Environmental Science
Water Research & Technology

Accepted Manuscript

This article can be cited before page numbers have been issued, to do this please use: C. Ortiz-Lopez, C. Bouchard and M. J. Rodriguez, *Environ. Sci.: Water Res. Technol.*, 2026, DOI: 10.1039/D6EW00235H.

View Article Online
DOI: 10.1039/D6EW00235H

Water impact statement

This research shows that integrating upstream water-quality data, river flow, and rainfall, greatly improves predictions of source-water turbidity. More accurate, timely predictions can help drinking-water facilities anticipate treatment challenges, supporting real-time early-warning systems during and after rainfall events and safeguarding water quality. These models outline a clear pathway toward proactive, climate-resilient water management.

Enhancing source water quality predictions to improve treatment by integrating watershed data on water quality, river flow and rainfall into interpretable machine learning algorithms.

View Article Online
DOI: 10.1039/D6EW00235H

Christian Ortiz-Lopez [a]*, Christian Bouchard [a], and Manuel Rodriguez [b]

[a] Centre de recherche en aménagement et développement, Université Laval, Québec, Canada.

[b] École supérieur d'aménagement du territoire et de développement régional, Université Laval, Québec, Canada.

* Corresponding Author: christian.ortiz-lopez.1@ulaval.ca

View Article Online
DOI: 10.1039/D6EW00235H

Abstract

Raw water quality used for municipal water treatment is impacted during and after rainfall events and the resultant changes in river flow. Recently, raw water quality parameters such as turbidity have been modeled and predicted using machine learning algorithms, based on environmental, hydrological, and meteorological information as input variables. Our research aims to integrate upstream water quality with river flow and watershed rainfall data into interpretable machine learning algorithms to enhance raw water turbidity predictions. Such predictions would allow water utility operators to anticipate the required adjustments during water treatment processes. First, we estimated lag-times between the upstream input variables of rainfall in watershed, river flow and raw water turbidity, and the output targeted variable of downstream raw water turbidity. Then, we used a XGBoost technique to predict raw water turbidity using upstream water quality along with river flow and watershed rainfall data. Finally, the overall importance of every input variable was estimated using a SHAP (SHapley Additive exPlanations) strategy. Results showed that the upstream raw water turbidity is the most important input variable, followed by river flow. Best performance metrics and time series visual inspection of modeled variables showed that integrating upstream raw water quality data leads to enhanced raw water predictions. These results could open possibilities for developing and implementing regional raw water quality modeling that can feed Weather-Event-Water-Treatment–Early-Warning-Systems (WEWT-EWS). Future research could improve raw water quality prediction horizons and include interannual data.

Keywords: Raw water quality, drinking water, interpretable machine learning algorithms, river flow, watershed rainfall.

View Article Online
DOI: 10.1039/D6EW00235H

1. Introduction

Raw water quality is a crucial aspect of drinking water treatment and production, along with environmental, economic and technical elements. Evidence shows that surface source water quality is increasingly degraded by industrial and agricultural activities, growing urbanization, and weather events such as wildfires, extreme heat and cold, droughts, superstorms, and heavy rainfalls and floods (1, 2). Several previous studies have shown that raw water quality is negatively impacted during and after rainfall in watersheds. Effects on raw water quality can last between a few hours and several days depending on whether the parameter targeted is particles (turbidity), natural organic matter (NOM), or microorganisms (3). Turbidity (Tu) is a critical indicator for drinking water treatment since its presence and removal directly impact physical and microbiological drinking water quality. It has been demonstrated that raw water Tu increases from baseline values after rainfall, due to the accompanying higher river flow (4-6). Those increases, usually called peaks, can last for hours.

Spatiotemporal variability of source water quality parameters, such as Tu, can affect drinking water treatment and production, and lag-times can be an issue in determining a correct and timely treatment response (7). The first lag-time is between the moment when raw water Tu starts to increase and the moment the required coagulant dosage adjustment is determined at the Drinking Water Treatment Plant (DWTP). Then, there is a second lag-time before verification that this dosage adjustment has been effective in treated or produced water. During these lag-times, potential non-optimal operating conditions and poor treatment performance due to insufficient consideration of changes in raw water quality may lead to higher risks to public health (8).

Early Warning Systems (EWS) have been developed to give decision-makers timely

View Article Online
DOI: 10.1039/D6EW00235H

alerts about contaminant events in surface waters (9, 10). Characterizing the water contamination event is a crucial function of an EWS. Nowadays, characterization can be carried out by means of modeling and forecasting, with machine learning emerging as a cost-effective and accurate tool to model and forecast raw water quality parameters such as Tu (7, 11, 12). Machine learning techniques take advantage of large amounts of data produced at DWTPs and/or hydrological and meteorological meter stations to find empirical relationships.

Previous studies have reported the relatively high accuracy of machine learning in predicting crucial raw water quality parameters, such as Tu. For instance, Alizadeh et al. (2018) Alizadeh, Kavianpour (13) implemented Artificial Neural Networks (ANN), Extreme Learning Machines (ELM), and Support Vector Regressions (SVR) to model Tu and other physical parameters in an estuary using upstream river flow as an input variable with a lag-time of up to three hours. Results found that using those lag-times lead to R^2 between 0.89 and 0.89 and a root mean square error (RMSE) between 1.04 and 1.51 nephelometric turbidity units (NTU) during the test period. In another study by Delpla et al. (2019) Delpla, Florea (9), ANN were employed to predict daily Tu in source water using rainfall and antecedent soil moisture conditions (antecedent dry days), obtaining a R^2 of 0.81 and a mean square error (MSE) of 1.08 NTU. Ahmed et al. (2021) (14) applied several machine learning techniques such as Decision Trees (DT), k-Nearest Neighbour (KNN), Logistic Regression (LogR), ANN, and Naive Bayes (NB) to classify Tu and other physical and microbiological parameters used to calculate a water quality index in a dam reservoir. They found that DT outperformed other models and demonstrated the need to aggregate other hydrological parameters such as rainfall and river flow. Zhang et al. (2021) (15) used Random Forest (RF) to model raw water Tu in a lake based on meteorological data such as wind field, air temperature, and rainfall, finding R^2

View Article Online
DOI: 10.1039/D6EW00235H

coefficients between 0.73 and 0.90. However, they did not consider lag-times in the input variables. Adedeji et al. (2022) (16) implemented SVR, ANN, RF, and XGBoost to model physical and microbiological variables such as total phosphorus, total nitrogen, suspended solids, dissolved oxygen, and fecal coliform bacteria. They used preset modeling scenarios that include environmental, hydrological and water quality input variables. One scenario included precedent Tu in the same location as the output target as an input variable. Ortiz-Lopez et al. (2023, 2024) (6, 17) developed and tested a methodology to include lagged rainfall and river flow as input variables to model raw water Tu and NOM (represented by UV absorbance) using ANN, SVR, RF, and XGBoost. Performance metrics show that XGBoost outperformed other machine learning techniques. However, modeling multiple Tu peaks remains to be improved. More recently, J. Chen et al. (2025) (18) predicted post-wildfire stream temperature and Tu using SVR and RF. For input variables, they used a number of antecedent moving averages of environmental, meteorological, and hydrological parameters. The best models showed performance metrics of R^2 between 0.87 and 0.89 and RMSE between 1.77 and 2.2 NTU using RF. Kemper et al. (2025) (19) forecasted raw water Tu in three different locations using the Gradient Boosting (GB) technique, with ten streamflow forecasts (precedent conditions and precedent peaks). Although upstream Tu was initially considered as a possible input variable, the final models only used river flow data. Models showed relatively high performance (R^2 of 0.7 and NSE between 0.48 and 0.49) when using GB. Recently, Zhang et al. (2025) (20) coupled a light GB model and a Long-Short Term Memory (LSTM) model to predict chemical oxygen demand (COD) using several input variables including upstream COD. The best model showed performance metrics of R2 of 0.826 for the stacked model and 0.802 for the LSTM. Yang et al. (2025) (21) developed a sub-daily machine learning approach for predicting eutrophication trends

View Article Online
DOI: 10.1039/D6EW00235H

using also upstream features as input values. GB and RF models presented the best performance metrics predicting Tu, total nitrogen and pH.

On the other hand, the use of data-driven and black box models for decision making can be problematic, since results cannot usually be interpreted. Several techniques have recently emerged to explain how predictions are performed by machine learning, such as linear models, neural networks and decision-tree-based techniques. For instance, the SHAP strategy can estimate the global importance of every input variable in the prediction of the output variable (22). This technique was recently used in several water quality modeling studies to explain and understand modeling results (23-26). According to the reported literature above, several machine learning techniques have been used to model raw water quality parameters such as Tu using hydrological and environmental variables. These studies favored the use of variables such as river flow, rainfall, number of dry or precipitation days, and water level, among others. However, including the targeted raw water variable measured upstream as an input variable has remained little explored until now, especially in the field of drinking water production.

Our research aims to demonstrate that modeling of raw water quality parameters such as Tu (especially peaks of turbidity) can be improved by integrating watershed upstream water quality as input information. Downstream DWTP turbidity models and predictive analyses can utilize upstream Tu measurements that are continuously monitored online. As a case study, we considered a watershed where the raw water source for at least three municipal water treatment intakes are supplied by the same river. We used the XGBoost machine learning technique (27) as a tool to estimate the empirical relationships between all input variables and the raw water Tu at the targeted intake. To verify whether adding upstream Tu to input variables improved raw water Tu modeling, we coupled this

View Article Online
DOI: 10.1039/D6EW00235H

technique with a SHAP strategy (22) that estimates the overall feature importance and local decision interpretation of the machine learning model. XGBoost is a robust and interpretable technique which aligns with SHAP strategy. The novelty of this study, beyond the improvement in raw water quality predictions, is the use of information produced at upstream DWTP as decision-support aids. Timely and accurate predictions could provide improved input for developing tools for drinking water treatment managers and operators to enhance response strategies to variations in raw water quality that affect treatment efficiency. To the best of our knowledge, this is the first study to develop raw water Tu modeling for drinking water production purposes that includes upstream raw water quality, river flow, and rainfall as input variables within the same framework.

