Journal of Hydrology 524 (2015) 587–596

Contents lists available at ScienceDirect

Journal of Hydrology

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

Gene expression models for prediction of longitudinal dispersion coefficient in streams

Ahmed M.A. Sattar [a,1], Bahram Gharabaghi [b,*]

[a] Department of Irrigation & Hydraulics, Faculty of Engineering, Cairo University, Giza, Egypt
[b] School of Engineering, University of Guelph, Guelph, Ontario NIG 2 W1, Canada

ARTICLE INFO

Article history:
Received 30 June 2014
Received in revised form 7 December 2014
Accepted 6 March 2015
Available online 14 March 2015
This manuscript was handled by Geoff Syme, Editor-in-Chief, with the assistance of Bellie Sivakumar, Associate Editor

Keywords:
Longitudinal dispersion coefficient
Contaminant spill
Evolutionary algorithms
Gene expression programming
Uncertainty analysis
Parametric analysis

SUMMARY

Longitudinal dispersion is the key hydrologic process that governs transport of pollutants in natural streams. It is critical for spill action centers to be able to predict the pollutant travel time and break-through curves accurately following accidental spills in urban streams. This study presents a novel gene expression model for longitudinal dispersion developed using 150 published data sets of geometric and hydraulic parameters in natural streams in the United States, Canada, Europe, and New Zealand. The training and testing of the model were accomplished using randomly-selected 67% (100 data sets) and 33% (50 data sets) of the data sets, respectively. Gene expression programming (GEP) is used to develop empirical relations between the longitudinal dispersion coefficient and various control variables, including the Froude number which reflects the effect of reach slope, aspect ratio, and the bed material roughness on the dispersion coefficient. Two GEP models have been developed, and the prediction uncertainties of the developed GEP models are quantified and compared with those of existing models, showing improved prediction accuracy in favor of GEP models. Finally, a parametric analysis is performed for further verification of the developed GEP models. The main reason for the higher accuracy of the GEP models compared to the existing regression models is that exponents of the key variables (aspect ratio and bed material roughness) are not constants but a function of the Froude number. The proposed relations are both simple and accurate and can be effectively used to predict the longitudinal dispersion coefficients in natural streams.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

Accidental spills are major threats to urban streams. The Ontario Ministry of the Environment's Spill Action Centre (SAC) documented 1030 spills to water courses during the year 2008 (Ministry of Environment, 2008). Contaminant spills in streams may occur for different reasons such as chemical transport accidents which occur while moving chemicals too close to streams, illegal dumping of contaminants, and sudden increases in untreated wastewater discharges (bypass) into a stream (Chin, 2013).

To more accurately simulate the travel time and the break-through curves of spilled contaminants, it is vital to predict longitudinal dispersion coefficients for different reaches of the stream during a range of flow conditions. The longitudinal dispersion coefficient (E) is known to depend on the bed material roughness (friction term), the aspect ratio (width-to-depth ratio), and the Froude number, which reflects the effect of the longitudinal slope (Disley et al., 2015).

Some theoretical and empirical models that have been widely considered, include those by Elder (1959), Fisher (1968, 1975), McQuivey and Keefer (1974), Liu (1977), Fisher et al. (1979), Iwasa and Aya (1991), Seo and Cheong (1998), Kashefipour and Falconer (2002) and Disley et al. (2015). Although these models include the same key input variables, their predictions of longitudinal dispersion vary significantly. Moreover, all of the models have shortcomings with respect to the inadequate representation of natural data and the presence of large prediction errors. Compared with natural stream measurements, all of the models yielded significant errors in prediction of the longitudinal dispersion coefficient. Therefore, a simpler and more reliable approach is required to predict longitudinal dispersion coefficient in urban streams (Disley et al., 2015).

1.1. Previous studies using GEP

Gene expression programming (GEP) has been recently employed by a number of researchers for developing complex

* Corresponding author. Tel.: +1 (519) 824 4120.
E-mail addresses: ahmoudy77@yahoo.com (A.M.A. Sattar), bgharaba@uoguelph.ca (B. Gharabaghi).
[1] On leave for German University in Cairo.

http://dx.doi.org/10.1016/j.jhydrol.2015.03.016
0022-1694/© 2015 Elsevier B.V. All rights reserved.

Nomenclature

a, b, c	constants used in developed GEP models (–)
D	index of agreement **(–)**
E	longitudinal dispersion coefficient (m^2/s)
EAs	evolutionary algorithms
ET	expression tree
E_{sn}	coefficient of efficiency (–)
E^*	dimensionless dispersion coefficient (–)
e_{ij}	Error in prediction (–)
$\bar{e}$	Mean prediction error (–)
F_r	Froude number (–)
f_j	GEP model fitness function (–)
GEP	gene expression programming
GP	genetic programming
g	gravitational acceleration (m/s^2)
H	average depth of channel (m)
h	head of gene (–)
i, j	counter indices
K, K'	gradients of the regression line through the origin (–)
MAD	mean absolute deviation (–)
MCS	Monte Carlo simulation
m, n	coefficient of determination of the regression line through the origin
n	number of cases
O_i	observed values (–)
$\bar{O}$	mean of observed values (–)
P_{ij}	value predicted by model (–)
$\bar{P}$	mean value of model predictions (–)
RRSE	relative squared error (–)
RMSE	root mean square error (–)
R_m	cross validation measures (–)
R^2	coefficient of correlation (–)
R_O^2	squared correlation coefficient through the origin between predicted and observed values (–)
$R_O'^2$	squared correlation coefficient through the origin between observed and predicted values (–)
S_c	marginal sensitivity coefficient (–)
S_e	standard deviation of the prediction errors (–)
S_n	normalized sensitivity coefficient (–)
s	longitudinal slope of the stream reach (m/m)
t_g	tail length of gene (–)
T_j	value observed (–)
$\bar{T}$	mean value of observed cases (–)
U	average stream velocity (m/s)
U^*	shear velocity (m/s)
W	average width of channel (m)
ρ	fluid density (kg/m^3)
μ	dynamic viscosity ($N\ m/s^2$)
σ	channel sinuosity (–)
$\emptyset$	expected solution (–)

