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

Ali Tafarojnoruz and Ahmad Sharafati

ABSTRACT

Sedimentation in storm sewers 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 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$_{\text{train}}$ = 9.2%) and testing (IM$_{\text{test}}$ = 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 to predict the densimetric Froude number in storm sewers.
- The generalized likelihood uncertainty estimation approach is used as a parameters estimator prior to the parametric formulas.
- The proposed model is validated against literature formulas.
- The proposed stochastic-based model is shown to be 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 severe conditions it might lead to network deterioration (Ab. Ghani & Md. Azamathulla 2011; Ebtehaj 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 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 et al. 2010).

The earliest and most common approach of the sewer design was based on a minimum flow velocity for scouring of the existing bed sediment. In this context, Camp (1942) suggested an equation for the minimum velocity to scour the deposited bed sediments \( V_{Camp} \) as follows:

\[
V_{Camp} = \frac{1}{n} \rho^\frac{1}{2} \sqrt{B \left( \frac{\rho_s}{\rho} - 1 \right) d}
\]  

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; \( \rho_s \) and \( \rho \) are the density of sediment and water, and \( R_h \) is 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 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 worth noting 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 originates 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 including 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 intelligence (AI) models to achieve more accurate predictions than the conventional regression-based formulas (Sharafati et al. 2019a). For instance, an AI technique, named group method of data handling (GMDH), has been optimized to estimate longitudinal dispersion coefficient in rivers (Najafzadeh & Tafarognoruz 2016). Application of other AI methodologies, like evolutionary polynomial regression (EPR), gene-expression programming (GEP), and model tree, 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 et al. 2013, 2014, 2017, 2018; Najafzadeh 2016; Kargar et al. 2019; Sharafati et al. 2019b).

According to Azamathulla & Ghani (2011), Azamathulla et al. (2012), Ebtehaj & Bonakdari (2013, 2017), and Najafzadeh & Bonakdari (2017), artificial neural networks (ANNs), adaptive neuro-fuzzy inference systems (ANFIS), GEP, and neuro-fuzzy GMDH were successfully utilized to predict velocity at limit of deposition. More specifically, some investigations devoted to applying EPR, support vector regression, and the firefly algorithm resulted in accurate sediment transport prediction in pipes (Ebtehaj & Bonakdari 2016; Montes 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 the design of 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 of 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 et al. 2010; Ebtehaj 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, AI 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 et al. 2018; Shojaeezadeh et al. 2018). The sediment transport process, whether in nature (e.g., rivers, coastal zones) or in manmade structures (e.g., pipelines, sewer networks) is inherently stochastic. Nevertheless, a 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 et al. 2014, 2016; Sharafati et al. 2018).

To quantify the uncertainty of random variables, several techniques such as generalized likelihood uncertainty estimation (GLUE) and sequential uncertainty fitting (SUFi) have already been proposed (Beven & Binley 1992; Abbaspour 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 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 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 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 et al. 1992; Beven 1993). With this advantage, Sharafati 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 include the most effective dimensionless groups. The performance of the new stochastic formulas 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 et al. 2010; Ebtehaj & Bonakdari 2013). Equation (2) briefly presents a functional relationship between velocity at limit of deposition ($V_L$) and the effective parameters through an unknown function, $f_1$:

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

(2)

where $g$ is acceleration due to gravity; $C_{vl}$ denotes sediment concentration at limit of deposition; $\lambda_s$ is the friction factor in the presence of sediment motion; $d_{50}$ represents median sediment grain size; $\nu$ is the kinematic viscosity of water.

