International Journal of Physical Sciences Vol. 7(29), pp. 5106 - 5117, 26 July, 2012
Available online at http://www.academicjournals.org/IJPS
DOI: 10.5897/IJPS12.269
ISSN 1992-1950 ©2012 Academic Journals

*Full Length Research Paper*

# Estimating suspended sediment concentration using neural differential evolution (NDE), multi layer perceptron (MLP) and radial basis function (RBF) models

**Mehdi Feyzolahpour*, Masoumeh Rajabi and Shahram Roostaei**

Physical Geography Department, Faculty of Humanities and Social Science, University of Tabriz, Tabriz, Iran.

Accepted 22 May, 2012

**Sediment transport has attracted the attention of engineers from various aspects and different methods have been used for its estimation. So, several experimental equations have been submitted by experts. Though the results of these methods have considerable differences with each other and with experimental observations, because the sediment measures have some limits, these equations can be used in estimating sediment load. With regard to the fact that Givichay River has high sediment production in the region, it is chosen as the study area. This river is one of the Qezeuzan River branches and through this river it joins to the Sefidrud River. Sefidrud dam is one of the most sediment receiver dams in the world and now more than half of the dam capacity has been filled with sediment. With regard to the fact that there are not enough sediment measure stations in this region, therefore different methods have been used. In this study, neural differential evolution (NDE) models are proposed to estimate suspended sediment concentration. NDE models are improved by combining two methods, neural networks and differential evolution. In the first part of the study, NDE model is trained using daily river flow and suspended sediment data belonging to Givi Chay River in northwest of Iran and various combinations of current daily stream flow and past daily stream flow, suspended sediment data are used as inputs to the NDE model so as to estimate current suspended sediment. In the second part of the study, the suspended sediment estimations provided by NDE model are compared with multi layer perceptron (MLP), radial basis function (RBF) and sediment rating curves (SRC) results. The Root mean squared error (RMSE) and the determination coefficient ($R^2$) are used as comparison criteria. Obtained results demonstrate that NDE are in good agreement with the observed suspended sediment concentration; while they depict better results than RBF, MLP and SRC methods. For example, in Givi Chay River station, the determination coefficient ($R^2$) is 0.9621 for NDE model, while it is 0.9114, 0.90 and 0.8447 for RBF, MLP and SRC models, respectively. However, for the estimation of maximum sediment peak, the NDE was mostly found to be better than the RBF and the other techniques. The results also indicate that the NDE may provide better performance than the RBF, MLP and SRC in the estimation of the total sediment load (Re = -26%).**

**Key words:** Givi Chay River, neural differential evolution, multi-layer perceptron model, radial basis function, sediment rating curves.

## INTRODUCTION

Sediment transport by rivers, from land to the seas and oceans are the most important ways of transferring materials in the earth and amounts to 20 to 52 billion tons of sediment transferred by rivers and deposits in resident waters of the world every year. Reducing fertile agricultural lands, quantitative and qualitative decline in pastures, increasing the amount of sediment in rivers and reducing useful life of dam's tanks are among the consequences of soil erosion. Every year large amounts of soil transferred by various factors from basins causes agricultural (the decrease in output unit level), economic (reduction in annual income) and social problems (immigration). Givichay River is not an exception. This river is one of the Qezeuzan River branches and through this river it joins to the Sefidrud River. Sefidrud dam is one of the most sediment receiver dams in the world and now more than half of the dam capacity is filled with sediment. To solve this problem this dam is emptied and filled again for different times from 1980. Also the construction of a few diversion dams in Shahrood and Qezeluzan River has begun. With regard to the fact that the great volume of Sefidrud dam is field by Shahrood and Qezeluzan Rivers sediments and more than 40% of the useful amount of the lake has been reduced, sediment removal method is used in this dam. This way, in addition to the problems such as water level falls, mass sediments slide in to the lake's front, erosion in the dam's walls and the loss in a large number of fish and it also has sediment evacuation problems in the end of the dam. With regard to all cases the necessity of measuring suspended sediment load is obvious.

The suspended sediment load of the stream is generally determined from direct measurement of the sediment load or from sediment transport equations. Although direct measurement is the most reliable method, it is very expensive and the sediment measurement cannot be done for as many streams as the measurement of water discharge. On the other hand, most of the sediment transport equations require detailed information on the flow and sediment characteristics (Ozturk et al., 2001). A number of attempts have been made to relate the amount of sediment transported by a river to flow conditions such as discharge, velocity and shear stress. However, none of the equations derived have received universal acceptance. Usually, either the weight or the concentration of sediment is related to the discharge. These two forms are often used interchangeably. Bean and Nassri (1988) examined load vs. discharge is misleading because the goodness of fit implied by this relationship is spurious. Instead they recommended that the regression link be established. The physically based models are based on the simplified partial differential equations of flow and sediment flux as well as on some unrealistic simplifying assumptions for flow and empirical relationships for erosive effects of rainfall and flow. Examples of such models are presented by Wicks and Bathurst (1996), Refsgaard (1997) and others. These highly sophisticated and complex models have components that correspond to physical processes. They are theoretically capable of accounting for the spatial variation of catchment properties as well as uneven distribution of precipitation and evapotranspira--tion. The model complexity should, however, be keyed to utilizable information about the catchment characteristic and density and frequency of the available input data. In particular, because the real spatial distribution of precipitation is not presently measurable for much of the world, process-oriented distributed models offer no practical advantage over lumped models and have many disadvantages (Guldal et al., 2001).

Artificial neural networks (ANN) are gaining popularity, especially over the last few years, in terms of hydrological applications. Since the early nineties, it has been successfully used in hydrology related areas such as rainfall runoff modeling, stream flow forecasting, ground water modeling, water quality, water management policy, precipitation forecasting, hydrologic time series, and reservoir operations. Most hydrologic processes exhibit a high degree of temporal and spatial variability and are further plagued by issues of non linearity of physical processes, conflicting spatial and temporal scales, and uncertainty in parameter estimates. The time and effort required in developing and implementing such complicated models may not be justified. Simpler neural network forecasts may therefore seem attractive as an alternative tool. White conceptual models are of importance in understanding hydrologic processes. There are many practical situations such as stream flow forecasting where the main concern is with making accurate prediction of specific watershed location. In such a situation, a hydrologist may prefer not to expand the time and efforts required to develop and implement a conceptual model and instead implement a simpler theoretic model. Traditionally, feed forward networks, where nodes in one layer are only connected to nodes in the next layer, have been used for prediction and forecasting applications. In the past decades, great strides have been made in conceptualizing the runoff and sediment yield processes from watersheds through modeling. Models are classified based on their comprehensiveness in representing the physical processes involved. With increasing comprehensiveness, models are classified as black-box models, conceptual models and physically based distributed models. The last of the three can be considered the better choice in a rigorous theoretical sense. However, the significant data need of such models and their marginally superior results compared to the others make them an unfavorable choice in operational hydrology (Gautam et al., 2000). Lumped conceptual models are favored, as they can be based on a sound conceptual framework due to their limited data need. But they require lengthy calibration and parameterization processes. Amongst the soft computing tools vis. genetic algorithm (GA), simulated annealing (SA), multivariate adaptive, regression splines (MARS) and ANN, The ANN, are most frequently used for hydrological modeling.

