Volume 2, Issue 3

Innovation Series Quarterly

https://journal.academicapress.org/aps/index.php/isq

Research Article

Determinants of Water Quality Degradation in the Langat River Basin Using Ordinal Logistic Modelling

Nur Faiqah Md Zoolkepli[1], Haslinda Ab Malek[2*], Siti Sarah Januri[3], Noryanti Nasir[4], S Sharifah Radiah Shariff[5] and Shahrul Nizam Ahmad[6]

1 Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA Negeri Sembilan Branch, Seremban Campus, 70300 Seremban, Negeri Sembilan, Malaysia; 2024977791@student.uitm.edu.my

2 Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA Negeri Sembilan Branch, Seremban Campus, 70300 Seremban, Negeri Sembilan, Malaysia; haslinda8311@uitm.edu.my

3 Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA Negeri Sembilan Branch, Seremban Campus, 70300 Seremban, Negeri Sembilan, Malaysia; sarahjanuri@uitm.edu.my

4 Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia; noryantinasir@uitm.edu.my

5 Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia; shari990@uitm.edu.my

6 Faculty of Applied Sciences, Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia; shahruln@uitm.edu.my

* Correspondence: haslinda8311@uitm.edu.my; 06-6342458.

Keywords:
water pollution
water quality
environmental health
ordinal logistic regression
Langat river basin

Abstract: *Water pollution refers to the contamination of water bodies by harmful substances, which adversely affects water quality and poses risks to aquatic ecosystems and human health. This study aims to identify the most significant factors influencing water quality in Langat River Basin using ordinal logistic regression. Notably, chemical oxygen demand (COD) emerged as the most critical factor (p-value of < 0.001), indicating a strong correlation with polluted water conditions. COD levels signify increased pollutants in the river, leading to degraded water quality. It is hoped that this study will enhance public awareness regarding water pollution on environmental and human health.*

1. INTRODUCTION

Water is one of the most valuable resources that is used every day to drink and keep clean. It is also used in farming, making energy and balancing the ecosystem. Public health, economic growth and life quality are all closely tied to having safe, clean and easy to access water. On the contrary, the global concern of water pollution is rapidly increasing. Waste from industries, agriculture, and especially households significantly pollute rivers, lakes and groundwater with domestic waste accounting for nearly 80% of water pollution (Norrrahim et al., 2021).

In Malaysia, chemical pollution in rivers is a common problem. Both human health and the ecosystem are seriously threatened by this pollution. The same pollution problems affect the Langat River Basin which flows through the heavily populated and quickly growing state of Selangor. The river originates in the hilly areas of Hulu Langat with the Langat River Basin covering the states of Selangor and Negeri Sembilan as well as the federal territories of Kuala Lumpur and Putrajaya, before flowing into the Straits of Malacca.

For many years, Malaysia has been monitoring and controlling the quality of river water. Since 1978, the Department of Environment (DOE) has been keeping an eye on the quality of the water, primarily to establish guidelines for identifying changes in the water and identifying sources of pollution. The IKA-JAS (Indeks Kualiti Air – Jabatan Alam Sekitar) system is currently used in Malaysia to monitor water quality. IKA-JAS evaluates pollution levels and classifies water quality based on the National Water Quality Standard and the water's intended use (Shamsuddin et al., 2022).

Water pollution was a serious environmental problem in the Langat River Basin. However, there was a lack of research that treated water quality classification as an ordinal variable. This could lead to less informative or potentially incorrect conclusions regarding the relative severity of water quality conditions and the influence of water quality parameters across different classes. Although water quality data is plentiful, the combined influence and relative significance of the factors involved in determining the quality of water are often unclear. This problem is especially prominent in river basins where the interaction of two or more pollutants makes it challenging to identify the dominating factors that define the level of water quality classification.

