Updating the neural network sediment load models using different sensitivity analysis methods: a regional application

Reza Asheghi, Seyed Abbas Hosseini, Mojtaba Saneie and
Abbas Abbaszadeh Shahri

ABSTRACT

The amount of transported sediment load by streams is a vital but high nonlinear dynamic process in water resources management. In the current paper, two optimum predictive models subjected to artificial neural network (ANN) were developed. The employed inputs were then prioritized using diverse sensitivity analysis (SA) methods to address new updated but more efficient ANN structures. The models were found through the 263 processed datasets of three rivers in Idaho, USA using nine different measured flow and sediment variables (e.g., channel geometry, geomorphology, hydraulic) for a period of 11 years. The used parameters were selected based on the prior knowledge of the conventional analyses in which the effect of suspended load on bed load was also investigated. Analyzed accuracy performances using different criteria exhibited improved predictability in updated models which can lead to an advanced understanding of used parameters. Despite different SA methods being employed in evaluating model parameters, almost similar results were observed and then verified using relevant sensitivity indices. It was demonstrated that the ranked parameters using SA due to covering more uncertainties can be more reliable. Evaluated models using sensitivity indices showed that contribution of suspended load on predicted bed load is not significant.

Key words | artificial neural network, sensitivity analysis, transported sediment loads, updated model

Reza Asheghi
Seyed Abbas Hosseini (corresponding author)
Mojtaba Saneie
Abbas Abbaszadeh Shahri
Department of Civil Engineering, Science and
    Research Branch,
Islamic Azad University,
Tehran,
Iran
E-mail: *abbas_hoseyni@srbiau.ac.ir*

INTRODUCTION

The transported sediments by rivers as a complicated set of processes between stream flow, geologic, geomorphic, and organic factors is an important but critical regionally specific concern in the hydrological perspective to realize how rivers work (e.g., Melesse *et al.* 2011; Hajbabaei *et al.* 2017; Sari *et al.* 2017; Jin *et al.* 2018). Such sediments can be very informative in assessment of engineering purposes (e.g., channels, reservoirs, and dams), geo-environmental and ecosystem impacts (e.g., protection of fish and wildlife habitats), and river basin management (e.g., soil erosion, transported sediments, and pollutants) (e.g., Kisi *et al.* 2012; Bouzeria *et al.* 2017; Jin *et al.* 2018). Thereby, prediction of

sediment loads has become an important issue in many countries in introducing schemes for river water monitoring. Modeling approaches are the common way to estimate transported sediment loads. However, the effects of the involved parameters due to model structure, hydrological, time-series inputs, geological, geomorphological, hydrological and hydraulic features on predicted sediment loads should be considered. The wide variety of involved parameters exhibit no accepted universal approach to predict all types of sediment loads (Ma *et al.* 2017; Leimgruber *et al.* 2018; Asheghi & Hosseini 2020). This indicates why several modeling tools for simulating sediment loads have

been developed and evaluated (e.g., Kisi *et al.* 2012; Bouzeria *et al.* 2017; Leimgruber *et al.* 2018; Asheghi & Hosseini 2020). However, development of transported sediment models often needs to identify the uncertainty and the sensitivity of system performance due to any changes in possible input data from that predicted. This process assists in reducing not only the level of uncertainty but also, to an extent, the practicability (Gevrey *et al.* 2003; Saltelli *et al.* 2008; Razavi & Gupta 2015; Vu-Bac *et al.* 2016; Asheghi & Hosseini 2020).

Almost all of the developed conventional and analytical predictive sediment load models have mainly been utilized by regression techniques relying on hydrologic engineering parameters or landscape features (e.g., Camenen & Larson 2008; Ahmad *et al.* 2010; Kumar 2012). Meanwhile, the deficiency of regression techniques in simulating the effects of auxiliary factors and the involved uncertainty of experimental tests as well as inaccurate prediction in the wide range of expanded data (Cao *et al.* 2016; Asheghi *et al.* 2019) should be considered. Therefore, the developed equations are regionally specific and thus their applicability for other areas can never be guaranteed. Owing to such drawbacks in the adopted equations, prediction of sediment loads using different variables is a challenging task in the field of computational hydrology (e.g., Melesse *et al.* 2011; Bouzeria *et al.* 2017; Jin *et al.* 2018). Despite increased computing power in creating more sophisticated mathematical models, identifying the most important parameters on predicted sediment loads using sensitivity analysis (SA) techniques can lead to generating more accurate predictive models for carried sediment by a river.

In recent years, soft computing and data mining techniques and, in particular, artificial neural networks (ANNs) have successfully been applied, not only to capture complex nonlinear predictive sediment load models but also to overcome the inefficiencies of conventional methods to produce more precise results (e.g., Kisi *et al.* 2012; Afan *et al.* 2015; Bouzeria *et al.* 2017; Toriman *et al.* 2018; Asheghi *et al.* 2019). The main goal of ANN technology in dynamic environments such as rivers is to build a system that can change, adapt, and convert the potentials to become computable using many different types of computer learning. In order to design adaptive models for the evolving complexity of dynamic environments, ANN-based models are indicated as an appropriate choice.

Such developed ANN-based models then can be analyzed by different SA methods to identify the importance of the input variables. The SA methods allow understanding the concept of scientific codes (Rabitz 1989) and play a crucial role to provide essential insights on model behavior, structure, and response to inputs (Razavi & Gupta 2015; Borgonovo & Plischke 2016; Jin *et al.* 2018). Subsequently, removing the less effective factors not only leads to simpler and cost-effective models but also reduces the design and analysis time (Storlie *et al.* 2009; Abbaszadeh Shahri 2016; Asheghi *et al.* 2019). This issue in water resource engineering is gaining more importance to explain the nonlinear relationships between the explicative and response variables of a problem (e.g., Bahremand & De Smedt 2008; Razavi & Gupta 2015; Leimgruber *et al.* 2018).

Applying SA compels the decision-maker to identify effective variables on forecasts and indicates the critical variables for which additional information may be obtained. It helps to expose inappropriate estimation and thus guides the decision-maker to concentrate on the relevant variables. Due to the influence of various uncertain parameters on transported sediment loads' behavior, there is a need to identify and rank the importance of input factors on model output.

In this paper, two ANN-based predictive models for suspended and bed load using compiled datasets from 11 years' measurements of three rivers in Idaho, USA were developed. Discharge (Q), mean grain size (D_{50}), slope (S), flow velocity (V), area (A), depth (d), width (W), and shear flow velocity (U^*) were the selected inputs according to prior knowledge of conventional analyses. The models were then updated using different SA methods and examined by means of different external sensitivity indices. The compared performances indicated the appropriate predictability level of the updated models which can lead to an advanced understanding of the parameters used for model improvement.

STUDY AREAS AND DATA SOURCE

The Main Fork Red River (MFRR), South Fork Red River (SFRR), and Little Slate Creek (LSC) are in the streams' category of the state of Idaho (Figure 1). The MFRR in
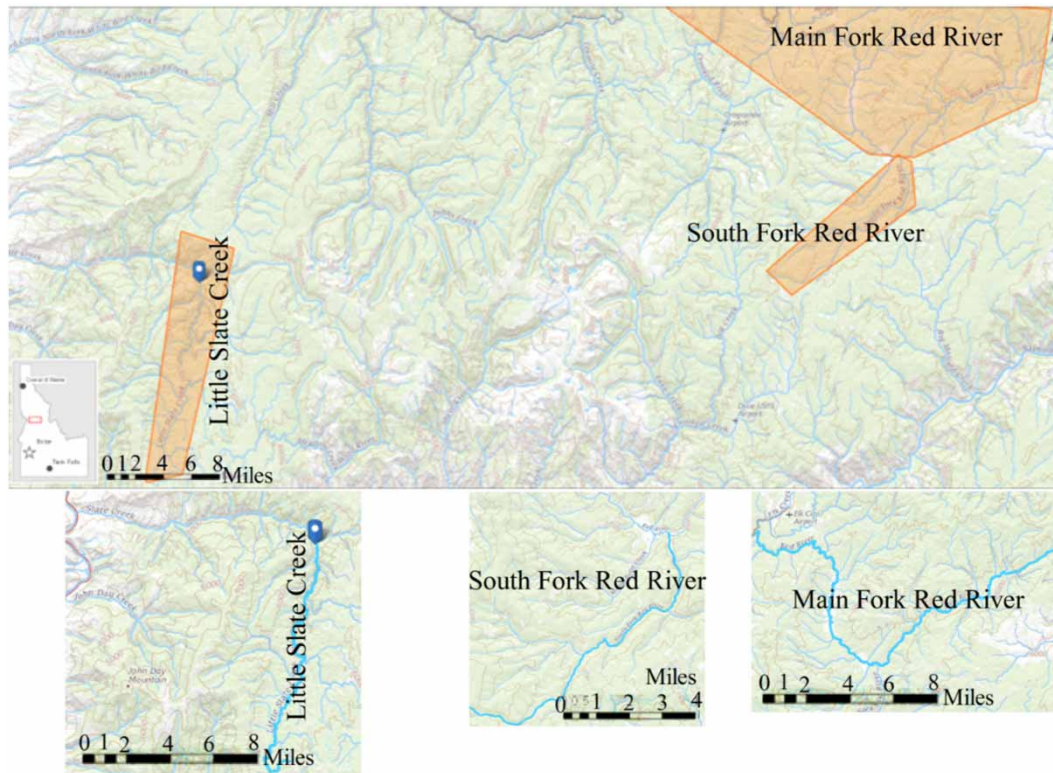
Figure 1 │ An overview of investigated rivers and connected streams (the maps have been generated by Idaho Fishing Planner (www.idfg.idaho.gov/ifwis/fishingplanner/mapcenter) and reproduced by the authors).

