



Kernel ridge regression model for sediment transport in open channel flow

Mir Jafar Sadegh Safari¹ · Shervin Rahimzadeh Arashloo²

Received: 4 July 2020 / Accepted: 1 December 2020 / Published online: 11 January 2021
© The Author(s), under exclusive licence to Springer-Verlag London Ltd. part of Springer Nature 2021

Abstract

Sediment transport modeling is of primary importance for the determination of channel design velocity in lined channels. This study proposes to model sediment transport in open channel flow using kernel ridge regression (KRR), a nonlinear regression technique formulated in the reproducing kernel Hilbert space. While the naïve kernel regression approach provides high flexibility for modeling purposes, the regularized variant is equipped with an additional mechanism for better generalization capability. In order to better tailor the KRR approach to the sediment transport modeling problem, unlike the conventional KRR approach, in this study the kernel parameter is directly learned from the data via a new gradient descent-based learning mechanism. Moreover, for model construction, a procedure based on Cholesky decomposition and forward-back substitution is applied to improve the computational complexity of the approach. Evaluation of the recommended technique is performed utilizing a large number of laboratory experimental data where the examination of the proposed approach in terms of three statistical performance indices for sediment transport modeling indicates a better performance for the developed model in particle Froude number computation, outperforming the conventional models as well as some other machine learning techniques.

Keywords Sediment transport · Open channel · Rigid boundary channel · Kernel ridge regression · Regularization

1 Introduction

Sediment transport in open channel flow is of paramount importance for the design of drainage and sewer pipe channels. In practice, it is quite essential to design a channel in which flow retains the sediment particles in motion without deposition. Deposition of sediment has numerous undesirable effects on the performance of open channels. For instance, it changes the shear stress and velocity distribution, and therefore, the transport capacity of the channel reduces as a result of a decrease in the channel cross section area [1]. As the most important

environmental impact, one may consider the deposited sediment at the bottom of the channel which is known as the major source of pollution in drainage and sewer systems [2].

In the past and in some cases till present, a number of different values of flow velocity or shear stress were used as the main criterion for conventional sewer and drainage system design. Design standards in many countries suggested a variety of velocity and shear stress values in the ranges of 0.3–1 m/s for the former, and 1–12.6 N/m² for the latter [3, 4]. Such an approach relies on experience, without deep theoretical justification and consequently suffers from different deficiencies as the basic effective factors of sediment transport are not taken into account. To this end, in the recent years, the conventional method is modified to account for a larger number of parameters affecting sediment deposition in open channel flow. Based on different sediment transport conditions, the fundamental design principles of incipient deposition, incipient motion and non-deposition were recommended [3–14]. The corresponding models were then developed considering the

✉ Mir Jafar Sadegh Safari
jafar.safari@yasar.edu.tr

Shervin Rahimzadeh Arashloo
s.rahimzadeh@cs.bilkent.edu.tr

¹ Department of Civil Engineering, Yaşar University, Izmir, Turkey

² Department of Computer Engineering, Bilkent University, Ankara, Turkey

hydraulic characteristics using the conventional multiple nonlinear regression technique using laboratory experimental data. In this context, due to their perfect suitability to the problem, machine learning techniques have been widely put into practice not only for environmental modeling problems but also in other applications. Because of the stochastic nature of sediment movement within the flow, development of an analytical model applicable to channel design is a crucial problem. Due to the turbulence in the fluid flow, sediment particles within the flow are randomly diffused in the fluid, that leads to difficulties in accurate detection of the sediment particle motion within the flow. Therefore, the transport of sediment particles in a turbulent flow has a stochastic nature. Recently, applicability of different machine learning algorithms is analyzed for the design of drainage systems. As an instance, gene expression programming (GEP) and adaptive neuro-fuzzy inference systems (ANFIS) were implemented by Ab Ghani and Azamathulla [15] and Azamathulla et al. [16] for sediment transport modeling, respectively. The most commonly applied machine learning approaches to sediment transport in drainage systems are reported in Table 1. It is seen from Table 1 that support vector machines (SVM) and its hybridized versions with wavelet and firefly algorithm (FA); artificial neural network (ANN) techniques of feed forward-back propagation (FFBP), radial basis function (RBF) and generalized regression (GR); integration of ANN with decision tree (DT) and evolutionary

algorithms (EA); particle swarm optimization (PSO), neuro-fuzzy-based group method of data handling (NF-GMDH), evolutionary polynomial regression (EPR), extreme learning machine (ELM), multigene genetic programming (MGGP), multivariate adaptive regression splines (MARS), fuzzy c-means-based adaptive neuro-fuzzy inference system (FCM-ANFIS), generalized structure group method of data handling (GS-GMDH) and evolutionary polynomial regression–multi-objective genetic algorithm (EPR-MOGA) are typical examples of machine learning techniques applied to sediment transport modeling. The majority of the studies mentioned above utilized limited ranges of experimental data for an analysis of the applicability of the available machine learning techniques in the literature to the problem.

This study advocates kernel ridge regression (KRR) for sediment transport modeling purposes. The kernel ridge regression approach may be considered as a member of the large family of kernel methods constituting a family of pattern analysis techniques, whose best-known approach is the support vector machines (SVM) [36]. The kernel methods owe their name to the utilization of a *kernel* function which enables them to operate in a very high (possibly infinite) dimensional space, known as reproducing kernel Hilbert space (RKHS) without the requirement to perform computations explicitly in that high dimensional space. This is realized through the usage of the *kernel trick* which facilitates computing the inner products of the

Table 1 Some machine learning studies on sediment transport modeling

References	Technique
Ab Ghani and Azamathulla [15]	GEP
Azamathulla et al. [16]	ANFIS
Ebtehaj et al. [17]	Wavelet–SVM
Safari et al. [18]	FFBP ANN, RBF ANN, GR ANN
Qasem et al. [19]	PSO- RBF
Ebtehaj et al. [20]	DT-ANN
Ebtehaj et al. [21]	SVM-FA
Ebtehaj and Bonakdari [22]	EL-ANN
Najafzadeh and Bonakdari [23]	NF-GMDH
Najafzadeh et al. [24]	EPR
Roushangar and Ghasempour [25]	ELM- SVM
Roushangar and Ghasempour [26]	EA-ANN
Safari et al. [27]	PSO
Safari and Danandeh Mehr [28]	MGGP
Wan Mohtar et al. [29]	FFBP ANN, RBF ANN
Safari and Shirzad [30]	PSO
Safari [31]	DT, GR ANN, MARS
Kargar et al. [32]	GEP, ANFIS
Safari et al. [33]	GEP, ELM, GS-GMDH, FCM-ANFIS
Ebtehaj et al. [34]	ELM
Montes et al. [35]	EPR-MOGA

mapped data points through a (commonly nonlinear) kernel function. In regression analysis, the high flexibility of the proposed approach provides a suitable mechanism to infer highly accurate models where the conventional regression methods would often fail. The objective function of the sediment transport model developed in the current study incorporates two terms. The first term measures data fidelity. That is, it measures how close the estimated value lies to its real value. However, the regression model is susceptible to over-fitting due to its high flexibility. In order to constrain the flexibility of the model and prevent possible over-fitting to the training data, a regularization term is included in the formulation which penalizes the norm of the inferred function in the RKHS. This in turn provides a mechanism to control the flexibility of the model to a desired degree. As will be discussed in the subsequent sections, in this work, all the parameters characterizing the KRR model are inferred directly from the data.

1.1 Contributions

The main contributions of this study may be summarized as follows:

1. While a variety of different machine learning approaches have been applied to the problem, we propose to model sediment transport via a nonlinear regression technique, called kernel ridge regression, a.k.a. KRR. Although the KRR approach has been around for some while and has been applied to different problems in other disciplines [37, 38], it has not been utilized in the context of environmental modeling problems for sediment transport modeling. In the general domain of machine learning, the two potentially successful groups to be applied to the problem may be considered to be those based on deep learning [39] and those based on kernel methods [40]. While the deep learning-based approaches have recently illustrated superior performance in different problems, their applicability is typically limited to the problems where a very large number (typically in the order of millions) of training instances are available. In contrast, kernel-based approaches may be considered as the top alternative in problems with moderate amount of training data such as sediment transport modeling. In this context, the superiority of our proposition is verified by evaluating the performance of the recommended approach on a large number of experimental tests, yielding state-of-the-art performance for sediment transport modeling;
2. A main difficulty attributed to the kernel-based methods is that of choosing an appropriate kernel function for the problem at hand. In this study, we opt for the

most widely used kernel function (RBF) and try to relax the difficulty to some extent by learning the kernel parameter directly from the data. In this context, we propose a modified KRR algorithm which unlike the conventional approach learns the kernel parameter directly from the data through a novel gradient-based learning scheme. Learning the model parameter directly from the data can be viewed as an important contribution as, in general, there exists no principled approach neither for choosing the kernel function in the KRR approach nor its parameters. In this context, the new learning approach relaxes the difficulties associated with choosing an appropriate kernel function by providing a data-driven kernel parameter learning mechanism;

3. We present a computationally efficient variant of the proposed modified KRR algorithm (applied to the sediment transport modeling) to improve its computational complexity based on Cholesky decomposition and forward-back substitution;
4. Last but not least, we provide an experimental examination of the recommended technique on a large number of laboratory experimental data and compare its performance to other alternative solutions.

1.2 Organization

The subsequent sections of the manuscript are organized as follows: In Sect. 2, we provide a review of the existing techniques to sediment transport modeling followed by a critical analysis of their performances where the proposed approach is put into context. In Sect. 3, the experimental data, the model and performance indices in the current study are introduced. Section 4 presents the proposed approach for sediment transport modeling. Sections 5–6 present the results of an experimental evaluation of the proposed approach and provide further discussions. Finally, in Sect. 7, conclusions are drawn.

