Journal of Hydrology 495 (2013) 175–185

Contents lists available at SciVerse ScienceDirect

Journal of Hydrology

journal homepage: www.elsevier.com/locate/jhydrol

A reduced-order adaptive neuro-fuzzy inference system model as a software sensor for rapid estimation of five-day biochemical oxygen demand

Roohollah Noori *, Salman Safavi, Seyyed Afshin Nateghi Shahrokni

Islamic Azad University, Malard Branch, Department of Civil & Environmental Engineering, Tehran, Iran

ARTICLE INFO

Article history:
Received 22 January 2013
Received in revised form 24 April 2013
Accepted 30 April 2013
Available online 10 May 2013
This manuscript was handled by Andras Bardossy, Editor-in-Chief, with the assistance of Fi-John Chang, Associate Editor

Keywords:
Uncertainty analysis
Reduced-order model
Five-day biochemical oxygen demand (BOD_5)
Proper orthogonal decomposition
Adaptive neuro-fuzzy inference system

SUMMARY

The five-day biochemical oxygen demand (BOD_5) is one of the key parameters in water quality management. In this study, a novel approach, i.e., reduced-order adaptive neuro-fuzzy inference system (ROANFIS) model was developed for rapid estimation of BOD_5. In addition, an uncertainty analysis of adaptive neuro-fuzzy inference system (ANFIS) and ROANFIS models was carried out based on Monte-Carlo simulation. Accuracy analysis of ANFIS and ROANFIS models based on both developed discrepancy ratio and threshold statistics revealed that the selected ROANFIS model was superior. Pearson correlation coefficient (R) and root mean square error for the best fitted ROANFIS model were 0.96 and 7.12, respectively. Furthermore, uncertainty analysis of the developed models indicated that the selected ROANFIS had less uncertainty than the ANFIS model and accurately forecasted BOD_5 in the Sefidrood River Basin. Besides, the uncertainty analysis also showed that bracketed predictions by 95% confidence bound and *d*-factor in the testing steps for the selected ROANFIS model were 94% and 0.83, respectively.

1. Introduction

One of the main indices for evaluating the surface water quality is the five-day biochemical oxygen demand (BOD_5). It is widely used in the measurement of the organic pollution in water resources systems. Currently available method for BOD_5 determination is very tedious (Beltran et al., 1998; Einax et al., 1998, 1999). Conventional BOD_5 tests take a long time, about five-day incubation period, and are too complex to use in process control. On the other hand, these methods are subject to various complicating factors such as the oxygen demand resulting from the respiration of algae in the sample and the possible oxidation of ammonia. In addition, presence of toxic substances in the sample may also affect the microbial activity leading to a reduction in the measured BOD_5 value. Furthermore, the laboratory conditions for BOD_5 determination usually differ from those in aquatic systems (Singh et al., 2009). Therefore, interpretation of BOD_5 results and their implications may be associated with large variations.

A way of dealing with the mentioned problems related to BOD_5 determination is to develop new online software sensors based on information from the existing other online measured parameters. Literature provides several approaches to develop the software sensors for online modeling of BOD_5 such as classic regression methods (Oliveira-Esquerre et al., 2004a; Dogan et al., 2008) and more sophisticated approaches based on artificial intelligence (AI) techniques (Zhu et al., 1998; Oliveira-Esquerre et al., 2004b; Dogan et al., 2008; Singh et al., 2009). It is demonstrated that the AI techniques are powerful tools for developing the software sensors (Chen and Chau, 2006; Wu et al., 2009) and have better performance than the classical regression methods (Singh et al., 2010, 2012). Therefore, some studies focused on the AI techniques for online prediction of BOD_5. Oliveira-Esquerre et al. (2002) proposed an estimation model based on neural network (NN) technique that could provide accurate predictions of BOD_5 in the output stream of biological wastewater treatment plant (WWTP). They suggested that NN model is capable to represent nonlinear relationship between input data and BOD_5 in WWTP. In another work by Hamed et al. (2004), two NN models for prediction of BOD and suspended solids concentrations were presented. They concluded that both NN models were capable to provide an efficient and a robust tool for predicting the WWTP performance. Oliveira-Esquerre et al. (2004b) predicted inlet and outlet BOD_5 of an aerated lagoon using functional-link NN and multilayer perceptron NN models. They also evaluated potential improvement of these models when

* Tel.: +98 9374320526; fax: +98 2166407719.
E-mail address: roohollahnoori@gmail.com (R. Noori).

0022-1694/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jhydrol.2013.04.052

partial least square is used to preprocess input data. Finally they reported that both NN models accurately predicted inlet and outlet BOD_5 parameters. Onkal-Engin et al. (2005) developed a relationship between sewage odor and BOD_5 by NN model and reported that the NN model could be used to classify the sewage samples collected from deferent locations of WWTP. Zhao and Chai (2005) presented a hybrid time-delay NN method with data pretreatment for BOD_5 forecasting in WWTP. They reported an enhancement in speed and accuracy when the simulation results were compared with a back propagation NN model. Dogan et al. (2008) compared NN and multi-linear regression models to estimate daily BOD_5 in the inlet of WWTP. Rustum et al. (2008) developed Kohonen self-organizing NN model for the rapid prediction of BOD_5. Findings indicated that the obtained results were in a good agreement with those measured using the conventional bioassay method. Dogan et al. (2009) investigated the abilities of NN model to improve the accuracy of the BOD estimation in the Melen River Basin, Turkey. In another work by Singh et al. (2009), two NN models were identified, validated, and tested for the computation of dissolved oxygen (DO) and BOD_5 concentrations, respectively, in the Gomti River. The model computed values of DO and BOD_5 by both NN models, which were in close agreement with their measured values in the river. Noori et al. (2011a) adapted a reduced-order NN model for online prediction of BOD_5. In another work, Noori et al. (2011b) proposed a reduced-order support vector machine (SVM) for online estimation of BOD_5. Yel and Yalpir (2011) applied a fuzzy-logic-based diagnosis system to determine the primary treatment effluent quality related to BOD in a WWTP. The resulting configuration proved a good modeling approach for the WWTP effluent quality prediction. Generally results from all mentioned studies proposed that the NN model is an alternative technique for online prediction of BOD_5. In addition, literature shows that adaptive neuro-fuzzy inference systems (ANFISs) technique for online prediction of BOD_5 has seen little progress than the NN model.

