

## Prediction of longitudinal dispersion coefficients in natural rivers using genetic algorithm

Rajeev Ranjan Sahay and Som Dutta

### ABSTRACT

A new expression for the prediction of longitudinal dispersion coefficient in natural rivers, using genetic algorithms, is proposed. The expression uses hydraulic and geometric characteristics of rivers, which are readily available. For performance evaluation, using published field data, results of coefficient prediction by the new expression and by the other reported expressions are compared. According to various performance indices, it is concluded that the new formula predicts the longitudinal dispersion coefficient more accurately. Sensitive analysis performed on input parameters indicates the ratio of the cross-sectional mean velocity to the bottom shear velocity to be the most influencing parameter for accurate prediction of the longitudinal dispersion coefficient.

**Key words** | coefficients, dispersion, genetic algorithms, pollutant flow, river flow, streams

**Rajeev Ranjan Sahay** (corresponding author)  
**Som Dutta**  
 Department of Civil Engineering,  
 Birla Institute of Technology,  
 Mesra, Ranchi 835215,  
 India  
 E-mail: rajeev\_sahay@yahoo.com

### NOMENCLATURE

$A$	cross-sectional area	$t$	time of observation
$AS$	absolute sensitivity	$U$	cross-sectional mean velocity
$a, b, c$	numerical constants	$u'$	difference between depth-averaged velocity $u(y)$ and cross-sectional mean velocity $U$
$C$	tracer/pollutant concentration	$U_*$	shear velocity
$CC$	coefficient of correlation	$W$	channel width
$DR$	discrepancy ratio	$X$	input variable
$g$	acceleration due to gravity	$y$	lateral co-ordinate of the observation point from left bank of the stream
$K$	longitudinal dispersion coefficient	$\varepsilon_t$	transverse and turbulent diffusion coefficient
$K_{\text{meas}}, K_{\text{pred}}$	measured and predicted longitudinal dispersion coefficients respectively	$\beta$	a function of the channel cross-section, shape and velocity distribution
$h$	local depth of flow	$\Delta Y, \Delta X$	deviations in output $Y$ caused by deviation in input $X$
$H$	mean depth of flow		
$N$	number of observation		
$R$	hydraulic radius		
$RE$	relative error		
$RS$	relative sensitivity		
$RMSE$	root mean square error		
$S$	slope of the total energy line in the downstream direction		
$SSQ$	sum of squares		
$S_{\text{meas}}, S_{\text{pred}}$	standard deviations of measured and predicted values		

### INTRODUCTION

When injected into a river, pollutants and effluents undergo dispersion in all three directions. In the first stage, the initial momentum of flow dilutes them vertically. In the second stage, they are mixed throughout the cross-section of the

stream by the advective transport process. In the third stage, when the cross-sectional mixing is complete, the process of longitudinal dispersion becomes the most important mechanism, erasing all longitudinal concentration gradients (Fischer *et al.* 1979). The longitudinal dispersion coefficient is an important factor for prediction of concentration variation of dispersed pollutants in the flow direction of a stream, and for explanation of one-dimensional (1D) pollutant transport phenomena. The concept of the longitudinal dispersion coefficient was first introduced by Taylor (1953, 1954). He derived the following equation for 1D dispersion in a laminar pipe flow:

$$\frac{\partial C}{\partial t} + U \frac{\partial C}{\partial x} = K \frac{\partial^2 C}{\partial x^2} \quad (1)$$

where  $K$  is the longitudinal dispersion coefficient at distance  $x$  from the point of injection of the pollutant,  $C$  is the cross-sectional average concentration,  $U$  is the mean longitudinal velocity and  $t$  is time of observation.

Elder (1959) extended Taylor's method and derived a simple expression for the estimation of the longitudinal dispersion coefficient for very wide channels, in which he assumed a logarithmic velocity profile in the vertical direction

$$K = 5.93 HU_* \quad (2)$$

where bottom shear friction velocity  $U_* = \sqrt{gRS}$ ,  $g$ ,  $R$  and  $S$  are acceleration due to gravity, hydraulic radius and slope of the total energy line in the downstream direction, respectively and  $H$  is the mean cross-sectional depth.

Although Equation (2) was an important breakthrough in the prediction of  $K$  in streams, there were large discrepancies between measured and predicted  $K$  values. Fischer (1967) argued that for producing longitudinal dispersion in a wide and straight natural channel with well-defined cross-section, the transverse profile of velocity is more important than the vertical profile. The following integral expression was developed:

$$K = -\frac{1}{A} \int_0^W hu' \int_0^y \frac{1}{\varepsilon_t h} \int_0^y hu' dy dy dy \quad (3)$$

where  $A$  is cross-sectional area,  $W$  is channel width at the free surface,  $h = h(y)$ , i.e. local depth of flow,  $u'$  is the

difference between local depth-averaged and cross-sectional mean velocities,  $\varepsilon_t$  is transverse, turbulent diffusion coefficient and  $y$  is lateral co-ordinate of the observation point from left bank of the stream.

