ORIGINAL ARTICLE

CrossMark

Sensitivity Analysis of Empirical and Data-Driven Models on Longitudinal Dispersion Coefficient in Streams

Hosein Nezaratian[1] · Javad Zahiri[1] [iD] · Seyed Mahmood Kashefipour[2]

Abstract

Longitudinal dispersion coefficient (LDC) is a key element in pollutant transport modeling in streams. Several empirical and data-driven models have been proposed to evaluate this parameter. In this study, sensitivity analysis was performed on four key parameters affecting the LDC including: channel width, flow depth, mean flow velocity and shear velocity. In addition, Monte Carlo simulation was used to generate new datasets and evaluate performance of LDC estimation methods based on uncertainty of input parameters. Sensitivity indices of the input parameters in selected empirical equations and differential evolution model follow almost the same trend, where mean flow velocity is the most sensitive parameter among input parameters and the prediction accuracy depends heavily on the value of this parameter. In above mentioned models, shear velocity had a negative value and a reverse effect on LDC estimation. Channel width and mean flow velocity have the highest sensitivity in M5 model for narrow and wide streams, respectively. Based on sensitivity indices, the efficiency of empirical and data-driven models in different conditions, according to uncertainties in the input parameters, has been investigated. Result of LDC estimation based on the data of Monte Carlo simulation, showed that most LDC estimation models have a high uncertainty for upper LDC values.

✉ Javad Zahiri
  zahiri_javad@yahoo.com

[1] Department of Water Engineering, Agricultural Sciences and Natural Resources University of Khuzestan, Ahvaz, Iran

[2] Department of Water Engineering, Shahid Chamran University of Ahvaz, Ahvaz, Iran

1 Introduction

In recent years much attention has been paid to the environment, especially river and lake pollution. Rivers and streams are usually receiving the outlet of sewage systems which may cause pollutant levels to rise (Haghiabi 2016). Pollutant dispersion is a key element in water quality modeling (Antonopoulos et al. 2015) and the longitudinal dispersion coefficient (LDC) is an important factor in stream pollution modeling due to its effect on pollutant mixing intensity. The pollution is affected by advective and dispersive processes, and is dispersed longitudinally, transversely and vertically (Seo and Cheong 1998). Most of the experimental studies on dispersion coefficient in streams are based on routing tracer concentration along the river (Atkinson and Davis 2000; Davis et al. 2000; Velísková et al. 2014; Disley et al. 2015; Parsaie and Haghiabi 2017). Estimation of the dispersion coefficient directly, using tracers in rivers, is very difficult, expensive and time-consuming. Mixing is a three-dimensional process near the pollution point source. After total mixing far from the injection point, only longitudinal dispersion is used to describe the dispersion phenomenon (Chatila 1997; Velísková et al. 2014; Haghiabi 2017). Therefore, to predict longitudinal dispersion, researchers have developed different equations based on experimental and field measurements. These equations use hydraulic and geometric parameters such as width of the channel, mean flow velocity, shear velocity and depth of flow (Alizadeh et al. 2017). Several researchers have used as input parameters the ratio of channel width to flow depth (W/H) and mean velocity to shear velocity (U/U_*) to estimate LDC based on their correlation with dimensionless dispersion coefficient (K_x HU_*) (Noori et al. 2017). Table 1 presents some empirical equations for LDC estimation. These equations have been derived using different methods and in some cases different set of data, and result in various performances based on stream conditions. In this regard, investigators have recently used data-driven models to estimate LDC. Some of data-driven models which include support vector machine, M5 tree algorithm, differential evolution, genetic algorithms and genetic programming, have been used by Azamathulla and Wu (2011), Etemad-Shahidi and Taghipour (2012), Li et al. (2013), Sahay and Dutta (2009) and Sattar and Gharabaghi (2015), respectively. Due to the use of LDC in devising water diversion strategies, designing treatment plants, intakes and outfalls, and studying the environment (Ho et al. 2002), an important step is the validation of LDC estimation models under different condition. It should be noted that, almost all empirical and data-driven models predict longitudinal dispersion coefficient with simplifying assumptions, which could affect the accuracy of the model results (Sahin 2014). For accurate estimation of water quality parameters, uncertainty and sensitivity analysis must be performed along with water quality modeling (Nakhaei and Etemad-Shahidi 2012). Quantification of the error in water quality models could be used as a first step in evaluation of risk assessment in water resources management and planning. Uncertainty of model-input, model-structure, model-parameter and measurement could be classified as different sources of uncertainty in water quality management (Radwan et al. 2004). Monte Carlo simulation in hydrologic models is widely used in uncertainty analysis (Mishra 2009). This method enables hydrologic modelers to study the effect of input parameter sets on uncertainty of water quality parameters (Pasha and Lansey 2010). Along with uncertainty analysis, sensitivity analysis assists modelers to evaluate output range based on range of input parameters, which could be used to determine the most effective parameters (Nakhaei and Etemad-Shahidi 2012). Non-dimensional sensitivity coefficient is used by hydrological and environmental scientists, to analyze multivariable models (McCuen 1974; Saxton 1975; Rana and Katerji 1998; Hupet and Vanclooster 2001; Gong et al. 2006).

Table 1 Empirical and data-driven models for estimating LDC

	Model	Formula	Category
Empirical (Statistical and Mathematical) Models	Elder (1959)	$K_x = 5.93 HU_*$	Mathematical
	McQuivey and Keefer (1974)	$K_x = 0.58 \left(\frac{H}{U_*}\right)^2 (UW)$	Statistical
	Fischer (1967)	$K_x = 0.011 \left(\frac{U^2 W^2}{HU_*}\right)$	Mathematical
	Liu (1977)	$K_x = \beta \left(\frac{W}{H}\right)^2 \left(\frac{U}{U_*}\right)^2 (HU_*) \quad \beta = 0.18 \left(\frac{U_*}{U}\right)^{1.5}$	Mathematical
	Seo and Cheong (1998)	$K_x = 5.915 \left(\frac{W}{H}\right)^{0.62} \left(\frac{U}{U_*}\right)^{1.428} (HU_*)$	Statistical
	Deng et al. (2001)	$K_x = \left(\frac{0.15}{8\varepsilon_{t_0}}\right) \left(\frac{W}{H}\right)^{5/3} \left(\frac{U}{U_*}\right)^2 (HU_*)$ $\varepsilon_{t_0} = 0.145 + \frac{1}{3520}\left(\frac{U}{U_*}\right)\left(\frac{W}{H}\right)^{1.38}$	Mathematical
	Kashefipour and Falconer (2002)	$K_x = \left[7.428 + 1.775 \left(\frac{W}{H}\right)^{0.62} \left(\frac{U_*}{U}\right)^{0.572}\right] (HU)\left(\frac{U}{U_*}\right)$	Statistical
	Zeng and Huai (2014)	$K_x = 5.4 \left(\frac{W}{H}\right)^{0.7} \left(\frac{U}{U_*}\right)^{0.13} HU$	Statistical
	Sahin (2014)	$K_x = 48 \left(\frac{U}{U_*}\right)^{0.47} R_h U$	Statistical
	Disley et al. (2015)	$K_x = 3.563 \left(\frac{U}{gH}\right)^{-0.4117} \left(\frac{W}{H}\right)^{0.6776} \left(\frac{U}{U_*}\right)^{1.0132} HU_*$	Statistical
Data-driven models	Sahay and Dutta (2009)	$K_x = 2 \left(\frac{W}{H}\right)^{0.96} \left(\frac{U}{U_*}\right)^{1.25} (HU_*)$	Genetic Algorithm (GA)
	Etemad-Shahidi and Taghipour (2012)	$If \left(\frac{W}{H}\right) \leq 30.6, K_x = 15.49 \left(\frac{W}{H}\right)^{0.78} \left(\frac{U}{U_*}\right)^{0.11} (HU_*)$ $If \left(\frac{W}{H}\right) > 30.6, K_x = 14.12 \left(\frac{W}{H}\right)^{0.61} \left(\frac{U}{U_*}\right)^{0.85} (HU_*)$	Model Tree (M5)
	Li et al. (2013)	$K_x = 2.282 \left(\frac{W}{H}\right)^{0.7613} \left(\frac{U}{U_*}\right)^{1.4713} (HU_*)$	Differential Evolution (DE)
	Sattar and Gharabaghi (2015)	$K_x = 8.45 \left(\frac{U}{U_*}\right)^{1.65} \left(\frac{W}{H}\right)^{0.5-0.514 Fr^{0.516}+\left(\frac{U}{U_*}\right)^{0.42}\left(\frac{U_*}{U}\right)} (HU_*)$	Gene Expression (GE)

This study presents the performance of empirical equations and data-driven models using statistical analysis. The novelty of this study lies in sensitivity analysis of accurate empirical and data-driven models, to determine the effect of each parameter on LDC estimation. Sensitivity indices showed which parameter in each model had the main role on LDC estimation. Also, uncertainty analysis was performed based on non-repeated, random data series produced by Monte Carlo simulation to investigate the behavior of empirical and data-driven models. This can determine the performance of each model considering uncertainty in input parameters.

2 Theory and Previous Studies

Transverse shear velocity and transverse mixing become in equilibrium after a certain time-scale (Taylor 1954). The simplified 1-D advection-dispersion equation for steady flow conditions has the following form (Etemad-Shahidi and Taghipour 2012):

$$\frac{\partial C}{\partial t} + U\frac{\partial C}{\partial x} = K_x\frac{\partial^2 C}{\partial x^2} \tag{1}$$

where C is the cross-sectional average concentration (kg/m^3), U is the mean flow velocity (m/s), x is the direction of the mean flow (m), t is the time (s), and K_x is the longitudinal dispersion coefficient (m^2/s) in the flow direction. Based on Rutherford (1994), important features of tracer profiles in laboratory and river channels can be illustrated using Eq. (1).

The LDC in streams is affected by a range of parameters. The most important parameters are density, viscosity, channel width, flow depth, mean flow velocity, shear velocity, bed slope, bed roughness, horizontal stream curvature (i.e., sinuosity), and bed shape factor (Seo and Cheong 1998; Guymer 1998; Etemad-Shahidi and Taghipour 2012). Due to the complexity of measuring these parameters, the researchers have generally applied the hydraulic and geometric parameters such as channel width, flow depth H, flow velocity U and shear velocity U_* which have important effects on LDC. Based on the equilibrium between longitudinal shear velocity and vertical turbulent diffusion, Elder (1959) used Taylor's results from pipes to open channels and derived the following equation to estimate the LDC (Deng et al. 2001):

$$K_x = 5.93 H U_* \tag{2}$$

where H and U_* represent the flow depth (m) and shear velocity (m/s), respectively.

