

# Predicting longitudinal dispersion coefficients in sinuous rivers by genetic algorithm

Rajeev Ranjan Sahay

Department of Civil Engineering, Birla Institute of Technology, Mesra, Ranchi-835215, India.  
Tel.: +91 9431382737. Fax: +91 651 2275401. E-mail: rajeev\_sahay@yahoo.com

**Abstract:** In the present work, existing empirical expressions for longitudinal dispersion coefficient of rivers ( $K$ ) are evaluated. They are found inadequate primarily because these expressions ignore the channel sinuosity, an important parameter representing a river's transverse irregularities that affect mixing process. Hence, a new expression for  $K$  is derived taking into account the sinuosity besides few of other hydraulic and geometric characteristics of a river. The model makes use of genetic algorithm (GA) on published field data. Based on several performance indices, the new expression is found superior to many existing expressions for estimating  $K$ . The sensitivity and error analysis conducted on parameters of the new expression show the channel sinuosity an important input for predicting  $K$  accurately. Any error in measurement of sinuosity would lead to significant deviation in the longitudinal dispersion coefficient in sinuous rivers.

**Keywords:** Dispersion coefficients; Dispersion; Genetic algorithm; Pollutant transport; Sinuous rivers.

## INTRODUCTION

When a conservative and non-buoyant pollutant is released into a river, physical processes such as advective transport and dispersion determine the movement and change in pollutant concentration. The pollutant transport process can be viewed as being composed of three stages. In the first stage, a pollutant is diluted by the flow in the channel because of its initial momentum. In the second stage, the pollutant is mixed throughout the cross-section of the river by turbulent transport processes. In the third stage, after the cross-sectional and vertical mixing is complete, longitudinal dispersion tends to erase any longitudinal variation in the pollutant concentration (French, 1986). The knowledge of reliable value of longitudinal dispersion coefficient is important for devising water diversion strategies, designing treatment plants, intakes and outfalls, and studying environmental impact due to injection of polluting effluents into streams (David et al., 2002).

A large disparity exists between dispersion coefficient values obtained for idealized and simplified systems (such as irrigation canals) and for natural rivers (Rutherford, 1994). Such a disparity suggests that the processes contributing to dispersion in streams are not well understood. Very few studies have recognized the role of the channel sinuosity that is defined as the ratio of the channel length to the valley length, in riverine dispersion. The channel sinuosity develops secondary flow which is transverse circulation induced by centrifugal force. A 3D hydrodynamic model is required to simulate dispersion accurately in meandering channels, however, because of rapid vertical and transverse mixing relative to longitudinal mixing, most rivers can be approximated as 1D systems. Taylor (1953, 1954), while deriving the following equation for 1D dispersion in a laminar pipe flow, introduced the concept of the longitudinal dispersion coefficient

$$A \frac{\partial C}{\partial t} + UA \frac{\partial C}{\partial x} = \frac{\partial}{\partial x} \left( KA \frac{\partial C}{\partial x} \right), \quad (1)$$

where  $C$  is cross-sectionally averaged concentration of pollutants ( $\text{g L}^{-1}$ ),  $A$  is cross-sectional area of the stream ( $\text{m}^2$ ),

$U$  is mean longitudinal flow velocity ( $\text{m s}^{-1}$ ),  $K$  is longitudinal dispersion coefficient at a distance of  $x$  from the point of injection of the pollutant ( $\text{m}^2 \text{s}^{-1}$ ) and  $t$  is time of observation (s). Eq. (1) is applicable only after the Fickian period during which a contaminant is uniformly mixed across the stream. The solution of Eq. (1), for instantaneous injection of mass  $M_0$  can be obtained for stable rivers with constant longitudinal dispersion coefficient as

$$C = \frac{M_0}{A\sqrt{4\pi Kt}} \exp \left[ \frac{-(x-Ut)^2}{4Kt} \right] \quad (\text{Sayre, 1968}) \quad (2)$$

The estimate of  $K$  by Eq. (2) requires concentration samples taken from a particular reach of the stream and the models' calibration and verification, such tracer investigation is expensive and rarely done for every stream. For that reason, many investigators have derived empirical expressions for  $K$  (Elder, 1959; Fischer et. al., 1979; Lisa and Aya, 1991; Thackston and Krenkel, 1967; Toprak and Cigizoglu, 2008).

## PREVIOUS STUDIES

Fisher et al. (1979) derived the following simple and approximate expression for longitudinal dispersion coefficient of streams

$$K = 0.011 \frac{U^2 W^2}{H U_*}, \quad (3)$$

where  $H$  is cross-sectional average flow depth (m),  $U_*$  is shear velocity ( $\text{m s}^{-1}$ ) and  $W$  is stream width (m).

Initially, Eq. (3) was widely used for its simplicity and theoretical background but the large deviation between the predicted values by it and the measured values has made the equation unpopular now. The deviation may be because no stream completely fulfills the assumptions inherent in the development of this equation.

