KSCE Journal of Civil Engineering (2017) 21(5):1956-1963

DOI 10.1007/s12205-016-1784-7

Water Engineering

pISSN 1226-7988, eISSN 1976-3808
www.springer.com/12205

Application of Model Tree and Evolutionary Polynomial Regression for Evaluation of Sediment Transport in Pipes

Mohammad Najafzadeh*, Daniele Biagio Laucelli, and Abdolreza Zahiri*****

Received September 23, 2015/Revised March 29, 2016/Accepted July 11, 2016/Published Online August 30, 2016

Abstract

Prediction of critical velocity for sediment deposition is a significant component in design of sewer pipes. Because of the abrupt changes in velocity and shear stress distributions, traditional equations based on regression analysis can fail in evaluating sediment transport efficiently. Therefore, different artificial intelligence approaches have been applied to investigate sediment transport in sewer pipes. This study proposes two different approaches to predict the critical velocity for sediment deposition in sewer networks: Model Tree (MT) and the Evolutionary Polynomial Regression (EPR), a hybrid data-driven technique that combines genetic algorithms with numerical regression. The hydraulic radius, average size of sediments, volumetric concentration, total friction factor, and non-dimensional sediment size were considered as input parameters to characterize sediment transport in clean sewer pipes. The present study implements data collected from different works in literature. The proposed modeling approaches are compared to some benchmark formulas from literature, and discussed from the accuracy and knowledge discovery points of view, highlighting the advantage of both proposed techniques. Results indicated that both techniques have similar accuracy in predictions, but EPR allows to physical validation of returned formulas, allowing identifying the most influent inputs on the phenomenon at stake.

Keywords: *model tree, evolutionary polynomial regression, sediment transport, sewer pipes, non-deposition conditions, traditional methods*

1. Introduction

In recent years, sediments transport phenomena in sewers has extensively attracted attention by hydraulic engineers. Due to the movements of materials, sediment transport processes in sewers were classified in two groups. The first is related to the deposition condition in which falling velocity of suspended sediment is higher than the average flow velocity in pipe. The second status is the non-deposition mode in which falling velocity of sediments is smaller than the average flow velocity. Movement of sediments in sewers produces several problems such as blockage, surcharging, compaction of sediments, cementation and decrease of effective cross-sectional area (Ab Ghani, 1993). For practical applications, sewer networks were designed only based on a critical velocity criterion (CIRIA, 1987). The main aim of this criterion is that sewers have to be free of any sediment deposition. Low flow criterion in field studies establish a constant minimum flow velocity for non-deposition status equal to 0.6 m/s. Furthermore, this assumption may be often affected by different aspects of environmental and hydraulic status such as bed roughness, sewer type, longitudinal slope of pipe, flow cross section, pollution types, incipient motion of sediment, and physical properties of deposition sediments (Vongvisessomjai *et al.*, 2010).

Several experimental studies were conducted to characterize the sediment transport processes in sewers for two major sediment transport conditions, namely rigid boundary (clean pipes) and loose boundary (pipes with deposited bed) (e.g., Mayerle *et al*., 1991; Ab Ghani, 1993; Nalluri and Ab Ghani, 1996; Arthur *et al*., 1999; Vongvisessomjai *et al*., 2010). Moreover, many aspects of sediment transport phenomenon in sewers are difficult to assess, because they pertains to the three-dimensional nature of this phenomenon. From experimental studies, empirical equations have been extracted for a wide range of different geometric and hydraulic status.

Existing equations based on experimental observations cannot exactly characterize different parameters on the bed load transport. Moreover, the use of the experimental procedures and equipment for field studies are time-consuming and expensive. Therefore, resorting to numerical techniques, such as Artificial Intelligence (AI) techniques, can be effective both for computational

*Assistant Professor, Dept. of Civil Engineering, Graduate University of Advanced Technology-Kerman, P.O. Box 76315-116, Kerman, Iran (Corresponding Author, E-mail: moha.najafzadeh@gmail.com)
**Assistant Professor, Dept. of Civil Engineering and Architecture, Technical University of Bari, Via E. Orabona 4, 71025, Bari, Italy (E-mail: d.laucelli@poliba.it)
***Assistant Professor, Dept. of Water Engineering, Gorgan University of Agricultural Sciences and Natural Resources, P.O. Box 386, Gorgan, Iran (E-mail: zahiri_reza@yahoo.com)

and economic reasons. Additionally, they have been already applied in several engineering fields, proving to be powerful approaches for analyzing different engineering problems. In particular, AI models such as artificial neural networks (ANNs), adaptive neuro-fuzzy inference systems (ANFIS), gene-expression programming (GEP), genetic algorithm (GA), and imperialist competitive algorithms (ICA) were utilized to evaluate sediment transport processes and limiting flow velocity for deposition (or non-deposition) status in sewers (Ab Ghani and Azamathulla, 2011; Azamathulla *et al*., 2012; Ebtehaj and Bonakdari, 2013; Ebtehaj and Bonakdari, 2014).

