

Bed material load estimation in channels using machine learning and meta-heuristic methods

Shahram Sahraei, Mohammad Reza Alizadeh, Nasser Talebbeydokhti and Maryam Dehghani

ABSTRACT

This study is trying to develop an alternative approach to the issues of sediment transport simulation. A machine learning method, named least square support vector regression (LSSVR) and a meta-heuristic approach, called particle swarm optimization (PSO) algorithm are used to estimate bed material load transport rate. PSO algorithm is utilized to calibrate the parameters involved in the model to facilitate a desirable simulation by LSSVR. Implementing on a set of laboratory and field data, the model is capable of performing more satisfactorily in comparison to candidate traditional methods. Similarly, the proposed method has a better performance than a specific version of decision tree method. To enhance the model, the variables are scaled in logarithmic form, leading to an improvement in the results. Thus, the proposed model can be an efficient alternative to conventional approaches for the simulation of bed material load transport rates providing comparable accuracy.

Key words | bed material load, particle swarm optimization, river channels, similarity theory, support vector regression

Shahram Sahraei (corresponding author)
Civil Engineering Department, Faculty of
Engineering,
University of Manitoba,
Winnipeg, MB,
Canada
E-mail: sahraeis@myumanitoba.ca

Mohammad Reza Alizadeh
Nasser Talebbeydokhti
Maryam Dehghani
Department of Civil and Environmental
Engineering, School of Engineering,
Shiraz University,
Shiraz,
Iran

NOTATION

C_t Concentration (ppm by weight)
 D Channel depth (m)
 D_* Dimensionless particle diameter (-)
 d_{50} Median particle diameter (m)
 Fr Densimetric Froude number (-)
 G Gravity acceleration ($m \cdot s^{-2}$)
 Q_w Water discharge ($m^3 \cdot s^{-1}$)
 R_* Shear Reynolds number (-)
 Re Reynolds number (-)
 S_f Channel slope (-)
 SG Specific gravity (-)
 T Temperature ($^{\circ}C$)
 u_* Shear velocity ($m \cdot s^{-1}$)
 V Flow velocity ($m \cdot s^{-1}$)
 V_{cr} Critical velocity ($m \cdot s^{-1}$)

W Channel width (m)
 τ_b Bed shear stress ($N \cdot m^{-2}$)
 τ_c Critical shear stress ($N \cdot m^{-2}$)
 γ_w Water specific weight ($N \cdot m^{-3}$)
 ω Sediment fall velocity ($m \cdot s^{-1}$)

INTRODUCTION

Sediment transport science deals with the interactions between flow and sediment particles. Sediment transport and sedimentation lead to the development of sedimentary bars, resulting in a decline in flood flow transportation, channel sedimentation, dam lifetime and reservoir storage capacity, corrosion in river structure installations, and

contaminant transport. Sediment transport is a highly complex phenomenon often subjected to theoretical, empirical, and semi-empirical treatment. Most theory-driven approaches, such as those by Meyer-Peter & Muller (1948), Bagnold (1966), and Einstein (1942) are based on ideal and simplified assumptions in which sediment transport rate is determined according to a small number of dominant factors. However, the obtained results from various approaches based on statistical analyses differ dramatically from each other as well as from field observations. Due to the nonlinearity and complexity of sediment transport and the variability of particle diameters, as well as the various forms the channel bed can take, data-driven methods may provide a better response to such problems.

Computational intelligence techniques, such as artificial neural networks (ANNs), fuzzy logic, genetic programming (GP), or a combination of these techniques, have been successfully used to solve complex problems associated with water resources management. Among them, ANN is the most widely used learning method. This method has two disadvantages. In the first place, its structure needs to be defined prior to training. Second, the regularization techniques, such as early termination and training with noise are somewhat restricted. Furthermore, ANNs can also be trapped in local minima in the process of model training (Smola 1996). These drawbacks do not exist in a support vector machine (SVM) approach. Therefore, in this study, a combination of least square support vector regression (LSSVR) and particle swarm optimization (PSO) algorithm is applied to the most important variables in sediment transport problems including channel geometry and hydraulic properties. In general, the calibration of LSSVR and kernel parameters is a challenging process. As such, PSO algorithm is used for the calibration of these parameters.

SVM is a newly developed learning technique that has gained enormous popularity in the areas of classification, pattern recognition, and regression. SVM works based on the structural risk minimization principle offering greater generalizability. SVM was initially introduced by Vapnik & Lerner (1963) and extended to nonlinear conditions by Cortes & Vapnik (1995). A great deal of previous research into SVM application in water resources has focused on flood forecasting, some of which are mentioned as follows. Having investigated a case study application of SVM in

flood prediction, Yu *et al.* (2006) reported the model accuracy in predicting water surface levels 1 to 6 hours after the flood. Sivapragasam *et al.* (2001) introduced a useful prediction technique based on singular spectrum analysis coupled with SVM. Compared with nonlinear prediction methods, their hybrid model enjoyed greater accuracy when it came to predicting hydrologic parameters. Yilin *et al.* (2006) used a combination of SVM and the shuffled complex evolution (SCE) optimization algorithm in forecasting long-term discharge and concluded that SVM satisfactorily predicts long-term discharge. Similarly, Asefa *et al.* (2006) offered an SVM-based method for hourly and seasonal flow prediction. In the same vein, Shuquan & Lijun (2007) used least square support vector machine (LSSVM) to predict both mid-term and long-term runoff. They concluded that LSSVM functions more effectively in comparison to ANN. Han *et al.* (2007) conducted research on a case study flood flow analysis and found that SVM excels any other benchmarking models that are being used for flood forecasting. Huang *et al.* (2010) used SVM to comprehensively assess flood disaster loss. Noori *et al.* (2011) investigated the impact of three input selection techniques on SVM efficiency for monthly flow prediction, showing that the pre-processing of input variables through principal component analysis and Gamma tests can improve SVM performance. In another investigation into the application of SVM and regression tree methods over a case study research on outflow forecasting, Sahraei *et al.* (2015) found that support vector regression (SVR) surpasses the regression tree in terms of outlet discharge prediction.

There is also a large volume of published studies describing the role of machine learning methods in estimating sediment discharge, one of which is the research done by Jain (2001). He developed an ANN-based sediment rating curve for two gauging sites on the Mississippi River and demonstrated that his model is more precise than the conventional sediment rating curve techniques in terms of sediment load calculation. Nagy *et al.* (2002) applied the ANN model to datasets measured in different rivers to develop an estimator for natural sediment discharge. Their results indicated that the new approach yielded better results compared to several commonly used formulas. Likewise, Bhattacharya *et al.* (2007), in a comparison with the existing approaches, asserted that adopting two machine

learning methods, namely, ANN and MT, for estimating bed load and total load transport rate can enhance the sediment model accuracy to some extent. However, the existing literature on the application of SVR over sediment load is extensive and focuses particularly on either meteorological data or streamflow as input variables for their model (Kisi et al. 2012; Afan et al. 2016). Having utilized SVR to forecast the suspended sediment concentration of rivers in the USA in 2006, Çimen showed that their method performs better than fuzzy logic and gradual evolution approaches (Çimen 2006). In the same way, Azamathulla et al. (2010) found that SVR technique enjoys superior performance in comparison to existing sediment transport methods in three rivers in Malaysia. Misra et al. (2009) demonstrated that simulated daily, weekly, and monthly runoff and sediment yield by SVR was significantly better than those yielded by ANN at model training, calibration, and verification.

Although some research has been carried out on SVR, none considered estimation of bed material load as an approach applicable to a wide range of flow and sediment characteristics. The main goal of this paper is to present a reliable approach for the estimation of bed material load if there is no concern about the data availability. The remainder of this paper is organized as follows. The next section details underlying methodology of this research and describes how LSSVR and PSO methods work. This is followed by a section presenting a discussion regarding the method's suitability to estimation of the sediment transport rate by comparing with the traditional methods. The final section presents the conclusions of the study.

MATERIALS AND METHODS

Data description and dimensional analysis

In this study, a combination of experimental and field datasets is considered. Each data point includes flow discharge (Q_w), channel width (W), channel depth (D), water surface slope (S_f), water temperature (T), median particle diameter (d_{50}), specific gravity (SG), and bed material concentration by weight (C_t). The source of data as well as their averages and standard deviations are presented in Table 1. It is preferred to use kinematic viscosity for two reasons (Yang

1996): (1) most of the traditional methods consider viscosity as decision variable rather than temperature and (2) variables are converted to dimensionless variables prior to modeling. As a result, it substitutes the temperature using Equation (1):

$$\nu = \frac{1.792 \times 10^{-6}}{1.0 + 0.0337 T + 0.000221 T^2} \quad (1)$$

where T is the water temperature in centigrade and ν is the kinematic viscosity in square meters per second.