Use of Equation (3) is not very popular among planners and engineers for two reasons. First, the estimate of  $K$  is not very accurate. The primary reason for the discrepancy is due to the fact that no natural channel completely meets the assumptions inherent in the development of Equation (3). Natural channels have bends, changes in shape, pools and many other irregularities, all of which contribute significantly to the dispersion process. Second, Equation (3) requires elaborate transverse profiles of velocity and cross-sectional geometry that are generally unavailable to the field engineers.

Realizing the difficulties in the use, Fischer *et al.* (1979) presented a simplified and approximate non-integral expression for Equation (3), which has the distinct advantage of estimating  $K$  in non-dimensional format from parameters that are usually known:

$$K = 0.011 \left(\frac{W}{H}\right)^2 \left(\frac{U}{U_*}\right)^2 HU_* \quad (4)$$

However, despite several previously mentioned advantages, Equation (4) does not provide an adequate prediction of  $K$ . In the derivation of Equation (4) (based on laboratory experiments and field tests), approximate values of the triple integration, the deviation between local depth-averaged and cross-sectional mean velocities and transverse turbulent diffusion coefficient were substituted in Equation (3). Here, it is understood that the deviation between local depth-averaged and cross-sectional mean velocities is not properly represented in the expression, which results in shear effects causing the anomaly between observed and predicted  $K$ .

Liu (1977) derived a dispersion expression in natural streams by considering the lateral velocity gradient as an input as well as other variables of Equation (3):

$$\frac{K}{HU_*} = \beta \left(\frac{U}{U_*}\right)^2 \left(\frac{W}{H}\right)^2 \quad (5)$$

Parameter  $\beta$  is a function of the channel cross-section, shape and the velocity distribution across the stream.

Godfrey & Frederick (1970) estimated  $\beta$  as

$$\beta = 0.18 \left( \frac{U}{U_*} \right)^{1.5} \quad (6)$$

Using the one-step Huber method for non-linear multi-regression on published data, Seo & Cheong (1998) gave the following empirical expression for the prediction of longitudinal dispersion coefficient  $K$ , demonstrating its superiority over other reported expressions:

$$\frac{K}{HU_*} = 5.915 \left( \frac{W}{H} \right)^{0.62} \left( \frac{U}{U_*} \right)^{1.428} \quad (7)$$

Deng *et al.* (2001) emphasized the importance of the transverse turbulent mixing ( $\varepsilon_t$ ), in addition to other variables of Equation (3), and derived expressions for  $\varepsilon_t$  and  $K$  as

$$\frac{K}{HU_*} = \frac{0.15}{8\varepsilon_t} \left( \frac{W}{H} \right)^{5/3} \left( \frac{U}{U_*} \right)^2 \quad (8)$$

where

$$\varepsilon_t = 0.145 + \left( \frac{1}{3520.0} \right) \left( \frac{W}{H} \right)^{1.38} \left( \frac{U}{U_*} \right) \quad (8a)$$

Based on dimensional and regression analyses, Kashefipour & Falconer (2002) developed another empirical expression for  $K$ :

$$\frac{K}{HU_*} = 10.612 \left( \frac{U}{U_*} \right)^2 \quad (9)$$

It is therefore apparent from the above that most of the investigators considered  $W$ ,  $H$ ,  $U$  and  $U_*$  as important input variables and have given their version of the expressions for estimation of longitudinal dispersion coefficient, which are essentially of the following dimensionless format

$$\frac{K}{HU_*} = a \left( \frac{W}{H} \right)^b \left( \frac{U}{U_*} \right)^c \quad (10)$$

Parameters  $a$ ,  $b$  and  $c$  vary with the investigator.

## OBJECTIVE OF THE PRESENT STUDY

Using genetic algorithms (GA), a new expression of the non-dimensional form of Equation (10) is proposed for the prediction of longitudinal dispersion coefficient. Data consisting of 65 sets of observations from 29 rivers in the United States, as shown in Table 1, has been used for illustration (Deng *et al.* 2001). The reason for selecting this data is twofold: (i) it represents a wide range of geometric ( $W$ ,  $H$ ) and flow ( $U$ ,  $U_*$ ) characteristics of natural streams; and (ii) it has been studied previously by other investigators, e.g. Seo & Cheong (1998) and Deng *et al.* (2001). Results from the proposed expression can therefore be compared to the previously reported results.

## GA MODEL FORMULATION AND APPLICATION

### Genetic algorithm (GA)

The genetic algorithm (GA), invented by Holland (1975), operates on the principle of natural evolution of genes. The working principle of GA can be found in the books of Goldberg (1989), Michalewicz (1992) or Deb (2002). Rather than starting from one initial guess, GA starts from a population of randomly generated guesses (called chromosomes or strings) which represent variables or components. Strings are made up of genes or substrings. Each string, a candidate solution, has its own fitness value based on an objective fitness value. The entire population of strings forms a generation. A set of genetic operators (selection, crossover and mutation) are employed on chromosomes of this generation to create chromosomes for next generation. This operation is repeated until a stopping criterion is reached. The criterion may be either the number of generations or the change in the fitness value of chromosomes between two consecutive generations. In general, the fitness values of later generations should improve, although we cannot expect the best solution to be found in the final generation. The structure of GA model is given in Figure 1.