Outcomes of these models demonstrated more accurate predictions of the critical velocity for sediment deposition in sewer networks than traditional models. Among the artificial intelligence models, methods such as Model Tree (MT) and Evolutionary Polynomial Regression (EPR) are known as powerful soft computing methodologies that can express functional relationships for the studied phenomenon as explicit mathematical expression based on input-output observations, thus allowing more physical insight of the problems.

MT and EPR methodologies have been applied for hydraulic engineering problems. The MT method was frequently used in hydrology and hydraulic engineering sciences such as flood forecasting (Solomatine and Xue, 2004), sediment transport (Bhattacharya and Solomatine, 2005), prediction of mean annual flood (Singh *et al.*, 2009), scour depth prediction around group piers (Etemad-Shahidi and Ghaemi, 2011; Ghaemi *et al*., 2013), and sediment yield estimation in rivers (Goyal, 2014).

The EPR paradigm has been applied in a wider range of civil and environmental engineering problems, as fluid dynamics (Giustolisi *et al*., 2008), prediction of scour depth downstream of grade-control structures (Laucelli and Giustolisi, 2011), groundwater system dynamics (Doglioni *et al*., 2010), soil behavior prediction (Rezania *et al*., 2010), pipe deterioration prediction for water networks (Savic *et al*, 2006), and material behavior modeling (Faramarzi *et al.*, 2014).

In this paper, the MT and EPR methodologies are tested on experimental data sets from literature, to predict the critical velocity for sediment deposition in sewers. Formulations retrieved by the proposed techniques are compared with some benchmark expressions from literature, and results are discussed in order to suggest possible advantages and drawbacks of their application in sewer analysis.

2. Explicit Expressions for Sediment Transport In Sewers

Novak and Nalluri (1975) have studied influences of smooth fixed bed on the sediment transport phenomena in rectangular and circular channels. They proposed two non-dimensional parameters of transport (ϕ) and flow (φ), respectively:

$$\phi = \frac{C_v V R_H}{\sqrt{(S_s-1)gd^3}} \quad (1)$$

$$\varphi = \frac{(S_s-1)gd}{R_H S} \quad (2)$$

in which, S_s, d, R_H, V, C_v, S, and g are specific density of bed sediment, average size of sediment, hydraulic radius of cross-section, critical velocity for sediment deposition, volumetric concentration, longitudinal slope of sewer, and gravitational acceleration, respectively. In addition, Novak and Nalluri (1975) expressed a relationship to predict the critical velocity for sediment deposition in sewer networks by using the resistance formula of Darcy-Weisbach and Eqs. (1) and (2) as:

$$\frac{V}{\sqrt{(S_s-1)gd}} = 1.77 C_v^{1/3} \cdot \left(\frac{R_H}{d}\right)^{1/3} \lambda_s^{-2/3} \quad (3)$$

in which λ_s is the total friction factor.

Mayerle *et al.* (1991) presented an empirical equation based on non-linear regression analysis for predicting the sediment transport in rectangular and circular channels using experimental data sets:

$$\frac{V}{\sqrt{(S_s-1)gd}} = 4.32 C_v^{0.23} \cdot \left(\frac{R_H}{d}\right)^{0.68} \quad (4)$$

Ab Ghani (1993) performed experimental investigations in condition of deposited and non-deposited sediments, using three types of pipes with various diameters (i.e., 154, 305, 450 mm), and assuming the volumetric concentration for sediments between 1 and 14542 ppm. The following equation, based on a regression technique, was proposed in his work:

$$\frac{V}{\sqrt{(S_s-1)gd}} = 3.08 D_{gr}^{-0.09} \cdot C_v^{0.21} \cdot \left(\frac{R_H}{d}\right)^{0.53} \cdot \lambda_s^{-0.4} \quad (5)$$

in which, D_{gr} is non-dimensional size of particles, written as:

$$D_{gr} = d\left(\frac{g(S_s-1)}{\upsilon^2}\right)^{1/3} \quad (6)$$

In which υ is the kinematic viscosity.

Nalluri and Ab Ghani (1996) conducted experiments for pipe diameter equal to 1 m or greater in a deposited bed. They presented following efficient equations:

$$\frac{V}{\sqrt{(S_s-1)gd}} = 1.77 C_v^{0.16} \cdot \left(\frac{d}{D}\right)^{-0.34} \cdot \left(\frac{W_b}{y}\right)^{-0.18} \cdot \lambda_s^{-0.31} \quad (7)$$

$$\lambda_s = 0.0014 C_v^{-0.04} \cdot \left(\frac{R_H}{d}\right)^{0.24} \cdot \left(\frac{W_b}{y}\right)^{0.34} \cdot D_{gr}^{0.54} \quad (8)$$

in which W_b is the width of deposited bed.

Arthur *et al.* (1999) proposed an empirical model for prediction of volumetric concentration using experimental data sets:

$$C_v = \frac{\lambda_s^3 V^5}{30.4(S_s-1)W_s^{1.5} A} \quad (9)$$

in which, W_s and A are settling velocity of the sediment particle and cross-section, respectively. Eq. (9) has been validated for bed shear stress of 1.07 Pa and can be utilized for part-full and full flow in pipes with the sediment concentration between 3 and

1700 ppm. In Arthur *et al.* (1999) experiments, the range of sediment size is from 0.16 to 0.37 mm and pipe diameters are 192, 290 and 445 mm.