2. Methodology

2.1. Study sites and data

Our case study watershed is the Chaudière River in the province of Québec, Canada. This watershed covers approximately 6,711 km^2 and consists of 66% forested areas, 17% agricultural zones, 11% wetland areas, 4% urbanized areas, and 2% water bodies. The Chaudière River is 195 km long, beginning in Mégantic Lake in the Appalachian Mountains, near the border between Canada and the United States and flowing into the Saint Lawrence River near Quebec City. The watershed is characterized by a temperate climate, with warm and very humid summers and cold winters. The average annual air temperature in the watershed ranges between 2.9 and 4.6°C (1991–2020 reference). Annual precipitation has ranged between 1,030 and 1,367 mm $year^{-1}$ over the past 40 years (28). Figure 1 presents the location of watershed, DWTPs and intakes, flow meter stations, and rain gauge stations used in this research.

View Article Online
DOI: 10.1039/D6EW00235H

In this case study, three DWTPs and intakes were considered: DWTP1 – Charny (CH), DWTP2 – Sainte-Marie (SM) and DWTP3 – Saint-Georges (SG), all of which take their raw water from Chaudière River. The raw water quality parameter targeted is Tu, which was measured at the three intakes by online analyzers. Data were collected from three river flow meters along the Chaudière River, Q1 – Saint-Lambert-de-Lauzon, Q2 – Saint-Georges and Q3 – Saint-Martin, and from four rain gauge stations, P1 – Scott, P2 – Saint-Severin, P3 – Saint-Georges and P4 – Saint-Ludger (29, 30). The data set included raw water quality and hydrological and meteorological information, with more than 4,000 hourly observations from April to October 2017. Due to technical issues, an average of 10.5% of observations were lost from the Tu series. On average, 60 gaps in observations lasting no longer than 24 hours and fewer than 5 gaps lasting no longer than 10 days were found in the Tu series. Flowmeter data and rainfall accumulation data were preprocessed by the Ministry of the Environment of the province of Quebec. We implemented several techniques to find and eliminate outliers (as described in (3)) and to fill in missing data, including Kalman smoothing with an ARIMA model (31) to impute missing data. The data collection period captured the behavior of raw water quality during different dry and rainy periods (between April and October). In a Nordic country, the period considered is when turbidity varies the most. During winter, early spring, and late autumn, turbidity variations are much weaker. This period (April to October) excludes winter and early spring and interannual variability. Hydrological and meteorological data from raw water were collected either hourly (precipitation) or sub-hourly (Tu and river flow) and transformed to hourly (average) frequency.

View Article Online
DOI: 10.1039/D6EW00235H

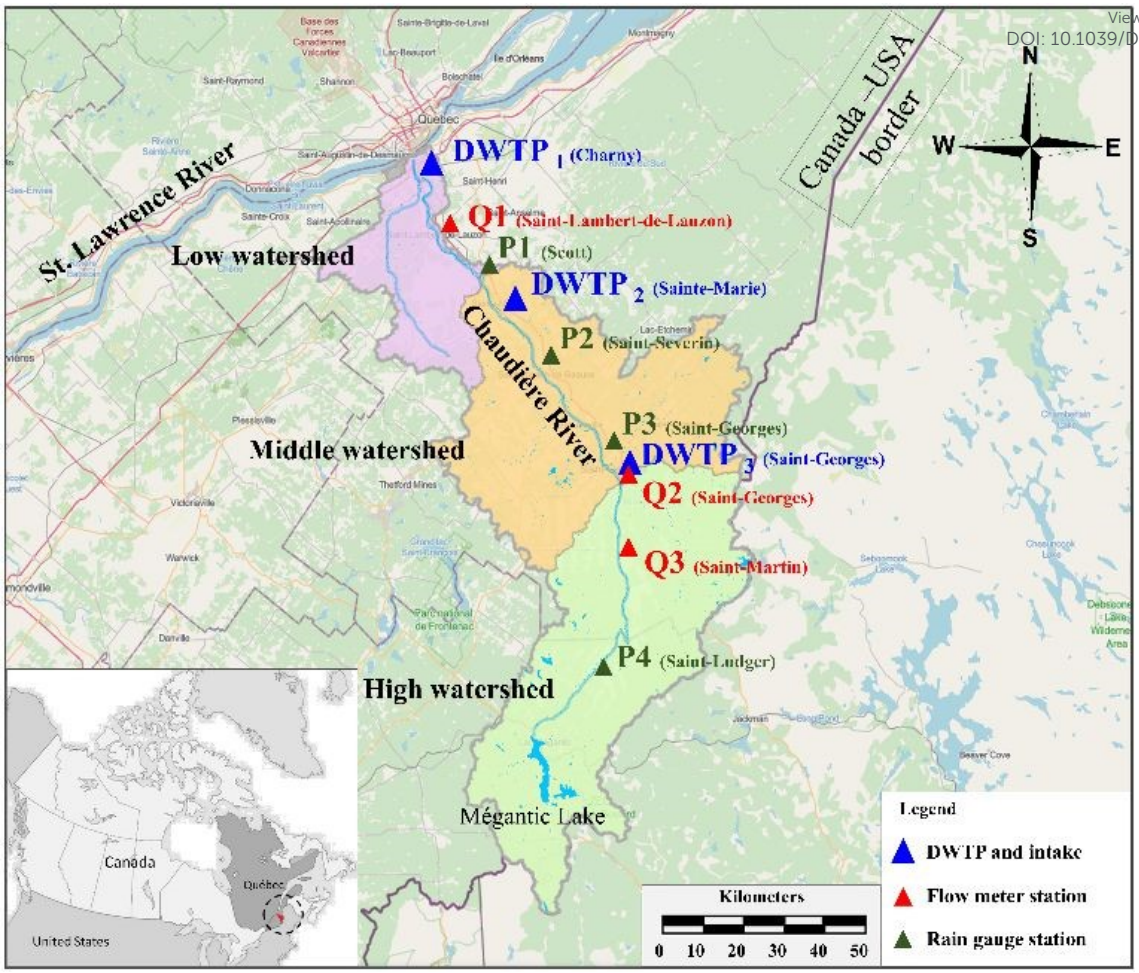

Figure 1. Location of case study watershed, showing drinking water treatment plants (DWTP) and intakes, flow meter stations and rain gauge stations.

Table 1 presents a summary of primary statistics for all the variables used in this study. Tu values observed throughout the Chaudière River were relatively low, although severe peaks did occur. Previous studies have shown that those Tu peaks (as well as peaks of other relevant raw water quality variables) are observed during and after rainfall events in the watershed (3, 6). We observed that maximum river flow values occurred in the month of April, a period when snowmelt and ice jams result in spring flooding. During the summer period, peak flow rates were around 200 to 400 m^3 s^{-1}. We also noted that river flow gradually increased from upstream to downstream (Q3 to Q1), because of various tributaries flowing into the mainstem. Time series plots of river flow, precipitation and raw water Tu are shown in Supplementary Material (Figures S1-3).

View Article Online
DOI: 10.1039/D6EW00235H

Table 1. Summary of primary statistics for all the variables used in this study. SD = standard deviation.

Data	Location	Variable	Symbol	Unit	Mean	Min.	Max.	S.D.
Raw water quality	Charny DWTP	Turbidity	Tu_{CH}	NTU	9.7	1.3	152.1	15.0
	Sainte-Marie DWTP	Turbidity	Tu_{SM}	NTU	7.4	0.6	100.0	13.5
	Saint-Georges DWTP	Turbidity	Tu_{SG}	NTU	15.1	4.6	225.1	20.0
Hydrological	Saint-Lambert municipality	River flow	Q_1	$m^3 s^{-1}$	121.3	8.5	1580.3	206.2
	Saint-Georges municipality	River flow	Q_2	$m^3 s^{-1}$	58.2	3.7	959.9	99.4
	Saint-Martin municipality	River flow	Q_3	$m^3 s^{-1}$	39.3	3.2	608.9	65.4
Meteorological	Scott municipality	Precipitation	P_1	mm h^{-1}	0.14	0.00	24.20	0.93
	Saint-Severin municipality	Precipitation	P_2	mm h^{-1}	0.15	0.00	19.20	0.80
	Saint-Georges municipality	Precipitation	P_3	mm h^{-1}	0.15	0.00	18.00	0.81
	Saint-Ludger municipality	Precipitation	P_4	mm h^{-1}	0.14	0.00	24.8	0.79

2.2. Conceptual models and input data definitions

Figure 2 shows three different models that were developed to predict Tu at three different water intakes, $DWTP_{1-(SG)}$, $DWTP_{2-(SM)}$, and $DWTP_{3-(CH)}$. Figure shows the conceptual models proposed in this study. They consist of the predictor (input) variables, which primarily include river flows and rainfall upstream of the targeted water intake, and the predicted variables, which include all raw water Tu measured at the water intake. Figurea shows the schematic locations of river flow meters (red labels), rain gauge stations (green labels) and DWTPs (blue labels) along the Chaudière River. Raw water Tu at the water intake is predicted using all upstream variables (river flow, rainfall, and raw water Tu) as shown in Figureb–d.

