

New formulations for prediction of velocity at limit of deposition in storm sewer based on a stochastic technique

Ali Tafarajnoruz and Ahmad Sharafati

ABSTRACT

Sedimentation in storm sewer strongly depends on velocity at limit of deposition. This study provides application of a novel stochastic-based model to predict the densimetric Froude number in sewer pipes. In this way, the Generalized Likelihood Uncertainty Estimation (GLUE) is used to develop two parametric equations, called GLUE based four-parameter (GBFP) and GLUE based two-parameter (GBTP) models to enhance the prediction accuracy of the velocity at the limit of deposition. A number of performance indices are calculated in training and testing phases to compare the developed models with the conventional regression-based equations available in the literature. Based on the obtained performance indices and some graphical techniques, the research findings confirm that a significant enhancement in prediction performance is achieved through the proposed GBTP compared with the previously developed formulas in the literature. To make a quantified comparison between the established and literature models, an index, called improvement index (IM), is computed. This index is a resultant of all the selected indices, and this indicator demonstrates that GBTP is capable of providing the most performance improvement in both training ($IM_{train} = 9.2\%$) and testing ($IM_{test} = 11.3\%$) phases, comparing with a well-known formula in this context.

Key words | densimetric froude number, generalized likelihood uncertainty estimation, prediction, storm sewer

HIGHLIGHTS

- The stochastic based model is developed predict the densimetric Froude number in storm sewer.
- GLUE approach is used as a parameters estimator prior the parametric formulas.
- The proposed model is validated against literature formulas.
- The proposed stochastic based model is exhibited a reliable and robust prediction methodology.

INTRODUCTION

Dry weather seasons may lead to a decrease of flow rate and water level, causing deposition of solid materials in many storm sewer networks. If the deposits remain for a relatively long duration within the sewer system, sediment characteristics change, and the deposits may permanently become consolidated. This phenomenon not only influences the sewer hydraulic resistance, flow properties, and the capacity of sediment motion, it may also significantly reduce the

flood conveyance capacity of the sewer system and even in sever condition it might lead to network deterioration (Ab. Ghani & Md. Azamathulla 2011; Ebtehaj Bonakdari *et al.* 2014).

Hydraulic designers usually take into account two main criteria in the design of sewer systems. These criteria are defined based on low discharge and high flow conditions: the sewer should convey the design floods; while, the

Ali Tafarajnoruz

Dipartimento di Ingegneria Civile,
Università della Calabria,
Cubo 42B, Rende,
Italy

Ahmad Sharafati (corresponding author)

Institute of Research and Development, Duy Tan
University,
Da Nang 550000,
Vietnam;
Faculty of Civil Engineering,
Duy Tan University,
Da Nang 550000,
Vietnam
and
Department of Civil Engineering, Science and
Research Branch,
Islamic Azad University,
Tehran,
Iran
E-mail: ahmadsharafati@duytan.edu.vn

sediment deposition in the network should be minimized during the low flow seasons. Combined sewer systems, which carry rainfall and wastewater, are mostly designed based on frequency and intensity of rainfall and runoff to flush out the deposited sediment of the dry weather periods (Vongvisessomjai Tingsanchali *et al.* 2010).

The earliest and most common attitude of the sewer design has been identified based on a minimum flow velocity to scouring of the existing of bed sediment. In this context, (Camp 1942) suggested an equation for the minimum velocity to scour the deposited bed sediments as follows:

$$V_{Camp} = \frac{1}{n} R_h^{\frac{1}{6}} \sqrt{B \left(\frac{\rho_s}{\rho} - 1 \right) d} \quad (1)$$

where B denotes a constant, which is 0.8 for adequate self-cleansing of sewers; n and d represent the Manning's roughness coefficient and sediment diameter; ρ_s and ρ are the density of sediment and water, and R_h being the hydraulic radius. Although some researchers followed the idea of the minimum flow velocity of self-cleansing sewers during the 1950s, another concept of scouring of the deposited sediments was later developed in 1970s taking into account the minimum bed shear stress for incipient of sediment motion (Novak & Nalluri 1975).

Further studies showed that the mechanism of sediment transport in sewers is a complex three-dimensional phenomenon and additional parameters than those assumed by (Camp 1942) should be considered to obtain acceptable results. It is worthy to note that the overall mechanism of sediment transport in sewer pipes is almost similar to what occurs in natural channels. Two main differences in the mechanism of sediment transport in alluvial channels and sewers are generally considered. First, the supply amount of sediments in a natural channel is practically unlimited and originated from the channel itself. In contrast, in sewers, the sediment supply rate is derived from the catchment of the sewer network. Second, the effective roughness of sewers may vary depending on whether the internal surface of sewer pipes is clean, or it is made up of deposits of sediment particles (Ab. Ghani 1993).

The available criteria were modified, and more reliable design methodologies were suggested based on non-deposition of sediment materials rather than scouring of the deposited sediment. To this end, several regression-based equations accounting more parameters have been derived during the past 40 years. Those parameters, as well as the

proposed equations, will be described in the following section.

Owing to the complex nature of some hydraulic phenomena, researchers in recent decades utilized several sophisticated artificial intelligent (AI) models to achieve more accurate predictions than the conventional regression-based formulas (Sharafati Haghbin *et al.* 2019). For instance, an AI technique, named group method of data handling (GMDH), has been optimized to estimate longitudinal dispersion coefficient in rivers (Najafzadeh & Tafarjnoruz 2016). Application of other AI methodologies, like evolutionary polynomial regression (EPR), gene-expression programming (GEP), and model tree (MT) have been widely evaluated by researchers in the field of hydraulic engineering to predict sediment transport, scouring around hydraulic structures, design of riprap stone size, etc. (Azamathulla & Zakaria 2011; Azamathulla 2012; Guven & Azamathulla 2012; Najafzadeh Barani *et al.* 2013; Najafzadeh Barani *et al.* 2014; Najafzadeh 2016; Najafzadeh Tafarjnoruz *et al.* 2017; Najafzadeh Rezaie-Balf *et al.* 2018; Kargar Safari *et al.* 2019; Sharafati Tafarjnoruz *et al.* 2019).

According to (Azamathulla & Ghani 2011; Azamathulla Ab. Ghani *et al.* 2012; Ebtehaj & Bonakdari 2013; Ebtehaj & Bonakdari 2017; Najafzadeh & Bonakdari 2017), artificial neural networks (ANNs), adaptive neuro-fuzzy inference systems (ANFIS), gene-expression programming (GEP), and Neuro-Fuzzy GMDH (NF-GMDH) were successfully utilized to predict velocity at limit of deposition. More specifically, some investigation devoted to applying Evolutionary Polynomial Regression (EPR), support vector regression (SVR), and the firefly algorithm (FFA) resulted accurate sediment transport prediction in pipes (Ebtehaj & Bonakdari 2016; Montes Berardi *et al.* 2020). Authors of these studies reported significant improvement in the prediction of the selected parameter comparing with the conventional regression-based formulas. However, some of the selected AI techniques are not capable of introducing a precise predictive method for practical purposes or in some cases; the proposed equation has an extremely complicated mathematical formulation.

Apart from use of a regression-based equation or an AI technique, some countries defined a single minimum non-deposition velocity magnitude for design purposes. For example, in the USA, the minimum velocity within the foul sewers must be 0.60 m/s; however, a larger value of 0.90 m/s should be considered in storm sewers. It is interesting to note that such values vary in other countries. Take the UK as another example; different values of 0.75 and 1 m/s were, respectively, adopted for storm and combined sewers.

A survey on available studies clarifies that a dozen of such velocity limits have already defined by different authorities, yet it is obvious that taking into account a single minimum velocity looks inadequate and unsafe to design storm sewer pipelines of various roughness, diameter, gradient and sediment properties. These criteria may be only acceptable for the design of small sewer networks (Vongvisessomjai Tingsanchali *et al.* 2010; Ebtehaj Bonakdari *et al.* 2014).

The mentioned methods and the available formulas in the design of non-deposition sediments have been developed according to the flow and sediment characteristics; while, engineers usually assess a combination of operation, maintenance and construction in the design of a sewer system. For example, to reduce the construction cost, a limited sediment disposition may be allowed within the sewer pipelines. This less conservative design significantly diminishes the slope of sewer lines and as a result, reduces the construction cost, even though it requires specific maintenance and operation.

All the methodologies, as mentioned earlier (i.e., regression-based equations, artificial intelligent techniques, pre-defined non-deposition velocity values) have been developed assuming that the 'sediment transport' is a deterministic phenomenon. This hypothesis stands in contrast to the physical nature of the sediment transport process. Most available environmental models are developed to simulate the mechanism of a system response to some influencing factors which are not necessarily unique. In this case, a stochastic methodology may offer a more reliable relationship between the system inputs and the model output (Sadegh Shakeri Majd *et al.* 2018; Shojaeezadeh Nikoo *et al.* 2018). Sediment transport process, whether in nature (e.g., rivers, coastal zones, etc.) or in manmade structures (e.g., pipelines, sewer networks, etc.) is inherently stochastic. Nevertheless, little research has been already conducted to take into account the intrinsic property of bed material motion which may considerably improve sediment transport prediction (Kleinhans & Van Rijn 2002; Dodaro Tafarjnoruz *et al.* 2014; Dodaro Tafarjnoruz *et al.* 2016; Sharafati Yasa *et al.* 2018).

