2013, Scienceline Publication

Journal of Civil Engineering and Urbanism

Volume 3, Issue 4: 136- 141(2013) (Received: January 24, 2013; Accepted: April 15, 2013; Published: July 30, 2013) ISSN-2252-0430

Sediment load estimation by MLR, ANN, NF and Sediment Rating Curve (SRC) in Rio Chama River

Mohamad Ali Ghorbani[1], Seyyed Hasan Hosseini[1]*, Mohamad Hasan Fazelifard[1], Habibeh Abbasi[2]

[1] *Department of Water Engineering, Tabriz University, Tabriz, Iran*
[2] *Azarpeymayesh Consulting Engineers Company, Tabriz, Iran*

*Corresponding Author's E-mail address:hosseini_hasan@yahoo.com

ORIGINAL ARTICLE

ABSTRACT: As the major part of river sediments is suspended sediment load, its estimation has important significance to manage of the water resources and environments. In this study, two conventional models: Sediment Rating Curve (SRC) and Multi Linear Regression (MLR) and two artificial intelligent models Artificial Neural Network (ANN) and Neuro-Fuzzy (NF) are applied to estimate suspended sediment load of the Rio Chama, a major tributary river of the Rio Grande, in the U.S. states of Colorado and New Mexico. Three statistical parameters–coefficients of determination (R^2), root mean square error (RMSE) and Nash-Sutcliffe efficiency (NSE) are used to compare the results of models. The results showed that ANN using only discharge as input and NF model using both discharge and sediment as inputs have better performance than other two models. Furthermore, in this study, mentioned models are applied to evaluate annual sediment load and the best results have been achieved from NF and ANN respectively. Results of this study may be useful in picking up the most suitable modeling approach for similar studies in other river basins.

Keywords: Artificial Neural Network; Neuro-Fuzzy; Regression Analysis; Sediment Rating Curve; Suspended Sediment Load.

INTRODUCTION

Prediction of suspended sediment load has an important significance in management of the water resources and environments. It is necessary to predict the amount of sediment in designing and operation of dams, intakes, channels, navigating in rivers, river training and determination of useful life time of reservoirs and hydroelectric equipment of dams. As directly measuring sediment load of rivers is time consuming and expensive, studies have been made to develop sediment rating curve (SRC), regression methods and artificial intelligence techniques for simulation processes with limited knowledge of the physics. In most rivers, sediments are mainly transported as suspended sediment load (Morris and Fan, 1997). Many models have been provided to simulate this phenomenon. Conventional sediment rating curve and regression models, in which the system is supposed to be static, are often used to estimate suspended sediment load of a river (Asselman, 2000; Jansen and Painter, 1974). Artificial neural network (ANN) which map the inputs to output without the need to identify the physics of a priori have been widely applied to hydrology field (ASCE task committee, 2000; Sahoo et al., 2006). Jain (2001) applied sediment rating curve and ANN method to predict relationship between the suspended sediment load and the river flow. Results of the mentioned study showed that the ANN method was capable to provide much better results than rating curve method.

Nagy et al. (2002) provided an ANN model with inputs of Froude number, water top width ratio and Reynolds number for the concentrations of suspended sediment load prediction. Comparison of the results indicated that the ANN model was more accurate in predicting sediment concentration in comparison with the other conventional models. Alp and Cigizoglu (2007) developed two different ANN methods to simulate relationship of suspended sediment load with precipitation and river flow by using hydro meteorological data. The obtained results showed that the provided models produced significantly better results than multi linear regression (MLR). Zhu et al. (2007) by using average of precipitations, precipitation intensity, temperature and flow discharge predicted suspended sediment load with ANN modeling the Long chuanjiang River. This model was capable to predict monthly suspended sediment load of flow accurately. And recently Melesse et al. (2011) compared Results from ANN model with results from multiple linear regressions (MLR), multiple non-linear regression (MNLR) and Autoregressive integrated moving average (ARIMA) for daily and weekly prediction of suspended sediment load in different period length of training and testing data. The results indicate ANN predictions for most simulations were superior compare to predictions using MLR, MNLR and ARIM. However, with the advent of Neuro-Fuzzy (NF) approach, better results are reported in literature for specific applications (Kisi, 2005; Cobaner et al., 2009; Rajaee et al., 2009; Kisi et al., 2009).