View Article Online
DOI: 10.1039/D6EW00235H

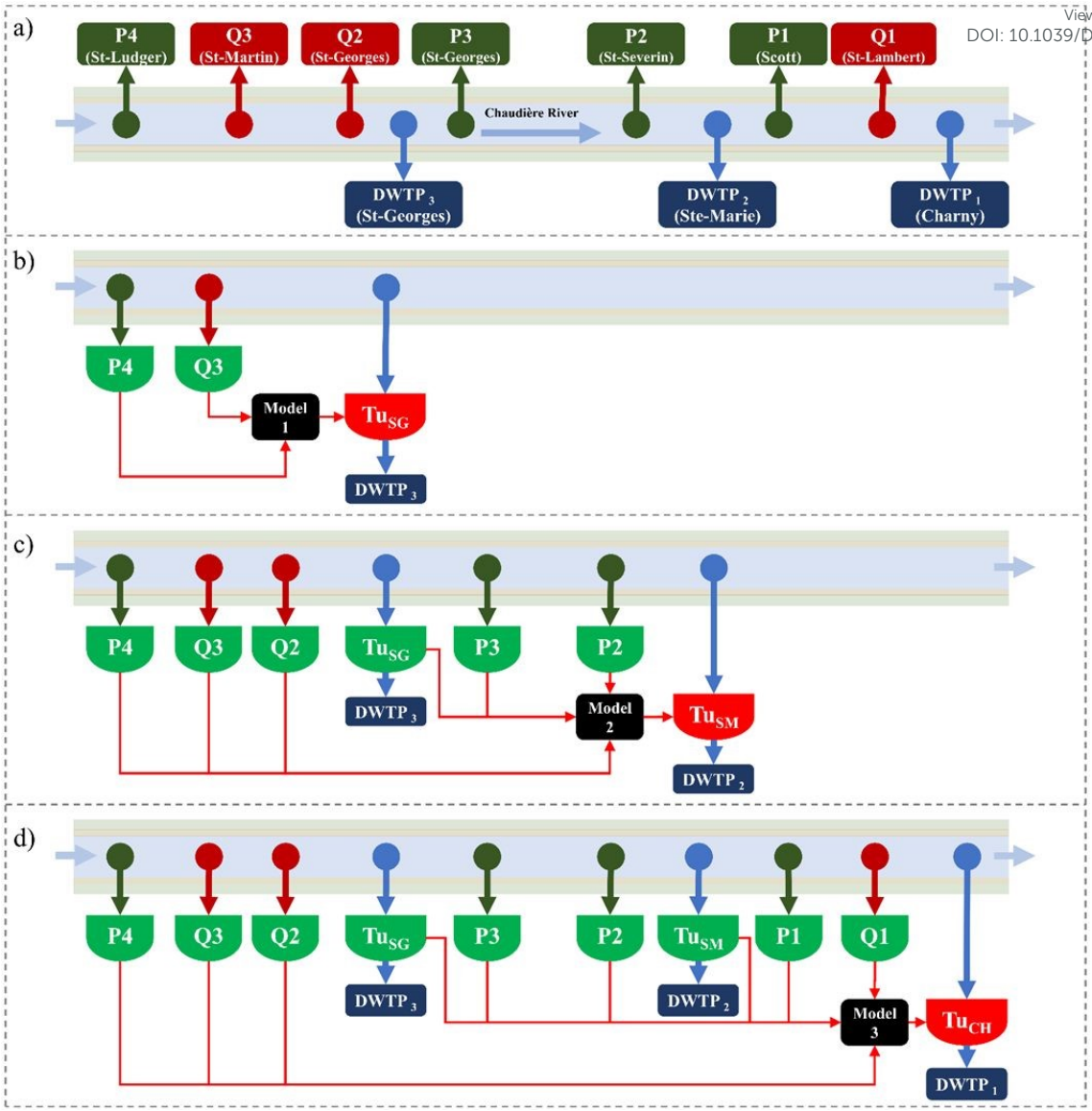

Figure 2. Conceptual models. (a) Conceptual location scheme of the DWTPs, flow meter stations and rain gauge stations in the source river; (b) Model 1, a conceptual model to predict Tu at $DWTP_3$; (c) Model 2, a conceptual model to predict Tu at $DWTP_2$; (d) Model 3, a conceptual model to predict Tu at $DWTP_1$.

2.3. General modeling methodology

Figure 3a shows a framework of the general methodology consisting of several sequential steps: (1) Data collection and preprocessing; (2) Input data selection; (3) Model optimization; (4) Model evaluation; and (5) Interpretability. Data collection and preprocessing, including cleaning and completing missing data, are described in Section 2.1.

View Article Online
DOI: 10.1039/D6EW00235H

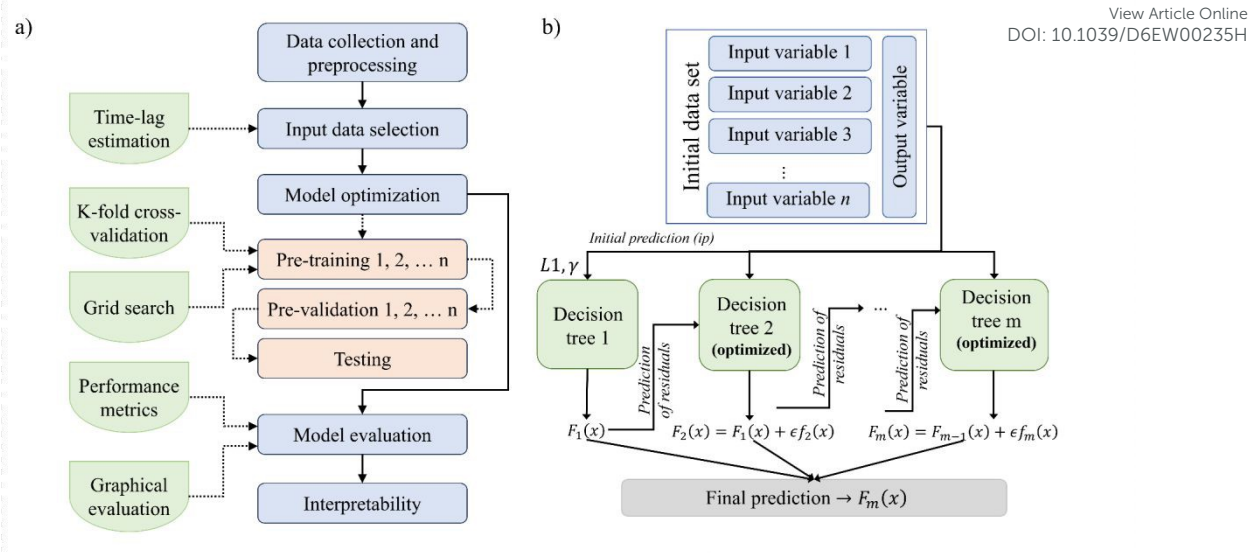

Figure 3. (a) Framework of the general modeling methodology. (b) Architecture of the XGBoost model.

2.3.1. Input data selection and model optimization

Because there are delays between rainfall events and the resulting changes in river flow and river water quality, it is necessary to estimate lag-times between input and output variables. We selected lag-times of input data using an approach already tested by Ortiz-Lopez et al., (2023, 2024) (6, 17), which calculates the Spearman's rank correlation coefficient (ρ) between the output variable and each of the input variables, with a lag-time of one hour. This process was repeated using time steps between one hour and 240 hours (10 days). For each input variable, the lag-time retained was the one with the highest correlation coefficient. Those lag-times were then estimated using the above-mentioned statistical approach, which is empirical and formally different from a hydraulic or physical process. Despite of the lag-times between every flowmeter station are not required in this study, those lag-times can be also estimated using the proposed statistical approach. Then, the statistically estimated lag-times can be compared with either reported data in the literature or simple calculations including typical flow velocities and travel distance. This approach using cross-correlations might lead to an underestimation of

View Article Online
DOI: 10.1039/D6EW00235H

hydraulics travel times. However, we consider the statistical approach sufficient for the purposes of raw water quality modelling which uses empirical techniques such as machine learning.

2.3.2. Machine learning technique

To develop and optimize the model, we selected the XGBoost machine learning technique, an optimized version of the Gradient Boosting (GB) library developed by Chen and Guestrin (2016) (27), as our modeling tool. The R xgboost software package (32, 33) was used to develop the XGBoost model, with the general architecture shown in Figure 3b. Building an XGBoost model involves several repeated steps that construct and optimize a decision tree. The first step is to make an initial prediction (ip) of the model and then calculate the residuals (r_i). Step two is to construct the decision tree, which is used and optimized throughout the entire algorithm. To construct the XGBoost regression tree: (1) similarity scores are calculated using a regularization parameter (L1) to reduce sensitivity to individual observations and reduce overfitting; (2) the gain is calculated to evaluate different thresholds; and (3) the constructed tree is "pruned" using a (γ) parameter, which is the minimum loss reduction required to make the next split on a leaf node. Once the single tree is pruned, new predictions are made ($f_{2,3,\ldots,m}(x)$) by scaling the initial predictions with a learning rate ε. The cumulated prediction made by each iteration corresponds to $F_{2,3,\ldots,m}(x)$. This process is repeated in a loop using the predicted residuals at each iteration until the maximum number of iterations or the desired minimum error is reached, yielding the final prediction ($F_m(x)$). We also tested a multi-linear regression (MLR) model as a baseline model to contextualize the results of the XGBoost model.

Data was randomly split using a training with a Cross-Validation (80%) and Testing

View Article Online
DOI: 10.1039/D6EW00235H

(20%) (CVT) approach (34). The model optimization involved using a 10-fold cross-validation approach for the training phase and grid search to find the optimized hyperparameters. After the training phase, the optimized model was tested using unseen data. The models were then evaluated using performance metrics to determine the quality of the predictions. Furthermore, predictions for Tu were compared to the observed data using scatter plots and time series graphs. Model predictive uncertainty was quantified using bootstrap prediction intervals (35, 36). We resampled the training dataset 1,000 times with replacement, refitting the XGBoost model for each bootstrap sample, and generating a distribution of predictions for each observation. The 5th and 95th percentiles of these bootstrap predictions defined the 90% prediction band, while the 50th percentile represented the median prediction. To obtain smooth and explainable uncertainty curves for visualization, the lower, median, and upper percentile estimates were further smoothed using a LOESS (locally weighted regression) technique, which reduced local variability and highlighted the overall trend of the uncertainty structure.

2.3.3. Model performance evaluation

Our research assessed the performance of predictive models using several metrics such as the Coefficient of determination, R^2 (Eq. 1), the Nash-Sutcliffe Efficiency (NSE) (37) (Eq. 2), the Root Mean Square Error (RMSE) (Eq. 3) and the Mean Absolute Error (MAE) (Eq. 4).