2 Sediment transport in rigid boundary open channels

Sediment transport in loose boundary channels has been widely investigated in the past. The basic methodology could be accessed in the universal sources such as Raudkivi [41] and Vanoni [42]. The critical difference between loose boundary and rigid boundary channels lies in the fact that the supply rate of sediment particle within the flow in loose boundary channels is gathered from channel itself, whereas in the case of rigid boundary channels, it depends on the wastewater catchment [7]. In other words, in the case of

loose boundary channels, the channel's boundary is erodible and once the flow reaches the critical shear stress or velocity value, it may change the cross section of the channel causing the eroded sediment particles from the channel boundary to enter the flow. In the rigid boundary channels, on the other hand, flow cannot erode the channel boundary to change its cross section shape, and accordingly, it is not a source of sediment particles. Sediment particles in the rigid boundary channels are gathered from surrounding area and depend on the catchment characteristics. Consequently, a direct application of loose boundary channels sediment transport models in rigid boundary channels is not recommended [43, 44]. Motivated by the channel design considerations, sediment transport studies typically focus on three basic sediment movement conditions of incipient deposition, incipient motion and non-deposition. It is well known that non-deposition condition of sediment transport is a reliable criterion for channel design purposes to keep the channel clean from sediment deposition [2]. The most prominent studies on non-deposition sediment transport are given in Table 2 and briefly reviewed next.

The earliest experimental studies in this context were conducted by Craven [45], Ambrose [46], Durand [47] and Durand and Condolios [48]. The models suggested incorporated particle Froude number and sediment volumetric concentration, while channel characteristic was missing in the proposed models. Robinson and Graf [49] suggested a different model for high sediment volumetric concentration and reported that flow mean velocity is directly proportional to the sediment volumetric concentration. Novak and Nalluri [43, 50] applied the models that were initially proposed for loose boundary channels to experimental data collected from rigid boundary channels and concluded that a lower velocity and shear stress level is needed in rigid boundary channels. May [51] investigated the impact of hydraulic parameters in sediment transport followed by their subsequent study, May et al. [52], where the authors investigated the sediment transport problem considering a larger number of hydraulic variables. Mayerle [53] performed experiments to determine non-deposition sediment transport condition followed by a re-analysis of the data by Mayerle et al. [5] which led to an acceptable agreement with Novak and Nalluri [43] based on the experimental data. Kithsiri [54] developed an alternative model considering the average bed shear stress and evaluated the applicability of the model to laboratory experimental data. Ab Ghani [7] studied non-deposition sediment transport in a variety of pipe sizes using a wide range of sediment sizes with the conclusion that the channel design velocity may increase in proportion to the pipe size, sediment concentration and channel bed roughness. Ackres et al. [55] and May et al. [56] evaluated the equations available in the

literature and found out that none of the equations gave acceptable results. Therefore, using the previously collected experimental data, a model similar to that of May et al. [52] was developed. Ota [8], Ota and Nalluri [9] and Ota and Perrusquia [10] studied sediment particle characteristics and developed different models for different pipe sizes. Vongvisessomjai et al. [12] evaluated the conventional Camp criterion and concluded that it overpredicts channel design velocity. Safari et al. [13] extended the sediment transport models developed previously by considering the cross section shape factor. In a subsequent study, Safari et al. [14] performed experiments in different cross section channels. In their work, channel geometry characteristics were utilized for cross section shape factor consideration, which is linked to the flow resistance. The performance of the developed model was examined on a large number of data sets. The aforementioned studies and those reported in Table 2 developed different models utilizing multiple nonlinear regression analyses. Most of these approaches considered particle Froude number ($Fr_p = V/\sqrt{gd(s-1)}$) as the dependent variable which consists of flow mean velocity, sediment median size and sediment relative specific mass. Further details on the experimental sediment transport studies may be found in Safari et al. [2].

Sediment transport models listed in Table 2 were suggested for rigid boundary channels. Behind theoretical consideration of sediment movement within the flow based on velocity or shear stress approaches, dimensional analysis was widely used for determination of the effective parameters involved in non-disposition condition of sediment transport. It can be understood from Table 2 that dimensionless particle Froude number (Fr_p) or volumetric sediment concentration (C_v) was considered as dependent parameter in the models developed. It is worthy to emphasize that dimensionless grain size parameter (D_{gr}) and relative particle size (d/R) were found as effective parameters indicating the flow and sediment characteristics. Although the most of the models summarized in Table 2 have ignored the channel characteristic in their structure, it is quite important to incorporate the channel friction factor (λ) as an independent parameter into the model to consider flow resistance in the modeling procedure.

While there has been a considerable amount of research for sediment transport modeling via a regression analysis, one may consider two general drawbacks for the existing techniques: (1) they either follow the observed training data too closely without making any prior assumptions on the function regressing the data or (2) the models deployed are too restrictive and inflexible which do not allow the training data to be fully utilized. The first category, despite

Table 2 Empirical sediment transport models for rigid boundary open channels

Reference	Model
Craven [45]	$\frac{V}{\sqrt{gd(s-1)}} \geq 6.37 C_v^{1/3} \left(\frac{d}{D}\right)^{-0.50}$
Ambrose [46]	$\frac{V}{\sqrt{2gD(s-1)}} = 2.45 C_v^{1/4}$
Durand [47], Durand and Condolios [48]	$\frac{V}{\sqrt{2gD(s-1)}} = 1.36 C_v^{0.136}$
Robinson and Graf [49]	$\frac{V}{\sqrt{2gD(s-1)}} = \frac{1.40 C_v^{0.11} d^{0.06}}{1 - \tan \theta}$
May [51]	$C_v = 0.0205 \left(\frac{D^2}{A}\right) \left(\frac{d}{R}\right)^{0.6} \left(\frac{V^2}{g(s-1)D}\right)^{1.5} \left(1 - \frac{V_{im}}{V}\right)^4$
May et al. [52]	$C_v = 0.0211 \left(\frac{Y}{D}\right)^{0.36} \left(\frac{D^2}{A}\right) \left(\frac{d}{R}\right)^{0.6} \left(\frac{V^2}{g(s-1)D}\right)^{1.5} \left(1 - \frac{V_{im}}{V}\right)^4$
Mayerle [53], Mayerle et al. [5]	$\frac{V}{\sqrt{gd(s-1)}} = 14.43 C_v^{0.18} D_{gr}^{-0.14} \left(\frac{d}{R}\right)^{-0.56} \lambda^{0.18}$
Kithsiri [54]	$\frac{\tau}{\rho g d(s-1)} = 12.93 C_v^{0.29} D_{gr}^{-0.21} \left(\frac{R}{d}\right)^{-0.98} \lambda^{1.50}$
Ab Ghani [7]	$\frac{V}{\sqrt{gd(s-1)}} = 3.08 C_v^{0.21} D_{gr}^{-0.09} \left(\frac{R}{d}\right)^{0.53} \lambda^{-0.21}$
Ackers et al. [55], May et al. [56]	$C_v = 0.0303 \left(\frac{D^2}{A}\right) \left(\frac{d}{D}\right)^{0.6} \left(\frac{V^2}{g(s-1)D}\right)^{1.5} \left(1 - \frac{V_{im}}{V}\right)^4$
Vongvisessomjai et al. [12]	$\frac{V}{\sqrt{gd(s-1)}} = 4.31 C_v^{0.226} \left(\frac{d}{R}\right)^{-0.616}$
Safari et al. [13]	$\frac{V}{\sqrt{gd(s-1)}} = 7.34 C_v^{0.13} D_{gr}^{-0.12} \left(\frac{d}{R}\right)^{-0.44} \beta^{-0.91}$
Safari et al. [14]	$\frac{V}{\sqrt{gd(s-1)}} = 4.83 C_v^{0.09} D_g^{-0.14} \left(\frac{d}{R}\right)^{-0.32} \left(\frac{P}{B}\right)^{0.20}$

V flow mean velocity; d sediment median size; g gravitational acceleration; s relative specific mass of sediment (ρ_s) to specific mass of fluid (ρ); C_v sediment volumetric concentration; D pipe diameter; θ angle of the channel bed to the horizontal; Y flow depth; λ channel friction factor; V_{im} flow mean velocity at incipient motion; D_{gr} dimensionless grain size parameter ($= (s-1)gd^3/\nu^2)^{1/3}$; ν fluid kinematic viscosity; τ average bed shear stress, β cross section shape factor ($= \sqrt{P/B}/1.31(B/D_h)^{-0.49}$), P wetted perimeter, B water surface width, D_h hydraulic depth of flow

achieving relatively low error rates on the training data, suffer from an over-fitting phenomenon and thus yield high error rates on the unseen test data. The second group, on the other hand, are unable to fully exploit the training samples and are subject to an under-fitting problem. In contrast, the proposed approach in the present study is able to find the right balance between following the training samples too closely or not being flexible enough to benefit from the training instances. The desired behavior, as we shall see, is explicitly encoded into the objective function of the proposed approach while its performance is validated on laboratory experimental data and compared against the existing solutions.

3 Methodology

The experimental data, model parameters and the performance indices used in the current study are discussed next.

3.1 Experimental data

The sediment transport experimental data compiled from the rigid boundary open channels are utilized in the current study where the non-deposition with a clean bed sediment transport condition is considered. To this end, four experimental data of May [6], Ab Ghani [7], Vongvisessomjai et al. [12] and Mayerle [53] are utilized for the modeling purposes. May [6] performed tests in a 450-diameter circular pipe utilizing 0.73-mm sediment size where the author determined the minimum non-deposition condition by satisfying the equilibrium of sediment deposition within a specific period of time. Through experiments he recorded different hydraulic characteristics such as flow velocity, discharge and flow depth. Ab Ghani [7] performed tests in three different pipe sizes having 154 mm, 305 mm and 405 mm diameters utilizing sediment sizes of 0.46–8.3 mm. Based on the experiments conducted, through increasing the amount of entered sediment into the flow, the minimum flow velocity for non-deposition condition was determined. Vongvisessomjai et al. [12]

Table 3 KRR training using Cholesky decomposition and forward-back method

Algorithm 1	
1:	n = number of training observations
2:	\mathbf{y} =expected responses
3:	Compute the kernel matrix \mathbf{K}
4:	$\mathbf{K} = \mathbf{R}^T \mathbf{R}$
5:	Find θ in $\mathbf{R}^T \theta = \mathbf{y}$
6:	Find α in $\mathbf{R} \alpha = \theta$
7:	Output α

performed different tests in two circular pipes having 100 and 150 mm diameters using sediment with sizes of 0.2–0.43 mm. Similar to the previous studies, by increasing the entered sediment in the flow, sediment deposition was achieved. Mayerle [53] conducted tests in circular and rectangular channels. Sediment particles with different sizes of 0.5–8.74 mm were used in the tests. The non-deposition condition of sediment transport was determined as the condition where sediments are transported in the flow without a tendency to deposit. The minimum flow velocity that satisfies such a condition is defined as the non-deposition velocity. Additional details regarding the experimental data and experimental procedures may be found in Safari et al. [2].

