"Analytical Study of Water Pollution Across Multiple Water Bodies"

Mohini Kulkarni[a], Anam Attar[a], Harshada Salve[a], Tukaram Navghare[a*]
[a]Department of Statistics, B.N. Bandodkar college of science, Thane,400601, Maharashtra, India.
E-mail : tvnavghare@vpmthane.org*(corresponding Author)*

Abstract:

Abstract Water is uniquely vulnerable to pollution. Known as a "universal solvent," water is able to dissolve more substances than any other liquid on earth. Living world depends on water in a variety of ways. This study works towards to evaluate the water quality of the water bodies (lakes, wetlands, tanks, and ponds) and identify the most influential parameters affecting them, and pinpoint polluted hotspots in India. Utilizing the dataset from the Central board pollution control spanning 2018-2022. Multiple linear regression model is applied to analyse the relationship between water quality index with various parameter including DO, BOD, pH etc. It is observe that fitted model give better perform R^2 value is closure to 1. Regression model achieved high accuracy with R^2 value ranging from 0.675 to 0.93. With DO, pH and conductivity emerging as significance predictor across all water body type. The predictive model could be integrated into real time monitoring system device to provide continuous update on water quality. Farmer can use predictive model to evaluate the safety of using tank or pond water for irrigation reducing crop damage and soil contamination. This study offers valuable insights for strengthening waste water management strategies and highlight the urgent need for targeted interventions in identified pollution hotspots.

Keywords: Multiple Linear Regression Model, Water Quality Index, Water pollution, Pollution Hotspot, Environmental Impact.

Introduction:

The term "Water Quality" includes the water column and the physical channel required to sustain aquatic life. The goal of the federal Clean Water Act, "To protect and maintain the chemical, physical and biological integrity of the nation's waters," establishes the importance of assessing both water quality and the habitat required for maintaining other aquatic organisms. The prominent source of surface water pollution is domestic sewage, industrial wastewater and agricultural run-off. So, we must turn to surface water quality studies in details. Application of fertilizers, pesticides, manure, and lime refuse dumps etc. is the main source of surface water and ground water pollution.

In this study, we tried to understand the impact of different water quality parameters on the Water Quality Index (WQI) is essential for accurately assessing the overall health of water bodies. By analysing these parameters, we can identify the key contributors to water quality deterioration, which is crucial for developing effective management strategies. This knowledge helps in monitoring water resources more efficiently, ensuring they meet safety standards for human consumption, agriculture, and ecosystem support. Ultimately, it enables informed decision-making for sustainable water resource management and environmental conservation. Then we find the pollution hotspots across India to prioritize the area who need urgent attention and remediation This helps in mitigating the negative

effects of pollution on **ecosystems, public health, and local communities.** Comprehensive assessment of global water resources, challenges, and management strategies**. The report highlights the** growing water crisis **due to** rapid urbanization, industrialization, climate change, and population growth**. It emphasizes that** inefficient water management, pollution, and lack of access to clean water disproportionately affect developing countries**, leading to** health crises, food insecurity, and economic instability.**[6]**
The APHA (2005) Standard Methods for the Examination of Water and Wastewater (21st Edition) **is a** comprehensive guide **developed by the** American Public Health Association (APHA), American Water Works Association (AWWA), and Water Environment Federation (WEF)**. This publication serves as a** global benchmark for water quality testing and analysis**, providing** standardized methodologies **for assessing** physical, chemical, and biological parameters **of water and wastewater.[3]**

Related Work:

Many research have been conduct on water quality across multiple water bodies. Kapani and V.et.al [8] examine water pollution in a specific lake in Karnataka, India, and assess its ecological impact. Seasonal variations significantly influence water quality, with monsoons worsening contamination. as it provides a regional case study demonstrating how lakes are more vulnerable to organic and microbial pollution. Rahman and C.et.al [13], research examines the seasonal variations and pollution effects on the water quality of the River Indus using Multiple Linear Regression (MLR) analysis. Multiple linear regression analysis successfully predicted the influence of multiple factors on overall water quality, providing insights into the dynamics of seasonal pollution. Pathak and Limaye's [12] findings support the application of statistical correlation and regression models in water quality studies and emphasize the necessity of sustainable pollution mitigation strategies. The study serves as a foundation for further research on the impact of human activities on river ecosystems and the development of predictive models for water pollution management.
Mulla .Fet.al [10] reinforces the significance of statistical model in water quality assessment by establishing correlations and regression equations among key physicochemical parameters. The findings demonstrate that parameters such as BOD, COD, TDS, and electrical conductivity are strongly interrelated, with pollution levels having a direct impact on dissolved oxygen and overall water quality. The application of multiple regression models provides a valuable predictive tool for monitoring pollution trends and identifying critical contamination sources. Agarwal [2] conducted a statistical analysis of water quality parameters using linear regression and correlation. This study finding indicated a strong positive correlation between BOD and COD, confirming that organic pollutants significantly contribute to oxygen demand in the river. Additionally, TDS and EC exhibited a high correlation, suggesting contamination from industrial effluents, agricultural runoff, and domestic waste study emphasized that statistical techniques provide valuable insights into pollution patterns, making them essential for monitoring water quality.
Dash . P.el.al [5] conducted a groundwater quality assessment in rural areas around the

industrial zone, demonstrating significant correlations between TDS, hardness, and chloride levels, indicating contamination from industrial and agricultural activities. Their regression analysis provided predictive insights into groundwater pollution trends, emphasizing the need for continuous monitoring and stricter pollution control policies .
Simonov et al.[14] conducted a comprehensive environmental model study to analyse and interpret river water monitoring data using advanced statistical and chemo metric techniques. Finding highlight the effectiveness of environmental model in identifying key pollution contributors. advanced statistical techniques into environmental monitoring programs allows for better decision-making in water resource management and pollution control strategies.
Kumar et.al [9] found a strong positive correlation between TDS and EC, indicating that higher dissolved solids contribute to increased conductivity, which can be a potential indicator of contamination. findings reinforced the importance of statistical correlation analysis in predicting water quality trends and developing effective monitoring systems for drinking water sources.
Agarwal S.et.al [1] their analysis highlighted that industrial effluents and domestic wastewater contribute to rising pollution levels, affecting the river's self-purification capacity. By applying linear regression models, the study successfully forecasted pollution trends, reinforcing the importance of statistical tools in environmental monitoring.

Methodology:

In this paper we use multiple linear regression like log transformation model this model are used to predict the unknown value and to identify the significant parameter. According to our objective we wanted to study that which parameters are highly affecting on WQI that's why we used multiple linear regression model to know the value in numeric form. Multiple regression model is use to predict the dependent variable($\boldsymbol{y_i}$)as water quality index. Independent variable as, Dissolve oxygen(x_1), pH(x_2),conductivity(x_3), BOD(x_4)., nitrite(x_5),faecal coliform(x_6)

$$\boldsymbol{y_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6}$$

Where; $\boldsymbol{y_i}$=Dependentvariable

- $\boldsymbol{x_i}$=Independentvariable (i=1,2,3,4,5,6)

- $\boldsymbol{\beta_0}$=y-intercept (constant term).

- Linearity the relationship between the independent variables and the dependent variable should be linear. Homoscedasticity the residuals should have constant variance at all levels of the independent variables. Multicollinearity can be detected using Variance Inflation Factor (VIF) .Independence the residuals should be independent of each other. Normality of residualthe residuals should be approximately normally distributed.

Figure 1.Shows the overall method of analysis right from deciding objective till the analysis and it's conclusion.