Fischer (1967) suggested that the transverse profile of the velocity is more important than the vertical profile for dispersion in natural streams and developed the following integral relation for the dispersion coefficient in natural streams having large width to depth ratios (Sahay 2011):

$$K_x = -\frac{1}{A}\int_0^W Hu'\int_0^y \frac{1}{\varepsilon_t H}\int_0^y Hu' \, dy dy dy \tag{3}$$

where A is cross-sectional area (m^2); y is the coordinate in the lateral direction (m); u' is the deviation of local depth velocity from the cross-sectional mean velocity (m/s); W is

channel width (m); and ε_t is the transverse, turbulent diffusion coefficient (m²/s). Due to the complexity of Eq. (3), Fischer (1967) developed the following simple and practical equation:

$$K_x = 0.011 \left(\frac{U^2 W^2}{H U_*} \right) \tag{4}$$

Seo and Cheong (1998) proposed an empirical expression based on one-step method developed by Huber (1981), which is a robust regression method, gives reasonably good estimation even in the presence of moderately bad leverage points. Seo and Cheong (1998) used 59 sets of data from 26 U.S. streams to develop the following equation and showed its superiority over existing expressions:

$$K_x = 5.915 \left(\frac{W}{H} \right)^{0.62} \left(\frac{U}{U_*} \right)^{1.428} (H U_*) \tag{5}$$

Deng et al. (2001) developed an analytical method based on Fischer's triple integral expression for estimation of LDC in rivers. They assumed that uniform-flow formula is valid for local depth-averaged parameters. Their equation is theoretically-based and clarifies the dispersion mechanism. Based on Deng et al. (2001), the velocity is the most sensitive parameter among all input parameters in Eq. (6); a change of 10% in this parameter causes significant variation in the LDC:

$$\begin{aligned} K_x &= \left(\frac{0.15}{8\, \varepsilon_{t_0}} \right) \left(\frac{W}{H} \right)^{5/3} \left(\frac{U}{U_*} \right)^2 (H U_*) \\ \varepsilon_{t_0} &= 0.145 + \frac{1}{3520} \left(\frac{U}{U_*} \right) \left(\frac{W}{H} \right)^{1.38} \end{aligned} \tag{6}$$

Kashefipour and Falconer (2002) established an equation for predicting the LDC in natural channels using 81 sets of field data in the U.S., by relating this process through dimensional and regression analysis to the main hydraulic parameters such as river depth, width, velocity and shear velocity. Kashefipour and Falconer (2002) applied multiple regression between parameter combinations, and a best fit simple equation was derived, as follows:

$$K_x = 10.612 \, (H U) \left(\frac{U}{U_*} \right) \tag{7}$$

Kashefipour and Falconer (2002) used a linear combination of Eq. (7) and Seo and Cheong's (1998) formulation to develop Eq. (8), which led to a further improved equation for predicting the LDC in streams:

$$K_x = \left[7.428 + 1.775 \left(\frac{W}{H} \right)^{0.62} \left(\frac{U_*}{U} \right)^{0.572} \right] (H U) \left(\frac{U}{U_*} \right) \tag{8}$$

Zeng and Huai (2014) showed that the product of water depth and cross-sectional mean flow velocity has a higher linear correlation with the LDC than the product of water depth and shear velocity. Therefore, with combination of the product of H and U and other two non-dimensional parameters, a new equation for longitudinal dispersion coefficient was proposed:

$$K_x = 5.4 \left(\frac{W}{H} \right)^{0.7} \left(\frac{U}{U_*} \right)^{0.13} H U \tag{9}$$

Sahin (2014) proposed an equation based on dimensional and least squares analysis, using 128 field data sets measured in 41 rivers in the U.S. as follows:

$$K_x = 48 \left(\frac{U}{U_*}\right)^{0.47} R_h U \qquad (10)$$

where R_h is the hydraulic radius (m), which was calculated assuming a rectangular channel section due to the lack of data on cross section shape (Sahin 2014). Disley et al. (2015) developed an equation to estimate LDC using combined data sets from five steeper head – water streams and 24 milder and larger rivers. This equation relates the LDC to hydraulic and geometric parameters of the stream and has been developed using multiple regression analysis:

$$K_x = 3.563 \left(\frac{U}{gH}\right)^{-0.4117} \left(\frac{W}{H}\right)^{0.6776} \left(\frac{U}{U_*}\right)^{1.0132} HU_* \qquad (11)$$

where g is a gravitational acceleration (m/s^2).

Data-driven models have widely been used by researchers to estimate LDC in streams. Sahay and Dutta (2009) developed an equation to estimate the LDC, using the datasets of Deng et al. (2001) and genetic algorithm:

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

Etemad-Shahidi and Taghipour (2012) derived two interpretable equations to estimate LDC using M5 tree algorithm and 149 datasets from rivers around the world:

$$If \left(\frac{W}{H}\right) \leq 30.6, \ K_x = 15.49 \left(\frac{W}{H}\right)^{0.78} \left(\frac{U}{U_*}\right)^{0.11} (HU_*) \qquad (13-a)$$

$$If \left(\frac{W}{H}\right) > 30.6, \ K_x = 14.12 \left(\frac{W}{H}\right)^{0.61} \left(\frac{U}{U_*}\right)^{0.85} (HU_*) \qquad (13-b)$$

Table 1 summarizes selected equations and models for LDC estimation.

3 Materials and Methods

3.1 Data and Statistical Analysis

A collection of distinctive datasets measured in different streams were used in this study (Fischer 1968; Yotsukura et al. 1970; Godfrey and Frederick 1970; McQuivey and Keefer 1974; Nordin and Sabol 1974; Rutherford 1994; Graf 1995; Seo and Cheong 1998; Disley et al. 2015). The datasets contained geometric and hydraulic characteristics, which include: channel width, flow depth, mean flow velocity, shear velocity and longitudinal dispersion coefficient (Appendix Table 8). Histograms of W, H, U, U_*, K_x, W/H, and U/U_* are illustrated in Fig. 1. The histogram of W/H implies that the studied cases varied from narrow rivers ($W/H < 10$) to very wide rivers ($W/H > 100$). The friction term in the form of
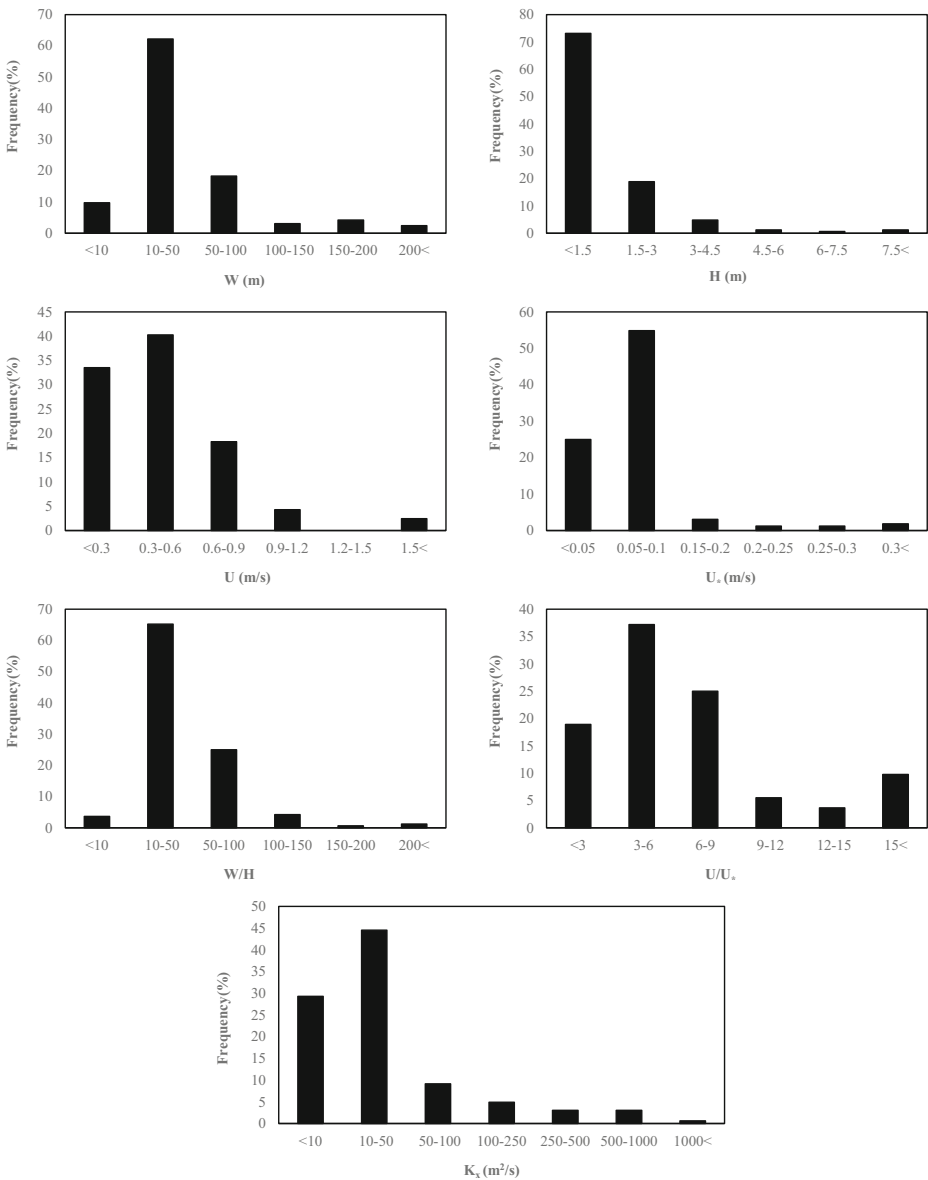
Fig. 1 Histograms of all parameters

U/U_* (Seo and Cheong 1998) can be considered as the hydrodynamic characteristic of the river bed (Etemad-Shahidi and Taghipour 2012). The statistical values of parameters are presented in Table 2.

3.2 Sensitivity Analysis

Geometric and hydraulic characteristics such as channel width, flow depth, mean flow velocity and shear velocity may have some uncertainties in their value

Table 2 Statistics of parameters used in this study

Statistical indicator	W (m)	H (m)	U (m/s)	U_* (m/s)	W/H	U/U_*	K_x HU_*	K_x (m²/s)
Maximum	253.60	8.20	1.73	0.55	403.75	20.25	37,140.74	1486.50
Minimum	1.40	0.14	0.03	0.00	2.20	0.77	3.08	0.20
Average	46.15	1.25	0.47	0.09	46.83	6.70	1116.41	76.11
Standard deviation	47.33	1.30	0.31	0.07	45.06	4.59	3478.76	173.96

estimation. Poor estimation procedures, tracer loss, or measurements made in the advective zone are examples of such uncertainties in K_x values (Etemad-Shahidi and Taghipour 2012).

Sensitivity analysis was employed in order to identify which parameters have more influence on the dimensionless longitudinal dispersion coefficient. Model sensitivity is the rate of change in one factor as output with respect to change in another factor as input while the other parameters are kept constant (McCuen 1973), or how the variation in the output of a model (numerical or other) can be apportioned, qualitatively or quantitatively, to different sources of variation of input parameters (Saltelli et al. 2004).

