Estimation of longitudinal dispersion coefficients in meandering streams using soft computing techniques

Ahmad Fathi[1*], Abbas Parsaie[2] and Mostafa Moghadam[3]

1- *Corresponding author: Assistant professor, Faculty of Water and Environmental Engineering, Shahid Chamran University of Ahvaz, Ahvaz, Iran. Email: a.fathi@scu.ac.ir
2- Assistant professor, Faculty of Water and Environmental Engineering, Shahid Chamran University of Ahvaz, Ahvaz, Iran.
3- M.Sc, Faculty of Water and Environmental Engineering, Shahid Chamran University of Ahvaz, Ahvaz, Iran.

ARTICLE INFO

Artlcle history:
Received:
Revised:
Accepted:
Available online:

Keywords:
Longitudinal dispersion coefficient; Meandering Rivers; Soft computing; GMDH model; Artificial Neural Networks

TO CITE THIS ARTICLE:

Abstract

Modeling and estimating the longitudinal dispersion coefficient (LDC) in rivers is crucial for reducing contamination and improving river water quality. In this study, the LDC in meandering rivers was modeled and estimated using soft-computing techniques, including the Group Method of Data Handling (GMDH) and Artificial Neural Network (ANN) methods. For this purpose, flow depth, channel width, mean velocity, shear velocity, and river path sinuosity factor were used as input parameters. Simulation results indicated that the GMDH model with two hidden layers—five neurons in the first and three in the second—predicted the longitudinal dispersion coefficient with satisfactory statistical accuracy (R^2=0.91 and RMSE=40.26). Similarly, the ANN model with a two hidden layers containing 10 and five neurons in the first and second hidden, respectively and a logarithmic sigmoid activation function provided reliable predictive performance (R^2 = 0.90 and RMSE=42.51). Analysis of the GMDH model structure revealed that the ratio of mean velocity to shear velocity is dominant in predicting the longitudinal dispersion coefficient in meandering rivers.

Introduction

Studying and modeling river water quality are crucial for protecting the aquatic environment and reducing the effects of various river pollutants on human health. The longitudinal dispersion coefficient (LDC) is one of the most important parameters used to model the propagation of pollution in rivers (Ramezani et al., 2019). To control contamination in rivers, the release of contaminants must be reasonably regulated. The self-purification capacity of rivers and other surface streams to disperse pollutants in longitudinal, transverse, and vertical directions is expressed by dispersion coefficients. The LDC can be used to model pollution transport in rivers and to estimate downstream pollution loads (Huai et al., 2018). This parameter plays a crucial role in the management and regulation of wastewater discharge into rivers. Understanding the mechanisms of pollution transport and dispersion in rivers with different geometries is essential for developing effective strategies to mitigate pollution impacts on public health and to enhance rivers' self-purification capacity (Cazelles et al., 1991). El Kadi Abderrezzak et al. (2015) and Parsaie and Haghiabi (2017a) investigated the role of the LDC in numerical modeling of pollution transmission in rivers. They stated that accurate prediction of the LDC significantly improves the accuracy of numerical simulation. The longitudinal dispersion coefficient, which depends on hydraulic flow variables (such as mean depth and velocity) in the river, fluid properties (such as viscosity), sediment transport (such as suspension), and river geometry (such as cross-sectional shape and path curvature), is highly sensitive and variable. It is noteworthy

that the LDC will vary greatly with the slightest change in river geometry and in river hydraulic conditions, so, given the complex and variable river conditions, there is considerable uncertainty in estimating the longitudinal dispersion coefficient (Zeng and Huai., 2014).

Field studies and tracer experiments are used to accurately estimate the longitudinal dispersion coefficient. The routing method proposed by Fischer for determining the longitudinal propagation coefficient from tracking data has been widely used to date. This method is based on the Frozen-cloud approximation, which requires a complex numerical solution.

In studies by Atkinson and Davis (2000), Davis and Atkinson (2000) and Davis et al. (2000) a case study of the River Severn used Rhodamine WT as a tracer to investigate the propagation of contamination. These tests are difficult, time-consuming, and costly, and the estimated longitudinal dispersion coefficient is valid only for the stream studied and the hydraulic conditions under which the tracer test was performed. If river conditions change, the test should be repeated.

Singh and Beck (2003) proposed a method for determining the longitudinal dispersion coefficient using tracking data and presented a relatively accurate solution to the one-dimensional advection-dispersion equation. They stated that this method has no Fisher-approximation error or numerical-resolution error.

The a study by Palancar et al. (2003), based on trace data and hydraulic flow parameters in two Spanish rivers, indicates that these two methods are more accurate at estimating the LDC than other methods.

Because tracking experiments in practical and engineering applications are time-consuming, difficult, and costly, empirical formulas based on hydraulic and geometric parameters have been proposed.

Reports of the limited accuracy of empirical formulas have been published for other rivers under varying conditions. These formulas are usable only within the range for which they are fitted and are inaccurate outside it. Therefore, investigators have turned to soft computing methods to develop a comprehensive model for predicting the LDC. In this regard, artificial neural network (ANN) (Kashefipour and Falconer, 2002; Noori et al., 2016), radial basis function neural network (RBF) (Parsaie and Haghiabi, 2017b), Genetic Programming (GP) (Azamathulla and Ghani, 2011), Support vector machine (SVM) (Azamathulla and Wu, 2011), adaptive neuro fuzzy inference system (ANFIS) (Noori et al., 2016; Riahi-Madvar et al., 2009), group method of data handling (GMDH) (Najafzadeh and Sattar, 2015a), gene expression programing (GEP) (Sattar and Gharabaghi, 2015), and Multivariate Adaptive Regression Splines (MARS) (Haghiabi, 2016, 2017) have been reported. Based on these reports, the accuracy of soft computing methods is much higher than that of the empirical formulas.

It appears that modeling and predicting the contaminant diffusion coefficient in rivers based on sinusoidal flow rates requires further study. For this purpose, this study aims to develop soft-computation models, including GMDH and artificial neural networks, to model and predict LDC in rivers in relation to the degree of seismicity along the river's path.

Material and Methods

The advection-dispersion equation (ADE) governs pollutant transport in rivers. Equation (1) presents the one-dimensional ADE without any source or sink terms (Chao et al. 2007).

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

Where C is the concentration of pollution, D_L is the Longitudinal dispersion coefficient, and u is the mean velocity of the river flow(Benedini and Tsakiris, 2013). Various factors affect the LDC, including hydraulic, fluid, and geometric parameters. Most of them are gathered in equation (2).

$$D_L = f_1(\rho, \mu, u, u_*, h, w, s_n) \quad (2)$$

Where D_L is the LDC, ρ is the fluid density; μ is the dynamic viscosity; u is the mean flow velocity, w is the cross-sectional area width; h is the flow depth; u_* is the flow shear velocity, and s_n is a sinusoidal factor.

