ISBN 978-83-927084-8-3
ISSN 0867-7964

A DECISION TREE SELF-CLEANSING MODEL FOR STORMWATER DRAINAGE SYSTEMS

Mirali Mohammadi[1], Mir-Jafar-Sadegh Safari[2], Babak Vaheddoost[2] and Hafzullah Aksoy[2]

[1] *Department of Civil Engineering, Urmia University, P O Box 165, Urmia, Iran.*
and The University College of Science & Technology (Elm-O-Fann), Urmia, Iran.
e-mail: *m.mohammadi@urmia.ac.ir*
[2] *Department of Civil Engineering, Istanbul Technical University, Istanbul, Turkey.*
e-mails: *safari@itu.edu.tr, vaheddoostb@itu.edu.tr, haksoy@itu.edu.tr*

Abstract: Experimental studies were carried out in order to investigate required self-cleansing velocity in no-deposition without deposited bed condition. Data from four different channels cross-sectional shape namely; trapezoidal, rectangular, circular and U-shaped and non-cohesive sands are used in this study. A Decision Tree Model (DTM) is developed as a classification and data estimation tool. The model established using experimental data and based on flow, fluid, sediment and channel characteristics. The DTM suggests typical self-cleansing velocity of 0.58 m/s for no-deposition without deposited bed condition. The performance of DTM is compared with the available self-cleansing models in the literature. The accuracy of no- deposition self-cleansing models in the literature are less than DTM due to the fact that they were especially developed for circular and rectangular channels. It is concluded, channel cross-section shape plays an important role in self-cleansing design criteria.

KEY WORDS: cross-sectional shape; Decision Tree Model; drainage systems; sediment transport; self-cleansing.

NOTATION

C_v ; volumetric sediment concentration
D_{gr} ; dimensionless grain size parameter
d ; sediment median size
g ; gravitational acceleration
n ; number of data
p_p ; proportion of positive examples in S
p_n ; proportion of negative examples in S
R ; hydraulic radius
r ; correlation coefficient
S ; set
S_c ; channel bed slope
s ; sediment relative density
V ; flow mean velocity

V^c	; calculated flow mean velocity
V^m	; measured flow mean velocity
$\overline{V^c}$	; average of calculated flow mean velocities
$\overline{V^m}$	; average of measured flow mean velocities
λ_s	; channel friction factor with the presence of sediment
ν	; kinematic viscosity of fluid
ρ	; fluid density
ρ_s	; sediment density
σ_c	; standard deviations of calculated flow mean velocities
σ_m	; standard deviations of measured flow mean velocities

1. INTRODUCTION

Continuous deposition of sediment causes to change the velocity and wall shear stress distribution in rigid boundary channels such as sewers, irrigation channels and stormwater drainage systems and significantly affects the carrying capacity and hydraulic resistance of the channel. Self-cleansing is a design criteria which is desired a condition in which sediment particles in motion must be transported through the flow, or deposited sediment at the channel bottom must be removed. According to this definition, self-cleansing-based models can be classified in two groups of “moving existing sediment on the bed” and “no-deposition”. The first group considers as “incipient motion” and “scouring” models (Ambrose, 1953; Novak & Nalluri, 1984; Safari et al., 2011). The second group of self-cleansing models considers as “no-deposition-without deposited bed, “no-deposition-with deposited bed” and “incipient deposition”. The first and second subgroups are the limiting or minimum velocity or shear stress for no-deposition condition (limit-of-deposition) that in former, channel has not deposited bed however, in the later, it allows for a small portion of sediment deposition at the bottom of sewer (maximum deposited bed is 1-2% of pipe diameter) thus reduces the channel slope (May, 1993; May et al., 1996; Ota & Nalluri, 2003). Incipient deposition is considered as a point that sediment particles in suspension begin to deposit and transport as bed load (Safari et al., 2015).

Machine learning methods recently are used to study sediment transport in drainage and sewer systems (Ab Ghani & Azamathulla, 2010; Safari et al., 2013; Ebtehaj & Bonakdari, 2014). In present study the application of Decision Tree Model (DTM) is investigated as a machine learning method.