NF modeling is another method that refers to the approach of applying different learning algorithms developed in the neural network literature to fuzzy

To cite this paper: M. A. Ghorbani, S. H. Hosseini, M. H. Fazelifard, H. Abbasi. 2013. Sediment load estimation by MLR, ANN, NF and Sediment Rating Curve (SRC) in Rio Chama River .*J. Civil Eng. Urban*., 3 (4): 136-141.
Journal homepage: http://www.ojceu.ir/main/

modeling or a fuzzy inference system. Kisi et al. (2005) studied the accuracy of an adaptive NF and ANN to simulate the concentration of sediment load with river flow. The results showed that NF model produced better performance than ANN, SRC and MLR models. Lohani et al. (2007) derived stage–discharge–sediment concentration relationships by using fuzzy logic, ANN and SRC. Fuzzy logic had better performance in comparing with other mentioned models. However, NF approach was not tested in their study. Cobaner et al. (2009) estimated suspended sediment concentration by an adaptive NF and neural network approaches using hydro meteorological data. The result of NF model was found to be better. Rajaee et al. (2009) simulated daily suspended sediment concentration by using ANN, NF, MLR and SRC models. Results from NF model were found to close to observed data as compared to that for other models. Kisi et al. (2009) applied adaptive NF computing technique, ANN and SRC to estimate monthly suspended sediment of two rivers in Turkey. NF computing technique showed better performance in predicting suspended sediment load of rivers.

Thus in the view of the superior performance of ANN and NF in different application cases, both the modeling approaches are adopted in this study to compare their performances in modeling the sediment load of the river Rio Chama. The river Rio Chama has an important significance to provide major part of drinking and industrial water in New Mexico State. Apart from these two artificial intelligence based approached, two traditional methods, which are being used in practice for quite long time, namely MLR and SRC, are also applied in the same river and the comparative performance are tested. Recommendation and conclusions are made for suitable modeling approaches to address such problems.

MATERIALS AND METHODS

Artificial neural networks

McCullon and Pitts (1943) are generally recognized as the designers of the first neural network (NN). ANN is a method that is inspired by the studies of the brain and nerve systems in Biological organisms. NNs have the capability of self-learning and automatic abstracting. Applying this technique may reduce the time of modeling the complex systems. ANNs are important alternatives to the traditional methods of data analysis and modeling.

The basic processing elements of NNs are called *artificial neurons*, or *simply neurons* or *nodes*. In as implified mathematical model of the neuron, the effects of the synapses are represented by connection weights that modulate the effect of the associated input signals, and the nonlinear characteristic exhibited by neurons is represented by a transfer function. The neuron impulse is then computed as the weighted sum of the input signals, transformed by the transfer function. The learning capability of an artificial neuron is achieved by adjusting the weights in accordance to the chosen learning algorithm. Theoretical background of ANN approach is well documented in the literature (e.g.ASCE Task Committee, 2000).

A typical artificial neuron and the modeling of a multilayered NN are illustrated in Figure 1. Referring to this figure, the signal flow from inputs x_1, ..., x_n is considered to be unidirectional, which are indicated by arrows,as is a neuron's output signal flow (O). The neuron outputsignal O is given by the following relationship:

$$O = f(net) = f(\sum_{j=1}^{n} w_j x_j) \quad (1)$$

Where w_j is the weight vector, and the function $f(net)$ is referred to as an activation (transfer) function. The variable *net* is defined as a scalar product of the weight and input vectors,

$$net = w^T x = w_1 x_1 + ... + w_n x_n \quad (2)$$

Where T is the transpose of a matrix, and, in the simplest case, the output value O is computed as

$$O = f(net) = \{ f(net) \quad if \ \ w^T x \geq \theta; \quad 0 \ otherwise \} \quad (3)$$

Where θ is called the threshold level or bias; and this type of node is called a *linear threshold unit*.

