Water Resour Manage (2013) 27:5245–5260
DOI 10.1007/s11269-013-0465-2

Differential Evolution for Prediction of Longitudinal Dispersion Coefficients in Natural Streams

Xiangtao Li · Huawen Liu · Minghao Yin

Received: 30 June 2013 / Accepted: 15 October 2013 /
Published online: 30 October 2013

Abstract Differential evolution (DE) is a population-based evolutionary algorithm widely used for solving multidimensional global optimization problems over continuous spaces, and has been successfully used to solve several kinds of problems. In this paper, a novel expression for the prediction of longitudinal dispersion coefficient in natural streams is proposed to minimize the sum-square error using differential evolution algorithm. The new expression considers the hydraulic and geometric characteristics of rivers. Datasets consisting 65 sets of observations from 29 rivers in the unite states are used to test the proposed algorithm, and results demonstrate the performance and applicability of the proposed differential evolution. Compared with the previous methods, the new expression using differential evolution is superior to other expressions. Moreover, 56.92 % of the prediction using the new expression lie with the $0.5<K_{pre}/K_{meas}<1.5$ that is better than other expressions.

Keywords Longitudinal dispersion coefficients · Differential evolution · Genetic algorithm · Meta-heuristic

1 Introduction

The longitudinal dispersion coefficients play an important role for predicting concentration variation of dispersed pollutants in the flow direction and explanting on-dimensional equations of motion (Fischer et al. 1979). Many researchers have proposed a number of methods to the understanding of mechanisms of longitudinal dispersion in rivers. After that, the idea of dispersion is used to the mixing in open channels and further to natural streams. Different methods have been proposed to find the longitudinal dispersion coefficient.

X. Li · M. Yin (✉)
College of Computer Science, Northeast Normal University, Changchun 130117, People's Republic of China
e-mail: Minghao.yin1@gmail.com

X. Li
e-mail: lixt314@nenu.edu.cn

H. Liu
College of Mathematics, Physics and Information Engineering, Zhejiang Normal University, Jinhua 321004, China

Accurate estimation of longitudinal dispersion coefficient is important in many different hydraulic problems including environmental engineering, estuaries problems and contaminants into river flows. In order to verify the quality of natural streams, the streams need to use 1D mathematical model to find the best value of longitudinal dispersion coefficient. There are several empirical equations for estimation of longitudinal dispersion coefficient in streams. Estimation of longitudinal dispersion coefficient in streams using the equation of Table 1 needs hydraulic and geometry of data sets. According to the above analyses in Table 1, these techniques can be categorized into two parts: exact methods and global search methods (Taylor 1953, 1954; Elder 1959; Fischer 1967; Liu 1977; Seo and Cheong 1998; Deng et al. 2001; Kashefipour and Falconer 2002; Sahay and Dutta 2009). From the paper (Sahay and Dutta 2009), we can find that the global search method can obtain better solutions than other methods. Therefore, the global search method is a better choose for prediction of longitudinal dispersion coefficient. However, this field of study is still in its early days, a large number of future researches are necessary in order to develop new global search methods for the prediction of longitudinal dispersion coefficient.

Recently, a differential evolution algorithm (Storn and Price 1997) is proposed as a simple and powerful population-based stochastic optimization method, which is originally motivated by the mechanism of natural selection. This algorithm searches solutions using three basic operators: mutation, crossover and greedy selection. Mutation is used to generate a mutant vector by adding differential vectors. After that, crossover operator generates the trial vector by combining the parameters of the mutated vector with the parameters of a parent vector selected from the population. Finally, according to the fitness value, selection determines which of the vectors should be chosen for the next generation by implementing one-to-one completion between the generated trail vectors and the corresponding parent vectors. In order to accelerate the convergence speed and avoid trapping in the local optima, several variations of DE have

Table 1 Empirical equation for estimation of longitudinal dispersion coefficient

Method	Equation
Taylor (1953, 1954)	$\frac{\partial C}{\partial t}+U\frac{\partial C}{\partial x}=K\frac{\partial^2 C}{\partial t^2}$
Elder (1959)	K=5.93 HU_*
Fischer (1967)	$K=-\frac{1}{A}\int_0^W hu'\int_0^y \frac{1}{\varepsilon_t h}\int_0^y hu'dydydy$
Fischer et al. (1979)	$K=0.11\left(\frac{W}{H}\right)^2\left(\frac{U}{U_*}\right)^2 HU_*$
Liu (1977)	$\frac{K}{HU_*}=\beta\left(\frac{W}{H}\right)^2\left(\frac{U}{U_*}\right)^2$
Seo and Cheong (1998)	$\frac{K}{HU_*}=5.915\left(\frac{W}{H}\right)^{0.62}\left(\frac{U}{U_*}\right)^{1.428}$
Deng et al. (2001)	$\frac{K}{HU_*}=\frac{0.15}{8\varepsilon_t}\left(\frac{W}{H}\right)^{5/3}\left(\frac{U}{U_*}\right)^2$
Kashefipour and Falconer (2002)	$\frac{K}{HU_*}=10.612\left(\frac{U}{U_*}\right)^2$
GA model (2009)	$\frac{K}{HU_*}=2\left(\frac{W}{H}\right)^{0.96}\left(\frac{U}{U_*}\right)^{1.25}$

been proposed to enhance the performance of the standard DE recently. Moreover, DE has been proved to be quite efficient when solving real-world problems (Brest et al. 2006; Rahnamayan et al. 2008; Qin et al. 2009; Li et al. 2011; Li and Yin 2012a, b, 2013, 2012a, b; Neri and Tirronen 2009; Zhang and Sanderson 2009; Qin et al. 2010; Vasan and Raju 2007). Brest et al. (2006) proposed a self adaptive parameter setting in differential evolution in order to avoid the manual parameter setting of F and CR. The parameter control technique is based on the self adoption of two parameters associated with the evolutionary process. Rahnamayan et al. (2008) proposed an opposition based differential evolution, as called ODE. The ODE algorithm consisted of a DE framework and two opposition based components: the former after the initial sampling and the latter after the survivor selection scheme. Qin et al. (2009) proposed a self adaptive DE algorithm (SaDE), in which both trail vector generation strategies and the associated control parameter values were gradually self-adaptive by learning from their previous experiences when generating promising solutions. Neri and Tirronen (2009) proposed the scale factor local search differential evolution. This algorithm is a differential evolution based memetic algorithm which employs, within a self adaptive scheme, two local search algorithms. These local search algorithms aim at detecting a value of the scale factor corresponding to an offspring with a high performance. A statistical analysis of the optimization results has been included in order to compare the results in terms of final solution detected and convergence speed.