To quantify the uncertainty of random variables, several techniques like GLUE; and Sequential Uncertainty Fitting (SUFI) have already proposed (Beven & Binley 1992; Abbas-pour Johnson *et al.* 2004). In the field of water resource engineering and sediment transport, GLUE has been proven as a useful tool to include the uncertainty in a wide range of predictions. In particular, in hydrological models, it may recognize the fundamental limitations in the rainfall-runoff process (Beven 1993; Freer Beven *et al.*

1996). For example, the GLUE methodology was employed to analyze the uncertainty of a fully distributed physically-based code in terms of streamflow prediction limits in France (Vázquez Beven *et al.* 2009). It has also been applied to flood estimation by fitting the model and identifying the model parameters in river flow prediction in Vietnam (Cu Ball *et al.* 2018). The GLUE technique does not assume only one optimal parameter set in deriving a model: this is one key issue and merit of this methodology (Duan Sorooshian *et al.* 1992; Beven 1993). Taking this advantage, (Sharafati Yasa *et al.* 2018) improved the prediction accuracy of wave-induced scour depth around a pipeline, assuming the scouring process and the attributable sediment transport are stochastic phenomena.

The main objective of the present investigation is to revise the most common sediment transport mathematical formulation in sewer pipelines utilizing the stochastic technique, GLUE. Two mathematical structures have been considered in the present study, which includes the most effective dimensionless groups. The performance of the new stochastic formulae in the prediction of non-deposition velocity limit is evaluated by means of several statistical indices.

DIMENSIONLESS PARAMETERS AND EMPIRICAL EQUATIONS

Previous studies revealed that cross-sectional flow geometry, sediment properties, and flow characteristics affect the velocity at limit of deposition (Ab. Ghani 1993; Vongvisessomjai Tingsanchali *et al.* 2010; Ebtehaj & Bonakdari 2013). Equation (2) briefly presents a functional relationship between V_L and the effective parameters through an unknown function, f_1 :

$$V_L = f_1(g, C_{vL}, R_h, \lambda_s, d_{50}, \rho, \rho_s, \nu) \quad (2)$$

where g being acceleration due to gravity; C_{vL} denotes sediment concentration at limit of deposition; λ_s is the friction factor in the presence of sediment motion; d_{50} represent median sediment grain size; ν is the kinematic viscosity of water.

Through the application of the well-known Buckingham π -theorem along with some mathematical manipulations, five variables in terms of dimensionless groups are obtained from Equation (2). The resulted dimensionless parameters can be expressed as a new unknown function, f_2 :

$$\frac{V_L}{\sqrt{g\Delta d_{50}}} = f_2\left(C_{vL}, \frac{R_h}{d_{50}}, \lambda_s, D_{gr}\right) \quad (3)$$

in which $V_L/\sqrt{g\Delta d_{50}} = F_{dL}$ denotes the densimetric Froude number at the limit of deposition; $\Delta = \frac{\rho_s}{\rho} - 1$ represents the

relative density of sediment in water; and $D_{gr} = d_{50} \left[\frac{g\Delta}{v^2} \right]^{1/3}$ is dimensionless grain size. λ_s is a number around the wall friction factor under the clear-water condition $\lambda_0 = 8gn^2/R_h^{1/3}$, even though two other dimensionless parameters, i.e. C_{vL} and D_{gr} , may slightly affect it. The following equation is recommended to calculate λ_s (Ab. Ghani 1993):

$$\lambda_s = 1.13 \lambda_0^{0.98} C_{vL}^{0.02} D_{gr}^{0.01} \quad (4)$$

Since the 1990s, several semi-empirical equations have been proposed to predict F_{dL} adopting all or some of the dimensionless groups in Equation (3). (Ab. Ghani 1993) took into account all the governing dimensionless groups of Equation (3) and proposed Equation (5). Later, (Vongvisessomjai Tingsanchali et al. 2010) [Equation (6)], and (Ebtehaj Bonakdari et al. 2014) [Equation (7)] assumed C_{vL} and $\frac{R_h}{d_{50}}$ are the main influencing parameters on F_{dL} . This hypothesis was originally introduced by (Mayerle Nalluri et al. 1991) who neglected λ_s and D_{gr} aiming at ‘a quick calculation of F_{dL} .

$$\frac{V_L}{\sqrt{g\Delta d_{50}}} = 3.08 D_{gr}^{-0.09} C_{vL}^{0.21} \left(\frac{R_h}{d_{50}} \right)^{0.53} \lambda_s^{-0.21} \quad (5)$$

$$\frac{V_L}{\sqrt{g\Delta d_{50}}} = 4.31 C_{vL}^{0.226} \left(\frac{R_h}{d_{50}} \right)^{0.616} \quad (6)$$

$$\frac{V_L}{\sqrt{g\Delta d_{50}}} = 4.49 C_{vL}^{0.21} \left(\frac{R_h}{d_{50}} \right)^{0.54} \quad (7)$$

DESCRIPTION OF DATASETS

To develop and evaluate the new model, and comparing its performance with those available in the literature, a large number of datasets, including 257 datasets were extracted from two experimental studies (Ab. Ghani 1993; Vongvisessomjai Tingsanchali et al. 2010). To train the proposed stochastic model, 193 datasets (approximately 75% of the entire data) were selected randomly. In comparison, the 64 datasets (25% of the available data) were allocated for the testing phase and verifying the proposed model. Table 1 furnishes the range of each dimensionless parameter employed for training and testing stages while their frequency is presented in Figure 1.

DEVELOPMENT OF GLUE BASED FORMULAS

GLUE is a stochastic approach to measure the model uncertainty, which was first introduced in 1992 (Beven & Binley 1992). Recently, GLUE has been used as parameters estimator in different engineering fields (Sharafati Yasa et al. 2018). To achieve the best parameters set, several random parameters sets would be generated using their probability distribution functions. Then, the generated parameters sets are stored as behavioral and non-behavioral samples based on the behavioral threshold value. To sort the behavioral samples, it is necessary to compute the re-scaled likelihood weights (w_r^i) of the of i^{th} set of the parameters as follows (Wang Frankenberger et al. 2006):

$$\begin{cases} w_r^i = \frac{w^i}{\sum_{j=1}^k w^j} \\ w^i = \exp\left(-\frac{RMSE^i}{\min(RMSE)}\right) \end{cases} \quad (8)$$

Table 1 | Statistical characteristics of the dimensionless parameters of the datasets

Parameter	Training phase					Testing phase				
	F_{dL}	C_{vL}	$\frac{R_h}{d_{50}}$	λ_s	D_{gr}	F_{dL}	C_{vL}	$\frac{R_h}{d_{50}}$	λ_s	D_{gr}
Minimum	1.35	0.76	4.46	0.013	5.06	1.27	5.00	3.86	0.016	5.06
Maximum	13.49	1450	188.3	0.053	207.6	9.44	1415	188.5	0.053	207.6
Average	4.09	211.8	49.30	0.026	78.46	4.04	242.9	45.61	0.026	78.83
SD	2.19	312.8	52.96	0.009	64.73	2.13	326.6	50.14	0.010	67.97
COV	0.54	1.48	1.07	0.355	0.82	0.53	1.34	1.10	0.375	0.86

Note: SD = standard deviation; COV = coefficient of variation.

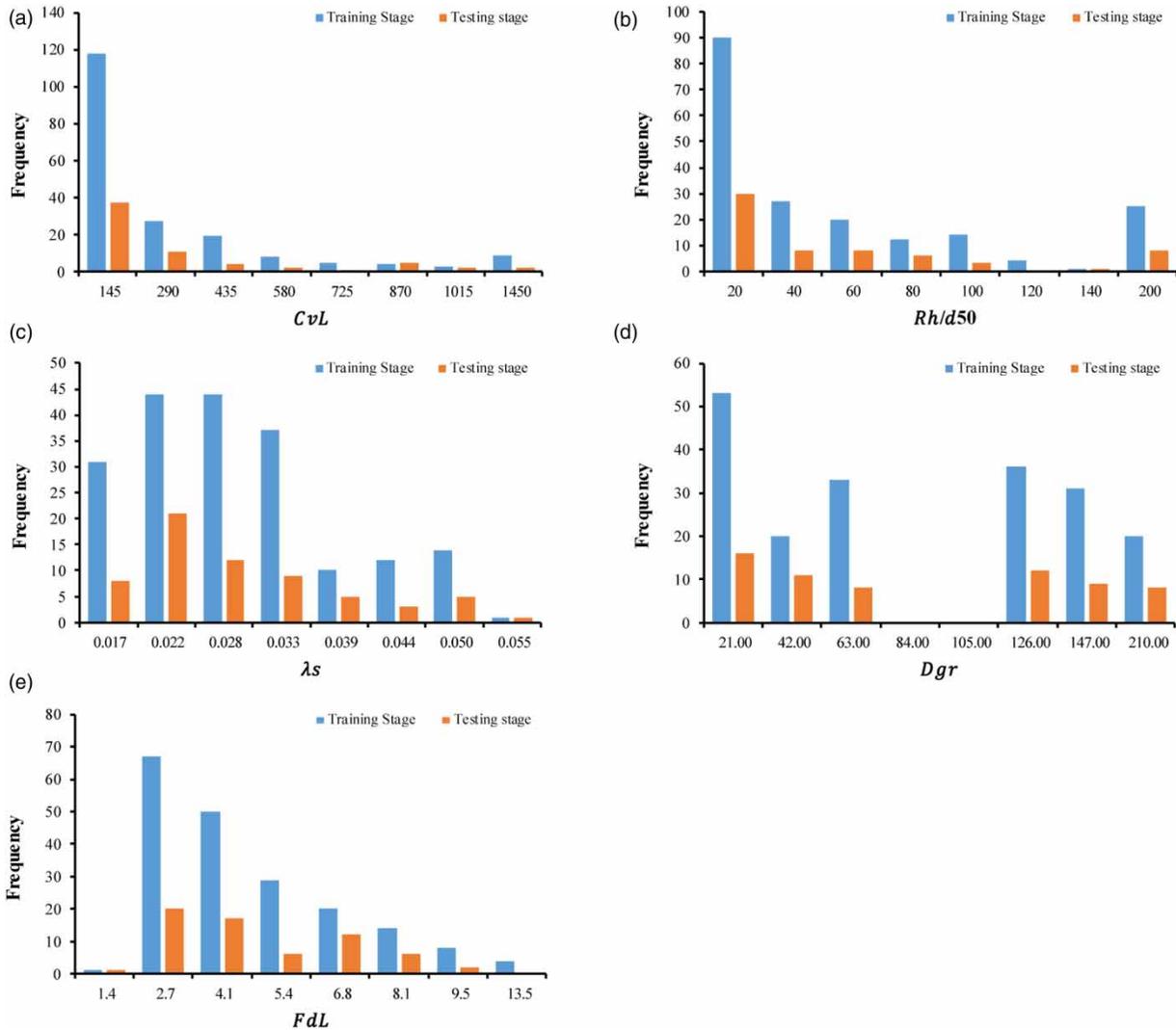


Figure 1 | Frequency of the experimental datasets in the training and testing phases. (a) C_{vL} , (b) $\frac{R_h}{d_{50}}$, (c) λ_s , (d) D_{gr} , (e) F_{dL} .

