Journal of Hydrology 598 (2021) 126452

Contents lists available at ScienceDirect

Journal of Hydrology

journal homepage: www.elsevier.com/locate/jhydrol

Research papers

Weighted instances handler wrapper and rotation forest-based hybrid algorithms for sediment transport modeling

Katayoun Kargar [a,*], Mir Jafar Sadegh Safari [b], Khabat Khosravi [c]

[a] Department of Civil Engineering, Ryerson University, Toronto, Ontario, Canada
[b] Department of Civil Engineering, Yaşar University, Izmir, Turkey
[c] Department of Watershed Management Engineering, Ferdowsi University of Mashhad, Mashhad, Iran

ARTICLE INFO

This manuscript was handled by Andras Bardossy, Editor-in-Chief, with the assistance of Fi-John Chang, Associate Editor

Keywords:
Hybrid models
Open channel
Optimization
Rotation forest
Sediment transport
Weighted instances handler wrapper

ABSTRACT

Sediment transport modeling has been known as an essential issue and challenging task in water resources and environmental engineering. In order to minimize the adverse impacts of the continues sediment deposition that is known as a main source of pollution in the urban area, the self-cleansing method is widely utilized for designing the sewer pipes to create a condition to keep the bottom of channel clean from sedimentation. In the present study, an extensive data range is utilized for modeling the sediment transport in non-deposition with clean bed condition. Regarding the effective parameters involved, four different scenarios are considered for the modeling. To this end, four standalone methods including the M5P, reduced error pruning tree (REPT), random forest (RF) and random tree (RT) and two hybrid models based on rotation forest (ROF) and weighted instances handler wrapper (WIHW) techniques are developed and result compared with three empirical equations. Based on the results, the hybrid WIHW-RT and WIHW-RF models provide better performance in particle Froude number estimation in comparison to other standalone and hybrid models. Performances of the most of the models are found accurate except RT and REPT standalone models. The outcomes revealed that the empirical models have considerable overestimation. Generally, hybrid data mining methods yield more precise estimations of sediment transport in contrast to the regression equations and standalone models. Particularly, both WIHW-RT and WIHW-RF models provide almost the same performances however, as WIHW-RT can better capture the extreme particle Froude number values, it slightly outperforms WIHW-RF. Promising findings of the current study may encourage the implementation of the recommended approaches in alternative hydrological problems.

1. Introduction

Sewers and urban drainage systems should be designed to overcome several problems related to the sediment deposition (Butler et al., 2003; Safari et al., 2018). Designing rigid boundary channels based on sediment transport principles is an important problem in hydraulics, hydrology, and environmental engineering. Sediment deposition has detrimental environmental effects such as sediment contamination with poisonous materials, decreasing the channel hydraulic capacity which in turn can cause changes in the shear stress and velocity distributions, roughness, and the channel cross-section shape (Ashley et al., 1992; Ackers et al., 1996; Ota and Nalluri, 2003). Therefore, the self-cleansing concept is used in designing the channels and sewer systems to solve the above-mentioned problems. Self-cleaning is known as a condition in which sediments continue to transport without deposition (May et al., 1996). This method is classified into three categories of incipient motion (Wan Mohtar et al., 2021), incipient deposition (Safari, 2020) and non-deposition with clean bed and with deposited bed (Ab Ghani, 1993; Kargar et al., 2019; Montes et al., 2020; Nalluri and Ghani, 1996; Safari and Aksoy, 2021; May, 1993). In the incipient motion, the sediment particles initiate to transport in the channel bed. The incipient deposition is explained as a state in which sediments (suspended in the flow) tend to deposit at the channel bed or convey as bed load (Loveless, 1992; Safari, 2020). At the non-deposition with clean bed state, the bottom of the channel remains clean from sediment deposition. This method is a conservative approach in designing of the small sewers (Ab Ghani, 1993; Montes et al., 2020). The non-deposition with deposited bed condition is applied for large pipes design. Although the small thickness of sediment is present at the bottom of the channel to decrease design slope, applying

* Corresponding author.
E-mail addresses: katayoun.kargar@ryerson.ca (K. Kargar), jafar.safari@yasar.edu.tr (M.J.S. Safari), khabat.khosravi@gmail.com, kh.khosravi@um.ac.ir (K. Khosravi).

https://doi.org/10.1016/j.jhydrol.2021.126452
Received 30 January 2021; Received in revised form 21 April 2021; Accepted 11 May 2021
Available online 14 May 2021

Nomenclature

b	bootstrap sample in RF
B	water surface width (m)
C_v	sediment volumetric concentration (-)
d	sediment median size (m)
D	circular pipe diameter (m)
DT	decision tree
D_{gr}	dimensionless grain size (-)
ELM	extreme learning machine
FFNN-ELM	feed-forward neural network-extreme learning machine
f	function symbol
Fr_p	particle Froude number (-)
FCM-ANFIS	fuzzy c-means based adaptive neuro-fuzzy inference system
g	gravitational acceleration (m s^{-2})
GEP	gene expression programming
GS-GMDH	generalized structure group method of data handling
GR	generalized regression neural network
MAE	mean absolute error (-)
MAPE	mean absolute percentage error (-)
MARS	multivariate adaptive regression splines
MGGP	multigene genetic programming
ML	machine learning
n	number of data
NF	neuro-fuzzy
NSE	Nash–Sutcliffe efficiency (-)
P	wetted perimeter (m)
PCC	Pearson correlation coefficient (-)
PBIAS	percentage of bias (-)
q_c	class prediction
R	hydraulic radius (m)
REPT	reduced error pruning tree
RF	random forest
RMSE	root mean square error (-)
ROF	rotation forest
RT	random tree
s	relative density of sediment to fluid (-)
sd	standard deviation
SDR	standard deviation reduction
S_c	stopping criterion
S_i	attributes values set
T_b	RF tree
U_c	leaf-within variance
V	flow mean velocity (m s^{-1})
W	rectangular channel bed width (m)
WIHW	weighted instances handler wrapper
x	input parameter
x_i^o	observed value
$\overline{x_i^o}$	average of observed values
x_i^p	predicted value
Y	flow depth (m)
λ	channel friction factor (-)
ν	kinematic viscosity (m^2 s^{-1})
ρ	water density (kg m^{-3})
ρ_s	sediment density (kg m^{-3})

this approach has no an adverse effect on the channel performance (Ota and Nalluri, 2003; Safari and Shirzad, 2019). Large channels should be designed with higher self-cleansing velocity and accordingly applying non-deposition with clean bed criterion is not an economical design method (Ab Ghani, 1993; May, 1993; Safari and Shirzad, 2019). Owing to the reason that the bed load transport is near the constant deposition situation, most of the studies investigated the bed load transport (Nalluri and Ghani, 1996; Ota and Nalluri, 2003; Safari and Aksoy, 2021). Due to the simple structure and using limited datasets, regression approaches were mostly implemented for the modeling. The main deficiencies of the existing empirical equations are that, they are over-fitted on the entire data, applying conventional best-fit regression approaches (Ebtehaj et al., 2020; Montes et al., 2021). This is the main reason that, aforementioned studies recommended empirical equations that have acceptable performances only on the datasets used for the model development as described by Safari et al. (2018). To this end, utilizing wide range of experimental data and applying robust machine learning (ML) algorithms can overcome deficiencies of the empirical equations.

The data mining and ML models have drawn considerable attention in solving complex problems related to the water and environmental engineering (Zhao et al., 2020; Zounemat-Kermani et al., 2020; Kao et al., 2020; Huang et al., 2021). Recently, these models have extensively been used to overcome difficulties that occur in numerical models and also limitations and high costs of laboratory models. Such algorithms establish a functional relationship between the parameters involved. As examples in application of ML techniques in sediment transport molding, Kargar et al. (2019) used two methods of gene expression programming (GEP) and neuro-fuzzy (NF) and demonstrated that the NF method has acceptable precision in sediment transport prediction. The GEP method was utilized by Roushangar and Ghasempour (2017) for bed load sediment modeling in sewer pipes to show that GEP models have superior performance when compared to the empirical equations. For the same purpose, generalized regression neural network (GR), decision tree (DT), and multivariate adaptive regression splines (MARS) methods were used by Safari (2019). According to the obtained results, MARS and GR models gave better results in contrast to the DT in particle Froude number estimation. Applying a new pareto-optimal method, which was established through multigene genetic programming (MGGP) technique, Safari and Danandeh Mehr (2018) recommended design tools for large sewers. A hybrid algorithm of feed-forward neural network-extreme learning machine (FFNN-ELM) was implemented by Ebtehaj et al. (2016) to show its superiority to the empirical models. It should be noticed that selecting of an algorithm which works well in all conditions is a challenging task in hydrological modeling. Generally a neuron–based algorithms like artificial neural network (ANN), support vector machine (SVM) and adaptive neuro-fuzzy inference system (ANFIS) have some drawbacks such as needing to a large dataset for learning, over-fitting problem, exact determination of weights in membership function, and hyper-parameters. Recently, novel ML algorithms have been explored to address some of the weaknesses of traditional ML methods. For instance, Hussain and Khan (2020) declared random forest (RF) has a higher performance than ANN and SVM for hydrological modeling. Novel types of ML approaches like tree, lazy, and rule based-algorithms can solve aforementioned issues and may outperform those traditional ML models.