Liu (1977), after analyzing several dispersion data of USA Rivers, emphasized the role of lateral velocity gradients in dispersion and derived an expression for  $K$  as

$$K = \mu \frac{U^2 W^2}{HU_*}, \quad (4)$$

where parameter  $\mu$  is a function of channel geometry and cross-sectional velocity distribution. Godfrey and Frederick (1970) obtained an estimate of  $\mu$  as

$$\mu = 0.18 \left( \frac{U_*}{U} \right)^{1.5}. \quad (5)$$

Applying the one-step Huber method of nonlinear multi-regression, Seo and Cheong (1998) derived the following empirical expression for predicting longitudinal dispersion coefficient and showed its superiority over other reported expressions of the time

$$\frac{K}{HU_*} = 5.915 \left( \frac{W}{H} \right)^{0.62} \left( \frac{U}{U_*} \right)^{1.428}. \quad (6)$$

Deng et al. (2001) emphasized the importance of local mixing coefficient ( $\epsilon_t$ ) in riverine pollutant transport in addition to the other influencing parameters of Eq. (3) and derived an expression for  $K$  in stable straight rivers as

$$\frac{K}{HU_*} = \frac{0.15}{8\epsilon_t} \left( \frac{W}{H} \right)^{5/3} \left( \frac{U}{U_*} \right)^2, \quad (7)$$

where local transverse mixing coefficient

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

Kashefipour and Falconer (2002), using dimensional and regression analyses on published river data, developed the following expression for  $K$  in natural streams

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

which can be re-written as

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

Applying genetic algorithm optimization technique (GA) on published field data, Sahay and Dutta (2009) developed an expression for  $K$  as

$$\frac{K}{HU_*} = 2 \left( \frac{W}{H} \right)^{0.96} \left( \frac{U}{U_*} \right)^{1.25}. \quad (10)$$

## MODEL DEVELOPMENT

The present study shows the above-mentioned expressions (Eqs. (3) to (10)) inadequate for forecasting  $K$  in sinuous rivers as these rivers are morphologically different from straight rivers. Unlike the case of a constant cross-sectional shape, the longitudinal and transverse momentum and solute transport rates are continually adjusting, never achieving equilibrium in a meandering channel. Studies show the magnitude of dispersion coefficient more in sinuous rivers than in straight rivers (Fukuoka and Sayre, 1973), which increases with increasing radius of curvature and decreases with increasing bend length and depth (Guymer, 1998). Tayfur and Singh (2005) and Sahay (2011) showed improvement in predictive accuracy of models for  $K$  when channel sinuosity is considered as an input, however, they did not give any expression for  $K$ . Hence, a new expression for  $K$  in the following non-dimensional format is tried in the present study

$$\frac{K}{HU_*} = \mu \left( \frac{W}{H} \right)^\beta \left( \frac{U}{U_*} \right)^\gamma S_i^\delta, \quad (11)$$

where  $S_i$  is the channel sinuosity, which can be determined from topographical sheets. Sinuosity is unity for straight rivers. The unknown coefficients  $\alpha$ ,  $\beta$ ,  $\gamma$  and  $\delta$  are determined employing genetic algorithm on published field data.

For deriving and verifying the new expression, data consisting of 61 sets of observation from 24 rivers in USA (Table 1), Tayfur and Singh, (2005) is utilized. Reasons for selecting this data are two: (1) it represent a wide range of geometric and flow characteristics of streams, and (2) it was used previously by other investigators (Deng et al., 2001; Sahay, 2011; Seo and Cheong, 1998 and Tayfur and Singh, 2005). Thus, results from the proposed and other expressions can be compared well. However, some of the datasets from Tayfur and Singh (2005) with missing values of channel sinuosity are excluded from this study. The Mississippi River is also ignored as its width is much larger than the average width of the other considered rivers. If included, performance of all models are affected. The empirical expression for parameters with such high values should be derived separately. Table 2 shows the average and range values of the hydro-meteorological characteristics of the rivers used in this study. Comparable datasets are chosen for deriving and verifying the new expression to avoid any bias.

Of the 61 published datasets of Table 1, 45 datasets are selected to derive the expression and the remaining 16 datasets are used to verify it. For performance evaluation, predictions by the new expression and Fischer et al. (1979), Liu (1977), Seo and Cheong (1998), Deng et al. (2001), Kashefipour and Falconer (2002) and Sahay and Dutta (2009) are compared. For brevity, these comparison models are denoted as Fischer, Liu, S-C, D-S-B, K-F and S-D respectively. The statistical indices used for comparison are the coefficient of correlation (CC), the root mean square error (RMSE), the discrepancy ratio (DR) and the % accuracy, which are defined as

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^N (K_p - K_m)^2}{N}}, \quad (12)$$

