Sediment Prediction Using ANN and Regression Approach

JUNAIDAH ARIFFIN**,** *Associate Professor, Faculty of Civil Engineering***,** *Universiti Teknologi MARA, 40450, Shah Alam, Selangor, Malaysia*

AMINUDDIN ABDUL GHANI, *Associate Professor, River Engineering and Urban Drainage Research Centre (REDAC), Universiti Sains Malaysia, Engineering Campus, Seri Ampangan, 14300 Nibong Tebal, Pulau Pinang, Malaysia*

NOR AZAZI ZAKARIA, *Associate Professor, REDAC, Universiti Sains Malaysia, Engineering Campus, Seri Ampangan, 14300 Nibong Tebal, Pulau Pinang, Malaysia*

AHMAD SHUKRI YAHYA, *Associate Professor, School of Civil Engineering, Universiti Sains Malaysia, Engineering Campus, Seri Ampangan, 14300 Nibong Tebal, Pulau Pinang, Malaysia*

ABSTRACT
In this paper, an attempt is made to predict sediment using two statistical techniques namely multiple linear regression (MLR) and artificial neural network (ANN) that uses feed-forward back-propagation algorithm. The main aim is to identify the best technique that could adequately predict sediment to a desired accuracy. The accuracy of the derived models is measured using the discrepancy ratio that is supported graphically. Discrepancy ratio is the ratio of the calculated values to the measured values. The performances of the derived models confirmed that the model derived using MLR technique gave a better prediction then the model derived using ANN for the specified flow range and sediment characterictics.

Keywords: Artificial neural network, multiple linear regressions, sediment prediction.

1 Introuction

Regression technique has received considerable attention as to its broad applications in solving engineering problems and other branches of science. The ultimate product of the analysis is the determination of a regression equation. The works by some investigators using the regression techniques for sediment prediction coefficients are those by Laursen (1958), Colby (1964), Shen and Hung (1972), Ackers and White (1973), Brownlie (1982), Chang *et.al* (1988), Karim and Kennedy (1990) and Yang (1996).

ANN is a mathematical model with highly connected structures similar to brain cells. It comprises a number of neurons arranged in different layers and it is a tool to facilitate hydraulic engineers in design and management practices. This mathematical model has a wide field of applications such as in the prediction of diaphragm wall deflection in deep excavation (Jan *et.al*, 2002) and sediment prediction in rivers (Nagy *et.al*, 2002). Zou *et.al* (2002) proposed a neural network embedded Monte Carlo (NNMC) approach to account for uncertainty in water quality modeling. The framework of their proposed method has three major parts namely a numerical water quality model, a neural network technique and Monte Carlo simulation. Several ANN models and learning procedures have been developed but this analysis is limited to the feed-forward back-propagation algorithm claimed to be the best approach for sediment prediction (Nagy *et.al*, 2002).

This paper highlights the regression techniques namely multiple regression and artificial neural network to predict sediment discharge using the selected flow and sediment discharge parameters. The best approach for sediment prediction is a model that predicts sediment discharge to a desired accuracy based on the graphical and physical analysis. Table 1 shows the sediment discharge parameters by some investigators, where C_V is the sediment concentration which is the dependent variable with different independent variables. The meanings of the symbols can be found in APPENDIX II.

Table 1 Sediment discharge approaches and relevant variables (Ariffin, 2004).

Researcher	Selected variables	
	Dependent variable	Independent variables
Engelund and Hansen	C_V	$\frac{\tau}{\gamma(S_s-1)d_{50}}, \frac{2gSy}{v^2}, \gamma_s(S_s-1)gd_{50}{}^3$
Yang	C_V	$\frac{VS}{\omega}, \frac{U^*}{\omega}, \frac{\omega d_{50}}{v}$
Nagy *et.al*	C_V	$\psi, \frac{\omega}{U_*}, \frac{h}{d_{50}}, F, R_*, \frac{h}{B}$

2 Data Used in Model Development

Data used in learning and verification process are taken from three rivers in Selangor namely Sungai Lui, Sungai Semenyih and Sungai Langat and 2002 data of Ibrahim and 2003 data of Drainage and Irrigation Department (Ariffin, 2004). The details and map of the study area can be found in Ariffin (2004). Range of river and sediment data used in the analysis and verification for the developed ANN model is as shown in Table 2.

Table 2 Range of data in learning and verification for the developed ANN model.

