VOL. 11, NO. 4, FEBRUARY 2016 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

MODELLING DAILY SUSPENDED SEDIMENTS OF A HYPER-CONCENTRATED RIVER IN MALAYSIA

M. R. Mustafa
Department of Civil Engineering, Universiti Teknologi Petronas, Tronoh, Perak, Malaysia
E-mail: raza.mustafa@petronas.com.my

ABSTRACT

Estimation of suspended sediments in hyper-concentrated rivers is prime important as it is highly desired in design and operation of hydraulic structures. In this study the application of Multiquadric basis function neural network for prediction of suspended sediment of a hyper-concentrated river was investigated. Five years daily time series data of discharge and suspended sediments from 1992 - 1996 at Bidor River in Perak, Malaysia was used to develop the prediction model. Several trials were made to investigate the appropriate number of hidden neurons. Performance of the model was evaluated by comparing the observed and predicted sediments with perfect line of agreement. Furthermore, root mean square error and coefficient of efficiency were also used as performance statistical measures for the model. The results showed that the model successfully predicted the suspended sediments with minimum error of (RMSE = 9.06, MAE = 6.0) and highest efficiency of (CE = 0.94). The performance of the model with previous models was also comparable. The results suggested the suitability of Multiquadric basis function neural network for modelling suspended sediments of hyper-concentrated river.

Keywords: basis function, modeling, neural network, river, sediment.

INTRODUCTION

Suspended sediments are the early phase of bed load material which moves with river water flow in suspension for a considerable time without any contact with the river bed. The amount of suspended sediments in the river flows is always undesirable since it directly affect the ecosystem, biodiversity, fluvial geomorphology as well as different engineering aspects particularly design and operation of hydraulic structures. Therefore, it is vital to predict accurate amount of suspended sediments in the river for appropriate index to assess prospective conditions of water resources management. So far, several studies have been suggested to predict the suspended sediments in rivers. However, due the very complex mechanism of sediment passage in the river and highly nonlinear behavior of the parameters affecting the amount of suspended sediments, these techniques are unable to produce accurate amount of the sediments (Lafdani, *et al.* 2013).

Recently, artificial neural network (ANN) techniques have emerged as a one of robust tool to predict complex and nonlinear behavior between the hydrological parameters. A number of applications of ANN have been observed for prediction of suspended sediments in rivers around the world (Kisi, *et al.* 2008, Mustafa, *et al.* 2011, Mustafa, *et al.* 2011, Kisi, *et al.* 2012, Kisi and Shiri 2012, Mustafa, *et al.* 2012, Lafdani, *et al.* 2013, Liu, *et al.* 2013). Kisi and Shiri 2012 estimated suspended sediments in Eel River, USA by using rainfall, stream flow and suspended sediments data. They predicted suspended sediments by ANN, adaptive neuro-fuzzy interference system and gene expression programing and found comparable results. Mustafa, *et al.* (2012) predicted suspended sediments in Malaysia by several multilayer perceptron neural network. They investigated the performance of several training algorithms and found that Levenberg Marquardt and Scaled Conjugate Gradient performed better than other algorithms. Liu, *et al.* (2013) estimated suspended sediments by sediment rating curve analysis, feed forward back propagation neural network and wavelet neural networks. The investigation of the performance of all the models suggested that wavelet neural network model performed better than rest of the models. Furthermore, daily suspended sediments have been predicted by machine learning approach (Dheeraj *et al.* 2016), combining deterministic modelling with neural networks (Oleg *et al.* 2015), and integrative neural network (Atieh *et al.* 2015).

All aforementioned studies suggested the successful applications of different training algorithms and types of neural networks for prediction of suspended sediments in rivers around the world. However, it was observed one of the robust basis functions Multiquadric basis function have not been investigated yet for prediction of river suspended sediments. Multiquadric basis function has ability to capture nonlinear behavior between the variables and could be one of the most accurate predictors of suspended sediment data. Therefore, the objective of this study is to predict the suspended sediments in river using Multiquadric basis function neural network.

