Desalination and Water Treatment
www.deswater.com

doi:10.5004/dwt.2018.22353

111 (2018) 125–133
April

Modeling trihalomethanes concentrations in water treatment plants using machine learning techniques

Jongkwan Park[a], Chan ho Lee[a], Kyung Hwa Cho[a], Seongho Hong[b], Young Mo Kim[c,*], Yongeun Park[d,*]

[a]*School of Urban and Environmental Engineering, Ulsan National Institute of Science and Technology, Ulsan 44919, Korea*
[b]*Department of Chemical Engineering, Soongsil University, 369 Sangdo-Ro, Dongjak-Gu, Seoul 156-743, Korea*
[c]*School of Environmental Science and Engineering, Gwangju Institute of Science and Technology (GIST), 123 Cheomdan-gwagiro, Buk-gu, Gwangju 61005, Korea, email: youngmo@gist.ac.kr (Y.M. Kim)*
[d]*School of Civil and Environmental Engineering, Konkuk University, Seoul 05029, Korea, email: yepark@konkuk.ac.kr (Y. Park)*

Received 25 November 2017; Accepted 17 January 2018

ABSTRACT

Water disinfection process in a water treatment process results in the formation of disinfection by-products (DBPs), including total trihalomethanes (TTHMs). It takes a relatively long time to estimate TTHMs concentration level in the water treatment plants; thereby it is impossible to timely control operation parameters to reduce the TTHMs concentration. Here, we developed a predictive model to quantify TTHMs concentration using conventional water quality parameters from six water treatment plants in Han River. Before the developing the model, self-organizing map (SOM) and artificial neural network (ANN) restored missing values in input and output parameters. Then, an ANN model was trained to predict TTHMs by using relevant water quality parameters investigated from Pearson correlation. Pearson Correlation test selected six significant input parameters such as temperature, algae, pre-middle chlorine, post chlorine, total chlorine, and total organic carbon. Based on five-fold jackknife cross-validation, the ANN models built using different types of input data showed different performance in training (range of R^2 from 0.62 to 0.92) and validation (range of R^2 from 0.62 and 0.80) steps. This study can be a useful tool for predicting TTHMs concentrations using conventional water quality data in drinking water treatment plants. Machine learning models can be readily developed and utilized by managers working with drinking waters.

Keywords: Trihalomethanes (THMs); Drinking water treatment plant; Han River; Machine learning technique

*Corresponding author.