Vongvisessomjai *et al.* (2010) have performed laboratory works in bed-load condition on two sewer pipes diameter (100 and 150 mm). They proposed the following empirical equation based on a non-linear regression model:

$$\frac{V}{\sqrt{(S_s-1)gd}} = 4.31C_v^{0.226}\cdot\left(\frac{R_H}{d}\right)^{-0.616} \quad (10)$$

Furthermore, Ab Ghani and Azamathulla (2010) represented an explicit equation for both conditions of deposited and non-deposited load sediments using GEP model:

$$\frac{V}{\sqrt{(S_s-1)gd}} = 1.425+\left(\frac{-0.41}{R_H/d}\right)+\left[\frac{\frac{C_v}{5.91}-1}{D_{gr}}\right] + \left\{\frac{0.014}{\lambda_s}+\lambda_s-8.34\lambda_s^{1.5}D_{gr}\cdot\frac{R_H}{d}\right\} \quad (11)$$

Azamathulla *et al*. (2012) expressed a regression model using the experimental data collection from Ab Ghani (1993) and Vongvisessomjai *et al.* (2010) based on a non-linear analysis to describe bed load transport for no deposition (clean pipes) conditions:

$$\frac{V}{\sqrt{(S_s-1)gd}} = 0.22D_{gr}^{-0.27}\cdot C_v^{0.16}\cdot\left(\frac{R}{d}\right)^{0.29}\cdot\lambda_s^{-0.51} \quad (12)$$

Ebtehaj and Bonakdari (2013) expressed a traditional model for non-deposited bed load condition using Ab Ghani (1993) experimental data sets as follows:

$$\frac{V}{\sqrt{(S_s-1)gd}} = 4.49C_v^{0.21}\cdot\left(\frac{R_H}{d}\right)^{0.54} \quad (13)$$

Equation (13) shows a higher accuracy in comparison with those obtained by May *et al.* (1996) and Vongvisessomjai *et al.* (2010) equations. In deposited bed status, Ebtehaj and Bonakdari (2013) presented a non-linear regressive model based on Ota *et al*. (1999) data sets:

$$\frac{V}{\sqrt{(S_s-1)gd}} = 2.5C_v^{0.375}\cdot\left(\frac{R_H}{d}\right)^{-0.766}\left(\frac{y_s}{D}\right)^{-0.258} \quad (14)$$

Equation (14) was compared with that proposed by Ab Ghani (1993) equation, resulting more accurate in predicting bed load sediment.

3. The Proposed Modelling Approaches

In this section, descriptions of the MT and EPR modelling approaches are presented.

3.1 Model Tree

Among the data mining techniques, model tree are frequently employed to solve the problem by dividing it into several sub-problems (sub-domains) and the result is a combination of these sub-problems. Classification trees classify data records by sorting them down the tree from the root node to some leaf nodes. The difference between the better-known classification trees and the MT technique is that the latter have a numeric value rather than a class label in connection with the leaves. In addition, MT splits the entire input or parameter domain into sub-domains and a linear multivariable regression model is applied for each of them (Quinlan, 1992; Wang and Witten, 1997; Etemad-Shahidi and Ghaemi, 2011). In this way, MT models can be applied to solve continuous class problems and obtain a structural representation of the data sets using the piecewise linear models to approximate nonlinear relationships. Furthermore, this algorithm was known as one of the most effective approach to express meaningfully physical insight of the phenomenon. The tree-building procedure within bunches of linear regression models and knowledge extraction from the structure for corresponding sub-domains is constructed. Based on the domain-splitting criterion, various approaches such as M5 model was frequently utilized to generalize the MT technique (Quinlan, 1992; Wang and Witten, 1997). Through the MT approach, the basic tree is firstly generated using the splitting criterion of the Standard Deviation Reduction (SDR) factor:

$$SDR = sd(E)-\sum_i\frac{|E_i|}{|E|}sd(E_i) \quad (15)$$

in which E, sd, and E_i are the set of examples (data records) that reach the node, the set that results from splitting the node according to the chosen attribute (parameter), and standard deviation, respectively. The M5 utilizes the *sd* parameter as an error measure of the class values that reach a node. Testing all parameters at a node, it computes the expected reduction in error and then selects the parameter that maximizes SDR. This process stops when the standard deviation reduction becomes less than a certain percent of the standard deviation of the original dataset or when only a few data records remain (Quinlan, 1992; Wang and Witten, 1997). Then, a linear regression model is developed for each sub-domain. Only the data in connection with the variables tested in that sub-domain are used in the regression. Other descriptions of the MT model were presented in literature (e.g., Etemed-Shahidi and Ghaemi, 2011; Ghaemi *et al*., 2013; Goyal, 2014).

3.2 Evolutionary Polynomial Regression

EPR can be defined as a non-linear global stepwise regression that provides symbolic formulas of models. Differently from the original stepwise regression of Draper and Smith (1998), EPR is non-linear because the relationships between variables may results into non-linear functions although they are linear with respect to regression parameters. It is global since the search for optimal model structure is based on the exploration of the entire space of models by leveraging a flexible coding of the candidate mathematical expressions.

