Soft Computing
https://doi.org/10.1007/s00500-021-05834-x

APPLICATION OF SOFT COMPUTING

A comparative study between dynamic and soft computing models for sediment forecasting

Sarita Gajbhiye Meshram[1] · Hamid Reza Pourghasemi[2] · S. I. Abba[3] · Ehsan Alvandi[4] · Chandrashekhar Meshram[5] · Khaled Mohamed Khedher[6,7]

Accepted: 20 April 2021

Abstract

Runoff–sediment process modeling is highly variable and nonlinear in nature. For sediment yield prediction, the difficulty of rainfall–runoff–sediment yield hydrological processes remains challenging. The present study uses a simple nonlinear dynamic (NLD) model to predict daily sediment yields, taking into account the degree of daily–sediment yield in catchment areas, and its findings were compared to three widely used models including artificial neural networks (ANN), support vector machine (SVM), and gene expression programming (GEP). The daily measured discharge–sediment data for 25 years were obtained from Shakkar Watershed; Central India as in the current study. The coefficient of correlation (CC), Nash-Sutcliff (NS), and root-mean-square error (RMSE) were employed to assess the performance of the models. The results show that the NLD model was found better than ANN, SVM, and GEP model. These models had correlation coefficient (CC = 0.975, 0.887, 0.843, and 0.901), root-mean-square error (RMSE = 0.748, 1.751, 1.961, and 1.545), and Nash–Sutcliffe efficiency (0.952, 0.784, 0.673, and 0.814) correspondingly. Hence, the NLD model can be used for predicting sediment. In order to implement appropriate measures of soil conservation in the watershed to reduce the sediment load in the river, predicting the sediment yield is very necessary to maximize the life of the structure.

Keywords Sediment yield · Runoff · Dynamic model · ANN · SVM · Gene expression programming

1 Introduction

Research on rainfall and runoff produced sediment-based problems would be very helpful in knowing the broad issue of soil degradation and soil erosion in an agricultural area like India, where there are growing pressures on soil and water resources from the inhabitants (Meshram et al. 2019a,b). The planning, designing, and evaluation of land conservation projects, reservoir design and management, environmental and water-pollution measures, and drought and flood control programs are mostly required in the case of information about a suspended sediment yield (wash load) (Meshram et al. 2018a). Information on suspended catchment sediment yields (wash load) is required on several occasions in order to schedule, plan and review land management systems, park design and operation, environmental and water pollution strategies, as well as drought and water control programs.

Various approaches have been proposed to predict soil loss and sediment transport under current and alternate

✉ Sarita Gajbhiye Meshram
gajbhiyesarita@gmail.com

1 Department for Management of Science and Technology Development, Ton Duc Thang University, Ho Chi Minh City, Vietnam

2 Department of Natural Resources and Environmental Engineering, College of Agriculture, Shiraz University, Shiraz, Iran

3 Faculty of Engineering Department of Civil Engineering, Baze University, Abuja, Nigeria

4 Department of Watershed Management, Gorgan University of Agricultural Sciences and Natural Resources, Gorgan, Iran

5 Department of Post Graduate Studies and Research in Mathematics, Jayawanti Haksar Government Post Graduation College of Chhindwara University, Betul, India

6 Department of Civil Engineering, College of Engineering, King Khalid University, Abha, Saudi Arabia

7 Department of Civil Engineering, High Institute of Technological Studies, Mrezgua University Campus, 8000 Nabeul, Tunisia

Published online: 22 May 2021

hydrological conditions. The specific needs for soil loss calculation are so varied that no single model is able to meet the results satisfactorily. Most of these models can be grouped into three broad categories including (i) empirical models based on empirical equations generally derived from field data such as sediment rating curve (Walling 1977), and regression models (Khosla 1953; Flaxman 1972; Singh 1973; Williams 1977; Renard 1980; Narayan and Babu 1983; Garde and Kothyari 1987); (ii) conceptual soil erosion models: Morgan–Morgan–Finney (MMF) and the revised MMF (RMMF) (Morgan 2001), and (iii) theoretical or physically based models; sediment component of SHE-SHESED (Wicks and Bathurst 1996), Chemicals, Runoff, and Erosion from Agricultural Management Systems-CREAMS (Kinsel, 1980), and Areal Non-point Source Watershed Response Simulation-ANSWERS (Beasley et al. 1980). The accessible soil loss measurements can be divided on the basis of storm-wise and annually analyses. Storm-wise models are sediment graph models (Mishra and Ravibabu 2009; Fazli and Noor 2013, 2014). However, soft data mining techniques use alternative simulation approaches distributed and physically oriented.

Black box models, such as ANN, have been promising in various studies owing to it is the ability to handle the chaotic and highly nonlinear system (Rao et al. 2014). Yet, one of the most important steps to evaluate dominant system input variables is when using a computational method. Several researchers demonstrated the potential applications of ANN by taking different input parameters in specific hydrological processes and water resources. The ANN models have also been stated not to be very accurately satisfactory. In most models, the rainfall–runoff and the rainfall–runoff–sediment processes are analyzed and simulated using only the initial data onto hydrological time series.

The runoff–sediment production method of the watersheds is very complicated (Meshram et al. 2018b). The time and spatial variability are extremely nonlinear. The event-based modeling has an important position in watershed management and development (Meshram et al. 2018c). Many models, such as black box, conceptual, and physical techniques, were developed in particular for the rainfall–runoff phase on the other hand, very few models for the reliable estimation of storm event sedimentgraphs. It is worth mentioning that there is a growing need for comprehensive information about river and sediment, due to various conservation, development, and useful utilization programs for all-natural resources including soil and water, over the past few centuries (Meshram et al. 2018d).

Different researchers in the distinct region of India have developed a dynamic model of runoff–sediment yields. Kumar and Das (2000) proposed a model for the simulation of daily runoff and sediment based on a dynamic system for Ramganga (India). In the Naula Watershed Ramganga basin (India), Pyasi and Singh (2001) studied a dynamic model of sediment yield on a weekly basis. In their research, the models focused on the idea to determine and assign the variable weighting to antecedent runoff and sediment occurrences were created for both linear and not linear annual sediment yield forecast. The new model of linear regression was developed by Panigrahi (2007) to estimate sediment yield, with a known value of runoff for Odisha Watersheds. For the Kushinagar Watershed of the Vamsadhara Rivers Catchment, Orissa (in India), a dynamic model of the sediment yields was developed by Ranjan et al. (2011). For this purpose, a linear and nonlinear modeling were developed for the estimation of sediment that involved the idea of defining and assigning the weight of the precipitation-runoff–sediment system for the antecedent event. Kumar et al. (2013) in the Barakar River basin, Jharkhand (India) Giridih Watershed, considered the current runoff and previous levels of runoff and sediment yield as the input variable and subsequently established nonlinear (log–log transformed) sediment yield model to assess the daily catchment sediment yield. For the nonlinear stochastic model, multiple regression coefficients were 0.873. Singh et al. (2019) developed new dynamic AES (advanced encryption standard) algorithm by key-dependent dynamic S-Box (substitution) using dynamic irreducible polynomial and affine constant. This analysis is done on grayscale and color images. Both the images are encrypted and decrypted by using standard AES and dynamic AES.

Artificial neural networks have been applied in almost all branches of science. ANNs are well known for their ability to model nonlinear systems, such as precipitation-runoff, streamflow, and time series analysis (ASCE, 2000). Specific applications are rainfall–runoff modeling (Chen et al., 2013; Tfwala et al., 2013), groundwater management and forecasting (Daliakopoulos et al., 2005; Nourani et al., 2008; Lee et al., 2012; Gorelick and Zheng, 2015; Adamowski and Chan, 2011), streamflow forecasting (Anctil et al., 2004; Besaw et al., 2010), rainfall forecasting (Chiang et al., 2004; Nasseri et al., 2008), suspended sediment prediction (Alp and Cigizoglu, 2007; Kisi and Shiri, 2012), and water quality management (Palani et al., 2008; Faruk, 2010).