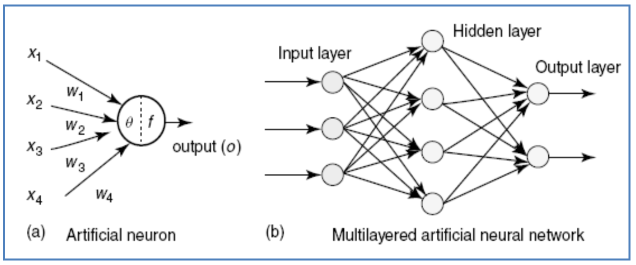

Figure 1. Architecture of an artificial neuron and a multilayered NN

The best-known examples of this technique occur in the backpropagation algorithm, the delta rule, and the perceptron rule. In unsupervised learning (or self-organization), a (output) unit is trained to respond to clusters of pattern within the input. In this paradigm, the system is supposed to discover statistically salient features of the input population. Unlike the supervised learning paradigm, there is no a priori set of categories into which the patterns are to be classified; rather, the system must develop its own representation of the input stimuli. Reinforcement learning is learning what to do – how to map situations to actions – so as to maximize a numerical reward signal. The learner is not told which actions to take, as in most forms of machine learning, but instead must discover which actions yield the most reward by trying them. In the most interesting and challenging cases, actions may affect not only the immediate reward, but also the next situation and, through that, all subsequent rewards. These two characteristics, trial-and error search and delayed reward are the two most important distinguishing features of reinforcement learning.

Neuro-fuzzy model

Neuro-fuzzy systems are fuzzy systems which use neural networks theory in order to determine their properties (fuzzy sets and fuzzy rules) by processing data samples. A fuzzy rule base consists of readable if-then statements that are almost natural language, but it cannot learn the rules itself. The two are combined in *nero-fuzzy system* in order to achieve readability and learning ability at the same time.

The neural network research started in the 1940s, and the fuzzy logic research in the 1960s, but the neuro-fuzzy research area is relatively new. The first book was probably by Kosko (1992). His ideas were implemented slightly earlier in the commercial tool TILGen (Hill et al., 1990), and in 1995 came the Fuzzy Logic Toolbox for MATLAB (Jang and Gulley, 1995), which includes a neuro-fuzzy method.

Sediment rating curve

A SRC is a relationship established between sediment concentration, *C*, and water discharge, *Q*, so that *C=f(Q)*, or between load, *L*, and discharge so that *L=f(Q)*. This relationship is in most cases defined as a power equation (Picouet et al., 2001): $L= aQ^b$ (4)
Where *a* and *b* are constants. As sediment concentration or load has a lognormal distribution it has been common practice to logtransform the data to obtain a normal distribution and to develop a linear regression equation on the logarithms using the least-squares method.

$$Log(L)= Log(a)+ b\times Log(Q) \quad (5)$$

Multiple linear regression model

Regression analysis is used when two or more variables are thought to be systematically connected by a linear relationship.

MLR applies to problems in which records have been kept of one variable, *y*,the dependent variable, and several other variables x_1, ..., x_k, the independent variables, and in which the objective requires the relationship between the variable *y* and the variables x_1, ..., x_k to be investigated. For any such record, the specific mathematical relationship (model) assumed is (Berk, 2004): $y= a+ b_1x_1+ b_2x_2+ ... +b_kx_k$ (6)
Where *a*, b_1, ..., b_k are constants and x_1, ..., x_k are the variables. Thus, it is assumed that *y* is linearly related to each of the independent variables and that each independent variable has an additive effect on *y*. Therefore, at this stage, we are assuming that x_1, ..., x_k do not interact amongst themselves in their effect on *y*.