northern Idaho forms a confluence with the SFRR in the Nez Perce National Forest and the watershed predominantly lies on metamorphic rocks. The LSC flow is also on land administered by the Nez Perce Forest Service, but the geology of the watershed is mostly intrusive igneous. A unified dataset from the primary information on flow records and sediment transport measurements was screened from the United States Department of Agriculture (USDA) and United State Geological Survey (USGS). The provided dataset covers a period of 11 years (1986–1997) for both suspended and bed load sediments, including 263 sets of discharge (Q), mean grain size (D_{50}), slope (S), river area (A), velocity (V), river depth (d), river width (W), and shear flow velocity (U^*). The components of the database were then categorized into channel geometry, geomorphological, and hydraulic sets. Descriptive statistics of the compiled datasets can be found in Table 1. Due to the wide range of precipitations in the recorded years and consequently significant observed variation in Q and A, higher standard deviation

for these factors are to be expected (Table 1). The datasets were normalized within the range of [0, 1] as a necessary step to improve the learning speed and model stability. To organize training, testing, and validation sets for ANN models, datasets were randomized into 55%, 25%, and 20%. These values were considered because in comparison with several different tested percentages they showed more accurate results.

MODELING BY ANN

The ANNs are recognized as applicable and robust computational models for predicting and classification purposes. Typically, such structures are configured by an appropriate combination of artificial neurons and activation functions to improve the quality of processed information (e.g., Kisi *et al.* 2012; Bouzeria *et al.* 2017; Toriman *et al.* 2018; Asheghi & Hosseini 2020). In each artificial neuron (Figure 2), input (x_i), weights ($w_{i,j}$), bias (b_i), activation function

Table 1 | Statistical description of input parameters of the used rivers to predict sediment loads

River	Variable	Mean	Mean SE	St. dev.	Min	Max	Skewness
MFRR	Q (ft³/s)	151.50	10.50	105.3	13.3	487	1.36
	D_{50} (mm)	1.343	0.06	0.634	0.54	5.279	2.6
	W (ft)	32.184	0.29	2.925	22	40.3	0.4
	V (ft/s)	2.867	0.089	0.903	0.99	5.01	0.22
	d (ft)	1.423	0.044	0.445	0.34	2.86	0.57
	A (ft²)	46.4	2.41	24.22	9.3	126	0.94
	S (ft/ft)	0.004	0.000001	0.000085	0.0038	0.0041	−0.06
	U^* (ft/s)	0.42	0.007	0.06938	0.204	0.612	−0.08
SFRR	Q (ft³/s)	109.78	9.95	93.85	7.25	458	1.69
	D_{50} (mm)	0.886	0.048	0.454	0.13	2.7	1.03
	W (ft)	26.942	0.364	3.432	20	40	1.28
	V (ft/s)	2.572	0.107	1.014	0.553	5.293	0.52
	d (ft)	1.281	0.045	0.423	0.37	2.28	0.46
	A (ft²)	37.37	1.66	15.69	10.3	78.95	0.94
	S (ft/ft)	0.0014	0.000005	0.000044	0.0013	0.00146	−0.13
	U^* (ft/s)	0.236	0.0044	0.042	0.125	0.32657	0.07
LSC	Q (ft³/s)	194.4	14.5	123.9	18.7	534	0.81
	D_{50} (mm)	1.118	0.147	1.26	0.42	6.65	3.73
	W (ft)	37.775	0.487	4.16	22	44	−1.37
	V (ft/s)	2.537	0.127	1.085	0.68	5.39	0.52
	d (ft)	1.619	0.044	0.374	0.81	2.67	0.18
	A (ft²)	69.3	2.28	19.46	27.1	112	−0.14
	S (ft/ft)	0.0261	0.000052	0.00044	0.025	0.0267	−0.26
	U^* (ft/s)	1.158	0.016	0.136	0.825	1.47	−0.21

Note: The units are according to US measurement system: SE, standard error; St. dev., standard deviation.
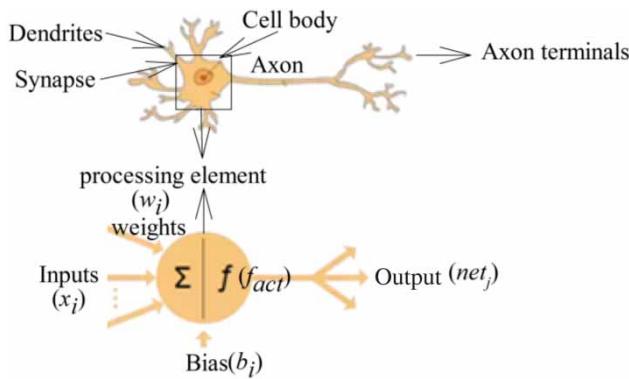

Figure 2 | Simplified comparison between artificial and biological neuron scheme.

(f_{act}), and output ($O_{i,j}$) are the involved components on information transferring. The data from the input layer are projected to the intermediate (hidden) layers while the final hidden layer projects the information to the output neurons.

The jth network output (net_j) using set of inputs $X = \{x_1, x_2, ..., x_n\}$ and corresponding adaptive weight of $w_{i,j}$ can be

expressed using the propagation function (f_{prop}) as:

$$net_j = f_{prop}(O_{i1}, ..., O_{in}, w_{i1,j}, ..., w_{in,j}) \rightarrow net_j$$
$$= \sum_{x \in X} (O_i . w_{i,j}) + b_i \tag{1}$$

where b_i denotes the bias which is a type of connection weight with a constant nonzero value and set up into all the neurons in the back-propagation and transfer functions except for the input layer. The activation state $a_j(t)$ explicitly is assigned to any given j^{th} neuron and transforms the net_j from the previous activation state $a_j(t-1)$ into a new $a_j(t)$ using:

$$a_j(t) = f_{act}(net_j(t), a_j(t-1), \theta_j) \tag{2}$$

where θ_j denotes the threshold value uniquely assigned to j^{th} neuron and marks the position of the maximum gradient value of the activation function. Then, the output value O_j

of the neuron j is calculated from its activation state a_j as:

$$f_{out}(a_j) = O_j \tag{3}$$

The specific error of each sample (Err_p) and root mean square error ($RMSE_{Err}$) between the input (x) and the actual output (y) for the kth output neuron can be defined by:

$$Err_p = \frac{1}{2}\sum_{k \in O}(x_k - y_k)^2 \; ; \; Err_{total} = \sum Err_p \tag{4}$$

$$RMSE_{Err} = \sqrt{\frac{\sum_{k \in O}(x_k - y_k)^2}{|O|}} \tag{5}$$

The weight $w_{i,j}$ due to activity of both neuron j and i is changed to ($\Delta w_{i,j}$):

$$\Delta w_{i,j} = \eta O_i a_j \tag{6}$$

where η is the learning rate.