Variables	**Range**
Relative roughness on the bed, $\frac{R}{d_{50}}$	73 – 1885
Ratio of shear velocity and fall velocity, $\frac{U^*}{W_s}$	0.218 – 3.725
Ratio of shear velocity and average velocity, $\frac{U^*}{V}$	0.1 – 0.463
Froude Number, $\frac{V^2}{gy}$	0.121 – 0.414
Sediment concentration by volume, C_V	0.0000384 – 0.001125

3 Model Development Techniques

3.1 Multiple linear regression

In the multiple linear regression technique, least-squares method has been used to determine the best estimate of the multiple regression equation. This method chooses the best-fitting model which minimizes the sum of squares of the distances between the observed values and those predicted by the fitted model. The best fit is when the sum of squares of deviations between the observed and the predicted values is the smallest.

3.2 Artificial neural network

A typical multi-layer ANN with one hidden layer was used in this study. A set of data was first fed directly in the network through the input neurons and subsequently the multi-layer perceptron produces the predicted results in the output layer. The output layer is determined by the architecture of the network. The hidden layer serves as links between the input and the output layers. In this study a three layer feed-forward back propagation algorithm as a functional mapping tool is used to model the relationship between the parameters.

The three distinctive characteristics of the model are

a) The model of each neuron in the network uses a non-linear activation function for modification of each neuron in the network. The nonlinear activation function is usually sigmoid function. The activation of each neuron in a hidden layer or output layer is calculated as

$$a_i = \sigma_{LTF}\left(\sum_j w_{ij} o_j\right) \tag{1}$$

Where a_i is the activation of neuron i, j is the set of neurons in the preceeding layer ; w_{ij} is the weight of the connection between neuron i and neuron j ; o_j is the output of neuron j ; and σ_{LTF} is the logistic transfer function.

$$\sigma_{LTF} = \frac{1}{1+e^{-x}} \tag{2}$$

b) The network consists of one or more hidden layers which enable the network to learn complex problems.
c) The network exhibits a high degree of connectivity determined by the synapses of the network.

The controlling parameters for sediment discharge obtained from the multiple linear regression analyses represent the input variables that were fed into the neural network model. Supervised learning technique was adopted for all the derived models. In the analyses, multi-layer perceptron was used to model the data for sediment prediction with error back-propagation as the algorithm. The first 162 sets of data were used for training the network. The alpha and the learning rate parameters were determined by calibration through several computer run tests.

4 Results

Table 2 summarizes the performance of the derived models tested against a total of 346 sets of field data. The accuracy of the models has been evaluated using the discrepancy ratios 0.5-2.0, 0.75-1.25, 0.5-1.5 and 0.25-1.75. The results show that the model yields an accuracy of 67.35%. The graph of Figure 1 indicates a strong positive association between the dependent and independent variables. The scatter of points along 45^0 straight-line indicates the suitability of the model. The regression equation derived is as follows,

$$C_V = 1.156\times10^{-5}\left(\frac{R}{d_{50}}\right)^{0.716}\left(\frac{U^*}{W_s}\right)^{-0.975}\left(\frac{U^*}{V}\right)^{0.507}\left(\frac{V^2}{gy}\right)^{0.524} \tag{3}$$

Table 2 Performance of the derived models.

Model	Discrepancy ratio	Overall performance on 346 data		
	0.5-2.0	0.75-1.25	0.5-1.5	0.25-1.75
Regression	67.35%	31.2%	55.1%	65.6%
ANN	64.8%	26.2%	46.1%	60.6%

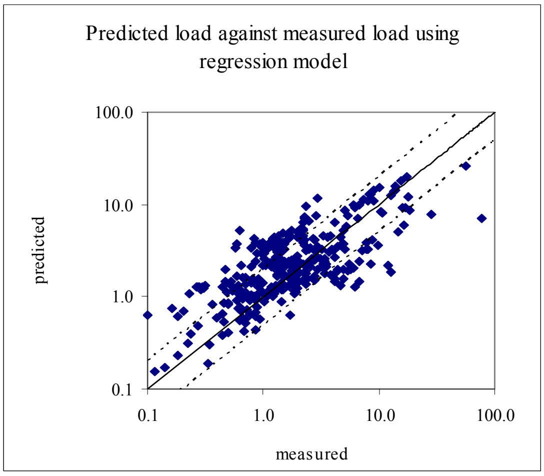

Figure 1 Predicted load against measured load using regression model.

The range of data in learning and verification for the developed ANN model is shown in Table 3. Figure 2 shows the derived ANN network architecture obtained after several computer runs. It is a 4-8-1 network consisting of 4 source nodes, 8 hidden neurons and 1 output neuron. The synaptic connections are fully connected. The input pattern constitutes the inputs to the nodes in layer (i) representing the independent variables $\left(\frac{R}{d_{50}}, \frac{U^*}{W_s}, \frac{V^2}{gy}, \frac{U^*}{V}\right)$. The output is the predicted loads or concentrations, C_V. The alpha (α) and the learning rate parameters η are 0.9 and 0.075 respectively. An additional 181 sets of data was added without target outputs C_V and the estimated values for concentration were estimated. The best fit of the obtained and given data for concentration in ANN model and the verified model yield prediction accuracies of 64.8% and 43.1%. The performances of the derived and verified ANN models are as shown in Figures 3 and 4 respectively. From the analyses the derived ANN model predict fairly for the first set of data but failed to adequately validate the model when tested against the second set of data. The expression for total load concentration for ANN model is given implicitly as follows:

$$C_V = f\left(\frac{R}{d_{50}}, \frac{U^*}{W_s}, \frac{V^2}{gy}, \frac{U^*}{V}\right) \tag{4}$$

Table 3 Range of data in learning and verification for the developed ANN model.

