ICA2025OSLO
14th IWA Specialised Conference on
Instrumentation, Control and Automation

# Title: Investigation on Improving Accuracy of Predicting Filtered Water Turbidity in Rapid Sand Filtration Processes

Hiroyuki Yamahara*, Seiichi Murayama, Suguru Yokoyama, Michiaki Kanadani

*1 Toshiba-cho, Fuchu-shi, Tokyo 183-8511, Japan, hiroyuki.yamahara@toshiba.co.jp

**Abstract:** In water purification plants with rapid sand filtration in Japan, the coagulation-sedimentation process removes most of the suspended solids from the raw water, and the remaining suspended components are removed in the subsequent filtration process. It is common practice to operate on the safe side by ensuring the removal during the coagulation-sedimentation process and reducing the load on the subsequent filtration process. This means that the filtration process is operated with some spare capacity. If the treatment capacity of the filtration process can be modelled according to changes in the inflow water volume and turbidity, it is possible to calculate the range of inflow water quality that the filtration process can handle. As a result, it is expected to optimize the operation by reducing the coagulant injection amount while maintaining the desired treated water quality through the coagulation-sedimentation and filtration processes by reducing the load on the coagulation-sedimentation process based on the calculation. Against this background, we are developing a filtration process model to predict the turbidity of filtered water and the head loss in the filtration process from the inflow water quality and filtration flow rate, aiming to optimize the coagulation-sedimentation and filtration processes in water purification plants with rapid sand filtration. This paper reports that the prediction accuracy of filtered water turbidity can be improved by correcting the prediction results of a prediction model expressed by physicochemical equations with a machine learning model.

**Keywords**: Rapid sand filtration; Turbidity prediction; Optimization; Machine learning

## 1. Introduction

Artificial intelligence and machine-learning technologies have demonstrated their ability to optimize, model, and automate water-based applications (Lowe et al. 2022). We aim to predict filtration behaviour and optimize operations in water purification plants by combining these technologies with methods based on physicochemical knowledge. In a previous report (Murayama et al., 2021), we have focused that the small suspended components and the residual coagulant components in the inflow water to the filtration process improve the turbidity removal performance in the filtration process based on physicochemical knowledge. The reports describe that it is possible to improve the prediction accuracy of filtered water turbidity by using a model that changes the filtration coefficient λ related to turbidity removal performance for each divided filtration layer according to the aluminium concentration derived from the coagulant in the inflow water to the filtration process. This paper reports the results of evaluating a method that improves the prediction accuracy by correcting values predicted based on the previously reported prediction model expressed by physicochemical equations with a machine learning model.

## 2. Experimental Data

We collected experimental data with an experimental installation capable of simulating a series of water purification processes as shown in Figure 1, using raw water from a water purification plant that sources water from a dam lake. We measured raw water turbidity, sedimentation water temperature, sedimentation water turbidity, head loss, and filtered water turbidity. The filtration tower is made of transparent PVC with an inner diameter of 7.5 cm and is a gravity-type downward flow with an open top. The filtration layer thickness was set to 80 cm. The PAC injection rate and filtration flow rate were controlled to arbitrary values according to the experimental conditions. Sedimentation water turbidity and filtered water turbidity were measured alternately at regular intervals using a single turbidimeter. Therefore, for times when values were missing, linear interpolation was performed using the preceding and following measurements.

For evaluation, we collected experimental data with seven cases under different conditions. Table 1 shows the average values of filtration flow rate, filtration speed, PAC injection rate, sedimentation water temperature, raw water turbidity, and sedimentation water turbidity for each

case. Each case consists of data taken every 30 minutes after backwashing a filter until head loss rises to 100-130 cm. The number of collected data samples for each case is also shown in Table 1. Cases 1-4 have sedimentation water turbidity of 0.57-0.74 degrees, indicating stable filtration is expected in the filtration process. On the other hand, cases 5-7 have higher raw water turbidity and lower PAC injection rates compared to cases 1-4, resulting in higher sedimentation water turbidity of 1.36-1.77 degrees.