The expressions achievable by EPR are made of a number of additive terms multiplied by as many coefficients (i.e., like for polynomials) as reported in the following general expression

$$\hat{\mathbf{Y}} = a_0 + \sum_{j=1}^{m} a_j \cdot (\mathbf{X}_1)^{\mathbf{ES}(j,1)} \cdot \ldots \cdot (\mathbf{x}_k)^{\mathbf{ES}(j,k)} \cdot f((\mathbf{X}_1)^{\mathbf{ES}(j,k+1)} \cdot \ldots \cdot (\mathbf{x}_k)^{\mathbf{ES}(j,2k)}) \quad (16)$$

where m is the maximum number of additive terms, $\mathbf{X}_i$ and $\hat{\mathbf{Y}}$ are model input and output variables, function f is chosen by the user and exponents of variables (i.e., $\mathbf{ES}(j, i)$) are selected from a set **EX** of candidates defined by the user (see Giustolisi and Savic (2006) for details).

The genetic algorithm is used to select the exponents $\mathbf{ES}(j, i)$ from among the values in set **EX**. This means that an integer coding of possible alternative exponents $\mathbf{ES}(j, i)$ is adopted to achieve non-linear relationships. It is worth noting that, if the set of exponents contains zero and $\mathbf{ES}(j,i) = 0$, the relevant input disappears from the final expression. Thus, although simple, structures like Eq. (16) are quite versatile and flexible to reproduce patterns in data.

A key point of the EPR model development strategy is that final expressions are linear with respect to coefficient a_j so that they are estimated using classical numerical regression (e.g., least squares). The parameter estimation is solved as a linear inverse problem in order to guarantee a two-ways (i.e., unique) relationship between each model structure and its parameters (Giustolisi and Savic, 2006; Giustolisi *et al.*, 2007). In terms of numerical regression strategy, EPR may produce a non-linear mapping of data (like that achievable by Artificial Neural Networks (Haykin, 1999) although with few constants to estimate and using linear regression for parameters estimation. These features, in turn, help in avoiding over-fitting to training data especially when the dataset is not large. Furthermore, prior assumptions on mathematical structures, functions (i.e., $f(\cdot)$) and number of parameters can be user's initial hypotheses for the automatic model construction. More details on the EPR working sequence are reported in Giustolisi and Savic (2006) and Laucelli and Giustolisi (2011).

The most significant upgrade of the initial EPR paradigm encompasses the multi-objective optimization strategy (i.e., EPR-MOGA) where accuracy of data reproduction and parsimony of model structures are simultaneously maximized (Giustolisi and Savic, 2009). Accuracy is evaluated in terms of determination coefficient (CoD), correlation coefficient (R), root mean square error (RMSE), as indicated in literature (Giustolisi and Savic, 2006; Azamathulla *et al.*, 2012), while parsimony refers to the number of variables and (or) additive terms of the mathematical expressions. Maximizing the parsimony of resulting formulas is aimed at facilitating the physical meaning of final expressions and, in turn, achieving a general description of the underlying phenomenon.

The search space in EPR-MOGA is defined by the user in terms of the base structure of mathematical expressions (e.g., as in Eq. (16) and type of function f), the maximum number of additive terms m, the cardinality of set **EX** of candidate exponents and number of candidate explanatory variables (i.e., k). Additional descriptions of the EPR model have been expressed in literature (Laucelli and Giustolisi, 2011).

4. Case Study

Laboratory investigations on the prediction of bed load and suspended load of sediment in sewers demonstrated that limiting (critical) velocity depends on the hydraulic flow conditions and geometric properties of the cross section (e.g., Ab Ghani, 1993; Vongvisessomjai *et al*., 2010; Ab. Ghani and Azamathulla, 2011; Azamathulla *et al*., 2012; Ebtehaj and Bonakdari, 2013).

Therefore, the relationship between the critical velocity and its influencing parameters can be characterized as follows:

$$V = f(D_{gr} \cdot C_v, R_H, \lambda_s, d, g, S_s - 1) \quad (17)$$

Further studies demonstrated that using non-dimensional parameters can produce more accurate predictions of sediment transport than those obtained using dimensional variables (Ab Ghani and Azamathulla, 2011; Azamathulla *et al*., 2012; Ebtehaj and Bonakdari, 2013). Therefore, according to the approaches adopted in the above-mentioned works, the following functional relationship was considered to develop the MT and EPR models:

$$DFr = \frac{V}{\sqrt{(S_s - 1)gd}} = f\left(D_{gr}, C_v, \frac{R_H}{d}, \lambda_s\right) \quad (18)$$

Equation (18) contains the grouped non-dimensional parameters retrieved from literature, which can be considered as the most important for predicting the densimetric Froude number (DFr), here assumed as representative of critical velocity V.

In this study, 221 datasets were collected from the databases published in Ab Ghani (1993) including small-scale experiments and large-laboratory scale. The training set is made of 167 data samples (about 75%) and the test set of 54 data samples (25%). For the sake of generality of the proposed models, data were selected in order to have different ranges of output values among the available data, as showed in Table 1. This means that having good accuracy on test data will indicate a good generalization ability of the returned models.