Study area and data

The daily discharge and sediment data obtained from a station on Rio Chama River in the U.S. state of Colorado and New Mexico which is located Latitude 32º 20´24 north and Longitude 90º 09´ 04" east was used for calibration and verification for all the models provided in this study. Basin area in this station is *201(Km²)*. Figure 2 and 3 show daily discharge and sediment time series, respectively. Table 1 shows statistical properties of recorded time series of river flow (in ft3/s) and sediment load (mg/l) in river Rio Chama.

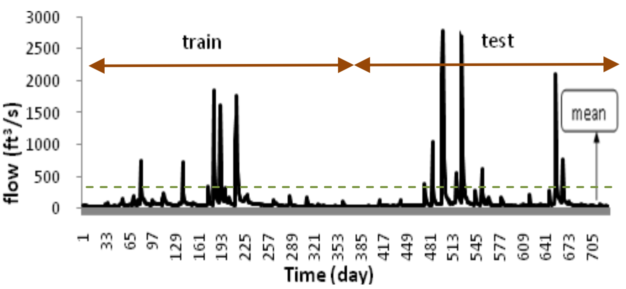

Figure 2. Daily discharge series time

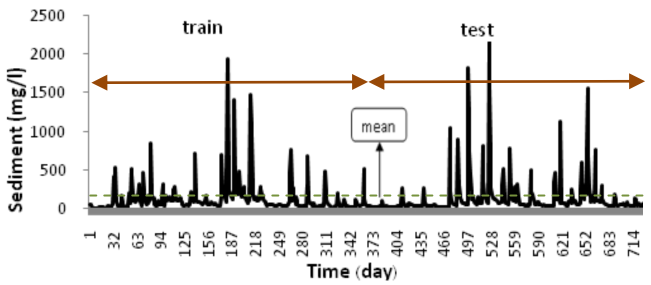

Figure 3. Daily sediment load series time

MODEL DEVELOPMENT

In this study, *MATLAB Software* was used to model suspended sediment load by ANN, NF, SRC, and *Datafit Software* was used to model regression analysis. Since river discharge and suspended sediment load time series in river Rio Chama had been recorded for two years, in order to evaluate the models in full year, one year (January 1994 - December 1994) data was used for training and the other year data (January 1995 - December 1995) was used for testing. Thus, 365 patterns were used for training and remaining 365 patterns are used for testing.

Three criteria, the root mean square error *(RMSE)*, the coefficient of determination (R^2) and Nash-Sutcliffe efficiency (*NSE*) have been used to assess the goodness of fit performance of the models:

$$CC=\sum_{i=1}^{N}\frac{\left[(x_{obs})_i-(\bar{x}_{obs})\right]\left[(x_{comp})_i-(\bar{x}_{comp})\right]}{\sqrt{\sum_{i=1}^{N}\left[(x_{obs})_i-(\bar{x}_{obs})\right]^2\sum_{i=1}^{N}\left[(x_{comp})_i-(\bar{x}_{comp})\right]^2}} \quad (7)$$

$$RMSE=\sqrt{\sum_{i=1}^{N}(x_{comp}-x_{obs})^2/N} \quad (8)$$

$$NSE=1-\sum_{i=1}^{N}(x_{obs}-x_{copm})\Big/\sum_{i=1}^{N}(x_{obs}-\bar{x}_{obs}) \quad (9)$$

Where *i* is an integer varying from 1 to *N*, x_{obs} and x_{comp} are the observed and computed suspended sediment load, respectively, the average value of the associated variable is represented with a bar above it and *N* is the total number of records. *RMSE* can provide a balanced evaluation of the goodness of fit of the model as it is more sensitive to the larger relative errors and the best coefficient will be zero (*RMSE=0*).R^2, which ranges from 0 to 1, is a statistical measure of how well the regression line close to the observed data and a coefficient of one ($R^2=1$) indicates that the regression line perfectly fits the observed data. *NSE*, that ranges between −∞ and 1.0, with *NSE=1* being the optimal value. Values between 0.0 and 1.0 are generally viewed as acceptable levels of performance, whereas values <0.0 indicates that the mean observed value is a better predictor than the simulated value, which indicates unacceptable performance.