3.2 Model parameters

Analyzing the sediment transport models given in Table 2, it can be seen that the sediment, fluid, flow and channel characteristics are commonly recognized as the most effective variables across different studies. Accordingly, the most important variables in sediment transport process in open channels flow for the design purposes may be selected as sediment-specific mass (ρ_s), sediment median size (d), sediment volumetric concentration (C_v) for expressing the sediment properties; hydraulic radius (R), flow mean velocity (V) and gravitational acceleration (g) as sediment features variables; kinematic viscosity (ν) and fluid specific mass (ρ) as fluid properties; and channel friction factor (λ) for including the channel properties. Accordingly, variables mentioned above can be expressed as

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

The parameters involved in the equation above are computed as the first stage of data preparation. The parameter at the left-hand side of Eq. (1) is the particle

Froude number (Fr_p) which is selected as model's output, while the C_v , D_{gr} , d/R and λ are chosen as inputs to the model. The entire dataset is divided into two non-overlapping splits for training and testing purpose. At the training stage, the unidentified relationship of dependent and independent parameters given in Eq. (1) is determined while at the testing phase the constructed model is examined on the left-out data set. From the available 375 data items, 300 are randomly chosen for training purposes and the rest for testing the model's accuracy. That is, 80% of the data is used for training and parameter adjusting purposes and the 20% for testing phase. The training set is further divided randomly into two disjoint sets where one is used for model training while the second partition is employed to set model parameters based on the performance (hold out). In order to give the parameters equal importance in the modeling process, the data are re-scaled to the range of [0–1] via dividing them by their maximum values.

3.3 Performance indices

The examination of model performance is of importance to determine its validity. To this end, three performance indices of concordance coefficient (CC), root mean square error ($RMSE$) and discrepancy ratio (DR) are utilized for model evaluation. CC calculates the concordance of the measured and calculated variables with the perfect value of unity and can be computed by

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

where σ_m and σ_c denote the standard deviations of the measured and calculated Fr_p , r is the correlation coefficient while $\overline{Fr_p^m}$ and $\overline{Fr_p^c}$ stands for the average of measured and calculated Fr_p , respectively. The $RMSE$ measures the model's error and can be computed by

$$RMSE = \sqrt{\frac{\sum_{i=1}^n \left(Fr_p^m - Fr_p^c\right)^2}{n}} \quad (3)$$

where Fr_p^m and Fr_p^c are the measured and the calculated Fr_p , respectively, and n is the number of data items. The discrepancy ratio (DR) calculates the percentage of model accuracy in a specific range and can be written by

$$DR = \frac{Fr_p^c}{Fr_p^m} \quad (4)$$

In the current study, DR is grouped in the 80–120%, 70–130% and 50–150% ranges.

4 Regularized regression in the Hilbert space

In order to fully characterize the model developed in this work for prediction purposes, we start the discussion with the linear regression, a commonly applied technique for function estimation and then illustrate how it may be generalized to a more advanced model to handle more difficult situations in terms of the complexity and nonlinearity of the data. The generalization to be discussed corresponds to the formulation of the linear regression technique in the RKHS: reproducing kernel Hilbert space (also known as feature space) so that the estimation function would ultimately act as a nonlinear mapping function with respect to the original data points. The kernel ridge regression (KRR) model development process is discussed in the subsequent sections and summarized as the flowchart given in Fig. 1.

4.1 Linear regression

Let us suppose that there exists a set of data items x_1, \dots, x_n in a space of d dimensions which are to be mapped onto real numbers $y_1, \dots, y_n \in \mathbb{R}$. In other words, each point $x_i \in \mathbb{R}^d$ shall be mapped onto the corresponding $y_i \in \mathbb{R}$ through a function. From now on, we shall refer to y_i 's as *responses*. A commonly applied technique to solve the mapping problem mentioned above is that of linear least squares regression. In this approach, it is assumed that a linear function $g(x_i) = \alpha^\top x_i$ is able to generate the responses y_i 's. The goal then becomes estimating the unknown parameters α [57]. The squared error incurred by the linear model with respect to the sample x_i would be $(y_i - \alpha^\top x_i)^2$. In an ideal scenario, this error should be zero. Consequently, in order to estimate the parameter α , a sum of squared errors is commonly minimized:

$$\begin{aligned} \min_g \text{Cost} &= \min_g \sum_{i=1}^n (y_i - g(x_i))^2 = \min_{\alpha} \sum_{i=1}^n (y_i - \alpha^\top x_i)^2 \\ &= \min_{\alpha} (\mathbf{y} - \mathbf{X}\alpha)^\top (\mathbf{y} - \mathbf{X}\alpha) \end{aligned} \quad (5)$$

where \mathbf{X} denotes the $n \times d$ data matrix, \mathbf{y} is the $n \times 1$ vector of expected responses and α is the $d \times 1$ vector of parameters characterizing the function. The optimal solution α may be determined by setting the derivative of the cost function w.r.t. α to zero:

$$\frac{\partial \text{Cost}}{\partial \alpha} = -2\mathbf{X}^\top \mathbf{y} + 2\mathbf{X}^\top \mathbf{X}\alpha = 0 \quad (6)$$

α would then be derived as

$$\alpha = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y} \quad (7)$$

Having determined α , the function $g(\cdot)$ at a given point z is then evaluated as

$$\begin{aligned} g(z) &= \alpha^\top z = \mathbf{y}^\top \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} z = \mathbf{y}^\top (\mathbf{X}\mathbf{X}^\top)^{-1} \mathbf{X}z \\ &= \mathbf{y}^\top (\mathbf{X}\mathbf{X}^\top)^{-1} [\langle z, x_1 \rangle, \dots, \langle z, x_n \rangle]^\top \\ &= [\langle z, x_1 \rangle, \dots, \langle z, x_n \rangle] (\mathbf{X}\mathbf{X}^\top)^{-1} \mathbf{y} \end{aligned} \quad (8)$$

where $\langle \cdot, \cdot \rangle$ denotes the dot (inner) product. Considering Eq. (8), $g(z)$ may be determined if $(\mathbf{X}\mathbf{X}^\top)^{-1}$ exists. As it will be seen shortly, in the nonlinear setting, the equivalence of this constraint can be readily satisfied by choosing a suitable *kernel function*.

An important aspect regarding Eq. (8) is that of representation of all terms involving the data points as dot products. Such a representation is highly useful in extending the linear regression technique into the reproducing kernel Hilbert space (RKHS) where one would ultimately achieve a regression function which is *nonlinear* w.r.t. the original data items in the d -dimensional space.

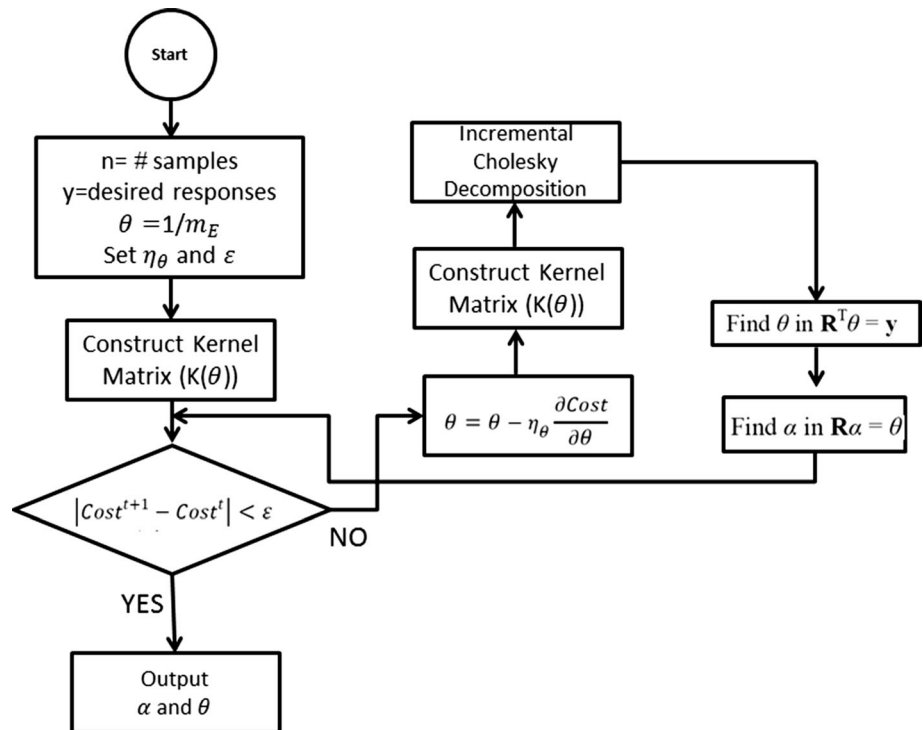
4.2 Linear regression in the RKHS

Despite the utility and simplicity of the linear regression which make it a suitable candidate for a wide range of applications, it suffers from a lack of representation capacity for some applications. In particular, when the data represents nonlinear characteristics, the preceding approach based on linear regression may not yield desirable results. An alternative solution in these cases is offered by a nonlinear regression method. In particular, the regression approaches formulated in the reproducing kernel Hilbert space provide a powerful alternative to other existing techniques. In a RKHS, the approach is to project the data into a high-dimensional (possibly infinite-dimensional) feature space in an implicit fashion with the hope that the conventional techniques would offer better solutions in this new space. The most well-known instance of these methods is offered by kernel machines [58–61].