### Problem formulation and application

Rearranging expression (10),  $K$  can be determined as

$$K = aW^b H^{1-b} U^c U_*^{1-c} \quad (10a)$$

**Table 1** | Measured and predicted longitudinal dispersion coefficients

S.N.	Stream	Longitudinal dispersion coefficient, $K$ ( $m^2/s$ )												
		Width, $W$ (m)	Depth, $H$ (m)	Velocity $U$ (m/s)	Sh. Vel., $U_s$ (m/s)	$W/H$	$U/U_s$	Measured value	Predicted by Equation (12)	Predicted by Equation (4)	Predicted by Equation (5)	Predicted by Equation (7)	Predicted by Equation (9)	Predicted by Equation (8)
1	Amita River	37.0	0.81	0.29	0.070	45.68	4.14	23.2	26.3	22.3	43.3	27.3	10.3	28.5
2		42.0	0.80	0.42	0.069	52.50	6.09	30.2	47.3	62.0	67.6	50.2	21.7	51.0
3	Antietam Creek, Md.	12.8	0.30	0.42	0.057	42.67	7.37	17.5	15.2	18.6	15.2	18.0	9.9	17.6
4		24.1	0.98	0.59	0.098	24.59	6.02	101.5	39.2	23.2	25.7	53.7	36.9	47.3
5		11.9	0.66	0.43	0.085	18.03	5.06	20.9	13.7	5.1	7.4	20.2	15.2	15.0
6		21.0	0.48	0.62	0.069	43.75	8.99	25.9	38.8	56.3	34.2	46.9	28.4	44.4
7	Bayou Anacoco	20.0	0.42	0.29	0.045	47.62	6.44	13.9	15.8	19.6	19.6	17.6	8.3	17.6
8		17.5	0.45	0.32	0.024	38.89	13.33	5.8	18.5	31.9	10.7	25.0	20.4	21.8
9		25.9	0.94	0.34	0.067	27.55	5.07	32.5	23.2	13.5	19.4	29.6	17.2	26.8
10		36.6	0.91	0.40	0.067	40.22	5.97	39.5	39.5	38.7	43.4	45.7	23.1	45.6
11	Bayou Barthol., La	33.4	1.40	0.20	0.031	23.86	6.45	54.7	18.8	11.3	11.3	26.3	19.2	23.0
12	Bear Creek, Colo.	13.7	0.85	1.29	0.553	16.12	2.33	2.9	39.1	7.3	33.6	52.3	27.1	28.1
13	Chattahoochee, Ga	75.6	1.95	0.74	0.138	38.77	5.36	88.9	147.1	127.9	168.6	169.3	82.1	168.8
14		91.9	2.44	0.52	0.094	37.66	5.53	166.9	126.8	109.5	137.7	148.2	74.5	146.7
15	Clinch River, Va	48.5	1.16	0.21	0.069	41.81	3.04	14.8	23.2	14.3	43.9	23.5	7.9	23.8
16		28.7	0.61	0.35	0.069	47.05	5.07	10.7	25.8	26.4	37.8	27.6	11.5	28.5
17		57.9	2.45	0.75	0.104	23.63	7.21	40.5	125.4	81.4	68.8	180.0	140.6	158.2
18		53.2	2.41	0.66	0.107	22.07	6.17	36.9	97.8	52.6	56.2	139.7	104.1	118.3
19	Comite River	13.0	0.26	0.31	0.044	50.00	7.05	7.0	11.2	15.6	13.7	12.4	6.0	12.3
20		16.0	0.43	0.37	0.056	37.21	6.61	13.9	16.4	16.0	15.4	19.9	11.2	19.4
21		15.7	0.23	0.36	0.039	68.26	9.23	69.0	16.6	39.2	22.9	17.4	8.1	15.8
22	Conoco. Creek, Md.	42.2	0.69	0.23	0.064	61.16	3.59	40.8	22.7	23.5	56.4	20.8	6.1	22.9
23		49.7	0.41	0.15	0.081	121.22	1.85	29.3	14.4	18.4	119.5	9.3	1.2	11.8
24		43.0	1.13	0.63	0.081	38.05	7.78	53.3	78.2	88.2	66.5	96.8	58.8	93.2
25	Copper Creep, Va.	16.7	0.49	0.20	0.080	34.08	2.50	16.8	7.3	3.1	13.0	7.7	2.6	6.9
26		18.3	0.38	0.15	0.116	48.16	1.29	20.7	5.0	1.9	20.9	4.2	0.8	4.0
27		16.8	0.47	0.24	0.080	35.74	3.00	24.6	9.2	4.8	15.0	9.8	3.6	9.3
28		19.6	0.84	0.49	0.101	23.33	4.85	20.8	25.1	12.0	18.3	33.8	21.2	28.4
29	Difficult Run, Va.	14.5	0.31	0.25	0.062	46.77	4.03	1.9	8.8	7.5	15.2	9.0	3.3	9.5
30	John Day River, Ore.	25.0	0.58	1.01	0.140	43.10	7.21	13.9	71.2	86.4	72.9	83.3	44.8	81.8
31		34.1	2.47	0.82	0.180	13.81	4.56	65.0	73.6	19.3	32.6	116.8	97.9	71.1
32	Little Pincy Creek	15.9	0.22	0.39	0.053	72.27	7.36	7.1	17.2	36.3	29.7	17.0	6.7	16.3
33	Minnesota River	80.0	2.74	0.03	0.002	29.20	14.17	22.3	9.2	12.4	3.8	13.9	14.0	12.1
34		80.0	2.74	0.14	0.010	29.20	14.43	34.9	38.1	51.9	15.5	57.7	58.8	49.9