where, w^i is the likelihood measure of i^{th} behavioral set and k is the number of behavioral samples. Having considered the uniform distribution as a prior probability distribution of each parameter, the posterior distributions of the parameters would be obtained using their re-scaled likelihood weights (Equation (8)) (Freni Mannina *et al.* 2009).

To obtain the GLUE based predictive models, a parametric equation is proposed in the form of the following equation:

$$\frac{V_L}{\sqrt{g\Delta d_{50}}} = a_1 D_{gr}^{a_2} C_{vL}^{a_3} \left(\frac{R_h}{d_{50}}\right)^{a_4} \lambda_s^{a_5} \quad (9)$$

where $a_1 \sim a_5$ are the random variables estimated using the GLUE algorithm. This model is named as “GLUE Based

With Four Parameters (GBFP)” which includes all the dimensionless parameters (D_{gr} , C_{vL} , $\frac{R_h}{d_{50}}$, λ_s).

To develop another alternative GLUE based model with a simple structure, the employed predictive (input) variables are ranked using the Gamma Test (GT) technique. GT selects the best input variables for nonlinear problems (Sharafati Asadollah *et al.* 2020). The best input variables are selected via computing an index named ϑ – ratio by means of a nonlinear approach. Consider a given dataset as the following form:

$$\{(x_i, y_i), 1 \leq i \leq M\} \quad (10)$$

where the x_i are the m-dimensional input vectors and y_i are the output values.

GT is originated from x_i in delta function (δ) expressed as follows:

$$\delta_M(k) = \frac{1}{M} \sum_{i=1}^M |x_{N[i,k]} - x_i|^2 \quad (1 \leq k \leq p) \quad (11)$$

where $|\dots|$ is the Euclidean distance and $x_{N[i,k]}$ is the k^{th} ($1 \leq k \leq p$) nearest neighbors for each x_i ($1 \leq i \leq M$).

Then the gamma function of the y_i would be computed using the following equation:

$$\gamma_M(k) = \frac{1}{2M} \sum_{i=1}^M |y_{N[i,k]} - y_i|^2 \quad (1 \leq k \leq p) \quad (12)$$

where $y_{N[i,k]}$ is the corresponding y for the k^{th} nearest neighbor of x_i in $\delta_M(k)$.

An estimate of the model output variance is provided by gamma statistic (Γ). This statistic would be computed by a regression line through p points of $[\delta_M(k), \gamma_M(k)]$ as follow:

$$\gamma = A\delta + \Gamma \quad (13)$$

The gamma statistic can be standardized by measuring another term, ϑ - ratio, which is defined as:

$$\vartheta - ratio = \frac{\Gamma}{\sigma^2(y)} \quad (14)$$

where $\sigma^2(y)$ is the output variance.

ϑ - ratio is in range of 0 to 1: the zero value indicates high accuracy in output prediction.

To obtain a simple GLUE based model, the values of ϑ - ratio from different two variate combinations are represented in Table 2.

Table 2 | Gamma-test statistics of different combinations of the dimensionless parameters

Statistic	ϑ - ratio	Parameters			
		C_{vL}	$\frac{R_h}{d_{50}}$	λ_s	D_{gr}
0.49	0.103	✓	✓		
0.68	0.144	✓			✓
1.06	0.225		✓		✓
1.20	0.253			✓	✓
1.43	0.302		✓	✓	
3.94	0.832	✓		✓	

Table 2 indicates that the most influencing parameters are C_{vL} and $\frac{R_h}{d_{50}}$. The achieved result of GT technique has good consistency with previous developed two parameters models (Vongvisessomjai Tingsanchali *et al.* 2010; Ebtehaj Bonakdari *et al.* 2014) which employed C_{vL} and $\frac{R_h}{d_{50}}$ as the main effective variables on F_{dL} . Hence, another parametric equation would be developed as follows:

$$\frac{V_L}{\sqrt{g\Delta d_{50}}} = b_1 C_{vL}^{b_2} \left(\frac{R_h}{d_{50}}\right)^{b_3} \quad (15)$$

where $b_1 \sim b_3$ are the random variables which are estimated by means of the GLUE algorithm. This model is named as “GLUE Based With Two Parameters (GBTP)” which comprises fewer but the most influencing parameters for quick predictions.

The prior probability density functions of the random variables are obtained using the cross-validation method. Parameters of prior functions are presented in Table 3. To compute the random variables using the GLUE algorithm, the posterior probability density functions of the random variables are obtained by Equation (8) and an approach which is demonstrated in Figure 2.

DESCRIPTION OF AGREEMENT AND ERROR INDICES

To make a comparison among proposed GLUE based models and previous alternative formulas, a number of agreement and error indices such as Scatter Index (SI), Mean Absolute Error (MAE), Nash-Sutcliffe coefficient

Table 3 | Parameters of prior distribution of the defined random coefficients

Random coefficients	PDF	PDF parameters	
		(Lower limit)	(Upper limit)
a_1	Uniform	0	5
a_2		-1	0
a_3		0	0.4
a_4		-1	1
a_5		-1	0
b_1		0	5
b_2		0	0.4
b_3		-1	1

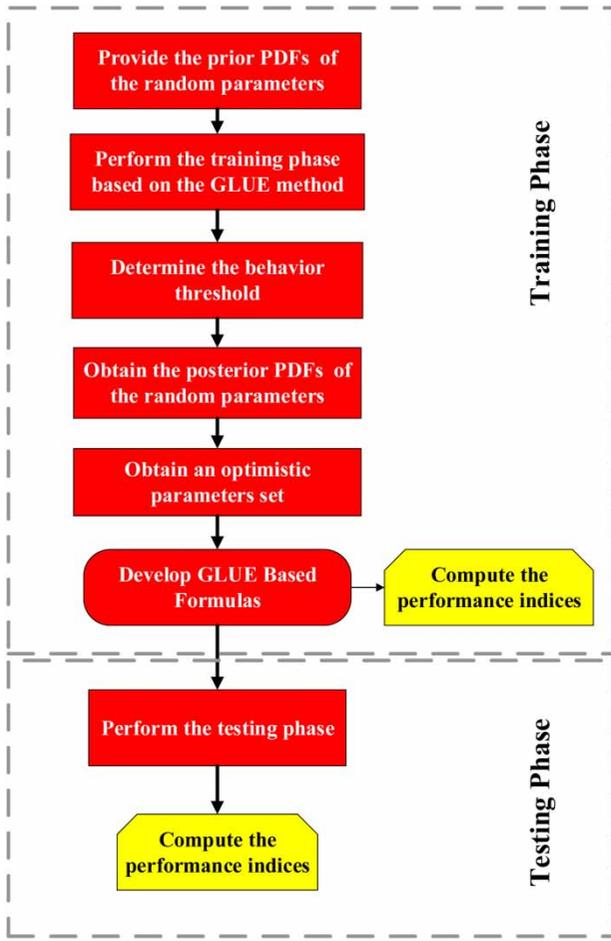


Figure 2 | Developing the GLUE based model for predicting the densimetric Froude number.

(NSE), Legates and McCabe’s Index (LMI), and Coefficient of determination (R^2) are computed as follows:

$$SI = \frac{\left(\sqrt{\frac{1}{n} \sum_{i=1}^n ((F_{dL})_{Obs,i} - (F_{dL})_{Sim,i})^2} \right)}{\overline{(F_{dL})_{Obs}}} \quad (16)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |(F_{dL})_{Obs,i} - (F_{dL})_{Sim,i}| \quad (17)$$

$$NSE = 1 - \left[\frac{\sum_{i=1}^n ((F_{dL})_{Obs,i} - (F_{dL})_{Sim,i})^2}{\sum_{i=1}^n ((F_{dL})_{Obs,i} - \overline{(F_{dL})_{Obs}})^2} \right] \quad (18)$$

$$LMI = 1 - \left[\frac{\sum_{i=1}^n |(F_{dL})_{Obs,i} - (F_{dL})_{Sim,i}|}{\sum_{i=1}^n |(F_{dL})_{Obs,i} - \overline{(F_{dL})_{Obs}}|} \right] \quad (19)$$

$$R^2 = \left(\frac{\sum_{i=1}^n ((F_{dL})_{Obs,i} - \overline{(F_{dL})_{Obs}})((F_{dL})_{Sim,i} - \overline{(F_{dL})_{Sim}})}{\sqrt{\sum_{i=1}^n ((F_{dL})_{Obs,i} - \overline{(F_{dL})_{Obs}})^2 \sum_{i=1}^n ((F_{dL})_{Sim,i} - \overline{(F_{dL})_{Sim}})^2}} \right)^2 \quad (20)$$

where the $(F_{dL})_{Obs,i}$ and $(F_{dL})_{Sim,i}$ represent the i^{th} measured and predicted densimetric Froude numbers at the limit of deposition, $\overline{(F_{dL})_{Obs}}$ and $\overline{(F_{dL})_{Sim}}$ are measured and predicted mean values of the densimetric Froude numbers at the limit of deposition. n is a total number of the datasets.

