**Academic Journal of Science and Technology**
ISSN: 2771-3032 | Vol. 21, No. 2, 2026

# Interpretable Dynamic Prediction and Risk Assessment of Filtered-Water Turbidity in a Drinking Water Treatment Plant

**Xufeng Liu**[1,*]

[1]College of Earth and Environmental Sciences, Lanzhou University, Lanzhou 730000, Gansu, China
[*]Corresponding author: liuxf2023@lzu.edu.cn

**Abstract:** Filtered-water turbidity at the study plant was usually low during 2025, but several short excursions coincided with changes in raw-water and hydraulic conditions. The practical problem was not simply to predict those values: operators also needed to know which measurements were driving a warning and whether the predicted value represented a meaningful operational risk. We analysed 4,380 records from January to December 2025 and used observations from January to March 2026 as a separate follow-up set. The available signals covered raw-water turbidity, pH and flow, alum dose, river level, clear-water well level and filtered-water turbidity. Missing entries were handled according to their frequency, while extreme observations were retained and flagged because they could represent genuine plant events. Predictor relevance was examined with Pearson correlation, mutual information, LASSO and randossm-forest importance. XGBoost and LightGBM were then combined in a weighted model, with SHAP values used to inspect individual contributions. An ARX formulation was fitted separately to examine delayed responses, and entropy-weighted fuzzy grading was used to express risk. Clear-water well level, river level and lagged raw-water flow emerged as the leading variables. The ensemble obtained R2 = 0.834 on the training data, whereas the extended ARX model obtained R2 = 0.800. Of the 2025 observations, 90.6% fell in the safe class; every follow-up observation was also classified as safe. The framework therefore links a turbidity estimate to both an explanation and an operational risk label.

**Keywords:** Filtered-water turbidity; drinking-water treatment; gradient-boosting ensemble; lag response; SHAP interpretation; operational risk grading.

## 1. Introduction

At a conventional drinking water plant, turbidity measured after filtration is one of the quickest checks on whether the treatment train is operating steadily. An increase may indicate that suspended material has passed through the filters and can also reduce the effectiveness of subsequent disinfection [1]. The signal is not easy to anticipate from one measurement alone. Raw-water quality, intake flow, river stage, dosing decisions and storage levels change on different time scales, while the effect of a coagulant adjustment may not appear at the filtered-water outlet until much later. Operators therefore face a timing problem as well as a prediction problem: a warning must arrive early enough to be useful, but it should also point to a plausible process cause.

Two broad modelling routes are available. Process-based equations retain a direct physical meaning, although the assumptions required for calibration may become restrictive when several operating variables change together. Data-driven models avoid some of those assumptions and have been used widely for water-quality assessment and management [2,3]. Their suitability nevertheless depends on the data structure. The records in this study form a moderate-sized table rather than an image or high-frequency sensor stream, so gradient-boosting trees offer a practical alternative to deeper networks. SHAP values can then be used to examine how a particular combination of measurements shifts a prediction. Time also needs explicit treatment. Water passes through successive units, and the effect of an upstream disturbance is therefore distributed over later observations; previous work has used artificial intelligence to study this recovery behaviour [4]. A model that aligns every input only with the concurrent turbidity value may conceal that delay.

The plant records were used to answer four linked operational questions. First, which routinely recorded variables carry the most information about filtered-water turbidity? Second, how much of the observed variation can be reproduced by an ensemble built from those measurements? Third, do alum dose, pH, flow and raw-water turbidity influence the outlet at different lags? Finally, can the numerical output be recast as a small set of warning categories that an operator can act on? Factor screening, boosted-tree prediction, ARX analysis and fuzzy grading were selected to address these questions. The use of machine learning at filter level is supported by recent plant-scale work on filter-effluent turbidity [5].

These analytical steps were kept separate because they answer different questions. The ensemble estimates turbidity, SHAP shows which current or lagged measurements move that estimate upward or downward, and the grading procedure places the result in an operational band. Used together, the outputs provide more than a single number. They indicate whether the estimate deserves attention and which part of the recent operating record should be checked first.

## 2. Data Sources and Preprocessing

### 2.1. Data Sources and Core Variables

Operational data were extracted from the monitoring system of an urban drinking water treatment plant. The development set comprised 4,380 observations recorded between January and December 2025 and contained 21

variables; measurements from January to March 2026 were reserved for follow-up validation. The variables used in the main analysis were filtered-water turbidity (FILT. NTU), raw-water turbidity (R/W NTU), raw-water pH (R/W PH), raw-water flow (R/W FLOW), alum dosage (ALUM), river level (RIVER LEVEL), clear-water well level (C/W WELL LEVEL), raw-water color (R/W CLR), and residual chlorine (CL2). Table 1 lists their definitions and analytical roles.