In the present study, sediment transport is modeled using four standalone models of M5P, reduced error pruning tree (REPT), random forest (RF) and random tree (RT) and, two techniques of rotation forest (ROF) and weighted instances handler wrapper (WIHW) are utilized for hybridization of standalone models. Conducted studies on this topic have indicated that innovative tactics with a probabilistic basis and considering non-linear relationships in the estimation of the sediment transport could increase the accuracy of the prediction. These approaches have not been evaluated so far; therefore, this study recommends novel approaches for sediment transport modeling.

Organization of the manuscript is as follows: In Section 2, existing

empirical models for self-cleansing channel design are briefly reviewed followed by description of the experimental procedure and data preparation. Explanation of the utilized ML algorithms and model performance criteria are also given in Section 2. In Section 3, different input combinations are assessed and then the best models are compared to their alternatives. Section 4 presents the discussion of the results highlighting the research question, analyzing the results, generalization of the main findings, comparing the obtained results with similar studies in the literature, possible practical application of the developed models, limitation of the current study and future research directions. Finally in Section 5 concluding remarks are explained.

2. Material and methods

For the sake of understanding the effective parameters involved in sediment transport for channel design consideration, firstly existing empirical models are briefly reviewed to explain the structure of the developed models. Secondly, utilized experimental data are described and data preparation procedure are given. Thirdly, basic information of the applied algorithms are explained and fourthly, model performance criteria are described for models performance evaluation.

2.1. Non-deposition with clean bed

Among variety of criteria for self-cleansing channel design, non-deposition with clean bed has been widely used (Montes et al., 2020; Safari and Aksoy, 2021; Vongvisessomjai et al., 2010). Applying this criterion, the required flow velocity or shear stress of the bed load should be considered to retain the sediment particles in motion within the flow in the non-deposition self-cleansing condition. As documented in Safari and Danandeh Mehr (2020), in the past (before 1990s in the UK and until present in some countries) as a conventional design method, a certain quantity of shear stress or velocity was considered (based solely on experience). Minimum velocity was mostly utilized from the range of 0.3 m/s to 1 m/s as a design criterion in USA, UK, Germany, and France. Furthermore, it changes for each country considering the sewer type (Mayerle, 1988). In this approach, some essential factors such as size of sewer and sediment properties are not considered. Moreover, shear stress approach is implemented in some European countries including Norway, Germany, Sweden and USA within the range of 1–12.6 N/m^2 (Vongvisessomjai et al., 2010). This method has some defects just like the minimum velocity method. In order to overcome these deficiencies, more hydraulic parameters are used in the modeling (Safari et al., 2018). Different models are recommended for bed and suspended loads. In this study, the bed load under the non-deposition condition is investigated.

Since 1990s, conventional design approaches mentioned above has been modified to consider more hydraulic parameters (Safari et al., 2018). Regarding the empirical equations reported in the literature (as described below), various effective parameters related to the properties of sediment, fluid, channel and flow play crucial roles in the modeling. For this reason, to develop a model, hydraulic radius (R), flow velocity (V), fluid kinematic viscosity (ν), acceleration due to gravity (g), water density (ρ), sediment density (ρ_S), sediment median size (d), volumetric concentration of sediment (C_v), and friction factor of channel (λ) are considered. Based on the models in the literature, the above-mentioned parameters can be given as (Mayerle et al., 1991)

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

where s is relative sediment density, f is function symbol and D_{gr} is dimensionless grain size which is expressed as follows (Ab Ghani, 1993)

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

The parameters given in right side of the Eq. (1) are selected as independent parameters and the left hand side is considered as a dependent parameter which is the particle Froude number (Fr_p). As examples from the relevant literature, Loveless (1992) performed experiments in the incipient deposition and non-deposition of bed load to examine the models suggested by Ackers and White (1973) and, May (1982). May (1982) reported that pipe diameter, sediment concentration, and size of sediment play an important role in non-deposition flow velocity. May (1993) developed models for bed load in the circular pipe channels and improve the previously reported models. Mayerle et al. (1991) organized tests in circular channels and recommended the following formula

$$\frac{V}{\sqrt{gd(s-1)}} = 14.43 C_v^{0.18} D_{gr}^{-0.14} \left(\frac{d}{R}\right)^{-0.56} \lambda^{0.18}. \tag{3}$$

Vongvisessomjai et al. (2010) performed experiments in circular channels to analyze the non-deposition condition and recommended following relationship

$$\frac{V}{\sqrt{gd(s-1)}} = 4.31 C_v^{0.226} \left(\frac{d}{R}\right)^{-0.616} \tag{4}$$

Ab Ghani (1993) conducted tests to examine the effect of roughness and pipe size. For this purpose, Ab Ghani (1993) used experimental data of Loveless (1992), May et al. (1989) and Mayerle (1988) together with his own data to suggest a bed load self-cleansing model as follows

$$\frac{V}{\sqrt{gd(s-1)}} = 3.08 C_v^{0.21} D_{gr}^{-0.09} \left(\frac{d}{R}\right)^{-0.53} \lambda^{-0.21} \tag{5}$$

Ab Ghani (1993) pointed out that if the channel size, sediment concentration, and roughness increase, the self-cleansing velocity rises. Montes et al. (2020) performed experiments in a larger pipe and recommended the following formula

$$\frac{V}{\sqrt{gd(s-1)}} = 4.79 \lambda^{0.058} C_v^{0.209} \left(\frac{d}{R}\right)^{-0.593} \tag{6}$$

Safari (2016) conducted experiments in different cross-section channels of trapezoidal, U-shape, circular, V-bottom and rectangular. Safari and Aksoy (2021) introduced the shape factor of P/B and recommended

$$\frac{V}{\sqrt{gd(s-1)}} = 4.38 C_v^{0.09} D_{gr}^{-0.14} \left(\frac{d}{R}\right)^{-0.32} \left(\frac{P}{B}\right)^{0.20} \tag{7}$$

where P is wetted perimeter and B water surface width. The self-cleansing models are established by means of regression approach establieshed on experimental data. The accuracy of models can be linked to the data range and implemented techniques for the modeling. Hence, several data sets with an extensive range of sediment and pipe properties are used to develop novel self-cleansing models applying robust ML algorithms in the current study.

2.2. Experimental data

For the purpose of modeling sediment transport under the condition of non-deposition with clean bed, four sets of data including Vongvisessomjai et al. (2010), Ab Ghani (1993), May (1993), and Mayerle (1988) are utilized which are available in Safari et al. (2018). The ranges of the 375 data taken from the aforementioned studies are shown in Table 1. Vongvisessomjai et al. (2010) organized tests using sediment sizes ranging of 0.2–0.43 mm in pipes having diameters of 100 mm and 150 mm. Ab Ghani (1993) performed experiments in 154 mm, 305 mm, and 450 mm-diameter circular channels utilizing seven various sediment sizes between 0.46 and 8.3 mm. In the tests of May (1993), sediment size of 0.73 mm and the circular channel with the diameter of 450 mm were utilized. Mayerle (1988) carried out tests in both rectangular

Table 1
Utilized data range.

	D or W (mm)	d (mm)	Y (mm)	λ (-)	C_v (ppm)	V (m/s)	No.
Mayerle (1988)	$D = 152$	0.50–8.74	28–122	0.016–0.034	20–1275	0.37–1.10	106
	$W = 311.5$–462	0.50–5.22	31–111	0.011–0.025	14–1568	0.41–1.04	105
Ab Ghani (1993)	$D = 154$–450	0.46–8.30	24–342	0.013–0.048	4–1450	0.24–1.21	110
May (1993)	$D = 450$	0.73	222–338	0.014–0.018	2–38	0.50–1.22	27
Vongvisessomjai et al. (2010)	$D = 100$–150	0.20–0.43	30–60	0.034–0.053	4–90	0.24–0.63	27

D: circular channel diameter; W: rectangular channel bed width; d: sediment median size; λ: channel friction factor; Y: flow depth; V: flow mean velocity; C_v: sediment volumetric concentration and No. number of data.

and circular cross-section channels. In the tests of rectangular channels with the 462 mm and 311.5 mm widths, five different sediment sizes of 0.5–5.22 mm and, in 152 mm-diameter circular channel, sediment sizes of 0.5–8.74 mm were used.