The efficiency of the proposed GLUE based models in comparison to the previous alternative formulas in both training and testing phases are quantified using Improvement index (IM) as follows (Sharafati Yasa *et al.* 2018):

$$IM_{train} = \frac{(IM_{train}^{SI} + IM_{train}^{MAE} + IM_{train}^{NSE} + IM_{train}^{LMI} + IM_{train}^{R^2})}{5} \quad (21)$$

$$IM_{test} = \frac{(IM_{test}^{SI} + IM_{test}^{MAE} + IM_{test}^{NSE} + IM_{test}^{LMI} + IM_{test}^{R^2})}{5} \quad (22)$$

$$IM_{train/test}^{SI} = \frac{(SI_{train/test}^{Model} - SI_{train/test}^{Ab. Ghani 1993})}{SI_{train/test}^{Ab. Ghani 1993}} \times 100 \quad (23)$$

$$IM_{train/test}^{MAE} = \frac{(MAE_{train/test}^{Model} - MAE_{train/test}^{Ab. Ghani 1993})}{MAE_{train/test}^{Ab. Ghani 1993}} \times 100 \quad (24)$$

$$IM_{train/test}^{NSE} = \frac{(NSE_{train/test}^{Model} - NSE_{train/test}^{Ab. Ghani 1993})}{NSE_{train/test}^{Ab. Ghani 1993}} \times 100 \quad (25)$$

$$IM_{train/test}^{LMI} = \frac{(LMI_{train/test}^{Model} - LMI_{train/test}^{Ab. Ghani 1993})}{LMI_{train/test}^{Ab. Ghani 1993}} \times 100 \quad (26)$$

$$IM_{train/test}^{R^2} = \frac{(R^2_{train/test}^{Model} - R^2_{train/test}^{Ab. Ghani 1993})}{R^2_{train/test}^{Ab. Ghani 1993}} \times 100 \quad (27)$$

where $SI_{train/test}^{Model}$, $MAE_{train/test}^{Model}$, $NSE_{train/test}^{Model}$, $LMI_{train/test}^{Model}$, and $R^2_{train/test}^{Model}$ are, respectively, the computed SI, MAE, NSE, LMI, and R^2 of the predictive models (i.e., the new GLUE based, and the previous formulas (Vongvisessomjai Tingsanchali *et al.* 2010; Ebtehaj Bonakdari *et al.* 2014)) in training or testing phases.

RESULTS AND DISCUSSION

This study aims to propose a novel formulation to provide a more accurate estimation of the densimetric Froude number at the limit of deposition in comparison with the available equations (Ab. Ghani 1993; Vongvisessomjai Tingsanchali *et al.* 2010; Ebtehaj Bonakdari *et al.* 2014). Owing to high randomness of the densimetric Froude number, a stochastic approach (herein, the GLUE methodology) is utilized to

tune the defined random variables. Ultimately, the accuracy of the proposed stochastically models are compared with the existing well-known formulas.

Development of GLUE based model for simulating the densimetric Froude number

To obtain the GLUE based models (i.e., GBFP and GBTP models) for simulating the densimetric Froude number, a number of parameters (i.e., $a_1 \sim a_5$, and $b_1 \sim b_3$) in Equations (9) and (15) are defined as random variables. These variables are estimated by means of the GLUE

approach, where the uniform distributions are assigned as their prior distributions (Table 3). The posterior distributions of the random variables are obtained using the 20,000 simulations (Figure 4) through the GLUE approach (Figure 3).

Figure 3 indicates that the coefficient and exponent of D_{gr} in GBFP model (i.e., a_1 and a_2) have negative skewness implying that the accuracy of Froude number estimation would be increased through the larger values of the D_{gr} parameter, while coefficients of the other variables C_{vL} , $\frac{R_h}{d_{50}}$, and λ_s (i.e., a_3 , a_4 , a_5 , and b_1 , b_2 , b_3) have relatively symmetric distributions. This implies that the prediction

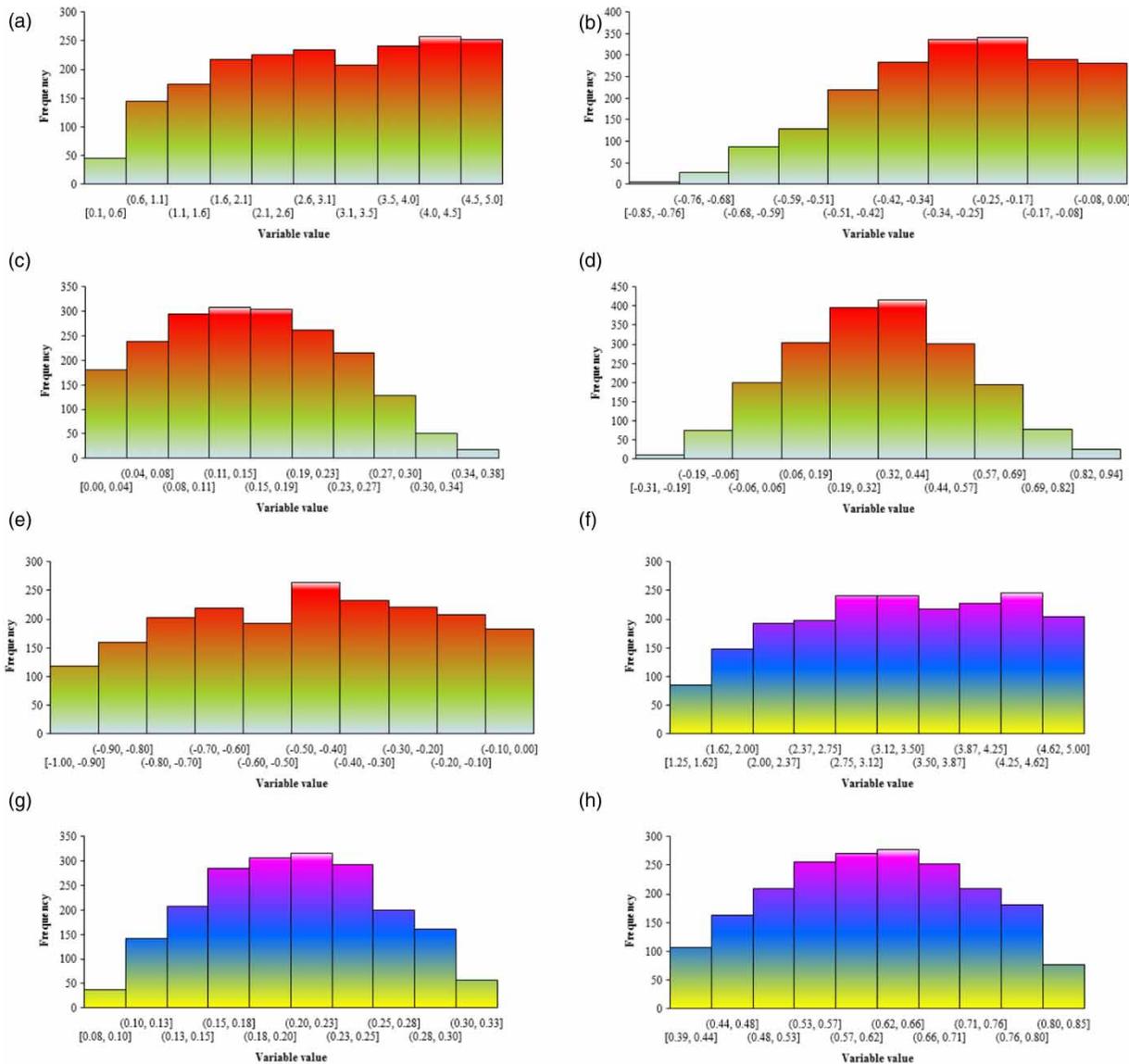


Figure 3 | Posterior distributions of the defined random coefficients in 4-Parameter and 2-Parameter GLUE models. (a) a_1 , (b) a_2 , (c) a_3 , (d) a_4 , (e) a_5 , (f) b_1 , (g) b_2 , (h) b_3 .