3. Models

We propose a combination of two models to predict filtered water turbidity. One is a prediction model, which derives base predicted values of filtered water turbidity by physicochemical equations. The other is a correction model, which corrects the base predicted values using machine learning.

First, we describe the prediction model. In the prediction model, a sand filter is divided into several layers and turbidity reduction in each layer is calculated with the filtration coefficient $\lambda$ different among layers. The higher the filtration coefficient $\lambda$, the more turbidity is removed. Equation 1 shows the relationship between the prediction model and experimental data items.

$$\begin{aligned} \boldsymbol{TB_{filter}} &= \boldsymbol{f}\left(\boldsymbol{TB_{sed}}, \boldsymbol{\lambda_{(i)}}, \cdots, \boldsymbol{\lambda_{(L)}}\right) \\ \boldsymbol{\lambda_{(i)}} &= \boldsymbol{f}\left(\boldsymbol{C_{Al}}, \boldsymbol{U}, \boldsymbol{T}, \boldsymbol{TB_{sed}}, \boldsymbol{TB_{in(i)}}\right) \\ \boldsymbol{C_{Al}} &= \boldsymbol{f}(\boldsymbol{I_{PACl}}, \boldsymbol{TB_{raw}}, \boldsymbol{TB_{sed}}) \\ \boldsymbol{TB_{out(i)}} &= \boldsymbol{f}\left(\boldsymbol{TB_{in(i)}}, \boldsymbol{\lambda_{(i)}}, \boldsymbol{z_{(i)}}\right) \end{aligned} \tag{1}$$

Here, $TB_{filter}$ is the predicted value of filtered water turbidity, $TB_{sed}$ is the sedimentation water turbidity, $\lambda_{(i)}$ is the filtration coefficient of the $i$-th layer, $L$ is the number of divided layers, $C_{Al}$ is the aluminium concentration derived from the coagulant in the sedimentation water, $U$ is the filtration speed, $T$ is the sedimentation water temperature, $TB_{in(i)}$ is the turbidity at the entrance of the $i$-th layer, $I_{PACl}$ is the PAC injection rate, $TB_{raw}$ is the raw water turbidity, $TB_{out(i)}$ is the turbidity at the exit of the $i$-th layer, and $z_{(i)}$ is the thickness of the $i$-th layer.

The predicted value of filtered water turbidity is calculated from the sedimentation water turbidity and the filtration coefficients of each layer. The filtration coefficients of each layer are calculated from the aluminium concentration derived from the coagulant in the sedimentation water, the filtration speed, the sedimentation water temperature, the sedimentation water turbidity, and the turbidity at the entrance of each layer. In this paper, the filtration layer is divided into five layers. The turbidity at the entrance of the first layer is the sedimentation water turbidity itself. The turbidity at the entrance of the second to fifth layers corresponds to the turbidity at the exit of the layer above. Therefore, the filtration coefficient and the turbidity at the exit are calculated sequentially from the first layer. The turbidity at the exit of the fifth layer corresponds to the predicted value of filtered water turbidity.

Next, we describe the correction model. Since the filtration coefficient is the most important in the prediction model described above, if the calculation formula of the filtration coefficient does not sufficiently represent the actual filtration behaviour, the error in the predicted value of filtered water turbidity tends to be large. Therefore, we hypothesize that correcting the error caused by the insufficient representation of the filtration coefficient calculation formula with data-driven methods will improve the prediction accuracy of filtered water turbidity. We consider the correction model shown in Equation 2.

$$\begin{aligned} \boldsymbol{TB_{correct}} &= \boldsymbol{f}(\boldsymbol{C_{Al}}, \boldsymbol{U}, \boldsymbol{T}, \boldsymbol{TB_{sed}}, \boldsymbol{TB_{in}}) \\ \boldsymbol{TB_{filter_crr}} &= \boldsymbol{TB_{filter}} + \boldsymbol{TB_{correct}} \end{aligned} \tag{2}$$