2.3. Data preparation and model scenarios

At the first step of the data preparation, utilizing four data sets used in this study, parameters given in Eq. (1) are computed. The left hand side of Eq. (1) as particle Froude number (Fr_p) is considered as a model output and parameters given in the right hand side (C_v, D_{gr}, d/R, λ) are incorporated into the model as inputs. Through the modeling, train and test stages are required. For this purpose, entire data are split in two sections. From entire 375 data, 80% of the data are utilized for training and the rest of data (20%) are considered for testing the model, randomly. Feasible models with respect to the inputs and output parameters are discovered in the training step and then, the accuracy of the established models is assessed on the rest of the dataset in the testing part. Table 2 shows a data range in the train and test parts. If an extensive data range is considered for the train part and a narrow range for the test part, the model cannot be an accurate and reliable tool. As can be seen in Table 2 ranges of data for each part (train and test) are selected almost equally for acquiring more reliable results.

After determination of the input variables based on the literature review, the next step is to identify the best input scenario as irrelevant or less effective input variables may led to lower prediction accuracy of the models. To meet this goal, first Pearson correlation coefficient (*PCC*) between input and output variables is applied. Next, several input scenarios are constructed based on *PCC*. A variable having the highest correlation with Fr_p is considered as the first input scenario. Thereafter, variables with the second highest *PCC* are added to the first input and scenario No. 2 is built. This approach is continued until all input variables involve in building the input scenarios. It is found that D_{gr} with a higher correlation coefficient of −0.678 is the most effective parameter among others for Fr_p prediction followed by d/R, C_v and λ having correlation coefficient of −0.649, −0.245 and −0.026, respectively. As there are four input variables, thus, according to this approach, four input scenario are constructed as shown in Table 3.

Table 2
Range of data in each part of training and testing.

Phase		C_v	D_{gr}	d/R	λ	Fr_p
Training	Max	0.002	215.591	0.380	0.053	13.529
	Min	0.000	5.059	0.005	0.011	1.311
	Mean	0.000	67.533	0.068	0.021	4.402
	St.D	0.000	59.961	0.073	0.008	2.321
	Skewnes	1.697	1.167	1.745	1.691	1.041
	Kurtosis	2.177	0.384	3.100	3.473	0.948
Testing	Max	0.001	215.591	0.416	0.047	10.483
	Min	0.000	5.059	0.006	0.011	1.298
	Mean	0.000	56.325	0.063	0.024	4.408
	St.D	0.000	57.638	0.084	0.011	2.035
	Skewnes	1.868	1.437	2.603	0.895	0.602
	Kurtosis	3.399	1.301	7.530	−0.510	−0.215

Table 3
Input and output parameters.

No.	Inputs	Output
1	D_{gr}	Fr_p
2	D_{gr}, d/R,	
3	D_{gr}, d/R, C_v	
4	D_{gr}, d/R, C_v, λ	

2.4. Models theory background

Machine learning algorithms are implemented in this study using Waikato Environment for Knowledge Analysis (WEKA 3.9) software (Hall et al., 2009).

2.4.1. M5Prime (M5P)

The M5P algorithm is a linear tree-based model that was first introduced by Quinlan (1992). This technique can be used in various fields of engineering (Heddam and Kisi, 2018; Khosravi et al., 2018). The M5P model is an adaptable method in which the decision tree that constructs the M5P can possess multivariate linear models. This technique has some advantages, for example, it can handle missing data without vagueness and, handling a large amount of data with large numbers of properties and dimensions (Zhan et al., 2011). Through several steps of building and smoothing the tree, the M5P tree model is developed. Constructing the tree is performed by dividing the input data into various subcategories. In the M5P method, in the growing procedure of the tree, the standard deviation reduction (*SDR*) is utilized to minimize the errors and reaching the best result. The *SDR* factor is defined as follows (Heddam and Kisi, 2018)

$$SDR = sd(S) - \sum_i \frac{|S_i|}{|S|} sd(S_i) \quad (8)$$

in which S is a set of examples, S_i is the subsets of examples, number of the examples is i to n, and $sd(S)$ is the standard deviation. The second step (tree pruning) is initiated to remove unnecessary sub-trees. The aim of this step is refraining data from the over-fitting matter that happen in the first step of building a tree. Moreover, in order to decrease the error, the attributes are reduced, separately. Compensating the discontinuities among adjoining linear models at the leaves of the pruned tree is accomplished in the third step of smoothing (Wang and Witten, 1997). Furthermore, the final model can be generated through combining leaf to root. The estimated leaf quantity throughout the smoothing step is filtered back to the root as its path.

2.4.2. Random forest (RF)

The random forest suggested by Breiman (2001) is a method for analyzing the regression and classifying problems. This technique is established on the composition of several decision trees and it is able to solve various problems in the field of water resources (Shiri, 2018; Yu et al., 2017; Zhao et al., 2018; Sadler et al., 2018). The weak performance of the traditional regression tree is due to the over-fitting on the train datasets. However, this deficiency is solved using the RF model which uses a random feature. Random features utilized a range of

variables in the structure of the tree in the process of growing. Then, for creating a powerful prediction according to the determined data set, produced trees are intermingled. The average value of estimated single outputs is considered as an output result. The difference of random forest with conventional regression tree is that, in the process of creating trees, the RF uses bootstrap samples as a substituted of whole training datasets. Dividing data in the decision tree method is done with the best splitter variable but, in the RF method, this process is accomplished by selecting the predictors randomly. Using the data without rescaling them is one of the features of the RF method. Moreover, several trees are used in the structure of the RF method and it indicates the accurate prediction of this technique. The RF model is formed over a random vector that produces the trees. Numerical values are the outcomes of tree predictors (Breiman, 2001).

By considering the training data as S with variables of P and N as records, it is planned to get the computation f for input x. Using two methods of bagging averages or bootstrap aggregation, the variance is decreased in prediction. The model is established for every bootstrap sample of $b = 1, 2, 3, \ldots, B$. In the first step, through selection of m variables from the set of P variables in random, the RF model is produced. Then, the best variables are selected from m variables, and in the last step, by splitting the node, two daughter nodes are created (Hastie et al., 2009).

Repeating this procedure is continued to minimize the node size n_{min} at every final node in order to develop a random forest tree (T_b). The estimation at the point of x is demonstrated as follows (Hastie et al., 2009)

$$f_{rf}^{B}(x) = \frac{1}{B}\sum_{b=1}^{B} T_b(x) \tag{9}$$

2.4.3. Random tree (RT)

RT method is similar to the classification and regression tree (CART); however, it has some differences in its operation where for choosing the subsets, it utilizes a ratio parameter. For creating the classification model, RT method predicts the quantities of the label based on the input attributes. The RT method's fast and flexible nature allows it to be used in various fields to solve a wide range of issues. The performance of the RT model is influenced by different parameters including randomly selected attribute number known as K value, defining the best instances process number known as size of batch, minimizing leaf total weight of the instance and minimum variance probability to demonstrate the minimum rate of variance on the entire data set.

2.4.4. Reduced error pruning tree (REPT)

The REPT method is considered as a kind of decision tree technique with more speed that decreases the error in the process of prediction by creating a decision tree (Mohamed et al., 2012). Initially, regression tree logic is used for building numerous trees in different iterations. Thereafter, the best tree giving the lesser error is chosen among numerous trees. After these two steps, in order to inhibit the over-fitting problems, the reduced error pruning (REP) method is utilized. Furthermore, this method is able to preserve its accuracy by minimizing the tree size. The REPT method utilizes a stopping criterion as the sum of squared errors, to construct a tree model having highest information. The stopping criterion is defined as (Quinlan, 1987):

$$S_c = \sum_{c \in leaves(RT)} q_c U_c \tag{10}$$

in which c is an element of leaves, U_c is the leaf-within variance and q_c is expressed as the class prediction.

2.4.5. Rotation forest (ROF)

ROF as an ensemble algorithm is utilized to construct sets of classifiers by the usage of independently built decision trees. This method has aimed to construct precise and various classifiers based on the concept of the random forest technique (Rodriguez et al., 2006). In this method, a training dataset for each classifier is selected from bootstrap samples (Lombardo et al., 2015). By taking out various sets of tree features, decision trees are trained autonomously (Chen et al., 2017). Furthermore, in order to provide the train sets for learning the base classifiers, the principal component analysis is implemented to every subset (Wold et al., 1987; Nguyen et al., 2017). Main advantages of this model are applying the random feature selection and data transmission techniques to improve the diversity of the decision tree and subsequently enhancing the achieved result. More information about this model are well-documented in Nguyen et al. (2017), Hong et al. (2018), and Pham et al. (2020).