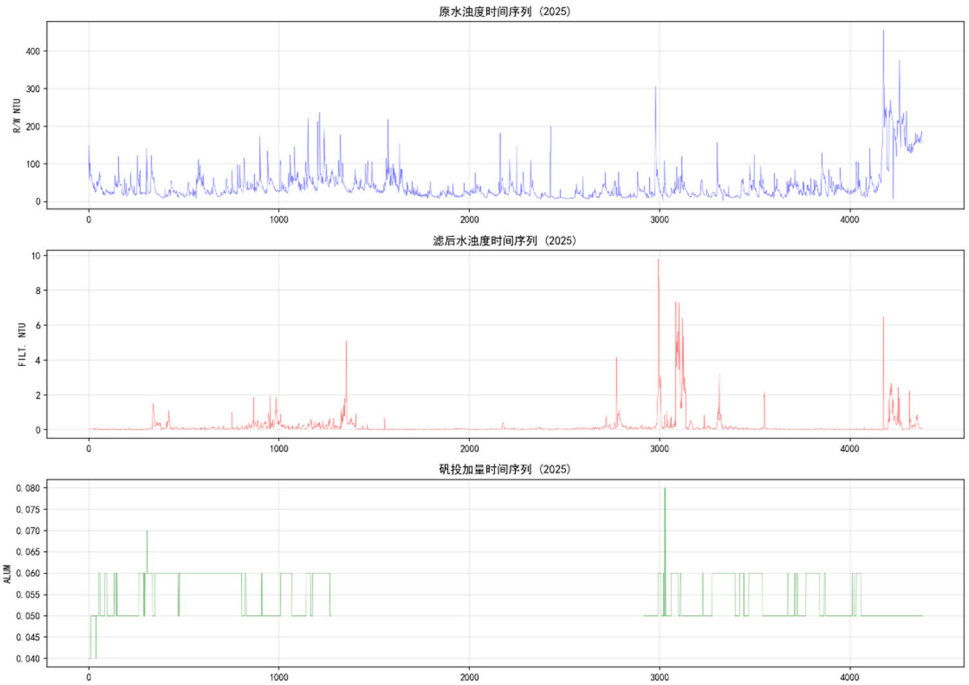


**Figure 1.** Recorded trajectories of raw-water turbidity, filtered-water turbidity and alum dosage during 2025.

The upper series in Figure 1 varied throughout the year and rose sharply near the end of the record. By comparison, filtered-water turbidity stayed close to its baseline for long periods and increased only in a limited number of episodes. This difference is consistent with substantial attenuation across the treatment train, but it also shows that the attenuation was not complete. The response distribution is strongly concentrated at low values, with relatively few excursions. Model fitting is therefore dominated by routine operation unless those rarer episodes are considered explicitly.

**Table 1.** Variables retained for modeling and their operational interpretation.

| Variable | Description | Role |
|---|---|---|
| FILT. NTU | Filtered-water turbidity | Target |
| R/W NTU | Raw-water turbidity | Water-quality input |
| R/W PH | Raw-water pH | Water-quality input |
| R/W FLOW | Raw-water flow | Hydraulic load |
| ALUM | Coagulant dosage | Control variable |
| RIVER LEVEL | River water level | Hydrological input |
| C/W WELL LEVEL | Clear-water well level | Process state |
| R/W CLR | Raw-water color | Water-quality input |

Table 1 identifies the target variable, raw-water inputs, hydraulic load, operational control variable, and process-state variables used in the subsequent factor screening and modeling.

## 2.2. Data Cleaning and Correlation Analysis

Because the raw spreadsheets from different months had slight differences in field names and column order, all monthly tables were first standardized into a unified format. Records with missing values in a few variables such as river level and raw-water flow were removed. Variables with more frequent missing values, including ALUM, CL2, and R/W PH, were filled by the median. Outliers were identified using the interquartile range method:

$$Q_1 - 1.5IQR \le x \le Q_3 + 1.5IQR \tag{1}$$

where $IQR = Q_3 - Q_1$. Since abnormal water-quality values may correspond to actual operational events, outliers were marked but not directly removed.