Therefore, to overcome the limitation of the above method, this research utilizes the ordinal logistic regression model to determine the most significant factors that influencing water quality through the ordered categories of water quality defined by DOE, which are Class I (natural environment), Class II (treatment required) and Class III (Extensive treatment required). In this way, it will be easier to interpret the results and assess the impact of the parameters on the water quality classifications. The results of the research are hoped to improve the understanding of water quality changes especially in the Langat River Basin.

2. LITERATURE REVIEW

The class of the water quality is usually determined based on the application of the Water Quality Index (WQI) where several physical, chemical and biological factors are combined in a single index parameter that reflects the overall quality of water for certain uses (Tyagi et al., 2013). In Malaysia, the DOE through a standardized WQI, measures the quality of rivers based on five important factors, which are the Biochemical Oxygen Demand (BOD), Chemical Oxygen Demand (COD), Ammonia Nitrogen (NH_3-N), Suspended Solids (SS) and Potential of Hydrogen (pH). Najah et al. (2021) proved the applicability of the WQI in recognizing the variation of water quality classes during the COVID-19 Movement Control Order (MCO) in Malaysia.

BOD expresses an important measure of pollution in water. The higher the BOD, the lower the DO level in water, making it harmful to marine life. BOD also expresses an indicator of the level of DO in milligrams needed in decomposing waste matter in water, making it an essential measure for ecological conditions (Adnan et al., 2023). A study was conducted in Malaysia by Ibrahim et al. (2023) utilizing a method of binary logistic regression to differentiate between acceptable and poor water quality in the context of the Selangor River. BOD is found to be most significantly influential factor at $p < 0.05$ and had a higher-ranking order as compared to all other variables. It indicates a strong impact within the logistic model. This further justifies the important role that BOD plays in monitoring, classification and management related to water quality in different kinds of aquatic ecosystems.

COD is considered a universal approach for estimating the level of organic pollution present in water bodies. According to Han et al. (2022), high COD levels can be a major reason for the low concentration of dissolved oxygen in water bodies. Similarly, Liu et al. (2020) stated that the quality of water can be estimated using the values of COD and BOD. Many of the linear regression models have illustrated the significant prediction capabilities of COD on different parameters of water quality at a p-value of less than 0.05 (Qasaimeh & Al-Ghazawi, 2020). These findings draw on the significance of COD in water quality studies. However, its importance varies depending on the modelling approach used.

SS is a critical parameter in water quality monitoring and is closely associated with ecological health. A large value of SS results in reduced water clarity, increased biological and chemical oxygen demand and ecological risk associated with habitat damage and increased fish deaths (Chen et al., 2023). Studies have shown the statistical significance of SS in water quality classification using logistic regression analysis. Ibrahim et al. (2023) carried out binary logistic regression analysis to determine the classification of the quality of the river waters of Selangor and found that SS was a highly significant factor exceeding other variables such as COD. There was an excellent fit and accuracy of 96% in the classification, establishing the significance of SS in determining the level of pollution. These results define SS as a physical pollutant and its importance as a marker in predictive models.

NH_3-N is an important indicator used to signify water pollution because of its toxicity to aquatic life. Pang et al. (2022) found that the use of fertilisers in agriculture, industry water discharge and river runoff all had an impact on the variation in the ammonia nitrogen concentration level in rivers. Furthermore, binary logistic regression analysis has determined that NH3-N is a highly significant variable in the classification of water quality. NH3-N is one of the most important variables in the model assessing the potability of water. This is because a relevant relationship has been established between the level of water safety and the concentration of NH3-N (Shi, 2024). According to this study, NH3-N is a significant factor that positively affects water potability as a potential predictive variable.

The pH value is one of the crucial indicators in the measurement of water quality because it determines the chemical composition, nutritional availability and metal solubility. Dewangan et al. (2023) further supported that pH values regulate the solubility of oxygen in water. Low pH values might reduce the concentration of oxygen in the water, resulting in the deterioration of aquatic life. Although pH is often included due to high correlation with other factors such as DO or BOD, it is a common parameter in calculating the WQI (Malek et al., 2021). Thus, while the role of pH is essential during water quality measurements or treatment, research surrounding its utilization as a predictive variable within regression models is insufficiently explored.