2.4.6. Weighted instances handler wrapper (WIHW)

The main advantage of the WIHW model is that, wrapper approach is used for weighting training instances (Karagiannopoulos et al., 2007). If the base classifier is not implementing the weka.core, WIHW algorithm uses resampling with weights technique. By default, it can control instance weights and the train dataset is crossed across to the base classifier. Meta-learning algorithms or base algorithms applied for the modeling use WIHW as a classifier and turn them into more robust learners. A certain parameter must be specified in the base classifier and determined the number of iterations for iterative schemes. WIHW is implemented as a classifier on the data to adjust a parameter. The unique parameters are based on the training data, which is a suitable method to be implemented on the test data. The WIHW is a generic wrapper around any classifier to enable weighted instances support (Joshuva and Sugumaran, 2019). Once the base classifier is not applying the interface and there are instance weights, WIHW implements resampling with weights. As a default option, the base classifier uses training data if could operate weights, although it may also apply resampling approaches together with weights.

2.5. Performance evaluation

In order to evaluate the models accuracy, the Nash–Sutcliffe efficiency (*NSE*), percentage of bias (*PBIAS*), root mean square error (*RMSE*), mean absolute error (*MAE*) and mean absolute percentage error (*MAPE*) are considered as given below (Nash and Sutcliffe, 1970; Yapo et al., 1996; Heddam and Kisi, 2018)

$$NSE = 1 - \frac{\sum_{i=1}^{n}\left(x_i^p - x_i^o\right)^2}{\sum_{i=1}^{n}\left(x_i^o - \overline{x_i^o}\right)^2} \tag{11}$$

$$PBIAS = \frac{\sum_{i=1}^{n}\left(x_i^o - x_i^p\right)}{\sum_{i=1}^{n} x_i^p} \times 100 \tag{12}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}\left(x_i^o - x_i^p\right)^2}{n}} \tag{13}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|x_i^o - x_i^p\right| \tag{14}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{x_i^o - x_i^p}{x_i^o}\right| \times 100 \tag{15}$$

in which x_i^o and x_i^p are the observed and predicted values, respectively, and n is the data number. The lesser *RMSE*, *MAE* and *MAPE*, and higher *NSE* values present the better results. The best result of *PBIAS* is the value closer to zero.

3. Results

3.1. Examination of studied scenarios

Among all variables for estimating the particle Froude number (Fr_p), the importance of parameters is investigated using the *PCC*. In this study, four scenarios are considered in order to estimate the output parameter (Fr_p). Owing to the results given above, the D_{gr} and λ are the most and less important independent parameters among others. In Fig. 1, the *RMSE* index is used to find the best model among all of the studied scenarios. As can be seen in Fig. 1, M5P (4), RF (4), RT (3), and REPT (3) with *RMSE* of 0.92, 0.74, 0.96, and 1.11 provide better results for different scenarios, respectively. Generally input scenario of 3 (D_{gr}, d/R, C_v,) and 4 (D_{gr}, d/R, C_v, λ) are found more effective than other input scenarios. This difference can be resulting from differences among model structure.

3.2. Compression of models

Efficiency of four standalone and eight hybrid ML models are examined in contrast to three empirical equations of Vongvisessomjai et al. (2010), Ab Ghani (1993) and Montes et al. (2020) based on different statistical error measurement indices. Comparison of the models is performed using box and whiskers plots as shown in Fig. 2, where whiskers are considered as stretched lines below and above the boxes. The lower part of the box consists of 25th percentile of data and the upper section of the box contains 75th percentile of the data. From Fig. 2, it is obvious that the box shapes and whiskers of both WIHW-RT and WIHW-RF models are analogous to each other and match with measured Fr_p. The obtained Fr_p values from those two methods are not perceptibly different except in predicting the maximum values of Fr_p where WIHW-RT model gives better results than the WIHW-RF model. It is shown in Fig. 2, almost all of the empirical models overestimate the Fr_p, and they provide similar performances. The maximum and minimum Fr_p values are diverse due to computing with different models. For example, the MIHW-M5P model calculates the maximum value of Fr_p higher than other models. When the measured middle line is considered, both WIHW-RT and WIHW-RF model's median lines are close to the measured one. Yet, in all models excluding three of them (M5P, ROF-RT, and WIHW-M5P), the calculated values are higher than the median measured line which reveals that these models have an overestimation. On the other hand, M5P, ROF-RT, and WIHW-M5P models indicate an underestimation. The alignment of the midline in the WIHW-RT model with the measured midline level indicates the accuracy of the WIHW-RT model.

Fig. 3 depicts the comparison between the measured and estimated Fr_p. In scatter plots of both WIHW-RT and WIHW-RF models, data are near to the bisector line that represents the precision of two models in estimating Fr_p. Furthermore, empirical equations of Vongvisessomjai et al. (2010), Ab Ghani (1993) and Montes et al. (2020) generate significant overestimation where most of the data are remained above the best fit line. Fig. 3 demonstrates that the hybrid algorithms provide better results in comparison to their standalone counterparts. For instance, the RT model does not provide acceptable results. But, when it combined with the WIHW and ROF algorithms, the estimation power of the RT model is improved, significantly. Moreover, the WIHW improves the performance of RT slightly more than the ROF technique. In order to identify the best method in estimating of Fr_p, demonstrating the quantitative information (Table 4) are essential. As shown in Table 4, although most of the models give acceptable results, the hybrid WIHW-RF and WIHW-RT models are able to create better outcomes and predict the Fr_p with high accuracy. Among recommended models, the RT

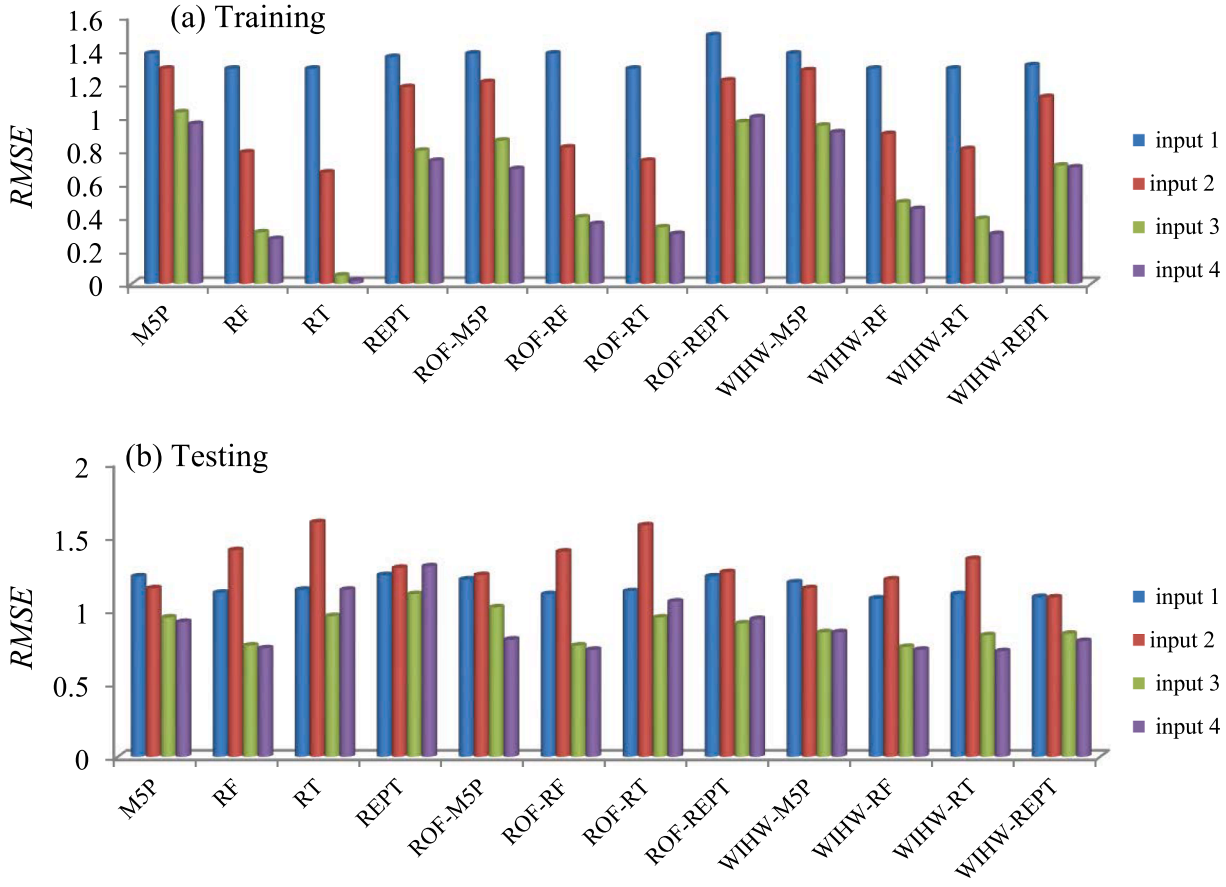

Fig. 1. Selection of the most effective input scenario.