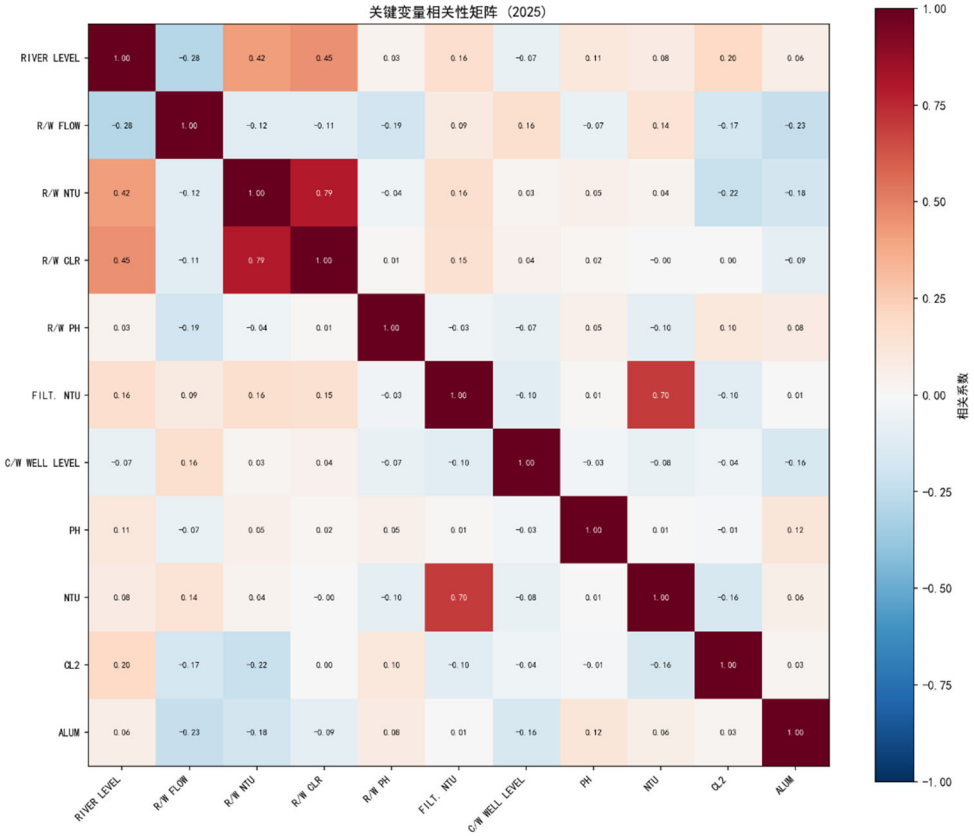


**Figure 2.** Pearson correlation matrix of key variable

Figure 2 presents the Pearson correlation matrix. Raw-water turbidity and raw-water color are positively correlated with filtered-water turbidity, while clear-water well level is negatively correlated. However, the overall linear correlations are weak, indicating the need for nonlinear models and dynamic variables.

# 3. Methods

## 3.1. Multi-Method Factor Screening

To improve the robustness of factor identification, four indicators were combined: Pearson correlation, mutual information, LASSO regression, and random-forest importance. Pearson correlation was used to measure linear dependence:

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{2}$$

Mutual information was used to capture nonlinear dependence:

$$I(X;Y) = \sum_{x}\sum_{y} p(x,y)\log\frac{p(x,y)}{p(x)p(y)} \tag{3}$$

To obtain a sparse set of predictors, LASSO estimation was applied with an L1 penalty:

$$\min_{\beta}\left\{\frac{1}{2n}\sum_{i}(y_i - \mathbf{x}_i^T\beta)^2 + \lambda\sum_{j}|\beta_j|\right\} \tag{4}$$

Random-forest importance was obtained from the accumulated decrease in node impurity attributed to each predictor. Each of the four screening outputs was rescaled before averaging, so no measure dominated simply because of its numerical range. The combined score was used as a ranking device rather than as a formal test. This distinction matters because the methods respond to different features of the data: Pearson correlation is sensitive to linear association, mutual information can retain nonlinear dependence without showing its sign, LASSO removes variables that add little once other predictors are present, and the forest score reflects repeated split decisions. Agreement across these views was taken as stronger evidence than a high value from any one of them. A related combination of weighting and machine-learning methods has been reported for water-quality prediction [6].