Study Area and Data Collection

Time series data of stream flow and suspended sediment were obtained from the Department of Irrigation and Drainage (DID) for the river “Sungai Bidor”. Sungai Bidor is located in the state Perak in Peninsular Malaysia (Figure-1). Bidor is located in the south of Tapah and the north of Sungkai along the North-South Expressway that leads to Padang Besar and Johor Bahru. Current survey on increasing population of Bidor

ARPN Journal of Engineering and Applied Sciences

suggested that the population may increase from 32,094 people to 44,564 people in 2020. The area is mainly occupied by plantation and substantial commercial activities related to agricultural business. In such conditions, increasing rate of the sediment generation from area and suspended sediment rate in the river is substantially increasing and consequently depleting the river capacity by sediment deposition (Competency 2006).

Figure-1. Location map of study area.

Statistical Analysis

Time series of daily stream discharge (Q) and suspended sediment (SS) data consists of five years from 1992 to 1996. However, due to few missing values, the actual number of data is 1766. Since, ANN networks consist of two stages as training and testing. Therefore, data analysis was performed after division of the data into training and testing datasets. About 80% of the data were used at the training stage and 20% of the data were used for testing the model. Table-1 shows a thorough comparison of different statistical measure to analysis the data variability and complexity. The mean, min and max of discharge (21.7 m^3/d, 3.5 m^3/d and 101 m^3/d) and suspended sediment (98.2 tons/d, 6 tons/d and 453.9 tons/d) during training stage are higher than the testing stage data (Q = 15 m^3/d, 6.3 m^3/d, 53.3 m^3/d and SS = 68.1 tons/d, 19 tons/d, 239.5 tons/d) respectively. Since, best performance of any ANN model can only be achieved if the min and max of the data are well trained during the training stage. Therefore, the division of the data were made in such a way that min and max of the data could be accommodated in training dataset.

Table-1. Statistical analysis of training and testing datasets.

Statistical Measures	Training		Testing	
	Q (m^3/d)	SS (tons/d)	Q (m^3/d)	SS (tons/d)
No. of Data	1409	1409	357	357
Mean	21.7	98.2	15.0	68.1
Max	101.0	453.9	53.3	239.5
Min	3.5	6.0	6.3	19.0
SD	13.5	60.7	7.5	35.0
Variance	183.3	3688.7	56.1	1226.8
Skew	1.8	1.8	1.8	1.4

Similarly, standard deviation (SD), variance and skewness coefficient shown in the Table-1 indicates that the training dataset contains more complexity and variability compared to testing dataset.

Neural Network Modelling

Generally, architecture of neural network consists of three numbers of layers known as input, hidden and output layer. Each layer is equipped with processing elements called as nodes or neurons. The neurons of each layer are connected with other layers through connection weights. Selection of appropriate number of neurons in each layer is vital to achieve optimal prediction model. Several studies on prediction of suspended sediments using neural network techniques proposed three number of inputs are appropriate in input layer (Mustafa, *et al.* 2011, Mustafa, *et al.* 2012, Mustafa, *et al.* 2012). Mustafa, *et al.* 2012 investigated appropriate number of input neurons by performing autocorrelation and cross-correlation analysis and suggested that two antecedent discharge with current discharge values are sufficient to achieve best ANN model. Therefore, in this study three numbers of input neurons has been used to establish the prediction model.

Optimal number of hidden neurons in the hidden layer or radial basis function (RBF) layer was investigated by trial and error procedure. Multiquadric radial basis function was used to capture the nonlinear behaviour between discharge and suspended sediments. A couple of trials were made and few of them are shown in Table-2. Selection of appropriate number of hidden neurons was made based on the minimum error produced during training and testing stages. Although, the large number of hidden neurons (i.e. 50, 70, 100 etc.) produced minimum error during training and testing stages but over-fitting problem was observed. However, it was observed that 22 number of hidden neurons produced minimum error during training and testing stages and do not experience with over-fitting problem. Therefore appropriate number of hidden neurons was selected as 22. Since, only one output (i.e. suspended sediment) is expected from the output layer, therefore the output layer

VOL. 11, NO. 4, FEBRUARY 2016 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

consists on one neuron only. The appropriate architecture then identified as 3 – 22 – 1 (input – hidden – output neurons).