In this paper, we will use the differential evolution for prediction of longitudinal dispersion coefficient. We employ the mutation operator, crossover operator, selection operator to generate the offspring. In the last part, the algorithm needs to input the generated population into the longitudinal dispersion coefficient according to the measure data and calculate the fitness function value. Data consisting of 65 sets of observations from 29 rivers in the unite states are used to demonstrate the performance and applicability of the proposed differential evolution. In order to demonstrate the advantages of the proposed design, the results obtained are compared with other state-of-the-art approaches. Experimental results demonstrate that the proposed method is better or at least comparable with previous method from literature when considering the quality of the solutions obtained.

2 Differential Evolution Algorithm and Application

2.1 Differential Evolution Algorithm

Differential Evolution (DE) is an algorithm proposed by Storn and Price (1997) based on vector operations to generate potential candidates to solve continuous optimization problems. The fundamental crucial idea behind DE is a scheme for producing trial parameter vectors. If the trail vector's fitness value is better than a predetermined population member, the better individual will be retained and be compared in the following generation.

The algorithm begins with a randomly initiated population which generates $NP*D$ matrix with uniform probability distribution random values. We can generate the jth component of the ith vector as

$$x_{j,i,0} = x_{j,\min} + rand_{i,j}[0,1]\cdot\left(x_{j,\max} - x_{j,\min}\right) \tag{1}$$

Where $rand_{i,j}[0,1]$ is a uniformly distribution random number between 0 and 1. i=1,…, NP and j=1,…,D. $x_{j,\max}$, $x_{j,\min}$ is the upper bound and lower bound of the jth column, respectively.

After initialization, mutation vectors $V_{i,G}$ are generated according to each population member or target vector $X_{i,G}$ in current population. In the standard DE algorithm, five different mutation strategies can be used with one of two different crossover methods, which are listed in the followings:

"DE/rand/1"

$$V_{i,G} = X_{r_1,G} + F\cdot\left(X_{r_2,G} - X_{r_3,G}\right)$$

"DE/best/1"

$$V_{i,G} = X_{best,G} + F\cdot\left(X_{r_1,G} - X_{r_2,G}\right)$$

"DE/current-to-best/1"

$$V_{i,G} = X_{i,G} + F\cdot\left(X_{best,G} - X_{i,G}\right) + F\cdot\left(X_{r_1,G} - X_{r_2,G}\right)$$

"DE/best/2"

$$V_{i,G} = X_{best,G} + F\cdot\left(X_{r_1,G} - X_{r_2,G}\right) + F\cdot\left(X_{r_3,G} - X_{r_4,G}\right)$$

"DE/rand/2"

$$V_{i,G} = X_{r_1,G} + F\cdot\left(X_{r_2,G} - X_{r_3,G}\right) + F\cdot\left(X_{r_4,G} - X_{r_5,G}\right) \tag{2}$$

Where $r_1,r_2,r_3,r_4,r_5 \in [1,\cdots,NP]$ are randomly chosen integers, and $r_1 \neq r_2 \neq r_3 \neq r_4 \neq r_5 \neq i$. F is a mutation control parameter which affects the disturbance added by the weight of different vectors. $X_{best,G}$ is the best individual with the best fitness in the current population at generation G. In this paper, we use the "DE/rand/1/bin" mutation method to generate the offspring vector.

In the crossover operation, a recombination of the offspring vector $V_{i,G}$ and the parent vector $X_{i,G}$ produce a trail vector $U_{i,G}=[u_{1,i,G},u_{2,i,G},u_{3,i,G},\cdots,u_{D,i,G}]$. Usually the binomial crossover is accepted, which is defined as follows:

$$u_{j,i,G} = \begin{cases} v_{j,i,G}, \left(rand_j[0,1] \leq CR\right) \ \ or \ \ (j = j_{rand}) \\ x_{j,i,G}, otherwise \end{cases} \tag{3}$$

where $j=[1,\cdots,D]$; $rand_j \in [0,1]$ is a random number between [0,1]; $j_{rand}=[1,\cdots,D]$ is the randomly chosen index, CR is the crossover rate $v_{j,i,G}$ is the difference vector of the jth particle in the ith dimension at the Gth iteration, and $u_{j,i,G}$ denotes the trail vector of the jth particle in the ith dimension at the Gth iteration.

Selection operator is used to choose the next population (i.e. G=G+1) between the trail population and the target population. The selection operation is described as:

$$\begin{aligned} X_{i,G+1} &= U_{i,G}, \quad \text{If } f\left(U_{i,G}\right) \leq f\left(X_{i,G}\right) \\ &= X_{i,G}, \quad \text{If } f\left(U_{i,G}\right) > f\left(X_{i,G}\right) \end{aligned} \tag{4}$$

The standard differential evolution algorithm can be described as the followings:

procedure Algorithm description of DE algorithm

```
begin
Set the generation counter G=0; and randomly initialize a population of
NP individuals X_i. Initialize the parameter F, CR
Evaluate the fitness for each individual in P.
while stopping criteria is not satisfied do
    for i= 1 to NP do
        select randomly a ≠ b ≠ c ≠ d ≠ i
        for j=1 to D do
            j_rand = ⌊rand(0,1) * D⌋
            If rand(0,1) ≤ CR or j== j_rand then
                u_{i,j} = x_{a,j} + F × (x_{b,j} − x_{c,j})
                /* five mutation strategies*/
            Else
                u_{i,j} = x_{i,j}
            end if
        end for
    end for
    for i=1 to NP do
        Evaluate the offspring u_i
        If u_i is better than X_i then
            X_i = u_i
        end if
    end for
    Memorize the best solution achieved so far
end while
end
```