A logical step in model development is the determination of the most important parameters affecting the model results. A 'sensitivity analysis' of these parameters could serve to help future studies (Hamby 1994). Computer models used in hydraulic engineering have been increased, and this has not been accompanied by a corresponding increase in sophistication of sensitivity analysis (Hall et al. 2009). Estimation of the risk by the coupling of hydrodynamic, structural reliability and impacts models causes additional motivation for improved sensitivity analysis (Dawson et al. 2005). However, without a systematic method to exploring the model response to inputs changes, model developers cannot discover reliable intuitions about the model behavior and interactions (Hall et al. 2009). Sensitivity analyses have been used to determine which parameter has the most effect on reducing output uncertainty, and/or which parameters are negligible and can be eliminated from the final model, and/or which inputs contribute most to output change, and/or which parameters are strongly correlated with the output, and/or what are the consequent results from changing each input parameter (Hamby 1994).

Input parameters for sensitivity analysis of LDC models were considered with their average value; one parameter was changed in a defined domain and this process continued for all of the remaining parameters. With this mechanism, the output variability was estimated based on insignificant modifications of each input parameter, and the model sensitivity to each parameter variation was predicted. A general LDC model can be defined as follows:

$$K_x = f(V_1, V_2, \ldots V_n) \tag{14}$$

where V_i represents input parameters. Based on Beven (1979), the variation of K_x can be written as:

$$K_x + \Delta K_x = f(V_1 + \Delta V_1, V_2 + \Delta V_2, \ldots, V_n + \Delta V_n) \tag{15}$$

Expanding Eq. (15) in Taylor series, and ignoring second-order terms, leads to:

$$\Delta K_x = \frac{\partial K_x}{\partial V_1} \Delta V_1 + \frac{\partial K_x}{\partial V_2} \Delta V_2 + \ldots + \frac{\partial K_x}{\partial V_n} \Delta V_n \tag{16}$$

where the differentials $\frac{\partial K_x}{\partial V_i}$ define the sensitivity of the estimated output to each model parameter. Let us set:

$$A_S = \frac{\partial K_x}{\partial V_i} \approx \frac{\Delta K_x}{\Delta V_i} \tag{17}$$

where A_s represents the absolute sensitivity of the output estimation to each input parameter. The differential analysis is typically much more demanding to implement than other sensitivity methods and yet provides only comparable results. Using sensitivity analysis as a partial derivative form is impractical due to its complexity (Gardner et al. 1981). In addition, when parameter variability takes realistic values this method which is valid for only small changes in parameter values will be impractical (Hamby 1994).

The magnitude of parameters in the LDC equation varied, therefore, the absolute form of sensitivity values from Eq. (17) are unsuitable for comparison of sensitivity values. So, relative sensitivity values were used to compare sensitivity values of input parameters (Mount et al. 2013) in the form:

$$R_s = \frac{\Delta K_x}{\Delta V_i} \frac{V_i}{K_x} \tag{18}$$

Relative changes or errors can be defined as in Saxton (1975):

$$R_E = \frac{\Delta K_x}{K_x} \tag{19}$$

where R_s is a dimensionless coefficient which demonstrates the percentage of the relative parameter change transmitted to the relative dependent parameter. This may be defined as the sensitivity coefficient, for example, a sensitivity coefficient of 0.2 means 10% change in V_i as an input parameter, would cause a 2% change in LDC ($\Delta K_x/K_x$) (Saxton 1975).

LDC models are affected by four input parameters which have a wide variation range in nature and a lot of real data are needed to investigate the performance of LDC estimation. A sensitivity and error analysis of the empirical and data-driven models are conducted for mean values of input and output parameters and on the assumption that the interaction between input parameters is negligible (Deng et al. 2001). Performance of selected models were evaluated by two approaches of changing each input individually and the whole ones by random and none-repeating dataset. In addition, global sensitivity based on Saltelli et al. (2008) has been performed to investigate the interaction between input parameters. Therefore, first-order sensitivity index (S_i) has been estimated for each input parameter. If the sum of all S_i was equal to 1, model is additive and there is not any interaction between input parameters (Saltelli et al. 2008).

In this study, for each input parameter, 100 random and none-repeating datasets were produced in a domain of ±10% and ±20% change of each input parameter for each available data series. To investigate the effect of all input parameters on LDC estimation, another random data series were produced based on ±10% and ±20% change on all parameters. These datasets have been used for selected models and the minimum and maximum LDC estimation for each data series of every dataset was derived to analyze the performance of models and derive the uncertainty curves. It is necessary to mention that for sensitivity and uncertainty analysis of each input parameter, the other input parameters were kept constant at their average values.

3.3 Model Validation

Performance of LDC models have been evaluated using statistical measures, including the mean absolute error (MAE), the root mean square error (RMSE), and the discrepancy ratio (DR) and the related accuracy. DR was defined by White et al. (1973) to evaluate the difference between measured and predicted values. If DR = 0, the predicted and measured values of the dispersion coefficient are identical, while the model overestimates the measured values of the dispersion coefficient when DR > 0, and underestimates them when DR < 0. Accuracy is defined as the proportion of numbers with DR between −0.3 and 0.3 in the total number of data (Seo and Cheong 1998):

$$MAE = \frac{1}{N}\sum|DR_i| \tag{20}$$

$$RMSE = \sqrt{\frac{1}{N}\sum(DR_i)^2} \tag{21}$$

$$DR = log\frac{K_{x_p}}{K_{x_m}} \tag{22}$$

where K_{x_p} and K_{x_m} are the predicted and measured LDC, respectively.

4 Results and Discussion

Sensitivity analysis, in addition to statistical analysis, helps the researchers to know limitations and advantages of LDC models. Statistical measures, including MAE, RMSE and DR of empirical and data-driven models are given in Table 3. Histogram of DR values for better comparison between models are also illustrated in Fig. 2.

Table 3 Comparison of the performance of the various models

	Model	DR < −0.3	−0.3 < DR < 0	0 < DR < 0.3	DR > 0.3	Accuracy	MAE	RMSE
Empirical	Elder (1959)	98.17	1.22	0.61	0.00	1.83	1.81	1.90
	McQuivey and Keefer (1974)	1.83	3.05	6.10	89.02	9.15	1.67	2.04
	Fischer (1967)	31.10	16.46	17.68	34.76	34.14	0.57	0.72
	Liu (1977)	15.85	22.56	27.44	34.15	50.00	0.43	0.58
	Seo and Cheong (1998)	17.07	17.07	29.27	36.59	46.34	0.42	0.58
	Deng et al. (2001)	21.34	18.90	26.22	33.54	45.12	0.42	0.56
	Kashefipour and Falconer (2002)	25.00	28.05	21.95	25.00	50.00	0.44	0.63
	Sahin (2014)	17.07	25.00	29.88	28.05	54.88	0.43	0.62
	Zeng and Huai (2014)	22.56	23.78	26.83	26.83	50.61	0.37	0.51
	Disley et al. (2015)	6.71	14.63	33.54	45.12	48.17	0.44	0.62
Data-driven	Genetic Algorithm (GA)	18.29	23.78	21.95	35.98	45.73	0.39	0.52
	Model Tree (M5)	16.46	26.83	32.93	23.78	59.76	0.33	0.44
	Differential Evolution (DE)	25.00	23.17	23.78	28.05	46.95	0.40	0.54
	Gene Expression (GE)	8.54	29.88	28.66	32.93	58.54	0.37	0.57
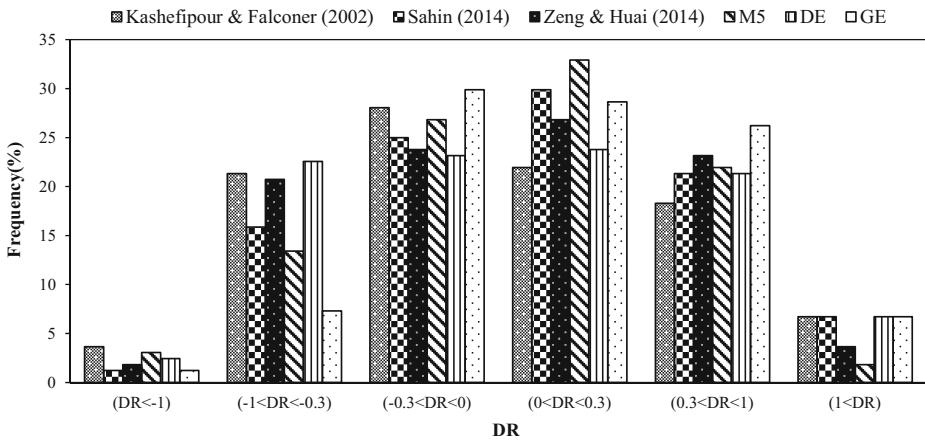
Fig. 2 Comparison of the DR values of different models

Table 3 results shows that Elder (1959) equation has the maximum error and minimum accuracy. This equation is suitable for streams with no transverse shear, but the accuracy of this equation illustrates the importance of transverse variation (Etemad-Shahidi and Taghipour 2012). DR < −0.3 for this model is about 98% and this demonstrates lower estimation of the LDC by Elder equation. McQuivey and Keefer (1974) model with RMSE equal to 2.04 and accuracy of 9.15% generally overestimates the LDC values in streams with 89% of DR > 0.3. Error criteria for Fischer (1967) decreased in comparison with Elder (1959) and McQuivey and Keefer (1974) and its accuracy has been improved. Sahin (2014) has the highest accuracy among all empirical models, followed by Zeng and Huai (2014), Liu (1977) and Kashefipour and Falconer (2002). Disley et al. (2015), Seo and Cheong (1998) and Deng et al. (2001) models with the accuracy of 48.17%, 46.34 and 45.12%, respectively, have relatively accurate estimation of LDC. Zeng and Huai (2014) has the lowest RMSE among all empirical formulas. Error estimation of LDC for the Kashefipour and Falconer (2002) model is more than the corresponding values for some of the empirical models but its perfect symmetry between lower and upper estimates make this model suitable for LDC estimation (Etemad-Shahidi and Taghipour 2012). Liu (1977), Seo and Cheong (1998) and Deng et al. (2001) overpredict the LDC by 2.15, 2.14 and 1.57 times, respectively, more than the underpredicted cases. However, for Kashefipour and Falconer (2002), the overpredicted and underpredicted cases are equal, which make the performance of this formula to be better on LDC estimation. This result is consistent with Etemad-Shahidi and Taghipour (2012) findings.

Genetic algorithm has the lowest accuracy among data-driven models, therefore, this model was eliminated from sensitivity analysis. Based on Table 3, M5 model has the highest accuracy and lowest RMSE among all LDC estimation models. Finally, based on statistical analysis three empirical equations, including Kashefipour and Falconer (2002), Sahin (2014) and Zeng and Huai (2014), and three data-driven models, including M5, GE and DE have been selected for sensitivity analysis.

As it was mentioned above, model sensitivity is the rate of change in LDC with respect to change in input parameters while the other parameters are kept constant; in other words, to investigate the direct effect of one parameter on LDC, the effect of other parameters should be neglected by keeping their values constant. The average of input parameters for LDC estimation are calculated from the existing datasets, and are presented in Table 4.