SVM is one of the most important machine learning methods utilized in hydrology and has proved to be an alternative to ANNs. The SVM is used for hydrologic forecasts, such as precipitation (Behzad et al., 2009; Okkan and Serbes, 2012), stream flows (Asefa et al., 2006; Liu et al., 2014), sediment (Misra et al., 2009; Ebtehaj et al., 2016), and groundwater fluctuations (Shiri et al., 2013;

Barzegar et al., 2017) and has been found to perform better than ANNs. Yoon et al. (2011) found that the prediction error of SVM for forecasting flood stage was less than of ANN models. Thongsuwan et al. (2020) developed a new deep learning model for classification problems. ConvXGB assessed not only on image data, but also on some general datasets, which used our data preprocessing module.

There were few research studies on the use of GEP in water resources engineering in the literature. Babovic and Keijzer (2002) explored the rainfall–runoff simulation using GEP. Cousin and Savic (1997); Savic et al. (1999); Drecourt (1999), and Crapper (1999, 2001) were the main applicants. GEP was used by Guven et al. (2008) to measure reference evapotranspiration and to use GEP sediment models were recently modeled by Aytek and Kisi (2008). Since the last decade, linear genetic programming (LGP) has been pronounced as a new robust method to solve wide range of modeling problems in water engineering and has been limitedly used in estimation hydrological parameters (Danandeh Mehr et al., 2013, 2018; Mehr et al., 2014; Olyaie et al., 2017; Abba et al., 2020a, b). Guven (2009) applied LGP, a variant of GP, and two versions of neural networks for prediction of daily flow of Schuylkill River in the USA and showed that the performance of LGP was moderately better than that of ANN. Danandeh Mehr et al. (2013) applied LGP in comparison with a neuro-wavelet technique in time series modeling of stream flow on Coruh River in Turkey. According to Mehdizadeh et al. (2020), one of the most important of standalone GEP could be attributed to resolving overfitting problem. Kisi et al. (2013) investigated the ability of GEP, ANFIS and ANN techniques in modeling DO concentration and showed that the GEP model performed better than the ANN and ANFIS models in modeling DO concentration. Londhe and Charhate (2010) used ANN, GP, and model trees (MT) to forecast river flow one day in advance at two stations in Narmada catchment of India. The results showed the ANNs and MT techniques performed almost equally well, but GP performed better than its counterparts. Martí et al. (2013) applied ANN and gene expression programming (GEP)-based models to estimate outlet DO in micro-irrigation sand filters. Due to the importance of this topic, there are still a need to employ several black-box and white-box approach for modeling of rainfall–runoff–sediment. This research aimed to develop a nonlinear dynamic (NLD) model for sediment yield prediction and to compare it with ANN, SVM, and GEP techniques.

2 Materials and method

2.1 Study area

Gadarwada gauging station is one of the gauged watershed of the Shakkar watershed (Fig. 1). The Shakkar River is a major stream of Narmada River. The Shakkar watershed lies between $22^{0}23'$-$23^{0}02'$ N latitude and $77^{0}25'$-$78^{0}52'$ E longitude. The total catchment area of this watershed is 2220 sq. km. The topography of the watershed is undulating. The climate of the Shakkar watershed is dry except

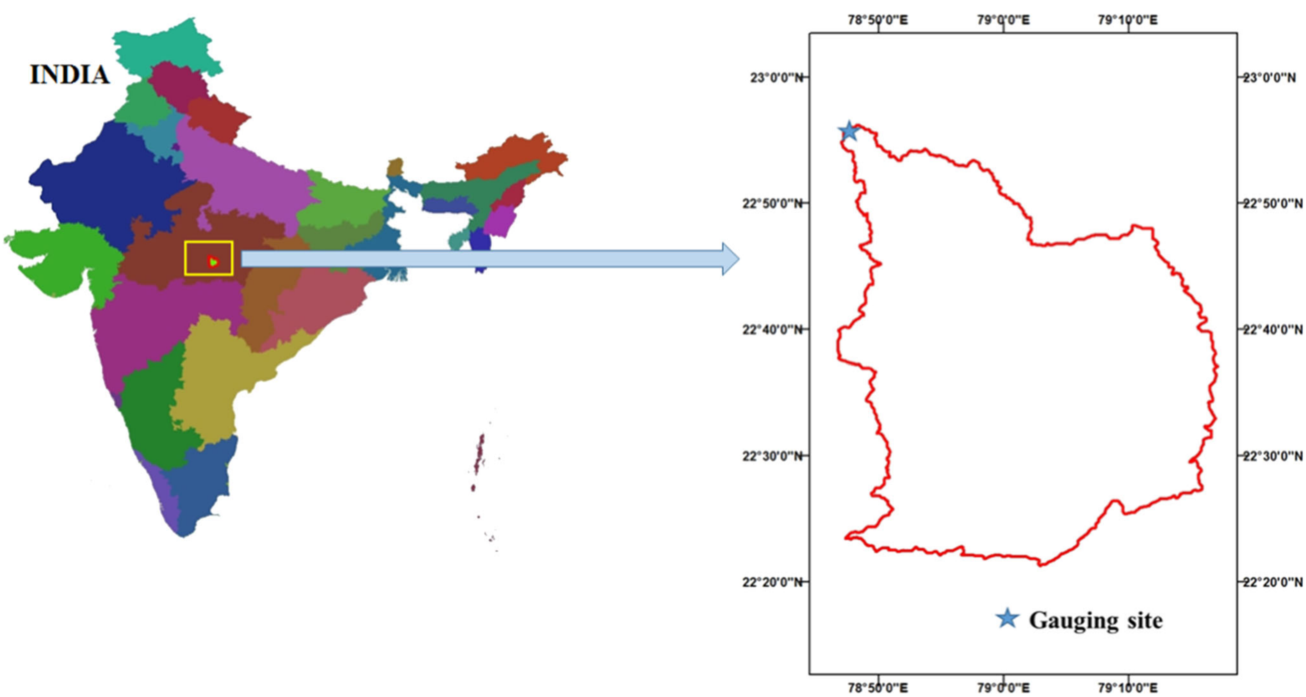


Fig. 1 Location map of the study area

Table 1 The statistical parameters of data used

Data	Period	Length of data	Minimum	Maximum	Median	Mean	Standard deviation
Discharge (Cumec)	June-Oct (1990)- June-Oct (2015)	3945	0	5850	42.33	109.91	271.03
Sediment yield (MT)	June-Oct (1990)- June-Oct (2015)	3945	0	2,682,247.08	263.53	10,131.12	68,135.42