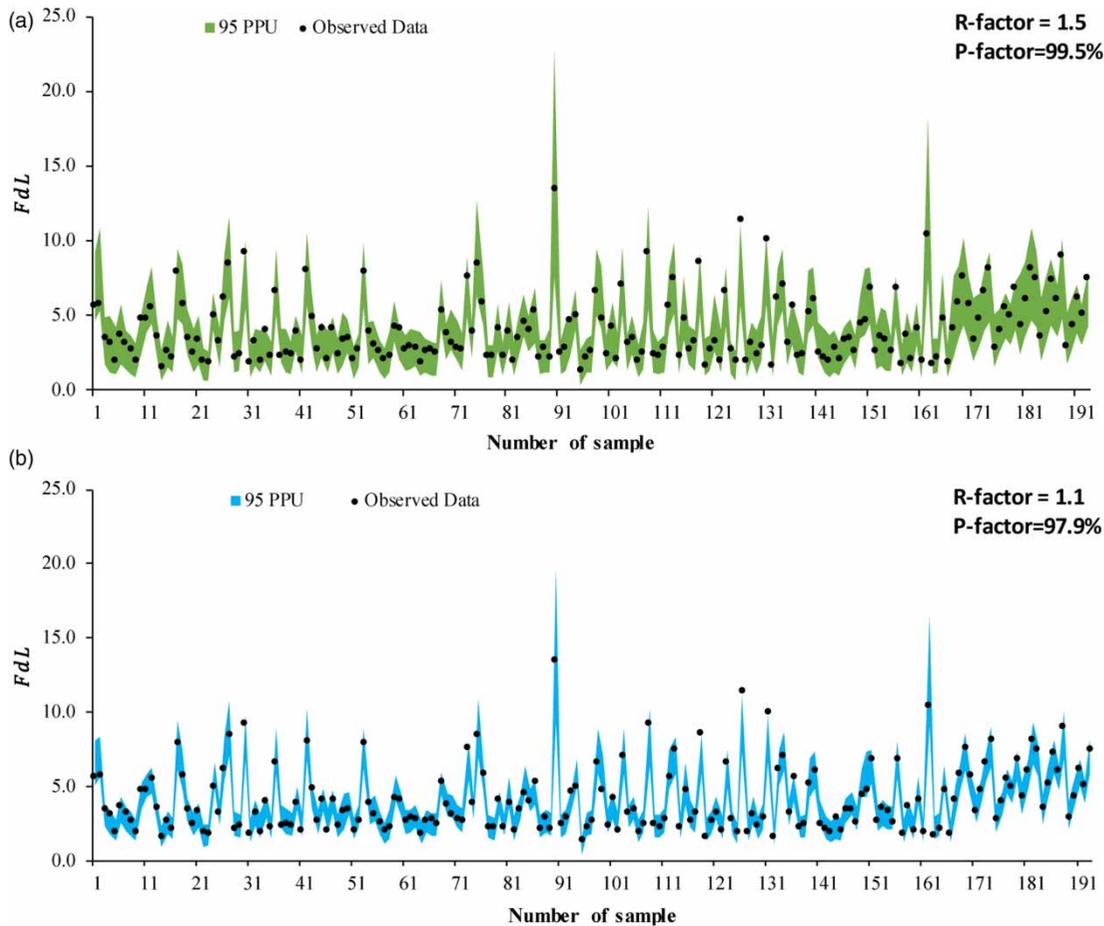


Figure 4 | Comparison between the generated 95PPU band of the GLUE based models and the experimental data in the training phase: (a) GBFP model, (b) GBTP model.

accuracy of the Froude number increases, taking into account the mean value of these parameters.

Using the obtained posterior distributions of the random coefficients, their optimistic values are obtained through the GLUE method. The proposed GLUE based formulas are expressed in the form of the following equations:

$$\frac{V_L}{\sqrt{g\Delta d_{50}}} = 3.06 D_{gr}^{-0.26} C_{vL}^{0.15} \left(\frac{R_h}{d_{50}}\right)^{0.34} \lambda_s^{-0.42} \quad (28)$$

$$\frac{V_L}{\sqrt{g\Delta d_{50}}} = 3.39 C_{vL}^{0.2} \left(\frac{R_h}{d_{50}}\right)^{0.62} \quad (29)$$

where Equations (28) and (29) are the proposed GBFP and GBTP models for predicting the densimetric Froude number at the limit of deposition.

Assessing the performance of the GLUE based models

To assess the reliability of the proposed GLUE based models, the corresponding 95 percent prediction uncertainty (95PPU) bands (i.e., the interval between the 97.5% and the 2.5% percentiles) are generated using the obtained 20,000 simulations in the training phase (Figure 4). The prediction reliability is measured using the P -factor index: the percentage of the bracketed experimental data by the 95 PPU band (Sharafati & Azamathulla 2018).

The P -factor value is an indicator to assess the stochastic models. The minimum acceptable range for the bracketed data by the 95 PPU band is 50% (Abbaspour Yang *et al.* 2007). In accordance to the GBFP and GBTP models prediction results in the training phase (Figure 4), it is clear that significant reliability is achieved in the predictability of both models (P -factor > 97%). In fact, the GBFP and GBTP models are potent to offer accurate densimetric Froude number estimations.

Another uncertainty index, R – factor, is related to the average thickness of 95 PPU band and standard deviation of the observed data (Sharafati & Azamathulla 2018). Figure 4 clarifies that the uncertainty in predicted densimetric Froude number associated with the GBFP model (R – factor = 1.5) is relatively more than the GBTP model (R – factor = 1.1).

The scatter plots (Figure 5) illustrate the consistency between the observed densimetric Froude number and the predicted values by the proposed GLUE based models as well as the previous investigations (Vongvisessomjai Tingsanchali *et al.* 2010; Ebtehaj Bonakdari *et al.* 2014)). Figure 5 indicates that a significant agreement is observed between the measured and predicted Froude number values by means of all the predictive models in both the training ($R^2 = 0.939 - 0.954$) and testing ($R^2 = 0.928 - 0.948$) phases. This agreement is stronger in small Froude numbers ($F_{dL} < 5$), whereas the points corresponding to the larger Froude numbers ($F_{dL} > 8$) are noticeably more spread. Besides, the GBTP points are located at the closest regions to the best-fit line, indicating the best agreement with the experimental data in training ($R^2 = 0.954$) and testing ($R^2 = 0.948$) phases in comparison with the other predictive models. Furthermore, the least agreement with the measured data is observed using the Ab. Ghani (Ab. Ghani 1993) equation.

The predictive models were also comparatively evaluated by the Taylor diagram (Figure 6). In this diagram, several metrics, e.g. correlation coefficient, standard deviation and the root mean are considered, simultaneously (Taylor 2001). From Figure 6, it is clear that in both phases, GBTP (yellow point) has a slightly shorter distance to the observed point (cyan point) while the point corresponding to the Ab. Ghani (Ab. Ghani 1993) results (blue point) has a relatively further distance from the observed point in comparison with the other models. Indeed, the Taylor diagram analysis confirms the highlighted results from the scatterplots.

Following the computed normalized $RMSE$, MAE , NSE , R^2 , and LMI indices, the heat maps over the training and testing phases (Figure 7) are obtained to provide a visual comparison among the performance of the predictive models.

From Figure 7, it is clear that all the cells relating to the GBTP model have dark blue color in both the training and testing stages, noting that the GBTP model offers the best performance indices. Besides, (Vongvisessomjai Tingsanchali *et al.* 2010) model is also ranked as the second adequate model in terms of MAE , R^2 , and LMI indices in training phase (Figure 7(a)) while in testing stage, the

GBFP model performs better than the remaining models, i.e. (Ab. Ghani 1993; Vongvisessomjai Tingsanchali *et al.* 2010; Ebtehaj Bonakdari *et al.* 2014) in terms of all indices, except the R^2 index (Figure 7(b)). In fact, in agreement with the scatter plots and the Taylor diagrams, heat maps also indicate that a significant prediction superiority is expected in the use of the GBTP model in comparison with the other predictive models.

The boxplots, respecting the percentiles (i.e., $Q_{25\%}$, $Q_{50\%}$ and $Q_{75\%}$) and the interquartile ranges (IQR) of the observed and predicted Froude number values, are exhibited in Figure 8 to quantify their variability in training and testing phases.

Figure 8 indicates that the closest values between the predicted and observed median ($Q_{50\%}$) of the Froude number in the training phase are obtained from the (Ebtehaj Bonakdari *et al.* 2014) and GBTP models, while the better performance to capture the observed $IQR = 2.81$ is evidenced in the GBTP model with $IQR = 2.92$. Relatively similar results are also obtained in testing phase, noting that the GLUE based models offered the best overall prediction performance. Indeed, that is a normal result where the input and output parameters have stochastic nature and stochastic-based models.

To quantify the performance improvement by the proposed models, the IM index [Equations (21) and (22)] is assessed in training and testing phases (Table 4). To this end, the (Ab. Ghani 1993) model is considered as the benchmark model. Hence, a positive value of IM means the model has better performance than the (Ab. Ghani 1993) model, and vice versa. Table 4 demonstrates that the highest improvement in all the performance indices (1.60–19.19%) as well as the IM index (9.2%) are obtained by GBTP over the training phase. Similar superiority is also achieved in testing phase, where the performance indices (2.16–24.65%) and the IM index (11.3%) are improved through the GBTP.