Let us assume a nonlinear projection $\phi: \mathbb{R}^d \rightarrow \mathcal{F}$ which induces the RKHS (a.k.a. the feature space \mathcal{F}). For a suitably chosen mapping, it is well known that an inner product $\langle \cdot, \cdot \rangle$ on the feature space \mathcal{F} can be expressed via a positive semidefinite kernel $\kappa(\cdot, \cdot)$ as $\langle \phi(x_i), \phi(x_j) \rangle = \kappa(x_i, x_j)$.

Being equipped with the necessary tools, extending the linear regression [62] to the RKHS [63] is straightforward. In this case, each x is projected onto $\phi(x)$ and one seeks a real function $g(\cdot)$ operating on $\phi(x_i)$'s of the feature space \mathcal{F} . That is, one is interested in $g(\phi(x)) = f(x)$ and thus, the regression problem to be solved in this case may be expressed as

Fig. 1 Flowchart of the core of the proposed learning approach



$$\begin{aligned} \min_g \sum_{i=1}^n (y_i - g(\phi(x_i)))^2 &= \min_f \sum_{i=1}^n (y_i - f(x_i))^2 \\ &= \min_{\beta} \sum_{i=1}^n (y_i - \beta^\top \phi(x_i))^2 \end{aligned} \quad (9)$$

Where as before, we have assumed a linear model. Note that the new linear model is defined in the RKHS rather than the original space of the data. The problem above may not be easily solved since β and $\phi(x_i)$ are very high (possibly infinite)-dimensional. Nevertheless, the problem may be more easily approached using inner products as follows. Since $f(z) = g(\phi(z))$, using Eq. (8), the relation for $f(z)$ may be written as

$$\begin{aligned} f(z) &= [\langle \phi(z), \phi(x_1) \rangle, \dots, \langle \phi(z), \phi(x_n) \rangle] (\phi(\mathbf{X}) \phi(\mathbf{X})^\top)^{-1} \mathbf{y} \\ &= [\kappa(z, x_1), \dots, \kappa(z, x_n)] \mathbf{K}^{-1} \mathbf{y} \end{aligned} \quad (10)$$

where we have used the fact that a dot product in the feature space \mathcal{F} can be equivalently written as $\langle \phi(x_i), \phi(x_j) \rangle = \kappa(x_i, x_j)$ and also used the symbol $\mathbf{K} = \phi(\mathbf{X}) \phi(\mathbf{X})^\top$. Note that \mathbf{K} denotes a matrix of inner products in the Hilbert space and thus its (i, j) th element is $\langle \phi(x_i), \phi(x_j) \rangle$. Thus, $K_{ij} = \kappa(x_i, x_j)$. Consequently, the evaluation of $f(\cdot)$ at an arbitrary point z may be viable if \mathbf{K}^{-1} exists.

Denoting $\alpha = \mathbf{K}^{-1} \mathbf{y}$, function $f(\cdot)$ may be expressed as

$$\begin{aligned} f(\cdot) &= [\kappa(\cdot, x_1), \dots, \kappa(\cdot, x_n)] \alpha \\ &= \sum_{i=1}^n \alpha_i \kappa(\cdot, x_i) \end{aligned} \quad (11)$$

In this case, if desired, parameter β in Eq. (9) may be derived as

$$\begin{aligned} \min_f \sum_{i=1}^n (y_i - f(x_i))^2 &= \min_{\alpha} \sum_{i=1}^n (y_i - \sum_{j=1}^n \alpha_j \kappa(x_i, x_j))^2 \\ &= \min_{\alpha} \sum_{i=1}^n (y_i - \sum_{j=1}^n \alpha_j \langle \phi(x_i), \phi(x_j) \rangle)^2 \\ &= \min_{\alpha} \sum_{i=1}^n (y_i - \left[\sum_{j=1}^n \alpha_j \phi(x_j) \right]^\top \phi(x_i))^2 \\ &= \min_{\beta} \sum_{i=1}^n (y_i - \beta^\top \phi(x_i))^2 \end{aligned} \quad (12)$$

and as a result, $\beta = \sum_{j=1}^n \alpha_j \phi(x_j)$.

Using Eq. (11), the responses on the data \mathbf{X} may be derived as $f(\mathbf{X}) = \mathbf{K} \alpha$. The resulting function $f(\cdot)$ corresponds to the optimal least squares regression solution in the Hilbert space which is linear with respect to $\phi(x_i)$'s and nonlinear w.r.t. the data points x_i 's. The problem may be expressed in a matrix form as

$$\min_{\alpha} \| \mathbf{K} \alpha - \mathbf{y} \|_2^2 \quad (13)$$

The relation above represents a *non-regularized* kernel regression. In this case, if \mathbf{K}^{-1} exists, function $f(\cdot) = \sum_{i=1}^n \alpha_i \kappa(\cdot, x_i)$ corresponds to an exact interpolation for the given data points. That is, it provides a zero error on the training data samples. In practice, however, an exact interpolation for the given training data may not be desirable as it may reduce the generalization capability of the method on any new and previously unseen data points. In such cases, a commonly applied technique is that of regularization which corresponds to imposing some constraint on function $f(\cdot)$. A commonly applied regularization is Tikhonov regularization [64] where one penalizes the norm of the inferred function $f(\cdot)$ [65]. This may be applied by introducing a new regularization term into the cost function as

$$\min_{\alpha} \|\mathbf{K}\alpha - \mathbf{y}\|_2^2 + \delta \|\mathbf{K}\alpha\|_2^2 = \min_{\alpha} \|\mathbf{K}\alpha - \mathbf{y}\|_2^2 + \delta \alpha^T \mathbf{K} \alpha \quad (14)$$

where $\|\mathbf{K}\alpha\|_2^2$ denotes the norm of function $f(\cdot)$ in the Hilbert space and parameter δ is the parameter controlling the regularization, and consequently the flexibility of the model and is typically determined via cross-validation on the training set.

In order to solve the minimization problem in Eq. (14), one may compute its derivative w.r.t. α and set it to zero:

$$\frac{\partial \text{Cost}}{\partial \alpha} = 2\mathbf{K}(\mathbf{K}\alpha - \mathbf{y}) + 2\delta \mathbf{K}\alpha = 0 \quad (15)$$

where we have used the property that \mathbf{K} is a symmetric matrix. The equation above reads

$$(\mathbf{K} + \delta \mathbf{I}_n)\alpha = \mathbf{y} \quad (16)$$

\mathbf{I}_n is the $n \times n$ identity matrix. α is then derived as

$$\alpha = (\mathbf{K} + \delta \mathbf{I}_n)^{-1} \mathbf{y} \quad (17)$$

Table 4 Incremental Cholesky decomposition for efficient solving of the system of linear equations

Algorithm 2	
1:	Set $\mathbf{R}^0 = \sqrt{\kappa(x_1, x_1)}$
2:	for $m = 2:n$ do
3:	$\mathbf{k}_{1m} = [\kappa(x_1, x_m), \dots, \kappa(x_{m-1}, x_m)]^T$
4:	Find \mathbf{r}_{1m} satisfying $\mathbf{k}_{1m} = \mathbf{R}^{(m-1) \times (m-1)T} \mathbf{r}_{1m}$
5:	$k_{mm} = \kappa(x_m, x_m)$
6:	$r_{mm} = \sqrt{k_{mm} - \mathbf{r}_{1m}^T \mathbf{r}_{1m}}$
7:	$\mathbf{R}^{(m-1) \times (m-1)} = \begin{pmatrix} \mathbf{R}^{(m-1) \times (m-1)} & \mathbf{r}_{1m} \\ 0 & r_{mm} \end{pmatrix}$
8:	end for
9:	$\mathbf{R} = \mathbf{R}_{m-1}$

In practice, the equation $\mathbf{y} = \mathbf{K}\alpha$ may be solved more efficiently for α using a Cholesky decomposition [66] along with the forward-back algorithm, as per Algorithm 1 shown in Table 3. A desirable property of such a solution lies in the fact that a Cholesky decomposition can be performed incrementally to enhance efficiency, discussed next.

