Title

Equation to Predict Riverine Transport of Suddenly Discharged Pollutants

Journal

Journal of Irrigation and Drainage Engineering, 142(11)

ISSN

0733-9437

Authors

Farhadian, Mostafa

Bozorg-Haddad, Omid

Seifollahi-Aghmiuni, Samaneh

et al.

Publication Date

2016-11-01

DOI

10.1061/(asce)ir.1943-4774.0001083

Peer reviewed

Equation to Predict Riverine Transport of Suddenly Discharged Pollutants

Mostafa Farhadian[1]; Omid Bozorg-Haddad[2]; Samaneh Seifollahi-Aghmiuni[3]; and Hugo A. Loáiciga, F.ASCE[4]

Abstract: Pollution discharge to rivers is the leading cause of freshwater pollution. This paper presents a method for predicting the concentrations of a pollutant that is suddenly released into a river. The prediction method is based on a first-order Gaussian function. The coefficients of the Gaussian function are calculated using three power equations. These equations have six parameters whose values are optimized with the genetic algorithm (GA). The results of this paper's case study are compared with those obtained with the artificial neural network method (ANN), genetic programming (GP), and from the analytical solution of the differential equation for riverine transport of suddenly released pollutants. The proposed equation is applied in a case study that confirms its suitability for estimating the concentration of pollutants downstream from a sudden release location in a river. The correlation coefficient (R^2) and the root mean square error (*RMSE*) of the training and testing data obtained from the application of the proposed equation increase and decrease approximately 14 and 50%, respectively, in comparison with the values associated with the application of the analytical advection-dispersion to calculate the transport of sudden pollution in rivers. **DOI: 10.1061/(ASCE)IR.1943-4774.0001083.**

Author keywords: Sudden pollution; Pollution propagation; Advection-dispersion equation; Genetic algorithm; Data mining tools.

Introduction

Riverine water quality has been degraded by inadvertent, intentional, and accidental discharge of pollutants in streams. The assessment of riverine water pollution is key to developing remedial and response actions (Hou et al. 2013). Water sampling is the most accurate method for establishing the degree of riverine water pollution. Yet, comprehensive spatial and temporal monitoring of riverine water quality is expensive in time and funding demands. For this reason, other indirect methods that rely on numerical prediction models of riverine water quality and pollutant transport have been developed to supplement measurement campaigns.

Matsuda (1979) developed a pollution prediction system to control the national waters and ports of Japan using the finite element method. Wen et al. (1998) used a multiobjective optimization model based on artificial neural network (ANN) to manage water quality in the Tou-Chen River in Taiwan. Demirdag et al. (2000) applied the ANN method to estimate values of the water quality parameters in the Gediz River in western Turkey. Drago et al. (2001) presented a three-dimensional numerical model for water eutrophication and pollution transmission. They used this model to analyze solid particles and pollutants discharged to water bodies. Suen and Eheart (2003) estimated the nitrate concentration in the Sangamon River of Illinois using the ANN. They compared the results of back propagation neural network (BPNN) and radial basis function neural network (RBFNN) methods with the results of the multiple regression analysis (MRA) method and the soil and water assessment tool (SWAT) considering the estimation accuracy of the nitrate concentration. Caplow et al. (2004) studied the pollution tracing and pollution advection and dispersion in the Hudson River, including several reservoirs under operation. The latter authors used a sulfur hexafluoride concentration as the pollutant and traced it for seven days by solving the advection-dispersion equation of pollution to evaluate the pollution distribution in the river. Their results indicated that the coefficients of advection and dispersion for sulfur hexafluoride were approximately 9.0 km/day and 17.3 m^2/s, respectively. El-Badia et al. (2005) identified the location and value of point-source pollution using mathematical models of river flow and contaminant transport. Jha et al. (2005) reported that fertilizers and chemical substances used on Indian farms cause nonpoint pollution sources and discharge of these pollutants to surface water and ground waters. They estimated input nutrients from three farms relying on 576 sets of water quality data. Riahi-Madvar et al. (2009) developed a method for estimating the coefficient of longitudinal dispersion in rivers using adaptive network-based fuzzy inference system (ANFIS), field measurements, and several statistical indices. The obtained results indicated that the developed method is more accurate than 12 other experimental methods used to calculate the dispersion coefficient. Noori et al. (2011) determined that the dispersion coefficient is the most important factor for modeling pollutant transport in surface waters. They developed a model based on the ANN for predicting the dispersion coefficient in natural waters. Tong and Deng (2015) established that identifying unknown sources of pollution is essential

[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, 31587 77871 Alborz, Iran.

[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, 31587 77871 Alborz, Iran (corresponding author). E-mail: OBHaddad@ut.ac.ir

[3]Ph.D. Candidate, Dept. of Irrigation and Reclamation Engineering, Faculty of Agricultural Engineering and Technology, College of Agriculture and Natural Resources, Univ. of Tehran, Karaj, 31587 77871 Alborz, Iran.

[4]Professor, Dept. of Geography, Univ. of California, Santa Barbara, CA 93106.