To derive the dimensionless parameter, the dimensional analysis technique based on Buckingham's theory was used (Seo and Cheong, 1998).

$$\frac{D_L}{\mathrm{hu}_*} = f_2\left(\frac{u}{u_*}, \frac{w}{h}, s_n\right) \qquad (3)$$

The preparation of soft computing models is based on data sets; therefore, about 61 data sets related to equation (3) were collected, and their ranges are listed in Table (1) (Sahay, 2013).

Table 1- collected data related to LDC in meandering streams (adapted from Sahay (2013))

w	H	V	u*	sin	w/h	u/u*	DL
12.8	0.3	0.42	0.057	1.4	42.67	7.37	17.5
24.1	0.98	0.59	0.098	2.25	24.59	6.02	101.5
11.9	0.66	0.43	0.085	2.25	18.03	5.06	20.9
21	0.48	0.62	0.069	1.26	43.75	8.99	25.9
20	0.42	0.29	0.045	1.41	47.62	6.44	13.9
17.5	0.45	0.32	0.024	1.41	38.89	13.33	5.8
25.9	0.94	0.34	0.067	1.41	27.55	5.07	32.5
36.6	0.91	0.4	0.067	1.41	40.22	5.97	39.5
33.4	1.4	0.2	0.031	2.46	23.86	6.45	54.7
13.7	0.85	1.29	0.553	1.08	16.12	2.33	2.9
75.6	1.95	0.74	0.138	1.27	38.77	5.36	88.9
91.9	2.44	0.52	0.094	1.57	37.66	5.53	166.9
48.5	1.16	0.21	0.069	1.25	41.81	3.04	14.8
28.7	0.61	0.35	0.069	1.14	47.05	5.07	10.7
57.9	2.45	0.75	0.104	1.14	23.63	7.21	40.5
53.2	2.41	0.66	0.107	1.14	22.07	6.17	36.9
13	0.26	0.31	0.044	1.31	50	7.05	7
16	0.43	0.37	0.056	1.31	37.21	6.61	13.9
15.7	0.23	0.36	0.039	1.31	68.26	9.23	69
42.2	0.69	0.23	0.064	2.25	61.16	3.59	40.8
49.7	0.41	0.15	0.081	2.25	121.22	1.85	29.3
43	1.13	0.63	0.081	1.31	38.05	7.78	53.3
16.7	0.49	0.2	0.08	2.54	34.08	2.5	16.8
18.3	0.38	0.15	0.116	2.54	48.16	1.29	20.7
16.8	0.47	0.24	0.08	2.54	35.74	3	24.6
19.6	0.84	0.49	0.101	1.26	23.33	4.85	20.8
14.5	0.31	0.25	0.062	1.09	46.77	4.03	1.9
25	0.58	1.01	0.14	1.08	43.1	7.21	13.9
34.1	2.47	0.82	0.18	1.89	13.81	4.56	65
15.9	0.22	0.39	0.053	1.13	72.27	7.36	7.1
183	2.33	0.89	0.066	1.35	78.54	13.48	465
197	3.11	1.53	0.078	1.35	63.34	19.62	892
201	3.56	1.28	0.084	1.35	56.46	15.24	837
48.7	0.55	0.26	0.052	1.28	88.55	5	37.8
93	0.71	0.16	0.046	1.28	130.99	3.48	41.4
51.2	0.65	0.62	0.044	1.28	78.77	14.09	29.6
97.5	1.15	0.32	0.058	1.61	84.78	5.52	119.8
40.5	0.41	0.23	0.04	1.61	98.78	5.75	66.5
86	2.93	1.2	0.53	1.3	29.35	2.26	153
64	0.76	0.67	0.268	1.3	84.21	2.5	34.8
36.8	0.87	0.13	0.054	2.2	42.3	2.41	15.5
253.6	1.62	0.61	0.032	1.2	156.54	19.06	143.8
161.5	3.96	0.29	0.06	1.44	40.78	4.83	130.5
152.4	3.66	0.45	0.057	1.44	41.64	7.89	227.6
155.1	1.74	0.47	0.036	1.24	89.14	13.06	177.7
116.4	1.65	0.58	0.054	1.19	70.55	10.74	131.3
160.3	2.32	1.06	0.054	1.17	69.09	19.63	308.9
14.2	0.5	0.13	0.037	2.53	28.4	3.51	12.8
12.2	0.51	0.23	0.03	2.05	23.92	7.67	14.7
21.3	0.93	0.36	0.035	1.47	22.9	10.29	24.2
32	0.5	0.24	0.038	1.38	64	6.32	52.2
203	1.35	0.39	0.065	1.13	150.37	6	92.9
31.4	0.81	0.48	0.072	1.46	38.77	6.67	45.1
29.9	0.4	0.34	0.02	1.46	74.75	17	44
15	0.59	0.27	0.08	1.75	25.42	3.38	10.3
44.2	1.37	0.99	0.142	1.56	32.26	6.97	184.6
85.3	2.38	1.74	0.153	1.56	35.84	11.37	464.6
59.4	1.1	0.88	0.119	1.18	54	7.39	41.8
68.6	2.16	1.55	0.168	1.18	31.76	9.23	162.6
70.1	2.35	0.43	0.101	2.17	29.83	4.26	111.5
71.6	3.84	0.76	0.128	2.17	18.65	5.94	260.1

Empirical formula

To date, many empirical formulas have been proposed to estimate the LDC. Some of the well-known LDC formulas are listed in Table (2).

Examination of the formulas presented in Table (2) shows that all of them have used hydraulic parameters, such as mean velocity (u) and shear velocity (u*), and geometrical parameters, such as flow depth (h) and flow surface width (w), to estimate LDC.

Table 2- Empirical equations for estimating the longitudinal dispersion coefficient

Row	Author	Equation
1	Elder (1959)	$D_L = 5.93 h u_*$
2	(McQuivey and Keefer, 1974)	$D_L = 0.58 \left(\frac{h}{u_*}\right)^2 uw$
3	(Fischer, 1968)	$D_L = 0.011 \frac{u^2 w^2}{h u_*}$
4	(Liu, 1977a)	$D_L = 0.55 \frac{w u_*}{h^2}$
5	(Liu, 1977b)	$D_L = 0.18 \left(\frac{u}{u_*}\right)^{0.5} \left(\frac{w}{h}\right)^2 h u_*$
6	(Iwasa, 1991)	$D_L = 2 \left(\frac{w}{h}\right)^{1.5} h u_*$
7	Seo and Cheong (1998)	$D_L = 5.92 \left(\frac{u}{u_*}\right)^{1.43} \left(\frac{w}{h}\right)^{0.62} h u_*$
8	(Koussis and Rodríguez-Mirasol, 1998)	$D_L = 0.6 \left(\frac{w}{h}\right)^2 h u_*$
9	Li et al. (1998)	$D_L = 5.92 \left(\frac{u}{u_*}\right)^{1.2} \left(\frac{w}{h}\right)^{1.3} h u_*$
10	(Deng et al., 2001)	$\frac{D_L}{h u_*} = \frac{0.15}{8\varepsilon_t} \left(\frac{w}{h}\right)^{5/3} \left(\frac{u}{u_*}\right)^2$ $\varepsilon_t 0.145 + \left(\frac{1}{3520.0}\right) \left(\frac{w}{h}\right)^{1.38} \left(\frac{u}{u_*}\right)$
11	Kashefipour and Falconer (2001)	$\frac{D_L}{h u_*} = 10.612 \left(\frac{u}{u_*}\right)^2$
12	Tavakollizadeh and Kashefipour (2007)	$D_L = 7.428 + 1.775 \left(\frac{u}{u_*}\right)^{1.752} \left(\frac{w}{h}\right)^{0.62} hu$
13	Rajeev and Dutta (2009)	$D_L = 10.612 \left(\frac{u}{u_*}\right) hu$
14	(Sahay and Dutta, 2009)	$\frac{D_L}{h u_*} = 2 \left(\frac{w}{h}\right)^{0.96} \left(\frac{u}{u_*}\right)^{1.25}$