Jain (2001) used a single ANN approach to establish

*Corresponding author. E-mail: m_mehdi_s2004@yahoo.com. Tel: +984413671922.

daily sediment discharge relationship and found that the ANN model could perform better than the rating curve. Tayfur (2002) developed an ANN model for sheet sediment transport and indicated that the ANN could performs as well as, in some cases better than the physically based models. Cigizoglu et al. (2005) investigated the accuracy of a single ANN in estimation and forecasting of daily suspended sediment data. Kisi (2004) used different ANN techniques for daily suspended sediment concentration prediction and estimation, and indicated that multi-layer perceptron could show better performance than the generalized regression neural networks and radial basis function. Kisi (2005) developed an ANN model for modeling daily suspended sediment and compared the ANN results with those of the rating curve. Cigizoglu et al. (2006) developed some methods to improve ANN performance in daily suspended sediment estimation. Tayfur and Guldal (2006) used MLP for predicting total suspended sediment from precipitation. Bhattacharya et al. (2005) provided an algorithm for developing a data using ANN published flume and field data from several researchers have been employed to build the ANN model. The predictive accuracy of the model was found to be better than well known sediment transport models such as Van Rija and Engeland and Hansen. Raghuwanshi et al. (2006) proposed an ANN model to runoff and sediment yield modeling in Nagwan watershed in India. A five year data set was employed for training and a two year data set was considered for testing the model. Also, other studies conducted in this field were by Cheng et al. (2002), Chau et al. (2005), Muttil and Chau (2006), Lin et al. (2006), Wu et al. (2009), Wang and Traore (2009) and Wang et al. (2009).

In this study, the performance of NDE, RBF, MLP and SRC models are analyzed for daily suspended sediment concentration prediction in the Givi Chay River in the northwest of Iran. The main aim of this study is to investigate capability and accuracy of conventional and artificial intelligence methods in suspended sediment concentration prediction. The paper presents a comparative study on convenient classic and new generation artificial intelligence approaches in suspended sediment concentration modeling. Therefore, it will be of particular utility to researcher that require time-series of suspended sediment concentration and do not have the resources to support sampling or turbidity monitoring and are deciding between various models that predict the needed data from discharge values.

## CASE STUDY

The Givi Chay River originates from southern Ardebil and flows in to the Ghezel Uzan River in Zanjan. The river is 240 km long, and the major landuse is forest and agriculture in its drainage area. Including its tributaries, Givi Chay River drains about 827 km$^2$. Elevations in the basin range from 2820 m above mean sea level at the headwaters of the Hero Chay River to 860 m above at the confluence with the Ghezel Uzan River in Firuzabad city. The climate of the basin is characterized as semiarid with average mean temperature of 6.7°C. Average precipitation in the study area is 444 mm and about 86 to 94% of it occurs in April and May. So the highest amount of sediment belongs to these months and more than 50% of region affected by erosion. In addition, this region is also affected by human factors such as population growth, rapid economic development, deforestation, intensified agriculture, construction of dam's and mining. For example, Forest coverage in 1965 was equivalent to 38% but in 2009 decreases to 21%. All these factors have increased sediments production. From 1992 both discharge and sediment parameters are increasing but the suspended sediment load has increased more quickly. Studies showed that there are nonlinear relations between discharge and sediment parameters. The reason of nonlinear relations is the effect of other factors like geology that can decrease or increase sediment production. Discharge and sediment parameters used as the entry of all artificial neural network models were used in this study. These two parameters can affect suspended sediment production. Melesse et al. (2011) realized that if the data has a lower distribution or be near normal distribution it will have better results. For this purpose to prevent the influence of discharge and sediment in dry period only wet period discharge and sediment was used. Research conducted by the researcher shows that rainfall parameter was less important between the input data. The main reason for this issue is the average estimated rainfall because of the vast basin and no rain measure stations so it has not enough efficiency. Therefore, in this investigation sediment and discharge parameters were used. Despite this, forms of entering data to the above models are different and different entrance structures will be test in them.

The main structure of the entrance data is $Q_t, Q_{t-1}, S_{t-1}$. Where Q is discharge, S and t are sediment and expected day respectively. The flow-sediment time series data of Givi Chay station at northwest of Iran (latitude$37^{\circ}, 41^{\prime}, 33^{''}$, longitude$48^{\circ}, 32^{\prime}, 26^{''}$) operated by the Ardebil regional water corporation were used in the study. Daily time series of river flow and sediment concentration for this station was downloaded from the web server of the Ardebil regional water corporation (www.arrw.ir).

In Givi Chay station for training NDE, RBF, MLP and SRC models wet period data were used from January to June 2009, the rest 180 days which were used for testing the model belongs to the January to June 2010. The statistical parameters of stream flow and sediment

**Table 1.** The daily statistical parameters of dataset for the Givi Chay station.

| Dataset | Data type | $\overline{x}$ | $S_x$ | $C_{Sx}$ | $x_{max}$ | $x_{min}$ |
|---|---|---|---|---|---|---|
| Training | Flow | 79 | 94 | 3.62 | 1029 | 7.6 |
| | Sediment | 387 | 815 | 5.23 | 7650 | 9 |
| Testing | Flow | 52 | 79 | 2.41 | 426 | 5.9 |
| | Sediment | 412 | 951 | 6.08 | 7300 | 8 |