4.3 Incremental Cholesky factorization

In the incremental Cholesky factorization, one aims to compute the Cholesky factorization of an $m \times m$ matrix under the assumption that the Cholesky factorization of the $(m-1) \times (m-1)$ sub-matrix is available. In other words, assuming that the Cholesky factorization of matrix $\mathbf{K}^{(m-1) \times (m-1)}$ associated with $m-1$ observations are available, the goal is to insert a single sample (x_m) into the training set and then compute the Cholesky factorization of matrix $\mathbf{K}^{m \times m}$ associated with the new data set. To this aim, the *Sherman's March* technique [67] can be applied as

$$\mathbf{K}^{m \times m} = \begin{pmatrix} \mathbf{K}^{(m-1) \times (m-1)} & \mathbf{k}_{1m} \\ \mathbf{k}_{1m}^T & k_{mm} \end{pmatrix} = \begin{pmatrix} \mathbf{R}^{(m-1) \times (m-1)T} & 0 \\ \mathbf{r}_{1m}^T & r_{mm} \end{pmatrix} \begin{pmatrix} \mathbf{R}^{(m-1) \times (m-1)} & \mathbf{r}_{1m} \\ 0 & r_{mm} \end{pmatrix} \quad (18)$$

where \mathbf{k}_{1m} is an $(m-1) \times 1$ vector given by

$$\mathbf{k}_{1m} = [\kappa(x_1, x_m), \dots, \kappa(x_{m-1}, x_m)]^T \quad \text{and} \\ k_{mm} = \kappa(x_m, x_m).$$

Equation (18) reads

Table 5 KRR training using Cholesky decomposition and forward-back method where the Gaussian kernel width is also learned

Algorithm 3	
1:	n = number of training observations
2:	\mathbf{y} =expected responses
3:	$\theta = 1/m_E$
4:	repeat
5:	$\theta = \theta - \eta_{\theta} \frac{\partial \text{Cost}}{\partial \theta}$
6:	$\mathbf{K} = \mathbf{K}(\theta)$
7:	$\mathbf{K} = \mathbf{R}^T \mathbf{R}$
8:	Find θ in $\mathbf{R}^T \theta = \mathbf{y}$
9:	Find α in $\mathbf{R}\alpha = \theta$
10:	until $ \text{Cost}^{t+1} - \text{Cost}^t < \varepsilon$
11:	Output α and θ

Table 6 Comparison of different models by means of *CC* and *RMSE*

Method	<i>CC</i>	<i>RMSE</i>
KRR (this study)	0.95	0.67
Mayerle et al. [3]	0.59	3.14
Ab Ghani [5]	0.79	1.51
Vongvisessomjai et al. [11]	0.84	1.27
GR [31]	0.90	0.83
DT [31]	0.86	1.03
MARS [31]	0.92	0.79
GS-GMDH [33]	0.93	0.76
GEP [33]	0.91	0.81
FCM-ANFIS [33]	0.92	0.79
ELM [33]	0.91	0.80

$$\mathbf{K}^{(m-1) \times (m-1)} = \mathbf{R}^{(m-1) \times (m-1)} \mathbf{R}^{(m-1) \times (m-1)}$$

$$\mathbf{k}_{1m} = \mathbf{R}_{m-1}^\top \mathbf{r}_{1m}$$

$$r_{mm} = \sqrt{k_{mm} - \mathbf{r}_{1m}^\top \mathbf{r}_{1m}} \quad (19)$$

In this respect, the first step is that of solving $\mathbf{k}_{1m} = \mathbf{R}^{(m-1) \times (m-1)} \mathbf{r}_{1m}$ for \mathbf{r}_{1m} followed by computing r_{mm} . The incremental approach has been found to improve the complexity of the Cholesky decomposition procedure from cubic to quadratic in number of samples [68].

The complete procedure for the incremental Cholesky factorization can be obtained by changing m from 1 to n (total number of samples), Algorithm 2 given in Table 4. Besides reducing the computational complexity, an interesting property of the incremental method is that it allows operation in the scenarios where the data comes in one at a time, i.e., data streams. In this latter case, once a new sample arrives, one only needs to update the new part of the kernel matrix \mathbf{K} . Additionally, as the Cholesky decomposition can be operated in an incremental scheme, the computations performed previously are completely utilized.

4.4 The Choice of kernel function $\kappa(\cdot, \cdot)$

Associated with all kernel methods is a kernel function. While there are alternative choices regarding the kernel $\kappa(\cdot, \cdot)$, a widely used one is the radial basis (RBF) function $\kappa(x_i, x_j) = \exp(-\theta \|x_i - x_j\|^2)$ where θ is the parameter characterizing the width of the kernel. θ may be learned from the data. The method proposed *Cost* w.r.t. θ first, the following matrices are defined:

$$\mathbf{F} = \mathbf{X}\mathbf{X}^\top \quad (20)$$

$$\mathbf{E} = (\mathbf{I} \circ \mathbf{F})\mathbf{1} + \mathbf{1}^\top (\mathbf{I} \circ \mathbf{F})^\top - 2\mathbf{F} \quad (21)$$

where \circ stands for the Hadamard (component-wise) product and $\mathbf{1}$ denotes a matrix of ones. The kernel matrix \mathbf{K} may then be expressed as

$$\mathbf{K} = \exp[-\theta \mathbf{E}]. \quad (22)$$

The derivative of *Cost* with respect to the kernel matrix \mathbf{K} is

$$\frac{\partial \text{Cost}}{\partial \mathbf{K}} = 2(\mathbf{K}\boldsymbol{\alpha} - \mathbf{y})\boldsymbol{\alpha}^\top + \delta \boldsymbol{\alpha} \boldsymbol{\alpha}^\top \quad (23)$$

The partial derivative of the objective function *Cost* with respect to θ is obtained as

$$\frac{\partial \text{Cost}}{\partial \theta} = \text{trace} \left(\frac{\partial \text{Cost}}{\partial \mathbf{K}} (-\mathbf{K} \circ \mathbf{E}) \right) \quad (24)$$

The new learning algorithm may then be expressed as per Algorithm 3 given in Table 5 where θ is initialized to the reciprocal of the average of \mathbf{E} .

4.5 The Choice of regularization parameter δ

As noted earlier, the regularization parameter controls the flexibility of the model. While a highly flexible model is instrumental in reducing the error on the training set, it may be undesirable from a generalization perspective. In this work, the regularization parameter is set using a cross-validation approach on the train set. For this purpose, the training set is randomly split into two non-overlapping sets where one is utilized for model training (training set) while the second partition (validation set) is employed to assess the performance. In order to set δ , a grid search is performed in the range $[1e-6, \dots, 1e+2]$. The value leading to the minimum error on the validation is chosen as the optimal parameter. The parameter is then kept fixed during the experiments on the test set.

4.6 Regression function evaluation

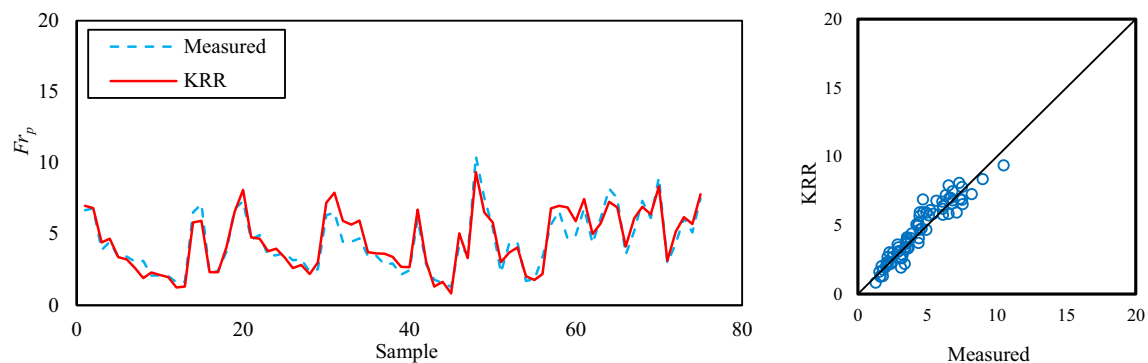
Once the parameters characterizing the nonlinear regression are found, in order to determine the output of the function $f(\cdot)$ on a test point z first $\mathbf{k}_z = [\kappa(x_1, z), \dots, \kappa(x_n, z)]^\top$ is computed followed by $f(z) = \mathbf{k}_z^\top \boldsymbol{\alpha}$.

4.7 Computational complexity

A number of observations regarding the computational complexity of the proposed approach are in order. A computationally expensive step of the proposed approach in the training stage is that of Cholesky factorization having cubic complexity in number of training observations. In the current study, the computational complexity of this

Table 7 Comparison of different models by means of *DR*

Method	<i>DR</i> Mean	Percentage of data in range		
		0.80–1.20	0.70–1.30	0.50–1.50
KRR (this study)	1.02	80.00	92.00	100.00
Mayerle et al. [3]	1.50	24.00	32.00	48.00
Ab Ghani [5]	1.17	62.67	66.67	78.67
Vongvisessomjai et al. [11]	1.15	64.00	70.67	85.33
GR [31]	0.98	78.67	90.67	98.67
DT [31]	0.96	73.33	89.33	96.00
MARS [31]	1.03	74.67	86.67	97.33
GS-GMDH [33]	1.05	69.33	89.33	100.00
GEP [33]	1.06	68.00	88.00	100.00
FCM-ANFIS [33]	1.05	72.00	85.33	98.67
ELM [33]	1.08	73.33	84.00	96.00

**Fig. 2** Comparison of measured and calculated Fr_p by KRR model

stage is moderated through an incremental approach, reducing the complexity to quadratic in number of training observations. Given the Cholesky factorization, the linear system of equations $\mathbf{y} = \mathbf{K}\boldsymbol{\alpha}$ has a quadratic complexity in number of samples. Other steps of the proposed approach include only matrix–vector multiplications and thus not computationally very demanding.

5 Experimental evaluation

The performance of Kernel Ridge Regression (KRR) model is examined on the test data by means of three performance indices of concordance coefficient (*CC*), root mean square error (*RMSE*) and discrepancy ratio (*DR*) and reported in Tables 6 and 7. Three models of Mayerle et al. [5], Ab Ghani [7] and Vongvisessomjai et al. [12] from the literature are included in the table to compare their performances with that of KRR. Furthermore, the results of seven techniques of generalized regression neural network (GR), decision Tree (DT) and multivariate adaptive regression splines (MARS) given by Safari [31] and

generalized structure group method of data handling (GS-GMDH), gene expression programming (GEP), fuzzy c-means-based adaptive neuro-fuzzy inference system (FCM-ANFIS) and extreme learning machine (ELM) developed by Safari et al. [33] taken from the literature are also included in the comparison. It is worth noting that all the benchmark results of GR, DT, MARS, GS-GMDH, GEP, FCM-ANFIS and ELM were established on the same datasets used in this study applying the same data split strategy.