However, the prediction capability of AI techniques such as NN and ANFIS strongly depends on the status of the training data. If there is noise and uncertainty in the training data, a problem of over-fitting often arises. Since AI techniques use only input and output data observed from the target system, it is necessary to extract required information from large and noisy input vectors through data preprocessing. Therefore, a large number of input vectors can be considered as one of the main common problems for modeling process using ANFIS technique (Kecman, 2005). The appropriate method for solving this problem is applying the reduced-order models. In this context, by means of ANFIS technique, various reduced-order ANFIS (ROANFIS) models have been developed for online prediction of BOD_5. To achieve this goal, the most popular approach is the proper orthogonal decomposition (POD) which would yield small degree-of-freedom models. Therefore, ROANFIS models based on POD approach can be effective tool to yield small degree-of-freedom model for on-line and real-time prediction of BOD_5.

Another problem in developing the software sensors by means of AI techniques is that the prediction results are associated with large uncertainties which makes online estimate more unreliable. It is pointed out although the previous studies using AI models overcame the shortcomings of traditional method for online estimation of BOD_5, but results of these models were not certain. Thus in application of the results of these AI models, an uncertainty analysis can be more effective. Literature shows that just a few methods have been proposed for determination of uncertainty in ANFIS model. Some of them are bootstrap and sandwich estimator (Tibshirani, 1994), maximum likelihood and Bayesian inference (Dybowski, 1997), and Mont-Carlo method (Marce et al., 2004). In this study, determination of uncertainty in ANFIS and ROANFIS techniques has been carried out using Mont-Carlo method, because it has not only better performance but also more novelty (Marce et al., 2004).

Considering the above-cited studies, the research aims to investigate: (1) the potential of ANFIS technique in online prediction of BOD_5; (2) the performance of ROANFIS models developed based on POD method; (3) the uncertainty of ANFIS and ROANFIS models for online estimation of BOD_5.

2. Materials and methods

2.1. Case study and data

The Sefidrood River Basin is selected for calibration and testing of the ANFIS and ROANFIS models. It is drained by the Sefidrood River and located in the northwest of Iran (Fig. 1). The basin area is 59,196 km^2 and located between Alborz and Zagros Mountains. The main branch of the Sefidrood River originates from the Zagros Mountains and flows down to the south of Caspian Sea. Data used in the present study were collected from 94 water quality monitoring stations along the Sefidrood River and its reaches in spring, summer, autumn, and winter seasons (Fig. 1). These datasets were collected by staff from the Department of Environment, Mahab Ghods consulting Engineers and Yekom Consulting Engineers. In this study, to develop proper ANFIS and ROANFIS models for online prediction of the maximum BOD_5, input vectors are selected as: location of stations {latitude (N) and longitude (E)}, dissolved oxygen (DO) {as minimum (DO_{min}), maximum (DO_{max}), and average (DO_{mean})}, electrical conductivity (EC) {as maximum (EC_{max}) and average (EC_{mean})}, nitrate (NO_3^-) {as maximum (NO^-_{3max}) and average (NO^-_{3mean})}, and total phosphorus (TP) {as maximum (TP_{max}) and average (TP_{mean})}.

2.2. ANFIS

"ANFIS modeling refers to the method of applying various learning techniques developed in the neural network literature to a Fuzzy Inference System (FIS) (Brown and Harris, 1994)". The FIS applies a fractional non-linear relationship to map its input space to the output space by a number of fuzzy if-then rules (Kisi and Ozturk, 2007; Moghaddamnia et al., 2009a,b; Wang et al., 2012; Chau et al., 2005). A FIS generally consist of five components; including fuzzification interface, rule base, data base, decision making unit, and defuzzification interface. FIS type selection is one of the main steps for ANFIS development. Different methods exist for developing the FIS. In this research, first order Sugeno FIS with two-fuzzy rules (Fig. 2) is used as follows:

$$\begin{aligned} &\text{Rule 1}: \quad \text{If } x \text{ is } A_1 \text{ and } y \text{ is } B_1; \text{then } f_1 = p_1x + q_1y + r_1 \\ &\text{Rule 2}: \quad \text{If } x \text{ is } A_2 \text{ and } y \text{ is } B_2; \text{then } f_2 = p_2x + q_2y + r_2 \end{aligned} \tag{1}$$

where A_1; A_2 and B_1; B_2 are the membership functions (MFs) for inputs x and y; respectively; and p_1; q_1; r_1 and p_2; q_2; r_2 are the parameters of the output function. Also, the output f is weighted average of the individual rule outputs. To implement these two rules, an equivalent ANFIS structure (Fig. 3) should be developed. In Fig. 3, characteristics of each layer are as follows (Jang, 1993):

Layer 1: All the nodes in this layer are adaptive nodes.

Layer 2: The nodes in this layer are fixed (not adaptive). These are labeled Π to indicate that they play the role of a simple multiplier. The output of each node is this layer represents the firing strength of the rule.

Layer 3: Nodes in this layer are also fixed nodes. These are labeled N to indicate that these perform a normalization of the firing strength from previous layer.