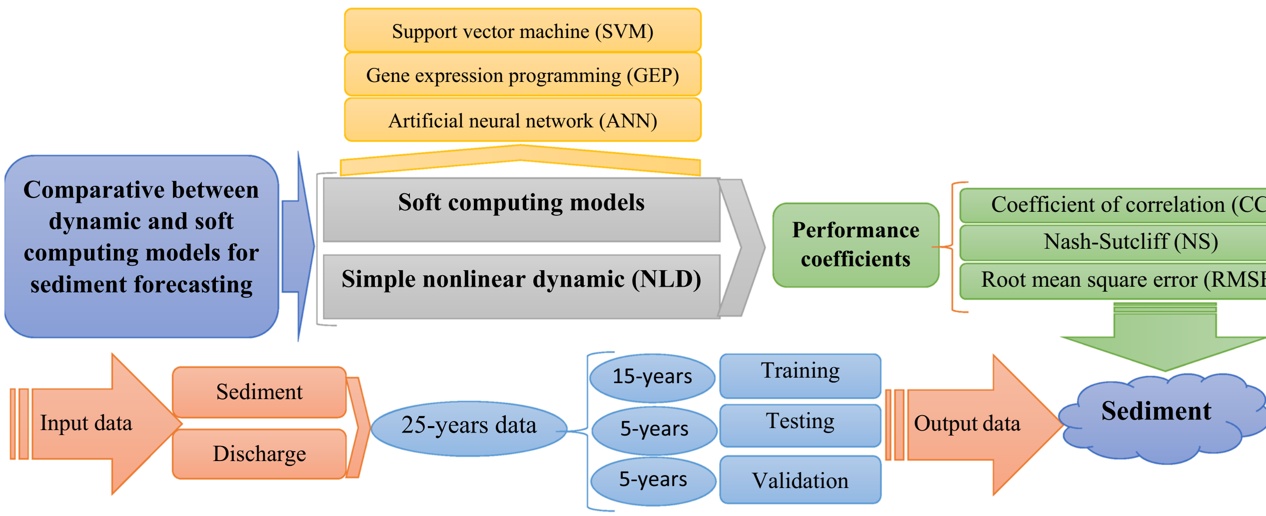


Fig. 2 Flowchart of methodology

the monsoon season. Rainfall mainly occurs from June to October by southwest monsoon (Gajbhiye, 2015). The soil in the watershed can be classified into clay to loamy texture. In 1990, the Central Water Commission (CWC) Bhopal began collecting the hydrological data at Gadarwada Station. During the 1990–2015 study, the daily sediment and runoff data were collected. Table 1 shows the statistical parameters for runoff and sediment data. Figure 2 displays the flow diagram of the technique.

2.2 Input–output data preparation and selection of network architecture parameter estimation

In training and testing/validation of the model, daily discharge and sediment flow data were used. Daily discharge/runoff and sediment output analysis showed that the more than two days of the previous hydrological values attained no substantial effect on sediment output today. Therefore, two and multiple regression equations for sediment prediction were established in this study a maximum lag value.

The data onto multi-layer networks are divided into 60, 20, and 20%, for training, verification, and testing, respectively (Tiwari and Chatterjee 2010; Liu et al., 2013). Thus, daily data onto 25 years (1990–2015) were used for sediment yield prediction. In this way, the data were divided into 60%, 20%, and 20% for training, verification, and testing of the models (ANN, SVM, and GEP), respectively.

2.3 Soft computing models

2.3.1 ANNs

ANN model consists of various processing components or neurons that are strongly interconnected. A collection of inputs to a set of outputs is the primary function in ANN's paradigms. The sigmoid function is the nonlinear activation function most commonly used in ANN. Due to its simplicity and its efficiency, multilayer perceptron feed-forward networks consisting of multiple neuron layers with a supervised learning process were used.

The ANN architecture comprises three layers, i.e., input, output, and hidden layers, as presented in Fig. 3.

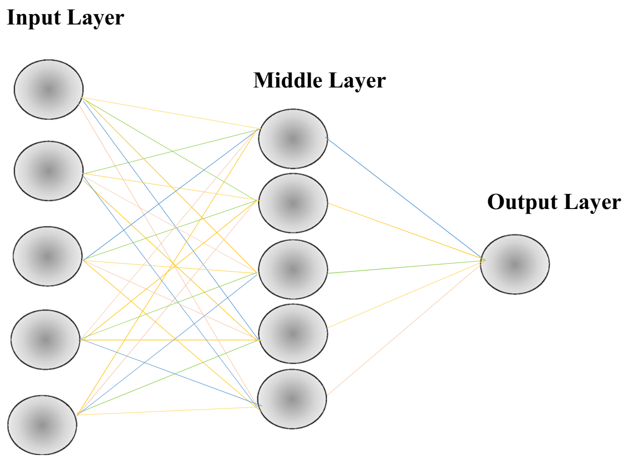


Fig. 3 A schematic structure of ANN

It can be given as below:

$$\gamma_j = f\left(\sum_i \omega_{ij}\mathcal{X}_{ij}\right) \tag{1}$$

where γ_j, $f(.)$, $\mathcal{X}_{ij}$, and ω_{ij} are the output of node j, the transfer function, the signal of input from i $^{\text{th}}$ node in the previous layer to jth node and weight between jth node and ith node, respectively. In this study, feed-forward ANN (FNN) was utilized by employing a Levenberg–Marquardt training algorithm.

Architecture is one of the main characteristics of a neural network layer, and it is proper selection served as a significant element for attaining the best model. Nonetheless, Shu and Ouarda (2007) proposed that less than twice as many hidden nodes be inputs. As such, trial and error methods have been used to determine the best-hidden neuron and architecture in this study.

The main advantage of MLP-ANNs is several kinds of ANN_S (Tiwari and Chatterjee, 2010), but it is less complex than other ANNs. Consequently, the sigmoid feed-forward feature for ANN training has been used (Khalil et al. 2011; Kisi 2011). In the current study, input–output pairs were added to the network of a chosen architecture for training and testing datasets. In order to prevent training or fitting of the model, a testing dataset was used for an early stop approach relevant to the epoch size.

2.3.2 SVM

The SVM devises a computationally proficient method for learning and isolating hyperplanes in a high-dimensional component space (Vapnik, 1995). SVM builds an N-dimensional hyperplane that isolates information into two classifications. SVM models are identified with neural systems. An SVM accomplishes higher arrangement rates in contrast to other characterization strategies. There are

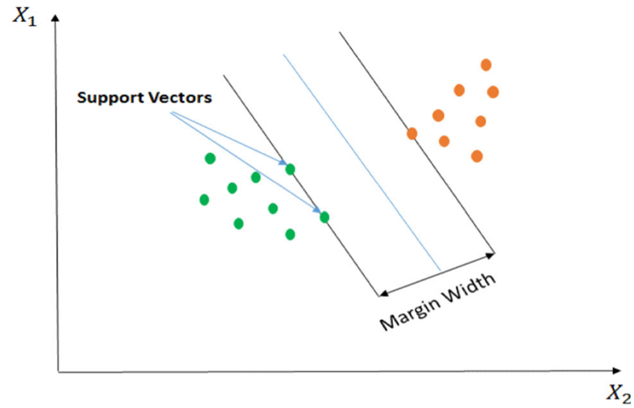


Fig. 4 A schematic representation of SVM (Corinna et al. 1995)

two types of SVM: straight SVM and non-direct SVM (Vapnik, 1999). The straight SVM was utilized here.

SVM is a cutting edge classifier and can speculate direct characterization limits in a multidimensional space (Cortes and Vapnik 1995; Cortes and Vapnik, 1998; Vapnik, 1998). SVM is configured by hyperplane (Cortes and Vapnik 1995) that implies the choice limits named "support vectors" (Fig. 4). The forecasting is done in light of these choice limits. In SVM, exactness is achieved in the estimation of class expectation concerning another informational index which is shaped using an ideal choice limit from the preparation information. Subsequently, the exactness rate is broken down by getting class precision (Cortes and Vapnik 1995). A basic interpretation of the SVM calculation is given afterwards. Given a training set $D = \{x_i\gamma_i\}_{i=1}^n$ with input vectors $x_i = \left(x_i^{(1)}, \ldots x_i^{(n)}\right)^T \in \mathbb{R}^n$ and target labels $\gamma_i \in \{-1, +1\}$, the SVM classifier, as indicated by Vapnik's unique plan, fulfills the accompanying conditions

$$w^T\varnothing(x_i) + \beta \geq +1 \quad if\gamma_i = +1 \tag{2}$$

$$w^T\varnothing(x_i) + \beta \leq -1 \quad if\gamma_i = -1 \tag{3}$$

Which is equivalent to

$$\gamma_i\left[w^T\varnothing(x_i) + \beta\right] \geq 1 \quad i = 1, 2, 3\ldots n$$

where w, β are the weight vector and the bias, respectively. Nonlinear function $\varnothing(.) : \mathbb{R}^2 \to \mathbb{R}^{n_k}$ maps info or estimation space to a high-dimensional, and potentially unending dimensional, highlight space (Min and Lee, 2005).