Finally, to quantify exhaustively accuracy performances of the proposed models and benchmarks, CoD, R and RMSE were used.

4.1 Implementation of MT

In the present study, in order to find a general mathematical formulation of a function for prediction of critical velocity for

Table 1. Ranges of Variables for Evaluating the Proposed Models

Variables	Ranges on training set	Ranges on test set
C_v	7.59×10^{-7} - 1.45×10^{-3}	2.00×10^{-6} - 7.34×10^{-4}
R/d	3.86 - 156.25	25.75 - 1884.72
λ_s	0.013 - 0.048	0.014 - 0.031
D_{gr}	0.85 - 97.48	0.85 - 23.49
DFr	1.25 - 5.56	5.62 - 26.48

Table 2. Weighting Coefficients of Linear Model for Prediction of Densimetric Froude Number

t_4	t_3	t_2	t_1	t_0	Linear Model Number
-0.4034	-90.241	0.0065	6284.2027	10.649	1
1.9169	-10.358	0.0437	457.607	1.9169	2
-0.0044	-10.358	0.039	457.6071	2.1255	3
-0.0044	-32.1097	0.1062	658.526	2.3766	4
3.0362	-10.746	0.0195	4070.379	3.0362	5

sediment deposition, the following expression has been considered:

$$DFr = t_0 + t_1 C_v + t_2 \frac{R_H}{d} + t_3 \lambda_s + t_4 D_{gr} \tag{19}$$

in which, t_0, t_1, t_2, t_3, and t_4 are constant variables. In this way, the proposed MT approach has four input and one-output parameters. MT technique was developed using 5 rules in form of linear equations. Weighting coefficients $\{t_0, t_1, t_2, t_3, t_4\}$ of the linear model (LMs) were presented in Table 2.

4.2 Implementation of EPR

The EPR application starts from the functional relationship in Eq. (18) for predicting the critical velocity, thus C_v, R_H/d, λ_s, and D_{gr} were considered as candidate inputs.

The range of exponents **EX** is [-3; -2.5; -2; -1.5; -1; -0.5; 0; 0.5; 1; 1.5; 2; 2.5; 3]; the maximum number of polynomial terms is set to $m = 3$, without assuming a bias a_0, and admitting only positive coefficients, i.e. $a_j > 0$. The optimization strategy made use of the following objective functions: (i) the maximization of model accuracy, (ii) the minimization of the number of actually used model inputs (i.e., whose exponent is not 0 in the resulting model structure). Different runs were performed using different options for the definition of $f(\cdot)$ in Eq. (16).

Among several models returned by EPR-MOGA-XL, the following model has been selected, as trade-off between accuracy (on the training set) and parsimony,

$$DFr = 0.404\left(\frac{R_H}{d}\right)^{0.5} + 23.25\left(\frac{R_H}{d}\right)^{0.5} C_v^{0.5} \tag{20}$$

The selected model contains inputs that are recursively present in all the returned models by EPR. This allows a possible identification of the most meaningful input variables among those available (Giustolisi and Savic, 2009). All calculations were performed using the software package EPR-MOGA-XL, working in the MS-Excel environment (Laucelli *et al*., 2012).

4.3 Results Discussion

The EPR model in Eq. (20) includes two out of four inputs, namely the volumetric concentration (C_v) and the ratio between the hydraulic radius of the cross-section and the average size of sediment (R/d). They are included in the model structure having both an exponent equal to 0.5 (square root) with a direct dependence on the target (DFr), as previously experimentally determined by Novak and Nalluri (1975), Ab Ghani (1993) and Azamathulla *et al.* (2012). Therefore, the selected formulation can be considered as consistent with the experimental evidences already found in literature. Additionally, due to the above-described features of the EPR modelling approach, the procedure highlighted the importance of the selected inputs among those available.

Table 3. Evaluation of Models Returned from the MOGA-EPR, MT, and Empirical Equations

Approach	Stage	CoD	R	RMSE
MOGA-EPR(Eq. 20)	Training	0.817	0.907	0.451
	Testing	0.891	0.965	1.77
MT	Training	0.724	0.88	0.556
	Testing	0.888	0.947	1.79
Eq. (3) (Novak and Nalluri, 1975)	Testing	-0.626	0.573	6.86
Eq. (4) (Mayerle *et al*., 1991)	Testing	-3.79	0.981	11.78
Eq. (5)(Ab. Ghani, 1993)	Testing	-19.46	0.984	24.35
Eq. (10) (Vongvisessmojai, 2010)	Testing	-3.4	0.5	11.289
Eq. (12) (Azamathulla *et al.*, 2012)	Testing	-2.49	0.98	10.06

Table 3 reports the accuracy measures for the proposed EPR and MT models and for the benchmark formulations from literature. Firstly looking at the accuracy to training data, both EPR and MT models are basically comparable, with a little advantage of the EPR model. Performances of results for training of the EPR and MT models are also presented in Fig. 1. This trend is also confirmed for the testing stage, were the differences among EPR and MT models are even lower than for training (see Table 3). Performances of results for testing of the EPR and MT models are presented in Fig. 2.