The implementation of ordinal logistic regression in recent water quality studies seems to be limited, as numerous studies now prefer machine learning and advanced computational methodologies (Evsyukov et al., 2024). Although of ordinal logistic regression offers a straightforward and interpretable structure that is well-suited for ordered categories of water quality, its adoption in current water modelling practices has been sparse. This gap shows that there may be a chance to use of ordinal logistic regression again in environmental research to categorize water quality levels according to predictors like BOD, COD, NH_3-N, SS and pH.

3. METHODOLOGY

The method that was employed in the investigation of the factors that affect the quality of the water in the Langat River Basin were discussed. The relationship between chosen parameters and class of water quality were explored.

3.1 Sources of Data

The data used in this study is secondary data gathered from the Malaysian Department of Environment from 2023 to 2024. A total of 312 samples were obtained which includes necessary parameters such as BOD, COD, NH_3-N, SS and pH. The data used in this study has been reflected over different points and helps in giving different classifications of water quality based on the Malaysian Water Quality Index.

3.2 Study Area

This study used data obtained from the Langat River Basin. The Langat River Basin is approximately 2350 km^2 in size and extends over diverse regions in the state of Selangor, namely Hulu Langat, Sepang and Kuala Langat and then discharges into the Straits of Melaka (Ahmed et al., 2020). In addition, it also extends over Negeri Sembilan. The Langat River Basin is a vital water source for almost one-third of people in Selangor, thus accentuating its importance as a water source for both domestic and industrial consumption. Notably, despite the vital and important role played by the basin, it normally faces pollution problems (Ahmed et al., 2022). The pollutants normally disrupt the activities of water treatment companies established in the Langat River Basin and thus raise important concerns about water management in the region.

3.3 Description of Variables

The data utilized in this research consists of water quality indicators gathered from the Langat River Basin. These indicators act as predictor variables to evaluate their effect on the overall water quality, which is regarded as an ordinal outcome variable according to recognized WQI classifications. Table 1 presents a summary of the variables incorporated in the analysis, including their scale of measurement and descriptions.

Table 1. Description of Variables

Variables	Scale of measurement	Description
Biochemical Oxygen Demand	Ratio	The amount of oxygen required by microorganisms to decompose organic matter in water.
Chemical Oxygen Demand	Ratio	The total quantity of oxygen needed to oxidize both organic and inorganic substances in water.
Suspended Solids	Ratio	Solid particles suspended in water that can affect clarity and water quality.
pH	Interval	A measure of acidity or alkalinity on a scale from 0 to 14
Ammonia Nitrogen	Ratio	Concentration of ammonia in water, typically from agricultural runoff or sewage discharge.
Water Quality Class	Ordinal	Water quality classification based on WQI levels

The dependent variable in this analysis is the water quality class, as defined by the DOE of Malaysia. Table 2 show the water quality data were classified into ordered categories namely Class I, Class II and Class III. It is based on the DOE water quality classification system. These classes represent decreasing levels of water quality, ranging from very good water quality (Class I) to polluted water (Class III).

Table 2. Water Quality Index (WQI) Classification

Class	**WQI Range**	**Description**
Class I	< 92.7	Conservation of natural environment
Class II	76.5 – 92.7	Conventional treatment required
Class III	51.9 – 76.5	Extensive treatment required

(Source: Department of Environment, 2022)

3.4 Theoretical Framework

Water quality levels were an overall measure determined by several essential physicochemical factors that indicate the condition of a water body. In this research, water quality acts as the dependent variable and was affected by five primary factors which are BOD, COD, NH_3-N, SS and pH. Figure 1 below demonstrates the connections among the variables.