Equation 1. Calculation of R^2

$$R^2 = \frac{\sum_{i=1}^{n} \left(\hat{y}_i - \bar{y}_i\right)^2}{\sum_{i=1}^{n} \left(y_i - \bar{y}_i\right)^2}$$

Equation 2. Calculation of Nash-Sutcliffe efficiency (NSE)

View Article Online
DOI: 10.1039/D6EW00235H

$$NSE = 1 - \frac{\sum_{i=1}^{n} (\hat{y}_i - y_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y}_i)^2}$$

Equation 3. Calculation of the root mean square error (RMSE)

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2}$$

Equation 4. Calculation of the mean absolute error (MAE)

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |y_i - \hat{y}_i|$$

Where y_i is the measured data (Tu), $\hat{y}_i$ is the predicted data, $\bar{y}_i$ is the average of measured data and n is the total number of data observations.

R2 values, which range between 0 and 1, give the percentage of the total variance of the observed data that is explained by the model. Values of the NSE coefficient range between minus infinity and 1, with 1 representing a perfect fit and an efficiency of less than zero indicating that the mean value of the observed data is a better predictor than the model. The RMSE provides the standard deviation of the model prediction error. We chose RMSE because it is reliable and gives a relatively high weight to errors. Finally, MAE was used to show the absolute difference between the actual and predicted values. However, this metric only provides information about the extent of the error and not the model validity.

2.3.4. Interpretability strategy

We adopted a SHAP (SHapley Additive exPlanations) strategy to explain individual predictions of machine learning and data-driven models based on the concept of

View Article Online
DOI: 10.1039/D6EW00235H

sensitivity analysis (22). This methodology decomposes a prediction into a sum of contributions from each of the model's predictor variables. SHAP strategy can be seen as an evolution of Shapley value theory (38) which explains predictions by assuming that each predictor value is a "player" in a game where the prediction is the payout (39).

The SHAP value of predictor ϕ_i represents its influence on the prediction (upstream raw water Tu), which is calculated as a weighted summation across all possible predictor combinations. It is calculated following Eq. 5 ((22) and adapted in (40)).

Equation 5. Calculation of SHAP values

$$\phi_i(val) = \sum_{S \subseteq x_1,\ldots,x_p \setminus x_i} \frac{|S|!(p-|S|-1)!}{p!}\left[val(S \cup x_i) - val(S)\right]$$

Where S is the set of input variables or predictors used in the model, $x = x_1,\ldots,x_p$ is the input variable vector of the observation to be explained, p is the number of input variables and $val(S \cup x_i)$ and val(S) are the predictions (Tu) trained on S and $S \setminus x_i$ (remove input x_i from S) respectively. Therefore, the SHAP value is calculated by aggregating marginal contributions from all possible combinations of inputs through a weighted average (41). A positive SHAP value indicates that a given predictor will push the prediction of the target variable (Tu) above its mean value (the mean value of the set of input variables or predictors used in the training stage of the model), while a negative SHAP value indicates that the predictor pushes the prediction of the variable below the mean. In regression models such as our raw water Tu prediction model, each prediction value j can be reproduced as the sum of the SHAP values of observation j and a fixed base value, for instance the mean of the Tu (see Eq. 6).

Equation 6. Model output in SHAP strategy

View Article Online
DOI: 10.1039/D6EW00235H

$$f(x) = \text{base value} + \text{sum}(\text{SHAP values}) \rightarrow Tu_j = \overline{Tu} + \sum_{i=1}^{P} \phi_{i,j}(val)$$

The average absolute SHAP values are used to rank the importance of different predictors, including upstream river flow, rainfall, and input Tu in predicting target Tu, as visualized in a barplot of SHAP input variable importance plot. Additionally, SHAP summary plots (beeswarm plots) map the importance of an input variable against its effect on target Tu. On summary plots, the position on the x-axis corresponds to the Shapley value. Overlapping points are stacked in the y-axis direction, to provide a sense of the distribution of the Shapley values for a given input variable (39).

3. Results and discussion

In this section, we begin by presenting the results of estimated lag-times of input variables used in Tu models. Then, we compare model results using scatter plots, performance metrics and time series plots. Finally, we present the model interpretability using SHAP strategy, followed by the discussion.

3.1. Lag-times of input variables

Table 2 shows the estimated overall lag-times (in hours) between every output variable (Tu_{SG}, Tu_{SM}, and Tu_{CH}) and input variable (Q1–Q3, P1–P4, Tu_{SG}, and Tu_{SM}) used in our models. In general terms, lag-times are longer for river flow than for precipitation. This can be explained because of the rainfall-runoff transformation process. Thus, effects of rainfall on watersheds are expected to take more time to affect raw water quality than effects of river flow. For model 3, which uses all the input variables to predict Tu_{CH}, some aspects can be highlighted. P1 and P2 rain gauge stations, which measure rainfall on the middle watershed, have similar lag-times (38 and 40 hours, respectively), while P3 and

View Article Online
DOI: 10.1039/D6EW00235H

P4, which measure rainfall on the high watershed, also have similar lag-times (82 and 95 hours, respectively). Likewise, the Q1 river flow meter, which cumulated flow from both the middle and the high watershed, had a lower lag-time (9 hours) than Q2 and Q3 (19 and 20 hours of lag-time respectively), which gauged only flow produced in the high watershed. The time-lags shown in Table 2 were used when incorporating input variables into the three proposed models (Tu_{SG}, Tu_{SM}, and Tu_{CH}). Despite not being reported here (since they are not necessary for the purposes of modeling), lag-times between all flowmeter station were estimated following the statistical approach and then compared with simple estimates of travel times (using typical flow velocities and travel distances). This statistical approach provided a sufficient approximation to hydraulic-based estimations.

Table 2. Estimated lag-time (in hours) between every output variable and every input variable included in its model.

Model	Output variable	Input variable								
		Q3	Q2	Q1	P4	P3	P2	P1	Tu_{SG}	Tu_{SM}
1	Tu_{SG}	4	NA	NA	29	NA	NA	NA	NA	NA
2	Tu_{SM}	14	15	NA	49	37	NA	NA	22	NA
3	Tu_{CH}	20	19	9	95	82	38	40	61	31

3.2. Turbidity models

Figure 4 scatterplots show the model performance by comparing predicted and observed Tu at the three DWTP intakes (Models 1–Tu_{SG}, 2–Tu_{SM}, and 3–Tu_{CH}). Both high and low Tu values were relatively well predicted.

View Article Online
DOI: 10.1039/D6EW00235H

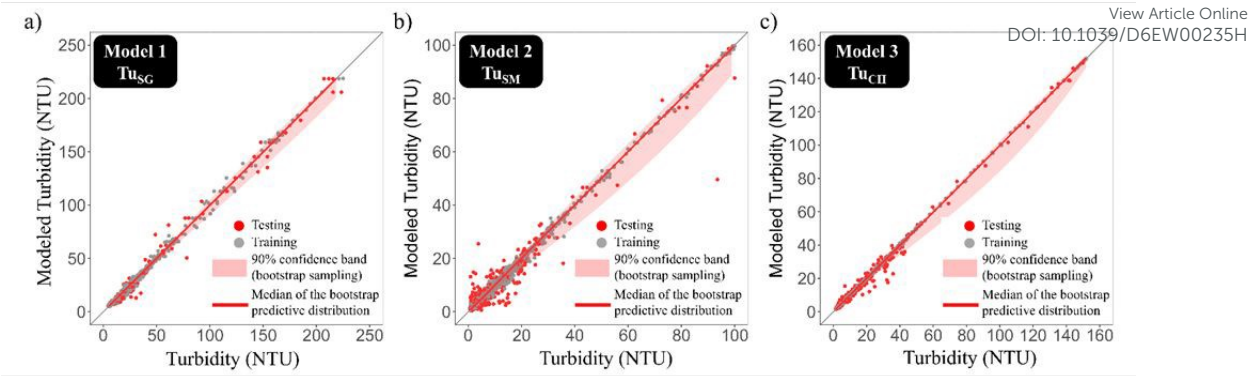

Figure 4. Model performance shown by comparing predicted and observed Tu (gray points, training dataset, and red points, testing dataset) at the three DWTPs for (a) Model 1–Tu_{SG}; (b) Model 2–Tu_{SM}, and (c) Model 3–Tu_{CH}. The red shaded band indicates the 90% prediction interval (5th–95th quantiles), and the median line (red line) shows the central tendency (50th percentile) of the bootstrap predictions.

Metrics show that all models performed reasonably well (Table 3). XGBoost models outperformed the baseline MLR models. R^2 and NSE metrics for the baseline MLR models are reasonably good. However, RMSE and MAE metrics (representing errors between observed and predicted Tu values) of MLR models indicate poorer predictive performance. In contrast, for XGBoost models, R^2 values are very high for all the three models ($R^2 = 0.98 - 0.99$) both in training and testing stages, indicating that observed data variability is well captured by every model. NSE are also very high both in training and testing stages suggesting that models have very good predictive skills. RMSE and MAE are higher in the testing stage than in the training stage, indicating a slight decrease in performance. We also observed that models using more input variables have lower errors in training stage ($RMSE_{model\,1} > RMSE_{model\,2} > RMSE_{model\,3}$). A similar trend can also be observed in the testing stage, for which errors in Model 2 are higher than errors in Model 3. However, errors in Model 1 are also lower than Model 2. The 50th percentile in the uncertainty analysis (red line in the three models on Figure 4) indicates a very small or null bias for all models, because they are almost perfectly aligned with 1:1 line (gray line). The width of the red band indicates that the models consistently

View Article Online
DOI: 10.1039/D6EW00235H

predict similar Tu values across bootstraps, showing high confidence. On the other hand, the confidence bands get slightly wider at the highest Tu. This likely reflects the greater variability and modeling difficulty associated with peak turbidity during or following rainfall events.

Table 3. Summary of model performance metrics R^2, Nash-Sutcliffe efficiency (NSE), root means square error (RMSE), mean absolute error (MAE). NTU = nephelometric turbidity units.