2.2 Formulation of the Problem and Application

Generally speaking, the expressions for estimation of longitudinal dispersion coefficient can be described as follows:

$$\frac{K}{HU_*} = a\left(\frac{W}{H}\right)^b \left(\frac{U}{U_*}\right)^c \tag{5}$$

The Eq. (5) can be converted as follows:

$$K = aW^b H^{1-b} U^c U^{1-c}$$

For this problem, the objective for the prediction of some parameters *a*, *b* and *c* within a DE model is to find the minimization of the sum-square error (SSE) between the actual and predicted dispersion coefficients. The sum-square error can be defined as follows:

$$\text{Min SSE} = \sum_{i=1}^{N} \left(K_{pre} - K_{meas}\right)^2 \quad \text{Or} \quad \text{Min SSE} = \sum_{i=1}^{N} \left(aW^b H^{1-b} U^c U^{1-c} - K_{meas}\right)^2 \tag{6}$$

where K_{pre} is the predicted longitudinal dispersion coefficient and K_{meas} is the actual longitudinal dispersion coefficient. The parameter N is the number of observations.

Obviously, this problem is a multi-dimension continuous optimization problem, where the predicted vector is K_{pre} and the optimization objective is to minimize SSE.

The estimation of longitudinal dispersion coefficient is not easy to find the suitable values of a, b and c to minimization of the sum-square error because of the unstable dynamic of the estimation of longitudinal dispersion coefficient. Moreover, due to multiple variables in the problem and multiple local search optima in the objective functions, traditional optimization cannot easy to trap in local optima. Therefore, we aim to solve this problem by proposing a differential evolution in this paper. The structure of DE for predicting longitudinal dispersion coefficients in natural rivers is given in Fig. 1.

From Fig. 1, the algorithm starts from a random population within the three bounds. An individual is consist of three parameters a, b, and c, which should be determined

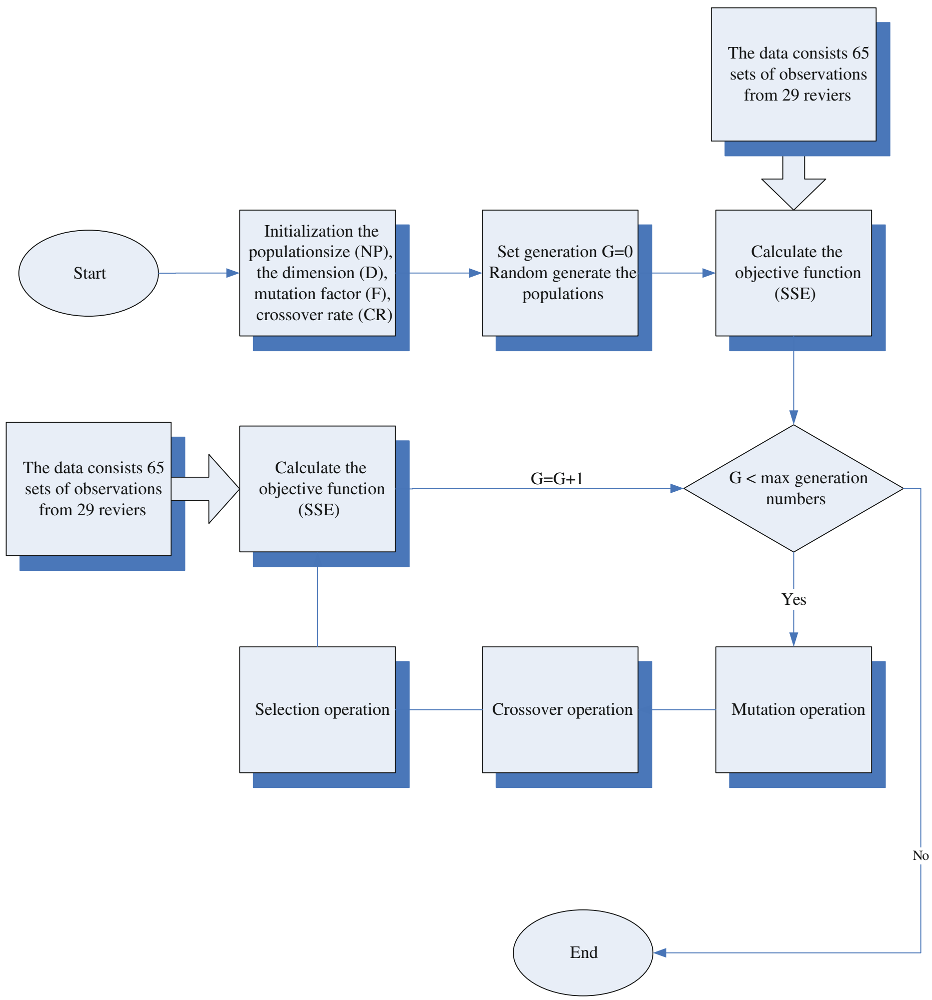

Fig. 1 The structure of DE for predicting longitudinal dispersion coefficients in natural rivers

in the expressions for estimation of longitudinal dispersion coefficient. Therefore the length of an individual is three. Then it needs to input the generated population into the longitudinal dispersion coefficient according to the measure data and calculate the fitness function value. The fitness function value employs the value of SSE. Then, we employ the mutation operator, crossover operator, selection operator to generate the offspring. In the last part, the algorithm needs to input the generated population into the longitudinal dispersion coefficient according to the measure data and calculate the fitness function value. The fitness function value employs the value of SSE. These operations are repeated until a stopping criterion is reached. In this paper, the criterion is the number of generation. Finally, the best *a*, *b*, *c* will be found in the final generation.