Group Method of Data Handling

The Group Method of Data Handling (GMDH) approach was introduced by Ivakhnenko (1971). GMDH is based on algebraic operations such as summation, multiplication, sorting, crossbreeding, and selection. The idea of crossbreeding and selection is similar to that of a genetic algorithm. Like other types of neural networks, the GMDH may involve one or more hidden layers (Figure 1). The GMDH is a flexible algorithm that can be hybridized with other metaheuristic algorithms such as genetic algorithms, particle swarm optimization (PSO), genetic programming, backpropagation, and Levenberg-Marquardt (Najafzadeh and Sattar, 2015; Najafzadeh et al., 2015; Najafzadeh and Barani, 2011). The GMDH method has been used to model hydraulic and hydrological phenomena. The GMDH is based on the Kolmogorov–Gabor polynomial, given in equation (4).

$$y = w_0 + \sum_{i=1}^{n} w_i x_i + \sum_{i=1}^{n}\sum_{j=1}^{m} w_{ij} x_i x_j + \sum_{i=1}^{n}\sum_{j=1}^{m}\sum_{k=1}^{k} w_{ij} x_i x_j x_k \quad (4)$$

In the above equation, is the weight and is a constant. To avoid system complexity, only the quadratic term of the Kolmogorov–Gabor polynomial is used. Therefore, to predict the desired output from the inputs, the following quadratic polynomial is used. In other words, equation (5) serves as the transfer function of neurons in the GMDH method.

$$\overline{y} = G(x_i, x_j) = w_0 + w_1 x_i + w_2 x_j + w_3 x_i^2 + w_4 x_j^2 + w_5 x_i x_j \quad (5)$$

The number of pairs is calculated using equation (6).

$$\frac{N_{inp}(N_{inp}-1)}{2} \quad (6)$$

In the GMDH preparation, at the first stage of development, each input pair is passed through equation (5). presents a schematic of the GMDH development process.

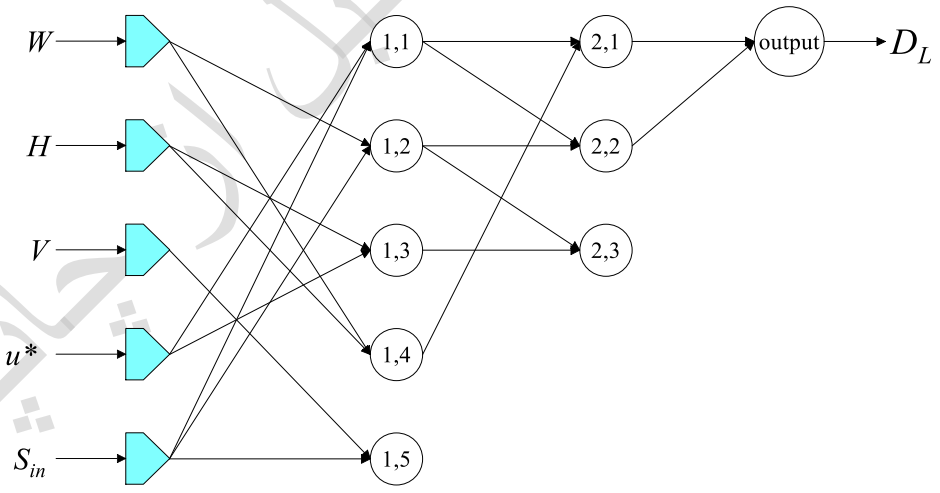


Fig. 1- The GMDH model was developed to estimate LDC in meandering streams

Unlike Artificial Neural Networks, which use hidden layers of artificial neurons, the GMDH network is built from sequential polynomial layers. Each layer contains a set of quadratic polynomials (partial descriptors) that combine pairs of input variables. Only the best-performing polynomials, selected by a criterion such as minimum prediction error, are retained and passed to the next layer. This crossbreeding and selection process continues until the model's performance no longer improves. In the GMDH architecture, each partial descriptor uses the second-order Kolmogorov-Gabor polynomial (Equation 5) as its transfer function. The polynomial form is fixed and does not vary across neurons; however, each polynomial's coefficients are independently optimized during training to best fit the relationship between the selected input pair and the target output. This combination of a fixed polynomial structure and data-driven coefficient optimization enables GMDH to model complex nonlinear relationships while maintaining interpretability.

Artificial neural network

The Multilayer Perceptron (MLP) is a feedforward artificial neural network widely used to model complex nonlinear relationships between input and output variables. It consists of an input layer, one or more hidden layers, and an output layer, with each layer comprising several processing units called neurons. Each neuron receives a weighted sum of its inputs and applies a nonlinear activation function to produce an output. Mathematically, the output of a neuron in the jth layer can be expressed as:

$$f\left(\sum_{i=1}^{n} w_{ij} x_i + b_j\right) = h_j \qquad (7)$$

where xi represents the input signals, wij denotes the connection weights, bj is the bias term, and f is a nonlinear activation function such as the sigmoid or hyperbolic tangent. The network training process involves adjusting the weights and biases to minimize the error between the predicted and observed outputs, typically using backpropagation with gradient descent. The MLP's strength lies in its universal approximation capability, enabling it to capture complex, nonlinear relationships between the independent (input) and dependent (output) variables. The optimal network architecture, including the number of hidden neurons, learning rate, and activation functions, was determined through trial and error and statistical performance evaluation (e.g., RMSE, R^2).

Results and discussion

Empirical formula evaluation

In this study, 14 empirical formulas for estimating the longitudinal dispersion coefficient (DL) in sinusoidal flows were evaluated using 61 experimental data points (Table 3). The evaluation criteria included RMSE, and coefficient of determination (R^2). The results showed that the formula proposed by Rajeev & Dutta (2009) was the most accurate, with the lowest error (RMSE = 60.68 m^2/s) and the high (R^2 = 0.89). The formula proposed by Seo and Cheong (1998)was ranked second the most accurate due to its higher RMSE (110.90). In contrast, older formulas, such as Elder (1959) and McQuivey & Keefer (1974), systematically underestimated or overestimated DL values and are not recommended for meandering flows. Analysis of the dependence of the error on the depth-to-width ratio (h/w) showed that the three best formulas had the lowest correlation with this geometric parameter, indicating their greater reliability over a wider range of hydraulic conditions. These findings emphasize that selecting an appropriate formula for estimating DL should take into account accuracy, computational complexity, and the specific conditions under study.