DTM (Quinlan, 1986) is a kind of decision support system (DSS) tools used for classification and data estimation in many branches. It uses uncertainty and available information in data sets for classification. In order to define information gain precisely, it is important to define a measure which used in information theory, called entropy that characterizes the impurity of an arbitrary collection of examples. The result of such modeling is usually a diagram named decision tree. In classification problems the goal is to fit a tree based patterns to respective classes based on previous observations from each class. Thus the output of the learning algorithm is one of a discrete set of possible classes rather than as in nonparametric regression (Orr, 1996).

DTM was also used for sediment transport studies. Bhattacharya et al. (2007) used M5 decision tree for sediment bed load transport. Goyal (2014) developed a decision tree based M5 model for estimation of sediment yield in watershed. Reddy & Ghimire (2009) used M5 model tree (MT) to predict suspended sediment loads in rivers.

In this study, experimental studies have been carried out for investigation of sediment transport in no-deposition condition. A DTM no-deposition self-cleansing model is established and compared with available models in the literature.

2. SELF-CLEANSING DESIGN BASED ON “NO-DEPOSITION”

It is the minimum flow velocity or minimum bed shear stress required for retention of sediment in motion within flow. This is the conventional design criteria where a single value of velocity (0.3-1 m/s) or shear stress (1-12.6 N/m^2) is used based on experience and without any theoretical rationalization. Available design criteria are reported by the Construction Industry Research and Information Association in the UK (CIRIA, 1986), Nalluri & Ab Ghani (1996) and Vongvisessomjai et al. (2010). Many important factors are missing in this method like quantity and type of sediment and sewer size. Rather than just using a single value, no-deposition design concept was further modified to use more parameters in the 1990s which resulted in case of no-deposition without deposited bed and with deposited bed design concepts (Vongvisessomjai et al., 2010). Safari et al. (2014 & 2015) used the concept of incipient deposition for the same purpose.

2.1. NO DEPOSITION WITHOUT DEPOSITED BED

In this design criteria channel has not deposited bed. The Models developed for both suspended load and bed load. Due to this study investigates sediment transport in bed load, the self-cleansing models that proposed for bed load sediment transport will be mentioned.

Mayerle et al. (1991) studied of bed load sediment transport in fixed bed channels for no-deposition without deposited bed condition. They conducted experiments in two rectangular tilting flumes with smooth and artificially roughened beds and in a tilting pipe channel with smooth bed. Mayerle et al. (1991) proposed

$$\frac{V}{\sqrt{gd(s-1)}} = 5.45 C_v^{0.15} D_{gr}^{-0.11} \left(\frac{d}{R}\right)^{-0.43} \tag{1}$$

and

$$\frac{V}{\sqrt{gd(s-1)}} = 4.32 C_v^{0.23} \left(\frac{d}{R}\right)^{-0.68} \tag{2}$$

for rectangular and circular cross-sections, respectively in which, *V* is the flow mean velocity, *d* is median particle size, *g* is gravitational acceleration, *s* is relative sediment density, C_v is volumetric sediment concentration, R is hydraulic radius and D_{gr} is dimensionless grain size parameters defined by

$$D_{gr} = \left(\frac{(s-1)gd^3}{\nu^2} \right)^{1/3} \tag{3}$$

in which ν is kinematic viscosity of fluid. Ab Ghani (1993) investigated bed load sediment transport of without deposition in pipes and suggested

$$\frac{V}{\sqrt{gd(s-1)}} = 3.08 C_v^{0.21} D_{gr}^{-0.09} \left(\frac{R}{d} \right)^{0.53} \lambda_s^{-0.21} \tag{4}$$

as no-deposition self-cleansing model in without deposited bed condition for circular channels, where λ_s is channel friction factor with the presence of sediment. Vongvisessomjai et al. (2010) studied no-deposition condition in suspended load and bed load. Experiments were conducted in two circular pipes and proposed

$$\frac{V}{\sqrt{gd(s-1)}} = 4.31 C_v^{0.226} \left(\frac{d}{R} \right)^{-0.616} \tag{5}$$

as bed load no-deposition self-cleansing model in without deposited bed condition for circular channels.

No-deposition self-cleansing models are semi-empirical or semi-theoretical and were developed using non-linear regression analysis. The aforementioned models developed based on velocity approach and especially for rectangular and/or circular channel sections.