This put in evidence how the proposed modelling approaches are able to produce formulations with a high generalization ability, also having input values out of the training range (see Table 1). This is not the case of the benchmark equations, that all showed worse accuracy from both measures standpoints (see

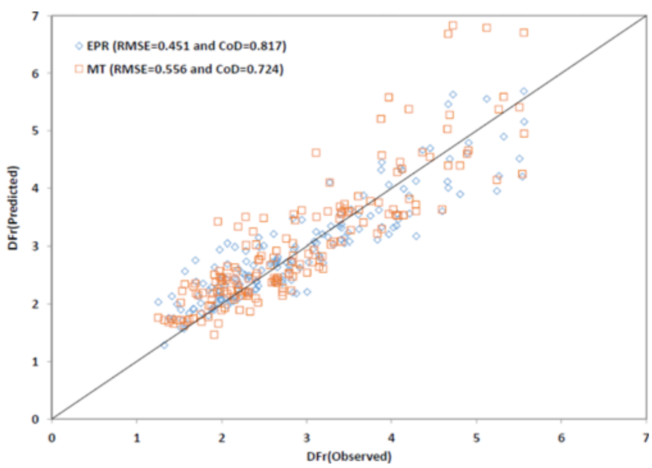

Fig. 1. Performances of Results for Training Stage of the MT and EPR Approaches

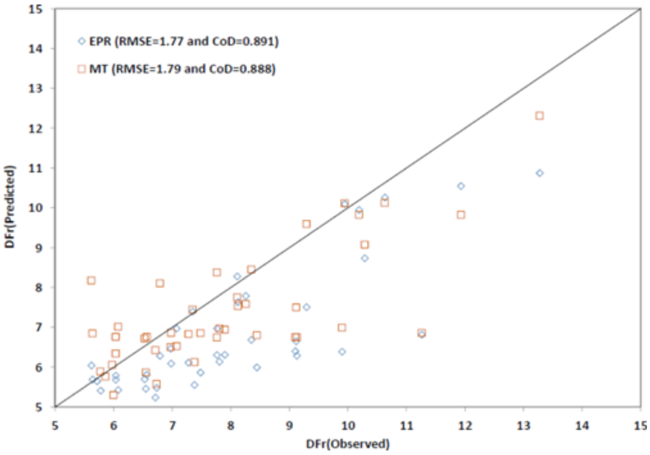

Fig. 2. Performances of Results for Testing Stage of the MT and EPR Approaches

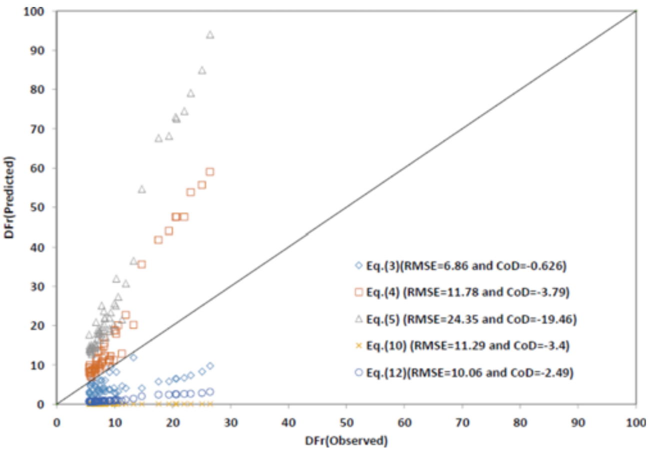

Fig. 3. Performances of the Literature Equations, using the Available Dataset

Table 3). This is evident also by Fig. 3, where prediction of traditional equations and proposed models were plotted. In particular, from Fig. 3, it can be said that all the densimetric Froude number values predicted by the Eq. (12) were under the best fit line. Table 3 indicated that Eq. (12) has quantitatively better performance (CoD = -2.49 and RMSE = 10.06) in comparison with empirical equations proposed by Vongvisessmojai *et al.* (2010) and Mayerle *et al.* (1991) whereas Eq. (12) fails to present the accurate prediction of densimetric Froude number in sewer pipes. This issue was illustrated in Fig. 3. Although the Eq. (12) was developed by all grouped-parameters (C_v, R_H/d, λ_s, and D_{gr}) but also lack of validation for Eq. (12) is related to the drawback of non-linear regression analysis.

However, it should be said that limitations of traditional approaches given by Novak and Nalluri (1975), Ab Ghani (1993), and Vongvisessmojai *et al*. (2010) are pertained to the presence of effective parameters and their data sets ranges. For instance, Mayerle *et al.* (1991) equation was validated only using the C_v and R_H/d parameters. Moreover, Ab Ghani (1993) equation has four non-dimensional parameters and lack of validation is due to restriction of dataset range.

5. Conclusions

In the present study, EPR and MT were used to predict the sediment transport phenomena in sewer pipes. Development of proposed models were carried out for both training and testing stages using four dimensional parameters including volumetric concentration, total friction factor, non-dimensional size of particles, and ratio of hydraulic depth of flow to pipe diameter. Furthermore, empirical equations proposed by Novak and Nalluri (1975), Mayerle *et al.* (1991), Ab Ghani (1993), Vongvisessmojai *et al.* (2010), and Azamathulla *et al*. (2012) were used for comparisons.