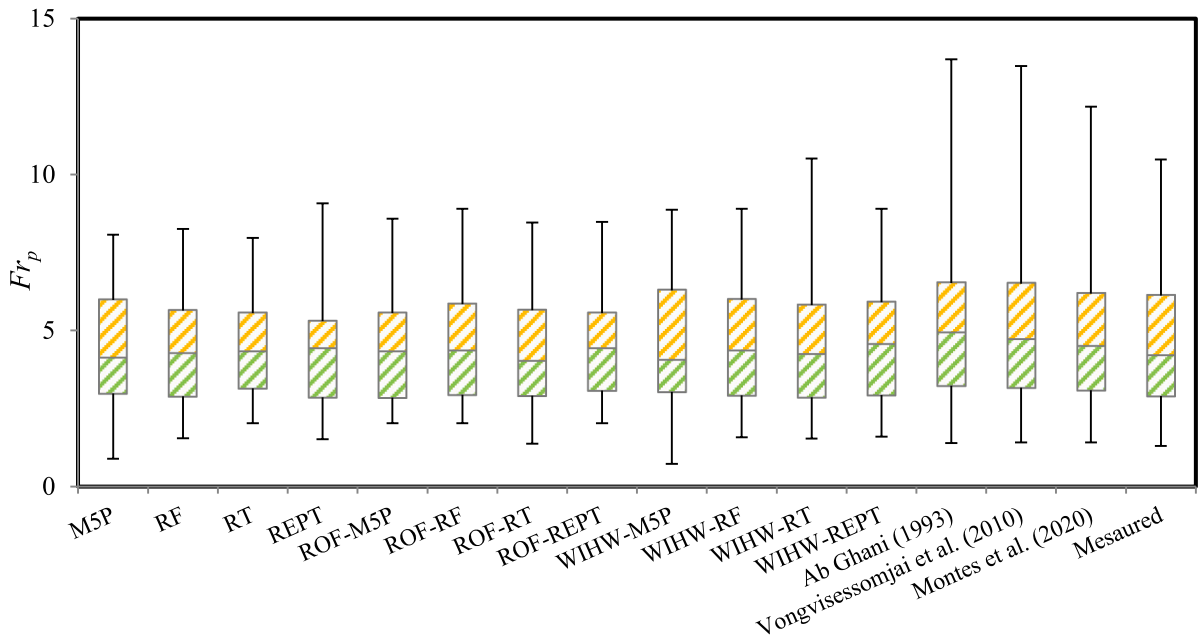

Fig. 2. Box plots of applied algorithms at testing stage.

standalone model had lowest precision with *RMSE* of 1.16, *MAE* of 0.54, *MAPE* of 19.52 and *NSE* of 0.67. Likewise, all empirical equations have low estimation power. The *PBIAS* metric shown in Table 4 indicates that the most of the models excluding WIHW-RT and WIHW-RF have noticeable overestimation. It is evident that the empirical equations give results with low precision and significant overestimation.

4. Discussion

Sediment transport modeling is a complex and important water resource engineering problem. In the traditional self-cleansing methods, the minimum velocity and shear stress, which have some imperfections, were used. In the traditional methods, essential factors including sediment size, flow depth, pipe size, slope of the channel and sediment volumetric concentration are missed. However, in order to produce a precise model, the combination of the best input parameters with various range of data should be determined. In this study, considering the effective parameters, the best input combination is selected.

To create the best model with high accuracy, most of the parameters with a wide range of data are considered and models are established using robust ML models including M5P, reduced error pruning tree (REPT), random forest (RF) and random tree (RT). Furthermore, two techniques of rotation forest (ROF) and weighted instances handler wrapper (WIHW) are implemented as optimizers to construct robust models. Several scenarios are utilized in order to estimate the Fr_p. Standalone models (M5P, RF, RT, and REPT) provide different outcomes due to their structure, intricacy, flexibility, ability in calculation, and preventing over-fitting problem. The hybrid models which are built using ROF and WIHW have high accuracy in Fr_p estimation because of their flexible and non-linear structure. Likewise, the accuracy of hybrid methods is completely related to the base model implemented.

Among four standalone models, RF is the best performing model and has some advantages like flexibility, simplicity, low bias, deal with unbalanced data and easy application. Overall, utilizing WIHW boosts the accuracy of the models. Based on the results, the hybrid WIHW-RF and WIHW-RT models yield the best performances. For further comparison of the models, Kargar et al. (2019) used two techniques of NF and GEP with the *RMSE* of 1.04 and 1.25, respectively. Ab Ghani and Azamathulla (2011) utilized the GEP method which is resulted in *RMSE* of 1.73. Although in both studies mentioned above, models had acceptable outcomes, the results of the methods used in this study are more accurate than the models reported in the literature. Moreover, Safari et al. (2019) implemented different methods for estimating the particle Froude number such as gene expression programming (GEP), extreme learning machine (ELM), generalized structure group method of data handling (GS-GMDH) and fuzzy c-means based adaptive neuro-fuzzy inference system (FCM-ANFIS) with the mean absolute percentage error *(MAPE)* values of 16.77, 16.40, 14.88, and 16.03, respectively. In this study two models of WIHW-RF and WIHW-RT provide lower *MAPE* values (12.69 and 12.25) demonstrating almost 20% improvement in their performances in contrast to the Safari et al. (2019) results.

Credibility of a channel design sediment transport model is significantly attributed to the range of data used and applied technique for the model development. Most of the available models in the relevant literature applied typical approaches for developing a design model. Most importantly, narrow data range and low number of data were utilized for the modeling. These are the main reasons that they fail to provide accurate results for different data sources. It seems that this study appropriately addressed aforementioned deficiencies of existing models utilizing wide data range and implementing robust ML modeling techniques.

As powerful tools for approximation of complex non-linear problems, ML techniques can be used without needing for deep understanding the physics of the problem, due to ML learns data behavior for construction of an intelligence model. These models are established based on the data to find possible relationship between dependent and independent parameters. However, once model parameters are selected based on the physics of the problems, the developed models give more reliable and accurate results. In this study, model parameters are chosen based to effective variables in sediment transport process to incorporate sediment, channel, fluid and flow parameters. It has to be mentioned that models having one-two input parameters failed to generate accurate results. It is found that sediment volumetric concentration and channel friction factor have poor linear correlation with particle Froude number, however incorporating these parameters into the models improved the model performance, significantly. It can be linked to the non-linear behavior of the sediment transport problem.

The importance of accurate computation of a design sediment transport model comes from a fact that, less accurate model may over- or underestimate design velocity. Most of the empirical equations

