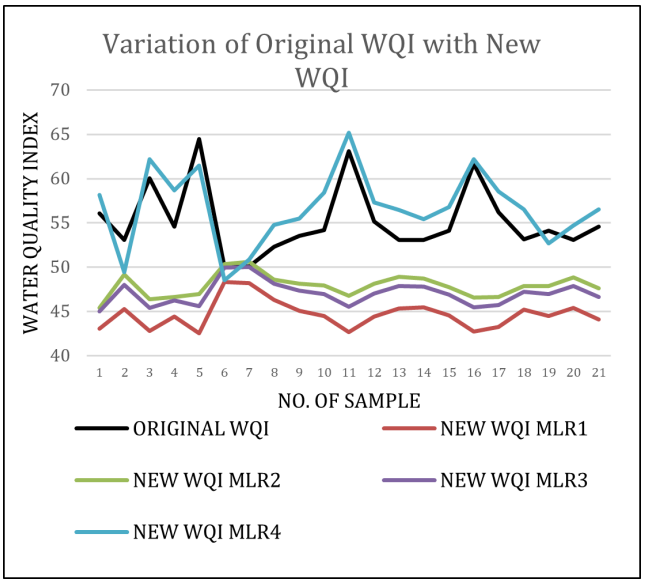


Figure 5. Variation of original and predicted values of water quality index of river Mutha.

(DO, COD, TS, TSS, TDS, Turbidity, and Hardness) all work together to produce a more accurate prediction, which emphasizes the significance of taking into account all pertinent characteristics while modelling WQI. But MLR Equation 4's slight overestimation indicates that the model might need to be further improved in order to better match the observed data. The variation of original and predicted values of WQI for 21 samples (year 2022) is shown in Figure 5.

6.0 Conclusion

The findings of this study reveal a mixed picture of the river's water quality. While certain parameters like pH and hardness fall within acceptable limits, others, such as Dissolved Oxygen (DO), Chemical Oxygen Demand (COD), Biological Oxygen Demand (BOD), and turbidity, raise serious concerns and require immediate attention. To ensure the protection of aquatic life and improve overall water quality, consistent monitoring, identifying sources of pollution, and applying effective control measures are essential steps moving forward. The Water Quality Index (WQI) proved to be a practical and insightful tool for evaluating the river water at the sampling site. According to the WQI results, the station exhibited poor water quality, underlining the need for urgent action. Statistical regression analysis also played a key role in this study, showing strong potential for accurately tracking and predicting water quality trends.

This research highlights how useful both the WQI and Multiple Linear Regression (MLR) models can be in understanding and forecasting the condition of the Mutha river. Importantly, the results emphasize the critical need for focused efforts to reduce pollution and restore the ecological balance of the river. Looking ahead, future studies should explore how long-term trends and seasonal changes influence water quality over time. Among the models tested, the MLR4 equation stood out for its strong performance, thanks to its inclusion of

all relevant parameters. On the other hand, the MLR1, MLR2, and MLR3 models tended to underestimate the WQI, suggesting that leaving out certain variables might compromise accuracy, even if the models are simpler to use. For real-world applications, especially when detailed data is available, the MLR4 model is the most reliable choice.