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

$$\frac{V_L}{\sqrt{gd_{50}}} = f_2\left(C_{vl}, \frac{R_h}{d_{50}}, \lambda_s, D_{gr}\right)$$

(3)
in which \( V_L / \sqrt{g \Delta d_{50}} = F_{dl} \) denotes the densimetric Froude number at the limit of deposition; \( \Delta = \rho_s / \rho \) represents the relative density of sediment in water; and
\[
D_{gr} = d_{50} \left( \frac{\Delta}{\rho_s^2} \right)^{1/3}
\]
is dimensionless grain size. \( \lambda_s \) is a number around the wall friction factor under the clear-water condition \( \lambda_0 = 8g n^2 / R_h^{1/3} \), even though two other dimensionless parameters, i.e., \( C_{vL} \) and \( D_{gr} \), may slightly affect it. The following equation is recommended to calculate \( \lambda_s \) (Ab. Ghani 1993):

\[
\lambda_s = 1.13 \lambda_0^{0.88} C_{vL}^{0.02} D_{gr}^{-0.01}
\]  

(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 et al. (2010) (Equation (6)), and Ebtehaj 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 et al. (1999) who neglected \( \lambda_s \) and \( D_{gr} \), aiming at a quick calculation of \( F_{dl} \):

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

(5)

\[
\frac{V_L}{\sqrt{g \Delta d_{50}}} = 4.51 C_{vL}^{0.226} \left( \frac{R_h}{d_{50}} \right)^{0.616}
\]  

(6)

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

(7)

### DESCRIPTION OF DATASETS

To develop and evaluate the new model, and compare its performance with those available in the literature, a large number of datasets, 257, were extracted from two experimental studies (Ab. Ghani 1993; Vongvisessomjai 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 shows 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 a parameters estimator in different engineering fields (Sharafati et al. 2018). To achieve the best parameter set, several random parameter sets would be generated using their probability distribution functions. Then, the generated parameter 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 \( w_i \) of the \( i \)th set of the parameters as follows (Wang et al. 2006):

\[
\left\{ \begin{array}{l}
    w_i = \frac{w^i}{\sum_{j=1}^{n} w^j} \\
    w^i = \exp \left( - \frac{\text{RMSE}^i}{\min (\text{RMSE})} \right)
\end{array} \right.
\]  

(8)

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

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Training phase</th>
<th>Testing phase</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( F_{at} )</td>
<td>( R_h / d_{50} )</td>
</tr>
<tr>
<td>Minimum</td>
<td>1.35</td>
<td>4.46</td>
</tr>
<tr>
<td>Maximum</td>
<td>13.49</td>
<td>188.3</td>
</tr>
<tr>
<td>Average</td>
<td>4.09</td>
<td>49.30</td>
</tr>
<tr>
<td>SD</td>
<td>2.19</td>
<td>52.96</td>
</tr>
<tr>
<td>COV</td>
<td>0.54</td>
<td>1.07</td>
</tr>
</tbody>
</table>

SD – standard deviation; COV – coefficient of variation.
where $w_i$ is the likelihood measure of $i$th behavioral set, $k$ is the number of behavioral samples, and $RMSE$ is the root mean square error. 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 et al. 2009).

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

$$VL = a_1 D_{gr}^{a_2} C_{vl}^{a_3} (\frac{R_h}{d_{50}})^{a_4} \lambda_s^{a_5}$$

(9)

where $a_1 - a_5$ 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 ($D_{gr}$, $C_{vl}$, $R_h$, $d_{50}$, $\lambda_s$).

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 et al. 2020). The best input variables are selected via computing an index named $\theta$-ratio by means of a nonlinear approach. Consider a given dataset as the following form:

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

(10)

where the $x_i$ are the m-dimensional input vectors, $y_i$ are the output values and $M$ is number of measured data.
GT is originated from \( x_i \) in delta function (\( \delta \)) expressed as follows:

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

where \( |...| \) is the Euclidean distance and \( x_{N[i,k]} \) is the kth (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:

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

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

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

\[
\gamma = A\delta + \Gamma
\] (13)

The gamma statistic can be standardized by measuring another term, \( \theta \)-ratio, which is defined as:

\[
\theta \text{-ratio} = \frac{\Gamma}{\sigma^2(y)}
\] (14)

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

\( \theta \)-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 \( \theta \)-ratio from different two variate combinations are represented in Table 2.