Variables	**Range**
Relative roughness on the bed, $\frac{R}{d_{50}}$	73 – 1885
Ratio of shear velocity and fall velocity, $\frac{U^*}{W_s}$	0.218 – 3.725
Ratio of shear velocity and average velocity, $\frac{U^*}{V}$	0.1 – 0.463
Froude Number, $\frac{V^2}{gy}$	0.121 – 0.414
Sediment concentration by volume, C_V	0.0000384 – 0.001125

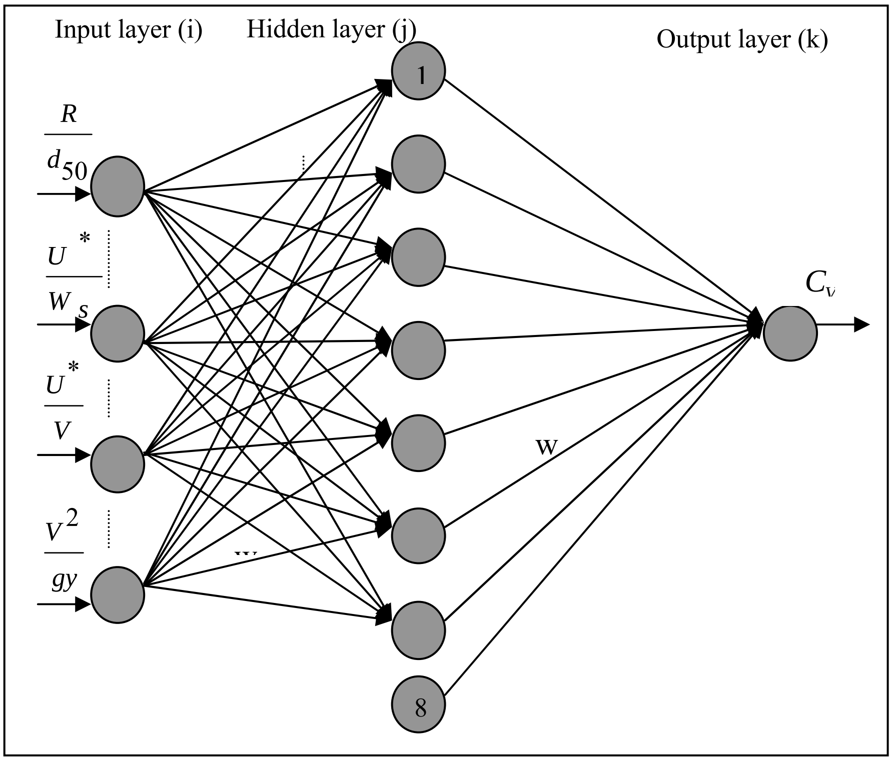

Figure 2 Derived ANN network

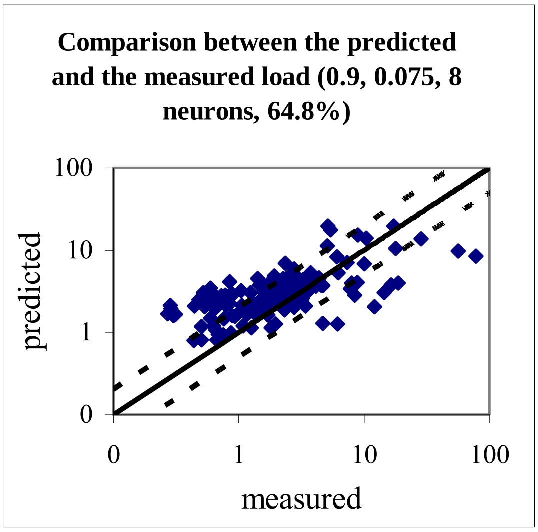

Figure 3 Predicted against measured load (ANN model with alpha = 0.9, η = 0.075 and 8 neurons in the hidden layer).