7.0 References

1. Achegave A, Rahate PB. Water quality monitoring of Mula- Mutha river in Pune region, Maharashtra, India. IERJ. 2017; 2(9):3731-3735.
2. American Public Health Association (APHA). Standard method for examination of water and wastewater. 23rd ed. AWWA; 2017.
3. Brown RM, McClelland NI, Deininger RA, Tozer RG. A water quality index-do we dare. Water and Sewage Work; 1970. p.117-110.
4. Cude CG. Oregon water quality index: A tool for evaluating water quality management effectiveness. J Am Water Res Assoc. 2001; 37(1):125-137. https://doi.org/10.1111/j.1752-1688.2001.tb05480.x
5. Chinmay D, More R, Shastri S. Effect of seasonal changes on water quality of river Mutha. Recent Trends in Civil Engineering. 2023; 274:707-715. https://doi.org/10.1007/978-981-19-4055-2_54
6. Fadtare VV, Mane TT. Studies on water pollution of Mula, Mutha and Pawana rivers in summer season in the Pune city region. Nature Environment and Pollution Technology. Technoscience Publications. 2007; 6(3):499-506.
7. Horton RK. An index number system for rating water quality. J Water Pollut Control Fed. 1965; 37:300-306.
8. Hussain M, Shivarajappa, Surinaidu L. Multiple linear regression model to predict the water quality parameters of Musi river near Hyderabad, India. IJCIET. 2022; 13(3):19-30. https://doi.org/10.17605/OSF.IO/UXBHD
9. More R, Shastri S. Assessment of water quality of river Mutha for onsite treatment of polluted river water. Advanced Modelling and Innovations in Water Resources Engineering. LNCE. Springer. 2021; 176:621-633.
10. Jagtap SS, Manivanan R. Water pollution status of Mula-Mutha rivers in Pune city: Review. IJTSRD. 2019; 4(1): 1080-1084.
11. Karmakar B, Singh MK. Assessment of water quality status of water bodies using water quality index and correlation analysis in and around industrial areas of west district, Tripura, India. Nat Environ Pollut Technol. 2021; 20(2):551-559. https://doi.org/10.46488/nept.2021.v20i02.011
12. Khatoon N, Khan AH, Rehman M, Pathak V. Correlation study for the assessment of water quality and its parameters of Ganga river, Kanpur, Uttar Pradesh, India. IOSR J Appl Chem. 2013; 5(3):80-90. https://doi.org/10.9790/5736-0538090
13. Kulkarni Eeshan, Shubham Aswale, Pratik Dhut, Atish Dhaygude, Sachin Shinde. Water quality assessment and mapping of parameters for Mutha river: A review. IJIRSET. 2018; 7(5).
14. Khatri N, Tyagi S, Rawtani D. Rural environment study for water from different sources in cluster of villages in Mehsana district of Gujarat. Environ Monit Assess. 2017; 190(10). https://doi.org/10.1007/s10661-017-6382-8
15. Hussein SH, Abed SA. Water quality index for Al-Gharraf river, Southern Iraq. Egypt J Aquat Res. 2017; 43(2):117-122. https://doi.org/10.1016/j.ejar.2017.03.001
16. Khatri N, Tyagi S, Rawtani D. Recent strategies for the removal of iron from water: A review. J Water Process Eng. 2017b; 19:291-304. https://doi.org/10.1016/j.jwpe.2017.08.015
17. Korde T, Dumbre N, Abhimane A, Bijwe S. Assessment for water quality of Mutha river using correlation. IJTSRD. 2020; 4(4):957-960.
18. Valentini1 M, Santos GBD, Vieira BM. Multiple Linear Regression analysis (MLR) applied for modeling a new WQI equation for monitoring the water quality of Mirim Lagoon, in the state of Rio Grande do Sul-Brazil. SN Appl Sci. 2021; 3(70). https://doi.org/10.1007/s42452-020-04005-1
19. Muthulakshmi L, Ramu A, Kannan N, Murugan A. Application of correlation and regression analysis in assessing ground water quality. Int J Chem Tech Res. 2013; 5(1):353-361
20. Priyadarshi N. Ganga river pollution in India- a brief report. Environment and Geology; 2009.
21. Shukla SK. The correlation and regression analysis of physiochemical water quality parameters of Wainganga river, India. National River Water Quality Laboratoty. Yamuna Basin Organisation. CWC. 2023.
22. Tyagi S, Rawtani D, Khatri N, Tharmavaram M. Strategies for nitrate removal from aqueous environment using nanotechnology: A review. J Water Process Eng. 2018; 21:84-95. https://doi.org/10.1016/j.jwpe.2017.12.005