One of the most important steps in developing a satisfactory forecasting model is the selection of the input variables. Different combinations of the antecedent sediment (S_{t-1}), antecedent river discharge (Q_{t-1}) and current river discharge (Q_t) are used to construct the appropriate input structure and current sediment (S_t) as output of models were selected. Table 2 shows 4

To cite this paper: M. A. Ghorbani, S. H. Hosseini, M. H. Fazelifard, H. Abbasi. 2013. Sediment load estimation by MLR, ANN, NF and Sediment Rating Curve (SRC) in Rio Chama River .*J. Civil Eng. Urban*., 3 (4): 136-141.
Journal homepage: http://www.ojceu.ir/main/

different combinations of discharge and sediment to have been considered as inputs of ANN, NF and MLR models, while SRC uses only previous sediment as model input.

Table 3 shows the best structure for each combination of inputs in *ANN* and *NF* models that have been obtained with try and error. Second column in this table indicates the number of triangular membership functions used for each input of NF models while Sugeno output membership functions, in all structures, are linear, e.g. in combination III, number of input membership functions of Q_t and S_{t-1} are respectively 2 and 3. Last column gives the number of nodes have been used for each hidden layer of ANN models with tansig transfer function while linear transfer function has been used in output layer.

The performance assessment of four models, ANN, NF, MLR and SRC for each combination of inputs are shown in Table 4. Table 4 shows that the best results of MLR model comes from combination I of discharge and sediment with $R^2=0.778$, $RMSE= 119.5$ *mg/l* and $NSE=0.752$ and for SRC model $R^2=0.754$, $RMSE= 320.2$*mg/l* and $NSE=0.788$. The best Results of ANN and NF models also have been derived using input combinations II and III, respectively. Table 4 shows the best result for each one of the models has been obtained in a special input combination.

RESULTS AND DISCUSSION

After modeling with each one of the models and calculating RMSE, R2 and NSE, the best combination of discharge and sediment was selected.R2, RMSE and NSE parameters of the best structure for ANN, NF, MLR and SRC models are shown in Fig 4. Comparison results of ANN model with other three models clears the superiority of ANN model that shows self-learning and automatic abstracting capability of this model while Kisi (2005) and Kisi & et al. (2008) studies indicate better performance of NF than ANN.As Lee and Han (2008) argued it maybe indicate difference in precise of the data and robustness of NF. Furthermore, like the previous studies (Kisi, 2005; Kisi et al., 2008), the results show NF model has better performance than MLR and SRC models. It is clear from the figure 4 that in comparison of SRC and MLR models, the assessment parameters are in conflict with each other and because of more importance of RMSE, it could be said, MLR performance is better than SRC. In fact SRC would be converted to a linear regression, after log transformation of data, there for in this case log transformation of data has been have an unfavorable effect on the results.

Observed and computed sediment load from four models are shown in Figure 5 and better fitness of ANN is seen. It is clear from Table 1 that both flow and sediment data series which were used in this study have high positive skewness *coefficient* and from 730 data just 24 percent of data are higher than mean value (129.136) that indicates the complexity of flow-sediment phenomenon. Moreover, minimum and maximum values of sediment data during training period are respectively 28 mg/l and 1860 mg/l while the in same during the testing period range is 18 mg/l and 2150 mg/l respectively. Thus, the maximum value for the training sediment data is lower than that during testing period, which may cause some extrapolation difficulties in prediction high sediment values. Maybe because of this, as it’s clear in the Fig 5, neither of the models exactly fit the extreme values of the picks although ANN, NF and MLR models show much better results than SRC. It could be seen peak-estimates of ANN and NF models sometimes are underestimate and sometimes are overestimate while SRC peak-estimates is always overestimate. Scatter plot of observed and predicted values from ANN, NF, MLR and SRC models also are showed in Figure 5 and overestimate of sediment peaks in SRC is clear. It is clear from this figure that the fit lines of the MLR, NF and ANN models are close to the ideal line. However, the ANN model has the highest R^2 value which implies that fit line of the ANN estimates closer to the observed data than those of the MLR and NF models although steep of fit line in NF and MLR models are closer to 1.