Table-2. Training and testing error vs. number of neurons.

Trial	Hidden	Training	Testing
1	4	0.01602	0.01200
2	5	0.01480	0.00889
3	6	0.01456	0.00710
4	7	0.01382	0.01044
5	8	0.01191	0.00919
6	9	0.00560	0.00566
7	10	0.00665	0.00720
8	11	0.00286	0.00351
9	12	0.00161	0.00193
10	13	0.00111	0.00135
11	14	0.00399	0.00348
12	15	0.00292	0.00355
13	16	0.00260	0.00304
14	17	0.00173	0.00234
15	18	0.00128	0.00209
16	19	0.00077	0.00127
17	20	0.00067	0.00071
18	21	0.00159	0.00246
19	22	0.00037	0.00041
20	23	0.00076	0.00122
21	24	0.00081	0.00101
22	25	0.00107	0.00119
23	50	0.00018	0.00035
24	70	0.00002	0.00010
25	100	0.00001	0.00006

Evaluation Criteria

Several statistical measures were used to evaluate the model performance in terms of least error, model efficiency, closeness of predicted and observed values. Root mean square error (RMSE) and mean absolute error (MAE) were adopted to testify the model in term of error produced by the model during training and testing stages. Zero value of RMSE and MAE indicate the ideal performance or perfection of the model which do not have any error between predicted and observed data. And the value close to zero is good. Whereas Nash–Sutcliffe coefficient of efficiency (CE) and coefficient of determination (R2) were used to show (i) competence of the model and (ii) agreement between the predicted and observed values respectively. The value of CE and R2 close to one shows nearly perfect predictions.

RESULTS AND DISCUSSION

Time series of observed and predicted suspended sediment during testing stage of RBF modelling is shown in Figure-2. The predicted suspended sediment followed the nonlinear behaviour of the observed data and adopted the complex pattern of the data. The predicted values are very close to the observed ones with exception of few values at the peaks. Although the model followed the pattern of observed data but some of the lowest values in the data (i.e. 19 July – 20 September) were not predicted so closely. Similar issues were also observed in some of previous studies on suspended sediments while using RBF modelling (Alp and Cigizoglu 2007, Mustafa, *et al.* 2011). Use of clustering techniques pooled with RBF modelling could be helpful to improve predictions at the low and high suspended sediment values.

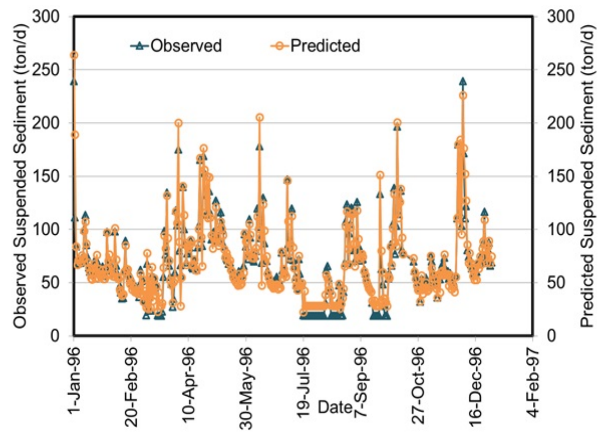

Figure-2. Time series of observed and predicted suspended sediment during testing stage.

A comparison of observed and predicted suspended sediments with perfect line of agreement during training stage of RBF modelling is shown in Figure-3. The trained values showed nearly a perfect agreement with the observed values by producing coefficient of determination close to one (R^2 = 0.9801). Nearly perfect agreement between the observed and trained data shows that the model was successfully trained during by minimizing the differences between the targets and outputs. Almost the complete set of the training data were found very close or overlying to the perfect line of agreement. None of the outlier was observed during training stage.

