Title
Machine learning techniques for real-time UV-Vis spectral analysis to monitor dissolved nutrients in surface water

Authors
Chuhong Fei, Xiang Cao, Cindy Hu, Claudia Wu, Eric Morris, Ting Liu, George Lampropoulos

Abstract for technical review purpose
Ultraviolet-visible (UV-Vis) spectroscopy is a well-established technique for real-time analyzing contaminants in finished drinking water and wastewater. However, it has struggled in surface water because surface water such as river water has more complex chemical compositions than drinking water and lower concentrations of nutrient contaminants such as nitrate. Previous spectrophotometric analysis using absorbance peak at UV region to estimate nitrate in drinking water performs poorly in surface water because of interference from suspended particles and dissolved organic carbon which absorb light along similar wavelengths. To overcome these challenges, the paper develops a machine learning approach to utilize the entire spectral wavelengths for accurate estimation of low concentration of dissolved nutrients from surface water background. The spectral training data used in this research are obtained by analyzing water samples collected from the US-Canada bi-nationally regulated Detroit River during agricultural seasons using A.U.G. Signals' dual channel spectrometer system. Confirmatory concentrations of dissolved nitrate in these samples are validated by laboratory analysis. Several commonly used supervised learning techniques including linear regression, Support Vector Machine (SVM), and deep learning using Convolutional Neural Network (CNN) and Long Short-term Memory (LSTM) are studied and compared in this work. The results conclude that the CNN deep learning model performs the best. By applying learning parameter optimization and cross validation, the CNN model is able to achieve a root-mean-squared-error (RMSE) of 0.64ppb, which demonstrates effectiveness of the machine learning approach and feasibility of real-time UV-Vis spectra analysis to monitor dissolved nutrient levels in the surface watersheds.

Summary for online and/or printed programs
Ultraviolet-visible spectroscopy is a well-established technique for finished drinking water and wastewater contaminant analysis. However, it has struggled in surface water because surface water has more complex chemical interference and lower concentrations of nutrient contaminants. To overcome these challenges, the paper develops a machine learning approach to utilize the entire spectral wavelengths for nitrate concentration estimation. Several commonly used learning techniques including linear regression, support vector machine, and deep learning are studied and compared. Our results conclude that the Convolutional Neural Network deep learning model performs the best, and it can achieve ppb level nitrate estimation accuracy.

Keywords
Machine learning, Deep learning, Convolutional Neural Network (CNN), Long Short-term Memory (LSTM), Support Vector Machine (SVM), UV-Vis spectroscopy, Surface water nutrient monitoring