As can be seen from Table 6, it is obvious that the proposed model outperforms the conventional regression models and the machine learning techniques of GR, DT, MARS, GS-GMDH, GEP, FCM-ANFIS and ELM and in terms of *CC* and *RMSE*. The proposed model gives a *CC* of 0.95 (close to unity) and an *RMSE* of 0.67 which demonstrate the high modeling capability of the KRR approach. It also provides much better performance than the GR, DT, MARS, GS-GMDH, GEP, FCM-ANFIS and ELM machine learning techniques. The conventional regression models give inferior results with higher *RMSE* values as compared with the proposed approach. The method presented in

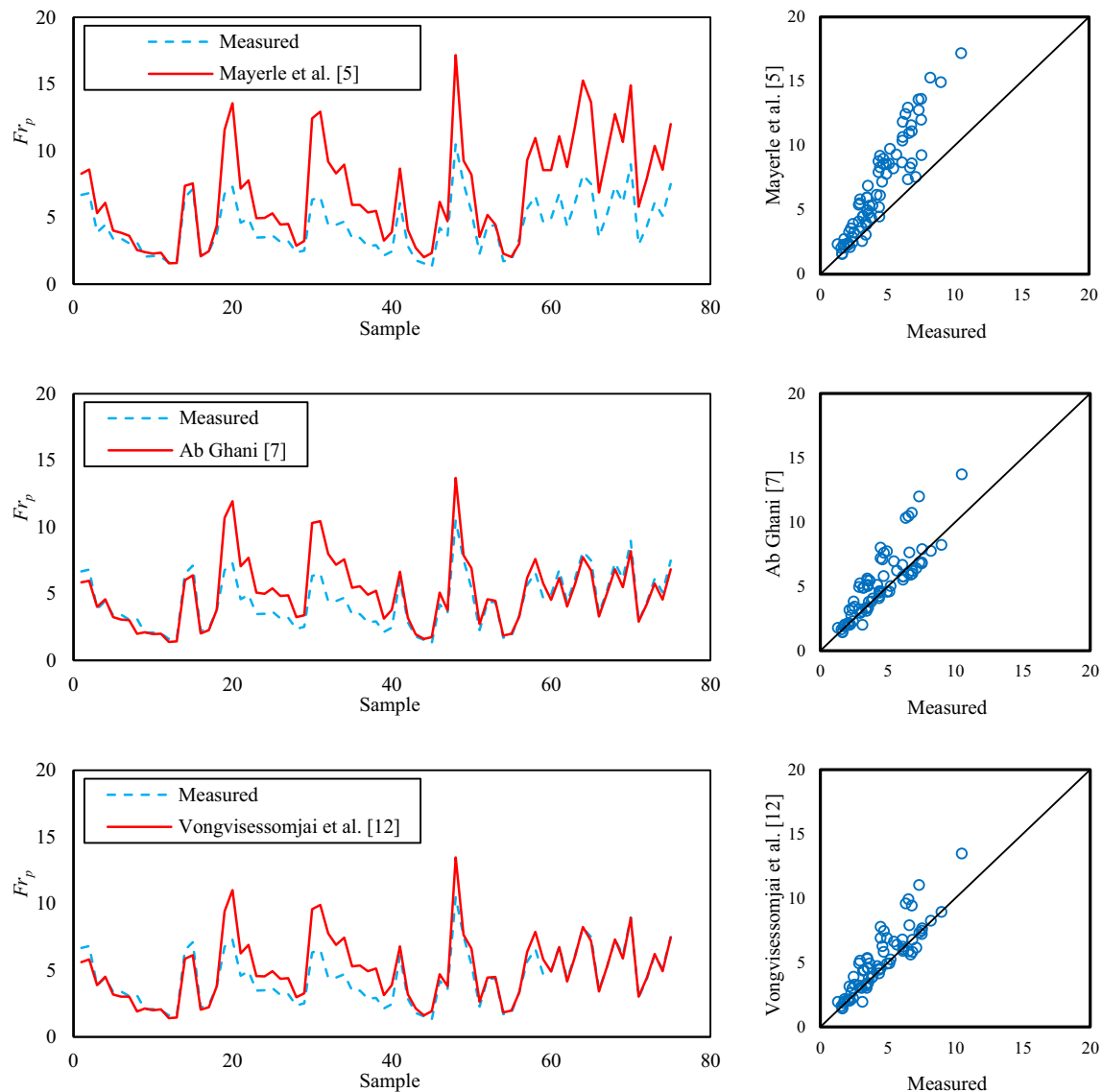


Fig. 3 Comparison of measured and calculated Fr_p by of Mayerle et al. [5], Ab Ghani [7] and Vongvisessomjai et al. [12] empirical models

Vongvisessomjai et al. [12] is found to perform better than other conventional regression models which may be attributed to utilizing a lower number of input parameters in comparison with the models in Mayerle et al. [5] and Ab Ghani [7]. In addition, the model in Mayerle et al. [5] provides inferior results with CC and $RMSE$ values of 0.59 and 3.14, respectively.

Table 7 illustrates that proposed approach provides a better performance in terms of DR with a mean of 1.02 close to unity and provides a superior performance as compared with the seven other machine learning techniques of GR, DT, MARS, GS-GMDH, GEP, FCM-ANFIS and ELM. The classical regression models typically over-estimate Fr_p where Vongvisessomjai et al. [12] gives more accurate results than Mayerle et al. [5] and Ab Ghani [7]

outcomes. The higher accuracy of the proposed approach is verified by the fact that 80% of the model outputs fall within the range of 0.80–1.20 of the measured Fr_p . The proposed approach gives DR results of 92%, for the range of 0.70–1.30, demonstrating that the suggested model in this study fits appropriately to the measured data. The inferior performances of the conventional regression models are also observed in terms of DR within the ranges of 0.80–1.20 and 0.70–1.30 of the measured particle Froude numbers. As a final conclusion of the statistical analysis based on DR , it should be mentioned that in the proposed approach particle Froude number is estimated more precisely with a DR of 100% within the range of 0.50–1.50 of the measured values.

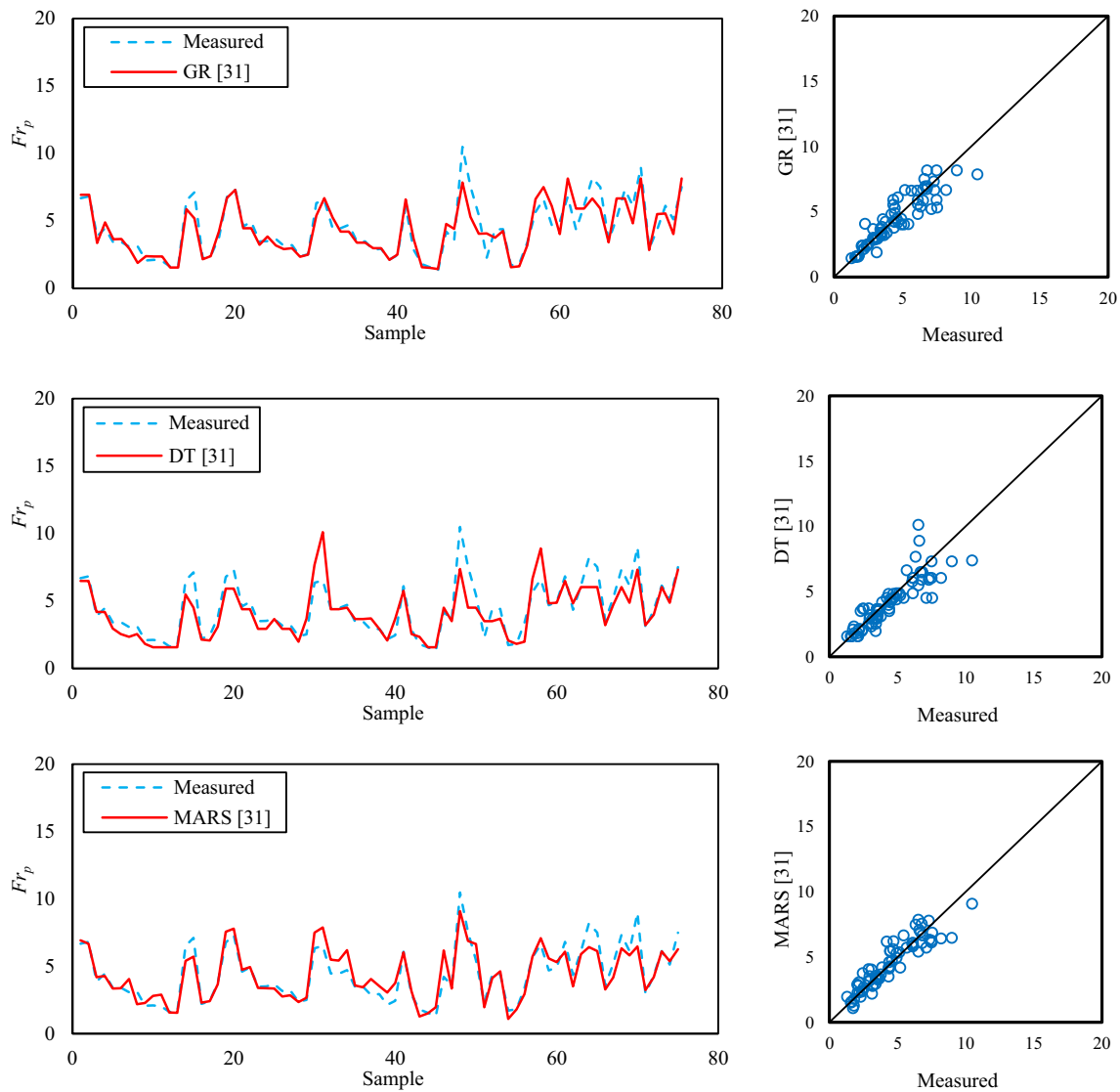


Fig. 4 Comparison of measured and calculated Fr_p by GR, DT and MARS models [31]

Figures 2, 3, 4 and 5 depict a comparison of the measured and calculated Fr_p through the scatter plot of goodness-of-fit for KRR, conventional regression models and GR, DT, MARS, GS-GMDH, GEP, FCM-ANFIS and ELM machine learning techniques. It may be seen from Fig. 2 that the estimated particle Froude numbers for the proposed approach match well with their measured counterparts where the results approximately remain on the bisector line with a small scatter. This in turn confirms the superior performance of the proposed approach without noticeable overestimation or underestimation. From Figs. 3, 4 and 5, it may be concluded that the conventional models of Mayerle et al. [5], Ab Ghani [7] and Vongvisessomjai et al. [12] significantly overpredict Fr_p with a large scatter. Although Ab Ghani [7] and Vongvisessomjai et al.'s [12] models outperform Mayerle et al. [5], their performances

are not as good as the proposed approach in this study. Similar to the results obtained based on statistical performance indices given in Tables 6 and 7, KRR outperforms seven machine learning techniques of GR, DT, MARS, GS-GMDH, GEP, FCM-ANFIS and ELM as shown Figs. 2 and 4 and 5. Among the machine learning techniques taken from the literature, MARS and GS-GMDH provide acceptable results and can be considered as alternatives to the developed model in this study.