In addition, it is required to determine and include critical shear stress, particle fall velocity, shear velocity, and critical velocity when dimensional analysis is carried out. Therefore, they have been calculated by equations in the literature (Yang 1996) and then applied to the PSO-LSSVR model as input components. Prasad (1991) showed that bed shear stress depends on the aspect ratio of the channel. They suggested that for aspect ratios of $W/D > 4$, the bed shear stress can be calculated using flow depth. Williams (1970) focused on smooth wall correction to obtain the bed shear stress for measured data with an aspect ratio of $W/D > 4$, which has been converted to SI unit system as follows:

$$\frac{\tau_b}{\gamma D S_f} = \frac{W^2}{W^2 + 0.055 D} \quad (2)$$

where τ_b is the bed shear stress and S_f is surface slope. Critical shear stress is the bed shear stress in the initiation of motion condition and Rao & Sreenivasulu (2006) have developed the following equation to approximately compute Shields' stress:

$$\frac{\tau_c}{(\gamma_s - \gamma) d_{50}} = \frac{0.12}{R_*^{1/3}} \exp(-0.16 R_*) + 0.05 \exp(-10/R_*) \quad (3)$$

where $R_* = (u_* d_{50})/\nu$ is shear Reynolds' number and u_* denotes shear velocity.

Once the required variables are extracted, it is first necessary to carry out dimensional analysis on all response and predictor variables shown in Equation (4):

$$C_t = f(Q, W, D, S_f, \nu, d_{50}, (SG - 1), \tau_b - \tau_c, g, \gamma_w) \quad (4)$$

Table 1 | Source of data and corresponding average and deviation for input variables (Brownlie 1981)

Experimental data ^a	Number of data	Index	Discharge (10 ⁻⁴ c-m/s)	Width (m)	Depth (m)	Surface slope (10 ⁻⁵ m/m)	Kinematic viscosity (10 ⁻⁸ m ² /s)	Median particle diameter (mm)	Specific gravity
Barton & Lin (1955)	30	Average	1,340	1.219	0.192	119	95.7	0.18	2.65
		Standard deviation	754	0	0.065	44	6.74	0	0
Vanoni & Brooks (1957)	21	Average	91	0.267	0.071	245	91.4	0.145	2.65
		Standard deviation	28	0	0.012	52	8.31	0	0
Kennedy & Brooks (1963)	9	Average	396	0.851	0.102	181.1	94.8	0.142	2.65
		Standard deviation	1.37	0	0.034	55.7	7.55	0	0
Nomicos (1957)	26	Average	88.3	0.267	0.073	236	90.9	0.149	2.65
		Standard deviation	32.5	0	1.18E-03	40	8.92	0.005	0
Soni (1980)	23	Average	55.9	0.2	0.065	406	83.3	0.32	2.65
		Standard deviation	20	0	0.021	175	1.97	0	0
Stein (1965)	57	Average	2,685	1.219	0.218	434	93.1	0.4	2.65
		Standard deviation	1,064	0	0.072	288	4.75	0	0
Taylor & Vanoni (1972)	6	Average	479	0.657	0.090	198	76.1	0.198	2.65
		Standard deviation	321	0.302	0.018	8	14.5	0.046	0
Vanoni & Brooks (1957)	15	Average	409	0.851	0.110	177	94.2	0.137	2.65
		Standard deviation	252	0	0.047	83.7	5.42	0	0
Willis <i>et al.</i> (1972)	96	Average	2,579	1.219	0.243	97.8	93.3	0.1	2.65
		Standard deviation	1,045	0	0.062	38.9	4.90	0	0
Willis (1979)	32	Average	307	0.360	0.128	394	96.4	0.54	2.65
		Standard deviation	109	0	0.015	221	17.2	0	0
Field data									
Brownlie (1981)	289	Average	488,790	58.264	0.649	119	114	0.384	2.65
		Standard deviation	667,180	25.879	0.486	33.7	21.4	0.184	0
Toffaletti (1968)	38	Average	1,433,290	120.64	0.825	80.5	103	0.304	2.65
		Standard deviation	719,450	46.119	0.280	3.5	7.49	0.031	0
Sum	642	Average	305,600	33.799	0.424	177	103	0.305	2.65
		Standard deviation	604,850	40.906	0.422	159	18.8	0.179	0

^aExperimental data are provided in Brownlie (1981).

As can be observed from Equation (4), sediment concentration serves as a dependent variable needing to be determined through simulation and the remainder work as explanatory variables. With this information, the dimensionless variables are then calculated as follows:

$$C_t = f\left(Fr, S_f, \frac{d_{50}}{D}, \frac{W}{D}, Re, (SG - 1), \frac{\tau_b - \tau_c}{\gamma_w(SG - 1)d_{50}}\right) \quad (5)$$

where C_t is bed material concentration by weight, Fr is the densimetric Froude number, S_f is surface slope, d_{50}/D is relative bed roughness, W/D is aspect ratio, Re is Reynolds' number, and $(\tau_b - \tau_c)/(\gamma_w(SG - 1)d_{50})$ is equivalent to the

dimensionless Shields' parameter (Shields 1936). As the term $(SG - 1)$ is nearly constant, it is excluded from the inputs. Details of the average and standard deviation (SD) of the dimensionless input and output variables can be observed in Table 2.

Support vector machine

SVM, in general, is a learning method that is utilized for the purpose of classification and regression in order to minimize error in classification and/or fitness functions. The method is based on constrained optimization theory, which is

Table 2 | Average and standard deviation of the dimensionless input and output parameters

Variables	C_t	Densimetric Froude	S_f	$\frac{d_{50}}{D}$	$\frac{W}{D}$	Re	$\frac{\tau_b - \tau_c}{\gamma_w(SG - 1)d_{50}}$
Average	2,580.6	16.96	0.00177	0.00131	65.78	467,941	2.654
Standard deviation	3,361.1	9.60	0.00159	0.00139	83.4	785,637	2.927

based on the structural risk minimization principle, and gives a general optimization response (Vapnik 1998). The aim of SVR is to diagnose an $f(X)$ function for training patterns (X) so that it has maximum margin from the training target values (Y). In other words, SVR is a model that fits a tube to the data with a radius of ϵ so that minimum error occurs in the test dataset. Let us first assume that a training set (T) is considered as follows to compare with its prediction values:

$$T = \{(\mathbf{X}_1, y_1), (\mathbf{X}_2, y_2), \dots, (\mathbf{X}_N, y_N)\} \ni \mathbf{X} \in IR^m, y \in IR \quad (6)$$

where \mathbf{X}_i is an m -dimensional vector, each of which belongs to a specific decision variable. y_i is the associated output variable and N is the number of samples. Second, Vapnik asserts that the term representing the complexity of set of functions must be minimized in order to minimize the test error according to Equation (7). Therefore, the SVR method uses a set of linear functions in the form of $f(\mathbf{X}) = \mathbf{w} \cdot \mathbf{X} + b$ (\mathbf{w} is the weight vector and b is a bias value) for prediction. Trivially, the predictions are different from the measurements and, in some cases, it is impossible to consider error values lower than ϵ . As such, a deviation from ϵ , known as slack variable (ξ_i^+ and ξ_i^-), is defined. The error value is then minimized using Equation (8). According to Equation (8), maximum margin requires minimizing the norm of weight vector:

$$\text{Testing Error} \leq \text{Training Error} + \text{Complexity of set of models} \quad (7)$$

$$\begin{aligned} \text{Minimize: } & \frac{1}{2}(\mathbf{w} \cdot \mathbf{w}) + C \sum_{i=1}^N (\xi_i^+ + \xi_i^-) \\ \text{Constraints: } & \begin{cases} \mathbf{w} \cdot \mathbf{X}_i + b - y_i \leq \epsilon + \xi_i^+ & i = 1, 2, 3, \dots, N \\ y_i - (\mathbf{w} \cdot \mathbf{X}_i + b) \leq \epsilon + \xi_i^- & i = 1, 2, 3, \dots, N \\ \xi_i^+ \geq 0, \quad \xi_i^- \geq 0 & i = 1, 2, 3, \dots, N \end{cases} \end{aligned} \quad (8)$$

where ϵ determines the estimated tube range and C controls the deviation error more than ϵ , both of which are above zero.