As mentioned, the purpose of this research is to estimate LDC in rivers using soft-computing models, including ANNs and GMDHs. These models are data-driven, so the first step in preparing and developing them is data preparation. The histogram of the collected data is shown in Figure (2). In this figure, the lines for the median and the mean are shown, as well.

Table 3- Statistical indices of empirical formula for the estimation of LDC

Row	Author(s)	RMSE[m^2/s]	R^2
1	(Elder, 1959)	199.33	0.08

2	McQuivey & Keefer (1974)	199.61	0.01
3	Fisher (1967)	867.44	0.47
4	Li et al. (1998)	199.31	0.07
5	Liu (1977)	173.55	0.44
6	Iwasa & Aya (1991)	146.13	0.29
7	Seo & Cheong (1998)	110.90	0.90
8	Koussis & Rodríguez-Mirasol (1998)	232.23	0.22
9	Li et al. (1998)	2414.87	0.79
10	Deng et al. (2001)	1433.07	0.28
11	Kasheffpour & Falconer (2001)	1090.51	0.37
12	Tavakollizadeh & Kasheffpour (2007)	3430.72	0.78
13	Rajeev & Dutta (2009)	60.68	0.89
14	Sahay & Dutta (2009)	1931.59	0.17

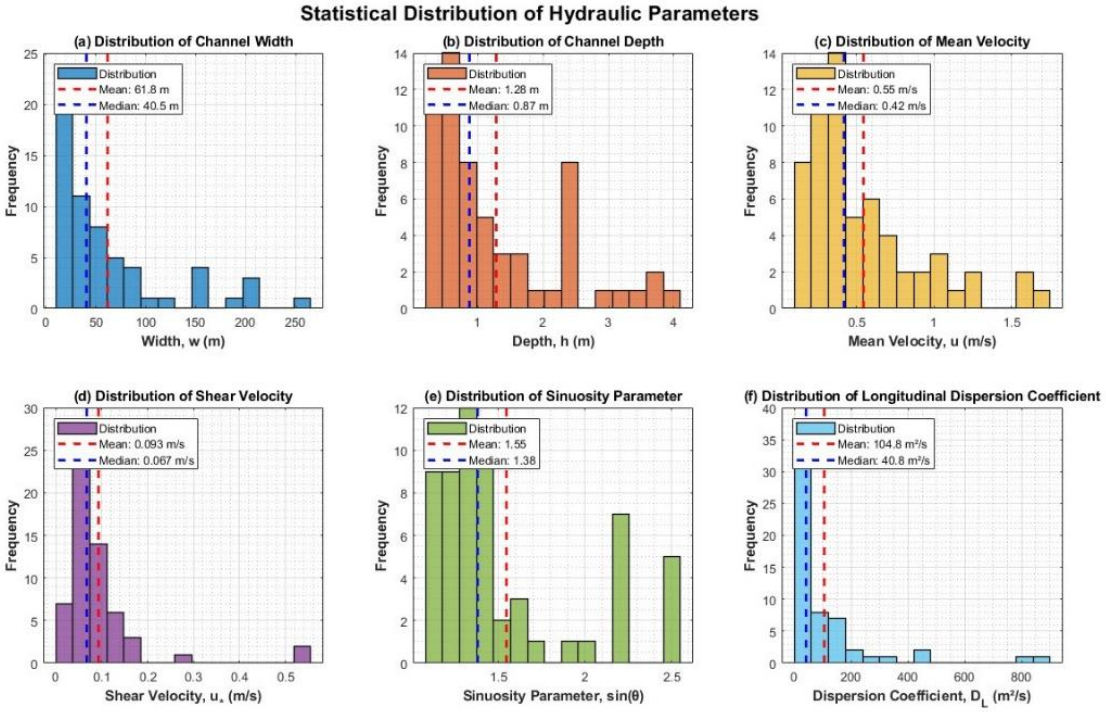


Fig. 2- Histogram of gathered data related to LDC

The purpose of data preparation is to collect, statistically review, and monitor data. The collected data total 61, and their statistical characteristics are presented in Table (1). The correlation between the D_L (LDC) and involved parameters is given in Table (4) and Figure (3). The distribution of LDC across the involved parameters is shown in Figure (4).

Table 4- Correlation between involved factors in the LDC in the meandering streams

Correlations	w	H	V	u*	sin	w/h	u/u*	DL
w	1/00							
H	0/67	1/00						
V	0/38	0/55	1/00					
u*	-0/05	0/22	0/54	1/00				
s_{in}	-0/29	-0/11	-0/37	-0/13	1/00			
w/h	0/54	-0/15	-0/16	-0/26	-0/26	1/00		

u/u*	0/55	0/27	0/43	-0/31	-0/38	0/29	1/00	
DL	0/69	0/66	0/65	0/03	-0/11	0/10	0/58	1/00