Network training basically aims to find the optimum weights using and updating procedure for $(n+1)^{th}$ pattern known as generalized delta rule (GDR):

$$\Delta w_{i,k} = -\eta \frac{\partial Err(W)}{\partial w_{i,k}} \tag{7}$$

$$w_{i,k}(n+1) = w_{i,k}(n) + \nabla w_{i,k}(n) \tag{8}$$

SENSITIVITY ANALYSES TO ASSESS MODEL PARAMETERS

In recent years, different SA techniques have been developed to evaluate quantitative models and address the contribution of parameters on produced output (Borgonovo & Plischke 2016). The SA methods, due to their ability in determining the effectiveness of input parameters on produced outputs, are important in a simulation process (Calver 1988; Saltelli *et al.* 2000, 2008). According to the literature (e.g., Jacomino & Fields 1997; Saltelli 2002; Borgonovo & Plischke 2016), the SA methods are categorized into quantitative techniques, graphical method,

sensitivity-index approach, and specified tailored mathematical models. These methods facilitate finding a simplified but robust calibrated model from a large number of parameters and identify important connections between observations and model output as well as ability in investigating the effect and impacts of the uncertainties in the output of a mathematical model (Wang *et al.* 2000; Saltelli 2002; Gevrey *et al.* 2003; Helton *et al.* 2006; Bahremand & De Smedt 2008).

The one-at-a-time (OAT/OFAT) method (Czitrom 1999), the local methods including adjoint modeling (Cacuci *et al.* 2005) and automated differentiation (Griewank 2000), scatter plots (Paruolo *et al.* 2013), regression analysis and variance-based methods (Sobol 1993), variogram-based methods (Haghnegahdar & Razavi 2017), screening (Campolongo *et al.* 2007), emulators (data-modeling/machine learning approaches) (Storlie *et al.* 2009), and probabilistic methods (Oakley & O'Hagan 2004; Vu-Bac *et al.* 2016) are some of the used or introduced SA methods.

In ANN-based models, the SA is conducted by analyzing adjusted weights through the equation method (EM) (Hashem 1992), weight magnitude analysis method (WMAM) (Garson 1991; Poh *et al.* 1998), variable perturbation method (VPM) (e.g., Gedeon 1997; Poh *et al.* 1998; Montaño & Palmer 2003; Zeng & Yeung 2003), partial derivative algorithm (PaD) (Dimopoulos *et al.* 1995), profile method (PM) (Lek *et al.* 1996), stepwise method (SM) (Sung 1998; Gevrey *et al.* 2003), and cosine amplitude method (CAM) (Ross 1995). Despite different suggested SA techniques, the PaD and the VPM have presented superior performance compared to other techniques based on the WMAM (Wang *et al.* 2000; Zeng & Yeung 2003). However, successes of the CAM in different engineering applications have also been approved (Abbaszadeh Shahri 2016; Abbaszadeh Shahri & Asheghi 2018; Abbaszadeh Shahri *et al.* 2019).

In EM, the influence of each input variable on the output (I_i) can be calculated as:

$$I_i = \sum_k O(1-O)w_{k1}^2 v_k^2(1-v_k^2)w_{ik}^1 \tag{9}$$

where w_{bc}^a denotes the weight from the bth node in the ath layer to the cth node in the next layer. O is the output

node and w_{k1}^2 expresses the outgoing weight of the kth node in the second layer. v_k^2 is the output value of the kth node in the second layer and w_{ik}^1 represents the connection weights between the ith and k^{th} nodes of the first and hidden layers.

Based on the WMAM, different equations have been proposed. Poh *et al.* (1998) indicated that by normalizing the connecting weights between input and hidden layers subjected to largest weight magnitude, the influence of variables on output then can be ranked as:

$$I_i = \sum_k \frac{w_{ik}^1}{\max All_{i,k} w_{ik}^1} \tag{10}$$

Garson (1991) pointed out that the importance of each parameter on the output (Q_{ik}) can be found through the connection weight between the input neuron i and the hidden neuron j (W_{ij}) and then hidden neuron j and the output neuron k (V_{jk}) for each of the hidden neurons of the network:

$$Q_{ik} = \frac{\sum_{j=1}^{L} \left(\left(\frac{w_{ij}}{\sum_{r=1}^{N} w_{rj}} \right) V_{jk} \right)}{\sum_{i=1}^{N} \left(\sum_{j=1}^{L} \left(\frac{w_{ij}}{\sum_{r=1}^{N} w_{rj}} \right) V_{jk} \right)} \tag{11}$$

where $\sum_{r=1}^{N} w_{rj}$ denotes the sum of the connection weights between the input neurons N and the hidden neuron j. Gevrey *et al.* (2003) showed that the relative contribution (RC) of each input on output can be calculated using the number of input (n_i) and hidden neurons (n_j) and corresponding weight to input neuron i and hidden neuron j (w_{ij}):

$$D = \frac{|w_{ij}|}{\sum_{i=1}^{n_i} |w_{ij}|} \rightarrow RC_i = \frac{\sum_{j=1}^{n_j} D_{ij}}{\sum_{j=1}^{n_j} \sum_{i=1}^{n_i} D_{ij}} \tag{12}$$

Conversely, the similarity between related parameters can be obtained by the CAM. In this method, all data pairs ($X = (x_i, y_j)$) are expressed in common X-space to provide a data array ($X = \{X_1, X_2, \ldots X_n\}$) in which each X_i is a vector of length m ($X_i = \{x_{i1}, x_{i2}, \ldots x_{im}\}$) and exhibits the dot product for the cosine function (Ross 1995). The assigned data pairs to a point in m-dimensional space needs to be described by m-coordinates. Therefore, the importance and membership value of each element of a

model in m-dimensional space (R_{ij}) in the form of a $m \times m$ matrix can be expressed by a pairwise comparison of two data samples (x_i and x_j) by:

$$R_{ij} = \frac{\sum_{k=1}^{m} (x_{ik} \times x_{jk})}{\sqrt{\sum_{k=1}^{m} x_{ik}^2 \sum_{k=1}^{m} x_{jk}^2}} \tag{13}$$

The PaD is a famous ANN-based SA technique that can identify the contribution of input changes on the output using Jacobian matrix of the partial derivatives of outputs with respect to inputs (Dimopoulos *et al.* 1995). The general formulation of PaD using the output variable (Y_j) and parameters (θ_i) for N_p number of parameters and N_v number of variables (model outputs) is expressed as:

$$PaD_{i,j} = \frac{\partial Y_j}{\partial \theta_i} \rightarrow i = 1, 2, \ldots N_p; j = 1, 2, \ldots N_v \tag{14}$$

In the case of ANN models, PaD can be expressed by:

$$PaD_{oi} = \frac{\partial O_k^p}{\partial x_i^p} \sum_{h=1}^{n_h} w_{jo} y_{ij} (1 - y_{jo}) w_{ij} \tag{15}$$

where y_{ij} is the output of jth neuron in respect to i^{th} input. w_{jo} and w_{ij} are the weights between the kth output neuron and j^{th} hidden neuron as well as i^{th} input and j^{th} hidden neuron, respectively. Then the sensitivity of p training samples of N total number of data variables for each input x_i on the output O_k is defined as:

$$S_i = \frac{1}{N} \sum_p \frac{\partial O_k^p}{\partial x_i^p} \tag{16}$$

The relative contribution of ith input variable (RC_i) on a specific output then can be determined by the sum of the squares of the partial derivatives (SSD) using:

$$RC_i = \frac{SSD_i}{\sum_i SSD_i} \rightarrow SSD_i = \sum_p \left(\frac{\partial O_k^p}{\partial x_i^p} \right)^2 \tag{17}$$

The PM introduced by Lek *et al.* (1996) aims to analyze the median of a particular input subjected to fixing of all

other inputs using dividing into five equal subintervals (scales) corresponding to minimum, quarter, half, three quarters, and maximum. The contribution of each input parameter then can be explained from the created profile of the median values against the corresponding subintervals. This procedure should be executed for all inputs to obtain a set of descriptive relative importance curves (Gevrey *et al.* 2003).

The SM is focused on examining a step-by-step procedure for adding or rejecting the input using an iterative loop. In the SM process, by blocking one-by-one of the input parameters and calculating the corresponding *MSE* of responses, the relative importance of each input variable is ranked. The parameter with the maximum *MSE* value is considered as the most important and can then be either removed from the model or use its mean value to find the contribution of other parameters (Gevrey *et al.* 2003). The SM can be organized into two forward and backward strategies. In the backward strategy, the *MSE* of each parameter is calculated using constructed ANN models consisting of all input parameters and then starting to block each input parameter while forward strategy works in the reverse way (Sung 1998).

The VPM is a common straightforward SA technique for ANN-based models which can be achieved by analyzing the output disturbance due to perturbed inputs. The VPM adjusts the input values of one variable while keeping all the other variables untouched (Gedeon 1997; Montaño & Palmer 2003). In the VPM, the direct small perturbation on each ANN input and the corresponding change in the outputs is measured, while EM and WMAM analyze indirect changes of ANN weights. The variance of the input parameter from 0 to 50% by steps of 5% can be implemented as perturbation and the generated outputs can be ranked based on the calculated *MSE* corresponding to each perturbed input (Gevrey *et al.* 2003).