3 Experimental Results

In this paper, data consisting of 65 sets of observations from 29 rivers (Deng et al. 2001) in the unite states are used to evaluate the performance and applicability of the proposed differential evolution. The experiment data is shown in Table 2. As can be seen in Table 2, *W* is the width of the rivers. *H* is the depth of the rivers. *U* is the mean longitudinal velocity. *U** denotes the bottom shear friction velocity. The variables *W*, *H*, *U*, *U** are the important input variables. The software is written in Matlab-7.9 language and experiments are made on a Pentium 3.0 GHz Processor with 1.0 GB of memory. Parameters in genetic algorithm are selected as follows: population size is 200, crossover probability is 0.9 and the mutation probability of 0.002. The algorithm has a total string length of 30 bits because the algorithm uses the binary string length of 10 for each variable. The following parameters of differential evolution algorithm have been used after a number of careful experimentation: The population size is 200, the number of generation is 200, the scale factor *F* and crossover probability CR is 0.5, and 0.9, respectively. The dimension is three because the representation of variables is real code. In this paper, "DE/rand/1/bin" was used.

Parameter ranges must be determined before the differential evolution process. The dimensions of the following ranges are used for the DE similar to the GA: the upper and lower of the variable is 0, 1023, respectively.

3.1 DE for Longitudinal Dispersion Coefficient

In order to show the effective and efficient of the new expression for longitudinal dispersion coefficient, we compared our new expression with other previous expression on the 65 sets of data.

After carrying out the DE algorithm, the best value of *a*, *b*, and *c* are: $a=2.2820$, $b=0.7613$, $c=1.4713$. Therefore, with the results yielded by the DE algorithm, the new expressions for estimation of longitudinal dispersion coefficient can be expressed as follows:

$$\frac{K}{HU_*} = 2.2820 \times \left(\frac{W}{H}\right)^{0.7613} \left(\frac{U}{U_*}\right)^{1.4713}$$

Or

$$K = 2.2820 \times W^{0.7613} H^{1-0.7613} U^{1.4713} U^{1-1.4713} \tag{7}$$

Table 2 Measured longitudinal dispersion coefficients (LDC)

S.N	Stream	Width W(m)	Depth H(m)	Velocity U(m/s)	Sh.vel U*	W/H	U/U*	LDC Measure vale
1	Amita River	37.0	0.81	0.29	0.070	45.68	4.14	23.2
2		42.0	0.80	0.42	0.069	52.50	6.09	30.2
3	Antietam creek Md	12.8	0.30	0.42	0.057	42.67	7.37	17.5
4		24.1	0.98	0.59	0.098	24.59	6.02	101.5
5		11.9	0.66	0.43	0.085	18.03	5.06	20.9
6		21.0	0.48	0.62	0.069	43.75	8.99	25.9
7	Bayou Anacoco	20.0	0.42	0.29	0.045	47.62	6.44	13.9
8		17.5	0.45	0.32	0.024	38.89	13.33	5.8
9		25.9	0.94	0.34	0.067	27.55	5.07	32.5
10		36.6	0.91	0.40	0.067	40.22	5.97	39.5
11	Bayou Barthol. La	33.4	1.40	0.20	0.031	23.86	6.45	54.7
12	Bear creek, Colo	13.7	0.85	1.29	0.553	16.12	2.33	2.9
13	Cheattahoochee, Ga	75.6	1.95	0.74	0.138	38.77	5.36	88.9
14		91.9	2.44	0.52	0.094	37.66	5.53	166.9
15	Clinch River, Va	48.5	1.16	0.21	0.069	41.81	3.04	14.8
16		28.7	0.61	0.35	0.069	47.05	5.07	10.7
17		57.9	2.45	0.75	0.104	23.63	7.21	40.5
18		53.2	2.41	0.66	0.107	22.07	6.17	36.9
19	Comit River	13.0	0.26	0.31	0.044	50.00	7.05	7.0
20		16.0	0.43	0.37	0.056	37.21	6.61	13.9
21		15.7	0.23	0.36	0.039	68.26	9.23	69.0
22	Conoco. Creek, Md.	42.2	0.69	0.23	0.064	61.16	3.59	40.8
23		49.7	0.41	0.15	0.081	121.22	1.85	29.3
24		43.0	1.13	0.63	0.081	38.05	7.78	53.3
25	Copper Creep, Va	16.7	0.49	0.20	0.080	34.08	2.50	16.8
26		18.3	0.38	0.15	0.116	48.16	1.29	20.7
27		16.8	0.47	0.24	0.080	35.74	3.00	24.6
28		19.6	0.84	0.49	0.101	23.33	4.85	20.8
29	Difficult Run, Va	14.5	0.31	0.25	0.062	46.77	4.03	1.9
30	John Day River, Ore.	25.0	0.58	1.01	0.140	43.10	7.21	13.9
31		34.1	2.47	0.82	0.180	13.81	4.56	65.0
32	Little Pincy Creek	15.9	0.22	0.39	0.053	72.27	7.36	7.1
33	Minnesota River	80.0	2.74	0.03	0.002	29.20	14.17	22.3
34		80.0	2.74	0.14	0.010	29.20	14.43	34.9
35	Missouri River	183.0	2.33	0.89	0.066	78.54	13.48	465.0
36		201.0	3.56	1.28	0.084	56.46	15.24	837.0
37	Monocacy River, Md	48.7	0.55	0.26	0.052	88.55	5.00	37.8
38		93.0	0.71	0.16	0.046	130.99	3.48	41.4
39		51.2	0.65	0.62	0.044	78.77	14.09	29.6
40		97.5	1.15	0.32	0.058	84.78	5.52	119.8
41		40.5	0.41	0.23	0.040	98.78	5.75	66.5
42	Muddy River	13.0	0.81	0.37	0.081	16.05	4.57	13.9

Table 2 (continued)