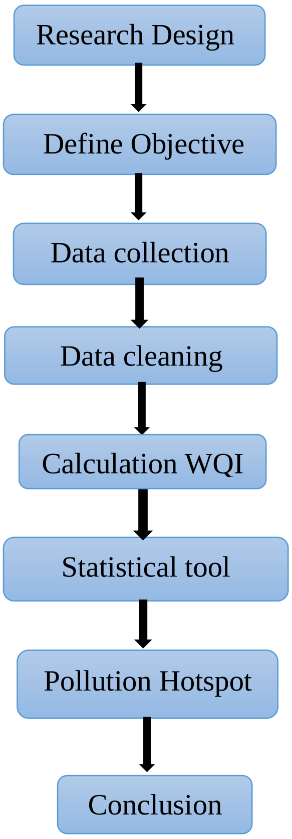

Fig 1.Flow chart of the study

Table1.Parameter and their standard values

Parameter	Standard ideal value (Sn)
pH	8.5
Dissolve oxygen	5
Nitrite	50
Faecal Coliform	200
Conductivity	300
Biological oxygen demand	3

Water Quality Index calculation of WQI was made using weighed Arithmetic index method [7]**(Brown el al 1972)**

$$W_n = \frac{K}{S_n} \quad n = 1, \ldots 6$$

Where; Wn: Unit weight of parameter.

Sn : Standard value of parameter .

K: Proportionality constant.

Water quality indexis calculate by the following equation:

$$WQI = \frac{\sum(W_n \times Q_n)}{\sum W_n}$$

Table 2. Rating of WQI

Range	Quality of water
0-25	Excellent
26-50	Good
51-75	Bad
76-100	Very bad
100 & above	Unfit

Dataset and software:

We have collected data from site collected from the **CBCP (Central Board pollution control)** website for the period 2018–2022 https://www.cpcb.nic.in . . It contain 1627observation. and 6 variable such as Dissolved Oxygen (DO),Biological Oxygen Demand (BOD),pH levels, Nitrite ,concentration, Total Coliform bacteria of water bodies(lakes, ponds, wetlands, and tanks).

This analysis done by software like Python ,PowerBI , Ms Excel , tableau .Python library and packages namely scikit-learn provides machine learning-based regression models with built-in optimization. Stats models offers detailed statistical insights, and matplotlib & seaborn help visualize regression result.

Result and discussion:

In this section we have analyse the finding .We use multiple linear regression model here we uncovered significant correlation between pollution parameter and water quality. Demonstrating pollution directly affects dissolve oxygen and water quality.

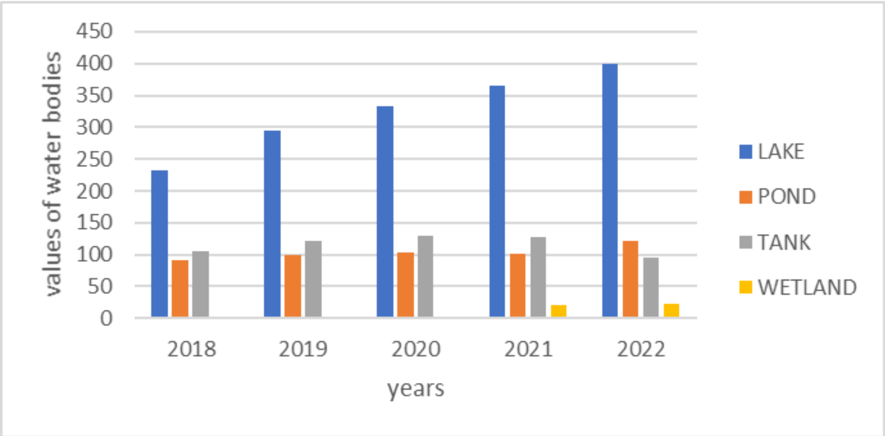

Fig.2 Water bodies count across the year

Figure 2 shows water body counts across the year in 2022 lake contain highest count.