Note. This manuscript was submitted on January 27, 2016; approved on April 13, 2016; published online on June 22, 2016. Discussion period open until November 22, 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.

for environmental protection and emergency actions. These authors developed two equations for the pollution location and value (the total mass of released pollution) based on the variable residence time equation. Their equations were used to detect 23 pollutant injection locations in five rivers in the United States. Heddam (2015) developed a dynamic evolving neural-fuzzy inference system (DENFIS) using artificial intelligence (AI) models for modeling riverine dissolved oxygen (DO) concentration. This latter author applied the developed model to hourly-observed data recorded by the United States Geological Survey (USGS) in the Klamath River. The qualitative indexes of the DENFIS system included pH, temperature, specific conductance, and water depth. The performance of DENFIS was evaluated with the root mean square error (*RMSE*), the mean absolute error (MAE), and other statistical indices, and the study results were compared with the results from linear and nonlinear regressions. Khorsandi et al. (2015) determined the location, concentration, and the timing of pollution injection to surface waters using data mining (ANN model) and optimization [GA and pattern search (PS)] methods. These authors applied the *CE-QUAL-W2* numerical model coupled with ANN to solve several hypothetical examples of pollution in a canal considering several types of pollution injection.

Calculation of the riverine transport of suddenly discharged pollutants is of utmost importance considering that it has not been fully addressed in the literature (Ahmadi et al. 2014; Ashofteh et al. 2013a, b, 2015a, c, b; Beygi et al. 2014; Bolouri-Yazdeli et al. 2014; Bozorg-Haddad et al. 2013, 2014, 2015b, a; Fallah-Mehdipour 2013b, a; Orouji et al. 2013, 2014; Shokri et al. 2013, 2014; Seifollahi-Aghmiuni et al. 2013; Soltanjalili et al. 2013). The concentration of river pollutants can be calculated solving the differential equations for pollution transport. The diffusion coefficient, a key factor in those equations, is difficult to estimate for streams of variable characteristics. This paper presents a method to calculate pollutant transport in rivers using data mining and a simulation method based on a first-order Gaussian function. The constant coefficients of the first-order Gaussian function are determined using three power equations. Each of these three power equations has two parameters that are optimized with the GA. A case study illustrates the applicability of the developed methodology.

Data Mining Methods

Data mining (sometimes called data or knowledge discovery) is the process of analyzing data from different perspectives and summarizing it in useful information that can be used to optimize performance of managed systems. Technically, data mining is the process of finding correlations or patterns among many fields in large relational databases. Data mining software is one of a number of analytical tools for analyzing data. ANNs and genetic programming (GP) are herein applied for estimating pollution concentrations in rivers.

Artificial Neural Network (ANN)

The ANN is a method commonly used in studies of water pollution. The ANN has been found to be a very capable method for data modeling. ANN is applied in this research for modeling riverine pollution transport. ANN is designed to mimic the human neural network system. It is an intelligent system that discovers relations among data through analysis. ANN is herein applied for riverine pollution transport. The two main varieties of ANNs are the multilayer perceptron (MLP) neural network and the general regression neural network (GRNN), which are applied in this study.

Multilayer Perceptron (MLP)

The MLP is one of the simplest and the most efficient layouts in ANN, whereby a MLP model consists of one input layer, one or more hidden layer(s), and one output layer. All of the neurons in one layer of the MLP network are connected to all neurons in the next layer producing a completely connected network. In MLP networks the number of neurons in each layer independent of the number of neurons in other layers.

The precision of data modeling increases by considering a sufficient number of neurons in the middle layers of the MLP neural network. The number of neurons in the middle layers should be increased when the complexity of data relationships is high, which can be done by a trial and error procedure. A simple or complex network can cause unsuitable training (under-fitting) or over training (over-fitting), respectively (Hosseini-Moghari et al. 2015). In the under-fitting conditions, the network learns very little about the existing relationships between training data which causes undesirable results in data training and testing. In the over-fitting conditions, the network models only the training data to learn about the relationship between inputs and outputs, and there is unsuitable prediction for new data. In other words, data training is well done, but testing is deficient.

General Regression Neural Network (GRNN)

The GRNN is a special type neural network that uses radial basis function, and its number of neurons exceeds those of MLP neural networks. GRNN is trained very fast and functions very well in problems with various types of input and output data. GRNN is a normalized radial basis function (RBF) network in which one neuron forms the hidden layer for each input data set. This network is able to estimate the parameters of the regression equations with high precision. GRNN has three layers: (1) the first receives inputs and data, (2) the second layer is a hidden one where radial functions are used in its neurons and the number of these neurons is equal to the number of input variables, and (3) the third is an output layer that collects the outputs from the second layer and generates the final outputs.

Genetic Programming (GP)

GP is an evolutionary method inspired by Darwin's theory of evolution that features crossover, mutation, and selection operators. GP implements mathematical and logical functions and operators as decision variables in addition to the common decision variables of a problem. It functions with a tree structure instead of a binary structure. Each chromosome in the initial population of the GP has a set of functions and terminals. The function set includes mathematic functions for summation, subtraction, multiplication, division, trigonometry, and others. The terminal set includes variables and constant numbers (Koza 1992). Mathematical equations are defined between different inputs and outputs using the GP, and it is applied to calculate outputs according to the input data. GP has been proven an efficient model for water resources system modeling in recent years. This study applies it to model riverine pollutant transport.