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-parameter models (Vongvisessomjai et al. 2010; Ebtehaj 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:

\[
\frac{V_L}{\sqrt{g\Delta d_{50}}} = b_1 C_{vl} \left( \frac{R_h}{d_{50}} \right)^{b_3}
\] (15)

where \( b_1 - b_3 \) 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 (PDFs) 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 the proposed GLUE-based models and the previous alternative formulas, a number of agreement and error indices such as scatter index (SI), mean absolute error (MAE), Nash–Sutcliffe coefficient.
where $(F_{dl})_{Obs,i}$ and $(F_{dl})_{Sim,i}$ represent the $i$th measured and predicted densimetric Froude numbers at the limit of deposition; $(F_{dl})_{Obs}$ and $(F_{dl})_{Sim}$ are measured and predicted mean values of the densimetric Froude numbers at the limit of deposition. $n$ is the 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 is quantified using the improvement index (IM) as follows (Sharafati et al. 2018):

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

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

$$IM_{\text{train}}^{SI} = \frac{(SI_{\text{train}}^{Ab\text{ Ghani 1993}} - SI_{\text{model train}}^{Model\text{ train}})}{SI_{\text{train}}^{Ab\text{ Ghani 1993}}} \times 100$$

$$IM_{\text{train}}^{MAE} = \frac{(MAE_{\text{train}}^{Ab\text{ Ghani 1993}} - MAE_{\text{model train}}^{Model\text{ train}})}{MAE_{\text{train}}^{Ab\text{ Ghani 1993}}} \times 100$$

$$IM_{\text{train}}^{NSE} = \frac{(NSE_{\text{train}}^{Model\text{ train}} - NSE_{\text{train}}^{Ab\text{ Ghani 1993}})}{NSE_{\text{train}}^{Ab\text{ Ghani 1993}}} \times 100$$

$$IM_{\text{train}}^{LMI} = \frac{(LMI_{\text{train}}^{Model\text{ train}} - LMI_{\text{train}}^{Ab\text{ Ghani 1993}})}{LMI_{\text{train}}^{Ab\text{ Ghani 1993}}} \times 100$$

$$IM_{\text{train}}^{R^2} = \frac{(R^2_{\text{model train}} - R^2_{\text{train}}^{Ab\text{ Ghani 1993}})}{R^2_{\text{train}}^{Ab\text{ Ghani 1993}}} \times 100$$

where $SI_{\text{train}}, MAE_{\text{train}}, NSE_{\text{train}}, LMI_{\text{train}},$ and $R^2_{\text{train}}$ 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 et al. 2010; Ebtehaj 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 et al. 2010; Ebtehaj 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 stochastic 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., $a_1$–$a_5$, and $b_1$–$b_5$) 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 $D_g$ 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 $D_g$ parameter, while coefficients of the other variables $C_{el}$, $R_h$, $\frac{D_5}{d_{50}}$, and $\lambda_s$ (i.e., $a_3$, $a_4$, $a_5$, and $b_1$, $b_2$, $b_3$) have relatively symmetric distributions. This implies that the prediction 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:

\[
\frac{V_L}{\sqrt{g\Delta d_{50}}} = 4.344 D_g^{0.088} C_{el}^{0.181} \left(\frac{R_h}{d_{50}}\right)^{0.48} \lambda_s^{0.092} \quad (28)
\]

\[
\frac{V_L}{\sqrt{g\Delta d_{50}}} = 3.733 C_{el}^{0.214} \left(\frac{R_h}{d_{50}}\right)^{0.61} \quad (29)
\]

where Equations (28) and (29), respectively, 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 (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 et al. 2007). In accordance with the GBFP and GBTP model 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 able to offer accurate densimetric Froude number estimations.

Another uncertainty index, $R$-factor, is related to the average thickness of the 95 PPU band and standard deviation of the observed data (Sharafati & Azamathulla 2018). Figure 4 clarifies that 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 ($F_d > 8$), whereas the points corresponding to the larger Froude numbers ($F_d < 8$) are noticeably more spread. Also, 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. 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 division, 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 diamond) has a relatively longer distance from the observed point in comparison with the other models. Indeed, the Taylor diagram analysis confirms the highlighted results from the scatter plots.

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 between 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. Furthermore, the Vongvisessomjai et al. (2010) model is also ranked as the second accurate 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 et al. 2010; Ebtehaj et al. 2014) in terms of all the 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, depicting 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 et al. (2014) and GBTP models, while a 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 the 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.