Results of global sensitivity based on Saltelli et al. (2008) are presented in Table 5. In this table, first-order sensitivity index (S_i) has been estimated for all selected models. Sum of S_i for LDC models has been estimated near 1, which implies that these models are additive with weak interaction between input parameters.

In this study, $\Delta V_i = 0.1 V_i$ has been used for estimation of relative sensitivity coefficient and relative error. Sensitivity analysis of selected empirical and data-driven models are presented in Table 6. Also, the rate of LDC changes based on $\pm 10\%$ and $\pm 20\%$ change of each input parameter with assuming the other parameters to be constant are illustrated in Fig. 3. For M5 models, which contain two equations, the dataset was divided into two domains and used for each equation based on model criteria. R_s is one of the most important sensitivity indicators for multivariable models. Parameters with the large amount of R_s have the great effect on LDC. The estimated error caused by changing each parameter on LDC is shown by R_E (Table 6).

Mean flow velocity has the maximum of R_s in Kashefipour and Falconer (2002), and when increasing 10% the velocity, the LDC value increases about 18%, as it is illustrated in Table 6. Shear velocity has an inverse effect on LDC, where increasing 10% the shear velocity leads to about 6.7% decrease in LDC. The effect of input parameters changes on LDC computed by Kashefipour and Falconer (2002) equation is presented in Fig. 3a. According to the contents of Table 6, mean flow velocity and channel width have the highest and lowest effect on LDC, respectively, according to Sahin (2014) model. Fig. 3b shows that the channel width has no influence on LDC using this model. Mean flow velocity, channel width and flow depth have the highest R_s based on Zeng and Huai (2014) model, respectively, (Table 6; Fig. 3c). This model has a lowest sensitivity to shear velocity among all empirical and data driven models.

The M5 algorithm proposed two piecewise equations for LDC estimation. For this reason, the sensitivity analysis was performed for two nonlinear equations of this model. Splitting value for W/H is approximately 30, close to the value obtained by Papadimitrakis and Orphanos (2004). In narrow rivers with $W/H \leq 30.6$, the importance of shear velocity and channel width are more than the flow depth and velocity,

Table 4 Average value of input parameters used for LDC estimation

Average data	All data	$W/H \leq 30.6$	$W/H > 30.6$
W (m)	46.15	28.92	56.89
H (m)	1.25	1.70	0.98
U (m/s)	0.47	0.48	0.45
U_* (m/s)	0.09	0.11	0.07

Table 5 First-order sensitivity index (S_i) of input parameters for selected models

Model	S_i				$\sum S_i$
	W	H	U	U$_*$	
Kashefipour and Falconer (2002)	0.12	0.08	0.26	0.50	0.96
Sahin (2014)	0.01	0.15	0.18	0.45	0.79
Zeng and Huai (2014)	0.32	0.11	0.45	0.05	0.93
Model Tree (M5)	0.45	0.08	0.24	0.08	0.85
Differential Evolution (DE)	0.11	0.02	0.19	0.64	0.96
Gene Expression (GE)	0.01	0.06	0.05	0.84	0.96

therefore W/H is more important than U/U_* on LDC estimation. In wider rivers, where $W/H > 30.6$, mean flow velocity has the highest value of R_s, hence U/U_* has the main effect on LCD (Table 6). A possible interpretation is that K_x may be less influenced by the W/H ratio in very wide rivers than in narrow rivers. This result is in agreement with Papadimitrakis and Orphanos (2004) and Etemad-Shahidi and Taghipour (2012) findings. In addition, Fig. 3d, e and g for M5 and GE models, show that the shear velocity has a direct effect on LDC which is not consistent with the empirical equations. The DE model developed by Li et al. (2013) behaves similarly to the empirical models, with the mean flow velocity having a highest effect and the shear velocity an inverse effect on LDC estimation (Table 6; Fig. 3f). It was also found that in GE model, the most effective parameters on LDC were the flow depth, the mean flow velocity, the channel width and the shear velocity, respectively, in descending order of importance. The shear velocity in GE model has an uncommon effect on LDC, as shown in Fig. 3g, where an increase and decrease of this parameter has the same impact on LDC. It should be noted that the models developed by GE have a complicated structure which reduce their interpretation in comparison with other models.

Table 6 Sensitivity analysis of selected empirical and data driven models

Input	Sensitivity indices	Kashefipour and Falconer (2002)	Sahin (2014)	Zeng and Huai (2014)	M5 Eq. (13-a)	M5 Eq. (13-b)	DE model	GE model
W	ΔV	4.615	4.615	4.615	2.892	5.689	4.615	4.615
	ΔK	1.207	0.273	3.393	2.32	3.41	3.43	2.033
	R_s	0.283	0.047	0.69	0.772	0.599	0.753	0.363
	R_E	0.028	0.005	0.069	0.077	0.06	0.075	0.036
H	ΔV	0.125	0.125	0.125	0.17	0.098	0.125	0.125
	ΔK	3.01	5.49	1.427	0.637	2.157	1.049	4.379
	R_s	0.706	0.944	0.29	0.212	0.379	0.23	0.781
	R_E	0.071	0.094	0.029	0.021	0.038	0.023	0.078
U	ΔV	0.047	0.047	0.047	0.048	0.045	0.047	0.047
	ΔK	7.676	8.749	5.593	0.317	4.807	6.862	3.645
	R_s	1.801	1.504	1.137	0.105	0.844	1.505	0.65
	R_E	0.18	0.15	0.114	0.011	0.084	0.151	0.065
U$_*$	ΔV	0.009	0.009	0.009	0.011	0.007	0.009	0.009
	ΔK	−2.864	−2.548	−0.606	2.662	0.82	−2.002	0.772
	R_s	−0.672	−0.438	−0.123	0.885	0.144	−0.439	0.138
	R_E	−0.072	−0.044	−0.012	0.081	0.014	−0.046	0.014
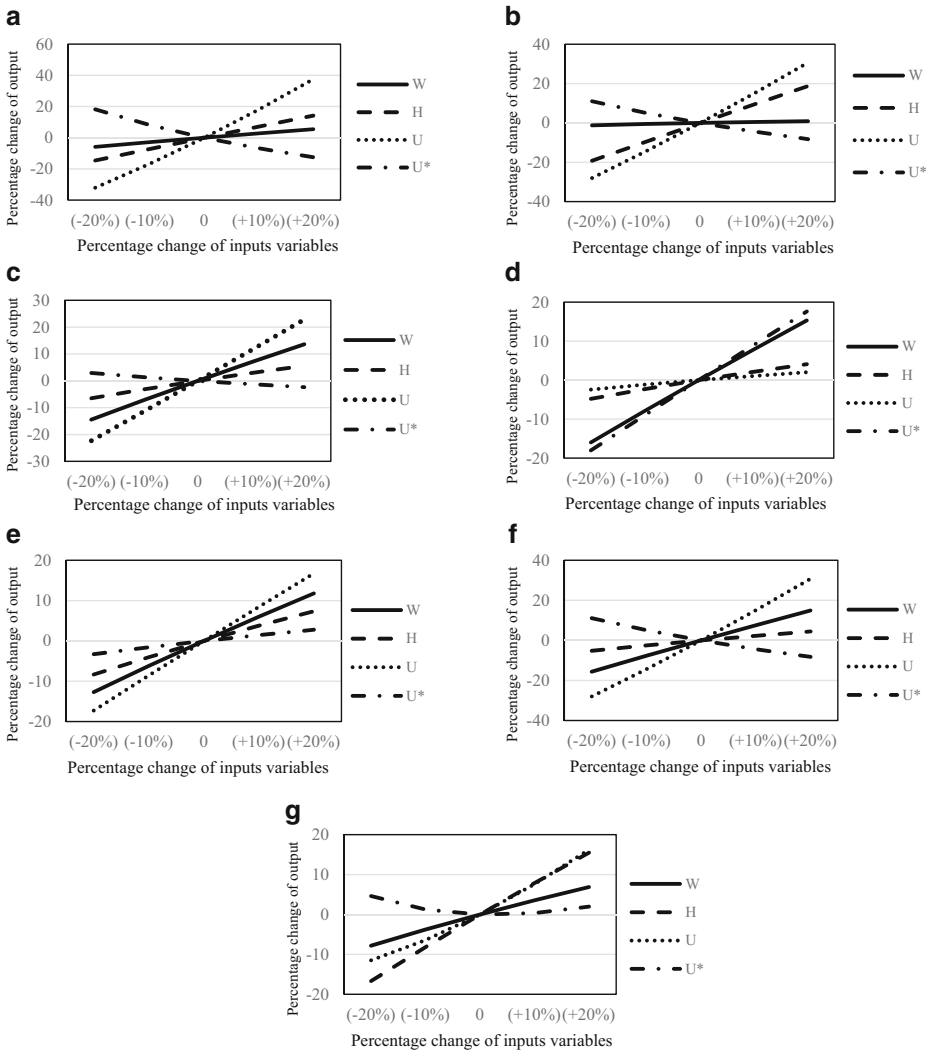
Fig. 3 Sensitivity analysis of (**a**) Kashefipour and Falconer (2002); (**b**) Sahin (2014); (**c**) Zeng and Huai (2014); (**d**) M5 Eq. (13-a); (**e**) M5 Eq. (13-b); (**f**) DE model; (**g**) GE model

Table 7 Relative sensitivity (R_s) of all models

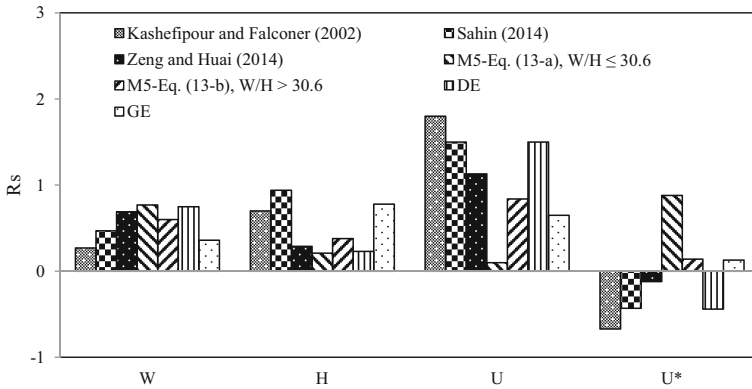| Models	R_s			
	W	H	U	U*
Kashefipour and Falconer (2002)	0.27	0.70	1.80	−0.67
Sahin (2014)	0.47	0.94	1.50	−0.43
Zeng and Huai (2014)	0.69	0.29	1.13	−0.12
M5-Eq. (13-a), $W/H \leq 30.6$	0.77	0.21	0.10	0.88
M5-Eq. (13-b), $W/H > 30.6$	0.60	0.38	0.84	0.14
DE	0.75	0.23	1.50	−0.44
GE	0.36	0.78	0.65	0.13