Generally, linear regression of data seems impossible since most phenomena are nonlinear in nature. In this circumstance, nonlinear SVR is performed. Input vectors (\mathbf{X}) are mapped into a space with higher dimensions so that linear regression can be performed in the mapped space. This is accomplished by defining feature function as can be seen in Equation (9). The constrained optimization formulation in Equation (8) can be easily converted to a Lagrangian form. As a result, the inner product of feature functions is directly substituted by the Kernel function under Mercer conditions since obtaining feature functions is a difficult task for different problems. By taking derivatives, the Lagrangian function is finally converted to a quadratic optimization form and Lagrangian multipliers (λ) are obtained. Vectors having $(\lambda_i^- - \lambda_i^+) \neq 0$ represent support vectors. Finally, making use of Equation (10), the estimation function is constructed by means of support vectors:

$$\mathbf{X} \Rightarrow \phi(\mathbf{X}) \therefore \mathbf{X}_i \cdot \mathbf{X}_j \Rightarrow \phi(\mathbf{X}_i) \cdot \phi(\mathbf{X}_j) \quad (9)$$

$$f(\mathbf{X}) = \sum_{i=1}^N (\lambda_i^- - \lambda_i^+) K(\mathbf{X}_i, \mathbf{X}) + b \quad (10)$$

In a similar way, the least square error version of SVR can be written by the reformulation of Equation (8) as follows (Suykens et al. 2002):

$$\text{Minimize: } \frac{1}{2}(\mathbf{w} \cdot \mathbf{w}) + \frac{\gamma}{2} \sum_{i=1}^N (e_i^2) \quad (11)$$

$$\text{Subject to: } y_i = (\mathbf{w} \cdot \phi(\mathbf{X}_i) + b) + e_i \quad i = 1, 2, 3, \dots, N$$

where γ is the regularization parameter that creates a trade-off between the uniformity of fitness curve and fitness error

minimization. The Lagrangian function of the constrained optimization problem can be written as follows:

$$L(\mathbf{w}, b, e, \lambda) = \frac{1}{2}(\mathbf{w} \cdot \mathbf{w}) + \frac{\gamma}{2} \sum_{i=1}^N (e_i^2) - \sum_{i=1}^N \lambda_i (\mathbf{w} \cdot \phi(\mathbf{X}_i) + b + e_i - y_i) \quad (12)$$

By derivation of Equation (12) with respect to \mathbf{w} , b , e , and λ , thereby substituting the terms in the equation, the optimization problem is converted to a linear programming and multipliers are obtained.

PSO algorithm

The original algorithm of PSO is inspired by the social behavior of birds trying to find their food in nature. This algorithm works based on the interaction and co-operation of the birds searching for food within an area (Kennedy & Eberhart 1995). The birds follow the one that is nearest to the food while simultaneously taking advantage of their earlier experiences finding the food (Kennedy & Eberhart 2001). The food can be found only in one point in the search space that the birds are not aware of. Each solution acts like a bird, which is called a particle in this paper. Each particle enjoys a merit value by means of optimizing the objective function. PSO first generates an initial random solution each of which has an n -dimensional position (\mathbf{X}) where n is the number of decision variables. As well, it assigns a velocity vector (\mathbf{V}) between the maximum and minimum allowable velocities to each particle. In order to produce a new population for the next generation and have progress in terms of convergence, PSO requires specifying some parameters, including inertia weight (w), inertia weight reduction factor (α), minimum inertia weight (w_{min}), personal and global acceleration values (c_1 and c_2), and maximum velocity reduction factor (β). In every generation, the personal (\mathbf{P}_p , the best position which every single particle has ever had since the beginning to the current generation) and global (\mathbf{P}_g , the best ever position among all particles) best particles are found to update the population for the next iteration.

The position of each particle can be improved using Equations (13) and (14):

$$\mathbf{V}_{k+1}^i = w\mathbf{V}_k^i + c_1 r_{1,k} (\mathbf{P}_k^p - \mathbf{X}_k^i) + c_2 r_{2,k} (\mathbf{P}_k^g - \mathbf{X}_k^i) \quad (13)$$

$$\mathbf{X}_{k+1}^i = \mathbf{X}_k^i + \mathbf{V}_{k+1}^i \quad (14)$$

where w is recommended to be taken between 0.8 and 1.4, $r_{1,k}$ and $r_{2,k}$ are the random uniform numbers between zero and one, and k is the iteration number. The velocity values are limited to $[-\mathbf{V}_{max}, \mathbf{V}_{max}]$ to guarantee the convergence of PSO to the optimal objective and either when there is no significant change in the value of objective function (Equation (15)) or after a specified number of generation the algorithm terminates:

$$\frac{F_{max} - F_{min}}{F_{min}} \leq \varepsilon \quad (15)$$

Model tree (M5P)

Model tree was initially developed by Quinlan (1992) and afterwards the methodology was modified so substantially that a new version called M5P was introduced (Wang & Witten 1997). M5P has been found very effective and efficient as a learning approach for the prediction of real values in large datasets (Etemad-Shahidi & Mahjoobi 2009). It recursively divides samples into two subsets in each node such that the intra-subset dissimilarity or inter-subset similarity of instances is minimized. Each attribute of all the values that reach a node from the root through the branch is examined for splitting purposes based on the calculation of the standard deviation of the values at that node as a dissimilarity measure. The attribute resulting in a maximum decrease of errors is considered as divider. Finally, the split halts if few examples remain or the values in nodes vary insignificantly (Chen 2006). Equation (16) represents an index for the standard deviation reduction (SDR):

$$SDR = S_d(T) - \sum_i \frac{|T_i|}{|T|} \times S_d(T_i) \quad (16)$$

Here, T denotes the subset that reached the node. T_i represents the new subset being created due to division based

on the selected attribute and S_d is the standard deviation (Wang & Witten 1997). Solomatine and Dulal were among the first to demonstrate the model trees as a substitute for rainfall-runoff modeling. Additional details can be found in the works of Solomatine & Dulal (2003).

Traditional methods

In spite of the complexity of sediment transport phenomenon, most conventional methods work on some idealized or simplified assumptions (Yang 1996). Currently, some of the most commonly used methods for measuring total bed material load are by Engelund & Hansen (1972), Ackers & White (1973), Graf (1971), and Yang (1972). Due to its inclusion in the modeling, Yang's (1972) methodology is briefly explained in the following. However, Yang's (1996) textbook gives complete details about the rationale for development of the aforementioned traditional methods.

Yang's unit stream power is dependent on four dimensionless parameters, which are listed in Equation (17). In this equation, ω is the fall velocity of sediment particle, u_* is shear velocity, S_f is the surface slope, V_{cr} is critical velocity, and V is flow velocity.

$$C_t = f\left(\frac{\omega d_{50}}{v}, \frac{u_*}{\omega}, \frac{(V - V_{cr})S_f}{\omega}, \frac{u_* d_{50}}{v}\right) \quad (17)$$

Among these variables, particle fall velocity and critical velocity are indeterminate and, thus, should be determined before insertion into the model. To determine critical velocity, Yang (1972) presented Equation (18). The equation presented by Soulsby (1997) is used to determine particle fall velocity. Soulsby proposed Equation (19) for the fall velocity of sediment particles, which can be computed using the dimensionless particle diameter (D_*). Equation (20) shows the method for dimensionless particle diameter calculation:

$$\frac{V_{cr}}{\omega} = \begin{cases} \frac{2.5}{\text{Log}\left(\frac{u_* d_{50}}{v}\right)} + 0.66 & \text{for } 1.2 < \frac{u_* d_{50}}{v} < 70 \\ 2.05 & \text{for } \frac{u_* d_{50}}{v} \geq 70 \end{cases} \quad (18)$$

$$\omega = \frac{v}{d_{50}} \left(\sqrt{10.36^2 + 1.049 D_*^3} - 10.36 \right) \quad (19)$$

$$D_* = \left(\frac{g(SG - 1)}{v^2} \right)^{1/3} d_{50} \quad (20)$$

Cross-validation

Cross-validation is a model validation technique for generalization of the results of statistical indices to an independent dataset. Cross-validation is used for random subsampling. The K -fold cross-validation has a general form in which the dataset is divided by K roughly equal parts. In this method, each sample is used $(K-1)$ times in the training process and only once in the testing process. For instance, suppose that the initial dataset is divided by two subsets. First, one of the subsets is used in the training and the other is treated as the test dataset. Then, the roles of the two datasets are reversed and the modeling process is repeated. That is to say, the part that was used for testing in the first stage is used for training in the second and vice versa. The result is calculated by averaging the outputs of both runs. This approach is referred to as two-fold cross-validation. As a result, for a general form, the process is repeated K times and final output values are then computed by taking average of the K output values. There has been a wide range of research in the field of water resources and the study by Kazeminezhad et al. (2010) is just one example that applied this technique in an ANN method for scour around pipelines.