3. EXPERIMENTAL APPARATUS AND PROCEDURE

Different experiments were conducted in the Hydraulics Laboratory of Istanbul Technical University, Turkey. A trapezoidal cross-section channel, having 12 m long with 30 cm bottom width was used. The inclined length of each side wall was 30 cm with outer angel 60^0. Water for the experiment was obtained from a tank supplied by the constant head water supply system of the laboratory. Uniform flow was established in the channel by adjusting the downstream tailgate. For each experiment, discharge was measured by an ultrasonic flow-meter. In total, ten channel bed slopes were set from 0.1% to 1%. The sediment was poured into the channel by a vibrant cube feeder that was installed at the beginning of the channel. Two kinds of uniform non-cohesive sediment were used in this study with median size of 0.15 and 0.83 mm, fine and coarse sands with geometric standard deviation of 1.36 and 1.21, respectively.

In this study data available in the literature are used as well. Loveless (1992) studied sediment transport in incipient deposition and no-deposition conditions in several cross-section channels namely; rectangular, circular and U-shaped. A detail of experimental apparatus and procedure may be found in Loveless (1992) and Safari et al. (2015).

4. APPLICATION OF DECISION TREE MODEL

In within-flow sediment transport, characteristics of flow, fluid, sediment and channel are of importance. Reviewing the no-deposition self-cleansing models based on velocity approach, no-deposition flow velocity (V), hydraulic radius (R), gravitational acceleration (g); fluid density (ρ) and kinematic viscosity (v), particles median size (d) and density (ρ_s), volumetric sediment concentration (C_v) and channel bed slope (S_c) are selected as most important parameters. Sediment relative density (s) is used instead of ρ and ρ_s and among the variables, kinematic viscosity (v) and acceleration gravity (g) are constant, therefore, eliminated to achieve the final expression of self-cleansing velocity as:

$$V = f(R, s, d, C_v, S_c) \tag{6}$$

The application of the DTM consisted of two steps of training (classification) and testing (estimation). At first, the algorithm trained with training data set, a test set is then used to verify the models and to estimate the expected performance. Among the data of 63 experiments (12 set of experiments performed in this study plus 51 set of experiments given by Loveless, 1992), so 53 experiments are used to train the model and tree generation and 10 experiments randomly are selected for estimation, respectively. In order to ensure that each variable is treated equally in a model, the data are normalized to the interval of [0, 1].

Given a set S, containing only positive and negative examples of some target concepts (a two class problem), entropy of the set S relative to this simple, binary classification is defined as:

$$Entropy(S) = -p_p\left(\log_2 p_p\right) - p_n\left(\log_2 p_n\right) \tag{7}$$

where, p_p is the proportion of positive examples in S and p_n is the proportion of negative examples in S. This procedure results in a systematic reduction of entropy downward from main branch to the leaves which are discrete answers to the model. A MATLAB based code “RegressionTree” is used for data treatment and calculation of positive and negative proportion of the sets while optimum size of the tree is developed by a binary tree. Priority of steps are selected so that, sets with the highest entropy are prior to those with lower amount of entropy. This procedure as DSS system helps to consider the most confident decision near leaves (i.e. containing lowest entropy). Result of such diagram can be translated step by step and is applicable experimentally from diagram. Figure 1 shows the results of the DTM procedure.

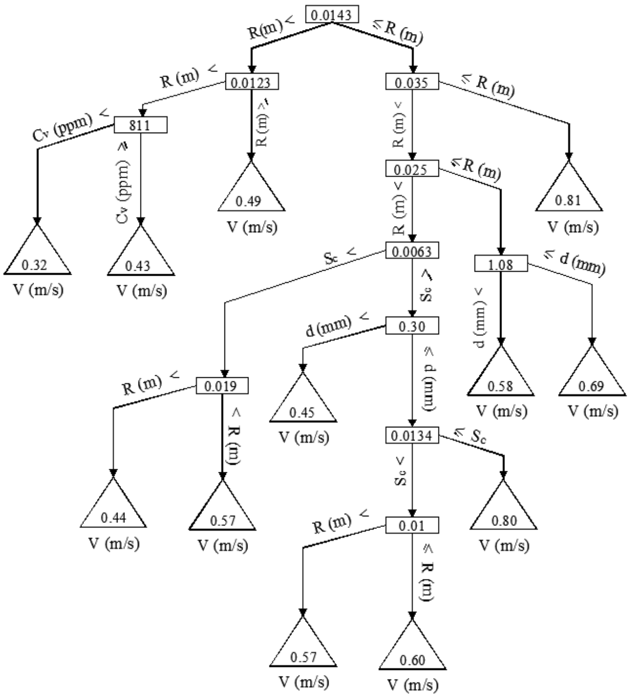

Fig. 1. A DTM algorithm.