Fig. 4 Histogram of relative sensitivity of all models

Table 7 and Fig. 4 illustrate R_s values of the selected empirical and data-driven models for better comparison. Mean flow velocity is the most sensitive parameter on empirical equations and some of data-driven models (M5-Eq. 13-b,
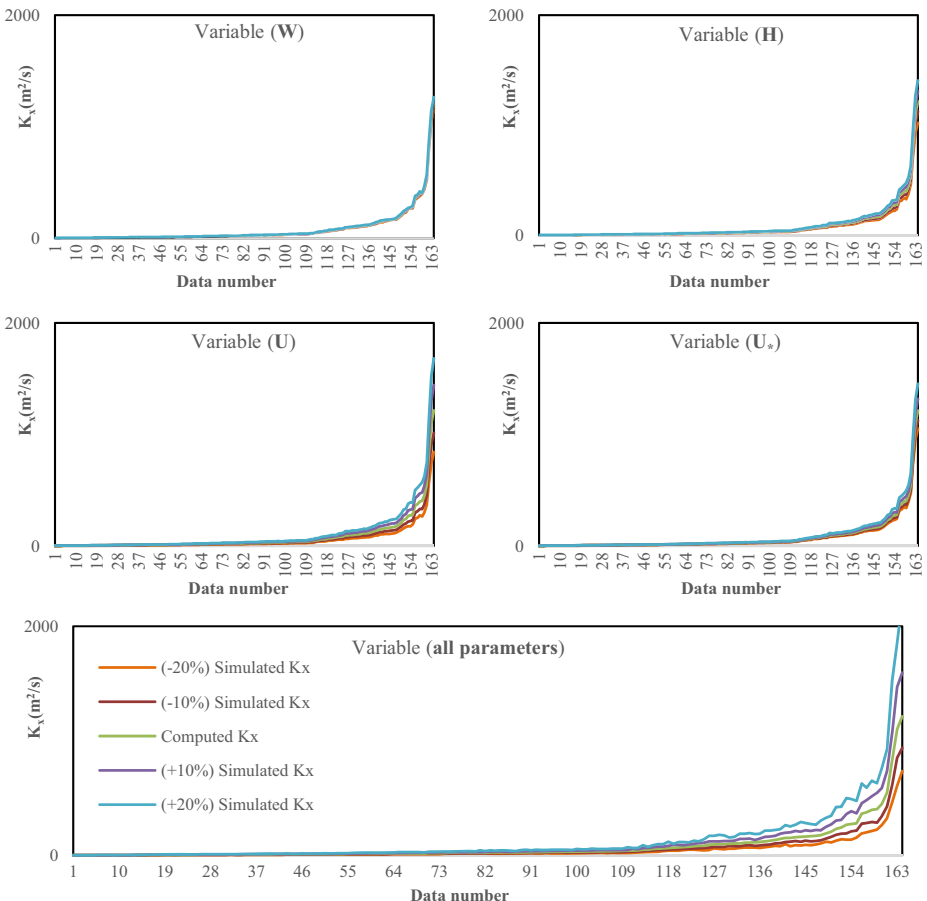

Fig. 5 Performance of Kashefipour and Falconer (2002) using Monte Carlo simulation

and DE), which is consistent with the findings of Deng et al. (2001) and Haghiabi (2017). Therefore, the performance of the formulas mentioned above, depends heavily on the velocity value. Comparison of the two equations developed by M5 model shows that the role of mean flow velocity becomes more pronounced in relatively wide rivers than narrow rivers, which is in agreement with Rutherford (1994) findings. In addition, shear velocity is the most sensitive parameter among the four input parameters in narrow streams. This has been reported by Tayfur and Singh (2005), who stated that ANN model can yield satisfactory predictions of the LDC in narrow streams, if the shear velocity was used as the only input parameter.

Table 7 shows that the channel width is the most sensitive parameter in M5 Eq. (13-a) model, while Kashefipour and Falconer (2002) has the least sensitivity to channel width. Flow depth has the most impact on Sahin (2014) formula, while M5-Eq. (13-a) has the least sensitivity to flow depth among all models of LDC estimation. Velocity has the highest effect on Kashefipour and Falconer (2002) and
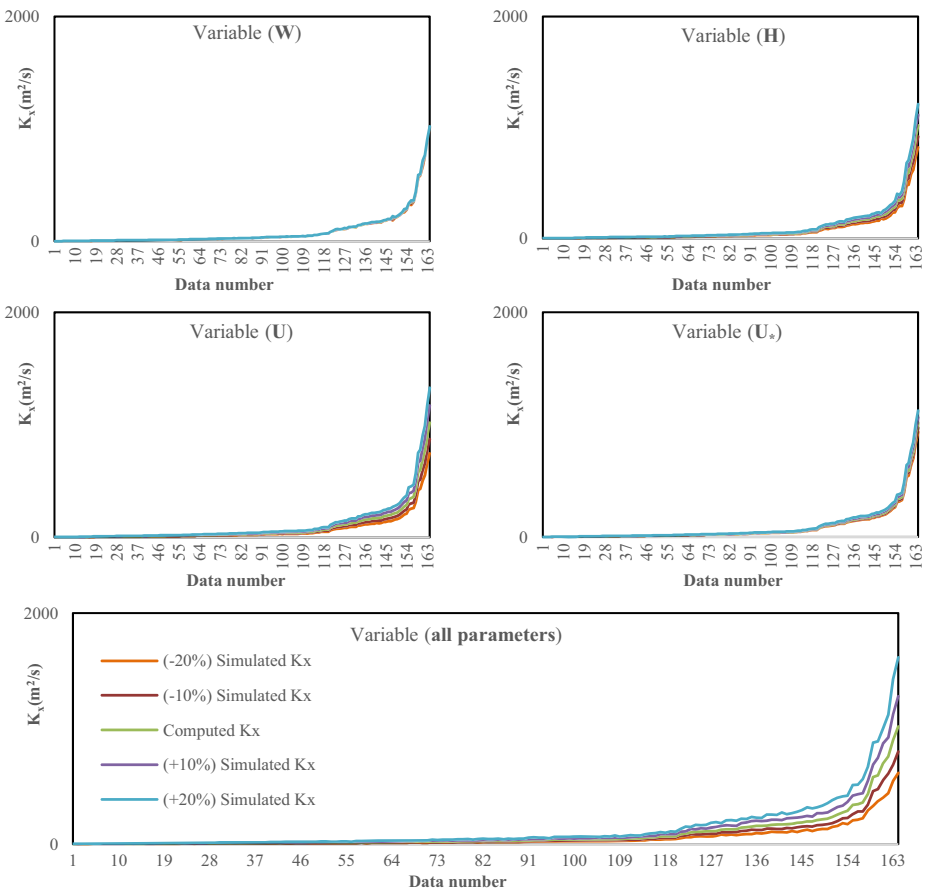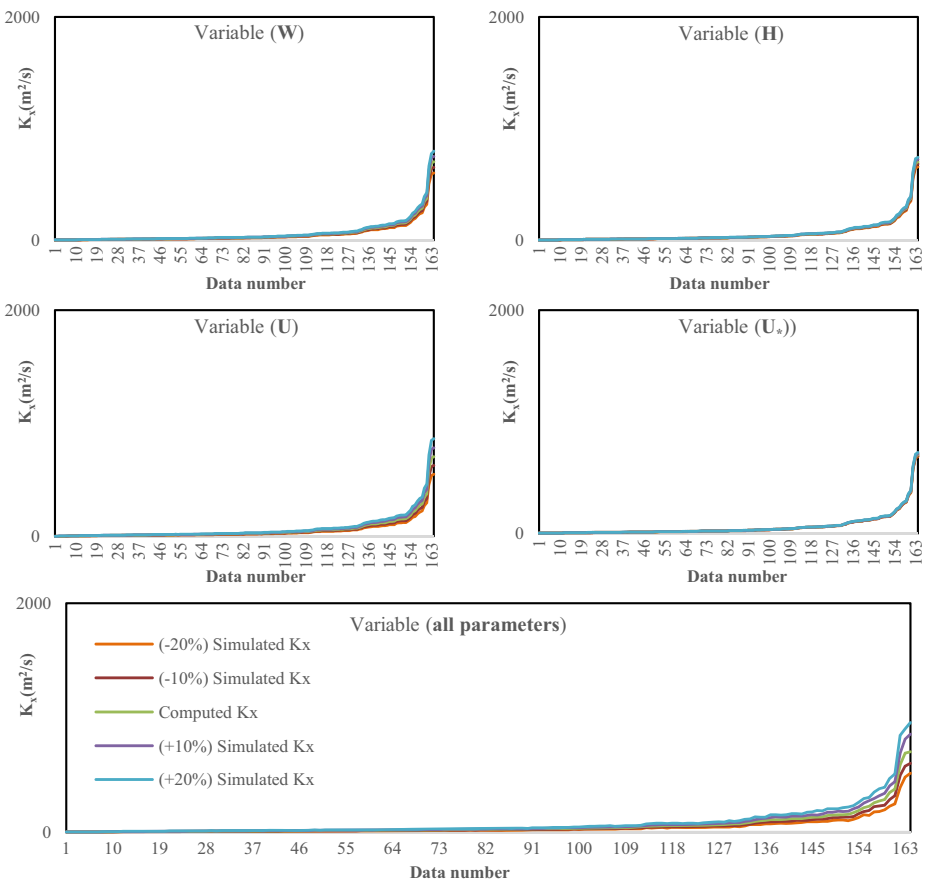

Fig. 6 Performance of Sahin (2014) using Monte Carlo simulation

this parameter is the least sensitive parameter in M5-Eq. (13-a). M5-Eq. (13-a) and Zeng and Huai (2014) have the most and the least sensitivity to shear velocity among others.

Uncertainty curves based on Monte Carlo simulation, which have been used to produce new data set based on ±10% and ±20% changes of each input parameter for the empirical and data-driven models are presented in Figs. 5, 6, 7, 8, 9 and 10. For interpretation of the figures, all data were arranged in descending order based on LDC values.