APPLYING THE SA TO UPDATE ANN MODELS

In this paper, the contribution of input variables in predicted suspended and bed loads were found through two developed optimum ANN-based models. The dependency of optimum network size to internal characteristics (e.g., training

algorithm, number of neurons, learning rate, activation function, architecture, regularization) implies that no *standard method nor* for programmatic network configuration neither to prevent the over-fitting problem is accepted (Ghaderi *et al.* 2019). To optimize the ANN models the organized procedure in Figure 3 using integration of trial-and-error methods with a developed code based on constructive techniques was followed. In this process, various training algorithms including quick propagation (QP), Levenberg–Marquardt (L-M), quasi Newton (QN) and momentum (MO) were used. As defined in Figure 3, different internal characteristics on numerous generated topologies were applied to avoid the overfitting problem and escape from local minima.

The QP as one of the most popular recognized back propagation training algorithms is based on the mathematical method of gradient descent, with appropriate results in most problems (Fahlman 1988). The L-M (Levenberg 1944; Marquardt 1963) is an advanced and fast non-linear optimization algorithm that can solve generic curve-fitting problems. However, it can only be used on networks with a single output unit or small networks because its
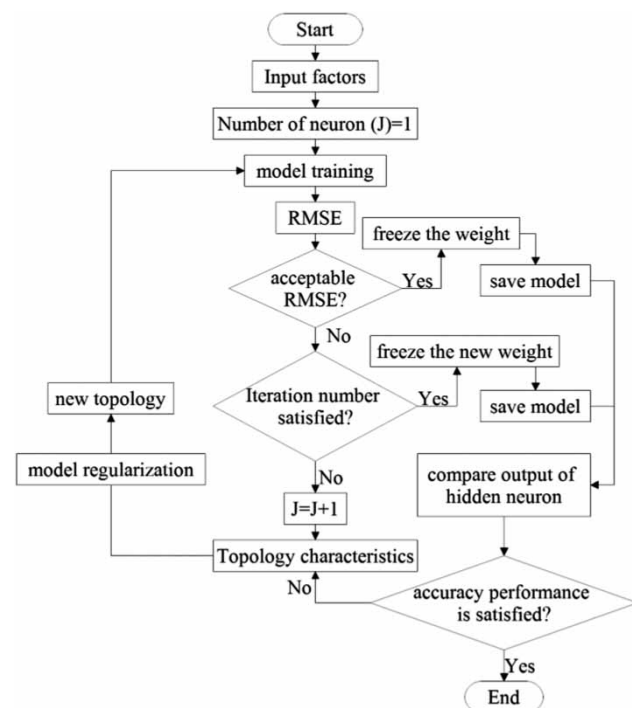

Figure 3 | The implemented procedure to find the optimum structure.

memory requirements are proportional to the square of the number of weights in the network. Moreover, L-M is specifically designed to minimize the sum of squares error and thus cannot be used for other types of network error. The QN (Bertsekas 1995) is a network training algorithm based on Newton's method to avoid the need to store computed Hessian matrix during each iteration and thus requires less memory and can be used for bigger networks. The MO as a well-known standard algorithm in the neural network community is designed to overcome some of the problems associated with standard back propagation training algorithm and is used to speed up convergence and maintain generalization performance (Swanston *et al.* 1994). The MO is a locally adaptive approach in which each weight remembers the most recent update and thus each weight is able to update independent of other weights (Wiegerinck *et al.* 1994). Two stopping criteria, the minimum root mean square error (*MRMSE*) and number of iterations, were employed. The number of iterations is replaced when *MRMSE* cannot be achieved. As presented in Figure 4(a) and 4(b), the *MRMSE* of applied training algorithms subjected to different activation functions against the number of neurons was found to be 11 and 12, which further should be organized in hidden layer(s). According to the defined procedure in Figure 3, numerous models with similar structures but different internal characteristics were examined and investigated by the absolute error (*AE*). The *AE* as the deviation between predicted and measured values corresponds to model quality and indicates the amount of physical error and uncertainty in a measurement (Abbaszadeh Shahri 2016). In Figure 4(c)–4(f), a sample of the carried out procedure to find the optimum topologies and corresponding calculated *AE*, as well as model predictability for suspended and bed loads, is presented. The characteristics of optimum structures using applied training algorithms are reflected in Table 2. It was observed that the 7-5-6-1 and 8-5-7-1 structures for suspended and bed loads can generate higher predictability than other tested models (Table 2; Figure 4(a) and 4(b)). The effect of input parameters on the predicted sediment loads were then identified using PaD, CAM, RC, and EM sensitivity analysis methods (Figure 5). Despite observed differences, the ranked parameters almost follow a similar trend. Accordingly, the *Q*, *V*, *d*,

and *A* for suspended load and the *Q*, *V*, D_{50}, *S*, and *d* on bed load were identified as the most effective factors.

On the basis of SA results, the least effective factors on predicted output can be removed. This procedure is not only able to update the model and reduce network size but may also lead to increasing the accuracy of prediction (Hamby 1994; Saltelli 2002; Gevrey *et al.* 2003; Helton *et al.* 2006; Saltelli *et al.* 2008; Razavi & Gupta 2015; Vu-Bac *et al.* 2016; Abbaszadeh Shahri *et al.* 2019; Asheghi *et al.* 2019). Therefore, the *S*, *U**, *W* and *Sus-load*, *W*, *A* as the three least effective factors were ignored. The results of updated models subjected to the most dominant identified variables (Figure 5) and the same randomized datasets through the defined procedure (Figure 3) are then reflected in Table 3 and Figure 6, respectively.

DISCUSSION AND VALIDATION

SA techniques can determine the performance indices, modeling hypothesis, and interaction of factors or group of factors with each other (Saltelli *et al.* 2000; Pappenberger *et al.* 2008). In nonlinear models, applying elementary effects (EE), first-order Sobol sensitivity index (SI) (Sobol 1993), and total sensitivity index (TSI) provide valuable information to quantify the sensitivity (Sobol 1993). The TSI measures the contribution to the output variance of X_i of the input factors including all interactions with any other input variables. The EE, SI, and TSI of a set of variables $X = \{X_1, X_2, ..., X_k\}$ on model Y are defined as:

$$EE_i = \frac{Y(X_1, X_2, ..., X_i + \Delta, ..., X_k) - Y(X_1, X_2, ..., X_k)}{\Delta}$$

(18)

$$SI = \frac{V_{X_i}(E_{X_{\sim i}}(Y|X_i)}{V(Y)}$$

(19)

$$TSI = 1 - \frac{V_{X_{\sim i}}(E_{X_i}(Y|X_{\sim i})}{V(Y)}$$

(20)

