

Non Linear Modeling of the Relationship Between Raw Water Turbidity, Accelerator Unit Turbidity, and Filtration Turbidity Using a Quadratic Interaction Model in a Water Treatment Plant

Nelvidawati^{a*}, Firdaus^b, M Andre Kurnia Festa^c

^{a,c} Environmental Engineering, Institut Teknologi Padang, Jl. Kandis Nanggalo, Padang, 25143, Indonesia

^b Telecommunication Engineering, Politeknik Negeri Padang, Address, Padang, 20362, Indonesia

^{a*} nelvidaus11@gmail.com

Abstract — This study aims to analyze the non-linear relationship between raw water turbidity, turbidity after the accelerator unit, and turbidity after the filtration unit at the Gunung Pangilun Water Treatment Plant (WTP) in Padang City using the Quadratic Interaction Model (QIM) approach. Turbidity data were collected from 1–5 June 2025 at one-hour intervals, resulting in 50 data points for each sampling location. The data processing stages included missing value inspection, outlier detection using the Interquartile Range (IQR) method, data consistency checking, relationship visualization among variables, and QIM-based modeling using Python. The results show that the QIM was unable to adequately represent the relationships among the variables, as indicated by an R-squared value of 0.145 and an adjusted R-squared of 0.048. All model parameters exhibited p-values greater than 0.05, indicating no statistically significant influence on turbidity after filtration. Model evaluation yielded an RMSE of 0.496 NTU and a MAPE of 19.34%, suggesting a moderate level of prediction accuracy. Additionally, the analysis identified 14% outliers and 9 inconsistent data points for each sampling location.

Keywords — Quadratic Interaction Model; turbidity; water treatment; accelerator; filtration; outliers; non-linear modelling; Gunung Pangilun WTP

Manuscript received 13 Dec. 2025; revised 15 Dec. 2025; accepted 21 Jan. 2026. Date of publication 03 Feb. 2026.

International Journal of Wireless And Multimedia Communications is licensed under a Creative Commons Attribution-Share Alike 4.0 International License.



I. INTRODUCTION

Assessment of drinking water quality is partly determined by turbidity, which serves as one of the primary parameters directly influencing the performance of water treatment processes [1]. During extreme weather events or climate-related changes, fluctuations in raw water turbidity often lead to variability in coagulation, flocculation, and filtration processes [2]. The accelerator unit, which functions as a sedimentation and initial floc formation system, and the filtration unit, which serves as the final stage of water treatment, both play crucial roles in ensuring that the treated water meets quality standards [3]. The relationships among water treatment process units are not always linear [4]. Complex response patterns frequently arise due to interactions between raw water characteristics, floc stability, sedimentation dynamics, and filter media conditions [5], making them difficult to adequately represent using conventional linear approaches.

Field observations conducted across various units of drinking water treatment plants indicate that turbidity levels after the accelerator unit do not always decrease proportionally to the raw water turbidity. In certain conditions, turbidity may even increase due to floc instability or suboptimal floc formation processes [6], [7]. Similarly, turbidity levels after the filtration unit are not always

consistent; although they are expected to decrease, occasional spikes may occur as a result of filter breakthrough, saturated filter media, or ineffective backwashing processes [8]. This variability suggests the presence of a non linear relationship involving interactions between raw water quality and the performance of each treatment unit [8].

The Quadratic Interaction Model (QIM) is a nonlinear polynomial approach capable of representing linear effects, quadratic terms, and interaction components among variables within an integrated framework, allowing the model to capture curvature and parameter interactions simultaneously [9]. This model provides a more comprehensive explanation of the response behavior of accelerator and filtration units under varying raw water characteristics and enables more accurate predictions to support operational decision-making in water treatment processes [6], [10], [11], [8], [12], [13].

This study aims to analyze the nonlinear relationship patterns between raw water turbidity, the accelerator unit, and the filtration unit. In addition, a predictive model based on the Quadratic Interaction Model is developed to describe the dynamics of turbidity reduction in each unit and to evaluate the performance of the resulting predictive model as a basis for understanding the performance of process units and supporting the optimization of water treatment plant operations. This research provides empirical contributions regarding the application of nonlinear statistical models in water treatment systems and offers an analytical approach

relevant to improving the reliability of treated water quality predictions and the operational effectiveness of clean water treatment installations.