## 3.2. Ensemble Prediction and SHAP Interpretation

XGBoost and LightGBM were used to construct the filtered-water turbidity prediction model. The XGBoost objective function is:

$$\mathrm{L}^{(t)} = \sum_{i} l\left(y_i, \hat{y}_i^{(t-1)} + f_t(\mathbf{x}_i)\right) + \Omega(f_t) \tag{5}$$

$$\Omega(f) = \gamma T + \frac{1}{2}\lambda \sum_j w_j^2 \tag{6}$$

LightGBM uses histogram-based gradient boosting and leaf-wise growth, making it suitable for structured tabular data. To reduce single-model bias, a weighted ensemble was constructed:

$$\hat{y}_{ens} = \alpha \hat{y}_{xgb} + (1-\alpha)\hat{y}_{lgb} \tag{7}$$

Model performance was evaluated using RMSE, MAE, and the coefficient of determination R2. To explain model output, SHAP was introduced. For an individual sample, the prediction can be decomposed as:

$$f(\mathbf{x}) = \phi_0 + \sum_j \phi_j \tag{8}$$

where $\phi_j$ denotes the marginal contribution of the $j$-th feature. Explainability is as important as prediction accuracy in practical plant operation. Recent water-quality studies have used SHAP to interpret black-box machine learning predictions and improve transparency [7,8].

### 3.3. Dynamic Lag Modeling

Filtered-water turbidity also responds to raw-water and operational variables with time delays. Therefore, cross-correlation between input$u_j(t)$ and target$u_j(t)$was calculated:

$$\rho_j(\tau) = corr(u_j(t-\tau), y(t)) \tag{9}$$

The lag $tau$ that maximizes $\rho_j(\tau)$ was selected as a candidate delay. An autoregressive exogenous-input model was then established:

$$y(t) = \sum_i a_i y(t-i) + \sum_j \sum_k b_{jk} u_j(t-\tau_j-k) + \varepsilon(t) \tag{10}$$

The ARX model uses both historical filtered-water turbidity and lagged exogenous inputs. Its parameters have clear engineering meanings: autoregressive terms reflect process inertia, exogenous terms represent the influence of raw-water and operational variables, and lag orders correspond to process transmission time.

### 3.4. Fuzzy Risk Classification

Finally, a four-level risk classification system was built using filtered-water turbidity as the core indicator. The classification criteria are listed in Table 2. Fuzzy comprehensive evaluation was applied with trapezoidal membership functions, and entropy weighting was used to determine indicator weights. Recent studies have used ensemble machine learning to improve water-quality index classification [9] and have coupled prediction with risk assessment for water-pollution identification [10].

**Table 2.** Risk classification criteria.

| Level | FILT. NTU range | Meaning |
|---|---|---|
| Safe | <= 0.5 | Stable and safe |
| Low risk | (0.5, 1.0] | Approaching threshold |
| Medium risk | (1.0, 2.0] | Beyond regular control target |
| High risk | > 2.0 | Significant abnormality |

For operational use, the continuous FILT. NTU values were mapped to the four bands in Table 2. These bands provide the basis for summarizing how often the plant operated within, or moved beyond, its normal turbidity range during the year.

