Decision tree (DT), generalized regression neural network (GR) and multivariate adaptive regression splines (MARS) models for sediment transport in sewer pipes

Mir Jafar Sadegh Safari

ABSTRACT

Sediment deposition in sewers and urban drainage systems has great effect on the hydraulic capacity of the channel. In this respect, the self-cleansing concept has been widely used for sewers and urban drainage systems design. This study investigates the bed load sediment transport in sewer pipes with particular reference to the non-deposition condition in clean bed channels. Four data sets available in the literature covering wide ranges of pipe size, sediment size and sediment volumetric concentration have been utilized through applying decision tree (DT), generalized regression neural network (GR) and multivariate adaptive regression splines (MARS) techniques for modeling. The developed models have been compared with conventional regression models available in the literature. The model performance indicators, showed that DT, GR and MARS models outperform conventional regression models. Result shows that GR and MARS models are comparable in terms of calculating particle Froude number and performing better than DT. It is concluded that conventional regression models generally overestimate particle Froude number for the non-deposition condition of sediment transport, while DT, GR and MARS outputs are close to their measured counterparts.

Key words | decision tree, generalized regression, multivariate adaptive regression splines, sediment transport, self-cleansing, sewer

INTRODUCTION

Drainage and sewer pipe channels are considered as rigid boundary channels. There are major differences between sediment transport in loose and rigid boundary channels as transport rate and flow characteristics in rigid beds differ from those of loose bed channels. In the case of rigid boundary channels, sediment particles brought from upstream are conveyed by the flow, and amount of sediment load within the flow should not exceed the channel sediment transport capacity, to prevent sediment deposition. Sediment deposition at the channel bottom changes modality of the velocity and shear stress distribution at the channel boundary. Hence, it has significant effects on hydraulic capacity of the channel.

In the traditional way, drainage systems were designed through determination of a certain value of velocity or shear stress based on experience (Nalluri & Ab Ghani 1996; Safari et al. 2018). It works well in some cases; however, the sediment characteristics and other aspects of the hydraulic behavior of the drainage systems are not taken into account, and thus it cannot be a reliable benchmark. Therefore, the design velocity or shear stress should be obtained considering the factors which have the greatest impact on sediment deposition, such as size and roughness of the channel, flow depth, sediment size and sediment volumetric concentration. Self-cleansing is a hydraulic design concept for drainage and sewer systems design which is defined as a flow condition that satisfies an equilibrium between amount of sediment deposition and erosion in the channel bottom (Safari et al. 2018). A considerable number of self-cleansing models were recommended in the literature, established on laboratory experimental data using multiple non-linear regression technique for the model development (Mayerle et al. 1991; Ab Ghani 1993; Nalluri & Ab Ghani 1996; Ota & Nalluri 2003; Vongvisessomjai et al. 2010; Ota & Perrusquia 2013). Regression models are important in defining the essential factors of the problem,
however, their computation ability is not as high as machine learning techniques (Safari & Danandeh Mehr 2018; Wan Mohtar et al. 2018).

Recently, machine learning techniques have been used as powerful tools for modeling of sewer systems problems such as wastewater quality (Qin et al. 2012; Granata et al. 2017), flood prediction in urban area (Duncan et al. 2013) and wastewater hydraulic problems (Granata & de Marinis 2017). As one of their basic features, machine learning techniques are capable of approximating any arbitrary functions and modeling nonlinear systems without primary assumptions. Modeling of sediment transport in drainage and sewer systems using machine leaning techniques has attracted interest of many researchers. For instance, Ebtehaj et al. (2016, 2017) applied a decision tree based on a radial basis function neural network and support vector machine coupled with firefly algorithm techniques on sediment transport in sewer pipes using three data sets available in the literature. Roushangar & Ghasempour (2017) utilized two data sets taken from the literature and developed gene expression programming models to predict sediment volumetric concentration. Wan Mohtar et al. (2018) evaluated the applicability of artificial neural network techniques for prediction of flow velocity at incipient motion of sediment in sewer pipes. Safari & Danandeh Mehr (2018) proposed sediment transport models applying a Pareto-optimal multigene genetic programming technique for design of large sewers.