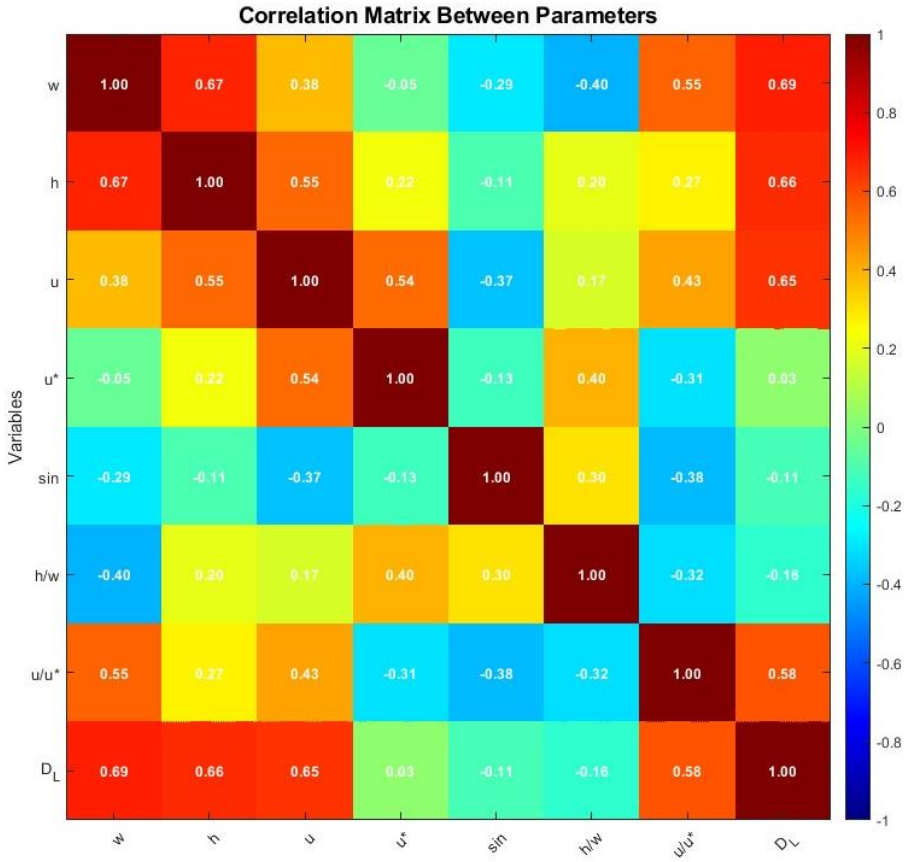


Fig. 3- Heat map of the correlation between LDC and involved parameters

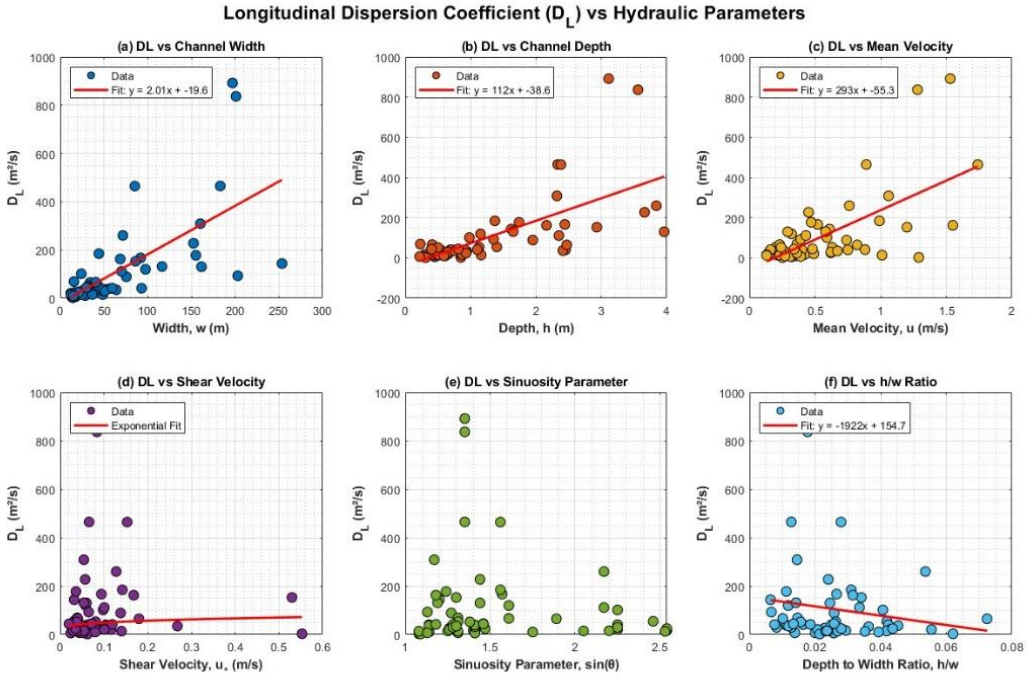


Fig. 4- Distribution of LDC versus the involved parameters

Based on the in-depth analysis of 61 field datasets from sinuous streams, the hydraulic and geometric parameters reveal four distinct hydraulic regimes: 1) Wide, shallow channels with low DL ($w > 100$ m, $h < 2$ m, DL < 150 m²/s), 2) Deep channels with moderate dispersion ($h > 2$ m, DL = 40–260 m²/s), 3) High-velocity turbulent flows ($u > 1$ m/s, $u^* > 0.1$ m/s, DL > 150 m²/s), and 4) Low-velocity flows with negligible dispersion ($u < 0.3$ m/s, DL < 30 m²/s). The pronounced right-skewness of the DL distribution (skewness coefficient = 2.8) indicates that extreme dispersion values are primarily generated under specific hydraulic conditions, with only 15% of the data (9 points) accounting for 65% of the total DL range. Multivariate analysis shows that the combined parameter $w \times u \times h$ has greater explanatory power ($R^2 = 0.89$) than any single variable, highlighting the multiplicative nature of dispersion processes. The velocity ratio u/u^* (turbulence indicator) exhibits a strong nonlinear correlation with DL—showing an approximately linear relationship ($R = 0.78$) for values below 10, but reaching saturation at higher values. The sinuosity parameter (sin), in addition to its overall negative correlation, shows the greatest variability in DL within the 1–2 range (moderately sinuous channels), indicating heightened system sensitivity to geometric changes in this interval. Conditional variance analysis reveals that prediction uncertainty is highest for channels with h/w ratios between 0.02 and 0.06 (DL variation up to 300 m²/s), whereas very wide ($h/w < 0.01$) or very deep ($h/w > 0.1$) channels exhibit more stable behavior. Cluster analysis segregates the data into three main groups: low-dispersion cluster (DL < 50 m²/s, 68% of data), moderate-dispersion cluster (DL = 50–200 m²/s, 23% of data), and high-dispersion cluster (DL > 200 m²/s, 9% of data)—each displaying distinct parametric relationships with the independent variables. Scale dependency is evident across the dataset: at small scales ($w < 50$ m), depth is the dominant factor ($R = 0.71$), whereas at large scales ($w > 100$ m), flow velocity is the primary driver ($R = 0.82$). Data anomalies include two exceptionally high DL values (892 and 837 m²/s) that deviate significantly from the linear model but align well with an exponential model, suggesting a threshold for transition into a super-dispersion regime. These findings not only confirm basic relationships but also unveil deeper layers of parametric interactions, scale dependencies, and distinct hydraulic regimes that are essential for developing more advanced, context-sensitive DL estimation models.

Development of ANN

In this section, the preparation and development of an ANN model for estimating LDC in meandering rivers are discussed. The development of soft computing models, including ANN, is based on collected data. Therefore, the first step in preparing and developing soft computing models is to split the data into two categories: training (80%) and testing (20%). The training data is used for calibration (training), and the testing data is used to verify the developed model. The purpose of preparing and developing the MLPNN model is to determine its internal structure, including the number of layers, the number of neurons, and the training method. To determine the network's internal architecture, a single-layer network with seven neurons, equal to the input variables, was initially constructed. After training the model, the network's efficiency was evaluated. In the formed perceptron, various types of well-known activation functions were examined, including: linear transfer function (purelin), radial basis function (radbas), logarithmic sigmoid transfer function (losig), and hyperbolic tangent sigmoid transfer function (tansig). In this post-processing step, the output neuron uses a linear activation function. The assignment of the three different transfer functions expressed in this network is shown in Figure (5). Developing an artificial neural network model is a trial-and-error process; however, drawing on the experiences of other researchers can greatly help achieve a network with a suitable structure. According to Tables (5), a single-layer neural network with 10 neurons and two layers, each with five tangent sigmoid activation functions, can achieve acceptable accuracy in estimating LDC in meandering rivers.