According to the Monte Carlo simulation (Figs. 5, 6, 7, 8, 9 and 10), Zeng and Huai (2014) equation had the lowest sensitivity to parameter changes, and showed a smooth curve. Uncertainty of this equation in comparison with other models, for all input parameters, was negligible. Two other empirical models, those by Kashefipour and Falconer (2002) and Sahin (2014), showed a high uncertainty for higher LDC values. LDC estimated for ±20% changes in all input parameters, for high LDC values, were more than the LDC calculated for the original data values at about 80 and 60% for Kashefipour and Falconer

Fig. 7 Performance of Zeng and Huai (2014) using Monte Carlo simulation
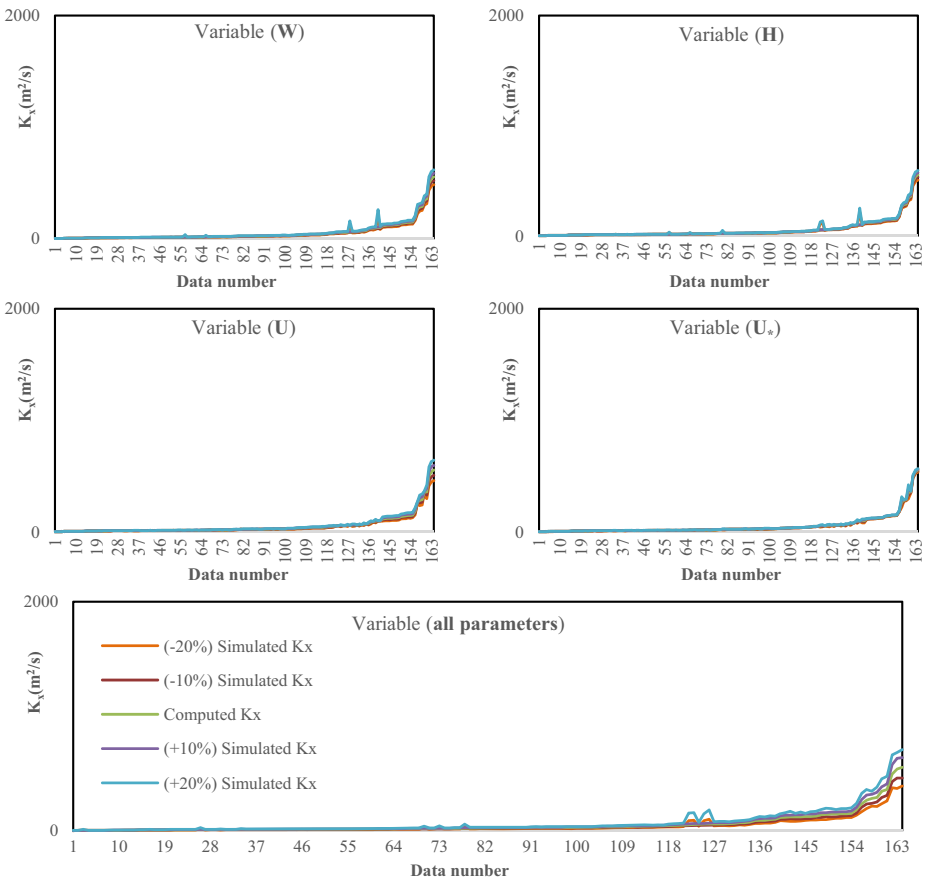
Fig. 8 Performance of M5 using Monte Carlo simulation

(2002) and Sahin (2014), respectively. This was also observed for DE and GE models. The M5 models are very sensitive to parameter changes, based on Fig. 8. The W/H ratio would change with variation in flow depth and channel width, and in some cases, the stream is converted from narrow to wide and vice versa. In such cases, the other equation of the M5 model would be used for the new dataset, which makes a high differences between LDC estimated for the original dataset and the new data generated by the Monte Carlo simulation for the same dataset (sudden jump). The GE model has a non-smooth uncertainty curve for high LDC values due to the model's complexity.

The results of the uncertainty analysis based on Monte Carlo simulation showed that the Zeng and Huai (2014) equation demonstrates less uncertainty to input parameter changes and is more reliable for estimation of LDC in comparison with the other models.
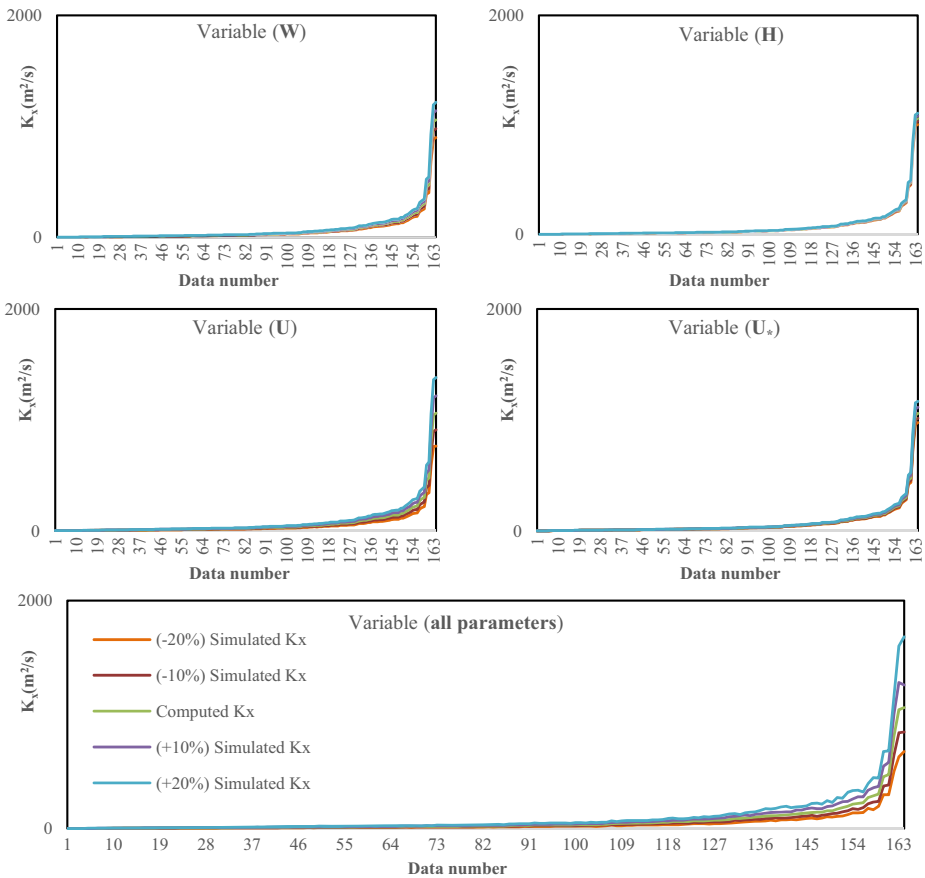
Fig. 9 Performance of DE using Monte Carlo simulation

5 Conclusions

Due to the complexity of measuring and the time-consuming tracer studies, empirical and data-driven models have been developed by many scientists to estimate the longitudinal dispersion coefficient, in order to apply in mathematical models for water quality modelling. Based on statistical analysis, M5 algorithm has the best accuracy and the least computational error compared to other studied empirical and data-driven models. Sensitivity analysis on selected empirical and data-driven models showed that the Kashefipour and Falconer (2002) equation has the least sensitivity to channel width, which could be used for rivers with variety in width or uncertainty in measuring this parameter. In this case, uncertainty in this parameter would have the least effect on LDC estimation. M5-Eq. (13-a), DE and Zeng and Huai (2014) may be used for conditions with fluctuations in flow depth. Also, M5 has the least sensitivity to velocity for narrow rivers which makes this model suitable for narrow streams with fluctuation in depth and velocity, such as meandering rivers. The results of Monte Carlo simulation showed that
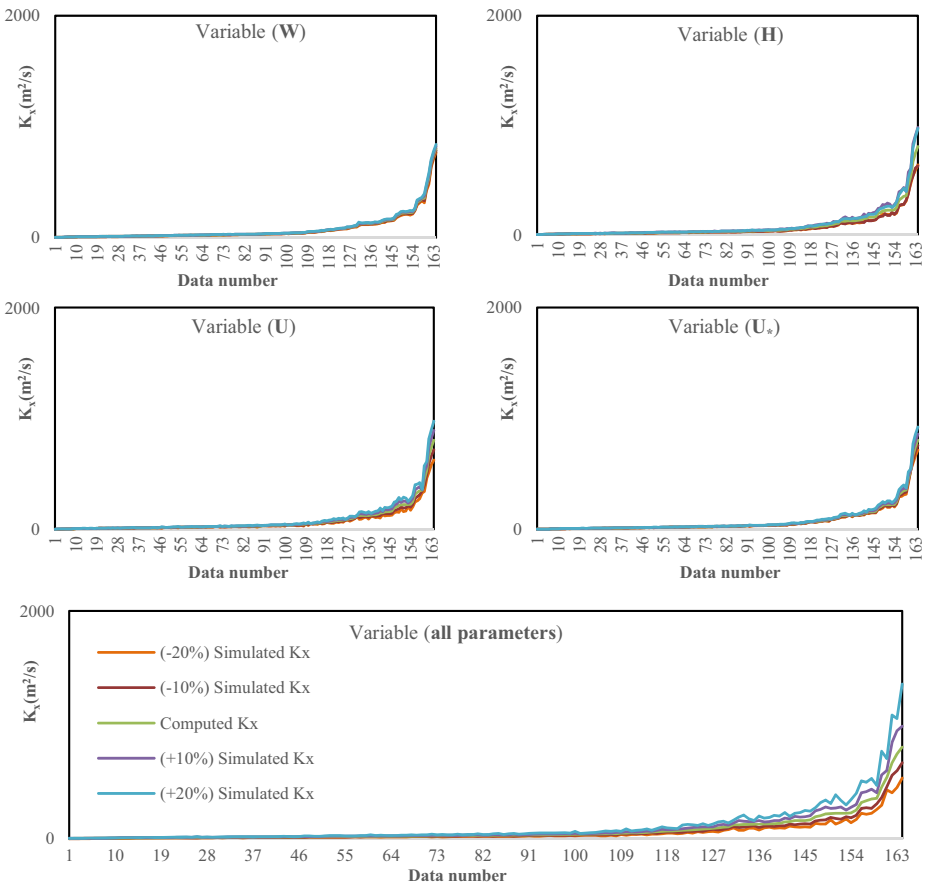
Fig. 10 Performance of GE using Monte Carlo simulation

uncertainty of Kashefipour and Falconer (2002), Sahin (2014), DE and GE models are very high for high LDC values. The M5 model results have some sudden jumps for W/H ratios around 30.6. It seems through this threshold (W/H=30.6), that the M5 model is forced to exploit two equations (Eqs. 13) for the original and the new dataset, making significant difference between the LDC estimated for the two datasets. Some jumps occurred in GE model for high LDC values due to the complexity of the equation of GE model. These jumps could have negative effect on performance of LDC estimation. However, this amount of measured data is a relatively small dataset to properly describe nonlinear processes. The effect of key factors will be better captured using empirical models by continuing to measure longitudinal dispersion in a wide range of streams (Gharabaghi and Sattar 2017).

Compliance with Ethical Standards

Conflict of Interest The authors declare that they have no conflict of interest.