As estimating of annual sediment load is important in reservoir management studies, the goodness of the models are assessed to estimate annual sediment load. Figure 6 shows the estimated annual values from all models. There isn't remarkable difference between ANN and NF models although the results of NF are a little better than ANN. As a result, in annual estimation, NF, ANN and MLR have better performance, respectively than the SRC. As shown inFigure6, while the ANN underestimates the observed values, the NF and MLR models overestimate and they can be used in water resources planning to have an optimistic and pessimistic estimation from ANN and NF, respectively.

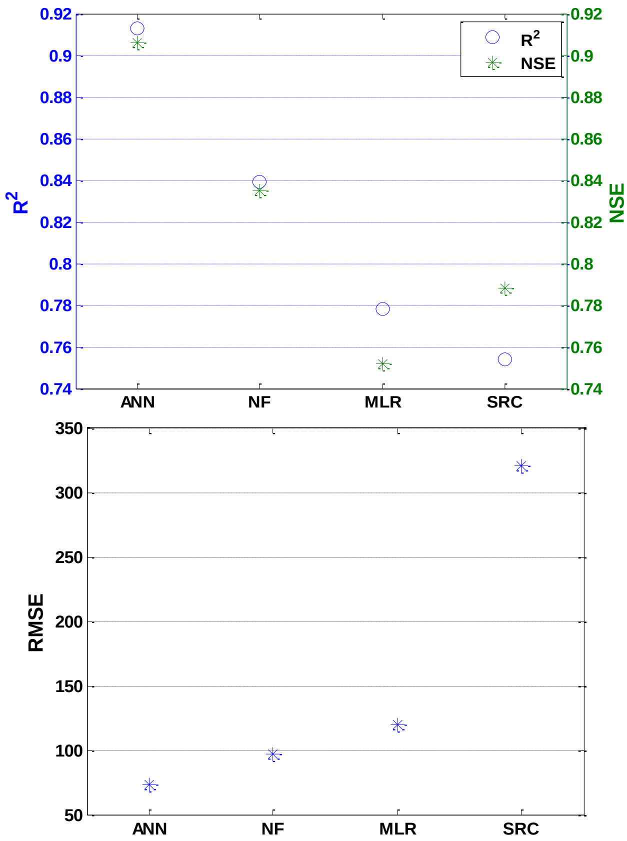

Figure 4. Comparison of R^2, NSE and RMSE values in four models

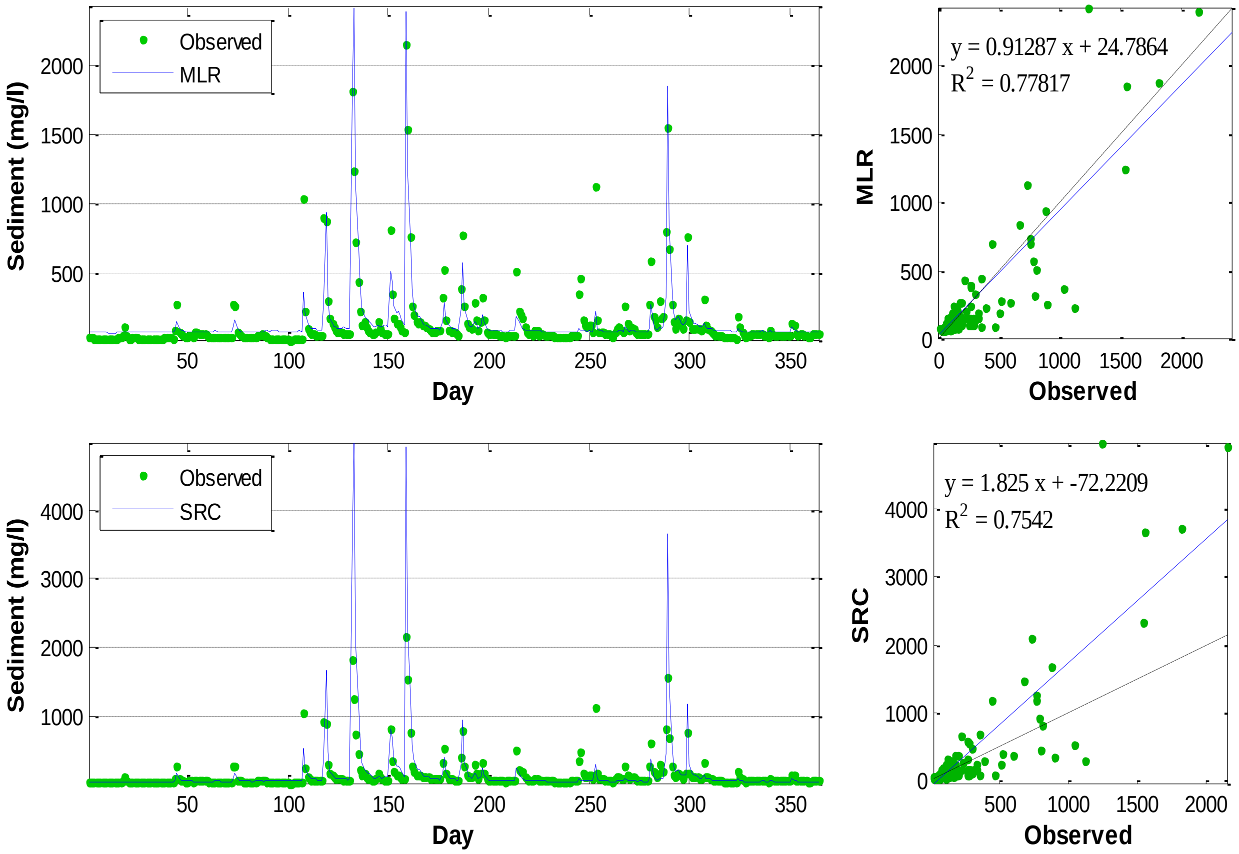

Figure 5 - Time series and Scatter plot of observed and predicted suspended sediment load from ANN, NF, MLR and SRC models

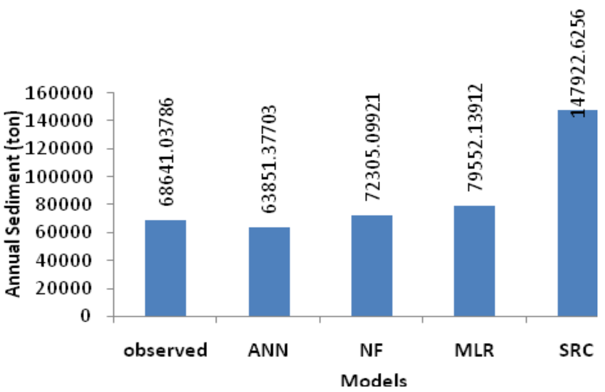

Figure 6: Estimated annual sediment load values from four models

CONCLUSION

This paper compares four different models artificial neural network, neuro-fuzzy, multiple linear regression and sediment rating curve in order to simulate daily sediment load and estimate annual sediment load of Rio Chama River. The results showed that ANN model has better performance than the NF, MLR and SRC models in the daily sediment load emulation, while in the estimation of annual sediment load NF model is the best, although ANN and MLR models with a low difference are also suitable. In addition in this condition ANN model for optimistic estimation and NF model for pessimistic estimation are better and sediment rating curve has worse performance. As a whole result, in this case ANN and NF models seem better than traditional methods of MLR and SRC, although Definitive conclusions need to more investigation in different regions.

REFERENCES

Alp M, Cigizoglu H K. (2007). Suspended sediment load simulation by two artificial neural network methods using hydro meteorological data. *Journal of Environmental Modeling & Software,* 22: 2–13.