Here, $TB_{correct}$ is the correction amount to the predicted value, $TB_{in}$ is the turbidity at the entrance of each layer, $TB_{filter_crr}$ is the corrected predicted value, and $TB_{filter}$ is the predicted value of the base prediction model. The corrected predicted value is obtained by adding the

correction amount $TB_{correct}$ derived by the correction model to the predicted value of the base prediction model.
The data items used in the calculation of the filtration coefficient shown in Equation 1 are used as explanatory variables for the correction model. The correction model is a model that derives the correction amount by combining measured values and internal calculation values. Among the explanatory variables, the filtration speed, sedimentation water temperature, and sedimentation water turbidity are measured values. The aluminium concentration derived from the coagulant and the turbidity at the entrance of the second and subsequent layers are internal calculation values obtained in the process of prediction by the prediction model. The measured values are expected to provide information on the state of the inflow water quality, and the internal calculation values are expected to suggest the error of the prediction.

## 4. Evaluation

We conducted cross-validation on the seven cases of experimental data collected to evaluate whether the correction model can improve the prediction accuracy of filtered water turbidity. For example, when evaluating case 1, we generate a correction model using the samples from cases 2-7 as training data and predict the filtered water turbidity for case 1. Then, we compare the predicted values before and after correction and check whether the error with the true value (measured value) is reduced. This process is repeated by changing the evaluation target to cases 1-7 sequentially.
Figure 2 shows the measured values (the true values) of filtered water turbidity and the predicted values before and after correction with line charts and shows the mean absolute prediction error before and after correction for each case with a table.
First, focusing on the predicted values before correction, cases 1-4 show values about 0.05 degrees lower than the measured values. In cases 5 and 6, the difference is even larger, with predicted values more than 0.1 degrees lower than the measured values. Case 7 shows predicted values close to the measured values, with a mean absolute prediction error of about 0.004 degrees. The prediction results before correction, especially in cases 5 and 6, suggest that the prediction model does not sufficiently represent the filtration behaviour under conditions of high sedimentation water turbidity and low PAC injection rates.
Next, focusing on the predicted values after correction, the prediction error in cases 1-4 is reduced from about 0.05 degrees before correction to about 0.01 degrees after correction. Similarly, in cases 5 and 6, the prediction error is reduced from more than 0.1 degrees before correction to about 0.03 degrees after correction. In case 7, the prediction error before correction is small at 0.004 degrees, so the correction amount should ideally be close to zero, but the prediction error worsened to 0.099 degrees after correction. The reason the correction model did not function as expected in case 7 is thought to be because the training data did not include situations similar to case 7. As mentioned earlier, cases 1-4 are similar in filtration behaviour, with low sedimentation water turbidity and sufficiently low filtered water turbidity. Compared to cases 1-4, cases 5-7 have higher sedimentation water turbidity and lower PAC injection rates. Cases 5-7 are similar in this respect. Among these, cases 5 and 6 are similar in filtration behaviour, including high filtered water turbidity. Case 7, however, has low filtered water turbidity, making its filtration behaviour different from cases 5 and 6. In other words, case 7 does not have other cases with similar filtration behaviour. In this evaluation, the correction model for case 7 was generated using cases 1-6 as training data, so it did not become a model that could obtain the desired correction amount for case 7. As a result, an intermediate correction amount between cases 1-4 and cases 5-6 was obtained, correcting the predicted value in the direction of increasing it.

## 5. Conclusion

This paper reported that the prediction accuracy of filtered water turbidity can be improved by correcting the prediction results of a prediction model expressed by physicochemical equations

with a machine learning correction model. This research, which aims to optimize process control in water purification plants, is an important topic of this conference. It is of great significance that the usefulness of the approach combining machine learning models with physicochemical models was evaluated using actual raw water data.