**Table 1.** Measured longitudinal dispersion coefficients at 61 sites of USA Rivers (Tayfur and Singh, 2005).

S.N.	Stream	Width, $W$ (m)	Depth, $H$ (m)	Velocity, $U$ ( $\text{m s}^{-1}$ )	Sh. Vel., $U_*$ ( $\text{m s}^{-1}$ )	Sinu., $S_i$	$W/H$	$U/U_*$	Meas. $K$ ( $\text{m}^2 \text{s}^{-1}$ )
1	Antietam Creek, Md.	12.8	0.30	0.42	0.057	1.40	42.67	7.37	17.5
2	Antietam Creek, Md.	24.1	0.98	0.59	0.098	2.25	24.59	6.02	101.5
3	Antietam Creek, Md.	11.9	0.66	0.43	0.085	2.25	18.03	5.06	20.9
4	Antietam Creek, Md.	21.0	0.48	0.62	0.069	1.26	43.75	8.99	25.9
5	Bayou Anacoco	20.0	0.42	0.29	0.045	1.41	47.62	6.44	13.9
6	Bayou Anacoco	17.5	0.45	0.32	0.024	1.41	38.89	13.33	5.8
7	Bayou Anacoco	25.9	0.94	0.34	0.067	1.41	27.55	5.07	32.5
8	Bayou Anacoco	36.6	0.91	0.40	0.067	1.41	40.22	5.97	39.5
9	Bayou Bartholomew, La	33.4	1.40	0.20	0.031	2.46	23.86	6.45	54.7
10	Bear Creek, Colo.	13.7	0.85	1.29	0.553	1.08	16.12	2.33	2.9
11	Chattahoochee River, Ga	75.6	1.95	0.74	0.138	1.27	38.77	5.36	88.9
12	Chattahoochee River, Ga	91.9	2.44	0.52	0.094	1.57	37.66	5.53	166.9
13	Clinch River, Va	48.5	1.16	0.21	0.069	1.25	41.81	3.04	14.8
14	Clinch River, Va	28.7	0.61	0.35	0.069	1.14	47.05	5.07	10.7
15	Clinch River, Va	57.9	2.45	0.75	0.104	1.14	23.63	7.21	40.5
16	Clinch River, Va	53.2	2.41	0.66	0.107	1.14	22.07	6.17	36.9
17	Comite River	13.0	0.26	0.31	0.044	1.31	50.00	7.05	7.0
18	Comite River	16.0	0.43	0.37	0.056	1.31	37.21	6.61	13.9
19	Comite River, La	15.7	0.23	0.36	0.039	1.31	68.26	9.23	69.0
20	Conococh. Creek, Md.	42.2	0.69	0.23	0.064	2.25	61.16	3.59	40.8
21	Conococh. Creek, Md.	49.7	0.41	0.15	0.081	2.25	121.22	1.85	29.3
22	Conococh. Creek, Md.	43.0	1.13	0.63	0.081	1.31	38.05	7.78	53.3
23	Copper Creep, Va.	16.7	0.49	0.20	0.080	2.54	34.08	2.50	16.8
24	Copper Creep, Va.	18.3	0.38	0.15	0.116	2.54	48.16	1.29	20.7
25	Copper Creep, Va.	16.8	0.47	0.24	0.080	2.54	35.74	3.00	24.6
26	Copper Creep, Va.	19.6	0.84	0.49	0.101	1.26	23.33	4.85	20.8
27	Difficult Run, Va.	14.5	0.31	0.25	0.062	1.09	46.77	4.03	1.9
28	John Day River, Ore.	25.0	0.58	1.01	0.140	1.08	43.10	7.21	13.9
29	John Day River, Ore.	34.1	2.47	0.82	0.180	1.89	13.81	4.56	65.0
30	Little Pincy Creek, Md.	15.9	0.22	0.39	0.053	1.13	72.27	7.36	7.1
31	Missouri River	183.0	2.33	0.89	0.066	1.35	78.54	13.48	465.0
32	Missouri River	197.0	3.11	1.53	0.078	1.35	63.34	19.62	892.0
33	Missouri River	201.0	3.56	1.28	0.084	1.35	56.46	15.24	837.0
34	Monocacy River, Md.	48.7	0.55	0.26	0.052	1.28	88.55	5.00	37.8
35	Monocacy River, Md.	93.0	0.71	0.16	0.046	1.28	130.99	3.48	41.4
36	Monocacy River, Md.	51.2	0.65	0.62	0.044	1.28	78.77	14.09	29.6
37	Monocacy River, Md.	97.5	1.15	0.32	0.058	1.61	84.78	5.52	119.8
38	Monocacy River, Md.	40.5	0.41	0.23	0.040	1.61	98.78	5.75	66.5
39	Nooksack River	86.0	2.93	1.20	0.530	1.30	29.35	2.26	153.0
40	Nooksack River	64.0	0.76	0.67	0.268	1.30	84.21	2.50	34.8
41	Powell River, Tenn.	36.8	0.87	0.13	0.054	2.20	42.30	2.41	15.5
42	Red River, La	253.6	1.62	0.61	0.032	1.20	156.54	19.06	143.8
43	Red River, La	161.5	3.96	0.29	0.060	1.44	40.78	4.83	130.5
44	Red River, La	152.4	3.66	0.45	0.057	1.44	41.64	7.89	227.6
45	Red River, La	155.1	1.74	0.47	0.036	1.24	89.14	13.06	177.7
46	Sabina River, La	116.4	1.65	0.58	0.054	1.19	70.55	10.74	131.3
47	Sabina River, La.	160.3	2.32	1.06	0.054	1.17	69.09	19.63	308.9
48	Sabina River, Tex.	14.2	0.50	0.13	0.037	2.53	28.40	3.51	12.8
49	Sabina River, Tex.	12.2	0.51	0.23	0.030	2.05	23.92	7.67	14.7
50	Sabina River, Tex.	21.3	0.93	0.36	0.035	1.47	22.90	10.29	24.2
51	Salt Creek, Nebr.	32.0	0.50	0.24	0.038	1.38	64.00	6.32	52.2
52	Susquehanna River	203.0	1.35	0.39	0.065	1.13	150.37	6.00	92.9
53	Tangipahoa River, La	31.4	0.81	0.48	0.072	1.46	38.77	6.67	45.1
54	Tangipahoa River, La	29.9	0.40	0.34	0.020	1.46	74.75	17.00	44.0
55	Tickfau River, La	15.0	0.59	0.27	0.080	1.75	25.42	3.38	10.3
56	Wind/Big. River, Wyo	44.2	1.37	0.99	0.142	1.56	32.26	6.97	184.6
57	Wind/Big. River, Wyo	85.3	2.38	1.74	0.153	1.56	35.84	11.37	464.6
58	Wind/Big. River, Wyo	59.4	1.10	0.88	0.119	1.18	54.00	7.39	41.8
59	Wind/Big. River, Wyo	68.6	2.16	1.55	0.168	1.18	31.76	9.23	162.6
60	Yadkin River, N.C.	70.1	2.35	0.43	0.101	2.17	29.83	4.26	111.5
61	Yadkin River, N.C.	71.6	3.84	0.76	0.128	2.17	18.65	5.94	260.1

Note: The shaded is a verification dataset.