Table 1 | (continued)

S.N.	Stream	Width, W (m)	Depth, H (m)	Velocity U (m/s)	Sh. Vel., U <sub>s</sub> (m/s)	W/H	U/U <sub>s</sub>	Longitudinal dispersion coefficient, K (m <sup>2</sup> /s)						
								Measured value	Predicted by Equation (12)	Predicted by Equation (4)	Predicted by Equation (5)	Predicted by Equation (7)	Predicted by Equation (9)	Predicted by Equation (8)
35	Missouri River	183.0	2.33	0.89	0.066	78.54	13.48	465.0	524.2	1897.5	627.0	559.2	296.7	438.6
36		201.0	3.56	1.28	0.084	56.46	15.24	837.0	865.2	2434.9	669.8	1055.3	736.9	848.4
37	Monocacy River, Md.	48.7	0.55	0.26	0.052	88.55	5.00	37.8	31.7	61.7	90.3	27.2	7.6	28.2
38		93.0	0.71	0.16	0.046	130.99	3.48	41.4	33.4	74.6	188.1	23.6	4.2	25.8
39		51.2	0.65	0.62	0.044	78.77	14.09	29.6	103.3	387.6	119.9	110.9	60.3	85.6
40		97.5	1.15	0.32	0.058	84.78	5.52	119.8	80.1	160.5	202.7	71.0	21.5	72.3
41		40.5	0.41	0.23	0.040	98.78	5.75	66.5	24.0	58.2	69.1	20.4	5.8	20.1
42	Muddy River	13.0	0.81	0.37	0.081	16.05	4.57	13.9	12.6	3.9	6.5	19.0	14.5	12.8
43		20.0	1.20	0.45	0.099	16.67	4.55	32.5	23.5	7.5	12.7	35.0	26.0	24.1
44	Nooksack River	86.0	2.93	1.20	0.530	29.35	2.26	153.0	221.2	75.4	362.4	240.0	84.5	195.8
45		64.0	0.76	0.67	0.268	84.21	2.50	34.8	90.3	99.3	411.1	69.7	13.5	82.7
46	Powell River, Tenn	36.8	0.87	0.13	0.054	42.30	2.41	15.5	10.3	5.4	23.5	9.9	2.9	9.9
47	Red River, La	161.5	3.96	0.29	0.060	40.78	4.83	130.5	119.7	101.6	156.4	133.0	58.9	134.6
48		152.4	3.66	0.45	0.057	41.64	7.89	227.6	198.1	248.0	182.9	238.3	138.0	230.3
49		155.1	1.74	0.47	0.036	89.14	13.06	177.7	231.6	933.2	323.7	235.3	113.3	181.4
50	Sabina River, La.	116.4	1.65	0.58	0.054	70.55	10.74	131.3	206.2	562.7	261.6	219.1	109.1	189.0
51	Sabina River, Tex.	14.2	0.50	0.13	0.037	28.40	3.51	12.8	4.4	2.0	5.0	5.2	2.4	4.6
52		12.2	0.51	0.23	0.030	23.92	7.67	14.7	8.2	5.7	4.4	11.9	9.5	10.5
53		21.3	0.93	0.36	0.035	22.90	10.29	24.2	24.2	19.9	9.9	37.5	36.5	32.7
54	Salt Creek, Nebr.	32.0	0.50	0.24	0.038	64.00	6.32	52.2	20.6	34.1	35.2	20.6	8.0	20.7
55	Susquehanna River	203.0	1.35	0.39	0.065	150.37	6.00	92.9	202.8	785.7	874.8	150.2	33.5	135.1
56	Tangipahoa River, La	31.4	0.81	0.48	0.072	38.77	6.67	45.1	41.8	42.8	40.7	50.1	27.5	49.1
57		29.9	0.40	0.34	0.020	74.75	17.00	44.0	34.7	142.1	33.2	39.3	24.5	28.7
58	Tickfau River, La	15.0	0.59	0.27	0.080	25.42	3.38	10.3	9.6	3.8	10.1	11.8	5.7	9.7
59	White River	67.0	0.59	0.35	0.044	113.56	7.95	30.2	65.2	233.0	170.0	55.8	17.4	48.5
60	Wind/Big, Wyo	44.2	1.37	0.99	0.142	32.26	6.97	184.6	123.8	108.3	96.2	158.9	100.3	151.0
61		85.3	2.38	1.74	0.153	35.84	11.37	464.6	472.4	665.5	283.9	638.4	499.8	577.8
62		59.4	1.10	0.88	0.119	54.00	7.39	41.8	147.0	229.6	186.8	160.0	76.0	156.7
63		68.6	2.16	1.55	0.168	31.76	9.23	162.6	322.8	342.7	200.1	437.8	327.8	406.1
64	Yadkin River, N.C.	70.1	2.35	0.43	0.101	29.83	4.26	111.5	75.6	42.1	78.4	91.3	45.7	83.9
65		71.6	3.84	0.76	0.128	18.65	5.94	260.1	151.1	66.3	75.0	227.2	183.9	177.2
	Max. Value	203.0	3.96	1.74	0.550	150.37	17.00	837.0	865.2	2434.9	874.8	1055.3	736.9	848.4
	Min. Value	11.9	0.22	0.03	0.002	13.81	1.29	1.9	4.4	1.9	3.8	4.2	0.8	4.0
	Avg. Value	53.8	1.24	0.49	0.090	48.86	6.75	80.5	89.0	169.1	110.6	104.5	63.0	91.5