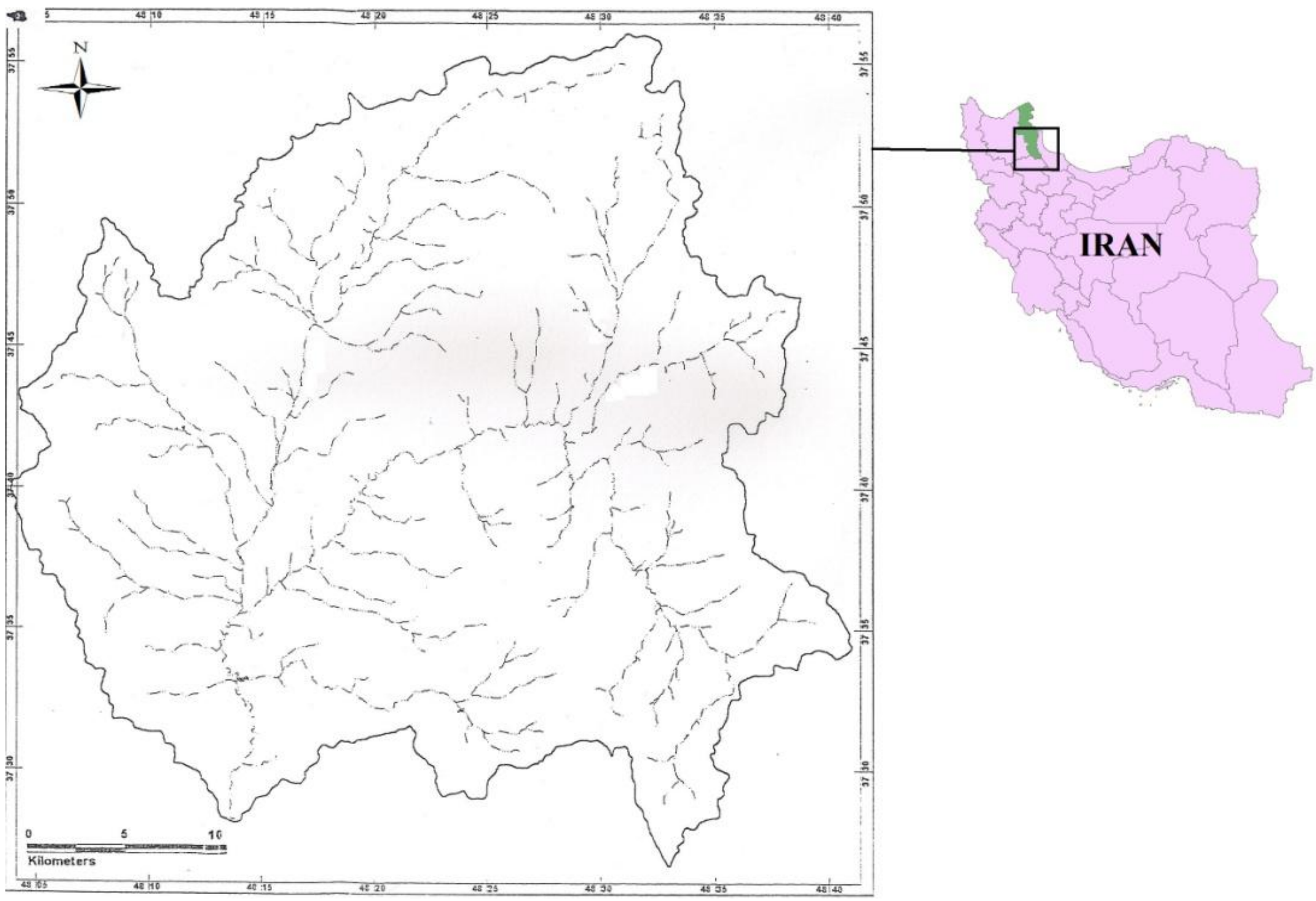


**Figure 1.** The location of Givichay station on the Givichay catchment.

concentration data of Givi Chay station are shown in Table 1. These statistics point out to the complexity of modeling suspended sediment behavior.

Figure 1 shows the drainage basins and sediment monitoring stations used in this study.

## METHODS

### Neural differential evolution (NDE)

The neural differential evolution algorithm can be classified as a floating point evolutionary optimization algorithm (Storn and Price, 1995, 1997; Lampinen, 2001). Generally, the function to be optimized, f, is of the form:

$$f(V): R^D \to R, \tag{1}$$

Here, R denotes the real numbers, and D is the number of parameters of the objective function, f (V) by optimizing the values of its parameters.

$$V = (V_{1,\ldots,}V_D), V \in R^D, \tag{2}$$

Where V denotes a vector composed of D objective function parameters. In the present study, objective function f (V) denotes the mean square errors between the observed and calculated sediment values and $v_i$is the parameter of the fuzzy subsets (membership functions). Usually, the parameters of the objective function are also subject to lower and upper boundary constraints, $v_i^{(L)}$ and$v_i^{(U)}$, respectively:

$$v_i^{(L)} \leq v_i \leq v_i^{(U)} \quad i = 1, \dots, D, \tag{3}$$

Where $v_i$ is the fuzzy subsets parameter and $v_i^{(L)}$and $v_i^{(U)}$ are lower and upper boundaries respectively.

As with all evolutionary optimization algorithms, NDE operates on a population, $P_G$ of candidate solutions, not just a single solution. These candidate solutions are the individuals of the population. In particular, NDE maintains a population, and G is the generation to which the population belongs.

$$P_G = \left(V_{1,G,\dots,}V_{NP,G}\right) \quad G = 0, \dots, G_{max}, \tag{4}$$

Additionally, each vector contains D real parameters (chromosomes of individuals):

$$V_{iG} = \left(v_{1,i,G,\dots,}v_{DiG}\right) \; i = 1, \dots, NP \quad G = 0, \dots, G_{max}, \tag{5}$$

To establish a starting point for optimum seeking, the population must be initialized. Often there is no knowledge available about the location of a global optimum other than the limits of the problem variables. Then a natural way to seed the initial population $P_{G=0}$ is with random values chosen from within the given boundaries:

$$V_{j,i,o} = rand_j[0,1]\left(v_j^{(U)} - v_j^{(L)}\right) + v_j^{(L)} \; i = 1, \dots, NP, j = 1, \dots, D, \tag{6}$$

Where $rand_j[0,1]$ denotes a uniformly distributed random value within the range: [0.0, 1.0] that is chosen for each new j.

NDE self-referential population reproduction scheme is different from other evolutionary algorithms. From the 1st generation onward, vectors in the current population $P_G$ are randomly sampled and combined to create candidate vectors for the subsequent generation, $P_{G+1} = U_{i,G+1} = u_{j,i,G+1}$, is generated as follows:

$$u_{j,i,G+1} = \left\{\begin{matrix}\left\{\left\{\begin{matrix} v_{j,r3,G+F(v_{j,r1,G}-v_{j,r2,G})if\ v_j^{(L)}<y_{j,i,G+1}<v_j^{(U)}} \\ otherwise\ rand_j[0,1](v_j^{(U)} - v_j^{(L)}\end{matrix}\right\} if\ rand_j[0,1] \leq CR\right\} \\ otherwise\ v_{j,i,G}\end{matrix}\right\}, \tag{7}$$

The randomly chosen indices, $r_1, r_2$and $r_3$are different from each other and also different from the running index, i. New random, integer values for, $r_1, r_2$and $r_3$are chosen for each value of the index I, that is for each individual.

F and CR are NDE control parameters. Like NP, both values remain constant during the search process. F is a real-valued factor in the range (0.0, 1.0] that scales the differential variations. The upper limit on F has been empirically determined.

CR is a real- valued crossover factor in the range [0.0, 1.0] that controls the probability that a trial vector parameter will come from the randomly chosen, mutated vector, $u_{j,i,G+1}$instead of from the current vector, $v_{j,i,G}$. Generally, both F and CR affect the convergence velocity and robustness of the search process. Suitable values for F, CR, and NP can usually be found by trial-and-error after a few tests using different values. Practical advice on how to select control parameters NP, F and CR can be found from research work carried out by Lampinen and Zelinka (2000).

NDE selection scheme also differs from other evolutionary algorithms. The population for the next generation,$P_{G+1}$, is selected from the current population, $P_G$, and the child population, according to the following rule:

$$V_{i,G+1} = \left\{\begin{matrix} U_{i,G+1\ if\ f(U_{i,G+1}) \leq f(V_{i,G})} \\ V_{i,G} \quad otherwise\end{matrix}\right\}, \tag{8}$$

Thus, each individual of the temporary population is compared with its counterpart in the current population. Assuming that the objective function is to be minimized, the vector with the lower objective function value wins a place in the next generations' population. As a result, all the individuals of the next generation are as good as or better than their counterparts in the current generation.