**Table 2.** Average and range values for geometric and hydraulic data of the considered rivers.

Dataset	$U/U_*$	$W/H$	$S_i$	$K \text{ (m}^2 \text{ s}^{-1}\text{)}$
	Min – Max – Avg	Min – Max – Avg	Min – Max – Avg	Min – Max – Avg
Whole	1.29 – 19.63 – 7.23	13.81 – 156.54 – 51.84	1.08 – 2.54 – 1.55	1.9 – 892.0 – 104.8
Deriving	1.29 – 19.06 – 7.10	18.65 – 156.54 – 52.75	1.09 – 2.54 – 1.57	1.9 – 837.0 – 104.5
Verification	2.33 – 19.63 – 7.60	13.81 – 131.00 – 49.28	1.08 – 2.53 – 1.48	2.9 – 892.0 – 105.4

$$CC = \frac{\left[ \sum_{i=1}^N (K_m - \bar{K}_m)(K_p - \bar{K}_p) \right]}{\sqrt{\sum_{i=1}^N (K_m - \bar{K}_m)^2 \sum_{i=1}^N (K_p - \bar{K}_p)^2}}, \quad (13)$$

$$DR = \log \frac{K_p}{K_m}, \quad (14)$$

$$\% \text{ accuracy} = \frac{100n}{N}, \quad (15)$$

where  $K_m$ ,  $K_p$  are the measured and predicted longitudinal dispersion coefficients respectively,  $\bar{K}_m$ ,  $\bar{K}_p$  are the mean measured and predicted longitudinal dispersion coefficients respectively,  $N$  is number of observations and  $n$  is number of predictions lying between 50% and 200% of the measured values (i.e., DR values lying between –0.3 and 0.3).

From Eq. (14),  $DR = 0$  suggests exact match between the measured and predicted values, otherwise, there is either an over prediction ( $DR > 0$ , i.e.,  $K_p > K_m$ ) or under prediction ( $DR < 0$ , i.e.,  $K_p < K_m$ ).

### Genetic algorithm

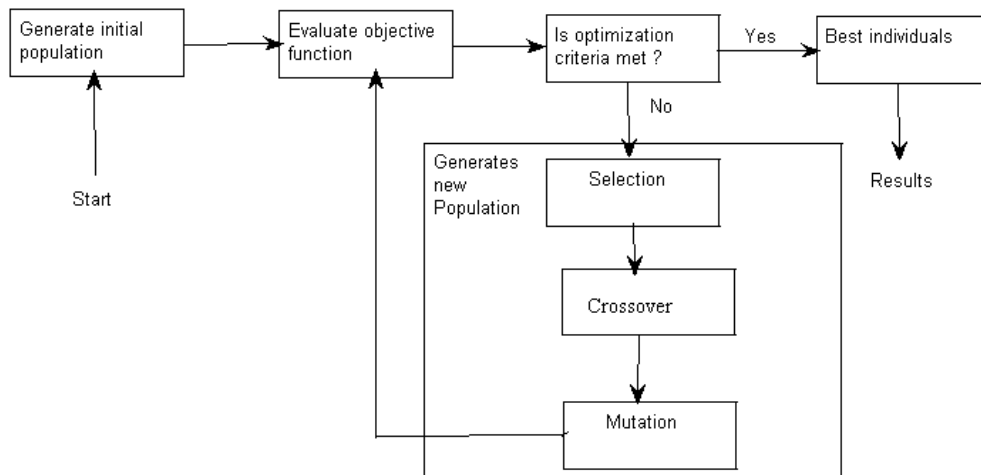
The new expression is derived using genetic algorithm on published dispersion data. GAs are search and optimization procedures motivated by the principles of natural genetics and natural selection. Using this analogy, a process involving selection, crossover, and mutation are used to evolve a

population of potential solutions for a variety of engineering problems towards improved solutions. These solutions satisfy the specified constraints while minimizing or maximizing the objective function. Goldberg (1989), Michalewicz (1996) and Deb (2002) have given good illustrations of GAs. The implementation of GA starts with initializing a population of candidate solutions (called chromosomes) which are randomly sampled from the feasible parameter space. In each generation, the individual chromosomes are selected through a fitness-based process, where the more fit chromosomes in the population are preferentially selected to reproduce new promising offspring using crossover and mutation operations. The crossover operator chooses ‘parent’ solutions and exchanges important building blocks of the two parent chromosomes to generate new ‘offspring’ solutions. The ‘offspring’ solutions are then randomly mutated to increase the diversity of the new population. The operation is repeated until a stopping criterion is reached. The criterion may be either the number of generations or the change in the fitness values of chromosomes between two consecutive generations. The working structure of GA model is given in Fig. 1.

