Water Quality Prediction Model Based Support Vector Machine Model for Ungauged River Catchment under Dual Scenarios

Abobakr Saeed Abobakr Yahya 1, Ali Najah Ahmed 1*, Faridah Binti Othman 2, Rusul Khaleel Ibrahim 2, Haitham Abdulmohsin Afan 2*, Amr El-Shafie 3, Chow Ming Fai 1, Md Shabbir Hossain 4, Mohammad Ehteram 5 and Ahmed Elshafie 2*

1 Institute of Energy Infrastructure (IEI), Universiti Tenaga Nasional, Kajang 43000, Selangor, Malaysia; ale.a.najah@gmail.com (A.S.A.Y.); Chowmf@uniten.edu.my (C.M.F.)
2 Department of Civil Engineering, Faculty of Engineering, University Malaya, Kuala Lumpur 50603, Malaysia; faridahothman@um.edu.my (F.B.O.); rusul.alqaisy@yahoo.com (R.K.I.); elshafie@um.edu.my (A.E.)
3 Civil Engineering Department El-Gazeera High Institute for Engineering Al Moqattam, Cairo 11311, Egypt; amrhuss63@gmail.com
4 Department of Civil Engineering, School of Energy, Geoscience, Infrastructure and Society, Heriot-Watt University, Putrajaya 62200, Malaysia; m.hossain@hw.ac.uk or realism007@gmail.com
5 Department of Water Engineering and Hydraulic Structures, Faculty of Civil Engineering, Semnan University, Semnan 35131-19111, Iran; mohammdehteram@semnan.ac.ir
* Correspondence: mahfoodh@uniten.edu.my (A.N.A.); haitham.afan@gmail.com (H.A.A.)

Received: 29 April 2019; Accepted: 15 May 2019; Published: 13 June 2019

Abstract: Water quality analysis is a crucial step in water resources management and needs to be addressed urgently to control any pollution that may adversely affect the ecosystem and to ensure the environmental standards are being met. Thus, this work is an attempt to develop an efficient model using support vector machine (SVM) to predict the water quality of Langat River Basin through the analysis of the data of six parameters of dual reservoirs that are located in the catchment. The proposed model could be considered as an effective tool for identifying the water quality status for the river catchment area. In addition, the major advantage of the proposed model is that it could be useful for ungauged catchments or those lacking enough numbers of monitoring stations for water quality parameters. These parameters, namely pH, Suspended Solids (SS), Dissolved Oxygen (DO), Ammonia Nitrogen (AN), Chemical Oxygen Demand (COD), and Biochemical Oxygen Demand (BOD) were provided by the Malaysian Department of Environment (DOE). The differences between dual scenarios 1 and 2 depend on the information from prior stations to forecast DO levels for succeeding sites (Scenario 2). This scheme has the capacity to simulate water-quality accurately, with small prediction errors. The resulting correlation coefficient has maximum values of 0.998 and 0.979 after the application of Scenario 1. The approach with Type 1 SVM regression along with 10-fold cross-validation methods worked to generate precise results. The MSE value was found to be between 0.004 and 0.681, with Scenario 1 showing a better outcome.

Keywords: support vector machine; water quality; dissolved oxygen

1. Introduction

Water plays a crucial role in environmental and social life. It occupies the largest area of our planet, and therefore, among all other natural resources, water resources gain the most special place. However, the ongoing increase in urbanization and industrialization processes generate wide ranges of hazardous contaminants that deteriorate the quality of river waters. The produced wastes are