1D Advection-Dispersion Equation of Pollutant Transport in Rivers

Pollutant transport in a river is described by the one-dimensional advection-dispersion equation [Eq. (1)] (Van Genuchten and Alves 1982). This equation assumes the complete mixing of a pollutant with flow depth in a river. Other assumptions are: (1) there is no

pollutant input and output along the river, (2) the dispersion and decay coefficients are constant, and (3) hydraulic (physical property) parameters of the river such as the velocity, flow, and section area are constant:

$$\frac{\partial c}{\partial t} = -u\frac{\partial c}{\partial x} + D\frac{\partial^2 c}{\partial x^2} - kc \tag{1}$$

in which c = pollutant concentration (mg/l); x = distance between prediction and pollution discharge points (m); t = time elapsed since pollution injection in the river (s); D = dispersion coefficient (m^2/s); u = river flow velocity (m/s); and k = coefficient of pollution decay (s^{-1}). Eq. (2) represents the analytical solution of Eq. (1) for the case of a sudden pollution injection (Farhadian et al. 2015):

$$c(x,t) = \frac{M}{2A\sqrt{\pi Dt}}\exp\left[\frac{-(x-ut)^2}{4Dt} - kt\right] \tag{2}$$

in which $c(x,t)$ = pollutant concentration at location x and at time t (g/l); M = sudden pollutant mass injected at time $t = 0$ (kg), and A = flow area of the river cross section (m^2). The dispersion and decay coefficients are dominant factors in predicting pollutant transport using Eq. (2). Their values can be determined with experimental study of the river or based on previous results obtained by previous researchers. One of the most common experimental equations is the Fischer equation (1975) [Eq. (3)]:

$$D = \frac{0.011u^2w^2}{h} \tag{3}$$

in which w = width of the flow section (m); and h = flow depth (m). The pollution concentrations downstream of the location of pollutant injection into the river can be estimated by calculating the dispersion coefficient with Eq. (3) and applying the calculated value in Eq. (1).

Proposed Equation

Consider a function whose shape is similar to the sudden-pollution transport equation. The first-order Gaussian function meets such a criterion, and is given by Eq. (4):

$$f(x) = a_1\exp\left[-\left(\frac{x-a_2}{a_3}\right)^2\right] \tag{4}$$

in which $f(x)$ = dependent variable; x = independent variable; and a_1, a_2 and a_3 = coefficients. The purpose is to match the graph of the first-order Gaussian function to the graph of pollutant concentration in a river. This can done for all locations along in a river by changing the values of the coefficients a_1, a_2 and a_3. Thus, the coefficients of the Gaussian function are determined for different locations of the river by relating pollutant concentration to the Gaussian function coefficients. It is necessary to establish a suitable relationship between the coefficients a_1, a_2, and a_3 and the distance from the point of pollutant injection in a river. These coefficients are considered to be dependent variables in power equations in which the distance from the point of pollutant injection is the independent variable according to Eqs. (5)–(7):

$$a_1 = p_1x^{p_2} \tag{5}$$

$$a_2 = p_3x^{p_4} \tag{6}$$

$$a_3 = p_5x^{p_6} \tag{7}$$

where $p_1, p_2, \ldots, p_6$ = constant coefficients. Eqs. (5)–(7) involve few constant coefficients with which to calculate a_1, a_2, and a_3.

According to Eqs. (5)–(7), the coefficients a_1, a_2, and a_3, depend on x, and the six p coefficients are a characteristics of the river. The coefficients a_1, a_2, and a_3 are calculated for each location x on the river by first determining the p coefficients. The coefficients p_i $i = 1, 2, \ldots, 6$ are determined with GA optimization. Thereafter, the graph of pollution at each location is obtained with the fitted first-order Gaussian function.

Genetic Algorithm (GA)

Determining suitable values of the six p coefficients is accomplished by minimizing the *RMSE* with the GA. The GA is a versatile optimization algorithm. This algorithm is inspired by evolutionary phenomena such as heredity and gene mutation. It generates a random population of all the decision variables of the problem at first, and then calculates the objective function for all the decision variables. Then, the algorithm selects several solutions according to their competence considering their desirability as measured by the objective function. The algorithm uses current possible solutions to produce new, improved, solutions for the next step of the optimization. This is an iterative process, and the algorithm improves the solutions in each step until reaching a stopping criterion.

Assessment Criteria

This study applies two performance criteria to assess the predictive skill of the method for calculating pollutant transport. These are the *RMSE* and R^2, which are used to evaluate the closeness between the observed and calculated pollutant concentrations. The *RMSE* is calculated as follows:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(c-\hat{c})^2} \tag{8}$$

where i = counter of concentrations; c = observed concentration (ppb); $\hat{c}$ = calculated concentration (ppb); and n = number of measuring points. The R^2 is given by Eq. (9):

$$R^2 = \frac{[\sum(c-\bar{c})(\hat{c}-\breve{c})]^2}{\sum(c-\bar{c})^2 \times \sum(\hat{c}-\breve{c})^2} \tag{9}$$

where $\bar{c}$ = average of the observed concentrations (ppb) and $\breve{c}$ = average of the calculated concentrations (ppb).

Case Study

A set of pollutant concentrations at different points of the river is required to apply this paper's methods. The results of pollutant tracking obtained by Atkinson and Davis (2000) are used in this study. Their test was done in the Severn River of England, downstream of the Llanidloes Bridge. These authors selected a river reach 13.7 km long without any tributaries and with high flow discharge, and repeated the test twice to complete the missing and questionable data in the second iteration. They used Rhodamine as the tracer, which does not decay and injected it rapidly simulating sudden injection into the river. Fig. 1 shows the region of the case study.