The criterion considered in the present study for determining optimal values of constant  $\alpha$ ,  $\beta$ ,  $\gamma$  and  $\delta$  within a GA framework is the minimization of the residual sum of squares errors ( $SSQ$ ) between the measured and predicted dispersion coefficients, i.e.,

$$\text{Min } SSQ = \sum_{i=1}^N (K_p - K_m)^2 \quad \text{or}, \quad (16)$$

$$\text{Min } SSQ = \sum_{i=1}^N \left[ \mu H U_* \left( \frac{W}{H} \right)^\beta \left( \frac{U}{U_*} \right)^\gamma S_i^\delta - K_m \right]^2. \quad (16a)$$

**Fig. 1.** Working structure of GA model.

Here, GA is implemented on MATLAB 7 toolbox adopting a binary code of representation for variables. A string-length of 10 for each variable is used which is sufficient for the range of values that these variables can attain. Thus, the total string-length of a chromosome is 40 bits. After trying many combinations of population size, crossover probability and mutation, a GA scheme with a population size 100, Gaussian crossover fraction 0.85, Gaussian mutation function with scale and shrink 1 each and reproduction with elite count 2 is found to yield the minimum SSQ. The corresponding values of the design coefficients found are  $\alpha = 2.0$ ,  $\beta = 0.72$ ,  $\gamma = 1.37$  and  $\delta = 1.52$ . Thus, the new derived expression for longitudinal dispersion coefficient in a natural stream is

$$\frac{K}{HU_*} = 2 \left( \frac{W}{H} \right)^{0.72} \left( \frac{U}{U_*} \right)^{1.37} S_i^{1.52} \quad (17)$$

After deriving the new expression, it is evaluated for its performance. Table 3 summarizes the performance of all models. It shows the new expression, i.e., Eq. (17), to have successfully been derived as prediction by it yields the minimum RMSE and the maximum CC among all the considered models. The predicted values of  $K$  by Eq. (17) can be seen evenly distributed around the ideal line, showing little bias for over or under prediction (Fig. 2). This expression successfully predicted the highest three  $K$ -values of  $892.0 \text{ m}^2 \text{ s}^{-1}$ ,  $837.0 \text{ m}^2 \text{ s}^{-1}$  and  $464.6 \text{ m}^2 \text{ s}^{-1}$  as  $897.1 \text{ m}^2 \text{ s}^{-1}$ ,  $720.8 \text{ m}^2 \text{ s}^{-1}$  and  $398.4 \text{ m}^2 \text{ s}^{-1}$  respectively, and the lowest three  $K$ -values of  $1.9 \text{ m}^2 \text{ s}^{-1}$ ,  $5.8 \text{ m}^2 \text{ s}^{-1}$  and  $7 \text{ m}^2 \text{ s}^{-1}$  as  $4.76 \text{ m}^2 \text{ s}^{-1}$ ,  $17.69 \text{ m}^2 \text{ s}^{-1}$  and  $8.42 \text{ m}^2 \text{ s}^{-1}$  respectively.

It is understood that the satisfactory prediction by the new expression may not be entirely due to the consideration of channel sinuosity and GA might have captured the input-output pattern better than earlier models, however, the results as a whole, do suggest improvement when sinuosity is taken into account.

## MODEL VALIDATION, RESULTS AND DISCUSSION

After deriving the new expression satisfactorily, it is evaluated for the verification and the whole datasets. Table 3 enumerates the performance indices of the models for these datasets. It shows the new expression a better predictive model than any other considered model as prediction by it has the least RMSE and the highest CC, both for the verification dataset and

the whole dataset. Models D-S-B, K-F and S-D follow the performance of the proposed model in that order. The Fischer model is found the least accurate with RMSE value as high as  $982.05 \text{ m}^2 \text{ s}^{-1}$  for the verification dataset and  $867.44 \text{ m}^2 \text{ s}^{-1}$  for the whole dataset. The performance of all models improves when extreme values ( $K > 100 \text{ m}^2 \text{ s}^{-1}$  and  $W/H > 50$ ) are ignored. The most significant improvements are found in Fischer, Liu and S-C models, with their respective RMSEs reducing from  $867.44 \text{ m}^2 \text{ s}^{-1}$ ,  $173.56 \text{ m}^2 \text{ s}^{-1}$ , and  $109.29 \text{ m}^2 \text{ s}^{-1}$  to  $21.15 \text{ m}^2 \text{ s}^{-1}$ ,  $21.18 \text{ m}^2 \text{ s}^{-1}$  and  $38.85 \text{ m}^2 \text{ s}^{-1}$  respectively. However, this reduction suggests their inadequacy for predicting  $K$  in very wide and shallow rivers or rivers with very high dispersion rates. On comparison, the proposed expression shows the least improvement with the omission of extreme values, making it a general solution, suitable for predicting  $K$  in all kinds of rivers.

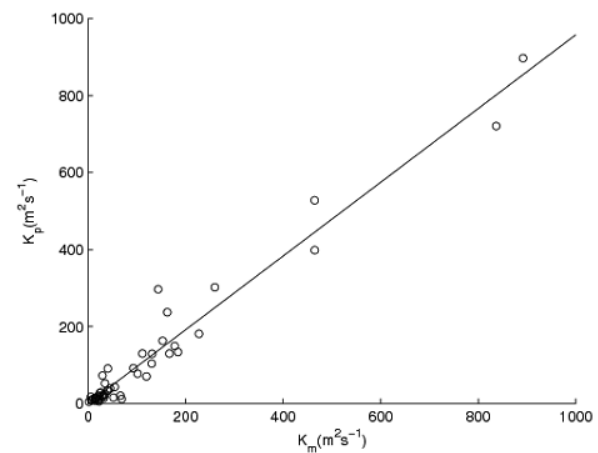


Fig. 2. Predicted vs. measured  $K$  (derivation dataset).