## Multi - layer perceptrons (MLP)

An MLP distinguishes itself by the presence of one or more hidden layers, with computation nodes called hidden neurons, whose function is to intervene between the external inputs and the network output in a useful manner. By adding hidden layer(s), the network is enabled to extract higher- order statistic. In a rather loose sense, the network acquires a global perspective despite its local connectivity due to the extra set of synaptic connections and the extra dimension of NN interconnections. The MLP can have more than one hidden layer; however, theoretical works have shown that a single hidden layer is sufficient for an ANN to approximate any complex nonlinear function (Cybenco, 1989; Hornik et al., 1989). Therefore, in this study, a one- hidden- layer MLP is used. Throughout all MLP simulations, the adaptive learning rates were used to speed up training. The numbers of hidden layer neurons were found using simple trial- and- error method in all applications. The MLP were trained using the Levenberg- Marquardt technique as this technique is more powerful than the conventional gradient descent techniques (Hagan and Menhaj, 1994; El-baker, 2003; Cigizoglu and Kisi, 2006).

### *The Levenberg- Marquardt algorithm*

While back – propagation with gradient descent technique is a steepest descent algorithm, the Levenberg- Marquardt algorithm is an approximation to Newton's method (Marquardt, 1963). If a function $V(x)$ is to be minimized with respect to the parameter vector x, then Newton's method would be:

$$\Delta x = -[\nabla^2 V(x)]^{-1} \nabla V(x), \tag{9}$$

Where $\nabla^2 V(x)$ is the Hessian matrix and $\nabla V(x)$ is the gradient. If v(x) reads:

$$V(x) = \sum e_i^2(x), \tag{10}$$

Then it can be shown that:

$$\nabla V(x) = J^T(x) e(x), \tag{11}$$

$$\nabla^2 V(x) = J^T(x) J(x) + S(x), \tag{12}$$

Where J(x) is the Jacobian matrix and,

$$S(x) = \sum e_i \nabla^2 e_i(x), \tag{13}$$

For the Gauss- Newton method it is assumed that S(x) = 0 and

Equation (11) become:

$$\Delta x = [J^T(x)j(x)]^{-1}J^T(x)e(x), \qquad (14)$$

The Levenberg- Marquardt modification to the Gauss- Newton method is:

$$\Delta x = [J^T(x)J(x) + MI]^{-1}J^T(x)e(x), \qquad (15)$$

The parameter M is multiplied by some factor (B) whenever a step would result in an increased V(x). When a step reduces V(x), M is divided by B. when the scalar M is very large the Levenberg-Marqurdt algorithm approximates the steepest descent method. However, when M is small, it is the same as the Gauss- Newton method. Since the Gauss- Newton method converges faster and more accurately towards an error minimum, the goal is to shift towards the Gauss-Newton method as quickly as possible. The value of M is decreased after each step unless the change in error is positive; that is the error increases. For the neural network-mapping problem, the terms in the Jacobian matrix can be computed by a simple modification to the back- propagation algorithm (Hagan and Menhaj, 1994).

### The radial basis function type neural networks (RBF)

The RBF network model is motivated by the locally tuned response observed in biological neurons. Neurons with a locally tuned response characteristic can be found in several parts of the nervous system, for example, cells in the visual cortex sensitive to bars oriented in a certain direction or other visual features within a small region of the visual field (Poggio and Girosi, 1990). These locally tuned neurons show response characteristic bounded to a small range of the input space. The theoretical basis of the RBF approach lies in the field of interpolation of multivariate functions. The objective of interpolating a set of tuples $(x^s, y^s)$with $x^s \in R^d$ is to find a function $F: R^d \to R$ with $F(x^s) = y^s$ for all$s = 1, \dots, N$. In the RBF approach, the interpolating function F is a linear combination of basis functions:

$$F(x) = \sum_{s=1}^{N} w_s \Phi(\|x - x^s\|) + p(x), \qquad (16)$$

Where $\|.\|$ denotes Euclidean norm, $w_{1,\dots,}w_N$ are real numbers, $\Phi$ a real valued function, and $p \in \prod_n^d$. a polynomial of degree at most n (fixed in advance) in d variables. The interpolation problem is to determine the real coefficients $w_{1,\dots,}w_N$ and the polynomial term $p = \sum_{l=1}^{D} a_1 p_j$ where$p_{1,\dots,}p_D$ , is the standard basis of $\prod_n^d$. and $a_{1,\dots,}a_D$ are real coefficients. The interpolation conditions are:

$$F(x^s) = y^s, s = 1, \dots, N, \qquad (17)$$

And

$$\sum_{s=1}^{N} w_s p_j(x^s) = 0, j = 1, \dots, D, \qquad (18)$$

The function $\Phi$is called a radial basis function if the interpolation problem has a unique solution for any choice of data point. In some cases the polynomial term in Equation (18) can be omitted, and by combining it with Equation (17), one obtains:

$$\Phi w = y, \qquad (19)$$

Where w $= (w_{1,\dots,}w_N)$, $y = (y^1, \dots, y^N)$, and $\Phi$ is a $N \times N$matrix defined by:

$$\Phi = (\Phi(\|x^k - x^s\|)), \qquad (20)$$

Provided the inverse of $\Phi$exists, the solution w of the interpolation problem can be explicitly calculated and has the form: $w = \Phi^{-1}y$. The most popular and widely used radial basis function is the Guassian basis function:

$$\Phi(\|x - c\| = e^{-\left(\frac{\|x-c\|}{2\sigma^2}\right)}, \qquad (21)$$

With peak at centre $c = R^d$and deceasing as the distance from the centre increases.

The solution of the exact interpolating RBF mapping passes through every data point $((x^s, y^s)$. In the presence of noise, the exact solution of the interpolation problem is typically a function oscillating between the given data points. An additional problem with the exact interpolation procedure is that the number of basis functions is equal to the number of data points and so calculating the inverse of the $N \times N$ matrix $\Phi$becomes intractable in practice. The interpretation of the RBF method as an artificial neural network consists of three layers: a layer of input neurons feeding the feature vectors in to the network; a hidden layer of RBF neurons, calculating the outcome of the basis functions and a layer of output neurons, calculating a linear combination of the basis function (Taurino et al., 2003). The RBF method does not perform the parameters learning as in the MLP networks, but just performs linear adjustment of the weights for the radial basis. This characteristic of the RBF method gives the advantage of very fast convergence without local minima, since its error function is always convex (Poggio and Girosi, 1990; Lee and Change, 2003).