Atkinson and Davis (2000) considered seven measurement stations ranging between 210 and 13,775 m from the injection point to estimate the tracer concentration. The results obtained in the first six stations are used in this study, whereas the results at the

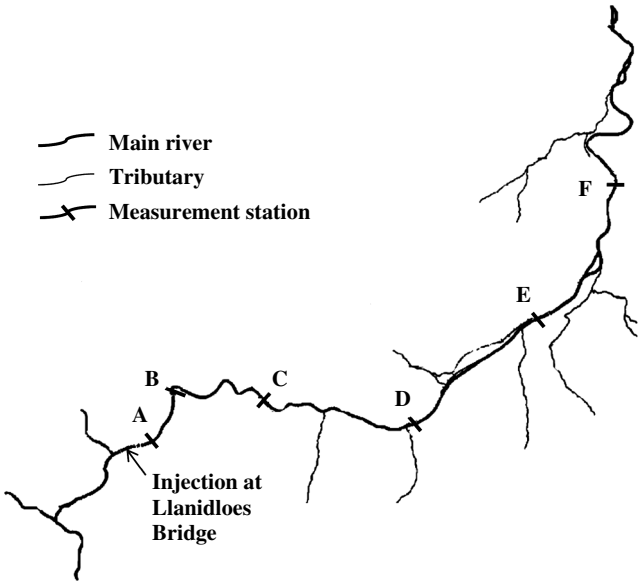

Fig. 1. Severn River in the study area (adapted from Atkinson and Davis 2000)

Table 1. Distance from the Injection Point and Hydraulic Characteristics of Measurement Stations

Station	Distance from injection point (m)	Area (m^2)	Flow (m^3/s)	Flow velocity (m/s)
A	210	10.62	7.33	0.69
B	1,175	9.13	7.03	0.77
C	2,875	10.81	7.24	0.67
D	5,275	10.58	7.51	0.71
E	7,775	22.56	9.25	0.41
F	10,275	13.80	9.80	0.71

seventh station are ignored attributable to poor accuracy of the concentration measurements at that station. The distance of these six stations from the injection point and their hydraulic characteristics are listed in Table 1.

After injecting the tracer, its concentration and the time of measurement were recorded as shown in Fig. 2. The tracer injection time determined starting time of sampling.

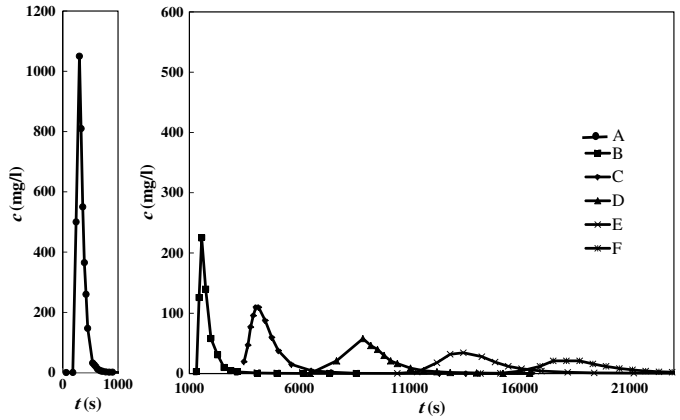

Fig. 2. Tracer concentration at six stations of the Severn River

Results and Discussion

The calculated results from the data mining method, the analytical equation, and the proposed method to estimate pollutant transport are summarized below.

Pollutant Concentration Estimation with the Data Mining Method

This section presents the results from the ANN and GP models.

Results of the ANN Method

The input and output data vectors were defined for the ANN. The locations of measurement stations (Table 1) and the measurement times (Fig. 2) are the input data to the ANN. The measured pollutant concentrations (Fig. 2) are its output data.

The data are divided into training, validation, and testing sets in the ANN, which can be done selectively or randomly. This model fits a suitable equation to the data using the training and validation data sets and evaluates the performance of the obtained equation using the testing data. The training and validation data sets are together called the training data, and the *RMSE* and R^2 assessment criteria are calculated for the training and testing sets separately. The training and testing data are selected from among all available data by the user when applying a selective division. Also, in the random division, a percentage of all the data is randomly determined for training, and the remaining data are considered as the testing data. Selective division of data was implemented for assessing the performance of the ANN in predicting the pollutant concentration, and 70% of the data were used for model training and the remaining data used as testing data. Specifically, 70% of the data were selected from the beginning of the data set for training, and the remaining 30% of the data were selected as the data set for testing progression from upstream stations to downstream stations in the river.

The MLP and GRNN networks are well suited for data modeling, and were implemented in this study with the *MATLAB* software. The number of neurons and layers in the MLP network and the number of parameters in the GRNN network were determined by trial and error. Then, the assessment criteria were calculated for the MLP and GRNN networks (Table 2).

It is seen in Table 2 that the ANN is not suitable for training data or testing data, and it is therefore for predicting the pollutant concentration in the river. The low predictive skill of the ANN is attributable to using location and time as the input data, only, because location and time change with pollutant transport in the river, whereas other parameters such as flow area, flow velocity, and dispersion and decay coefficients are constant along the river or exhibit independent changes.

Table 2. Results of the Assessment Criteria for ANN, GP, Analytical Equations, and Proposed Equation

	Training data		Testing data	
Model	R^2	RMSE	R^2	RMSE
ANN-MLP	0.34	128.6	0.69	148.6
ANN-GRNN	0.87	64.6	0.00	13.3
GP	0.81	76.7	0.09	19
Analytical equation using Fischer's dispersion	—	—	0.85	44.5
Analytical equation with optimized dispersion	0.98	29.6	0.69	52.6
Proposed equation	0.99	33.7	0.72	30.3

Results of the GP Method

The input and output data for GP and ANN are the same. Data were divided into two main groups for training (70% of all the data) and testing (the remaining 30% of all the data) in GP, similar to the ANN method. The *GENEXPRO* software developed by Ferreira (2001) was used to implement the GP method. The results for the performance criteria of the GP are listed in Table 2. It is seen in Table 2 that the accuracy of GP for modeling pollutant transport is poor, and its results are similar to those from the ANN. Therefore, these two data mining methods were not suitable for modeling the transport of sudden pollutant injection in the study river.