S.N	Stream	Width W(m)	Depth H(m)	Velocity U(m/s)	Sh.vel U*	W/H	U/U*	LDC Measure vale
43		20.0	1.20	0.45	0.099	16.67	4.55	32.5
44	Nooksach River	86.0	2.93	1.20	0.530	29.35	2.26	153.0
45		64.0	0.76	0.67	0.268	84.21	2.50	34.8
46	Powell River, Tenn	36.8	0.87	0.13	0.054	42.30	2.41	15.5
47	Red River, La	161.5	3.96	0.29	0.060	40.78	4.83	130.5
48		152.4	3.66	0.45	0.057	41.64	7.89	227.6
49		155.1	1.74	0.47	0.036	89.14	13.06	177.7
50	Sabina River, La	116.4	1.65	0.58	0.054	70.55	10.74	131.3
51	Sabina River, La	14.2	0.50	0.13	0.037	28.40	3.51	12.8
52		12.2	0.51	0.23	0.030	23.92	7.67	14.7
53		21.3	0.93	0.36	0.035	22.90	10.29	24.2
54	Salt Creek, Nebr.	32.0	0.50	0.24	0.038	64.00	6.32	52.2
55	Susquehanna River	203.0	1.35	0.39	0.065	150.37	6.00	92.9
56	Tangipahoa River, La	31.4	0.81	0.48	0.072	38.77	6.67	45.1
57		29.9	0.40	0.34	0.020	74.75	17.00	44.0
58	Tickfau River, La	15.0	0.59	0.27	0.080	25.42	3.38	10.3
59	White River	67.0	0.59	0.35	0.044	113.56	7.95	30.2
60	Wind/big, Wyo	44.2	1.37	0.99	0.142	32.26	6.97	184.6
61		85.3	2.38	1.74	0.153	35.84	11.37	464.6
62		59.4	1.10	0.88	0.119	54.00	7.39	41.8
63		68.6	2.16	1.55	0.168	31.76	9.23	162.6
64	Yadkin River, N.C.	70.1	2.35	0.43	0.101	29.83	4.26	111.5
65		71.6	3.84	0.76	0.128	18.65	5.94	260.1
	Max	203	3.96	1.74	0.550	150.37	17.00	837.0
	Min	11.9	0.22	0.03	0.002	13.81	1.29	1.9
	Avg	53.8	1.24	0.49	0.090	48.86	6.75	80.5

The best experimental results obtained from the DE are compared with those obtained by using other expressions including Fischer et al. (1979), Liu (1977), Seo and Cheong (1998), Deng et al. (2001), Kashefipour and Falconer (2002) and GA model (Sahay and Dutta 2009) in terms of the value of SSE, in Tables 3 and 4, which shows that DE successes in finding the best solutions for the test methods.

3.2 Model Validation

In this section, we will show the effective of the proposed algorithm using different performance indices including root mean square errors (RMSE), coefficients of correlation (CC), average relative error to the measure value (ARE) and accuracy. These performance indices can be described as follows:

Table 3 Predicted longitudinal dispersion coefficients (LDC)

S.N	Longitudinal dispersion coefficients						
	Fisher (1979)	Liu (1977)	Seo and Cheong (1998)	Deng et al. (2001)	Kashefipour and Falconer (2002)	GA model (2009)	DE model
1	22.3	43.3	27.3	10.3	28.5	26.3	**19.2**
2	62.0	67.6	50.2	21.7	51.0	47.3	**36.7**
3	18.6	15.2	18.0	9.9	17.6	15.2	**12.8**
4	23.2	25.7	53.7	36.9	47.3	39.2	**35.2**
5	5.1	7.4	20.2	15.2	15.0	13.7	**12.6**
6	56.3	34.2	46.9	28.4	44.4	38.8	**34**
7	19.6	19.6	17.6	8.3	17.6	15.8	**12.7**
8	31.9	10.7	25.0	20.4	21.8	18.5	**18.1**
9	13.5	19.4	29.6	17.2	26.8	23.2	**19.5**
10	38.7	43.4	45.7	23.1	45.6	39.5	**32.1**
11	11.3	11.3	26.3	19.2	23.0	18.8	**17.2**
12	7.3	33.6	52.3	27.1	28.1	39.1	**30.9**
13	127.9	168.6	169.3	82.1	168.8	147.1	**117.6**
14	109.5	137.7	148.2	74.5	146.7	126.8	**102.6**
15	14.3	43.9	23.5	7.9	23.8	23.2	**16.1**
16	26.4	37.8	27.6	11.5	28.5	25.8	**19.6**
17	81.4	68.8	180.0	140.6	158.2	125.4	**118.1**
18	52.6	56.2	139.7	104.1	118.3	97.8	**90.3**
19	15.6	13.7	12.4	6.0	12.3	11.2	**9.1**
20	16.0	15.4	19.9	11.2	19.4	16.4	**13.9**
21	39.2	22.9	17.4	8.1	15.8	16.6	**13.4**
22	23.5	56.4	20.8	6.1	22.9	22.7	**15.1**
23	18.4	119.5	9.3	1.2	11.8	14.4	**7.2**
24	88.2	66.5	96.8	58.8	93.2	78.2	**68.2**
25	3.1	13.0	7.7	2.6	6.9	7.3	**5.1**
26	1.9	20.9	4.2	0.8	4.0	5.0	**2.8**
27	4.8	15.0	9.8	3.6	9.3	9.2	**6.6**
28	12.0	18.3	33.8	21.2	28.4	25.1	**21.7**
29	7.5	15.2	9.0	3.3	9.5	8.8	**6.4**
30	86.4	72.9	83.3	44.8	81.8	71.2	**59.5**
31	19.3	32.6	116.8	97.9	71.1	73.6	**69.8**
32	36.3	29.7	17.0	6.7	16.3	17.2	**13.1**
33	12.4	3.8	13.9	14.0	12.1	9.2	**8.1**
34	51.9	15.5	57.7	58.8	49.9	38.1	**41.4**
35	1897.5	627.0	559.2	296.7	438.6	524.2	**446.7**
36	2434.9	669.8	1055.3	736.9	848.4	865.2	**809.4**
37	61.7	90.3	27.2	7.6	28.2	31.7	**21.2**
38	74.6	188.1	23.6	4.2	25.8	33.4	**19.1**
39	387.6	119.9	110.9	60.3	85.6	103.3	**88.9**
40	160.5	202.7	71.0	21.5	72.3	80.1	**55.2**
41	58.2	69.1	20.4	5.8	20.1	24.0	**16.2**

Table 3 (continued)