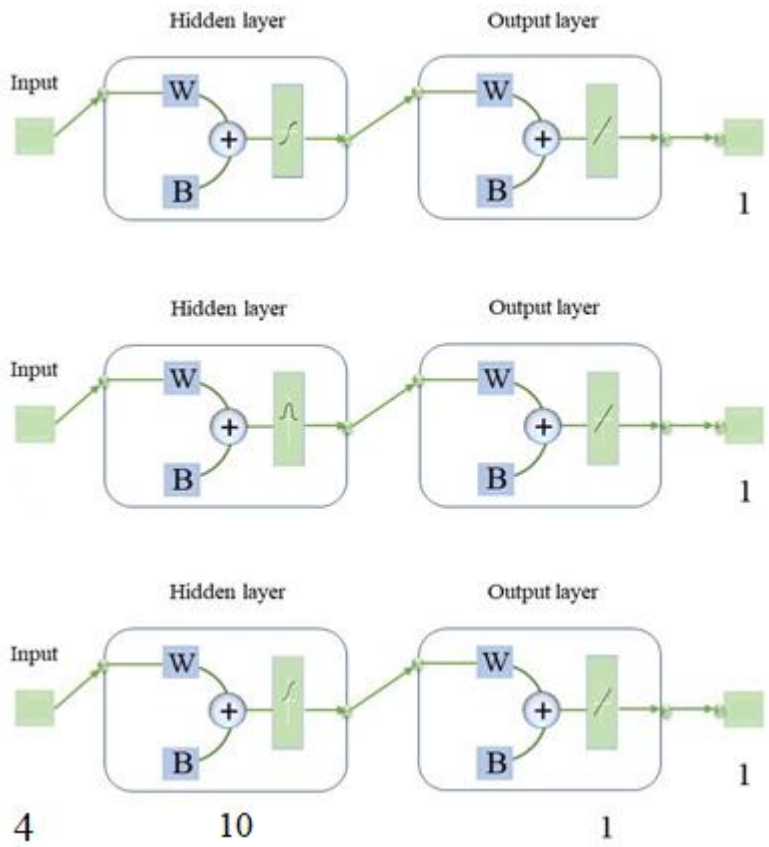


Fig. 5- Different activation functions in the structure of the MLP neural network model

The performance of the developed MLPNN model during training and testing is shown in Figure (6). Its statistical index in the training phase is R^2=0.93, RMSE=35.43 and in the testing phase is R^2=0.90, RMSE=42.51. As shown in these figures, the developed neural network model has estimated the minimum and maximum values of the diffusion coefficient with appropriate accuracy.

Table 5- Structure and performance of the MLPNN model in the development phase

N-H-L	N-N&A-FT	S- h-L	Train-R^2	Test-R^2
1	5-tansig	-	0.55	0.77
1	8-tansig	-	0.83	0.75
1	5-logsig	-	0.86	0.85
1	8-logsig	-	0.83	0.72
1	10-logsig	-	0.93	0.88
2	10-logsig	5-tansig	0.93	0.90
2	12-tansig	5-tansig	0.92	0.95

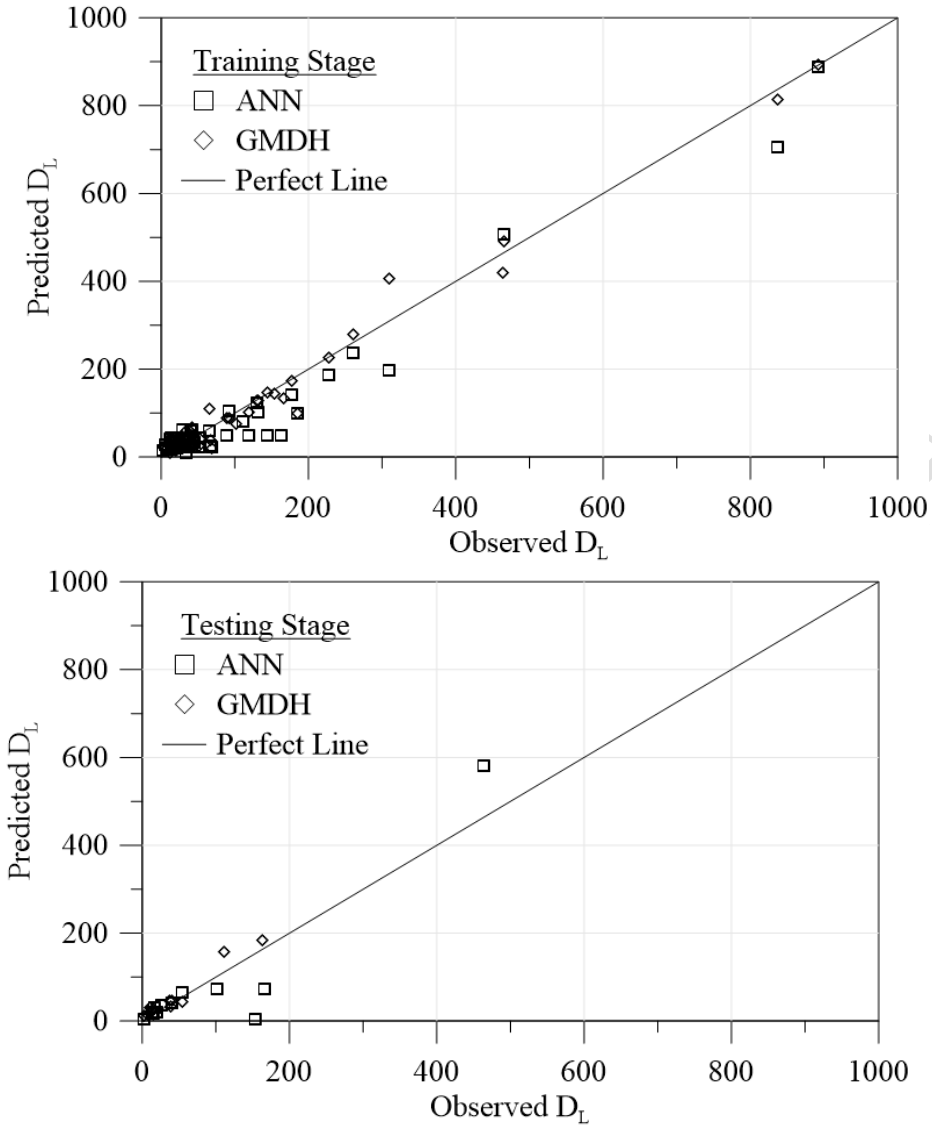


Fig. 6- Performance of the ANN and GMDH in the training and testing phases to estimate LDC in meandering streams.

Development of GMDH

To develop the GMDH model, the same dataset was used for both training and testing. The structure of the GMDH model developed to estimate the LDC in meandering streams is shown in Figure (1). As shown in this figure, the GMDH network has two hidden layers, with the first containing five and three neurons. The performance of this model during the training and testing are shown in Figure (6). The statistical indices of this model in the training phase are R^2=0.93,RMSE=34.73 and in the testing phase are R^2=0.91,RMSE=40.26 .