Model	Output	Inputs	R2		NSE		RMSE		MAE	
			Training	Testing	Training	Testing	Training (NTU)	Testing (NTU)	Training (NTU)	Testing (NTU)
Model 1 (XGBoost)	Tu_{SG}	P4, Q3	0.99	0.99	0.99	0.99	0.71	2.00	0.44	0.73
Model 2 (XGBoost)	Tu_{SM}	P4, P3, P2, Q3, Q2, Tu_{SG}	0.99	0.98	0.99	0.95	0.68	2.53	0.43	1.21
Model 3 (XGBoost)	Tu_{CH}	P4, P3, P2, P1, Q3, Q2, Q1, Tu_{SG}, Tu_{SM}	0.99	0.99	0.99	0.99	0.22	1.33	0.15	0.58
Model 1 (MLR)	Tu_{SG}	P4, Q3	0.98	0.98	0.96	0.97	3.92	3.82	1.32	1.32
Model 2 (MLR)	Tu_{SM}	P4, P3, P2, P1, Q3, Q2, Q1, Tu_{SG}, Tu_{SM}	0.91	0.94	0.83	0.88	5.43	4.06	2.73	2.73
Model 3 (MLR)	Tu_{CH}	P4, P3, P2, P1, Q3, Q2, Q1, Tu_{SG}, Tu_{SM}	0.84	0.81	0.7	0.66	8.25	8.49	3.97	3.97

Figure 5 shows time series plots of predicted and observed Tu at the three DWTPs. According to the performance metrics shown in Table 3, we found that all three models predicted data (both training and testing stages) that are generally near to observed data. They show that the three models are able to predict both low and high values throughout the year (from May to October). The best fit is observed for Model 3, which has the most input variables, as discussed above. Model 2 has several predicted data points further from the observed data points, which agrees with RMSE and MAE performance metrics. In particular, Model 2 overestimated several Tu peaks, especially between June/July and mid-September (Figure 5b), possibly because there are two dams regulating flow upstream of the Sainte-Marie water intake (see Figure 1), potentially influencing particles settling. This remains an untested hypothesis, and we currently lack the tools to evaluate

it.

View Article Online
DOI: 10.1039/D6EW00235H

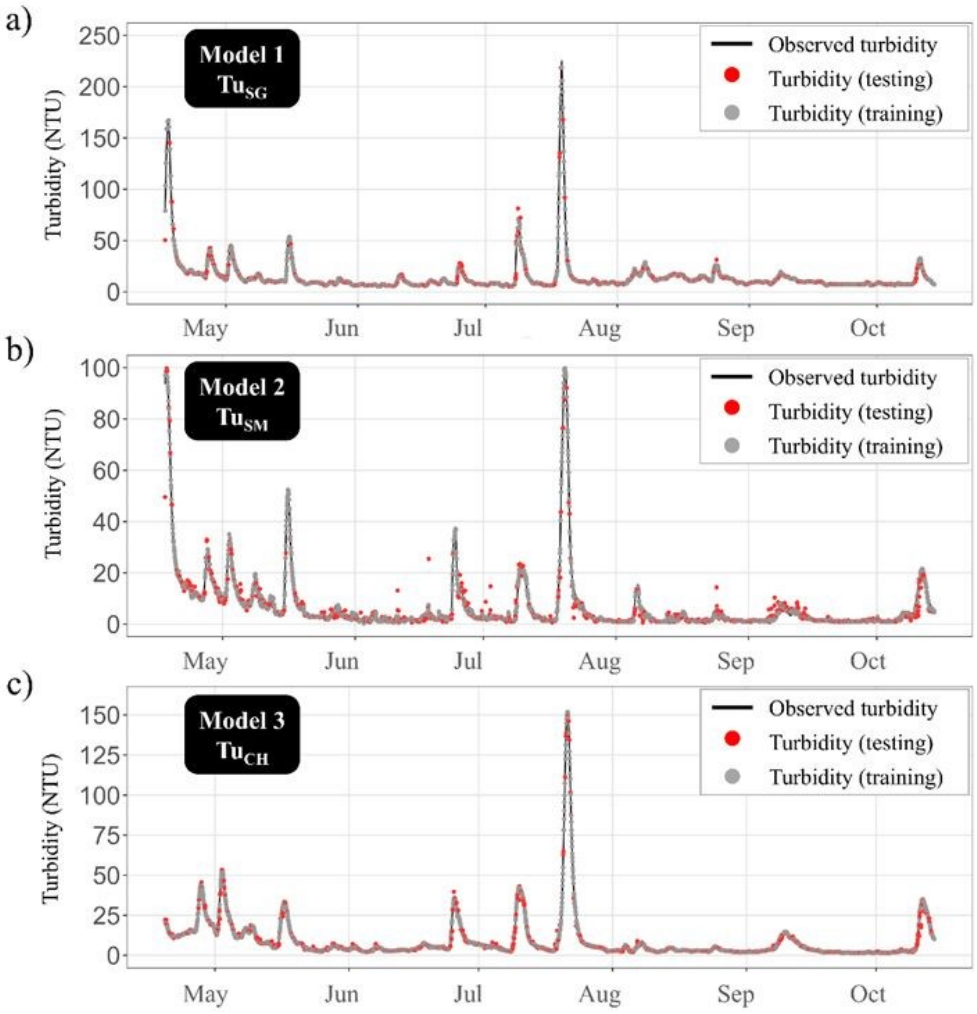

Figure 5. Time series plots of predicted and observed Tu at the three DWTPs for (a) Model 1–Tu_{SG}; (b) Model 2–Tu_{SM}), and (c) Model 3–Tu_{CH}.

3.3. Interpreting model prediction using SHAP strategy

Figures 6, 7, and 8, show an overall interpretation of the models using the SHAP strategy based on input variable importance (bar plots) and summary plots (beeswarm plots) (See 2.3.4, Interpretability strategy). For Model 1 (Tu_{SG}), Figure 6a shows that River flow – Q3 is the most important input variable, contributing an average of 6.9 NTU to the model, whereas Precipitation – P4 contributes in 2.7 NTU. In Figure 6b, we observe that higher River flow – Q3 values (red points) have higher SHAP values, leading to high Tu_{SG}

View Article Online
DOI: 10.1039/D6EW00235H

values. Therefore, higher River flow – Q3 values are more important for predicting raw water Tu_{SG}. In contrast, both high and low Precipitation – P4 values (blue and red points) have low SHAP values, leading to low Tu_{SG} values. The contribution of Precipitation – P4 to Tu_{SG} is less significant than River flow – Q3. We also observe that many SHAP values from both River flow – Q3 and Precipitation – P4 are stacked near the 0 value. These observations correspond to lower values of Q3 and P4, meaning that lower river flow and lower or zero precipitation provide insufficient information for predicting high TuSG values, since they tend to pull the raw water Tu_{SG} predictions below the average.

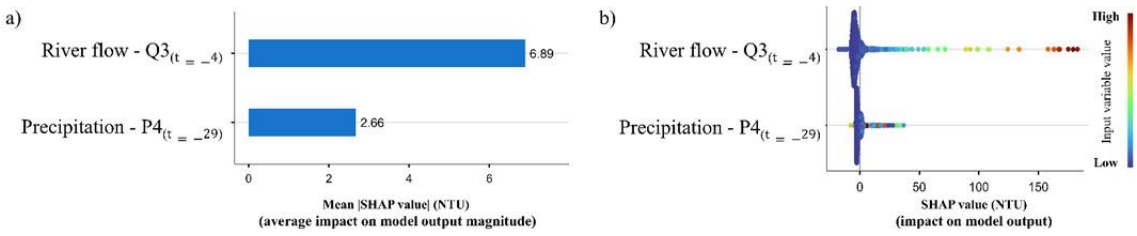

Figure 6. Interpretation using SHAP strategy of the best model 1 based on (a) input variable importance and (b) summary plot.

For Model 2–Tu_{SM} (Figure 7a), we observe that River flow – Q3 is also the most important input variable Tu_{SG} is the second most important, and Precipitation – P4 is the least important (and also the furthest from the Ste-Marie DWTP). Therefore, higher River flow – Q3 and Q2, and higher Tu_{SG} values are the most important for predicting raw water Tu_{SM}. Average impact on Tu_{SM} is 2.7 NTU for River flow – Q3 and 2.6 NTU for Tu_{SG}, whereas Precipitation – P4 has an impact of 0.939 NTU. Figure 7b shows that high Tu_{SG} values and River flow – Q3 and Q2 (red points), have higher SHAP values leading to high Tu_{SM} values. In contrast, both high and low Precipitation – P4 values (blue and red points) have low SHAP values leading to low Tu_{SM} values. We also observe that many SHAP values from River flow – Q3 and Q2, and Tu_{SG} are stacked near 0. Those observations correspond to lower values of Q3, Q2 and Tu_{SG}, meaning that lower river flow and lower or zero precipitation give insufficient information for predicting Tu_{SM}

high values, since they tend to pull the raw water Tu_{SM} predictions below the average.

View Article Online
DOI: 10.1039/D6EW00235H

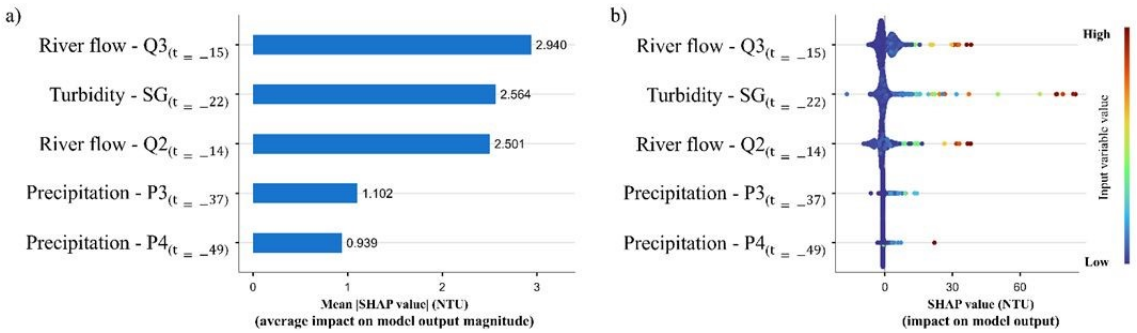

Figure 7. Interpretation using SHAP strategy of the best model 2 based on (a) input variable importance and (b) summary plot.

In addition, a second group of points for River flow – Q3 are observed to stack above the zero SHAP value (Figure 7b), thus pulling the raw water Tu_{SM} predictions above the average. This may account for the raw water Tu_{SM} peak values that are observed to be overestimated (pulled up) in Figure 5, as some River flow – Q3 observations are causing an overestimation of Tu_{SM} peak values, which decreases the Model 2 performance metrics. In contrast to Tu_{SG} and River flow – Q3 and Q2, neither Precipitation – P4 nor P3 contribute significantly to Tu_{SM}.