Most of the studies of sediment transport at non-deposition condition in the literature used two or three data sets for the model development. Through applying three machine learning techniques, decision tree (DT), generalized regression neural network (GR) and multivariate adaptive regression splines (MARS), this study utilizes four different pipe sizes and proposed a decision tree based on a radial basis function neural network techniques for prediction of bed load sediment transport. Most of the developed models in the literature were established applying multiple non-linear regression technique. Utilizing machine learning methods for providing models with better computational capabilities seems to be necessary.

OVERVIEW OF SELF-CLEANSING SEWER PIPE DESIGN

Considering sediment transport principles, open channels are designed based on the self-cleaning concept. In this respect, Safari et al. (2018) classified self-cleansing models into two main groups of ‘bed sediment motion’ and ‘non-deposition’. The former includes ‘incipient motion’ and ‘scouring’ models (Safari et al. 2018) while the latter covers ‘non-deposition without deposited bed’, ‘non-deposition with deposited bed’ and ‘incipient deposition’ models (May 1993; Nalluri & Ab Ghani 1996; Aksoy et al. 2017; Safari et al. 2017). Basic definitions and comparison of existing models are reported in Safari et al. (2018). This study investigates the non-deposition without deposited bed condition of sediment transport.

As examples from the literature, Mayerle (1988) and Mayerle et al. (1993) studied bed load sediment transport at non-deposition without deposited bed condition and recommended

$$\frac{V}{\sqrt{gd(s - 1)}} = 14.43C_{v}^{0.18}D_{gr}^{-0.14}(\frac{d}{R})^{-0.56} \lambda^{0.18}$$

(1)

for a sewer pipe channel, where $V$ is flow mean velocity, $g$ gravitational acceleration, $d$ sediment median size, $s$ relative specific mass of sediment to fluid, $C_{v}$ sediment volumetric concentration, $R$ hydraulic radius, $\lambda$ channel friction factor calculated by Darcy-Weisbach equation and $D_{gr}$ dimensionless grain size parameter, respectively defined as

$$\lambda = \frac{8gRS}{V^{2}}$$

(2)

$$D_{gr} = \left(\frac{(s - 1)gd^{5}}{v^{4}}\right)^{1/3}$$

(3)

where $S$ is channel bed slope and $v$ fluid kinematic viscosity. The left-hand side of Equation (1) is known as the particle Froude number ($F_{rp}$) and $d/R$ is relative particle size. Ab Ghani (1993) studied sediment transport in three different pipe sizes and proposed

$$\frac{V}{\sqrt{gd(s - 1)}} = 3.08C_{v}^{0.21}D_{gr}^{-0.09}(R/d)^{0.53} \lambda^{-0.21}$$

(4)

as a bed load non-deposition without deposited bed self-cleansing model. Vongvisessomjai et al. (2010) studied bed load and suspended load at the same sediment transport condition and developed a self-cleansing model considering fewer number of parameters as

$$\frac{V}{\sqrt{gd(s - 1)}} = 4.31C_{v}^{0.226}(\frac{d}{R})^{-0.616}$$

(5)

for bed load sediment transport. Most of the developed models in the literature were established applying multiple non-linear regression technique. Utilizing machine learning methods for providing models with better computational capabilities seems to be necessary.
METHODOLOGY

Experimental data

In this study, laboratory experimental data performed in rigid boundary channels at non-deposition without deposited bed condition of sediment transport are used. Four data sets including Mayerle (1988), May (1993), Ab Ghani (1993) and Vongvisessomjai et al. (2010) are used in the modeling. Ranges of experimental data are shown in Table 1. Experimental condition and rigs are given briefly as follows.