ASCE Task Committee. (2000). Artificial neural networks in hydrology. I: Preliminary Concepts. Journal Hydrologic Engineering, 5(2): 115–123.

Asselman N E M. (2000). Fitting and interpretation of sediment rating curves. *Journal of Hydrology*, 234: 228–48.

Berk R A. (2004). Regression Analysis: A Constructive Critique. Sage Publications, Thousand Oaks.

Cobaner M, Unal B, Kisi O. (2009). Suspended sediment concentration estimation by an adaptive neuro-fuzzy and neural network approaches using hydro meteorological data. *Journal of Hydrology*, 367: 52–61.

Hill G, Horstkotte, E, Teichrow J. (1990). Fuzzy-C development system-users manual. Togai Infralogic, 30 Corporate Park, Irvine, CA 92714, USA.

Jain S K. (2001). Development of integrated sediment rating curves using ANNs. *Journal of Hydraulic Engineering*, 127 (1): 30–37.

Jang J S R, Gulley N. (1995). Fuzzy logic toolbox, the math works Inc., 24 Prime Park way, Natick, Mass, 01760-1500.

Jansen I M L, Painter R B. (1974). Predicting sediment yield from climate and topography. *Journal of Hydrology*, 21: 371-380.

Kisi O, Haktanir T, Ardiclioglu M, Ozturk, O, Yalcin, E, Uludag S. (2009). Adaptive neuro-fuzzy computing technique for suspended sediment estimation. *Advances in Engineering Software journal*, 40: 438–444.

Kisi O, Yuksel I, Dogan E. (2008). Modelling daily suspended sediment of riversin Turkey using several data-driven techniques. Hydrological Sciences Journal, 53(6): 1270-1285.

Kisi O. (2005). Suspended sediment estimation using neuro-fuzzy and neural network approaches. *Hydrological Sciences Journal*,50(4): 683-696.

Kosko B. (1992). Neural network and fuzzy systems. A dynamic systems approach to machine intelligence. Prentice–Hall, Englewood Cliffs.

Lee H X, Han D. (2008). Exploration of neuro-fuzzy models in real time flood forecasting. Proceedings of Artificial Intelligence and Pattern Recognition, 7–10 July 2008: 264–268

Lohani A K, Goel N K, Bhatia K K S. (2007). Deriving stage–discharge–sediment concentration relationships using fuzzy logic. Hydrolog. Sci. J., 52 (4): 793–807.

McCulloch W S, Pitts W. (1943). A logical calculus of the ideas immanent in nervous activity, B. Math. Biol. 5 (4): 115-133.

Melesse A M, Ahmadb S, McClain M E, Wangc X, Lim Y H. (2011). Suspended sediment load prediction of river systems: An artificial neural network approach. Agr. Water Manage, 98 (5): 855–866.

Morris G L, Fan J. (1997). Reservoir Sedimentation Handbook, McGraw-Hill, New York.

Nagy H M, Watanabe K, Hirano M. (2002). Prediction of load concentration in rivers using artificial neural network model. *Journal of Hydraulic Engineering*, 128(6): 588–95.

Picouet C, Hingray B, Olivry J C. (2001). Empirical and conceptual modelling of the suspended sediment dynamics in a large tropical African river: the Upper Niger river basin. *Journal of Hydrology*, 250 (1-4): 19-39.

Rajaee T, Mirbagheri S A, Zounemat-Kermani M, Nourani V. (2009). Daily suspended sediment concentration simulation using ANN and neuro-fuzzy models. *Science of the Total Environment journal*, 407: 4916–4927.

Sahoo G B, Ray C, Mehnert E, Keefer D A. (2006). Application of artificial neural networks to assess pesticide contamination in shallow groundwater. *Journal of Science Total Environment*, 367: 234–251.

Zhu Y M, Lu X X, Zhou Y. (2007). Suspended sediment flux modeling with artificial neural network: an example of the Longchuanjiang River in the Upper Yangtze Catchment, China. *Journal of Geomorphology*, 84: 111–125.