Results from the Analytical Equation

Two procedures were implemented for calculating the dispersion coefficient. The first is based on Fischer's equation (1975), and the second relies on GA optimization whereby the dispersion coefficient was considered as the decision variable. The purpose of optimization with GA is achieving good resemblance between concentrations calculated with the analytical equation and observed ones available in the case study.

Analytical Equation Using Fischer's Dispersion Coefficient

The dispersion coefficient is calculated using the Eq. (3) and its value is applied in the analytical equation. The downstream pollutant concentration is estimated considering the location and measurement time in each of six stations. The R^2 and *RMSE* criteria were calculated for the predicted concentrations and were equal to 0.85 and 44.5, respectively (these are listed in Table 2).

It is not necessary to divide the data set into training and testing sets using the analytical equation because this equation only requires the values of the parameters and does not require training. Therefore, all the data were used for testing of the analytical equation, and the assessment criteria are considered only for testing of this equation. Values of these criteria show that the analytical equation yields better results than the data mining methods even though only 30% of the data were used for testing the data mining method. Also, the accuracy of the analytical equation is poor for estimating pollutant concentration even though it produced better results than those of the ANN and GP methods. The observed and calculated concentrations obtained from the analytical equation are shown in Fig. 3. In Fig. 3 and other following figures, the peak of the pollutant graph at each station is specified with the name of the related station to identify the observed and calculated pollution graphs of each station.

Fig. 3 shows that the analytical equation yielded a suitable estimation of pollutant concentration only at station A using the dispersion coefficient obtained from the Fischer (1975) equation. The calculated concentrations at other stations are very different from the observed ones in magnitude and occurrence time. The maximum difference in concentration value and time between the calculated and observed concentrations occurred at stations E and F. These results show that the analytical equation has undesirable performance for estimating the pollutant concentration in the case study using the Fischer dispersion coefficient, which is the result of numerous changes in the physical and hydraulic characteristics along the river path.

Analytical Equation Using an Optimized Dispersion Coefficient

The dispersion coefficient was herein considered as a decision variable, and its value was optimized for each river reach. These reaches encompass the distance between the injection point and station A and the distances between the six considered stations that define six reaches. Therefore, six optimal values of the dispersion coefficient were determined. The optimization minimized the *RMSE* between the observed and calculated concentrations. All possible values of the dispersion coefficients were evaluated with this procedure, and the value which causes the maximum similarity between the observed and calculated concentrations was selected.

The data for stations A and D were used as training data, and data for the other stations were used as testing data and yielded the best results using a trial and error procedure. Therefore, the value of the dispersion coefficient was determined with the GA to minimize the value of the *RMSE* between the observed and calculated concentrations at stations A and D. The pollution concentration at each station was calculated with the optimized value of the dispersion coefficient in Eq. (2) (the analytical equation). The values of the assessment criteria for training and testing data are listed in Table 2. Also, the observed and calculated concentrations using this method are shown in Fig. 4. The results indicate that considering a wide range of variation for the dispersion coefficient and selecting an optimal value for it with the optimization model produces better results than those calculated with other experimental equations

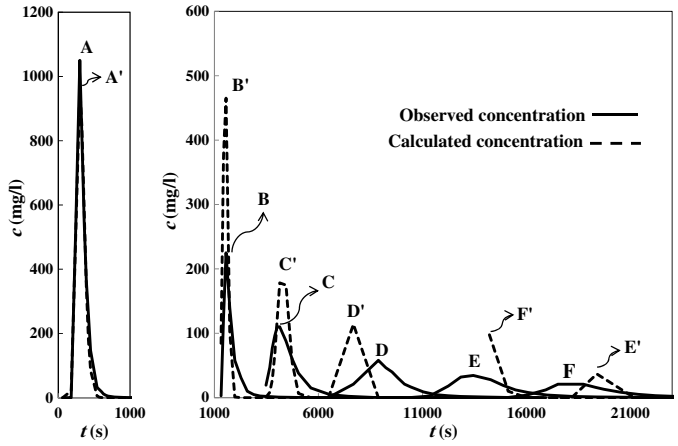

Fig. 3. Graph of observed and calculated pollutant concentrations using the analytical equation with optimized dispersion coefficients

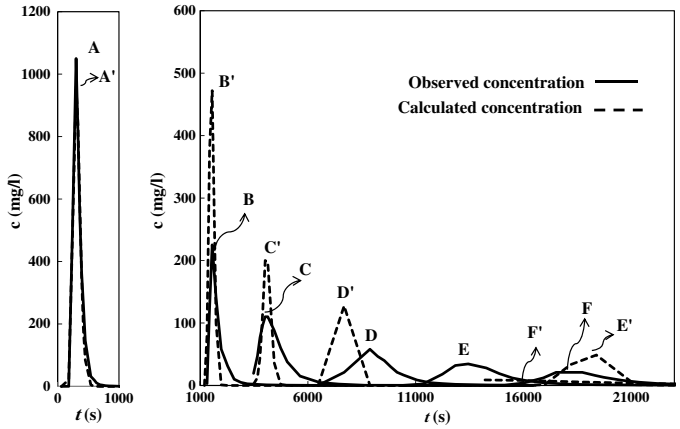

Fig. 4. Graph of observed and calculated pollution concentrations using the analytical equation with optimized dispersion coefficients

developed for calculating the dispersion coefficient. Fig. 4 demonstrates that using the optimized dispersion coefficient is effective in increasing the accuracy of the calculated pollutant concentrations using the analytical equation. The improvement was observed at all stations, but still the overall accuracy of this method is questionable. The inability of the optimized dispersion coefficient in yielding accurate concentrations shows that, unlike the results of previous studies, the dispersion coefficient does not affect the results of mathematical equations for pollutant transport to achieve calculated concentrations that are suitably close to observed ones.