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 the 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., \(D_{gr}\), \(C_{TL}\), \(R_{h}\), \(\lambda_{s}\)) changes on the target variable; a
Figure 5 | Scatter plots of the predicted versus the observed densimetric Froude number: (a) training phase, (b) testing phase.
sensitivity analysis is performed. In this way, an index named relative coefficient of variation (RCV) is introduced as follows:

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

where $x_i$ and $y$ are the $i$th input variable and target variable, respectively. $CV$ is the coefficient of variation, and $n$ is the total number of input variables. The target variable is most affected by an input variable that 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 are 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.

Figure 6 | Normalized Taylor diagrams of the predicted and the observed densimetric Froude number: (a) training phase, (b) testing phase. Please refer to the online version of this paper to see this figure in colour: http://dx.doi.org/10.2166/wst.2020.321.
From Figure 9, it is clear that the highest values of the RCV are found respectively, in the \( R_{h}/d_{50} \) (RCV = 0.55), and \( C_{L} \) (RCV = 0.29). It means that the non-deposition velocity limit is most affected by those variables. This finding is consistent with those obtained by Mayerle et al. (1991).

Although the formulas provided by the GLUE approach for predicting the non-deposition velocity limit are simple as the empirical equations, those formulas preserve the stochastic nature of the deposition phenomenon. Hence, the primary advantage of the GLUE-based approach is simple modeling with acceptable accuracy. At the same time, the AI models comprise a complicated structure with many unknown parameters, and, thus, accurate tuning of those parameters is a crucial issue. Furthermore, it is easy to identify the attributes of GLUE modeling 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 the 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 GT 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 is not rough enough to affect sediment transport. Moreover, the other less

<table>
<thead>
<tr>
<th>Model</th>
<th>Phase</th>
<th>Performance indices</th>
<th>Improvement (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>SI</td>
<td>MAE</td>
</tr>
<tr>
<td>Ab. Ghani (1995)</td>
<td>Training</td>
<td>0.139</td>
<td>0.396</td>
</tr>
<tr>
<td>Vongvisessomjai et al. (2010)</td>
<td>Training</td>
<td>0.129</td>
<td>0.346</td>
</tr>
<tr>
<td>Ebtehaj et al. (2014)</td>
<td>Training</td>
<td>0.155</td>
<td>0.426</td>
</tr>
<tr>
<td>GBFP</td>
<td>Training</td>
<td>0.124</td>
<td>0.352</td>
</tr>
<tr>
<td>GBTP</td>
<td>Training</td>
<td>0.115</td>
<td>0.320</td>
</tr>
<tr>
<td>Ab. Ghani (1995)</td>
<td>Testing</td>
<td>0.161</td>
<td>0.426</td>
</tr>
<tr>
<td>Vongvisessomjai et al. (2010)</td>
<td>Testing</td>
<td>0.145</td>
<td>0.404</td>
</tr>
<tr>
<td>Ebtehaj et al. (2014)</td>
<td>Testing</td>
<td>0.144</td>
<td>0.438</td>
</tr>
<tr>
<td>GBFP</td>
<td>Testing</td>
<td>0.133</td>
<td>0.382</td>
</tr>
<tr>
<td>GBTP</td>
<td>Testing</td>
<td>0.121</td>
<td>0.343</td>
</tr>
</tbody>
</table>

Figure 9 | Variation of densimetric Froude number against changes of (a) $D_{gr}$, (b) $CvL$, (c) $R_h/d_{50}$, (d) $\lambda_s$. 
effective parameter, i.e., dimensionless grain size, particularly expresses the effect of sediment size on the 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 evaluate the results of the GT, two separate equations are derived based on the 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 able to offer reliable predictions. In contrast, the R-factor analysis reveals that the prediction uncertainty by the GBTP model (Equation (29)) is considerably less than the GBFP model. Other analyses by several error indices, e.g., SI, LMI, NSE, support the superiority of the GBTP (Equation (29)) with respect to the four-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 enhance the prediction accuracy and reliability.

DATA AVAILABILITY STATEMENT

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

REFERENCES


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