## 4. Results and Analysis

### 4.1. Key Influencing Factors

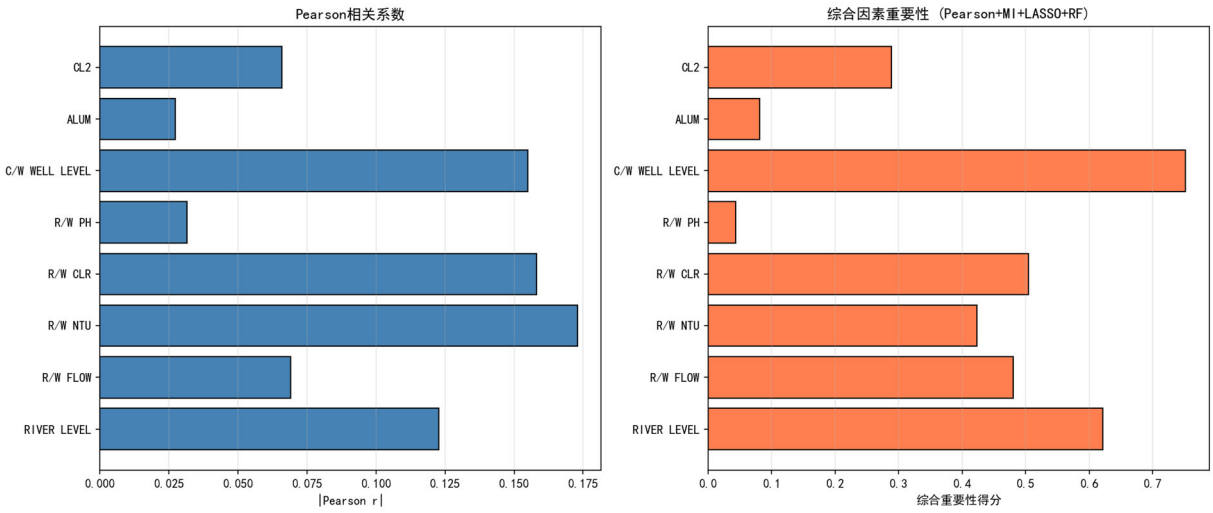


**Figure 3.** Combined ranking of variables associated with filtered-water turbidity.

The ranking obtained from the four screening measures is plotted in Figure 3. Clear-water well level had the largest combined score, followed by river level, raw-water flow at a six-step lag, raw-water color and current raw-water flow. Current raw-water turbidity was not the dominant term. Instead, the ranking gives substantial weight to variables that describe storage status, hydraulic movement and recent operating history. Table 3 provides the values used for this comparison.

**Table 3.** Overall importance scores for the five leading predictors.

| Rank | Variable | Main relation | Score |
|---|---|---|---|
| 1 | C/W WELL LEVEL | r = -0.155 | 0.752 |
| 2 | RIVER LEVEL | r = 0.123 | 0.622 |
| 3 | R/W FLOW_lag6 | Lag 6 steps | 0.531 |
| 4 | R/W CLR | r = 0.158 | 0.505 |
| 5 | R/W FLOW | r = 0.069 | 0.481 |

With a score of 0.752, clear-water well level was separated from the remaining predictors. This variable records the downstream storage state and may also capture how the plant is being balanced at the time of measurement. River level represents an external hydrological condition, whereas the lagged-flow term points to transport through the treatment train. The lower position of raw-water turbidity should not be interpreted as a lack of influence. It is more consistent with the fact that sedimentation and filtration weaken part of the upstream signal before water reaches the filtered-water monitoring point.

## 4.2. Prediction Performance

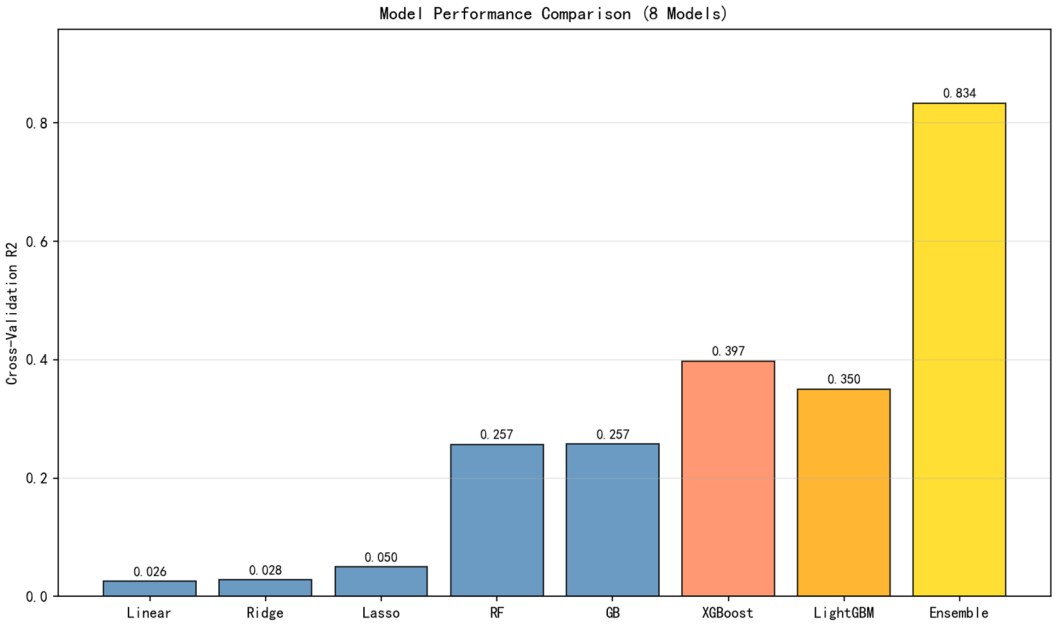


**Figure 4.** Predictive performance of the eight models evaluated in this study.

As seen in Figure 4, the two boosting algorithms performed better than linear regression, Ridge, LASSO, random forest, and conventional gradient boosting. Combining XGBoost and LightGBM produced the highest training R2 (0.834). This gain is plausible because tree ensembles can represent thresholds and interactions that are difficult to capture with a single linear relationship. The result nevertheless refers to model fit on the training data and should be interpreted together with the follow-up validation results.