Experimental setup

In this study, the estimation and parameter calibration processes were codified in MATLAB® environment. LSSVMlab toolbox (Least squares support vector machine Lab (LS-SVMlab) toolbox 2011), which is available for free, was partly used for the purpose of estimation and PSO algorithm was codified and combined with LSSVR. The ten-fold cross-validation technique was then applied to the data, and is the prevalent technique compared to other models and has been considered as the model of choice in recent studies. In summary, data were shuffled at the beginning, thereby dividing by ten approximately

equal parts. Each time, one subcategory of the dataset was set aside for testing the model. The other nine folds were used for model training and calibration of LSSVR parameters. These parameters include the regularization parameter (γ) and the one related to the kernel parameter, which differs depending on the type of kernel. After pre-processing and normalization of the data, the learning process was carried out and the parameters were calibrated using the validation dataset. As previously stated, PSO algorithm was adopted for the purpose of calibration to facilitate the identification of the magnitude of LSSVR parameters. When various kernel functions were considered, it became clear that the radial basis function (RBF) had the best generalization among all other available functions. The search space embraced a range from 0 to 50,000 for both regularization and RBF parameters. A flowchart representing the combined model is shown in Figure 1. In general, the hybrid model of PSO-LSSVR requires a larger computational budget than when each of the methods are used individually for simulation purposes. However, the combined model may be more reliable due to the enhanced accuracy it offers. PSO has pre-determined parameters, which must be entered prior to execution. Table 3 shows the pre-set magnitudes of PSO parameters, including inertia weight and its corresponding minimum amount, inertia weight reduction factor, personal and swarm learning rate, maximum velocity reduction factor, critical probability, and termination condition. The PSO parameter

selection is based on the recommended values and ranges in the literature (Shi & Eberhart 1998; Rini et al. 2011). The learning rate of the model was then analyzed via the testing dataset (the last fold). This process was repeated ten times in total and, finally, the average of the results was compared with the measurements and the statistical indices were provided.

In a similar way, ten-fold cross-validation technique was used for validating the model tree in order to assess the ability of the M5P model. Multiple models were constructed. However, as is seen in the following, the best model has resulted in five equations (together demonstrated as Equation (21)). The equations indicate that M5P is consistent and efficient to estimate bed material load concentration via simple linear relationships.

$$\begin{aligned}
 Fr &\leq 18.086: \\
 &|S \leq 0.002: LM_2[345/29.149\%] \\
 &|S > 0.002: \\
 &|S \leq 0.004: LM_1[77/19.566\%] \\
 Fr &> 18.086: \\
 &|S \leq 0.001: LM_4[93/58.87\%] \\
 &|S > 0.001: \\
 &|\frac{W}{D} \leq 12.236: LM_3[74/37.158\%] \\
 &|\frac{W}{D} > 12.236: LM_5[53/100\%]
 \end{aligned}
 \tag{21}$$

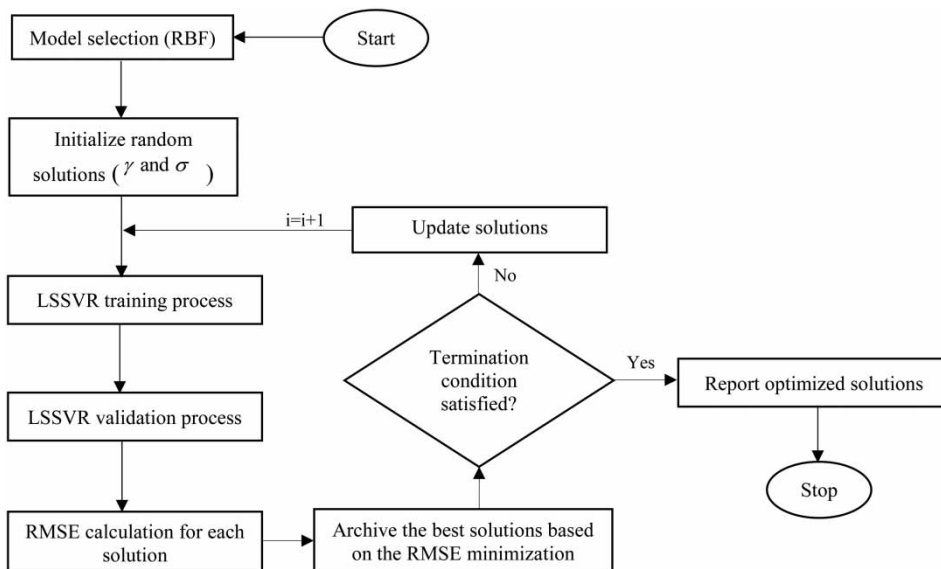


Figure 1 | Flowchart presentation of the training and optimization processes of PSO-LSSVR.

Table 3 | Predetermined values of the PSO parameters (Shi & Eberhart 1998; Rini et al. 2011)

Parameter	α	β	γ	w	w_{min}	c_1	c_2	Critical probability (P_{cr})	Termination condition (ϵ)
Value	0.9	0.9	0.4	1	0.4	2	2	0.1	0.01

where

$$LM_1: C_t = 188.993Fr + 682753.307S \\ - 18297.201\left(\frac{d_{50}}{D}\right) + 0.420\left(\frac{W}{D}\right) + 0.000Re \\ - 173.365\left(\frac{\tau_b - \tau_c}{\gamma_w(SG - 1)d_{50}}\right) - 1853.121$$

$$LM_2: C_t = 140.467Fr + 849309.993S - 26119.222\left(\frac{d_{50}}{D}\right) \\ + 1.976\left(\frac{W}{D}\right) + 0.001Re \\ - 191.849\left(\frac{\tau_b - \tau_c}{\gamma_w(SG - 1)d_{50}}\right) - 1529.854$$

$$LM_3: C_t = 493.349Fr + 2710196.341S \\ - 1834247.376\left(\frac{d_{50}}{D}\right) + 2.130\left(\frac{W}{D}\right) - 0.003Re \\ - 1137.777\left(\frac{\tau_b - \tau_c}{\gamma_w(SG - 1)d_{50}}\right) - 8074.967$$

$$LM_4: C_t = 91.029Fr + 1567494.646S \\ + 1918041.398\left(\frac{d_{50}}{D}\right) + 0.347\left(\frac{W}{D}\right) \\ - 15.387\left(\frac{\tau_b - \tau_c}{\gamma_w(SG - 1)d_{50}}\right) - 1739.678$$

$$LM_5: C_t = 3797.152$$

Performance metrics

To assess how the models in this paper perform, the following statistical indices (Equations (22)–(26)) are employed. They include the Nash–Sutcliffe efficiency (E), correlation coefficient (R), $RMSE$, scatter index (SI), and relative error (RE). E is a well-known index in water resources modeling and it has a range of $(-\infty, 1]$. An efficiency of one pertains to an excellent match between the predictions and observations, while a negative value corresponds to a model with a predictive power of worse than the observed mean. R is a measure of linear correlation between the predictions

and observations. RE represents the absolute bias. $RMSE$ shows the model's error of estimation and when it is normalized by the mean measured value is referred to as the scatter index (SI):

$$E = 1 - \frac{\sum (y - \hat{y})^2}{\sum (y - \bar{y})^2} \quad (22)$$

$$R = \left(n \sum (y \hat{y}) - \sum (y) \cdot \sum (\hat{y}) \right) \\ \times \left/ \left(\sqrt{n \sum (y^2) - (\sum y)^2} \cdot \sqrt{n \sum (\hat{y}^2) - (\sum \hat{y})^2} \right) \right. \quad (23)$$

$$RMSE = \sqrt{\sum (\hat{y} - y)^2 / n} \quad (24)$$

$$SI = (RMSE / \bar{y}) \times 100 \quad (25)$$

$$RE = \frac{1}{n} \cdot \sum |\hat{y} - y| / \bar{y} \quad (26)$$

where \hat{y} and y , respectively, denote the estimated and measured sediment concentration and n is the number of data. These metrics can generally give the analyst an insight into the model performance and robustness.