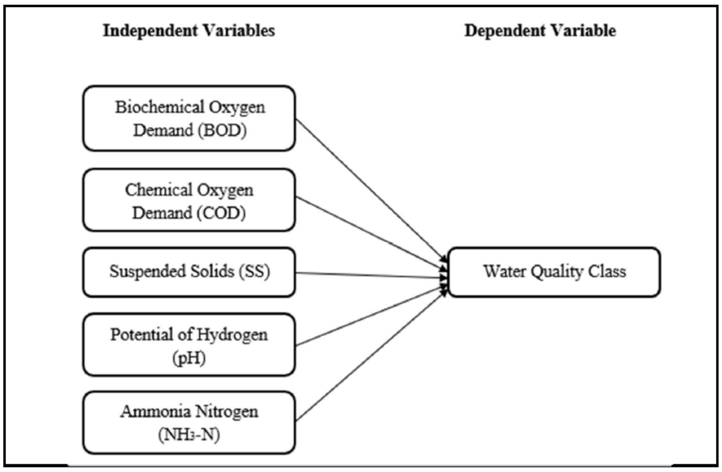

Figure 1. Theoretical Framework.

3.5 Ordinal Logistic Regression Analysis

Ordinal Logistic Regression is an often-used technique for analyzing an ordinal dependent variable. The categorical variable with an ordered category but an unknown distance between categories. It is especially valuable when the outcome of interest is ranked, such as various levels of water quality. Ordinal logistic regression focuses on the log-odds of the cumulative probabilities of being in or below a specific category of the dependent variable. This model consists of one ordinal dependent variable and several independent variables, which may be either continuous or categorical.

The general form of the cumulative logit model for ordinal logistic regression is given by:

$$\log\left(\frac{P(Y \leq j)}{P(Y > i)}\right) = \theta_j - (\beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_5 x_5) \quad (1)$$

Where:

Y is water quality class

x_1 is BOD

x_2 is COD

x_3 is NH3-N

x_4 is SS

x_5 is pH

β_1, β_2, … , β_5 are the regression coefficients representing the effect of each water quality parameter on the water quality class.

θ_j is the threshold (intercept) for the j^{th} cumulative logit, representing the cut-off point between adjacent water quality categories

$log\left(\frac{P(Y\leq j)}{P(Y>i)}\right)$ is the log-odds of being in category j or below versus above

3.5.1 Model Evaluation

The goodness of effectiveness and appropriateness of the models used were tested by different evaluation techniques. Model adequacy could be tested with the help of the Likelihood Ratio Test, the Goodness of Fit Test and the Pseudo R-Squared Test.

3.5.2 Parameter Estimates

In the case of the ordinal logistic regression model, the parameter estimates indicated the strength of the relationship for each predictor variable (BOD, COD, NH3-N, SS and Ph) with regard to the odds or chances of being in a higher or lower category for the dependent variable (Water Quality Class). Positive parameter estimates indicated that with an increase in the predictor variable, the chances are high to be in a higher category for the dependent variable. The negative parameter estimate indicated that with the increase in the value of the predictor variable, the chances are high to be in the lower category for the dependent variable. Each parameter estimate was followed by the standard error value, the Wald Statistic value and the p-value that aided in understanding the significance of the variable regarding the model. If the parameter estimates were significance at a level below 0.05, then it indicated that the predictor variable influenced the model.

4. FINDINGS

In this section, the findings of the study and the results are briefly discussed. The ordinal logistic regression was applied to achieve the research objective.