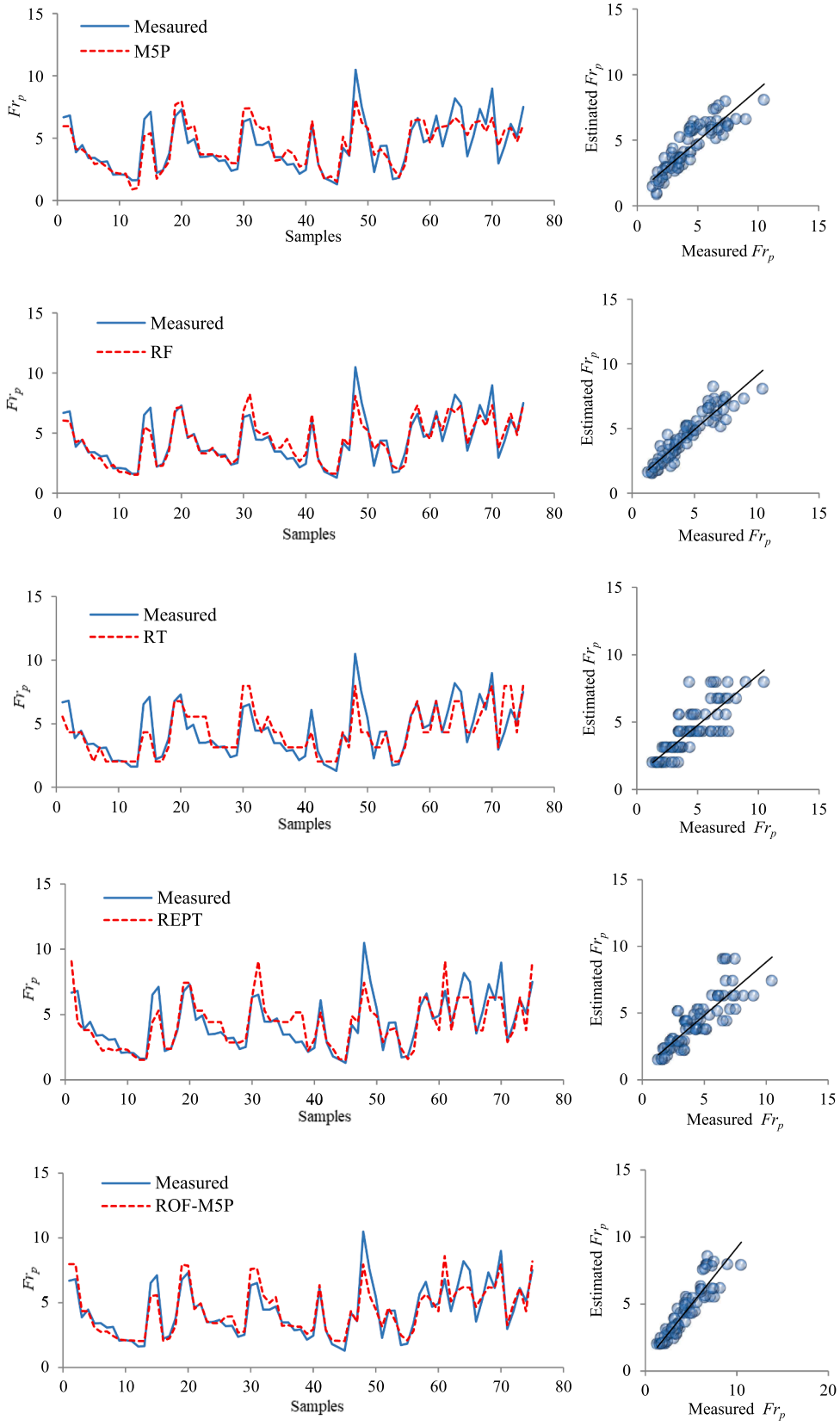

Fig. 3. Comparison of observed and estimated Fr_p.

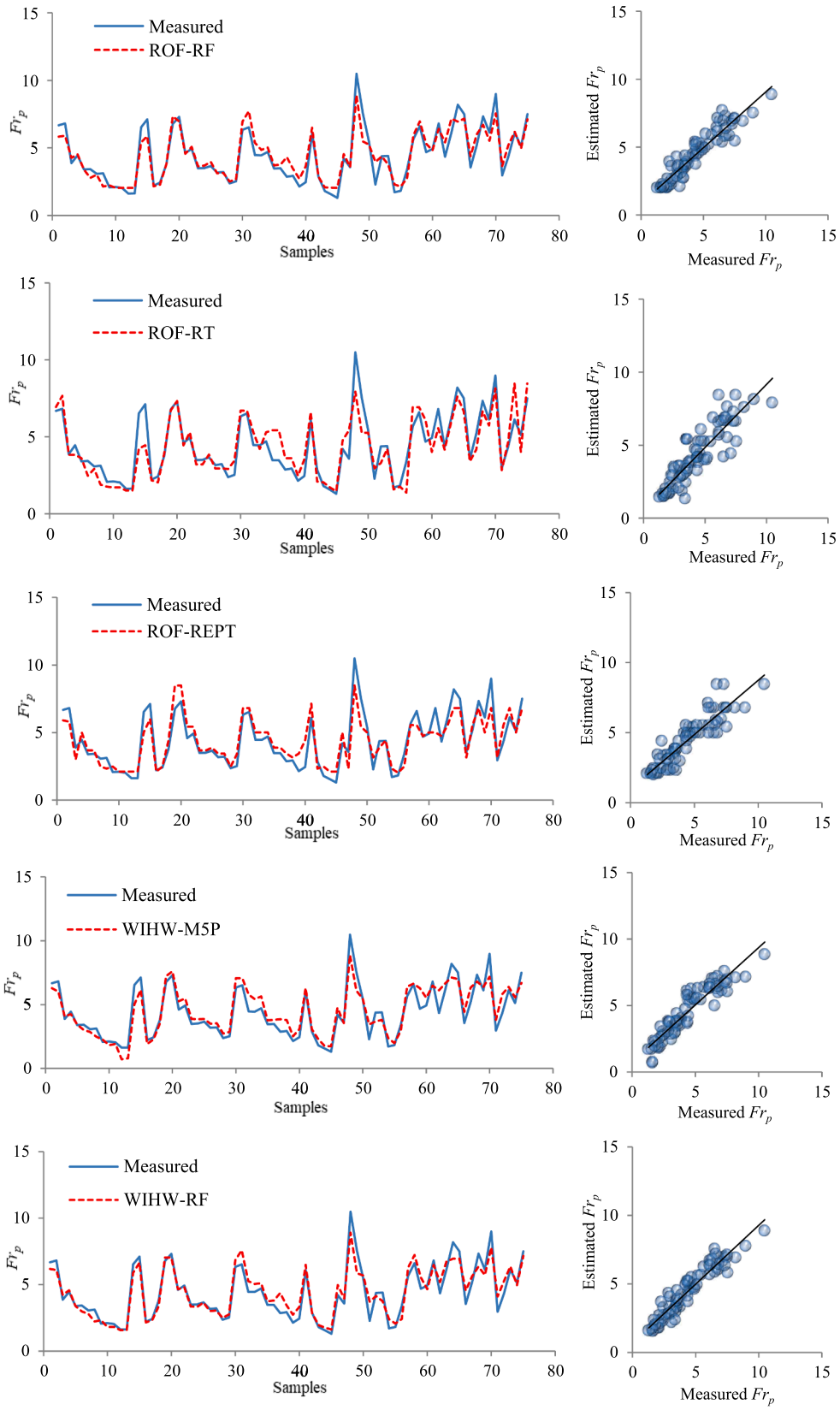

Fig. 3. (*continued*).

overestimate Fr_p and accordingly design flow velocity. Design of a channel through an overestimated model gives higher design velocity and accordingly steeper bed slope. Therefore, it cannot be an economic design criterion. On the other hand, if an underestimated model is applied for the same purpose, design velocity will be lower and then, sediment deposition and early overflow will take place. This study works out this problem through developing accurate sediment transport models. Consequently results are satisfactorily to recommend models as