The criterion for the prediction of unknown parameters  $a$ ,  $b$  and  $c$  within a GA model framework is the minimization of the residual sum of squares (SSQ) between measured and predicted dispersion coefficients, i.e.

$$\text{Min SSQ} = \sum_{i=1}^N (K_{\text{pred}} - K_{\text{meas}})^2 \quad (11)$$

or

$$\text{Min SSQ} = \sum_{i=1}^N (aW^b H^{1-b} U^c U_*^{1-c} - K_{\text{meas}})^2 \quad (11a)$$

where  $K_{\text{meas}}$ ,  $K_{\text{pred}}$  are the measured and predicted longitudinal dispersion coefficients, respectively, and  $N$  is number of observations.

The GA model is implemented by adopting a binary code for representation of variables. To estimate the three design coefficients  $a$ ,  $b$  and  $c$  in the objective function, as given in Equation (11a), a binary string length of 10 for each variable is selected. This string length is sufficient for the range of values which these variables can attain. There is therefore a total string length of 30 bits. Many combinations of population size, crossover probability and mutation are tried. A population size of 200, crossover probability of 0.9 and mutation probability of 0.002 are found to give the minimum value of SSQ.

The best values of  $a$ ,  $b$  and  $c$  found from the GA run are:  $a = 2.0$ ,  $b = 0.96$  and  $c = 1.25$ . Thus, based on GA, the new expression for longitudinal dispersion coefficient is

$$\frac{K}{HU_*} = 2 \left( \frac{W}{H} \right)^{0.96} \left( \frac{U}{U_*} \right)^{1.25} \quad (12)$$

## MODEL VALIDATION

In order to evaluate the performance of the new expression for longitudinal dispersion coefficient against other reported expressions, the 65 sets of data (Deng *et al.* 2001) were utilized. The other expressions considered here for comparison are those by Fischer (1975), Liu (1977), Seo & Cheong (1998), Deng *et al.* (2001) and Kashefipour & Falconer (2002). Performance indices considered here for the comparison of results from

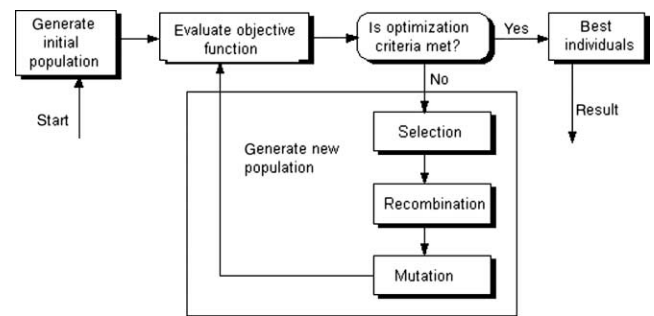


Figure 1 | Flow-chart of GA model.

different expressions are root mean square errors (RMSE), coefficients of correlation (CC) and discrepancy ratios (DR). RMSE, CC and DR (Kashefipour & Falconer 2002) are defined as follows:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^N (K_{\text{pred}} - K_{\text{meas}})^2}{N}} \quad (13)$$

$$\text{CC} = \frac{\left[ \sum_1^N K_{\text{pred}} K_{\text{meas}} - \sum_1^N K_{\text{pred}} \sum_1^N K_{\text{meas}} \right]}{NS_{\text{pred}} S_{\text{meas}}} \quad (14)$$

$$\text{DR} = \log \frac{K_{\text{pred}}}{K_{\text{meas}}} \quad (15)$$

where  $S_{\text{meas}}$ ,  $S_{\text{pred}}$  represent standard deviations of measured and predicted values.