where EE_i represents the elementary effects of each variable *i*. Δ shows the step change in the discrete variable X_i. $Y(X_1, X_2, ..., X_k)$ is the model output that should be
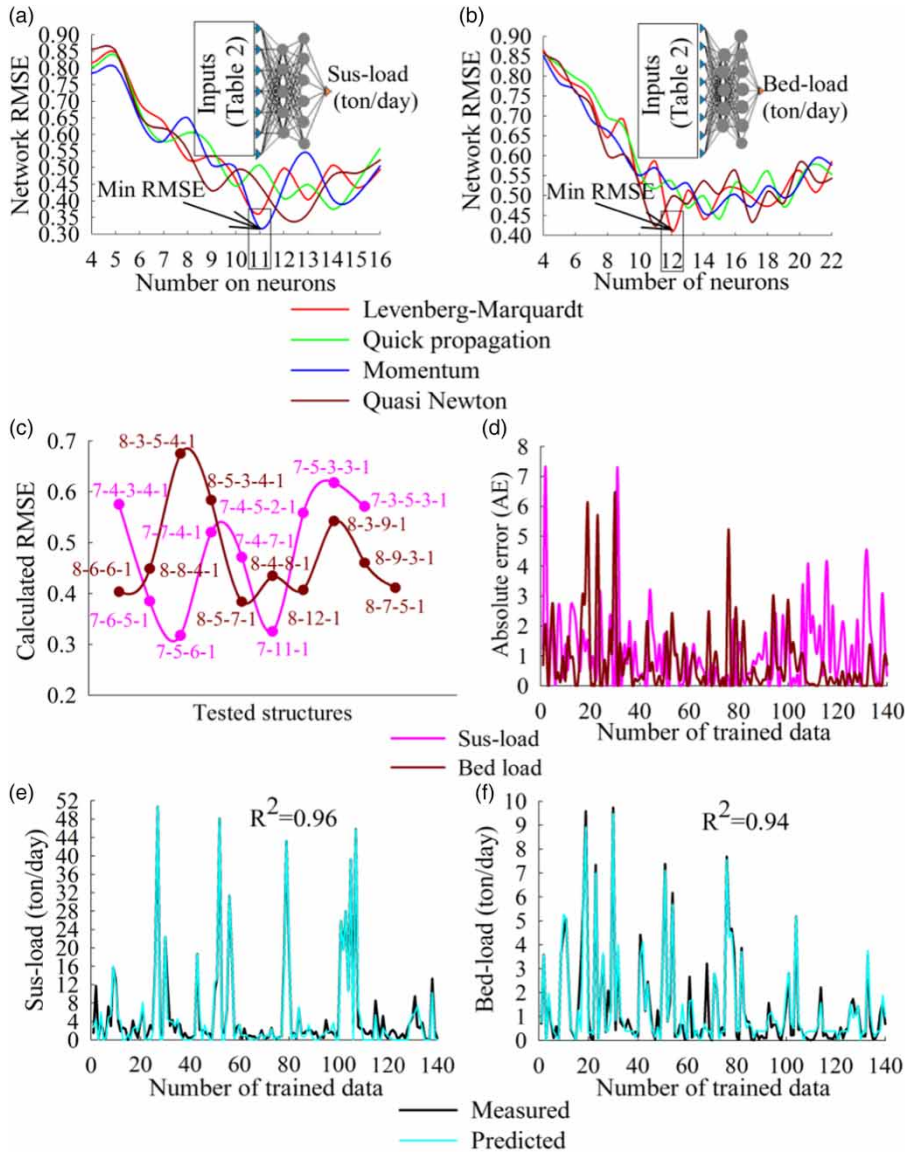
Figure 4 | Variation of network *RMSE* subjected to different training algorithms and number of neurons for (a) suspended and (b) bed loads. A series of tested structures and corresponding *AE* to find the optimum models (c) and (d). Comparison of measured and predicted values in training stage using introduced optimum models (e) and (f).

fixed for each calculated EE_i. V is the variance and $X_{\sim i}$ are all parameters but X_i and E represent the average. $V(Y)$ denotes the unconditional variance of the quantity of interest. The term $V_{X_i}(E_{X\sim i}(Y|X_i)$ as the variance of conditional expectation is the first-order effect of X_i on Y and is the variation of the average Y when fixing X_i at different values while varying the other parameters.

Each EE_i using SI then can be characterized by the mean value and its standard deviation whereas high EE_i

values indicate more impact on the model Y. The lower SI value also shows the less variability in the output Y and consequently is more robust to variations in the model parameters (Gan *et al.* 2014; Song *et al.* 2015; Jin *et al.* 2018). The scatter plots of mean and standard deviation values of SI using validation datasets for suspended and bed load are presented in Figure 7. The closer to the (0, 0) coordinate is interpreted as more robust method to capture the changes of parameters.

Table 2 | Characteristics and internal properties of optimum models

Training algorithm	Number of neurons	Corresponding structure	MRMSE	Activation function	
				Hidden layer	Output
Suspended load→ Inputs: Q, S, V, d, W, U^*, A					
QP	14	7-6-8-1	0.375	hyperbolic tangent	logistic
L-M	11	7-11-1	0.361	logistic	logistic
QN	12	7-5-7-1	0.350	hyperbolic tangent	hyperbolic tangent
MO	*11*	*7-5-6-1*	*0.317*	*hyperbolic tangent*	*logistic*
Bed load→ Inputs: $Q, S, V, d, W, D_{50}, A, Sus\text{-}load$					
QP	15	8-6-9-1	0.441	hyperbolic tangent	logistic
L-M	*12*	*8-5-7-1*	*0.383*	*logistic*	*logistic*
QN	11	8-11-1	0.403	hyperbolic tangent	hyperbolic tangent
MO	14	8-9-5-1	0.426	logistic	hyperbolic tangent
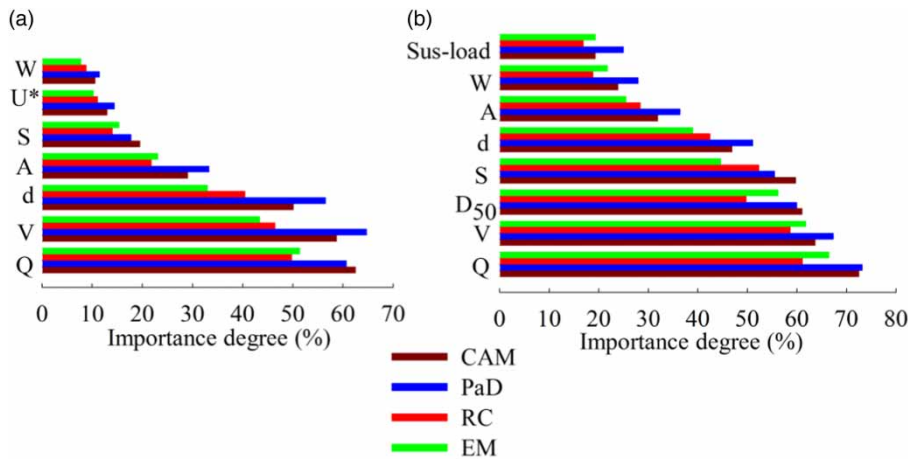
Figure 5 | The calculated effectiveness of input variables using CAM, PaD, RC, and EM techniques for (a) suspended load and (b) bed load.

Table 3 | Comparison of optimum and updated models

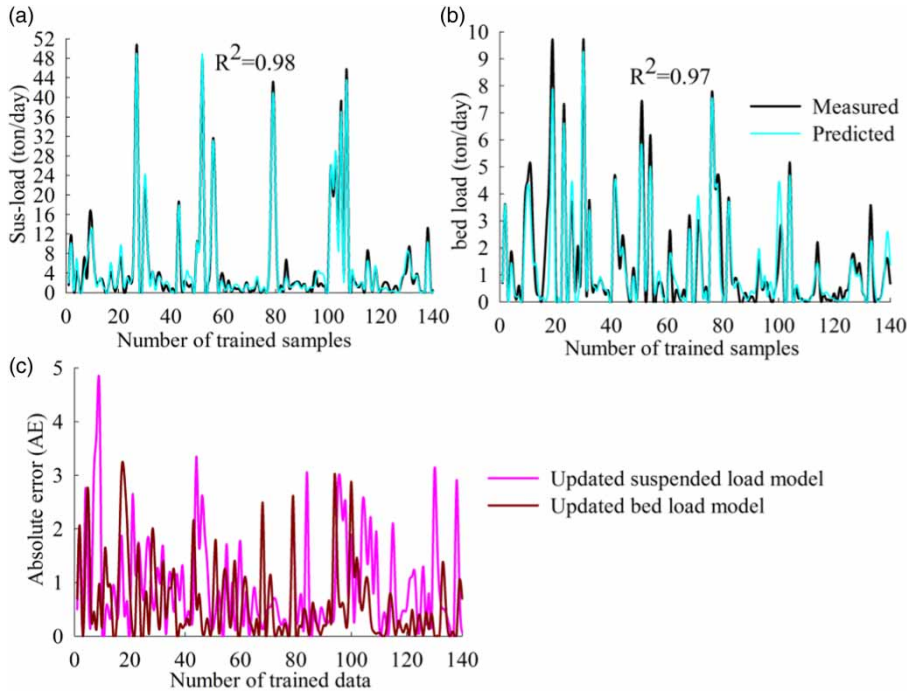|	Model	Topology	Activation function	Training algorithm	MRMSE
Suspended load	*optimum*	7-5-6-1	*hidden layer: hyperbolic tangent* *output: logistic*	MO	0.317
	updated	*4-6-1*	*hidden layer: logistic* *output: logistic*	**MO**	*0.198*
Bed load	*optimum*	8-5-7-1	*hidden layer: logistic* *output: logistic*	L-M	0.383
	updated	*5-8-1*	*hidden layer: logistic* *output: hyperbolic tangent*	QN	*0.201*