For Model 3–Tu_{CH} (Figure 8a), we observe that Tu_{SM} (the turbidity at the closest DWTP) is the most important input variable, followed by River flow – Q1 (the closest flow meter station), Tu_{SG}, and River flow – Q2 and Q3. Therefore, higher the River flow – Q3, Q2, Q1, and higher Tu_{SG} and Tu_{SM} values, are the most important for predicting raw water Tu_{CH}. The least important variables are the four Precipitation variables. The average impact on Tu_{CH} values are 2.9 NTU for Tu_{SM}, 2.0 NTU for River flow – Q1, and 1.6 NTU for Tu_{SG}. Figure 8b shows that high Tu_{SG} and Tu_{SM} values (red points) have higher SHAP values leading to high Tu_{CH} values. This is also true for Q1 and Q2, although their highest SHAP values (i.e., impact) are lower. None of the Precipitation variables contribute significantly to Tu_{CH}.

View Article Online
DOI: 10.1039/D6EW00235H

Similarly to Models 1 and 2, we observe that many SHAP values from River flow – Q3, Q2, Q1 and Tu_{SG} and Tu_{SM} are stacked near 0. These observations correspond to lower variable values of Q3, Q2, Q1, Tu_{SG}, and $Tu_{SM,}$ meaning that lower river flow, lower upstream raw water Tu and lower or zero precipitation provide no information for predicting Tu_{CH} high values, since they pull the raw water Tu_{CH} predictions below the average.

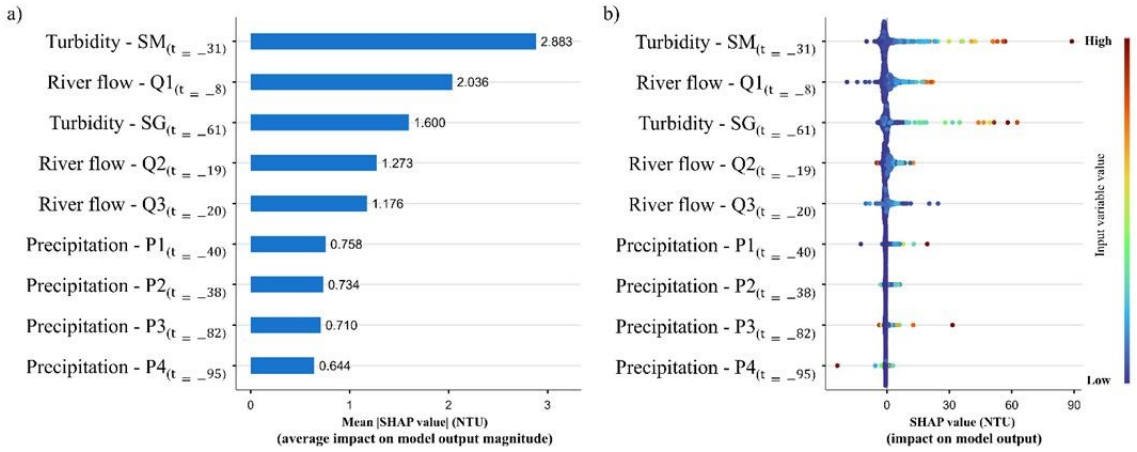

Figure 8. Interpretation using SHAP strategy of the best model 3 based on (a) input variable importance and (b) summary plot.

After interpreting the global influence of input variables on raw water Tu predictions, we selected the most extreme Tu events to analyze how input variables affected the predictions of Tu_{CH} (Model 3). Figure 9 presents the interpretation using the SHAP approach for four extreme turbidity events. In this figure, the Shapley value explanation is shown on the left, and the Tu time series on the right. As shown in Figure 9, both upstream turbidity and river flow generally had a positive impact on Tu_{CH} (except for the first event, where Q2 had a negative effect). Tu_{SM} is the input variable that has the greatest influence on Tu_{CH} predictions, except in the third event, where it is the second most important variable. The second most influential variable is either Q1 or Tu_{SG}. In contrast, precipitation shows a negative, very low (between 0 and 1.5 NTU), or negligible impact on Tu_{CH} predictions. The limited influence of precipitation on Tu predictions may be

View Article Online
DOI: 10.1039/D6EW00235H

explained by the nature of processes occurring in the watershed. River flow can act as an integrating variable of rainfall–runoff processes, as it reflects the cumulative effect of precipitation over the watershed. Furthermore, variations in flow may indicate the dynamics of particle transport into the river, whereas precipitation alone only represents the amount of rainfall. This remains an untested hypothesis, thus further research is needed to clarify the low impact of precipitation on Tu predictions.

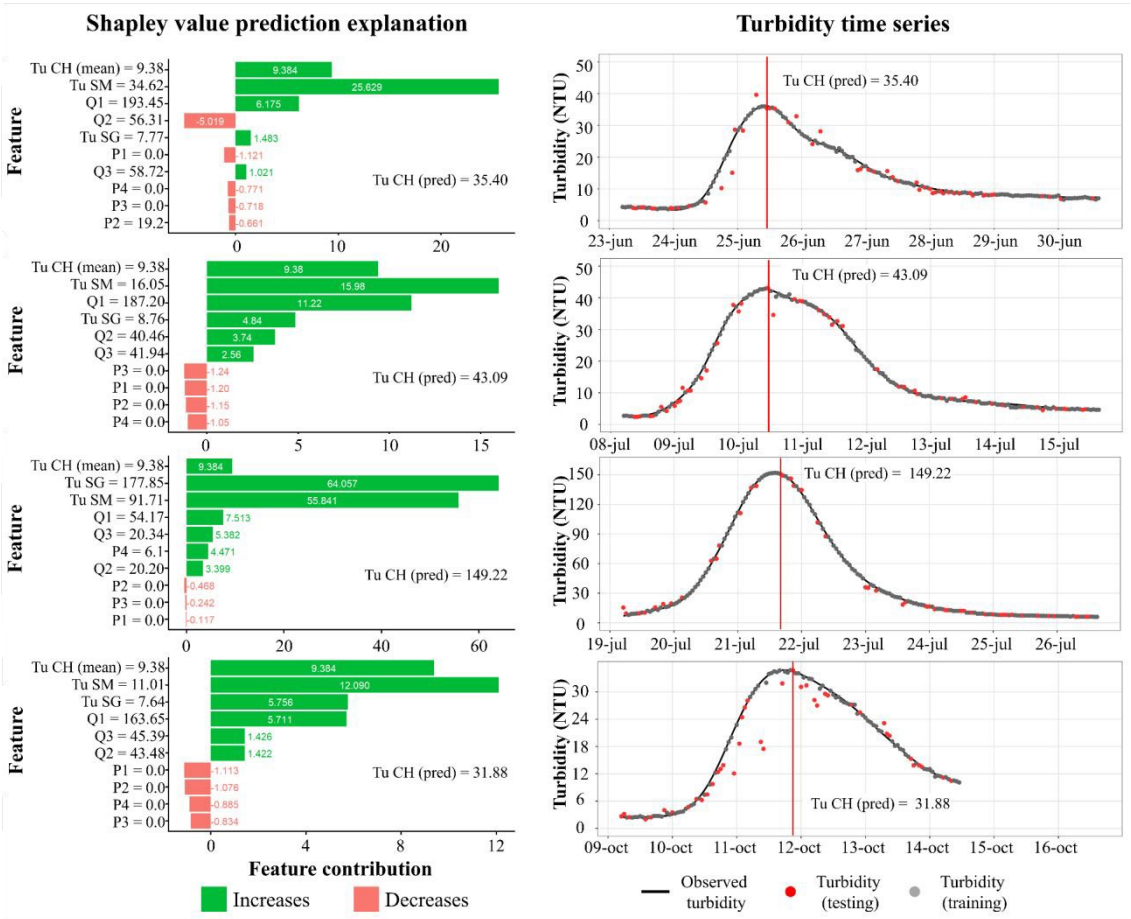

Figure 9. Interpretation using SHAP strategy for four extreme turbidity events. Shapley value prediction explanation (left). Turbidity time series (right).

3.4. Sensitivity analysis

A sensitivity analysis was conducted to evaluate how removing one or a group of variables impacted model predictions. This analysis was performed by removing each

View Article Online
DOI: 10.1039/D6EW00235H

variable individually, as well as groups of similar variables, for Model 3, which predicts Tu_{CH}. A summary of the performance metrics is presented in Table 4. When removing each flow variable (Q1–Q3) one at a time, no major differences were observed compared to the original model. However, removing Q1 (the flowmeter nearest the DWTP—Charny, where Tu_{CH} is predicted) shows the largest variation in performance metrics. In contrast, removing all three flow variables (Q1–Q3) results in a considerable decrease in model performance. Removing precipitation variables one at a time, and even removing all four variables (P1–P4) simultaneously, does not lead to a noticeable decrease in performance. This is consistent with the results presented in Figures 6, 7, and 8, where precipitation shows the lowest average impact on model output magnitude. Therefore, precipitation may be omitted as an input variable in raw water Tu modeling without impacting much the model performance. In contrast, upstream raw water Tu is a more important predictor for downstream raw water Tu. When the raw water Tu from a single upstream station is omitted, predictive performance remains high. However, when both upstream turbidity series are excluded, performance decreases and predictions deteriorate. This confirms the importance of including upstream raw water Tu as an input variable in turbidity modeling and prediction.