**Table 4.** Selected observed and predicted values from the February 2026 validation data.

| Time | R/W NTU | Observed FILT. NTU | Predicted FILT. NTU |
|---|---|---|---|
| 7:00 | 12 | 0.12 | 0.073 |
| 9:00 | 23 | 0.25 | 0.066 |
| 11:00 | 24 | 0.18 | 0.065 |
| 13:00 | 24 | 0.15 | 0.098 |
| 15:00 | 50 | 0.13 | 0.000 |

Table 4 lists several observations from the February 2026 follow-up set. For the lower and more typical turbidity values, the predictions remained within the range expected for routine operation, although they did not reproduce every short-term change. The clearest mismatch occurred at 15:00, when raw-water turbidity rose to 50 NTU but the predicted filtered-water value fell to 0.000 NTU. This underestimation indicates that the current predictors do not fully describe stronger disturbances. Rainfall, water temperature, individual filter condition, and backwash history are likely to be useful additions when the objective is to predict such events.

## 4.3. SHAP-Based Model Interpretation

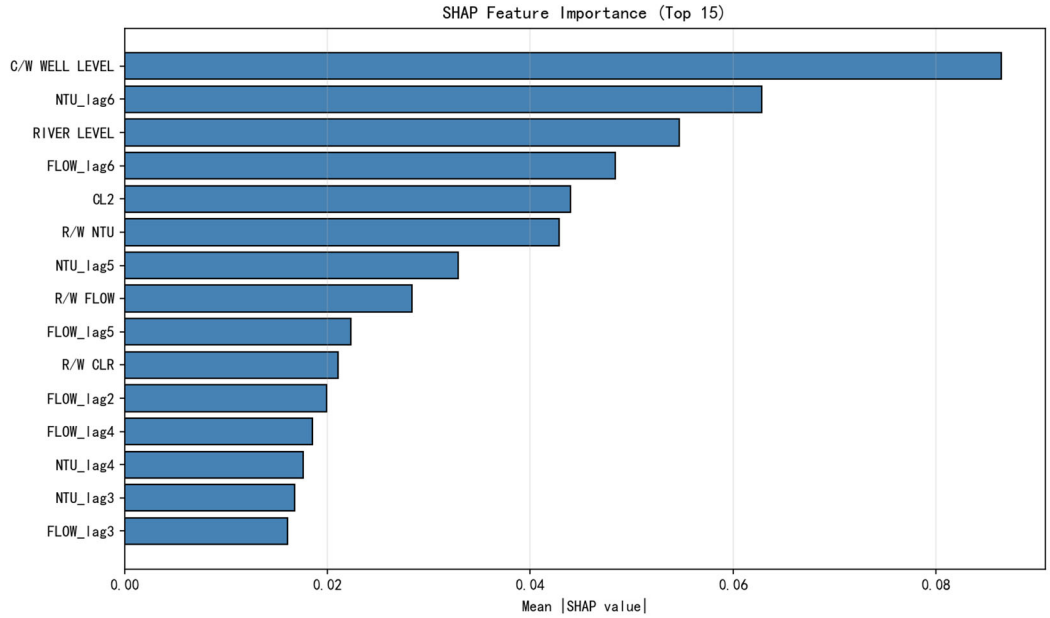


**Figure 5.** SHAP feature-importance ranking.

Figure 5 gives a similar, although not identical, picture. C/W WELL LEVEL produced the largest mean absolute SHAP value, and several lagged turbidity and flow terms were also prominent. Thus, the fitted ensemble did not base its output on raw-water turbidity alone; recent plant state and preceding measurements altered the prediction materially. For an unusually high estimate, the SHAP profile can be read as a case-specific checklist. An operator could first review the well level and then compare the recent flow and turbidity history to determine whether the warning reflects a hydraulic change, an upstream disturbance, or both.

## 4.4. Dynamic Lag Response

Cross-correlation analysis indicates that the optimal lags of R/W NTU and R/W FLOW are 0 steps, while the optimal lags of ALUM and R/W PH are approximately 12 and 9 steps, respectively. The detailed lag relationships are summarized in Table 5. With a two-hour sampling interval, the effect of ALUM appears at a time scale of about 24 hours, which is consistent with the process pathway from coagulant adjustment through sedimentation and filtration to the filtered-water side.