II. METHODOLOGY

The research was conducted using a quantitative observational experimental approach to analyze the nonlinear relationships among raw water turbidity, turbidity after the accelerator unit, and turbidity after the filtration unit at the Gunung Pangilun Water Treatment Plant (IPA), Padang City. The mathematical model employed is the Quadratic Interaction Model (QIM), a second order nonlinear regression model that incorporates interaction effects among variables. The data used in this study were collected from the raw water unit, the accelerator unit, and the filtration unit. Data collection was carried out from June 1 to June 5, 2025, from 08:00 to 17:00 Western Indonesian Time (WIB) at one-hour intervals. A total of 50 turbidity measurements were obtained from each unit. Data processing was performed using the Python programming language on Google Colaboratory through several stages, namely:

1. Missing data inspection

This step serves as the initial stage of data cleaning to ensure that all variables used in the modeling process contain sufficient and complete information. The purpose of this step is to identify the presence of missing values. Missing data can introduce bias, reduce the quality of analysis, and affect the accuracy of the modeling results [14]. The inspection was carried out using descriptive statistical functions and data-profiling methods in Python.

2. Outlier detection and treatment

This stage aims to ensure that the dataset used in the modeling process exhibits a reasonable distribution pattern and does not contain extreme values that could affect the stability and accuracy of the Quadratic Interaction Model. Outliers were identified using the interquartile range (IQR) method. When detected, each outlier was evaluated to determine whether it represented a recording error, a process anomaly, or a valid operational condition. If the outlier originated from measurement errors or was inconsistent with the operational pattern of the water treatment plant, the data point was removed. If the outlier represented realistic process variability, it was retained so that the model could accurately capture the actual dynamics of the treatment system. This stage ensures that the dataset is clean, representative, and ready for nonlinear modeling without the risk of distortion from inappropriate extreme values [15].

3. Data consistency check

The data consistency check was conducted to ensure that turbidity values at each stage of the treatment process follow a logical pattern and are aligned with the operational flow of the water treatment system. In general, turbidity levels are expected to decrease from upstream to downstream ($T_{\text{raw}} \geq T_{\text{accelerator}} \geq T_{\text{filtration}}$).

If anomalies occur such as turbidity after filtration being higher than after the accelerator unit, turbidity after the accelerator being higher than raw water, negative values, or sudden unrealistic spikes then the data are classified as inconsistent [15].

4. Relationship visualization

The purpose of this stage is to observe the relationship patterns among raw water turbidity, turbidity after the accelerator unit, and turbidity after the filtration unit. The visualization method used is a pairplot (scatter matrix) [16], which displays pairwise scatter plots as well as the distribution of each variable along the diagonal.

5. Nonlinear modeling using the Quadratic Interaction Model (QIM)

This stage involves constructing a quadratic model capable of capturing linear effects, quadratic effects, and interaction terms [17] among raw water turbidity, turbidity after the accelerator unit, and to predict turbidity after the filtration unit. The model used is:

$$\hat{Y} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1^2 + \beta_4 X_2^2 + \beta_5 X_1 X_2$$

with:

X_1 = Raw water turbidity

X_2 = Turbidity after the accelerator unit

X_3 = Turbidity after the filtration unit

6. Model evaluation was performed using the Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE).

RMSE is used to measure the average magnitude of prediction errors in the same unit as the target variable [18], thereby providing an indication of how far the model's predicted values deviate from the actual turbidity values after the filtration unit. A lower RMSE indicates that the model has good predictive accuracy and is able to effectively represent the non-linear relationship between raw water turbidity, turbidity after the accelerator unit, and turbidity after the filtration unit.

7. MAPE is used to determine whether the model tends to produce predictions that are proportional to the magnitude of the actual turbidity values [19][20], although this metric is sensitive to very small actual values. The combination of these two metrics provides a comprehensive evaluation of the model's accuracy and consistency. Furthermore, they can be used to assess the suitability of the model for applications in monitoring and optimizing water treatment unit operations.