**Table 1** Experimental conditions.

| case no. | filtration flow rate (mL/min) | filtration speed (m/day) | PAC injection rate (mg/L) | sedimentation water temperature (°C) | raw water turbidity (degree) | sedimentation water turbidity (degree) | number of samples |
|---|---|---|---|---|---|---|---|
| 1 | 377 | 123 | 15 | 12.9 | 1.26 | 0.57 | 49 |
| 2 | 281 | 92 | 15 | 12.3 | 1.16 | 0.74 | 56 |
| 3 | 379 | 124 | 20 | 18.6 | 2.52 | 0.65 | 65 |
| 4 | 380 | 124 | 10 | 18.3 | 2.62 | 0.70 | 83 |
| 5 | 380 | 124 | 8 | 6.5 | 4.57 | 1.77 | 132 |
| 6 | 379 | 124 | 10 | 6.1 | 4.50 | 1.54 | 127 |
| 7 | 473 | 154 | 12 | 5.8 | 4.56 | 1.36 | 108 |

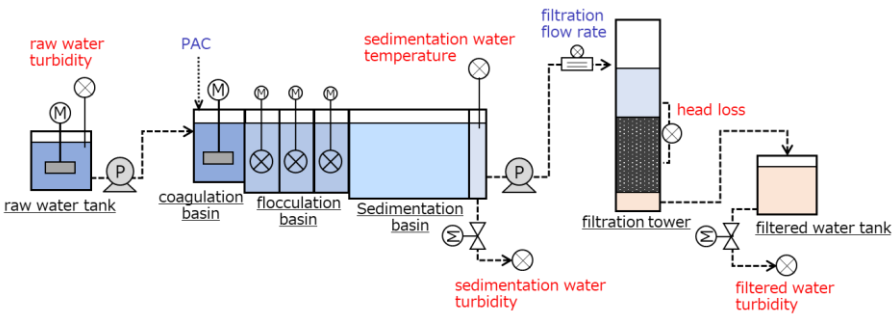


**Figure 1** Experimental installation.

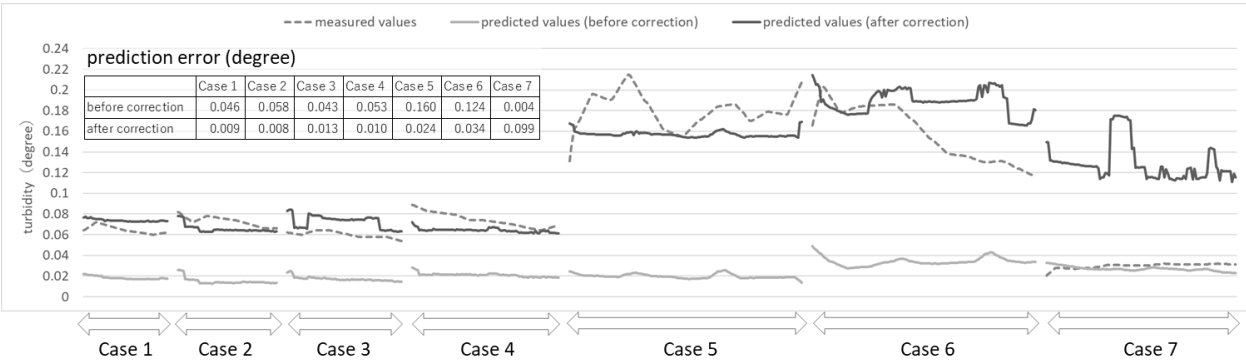


**Figure 2** Results of correction for the prediction model outputs.

## REFERENCES

Lowe, M., Qin, R. & Mao, X. 2022 A review on machine learning, artificial intelligence, and smart technology in water treatment and monitoring. *Water.* 14(9), 1384. https://doi.org/10.3390/w14091384

Murayama, S., Fukuda, M., Menju, T., Yokoyama, S., Kanadani, M. & Kurokawa, F. 2021 Modeling of turbidity in rapid filtration. *Journal of EICA.* 26(2/3), 16–19. (in Japanese)