### Sediment rating curves (SRC)

A rating curve consists of a graph or equation, relating sediment discharge or concentration to stream discharge, which can be used to estimate sediment loads from the stream flow record. The sediment rating curve generally represents a functional relationship of the form:

$$s = aQ^b, \qquad (22)$$

In which Q is stream discharge and S is either suspended sediment load (Sandy, 1990). Values of a and b for a particular stream are determined from data via a linear regression between (logs) and (logQ). After log- transformation to the arithmetic domain and application of the Ferguson (1986) correction factor, the sediment load occurring at a specific discharge can be estimated using the following expression:

$$S = CF.a.Q^b, \qquad (23)$$

Where CF is the log- transformation bias correction factor. Specifically,

$$CF = e^{2.65\sigma^2}, \qquad (24)$$

Where e is the exponential function and $\sigma$ is the standard error of the regression equation.

## APPLICATION AND RESULTS

### Model performance

Some conventional evaluations such as correlation

**Table 2.** The final architectures and RMSE and $R^2$ statistics of the NDE model.

| NDE model inputs | NDE structure | RMSE (mg/l) | $R^2$ |
|---|---|---|---|
| $Qr_t$ | NDE (1,3,1) | 187 | 0.9621 |
| $Qr_t$ and $Qr_{t-1}$ | NDE (2,2,1) | 208 | 0.9267 |
| $Qr_t$ and $Sr_{t-1}$ | NDE (2,2,1 | 219 | 0.8439 |
| $Qr_t, Qr_{t-1}$ and $Sr_{t-1}$ | NDE (3,2,1) | 228 | 0.8226 |

coefficient $(\rho)$, coefficient of determination $R^2$, sum of square error, and root mean square error (RMSE) were critically reviewed by Legates and McCabe (1999). They indicated that the correlation coefficient is unsuitable for model evaluation. Legates and McCabe proposed that a perfect evaluation of model performance should contain at least one good-of-fit or relative error measure for example $R^2$ and at least one absolute error measure for example RMSE. Nash and Sutcliffe (1970) described $R^2$, which has e range of minus infinity to 1, with higher values describing better agreement. In this paper, the performance of the models was evaluated utilizing $R^2$ and RMSE. In brief, the models predictions are optimum if $R^2$ and RMSE are found to be close to 1 and 0, respectively. The $R^2$ parameter clarifies relation between observed and predicted values and RMSE evaluates the residual between observed and predicted suspended sediment concentration. The RMSE are defined as follows:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left[S_{t\ measured} - S_{t\ predicted}\right]^2}, \qquad (25)$$

## Suspended sediment concentration estimation using neural differential evolution (NDE)

The input combinations used in this application to estimate suspended sediment values for Givi Chay River are $(i) Qr_t, (ii) Qr_t, Qr_{t-1}, (iii) Qr_t$ and $Sr_{t-1}$; and $(iv) Qr_t, Qr_{t-1}$ and $Sr_{t-1}$, where $Qr_t$ and $Sr_t$ represent, respectively, the stream flow and sediment concentration at day t.

A program code was written in MATLAB language for the NDE model simulations. Different NDE architectures were tried using this code and appropriate model structures were determined for each input combination. Then, the NDE models were tested and the results were compared by mean of RMSE and $R^2$ statistics. The RMSE and $R^2$ statistics of each NDE model in test period are given in Table 2. The final architectures of the NDE model found after many trials are also provided in this table. For the last input combination, the NDE (3.2.1) has 3, 2 and 1 nodes for the input, hidden and output layers, respectively. According to the results, it can be observe that by first combined entry NDE model has lowest RMSE (187) and the highest $R^2$ (0.9621) values.

## Comparisons of different models versus NDE

In the second part of the study, in order to assess the ability of NDE model in relative to different ANN computing techniques, RBF and MLP models are developed using the best input combination selected in the first application. The other model considered in the study for comparison is SRC.

Pre-processing of the data is usually required before presenting the data samples to the neural network models when the neurons have a transfer function with bounded range. The reasons for scaling of the data samples can be described as to initially equalize the importance of variables and to improve interpretability of network weights (Masters, 1993; Goh, 1995). In this study, the data are scaled in the range of 0.01 and 0.99 by using the following equation:

$$x_s = \frac{0.99(x_i - x_{min})}{(x_{max} - x_{min})} + 0.01, \qquad (26)$$

In Equation (26), $x_s$ is the scaled input value, $x_i$ is the actual unscaled observed input value, $x_{min}$ and $x_{max}$ refer to the minimum and maximum values of the data, respectively.

Two different ANN models, namely RBF and MLP are developed for estimating sediment concentration. For this purpose, two different program codes, including neural networks toolbox, were written in MATLAB language for the RBF and MLP simulations. Different ANN architectures were tried using these codes and the appropriate model structures were determined. Different number of hidden layer neurons and spread parameters was tried for the RBF model in the study. The number of unit for the single hidden layer and the spread parameter value providing the best testing performance of the RBF was found to be 17 and 0.39, respectively. The optimum hidden nodes numbers of the MLP was found to be four after employing trial and error method.

The SRC techniques were applied to the training data set. The following formulas were found to offer the best statistical measures for fit of training dataset,

**Table 3.** The testing performance of the NDE, MLP, RBF and SRC.

| Models | NDE | RBF | MLP | SRC |
|---|---|---|---|---|
| RMSE (mg/L) | 187 | 318 | 330 | 404 |
| $R^2$ | 0.9621 | 0.9114 | 0.90003 | 0.8447 |

respectively:

$$S_t = 1.327 \times (4.268 Q_t^{0.845}), \qquad (27)$$

In which Q is stream discharge and S is suspended sediment concentration.

The NDE model is compared with the MLP, RBF and SRC models in Table 3. The table indicates that the NDE whose input is the current discharge has the smallest RMSE (187 mg/L) and highest $R^2$ (0.9621). It is obvious from Table 3 that the NDE model performs better than the MLP, RBF and SRC models.

A comparison of the observed and estimated suspended sediment concentration in the test period is shown in Figures 2 and 3, in hydrograph and scatter plot form. It can be seen from the hydrographs that the NDE estimates closely follow the observed values.

This is also confirmed by the scatter plots. It can be clearly seen from the scatter plots that the NDE has a higher $R^2$ value (0.9621) than the MLP, RBF and SRC models.

The sediment peak- estimates of the models are compared in Table 4. In general, the NDE model gave better estimates of peak sediment concentration values than the other models. The NDE, RBF, MLP and SRC models predicted the maximum peak as 6549, 5982, 5732 and 5329 with underestimations of 10, 18, 21 and 27%, respectively. As can be seen from Table 4 the SRC give poor estimate for the other peak, too.

The estimation of total sediment load obtained from the estimated suspended sediment concentration values is also considered for comparison due to its importance in reservoir management. The total estimated sediment amounts in test period are given in Table 5. The estimates of the NDE, RBF, MLP and SRC are 26, 35, 52 and 66% lower than the observed value (76842 ton), respectively.

To evaluate the robustness of the NDE, MLP, RBF and SRC models using Akaike information criterion (AIC) defined by Akaike (1974) is utilized:

$$AIC = N \times \ln(RMSE) + 2k, \qquad (28)$$

Where N is the number of samples in the testing set and k is the number of model parameters or weights. Equation (28) indicates that the values of AIC increases as the number of model parameters (k) increases, but if the RMSE of the model is much lower than that of another models, its AIC may be lower despite its relatively layer network size. The AIC values of the models for the test period are given in Table 6. Table 6 shows that NDE has the lowest AIC values for the Givi Chay River.