The proposed MT and EPR models showed to be comparable from the accuracy point of view, outperforming the benchmark formulations from literature, which over (or under) predict the densimetric Froude number, possibly due to the presence of effective parameters and their data sets ranges.

In particular, some strong points need to be underlined:

1. The proposed application of MT and EPR modelling approach is a challenging test, due the different ranges of output data used for the training and test of the proposed models. This is confirmed by the fact that all the formulation from literature heavily failed in adequately predicting the densimetric Froude number;
2. For the design of sewer networks, if all the analyzed inputs are available, both modelling approaches can be reasonably applied, also out of the used training range (see Table 1);
3. If few inputs are available, the only approach that can be adopted is the EPR, which additionally, due to its intrinsic features, highlighted the importance of some inputs (R/d and C_v) over those available. Additionally, this can be of help for future experiments on sewer sediment transport or for monitoring of existing sewers, limiting the expense for the observing and modelling the phenomenon.

Notations

A = Cross-sectional average flow velocity;
a = Weighting coefficient of EPR models;
CoD = Coefficient of determination;
C_v = Volumetric concentration;
d = Average size of sediment;
D_{gr} = Non-dimensional size of particles
DFr = Densimetric Froude number;
ES = Selected exponents of variables;
EX = Candidate exponents used in EPR development;
f = Function is chosen by the user in EPR development;
g = Gravitational acceleration;
M = Number of samples in training and testing sets;
m = Maximum number of additive terms;
R = Correlation coefficient;
R_H = Hydraulic radius of cross-section;
$RMSE$ = Root means square error;

S = Longitudinal slope of sewer;
sd = Standard deviation;
SDR = Standard deviation reduction factor;
S_s = Specific density of bed sediment;
t = Weighting coefficient used in MT approach;
V = Critical velocity for sediment deposition;
w_b = Dimensionless transverse mixing coefficient;
w_s = Settling velocity of the sediment particle;
$\mathbf{X}$ = Model input;
$\hat{\mathbf{Y}}$ = Output variables;
λ_s Total friction factor;
ϕ = Parameter of transport;
φ = Parameter of flow;
υ = Kinematic viscosity;
Σ = Summation operator;

References

American Society of Civil Engineers (ASCE). "Water pollution control federation, design and construction of sanitary and storm sewers." *American Society of Civil Engineers Manuals and Reports on Engineering Practices,* No. 37, 1970.

Arthur, S., Ashley, R., Tait, S., and Nalluri, C. (1999). "Sediment transport in sewers-a step towards the design of sewers to control sediment problems." *P. I. Civil Eng.Water,* Vol.136, No. 1, pp. 9-19, DOI: 10.1680/iwtme.1999.31264.

Ab Ghani, A. (1993). *Sediment transport in sewers*, PhD Thesis, University of Newcastle upon Tyne, UK.

Ab Ghani, A. and Azamathulla, H. Md (2011). "Gene-expression programming for sediment transport sewer pipe systems." *J. Pipeline. Syst. Eng. Prac.,* ASCE. Vol. 2, No. 3, pp. 102-106, DOI: 10.1061/(ASCE)PS.1949-1204.0000076.

Azamathulla, H. Md., Ghani, A. A., and Seow, Y. F. (2012). "ANFIS-based approach for predicting sediment transport in clean sewer." *Applied Soft Comput.,* Vol. 12, No. 3, pp. 1227-1230, DOI: 10.1016/j.asoc.2011.12.003.

Bhattacharya, B. and Solomatine, D. P. (2005). "Neural networks and M5 model trees in modelling water level-discharge relationship." *Neurocomp.*, Vol. 63, pp. 381-396, DOI: 10.1016/j.neucom.2004.04.016.

Doglioni, A., Mancarella, D., Simeone, V., and Giustolisi, O. (2010) "Inferring groundwater system dynamics from hydrological time-series data." *Hydrol. Sci. J.,* Vol. 55, No. 4, pp. 593-608, DOI: 10.1080/02626661003747556.

Draper, N. R. and Smith, H. (1998). *Applied regression analysis*, Wiley & Sons, New York.

Ebtehaj, I. and Bonakdari, H. (2013). "Evaluation of Sediment Transport in Sewer Using Artificial Neural Network." *Eng. Appl. Comput. Fluid Dynamics.,* Vol. 7, No. 3, pp. 382-392, DOI: 10.1080/19942060.2013.11015479.

Ebtehaj, I. and Bonakdari, H. (2014). "Performance evaluation of adaptive neural fuzzy inference system for sediment transport in sewers." *Water Resources Management,* Vol. 28, No. 13, pp. 4765-4779.

Etemad-Shahidi, A. and Ghaemi, N. (2011). "Model tree approach for prediction of pile groups scour due to waves." *Ocean Engineering*, Vol. 38, pp. 1522-1527, DOI: 10.1016/j.oceaneng.2011.07.012.

Faramarzi, A., Alani, A. M., and Javadi, A. A. (2014) "An EPR-based self-learning approach to material modelling." *Computers & Structures*, Vol. 137, pp. 63-71, DOI: 10.1016/j.compstruc.2013.06.012.