Appendix

Table 8 Geometric, hydraulic and dispersion coefficient datasets used in this study

No	Stream	W (m)	H (m)	U (m/s)	U_* (m/s)	K_x (m²/s)
1	Copper Creek, VA (below gauge)	15.900	0.490	0.210	0.079	19.520
2	Copper Creek, VA (below gauge)	18.300	0.840	0.520	0.100	21.400
3	Copper Creek, VA (below gauge)	16.200	0.490	0.250	0.079	9.500
4	Clinch River, TN (below gauge)	46.900	0.860	0.280	0.067	13.930
5	Clinch River, TN (below gauge)	59.400	2.130	0.860	0.104	53.880
6	Clinch River, TN (below gauge)	53.300	2.090	0.790	0.107	46.450
7	Copper Creek, VA (above gauge)	18.600	0.390	0.140	0.116	9.850
8	Power River, TN	33.800	0.850	0.160	0.055	9.500
9	Clinch River, VA	36.000	0.580	0.300	0.049	8.080
10	Green and Duwamish	21.770	1.580	0.310	0.058	6.500
11	Green and Duwamish	29.610	1.080	0.360	0.048	0.500
12	Bayou Anacoco	19.800	0.410	0.290	0.044	13.940
13	Nooksack River	86.000	2.940	1.200	0.514	153.290
14	Antietam Creek	15.800	0.390	0.320	0.060	9.290
15	Antietam Creek	19.800	0.520	0.430	0.069	16.260
16	Antietam Creek	24.400	0.710	0.520	0.081	25.550
17	Monocacy River	35.100	0.320	0.210	0.040	4.650
18	Monocacy River	36.600	0.450	0.320	0.050	13.940
19	Monocacy River	47.500	0.870	0.440	0.070	37.160
20	Missouri River	182.900	2.230	0.930	0.065	464.520
21	Missouri River	201.200	3.560	1.270	0.082	836.130
22	Missouri River	196.600	3.110	1.530	0.077	891.870
23	Wind/Bighom River	67.100	0.980	0.880	0.110	41.810
24	Elkhom River	32.600	0.300	0.430	0.046	9.290
25	Elkhom River	50.900	0.420	0.460	0.046	20.900
26	John day River	25.000	0.560	1.010	0.137	13.940
27	Comite River	12.500	0.260	0.310	0.043	6.970
28	Comite River	15.800	0.410	0.370	0.055	13.940
29	Amite River	36.600	0.810	0.290	0.068	23.230
30	Amite River	42.400	0.800	0.420	0.068	30.190
31	Sabine River	103.600	2.040	0.560	0.054	315.870
32	Sabine River	127.400	4.750	0.640	0.081	668.900
33	Muddy Creek	13.400	0.810	0.370	0.077	13.940
34	Muddy Creek	19.500	1.200	0.450	0.093	32.520
35	Sabine River,Texas	35.100	0.980	0.210	0.041	39.480
36	White River	67.100	0.550	0.350	0.044	30.190
37	Chattahoochee River	65.500	1.130	0.390	0.075	32.520
38	Susquehanna River	202.700	1.350	0.390	0.065	92.900
39	Antietam Creek	10.970	0.520	0.210	0.075	17.500
40	Antietam Creek	23.470	0.700	0.520	0.101	101.500
41	Antietam Creek	24.990	0.450	0.410	0.081	25.900
42	Antietam Creek	12.800	0.300	0.420	0.057	17.500
43	Antietam Creek	24.080	0.980	0.590	0.098	101.500
44	Antietam Creek	11.890	0.660	0.430	0.085	20.900
45	Antietam Creek	21.030	0.480	0.520	0.069	25.900
46	Monocacy River	48.700	0.550	0.260	0.050	37.800
47	Monocacy River	92.960	0.710	0.160	0.050	41.400
48	Monocacy River	51.210	0.650	0.620	0.040	29.600
49	Monocacy River	97.540	1.150	0.320	0.058	119.800
50	Monocacy River	49.990	0.950	0.320	0.075	29.600
51	Monocacy River	33.530	0.580	0.160	0.041	66.500
52	Monocacy River	40.540	0.410	0.230	0.040	66.500
53	Conococheague Creek	42.210	0.690	0.230	0.064	40.800
54	Conococheague Creek	49.680	0.410	0.150	0.081	29.300
55	Conococheague Creek	42.980	1.130	0.630	0.081	53.300

No	Stream	W (m)	H (m)	U (m/s)	U_* (m/s)	K_x (m²/s)
56	Conococheague Creek	43.280	0.690	0.220	0.064	40.800
57	Conococheague Creek	63.700	0.460	0.100	0.056	29.300
58	Conococheague Creek	59.440	0.760	0.680	0.072	53.300
59	Chattahoochee River	75.600	1.950	0.740	0.138	88.900
60	Chattahoochee River	91.900	2.440	0.520	0.094	166.900
61	Chattahoochee River	99.970	2.500	0.300	0.105	166.900
62	Salt Creek	32.000	0.500	0.240	0.038	52.200
63	Difficult Run	14.500	0.310	0.250	0.062	1.900
64	Difficult Run	11.580	0.400	0.220	0.087	1.900
65	Bear Creek	13.700	0.850	1.290	0.553	2.900
66	Little Pincy Creek	15.900	0.200	0.390	0.053	7.100
67	Bayou Anacoco	17.500	0.450	0.320	0.024	5.800
68	Bayou Anacoco	25.900	0.940	0.340	0.067	27.600
69	Bayou Anacoco	36.600	0.910	0.400	0.067	40.200
70	Comite River	15.700	0.200	0.360	0.040	69.000
71	Comite River	6.100	0.490	0.250	0.058	69.000
72	Bayou Bartholomew	33.400	1.400	0.200	0.030	54.700
73	Bayou Bartholomew	37.490	2.070	0.100	0.040	54.700
74	Amite River	21.300	0.500	0.540	0.027	501.400
75	Amite River	46.020	0.530	0.410	0.043	501.400
76	Tickfau River	14.900	0.590	0.270	0.080	10.300
77	Tickfau River	41.450	1.040	0.070	0.090	10.300
78	Tangipahoa River	31.400	0.810	0.480	0.072	45.100
79	Tangipahoa River	29.900	0.400	0.340	0.020	44.000
80	Tangipahoa River	42.980	1.280	0.260	0.068	45.100
81	Tangipahoa River	31.700	0.760	0.360	0.053	44.000
82	Red River	253.600	0.810	0.480	0.072	45.100
83	Red River	161.500	0.400	0.340	0.020	44.000
84	Red River	152.400	1.620	0.610	0.032	143.800
85	Red River	155.100	3.960	0.290	0.060	130.500
86	Red River	248.110	4.820	0.310	0.065	143.800
87	Sabine River, LA	116.400	3.660	0.450	0.057	227.600
88	Sabine River, LA	160.300	1.740	0.470	0.036	177.700
89	Sabine River, TX	14.200	1.650	0.580	0.054	131.300
90	Sabine River, TX	12.200	2.320	1.060	0.054	308.900
91	Sabine River, TX	21.300	0.500	0.130	0.037	12.800
92	Sabine River, TX	21.640	0.610	0.080	0.042	12.800
93	Sabine River, TX	17.370	1.230	0.040	0.050	14.700
94	Sabine River, TX	31.390	1.430	0.130	0.041	24.200
95	Wind/Bighom Rivers	44.200	1.400	0.990	0.140	184.600
96	Wind/Bighom Rivers	85.300	2.400	1.730	0.150	464.600
97	Copper Creek	16.700	0.500	0.200	0.080	16.800
98	Clinch River	48.500	1.200	0.210	0.070	14.800
99	Copper Creek	18.300	0.400	0.150	0.120	20.700
100	Powell River	36.800	0.900	0.130	0.050	15.500
101	Clinch River	28.700	0.600	0.350	0.070	10.700
102	Copper Creek	19.600	0.800	0.490	0.100	20.800
103	Clinch River	57.900	2.500	0.750	0.100	40.500
104	Conchelaa Canal	24.700	1.600	0.660	0.040	5.900
105	Clinch River	33.530	0.780	0.190	0.049	10.700
106	Clinch River	55.780	2.260	0.690	0.099	36.930
107	Clinch River	53.200	2.400	0.660	0.110	36.900
108	Coachell Canal, CA	23.770	1.600	0.670	0.040	5.960
109	Coachell Canal,CA	24.990	1.540	0.660	0.037	5.920
110	Copper Creek	16.800	0.500	0.240	0.080	24.600
111	Missouri River	180.600	3.300	1.620	0.080	1486.500
112	Bayou Anacoco	25.900	0.900	0.340	0.070	32.500
113	Bayou Anacoco	36.600	0.900	0.400	0.070	39.500

No	Stream	W (m)	H (m)	U (m/s)	U_* (m/s)	K_x (m²/s)
114	Nooksack River	64.000	0.800	0.670	0.270	34.800
115	Wind/Bighom Rivers	59.400	1.100	0.880	0.120	41.800
116	Wind/Bighom Rivers	68.600	2.200	1.550	0.170	162.600
117	John Day River	34.100	2.500	0.820	0.180	65.000
118	Yadkin River	70.100	2.400	0.430	0.100	111.500
119	Yadkin River	71.600	3.800	0.760	0.130	260.100
120	Colorado River	106.100	6.100	0.790	0.088	181.000
121	Colorado River	71.600	8.200	1.200	0.337	243.000
122	Albert	100.000	4.400	0.029	0.002	0.200
123	Dessel-Herentals	35.000	2.500	0.037	0.002	0.200
124	Yuma Mesa A	7.600	3.450	0.680	0.047	0.500
125	Bocholt-Dessel	35.000	2.500	0.107	0.006	1.400
126	Villemsvaart	34.000	2.500	0.130	0.008	1.700
127	Chicago Ship Canal	49.000	8.070	0.270	0.019	3.000
128	Irrigation	1.400	0.190	0.380	0.110	9.600
129	Irrigation	1.500	0.140	0.330	0.100	1.900
130	Puneha	5.000	0.280	0.260	0.210	7.200
131	Kapuni	9.000	0.300	0.370	0.150	8.400
132	Kapuni	10.000	0.350	0.530	0.170	12.400
133	Manganui	20.000	0.400	0.190	0.180	6.500
134	Waiongana	13.000	0.600	0.480	0.240	6.800
135	Stony	10.000	0.630	0.550	0.300	13.500
136	Waiotapu	11.400	0.750	0.410	0.061	8.000
137	Manawatu	59.000	0.720	0.370	0.070	32.000
138	Manawatu	63.000	1.000	0.320	0.094	22.000
139	Manawatu	60.000	0.950	0.460	0.092	47.000
140	Tarawera	25.000	1.210	0.730	0.084	27.000
141	Tarawera	20.000	1.920	0.620	0.123	11.500
142	Tarawera	25.000	1.380	0.770	0.091	20.500
143	Tarawera	25.000	1.400	0.780	0.091	15.500
144	Tarawera	25.000	1.570	0.830	0.096	18.000
145	Tarawera	85.000	2.600	0.690	0.060	52.000
146	Waikato	120.000	2.000	0.640	0.050	67.000
147	Miljacka	11.000	0.290	0.350	0.058	2.700
148	Upper Tame	9.900	0.830	0.460	0.090	5.500
149	Upper Tame	9.900	0.920	0.520	0.100	5.100
150	Credit River	11.210	0.240	0.660	0.160	5.350
151	Credit River	10.990	0.210	0.490	0.140	6.050
152	Credit River	10.820	0.200	0.410	0.130	2.350
153	Credit River (West Credit)	5.570	0.310	0.260	0.090	2.440
154	Credit River (West Credit)	5.540	0.300	0.150	0.090	3.020
155	Credit River (West Credit)	5.100	0.250	0.140	0.090	2.530
156	Credit River (Main Credit)	21.640	0.560	0.380	0.090	10.200
157	Credit River (Main Credit)	20.740	0.480	0.290	0.090	7.370
158	Credit River (Main Credit)	20.310	0.450	0.260	0.080	6.000
159	Credit River (Main Credit)	20.970	0.310	0.980	0.110	8.180
160	Credit River (Main Credit)	19.390	0.260	0.740	0.090	5.510
161	Credit River (Main Credit)	17.170	0.220	0.520	0.090	7.130
162	Credit River (Black Creek)	3.770	0.270	0.230	0.120	2.350
163	Credit River (Black Creek)	3.570	0.200	0.150	0.100	1.900
164	Credit River (Black Creek)	3.580	0.190	0.140	0.100	1.140