Another performance indicator, discrepancy ratio (DR), which shows proximity of the predicted values with the measured values, for the new model is found superior as its range of  $-0.76$  to  $0.94$  is found the best among all the considered models. The DR ranges of other models are skewed towards either underestimation (K-F) or overestimation (Fischer, Liu, S-C, D-S-B and S-D). If % accuracy of a model is defined as the percentage of the predicted values lying between 50% and 200% of the measured values, i.e., DR values

Table 3. Performance indices of the models.

Prediction equations	Derivation dataset		Verification dataset		Whole dataset				
	CC	RMSE ( $\text{m}^2 \text{ s}^{-1}$ )	CC	RMSE ( $\text{m}^2 \text{ s}^{-1}$ )	CC	RMSE ( $\text{m}^2 \text{ s}^{-1}$ )	DR Range	RMSE ( $\text{m}^2 \text{ s}^{-1}$ ) ( $K > 100$ and $W/H > 50$ ignored)	Accuracy (%)
GA (Present study)	0.96	41.08	0.99	33.51	0.97	39.24	$-0.76$ to $0.94$	17.99	79
Fischer	0.51	822.85	0.96	982.05	0.69	867.44	$-1.04$ to $1.55$	21.15	39
Liu	0.56	196.47	0.94	79.28	0.66	173.56	$-1.78$ to $1.95$	21.18	46
S-C	0.94	86.43	0.98	156.61	0.95	109.29	$-1.77$ to $1.90$	38.85	39
D-S-B	0.94	55.50	0.98	69.47	0.95	59.48	$-0.72$ to $0.99$	30.30	64
K-F	0.93	60.01	0.97	62.46	0.94	60.67	$-1.42$ to $0.97$	23.04	56
S-D	0.91	74.88	0.98	94.99	0.93	80.64	$-0.62$ to $1.13$	23.70	62

Note: Fischer is Fischer et al. (1979), Liu is Liu (1977), S-C is Seo and Cheong (1998), D-S-B is Deng et al. (2001), K-F is Kashefipour and Falconer (2002), and S-D is Sahay and Dutta (2009).

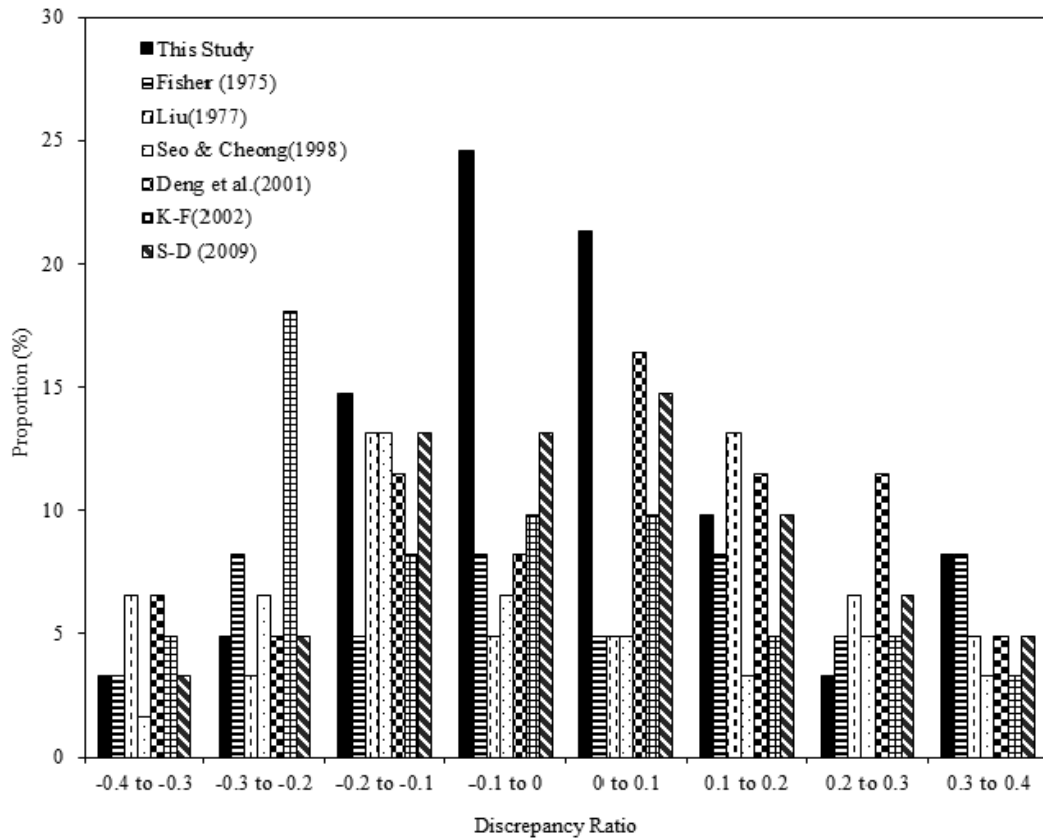


Fig. 3. Comparison of discrepancy ratio by models.