RESULT

The turbidity dataset used in this study consists of three variables: raw water turbidity, turbidity after undergoing coagulation, flocculation, and sedimentation processes in the accelerator unit, and turbidity after the filtration process, with a total of 50 initial samples. The completeness check indicated that no missing values were found in any of the variables, as shown in Figure 1, allowing the dataset to proceed to the subsequent analysis stages.

```

▶ #3.Check Missing Values
print("\nMissing values per kolom:")
print(df.isnull().sum())

...
Missing values per kolom:
Raw      0
Accel    0
Filter   0
dtype: int64

```

Figure 1. Python output of the missing value inspection

Outlier detection was performed using the Interquartile Range (IQR) method to evaluate the distribution of turbidity data at the three monitoring points, as shown in Figure 2.

```

▶ #4.Outlier Detection (IQR Method)
def detect_outliers(col):
    Q1 = df[col].quantile(0.25)
    Q3 = df[col].quantile(0.75)
    IQR = Q3 - Q1
    lower = Q1 - 1.5 * IQR
    upper = Q3 + 1.5 * IQR
    return df[(df[col] < lower) | (df[col] > upper)]

print("\nOutlier Raw:")
print(detect_outliers("Raw"))

print("\nOutlier Accel:")
print(detect_outliers("Accel"))

print("\nOutlier Filter:")
print(detect_outliers("Filter"))

Outlier Raw:
   Raw  Accel  Filter
...  3  2.43  3.10  1.42
   4  2.44  3.11  1.32
   5  2.54  3.25  1.50
  10  7.43  4.27  1.17
  17  5.18  2.17  1.34
  18  5.10  3.78  1.39
  19  4.98  3.70  1.45

Outlier Accel:
Empty DataFrame
Columns: [Raw, Accel, Filter]
Index: []

Outlier Filter:
   Raw  Accel  Filter
24  3.74  2.56  2.35
28  4.17  3.70  3.68
29  4.21  3.69  3.65
47  3.34  1.94  2.36
48  4.25  1.95  2.36
49  4.13  1.95  2.45

```

Figure 2. Outlier detection using Python

The analysis results indicate the presence of extreme values amounting to 14%. These outlier values primarily appeared during sudden increases in raw water turbidity and variations in the efficiency of the coagulation flocculation and filtration processes, which represent real phenomena in surface-water based treatment plant operations. Although statistically identified as outliers, all values were retained in the dataset because they were deemed to reflect actual operational dynamics rather than recording errors or instrumentation anomalies. The decision to retain these outliers ensures that the Quadratic Interaction Model (QIM) can more comprehensively capture process variability under both stable and extreme conditions, thereby producing prediction models that are more relevant for operational

monitoring, control, and optimization in water treatment processes.

A total of nine (9) inconsistent data points were identified for each parameter raw water turbidity, turbidity after the accelerator unit, and turbidity after the filtration unit as shown in Figure 3.

```

▶ #5.Check Data Consistency
#For example: Turbidity should decrease from Raw → Accel → Filter
df["Consistent"] = (df["Raw"] >= df["Accel"]) & (df["Accel"] >= df["Filter"])
print("\nInconsistent Data:")
print(df[df["Consistent"] == False])

...
Inconsistent Data:
   Raw  Accel  Filter  Consistent
2  3.44  3.54  1.56         False
3  2.43  3.10  1.42         False
4  2.44  3.11  1.32         False
5  2.54  3.25  1.50         False
20 3.99  5.32  1.77         False
21 3.56  3.65  1.87         False
47 3.34  1.94  2.36         False
48 4.25  1.95  2.36         False
49 4.13  1.95  2.45         False

```

Gambar 3. Check Data Consistency

The patterns observed in the consistency check indicate instability within each unit operation and water treatment process, suggesting that the underlying causes may be systemic. For raw water, inconsistencies may result from environmental fluctuations such as rainfall intensity, changes in source water quality, variations in inflow discharge, or external events such as erosion and surface runoff. Inconsistencies after the accelerator unit may be associated with variability in coagulation flocculation process control, including unstable coagulant dosage, pH fluctuations, mixing velocity, or suboptimal floc formation efficiency. For the filtration unit, inconsistencies may arise due to filter media performance, improper backwashing, fouling, or non-uniform hydraulic distribution.