**Table 5.** Lag relationships between input variables and filtered-water turbidity.

| Input variable | Optimal lag | Approx. time | Correlation |
|---|---|---|---|
| R/W NTU | 0 steps | 0 h | 0.157 |
| R/W FLOW | 0 steps | 0 h | 0.087 |
| R/W PH | 9 steps | 18 h | -0.037 |
| ALUM | 12 steps | 24 h | -0.064 |

As shown in Table 5, ALUM has the longest lag among the listed variables. The extended ARX model achieved an R2 of 0.800 and an RMSE of 0.319. Granger causality testing further showed that R/W FLOW has significant predictive value for filtered-water turbidity at lag 7, with p = 7.96e-5. This suggests that hydraulic load changes are not only correlated with filtered-water turbidity but also contain dynamic predictive information.

## 4.5. Water-Quality Risk Assessment

Entropy weighting showed that FILT. NTU had the highest weight, 0.553. Treated-water turbidity-related indicators had a weight of 0.295, R/W NTU had a weight of 0.141, and CL2 had a relatively small weight of 0.011. Thus, filtered-water turbidity is the dominant risk-assessment indicator, while raw-water turbidity provides auxiliary explanatory information.

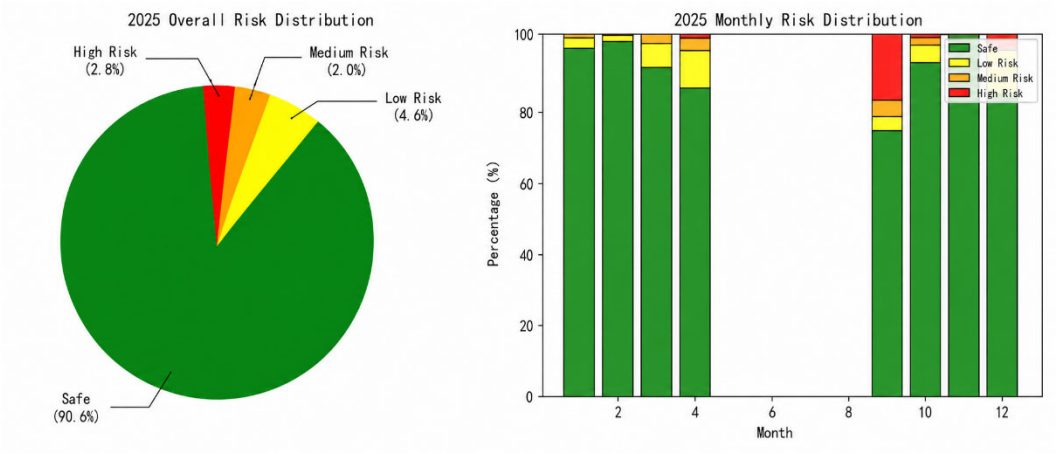


**Figure 6.** Distribution of water-quality risk levels in 2025.

As shown in Figure 6, safe samples accounted for 90.6% of the 2025 dataset, while low-risk, medium-risk, and high-risk samples accounted for 4.6%, 2.0%, and 2.8%, respectively. The proportion of high-risk samples was relatively higher in September, possibly because of rainy-season raw-water disturbances. KMeans clustering verified that the centers of the safe, low-risk, medium-risk, and high-risk clusters were approximately 0.104, 0.570, 2.515, and 5.693 NTU, respectively, which is consistent with the proposed risk intervals in Table 2. The follow-up FILT. NTU values from January to March 2026 ranged from approximately 0.04 to 0.25 NTU and were all classified as safe.

# 5. Discussion

## 5.1. Engineering Interpretation

The dominant predictors help explain why filtered-water turbidity cannot be treated as an immediate copy of raw-water turbidity. Clear-water well level describes the condition on the downstream side of the filters, river level represents the external source-water setting, and lagged flow carries information about hydraulic transport. Their joint importance suggests that a given raw-water turbidity value can lead to different filtered-water responses under different storage and flow conditions. SHAP adds a sample-level view to this general ranking, allowing a high prediction to be linked to the measurements that actually raised it.

The estimated delay for alum is also operationally relevant. A dose change and the resulting filtered-water response should not be paired at the same timestamp because water must pass through coagulation, sedimentation and filtration before the effect is observed. The ARX model retains this sequence through lagged exogenous terms and the recent history of filtered-water turbidity. Its structure is simpler than the boosted ensemble, but that simplicity makes the timing of the fitted response easier to inspect.

## 5.2. Limitations and Future Work

Several constraints should be considered when reading the results. All development records came from one treatment plant, so the ranking and model coefficients may change at plants with different source water or unit configurations. The monitoring export also lacked rainfall, temperature, filter-specific condition, backwash timing and detailed adjustment logs. Their absence is particularly important for the short, high-turbidity episodes that the model underestimated. In addition, the follow-up period covered only the first three months of 2026 and contained no unsafe samples. Validation over a longer period, across seasons and at additional plants is therefore needed. Future updates should incorporate hydrometeorological inputs and allow the model to be recalibrated as new operating data accumulate.