RESULTS AND DISCUSSION

Three models were basically developed in this research. In the first model, the dimensionless variables in Equation (5) were normalized and then entered in the hybrid estimator. The associated results appear in Figure 2 representing the error histogram of data. The abscissa axis in Figure 2 shows the relative error (RE) and the ordinate axis depicts the data frequency. It is apparent from the figure that the simulation error proved to be high. However, it was still

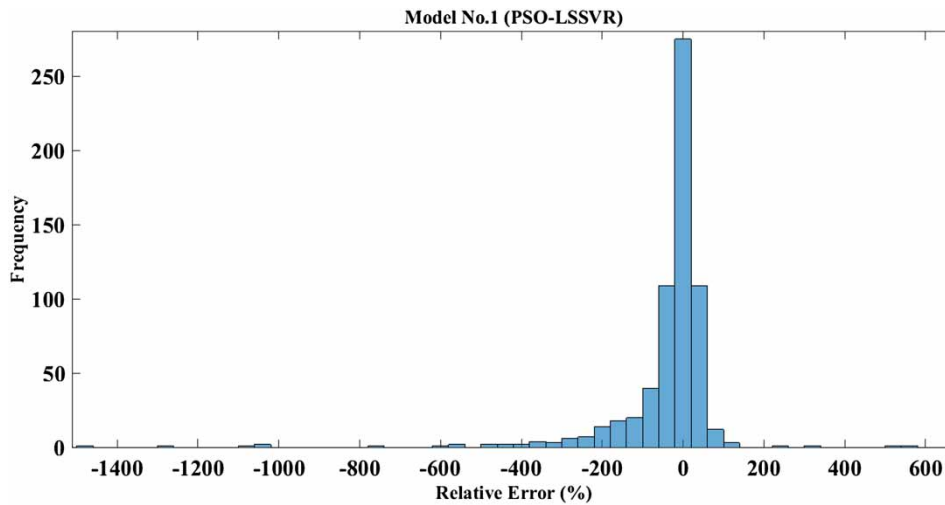


Figure 2 | Histogram of errors related to testing datasets of the initial model.

more satisfactory than traditional methods (see Table 4) since the majority of data lie in the range of %–200 and %200 and this is convincing for the phenomenon of sediment transport. The reason why there still occurs great errors in a few number of data points is the existence of considerable diversity in sediment concentration (ranging from 0 to approximately 40,000 mg/liter). Due to this, LSSVR attributes more error to the samples having lower sediment concentration as it is focusing upon the minimization of the sum of squared errors rather than relative errors in the optimization problem.

At the first stage of simulation, it is attempted to portray the functionality of the hybrid PSO-LSSVR model versus M5P. The scatter diagram in Figure 3 compares the proposed model with the M5P method and depicts

the simulated bed material load concentration versus the measured concentrations. As the figure illustrates, the PSO-LSSVR model fits the data to a relatively high degree and confirms its superiority over the M5P model (see the statistical indices in Table 4). Therefore, we refrained from further simulations by M5P for the next models. It is noteworthy that a handful of data with low concentrations were predicted as zero by PSO-LSSVR and M5P. Hence, they do not appear in the logarithmic scale.

The results obtained from the first model are then compared with those of several well-known approaches, including Engelund & Hansen (1972), Ackers & White (1973), Graf (1971), and Yang (1972). As is shown in Figure 4, Engelund & Hansen's (1972) approach overestimates bed material load transport rate. By contrast, Graf (1971) and

Table 4 | Statistical comparison of the results of PSO-LSSVR with those of model tree and traditional approaches

		<i>R</i>	<i>E</i>	<i>RMSE</i>	<i>RE</i>	<i>SI</i>	<i>RE < 0.4</i>
1st type of PSO-LSSVR	All data	0.9605	0.9217	940.38	23.30%	36.44%	62.46%
2nd type of PSO-LSSVR	All data	0.9662	0.9321	876.07	21.62%	33.95%	64.64%
3rd type of PSO-LSSVR	All data	0.9558	0.9100	1,008.5	21.80%	39.08%	71.50%
M5P		0.9059	0.8182	1,428.1	34.07%	55.34%	53.20%
Ackers & White (1973)		0.744	0.2342	2,941.4	52.48%	113.98%	38.63%
Engelund & Hansen (1972)		0.676	–73.232	28,958	515.99%	1122.1%	2.96%
Graf (1971)		0.397	–0.3041	3,838.2	89.35%	148.73%	8.41%
Yang (1972)		0.869	0.7157	1,792.1	41.34%	69.44%	43.93%

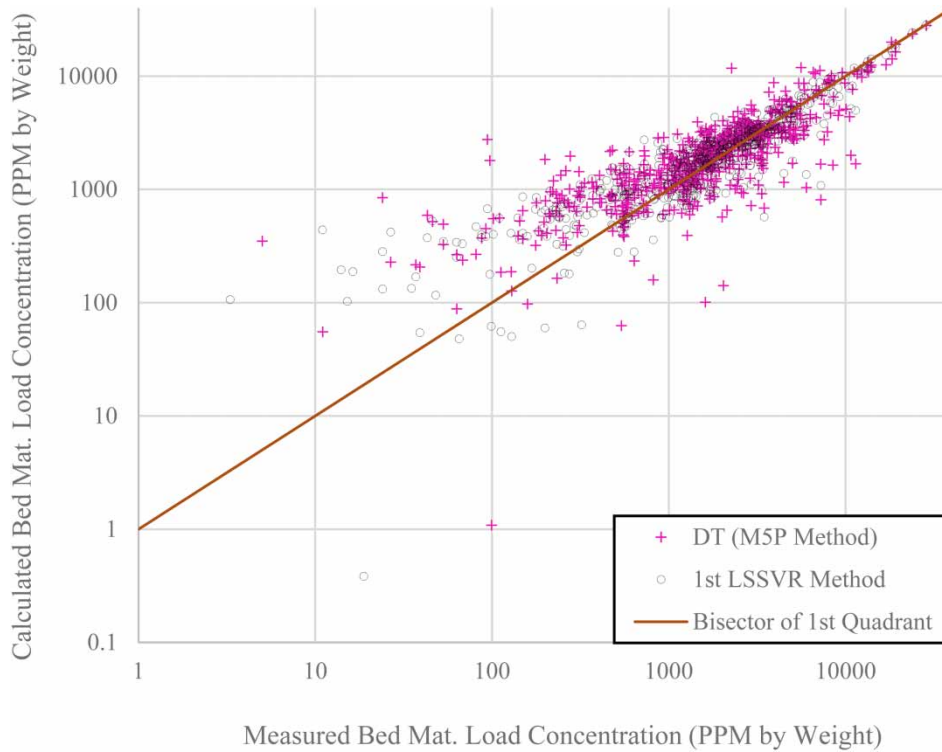


Figure 3 | Comparing the initial model of PSO-LSSVR with MSP method.

Ackers & White's (1973) approaches underestimate sediment concentration. This inability may be due to the methods' range of validity as Engelund & Hansen (1972) and Ackers & White's (1973) methods are respectively valid for $d_{50} > 0.15$ mm and $Fr < 0.8$. On the other hand, Yang's unit stream power approach has the best performance among all the traditional approaches, enjoying better correlation with the observed data. The PSO-LSSVR model demonstrates greater density around the bisector of the first quadrant of the Cartesian system in Figure 4 which means it estimates bed material load transport rate more accurately than traditional methods. However, Yang's unit stream power method is comparable with the PSO-LSSVR model in terms of estimating concentrations less than 50 ppm. Equation (27) shows a single execution of the model without utilizing cross-validation technique. In this equation, X_i denotes the vector of decision variables for the training and validation datasets. λ_i is the Lagrangian coefficient belonging to the i -th sample. X is the vector of decision variables for the test sample. σ^2 , n , and b are the RBF parameter, the total number of training datasets, and

bias term, respectively. It is meaningless to present the optimized values for these parameters because one may have a different dataset and use a specific shuffling procedure on their data, thus producing values specific to their model.

$$f(X) = \sum_{i=1}^n \lambda_i \exp\left(-\frac{|X - X_i|^2}{\sigma^2}\right) + b \quad (27)$$

Since Yang's method is capable of estimating bed material concentration, it may be beneficial to use its dimensionless variables (Equation (17)) in the proposed model to achieve a better estimation of the bed material load. By obtaining the fall velocity and critical velocity variables and adding the four dimensionless parameters of Yang's equation to the first model, further simulations based upon ten-fold cross-validation technique were conducted. There resulted in an improvement, to some extent, although it was not significant. Consequently, showing its visual comparison with the first model was avoided. Nevertheless, the model was still suffering from some weaknesses despite adding Yang's