Giustolisi, O. and Savic, D. A. (2006). "A symbolic data-driven technique based on evolutionary polynomial regression." *J. of Hydroinformatics*, Vol. 8, pp. 207-222, DOI: 10.2166/hydro.2006.020.

Giustolisi, O., Doglioni, A., Savic, D. A., and Webb, B. (2007). "A multi-model approach to analysis of environmental phenomena." *Environmental Modelling & Software*, Vol. 22, pp. 674-682, DOI: 10.1016/j.envsoft.2005.12.026.

Giustolisi, O. and Savic, D. A. (2009). "Advances in data-driven analyses and modelling using EPR-MOGA." *J. of Hydroinformatics*, Vol. 11, pp. 225-236, DOI: 10.2166/hydro.2009.017.

Goyal, M. K. (2014). "Modeling of sediment yield prediction using m5 model tree algorithm and wavelet regression." *Water Resources Management*, Vol. 28, No. 7, pp. 1991-2003, DOI: 10.1007/s11269-014-0590-6.

Ghaemi, N., Etemad-Shahidi, A., and Ataie-Ashtiani, B. (2013). "Estimation of current-induced pile groups scour using a rule based method." *Journal of Hydroinformatics*, IWA, Vol. 15, pp. 516-528, DOI: 10.2166/hydro.2012.175.

Haykin, S. (1999). *Neural networks: A comprehensive foundation* (2nd edition), Prentice-Hall Inc., Englewood Cliffs, New Jersey, USA.

Laucelli, D., Berardi, L., Doglioni, A., and Giustolisi, O. (2012). "EPR-MOGA-XL: An excel based paradigm to enhance transfer of research achievements on data-driven modeling." *Proceedings of 10th International Conference on Hydroinformatics HIC 2012,* 14-18 July, Hamburg, Germany, R. Hinkelmann, M.H. Nasermoaddeli, S.Y. Li-ong, D. Savic, P. Fröhle (Eds).

Laucelli, D. and Giustolisi, O. (2011). "Scour depth modelling by a multi-objective evolutionary paradigm." *Environmental Modeling & Software*, Vol. 26, No. 4, pp. 498-509, DOI: 10.1016/j.envsoft.2010.10.013.

Mayerle, R., Nalluri, C., and Novak, P. (1991). "Sediment transport in rigid bed conveyances." *J. Hydraul. Res.,* Vol. 29, No. 4, pp. 475-496, DOI: 10.1080/00221689109498969.

Nalluri, C., Ghani, A. A., and El-Zaemey, A. K. S. (1994). "Sediment transport over deposited beds in sewers." *Water Sci. Tech.,* Vol. 29, pp. 125-133.

Nalluri, C. and Ab. Ghani, A. (1996). "Design options for self-cleansing storm sewers." *Water Sci. Tech.,* Vol. 33, No. 9, pp. 215-220, http://www.iwaponline.com/wst/03309/wst033090215.htm.

Novak, P. and Nalluri, C. (1975). "Sediment transport in smooth fixed bed channels." *J. Hydraul. Div.,* ASCE. Vol. 101, No. 9, pp. 1139-1154, http://cedb.asce.org/cgi/WWWdisplay.cgi?6183.

Quinlan, J. R. (1992). "Learning with continuous classes." *Adams, Sterling, editors. Proceedings of AI'92. World Scientific,* pp. 343-348.

Rezania, M., Javadi, A. A., and Giustolisi, O. (2010). "Evaluation of liquefaction potential based on CPT results using evolutionary polynomial regression." *Computers and Geotechnics*, Vol. 37, pp. 82-92, DOI: 10.1162/106365600568158.

Savic, D. A., Giustolisi, O., Berardi, L., Shepherd, W., Djordjevic, S., and Saul, A. (2006). "Modelling sewer failure by evolutionary computing." *Water Management Journal*, Vol. 159, No. 2, pp. 111-118, DOI: 10.1680/wama.2006.159.2.111.

Solomatine, D. P. and Xue, Y. P. (2004). "M5 model trees and neural networks: Application to flood forecasting in the upper reach of the Huai River in China." *J. Hydrol. Eng*., Vol. 9, No. 6, pp. 491-501, DOI: 10.1061/(ASCE)1084-0699(2004)9:6(491).

Singh, K. K., Pal, M., and Singh, V. P. (2009). "Estimation of mean annual flood in indian catchments using backpropagation neural network and M5 model tree." *Water Resources Management*, Vol. 24, No. 10, pp. 2007-2019, DOI: 10.1007/s11269-009-9535-x.

Wang, Y. and Witten, I. H. (1997). "Induction of model trees for predicting continuous lasses." *Proceedings of the Poster Papers of the European Conference on Machine Learning, University of Economics*, Faculty of Informatics and Statistics, Prague.

Vongvisessomjai, N., Tingsanchali, T., and Babel, M. S. (2010). "Non-deposition design criteria for sewers with part-full flow." *Urban Water J.,* Vol. 7, No. 1, pp. 61-77, DOI: 10.1080/15730620903242824.