Application is led with a vector which can be part direct as 1 or—1 with a class portrayal in view of Eqs. 2 and 3 (Cortes and Vapnik 1998). The hyper-planes of specimens are not discovered by only adjoining the line. For superior speculation, they should persevere inside a specific separation. The separation of the closest specimens on the two

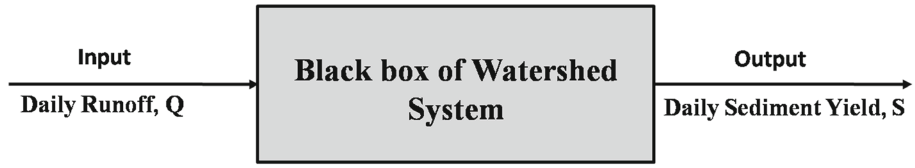


Fig. 5 Schematic representation of runoff and sediment yield process in a watershed

sides of the limit is the edge which ought to be as high as feasible for ideal speculation (Cortes and Vapnik, 1998).

2.3.3 GEP

GEP was suggested as the genetic algorithm genotype/phenotype (Ferreira, 2001). The GEP algorithm method shares many common steps with other evolutionary algorithms. For example, it begins by developing an initial population until the termination criterion is achieved. Instead, according to the preset fitness function, the best chromosome is selected as its final production.

GEPs are genetically modified evolutionary algorithms (Koza, 1999; Parsaie et al., 2017). The GEP computer programs, then expression trees (ETs), all are held in linear chromosomes. ETs are advanced computer programs that solve a specific problem and are selected for fitness-based solutions.

GEPs is a complete system of the kind phenotype that distinguishes the genotype from the phénotype. On the other side, GPs are replicators. GEP solving potential is therefore greatly improved with respect to GP (Ferreira, 2001). By initializing the population in which chromosomes of the individual are generated randomly, GEP begins the search solution. Such chromosomes are independently tested and replicated with modifications depending on their fitness role. The breeding process is repeated until GEP has been established or a solution has been found in a predefined number of generations. Gene expression programming mathematical operators used:

$$\mathrm{F}_1 = \{+, -, *, /, \sqrt{}, Exp, Ln^2, \sqrt[3]{}, Sin, Cos, Atan\} \tag{4}$$

$$\mathrm{F}_2 = \{+, -, *, /\} \tag{5}$$

$$\mathrm{F}_3 = \{+, -, *, /, x^2\} \tag{6}$$

$$\mathrm{F}_4 = \{+, -, *, /, x^3\} \tag{7}$$

2.4 Nonlinear Dynamic model (NLD)

The functional presentations of dynamic-invariant models for runoff–sediment yield [Eqs. (8) and (9)] are as follows:

$$S_t = f(Q_t, Q_{t-1}, Q_{t-2}, \ldots Q_{t-n}.S_{t-1}, S_{t-2}, S_{t-3} \ldots S_{t-n}) \tag{8}$$

In the logarithmic form,

$$\ln S_t = \ln \mathcal{K}_0 + \sum_{i=0}^{n} \mathcal{K}_{Q_i} \ln Q_{t-i} + \sum_{i=1}^{n} \mathcal{K}_{S_i} \ln S_{t-i} \tag{9}$$

where is the respective coefficient of the watershed's lumped results. t' represents the current parameter time value, and t-1, t-2…, t-n are the previous 1, 2…, n time lags within days. The $\mathcal{K}$ coefficients have been determined through the multiple step-by-step regression analysis to identify the sensitivity of these variables. The system analogy of it is expressed as in Fig. 5, where: Q and S are daily runoff and sediment yield, respectively.

In India, especially, the departmental organizations are the majority of watershed managers. Staff are unfamiliar with modern tools that require skills. Nevertheless, nonlinear models for adaptive use can be run with SPSS and/or excel sheet statistic code. In this analysis, therefore, the simple nonlinear dynamic approach has been put more emphasis. For ANN, SVM, and GEP, various combinations of the number of input parameters are chosen, equal to or lower than the maximum number of model parameters calculated in a nonlinear dynamic system by phase regression. Table 2 has been shown the different combination of input for the initial modelling. Then, the three models were compared with a nonlinear dynamic model for runoff–sediment processes to get a better option.

2.5 Parameter estimation

Data have been evaluated in Windows edition 16.0 using the SPSS Statistical Package. One-ways multivariate analyses were used to depart from the Q_{t-2}, Q_{t-1}, Q_t, S_{t-1} and S_{t-2} factors for multi-step regression. In the case of runoff–sediment yield relationship, sediment yield (S_t) as the dependent variable and Q_{t-2}, Q_{t-1}, Q_t, S_{t-1}, S_{t-2} were independent variables.

2.6 Qualitative evaluation of model performance

In this study, three error measures were utilized to assess the quality of prediction models: correlation coefficient (CC), RMSE, and Nash-Sutcliff coefficient (NS) as follows:

Table 2 Different combinations of discharge and sediment for the initial modeling

1	Qt-2	Discharge of two days ago
2	Qt-1	Discharge of previous day
3	Qt	Discharge of current day
4	St-1	sediment of previous day
5	St-2	sediment of two days ago

$$CC = \frac{\sum_{i=1}^{N}(O_i - O_o)(C_i - C_c)}{\sqrt{\sum_{i=1}^{N}(O_i - O_o)^2 \sum_{i=1}^{N}(C_i - C_c)^2}} \tag{10}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(O_i - C_i)^2} \tag{11}$$

$$NS = 1 - \frac{\sum_{i=1}^{N}(O_i - C_i)^2}{\sum_{i=1}^{N}(O_i - C_c)^2} \tag{12}$$

where, O_i and C_i are the observed and calculated values in time step i, respectively. N The number of data, O_o, and C_c are the mean of observed and calculated values, respectively.

3 Results and discussion

Using the daily discharge and sediment yield data series for Shakkar watershed located in the Central India, the NLD, ANN, SVM, and GEP have been developed and evaluated for sediment yield prediction. The entire data were split into training (60%), verification (20%), and testing (20%) sub datasets and MATLAB software was implemented for model construction.

The effectiveness of the models was examined by comparing the forecasted and observed sediment yield data values at the study location. The performance of the models in prediction of sediment yield was estimated using a set of statistical metrics (NS, RMSE, and CC) and visual inspection of data through scatter plots.

3.1 Model development for runoff–sediment yield process

The all four models (NLD, ANN, SVM, and GEP) were developed by considering the daily dataset from 1990 to 2015.

The nonlinear dynamic (NLD) model with the highest R^2 (0.91) was developed as follows for runoff–sediment yield modeling on the basis of the stepwise regression:-

$$\ln S_t = -0.502 + 1.508 \ln Q_t - 0.976 \ln Q_{t-1} - 0.336 \ln Q_{t-2} + 0.647 \ln S_{t-1} + 0.193 \ln S_{t-2} \tag{13}$$

where Q (Cumec) and S (Kg/Sec). As mentioned above, the maximum number of inputs to the ANN, SVM and GEP models for these five independent variables are chosen.