# 6. Conclusions

This study developed an interpretable dynamic prediction and risk assessment method for filtered-water turbidity based on monitoring data from an urban drinking water treatment plant. The main conclusions are as follows.

(1) Clear-water well level, river level, lagged raw-water flow, raw-water color, and current raw-water flow are the main factors affecting filtered-water turbidity, with clear-water well level having the highest comprehensive importance.

(2) The XGBoost-LightGBM weighted ensemble achieved good prediction performance, with a training R2 of 0.834. SHAP analysis verified the contribution of key variables and provided operational interpretability.

(3) The ARX dynamic model revealed lagged responses among input variables. The effect of ALUM appeared at approximately a 24-hour time scale, and the extended ARX model achieved an R2 of 0.800.

(4) The fuzzy comprehensive evaluation and entropy-weighted risk model converted continuous turbidity values into operational risk levels. In the 2025 dataset, safe samples accounted for 90.6%, and all follow-up samples from January to March 2026 were classified as safe.

Overall, the proposed method combines prediction accuracy, interpretability, and risk expression, and can provide auxiliary decision support for online monitoring, abnormal-warning, and operational adjustment in drinking water treatment plants.

# References

[1] World Health Organization. (2022). Guidelines for drinking-water quality: Fourth edition incorporating the first and second addenda. World Health Organization. https://www.who.int/publications/i/item/9789240045064

[2] Zhu, M., Wang, J., Yang, X., Zhang, Y., Zhang, L., Ren, H., Wu, B., & Ye, L. (2022). A review of the application of machine learning in water quality evaluation. *Eco-Environment*

*& Health*, 1(2), 107–116. https://doi.org/10.1016/j.eehl.2022.06.001

[3] Cojbasic, S., Dmitrasinovic, S., Kostic, M., Turk Sekulic, M., Radonic, J., Dodig, A., & Stojkovic, M. (2023). Application of machine learning in river water quality management: A review. *Water Science and Technology*, 88(9), 2297–2308. https://doi.org/10.2166/wst.2023.331

[4] Park, J., Ahn, J., Kim, J., & Yoon, Y., Park, J. (2022). Prediction and interpretation of water quality recovery after a disturbance in a water treatment system using artificial intelligence. *Water*, 14(15), Article 2423. https://doi.org/10.3390/w14152423

[5] Kwarko-Kyei, J., Tornyeviadzi, H. M., & Seidu, R. (2025). A machine learning approach to predicting the turbidity from filters in a water treatment plant. *Water*, 17(20), Article 2938. https://doi.org/10.3390/w17202938

[6] Wang, X., Li, Y., Qiao, Q., Tavares, A., & Liang, Y. (2023). Water quality prediction based on machine learning and comprehensive weighting methods. *Entropy*, 25(8), Article 1186. https://doi.org/10.3390/e25081186

[7] Makumbura, R. K., Mampitiya, L., Rathnayake, N., Meddage, D. P. P., Henna, S., Dang, T. L., Hoshino, Y., & Rathnayake, U. (2024). Advancing water quality assessment and prediction using machine learning models coupled with explainable artificial intelligence. *Results in Engineering*, 23, 102831. https://doi.org/10.1016/j.rineng.2024.102831

[8] Choudhary, R., Kumar, A., Priyadharsini, C., Naik, M. M., Choudhury, M., & Khan, N. A. (2025). Predicting water quality index using stacked ensemble regression and SHAP based explainable artificial intelligence. *Scientific Reports*, 15, Article 31139. https://doi.org/10.1038/s41598-025-09463-4

[9] Rahman, A., Syeed, M. M. M., Karim, M. R., Fatema, K., Khan, R. H., & Uddin, M. F. (2025). An optimized ensemble ML-WQI model for reliable water quality prediction by minimizing the eclipsing and ambiguity issues. *Applied Water Science*, 15, Article 113. https://doi.org/10.1007/s13201-025-02450-0

[10] Ruan, J., Cui, Y., Meng, D., Wang, J., Song, Y., & Mao, Y. (2023). Integrated prediction of water pollution and risk assessment of water system connectivity based on dynamic model average and model selection criteria. *PLOS ONE*, 18(10), Article e0287209. https://doi.org/10.1371/journal.pone.0287209