Calculated Pollutant Concentration with the Proposed Equation

The proposed equation is a dimensionless equation based on finding six constant coefficients [Eqs. (5)–(7)]. The observed concentrations at stations A and D served as the training data to determine the values of these coefficients. The data of the other stations were used for testing of the proposed equation. The GA was used to optimize the values of these six coefficients by minimizing the *RMSE* between the observed and calculated concentration at stations A and D. The a_1, a_2, and a_3 coefficients are calculated for all stations [Eqs. (5)–(7)] based on the calculated values of the six p coefficients and the distance of each station from the injection point (Table 1). Thereafter, the pollution concentration at each station was calculated considering the a_1, a_2, and a_3 coefficients in the Gaussian function and the concentration measurement time from the Atkinson and Davis study (2000) as the input to this function. The assessment criteria for the developed equation were calculated according to the pollutant concentrations at the training and testing stations (Table 2). Also, Fig. 5 shows the observed and calculated pollution concentrations using the proposed equation. Fig. 5 demonstrates that there is a suitable correlation between the observed and calculated concentrations at stations A and D. Also, the results of the proposed equation are much better than those from the data mining methods and the analytical equation according to the testing stations. The proposed equation does not consider the hydraulic parameters that change along the 13.7 km study reach. Yet, the proposed equation acceptably estimated the riverine pollutant concentration by determining appropriate values for six p coefficients. This demonstrates the predictive skill of the proposed equation for estimating riverine pollutant concentration.

The flow in the river of our case study is slightly increased by lateral flows. In addition, the concentration measurements had insufficient accuracy at several stations (Atkinson and Davis 2000). These and other factors affected the measurement accuracy of observed concentrations by Atkinson and Davis (2000). This is obvious in the graph of observed concentrations at all stations, especially in stations D, E, and F whose peak values are not bell shaped. This lack of accuracy in measuring the observed concentrations rendered the Gaussian function and the calculated coefficients unable to completely match the hydrograph of the calculated concentrations to the observed ones. Therefore, the accuracy of the calculated concentrations by the proposed equation was decreased (Fig. 5).

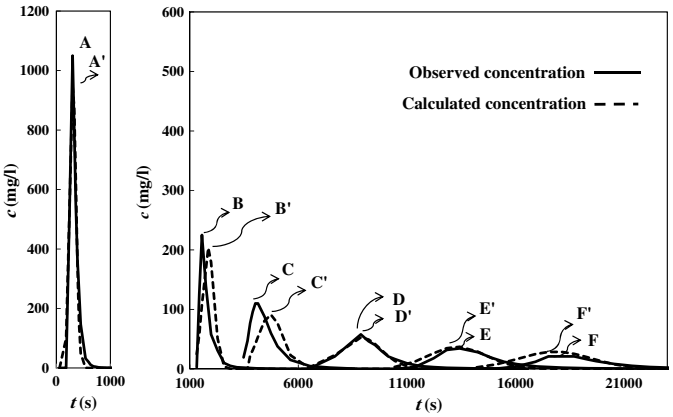

Fig. 5. Graph of observed and calculated pollution concentrations using the proposed equation

All methods of estimating the pollutant concentration have some strengths and weaknesses. The weakness of the proposed equation is its requirement to have concentration and time data of pollution graphs downstream of the injection point to calculate the values of the six p coefficients. The number of required data must be sufficient to prevent decreasing the accuracy of the calculated concentrations when the hydraulic parameters are variable. Thus, it would be necessary to measure these data for each river and use them for predicting pollutant concentrations when a sudden release occurs.

The analytical equation is commonly used for modeling riverine pollutant transport. The analytical equation with optimized dispersion coefficient is a procedure developed in this study. The R^2 and RMSE obtained with the analytical equation equaled 0.85 and 44.5, respectively, and they were equal to 0.92 and 44.1 with the analytical equation with an optimized dispersion coefficient. The R^2 and RMSE obtained with the proposed equation equaled 0.99 and 22.16, respectively. Therefore, the improvement of R^2 and RMSE with the proposed equation relative to the analytical equation are equal to 14 and 50%, respectively, and are equal to 7 and 50%, respectively, compared with the analytical equation with an optimized dispersion coefficient.

The proposed equation requires only two parameters of location and time for estimating pollutant concentration at each point downstream of the injection point after determining the values of the six p coefficients, which is another advantage of the proposed equation. This is valuable in comparison with the mathematical equations and especially with computer simulation models that require values of many parameters such as the flow area, flow velocity, type of river bed, dispersion and decay coefficients, slope, water temperature, insolation, wind speed, and so forth.

Concluding Remarks

A mathematical equation using first-order Gaussian function for estimating pollution concentration in rivers was proposed in this study. This equation has three variable parameters, and the values of these parameters, through three equations, are optimized using GA. To evaluate the effectiveness of the proposed equation, a case study that includes results of a pollutant tracking experiment in the river, was used. In addition to the proposed equation, ANN, GP, and mathematical equations were used to compare the results obtained from the proposed equation with other important methods of estimating pollution concentration in riverine environments.