Standardization to improve integrity and reduce the redundancy of data was performed prior to models design. Various single layer networks and two hidden layer networks were trained up to the maximum iterations or periods of 2000, chosen using the minimum values for the root-mean-square error and maximum values for CC and NS, with different combinations of hidden neurons and the best suited network (Agarwal et al. 2006). Th e test datasets used to study the best-performing model by the observations and the simulations of runoff and sediment yield were saved after the training cycle had satisfactorily been completed. For the contrast with the respective observed values, the normalized production values have been reversed to the forecasted values of runoff and sediment yield. In terms of CC, RMSE (Agarwal, 2007), and efficiency (NS) model (Nash and Sutcliffe 1970; Rao et al., 2014), the results of the built models were evaluated. Elkiran et al. (2019); Legates and McCabe, (1999) suggested that at least one' good-of-fit' (i.e. NS) and at least one absolute measure of error (e.g., RMSE) should be included in a systematic model performance assessment. In ANN model, trial and error identification of proper nodes in the hidden layers is crucial for the determination of the best structure; hence, the 5-input network structures 5-5-1 is observed as a better one than the other networks based on performance criterion (Table 3), i.e. 5 inputs, 1 hidden layer of 5 neurons and 1 output period. However, the SVM model showed better performances in comparison with

Table 3 Structure of neural networks and statistical indicators in the training and validation phase

Row	Structure	Training			Validation		
		NS	RMSE	CC	NS	RMSE	CC
1	3-2-1	0.544	4.347	0.678	0.617	3.495	0.685
2	3-5-1	0.583	4.156	0.727	0.675	3.261	0.749
3	4-2-1	0.585	3.152	0.734	0.674	2.981	0.751
4	4-5-1	0.593	3.134	0.745	0.686	2.853	0.768
5	5-2-1	0.636	2.691	0.787	0.760	1.892	0.862
6	5-5-1	0.632	2.652	0.782	0.784	1.751	0.887
7	5-7-1	0.639	2.664	0.791	0.768	1.817	0.871

Table 4 Results of the three kernel types used in the SVM model for training and validation step data

Kernel	Training			Validation		
	NS	RMSE	CC	NS	RMSE	CC
Polynomial	0.540	3.487	0.739	0.661	2.223	0.823
Linear	0.543	3.484	0.741	0.664	2.209	0.829
Radial Basis Function	0.549	3.480	0.745	0.673	1.961	0.843

Table 5 Gene expression programming equations for the four sets of selected mathematical operators

Operator	GEP equations
F_1	$S(t) = \text{Atan}\left[S(t-1) - Q(t)^4\right) * \sin(-1.72)^2)] + \sqrt{\sqrt{Q(t)} * S(t-1)}$
F_2	S(t) = 0.17S(t-1) + 0.055Q(t-1) + 0.09Q(t) + 1.31
F_3	S(t) = 8.12S(t-1) + S(t-1)—0.75Q(t)
F_4	S(t) = -0.62S(t-1) + 0.23Q(t) + 0.37

Table 6 Results of the gene expression programming method using four mathematical function sets

Operator	Training			Validation		
	NS	RMSE	CC	NS	RMSE	CC
F_1	0.564	2.126	0.761	0.814	1.545	0.901
F_2	0.551	2.401	0.746	0.668	2.321	0.815
F_3	0.542	3.264	0.733	0.679	2.216	0.826
F_4	0.542	3.216	0.737	0.703	2.032	0.837

linear and polynomial kernel based on the kernel (Radial Basis Function) (Table 4).

In the GEP model, the mathematical expressions for the forecast sediment S (representing S(t)) are given in terms of antecedent sediment, e.g., S(t-1) and antecedent discharge, e.g., Q(t-1), Q(t-2), present discharge Q(t) are summarized in Table 5. Table 6 summarizes the four mathematical function sets of the GEP model. It is noteworthy that the forecast sediment for both the training and testing sets, operator F1 were found to be performing better compared to other operators. A visual assessment of the predicted and observed sediment during calibration (1990–2005) (Fig. 6) shows that the NLD predicted sediment had the best fit, followed by GEP, ANN and that the SVM fit was the worst.

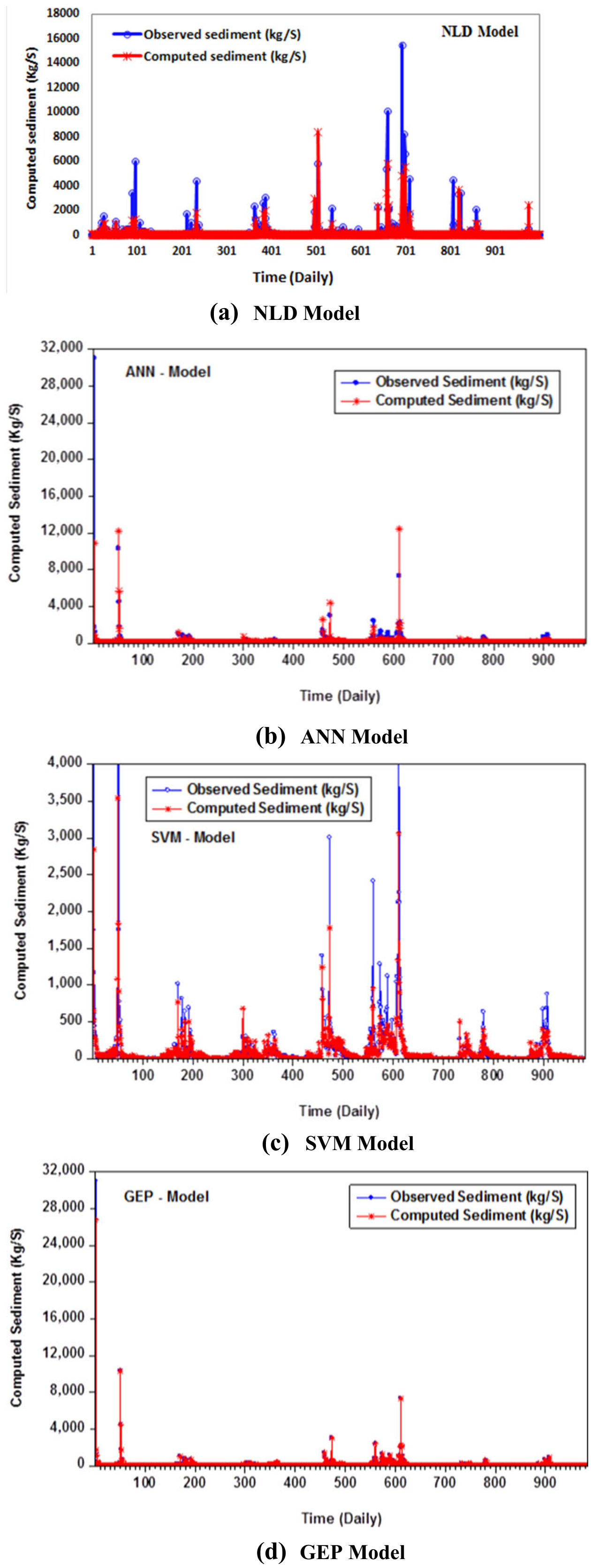
All the four models (NLD, ANN, SVM, and GEP) were tested and verified for their applicability in the study area by applying them on the daily sediment yield data series individually for successive years. The comparison of observed value and predicted value through developed model using second dataset (2006–2015) along with the graphical testing as shown in Fig. 7. The value of qualitative parameters for model developed using the datasets of 1990–2005 is also given in Table 7. It was observed that model perform well for prediction of daily sediment yield which is the necessity for a successful soil conservation programs.

3.2 Model validation for runoff–sediment yield process

All the models (NLD, ANN, SVM, and GEP) were validated for the daily discharge-sediment yield data series from the years 1990 to 2015. The value of the (R^2) coefficient of determinations for all the models was observed to be 0.93, 0.88, 0.84, and 0.90, respectively. R^2 is optimized for variations between mean and variance of measured and expected quantities; it is prone to outliers and must not be utilized exclusively for examination of developed models (Legates and McCabe 1999; Shiri and Kisi 2012). Therefore, alternative error measurement indices (NS, RMSE, and CC) were used for model performance evaluation. The values of NS and RMSE of NLD model were found of (95.20 and 0.748), ANN model were (78.40 and 1.751), SVM model were (67.30 and 1.961), GEP model were (81.40 and 1.545), respectively (Table 7). The NLD and SVM models proved the greatest and least predictive capability, taking into account all the evaluation metrics together. The correlation coefficient (CC) of all model is found better during the testing process than the respective training phase. Validation statistics of all the best-fit models were found to satisfy the criteria of a good model.