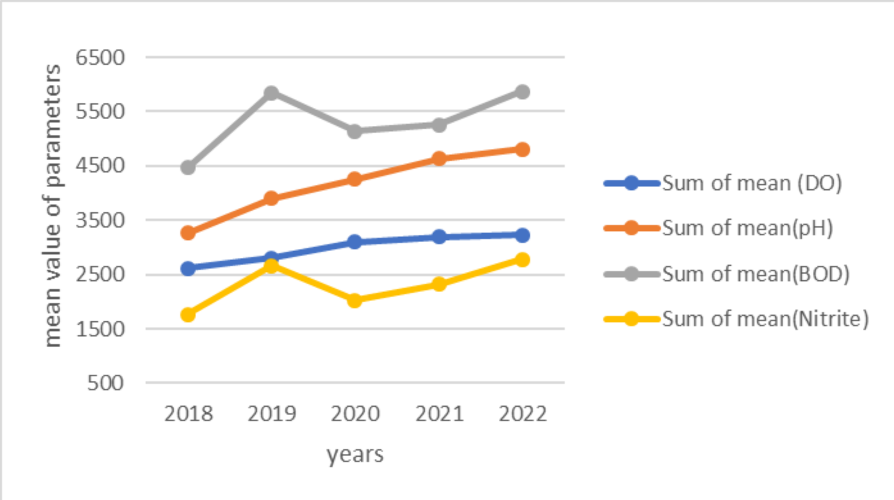

Fig.3.Variation in parameter across the year

Figure 3 .Shows trend analysis, we observe that DO have increasing line in specific limit than others but BOD have several changes across the years. PH also increasing line in years which can be harmful by future's perspective

Water Body	R^2	MLR Model	Significant Predictors
Lake	0.857	$y = 5.2518 + 0.1904(x_1) - 0.0819(x_2) + 0.2036(x_3) + 2.0183(x_6)$	Dissolved oxygen, pH, Conductivity, coliform
Pond	0.675	$y = 4.4156 + 0.2245(x_1) - 0.1144(x_2) + 0.2702(x_3) - 0.1096(x_4) + 0.8428(x_6)$	Dissolved oxygen, pH, Conductivity, BOD, coliform
Tank	0.899	$y = 5.6979 + 0.2322(x_1) - 0.1554(x_2) + 0.1664(x_4) + 0.2574(x_5) + 2.1211x_6$	Dissolved oxygen, pH, BOD, Nitrite coliform,
Wetland	0.93	$y = 4.4021 + 0.0761(x_1) + 0.4654(x_2)$	Dissolved oxygen, Conductivity.

Table 3. Multiple linear regression model table

Table 3 shows that ,For lake R^2 =0.857 strong predicting model as average DO level at lake were below 5mg/l indicating poor oxygenation BOD exceeded 3mg/l , suggesting high organic pollution .R^2value given in the table presenting the good fit model. Significant predictors are the parameters which are highly affecting the WQI with respect to their water body. In MLR analysis, WQI is influenced by independent variables differently across water bodies. In lakes, a unit increase in DO, conductivity, and coliform raises WQI by 0.1904, 0.2036, and 2.0183, while a unit decrease in pH lowers it by 0.0819. In ponds, DO, conductivity, and coliform increase WQI by 0.2245, 0.2702, and 0.8428, whereas a unit drop in pH and BOD decreases it by 0.1144 and 0.1096. In tanks, DO, BOD, nitrite, and coliform raise WQI by 0.2322, 0.1664, 0.2574, and 2.1211, while pH reduction lowers it by 0.1554. In wetlands, a unit rise in DO and conductivity increases WQI by 0.0761 and 0.4654, respectively.

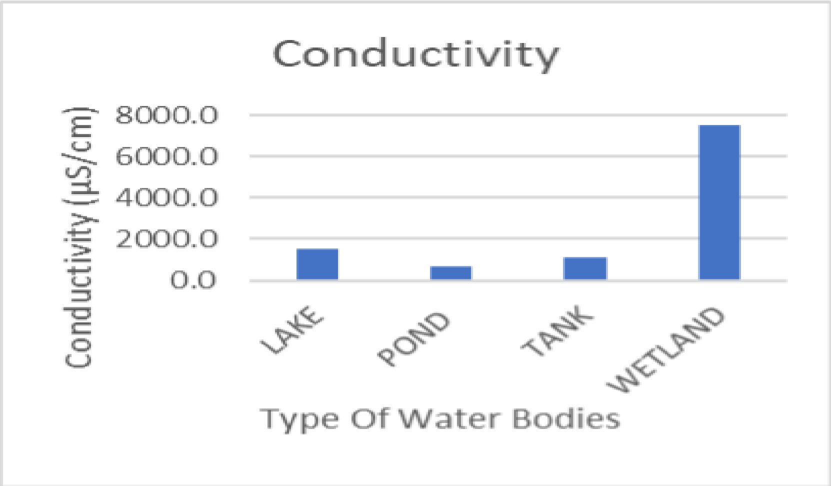

Fig 4. Conductivity in water bodies

Figure 4. Wetland have significantly higher conductivity compared to other. Conductivity of wetland is 7547µs/cm. May have higher concentration of dissolved ion which could result of natural process or agricultural runoff.