To assess the impact of the input variable (e.g., D_{gr} , C_{vL} , $\frac{R_h}{d_{50}}$, λ_s) changes on the target variable; a sensitivity analysis is performed. In this way, an index named relative coefficient of variation (RCV) is introduced as follow:

$$RCV(x_i, y) = \frac{CV_y}{CV_{x_i}} \quad i = 1 \dots n \quad (30)$$

where x_i and y are the i^{th} input variable and target variable, respectively. CV is the coefficient of variation, and n is a

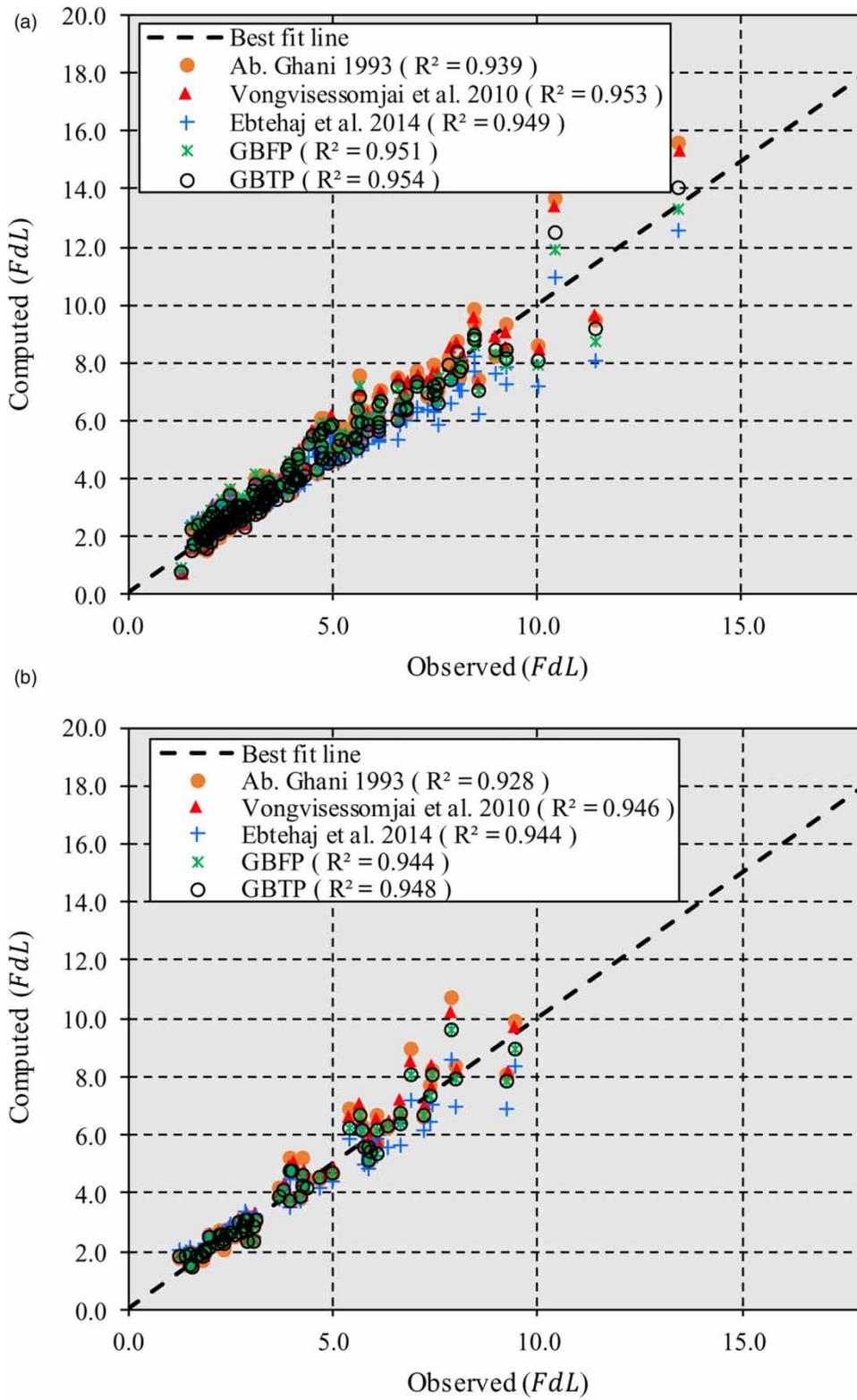


Figure 5 | Scatter plots of the predicted versus the observed densimetric Froude number. (a) training phase, (b) testing phase.

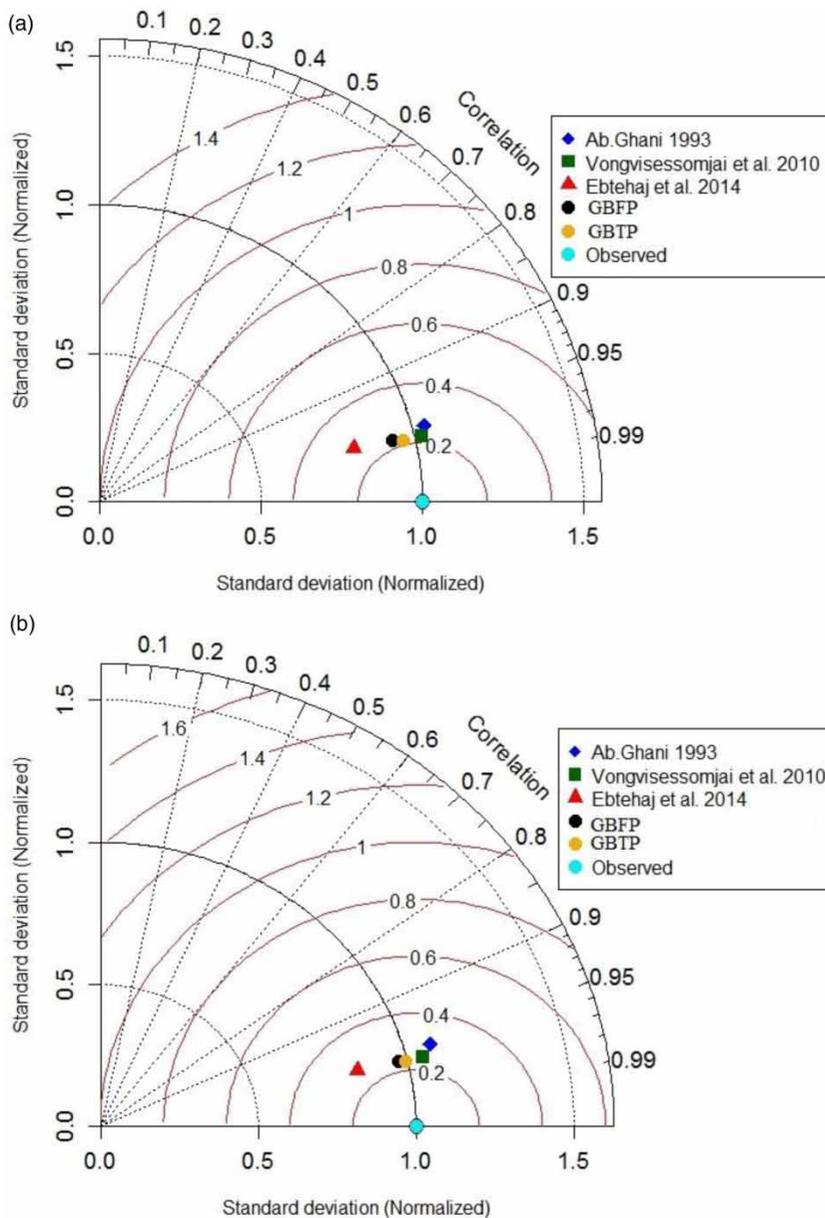


Figure 6 | Normalized Taylor diagrams of the predicted and the observed densimetric Froude number: (a) training phase, (b) testing phase.

total number of input variables. The target variable is most affected by an input variable which provides the highest *RCV*. To obtain the *RCV* value for each input variable, 100 random samples are generated using the Monte Carlo simulation. In this way, the PDFs of the input variables is assumed to be a uniform distribution. The changes in the non-deposition velocity limit against the variation of the input variables are presented in Figure 9.

From Figure 9, it is clear that the highest values of the *RCV* are found respectively, in the $Rh/d50$ ($RCV = 0.55$), and CvL ($RCV = 0.29$). It means that the non-deposition velocity limit

is most affected by those variables. This finding is consistency with those obtained by (Mayerle Nalluri *et al.* 1991).

Although the formulas provided by GLUE approach for predicting the non-deposition velocity limit are simple as the empirical equations, those formulas preserve the stochastic nature of deposition phenomenon. Hence, the primary advantage of the GLUE-based is simple modeling with acceptable accuracy. At the same time, the AI models comprise the complicated structure with many unknown parameters, and thus, accurate tuning of those parameters is a crucial issue in AI modelling. Furthermore, it is easy

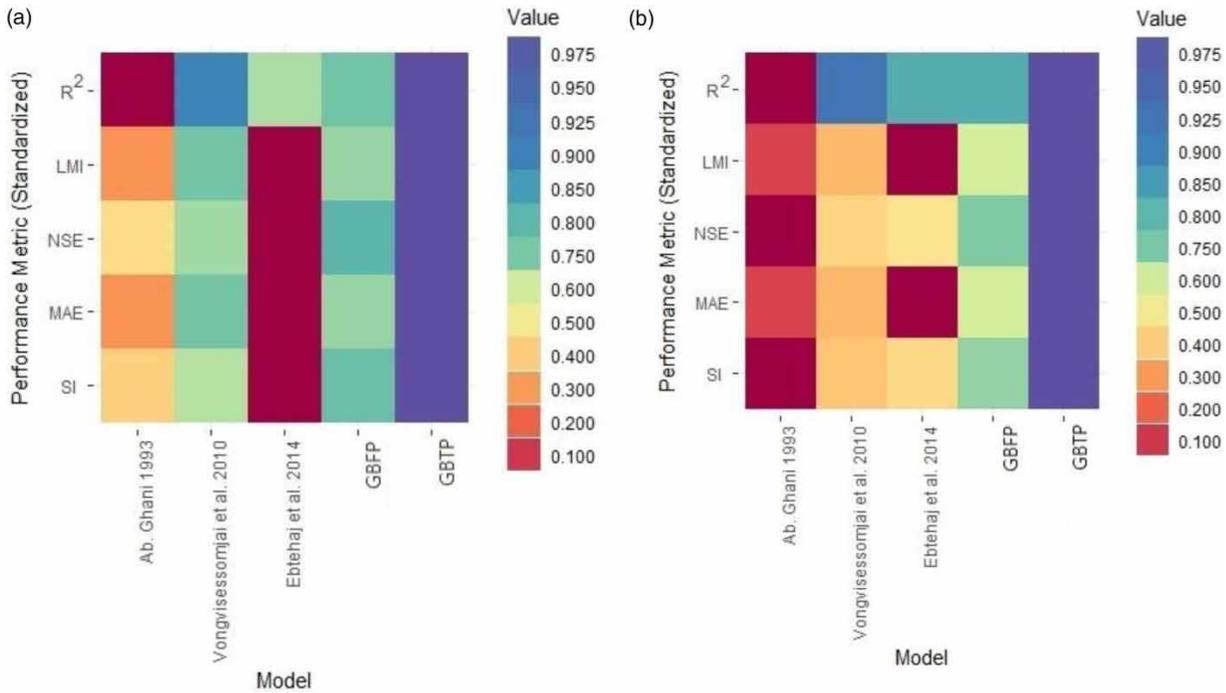


Figure 7 | Heat maps of the predictive models of the densimetric Froude number: (a) training phase, (b) testing phase.

