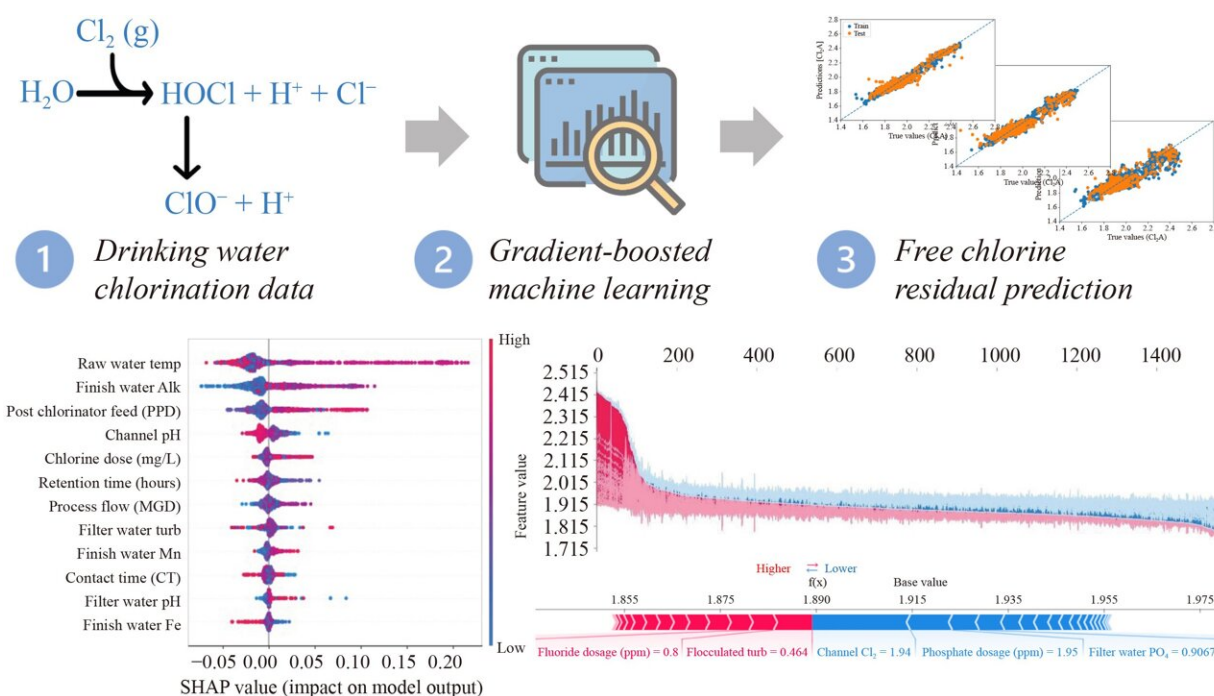


Inexpensive water-treatment monitoring process powered by machine learning

November 6 2023



Credit: *Frontiers of Environmental Science & Engineering* (2023). DOI: 10.1007/s11783-024-1777-6.

<https://journal.hep.com.cn/fese/EN/10.1007/s11783-024-1777-6>

Small, rural drinking water treatment (DWT) plants typically use only chlorine to implement the disinfection process. For these plants, free chlorine residual (FCR) serves as a key performance measure for disinfection. The FCR is stated as the concentration of free chlorine

remaining in the water, after the chlorine has oxidized the target contaminants.

In practice, the FCR is determined by plant operators based on their experience. Specifically, operators choose a dose of [chlorine](#) to achieve a satisfactory FCR concentration, but often have to make an estimate of the chlorine requirements.

The challenge of determining an accurate FCR has led to the use of advanced FCR prediction techniques. In particular, [machine learning](#) (ML) algorithms have proven effective in achieving this goal. By identifying correlations among numerous variables in [complex systems](#), successful ML implementation could accurately predict FCR, even from cost-effective, low-tech monitoring data.

In a new study [published](#) in *Frontiers of Environmental Science & Engineering*, the authors implemented a gradient boosting (GB) ML [model](#) with categorical boosting (CatBoost) to predict FCR. GB algorithms, including CatBoost, accumulate decision trees to generate the prediction function.

The input data was collected from a DWT plant in Georgia in the U.S., and included a wide variety of DWT monitoring records and operational process parameters. Four iterations of a generalized modeling approach were developed, including (1) base case, (2) rolling average, (3) parameter consolidation, and (4) intuitive parameters.

The research team also applied the SHapely Additive explanation (SHAP) method to this study. SHAP is an [open-source software](#) for interpreting ML models with many input parameters, which allows users to visually understand how each parameter affects the prediction function. We can study the influence of each parameter on the predicted output, by calculating its corresponding SHAP value. For example, the

SHAP analysis ranks the channel Cl_2 as the most influential parameter.

Of all four iterations, the fourth and final iteration considered only intuitive, physical relationships and water quality measured downstream from filtration. The authors summarized the comparative performance of the four ML modeling iterations. According to them, the key findings are: 1) with a sufficient number of related input parameters, ML models can produce accurate prediction results; 2) ML models can be driven by correlations that may or may not have a physical basis; 3) ML models can be analogous to operator experience.

Looking forward, the research team suggests that future studies should explore expanding the applicability domain. For example, the data set analyzed was limited to only one full year. Therefore, greater data availability is expected to broaden the applicability domain and improve the predictivity.

More information: Wiley Helm et al, Development of gradient boosting-assisted machine learning data-driven model for free chlorine residual prediction, *Frontiers of Environmental Science & Engineering* (2023). DOI: [10.1007/s11783-024-1777-6](https://doi.org/10.1007/s11783-024-1777-6).
journal.hep.com.cn/fese/EN/10.1007/s11783-024-1777-6

Provided by Frontiers Journals

Citation: Inexpensive water-treatment monitoring process powered by machine learning (2023, November 6) retrieved 8 May 2024 from <https://techxplore.com/news/2023-11-inexpensive-water-treatment-powered-machine.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.