S.N	Longitudinal dispersion coefficients						
	Fisher (1979)	Liu (1977)	Seo and Cheong (1998)	Deng et al. (2001)	Kashefipour and Falconer (2002)	GA model (2009)	DE model
42	3.9	6.5	19.0	14.5	12.8	12.6	**11.6**
43	7.5	12.7	35.0	26.0	24.1	23.5	**21.5**
44	75.4	362.4	240.0	84.5	195.8	221.2	**154.1**
45	99.3	411.1	69.7	13.5	82.7	90.3	**52.3**
46	5.4	23.5	9.9	2.9	9.9	10.3	**6.8**
47	101.6	156.4	133.0	58.9	134.6	119.7	**92.6**
48	248.0	182.9	238.3	138.0	230.3	198.1	**170**
49	933.2	323.7	235.3	113.3	181.4	231.6	**191.3**
50	562.7	261.6	219.1	109.1	189.0	206.2	**170.7**
51	2.0	5.0	5.2	2.4	4.6	4.4	**3.4**
52	5.7	4.4	11.9	9.5	10.5	8.2	**7.8**
53	19.9	9.9	37.5	36.5	32.7	24.2	**24.9**
54	34.1	35.2	20.6	8.0	20.7	20.6	**15.5**
55	785.7	874.8	150.2	33.5	135.1	202.8	**127**
56	42.8	40.7	50.1	27.5	49.1	41.8	**35.2**
57	142.1	33.2	39.3	24.5	28.7	34.7	**31.5**
58	3.8	10.1	11.8	5.7	9.7	9.6	**7.6**
59	233.0	170.0	55.8	17.4	48.5	65.2	**45.9**
60	108.3	96.2	158.9	100.3	151.0	123.8	**108.8**
61	665.5	283.9	638.4	499.8	577.8	472.4	**453.1**
62	229.6	186.8	160.0	76.0	156.7	147.0	**118.1**
63	342.7	200.1	437.8	327.8	406.1	322.8	**303.1**
64	42.1	78.4	91.3	45.7	83.9	75.6	**60.6**
65	66.3	75.0	227.2	183.9	177.2	151.1	**143.1**
Max	2434.9	874.8	1055.3	736.9	848.4	865.2	**809.4**
Min	1.9	3.8	4.2	0.8	4.0	1.4	**2.8000**
Avg	169.1	110.6	104.5	63.0	91.5	89.0	**75.6908**

Bold entries are the values of the differential evolution

$$
\begin{aligned}
RMSE &= \sqrt{\frac{\sum_{i=1}^{N}\left(K_{pre}-K_{meas}\right)^2}{N}} \\
CC &= \frac{\left[\sum_{1}^{N} K_{pre}K_{meas}-\sum_{1}^{N} K_{pre}\sum_{1}^{N} K_{meas}\right]}{N S_{pre} S_{meas}} \\
ARE &= \frac{\sum_{i=1}^{N}\left(\frac{K_{pre}-K_{meas}}{K_{meas}}\right)}{N}\times 100\ \% \\
\text{Accuracy} &= 0.5 < \frac{K_{pre}}{K_{meas}} < 1.5
\end{aligned}
\tag{8}
$$

Table 4 Comparison of performance indices of models

Model	Fitness value	RMSE	RMSE (K >100 m^{2}s^{-1} ignored)	CC	K_{pred}/K_{meas} range	ARE	Predicted Accuracy (%)
DE	**9.8691e+004**	**38.9658**	**27.0216**	**0.9556**	0.1350 to 10.6586	**33.06**	**56.92**
GA (2009)	1.3042e+005	44.7931	35.3009	0.9495	0.2406 to13.4828	37.81	53.85
Fisher (1979)	6.2241e+006	309.4445	120.2130	0.8661	0.0918 to13.0946	146.01	36.92
Liu (1977)	1.0880e+006	129.3777	131.0187	0.6589	0.1704 to 11.8132	79.38	41.54
Seo and Cheong (1998)	2.7379e+005	64.9010	41.3366	0.9524	0.2029 to 18.0345	48.65	46.15
Dong et al. (2001)	1.7139e+005	51.3492	29.1236	0.9298	0.0386 to 9.3448	44.01	49.32
Kashefipour and Falconer (2002)	1.5821e+005	49.3356	36.2527	0.9409	0.1932 to 9.6897	37.44	55.38

Bold results represent that our algorithm is better than other methods

Where S_{pre} and S_{meas} denote standard deviations of the measured and predicted values. The predicted value of longitudinal dispersion coefficient using the new expression and other previous expression is shown in Table 3. Following the Accuracy $= 0.5 < \frac{K_{pre}}{K_{meas}} < 1.5$, it shows that Accuracy =1 denotes exact matching between measured values and predicted values. While Accuracy >1, it denotes an overproduction value, otherwise, it represents an underproduction (Accuracy <1). Figure 2 shows a comparison between the predicted coefficients

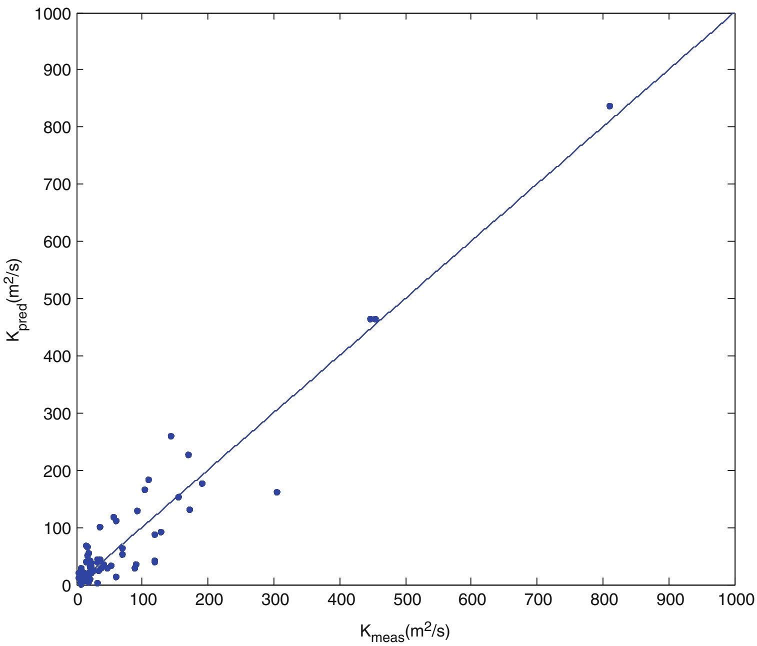

Fig. 2 Comparison of predicted dispersion coefficient from the new expression and the measure coefficient