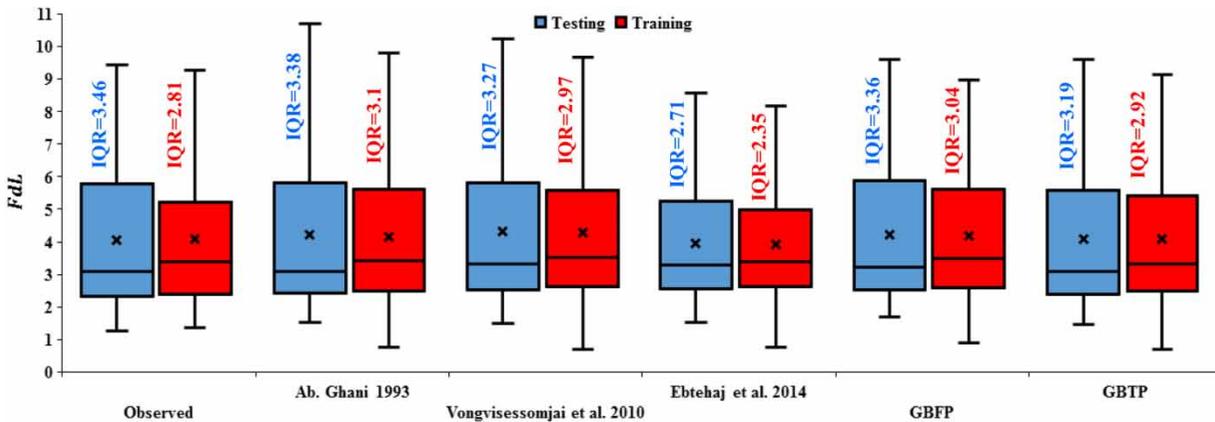


Figure 8 | Boxplots of the observed densimetric Froude number against the different predictive models results.

to identify the attributes of GLUE modelling while it is a difficult task in AI modeling due to its close-box nature.

CONCLUSION

Clean sewer design is a key issue for engineers. Deposition and consolidation of solid materials in sewer pipes may reduce the system efficiency or even block a part of the network. The velocity at limit of sediment deposition is embedded in densimetric Froude number, a function of

dimensionless grain size, friction factor in the presence of sediment motion, sediment concentration, and the ratio of hydraulic radius over the median sediment size. Fundamental analysis is made through the Gamma Test to identify the most effective parameters. Results reveal that among the four mentioned dimensionless parameters, sediment concentration at limit of deposition and the ratio of hydraulic radius over median sediment size are the most effective parameters on the prediction of velocity at limit of deposition. The internal part of many sewer pipes are not rough enough to affect sediment transport. Moreover, the other

Table 4 | Performance indices of the proposed and the previous predictive models

Model	Phase	Performance indices					Improvement (%)					
		SI	MAE	NSE	LMI	R ²	SI	MAE	NSE	LMI	R ²	IM
Ab. Ghani (1993)	Training	0.139	0.396	0.933	0.772	0.939	–	–	–	–	–	–
Vongvisessomjai <i>et al.</i> (2010)		0.129	0.346	0.942	0.801	0.953	7.05	12.63	0.96	3.76	1.49	5.2
Ebtehaj <i>et al.</i> (2014)		0.155	0.426	0.916	0.755	0.949	–11.82	–7.58	–1.82	–2.20	1.06	–4.5
GBFP		0.124	0.352	0.946	0.797	0.951	10.76	11.11	1.39	3.24	1.28	5.5
GBTP		0.115	0.320	0.954	0.816	0.954	17.28	19.19	2.25	5.69	1.60	9.2
Ab. Ghani (1993)		Testing	0.161	0.426	0.906	0.763	0.928	–	–	–	–	–
Vongvisessomjai <i>et al.</i> (2010)	0.145		0.404	0.923	0.775	0.946	9.55	5.16	1.88	1.57	1.94	4.0
Ebtehaj <i>et al.</i> (2014)	0.144		0.438	0.925	0.756	0.944	10.63	–2.82	2.10	–0.92	1.72	2.1
GBFP	0.133		0.382	0.936	0.787	0.944	17.41	10.33	3.31	3.15	1.72	7.2
GBTP	0.121		0.343	0.946	0.809	0.948	24.65	19.48	4.41	6.03	2.16	11.3

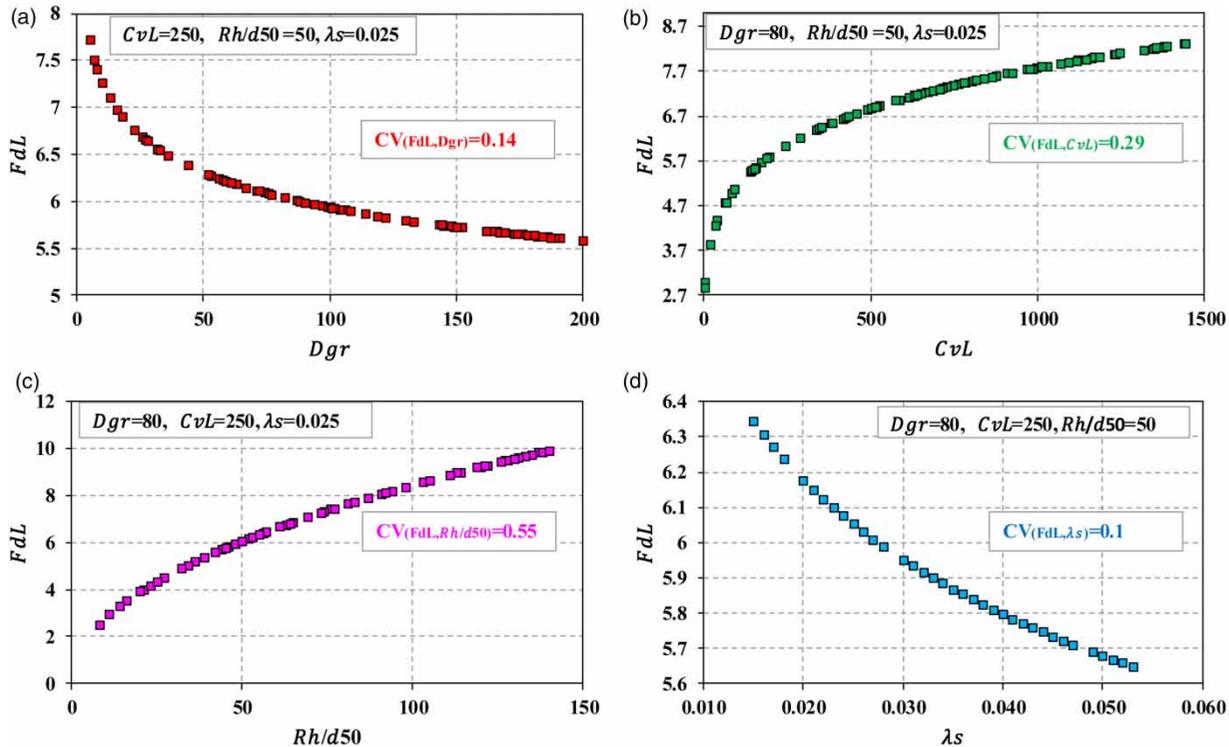


Figure 9 | Variation of densimetric Froude number against changes of (a) D_{gr} , (b) CvL , (c) $Rh/d50$, (d) λ_s .

less effective parameter, i.e. dimensionless grain size, particularly expresses the effect of sediment size on densimetric Froude number. Noting that median sediment size is also included in the ratio of hydraulic radius over median sediment size, one may conclude that the dimensionless grain size is a redundant parameter in the prediction of velocity at limit of deposition.

To ensure about the results of the Gamma Test, two separate equations are derived based on the Generalized Likelihood Uncertainty Estimation (GLUE) methodology: 1) GBFP (Equation (28)) and 2) GBTP (Equation (29)). These two formulas consist of all the four influencing parameters and the selected two variables, respectively. Further analysis by means of P – factor technique assured

that both models are potent to offer reliable predictions. In contrast, the R – factor analysis reveals that the prediction uncertainty by GBTP model (Equation (29)) is considerably less than the GBFP model. Other analyses by several error indices, e.g. SI, LMI, NSE, etc., support the superiority of the GBTP (Equation (29)) with respect to the 4-parameter GLUE based formula (i.e., GBFP), as well as the conventional regression-based equations in this context.

From the methodology used in this study, a new equation was derived for the estimation of velocity at the limit of deposition for sewer design purposes. The suggested equation is more accurate than the available regression-based formulas in the literature. Thus, one may also employ this technique in some other predictions if the phenomena are not fully deterministic. For example, local scouring is an interaction result of two stochastic processes (i.e., transient vortices and sediment motion) and use of a stochastic method may lead to enhance the prediction accuracy and reliability.