Comparison of ANN with GMDH

In this section, the MLPNN and GMDH models are compared in terms of structure and statistical indicators. In the MLPNN model, all input variables participate in the first layer, and to increase the accuracy of modeling, the number of neurons and the activation functions of the neurons can be changed, but in the GMDH model, only two inputs can be introduced to each neuron, and the activation function is polynomial and unchangeable. In the MLPNN model, the most effective parameters cannot be determined solely from the network structure. In contrast, in the GMDH model (because the network structure is developed

through elite selection), they can be identified during the network's development. The statistical indicators of the MLPNN and GMDH models for estimating LDC in meandering streams during the training and the testing stages are almost same.

Comparison with previous research

In this section, the results of the developed models are compared with previous research. Regarding the estimation of the LDC in meandering rivers, only Sahay (2013) research is available, in which he achieved high modeling accuracy using a genetic programming model. More researchers have researched the estimation of the LDC in straight rivers (Soltani-Gerdefaramarzi et al., 2015) developed a neurofuzzy model to estimate the LDC in straight rivers. (Zahiri and Sabzkouhi, 2020) reported the indicators for modeling and estimating LDC in straight rivers using the M5 model and genetic programming. Sahay and Dutta (2009) accurately estimated the LDC in rivers using a genetic programming model. Kashefipour and Falconer (2002) accurately predicted LDC in rivers using an artificial neural network model. Haghiabi (2016 and 2017) modeled and estimated the diffusion coefficient in straight rivers using the MARS model, achieving good statistical indicators. (Balf et al., 2018) modeled and estimated the LDC in rivers using the evolutionary polynomial regression method, which achieved accuracy.

Conclusion

Studying, understanding, modeling, and estimating the longitudinal dispersion coefficient in rivers can significantly reduce pollution and improve river water quality. In this study, an attempt has been made to model and estimate the LDC in meandering rivers using soft-computing methods, including ANN and GMDH. The empirical formulas proposed in this case were also evaluated. The most important results obtained are as follows.

1- Factors involved in LDC in meandering rivers include depth, width, average flow velocity, sinuosity coefficient, and shear velocity.

2- Among the evaluated empirical formulas, the formulas of Rajeev & Dutta (2009), Seo & Cheong (1998) were the most accurate.

3- The ANN model developed in this study has ten neurons, and its activation function is of the logarithmic sigmoid type. Other activation functions, including tangent sigmoid, also showed similar performance. Adding more neurons and more hidden layers did not increase the accuracy of modeling and estimating the LDC. The accuracy of the MLPNN model in estimating the LDC in meandering rivers in the test phase is $R^2=0.90$, RMSE=42.51 .

4- The structure of the GMDH in estimating the LDC in meandering rivers has a higher number of hidden layers than the MLPNN model. The GMDH model used to estimate LDC in meandering rivers has two hidden layers. The statistical indices of the GMDH model in the testing phase are $R^2=0.91$,RMSE=40.26 .

5- Comparison of the developed models with previous research shows that the developed models in meandering rivers are at the level of other soft computing models and, in some cases, are better.

References

1- Atkinson, T.C. and Davis, P.M., 2000. Longitudinal dispersion in natural channels: l. Experimental results from the River Severn, UK. *Hydrology and Earth System Sciences*, *4*(3), pp.345-353. https://doi.org/10.5194/hess-4-345-2000, 2000.

2- Azamathulla, H.M. and Ghani, A.A., 2011. Genetic programming for predicting longitudinal dispersion coefficients in streams. *Water resources management*, *25*(6), pp.1537-1544.doi:10.1007/s11269-010-9759-9.

3- Azamathulla, H.M. and Wu, F.C., 2011. Support vector machine approach for longitudinal dispersion coefficients in natural streams. *Applied Soft Computing*, *11*(2), pp.2902-2905.doi:http://dx.doi.org/10.1016/j.asoc.2010.11.026.

4- Balf, M.R., Noori, R., Berndtsson, R., Ghaemi, A. and Ghiasi, B., 2018. Evolutionary polynomial regression approach to predict longitudinal dispersion coefficient in rivers. *Journal of Water Supply: Research and Technology—AQUA*, *67*(5), pp.447-457. https://doi.org/10.2166/aqua.2018.021

5- Benedini, M. and Tsakiris, G., 2013. *Water quality modelling for rivers and streams*. Springer Science & Business Media.10.1007/978-94-007-5509-3

6- Cazelles, B., Fontvieille, D. and Chau, N.P., 1991. Self-purification in a lotic ecosystem: a model of dissolved organic carbon and benthic microorganisms dynamics. *Ecological modelling*, *58*(1-4), pp.91-117. doi:https://doi.org/10.1016/0304-3800(91)90031-U.

7- Chao, X., Jia, Y., Shields Jr, F.D., Wang, S.S. and Cooper, C.M., 2007. Numerical modeling of water quality and sediment related processes. *Ecological modelling*, *201*(3-4), pp.385-397. https://doi.org/10.1016/j.ecolmodel.2006.10.003.

8- Davis, P.M. and Atkinson, T.C., 2000. Longitudinal dispersion in natural channels: 3. An aggregated dead zone model applied to the River Severn, UK. *Hydrology and Earth System Sciences*, *4*(3), pp.373-381. https://doi.org/10.5194/hess-4-373-2000, 2000.

9- Davis, P.M., Atkinson, T.C. and Wigley, T.M.L., 2000. Longitudinal dispersion in natural channels: 2. The roles of shear flow dispersion and dead zones in the River Severn, UK. *Hydrology and Earth System Sciences*, *4*(3), pp.355-371. https://doi.org/10.5194/hess-4-355-2000,

10- Deng, Z.Q., Singh, V.P. and Bengtsson, L., 2001. Longitudinal dispersion coefficient in straight rivers. *Journal of hydraulic engineering*, *127*(11), pp.919-927.https://doi.org/10.1061/(ASCE)0733-9429(2001)127:11(919).

11- Abderrezzak, K.E.K., Ata, R. and Zaoui, F., 2015. One-dimensional numerical modelling of solute transport in streams: The role of longitudinal dispersion coefficient. *Journal of Hydrology*, *527*, pp.978-989.doi:http://dx.doi.org/10.1016/j.jhydrol.2015.05.061.

12- Elder, J., 1959. The dispersion of marked fluid in turbulent shear flow. *Journal of fluid mechanics*, *5*(4), pp.544-560. DOI: https://doi.org/10.1017/S0022112059000374.

13- Fischer, H.B., 1968. *Methods for predicting dispersion coefficients in natural streams: with applications to lower reaches of the green and Duwamish Rivers, Washington* (Vol. 582). US Government Printing Office.

14- Haghiabi, A.H., 2016. Prediction of longitudinal dispersion coefficient using multivariate adaptive regression splines. *Journal of Earth System Science*, *125*(5), pp.985-995. doi:10.1007/s12040-016-0708-8.

15- Haghiabi, A.H., 2017. Modeling river mixing mechanism using data driven model. *Water Resources Management*, *31*(3), pp.811-824.doi:10.1007/s11269-016-1475-7.

16- Huai W, Shi H, Song S, Ni S 2018. A simplified method for estimating the longitudinal dispersion coefficient in ecological channels with vegetation. *Ecological Indicators* 92:91-98. doi:https://doi.org/10.1016/j.ecolind.2017.05.015

17- Ivakhnenko, A.G., 2007. Polynomial theory of complex systems. *IEEE transactions on Systems, Man, and Cybernetics*, (4), pp.364-378.doi:10.1109/TSMC.1971.4308320.