4.1 Ordinal Logistic Regression

To fulfil the necessary assumptions, the dependent variable must initially be evaluated at the ordinal level. Ordinal data refers to variables with two or more categories that can be arranged in a significant order or ranking. This study concentrated on evaluating water quality conditions, so the water quality which is the dependent variable was organized in an ordered sequence. The categories employed included Class I, Class II and Class III. This ordering enables the analysis to represent the transition from no treatment required to treatment required, thus making it appropriate for ordinal logistic regression. Moreover, the assumption of no multicollinearity among the independent variables was fulfilled in this study. Each independent variables which are BOD, COD, NH_3-N, SS and pH has a consistent effect on water quality across all classes. Moreover, the Proportional odds assumption of ordinal logistic regression is also met.

4.1.1 Fitness of Model

The fitness of the model refers to the ordinal logistic regression model's capacity to accurately describe the observed data. It presents the model fitting and goodness-of-fit measures used to determine if the proposed model is suitable for the dataset. The significant values of the model fitting and Deviance tests are used to determine the adequacy of the model.

Table 3. Fitness of The Model

Model	Sig
Model Fitting	< 0.001
Deviance	1.000

Table 3 reveals a p-value ($p < 0.0001$) of model fitting less than the significant value (0.05). It indicates that the ordinal logistic regression model fits the data better than the null model. This shows that the included water quality parameters significantly improve the model. The Deviance test yields a p-value of 1.000 which is above 0.05. The non-significant Deviance result means that the model predictions closely match the observed data. It implies that the model fits the data well and can be reliably used for analysis. As a result, the model can be considered to have an acceptable goodness of fit.

4.1.2 Pseudo R-Square

Pseudo R-Square values are used in ordinal logistic regression to estimate the proportion of variation in the dependent variable that can be explained by independent variables.

Table 4. Pseudo R-Square

Cox and Snell	.748
Nagelkerke	.878
McFadden	.721

According to Table 4, the Pseudo R-Square values show that the ordinal logistic regression model has a high explanatory power. The Cox and Snell R-Squared value of 0.748 indicates that the independent factors account for a significant amount of the variation in water quality classification. Similarly, the Nagelkerke R-Square value of 0.878 indicates a high level of explained variation. It implies that the model is effective in predicting water quality categories. The McFadden R-Square value is 0.721. It is considered suggestive of a solid model fit. These results show that the fitted model explains a large proportion of the variability in water quality and is highly suitable for the analysis.

4.1.3 Parameter Estimates

Parameter estimates of the ordinal logistic regression model show the link between water quality parameters and water quality classification. The estimates and their significance values are used to determine the direction and strength of each independent variable's influence on the ordinal dependent variable. Table 5 shows that Class=0 have p-value = 0.002 and Class=1 have p-value = 0.017. Both ordinal categories are statistically significant, indicating that the classification is valid. Therefore, the response variable which is water quality classification is meaningful and appropriate for the analysis.

Table 5. Parameter Estimates

		Estimate	**Sig.**
Threshold	CLASS=0	-47.499	.002
	CLASS=1	-35.932	.017
Location	BOD	-.163	<.001
	COD	-.192	<.001
	NH3-N	-.120	<.001
	SS	-.109	<.001
	pH	.051	.723

Table 5 shows BOD (p-value = <0.001), COD (p-value = <0.001), NH3-N (p-value = <0.001) and SS (p-value = <0.001) were statistically significant toward water quality. These are because their significance values were less than the significance level of 0.05. However, pH (p-value = 0.723) was not a significant predictor for this study because the p-value is greater than 0.05. It can be assumed that pH does not affect the water quality. This finding is consistent with previous studies done by Salvador et al. (2024) and Hashim et al. (2023) which also reported that pH was not a significant factor in influencing water quality classification. Therefore, pH does not contribute significantly to variations in water quality compared to other parameters.

5. DISCUSSION

BOD is an index of the level of biodegradable organic materials present in domestic sewage, agricultural runoff and industrial effluents. It is a significant predictor of water quality. One unit increase in BOD is associated with a decrease in the log odds of being in a good water quality class versus a poorer water quality class by 0.163. This indicates that higher BOD levels increase the likelihood of deterioration in water quality, as elevated BOD reflects higher levels of organic pollution that reduce dissolved oxygen availability in water bodies. In addition, COD is also a significant predictor of water quality. A one unit increase in COD is associated with a decrease in the log odds of being in a good water quality class versus a poorer class by 0.192. This suggests higher COD concentrations. It represents increased levels of oxidizable pollutants contribute significantly to poorer water quality conditions.

