Simulation of Methyl Tertiary Butyl Ether Concentrations in River-Reservoir Systems Using Support Vector Regression

Mahyar Aboutalebi, M.ASCE[1]; Omid Bozorg-Haddad[2]; and Hugo A. Loáiciga, F.ASCE[3]

Abstract: Mathematical and numerical models are used to simulate the transport of pollutants released into a water body. Such simulations can be computationally burdensome, however. One approach to overcome computational burdens associated with the simulation of pollutant transport is to use data-mining tools. The aim of this study is to simulate the concentration of methyl tertiary butyl ether (MTBE) at various locations within a river-reservoir system using the support vector regression (SVR) data-mining tool. The SVR tool is optimized by means of a genetic algorithm (GA). This paper's results indicate that the developed and optimized SVR tool is more accurate than artificial neural networks (ANN) and genetic programming (GP) when judged by the correlation coefficient of regression analysis (R^2). **DOI: 10.1061/(ASCE)IR.1943-4774.0001007.**

Author keywords: Simulation; Data mining; Support vector regression; Genetic algorithm.

Introduction

Many recent publications on water resources have dealt with topics such as reservoir operation (Ahmadi et al. 2014; Bolouri-Yazdeli et al. 2014; Ashofteh et al. 2013a, 2015a), groundwater resources (Bozorg-Haddad et al. 2013; Fallah-Mehdipour et al. 2013b), conjunctive use operation (Fallah-Mehdipour et al. 2013a), design-operation of pumped-storage and hydropower systems (Bozorg-Haddad et al. 2014), flood management (Bozorg-Haddad et al. 2015b), water project management (Orouji et al. 2014; Shokri et al. 2014), hydrology (Ashofteh et al. 2013b), qualitative management of water resources systems, (Orouji et al. 2013; Bozorg-Haddad et al. 2015a), water distribution systems (Seifollahi-Aghmiuni et al. 2013; Soltanjalili et al. 2013; Beygi et al. 2014), agricultural crops (Ashofteh et al. 2015c), sedimentation (Shokri et al. 2013), and algorithmic developments (Ashofteh et al. 2015b). Yet, very few water-resources publications have dealt with the simulations of contaminants [methyl tertiary butyl ether (MTBE) is a case in point] in river-reservoir systems using support vector machine (SVM).

The sudden release of toxic chemicals into water bodies poses unique challenges for emergency response due to their unexpected occurrence and rapid transport in water bodies (Hou et al. 2014). There are water-quality models that can be used to simulate the transport of pollutants released to water bodies, which is one way to respond to sudden pollutant releases using real-time approaches. Yet, those models may be computationally burdensome. One approach to reduce the computational burden associated with the simulation of pollutant transport in water resources systems is by resorting to data-mining tools. Data-mining tools have been used in several studies involving the simulation of hydrologic phenomena. One powerful data-mining tool, however, the support vector machine (SVM), has received increasing use in the simulation of quantity and quality phenomena related to water resources.

Concerning the application of data-mining tools to quantify phenomena, Savic et al. (1999) used genetic programming (GP) to model runoff in the Kirkton basin in Scotland. The latter authors compared the performance of GP with that of artificial neural networks (ANN). Their results showed better accuracy for GP than ANN. Asefa et al. (2006) used SVM to predict the hourly and seasonal inflow in the Sevier River basin, in Utah in the United States. Sivapragasam et al. (2007) evaluated the accuracy of inflow prediction in the operation of the Kovilar and Priyar reservoirs in India to supply agricultural water. They used GP to predict inflows. Behzad et al. (2009) evaluated the performance of ANN and SVM in prediction of the Bakhtiari River runoff in Iran and showed that SVM is more accurate in predicting runoff than ANN and ANN-GP methods. Wang et al. (2009) compared the performance of different methods of monthly inflow prediction in two rivers in China. They applied the autoregressive moving average (ARMA), ANN, adaptive neural-based fuzzy inference system (ANFIS), GP, and SVM. Their results indicated that GP, ANFIS, and SVM had better results (that is, smaller errors) than several other methods. Yoon et al. (2011) applied ANN and SVM models to predict the groundwater level in coastal aquifers in Korea. Their results showed better performance for SVM than ANN. Wei (2012) coupled kernel wavelet function with SVM to predict the water level in a measuring station of the Tanshui River in China. The wavelet SVM performed better in predictions than the SVM coupled with a Gaussian kernel. Maity et al. (2013) applied the SVM and auto-regressive integrated moving average (ARIMA) method to predict monthly river inflow in the Mahanadi River in India. Their results indicated that SVM had better predictive accuracy than ARIMA. More comprehensive reviews and comments on application of SVM are

[1]M.Sc. Graduate, Dept. of Irrigation and Reclamation Engineering, Faculty of Agricultural Engineering and Technology, College of Agriculture and Natural Resources, Univ. of Tehran, Karaj, 1437835693 Tehran, Iran. E-mail: Aboutalebi@ut.ac.ir

[2]Associate Professor, Dept. of Irrigation and Reclamation Engineering, Faculty of Agricultural Engineering and Technology, College of Agriculture and Natural Resources, Univ. of Tehran, Karaj, 3158777871 Tehran, Iran (corresponding author). E-mail: OBHaddad@ut.ac.ir

[3]Professor, Dept. of Geography, Univ. of California, Santa Barbara, CA 93106. E-mail: Hugo.Loaiciga@ucsb.edu

Note. This manuscript was submitted on March 12, 2015; approved on November 19, 2015; published online on March 16, 2016. Discussion period open until August 16, 2016; separate discussions must be submitted for individual papers. This paper is part of the ***Journal of Irrigation and Drainage Engineering***, © ASCE, ISSN 0733-9437.