Table 4. Summary of performance metrics of the sensitivity analysis

Removed variable(s)	R2		NSE		RMSE		MAE	
	Training	Testing	Training	Testing	Training (NTU)	Testing (NTU)	Training (NTU)	Testing (NTU)
Q1	0.99	0.99	0.98	0.99	1.89	0.87	0.91	0.52
Q2	0.99	0.99	0.99	0.99	0.72	0.66	0.76	0.43
Q3	0.99	0.99	0.98	0.99	1.75	0.67	0.78	0.43
P1	0.99	0.99	0.99	0.99	1.69	0.66	0.76	0.43
P2	0.99	0.99	0.99	0.99	1.70	0.66	0.75	0.43
P3	0.99	0.99	0.99	0.99	1.74	0.68	0.77	0.45
P4	0.99	0.99	0.99	0.99	1.73	0.66	0.76	0.42
Tu_{SM}	0.99	0.99	0.97	0.99	2.35	1.00	1.06	1.11
Tu_{SG}	0.99	0.99	0.98	0.98	2.05	0.83	0.96	0.58

View Article Online
DOI: 10.1039/D6EW00235H

Removed variable(s)	R2		NSE		RMSE		MAE	
	Training	Testing	Training	Testing	Training (NTU)	Testing (NTU)	Training (NTU)	Testing (NTU)
Q1, Q2, Q3	0.87	0.89	0.75	0.98	7.32	2.01	2.32	1.11
P1, P2, P3, P4	0.99	0.99	0.99	0.99	1.72	0.69	0.78	0.45
Tu_{SM}, Tu_{SG}	0.68	0.75	0.43	0.50	10.98	7.50	3.19	2.60

3.5. Discussion

We can infer several findings from the lag-times shown in Table 2. Overall, lag-times for river flow are lower than for precipitation for all models (1–Tu_{SG}, 2–Tu_{SM} and 3–Tu_{CH}), as mentioned by Ortiz-Lopez et al. (2023) (6). While the effects of precipitation events are mediated through the rainfall-runoff process before they can be measured in water intakes, river flow has a more direct and rapid influence on raw water Tu. As expected, Model 3 (Tu_{CH}), which uses all the available input variables (Q1, P1 and P2 representing low and middle watershed, along with Q2, Q3, P3 and P4 representing the high watershed) had the highest performance metrics. The lag-times observed between upstream raw water Tu and Tu at the target intake likely reflect the combined influence of river flow dynamics (rapid and direct) and rainfall-driven processes (more prolonged and indirect). Our study found that the further the locations of the input variable measurement stations (rainfall, river flow rates, and Tu upstream) are from the location of the measurement of the target variable (raw water Tu), the greater the lag-time. These time-lagged variables were then used in a machine learning technique to model raw water Tu in a water intake.

Based on the modeling results shown in Figure 4, the XGBoost technique demonstrated a very high ability to predict the raw water Tu peaks in our case study. The very high R2 performance metric shows that all three models (Model 1–Tu_{SG}, Model 2–Tu_{SM} and Model 3–Tu_{CH}) can explain 99% of the variability of observed Tu. High values of the NSE metric also indicate a very good prediction performance, showing a very low

View Article Online
DOI: 10.1039/D6EW00235H

deviation between measured and modeled Tu values. Relatively low RMSE and MAE values also show low deviation between modeled and observed Tu values and are similar to or smaller than minimal values of raw water Tu at the three respective water intakes (Table 1). However, Model 2 performed slightly less well than Models 1 and 3. The main implication of the results found in this study is that including upstream raw water Tu in addition to River flow as input variables enhances prediction of raw water Tu peaks. Considering the uncertainty levels and confidence bands (Figure 4), Models 1 and 3 could be used in monitoring and decision-making systems, for example, for operational adjustments, anticipating raw water changes, and staff preparation. Since uncertainty moderately increases for higher and peak values, Models 1 and 3 should be used with caution in critical and automated operational decisions. Therefore, although the models are useful as a supporting tool, their use as the sole criterion for critical operational decisions is not recommended. Better anticipation of raw water Tu peaks by DWTP managers should contribute to an appropriate and timely preparation of DWTP operations.

Up until now, the importance of input variables on raw water Tu modeling has remained obscured by the black-box nature of machine learning models. Moreover, the impact of the range (high or low) of input variable values on raw water Tu predictions was unknown. However, the SHAP strategy reveals the importance of each input variable in the implemented XGBoost technique. As expected, in Model 1–Tu_{SG} (only two input variables), river flow is more correlated with raw water Tu than precipitation. Moreover, high River flow – Q3 values contributed the most (over 150 NTU, dark red points in Figure 6b) to predicting Tu_{SG}. In Model 2–Tu_{SM} (five input variables), the most important input variable was River flow – Q3, followed by Tu_{SG} and River flow – Q2 (both these contributing a small difference). In this case, although high Tu_{SG} values contributed the

View Article Online
DOI: 10.1039/D6EW00235H

most (over 60 NTU, dark red points in Figure 7b) to Tu_{SM} prediction, the stacked dark blue points to the right of 0 NTU for River flow – Q3 also contributed positively to TuSM prediction by pulling the raw water Tu_{SM} predictions above the average. In Model 3–Tu_{CH} (nine input variables) the most important input variable was Tu_{SM} and then River flow – Q1. In this case, high values for both Tu_{SG} and Tu_{SM} (over 60 NTU, dark red points, in Figure 8b) improved Tu_{CH} prediction. Those findings indicate the most useful values for predicting raw water Tu peaks, especially during and after rainfalls.

Studies conducted by Ortiz-Lopez et al., (2023, 2024) (6, 17) to predict raw water Tu at the Charny DWTP did not consider upstream information on raw water Tu and had lower performance metrics than this current study: $R^2 = 0.87$ and NSE = 0.75 for a Tu model in the testing stage using the SVR technique 6 and $R^2 = 0.81$ and NSE = 0.65 for a Tu model in the testing stage using XGBoost technique (17). These findings demonstrate the relevance of including upstream information on water quality at the watershed scale, in addition to river flow and rainfall, as input variables for targeted Tu modeling, as was done in our Model 3. We observed the high importance given to upstream raw water Tu by the machine learning technique, as highlighted by the SHAP strategy. An interpretability strategy, such as SHAP is needed to reveal the impact on the raw water Tu predictions, not only of every input variable, but also of their higher or lower value ranges. Model 3 clearly outperformed other models in this case study, based on its performance metrics. We also suggest that rainfall could be omitted as an input variable in Tu modeling, given its low importance estimated by the SHAP strategy in the three models.

Another important aspect to consider is the feasibility of having all input variables available in real-world applications for EWS. Decision-support approaches based on

View Article Online
DOI: 10.1039/D6EW00235H

predictive models that incorporate upstream Tu information must be continuously supplied by DWTP. In the case of upstream Tu, as demonstrated in the sensitivity analysis, the model is able to provide predictions when one of the two Tu inputs is missing; however, its performance decreases when both turbidity inputs are unavailable. Continuous monitoring of raw water Tu is always present in a DWTP. These data are send continuously (online) to the SCADA (Supervisory Control and Data Acquisition) system. The system should therefore include a real-time data transmission framework to enable immediate data acquisition. In addition, continuous maintenance of sensors is required to ensure higher reliability. The same considerations apply to the flow variable measured upstream of the DWTP where Tu is to be predicted. Data transmission must be almost instantaneous, and mechanisms to correct flow measurements should be implemented. As shown in the sensitivity analysis, the omission of one or several flow series does not result in a significant loss of predictive performance in the models.

Some limitations of this study may be related to the prediction time horizon. This approach allows us to predict the raw water Tu at a specific moment, using time-lagged input variables. Thus, the greatest time step at which the raw water Tu can be predicted corresponds to the smallest lag-time among all input variables (4 hours for Model 3, 14 hours for Model 2, and 8 hours for Model 1). These prediction time horizons might be too tight for application in Early Warning Systems. Other modeling approaches might produce longer prediction time horizons, for instance using time series forecasting or more powerful time-series based machine learning techniques such as Long-Short Term Memory (LSTM) networks. In addition, moisture and dry precedent conditions were not considered as predictors because the related data were not available. Since previous studies have used moisture and dry precedent conditions, we recommend including these variables in future research. Since the Chaudière River has some dams upstream of two

View Article Online
DOI: 10.1039/D6EW00235H

of the DWTP of this study, we also recommend including an analysis of dam operations in future studies.

Future research could improve the robustness of raw water Tu predictions by extending prediction horizons and using more complex machine learning models. More data should be collected at water intakes to improve the learning space for training stages of models. This could also open the possibility to test a different data-splitting approach for training machine learning models. For example, Zhu et al. (2023) (34) propose using a block splitting approach that combines chronological and random split. Other crucial parameters in drinking water treatment and production should be measured, collected and modeled, such as natural organic matter (NOM). Surrogate parameters for NOM, such as, UV absorbance, have been already modeled and predicted in a water intake using watershed information. Prediction of NOM at water intakes might be improved by including upstream NOM measurements as input variables in machine learning models.

4. Conclusion

This research demonstrated that the predictability of the targeted parameter (raw water Tu) can be enhanced by adding information about that same parameter from locations upstream of the water intake at the watershed scale (for instance, in other municipal water intakes) to use as input variables for machine learning models. Other upstream water quality variables, such as river flow and watershed rainfall data, also enhance prediction of raw water Tu, as demonstrated by a time series of modeled raw water Tu and the performance metrics of the XGBoost models. Interpreting the predictions by means of a SHAP strategy allowed us to evaluate not only the significance of every input variable, but also of the effect of high or low value ranges of these variables on the raw water Tu predictions. Overall, upstream raw water Tu and river flow were the most important input

View Article Online
DOI: 10.1039/D6EW00235H

variables to model the targeted variable.

Our findings could help decision-making for operators in DWTP. Accurate and timely predictions of raw water Tu peaks during and after rainfall events are critical information for anticipating sudden changes in raw water quality and for timely adaptation of DWTP operations, such as coagulation management. Results show predictions horizon between at least 4 hours (in the predictive model with fewer input variables) and 8 hours (in the predictive model with more input variables). This is particularly relevant for watersheds with several water intakes supplied by the same river and that are affected by frequent intensive rainfall events.

This research highlights the importance of estimating lag-times when modeling a raw water parameter, such as Tu, and using time series of either hydrological, meteorological and other upstream raw water parameters as input variables. Identification of such lag-times between input and output variables is a critical methodological component when selecting input variables. As previously demonstrated, time-lagged input variables are better at representing the delays between changes in rainfall, river flow, upstream water quality and the targeted water quality parameter.