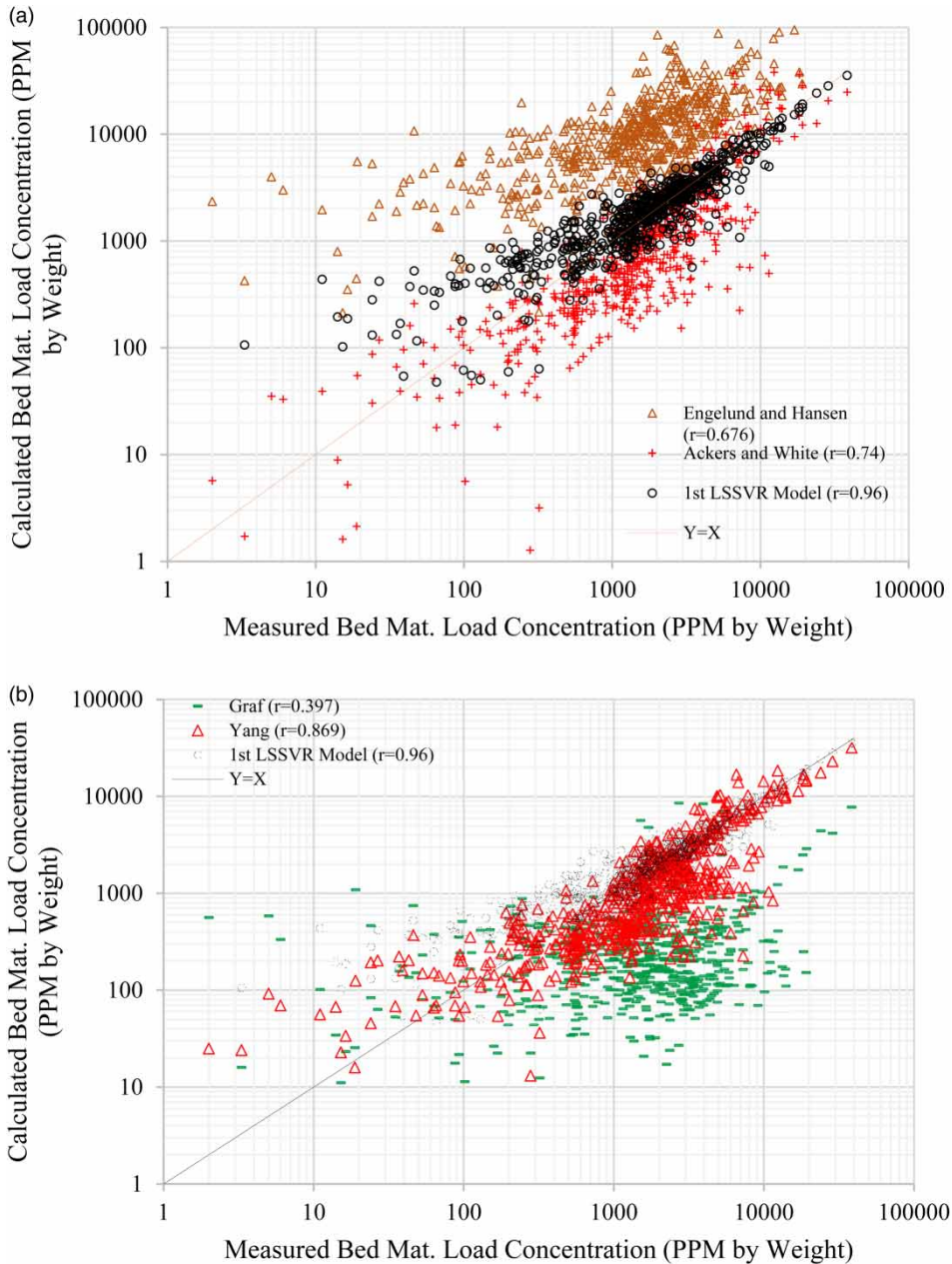


Figure 4 | Comparison of the initial model of PSO-LSSVR with (a) Engelund & Hansen (1972) and Ackers & White (1973) and (b) Graf (1971) and Yang (1972).

dimensionless variables. This is due to the nature of the objective function that is considered for finding the optimized model. The most striking conclusion to emerge from the results of the first and second models is that the pair of *RMSE* and correlation coefficient/NSE cannot solely explain the intricacies of the model. In other words, Nash–Sutcliffe, correlation coefficient, and

root mean squared errors are all biased towards the high values of sediment concentration when they are considered as the objective function for tuning LSSVR parameters. As a result, there were still noticeable oscillations in the estimation of the samples with low concentration values. By exploring the theoretical background of sediment hydraulics, and reviewing the

related literature, it is observed that most equations have been presented logarithmically. Yang's equation is a clear-cut example to illustrate the point. By doing so, the difference between the maximum and minimum sediment concentrations decreases and the model can emphasize low concentrations as well. Thus, the predictor and response variables were logarithmically scaled and ten further simulations were applied. The outcomes were converted to their initial scales at the end of the simulation process. Finally, the results were compared with that of the secondary model in Figure 5. The resulting function of a trained PSO-LSSVR following the procedures of the third model can be seen in Equation (28). Note that the logarithms are at base 10:

$$f(X) = 10^{\sum_{i=1}^n \lambda_i \exp(-((\text{Log}(X) - \text{Log}(X_i))^2 / \sigma^2)) + b} \quad (28)$$

As shown in Table 4, all three simulated models performed far better than traditional approaches; this is evident in the high correlation between the predicted and observed values. Moreover, the resulting *RMSE*, *SI*, and *RE* for the combined model are less than those for the traditional methods. The statistical indices attest to

the superiority of the PSO-LSSVR method over other approaches introduced in this study, thus making this model a viable alternative for calculation of sediment loads.

As mentioned before, adding the four dimensionless variables of Yang's method to the PSO-LSSVR model partially improved the results of the model. When the dependent and independent variables were logarithmically inserted in the model, *RMSE* and *SI* were slightly increased and the correlation and *NSE* approximately remained constant. By contrast, the percentage of data whose single *REs* were lower than 0.4 increased, which led to a decline in error dispersion and better distribution of errors in the histogram. Therefore, the *RE* of each sample must be taken into consideration in conjunction with the statistical performance metrics and the conclusion should be drawn by observing the histogram of errors. In the second model, the estimation accuracy of almost all the samples with high concentration values was satisfactory, but samples containing low concentrations were highly fault tolerant. In the final model, the histogram of relative errors was compressed such that 71.5% of the dataset had a *RE* of less than 0.4, whereas 64.64% of the data fell within this range in the

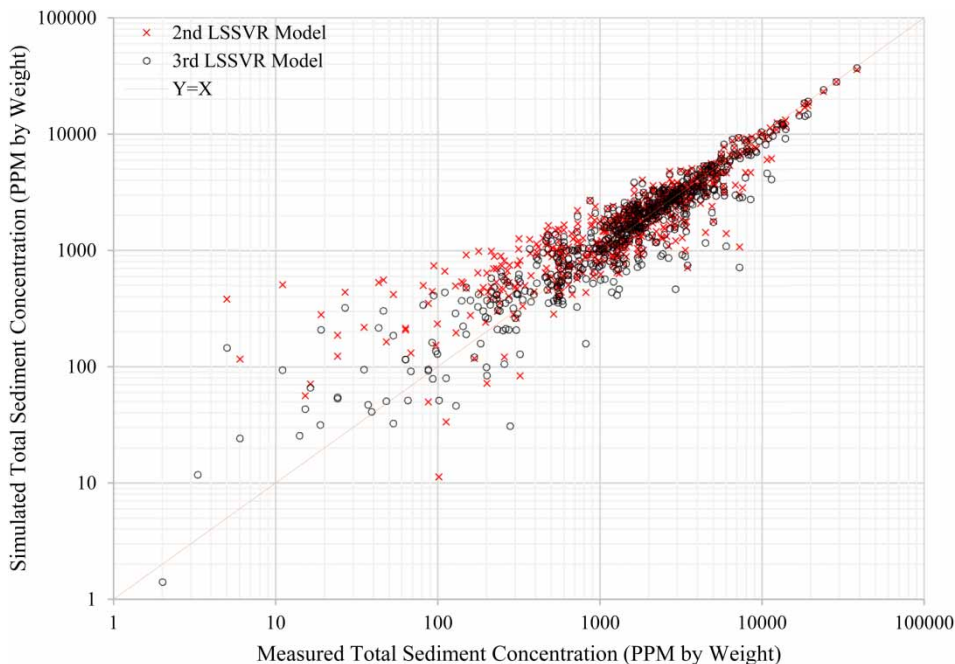


Figure 5 | Comparing the second model of PSO-LSSVR with the final model (logarithmic scale).

second model. According to Figure 6, a large percentage of samples with great errors has interestingly declined and resulted in a denser error distribution in the final model (see Table 4 and Figures 6 and 7). Figures 8 and 9 present the error histograms of Engelund & Hansen (1972) and Yang (1972) methods. It can be observed that the errors' magnitude was dramatically greater than the final version of PSO-LSSVR. By looking at the figures, it is seen that Engelund and Hansen's method yielded outlier predictions during the estimation of bed sediment concentrations.

Yang's method can be a relatively good way to estimate the concentration of bed sediments although it is not comparable with the hybrid model of PSO-LSSVR.