When linked to the earlier finding that 14% of the data were identified as outliers, these inconsistencies represent the actual dynamics of water treatment processes, which do not always operate under steady state conditions. Therefore, extreme values were retained in the analysis to ensure that the model reflects real operational phenomena, particularly because the non linear model used is sensitive to extreme variations as part of the inherent characteristics of field systems. The visualization of the relationships among the variables is presented in Figure 4.

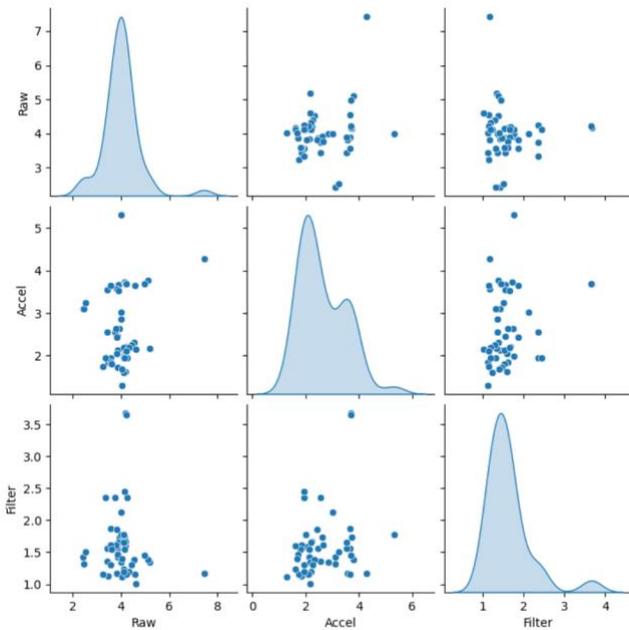


Figure 4. Relationship Visualization Among Variables

The scatter matrix visualization illustrates the distribution patterns and relationships among the three turbidity parameters: raw water turbidity (Raw), turbidity after the accelerator unit (Accel), and turbidity after filtration (Filter). On the diagonal plots, the distribution profiles appear relatively symmetrical with a dominant unimodal peak, although extreme values are visible on the right side, indicating the presence of outliers. The turbidity distribution after the accelerator unit is more dispersed and tends to be right-skewed, suggesting unstable coagulation flocculation performance at certain times. Meanwhile, the turbidity distribution after filtration shows a narrower spread compared with the other two variables, demonstrating that the filtration unit still functions as the primary control stage in reducing incoming variability from previous processes.

In the pairwise scatterplots, the relationship between raw water turbidity and accelerator turbidity does not exhibit a strong linear pattern; instead, a scattered distribution indicates that the accelerator unit does not always respond proportionally to the raw water condition. This may reflect the influence of operational factors such as coagulant dosage, pH, or inconsistent mixing intensity. Conversely, the relationship between accelerator turbidity and filtration turbidity shows a clearer decreasing trend, although still non-linear, supporting the notion that filtration plays a key role as the final polishing stage even when variability from upstream processes remains high. The scatterplot between raw water turbidity and filtration turbidity shows a very weak relationship, reinforcing that the treatment system provides substantial separation between initial and final water quality conditions.

Overall, the visualization confirms that the relationships among variables are non linear with considerable process variability, particularly in the upstream stages of the system. Therefore, the use of the Quadratic Interaction Model (QIM) in this study is considered appropriate for capturing the

complex interaction patterns among turbidity variables at each treatment unit.