Mayerle (1988) performed experiments in rectangular and circular channels. Through the experiments, after establishing a uniform flow in the channel, the feeder sediment discharge was gradually increased to allow for the deposition of sediment. Flow and sediment discharge were kept constant for 20–30 minutes to satisfy a steady rate of sediment transport and hydraulic condition. May (1993) conducted experiments in a large concrete circular channel. The sediment was stored within a container and a mixture of water and sediment was pumped through a pump into the channel. The required flow discharge for the specified depth and velocity was determined. After obtaining the equilibrium condition and measuring the flow depth and sediment discharge, the flow was suddenly cut off. Then, without deposited bed sediment deformation, water was gradually removed from the channel.

Ab Ghani (1993) performed experiments in three circular channels with different sizes. The sediment was added to the channel by a sediment feeder in increasing state, until it reached the limit of deposition condition. After the deposition condition occurred, the sediment discharge was measured for 15 minutes to ensure that its value was constant to determine the limit of deposition condition. Vongvisessomjai et al. (2010) carried out experiments in two circular channels. The flow velocity was measured at three different depths and the average of them was considered as the mean flow velocity. The sediment discharge was entered into the channel through the sediment feeding system. When the amount of sediment discharge was increased, a deposition occurred in the bottom of the channel.

Data preparation

In order to investigate the sediment transport in sewer and drainage systems, flow, fluid, sediment and channel characteristics should be considered. To this end, as flow characteristics, flow velocity ($V$), hydraulic radius ($R$) and gravitational acceleration ($g$) and as fluid characteristics, fluid specific mass ($\rho$) and kinematic viscosity ($\nu$) are considered. Sediment median size ($d$), sediment volumetric concentration ($C_v$) and sediment specific mass ($\rho_s$) as sediment characteristics and channel friction factor ($\lambda$) as channel characteristic are selected. Considering the aforementioned variables, the following relationship can be expressed:

$$f(V, R, g, \nu, \rho_s, d, C_v, \lambda) = 0$$

(6)

Looking to the sediment transport models mentioned before, the variables given in Equation (6) can be expressed in the form of dimensionless parameters as follows:

$$\frac{V}{\sqrt{gd(s – 1)}} = f(C_v, D_{gr}, \frac{d}{R}, \lambda)$$

(7)

where the left side of Equation (7), the particle Froude number ($Fr_p$), is considered as the dependent parameter and output of the model, due to it consisting of flow velocity ($V$) as a desired variable for channel design, and the parameters given on the right side of Equation (7) are

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Ranges of experimental data</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$D$ or $W$ (mm)</td>
</tr>
<tr>
<td>Mayerle (1988)</td>
<td>$D = 152$</td>
</tr>
<tr>
<td></td>
<td>$W = 311$–462</td>
</tr>
<tr>
<td>May (1993)</td>
<td>$D = 450$</td>
</tr>
<tr>
<td>Ab Ghani (1993)</td>
<td>$D = 154$–450</td>
</tr>
<tr>
<td>Vongvisessomjai et al. (2010)</td>
<td>$D = 100$–150</td>
</tr>
</tbody>
</table>

$D$, circular channel diameter; $W$, rectangular channel bed width; $d$, sediment median size; $\lambda$, channel friction factor; $Y$, flow depth; $V$, flow mean velocity; $C_v$, sediment volumetric concentration.
considered as independent parameters and the input of the model. As the first step of data preparation, all parameters given in Equation (7) are calculated.

In the sediment transport technology, validity of a model significantly depends on the ranges of the experimental data used in the model development. Indeed, ranges of three variables of pipe size, sediment median size and sediment volumetric concentration are quite important. For this reason, all four data sets are used for the model development to cover wide ranges of different variables. Each of four data sets has a unique data range. For example, as given in Table 1, Mayerle (1988) data cover a wide range of $C_v$ and used two different cross-section channels as circular and rectangular; May (1993) performed experiments in a large pipe with low $C_v$ range; Ab Ghani (1993) performed experiments in three different pipe sizes and, finally, Vongvisessomjai et al. (2010) data cover small pipe size with low $C_v$ range.