3.3 Comparison of prediction models

We investigated the ability of the NLD model for sediment yield prediction and compared with ANN, SVM, and GEP



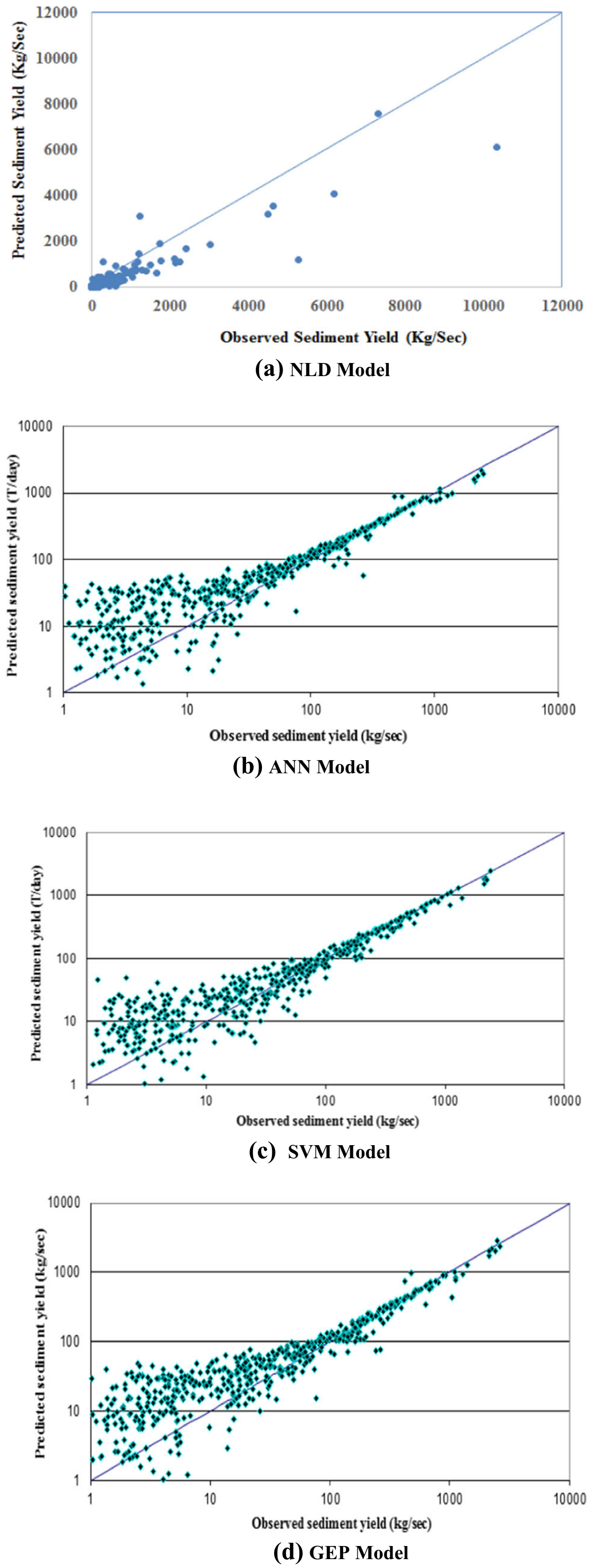
◀**Fig. 6** Distribution graph of observation and calculation values during calibration stage: **a** NLD model, **b** ANN model, **C** SVM model and **d** GEP model

models. The accuracies of the ANN, SVM, GEP, and NLD models were compared using CC, RMSE, and NSE. Table 7 presents the training and test results of the Shakkar watershed. The direct comparison of the results demonstrated the NLD model outperform the three soft computing models, this is not surprising owing to the promising ability in the various literature of hydro-environmental processes. This can be proved by comparing a visual graph of radar chart (spider) that indicates the predictive accuracy in term of CC. The main advantage of the use of a radar chart is to allow for multivariable quantitative analytics and to show the highest and lowest values of the variables in the dataset (Fig. 8).

Even though soft computing model yielded unsatisfactory results in terms of NSE and RMSE but still displayed the remarkable consideration with regard to CC, it can be seen that the CC obtained for all the models was found to be greater than 0.7, which follows to the conclusion reached by Legates and McCabe, (1999); Pham et al. (2019) that CC values higher than 0.70 are considered acceptable; thus the results of all four models are acceptable (Table 5). A closer investigation of the predictive results shows that the prediction accuracy was achieved in the following hierarchical order: NLD > ANN > SVM > GEP in both training and testing phases. The following conclusion can be drawn; NLD is the reliable tool for the prediction of sediment in Gadarwada gauging station, secondly, considering the single model's results and the variation in the models. It can be inferred that more experiments are still necessary to develop the level of agreement between observing and predicting datasets with other soft computing devices.

Besides the above comparison, the quantitative assessment of the four models was carried in testing phase. From the assessment, it can be observed that with regard to goodness-of-fit (NS) NLD model increased the prediction accuracy of soft computing in the following order: SVM (27% and 36%), ANN (16% and 28%), and GEP (13% and 34%), in both training and testing, respectively. It is evident that RMSE range of NLD was found below the range of 0 while the soft computing was with the range of (1 to above 3).

In Table 7, it is shown that NLD model acquired the best CC, RMSE, and NS in the training and testing stages. Results indicated that the NLD model might provide an alternative to the ANN, SVM, and GEP models for predicting sediment yield. Although the AI-based models have been applied and demonstrated promising application in



◀ **Fig. 7** Distribution graph of observation and calculation values during validation stage: **a** NLD model, **b** ANN model, **C** SVM model and **d** GEP model

Table 7 Comparative performance of optimal models

Model	Training			Validation		
	NS	RMSE	CC	NS	RMSE	CC
DLN	0.913	0.929	0.958	0.952	0.748	0.975
SVM	0.549	3.480	0.745	0.673	1.961	0.843
ANN	0.632	3.652	0.792	0.784	1.751	0.887
GEP	0.564	2.126	0.761	0.814	1.545	0.901

many fields of scientific research, still there are some notable challenges that ware attributed to AI-based models. For example, the main drawback of the ANN algorithm model is the poor generalization ability, lack of strict design programs with theoretical foundation, and difficult to control the training process, and slow convergence and issues related to inefficiency. While for SVR, the training data must be manually labelled and the three parameters of SVR model should be adjusted with previous knowledge (Pham et al., 2019). On the hand, nonlinear model (GEP) displayed successful application with high performance accuracy (Danandeh Mehr et al., 2013). According to Mehdizadeh et al. (2020), one of the most important of standalone GEP could be attributed to resolving overfitting problem.

As the literature review shows, there is no unique model to be superior to others in all cases and the performances of different models may be different according to condition of each hoplological system. Therefore, it is tested and verified that the combination of outputs (from different models) through an ensemble method may lead to more accurate results. The idea of such an ensemble model has been already used at different fields of engineering, environmental and water quality modelling (Abba et al. 2020a, b). It is therefore suggested that in order to improve the accuracy of the prediction models an ensemble learning approach should be proposed. Ensemble learning is a machine learning to combine the process of multiple predictors in order to enhance the final performance (Abba et al., 2019; Usman et al., 2020). Ensemble techniques were proved to produce more accurate results than a single model. Ensemble techniques have already applied in several fields such as web ranking, classification and clustering, time series, and regression modeling (Kazienko et al., 2013).