The estimation results of the Quadratic Interaction Model (QIM), as shown in Figure 5, indicate that the relationships among raw water turbidity (RAW), turbidity after the accelerator unit (Accel), and turbidity after filtration (Filter) cannot be explained significantly by the second-order non-linear model used. The R-squared value of 0.145 and the adjusted R-squared value of 0.048 indicate that the model explains only about 14.5% of the variation in post-filtration turbidity, while the remaining variation is influenced by other factors not included in the model. Additionally, the Prob(F-statistic) value of 0.212 (> 0.05) shows that the overall model is not statistically significant.

```
#7.Creating a Quadratic Interaction Model
df["Raw2"] = df["Raw"]**2
df["Accel2"] = df["Accel"]**2
df["RawAccel"] = df["Raw"]*df["Accel"]

formula = "Filter ~ Raw + Accel + Raw2 + Accel2 + RawAccel"
model = ols(formula, data=df).fit()

print("\nHasil regresi Quadratic Interaction Model:")
print(model.summary())

...
Hasil regresi Quadratic Interaction Model:
OLS Regression Results
=====
Dep. Variable:      Filter    R-squared:      0.145
Model:              OLS      Adj. R-squared: 0.048
Method:             Least Squares    F-statistic:    1.492
Date:               Sun, 16 Nov 2025    Prob (F-statistic): 0.212
Time:               12:52:35          Log-Likelihood: -35.924
No. Observations:   50              AIC:            83.85
Df Residuals:       44              BIC:            95.32
Df Model:           5
Covariance Type:   nonrobust
=====
               coef    std err          t      P>|t|    [0.025    0.975]
-----
Intercept    0.2805    1.955    0.144    0.887    -3.659    4.220
Raw          0.6659    0.516    1.291    0.203    -0.373    1.705
Accel       -0.0813    0.825   -0.099    0.922    -1.745    1.582
Raw2        -0.1479    0.077   -1.921    0.061    -0.303    0.007
Accel2      -0.0722    0.088   -0.825    0.414    -0.249    0.104
RawAccel    0.1776    0.186    0.954    0.345    -0.197    0.553
=====
Omnibus:         29.575    Durbin-Watson:    0.894
Prob(Omnibus):   0.000    Jarque-Bera (JB): 58.053
Skew:            1.798    Prob(JB):         2.48e-13
Kurtosis:        6.865    Cond. No.:        654.
=====
Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

Gambar 5. Quadratic Interaction Model

Individually, all variables including the linear components (Raw, Accel), the quadratic components (Raw², Accel²), and the interaction term (Raw, Accel) have p-values greater than 0.05, indicating that none of them exert a statistically significant influence on post-filtration turbidity. Although the Raw coefficient has a positive value (0.6659), which physically suggests that higher raw water turbidity tends to increase the likelihood of higher turbidity after filtration, this relationship is not statistically significant. The interaction coefficient between Raw and Accel is also insignificant, indicating that the combined effect of both variables does not meaningfully influence filtration performance.

Diagnostic tests reveal abnormalities in the residual distribution, as indicated by the Omnibus and Jarque Bera statistics with p-values of 0.000, skewness of 1.798, and kurtosis of 6.865, confirming that the residuals are not normally distributed. This reinforces earlier indications of anomalies in the dataset, including the presence of outliers (14%) and inconsistent data at all three sampling points (nine

data points each), which hinder the model's ability to capture the underlying relationship. The Durbin Watson value of 0.894 (< 1.5) also indicates potential positive autocorrelation, which commonly occurs in time-series data from water treatment plant operations.

Overall, the regression results show that the physical process structure at the Gunung Pangilun Water Treatment Plant (IPA) is not adequately captured by the quadratic interaction mathematical pattern. This is likely due to more complex process dynamics such as variability in coagulation flocculation performance, flow rate fluctuations, coagulant dosage changes, and filter media efficiency combined with the presence of outliers, inconsistent data, and short term operational fluctuations. Thus, the QIM is not suitable for representing the relationships among the three variables, and alternative modeling approaches, such as process-based modeling, should be considered.

The evaluation results of the Quadratic Interaction Model (QIM), shown in Figure 6, indicate a Root Mean Square Error (RMSE) of 0.496 NTU, meaning that the average deviation of model predictions from the actual post filtration turbidity values is approximately 0.5 NTU. Although this RMSE is relatively small in absolute terms given that post filtration turbidity at the plant generally ranges from 0.8 to 3 NTU it still reflects the model's limited ability to accurately capture all process variations. Furthermore, the Mean Absolute Percentage Error (MAPE) of 19.34% indicates that, on average, the model's prediction error is about 19% of the actual value, placing the model's performance in the "fair accuracy" category. Typically, MAPE below 10% is considered excellent, 10–20% acceptable, 20–50% poor, and above 50% unacceptable.

These moderate RMSE and MAPE values align with the regression results showing low R-squared values and nonsignificant model coefficients. Together, these metrics confirm that the QIM can represent only a small portion of the filtration process dynamics, while much of the variability in the water treatment system is influenced by other factors such as coagulation–flocculation efficiency, flow fluctuations, changes in raw water characteristics, and the presence of outliers and inconsistent data reaching 14%. Therefore, the model requires improvement, either through selecting a more adaptive non-linear modeling approach or enhancing the model with additional process variables to improve predictions of post filtration turbidity.