The results of the proposed equation, using the assessment criteria R^2 and RMSE, were compared with the results of the methods of ANN and GP and the mathematical equation to estimate pollution concentration in the river. The results showed that the proposed equation has better results than all the above methods judged by its R^2 and RMSE, achieving improvements of 7 and 50%, respectively. Therefore, the equation proposed in this study is a practical, accurate, and straightforward approach to estimate pollution concentration caused by suddenly discharged pollutants in rivers.

References

Ahmadi, M., Bozorg-Haddad, O., and Mariño, M. A. (2014). "Extraction of flexible multi-objective real-time reservoir operation rules." *Water Resour. Manage.*, 28(1), 131–147.

Ashofteh, P. S., Bozorg-Haddad, O., and Loáiciga, H. A. (2015a). "Evaluation of climatic-change impacts on multiobjective reservoir operation with multiobjective genetic programming." *J. Water Resour. Plann. Manage.*, 10.1061/(ASCE)WR.1943-5452.0000540, 04015030.

Ashofteh, P.-S., Bozorg-Haddad, O., Akbari-Alashti, H., and Mariño, M. A. (2015b). "Determination of irrigation allocation policy under climate change by genetic programming." *J. Irrig. Drain. Eng.*, 10.1061/(ASCE)IR.1943-4774.0000807, 04014059.

Ashofteh, P.-S., Bozorg-Haddad, O., and Mariño, M. A. (2013a). "Climate change impact on reservoir performance indices in agricultural water supply." *J. Irrig. Drain. Eng.*, 10.1061/(ASCE)IR.1943-4774.0000496, 85–97.

Ashofteh, P.-S., Bozorg-Haddad, O., and Mariño, M. A. (2013b). "Scenario assessment of streamflow simulation and its transition probability in future periods under climate change." *Water Resour. Manage.*, 27(1), 255–274.

Ashofteh, P.-S., Bozorg-Haddad, O., and Mariño, M. A. (2015c). "Risk analysis of water demand for agricultural crops under climate change." *J. Hydrol. Eng.*, 10.1061/(ASCE)HE.1943-5584.0001053, 04014060.

Atkinson, T. C., and Davis, P. M. (2000). "Longitudinal dispersion in natural channels. L: Experimental results from the River Severn, U.K." *Hydrol. Earth Syst. Sci. Discuss.*, 4(3), 345–353.

Beygi, S., Bozorg-Haddad, O., Fallah-Mehdipour, E., and Mariño, M. A. (2014). "Bargaining models for optimal design of water distribution networks." *J. Water Resour. Plann. Manage.*, 10.1061/(ASCE)WR.1943-5452.0000324, 92–99.

Bolouri-Yazdeli, Y., Bozorg-Haddad, O., Fallah-Mehdipour, E., and Mariño, M. A. (2014). "Evaluation of real-time operation rules in reservoir systems operation." *Water Resour. Manage.*, 28(3), 715–729.

Bozorg-Haddad, O., Ashofteh, P.-S., Ali-Hamzeh, M., and Mariño, M. A. (2015a). "Investigation of reservoir qualitative behavior resulting from biological pollutant sudden entry." *J. Irrig. Drain. Eng.*, 10.1061/(ASCE)IR.1943-4774.0000865, 04015003.

Bozorg-Haddad, O., Ashofteh, P.-S., and Mariño, M. A. (2015b). "Levee's layout and design optimization in protection of flood areas." *J. Irrig. Drain. Eng.*, 10.1061/(ASCE)IR.1943-4774.0000864, 04015004.

Bozorg-Haddad, O., Ashofteh, P.-S., Rasoulzadeh-Gharibdousti, S., and Mariño, M. A. (2014). "Optimization model for design-operation of pumped-storage and hydropower systems." *J. Energy Eng.*, 10.1061/(ASCE)EY.1943-7897.0000169, 04013016.

Bozorg-Haddad, O., RezapourTabari, M. M., Fallah-Mehdipour, E., and Mariño, M. A. (2013). "Groundwater model calibration by meta-heuristic algorithms." *Water Resour. Manage.*, 27(7), 2515–2529.

Caplow, T., Schlosser, P., and Ho, D. T. (2004). "Tracer study of mixing and transport in the upper Hudson River with multiple dams." *J. Environ. Eng.*, 10.1061/(ASCE)0733-9372(2004)130:12(1498), 1498–1506.

CE-QUAL-W2 [Computer software]. DSI, Inc.

Demirdag, O., Yurdusev, M. A., and Solmaz, B. (2000). "Application of artificial neural networks to the estimation of water quality parameters of river Gediz." *Building Partnerships*, ASCE, Minneapolis, 1–5.

Drago, M., Cescon, B., and Iovenitti, L. (2001). "A three-dimensional numerical model for eutrophication and pollutant transport." *Ecol. Modell.*, 145(1), 17–34.

El-Badia, A., Ha-Duong, T., and Hamdi, A. (2005). "Identification of a point source in a linear advection–dispersion–reaction equation: Application to a pollution source problem." *Inverse Prob.*, 21(3), 1121–1136.

Fallah-Mehdipour, E., Bozorg-Haddad, O., and Mariño, M. A. (2013a). "Extraction of optimal operation rules in aquifer-dam system: A genetic programming approach." *J. Irrig. Drain. Eng.*, 10.1061/(ASCE)IR.1943-4774.0000628, 872–879.

Fallah-Mehdipour, E., Bozorg-Haddad, O., and Mariño, M. A. (2013b). "Prediction and simulation of monthly groundwater levels by genetic programming." *J. Hydro-Environ. Res.*, 7(4), 253–260.