From Equation (15), it follows that  $\text{DR} = 0.0$  suggests exact matching between measured and predicted values. Otherwise, there is either an overprediction ( $\text{DR} > 0$  i.e.  $K_{\text{pred}} > K_{\text{meas}}$ ) or underprediction ( $\text{DR} < 0.0$  i.e.  $K_{\text{pred}} < K_{\text{meas}}$ ).

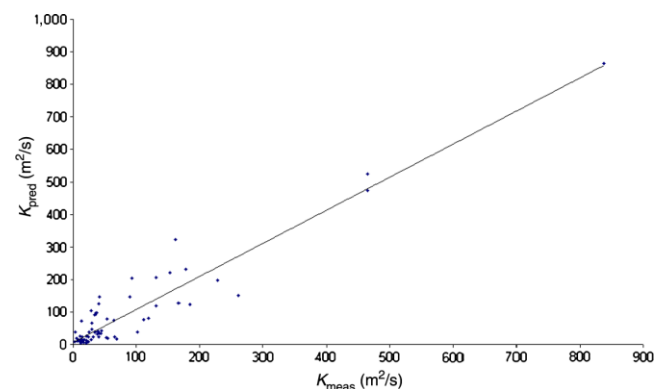


Figure 2 | Comparison of predicted dispersion coefficient from the new expression and the measured coefficient.

**Table 2** | Comparison of performance indices of models

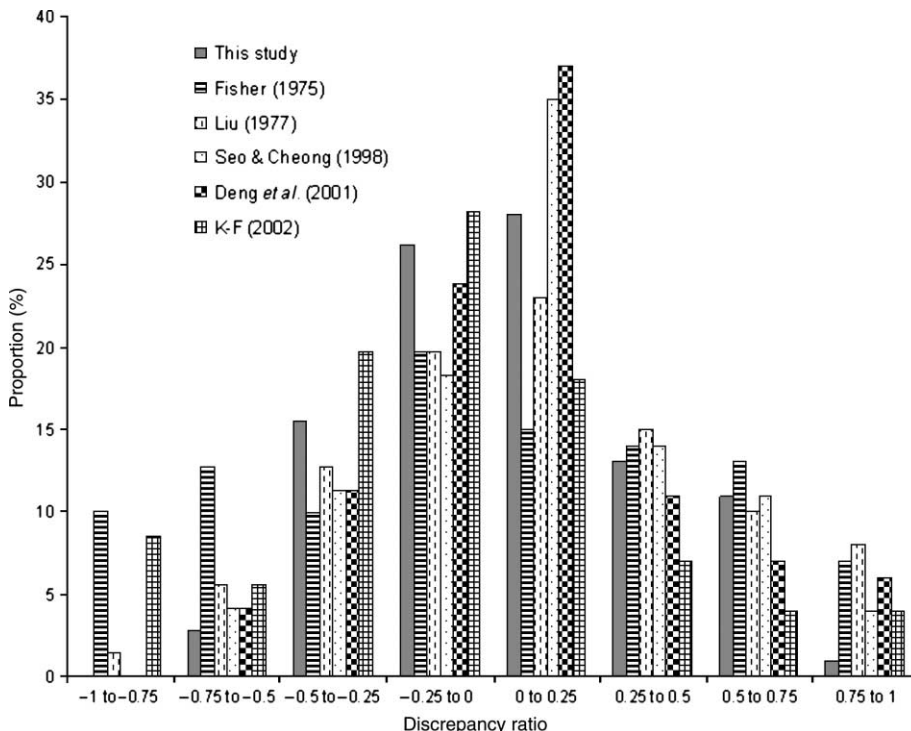
Model	RMSE	RMSE ( $K > 100 \text{ m}^2 \text{ s}^{-1}$ ignored)	CC	DR range	Accuracy (%)
GA	45	36	0.95	-0.60 to 1.08	84
Fischer (1975)	309	120	0.86	-0.80 to 1.56	29
Liu (1977)	129	113	0.66	-0.77 to 1.17	72
Seo & Cheong (1998)	65	43	0.94	-0.70 to 1.26	78
Deng <i>et al.</i> (2001)	51	37	0.94	-0.72 to 0.99	81
Kashefipour & Falconer (2002)	49	31	0.93	-1.42 to 0.97	55