Furthermore, NH_3-N shows a significant negative effect on water quality classification. A one unit increase in NH_3-N decreases the log odds of being in a good water quality class by 0.120. Elevated ammonia nitrogen levels are commonly associated with domestic and industrial wastewater discharges which can adversely affect aquatic ecosystems and overall water quality. Similarly, SS is a significant predictor of water quality. A one unit increase in SS is associated with a decrease in the log odds of being in a good water quality class by 0.109. High concentrations of suspended solids can reduce light penetration and transport pollutants, thereby contributing to the degradation of water quality.

Among these variables, COD is identified as the most significant factor influencing water quality, as it has the largest absolute coefficient value (0.192) and a highly significant p-value ($p < 0.001$). This suggests that COD has the strongest effect in increasing the probability of water quality shifting from good quality water to poor quality water. This finding is consistent with previous studies done by Qasaimeh & Al-Ghazawi (2020) which also identified COD as a significant predictor of water quality. However, this result differs from the findings done by Hridoy et al. (2025), who found that COD was not statistically significant, while BOD played as most significant factor that affect water quality.

6. CONCLUSION

Water quality must be monitored because any deterioration in water can pose threats to human health as well as aquatic life in water. Therefore, water quality data collected in the Langat River Basin between 2023 and 2024 was used in this study. The study seeks to establish the most contributing factor affecting the water in the Langat River Basin.

In conclusion, the objective of this study, which is to identify the most significant factors that influence water quality classification has been achieved. Since all selected water quality parameters satisfied the multicollinearity assumptions, all variables were retained in the analysis. BOD, COD, NH_3-N and SS were identified as significant factors influencing water quality classification. However, pH was found to be not a significant factor influencing the water quality class. Among all significant variables, COD was identified as the most significant factor because it has the largest absolute coefficient value (0.192) and small p-value ($p < 0.001$). It has the strongest influence on the water quality classification.

It is recommended for future studies to consider other water quality parameters such as heavy metals, nutrients, and microbiological indicators. It is to ensure a comprehensive water quality assessment outcome. By considering these parameters, other possible sources of water quality impairment beyond the capability of basic physicochemical parameters are identified. As a result, a more comprehensive understanding of the factors influencing water quality classification can be gained. Thus, it produces more precise and trustworthy outcomes.

Acknowledgments: This research was supported by Universiti Teknologi MARA (UiTM) and funded under UiTM internal Grant No. 600-RMC/GPM LPHD 5/3 (077/2023). This support is gratefully acknowledged. The authors would like to thank other researchers, lecturers and friends for ideas and discussion.

References

Adnan, R. M., Dai, H. L., Kisi, O., Heddam, S., Kim, S., Kulls, C., & Zounemat-Kermani, M. (2023). Modelling biochemical oxygen demand using improved neuro-fuzzy approach by marine predators algorithm. *Environmental Science and Pollution Research*, 30(41), 94312-94333.

Ahmed, M. F., Mokhtar, M. B., & Alam, L. (2020). Factors influencing people's willingness to participate in sustainable water resources management in Malaysia. *Journal of Hydrology: Regional Studies*, 31, 100737.

Chen, Z., Liu, X., Cheng, X., & Guo, Z. (2023). Dissipation of eutrophic substances in grass carp aquaculture pond water by ozone. *Water*, 15(18), 3167.