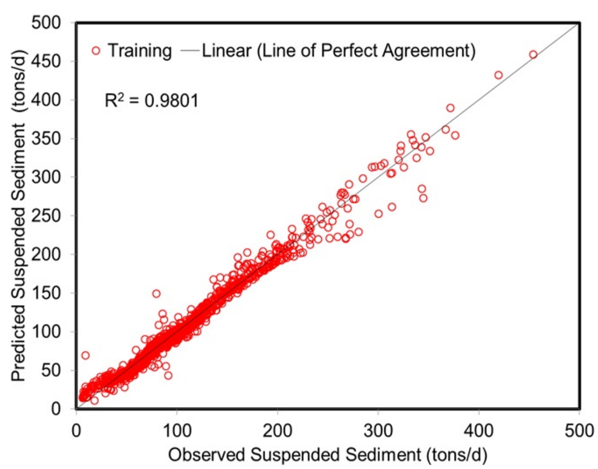

Figure-3. Comparison of observed and predicted sediment with perfect line of agreement during training stage.

Figure-4 shows a comparison between the predicted and observed suspended sediment data with the perfect line of agreement during the model's testing stage. The predicted values showed a good agreement with the observed sediment data and produced coefficient of determination close to one (R2 = 0.9367). The suspended sediment values which were not accurately predicted at the high and low suspended sediment data (Figur-2) can be clearly seen in Figure-4 in the form of outliers. Although these outliers highlight the limitation of the model but still the majority of the data lies on the perfect line of agreement showing the acceptability of the model.

A summary of performance statistics of the model during training and testing stages is shown in Table-3. The performance statistics showed that the model produced RMSE = 0.56; MAE = 5.48 and CE = 0.98 during training and RMSE = 9.06; MAE = 6.00 and CE = 0.94 during testing stage. The error in both stages is far less than the mean value of sediment during training 98.2 tons/d and testing 68.1 tons/d (Table-1). The less error and high efficiency during training stage as compared to the testing stage may be due to the length of data span used in the training (1409) and testing (357). Additionally, input and target data are presented in the training stage to minimize the training error; conversely, only input data is presented during the testing stage.

Table-3. Performance statistics of the model.

Performance Measures	Training	Testing
RMSE	8.56	9.06
MAE	5.48	6
CE	0.98	0.94

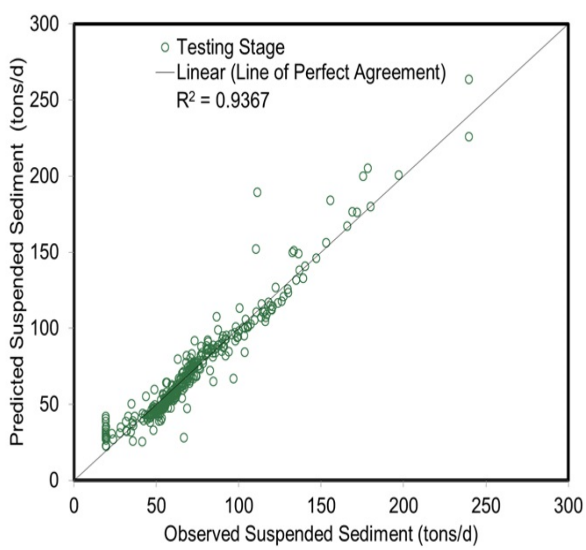

Figure-4. Comparison of observed and predicted sediment with perfect line of agreement during testing stage.

It is obvious that the results showed that the neural network using Multiquadric basis function is suitable for estimation of suspended sediment in rivers. The results produced by Multiquadric RBF model are comparable with previous studies (Alp and Cigizoglu 2007, Mustafa, Isa *et al.* 2011, Mustafa, Rezaur *et al.* 2012) which suggest appropriateness of the model. Furthermore, it appeared from the results that Multiquadric basis function can be used as an alternative to the other basis functions (i.e. Gaussian) for prediction of suspended sediments in hyper – concentrated rivers.