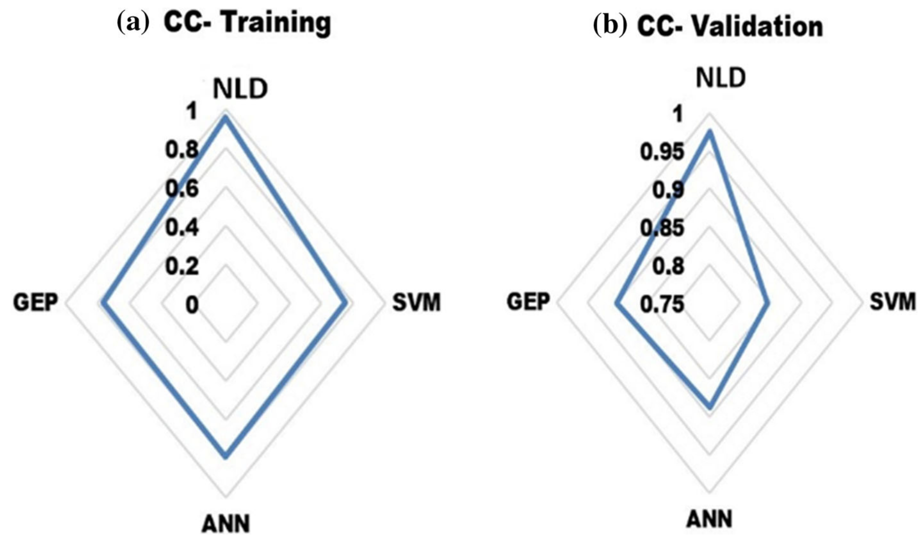


Fig. 8 Radar chart showing the CC for models' comparison in both training and validation

4 Conclusions

The prime aim of this research was to predict the daily sediment yield at Shakkar watershed, Central India, by employing the NLD, and three different soft computing techniques i.e., ANN, SVM, and GEP. The prediction accuracy of these models was estimated using statistical measures and graphical examination. Daily discharge and sediment yield data of one-two ahead historical records are used for the modeling. Different input combinations were examined on all studied models to select the best scenario for further analysis. Comparison of the developed models based on the variety of statistical error measurement indices showed that the NLD and ANN techniques provide better performance for estimating the daily sediment yield and have been performed as the best-ranked 1st and 2nd models, respectively.

Acknowledgements The Authors extend their thanks to the Deanship of Scientific Research at King Khalid University for funding this work through the small research groups under grant number RGP. 1/372/42.

Funding This research work was supported by the Deanship of Scientific Research at King Khalid University under Grant number RGP. 1/372/42.

Declarations

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

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

References

Abba SI, Abdulkadir RA, Gaya MS, Saleh MA, Esmaili P, Jibril MB (2019) Neuro-fuzzy ensemble techniques for the prediction of turbidity in water treatment plant. 2nd International Conference of the IEEE Nigeria Computer Chapter. Nigeria Comput Conf 2019:1–6. https://doi.org/10.1109/NigeriaComputConf45974.2019.8949629

Abba SI, Elkiran G, Nourani V (2020) Non-linear Ensemble Modeling for Multi-step Ahead Prediction of Treated COD in Wastewater Treatment Plant. In: Aliev R., Kacprzyk J., Pedrycz W., Jamshidi M., Babanli M., Sadikoglu F. (eds) 10th International Conference on Theory and Application of Soft Computing, Computing with Words and Perceptions - ICSCCW-2019. ICSCCW 2019. Advances in Intelligent Systems and Computing, vol 1095. Springer, Cham.

Abba SI, Hadi SJ, Sammen SS, Salih SQ, Abdulkadir RA, Pham QB, Yaseen ZM (2020) Evolutionary computational intelligence algorithm coupled with self-tuning predictive model for water quality index determination. J Hydrol 587:124974. https://doi.org/10.1016/j.jhydrol.2020.124974

Adamowski J, Chan HF (2011) A wavelet neural network conjunction model for groundwater level forecasting. J Hydrol 407(1–4):28–40

Agarwal A, Mishra SK, Ram S, Singh JK (2006) Simulation of runoff and sediment yield using artificial neural networks. Biosyst Eng 94(4):597–613

Agarwal BL (2007). Basic statistics. New Age International (P) Ltd., Publishers, New Delhi, 763 PP.

Alp M, Cigizoglu HK (2007) Suspended sediment load simulation by two artificial neural network methods using hydro-meteorological data. Environ Modell Softw 22(1):2–13

Anctil F, Michel C, Perrin C, Andréassian V (2004) A soil moisture index as an auxiliary ANN input for stream flow forecasting. J Hydrol 286(1–4):155–167

ASCE task committee on application of Artificial Neural Networks in hydrology (2000) Artificial neural networks in hydrology 2: hydrologic applications. J Hydrol Eng 5(2):124–137

Asefa T, Kemblowski M, McKee M, Khalil A (2006) Multi-time scale stream flow predictions: The support vector machines approach. J Hydrol 318(1–4):7–16

Barzegar R, Moghaddam AA, Adamowski J (2017) Comparison of machine learning models for predicting fluoride contamination in groundwater. Stoch Environ Res Risk Assess 31:2705–2718. https://doi.org/10.1007/s00477-016-1338-z

Beasley DB, Huggins LF, Monke EJ (1980) ANSWERS: a model for watershed planning. Trans ASAE 23:938–944

Behzad M, Asghari K, Eazi M, Palhang M (2009) Generalization performance of support vector machines and neural networks in runoff modeling. Expert Syst Appl 36(4):7624–7629

Besaw LE, Rizzo DM, Bierman PR, Hackett WR (2010) Advances in ungauged streamflow prediction using artificial neural networks. J Hydrol 386(1–4):27–37

Chen SM, Wang YM, Tsou I (2013) Using artificial neural network approach for modeling rainfall–runoff due to typhoon. J Earth Syst Sci 122:399–405. https://doi.org/10.1007/s12040-013-0289-8

Chiang YM, Chang LC, Chang FJ (2004) Comparison of static-feed forward and dynamic-feedback neural networks for rainfall–runoff modeling. J Hydrol 290(3–4):297–311

Corinna C, Vapnik VN (1995) Support-vector networks. Mach Learn 20(3):273–297

Daliakopoulos IN, Coulibaly P, Tsanis IK (2005) Groundwater level forecasting using artificial neural networks. J Hydrol 309(1):229–240

Dhruvnarayana VV, Babu R (1983) Estimation of soil erosion in India. J Irrig Drain Eng 109(4):419–434

Danandeh Mehr A, Kahya E, Olyaie E (2013) Streamflow prediction using linear genetic programming in comparison with a neuro-wavelet technique. J Hydrol 505:240–249. https://doi.org/10.1016/j.jhydrol.2013.10.003

Danandeh Mehr A, Nourani V, Kahya E, Hrnjica B, Sattar AMA, Yaseen ZM (2018) Genetic programming in water resources engineering: a state-of-the-art review. J Hydrol. https://doi.org/10.1016/j.jhydrol.2018.09.043

Ebtehaj I, Bonakdari H, Shamshirband S, Mohammadi K (2016) A combined support vector machine-wavelet transform model for prediction of sediment transport in sewer. Flow Meas Instrum 47:19–27

Faruk DO (2010) A hybrid neural network and ARIMA model for water quality time series prediction. Eng Appl Artif Intell 23(4):586–594

Fazli S, Noor H (2013) Storm-wise sediment yield prediction using hillslope erosion model in semi-arid abundant lands. Soil Water Res. https://doi.org/10.17221/27/2012-SWR

Flaxman EM (1972) Predicting sediment yield in Western United State. Journal of Hydraulics Engineering ASCE 98(12):2073–2085

Fazli S, Noor H (2014) Prediction of storm-wise soil erosion in dryland farming using a hillslope erosion model. Agric Consp Sci. 79(3):145–149

Garde RJ, Kothari UC (1987) Sediment yield estimation. J Irrig Power (India) 44(3):97–123

Gorelick SM, Zheng C (2015) Global change and the groundwater management challenge. Water Resour Res 51(5):3031–3051

Khosla AN (1953) Silting of reservoir, CBIP. Publication No, N. Delhi, p 51

Kazienko P, Lughofer E, Trawiński B (2013) Hybrid and ensemble methods in machine learning JUCS special issue. J Univ Comput Sci 19(4):457–461

Kinsel WG (1980) A field scale model for chemicals, runoff, and erosion from agricultural management systems. US Dept Agric Conserv Res Rept 26.