Figure 6 │ Results of updated models subjected to training procedure for (a) suspended load, (b) bed load, and (c) calculated *AE*.
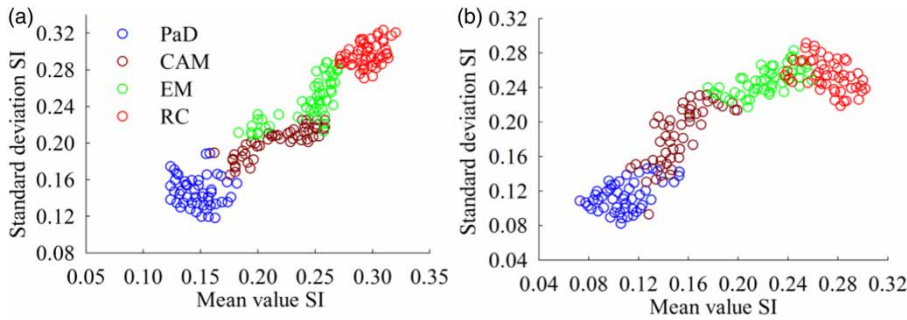

Figure 7 │ Comparison of robustness of used SA methods in this study for (a) suspended load and (b) bed load.
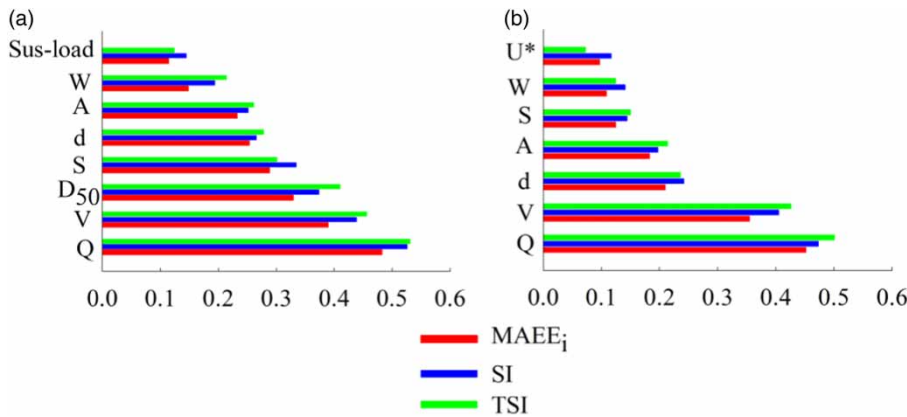

Figure 8 │ The calculated sensitivity indices of input variables on predicted bed (a) and suspended load (b).

The man absolute values of EE_i (MA_{EE_i}) can be expressed as:

$$MA_{EE_i} = \frac{1}{n}\sum_{i=1}^{n}|EE_j^{(i)}| \tag{21}$$

where $EE_j^{(i)}$ denotes the EE of jth variable at the ith repetition. The larger value of MA_{EE_i} shows more influence and contribution of jth input on output. As presented in Figure 8, the ranked MA_{EE_i}, SI and TSI indices (Saltelli *et al.* 2008) for both suspended and bed loads indicate similar trend as the used SA techniques.

In intelligence models, performance measurement is an essential task. The AUC-ROC (area under curve-receiver operating characteristics) curve is one of the most important evaluation metrics in illustrating the diagnostic ability of a classifier system, where the higher the AUC, the higher predictability in the model. The ROC curve is created by plotting the true positive rate (TPR) against the false positive rate (FPR) at various threshold settings.

TPR is the measured percentage of actual positives which are correctly identified. In statistics, when performing multiple comparisons, a false positive ratio is the probability of falsely rejecting the null hypothesis for a particular test. The FPR usually refers to the expectancy of the false positive ratio. As presented in Figure 9, increased AUC-ROC of the updated models is an indicator of model improvement. This also implies that by removing the least effective factors the predictability of models has been increased.
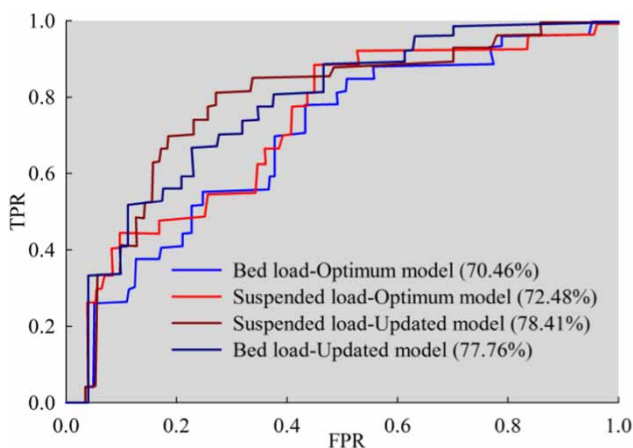

Figure 9 | AUC-ROC of the optimum and updated models.

The capability of updated and optimum models in covering a new set of feed data points can be interpreted by confidence intervals and prediction bands. These factors can reflect the region of uncertainties in the predicted or single additional observation values over a range of independent variables. Therefore, aggregation of data in a higher percentage of these factors indicates better model performance. The reflected results of updated models for suspended and bed loads subjected to 95% confidence interval and prediction bands using validation datasets assign higher predictability and consequently better performance than optimum structures (Figure 10(a) and 10(c)). To evaluate the accuracy performance, the calculated residuals (CR) (Figure 10(e) and 10(f)), measured and predicted values (Figure 10(b) and 10(d)), as well as $MRMSE$ and R^2 (Table 3) were compared to each other. The CR is the difference between the measured and predicted values and thus better performance can be found in higher values of R^2 as well as lower CR and $MRMSE$ (Figure 10 and Table 3). Decreasing the tolerances of CR and $MRMSE$, as well as increasing R^2, is obvious evidence of improvement in the predictability level of updated models (Figure 10(e) and 10(f) and Tables 2 and 3).

CONCLUSION

Modeling of sediment loads as a very complex nonlinear behavior is a difficult task in river engineering. In the current paper, two predictive ANN-based models for suspended and bed loads of three rivers in Idaho, USA were successfully developed and examined. These models, using nine input parameters, covered the channel geometry, geomorphological features, and hydraulic characteristics. To overcome the complexity of the introduced models, four different SA methods, CAM, EM, RC, and PaD, were applied and two updated models with smaller size using the highest ranked inputs were introduced. It was observed that the best performance before applying the SA methods decreased from 11 and 12 neurons to 6 and 8, respectively. Accordingly, the calculated $MRMSE$ values for suspended (0.317) and bed (0.383) loads after updating were reduced to 0.198 and 0.201. This implies a 37.54% and 47.54% decreasing of $MRSME$ in the updating process of suspended
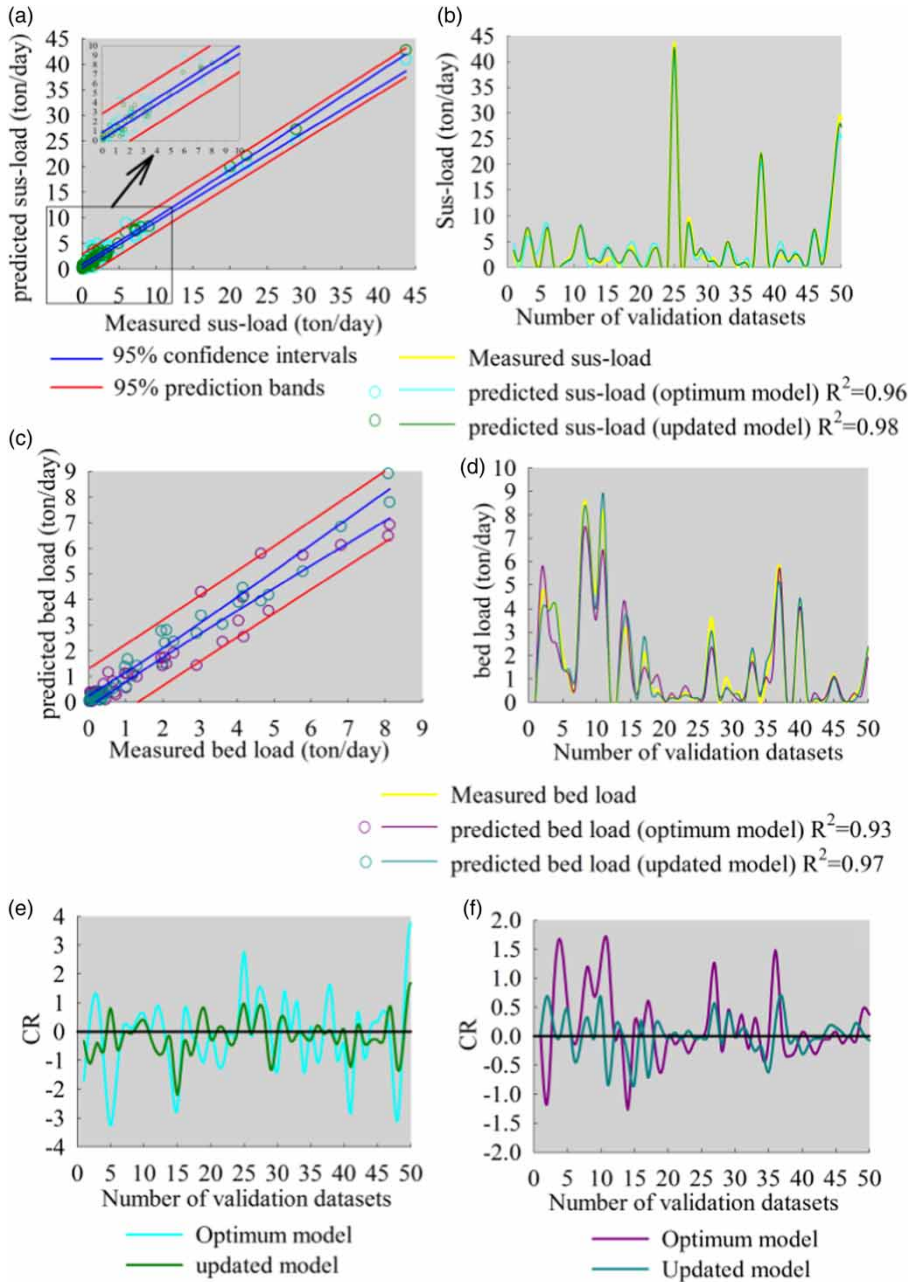
Figure 10 │ Results of optimized models based on 95% confidence interval and prediction bands of randomized datasets (a and c), predictability level suspended and bed loads using validation datasets (b and d), and compared CR of optimum and updated models for suspended (e) and bed loads (f).