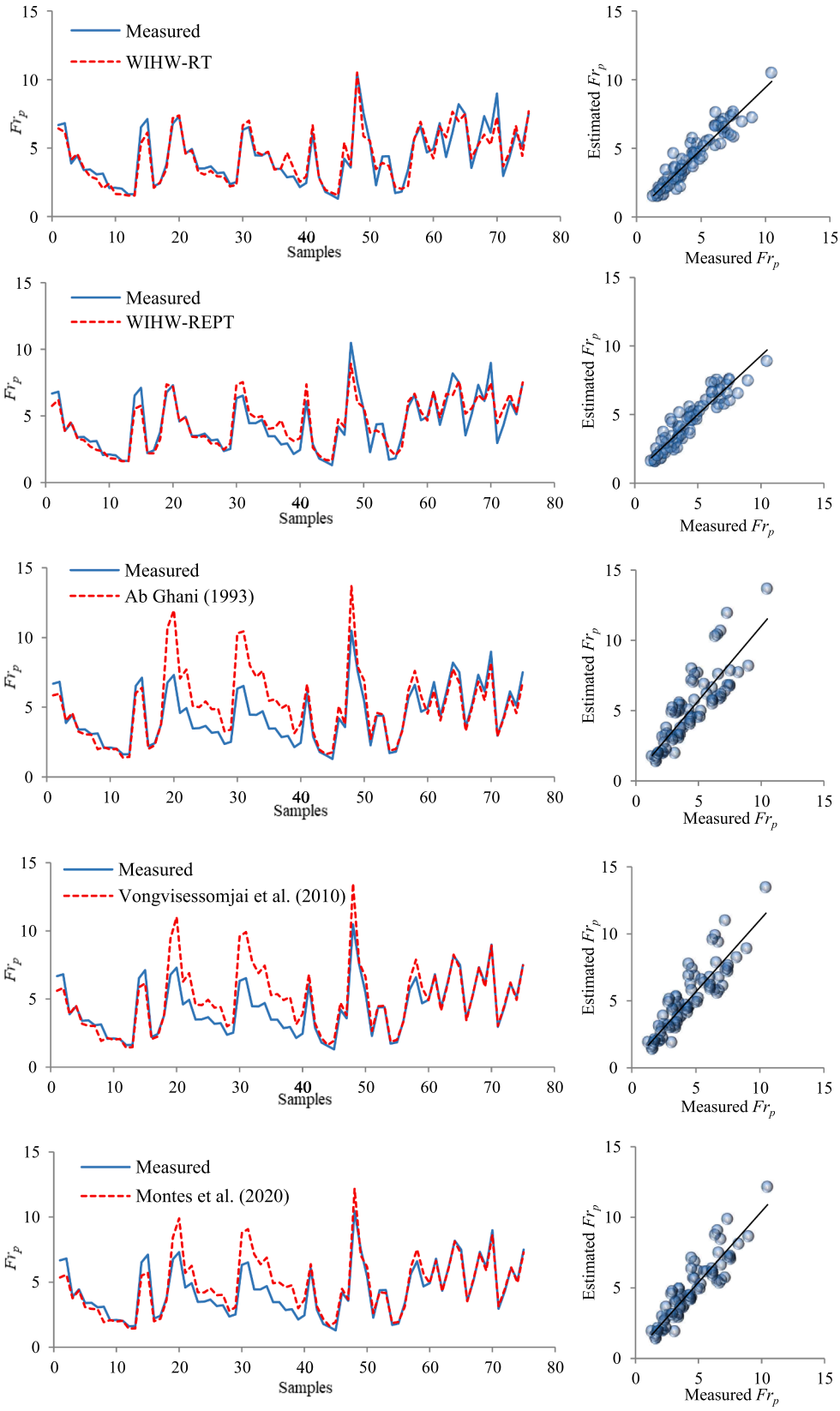

Fig. 3. (*continued*).

practical and reliable tools for channel design.

This study accomplished utilizing the laboratory experimental data and using non-cohesive sediment in smooth channels, to this end, extension of the present study could be the usage of field data, in rough channels and incorporating cohesive sediment characteristics into the models. Additionally, in future studies, alternative ML algorithms with various optimization techniques can be implemented for the sediment transport modeling.

Table 4
Comparison of models based on *NSE*, *PBIAS*, *RMSE*, *MAE* and *MAPE*.

Models	*NSE*	*PBIAS*	*RMSE*	*MAE*	*MAPE*
M5P	0.79	−1.49	0.92	0.73	17.46
RF	0.86	−0.01	0.74	0.54	12.86
RT	0.67	2.05	1.16	0.84	19.52
REPT	0.69	1.26	1.12	0.82	18.31
ROF-M5P	0.84	−0.56	0.80	0.61	14.25
ROF-RF	0.88	−1.52	0.70	0.52	13.42
ROF-RT	0.78	1.76	0.96	0.69	16.25
ROF-REPT	0.82	−0.05	0.85	0.67	16.89
WIHW-M5P	0.86	−2.85	0.75	0.60	15.57
WIHW-RF	0.90	−1.28	0.64	0.50	12.69
WIHW-RT	0.89	0.85	0.66	0.48	12.25
WIHW-REPT	0.87	−2.06	0.73	0.55	14.01
Ab Ghani (1993)	0.44	−15.4	1.51	1.01	23.48
Vongvisessomjai et al. (2010)	0.60	−14.11	1.27	0.83	19.97
Montes et al. (2020)	0.76	−8.75	0.99	0.68	16.73

5. Conclusions

Applying variety of ML algorithms both standalone and hybrid models are applied for particle Froude number computation for the case of non-deposition with clean bed criterion. Determining the best combination of input variables and optimal values of the operator are the most important stages in preparing a precise model. Through considering effective variables involved as inputs, four scenarios were composed. According to the results, models having four input parameters give better performance which is reasonable based on the hydrological consideration of the problem where flow, channel, fluid and sediment properties are embedded into the models. Hybrid models provided better results when compared with the standalone counterparts (M5P, RF, RT, and REPT). Obtained results demonstrated the precise performances of WIHW-RT and WIHW-RF in contrast to their alternatives. However, the WIHW-RT model slightly outperforms WIHW-RF. Furthermore, ML models give better results that than empirical regression equations reported in the relevant literature. Regression models overestimate the particle Froude number which can be linked to the fact that, they were over-fitted on the limited data ranges. As a result, for the purpose of developing a powerful model for the estimation of sediment transport, novel and robust techniques accompanied by an extensive data range are required. Accordingly, the models developed in this study can be used as applicable and reliable tools in practice.

CRediT authorship contribution statement

Katayoun Kargar: Investigation, Supervision, Validation, Visualization, Writing - original draft. **Mir Jafar Sadegh Safari:** Conceptualization, Data curation, Formal analysis, Methodology, Resources, Validation, Visualization, Writing - review & editing. **Khabat Khosravi:** Conceptualization, Formal analysis, Methodology, Software.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors would like to express sincerest appreciation to Editor-in-Chief, Associate Editor and two anonymous reviewers for their highly insightful comments that improved the quality of this manuscript.

References

Ab Ghani, A., 1993. Sediment Transport in Sewers (Ph. D Thesis). University of Newcastle Upon Tyne, UK.
Ab Ghani, A., Azamathulla, H.M., 2011. Gene-expression programming for sediment transport in sewer pipe systems. J. Pipeline Syst. Eng. Pract. 2 (3), 102–106.
Ackers, P., White, W.R., 1973. Sediment transport: new approach and analysis. J. Hydraulics Division 99 (hy11).
Ackers, J.C., Butler, D., May, R.W.P., 1996. Design of sewers to control sediment problems. Construction Industry Research and Information Association, London, pp. 1–181.
Ashley, R.M., Wotherspoon, D.J.J., Coghlan, B.P., McGregor, I., 1992. The erosion and movement of sediments and associated pollutants in combined sewers. Water Sci. Technol. 25 (8), 101–114.
Breiman, L., 2001. Random forests. Mach. Learn. 45 (1), 5–32.
Butler, D., May, R., Ackers, J., 2003. Self-Cleansing Sewer Design Based on Sediment Transport Principles. J. Hydraul. Eng. 129 (4), 276–282.
Chen, W., Shirzadi, A., Shahabi, H., Ahmad, B.B., Zhang, S., Hong, H., Zhang, N., 2017. A novel hybrid artificial intelligence approach based on the rotation forest ensemble and naïve Bayes tree classifiers for a landslide susceptibility assessment in Langao County, China. Geomatics, Natural Hazards and Risk 8 (2), 1955–1977.
Ebtehaj, I., Bonakdari, H., Safari, M.J.S., Gharabaghi, B., Zaji, A.H., Madavar, H.R., Khozani, Z.S., Es-haghi, M.S., Shishegaran, A.D., Mehr, A., 2020. Combination of sensitivity and uncertainty analyses for sediment transport modeling in sewer pipes. Int. J. Sedim. Res. 35 (2), 157–170.
Ebtehaj, I., Bonakdari, H., Shamshirband, S., 2016. Extreme learning machine assessment for estimating sediment transport in open channels. Engineering with Computers 32 (4), 691–704.
Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.H., 2009. The WEKA Data Mining Software: An Update. SIGKDD Explorations 11 (1).
Hastie, T., Tibshirani, R., Friedman, J., 2009. The elements of statistical learning: data mining, inference, and prediction. Springer Science & Business Media.
Heddam, S., Kisi, O., 2018. Modelling daily dissolved oxygen concentration using least square support vector machine, multivariate adaptive regression splines and M5 model tree. J. Hydrol. 559, 499–509.
Hong, H., Liu, J., Bui, D.T., Pradhan, B., Acharya, T.D., Pham, B.T., Ahmad, B.B., 2018. Landslide susceptibility mapping using J48 Decision Tree with AdaBoost, Bagging and Rotation Forest ensembles in the Guangchang area (China). Catena 163, 399–413.
Huang, C.C., Chang, M.J., Lin, G.F., Wu, M.C., Wang, P.H., 2021. Real-time forecasting of suspended sediment concentrations reservoirs by the optimal integration of multiple machine learning techniques. J. Hydrol.: Reg. Stud. 34, 100804.
Hussain, D., Khan, A.A., 2020. Machine learning techniques for monthly river flow forecasting of Hunza River, Pakistan. Earth Sci. Inf. 13, 939–949.
Joshuva, A., Sugumaran, V., 2019. Selection of a meta classifier-data model for classifying wind turbine blade fault conditions using histogram features and vibration signals: a data-mining study. Progress in Industrial Ecology, an International Journal 13 (3), 232–251.
Karagiannopoulos, M., Anyfantis, D., Kotsiantis, S., Pintelas, P., 2007. In: September). A wrapper for reweighting training instances for handling imbalanced data sets. Springer, Boston, MA, pp. 29–36.
Kargar, K., Safari, M.J.S., Mohammadi, M., Samadianfard, S., 2019. Sediment transport modeling in open channels using neuro-fuzzy and gene expression programming techniques. Water Sci. Technol. 79 (12), 2318–2327.
Kao, I.F., Zhou, Y., Chang, L.C., Chang, F.J., 2020. Exploring a Long Short-Term Memory based Encoder-Decoder framework for multi-step-ahead flood forecasting. J. Hydrol. 583, 124631.
Khosravi, K., Mao, L., Kisi, O., Yaseen, Z.M., Shahid, S., 2018. Quantifying hourly suspended sediment load using data mining models: case study of a glacierized Andean catchment in Chile. J. Hydrol. 567, 165–179.
Lombardo, L., Cama, M., Conoscenti, C., Märker, M., Rotigliano, E.J.N.H., 2015. Binary logistic regression versus stochastic gradient boosted decision trees in assessing landslide susceptibility for multiple-occurring landslide events: application to the 2009 storm event in Messina (Sicily, southern Italy). Nat. Hazards 79 (3), 1621–1648.
Loveless, J.H., 1992. Sediment transport in rigid boundary channels with particular reference to the condition of incipient deposition (Doctoral dissertation. King's College London (University of London)).
May, R.W.P., 1982. Sediment transport in sewers. Hydraulic Research Station Wallingford.
May, R., Brown, P., Hare, G., & Jones, K. (1989). Self-cleansing conditions for sewers carrying sediment.
May, R. (1993). Sediment transport in pipes, sewers and deposited beds.
May, R.W., Ackers, J.C., Butler, D., John, S., 1996. Development of design methodology for self-cleansing sewers. Water Sci. Technol. 33 (9), 195–205.
Mayerle, R., 1988. Sediment Transport in Rigid Boundary Channels (Ph. D. Thesis). University of Newcastle upon Tyne, UK.
Mayerle, R., Nalluri, C., Novak, P., 1991. Sediment transport in rigid bed conveyances. J. Hydraul. Res. 29 (4), 475–495.
Mohamed, W. N. H. W., Salleh, M. N. M., & Omar, A. H. (2012, November). A comparative study of reduced error pruning method in decision tree algorithms. In 2012 IEEE International conference on control system, computing and engineering (pp. 392-397). IEEE.
Montes, C., Kapelan, Z., Saldarriaga, J., 2021. Predicting non-deposition sediment transport in sewer pipes using Random forest. Water Res. 189, 116639.