Despite the number of model weights being higher than SRC, SRC estimates gave worse AIC values. The AIC values of NDE model was slightly lower than the values of the other models.

Overall, the NDE models which combine the two methods, ANN and DE, seem to perform better than the RBF, MLP and SRC models in establishing a rating relationship between suspended sediment and flow. Such problems frequently arise in a non linear manner. However, the SRC technique assumes a linear relationship between the log of sediment and the log of stream flow values since the SRC is obtained by establishing linear regression between the logarithm transformation of the sediment and flow data. This model requires that the variable be normally distributed. It is clear from Table 1 that the stream flow and sediment data have quite a scattered distribution. In view of the complexity of the problem, therefore, the SRC technique is not adequate. The main advantages of using ANN models are their flexibility and ability to model nonlinear relationships. However, the ANN models use back propagation methodology for adjusting the membership function parameters and weights, respectively (Jang, 1993). In back propagation methodology, it is very easy for the training process to get trapped in a local minimum (Kumar et al., 2002; Sudheer et al., 2003). The evolutionary algorithms (EA) belong to a class of search methods with remarkable balance between exploitation of the best solutions and exploration of the search space. They combine elements of stochastic and directed search and, therefore, are more robust than existing direct search methods, providing the global optimum without becoming trapped in local optima (Mantoglu et al., 2004; Karterakis et al., 2007). The NDE proposed in this study uses DE. Unlike conventional evolutionary algorithmswhich depend on a predefined probability distribution function for the mutation process, DE uses the differences of randomly sampled pairs of objective vectors for its mutation process. Consequently, the object vectors differences will pass the objective functions topographical information toward the optimization process, and therefore provide more efficient global optimization capability (Storn et al., 1995, 1997).

## Conclusions

In current study, suspended sediment concentration were

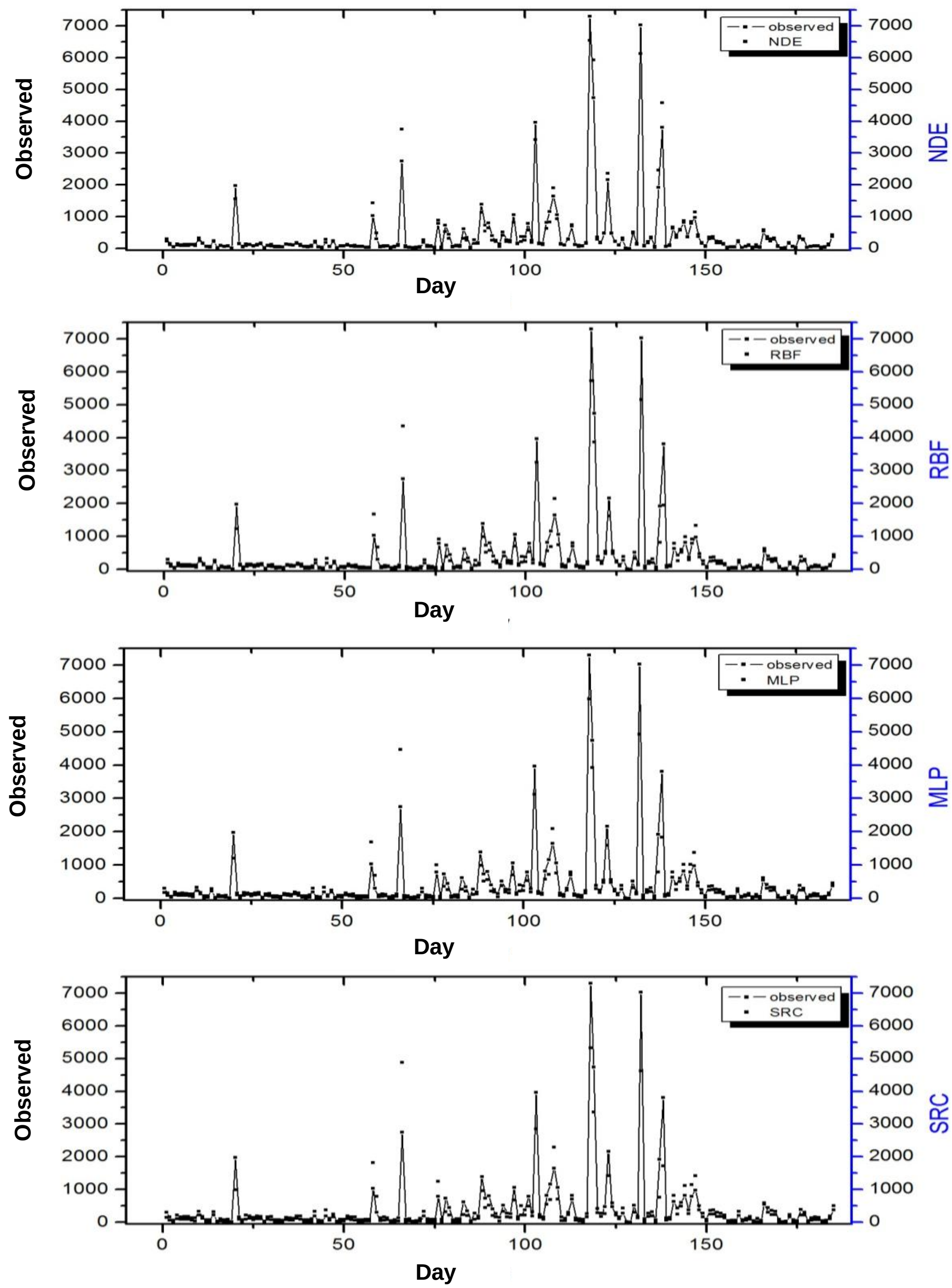


**Figure 2.** A comparison of the observed and estimated suspended sediment concentration by the NDE, MLP, RBF and SRC models for the test period.

estimated by an neural differential evolution (NDE) and two different neural network approach using different combination of hydrological variables (stream flow) and antecedent suspended sediment concentrations. In the first part of the study, several input combinations including daily stream flow and suspended sediment concentration of previous days are used as inputs to the NDE model to estimate current suspended sediment concentration. It is observed that in NDE model the structure with one entry layer, 3 hidden layers and 1