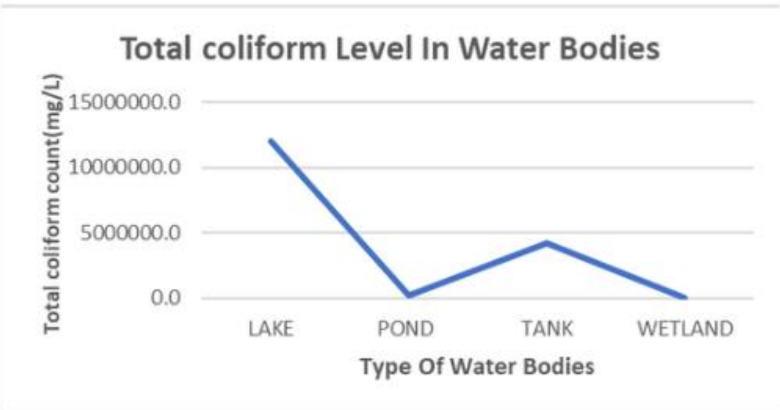

Fig 5. Coliform level in water bodies

Figure 5. The lake are receiving the most organic and microbiological pollution possibly from untreated sewage or urban waste. Serious threat to public and aquatic ecosystem.

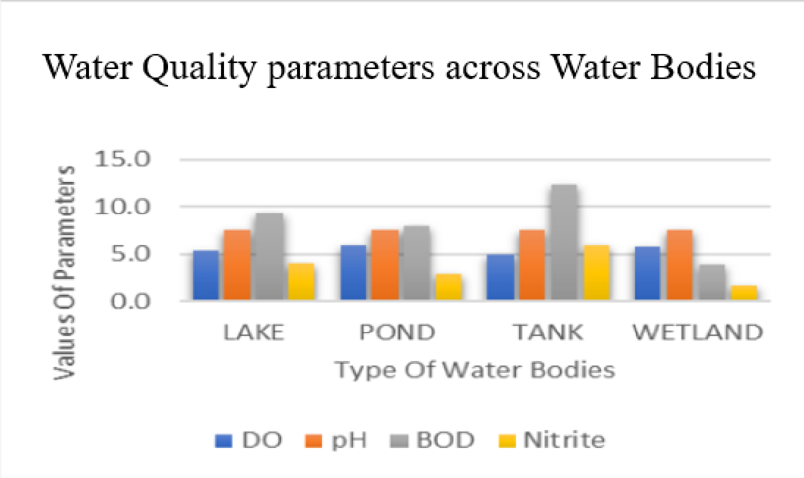

Fig 6. Water Quality parameter across water bodies.

Figure 6. Lake show low DO level and higher BOD which indicates organic pollution. Wetland demonstrates better natural filtration. Lake and Tank are more suspectable to contamination of human activities. Tank and lake have high BOD i.e. 12.4mg/l, 9.3mg/l respectively

Fig 7. Pollution hotspots

Figure7. Shows the critical water pollution levels, marked as red region indicating high organic matter content and oxygen depletion, harmful for aquatic ecosystems.The hotspots cover parts of southern, central, and eastern India**.**Low pH (<7) Suggestacidic water conditions, potentially due to industrial effluents or acid rain.High BOD (>6)Indicates excessive organic pollutants, likely from untreated sewage, agricultural runoff, or industrial waste.Low DO (<2) Reflects oxygen depletion, threatening aquatic life, often caused by eutrophication or organic waste decomposition.