Figure 1 indicates that variable of hydraulic radius (R) makes the most uncertainty in the model. It can be attributed to the channel cross-section shape. Considering a certain cross-section area and discharge, channel having larger wetted perimeter will provide smaller hydraulic radius. In the first stage a threshold of R=0.0143 m, is used for classification. It reveals that for values of R lower than 0.0143 m, flow mean velocity is dependent only on hydraulic radius (R) and sediment concentration (C_v). The significance of sediment median diameter (d) is first revealed in $R \geq 0.025$ m. Interestingly, DTM proposed self-cleansing velocity of 0.58 m/s for d <1.08 mm. It is a reasonable value of self-cleansing velocity for sediment particle size of near to 1 mm, according to the literature (0.3-1 m/s). As the results of DTM, for $R < 0.025$ m, channel bed slope (S_c) play an important role in DTM algorithm. However, in lower bed slopes the flow mean velocity is dependent on sediment size again.

5. COMPARISON OF MODELS

The results obtained by DTM are compared with the models given by Mayerle et al. (1991), Eqs. (1-2), Ab Ghani (1993), Eq. (4) and Vongvisessomjai et al. (2010), Eq. (5). Models performances are evaluated by two statistical performance criteria; the mean absolute percentage error (*MAPE*) and concordance coefficient (*CC*). The *MAPE* which gives the model accuracy as the percentage is defined by

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{\left(V^c\right)_i - \left(V^m\right)_i}{\left(V^m\right)_i}\right| \times 100 \tag{8}$$

in which V^c and V^m, respectively, are the calculated and measured flow mean velocity in no-deposition condition; *n* is the number of data, *CC* is the concordance between measured and calculated values and has the range from -1 to 1, with perfect agreement at 1. It is computed by

$$CC = \frac{2r\sigma_m\sigma_c}{\sigma_m^2 + \sigma_c^2 + \left(\overline{V^m} - \overline{V^c}\right)^2} \tag{9}$$

where *r* is a correlation coefficient, σ_m and σ_c are standard deviations of measured and calculated flow mean velocities, respectively and $\overline{V^m}$ and $\overline{V^c}$ are average of the measured and calculated flow mean velocities.

MAPE and CC calculated for each model as given in Table 1. Considering the results in Table 1 based on *MAPE* and *CC* demonstrates that DTM has a good performance for predicting flow mean velocity in no-deposition condition with *MAPE* and *CC* of 9.71% and 0.96, respectively. Among regression models, Eq. (4) of Ab Ghani (1993) has acceptable performance with *MAPE* and *CC* of 19.57% and 0.85, respectively. Eq. (1) of Mayerle et al. (1991) which proposed for rectangular cross-section channels, and Eq. (5) of Vongvisessomjai et al. (2010) have intermediate performances with *MAPE* of 25.58% and 34.90 %; *CC* of 0.61 and 0.64, respectively. Eq. (2) of Mayerle et al. (1991) proposed for circular cross-section channels has a poor performance with *MAPE* and *CC* of 56.50 % and 0.45, respectively.

Table 1 Performance of models based on *MAPE* and *CC*

Model	DTM	Eq. 1, (Rec.)	Eq. 2, (Cir.)	Eq. 4, (Cir.)	Eq. 5, (Cir.)
MAPE (%)	9.71	25.58	56.50	19.57	34.90
CC	0.96	0.61	0.45	0.85	0.64

Rec. Rectangular, Cir. Circular

The goodness-of-fit is also examined by the scatter plots of the measured and calculated flow mean velocity in no-deposition condition (Fig. 2). It is seen from Figure (2a) that calculated flow mean velocities by DTM match with the measured velocities. Generally, regression models overestimate flow mean velocity in comparison with DTM. As Figure (2d) indicates, although model of Ab Ghani (1993) slightly overestimates flow mean velocity in no-deposition condition, a few data remained away and some of the data fall on the bisector line.