Two stages of training and testing are adopted for modeling. From 375 experimental data, 300 data for training and 75 data for testing are selected randomly. In order to treat parameters equally in the modeling process, they are converted to values between 0 and 1, called normalization, by dividing all of the parameters into their maximum values. In the training phase, the relationship between the dependent and independent variables is found, and in the testing stage, the computational ability of the model is evaluated on an unused data set.

Performance indices

The evaluation of model accuracy is the most important factor in determining the validity of the proposed model. Hence, for evaluating the performances of the driven models, three statistical indices namely, root mean square error (RMSE), mean absolute percentage error (MAPE) and concordance coefficient (CC) are used. The RMSE shows the difference between the measured and computed particle Froude numbers and is defined as

$$
\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (F_{p,m}^i - F_{p,c}^i)^2}
$$

(8)

where $F_{p,m}^i$ and $F_{p,c}^i$ respectively are measured and computed particle Froude numbers and $n$ is number of data. The MAPE is obtained by comparing the absolute error between each measured and computed data in percentage. In order to find the accuracy of the established model, the computed RMSE values should be compared with the maximum or mean of measured values. Therefore, MAPE is also used in this study. MAPE is calculated by

$$
\text{MAPE} = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{F_{p,m}^i - F_{p,c}^i}{F_{p,c}^i} \right| \times 100
$$

(9)

CC shows the concordance between measured and calculated data. It ranges from 1 to −1 and has the best value of 1. CC is calculated by

$$
\text{CC} = \frac{2r\sigma_m\sigma_c}{\sigma_m^2 + \sigma_c^2 + (F_{p,m} - F_{p,c})^2}
$$

(10)

where $r$ is correlation coefficient, $\sigma_m$ and $\sigma_c$ are standard deviation, while the $F_{p,m}^i$ and $F_{p,c}^i$ are the average of measured and computed particle Froude numbers, respectively.

RESULTS

Decision tree

DT is a classification and data prediction tool which is used broadly in water resources engineering problems as a decision support system machine (Quinlan 1986). Existing information in a data set is used for classification. It uses entropy theory for characterizing the impurity on a data set to obtain accurate information. Results are presented as a diagram called a decision tree. To this end, a tree-based model fits to particular classes with respect to the primary information on each class. Therefore, output of the training algorithm is a kind of discrete class, as DT is a non-parametric regression technique. Considering a two class problem as set $S_d$ which consists only of positive and negative samples, the entropy of set $S_d$ is defined as

$$
\text{Entropy}(S_d) = -p_p (\log_2 p_p) - p_n (\log_2 p_n)
$$

(11)

where $p_p$ and $p_n$ are respectively positive and negative sample proportions in $S_d$. It provides discrete answers for the model from branch to leaves as a systematic entropy reduction. A MATLAB code is utilized for application of DT in which tree optimum size is generated through a binary tree. In DT application, prior sets have higher entropy in comparison with those which have lower entropy. Having the lowest entropy in leaves makes it a confident decision which is based on the decision support system machine. A flowchart of the DT modeling is shown
in Figure 1. The top to down induction of the DT shows the effectiveness degree of variables in the modeling. An example of an interpretation of Figure 1 is given as follows.