18- Iwasa, Y., 1991. Predicting longitudinal disperdsion coefficient in open-channel flows. *Environmental Hydraulics I* 505–510.

19- Kashefipour, S.M. and Falconer, R.A., 2002. Longitudinal dispersion coefficients in natural channels. *Water Research*, *36*(6), pp.1596-1608.doi:http://dx.doi.org/10.1016/S0043-1354(01)00351-7.

20- Koussis, A.D. and Rodríguez-Mirasol, J., 1998. Hydraulic estimation of dispersion coefficient for streams. *Journal of hydraulic Engineering*, *124*(3), pp.317-320.https://doi.org/10.1061/(ASCE)0733-9429(1998)124:3(317).

21- Li, Z.H., Huang, J. and Li, J. 1998. Preliminary study on longitudinal dispersion coefficient for the gorges reservoir. In: *Proceedings of the Seventh International Symposium Environmental Hydraulics*, 16–18 December. Hong Kong, China.

22- Liu, H., 1977. Predicting dispersion coefficient of streams. *Journal of the Environmental Engineering Division*, *103*(1), pp.59-69.https://doi.org/10.1061/JEEGAV.0000605

23- Liu, H., 1977b. Predicting Dispersion Coefficient of Streams. *Journal of the Environmental Engineering Division*, 103, 59–69. https://doi.org/10.1061/JEEGAV.0000605

24- McQuivey, R.S., Keefer, T.N., 1974. Simple Method for Predicting Dispersion in Streams. *Journal of the Environmental Engineering Division*,. 100, 997–1011. https://doi.org/10.1061/JEEGAV.0000247.

25- Najafzadeh, M. and Barani, G.A., 2011. Comparison of group method of data handling based genetic programming and back propagation systems to predict scour depth around bridge piers. *Scientia Iranica*, *18*(6), pp.1207-1213.doi:http://dx.doi.org/10.1016/j.scient.2011.11.017.

26- Najafzadeh, M., Barani, G.A. and Hessami-Kermani, M.R., 2015. Evaluation of GMDH networks for prediction of local scour depth at bridge abutments in coarse sediments with thinly armored beds. *Ocean Engineering*, *104*, pp.387-396.doi:https://doi.org/10.1016/j.oceaneng.2015.05.016.

27- Najafzadeh, M. and Sattar, A.M., 2015. Neuro-fuzzy GMDH approach to predict longitudinal dispersion in water networks. *Water Resources Management*, *29*(7), pp.2205-2219.doi:10.1007/s11269-015-0936-8.

28- Noori, R., Deng, Z., Kiaghadi, A. and Kachoosangi, F.T., 2016. How reliable are ANN, ANFIS, and SVM techniques for predicting longitudinal dispersion coefficient in natural rivers?. *Journal of Hydraulic Engineering*, *142*(1), p.04015039.doi:10.1061/(asce)hy.1943-7900.0001062.

29- Palancar, M.C., Aragón, J.M., Sánchez, F. and Gil, R., 2003. The determination of longitudinal dispersion coefficients in rivers. *Water environment research*, *75*(4), pp.324-335. https://doi.org/10.2175/106143003X141132.

30- Parsaie, A. and Haghiabi, A.H., 2017a. Computational modeling of pollution transmission in rivers. *Applied water science*, *7*(3), pp.1213-1222.. doi:10.1007/s13201-015-0319-6.

31- Parsaie, A. and Haghiabi, A.H., 2015. Predicting the longitudinal dispersion coefficient by radial basis function neural network. *Modeling Earth Systems and Environment*, *1*(4), p.34.doi:10.1007/s40808-015-0037-y.

32- Parsaie, A. and Haghiabi, A.H., 2017. Numerical routing of tracer concentrations in rivers with stagnant zones. *Water Science and Technology: Water Supply*, *17*(3), pp.825-834.doi:10.2166/ws.2016.175.

33- Ramezani, M., Noori, R., Hooshyaripor, F., Deng, Z. and Sarang, A., 2019. Numerical modelling-based comparison of longitudinal dispersion coefficient formulas for solute transport in rivers. *Hydrological Sciences Journal*, *64*(7), pp.808-819. doi:10.1080/02626667.2019.1605240.

34- Riahi-Madvar, H., Ayyoubzadeh, S.A., Khadangi, E. and Ebadzadeh, M.M., 2009. An expert system for predicting longitudinal dispersion coefficient in natural streams by using ANFIS. *Expert Systems with Applications*, *36*(4), pp.8589-8596.doi:http://dx.doi.org/10.1016/j.eswa.2008.10.043.

35- Sahay, R.R., 2013. Predicting longitudinal dispersion coefficients in sinuous rivers by genetic algorithm. *Journal of Hydrology and Hydromechanics*, *61*(3), p.214.

36- Sahay, R.R. and Dutta, S., 2009. Prediction of longitudinal dispersion coefficients in natural rivers using genetic algorithm. *Hydrology Research*, *40*(6), pp.544-552. https://doi.org/10.2166/nh.2009.014

37- Sattar AMA, Gharabaghi B 2015. Gene expression models for prediction of longitudinal dispersion coefficient in streams. *Journal of Hydrology* 524 (0):587-596. doi:http://dx.doi.org/10.1016/j.jhydrol.2015.03.016.

38- Seo, I.W., Cheong, T.S., 1998. Predicting Longitudinal Dispersion Coefficient in Natural Streams. *Journal of Hydraulic Engineering* 124, 25–32. https://doi.org/10.1061/(ASCE)0733-9429(1998)124:1(25).

39- Singh, S.K. and Beck, M.B., 2003. Dispersion coefficient of streams from tracer experiment data. *Journal of environmental engineering*, *129*(6), pp.539-546. https://doi.org/10.1061/(ASCE)0733-9372(2003)129:6(539.

40- Soltani-Gerdefaramarzi, S., Taghizadeh-Mehrjerdi, R., Ghasemi, M., 2015. Prediction of Longitudinal Dispersion Coefficient in Natural Streams using Soft Computing Techniques. *Iranian Journal of Soil and Water Research*, 46, 385–394. https://doi.org/10.22059/ijswr.2015.56728.

41- Tavakollizadeh, A. and Kashefipur, S. M. 2007.Effects of dispersion coefficient on quality modeling of surface waters. In: *Proceedings of the Sixth International Symposium River Engineering*, 16–18.October, Ahvaz, Iran, pp 67–78. (In Farsi).

42- Zahiri, J. and Sabzkouhi, A.M., 2020. Uncertainty and accuracy analysis of longitudinal dispersion coefficient estimation methods in rivers.

43- Zeng Y, Huai W 2014. Estimation of longitudinal dispersion coefficient in rivers. *Journal of Hydro-environment Research*, 8 (1):2-8. doi:http://dx.doi.org/10.1016/j.jher.2013.02.005

پذیرش نهایی - قبل از چاپ