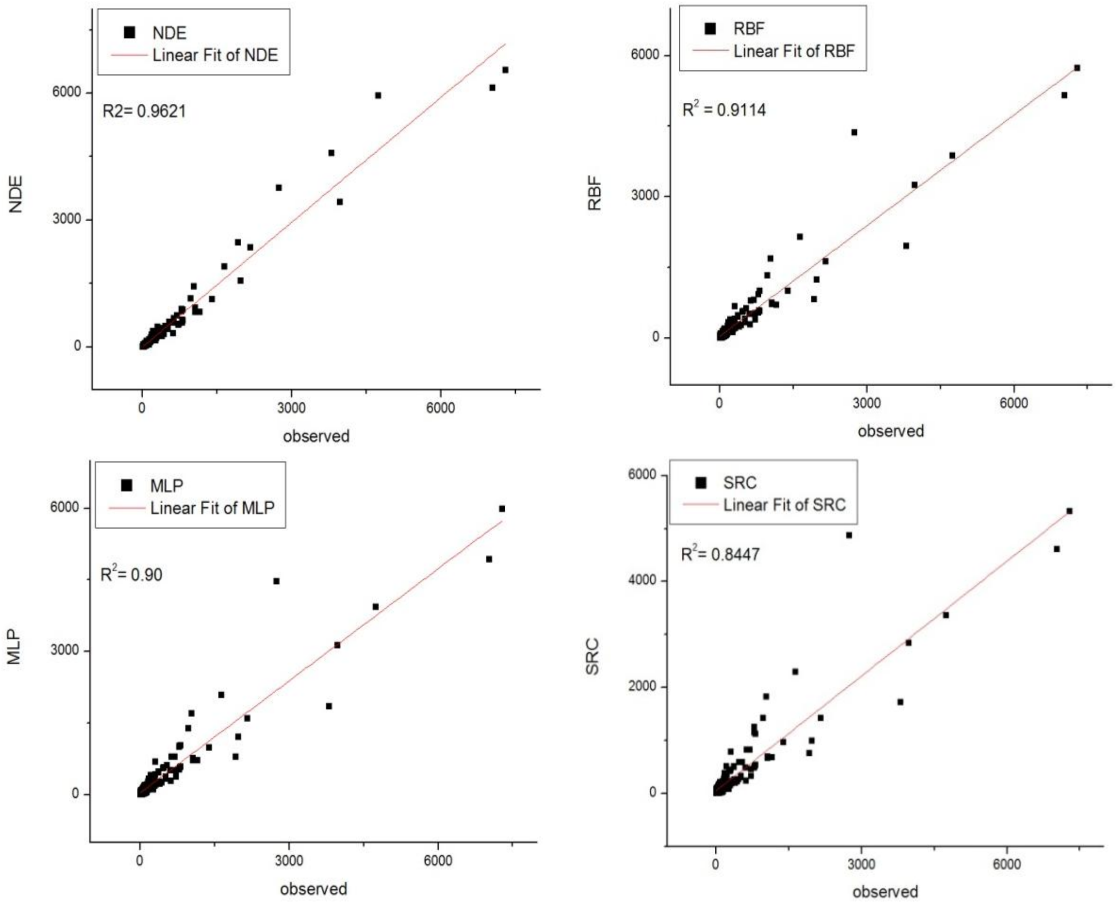


**Figure 3.** Scatter plots of observed and predicted suspended sediment by the NDE, MLP, RBF and SRC models for the test period.

**Table 4.** The comparison of the NDE, MLP, RBF and SRC peak-estimation for the test phase.

| Observed sediment peaks (>3000 mg/L) | NDE | RBF | MLP | SRC | Relative error (%) | | | |
|---|---|---|---|---|---|---|---|---|
| | | | | | NDE | RBF | MLP | SRC |
| 7300 | 6549 | 5982 | 5732 | 5329 | -10 | -18 | -21 | -27 |
| 7036 | 6132 | 5153 | 4923 | 4617 | -12 | -26 | -30 | -38 |
| 4743 | 5937 | 3867 | 3927 | 3364 | 25 | -17 | -18 | -29 |
| 3970 | 3421 | 3245 | 3128 | 2843 | -5 | -18 | -21 | -28 |
| 3806 | 4576 | 1945 | 1837 | 1721 | -20 | 48 | 51 | 54 |

**Table 5.** Estimated total sediment amounts in test period.

| Parameters | Observed | NDE | RBF | MLP | SRC |
|---|---|---|---|---|---|
| Estimate (ton) | 76842 | 97313 | 104243 | 117539 | 128042 |
| Relative error (%) | | -26 | -35 | -52 | -66 |

output layer has the smallest RMSE (187 mg/l) and the highest $R^2$ (0.9621). In the second part of the study, the accuracy of the NDE model was compared with two different ANN computing techniques, MLP and RBF for the best input combination obtained in the first part of the study. The SRC models were also considered for the comparison. The comparison results reveal that the NDE model performs better than the ANN and SRC models in

**Table 6.** AIC values of NDE, RBF, MLP and SRC.

| Model | RMSE (mg/L) | k | AIC |
|---|---|---|---|
| NDE | 187 | 7 | 955 |
| RBF | 318 | 11 | 1059 |
| MLP | 330 | 6 | 1055 |
| SRC | 404 | 3 | 1086 |

daily suspended sediment concentration estimation. The ANN models also provided better estimates than the SRC. After estimating the sediment load and comparing the results from each of the models, quantities that are higher than 3000 mg/L compared with each other. Relative error (RE) had been used for validating the accuracy of models and it is observed that the NDE model had better efficiency in estimating sediment load. For example, this model estimated the 7300 mg/L sediment amount equal to 6549 mg with relative error -10. This is when the above quantities for the RBF, MLP and SRC models are estimated as -18, -21 and -27. Finally in order to determine the RMSE parameter ability the Akaike information system were used in validating each of the models. Investigations showed that the NDE model with 955 Akaike value had the best capability.

Among the ANNs methods, in general, the RBF model was found to be slightly better than those of the MLP method in setting up suspended sediment concentration-hydrological relationship. In general, the NDE model can be considered to be relatively superior to the ANN and SRC models.

The superiority of ANNs over conventional methods in the simulation of sediment load series is evident because the ANNs are able to capture the nonlinear dynamics and generalize the structure of the whole data set. They are a flexible alternative and standard ANN software can be used to construct intricate multi-purpose nonlinear solutions. The method has no limitations in the form of fixed assumptions or formal constraints. The neural network has a distributed processing structure.

The employment of ANN algorithms other than NDE, RBF and MLP should also be investigated in future studies in order to obtain a better fit to the observed data and to remove the negative value production. On the other hand, further information about the hydrological data could also enrich the input data sets of ANNs.

The prediction of suspended sediment loads carries significance for water resource projects like dam reservoir constructions. Therefore, the results of this study, which show ANNs are an important tool in suspended sediment load simulation, could be considered as progress for the solution of this problems.

## REFERENCES

Akaike H (1974). A new look at the statistical model identification. IEEE Trains. Automat. Contr. 19(6):716-723.

Bhattacharya B, Price R, Solomatine D (2005). Data- driven modeling in the context of sediment transport. Phys chem. Earth 30:297-302.

Chau KW, Wu CL, Li YS (2005). Comparison of several flood forecasting models in Yangtze River. J. Hydrol. Eng. ASCE 10(6): 485-491.

Cheng CT, Ou CP, Chau KW (2002). Combining a fuzzy optimal model with a genetic algorithm to solve multiobjective rainfall-runoff model calibration. J. Hydrol. 268(4):72-86.

Cigizoglu HK, Kisi O (2005). Flow prediction by three back propagation techniques using K-fold partitioning of neural network traing data. Nordic Hydrol. 36(1):1-16.