Farhadian, M., Haddad, O. B., Seifollahi-Aghmiuni, S., and Loáiciga, H. A. (2015). "Assimilative capacity and flow dilution for water quality protection in rivers." *J. Hazard. Toxic Radioact. Waste*, 10.1061/(ASCE)HZ.2153-5515.0000234, 04014027.

Ferreir, C. (2001). "Gene expression programming: A new adaptive algorithm for solving problems." *Complex Syst.*, 13(2), 87–129.

Fischer, H. B. (1975). "Discussion of simple method for predicting dispersion in streams." *J. Environ. Eng.Div.*, 101(3), 453–455.

GENEXPRO [Computer software]. GeneXpro Inc., Brossard, QC, Canada.

Heddam, S., and Dechemi, N. (2015). "A new approach based on the dynamic evolving neural-fuzzy inference system (DENFIS) for modelling coagulant dosage (Dos): Case study of water treatment plant of Algeria." *Desalin. Water Treat.*, 53(4), 1045–1053.

Hosseini-Moghari, S. M., and Araghinejad, S. (2015). "Monthly and seasonal drought forecasting using statistical neural networks." *Environ. Earth Sci.*, 74(1), 397–412.

Hou, D., Song, X., Zhang, G., Zhang, H., and Loáiciga, H. A. (2013). "An early-warning and control system for urban, drinking, water quality protection: China's experience." *Environ. Sci. Pollut. Res.*, 20(7), 4496–4508.

Jha, R., Ojha, C. S., and Bhatia, K. K. (2005). "Estimating nutrient outflow from agricultural watersheds to the river Kali in India." *J. Environ. Eng.*, 10.1061/(ASCE)0733-9372(2005)131:12(1706), 1706–1715.

Khorsandi, M., Bozorg Haddad, O., and Mariño, M. A. (2015). "Application of data-driven and optimization methods in identification of location and quantity of pollutants." *J. Hazard. Toxic Radioact. Waste*, 10.1061/(ASCE)HZ.2153-5515.0000238, 04014031.

Koza, J. R. (1992). *Genetic programming: On the programming of computers by means of natural selection*, Vol. 1, MIT Press, Boston.

MATLAB [Computer software]. MathWorks, Natick, MA.

Matsuda, Y. (1979). "A water pollution prediction system by the finite element method." *Adv. Water Resour.*, 2(1), 27–34.

Noori, R., Karbassi, A. R., Mehdizadeh, H., Vesali-Naseh, M., and Sabahi, M. S. (2011). "A framework development for predicting the longitudinal dispersion coefficient in natural streams using an artificial neural network." *Environ. Progress Sustainable Energy*, 30(3), 439–449.

Orouji, H., Bozorg-Haddad, O., Fallah-Mehdipour, E., and Mariño, M. A. (2013). "Modeling of water quality parameters using data-driven models." *J. Environ. Eng.*, 10.1061/(ASCE)EE.1943-7870.0000706, 947–957.

Orouji, H., Bozorg-Haddad, O., Fallah-Mehdipour, E., and Mariño, M. A. (2014). "Extraction of decision alternatives in project management: Application of hybrid PSO-SFLA." *J. Manage. Eng.*, 10.1061/(ASCE)ME.1943-5479.0000186, 50–59.

Riahi-Madvar, H., Ayyoubzadeh, S. A., Khadangi, E., and Ebadzadeh, M. M. (2009). "An expert system for predicting longitudinal dispersion coefficient in natural streams by using ANFIS." *Expert Syst. Appl.*, 36(4), 8589–8596.

Seifollahi-Aghmiuni, S., Bozorg-Haddad, O., and Mariño, M. A. (2013). "Water distribution network risk analysis under simultaneous consumption and roughness uncertainties." *Water Resour. Manage.*, 27(7), 2595–2610.

Shokri, A., Bozorg-Haddad, O., and Mariño, M. A. (2013). "Reservoir operation for simultaneously meeting water demand and sediment flushing: A stochastic dynamic programming approach with two uncertainties." *J. Water Resour. Plann. Manage.*, 10.1061/(ASCE)WR.1943-5452.0000244, 277–289.

Shokri, A., Bozorg-Haddad, O., and Mariño, M. A. (2014). "Multi-objective quantity-quality reservoir operation in sudden pollution." *Water Resour. Manage.*, 28(2), 567–586.

Soltanjalili, M., Bozorg-Haddad, O., and Mariño, M. A. (2013). "Operating water distribution networks during water shortage conditions using hedging and intermittent water supply concepts." *J. Water Resour. Plann. Manage.*, 10.1061/(ASCE)WR.1943-5452.0000315, 644–659.

Suen, J. P., and Eheart, J. W. (2003). "Evaluation of neural networks for modeling nitrate concentrations in rivers." *J. Water Resour. Plann. Manage.*, 10.1061/(ASCE)0733-9496(2003)129:6(505), 505–510.

Tong, Y., and Deng, Z. (2015). "Moment-based method for identification of pollution source in rivers." *J. Environ. Eng.*, 10.1061/(ASCE)EE.1943-7870.0000683, 04015026.

vanGenuchten, M. Th., and Alves, W. J. (1982). "Analytical solutions of the one-dimensional convective-dispersive solute transport equation." *Technical Bulletin No. 1661*, U.S. Dept. of Agriculture, Washington, DC, 151.

Wen, C. G., and Lee, C. S. (1998). "A neural network approach to multiobjective optimization for water quality management in a river basin." *Water Resour. Res.*, 34(3), 427–436.