Table 4. Measured and predicted longitudinal dispersion coefficients by the models (verification dataset).

Stream	Obs. value	Present, Eq. (13)	Fischer, Eq. (3)	Liu, Eq. (4)	S-C, Eq. (6)	K-F, Eq. (9)	D-S-B, Eq. (7)	S-D, Eq. (10)	T-S (ANN)
1 Antietam Creek, Md.	20.9	28.6	5.1	7.4	20.2*	15.2	15.0	13.7	26.8
2 Bear Creek, Colo.	2.9	25.1	7.3*	33.6	52.2	27.1	28.1	39.1	39.2
3 Chattahooch. River, Ga	88.9	108.2	127.9	168.6	169.1	82.1*	168.8	147.1	77.6
4 Clinch River, Va	10.7	15.3	26.4	37.8	27.6	11.5*	28.5	25.8	26.9
5 Clinch River, Va	36.9	70.9	52.6*	56.2	139.6	104.1	118.3	97.8	76.6
6 Conococh. Creek, Md.	53.3	63.3	88.2	66.5	96.7	58.8*	93.2	78.2	43
7 John Day River, Ore.	13.9	41.3*	86.4	72.9	83.3	44.8	81.8	71.2	45.2
8 John Day River, Ore.	65.0	124.0	19.3	32.6	116.7	97.9	71.1*	73.6	77.2
9 Missouri River	892	897.1*	4119.6	776.0	1317.3	990.5	952.1	1074.7	763.4
10 Monocacy River, Md.	37.8	19.2	61.7	90.3	27.1	7.6	28.2	31.7*	27.1
11 Monocacy River, Md.	41.4	17.8	74.6	188.1	23.5	4.2	25.8	33.4*	31.4
12 Powell River, Tenn.	15.5	15.5*	5.4	23.5	9.9	2.9	9.9	10.3	25.3
13 Sabina River, La.	308.9	397.3	2535.1	477.0	718.7	512.3	509.0	603.8	346.6*
14 Sabina River, Tex.	12.8	9.5*	2.0	5.0	5.2	2.4	4.6	4.4	21.9
15 Tangipahoa River, La	44.0	30.9	142.1	33.2	39.2*	24.5	28.7	34.7	26.5
16 Wind/Big. River, Wyo	41.8	92.8	229.6	186.8	159.9	76.0	156.7	147.0	59.7*
No. of closest estimates		4	2	0	2	3	1	2	3

Note: \* is closest estimate to the measured value, Fischer is Fischer et al. (1979), Liu is Liu (1977), S-C are Seo and Cheong (1998), D-S-B is Deng et al. (2001), and K-F are Kashefipour and Falconer (2002), S-D are Sahay and Dutta (2009) and T-S are Tayfur and Singh (2005).

lying between  $-0.3$  and  $0.3$ , then, Table 3 shows that 79% of the prediction by the new expression is accurate, followed by D-S-B, K-F and S-D models with 64%, 62% and 56% accuracies respectively. The Fischer is again found to be the least accurate model with only 39% accuracy.

Fig. 3 shows percentage of predicted values of dispersion coefficient falling into different discrepancy brackets. The objective is to show how the predicted values compare with their measured values for the entire dataset. It shows an almost even distribution for predictions from the new expression, suggesting little bias for under or over prediction. On

comparison, predictions by other models are skewed toward either the positive or the negative side.

Table 4 shows the measured and predicted values of  $K$  for the verification dataset. The highest number of predicted values by the proposed expression is found closest to the measured values. While 4 out of 16 estimates by the new model are closest to the measured values, T-S, K-F, S-C, S-D, D-S-B, and Fischer have only 3, 3, 2, 2, 1 and 1 predictions respectively closest to the measured values.

## SENSITIVITY ANALYSIS

A sensitivity and error analysis is carried out on the mean values of input and output parameters of the new expression (Eq. (17)). The average values of 61 sets of data in Table 1 are taken as  $W = 61.79$  m,  $H = 1.28$  m,  $U = 0.54$  m s<sup>-1</sup>,  $U_* = 0.093$  m s<sup>-1</sup>, and  $S_i = 1.55$ . The indices used to compare sensitivities and error are defined as

$$\text{Absolute Sensitivity (AS)} = \frac{\partial K}{\partial \theta} \approx \frac{\Delta K}{\Delta \theta}$$

$$\text{Relative Sensitivity (RS)} = \frac{\Delta K}{\Delta \theta} \frac{\theta}{K}$$

$$\text{Relative Error (RE)} = \frac{\Delta K}{K}, \quad (18)$$

where  $\Delta K$  is the deviation in  $K$  caused by the variation in input  $\theta$  ( $W$ ,  $H$ ,  $U$ ,  $U_*$ , or  $S_i$ ) by an amount  $\Delta \theta$ . Incrementing each input by 10%, i.e.,  $\Delta \theta = 0.1\theta$ ,  $\Delta K$ , AS, RS and RE are determined and given in Table 5. Table 5 shows the channel sinuosity causing the greatest deviation in  $K$ , illustrating its importance in riverine solute transport. Any inaccuracy in measurement of  $S_i$  would lead to significant error in  $K$ . The flow velocity  $U$  has almost similar impact on  $K$ , followed by the channel width  $W$ , flow depth  $H$  and shear velocity  $U_*$ . But as far as AS, i.e., rate of change in  $K$  with change in  $\theta$  is concerned,  $U$  is the most sensitive input followed by  $S_i$ ,  $U_*$ ,  $H$  and  $W$ .