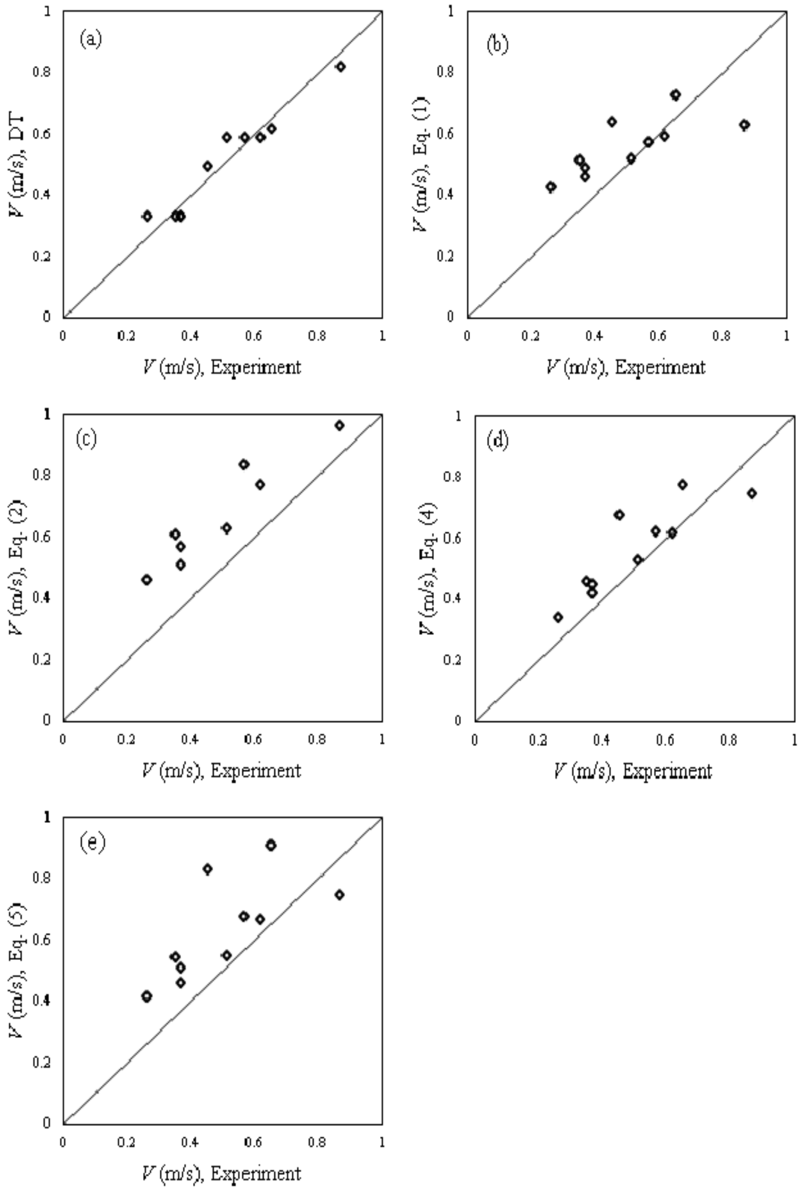

Fig. 2 Comparison between measured in experiments and calculated flow mean velocity by DTM and regression models

It should be mentioned that regression models were developed for a specific channel cross-section. However, in this study data from four different cross-sections are used for developing a DTM. On the other hand, the performance of regression models evaluated by testing data set which belongs to four different cross-section channels. It is a reason that why regression models have poor performances in comparison with DTM. The results demonstrate that channel cross-section is an important parameter in sediment transport in rigid boundary channels.

6. CONCLUSIONS

The application of DTM on sediment transport as self-cleansing design criteria are investigated based on velocity approach. DTM is a model developed by considering flow, fluid, sediment and channel characteristics such as flow mean velocity, hydraulic radius, sediment relative density, median size, volumetric sediment concentration and channel bed slope. As a result of models evaluation, DTM is found superior to regression self-cleansing models. The self-cleansing models in the literature are applicable only to circular and rectangular cross-sections and they have less accuracy in comparison with DTM established based on the data obtained from four different channel cross-sectional shapes. Considering DTM classification algorithm, hydraulic radius makes the most uncertainty in the model and is the most important parameter in velocity approach. The results achieved from models evaluation and DTM classification algorithm demonstrates the importance of channel cross-section shape. It may be concluded that DTM has a higher capability to calculate no-deposition self-cleansing flow mean velocity independent from channel cross-section shape.

ACKNOWLEDGEMENTS

This study was supported by the Scientific and Technological Research Council of Turkey (TUBITAK) under the projects 113M062.