and bed load predictions, which show a more superior performance than optimum models. Furthermore, decreasing the CR and AE as well as increasing the R^2 values (2.04% in suspended and 3.1% in bed load) exhibited robust improvement in predictability of the updated models. Accordingly, the interpreted confidence and prediction

intervals due to the presence of high aggregation of data in a more shrunk region of uncertainties demonstrated better consistency in updated models. Furthermore, comparing the performance of models using AUC-ROC as one of the most important evaluation metrics showed 9.39% and 7.56% improvement in accuracy level of bed and suspended

loads, respectively. Such increasing in the covered AUC-ROC, as an indicator, confirmed that the predictability of updated models by removing the least effective factors can significantly be enhanced.

Although the contribution of input parameters on output according to the used SA techniques showed similar trend, the analyses indicated that the results of PaD and CAM were more reliable than EM and RC. The results of the applied SA methods were then verified using $MAEE_i$, SI, and TSI indices and similarly the Q, V, S, d, and D_{50} (for bed load) and Q, V, A, and d (for suspended load) were recognized as the most effective factors on transported sediment loads. The influence of U^* and W were evaluated as the least effective. To have insight and better understanding in transported sediment process, the effect of suspended load on bed loads was also considered. The applied SA methods showed that the effect of suspended load on bed load is not significant and thus can be ignored in bed load predictions.

The results of this study in distinguishing the critical and effective variables on dynamic nonlinear forecasts will assist decision-makers to know which additional information may need to be obtained. Therefore, appropriate decisions can help in strengthening the model and guide the decision-makers to concentrate on the relevant variables.

REFERENCES

Abbaszadeh Shahri, A. 2016 An optimized artificial neural network structure to predict clay sensitivity in a high landslide prone area using piezocone penetration test (CPTu) data – a case study in southwest of Sweden. *Geotech. Geol. Eng.* **34** (2), 745–758.

Abbaszadeh Shahri, A. & Asheghi, R. 2018 Optimized developed artificial neural network based models to predict the blast-induced ground vibration. *Innovative Infrastruct. Solutions* **3**, 34. doi:10.1007/s41062-018-0137-4.

Abbaszadeh Shahri, A., Spross, J., Johansson, F. & Larsson, S. 2019 Landslide susceptibility hazard map in southwest Sweden using artificial neural network. *Catena* **183**, 104225. doi:10.1016/j.catena.2019.104225.

Afan, H. A., El-Shafie, A., Yaseen, Z. M., Hameed, M. M., Mothar, W. H. M. W. & Hussain, A. 2015 ANN based sediment prediction model utilizing different input scenarios. *Water Resour. Manage.* **29** (4), 1231–1245. doi:10.1007/s11269-014-0870-1.

Ahmad, M. M., Ghumman, A. R., Ahmad, S. & Nisar, H. H. 2010 Estimation of a unique pair of Nash model parameters: an optimization approach. *Water Resour. Manage.* **24** (12), 2971–2989.

Asheghi, R. & Hosseini, S. A. 2020 Prediction of bed load sediments using different artificial neural network models. *Front. Struct. Civ. Eng.* doi:10.1007/s11709-019-0600-0.

Asheghi, R., Abbaszadeh Shahri, A. & Khorsand Zak, M. 2019 Prediction of uniaxial compressive strength of different quarried rocks using metaheuristic algorithm. *Arabian J. Sci. Eng.* **44** (10), 8645–8659. doi:10.1007/s13369-019-04046-8.

Bahremand, A. & De Smedt, F. 2008 Distributed hydrological modeling and sensitivity analysis in Torysa watershed, Slovakia. *Water Resour. Manage.* **22** (3), 293–408. doi:10.1007/s11269-007-9168-x.

Bertsekas, D. P. 1995 *Nonlinear Programming*. Athena Scientific, Belmont, MA, USA.

Borgonovo, E. & Plischke, E. 2016 Sensitivity analysis: a review of recent advances. *Eur. J. Oper. Res.* **248**, 869–887.

Bouzeria, H., Ghenim, A. N. & Khanchoul, K. 2017 Using artificial neural network (ANN) for prediction of sediment loads, application to the Mellah catchment, northeast Algeria. *J. Water Land Dev.* **33** (IV–VI), 47–55. doi:10.1515/jwld-2017-0018.

Cacuci, D. G., Ionescu-Bujor, M. & Navon, M. 2005 *Sensitivity and Uncertainty Analysis: Applications to Large-Scale Systems*, Vol. II. CRC Press, Boca Raton, FL, USA.

Calver, A. 1988 Calibration, sensitivity and validation of a physically-based rainfall-runoff model. *J. Hydrol.* **103**, 103–115.

Camenen, B. & Larson, M. 2008 A general formula for non-cohesive bed load sediment transport. *J. Coast. Res.* **24** (3), 615–627.

Campolongo, F., Cariboni, J. & Saltelli, A. 2007 An effective screening design for sensitivity analysis of large models. *Environ. Modell. Software* **22**, 1509–1518. doi:10.1016/j.envsoft.2006.0. 004.

Cao, M., Alkayem, N. F., Pan, L. & Novák, D. 2016 Advanced methods in neural networks-based sensitivity analysis with their applications in civil engineering. In: *Artificial Neural Networks – Models and Applications* (J. L. G. Rosa, ed.). INTECH Press, pp. 335–354. doi:10.5772/64026.

Czitrom, V. 1999 One-factor-at-a-time versus designed experiments. *Am. Stat.* **53**, 126–131.

Dimopoulos, Y., Bourret, P. & Lek, S. 1995 Use of some sensitivity criteria for choosing networks with good generalization ability. *Neural Processes Lett.* **2** (6), 1–4. doi:10.1007/BF02309007.

Fahlman, S. E. 1988 Faster-learning variations of backpropagation: An empirical study. In: *Proceedings of the 1988 Connectionist Models Summer School* (D. Touretzky, G. E. Hinton & T. J. Sejnowski, eds). Morgan Kaufmann Publishers, San Mateo, CA, USA, pp. 38–51.

Gan, Y., Duan, Q., Gong, W., Tong, C., Sun, Y., Chu, W., Ye, A., Miao, C. & Di, Z. 2014 A comprehensive evaluation of

various sensitivity analysis methods: a case study with a hydrological model. *Environ. Modell. Software* **51**, 269–285.

Garson, G. D. 1991 Interpreting neural-network connection weights. *AIExpert* **6**, 46–51.

Gedeon, T. D. 1997 Data mining of inputs: analysing magnitude and functional measures. *Int. J. Neural Syst.* **8** (2), 209–218. doi:10.1142/S0129065797000227.