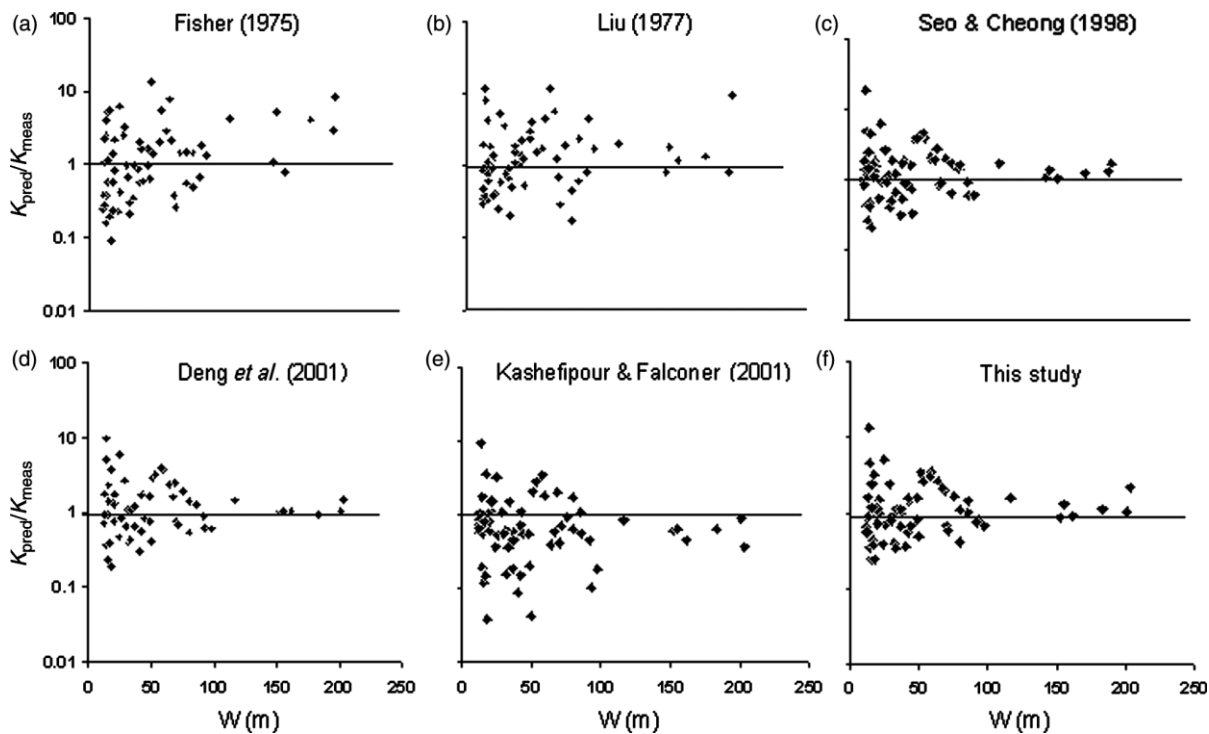
The predicted values of longitudinal dispersion coefficient from the new expression and other expressions, along with measured values, are given in Table 1. Figure 2 shows a comparison between the predicted coefficients from the GA model and the measured values. Table 2 shows the performance indices i.e. RMSE, CC and DR of the different expressions. As can be seen from this table, the new expression is superior to the other expressions as RMSE is the lowest and CC the highest.

The expressions given by Seo and Cheong, Deng *et al.* and Kashefipour and Falconer also perform well. The Fischer expression is found to be the most unsatisfactory. When the

large  $K$  values (i.e.  $K > 100 \text{ m}^2 \text{ s}^{-1}$ ) are not included in the analysis, the performances of all models improved. In this case, the GA model appears to be superior to all expressions except the Kashefipour and Falconer expression, which has a slightly lower RMSE. Table 2 shows that the DR values of the GA model range from -0.6 to 1.08, which suggests slight skewness towards the positive side. However, the positive skewnesses of Fischer and Seo and Cheong expressions are more pronounced whereas the Kashefipour and Falconer model is skewed towards the negative side.

A comparison of the percentage proportion of the predicted values falling into a different discrepancy bracket

**Figure 3** | Comparison of discrepancy ratio.



**Figure 4** | Ratio of predicted and measured dispersion coefficient versus stream width.

is seen in Figure 3. It shows an almost even distribution for the values from the new expression. If the accuracy of an expression can be defined as the percentage of discrepancy ratio values falling between  $-0.5$  and  $0.5$ , then it can be seen from Table 2 that 84% of the predicted values from the new expression are accurate, a percentage which is the highest among all expressions.

Figure 4 is a plot showing the variation of dispersion coefficient with the width of the river, demonstrating that GA prediction improves as the width increases.

### Sensitivity analysis

A sensitivity and error analysis of the new expression, for the prediction of longitudinal dispersion coefficient in the

**Table 3** | Sensitivity and error analysis of the new expression ( $X$ : input parameter;  $\Delta X$ : 10% increment in  $X$ ;  $\Delta Y$ : % change in output ( $K/HU_*$ ); AS: absolute sensitivity; RS: relative sensitivity; RE: relative error)

$X$	$\Delta X$	$\Delta Y$	AS	RS	RE
$W/H$	4.886	87.26	17.86	0.96	0.096
$U/U_*$	0.675	115.14	170.58	9.16	0.126

dimensionless format, is carried out for the mean values of input and output parameters. The analysis is based on the assumption that errors are independent of changes in inputs. The average values of dimensionless input parameters  $W/H$  and  $U/U_*$ , for the data given in Table 1, are 48.86 and 6.75 respectively.