```
#8. Model Evaluation
y_true = df["Filter"]
y_pred = model1.predict(df)

rmse = np.sqrt(mean_squared_error(y_true, y_pred))
mape = mean_absolute_percentage_error(y_true, y_pred) * 100

print("\nRMSE:", rmse)
print("MAPE (%)ate", mape)

...
RMSE: 0.49635840755968463
MAPE (%): 19.339499285838684
```

Figure 6. Evaluation Model

CONCLUSION

This study analyzes the nonlinear relationship among raw water turbidity, turbidity after the accelerator unit, and

turbidity after the filtration unit at the Gunung Pangilun Water Treatment Plant using the Quadratic Interaction Model (QIM) approach. The results show that the quadratic model with interaction terms is not able to adequately represent the dynamics of the water treatment process. The R-squared value of 0.145 and the Adjusted R-squared value of 0.048 indicate that only a small portion of the variation in filtrated water turbidity can be explained by the model. All model parameters linear, quadratic, and interaction have p-values greater than 0.05, indicating no statistically significant influence on filtration quality. Model evaluation also produced an RMSE of 0.496 NTU and a MAPE of 19.34%, suggesting a moderate level of prediction accuracy and aligning with the overall weak model performance.

The findings further show that 14% of the data were identified as outliers, along with inconsistencies in 9 data points at each sampling location, which contributed to the model's low ability to capture the relationships among variables. This condition reflects the complexity of coagulation–flocculation and filtration processes in the field, influenced by variability in raw water characteristics, changes in flow rate, mixing efficiency, and operational instability of the units during the observation period.

Overall, the QIM does not provide a strong representation of the relationships among process variables in the water treatment system at the Gunung Pangilun WTP. To obtain a more accurate and robust predictive model, further research is recommended using nonlinear approaches such as machine learning models, robust regression, or physics-based process modeling, complemented by additional operational variables such as coagulant dosage, flow rate, pH, and jar test parameters. These approaches have the potential to provide a more comprehensive understanding for optimizing clean water treatment processes in the future.

REFERENCES

- [1] F. K. K. A, A. D. K. B, and C. A. P. b, V.G. Papadakis a, "Revisiting of coagulation-flocculation processes in the production of potable water," *J. Water Process Eng.*, vol. 27, no. February 2019, pp. 193–204, 2019, doi: <https://doi.org/10.1016/j.jwpe.2018.12.007>.
- [2] K. S. Nemani, S. Peldszus, and P. M. Huck, "Practical Framework for Evaluation and Improvement of Drinking Water Treatment Robustness in Preparation for Extreme-Weather-Related Adverse Water Quality Events," *ACS EST Water*, vol. 3, no. 5, pp. 1305–1313, 2023, doi: [10.1021/acsestwater.2c00627](https://doi.org/10.1021/acsestwater.2c00627).
- [3] A. Chiavola, C. Di Marcantonio, M. D'Agostini, S. Leoni, and M. Lazzazzara, "A combined experimental-modeling approach for turbidity removal optimization in a coagulation–flocculation unit of a drinking water treatment plant," *J. Process Control*, vol. 130, 2023, doi: [10.1016/j.jprocont.2023.103068](https://doi.org/10.1016/j.jprocont.2023.103068).
- [4] S. Safeer *et al.*, "A review of artificial intelligence in water purification and wastewater treatment: Recent advancements," *J. Water Process Eng.*, vol. 49, no. October 2022, p. 102974, 2022, doi: <https://doi.org/10.1016/j.jwpe.2022.102974>.