REFERENCES

1. Ab Ghani, A. 1993. Sediment transport in sewers. PhD Thesis, University of Newcastle Upon Tyne, UK.
2. Ab Ghani, A., Azamathulla, H. M., 2010. Gene-expression programming for sediment transport in sewer pipe systems. Journal of Pipeline Systems Engineering and Practice, 2(3), 102-106.
3. Ambrose, H.H. 1953. The transportation of sand in pipes free surface flow. Proceeding of the Fifth Hydraulic Conference, Bulletin 34, State University of Iowa Studies in Engineering, Iowa State University, Iowa.
4. Bhattacharya, B., Price, R.K., Solomatine, D.P. 2007. Machine learning approach to modeling sediment transport. Journal of Hydraulic Engineering, 133(4), 440-450.
5. Butler, D., May, R., Ackers, J. 2003. Self-Cleansing Sewer Design Based on Sediment Transport Principles. Journal of Hydraulic Engineering 129(4), 276-282.
6. CIRIA, 1986. Sediment movement in combined sewerage and storm-water drainage systems. Phase 1. Project report. London: CIRIA research project No. 336.
7. Ebtehaj, I., Bonakdari, H. 2014. Performance Evaluation of Adaptive Neural Fuzzy Inference System for Sediment Transport in Sewers. Water Resources Management, 28(13), 4765-4779.
8. Goyal, M.K. 2014. Modeling of sediment yield prediction using M5 model tree algorithm

and wavelet regression. Water resources management, 28(7), 1991-2003.

9. Loveless, J.H. 1992. Sediment transport in rigid boundary channels with particular reference to the condition of incipient deposition. PhD Thesis, University College of London.
10. May, R.W.P. 1993. Sediment transport in pipes and sewers with deposited beds. Technical Report, Hydraulic Research Ltd., Report SR 320, Wallingford, UK.
11. May, R.W.P., Ackers, J. C., Butler, D., John, S. 1996. Development of design methodology for self-cleansing sewers. Water Science and Technology 33(9), 195-205.
12. Mayerle, R., Nalluri, C., Novak, P. 1991. Sediment transport in rigid bed conveyances. Journal of Hydraulic Research 29(4), 475-495.
13. Nalluri, C., Ab Ghani, A. 1996. Design options for self-cleansing storm sewers. Water Science and Technology 33(9), 215-220.
14. Novak, P., Nalluri, C. 1984 Incipient motion of sediment particles over fixed beds. Journal of Hydraulic Research 22(3), 181-197.
15. Orr M.J. 1996. Introduction to radial basis function networks. Center for Cognitive Science, University of Edinburgh; (Technical report).
16. Ota, J.J., Nalluri, C. 2003. Urban storm sewer design: Approach in consideration of sediments. Journal of Hydraulic Engineering 129(4), 291-297.
17. Quinlan, J.R. 1986. Induction of decision trees. Machine Learning, 1, 81–106.
18. Reddy, M.J., Ghimire, B.N. 2009. Use of model tree and gene expression programming to predict the suspended sediment load in rivers. Journal of Intelligent Systems, 18(3), 211-228.
19. Safari, M.J.S., Mohammadi, M., Manafpour, M. 2011. Incipient motion and deposition of sediment in rigid boundary channels. Proceedings 15th International Conference on Transport & Sedimentation of Solid Particles, 6-9 September, Wroclaw, Poland, 63-75.
20. Safari M.J.S., Aksoy H. & Mohammadi M., 2013. Application of ANN for the analysis of sediment incipient deposition in rigid boundary channels. Proceedings 16th International Conference on Transport & Sedimentation of Solid Particles, 18-20 September, Rostock, Germany. 53-62.
21. Safari M.J.S., Mohammadi, M., Gilanizadehdizaj, G. 2014. On the effect of cross sectional shape on incipient motion and deposition of sediments in fixed bed channels. Journal of Hydrology and Hydromechanics 62(1), 75-81.
22. Safari, M.J.S., Aksoy, H., Mohammadi, M. 2015. Incipient deposition of sediment in rigid boundary open channels. Environmental Fluid Mechanics DOI: 10.1007/s10652-015-9401-8 (in press).
23. Vongvisessomjai, N., Tingsanchali, T., Babel, M.S. 2010. Non-deposition design criteria for sewers with part-full flow. Urban Water Journal 7(1), 61-77.