Conclusion:

Correlations are established between water parameters and Water Quality using regression equation and these equations can be used to predict the water Quality index across multiple water bodies.We have tested the accuracy rate of models as which emphasize reliability of model. For lake, tank and wetland it is85.7%, 89.9% and 93% respectively. So the model can be applied in general to any water body.So that some preventive action can be taken before the detailed investigation and water pollution level can be controlled to a certain extent.The hotspots in the Fig.7 cover parts of southern, central, and eastern India, including **Karnataka, Andhra Pradesh, Telangana, Tamil Nadu, and Kerala**. The highlighted portion shows the critical water pollution trends across different water bodies. The MLR model presented in the paper can be applied for predicting water quality index given the independent variables. Wetlands, despite their higher conductivity, act as natural filters, while lakes and tanks face severe pollution challenges due to high organic and microbial contamination. These findings reinforce the necessity for **enhanced wastewater management, stricter regulations, and conservation of natural filtration ecosystems.**

Reference:

1. Agarwal, A., &Saxena, M. (2011). Assessment of pollution by physicochemical water parameters using regression analysis: A case study of Gaganriver at Moradabad-India. Advances in Applied Science Research, 2(2), 185-189.
2. Agarwal, M., & Agarwal, A. (2013). Linear regression and correlation analysis of water quality parameters: A case study of River Kosi at district Rampur, India. International Journal of Innovative Research in Science, Engineering and Technology, 2(12), 7273-7279.
3. American Public Health Association. (1926). Standard methods for the examination of water and wastewater (Vol. 6). American Public Health Association.
4. Dash, T. R., Dash, P. C., &Patra, H. K. (2006). A correlation and regression study on the ground water quality in rural areas around Angul-Talcher industrial zone. Indian Journal of Environmental Protection, 26(6), 550.
5. Hiwasaki, L. (2011). ‘Water for life’… water for whose life? Water, cultural diversity and sustainable development in the United Nations. In *Water, cultural diversity, and global environmental change: Emerging trends, sustainable futures?* (pp. 509-531). Dordrecht: Springer Netherlands.https://doi.org/10.1007/978-94-007-1774-9_34
6. Kapani, K., Charantimath, N. V., Chikkanarayanaswamy, P., & Jayaramaiah, U. (2024). Assessing the water quality of Vatadahosahalli Lake in Chikkaballapura district, Karnataka, India. HydroResearchDOI:10.1016/j.hydres.2024.06.001
7. Kumar, N., & Sinha, D. K. (2010). Drinking water quality management through correlation studies among various physicochemical parameters: A case study. International journal of environmental sciences, 1(2), 253-259
8. Mulla, J. G., Farooqui, M., Zaheer, A. 2007. A correlation and regression equations among water quality parameters. Int. J. chem. sci., 5(2): 943-952.
9. Palit, Debnath&Mondal, Saikat&Chattopadhyay, Pinaki. (2019). Analyzing Water Quality Index of Selected Pit-Lakes of Raniganj Coal Field Area, India. 1167-1175.
10. Pathak, H. E. M. A. N. T., &Limaye, S. (2011). Interdependencey between Physicochemical Water Pollution Indicators: A Case Study of River Babus, Sagar, Mp, India. *Annals of the University of Oradea, Geography Series/AnaleleUniversitatii din Oradea, SeriaGeografie*, *21*(1), 23-29.
11. Rahman, A., & Chughtai, M. O. (2014). Reginol interpretation of river Indus water quality data using regression model. African Journal of Environmental Science and

Technology, 8(1), 86-90DOI: 10.5897/AJEST2013.1610

12. PRADHAN, S. J. (May 2015). LINEAR REGRESSION AND CORRELATION ANALYSIS OF WATERQUALITY PARAMETERS OF DAYA CANAL AT BHUBANESWAR,ODISHA. *ResearchGate*
13. Simeonov, V., Einax, J. W., Stanimirova, I., & Kraft, J. (2002). Environmetricmodeling and interpretation of river water monitoring data. Analytical and bioanalytical chemistry, 374, 898-905.https://doi.org/10.1007/s00216-002-1559-5

~*~*~*~*~*~*~