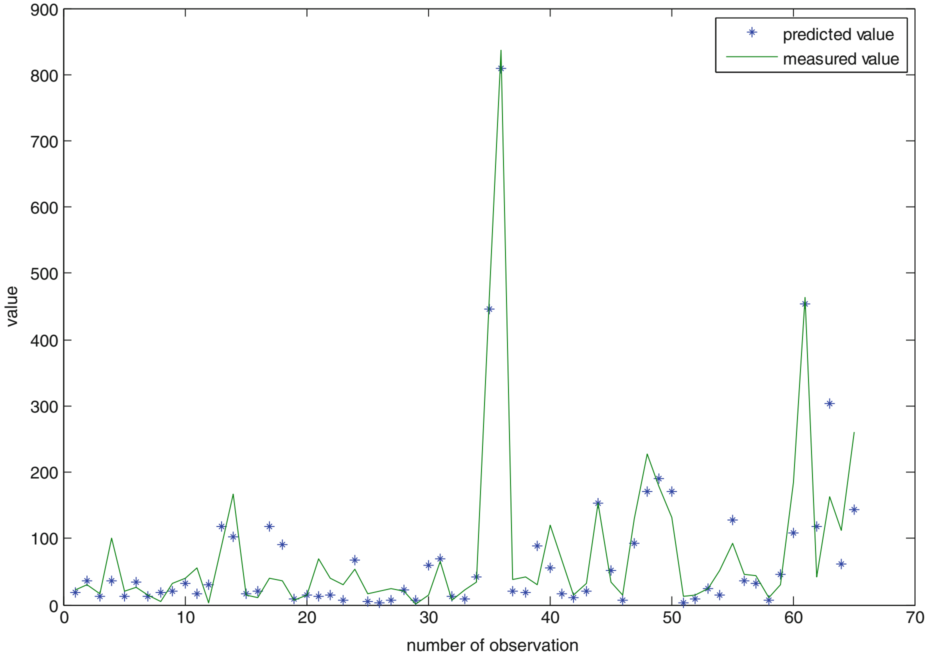

Fig. 3 Regression of the predicted value using new expression and the actual values

using the DE model and the actual values. The performance indices for different expressions are shown in Table 4. As can be seen in Table 4, we can find that the new expression using DE model is better than other expressions as RSME and ARE the lower and CC is the highest. Figure 3 shows the regression of the predicted value using new expression and the actual values.

From the Table 4, we can find that the Fisher expression give the most unsatisfactory relative to other expression. In the 4th column, when the measured value K value is larger 100 that are not included as the new data. We can find the RMSE of all models can be enhanced. In this new data, the new expression using DE model can also produce better solutions than other expressions. Table 4 also shows the average relative error to the measure value (ARE). As can be seen in Table 4, the new expression can obtain the best value of 33.06 % than other expressions. The Kashefipour and Falconer can give the second ranking of the ARE. Figure 4 show the average relative error to the measure value of different expression. From this figure, it can demonstrate the DE model is the best expression. The range of the $K_{\mathrm{pred}}/K_{\mathrm{meas}}$ also can be found in Table 4. Table 4 shows the $K_{\mathrm{pred}}/K_{\mathrm{meas}}$ value of the DE model ranges from 0.1350 to 10.6586, which trend in the positive direction.

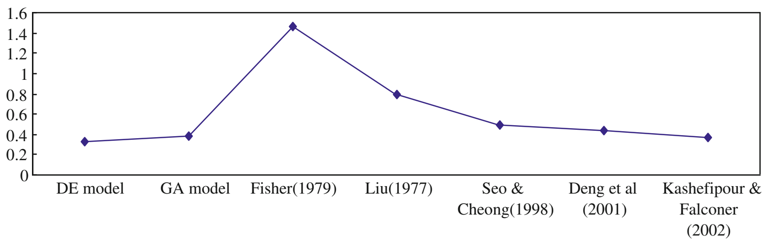

Fig. 4 The average relative error to the measure value of different expression

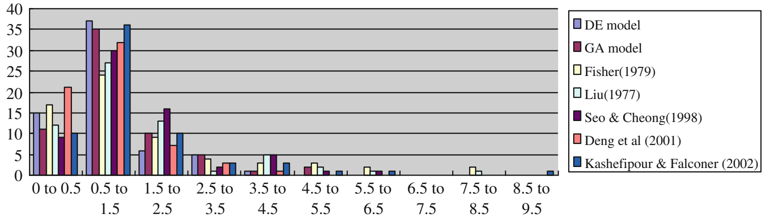

Fig. 5 Comparison of the percentage proportion of the predicted value with the different distribution

However, the positive of Fischer and Seo & Cheong's expressions are more satisfaction whereas the Dong's model trends the negative side. Figure 5 shows a comparison of the percentage proportion of the predicted value with the different distribution. From the figure, we can find that it shows a completed distribution for the value of the new expression.

In this paper, we defined the accuracy of a new expression is within the range $0.5 < \frac{K_{pre}}{K_{meas}} < 1.5$. The experiment results of the accuracy are also shown in Table 4. As can be seen in Table 4, we can find that the DE model obtains 56.92 % of the predicted accuracy from the new expression which is the highest among all expressions. Figure 6 shows the variation dispersion coefficient according to the width of river. It demonstrates that the predicted accuracy of DE enhances as the width increases.