Gevrey, M., Dimopoulos, I. & Lek, S. 2003 Review and comparison of methods to study the contribution of variables in artificial neural network models. *Ecol. Modell.* **160** (3), 249–264.

Ghaderi, A., Abbaszadeh Shahri, A. & Larsson, S. 2019 An artificial neural network based model to predict spatial soil type distribution using piezocone penetration test data (CPTu). *Bull. Eng. Geol. Environ.* **78**, 4579–4588. doi:10.1007/s10064-018-1400-9.

Griewank, A. 2000 *Evaluating Derivatives, Principles and Techniques of Algorithmic Differentiation*. SIAM, Philadelphia, PA, USA.

Haghnegahdar, A. & Razavi, S. 2017 Insights into sensitivity analysis of earth and environmental systems models: on the impact of parameter perturbation scale. *Environ. Modell. Software* **95**, 115–131. doi:10.1016/j.envsoft.2017.03.031.

Hajbabaei, E., Hosseini, S. A. & Sanei, M. 2017 Bed load pickup rate and flow resistance for turbid flow on a movable plane bed. *Environ. Process.* **4** (1), 255–272.

Hamby, D. M. 1994 A review of techniques for parameter sensitivity analysis of environmental models. *Environ. Monit. Assess.* **32**, 135–154.

Hashem, S. 1992 Sensitivity analysis for feedforward artificial neural networks with differentiable activation functions. In: *Proceedings of the 1992 International Joint Conferences on Neural Networks*, Vol. 1, Baltimore, MD, USA, pp. 419–424.

Helton, J. C., Johnson, J. D., Salaberry, C. J. & Storlie, C. B. 2006 Survey of sampling based methods for uncertainty and sensitivity analysis. *Reliab. Eng. Syst. Saf.* **91**, 1175–1209. doi:10.1016/j.ress.2005.11.017.

Jacomino, V. M. F. & Fields, D. E. 1997 A critical approach to the calibration of a watershed model. *J. Am. Water Resour. Assoc.* **33** (1), 143–154.

Jin, S., Liu, H., Ding, W., Shang, H. & Wang, G. 2018 Sensitivity analysis for the inverted siphon in a long distance water transfer project: an integrated system modeling perspective. *Water* **10**, 292. doi:10.3390/w10030292.

Kisi, O., Hosseinzadeh Dailr, A., Cimenc, M. & Shiri, J. 2012 Suspended sediment modeling using genetic programming and soft computing techniques. *J. Hydrol.* **450–451**, 48–58.

Kumar, B. 2012 Neural network prediction of bed material load transport. *Hydrol. Sci. J.* **57** (5), 956–966. doi:10.1080/02626667.2012.687108.

Leimgruber, J., Krebs, G., Camhy, D. & Muschalla, D. 2018 Sensitivity of model-based water balance to low impact development parameters. *Water* **10**, 1838. doi:10.3390/w10121838.

Lek, S., Belaud, A., Baran, P., Dimopoulos, I. & Delacoste, M. 1996 Role of some environmental variables in trout abundance models using neural networks. *Aquat. Living Resour.* **9** (1), 23–29.

Levenberg, K. 1944 A method for the solution of certain non-linear problems in least squares. *Q. J. Appl. Math.* **2**, 164–168.

Ma, H., Nittrouer, J. A., Naito, K., Fu, X., Zhang, Y., Moodie, A. J., Wang, Y., Wu, B. & Parker, G. 2017 The exceptional sediment load of fine-grained dispersal systems: example of the Yellow River, China. *Sci. Adv.* **3** (5), e1603114. doi:10.1126/sciadv.1603114.

Marquardt, D. W. 1963 An algorithm for least-squares estimation of nonlinear parameters. *J. Soc. Ind. Appl. Math.* **11** (2), 431–441.

Melesse, M. A., Ahmad, S., McClain, M. E., Wang, X. & Lim, Y. H. 2011 Suspended sediment load prediction of river systems: an artificial neural network approach. *Agric. Water Manage.* **98**, 855–866.

Montaño, J. J. & Palmer, A. 2003 Numeric sensitivity analysis applied to feedforward neural networks. *Neural Comput. Applic.* **2** (2), 119–125.

Oakley, J. & O'Hagan, A. 2004 Probabilistic sensitivity analysis of complex models: a Bayesian approach. *J. R. Stat. Soc. B* **66**, 751–769. doi:10.1111/j.1467-9868.2004.05304.x.

Pappenberger, F., Beven, K. J., Ratto, M. & Matgen, P. 2008 Multi-method global sensitivity analysis of flood inundation models. *Adv. Water Resour.* **31**, 1–14.

Paruolo, P., Saisana, M. & Saltelli, A. 2013 Ratings and rankings: Voodoo or Science? *J. R. Stat. Soc. Series A* **176** (3), 609–634. doi:10.1111/j.1467-985X.2012.01059.x.

Poh, H. L., Yao, J. T. & Jasic, T. 1998 Neural networks for the analysis and forecasting of advertising and promotion impact. *Int. J. Intell. Syst. Account., Finance Manage.* **7** (4), 253–268.

Rabitz, H. 1989 Systems analysis at the molecular scale. *Science* **246**, 221–226.

Razavi, S. & Gupta, H. V. 2015 What do we mean by sensitivity analysis? The need for comprehensive characterization of 'global' sensitivity in earth and environmental systems models. *Water Resour. Res.* **51**, 3070–3092. doi:10.1002/2014WR016527.

Ross, T. J. 1995 *Fuzzy Logic with Engineering Applications*. McGraw-Hill, New York, USA.

Saltelli, A. 2002 Sensitivity analysis for importance assessment. *Risk Anal.* **22** (3), 1–12. doi:10.1111/0272-4332.00040.

Saltelli, A., Chan, K. & Scott, E. M. 2000 *Sensitivity Analysis*. Wiley Series in Probability and Statistics, John Wiley & Sons, New York, USA.

Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M. & Tarantola, S. 2008 *Global Sensitivity Analysis: The Primer*. John Wiley & Sons, Chichester, UK.

Sari, V., Castro, N. M. D. R. & Pedrollo, O. C. 2017 Estimate of suspended sediment concentration from monitored data of turbidity and water level using artificial neural networks.

Water Resour. Manage. **31** (3). doi:10.1007/s11269-017-1785-4.

Sobol, I. 1993 Sensitivity analysis for non-linear mathematical models. *Math. Model. Comput. Exp.* **1**, 407–414.

Song, X., Zhang, J., Zhan, C., Xuan, Y., Ye, M. & Xu, C. 2015 Global sensitivity analysis in hydrological modeling: review of concepts, methods, theoretical framework, and applications. *J. Hydrol.* **523**, 739–757.

Storlie, C. B., Swiler, L. P., Helton, J. C. & Sallaberry, C. J. 2009 Implementation and evaluation of nonparametric regression procedures for sensitivity analysis of computationally demanding models. *Reliab. Eng. Syst. Saf.* **94** (11), 1735–1763. doi:10.1016/j.ress.2009.05.007.

Sung, A. H. 1998 Ranking importance of input parameters of neural networks. *Expert Syst. Appl.* **15** (3–4), 405–411.

Swanston, D. J., Bishop, J. M. & Mitchell, R. J. 1994 Simple adaptive momentum: new algorithm for training multiplayer perceptrons. *Electron Lett.* **30**, 1498–1500.

Toriman, E., Jaafar, O., Maru, R., Arfan, A. & Ahmar, A. S. 2018 Daily suspended sediment discharge prediction using multiple linear regression and artificial neural network. *J. Physics: IOP Conf. Series* **954**, 012030. doi:10.1088/1742-6596/954/1/012030.

Vu-Bac, N., Lahmer, T., Zhuang, X., Nguyen-Thoi, T. & Rabczuk, T. 2016 A software framework for probabilistic sensitivity analysis for computationally expensive models. *Adv. Eng. Software* **100**, 19–31.

Wang, W., Jones, P. & Partridge, D. 2000 Assessing the impact of input features in a feedforward neural network. *Neural Comput. Applic.* **9** (2), 101–112. doi:10.1007/ PL00009895.

Wiegerinck, W., Komoda, A. & Heskes, T. 1994 Stochastic dynamics of learning with momentum in neural networks. *J. Phys. A* **27**, 4425–4437.

Zeng, X. & Yeung, D. S. 2003 A quantified sensitivity measure for multilayer perceptron to input perturbation. *Neural Comput.* **15** (1), 183–212. doi:10.1162/089976603321043757.