Department of Environment. (2022). Department of Environment – Ministry of Environment and Water. doe.gov.my. https://www.doe.gov.my/en/utamaenglish/

Dewangan, S. K., Toppo, D. N., & Kujur, A. (2023). Investigating the impact of pH levels on water quality: An experimental approach. *International Journal for Research in Applied Science and Engineering Technology*, 11(9), 756-759.

Evsyukov, D., Glinscaya, A., Kukartsev, A., Volneikina, E., & Kukartseva, S. (2024). Analyzing the influence of parameters on water quality using logistic regression. *BIO Web of Conferences*, 130, 03001. EDP Sciences.

Han, X., Chen, X., Ma, J., Chen, J., Xie, B., Yin, W., ... & Huang, F. (2022). Discrimination of chemical oxygen demand pollution in surface water based on visible near-infrared spectroscopy. *Water*, 14(19), 3003.

Hashim, A. R., Kamaruddin, S. A., Tajam, J., Buyong, F., Che Ismail, C. Z., Lee Abdullah, A., ... & Anscelly, A. (2023). Assessment of the current spatialtemporal variations of salinity on the surface waters of Kuala Perlis, Perlis. *Science Letters (ScL)*, 17(2), 66-80.

Hridoy, M. A. A. M., Shawkat, A. I., Bordin, C., Acharjee, M. R., Masood, A., Baki, A. O., & Al Mamun, M. A. (2025). Advanced machine learning models for accurate water quality classification and WQI prediction: Implications for aquatic disease risk management. *Science of the Total Environment*, 1008, 180965.

Ibrahim, N., Abd Rahman, H. A., Azran, A. A., Mohd Faddillah, M. A., & Mohd Qamarudin, M.A.Q. (2023). Prediction of water quality for the Selangor rivers using data mining approach. *Journal of Sustainability Science and Management*, 18(9), 171–183.

Liu, C., Morrison, C., Jia, J., Xu, X., Liu, M., & Zhang, S. (2020). *Does the Nitrification-Suppressed BOD5 Test make sense?*.

Malek, N. H. A., Yaacob, W. F. W., Nasir, S. A. M., & Shaadan, N. (2021). The effect of chemical parameters on water quality index in machine learning studies: A meta-analysis. *Journal of Physics: Conference Series*, 2084(1), 012007.

Najah, A., Teo, F. Y., Chow, M. F., Huang, Y. F., Latif, S. D., Abdullah, S., ... & El-Shafie, A. (2021). Surface water quality status and prediction during movement control operation order under COVID-19 pandemic: Case studies in Malaysia. *International Journal of Environmental Science and Technology*, 18(4), 1009-1018.

Norrrahim, M. N. F., Kasim, N. A. M., Feizal Knight, V., Misenan, M. S. M., Janudin, N., Shah, N. A. A., Kasim, N., Yusoff, W. Y. W., Noor, S. A. M., Hasnawati Jamal, S., Khim Ong, K., & Yunus, W. M. Z. W. (2021).

Pang, J., Lin, K., Gan, W., Hu, S., & Luo, W. (2022). Contamination assessment and source analysis of urban waterways based on Bayesian and principal component analysis: A case study of Fenjiang River. *Water*, 14(18), 2912.

Qasaimeh, A., & Al-Ghazawi, Z. (2020). Regression modeling for rapid prediction of wastewater BOD5. *Desalination and Water Treatment*, 201, 165-172.

Salvador, T. D., Dos Santos, F. S., & Yokoyama, L. (2024). Assessment of water quality of the Sao Goncalo River in the urban section of Sao Domingos do Norte- Es. *Earth Envi Scie Res & Rev*, 7(2), 01-12.

Shamsuddin, I. I. S., Othman, Z., & Sani, N. S. (2022). Water quality index classification based on machine learning: A case from the Langat River Basin model. *Water*, 14(19), 2939.

Tyagi, S., Sharma, B., Singh, P., & Dobhal, R. (2013). Water quality assessment in terms of water quality index. *American Journal of Water Resources*, 1(3), 34-38.