Following the gray boxes in Figure 1 helps make this interpretation understandable. The model selects $D_{gr}$ value of 37.1 for the first classification. Substituting the value of 37.1 in Equation (3) gives $d = 1.5$ mm. At the left top branch of DT shown in Figure 1, the next $D_{gr}$ value is 18.3 which gives $d = 0.75$ mm. In most sewer design standards, sediment median size of 1 mm is selected. To this end, at the second level, the right-hand side with $D_{gr}$ higher than 18.3 is selected. According to CIRIA (1986) sediment volumetric concentration in stormwater drainage systems design for bed load transport is equal to 10–200 ppm depending on low to high concentration. Therefore, at the third level of the DT, $C_v$ less than 283 ppm is selected. At the next level, $C_v$ less than 150 ppm is considered to obtain results for medium concentration condition. As drainage systems are a kind of rigid boundary channel, the value of $\lambda$ may be less than 0.02. Hence, the left branch at the fourth level is selected. Taking $C_v$ higher than 68 ppm at the fifth level, $Fr_p$ of 4.7 is obtained. As mentioned previously, the lefthand side of Equation (7) corresponds to $Fr_p$. Taking $Fr_p = 4.7$, $d = 1$ mm and assuming $s = 2.65$, the value of flow velocity $V = 0.6$ m/s is obtained. It is an interesting result as it is the design velocity in many sewer design standards in several countries (CIRIA 1986; Vongvisessomjai et al. 2010).

**Generalized regression neural network**

The GR is used to estimate continuous variables, similar to the standard regression method. The GR finds an ideal function between input and output vectors without using an iterative procedure (Specht 1991). Rapid learning and convergence into optimal regression are the basic advantages of using GR in the modeling. GR is structured in four layers: input, pattern, summation and output layer, as shown in Figure 2. The number of input variables ($x_i$) is the same as the number of neuron in the input layer. Since each layer is connected to the next layer through the weighing vectors between the neurons it can be called a fully connected network. This method is used to find a nonlinear relation for the variables given in Equation (7). Each neuron in the pattern layer is connected to each neuron in the summation layer, in which the sum of weighted outcomes is
calculated at the summation layer. When the weights are set, the GR computes outputs (Y). A MATLAB code is used for application of GR.

As the spread parameter plays an important role in the design of the GR structure, GR models were calibrated by considering the different spread parameters. Through a trial and error procedure, 15 different spread parameters between 0.001 and 1 are tested. Table 2 lists the RMSE, MAPE and CC of each model based on different spread parameters. Table 2 indicates that the GR gives acceptable results for the spreads between 0.001 and 0.1, where the MAPE values are below 20% and the CC values are above 80%. However, for models with spreads larger than 0.1 the error increases significantly. The model no. 1, with spread of 0.001 with RMSE of 0.83, MAPE of 12.66 and CC of 0.90 is found as best model in the application of GR.

### Multivariate adaptive regression splines

MARS is a nonparametric model that finds a nonlinear relationship between dependent and independent variables through adjunction of several linear models. The modeling procedure consists of forward and backward steps. Appropriate independent variables are selected at forward stepwise while in backward stepwise unessential variables are removed to prevent over-fitting of the model and increase the model performance (Friedman 1991; Sharda et al. 2008). The MARS model is established based on basis functions in which variable X is projected to variable Y through applying one of the following basis functions:

\[
y = \max(0, X - c) \quad (12)
\]

\[
y = \max(0, c - X) \quad (13)
\]

where c is a threshold value. In the application of MARS, the desired number of basis functions is selected at the first step. At forward stepwise the defined number of basis functions are applied in the model, and in the backward phase, the model is simplified by removing less important basis functions. In this study, the maximum number of basis functions was set as 30, while the best model performance is obtained with 12 basis functions. In order to retain the continuity of the basis functions, adjacent splines are intersected at knots. The final form of the MARS model is defined by

\[
f = \sum_{i=1}^{n} a + cB_i(X) \quad (14)
\]

where n is the number of basis functions, a and c are constants and \( B_i(X) \) is the basis function. The Salford Predictive Modeler (SPM) is used in the application of MARS, and the following equation is obtained as a bed load self-cleansing model:

\[
Fr_p = 7.26 - 1.75 \times \max(0, d/R - 0.12) + 2 \times \max(0, 0.12 - d/R) + 15.89 \times \max(0, C_v - 0.44) - 16.42 \times \max(0, 0.44 - C_v) + 0.47 \times \max(0, D_{gr} - 0.29) - 7.25 \times \max(0, \lambda - 0.3) - 16.03 \times \max(0, C_p - 0.01) + 5.7 \times \max(0, D_{gr} - 0.12) - 4.33 \times \max(0, D_{gr} - 0.08) + 0.43 \times \max(0, \lambda - 0.59) + 6.75 \times \max(0, \lambda - 0.28) + 1.67 \times \max(0, d/R - 0.07) \quad (15)
\]
Through the modeling process, MARS evaluates the importance of the input variables on computing of the dependent variable in Equation (15). The result is shown as a pie chart in Figure 3. It is found that $C_v$ and $d/R$ have almost similar impacts on the particle Froude number estimation with 29% and 28% of contributions, respectively. Although all independent variables have more than 20% contribution in the modeling process, dimensionless grain size parameter ($D_{gr}$) and channel friction factor ($\lambda$) are less important than $C_v$ and $d/R$.

Comparison of models

In order to evaluate the efficiency of the developed models in this study, three conventional regression models from the literature are selected for comparison. Mayerle et al. (1991), Ab Ghani (1995) and Vongvisessomjai et al. (2010) models are compared with DT, GR and MARS models developed in this study on the testing data set. The models are evaluated based on three statistical indices, RMSE, MAPE and CC given in Table 3, and also with box-and-whisker and scatter plots shown in Figures 4 and 5.

As can be found from Table 3, DT, GR and MARS models outperform all conventional regression models. Despite there being no significant differences between the performance of developed models in this study, the MARS model with RMSE, MAPE and CC of 0.79, 13.79 and 0.92, respectively, is selected as the best model based on statistical indices. The GR model gets second rank among models with RMSE, MAPE and CC of 0.83, 12.56 and 0.90, respectively.

Among conventional regression models, Vongvisessomjai et al. (2010) performed better than Mayerle et al. (1991) and Ab Ghani (1995) models. The Mayerle et al. (1991) model gives worst results with MAPE of more than 50%.

In Figure 4 the box-and-whisker plot is used to compare the measured particle Froude number values with computed corresponding values obtained by DT, GR, MARS and models proposed by Mayerle et al. (1991), Ab Ghani (1995) and Vongvisessomjai et al. (2010). In Figure 4 the upper and lower line of the box are the 25th and 75th percentile of the data, respectively. The height of the box is the inter-quartile range. The line at the middle of the box is related to the median of the data. Extended lines above and below of the box are whiskers.

As is shown in Figure 4, whiskers and box shapes of GR and MARS methods are similar to each other and provide results close to the measured values. Particle Froude number obtained by GR and MARS methods are not noticeably different from each other; however, MARS gives slightly better performance than GR. The notches in boxes of Ab Ghani (1995) and Vongvisessomjai et al. (2010) models are approximately overlapping at the same level, which indicates that both models have similar performances. Although, GR and MARS results match well with measured values, the maximum and minimum particle Froude numbers calculated by different models are obviously different from each other. For instance, the maximum particle Froude number calculated by the Mayerle et al. (1991) model is considerably higher than those calculated by other models. When the median particle Froude number is considered, it can be found that in both the GR and MARS models, median particle Froude number is close to the measured median line level; however, median particle Froude numbers calculated by Mayerle et al. (1991), Ab Ghani (1995) and Vongvisessomjai et al. (2010) are considerably higher than those measured in the

### Table 3 | Comparison of existing equations in the literature with DT, GR and MARS models in terms of RMSE, MAPE and CC