CONCLUSION

This study presented an alternative LSSVR method for estimating sediment transport rate in channels. This method was coupled with PSO algorithm to calibrate the

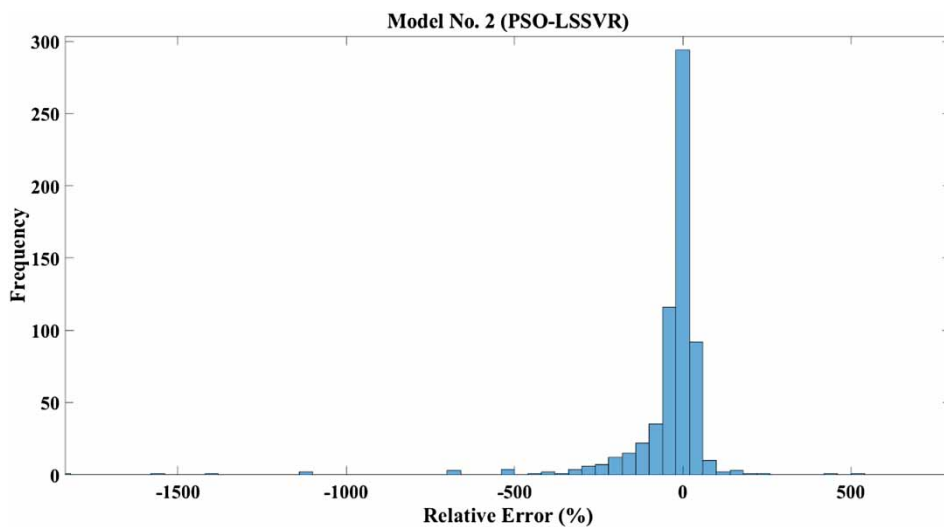


Figure 6 | Histogram of errors in the second model.

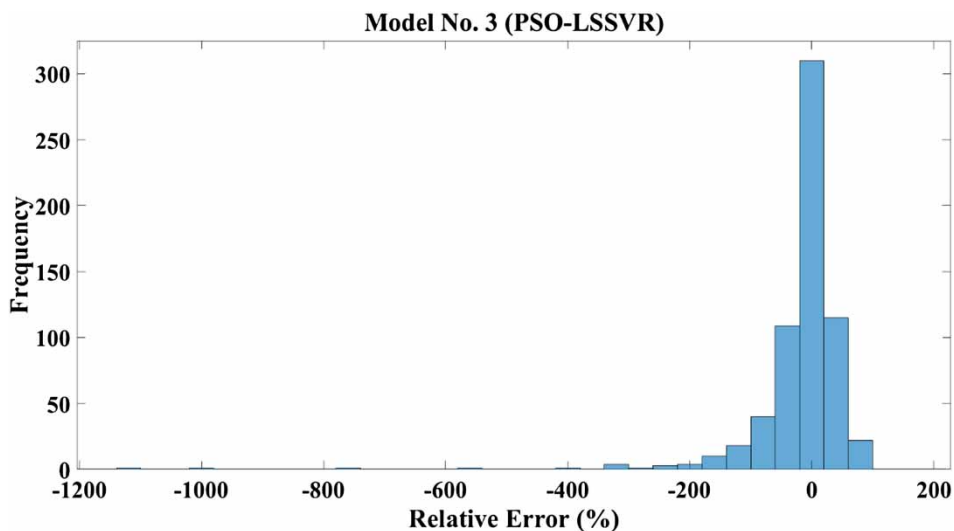


Figure 7 | Histogram of errors in the final model.

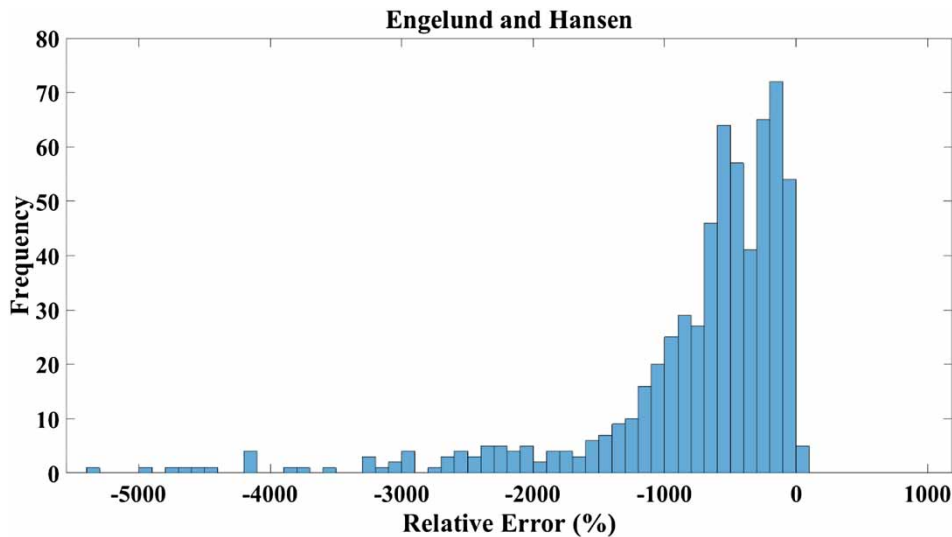


Figure 8 | Histogram of errors in the Engelund & Hansen (1972) method.

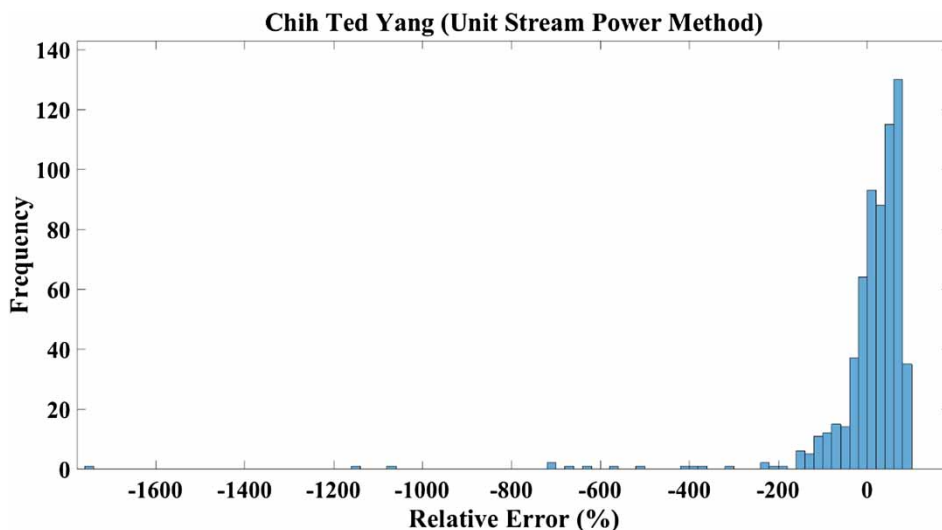


Figure 9 | Histogram of errors based on the Yang (1972) approach.

regularization and kernel parameters in LSSVR. The data were converted to a dimensionless form and then normalized. The cross-validation technique was applied to the data and a primary PSO-LSSVR model was created. It was found that the hybrid model had the best performance compared with those obtained from M5P and traditional methods. However, the Yang approach was still relatively superior for samples with low concentrations. As a result, its dimensionless variables were added to the hybrid

model leading to an improvement in the model training. Due to the wide range of concentration, all the data were converted into a logarithmic scale. When trained, it was found that a higher percentage of data had a *RE* of less than 0.4 and estimation of samples with low concentration became less volatile resulting in a denser error histogram. This study demonstrates a rigorous methodology for bed material load estimation and highlights key performance issues that may influence modeling in sediment transport.

Overall, PSO-LSSVR, developed as part of this study, shows great potential as an efficient, reliable, and easy-to-use machine learning approach for sediment transport rate estimation.