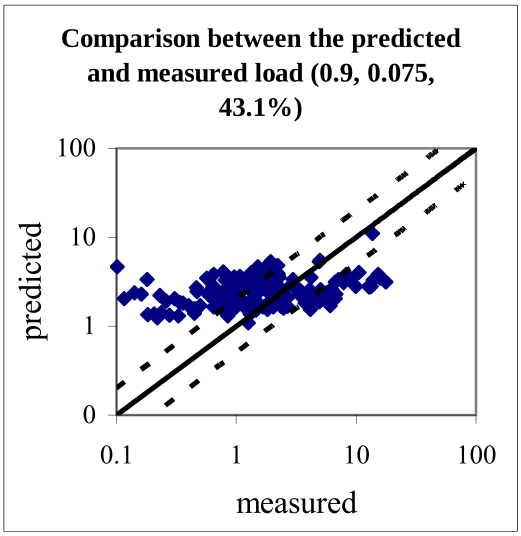

Figure 4 Predicted against measured load (Verified ANN model with alpha = 0.9, η = 0.075 and 8 neurons in the hidden layer) on 181 sets of data).

5 Conclusion

The results of the analyses indicate that the model derived using MLR technique gave a better prediction accuracy compared to model derived using ANN. The validity of this model has been confirmed graphically and physically. ANN model failed to predict sediment fairly for the 181 sets of data and the graph indicates poor association between the predicted and the measured concentrations.

Acknowledgement

The authors wished to express their sincere gratitude and appreciation to Professor Zekai Sen from the Istanbul Technical University for his critical and constructive comments on their paper.

This following symbols are used in this paper

Symbol	Description
B	River width
C_V	Volumetric concentration of sediment(ppm)
d_{50}	Sediment diameter where 50% of bed material is finer
F	Froude Number
g	Acceleration due to gravity (9.812 m/s^2)
h, y	Flow depth
R	Hydraulic radius
R_*	Shear velocity Reynolds Number
S	Energy slope
S_s	Specific gravity of sediment
τ	Mean bed shear stress
U_*, U^*	Shear stress $=\sqrt{gRS}$ or $\sqrt{gDS}$ or shear velocity
V	Average flow velocity (m/s)
ω	Fall velocity of sediment particles (m/s)
γ, γ_s	Specific weight of water and sediment (N/m^3)
ν	Kinematic viscosity
$\dfrac{R}{d_{50}}$	Relative roughness on the bed
$\dfrac{U^*}{W_s}$	Ratio of shear velocity and fall velocity
$\dfrac{U^*}{V}$	Ratio of shear velocity and average velocity
$\dfrac{V^2}{gy}$	Froude Number

References

1. Ackers, P., and White, W.R. (1973). "Sediment Transport: New Approach and Analysis." Journal of the Hydraulics Division, ASCE, 2041-2060.
2. Ariffin, J. (2004). "Development of Sediment Transport Models for Selected Rivers in Malaysia using Regression Analysis and Artificial Neural Network." Ph.D Thesis, University of Science Malaysia, Penang, Malaysia.
3. Brownlie, W. (1982). "Prediction of Flow Depth and Sediment Discharge in Open Channels." Reports of the California Institute of Technology, Pasadena, CA 91125, Report No. NSF/CEE-82090., 73-154.
4. Chang, F.M., Simons, D.B., and Richardson, E.V. (1965). "Total Bed-Material Discharge in Alluvial Channels." U.S. Geological Survey Water-Supply Paper 1498-I.
5. Colby, B.R. (1964). "Practical Computations of Bed Material Discharge." Journal of the Hydraulics Division, ASCE, Vol. 90, No. HY2.
6. Jan, J.C., Hung, S.L., Chi, S.Y., and Chern, J.C. (2002). "Neural Network Forecast Model in Deep Excavation." Journal of Computing in Civil Engineering., ASCE, 16(1), 59-65.
7. Karim, M.F., and Kennedy,J.F. (1990). "Menu of Coupled Velocity and Sediment- Discharge Relationship For

River." Journal of Hydraulic Engineering., ASCE, 116(8), 987-996.
8. Laursen, E.M. (1958). "The Total Sediment Load of Streams." Journal of the Hydraulics Division, ASCE., Proc. Paper 1530, 1–36.
9. Nagy, H.M., Watanabe, K., and Hirano, M. (2002). "Prediction of Sediment Load Concentration In Rivers Using Artificial Neural Network Model." Journal of Hydraulic Engineering, ASCE, 128(6), 588-595.
10. Shen,H.W., and Hung, C.S. (1972). "An Engineering Approach To Total Bed Material Load By Regression Analysis." Proceeding Sedimentation Symposium, Chapter 14, 14.1-14.7.
11. Zou, R., Lung, W.S., and Guo, H. (2002). "Neural Network Embedded Monte Carlo Approach For Water Quality Modeling Under Input Information Uncertainty." Journal of Computing in Civil Engineering, ASCE, 16(2), 135-142.