Absolute sensitivity (AS), relative sensitivity (RS) and relative error (RE) are defined as (ASCE 1996):

$$AS = \frac{\partial Y}{\partial X} \approx \frac{\Delta Y}{\Delta X}; \quad RS = \frac{\Delta Y X}{\Delta X Y}; \quad RE = \frac{\Delta Y}{Y} \quad (16)$$

where output  $Y = (K/HU_*)$  and  $\Delta Y$  is the deviation in  $Y$  caused by the deviation  $\Delta X$  in the input  $X$ . Investigation of the sensitivity and of the error is carried out by incrementing each input parameter by 10%. The result of the investigation is shown in Table 3. It can be seen from Table 3 that the greatest variation in the output  $K/HU_*$  is caused by the variation in the input  $U/U_*$ , which is almost ten times that caused by the variation in the input  $W/H$ .



## CONCLUSION

The genetic algorithm approach was used to derive a new expression for prediction of the longitudinal dispersion coefficient in natural rivers. The expression makes use of few geometric (river width, flow depth) and hydraulic parameters (cross-sectional average and shear velocities), commonly known to engineers and planners. A performance evaluation of the new expression is carried out by comparing the predictions from the new formula with other reported expressions, using published data from 29 rivers in US. The comparison study shows that the new expression has the least root mean square error, the highest coefficient of correlation and the best discrepancy ratio. More than 53% of the predictions using the new expression lie within the range  $0.5 < K_{\text{pred}}/K_{\text{meas}} < 1.5$ . The expression is found to be especially suited to wide rivers, where predictions are very close to the measured dispersion coefficients. Sensitivity analysis, conducted on the new expression, suggests that the ratio of the cross-sectional mean velocity to the shear velocity is the most influencing parameter for the accurate prediction of the longitudinal dispersion coefficient.

## ACKNOWLEDGEMENTS

The authors gratefully acknowledge the suggestions of the two anonymous reviewers for improving the language and the content of the article.

## REFERENCES

- ASCE Task Committee 1996 *Hydrology Handbook of Management Group D*, 2nd edition. ASCE, New York, pp. 752–765.
- Deb, K. 2002 *Multi-Objective Optimization Using Evolutionary Algorithms*. John Wiley and Sons, Asia.
- Deng, Z. Q., Singh, V. P. & Bengtsson, L. 2001 **Longitudinal dispersion coefficient in straight rivers**. *J. Hydraulic Eng.* **127**(11), 919–927.
- Elder, J. W. 1959 **The dispersion of a marked fluid in turbulent shear flow**. *J. Fluid Mech.* **5**(4), 544–560.
- Fischer, B. H. 1967 **The mechanics of dispersion in natural streams**. *J. Hydr. Div. ASCE* **93**(6), 187–216.
- Fischer, B. H. 1975 **Discussion of ‘Simple method for predicting dispersion in stream’ by R.S. McQuivey and T.N. Keefer**. *J. Environ. Eng. Div.* **101**(3), 453–455.
- Fischer, B. H., List, E. J., Koh, R. C. Y., Imberger, J. & Brooks, N. H. 1979 *Mixing in Inland and Coastal Waters*. Academic, New York, pp. 104–138.
- Godfrey, R. G. & Frederick, B. J. 1970 *Stream dispersion at selected sites*. US Geological Survey. Paper 433-K, Washington, DC.
- Goldberg, D. E. 1989 *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison-Wesley, New York.
- Holland, J. H. 1975 *Adaptation in Natural and Artificial Systems*. University of Michigan Press, Ann Arbor, Mich.
- Kashefipour, S. M. & Falconer, R. A. 2002 **Longitudinal dispersion coefficients in natural channels**. *Water Res.* **36**(6), 1596–1608.
- Liu, H. 1977 **Predicting dispersion coefficient of streams**. *J. Environ. Eng. Div.* **103**(1), 59–69.
- Michalewicz, Z. 1992 *Genetic Algorithm + Data Structures = Evolutionary Programs*. Springer, New York.
- Seo, I. W. & Cheong, T. S. 1998 **Predicting longitudinal dispersion coefficient in natural streams**. *J. Hydraulic Eng.* **124**(1), 25–32.
- Taylor, G. I. 1953 **Dispersion of soluble matter in solvent flowing slowly through a tube**. *Proc. R. Soc. London Ser. A* **219**, 186–203.
- Taylor, G. I. 1954 **The dispersion of matter in turbulent flow through a pipe**. *Proc. R. Soc. London Ser. A* **223**, 446–468.

First received 2 February 2009; accepted in revised form 11 June 2009. Available online November 2009