**Table 5.** Sensitivity and error analysis of new expression (Eq. (17)).

Par	$\theta$	$\Delta \theta$	$\Delta K$	AS	RS	RE
$W$	61.79	6.18	5.98	0.97	0.71	0.071
$H$	1.28	0.13	2.27	17.77	0.27	0.027
$U$	0.54	0.05	11.73	217.26	1.39	0.139
$U_*$	0.09	0.01	0.57	60.99	0.07	0.007
$S_i$	1.55	0.16	13.11	84.59	1.56	0.156

Note:  $\theta$  is input parameter,  $\Delta \theta = 0.1 \theta$ ,  $\Delta K$  is change in  $K$  due to change in  $\theta$  by  $\Delta \theta$ , AS is absolute sensitivity, RS is relative sensitivity, and RE is relative error.

## CONCLUSION

In spite of many studies dealing with effect of sinuosity on dispersion coefficient, there is still considerable uncertainty at accurate determination of the longitudinal dispersion coefficient value for the specific natural conditions. In the present work, a new empirical expression for  $K$  has been derived employing genetic algorithm on published field data which takes into account the river sinuosity in addition to the other readily available bulk hydraulic and geometric characteristics of rivers.

The performance of the new expression is compared with those of Fischer et al. (1979), Liu (1977), Seo and Cheong (1998), Deng et al. (2001), Kashefipour and Falconer (2002), and Sahay and Dutta (2009). Based on various performance indices, the new expression is found to give better estimate of  $K$ . About 79% of the predicted values by the new expression lie between 50% and 200% of the measured values, the highest among all the considered models. Fischer et al. (1979) is found to yield the least satisfactory estimate. The sensitivity and error analysis carried out on the average values of the parameters of the new expression find  $K$  sensitive to the channel sinuosity. An error of 10% in measurement of sinuosity leads to 13.11 % deviation in the longitudinal dispersion coefficient in sinuous rivers.

## REFERENCES

- David, T.H., Peter, S., Theodore, C., 2002. Determination of longitudinal dispersion coefficient and net advection in the Tidal Hudson River with a large-scale, high resolution SF6 tracer release experiment. *Environ. Sci. Technol.*, 36, 3234–3241.
- 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., List, E.J., Koh, R.C.Y., Imberger, J., Brooks, N.H., 1979. *Mixing in Inland and Coastal Waters*. Academic Press, New York, 104–138.
- French, R.H., 1986. *Open-Channel Hydraulics*. McGraw-Hill Book Company, New York.
- Fukuoka, S., Sayre, W.W., 1973. Longitudinal dispersion in sinuous channels. *J. Hydraulic Div.*, 99, 1, 195–217.
- Godfrey, R.G., Frederick, B.J., 1970. *Stream Dispersion at Selected Sites*. U.S. Geological Survey, Washington, D.C., Paper 433-K.
- Goldberg, D.E., 1989. *Genetic Algorithms*. In: *Search, Optimization and Machine Learning*. Addison-Wesley, New York.
- Guymer, I., 1998. Longitudinal dispersion in a sinuous channel with changes in shape. *J. Hydraulic Eng.*, 124, 1, 33–40.
- Iisa, Y., Aya, S., 1991. Predicting longitudinal dispersion coefficient in open channel flows. *Proceedings of International Symposium on Environmental Hydraulics*, Hong Kong, 505–510.
- Kashefipour, M.S., Falconer, R.A., 2002. Longitudinal dispersion coefficients in natural channels. *Water Resources Research*, 36, 6, 1596–1608.
- Liu, H., 1977. Predicting dispersion coefficient of streams. *J. of Environment Engineering Division*, 103, 1, 59–69.
- Michalewicz, Z., 1996. *Genetic Algorithm + Data Structures = Evolutionary Programs*. Springer, New York.
- Rutherford, J.C., 1994. *River Mixing*. John Wiley & Sons, New York.
- Sahay, R.R., 2011. Prediction of longitudinal dispersion coefficients in natural rivers using artificial neural network. *Environ. Fluid Mechanics*, 11, 247–261.
- Sahay, R.R., Dutta, S., 2009. Prediction of longitudinal dispersion coefficients in natural rivers using genetic algorithm. *Hydrology Research*, 40(6), 544–552.
- Sayre, W.W., 1968. Dispersion of mass in open channel flow. *Colorado State University, Hydraulic Paper*, 3(4), 73.

- Seo, I.W., Cheong, T.S., 1998. Predicting longitudinal dispersion coefficient in natural streams. *J. Hydraulic Eng.*, 124, 1, 25–32.
- Tayfur, G., Singh, V.P., 2005. Predicting longitudinal dispersion coefficient in natural streams by artificial neural network. *J. Hydraulic Eng.*, 131, 11, 991–1000.
- 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*, 223A, 446–468.
- Thackston, E.L., Krenkel, P.A., 1967. Longitudinal mixing in natural streams. *J. Sanitary Engineering Div., Proc. Soc.*, 93, SAS, 67–90.
- Toprak, Z.F., Cigizoglu, H.K., 2008. Predicting longitudinal dispersion coefficient in natural streams by artificial intelligence methods. *Hydrological Processes*, 22, 4106–4129.

Received 13 September 2011

Accepted 6 December 2012