CONCLUSIONS

In this study, a relationship between water discharge and suspended sediment in Bidor River, Perak, Malaysia was investigated using Multiquadric radial basis function neural network. The model successfully predicted the suspended sediments by capturing the exact pattern of nonlinear behaviour of the data. It was seen that increasing the number of hidden neurons although decrease the network error but a high number of hidden neurons may leads the network to over-fitting problem. The network configuration identified with 3 inputs, 22 hidden and 1 output neurons (3-22-1) was found suitable to map the relationship between the inputs and targets. The results recommend the application of Multiquadric basis function neural network for prediction of river suspended sediments and may be explored for other hydrological variables where it has not been used yet.

ACKNOWLEDGEMENTS

The authors is thankful to the Universiti Teknologi Petronas for providing facilities and funding to conduct this study.

REFERENCES

[1] Alp, M. and H. K. Cigizoglu (2007). Suspended sediment load simulation by two artificial neural network methods using hydrometeorological data. Environmental Modelling & Software 22(1): 2-13.

[2] Atieh M. , S.L. Mehltretter, B. Gharabaghi, and R. Rudra (2015). Integrative neural networks model for prediction of sediment rating curve parameters for ungauged basins. Journal of Hydrology 531, 1095 - 1107.

[3] Competency, D. o. S. S. (2006). A report on Bidor region by students of klang valley vampus.

[4] Dheeraj Kumara, Ashish Pandeya, Nayan Sharmaa, and Wolfgang-Albert Flügel (2016). Daily suspended sediment simulation using machine learning approach. CATENA 138, 77–90.

[5] Kisi, O., A. H. Dailr, M. Cimen and J. Shiri (2012). Suspended sediment modeling using genetic programming and soft computing techniques. Journal

of Hydrology 450: 48-58.

[6] Kisi, O. and J. Shiri (2012). River suspended sediment estimation by climatic variables implication: comparative study among soft computing techniques. Computers & Geosciences 43: 73-82.

[7] Kisi, O., I. Yuksel and E. Dogan (2008). Modelling daily suspended sediment of rivers in Turkey using several data-driven techniques/Modélisation de la charge journalière en matières en suspension dans des rivières turques à l'aide de plusieurs techniques empiriques. Hydrological sciences journal 53(6): 1270-1285.

[8] Lafdani, E. K., A. M. Nia and A. Ahmadi (2013). Daily suspended sediment load prediction using artificial neural networks and support vector machines. Journal of Hydrology 478: 50-62.

[9] Liu, Q.-J., Z.-H. Shi, N.-F. Fang, H.-D. Zhu and L. Ai (2013). Modeling the daily suspended sediment concentration in a hyperconcentrated river on the Loess Plateau, China, using the Wavelet–ANN approach. Geomorphology 186: 181-190.

[10] Mustafa, M., M. Isa and R. Rezaur (2011). A comparison of artificial neural networks for prediction of suspended sediment discharge in river- a case study in Malaysia. World Academy of Science, Engineering and Technology 81: 372-376.

[11] Mustafa, M., M. Isa and R. Rezaur (2012). Artificial neural networks modeling in water resources engineering: infrastructure and applications. World Academy of Science, Engineering and Technology 62: 341-349.

[12] Mustafa, M., R. Rezaur, S. Saiedi and M. Isa (2012). River suspended sediment prediction using various multilayer perceptron neural network training algorithms—a case study in Malaysia. Water resources management 26(7): 1879-1897.

[13] Mustafa, M. R. U., M. H. Isa and R. R. Bhuiyan (2011). Prediction of river suspended sediment load using radial basis function neural network-a case study in Malaysia. National Postgraduate Conference (NPC), 2011, IEEE.

[14] Oleg Makarynskyy, Dina Makarynska, Matthew Rayson, and Scott Langtry (2015). Combining deterministic modelling with artificial neural networks for suspended sediment estimates. Applied Soft Computing 35, 247-255.