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

Alizadeh MJ, Ahmadyar D, Afghantoloee A (2017) Improvement on the existing equations for predicting longitudinal dispersion coefficient. Water Resour Manag 31:1777–1794

Antonopoulos VZ, Georgiou PE, Antonopoulos ZV (2015) Dispersion coefficient prediction using empirical models and ANNs. Environmental Processes 2:379–394. https://doi.org/10.1007/s40710-015-0074-6

Atkinson TC, Davis PM (2000) Longitudinal dispersion in natural channels: l. experimental results from the river Severn, U.K. Hydrol Earth Syst Sci Discuss 4:345–353

Azamathulla HM, Wu F-C (2011) Support vector machine approach for longitudinal dispersion coefficients in natural streams. Appl Soft Comput 11:2902–2905

Beven K (1979) A sensitivity analysis of the Penman-Monteith actual evapotranspiration estimates. J Hydrol 44:169–190. https://doi.org/10.1016/0022-1694(79)90130-6

Chatila GJ (1997) Modeling of pollutant transfer in compound open channels. PhD dissertation. University of Ottawa, Ontario

Davis PM, Atkinson TC, Wigley TML (2000) Longitudinal dispersion in natural channels: 2. the roles of shear flow dispersion and dead zones in the river Severn, U.K. Hydrol Earth Syst Sci Discuss 4:355–371

Dawson R, Hall J, Sayers P, Bates P, Rosu C (2005) Sampling-based flood risk analysis for fluvial dike systems. Stoch Env Res Risk A 19:388–402

Deng Z-Q, Singh VP, Bengtsson L (2001) Longitudinal dispersion coefficient in straight rivers. J Hydraul Eng 127:919–927. https://doi.org/10.1061/(ASCE)0733-9429(2001)127:11(919)

Disley T, Gharabaghi B, Mahboubi AA, McBean EA (2015) Predictive equation for longitudinal dispersion coefficient. Hydrol Process 29:161–172. https://doi.org/10.1002/hyp.10139

Elder JW (1959) The dispersion of marked fluid in turbulent shear flow. J Fluid Mech 5:544–560. https://doi.org/10.1017/S0022112059000374

Etemad-Shahidi A, Taghipour M (2012) Predicting longitudinal dispersion coefficient in natural streams using M5' model tree. J Hydraul Eng 138:542–554

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

Fischer HB (1968) Dispersion predictions in natural streams. J Sanit Eng Div 94:927–944

Gardner RH, O'neill RV, Mankin JB, Carney JH (1981) A comparison of sensitivity analysis and error analysis based on a stream ecosystem model. Ecol Model 12(3):173–190

Gharabaghi B, Sattar AMA (2017) Empirical models for longitudinal dispersion coefficient in natural streams. J Hydrol. https://doi.org/10.1016/j.jhydrol.2017.01.022

Godfrey RG, Frederick BJ (1970) Stream dispersion at selected sites, US Government printing office, Washington

Gong L, Cy X, Chen D, Halldin S, Chen YD (2006) Sensitivity of the Penman–Monteith reference evapotranspiration to key climatic variables in the Changjiang (Yangtze River) basin. J Hydrol 329:620–629. https://doi.org/10.1016/j.jhydrol.2006.03.027

Graf JB (1995) Measured and predicted velocity and longitudinal dispersion at steady and unsteady flow, Colorado River, Glen Canyon Dam to Lake Mead. J Am Water Resour Assoc 31:265–281. https://doi.org/10.1111/j.1752-1688.1995.tb03379.x

Guymer I (1998) Longitudinal dispersion in sinuous channel with changes in shape. J Hydraul Eng 124:33–40. https://doi.org/10.1061/(ASCE)0733-9429(1998)124:1(33)

Haghiabi AH (2016) Prediction of longitudinal dispersion coefficient using multivariate adaptive regression splines. J Earth Syst Sci 125:985–995

Haghiabi AH (2017) Modeling river mixing mechanism using data driven model. Water Resour Manag 31:811–824

Hall JW, Boyce SA, Wang Y, Dawson RJ, Tarantola S, Saltelli A (2009) Sensitivity analysis for hydraulic models. J Hydraul Eng 135:959–969

Hamby DM (1994) A review of techniques for parameter sensitivity analysis of environmental models. Environ Monit Assess 32:135–154. https://doi.org/10.1007/bf00547132

Ho DT, Schlosser P, Caplow T (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

Huber PJ (1981) Robust Statistics. John Wiley & Sons, Inc., New York

Hupet F, Vanclooster M (2001) Effect of the sampling frequency of meteorological variables on the estimation of the reference evapotranspiration. J Hydrol 243:192–204

Kashefipour SM, Falconer RA (2002) Longitudinal dispersion coefficients in natural channels. Water Res 36: 1596–1608. https://doi.org/10.1016/S0043-1354(01)00351-7

Li X, Liu H, Yin M (2013) Differential evolution for prediction of longitudinal dispersion coefficients in natural streams. Water Resour Manag 27:5245–5260. https://doi.org/10.1007/s11269-013-0465-2

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

McCuen RH (1973) The role of sensitivity analysis in hydrologic modeling. J Hydrol 18:37–53. https://doi.org/10.1016/0022-1694(73)90024-3

McCuen RH (1974) A sensitivity and error analysis of procedures used for estimating evaporation. J Am Water Resour Assoc 10:486–497

McQuivey RS, Keefer TN (1974) Simple method for predicting dispersion in streams. J Environ Eng Div 100: 997–1011

Mishra S (2009) Uncertainty and sensitivity analysis techniques for hydrologic modeling. J Hydroinf 11:282–296

Mount NJ, Dawson CW, Abrahart RJ (2013) Legitimising data-driven models: exemplification of a new data-driven mechanistic modelling framework. Hydrol Earth Syst Sci 17:2827–2843

Nakhaei N, Etemad-Shahidi A (2012) Applying Monte Carlo and classification tree sensitivity analysis to the Zayandehrood River. J Hydroinf 14:236–250

Noori R, Ghiasi B, Sheikhian H, Adamowski JF (2017) Estimation of the dispersion coefficient in natural rivers using a granular computing model. J Hydraul Eng 143(5):04017001

Nordin CF, Sabol GV (1974) Empirical data on longitudinal dispersion in rivers. US Geological Survey, Water Resources Investigations, Report No. 74-20, p 332

Papadimitrakis I, Orphanos I (2004) Longitudinal dispersion characteristics of rivers and natural streams in Greece. Water Air Soil Poll Focus 4:289–305. https://doi.org/10.1023/b:wafo.0000044806.98243.97

Parsaie A, Haghiabi AH (2017) Computational modeling of pollution transmission in rivers. Appl Water Sci 7: 1213–1222. https://doi.org/10.1007/s13201-015-0319-6

Pasha M, Lansey K (2010) Effect of parameter uncertainty on water quality predictions in distribution systems-case study. J Hydroinf 12:1–21

Radwan M, Willems P, Berlamont J (2004) Sensitivity and uncertainty analysis for river quality modelling. J Hydroinf 6:83–99

Rana G, Katerji N (1998) A measurement based sensitivity analysis of the Penman-Monteith actual evapotranspiration model for crops of different height and in contrasting water status. Theor Appl Climatol 60:141–149. https://doi.org/10.1007/s007040050039

Rutherford JC (1994) River Mixing. John Wiley & Sons, ltd, Chichester

Sahay RR (2011) Prediction of longitudinal dispersion coefficients in natural rivers using artificial neural network. Environ Fluid Mech 11:247–261. https://doi.org/10.1007/s10652-010-9175-y

Sahay RR, Dutta S (2009) Prediction of longitudinal dispersion coefficients in natural rivers using genetic algorithm. Hydrol Res 40:544–552. https://doi.org/10.2166/nh.2009.014

Sahin S (2014) An empirical approach for determining longitudinal dispersion coefficients in rivers. Environmental Processes 1:277–285. https://doi.org/10.1007/s40710-014-0018-6

Saltelli A, Tarantola S, Campolongo F, Ratto M (2004) Sensitivity analysis in diagnostic modelling: Monte Carlo filtering and regionalised sensitivity analysis, Bayesian uncertainty estimation and global sensitivity analysis. In: Sensitivity Analysis in Practice. John Wiley & Sons, ltd, pp 151–192. https://doi.org/10.1002/0470870958.ch6

Saltelli A, Ratto M, Andres T, Campolongo F, Cariboni J, Gatelli D, Saisana M, Tarantola S (2008) Global sensitivity analysis: the primer. John Wiley & Sons, ltd, Chichester

Sattar AMA, Gharabaghi B (2015) Gene expression models for prediction of longitudinal dispersion coefficient in streams. J Hydrol 524:587–596. https://doi.org/10.1016/j.jhydrol.2015.03.016

Saxton KE (1975) Sensitivity analyses of the combination evapotranspiration equation. Agric Meteorol 15:343–353. https://doi.org/10.1016/0002-1571(75)90031-X

Seo IW, Cheong TS (1998) Predicting longitudinal dispersion coefficient in natural streams. J Hydraul Eng 124: 25–32. https://doi.org/10.1061/(ASCE)0733-9429(1998)124:1(25)

Tayfur G, Singh VP (2005) Predicting longitudinal dispersion coefficient in natural streams by artificial neural network. J Hydraul Eng 131(11):991–1000

Taylor G (1954) The dispersion of matter in turbulent flow through a pipe. Proc R Soc Lond A Mat Sci 223:446–468. https://doi.org/10.1098/rspa.1954.0130

Velísková Y, Sokáč M, Halaj P, Koczka Bara M, Dulovičová R, Schügerl R (2014) Pollutant spreading in a small stream: a case study in Mala Nitra canal in Slovakia. Environmental Processes 1:265–276

White W, Milli H, Crabbe A (1973) Sediment transport: an appraisal methods, Vol. 2: Performance of theoretical methods when applied to flume and field data. Hydr Res Station Rep. N. IT 119, Wallingford

Yotsukura N, Fischer HB, Sayre WW (1970) Measurement of mixing characteristics of the Missouri River between Sioux City, Iowa, and Plattsmouth, Nebraska. No. 1899-G. USGPO

Zeng Y, Huai W (2014) Estimation of longitudinal dispersion coefficient in rivers. J Hydro Environ Res 8:2–8