REFERENCES

- Ackers, P. & White, W. 1973 Sediment transport: new approach and analysis. *Journal of the Hydraulic Division (ASCE)* **99** (HY11), 2041–2060.
- Afan, H. A., El-shafie, A., Mohtar, W. H. M. W. & Yaseen, Z. M. 2016 Past, present and prospect of an Artificial Intelligence (AI) based model for sediment transport prediction. *Journal of Hydrology* **541**, 902–913.
- Asefa, T., Kemblowski, M., McKee, M. & Khalil, A. 2006 Multi-time scale stream flow predictions: the support vector machines approach. *Journal of Hydrology* **318** (1–4), 7–16. doi:10.1016/j.jhydrol.2005.06.001.
- Azamathulla, H., Ghani, A., Chang, C., Abu Hasan, Z. & Zakaria, N. 2010 Machine learning approach to predict sediment load. *Clean Soil Air Water* **38** (10), 969–976. doi:10.1002/clen.201000068.
- Bagnold, R. A. 1966 *An Approach to the Sediment Transport Problem From General Physics*. Geological Survey Professional Paper No. 422-I, US Government Printing Office, Washington, DC, USA.
- Bhattacharya, B., Price, R. K. & Solomatine, D. P. 2007 Machine learning approach to modeling sediment transport. *Journal of Hydraulic Engineering (ASCE)* **133** (4), 440–450. doi:10.1061/(ASCE)0733-9429(2007)133:4(440).
- Brownlie, W. R. 1981 *Compilation of Alluvial Channel Data: Laboratory and Field*. California Institute of Technology, California. Report No. KH-R-43B.
- Chen, Z. 2006 *Reduced-Parameter Modeling for Cost Estimation Models*. PhD Thesis, University of Southern California.
- Çimen, M. 2006 Estimation of daily suspended sediments using support vector machines. *Hydrological Sciences Journal* **53** (3), 656–666. doi:10.1623/hysj.53.3.656.
- Cortes, C. & Vapnik, V. 1995 Support-vector networks. *Machine Learning* **20** (3), 273–297. doi:10.1007/BF00994018.
- Einstein, H. A. 1942 Formulas for the transportation of bed load. *Transactions of the Society of Civil Engineers* **107**, 561–597.
- Engelund, F. & Hansen, E. 1972 *A Monograph on Sediment Transport in Alluvial Streams*. Teknisk Forlag, Copenhagen, Denmark.
- Etemad-Shahidi, A. & Mahjoobi, J. 2009 Comparison between M5' model tree and neural networks for prediction of significant wave height in Lake Superior. *Ocean Engineering* **36** (15–16), 1175–1181.
- Graf, W. 1971 *Hydraulics of Sediment Transport*. McGraw-Hill, New York, USA.
- Han, D., Chan, L. & Zhu, N. 2007 Flood forecasting using support vector machines. *Journal of Hydroinformatics* **9** (4), 267–276. doi:10.2166/hydro.2007.027.
- Huang, Z., Zhou, J., Song, L., Lu, Y. & Zhang, Y. 2010 Flood disaster loss comprehensive evaluation model based on optimization support vector machine. *Expert Systems with Applications* **37** (5), 3810–3814.
- Jain, S. K. 2001 Development of integrated sediment rating curves using ANNs. *Journal of Hydraulic Engineering (ASCE)* **127** (1), 30–37.
- Kazeminezhad, M. H., Etemad-Shahidi, A. & Yeganeh Bakhtiari, A. 2010 An alternative approach for investigation of the wave-induced scour around pipelines. *Journal of Hydroinformatics* **12** (1), 51–65. doi:10.2166/hydro.2010.042.
- Kennedy, J. & Eberhart, R. 1995 Particle Swarm Optimization. In: *IEEE International Conference on Neural Networks*, Nov/Dec, 4, 1942–1948, Perth, WA, Australia.
- Kennedy, J. & Eberhart, R. 2001 *Particle Swarm Optimization*. Academic Press, San Francisco, CA, USA.
- Kisi, O., Dailr, A. H., Cimen, M. & Shiri, J. 2012 Suspended sediment modeling using genetic programming and soft computing techniques. *Journal of Hydrology* **450–451**, 48–58. http://dx.doi.org/10.1016/j.jhydrol.2012.05.031.
- Least squares support vector machine Lab (LS-SVMlab) toolbox 2011 *Matlab/C Implementations for A Number of LS-SVM Algorithms*. http://www.esat.kuleuven.be/sista/lssvmlab/.
- Meyer-Peter, E. & Muller, R. 1948 Formula for bed load transport. In: *Proceedings of the 2nd Meeting, International Association for Hydraulic Structures Research*, Vol. 6, IAHR, Stockholm, Sweden.
- Misra, D., Oommen, T., Agarwal, A., Mishra, S. & Thompson, A. 2009 Application and analysis of support vector machine based simulation for runoff and sediment yield. *Biosystems Engineering* **103** (4), 527–535. doi:10.1016/j.biosystemseng.2009.04.017.
- Nagy, H. M., Watanabe, K. & Hirano, M. 2002 Prediction of sediment load concentration in rivers using artificial neural network model. *Journal of Hydraulic Engineering* **128** (6), 588–595. doi:10.1061/(ASCE)0733-9429(2002)128:6(588).
- Noori, R., Karbassi, A., Moghaddamnia, A., Zokaei-Ashtiani, M., Farokhnia, A. & Ghafari Gousheh, M. 2011 Assessment of input variables determination on the SVM model performance using PCA, Gamma test, and forward selection techniques for monthly stream flow prediction. *Journal of Hydrology* **401** (3–4), 177–189. doi:10.1016/j.jhydrol.2011.02.021.
- Prasad, V. 1991 *Velocity, Shear and Friction Factor Studies in Rough Rectangular Open Channels for Super Critical Flow*. PhD Thesis, Indian Institute of Science, Bangalore, India.
- Quinlan, R. J. 1992 Learning with continuous classes. In: *Proceedings of the 5th Australian Joint Conference on Artificial Intelligence*, 16–18 November, Hobart, Tasmania, pp. 343–348.
- Rao, A. & Sreenivasulu, G. 2006 Design of plane sediment bed channels at critical condition. *ISH Journal of Hydraulic Engineering* **12** (2), 94–117. doi:10.1080/09715010.2006.10514834.

- Rini, D. P., Shamsuddin, S. M. & Yuhaniz, S. S. 2011 **Particle swarm optimization: technique, system and challenges**. *International Journal of Computer Applications* **14** (1), 19–26.
- Sahraei, Sh., Zare Andalani, S., Zakermoshfegh, M., Nikeghbal Sisakht, B., Talebbeydokhti, N. & Moradkhani, H. 2015 Daily discharge forecasting using least square support vector regression and regression tree. *Scientia Iranica. Transaction A, Civil Engineering* **22** (2), 410–422.
- Shi, Y. H. & Eberhart, R. C. 1998 **A modified particle swarm optimizer**. In: *IEEE International Conference on Evolutionary Computation Proceedings*, pp. 69–73. IEEE Press, Piscataway, NJ, USA. doi:10.1109/ICEC.1998.699146.
- Shields, A. F. 1956 Application of similarity principles and turbulence research to bed-load movement. *Mitteilungen der Preussischen Versuchsanstalt für Wasserbau und Schiffbau* Berlin, Germany, **26**, 5–24.
- Shuquan, L. & Lijun, F. 2007 Forecasting the runoff using least square support vector machine. In: *Proceedings of the International Conference on Agriculture Engineering*, 20–22 October, Baoding, China.
- Sivapragasam, C., Liong, S. Y. & Pasha, M. 2001 Rainfall and runoff forecasting with SSA-SVM approach. *Journal of Hydroinformatics* **3** (3), 141–152.
- Smola, A. J. 1996 *Regression Estimation with Support Vector Learning Machines*. MSc Thesis, Technische Universität München, Munich, Germany.
- Solomatine, D. P. & Dulal, K. N. 2003 Model trees as an alternative to neural networks in rainfall-runoff modelling. *Hydrological Sciences Journal (Journal Des Sciences Hydrologiques)* **48** (3), 399–411.
- Soulsby, R. L. 1997 *Dynamics of Marine Sands*. Thomas Telford, London, UK.
- Suykens, J. A. K., Van Gestel, T., De Brabanter, J., De Moor, B. & Vandewalle, J. 2002 *Least Squares Support Vector Machines*. World Scientific Publishing, Singapore.
- Vapnik, V. N. 1998 *Statistical Learning Theory*. Springer, New York, USA.
- Vapnik, V. & Lerner, A. 1963 Pattern recognition using generalized portrait method. *Automation and Remote Control* **24**, 774–780.
- Wang, Y. & Witten, I. H. 1997 Induction of model trees for predicting continuous classes. In: *Poster Papers of the 9th European Conference on Machine Learning*, Springer.
- Williams, G. 1970 *Flume Width and Water Depth Effects in Sediment Transport Experiments*. US Geological Survey, Professional Paper, 562-H, 37 pp.
- Yang, C. T. 1972 Unit stream power and sediment transport. *Journal of the Hydraulic Division (ASCE)* **98** (HY10), 1805–1826.
- Yang, C. T. 1996 *Sediment Transport Theory and Practice*. McGraw-Hill, Singapore.
- Yilin, J., Cheng, C. T. & Chau, K.-W. 2006 **Using support vector machines for long-term discharge prediction**. *Hydrological Sciences Journal* **51** (4), 599–612. doi:10.1623/hysj.51.4.599.
- Yu, P., Chen, S. & Chang, I. 2006 **Support vector regression for real-time flood stage forecasting**. *Journal of Hydrology* **328** (3–4), 704–716. doi:10.1016/j.jhydrol.2006.01.021.

First received 11 March 2016; accepted in revised form 31 July 2017. Available online 12 September 2017