Results obtained in this study open possibilities for developing regional early warning systems where raw water quality information from upstream water intakes could be used to predict conditions at downstream water intakes. However, this study is limited to the summer period (when turbidity varies the most) and observations from only one year and could require more observations to be applied in an EWS. Forthcoming work should include an interannual analysis of raw water quality and hydrometeorological information. Future research on raw water quality modeling should be focused on increasing prediction horizons to provide timely and accurate input to an EWS. New

developments using regional models could focus on the development of decision support tools to help DWTP operators adjust water treatment processes according to the predicted raw water quality parameters.

Author contributions

Christian Ortiz-Lopez: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization. **Christian Bouchard**: Visualization, Supervision. **Manuel Rodriguez**: Funding acquisition, Visualization, Supervision. **Christian Ortiz-Lopez**: Writing – original draft. **Manuel Rodriguez**: Writing – review & editing. **Christian Bouchard**: Writing – review & editing.

Declaration of competing interest

There are no conflicts to declare.

Data availability

The data that support the findings of this study are available from the corresponding author, Dr. Christian Ortiz-Lopez, upon reasonable request.

Acknowledgements

This study was funded by the NSERC (Natural Sciences and Engineering Research Council of Canada) Drinking Water Research Chair at Laval University, whose main partners are the municipalities of Quebec and Lévis, Avensys Solutions, Agiro and WaterShed Monitoring. Authors also especially thank the drinking water services offices of the municipalities of Sainte-Marie and Saint-Georges for providing us with water intake data. The authors thank Ms. Mary Thaler for English editing.

References

1. Khan SJ, Deere D, Leusch FDL, Humpage A, Jenkins M, Cunliffe D. Extreme weather events: Should drinking water quality management systems adapt to changing risk profiles? Water Research. 2015;85:124–36.

2. Raseman WJ, Kasprzyk JR, Rosario-Ortiz FL, Stewart JR, Livneh B. Emerging investigators series: a critical review of decision support systems for water treatment: making the case for incorporating climate change and climate extremes. Environmental Science: Water Research & Technology. 2017;3:18–36.

3. Delpla I, Bouchard C, Dorea C, Rodriguez MJ. Assessment of rain event effects on source water quality degradation and subsequent water treatment operations. Science of The Total Environment. 2023;866:161085–.

4. Fukushima T, Kitamura T, Matsushita B. Lake water quality observed after extreme rainfall events: implications for water quality affected by stormy runoff. SN Applied Sciences. 2021;3(11):1–15.

5. Jia Z, Chang X, Duan T, Wang X, Wei T, Li Y. Water quality responses to rainfall and surrounding land uses in urban lakes. Journal of Environmental Management. 2021;298:113514–.

6. Ortiz-Lopez C, Torres A, Bouchard C, Rodriguez M. A methodology for integrating time-lagged rainfall and river flow data into machine learning models to improve prediction of quality parameters of raw water supplying a treatment plant. Journal of Hydroinformatics. 2023;25(6):2406–26.

7. Ortiz-Lopez C, Bouchard C, Rodriguez M. Machine learning models with potential application to predict source water quality for treatment purposes: a critical review. Environmental Technology Reviews. 2022;11(1):118–47.

8. Khan SJ, Deere D, Leusch FDL, Humpage A, Jenkins M, Cunliffe D, et al. Lessons and guidance for the management of safe drinking water during extreme weather events. Environmental Science: Water Research & Technology. 2017;3(2):262–77.

9. Delpla I, Florea M, Rodriguez MJ. Drinking Water Source Monitoring Using Early Warning Systems Based on Data Mining Techniques. Water Resources Management. 2019;33(1):129–40.

10. Quansah JE, Engel B, Rochon GL. Early Warning Systems: A Review. Journal of Terrestrial Observation. 2010;2(2).

11. Rajaee T, Khani S, Ravansalar M. Artificial intelligence-based single and hybrid models for prediction of water quality in rivers: A review. Chemometrics and Intelligent Laboratory Systems. 2020;200:103978–.

12. Tiyasha, Tung TM, Yaseen ZM. A survey on river water quality modelling using artificial intelligence models: 2000–2020. Journal of Hydrology. 2020;585:124670–732.

13. Alizadeh MJ, Kavianpour MR, Danesh M, Adolf J, Shamshirband S, Chau KW. Effect of river flow on the quality of estuarine and coastal waters using machine learning models. Engineering Applications of Computational Fluid Mechanics. 2018;12(1):810–23.

14. Ahmed M, Mumtaz R, Zaidi SMH. Analysis of water quality indices and machine learning techniques for rating water pollution: a case study of Rawal Dam, Pakistan. Water Supply. 2021;21(6):3225–50.

15. Zhang Y, Yao X, Wu Q, Huang Y, Zhou Z, Yang J, et al. Turbidity prediction of lake-type raw water using random forest model based on meteorological data: A case study of Tai lake, China. Journal of Environmental Management. 2021;290:112657–.

16. Adedeji IC, Ahmadisharaf E, Sun Y. Predicting in-stream water quality constituents at the watershed scale using machine learning. Journal of Contaminant Hydrology. 2022;251:104078–.

17. Ortiz-Lopez C, Bouchard C, Rodriguez M. Ensemble machine learning using hydrometeorological information to improve modeling of quality parameter of raw water supplying treatment plants. Journal of Environmental Management. 2024;362:121378–.

18. Chen J, Chang H. Predicting Post-Wildfire Stream Temperature and Turbidity: A Machine Learning Approach in Western U.S. Watersheds. Water. 2025;17(3).

19. Kemper JT, Underwood KL, Hamshaw SD, Davis D, Siemion J, Shanley JB, et al. Leveraging High-Frequency Sensor Data and U.S. National Water Model Output to Forecast Turbidity in a Drinking Water Supply Basin. Journal of the American Water Resources Association. 2025;61(2):e70011–e.

20. Zhang K, Xia R, Wang Y, Chen Y, Wang X, Dou J. Stack Coupling Machine Learning Model Could Enhance the Accuracy in Short-Term Water Quality Prediction. Water. 2025;17(19).

21. Yang S, Liu Y, Chen Q, Ren Z, Jing Z, Wang Y, et al. Prediction of Eutrophic Water Quality in the Daluxi River Based on a Multi-Scale Feature Extraction and Hybrid Screening Strategy. 2025.

22. Lundberg SM, Lee S-I, editors. A Unified Approach to Interpreting Model Predictions. NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems; 2017; Long Beach, CA: Curran Associates Inc.

23. Gharehbaghi A, Heddam S, Mehdizadeh S, Kim S. Development of interpretable intelligent frameworks for estimating river water turbidity. Engineering Applications of Computational Fluid Mechanics. 2025;19(1).

24. Xiao F, Zhang R, Jian Z, Liu W, Sun T, Pang W, et al. Using ensemble machine learning to predict and understand spatiotemporal water quality variations across diverse watersheds in coastal urbanized areas. Ecological Indicators. 2025;178:113976–.

25. Kruk M. SHAP-NET, a network based on Shapley values as a new tool to improve

View Article Online
DOI: 10.1039/D6EW00235H

the explainability of the XGBoost-SHAP model for the problem of water quality. Environmental Modelling and Software. 2025;188:106403–.

26. Soleymani Hasani S, Arias ME, Nguyen HQ, Tarabih OM, Welch Z, Zhang Q. Leveraging explainable machine learning for enhanced management of lake water quality. Journal of Environmental Management. 2024;370:122890–.

27. Chen T, Guestrin C, editors. XGBoost: A Scalable Tree Boosting System. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16) Association for Computing Machinery; 2016: Association for Computing Machinery.

28. COBARIC. Plan directeur de l'eau de la ZGIE Chaudière. Comité de bassin de la rivière Chaudière; 2024 2024/3//.

29. Ministère de l'environnement et de la lutte contre les changements climatiques M. Expertise hydrique, Niveaux et débits et niveaux en temps réel dans les cours d'eau. Québec: Direction de l'expertise hydrique, Banque de données hydriques (BDH) 2020 [Available from: https://www.cehq.gouv.qc.ca/hydrometrie/index.htm

30. Ministère de l'environnement et de la lutte contre les changements climatiques M. Données du Réseau de surveillance du climat du Québec. Québec: Direction de la qualité de l'air et du climat. 2020 [Available from: https://www.environnement.gouv.qc.ca/climat/surveillance/reseau-parametres.asp

31. Hyndman RJ, Khandakar Y. Automatic Time Series Forecasting: The forecast Package for R. Journal of Statistical Software. 2008;27(3):1–22.

32. R-Core-Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2025.

33. Chen T, He T, Benesty M, Khotilovich V, Tang Y, Cho H, et al. xgboost: Extreme Gradient Boosting. 2025.

34. Zhu J-J, Yang M, Ren ZJ. Machine Learning in Environmental Research: Common Pitfalls and Best Practices. Environmental Science and Technology. 2023;57(46):17671–89.

35. Davison AC, Hinkley DV. Bootstrap Methods and their Application: Cambridge University Press; 1997 1997/10//.

36. Efron B, Tibshirani RJ. An Introduction to the Bootstrap: Chapman and Hall/CRC; 1994 1994/5//.

37. Nash JE, Sutcliffe JV. River flow forecasting through conceptual models part I: A discussion of principles. Journal of Hydrology. 1970;10:282–90.

38. Shapley LS. Notes on the n-Person Game - II: The Value of an n-Person Game. ASTIA Document. 1951.

39. Molnar C. Interpretable machine learning: a guide for making black box models explainable: Christoph Molnar; 2025. Available from:

View Article Online
DOI: 10.1039/D6EW00235H

https://christophm.github.io/interpretable-ml-book

40. Schäfer B, Beck C, Rhys H, Soteriou H, Jennings P, Beechey A, et al. Machine learning approach towards explaining water quality dynamics in an urbanised river. Scientific Reports. 2022;12(1):12346.

41. Li L, Qiao J, Yu G, Wang L, Li HY, Liao C, et al. Interpretable tree-based ensemble model for predicting beach water quality. Water Research. 2022;211:118078–.

View Article Online
DOI: 10.1039/D6EW00235H

View Article Online
DOI: 10.1039/D6EW00235H

Data availability statement

The data that support the findings of this study are available from the corresponding author, Dr. Christian Ortiz-Lopez, upon reasonable request.