- [5] O. Bello, Y. Hamam, and K. Djouani, "Coagulation process control in water treatment plants using multiple

- model predictive control,” *Alexandria Eng. J.*, vol. 53, no. 4, pp. 939–948, 2014, doi: 10.1016/j.aej.2014.08.002.
- [6] C. W. K. Khedher, M., Awad, J., Donner, E., Drigo, B., Fabris, R., Harris, M., Braun, K., & Chow, “Using the Flocculation Index to optimise coagulant dosing during drinking water treatment,” *J. Water Process Eng.*, vol. 51, no. February 2023, 103394, 2023, doi: <https://doi.org/10.1016/j.jwpe.2022.103394>.
- [7] K. Saxena, U. Brighu, and A. Choudhary, “Experimental investigation and modelling the effect of humic acid on coagulation efficiency for sludge blanket clarifier,” *Chemosphere*, vol. 266, no. March 2021, p. 128958, 2021, doi: <https://doi.org/10.1016/j.chemosphere.2020.128958>.
- [8] Z. Gao, Z. Ding, J. Liu, and W. Liu, “Experimental study on the influencing factors of permeability of sand media,” *Hydrogeol. Eng. Geol.*, vol. 51, no. 6, pp. 8–17, 2024, doi: 10.16030/j.cnki.issn.1000-3665.202404025.
- [9] J. An *et al.*, “Machine Learning-Based Prediction of Coagulant Dosing in Drinking Water Treatment Plants Using Polynomial Regression with Lasso Regularization,” pp. 1–17, 2025.
- [10] M. Yateh, G. Lartey-Young, F. Li, M. Li, and Y. Tang, “Application of Response Surface Methodology to Optimize Coagulation Treatment Process of Urban Drinking Water Using Polyaluminium Chloride,” *Water (Switzerland)*, vol. 15, no. 5, pp. 1–13, 2023, doi: 10.3390/w15050853.
- [11] X. Ji, Z. Li, M. Wang, Z. Yuan, and L. Jin, “Response Surface Methodology Approach to Optimize Parameters for Coagulation Process Using Polyaluminum Chloride (PAC),” *Water (Switzerland)*, vol. 16, no. 11, 2024, doi: 10.3390/w16111470.
- [12] S. Moradi *et al.*, “Forecasting and Optimizing Dual Media Filter Performance via Machine Learning,” *Water Res.*, vol. 235, no. December 2022, p. 119874, 2023, doi: 10.1016/j.watres.2023.119874.
- [13] A. A. K. Dhruvo, M. Jannat, and M. S. Hossain, “Enhancing the performance of coagulants for wastewater treatment by varying and optimizing the experimental parameters,” *J. Water Process Eng.*, vol. 55, no. October 2023, 104144, 2023, doi: <https://doi.org/10.1016/j.jwpe.2023.104144>.
- [14] D. R. Roderick Little, *Statistical Analysis with Missing Data, Third Edition*. WILEY, 2019. doi: 10.1002/9781119482260.
- [15] I. F. Ilyas and X. Chu, *Data Cleaning*. ACM Books, 2019.
- [16] I. Raza *et al.*, “Geospatial interpolation and hydrogeochemical characterization of alluvial aquifers in the Thal Desert, Punjab, Pakistan,” *PLoS One*, vol. 19, no. 8, 2024, doi: 10.1371/journal.pone.0307025.
- [17] R. P. Duncan and B. J. Kefford, “Interactions in statistical models: Three things to know,” *Methods Ecol. Evol.*, vol. 12, no. 12, pp. 2287–2297, 2021, doi: 10.1111/2041-210X.13714.
- [18] H. Khoshvaght, R. R. Permala, A. Razmjou, and M. Khiadani, “A critical review on selecting performance evaluation metrics for supervised machine learning models in wastewater quality prediction,” *J. Environ. Chem. Eng.*, vol. 13, no. 6, p. 119675, 2025, doi: 10.1016/j.jece.2025.119675.
- [19] S. Kim and H. Kim, “A new metric of absolute percentage error for intermittent demand forecasts,” *Int. J. Forecast.*, vol. 32, no. 3, pp. 669–679, 2016, doi: 10.1016/j.ijforecast.2015.12.003.
- [20] A. de Myttenaere, B. Golden, B. Le Grand, and F. Rossi, “Mean Absolute Percentage Error for regression models,” *Neurocomputing*, vol. 192, no. 5 June 2016, pp. 38–48, 2016, doi: <https://doi.org/10.1016/j.neucom.2015.12.114>.