Montes, C., Vanegas, S., Kapelan, Z., Berardi, L., Saldarriaga, J., 2020. Non-deposition self-cleansing models for large sewer pipes. Water Sci. Technol. 81 (3), 606–621.
Nalluri, C., Ghani, A.A., 1996. Design options for self-cleansing storm sewers. Water Sci. Technol. 33 (9), 215–220.
Nash, J.E., Sutcliffe, J.V., 1970. River flow forecasting through conceptual models part I—A discussion of principles. J. Hydrol. 10 (3), 282–290.
Nguyen, Q.K., Tien Bui, D., Hoang, N.D., Trinh, P.T., Nguyen, V.H., Yilmaz, I., 2017. A novel hybrid approach based on instance based learning classifier and rotation forest ensemble for spatial prediction of rainfall-induced shallow landslides using GIS. Sustainability 9 (5), 813.
Ota, J.J., Nalluri, C., 2003. Urban storm sewer design: Approach in consideration of sediments. J. Hydraul. Eng. 129 (4), 291–297.
Pham, B.T., Prakash, I., Dou, J., Singh, S.K., Trinh, P.T., Tran, H.T., Bui, D.T., 2020. A novel hybrid approach of landslide susceptibility modelling using rotation forest ensemble and different base classifiers. Geocarto International 35 (12), 1267–1292.
Quinlan, J.R., 1987. Simplifying decision trees. Int. J. Man Mach. Stud. 27 (3), 221–234.
Quinlan, J.R., 1992. Learning With Continuous Classes. 5th Australian Joint Conference on Artificial Intelligence.
Rodriguez, J.J., Kuncheva, L.I., Alonso, C.J., 2006. Rotation forest: A new classifier ensemble method. IEEE Trans. Pattern Anal. Mach. Intell. 28 (10), 1619–1630.
Roushangar, K., Ghasempour, R., 2017. Prediction of non-cohesive sediment transport in circular channels in deposition and limit of deposition states using SVM. Water Sci. Technol. Water Supply 17 (2), 537–551.
Sadler, J.M., Goodall, J.L., Morsy, M.M., Spencer, K., 2018. Modeling urban coastal flood severity from crowd-sourced flood reports using Poisson regression and Random Forest. J. Hydrol. 559, 43–55.
Safari, M.J.S., 2016. Self-cleansing drainage system design by incipient motion and incipient deposition-based models (Doctoral dissertation. PhD Thesis. Istanbul Technical University, Turkey).
Safari, M.J.S., 2019. Decision tree (DT), generalized regression neural network (GR) and multivariate adaptive regression splines (MARS) models for sediment transport in sewer pipes. Water Sci. Technol. 79 (6), 1113–1122.
Safari, M.J.S., 2020. Hybridization of multivariate adaptive regression splines and random forest models with an empirical equation for sediment deposition prediction in open channel flow. J. Hydrol. 590, 125392.
Safari, M.J.S., Aksoy, H., 2021. Experimental analysis for self-cleansing open channel design. J. Hydraul. Res. 59 (3), 500–511. https://doi.org/10.1080/00221686.2020.1780501.
Safari, M.J.S., Danandeh Mehr, A., 2018. Multigene genetic programming for sediment transport modeling in sewers for conditions of non-deposition with a bed deposit. Int. J. Sedim. Res. 33 (3), 262–270.
Safari, M.J.S., Danandeh Mehr, A., 2020. Design of smart urban drainage systems using evolutionary decision tree model. IOT Technologies in Smart-Cities: From Sensors to Big Data. Security and Trust 131.
Safari, M.J.S., Ebtehaj, I., Bonakdari, H., Es-haghi, M.S., 2019. Sediment transport modeling in rigid boundary open channels using generalize structure of group method of data handling. J. Hydrol. 577, 123951.
Safari, M.J.S., Shirzad, A., 2019. Self-cleansing design of sewers: Definition of the optimum deposited bed thickness. Water Environ. Res. 91 (5), 407–416.
Safari, M.J.S., Mohammadi, M., Ab Ghani, A., 2018. Experimental studies of self-cleansing drainage system design: a review. J. Pipeline Syst. Eng. Pract. 9 (4), 04018017.
Shiri, J., 2018. Improving the performance of the mass transfer-based reference evapotranspiration estimation approaches through a coupled wavelet-random forest methodology. J. Hydrol. 561, 737–750.
Vongvisessomjai, N., Tingsanchali, T., Babel, M.S., 2010. Non-deposition design criteria for sewers with part-full flow. Urban Water J. 7 (1), 61–77.
Wan Mohtar, H.M.W., Ab Ghani, A., Safari, M.J.S., Taib, A.M., Haitham, A.F.A.N., Ahmed, E.S., 2021. Sediment Incipient Motion in Sewer with a Bed Deposit. Teknik Dergi 33 (1). https://doi.org/10.18400/tekderg.572529.
Wang, Y., Witten, I.H., 1997. Induction of model trees for predicting continuous classes. In: In 9th European Conference on Machine Learning, pp. 128–137.
Wold, S., Esbensen, K., Geladi, P., 1987. Principal component analysis. Chemometrics and intelligent laboratory systems 2 (1–3), 37–52.
Yapo, P.O., Gupta, H.V., Sorooshian, S., 1996. Automatic calibration of conceptual rainfall-runoff models: sensitivity to calibration data. J. Hydrol. 181 (1–4), 23–48.
Yu, P.S., Yang, T.C., Chen, S.Y., Kuo, C.M., Tseng, H.W., 2017. Comparison of random forests and support vector machine for real-time radar-derived rainfall forecasting. J. Hydrol. 552, 92–104.
Zhan, C., Gan, A., Hadi, M., 2011. Prediction of lane clearance time of freeway incidents using the M5P tree algorithm. IEEE Trans. Intell. Transp. Syst. 12 (4), 1549–1557.
Zhao, G., Pang, B., Xu, Z., Xu, L., 2020. A hybrid machine learning framework for real-time water level prediction in high sediment load reaches. J. Hydrol. 581, 124422.
Zhao, W., Sánchez, N., Lu, H., Li, A., 2018. A spatial downscaling approach for the SMAP passive surface soil moisture product using random forest regression. J. Hydrol. 563, 1009–1024.
Zounemat-Kermani, M., Mahdavi-Meymand, A., Alizamir, M., Adarsh, S., Yaseen, Z.M., 2020. On the complexities of sediment load modeling using integrative machine learning: Application of the great river of Loíza in Puerto Rico. J. Hydrol. 585, 124759.