Cigizoglu HK, Kisi O (2006). Methods to improve the neural network performance in suspended sediment estimation. J. Hydrol. 317:221-238.

Cybenco G (1989). Approximation by superposition of a sigmoidal function. Mathematics of control. Signals syst. 2:303-314.

EL-baker MY (2003). Feed forward neural networks modeling for K-P interaction. Chaos Solu. Fractals 18(5):995-1000.

Ferguson RI (1986). River loads underestimated by rating curves. Water Res. Res. 22(1):74-76.

Gautam MR, Watanabe K, Saegusa H (2000). Runoff analysis in humid forest catchment with artificial neural network. J. Hydrol. 235(1-2):117-136.

Goh ATC (1995). Back-propagation neural networks for modeling complex systems. Artif. Intell. Eng. 9:143-151.

Guldal V, Muftuglu RF (2001). 2D unit sediment graph theory. J. Hydrol. Eng. 6(2):132-140.

Hagan MT, Menhaj MB (1994). Training feed forward networks with the Marquardt algorithm. IEEE Trans Neural Netw. 6:861-867.

Hornik K, Stinchcombe M, White H (1989). Multilayer feed forward networks are universal approximators. Neural Netw. 2:359-366.

Jain SK (2001). Development of integrated sediment rating curves using ANNs. J. Hydraul. Eng. ASCE 127(1):30-37.

Jang J (1993). ANFIS: adaptive- network-based fuzzy inference system. IEEE Transactions on systems, Man and cybernetics 23(3):665-685.

Karterakis SM, Karatzas P, nikolos IK, Papadopoulou MP (2007). Application of linear programming and differential evolutionary optimization methodologies for the solution of coastal subsurface water management problems subject to environmental criteria. J. Hydrol. 342(3-4):270-282.

Kisi O (2004). Multi- layer perceptrons with Levenberg- marquardt optimization algorithm for suspended sediment concentration prediction and estimation. Hydrol. Sci. J. 49(6):1025-1040.

Kisi O (2005). Suspended sediment estimation using neuro- fuzzy and neural network approaches. Hydrol. Sci. J. 50(4):683-696.

Kumar M, Raghuwanshi NS, Singh R, Wallender W, Pruitt WO (2002). Estimating evapotranspiration using artificial neural network. J. Irrig. Drain. Eng. 128(4):224-233.

Lampinen J, Zelinka I (2000). On stagnation of the differential evolution algorithm. In: Osmera, Pavel(Ed), proceedings of MENDEL 200, 6th international Mendel conference on soft computing, June 7-9 2000, Brno, Czech Republic. Brno university of Technology, Faculty of Mechanical Engineering, Institute of Automation and computer science, Brno, Czech Republic, pp. 76-83.

Lampinen J (2001). Solving problems subject to multiple non linear constraints by the differential evolution. In: proceeding of MENDEL 2001, 7th international conference on soft computing, June 6-8 2001, brno, Czech Republic, Brno university of technology, faculty of Mechanical engineering, Institute of Automation and computer science, Brno, Czech Republ. pp. 50-57.

Lee GC, Change SH (2003). Radial basis function networks applied to DNBR calculation in digital core protection systems. Ann. Nucl. Energ. 30:1561-1572.

Legates DR, Mc Cabe GJ (1999). Evaluating the use of goodness- of- fit measures in hydrologic and hydro climatic model validation. Water Resour. Res. 35(1):233-41.

Lin JY, Cheng CT, Chau KW (2006). Using support vector machines for long-term discharge prediction. Hydrol. Sci. J. 51(4):599-612.

Mantoglu A, Papuntoniou M, Giannoulopoulos P (2004). Management of coastal aquifers based on nonlinear optimization and evolutionary algorithms. J. Hydrol. 297(1-4):209-228.

Marquardt D (1963). An algorithm for least squares estimation of non-linear parameters. J. Soc. Ind. Appl. Math. 23:431-441.

Masters T (1993). Practical Neural network Recipes C++. Academic press, San Diego, California, USA.

Bean EA, Nassri S (1988). Uncertainty in suspended sediment transport curves, J. Hydrol. Eng. ASCE 114(1):63-74.

Melesse AM, Ahmad S, Mcclain ME, Wang X, Lim YH (2011). Suspended sediment load prediction of river system: an artificial neural network approach. Agric. Water Manage. 98:855-866.

Muttil N, Chau KW (2006). Neural network and genetic programming for modeling coastal algal blooms. Int. J. Env. Pollut. 28(4):223-238.

Nash JE, Sutcliff JV (1970). River flow forecasting through conceptual models part I-a discussion of principles. J. Hydrol. 10(3):282-90.

Ozturk F, Apaydin H, Walling DE (2001). Suspended sediment loads through Flood events for streams of sakarya Basin. Turkish J. Eng. Env. TUBITAK 25:643-650.

Poggio T, Girosi F (1990). Regularization algorithms for learning that are equivalent to multilayer networks. Science 2247:978-982.

Raghuwanshi N, Singh R, Reddy L (2006). Runoff and sediment yield modeling using artificial neural networks: upper Siwane River, India. J. Hydrol. Eng. 11(1):71-79.

Refsgaard JC (1997). Parameterization, calibration and validation of distributed hydrological models. J. Hydrol. 198:69-97.

Sandy R (1990). Statistics for Business and Economics. Mc Graw Hill publishing. New York pp. 1117.

Storn R, Price KV (1995). Differential Evolution- A simple and Efficient Adaptive scheme for Global optimization over continuous spaces, Technical Report, TR-95-012, ICSI, March.

Storn R, Price KV (1997). Differential evolution- a simple and efficient heuristic for global optimization and continuouse space. J. Global Optim. 11(4):341-359.

Sudheer KP, Gosain AK, Ramasastri KS (2003). Estimating actual evapotranspiration from limited climatic data, using neural computing technique. J. Irrig. Drain. Eng. ASCE 129(3):214-218.

Taurino AM, Distante C, Siciliano P, Vasanelli L (2003). Quantitative and qualitative analysis of VOCs mixtures by means of a microsensors array and different evaluation methods. Sens. Actuators. 93:117-125.

Tayfur G (2002). Artificial neural networks for sheet sediment transport. Hydrol. Sci. J. 47(6):879- 892.

Tayfur G, Guldal V (2006). Artificial neural networks for estimating daily total suspended sediment in natural streams, Nordic Hydrol. 37:69-79.

Wang WC, Chau KW, Cheng CT, Qiu L (2009). A comparison of performance of several artificial intelligence methods for forecasting monthly discharge time series. J. Hydrol. 374(4):294-306.

Wang YM, Traore S (2009). Time- lagged recurrent network for forecasting episodic event suspended sediment load in typhoon prone area. Int. J. Phy. Sci. 4(9):519-528.

Wicks JM, Bathurst JC (1996). SHESED: a physically based, distributed erosion and sediment yield component for the SHE hydrological modeling system. J. Hydrol. 175:213-238.

Wu CL, Chau KW, Li YS (2009). Methods to improve neural network performance in daily flows prediction. J. Hydrol. 372(4):80-93.