DATA AVAILABILITY STATEMENT

Data cannot be made publicly available; readers should contact the corresponding author for details.

REFERENCES

- Abbaspour, K. C., Johnson, C. A. & van Genuchten, M. T. 2004 Estimating uncertain flow and transport parameters using a sequential uncertainty fitting procedure. *Vadose Zone Journal* **3** (4), 1340–1352.
- Abbaspour, K. C., Yang, J., Maximov, I., Siber, R., Bogner, K., Mieleitner, J. & Zobrist, J. 2007 Modelling hydrology and water quality in the pre-alpine/alpine thur watershed using SWAT. 413–430.
- Ab. Ghani, A. 1993 Sediment Transport in Sewers.
- Ab. Ghani, A. & Md. Azamathulla, H. 2011 Gene-Expression programming for sediment transport in sewer pipe systems. *Journal of Pipeline Systems Engineering and Practice* **2** (3), 102–106.
- Azamathulla, H. M. 2012 Gene expression programming for prediction of scour depth downstream of sills. *Journal of Hydrology*.
- Azamathulla, H. M., Ab. Ghani, A. & Fei, S. Y. 2012 ANFIS-based approach for predicting sediment transport in clean sewer. *Applied Soft Computing* **12** (3), 1227–1230.
- Azamathulla, H. M. & Ghani, A. A. 2011 ANFIS-based approach for predicting the scour depth at culvert outlets. *Journal of Pipeline Systems Engineering and Practice*.
- Azamathulla, H. M. & Zakaria, N. A. 2011 Prediction of scour below submerged pipeline crossing a river using ANN. *Water Science and Technology*.
- Beven, K. 1993 Prophecy, reality and uncertainty in distributed hydrological modelling. *Advances in Water Resources*.
- Beven, K. & Binley, A. 1992 The future of distributed models: model calibration and uncertainty prediction. *Hydrological Processes* **6** (3), 279–298.
- Camp, T. 1942 Minimum velocities for sewers final report committee to study limiting velocities of flow in sewers. *Journal of Boston Society of Civil Engineers* **29**.
- Cu, P., Ball, J. & Dao, N. 2018 Uncertainty estimation using the glue and Bayesian approaches in flood estimation: a case study – Ba river, Vietnam. *Water* **10** (11), 1641.
- Dodaro, G., Tafarjnoruz, A., Calomino, F., Gaudio, R., Stefanucci, F., Adduce, C. & Sciortino, G. 2014 An experimental and numerical study on the spatial and temporal evolution of a scour hole downstream of a rigid bed. In: *Proceedings of the International Conference on Fluvial Hydraulics, RIVER FLOW 2014, Lausanne, Switzerland*. Taylor and Francis Group plc, 1415–1422.
- Dodaro, G., Tafarjnoruz, A., Sciortino, G., Adduce, C., Calomino, F. & Roberto, G. 2016 Modified einstein sediment transport method to simulate the local scour evolution downstream of a rigid bed. *Journal of Hydraulic Engineering* **142** (11).
- Duan, Q., Sorooshian, S. & Gupta, V. 1992 Effective and efficient global optimization for conceptual rainfall-runoff models. *Water Resources Research* **28** (4), 1015–1031.
- Ebtehaj, I. & Bonakdari, H. 2013 Evaluation of sediment transport in sewer using artificial neural network. *Engineering Applications of Computational Fluid Mechanics* **7** (3), 382–392.
- Ebtehaj, I. & Bonakdari, H. 2016 A support vector regression-firefly algorithm-based model for limiting velocity prediction in sewer pipes. *Water Science and Technology* **73** (9), 2244–2250.
- Ebtehaj, I. & Bonakdari, H. 2017 No-deposition sediment transport in sewers using gene expression programming. *Journal of Soft Computing in Civil Engineering* **1** (1), 29–53.
- Ebtehaj, I., Bonakdari, H. & Sharifi, A. 2014 Design criteria for sediment transport in sewers based on self-cleansing concept. *Journal of Zhejiang University SCIENCE A* **15** (11), 914–924.
- Freer, J., Beven, K. & Ambrose, B. 1996 Bayesian estimation of uncertainty in runoff prediction and the value of data: an application of the GLUE approach. *Water Resources Research* **32** (7), 2161–2173.
- Freni, G., Mannina, G. & Viviani, G. 2009 Urban runoff modelling uncertainty: comparison among Bayesian and pseudo-Bayesian methods. *Environmental Modelling & Software* **24** (9), 1100–1111.
- Guyen, A. & Azamathulla, H. M. 2012 Gene-expression programming for flip-bucket spillway scour. *Water Science and Technology*.
- Kargar, K., Safari, M. J. S., Mohammadi, M. & Samadianfar, S. 2019 Sediment transport modeling in open channels using neuro-fuzzy and gene expression programming techniques. *Water Science and Technology*.

- Kleinhans, M. G. & Van Rijn, L. C. 2002 Stochastic prediction of sediment transport in sand-gravel bed rivers. *Journal of Hydraulic Engineering*.
- Mayerle, R., Nalluri, C. & Novak, P. 1991 Sediment transport in rigid bed conveyances. *Journal of Hydraulic Research* **29** (4), 475–495.
- Montes, C., Berardi, L., Kapelan, Z. & Saldarriaga, J. 2020 Predicting bedload sediment transport of non-cohesive material in sewer pipes using evolutionary polynomial regression – multi-objective genetic algorithm strategy. *Urban Water Journal* **17** (2), 154–162.
- Najafzadeh, M. 2016 Neurofuzzy-Based GMDH-PSO to predict maximum scour depth at equilibrium at culvert outlets. *Journal of Pipeline Systems Engineering and Practice*.
- Najafzadeh, M., Barani, G. A. & Azamathulla, H. M. 2014 Prediction of pipeline scour depth in clear-water and live-bed conditions using group method of data handling. *Neural Computing and Applications* **24** (3–4), 629–635.
- Najafzadeh, M., Barani, G. A. & Kermani, M. R. H. 2013 Abutment scour in clear-water and live-bed conditions by GMDH network. *Water Science and Technology*.
- Najafzadeh, M. & Bonakdari, H. 2017 Application of a neuro-Fuzzy GMDH model for predicting the velocity at limit of deposition in storm sewers. *Journal of Pipeline Systems Engineering and Practice* **8** (1), 06016003.
- Najafzadeh, M., Rezaie-Balf, M. & Tafarjnoruz, A. 2018 Prediction of riprap stone size under overtopping flow using data-driven models. *International Journal of River Basin Management*.
- Najafzadeh, M. & Tafarjnoruz, A. 2016 Evaluation of neuro-fuzzy GMDH-based particle swarm optimization to predict longitudinal dispersion coefficient in rivers. *Environmental Earth Sciences* **75** (2).
- Najafzadeh, M., Tafarjnoruz, A. & Lim, S. Y. 2017 Prediction of local scour depth downstream of sluice gates using data-driven models. *ISH Journal of Hydraulic Engineering* **23** (2), 195–202.
- Novak, P. & Nalluri, C. 1975 Sediment transport in smooth fixed bed channels. *ASCE Journal of the Hydraulics Division* **101** (9), 1139–1154.
- Sadegh, M., Shakeri Majd, M., Hernandez, J. & Haghghi, A. T. 2018 The quest for hydrological signatures: effects of data transformation on Bayesian inference of watershed models. *Water Resources Management* **32** (5), 1867–1881.
- Sharafati, A., Asadollah, S. B. H. S. & Hosseinzadeh, M. 2020 The potential of new ensemble machine learning models for effluent quality parameters prediction and related uncertainty. *Process Safety and Environmental Protection*.
- Sharafati, A. & Azamathulla, H. M. 2018 Assessment of Dam overtopping reliability using SUFI based overtopping threshold curve. *Water Resources Management* **32** (7).
- Sharafati, A., Haghbin, M., Motta, D. & Yaseen, Z. M. 2019 The application of soft computing models and empirical formulations for hydraulic structure scouring depth simulation: a comprehensive review, assessment and possible future research direction. *Archives of Computational Methods in Engineering* 1–25.
- Sharafati, A., Tafarjnoruz, A., Shourian, M. & Yaseen, Z. M. 2019 Simulation of the depth scouring downstream sluice gate: the validation of newly developed data-intelligent models. *Journal of Hydro-Environment Research*.
- Sharafati, A., Yasa, R. & Azamathulla, H. M. 2018 Assessment of stochastic approaches in prediction of wave-Induced pipeline scour depth. *Journal of Pipeline Systems Engineering and Practice* **9** (4), 4018024.
- Shojaeezadeh, S. A., Nikoo, M. R., McNamara, J. P., AghaKouchak, A. & Sadegh, M. 2018 Stochastic modeling of suspended sediment load in alluvial rivers. *Advances in Water Resources* **119**, 188–196.
- Taylor, K. E. 2001 Summarizing multiple aspects of model performance in a single diagram. *Journal of Geophysical Research: Atmospheres* **106** (D7), 7183–7192.
- Vázquez, R. F., Beven, K. & Feyen, J. 2009 GLUE based assessment on the overall predictions of a MIKE SHE application. *Water Resources Management*.
- Vongvisessomjai, N., Tingsanchali, T. & Babel, M. S. 2010 Non-deposition design criteria for sewers with part-full flow. *Urban Water Journal* **7** (1), 61–77.
- Wang, X., Frankenberger, J. R. & Klavivko, E. J. 2006 Uncertainties in DRAINMOD predictions of subsurface drain flow for an indiana silt loam using the GLUE methodology. *Hydrological Processes: An International Journal* **20** (14), 3069–3084.

First received 12 May 2020; accepted in revised form 28 June 2020. Available online 10 July 2020