Kisi O, Shiri J (2012) River suspended sediment estimation by climatic variables implication: Comparative study among soft computing techniques. Comput Geosci 43:73–82

Kisi O, Akbari N, Sanatipour M, Hashemi A, Teimourzadeh K, Shiri J (2013) Modeling of dissolved oxygen in river water using artificial intelligence techniques. J Environ Inf 22(2):92–101. https://doi.org/10.3808/jei.201300248

Kumar A, Das G (2000) Dynamic Model of Daily Rainfall, runoff and sediment yield for a Himalayan Watershed. J Agric Eng Res 75(2):189–193

Kumar R, Chandola VK, Nema AK, Singh RM (2013) Dynamic model of runoff-sediment yield for giridih watershed of Barakar river basin. Jharkhand. Indian J Soil Conserv 41(2):115–120

Lee S, Song KY, Kim Y (2012) Regional groundwater productivity potential mapping using a geographic information system (GIS) based artificial neural network model. Hydrogeol J 20:1511–1527. https://doi.org/10.1007/s10040-012-0894-7

Liu H, Guo H, Zhang L (2014) SVM-Based Sea Ice Classification Using Textural Features and Concentration From RADARSAT-2 Dual-Pol ScanSAR Data. IEEE J Sel Top Appl Earth Obs Remote Sens 8(4):1601–1613

Londhe S, Charhate S (2010) Comparaison de techniques de modélisation conditionnée par les données pour la prévision des débits fluviaux. Hydrol Sci J 55(7):1163–1174. https://doi.org/10.1080/02626667.2010.512867

Meshram SG, Ghorbani MA, Deo RC, Kashani MH, Meshram C, Karimi V (2019a) New approach for sediment yield forecasting with a two-phase feedforward neuron network-particle swarm optimization model integrated with the gravitational search algorithm. Water Resour Manag 33:2335–2356

Martí P, Shiri J, Duran-Ros M, Arbat G, de Cartagena FR, Puig-Bargués J (2013) Artificial neural networks vs. Gene Expression Programming for estimating outlet dissolved oxygen in micro-irrigation sand filters fed with effluents. Comput Electron Agric 99:176–185. https://doi.org/10.1016/j.compag.2013.08.016

Mehdizadeh S, Ahmadi F, Mehr AD, Jafar M, Safari S (2020) Drought modeling using classic time series and hybrid wavelet-gene expression programming models. J Hydrol https://doi.org/10.1016/j.jhydrol.2020.125017

Mehr AD, Kahya E, Özger M (2014) A gene – wavelet model for long lead time drought forecasting. J Hydrol 517:691–699. https://doi.org/10.1016/j.jhydrol.2014.06.012

Meshram SG, Alvandi E, Singh VP, Meshram C (2019b) Comparison of AHP and fuzzy AHP models for prioritization of watersheds. Soft Comput. https://doi.org/10.1007/s00500-019-03900-z

Meshram SG, Ghorbani MA, Shamshirband S, Karimi V, Meshram C (2018a) River flow prediction using hybrid PSOGSA algorithm based on feed-forward neural network. Soft Comput. https://doi.org/10.1007/s00500-018-3598-7

Meshram SG, Powar PL, Meshram C (2018b) Comparasion of cubic, quadratic and quintic splines for soil erosion modelling. Appl Water Sci 8:173. https://doi.org/10.1007/s13201-018-0807-6

Meshram SG, Powar PL, Singh VP, Meshram C (2018c) Application of cubic spline in soil erosion modelling from Narmada Watersheds. India Arab J Geosci 11:362. https://doi.org/10.1007/s12517-018-3699-8

Meshram SG, Singh SK, Meshram C, Deo RC, Ambade B (2018d) Statistical evaluation of long term time series of rainfall in concurrence with agriculture and water resources of ken river basin. Central India Theor Appl Climatol 134(3–4):1231–1243

Mishra P, Ravibabu R (2009) Simulation of storm sediment yield from an agricultural watershed using MUSLE, remote sensing and geographic information systems. J Soil Water Conserv 8(3):12–21

Misra D, Oommen T, Agarwal A, Mishra SK, Thompson AM (2009) Application and analysis of support vector machine based simulation for runoff and sediment yield. Biosys Eng 103(4):527–535

Nasseri M, Asghari K, Abedini MJ (2008) Optimized scenario for rainfall forecasting using genetic algorithm coupled with artificial neural network. Expert Syst Appl 35(3):1415–1421

Nash JE, Shutcliff JV (1970) River flow forecasting through conceptual models-I. J Hydrol 10:282–290

Nourani V, Mogaddam AA, Nadiri AO (2008) An ANN-based model for spatiotemporal groundwater level forecasting. Hydrol Processess 22(26):5054–5066

Okkan U, Serbes ZA (2012) Rainfall–runoff modeling using least squares support vector machines. Environmetrics 23(6):549–564

Olyaie E, Zare Abyaneh H, Danandeh Mehr A (2017) A comparative analysis among computational intelligence techniques for dissolved oxygen prediction in Delaware River. Geosci Front 8(3):517–527. https://doi.org/10.1016/j.gsf.2016.04.007

Palani S, Liong SY, Tkalich P (2008) An ANN application for water quality forecasting. Mar Pollut Bull 56(9):1586–1597

Panigrahi B (2007) Effect of moisture conservation measures on runoff, soil loss and yield of upland rice. J Agric Eng 44(3):121–127

Pham QB, Abba SI, Usman AG, Thi N, Linh T et al (2019) Potential of hybrid data-intelligence algorithms for multi-station modelling of rainfall. Water Resour Manage 33:5067–5087

Pyasi SK, Singh JK (2001) Weekly sediment yield dynamic model for Naula watershed of Ramganga reservoir. J Agric Eng 38(4):58–65

Ranjan V, Nema AK, Singh A, Bisen Y (2011) Modeling of runoff sediment yield for Kashinagar watershed. Indian J Soil Conserv 39(3):183–187

Rao YRS, Krisha B, Venkatesh B (2014) Wavelet based neural networks for daily stream flow forecasting. Int J Emerg Technol Adv Eng 4(1):307–317

Renard KG (1980) Estimating erosion and sediment yield from rangeland. Proceeding ASCE Symposium on Watershed Management, Australia, Institution of Engineers, pp 162–175

Shiri J, Kisi O, Yoon H, Lee KK, Nazemi AH (2013) Predicting groundwater level fluctuations with meteorological effect implications: a comparative study among soft computing techniques. Comput Geosci 56:32–44

Singh VP (1973) Discussion of "Predicting sediment yield in Western United States." J Hydraul Div 99(10):1891–1894

Singh A, Agarwal P, Chand M (2019). Image encryption and analysis using dynamic AES. In: 5th International Conference on Optimization and Applications (ICOA), Kenitra, Morocco, pp 1–6. https://doi.org/10.1109/ICOA.2019.8727711

Tfwala SS, Wang YM, Lin YC (2013) Prediction of missing flow records using multilayer perceptron and coactive neurofuzzy inference system. Sci World J, Article ID 584516. https://doi.org/10.1155/2013/584516

Thongsuwan S, Jaiyen S, Padcharoen A, Agarwal P (2020) ConvXGB: A new deep learning model for classification problems based on CNN and XGBoost. Nucl Eng Technol. https://doi.org/10.1016/j.net.2020.04.008

Usman AG, Işik S, Abba SI (2020) A novel multi-model data-driven ensemble technique for the prediction of retention factor in HPLC method development. Chromatographia. https://doi.org/10.1007/s10337-020-03912-0

Walling DE (1977) Assessing the accuracy of suspended sediment rating curve for a small Basin. Water Resour Res 13(3):531–538

Wicks JM, Barthurst JC (1996) SHESED: a physically based, distributed erosion and sediment for the SHE hydrological modelling system. J Hydrol 175:213–238

Williams JR (1977) Sediment delivery ratio determination with sediment and runoff models. IAHS Pub No 122:168–178

Yoon H, Jun SC, Hyun Y, Bae GO, Lee KK (2011) A comparative study of artificial neural networks and support vector machines for predicting groundwater levels in a coastal aquifer. J Hydrol 396:128–138

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