6 Discussion

As noted earlier, the proposed approach in this study corresponds to a linear regression model constructed in a different space (RKHS) other than the original space of the

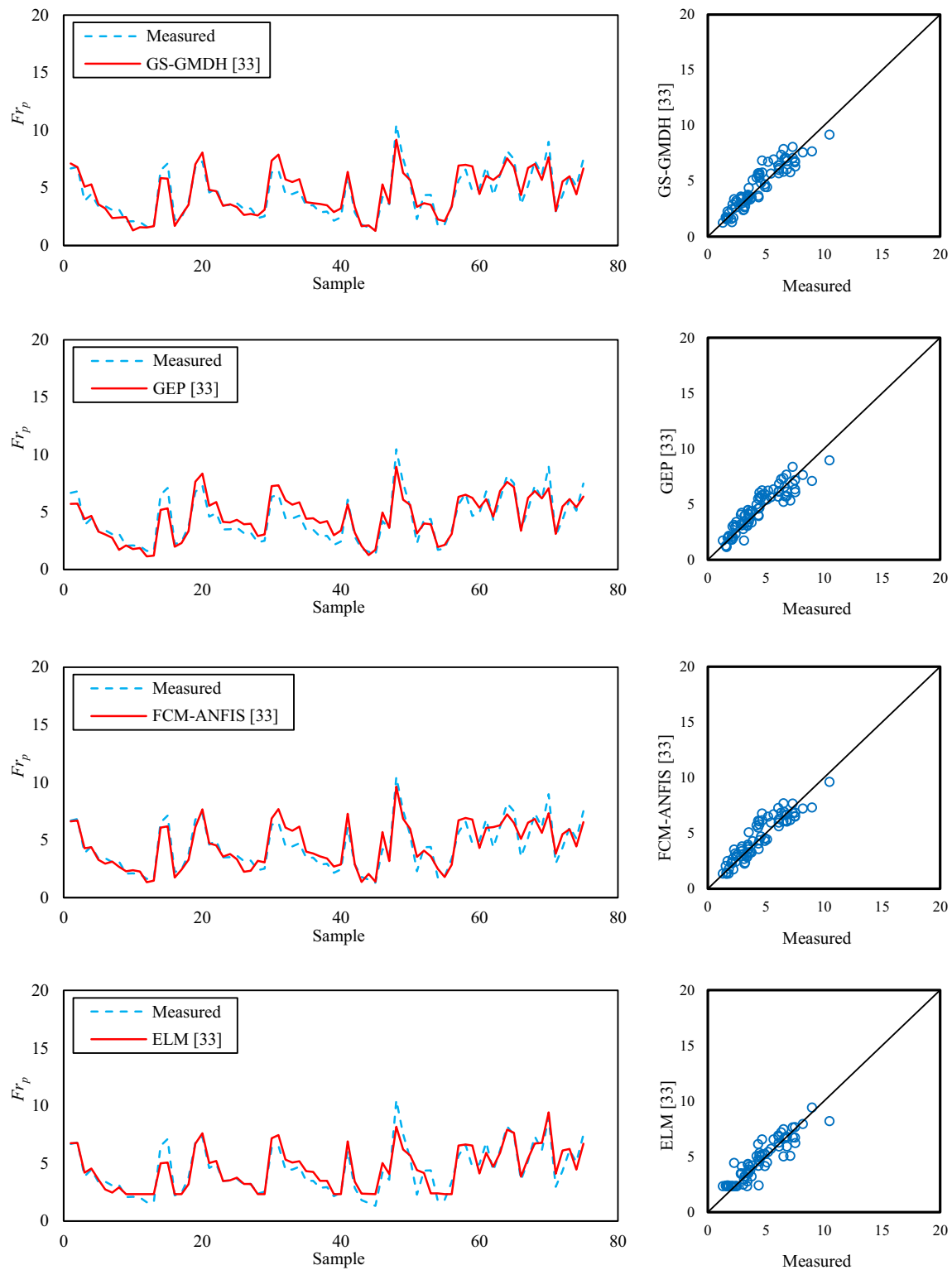


Fig. 5 Comparison of measured and calculated Fr_p by GS-GMDH, GEP, FCM-ANFIS and ELM models [33]

data characterized by a nonlinear kernel function. This in turn facilitates the estimation of a more accurate regression model when the data is highly nonlinear. A number of

different design choices should be considered when formulating the proposed approach. The first one being the choice of the kernel function. In the current study, the RBF

kernel function is considered for this purpose. Nevertheless, other choices may also be considered which are left as a subject of future investigation. As mentioned earlier, in this study, the regularization parameter controlling the flexibility of the model while preventing possible over-fitting to the training data is determined via cross-validation in training phase. Accordingly, it requires a large number of training observations for parameter tuning. While in this work the total number of training samples was set to 300, increasing the number of training observations is expected to yield better estimates for the regularization parameter. Moreover, any increase in the number of training samples is also expected to yield a more robust model. A further observation regarding regression in the kernel space is the choice of the regularization imposed on the function. In this work, we have opted for the widely used Tikhonov regularization. Nevertheless, other choices including sparsity of the solution may be also considered in the future.

Based on aforementioned statistical analysis, the proposed approach outperforms other commonly used regression models as well as some other commonly employed machine learning techniques of GR, DT, MARS, GS-GMDH, GEP, FCM-ANFIS and ELM. There is a considerable amount of work in the literature on modeling sediment transport applying the existing machine learning techniques. However, most of them use limited data ranges. The model developed in the current study was learned on a wide range of data taking from four well-known sources. More importantly, a novel approach where all the parameters characterizing the regression function are learned from the data is proposed. Based on the experimental analysis of the proposed approach and other well-known methods in the literature, it may be concluded that the proposed approach has a better capacity for sediment transport modeling.

This study used laboratory experimental data for sediment transport modeling which can be considered as main limitation of the study in the real cases. In this regard, utilizing field data is of importance to construct more reliable practical models. As the directions for future investigation, one may consider other regularization schemes in the context of proposed approach. One possibility in this direction is that of imposing a sparsity constraint on the solution. Other future research directions include evaluation of different kernel functions or learning the kernel function as a convex combination of multiple base kernels. Evaluation of the proposed approach on other environmental modeling problems may be considered as an alternative further direction for investigation.

7 Conclusions

In the current study, the regularized kernel ridge regression (KRR) technique is proposed as a robust method for sediment transport modeling. It was illustrated that the KRR technique corresponds to a linear regression in the RKHS. For the sake of preventing an over-fitting to the data, a *regularized* variant of the model was employed where the regularization parameter was determined via cross-validation in the training phase. In the formulation of the KRR approach, an RBF kernel was deployed the parameter of which was directly learned from the data via a learning approach based on gradient descent. Moreover, it was shown that training the model may be performed more efficiently via a Cholesky decomposition and forward-back substitution. In order to develop a reliable model, a wide range of laboratory data were utilized for the modeling purposes. The sediment, fluid, flow and channel characteristic are considered for modeling by incorporating the dimensionless parameters of channel friction factor, grain size parameter, sediment concentration and relative particle size to compute the particle Froude number. By comparing the accuracy of the proposed approach with the available models it was revealed that the proposed methodology performs better than other models available in the literature. More specifically, the proposed approach was shown to outperform some other machine learning techniques including GR, DT, MARS, GS-GMDH, GEP, FCM-ANFIS and ELM. More importantly, in the proposed approach, the non-deposition particle Froude number was estimated precisely with mean *DR* and *CC* values close to unity and lower *RMSE* value without noticeable underestimation or overestimation as compared with other available models. In the context of environmental and water resources engineering studies, this study proposed the use of the modified regularized kernel regression for the first time. It is expected that other environmental modeling problems may equally benefit from a similar modeling technique. Future directions of research motivated by the current study include examination of alternative regularization techniques for the KRR approach as well as evaluation of the current as well as other kernel-based techniques for environmental modeling purposes.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