relations between experimental data as an efficient alternative to traditional regression and machine learning methods. GEP involves computer programs of different sizes and shapes encoded in linear chromosomes of a fixed length. GEP chromosomes are composed of multiple genes, with each gene encoding a smaller sub-program (Ferreira, 2001). GEP has been recently used successfully to solve hydraulic engineering problems, e.g., scour prediction downstream of hydraulic structures (Guven and Gunal, 2008), dispersion coefficient in natural streams (Azamathulla and Wu, 2011), dispersion coefficient in pipes (Sattar, 2014), prediction of dam breach parameters (Sattar, in press), prediction of transverse mixing (Azamathulla and Ahmad, 2012), and prediction of scour depth downstream of sills (Azamathulla, 2012).

Table 1
Empirical models for longitudinal dispersion coefficient.

Researcher	Empirical equation
Fischer (1975)	$\frac{E}{U^*H} = 0.011\left(\frac{W}{H}\right)^2\left(\frac{U}{U^*}\right)^2$
Liu (1977)	$\frac{E}{U^*H} = 0.18\left(\frac{W}{H}\right)^2\left(\frac{U}{U^*}\right)^{0.5}$
Koussis and Rodríguez-Mirasol (1998)	$\frac{E}{U^*H} = 0.6\left(\frac{W}{H}\right)^2$
Iwasa and Aya (1991)	$\frac{E}{U^*H} = 2\left(\frac{W}{H}\right)^{1.5}$
Seo and Cheong (1998)	$\frac{E}{U^*H} = 5.195\left(\frac{W}{H}\right)^{0.62}\left(\frac{U}{U^*}\right)^{1.428}$
Deng et al. (2001)	$\frac{E}{U^*H} = 0.15\left(\frac{1}{8\left(0.145+\left(\frac{1}{3520}\right)\left(\frac{W}{H}\right)^{1.38}\left(\frac{U}{U^*}\right)\right)}\right)\left(\frac{W}{H}\right)^{1.667}\left(\frac{U}{U^*}\right)^2$
Kashefipour and Falconer (2002)	$\frac{E}{U^*H} = 10.612\left(\frac{U}{U^*}\right)^2$
Rajeev and Dutta (2009)	$\frac{E}{U^*H} = 2\left(\frac{W}{H}\right)^{0.96}\left(\frac{U}{U^*}\right)^{1.25}$
Azamathulla and Wu (2011)	$\frac{E}{U^*H} = e^{e^{cos(U/U^*)+((U/U^*)^2/((W/H)+3.956))}} + \frac{sin((\frac{W}{H})(\frac{U}{U^*}))(\frac{W}{H})(\frac{U}{U^*})}{e^{sin(W/H)}} + \frac{(U/U^*)}{1.037} - \frac{10.76(W/H)}{U/U^*-11.38}$
Etemad-Shahidi and Taghipour (2012)	$\frac{E}{U^*H} = 15.49\left(\frac{W}{H}\right)^{0.78}\left(\frac{U}{U^*}\right)^{0.11}$ $if\ W/H \leqslant 30.6$ $\frac{E}{U^*H} = 14.12\left(\frac{W}{H}\right)^{0.61}\left(\frac{U}{U^*}\right)^{0.85}$ $if\ W/H > 30.6$
Sahay (2013)	$\frac{E}{U^*H} = 2\left(\frac{W}{H}\right)^{0.72}\left(\frac{U}{U^*}\right)^{1.37}S_i^{1.52}$
Disley et al. (2015)	$\frac{E}{U^*H} = 3.563F_r^{-0.4117}\left(\frac{W}{H}\right)^{0.6776}\left(\frac{U}{U^*}\right)^{1.0132}$

Therefore, the main objective of this study is to develop a more accurate prediction model for the longitudinal dispersion coefficient using 150 published data sets of geometric and hydraulic parameters in natural streams in the United States, Canada, Europe, and New Zealand.

1.2. Available prediction models

Since the 1950s many researchers have developed empirical equations for the longitudinal dispersion coefficient (E), including Elder (1959), McQuivey and Keefer (1974), Fischer (1975), Liu (1977), Iwasa and Aya (1991), Koussis and Rodríguez-Mirasol (1998), Li et al. (1998), Seo and Cheong (1998), Deng et al. (2001), Kashefipour and Falconer (2002) and Disley (2010). These empirical models are presented in Table 1.

2. Materials and methods

2.1. Gene expression programming

Evolutionary algorithms (EAs) are a class of problem-solving techniques based on the Darwinian theory of evolution by "natural selection" and involve searching within a population of solutions for the "fittest" solution. A possible and acceptable solution, i.e., a member of the population, is called an individual. Each iteration of an EA includes a competitive selection that weeds out poor solutions through the evaluation of a fitness value that indicates the quality of the individual solution to the problem. Gene expression programming (GEP) was invented by Ferreira in 1999, and is the natural development of EAs.

The great insight of GEP was the invention of chromosomes capable of representing any expression tree; GEP greatly surpasses the genetic programming (GP) system (Ferreira, 2001). In GEP, complex relations are encoded in simpler, linear structures of a fixed length called chromosomes. The chromosomes consist of a linear symbolic string of a fixed length composed of one or more

Download English Version:

https://daneshyari.com/en/article/6411474

Download Persian Version:

https://daneshyari.com/article/6411474

Daneshyari.com