4 Conclusions

This paper illustrated the application of differential evolution to produce a new expression for prediction of the longitudinal dispersion coefficient in natural rivers. We employ the

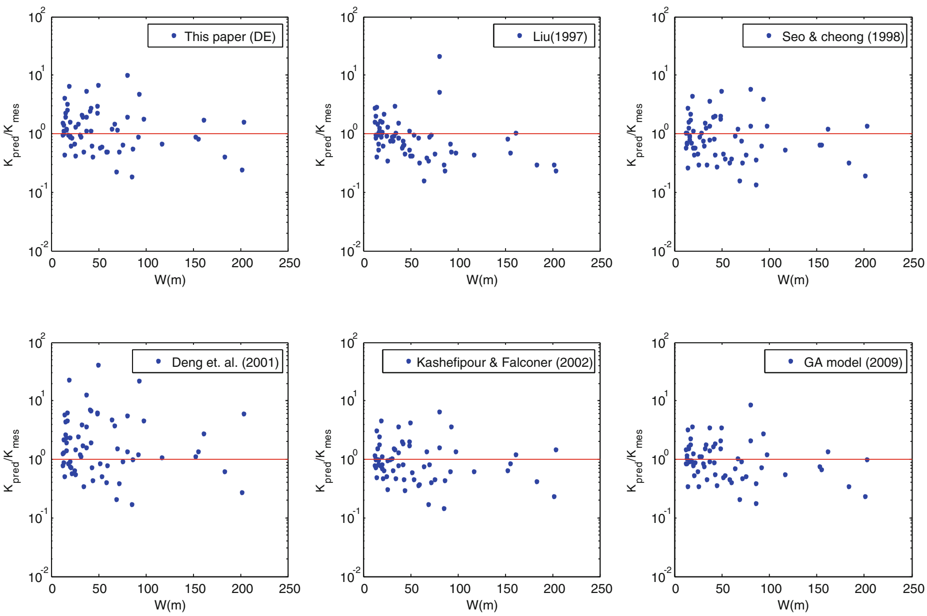

Fig. 6 The variation dispersion coefficient according to the width of river

mutation operator, crossover operator, selection operator to generate the offspring. The effectiveness of the proposed algorithm is demonstrated on data consisting of 65 sets of observations from 29 rivers in the unite states. The DE algorithm has the ability to find the least root mean square error, the highest coefficient of correlation, the least average relative error to the measure value, and the highest accuracy, and has better convergence characteristics and computational efficiency. The comparison of the results with other methods reported in the literature show the superiority of the proposed method and its potential for prediction of the longitudinal dispersion coefficient. From the results obtained, it can be concluded DE algorithm is a promising technique for prediction of the longitudinal dispersion coefficient. In the future work, we will use other enhance DE algorithms for prediction of the longitudinal dispersion coefficient.

Acknowledgments This research is fully supported by Opening Fund of Top Key Discipline of Computer Software and Theory in Zhejiang Provincial Colleges at Zhejiang Normal University under Grant No. ZSDZZZZXK37 and the Fundamental Research Funds for the Central Universities Nos. 11CXPY010. Guangxi Natural Science Foundation (No. 2013GXNSFBA019263), Science and Technology Research Projects of Guangxi Higher Education (No.2013YB029), Scientific Research Foundation of Guangxi Normal University for Doctors.

References

Brest J, Greiner S, Boskovic B, Mernik M, Zumer V (2006) Self adapting control parameters in differential evolution: a comparative study on numerical benchmark problems. IEEE Trans Evol Comput 10(6):647–657

Deng ZQ, Singh VP, Bengtsson L (2001) Longitudinal dispersion coefficient in straight rivers. J Hydraulic Eng 127(11):919–927

Elder JW (1959) The dispersion of a marked fluid in turbulent shear flow. J Fluid Mech 5(4):544–560

Fischer BH (1967) The mechanics of dispersion in natural streams. J Hydr Div ASCE 93(6):187–216

Fischer BH, List EJ, Koh RCY, Imberger J, Brooks NH (1979) Mixing in inland and coastal waters. Academic, New York, pp 104–138

Kashefipour SM, Falconer RA (2002) Longitudinal dispersion coefficients in natural channels. Water Res 36(6):1596–1608

Li X, Yin M (2012a) Hybrid differential evolution with artificial bee colony and its application for design of a reconfigurable antenna array with discrete phase shifters. IET Microw, Antennas Propag 6(14):1573–1582

Li X, Yin M (2012b) Application of differential evolution algorithm on self-potential data. PLoS One 7(12): e51199

Li X, Yin M (2013) An opposition-based differential evolution algorithm for permutation flow shop scheduling based on diversity measure. Adv Eng Softw 55:10–31

Li X, Wang J, Zhou J, Yin M (2011) A perturb biogeography based optimization with mutation for global numerical optimization. Appl Math Comput 218(2):598–609

Liu H (1977) Predicting dispersion coefficient of streams. J Environ Eng Div 103(1):59–69

Neri F, Tirronen V (2009) Scale factor local search in differential evolution. Memet Comput J 1(2):153–171

Qin AK, Huang VL, Suganthan PN (2009) Differential evolution algorithm with strategy adaptation for global numerical optimization. IEEE Trans Evol Comput 13(2):398–417

Qin H, Zhou JZ, Lu YL, Li YH, Zhang YC (2010) Multi-objective cultured differential evolution for generating optimal trade-offs in reservoir flood control operation. Water Resour Manag 24(11):2611–2632

Rahnamayan S, Tizhoosh HR, Salama MMA (2008) Opposition-based differential evolution. IEEE Trans Evol Comput 12(1):64–79

Sahay RR, Dutta S (2009) Prediction of longitudinal dispersion coefficients in natural rivers using genetic algorithm. Hydrol Res 40(6):544–552

Seo IW, Cheong TS (1998) Predicting longitudinal dispersion coefficient in natural streams. J Hydraulic Eng 124(1):25–32

Storn R, Price K (1997) Differential evolution-a simple and efficient heuristic for global optimization over continuous space. J Glob Optim 11(1):341–359

Taylor GI (1953) Dispersion of soluble matter in solvent flowing slowly through a tube. Proc R Soc Lond Ser A 219:186–203

Taylor GI (1954) The dispersion of matter in turbulent flow through a pipe. Proc R Soc Lond Ser A 223:446–468

Vasan A, Raju KS (2007) Application of differential evolution for irrigation planning: an Indian case study. Water Resour Manag 21(8):1393–1407

Zhang JQ, Sanderson AC (2009) JADE: adaptive differential evolution with optional external archive. IEEE Trans Evol Comput 13(5):945–958