- De Sutter R, Rushforth P, Tait S, Huygens M, Verhoeven R, Saul A (2003) Validation of existing bed load transport formulas using in-sewer sediment. *J Hydraul Eng* 129(4):325–333
- Safari MJS, Mohammadi M, Ab Ghani A (2018) Experimental studies of self-cleansing drainage system design: a review. *J Pipeline Syst Eng* 9(4):04018017
- CIRIA (1986) Sediment movement in combined sewerage and storm-water drainage systems." Phase 1. Project report. CIRIA research project No. 336, London
- Safari MJS (2016) Self-cleansing drainage system design by incipient motion and incipient deposition-based models. PhD thesis, Istanbul Technical University, Turkey
- Mayerle R, Nalluri C, Novak P (1991) Sediment transport in rigid bed conveyances. *J Hydraul Res* 29(4):475–495
- May RWP (1993) Sediment transport in pipes and sewers with deposited beds. Technical Report, Hydraulic Research Ltd., Report SR 320, Wallingford, UK
- Ab Ghani A (1993) Sediment transport in sewers. PhD thesis, University of Newcastle upon Tyne
- Ota JJ (1999) Effect of particle size and gradation on sediment transport in storm sewers. PhD thesis, University of Newcastle Upon Tyne, UK
- Ota JJ, Nalluri C (2003) Urban storm sewer design: approach in consideration of sediments. *J Hydraul Eng* 129(4):291–297
- Ota JJ, Perrusquia GS (2013) Particle velocity and sediment transport at the limit of deposition in sewers. *Water Sci Technol* 67(5):959–967
- Butler D, May R, Ackers J (2003) Self-cleansing sewer design based on sediment transport principles. *J Hydraul Eng* 129(4):276–282
- Vongvisessomjai N, Tingsanchali T, Babel MS (2010) Non-deposition design criteria for sewers with part-full flow. *Urban Water J* 7(1):61–77
- Safari MJS, Aksoy H, Unal NE, Mohammadi M (2017) Non-deposition self-cleansing design criteria for drainage systems. *J Hydro-environ Res* 14:76–84
- Safari MJS, Aksoy H (2020) Experimental analysis for self-cleansing open channel design. *J Hydraul Res*. <https://doi.org/10.1080/00221686.2020.1780501>
- Ab Ghani A, Azamathulla HM (2010) Gene-expression programming for sediment transport in sewer pipe systems. *J Pipeline Syst Eng* 2(3):102–106
- Azamathulla HM, Ab Ghani A, Fei SY (2012) ANFIS-based approach for predicting sediment transport in clean sewer. *Appl Soft Comput* 12:1227–1230
- Ebtehaj I, Bonakdari H, Shamshirband S, Mohammadi K (2015) A combined support vector machine-wavelet transform model for prediction of sediment transport in sewer. *Flow Meas Instrum* 47:19–27
- Safari MJS, Aksoy H, Mohammadi M (2016) Artificial neural network and regression models for flow velocity at sediment incipient deposition. *J Hydrol* 541:1420–1429
- Qasem SN, Ebtehaj I, Bonakdari H (2017) Potential of radial basis function network with particle swarm optimization for prediction of sediment transport at the limit of deposition in a clean pipe, Sustai. *Water Resour Manag* 3:391–401
- Ebtehaj I, Bonakdari H, Zaji AH (2016) An expert system with radial basis function neural network based on decision trees for predicting sediment transport in sewers. *Water Sci Technol* 74(1):176–183
- Ebtehaj I, Bonakdari H, Shamshirband S, Ismail Z, Hashim R (2016) New approach to estimate velocity at limit of deposition in storm sewers using vector machine coupled with firefly algorithm. *J Pipeline Syst Eng* 20:04016018
- Ebtehaj I, Bonakdari H (2016) Assessment of evolutionary algorithms in predicting non-deposition sediment transport. *Urban Water J* 13:499–510
- Najafzadeh M, Bonakdari H (2016) Application of a neuro-fuzzy GMDH model for predicting the velocity at limit of deposition in storm sewers. *J Pipeline Syst Eng* 8:06016003
- Najafzadeh M, Laucelli DB, Zahiri A (2017) Application of model tree and evolutionary polynomial regression for evaluation of sediment transport in pipes. *KSCE J Civ Eng* 21:1956–1963
- 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* 17:537–551
- Roushangar K, Ghasempour R (2017) Estimation of bed load discharge in sewer pipes with different boundary conditions using an evolutionary algorithm. *Int J Sediment Res* 32(4):564–574
- Safari MJS, Shirzad A, Mohammadi M (2017) Sediment transport in deposited bed sewers: unified form of May's equations using the particle swarm optimization algorithm. *Water Sci Technol* 76(4):992–1000
- Safari MJS, Danandeh Mehr A (2018) Multigene genetic programming for sediment transport modeling in sewers at non-deposition with deposited bed condition. *Int J Sediment Res* 33(3):262–270
- Wan Mohtar WHM, Afan H, El-Shafie A, Bong CHJ, Ab Ghani A (2018) Influence of bed deposit in the prediction of incipient sediment motion in sewers using artificial neural networks. *Urban Water J* 15(4):296–302
- Safari MJS, Shirzad A (2019) Self-cleansing design of sewers: definition of the optimum deposited bed thickness. *Water Environ Res* 91(5):407–416
- Safari MJS (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
- Kargar K, Safari MJS, Mohammadi M, Samadianfard S (2019) Sediment transport modeling in open channels using neuro-fuzzy and gene expression programming techniques. *Water Sci. Tech.* 79(12):2318–2327
- Safari MJS, Ebtehaj I, Bonakdari H, Es-haghi MS (2019) Sediment transport modeling in rigid boundary open channels using generalize structure of group method of data handling. *J Hydrol* 577:123951
- Ebtehaj I, Bonakdari H, Safari MJS, Gharabaghi B, Zaji AH, Madavar HR, Khozani ZS, Es-haghi MS, Shishegaran A, Danandeh Mehr A (2020) Combination of sensitivity and uncertainty analyses for sediment transport modeling in sewer pipes. *Int J Sediment Res* 35(2):157–170
- Montes C, Berardi L, Kapelan Z, Saldarriaga J (2020) Predicting bedload sediment transport of non-cohesive material in sewer pipes using evolutionary polynomial regression–multi-objective genetic algorithm strategy. *Urban Water J* 17(2):154–162
- Cortes C, Vapnik V (1995) Support vector networks. *Mach Learn* 20:273–297
- An S, Liu W, Venkatesh S (2007) Face recognition using kernel ridge regression. In: IEEE conference on computer vision and pattern recognition, Minneapolis, MN, pp 1–7
- He J, Ding L, Jiang L, Ma L (2014) Kernel ridge regression classification. In: International joint conference on neural networks (IJCNN), Beijing, pp 2263–2267
- Goodfellow I, Bengio Y, Courville A (2016) Deep learning. MIT Press, Cambridge. ISBN 9780262035613
- Hofmann T, Schölkopf B, Smola AJ (2008) Kernel methods in machine learning. *Ann Stat* 36:1171–1220

41. Raudkivi AJ (1990) Loose boundary hydraulics. Pergamon Press, Oxford
42. Vanoni VA (2006) Sedimentation engineering. ASCE, New York
43. Novak P, Nalluri C (1984) Incipient motion of sediment particles over fixed beds. *J Hydraul Res* 22(3):181–197
44. Safari MJS, Aksoy H, Unal NE, Mohammadi M (2017) Experimental analysis of sediment incipient motion in rigid boundary open channels. *Environ Fluid Mech* 17(6):1281–1298
45. Craven JP (1953) The transportation of sand in pipes—full pipe flow. In: Proceedings. Of the fifth hydraulics conference, bulletin 34, State University of Iowa Studies in Engineering, Iowa State University, Iowa
46. Ambrose HH (1953) The transportation of sand in pipes free surface flow. In: Proceeding of the fifth hydraulic conference, bulletin 34, State University of Iowa Studies in Engineering, Iowa State University, Iowa
47. Durand R (1953) Basic relationships of the transportation of solids in pipes. In: Experimental research, international association for hydraulic research, fifth congress, Minneapolis
48. Durand R, Condolois EI (1956) Donneés techniques sur le refoulement hydraulique des matériaux solides en conduite. *Rev. L'Industrie Mineral*, Special Number F, June
49. Robinson MP, Graf WH (1972) Pipelining of low concentration sand-water mixtures. *J Hydr Eng Div ASCE* 98 (HY7) 1221–1240
50. Novak P, Nalluri C (1975) Sediment transport in smooth fixed bed channels. *J Hydraul Div ASCE* 101(HY9):1139–1154
51. May RWP (1982) Sediment transport in sewers. Report No. IT 222, Hydraulic Research Station, Wallingford
52. May RWP, Brown PM, Hare GR, Jones KD (1989) Self-cleansing conditions for sewers carrying sediment. Report SR, 221
53. Mayerle R (1988) Sediment transport in rigid boundary channels. PhD thesis, University of Newcastle upon Tyne, UK
54. Kithsiri MMAU (1990) Sediment transport in rectangular channels with rough rigid beds. PhD thesis, University of Newcastle upon Tyne (1990)
55. Ackers JC, Butler D, May RWP (1996) Design of sewers to control sediment problems. Construction Industry Research and Information Association (CIRIA) Rep. No. 141, London. pp 1–181
56. May RW, Ackers JC, Butler D, John S (1996) Development of design methodology for self-cleansing sewers. *Water Sci Technol* 33(9):195–205
57. Alpaydin E (2010) Introduction to machine learning 2e. MIT Press, Cambridge
58. Baudat G, Anouar F (2000) Generalized discriminant analysis using a kernel approach. *Neural Comput* 12(10):2385–2404
59. Cortes C, Vapnik V (1995) Support-vector networks. *Mach Learn* 20(3):273–297
60. Mika S, Ratsch G, Weston J, Scholkopf B, Mullers KR (1999) Fisher discriminant analysis with kernels. In: Neural networks for signal processing IX: Proceedings of the 1999 IEEE signal processing society workshop (cat. No. 98th8468). IEEE. (1999), pp 41–48
61. Scholkopf B, Smola AJ (2001) Learning with kernels: support vector machines, regularization, optimization, and beyond. MIT Press, Cambridge
62. Ethem A (2010) Introduction to machine learning, 2nd edn. MIT Press, Cambridge
63. Berline A, Thomas C (2004) Reproducing kernel hilbert spaces in probability and statistics. Kluwer Academic Publishers, Dordrecht
64. Tikhonov AN (1943) Об устойчивости обратных задач [On the stability of inverse problems]
65. Saunders C, Gammerman A, Vovk V (1998) Ridge regression learning algorithm in dual variables. In: Proceedings of the 15th international conference on machine learning (ICML98), Madison-Wisconsin. pp 515–521
66. Golub GH, Van Loan CF (1996) Matrix computations, 3rd edn. Johns Hopkins, Baltimore. ISBN 978-0-8018-5414-9
67. Stewart GW (1998) Matrix algorithms: basic decompositions, vol 1. SIAM, Philadelphia
68. Cai D, He X, Han J (2011) Speed up kernel discriminant analysis. *VLDB J* 20(1):21–33

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.