<table>
<thead>
<tr>
<th>Model</th>
<th>RMSE</th>
<th>MAPE</th>
<th>CC</th>
</tr>
</thead>
<tbody>
<tr>
<td>DT</td>
<td>1.03</td>
<td>15.18</td>
<td>0.86</td>
</tr>
<tr>
<td>GR</td>
<td>0.83</td>
<td>12.56</td>
<td>0.90</td>
</tr>
<tr>
<td>MARS</td>
<td>0.79</td>
<td>13.79</td>
<td>0.92</td>
</tr>
<tr>
<td>Mayerle et al. (1991)</td>
<td>3.14</td>
<td>50.61</td>
<td>0.59</td>
</tr>
<tr>
<td>Ab Ghani (1995)</td>
<td>1.51</td>
<td>23.48</td>
<td>0.79</td>
</tr>
<tr>
<td>Vongvisessomjai et al. (2010)</td>
<td>1.27</td>
<td>19.97</td>
<td>0.84</td>
</tr>
</tbody>
</table>
experiments. It is seen from Figure 4 that in conventional regression models, the median line is not placed at the middle of the box, which can be linked to the fact that calculated particle Froude numbers are higher than their corresponding measured values. The median line in the MARS model is located at the same level as the measured median line, which demonstrates the precision of the MARS model. In DT and conventional regression models, the median line is located lower and higher than the measured median line, respectively, indicating underestimation and overestimation in the models. On the other hand, the median line in GR and MARS models are inconsistent.
with the measured median line. Results indicate that the MARS model provides better performance than the other models.

Comparison of the measured and computed particle Froude numbers in Figure 5 reveals that, for GR and MARS models, data are close to the best fit line, which indicates their higher accuracy in comparison with other models. Although the DT model provides results better than conventional regression models, its performance is not as high as GR and MARS models. Ab Ghani (1993) and Vongvisessomjai et al. (2010) models provide relatively similar results; however, it should be noted that conventional regression models overestimate particle Froude number, which is more evident for the Mayerle et al. (1991) model.

The traditional self-cleansing sewer design method based on the minimum velocity and shear stress has deficiencies, because many important factors such as the quantity and the type of sediment are missing. Therefore, it is important to consider a higher number of parameters in developing more precise self-cleansing models. It is worth mentioning that, for developing a robust model, two factors should be considered: range of experimental data and utilized technique. Ranges of experimental data used in this study seem to be adequate as they cover wide ranges of sediment size, pipe size, flow depth, sediment volumetric concentration and channel bed slope. Furthermore, it is found that DT, GR and MARS techniques are powerful tools for establishing a relationship for computing the non-deposition particle Froude number based on flow, fluid, sediment, and channel characteristics. Therefore, they can be used as parsimonious models for sewer pipe design in practice.

CONCLUSIONS

This study successfully applies DT, GR and MARS techniques for sediment transport modeling in sewer pipes using wide ranges of laboratory experimental data. Dimensionless parameters of sediment volumetric concentration, relative particle size, grain size and channel friction factor are considered as input of the models for calculating the particle Froude number at non-deposition condition of sediment transport. DT provides results close to the sewer design standards with design velocity of 0.6 m/s for $d = 1$ mm and $C_v$ less than 150 ppm. DT, GR and MARS models perform much better than all regression models available in the literature. Results obtained by GR and MARS are almost similar; however, MARS gives slightly better performance in comparison with GR. Conventional regression models overestimate particle Froude number considerably, while DT slightly underestimates with a small scatter. As a result, it is found that the MARS technique is a reliable tool for calculating particle Froude number at non-deposition condition of sediment transport in sewer pipes. In order to overcome the deficiency of the conventional self-cleansing design criterion based on a single velocity value, a higher number of hydraulic parameters such as fluid, flow, sediment and channel characteristics must be considered for sewer pipe design. Moreover, for developing a robust sediment transport model, a powerful technique with a wide range of experimental data should be used. Hence, it is considered that the developed models in this study can successfully be implemented as a precise, practical and parsimonious tool for self-cleansing sewer pipe design.

ACKNOWLEDGEMENTS

Special gratitude goes to two anonymous reviewers for their constructive comments, which improved the quality of the manuscript.

REFERENCES


