Journal of Basic and Applied Engineering Research
p-ISSN: 2350-0077; e-ISSN: 2350-0255; Volume 3, Issue 3; January-March, 2016, pp. 229-233

http://www.krishisanskriti.org/Publication.html

Modeling Stage-Discharge Relationship in a Tidal River using Artificial Neural Network Models

Susmita Bakshi[1] and Kalyan Kumar Bhar[2]

[1]*Dept.of Civil Engineering Meghnad Saha Institute of Technology Kolkata-150, India*
[2]*Dept.of Civil Engineering IIEST Shibpur Howrah-711103, India*
E-mail: [1]bakshi.susmita@gmail.com, [2]kalyan@becs.ac.in

Abstract—*Stage–discharge relationships for coastal rivers having tidal flow are not unique. They are nonlinear and show multiple loops. Conventional regression methods which are based on simple one to one stage-discharge relationships oversimplify the underlying complex physical processes and result in significant errors in the estimated discharges. In this study artificial neural network models are used to model such complex relationships. Stage and discharge data from a river reach having tidal effect were used for training, validation and simulation of the Artificial Neural Network models.*

1. INTRODUCTION

Determination of flow rate accurately in rivers is very important for a number of hydrologic applications such as water resources planning and management, water, design of different hydraulic structures. However, collecting data for measurement of streamflow discharge continuously is costly and challenging, especially during high flood conditions. In practice, water stages are easier and less expensive to measure. They are recorded and converted into discharges by using a pre-established stage–discharge relationship. These relationships are usually referred to as rating curves. However, the stage–discharge relationship is not always simple and unique because discharge is not a function of stage only. It depends on other parameters like water surface slope, channel geometry, channel bed roughness, and unsteadiness of flow. In some cases, a combination of these factors may result in non-unique relationship that is usually reflected as multiple loops in the observed measurements of stage and corresponding discharge.

There are two different approaches for modeling stage-discharge relationships. In the first approach the unsteady, non-uniform flow equations are solved numerically provided accurate information on channel geometry and boundary conditions are available. The other approach is data driven modeling which is basically a method of nonlinear regression. Artificial Neural Network(ANN) is an example of such approach.

2. STUDY AREA AND DATA

In this study the stage and discharge data of the river Sacramento has been used [3]. The Sacramento River is tributary to San Francisco Bay, an arm of the Pacific Ocean, and during periods of low flow, tidal effects extend upstream beyond the city of Sacramento for at least 40km. The discharge measurements were made by USGS at intervals of about 1.25 hours during the course of 12 daily tidal cycles in the years 1957-60. The streamflow measuring section is at the site of the stage recorder in the city of Sacramento; the auxiliary stage recorder is 17.4km downstream near the town of Freeport. Local inflow into the 17.4km reach of channel is negligible. The reach itself is located far enough upstream on the estuary so that no reversal of flow occurs there. However, when upland discharge (streamflow) into the estuary was less than about 849.5cumec, the discharge was affected by tidal action, and unsteady flow exists in the reach. The relative magnitude of the tidal effect in the reach increases with decrease in the upland flow and with increase in the range in elevation between high and low tides at the mouth of San Francisco Bay. Twelve series of discharge measurements, made during the years 1957-60, were available. Each series of measurements extended over a period of about 33 hours in order to include one complete lunar day (approximately 24.8 hours), and in the course of each series, about 25 discharge measurements were made. As the data was not recorded at uniform time intervals, the data was regularized using the MATLAB time series tools.

3. DESIGN OF ANN MODELS

3.1. Artificial Neural Networks

Artificial Neural Networks (ANNs) are a biologically inspired computational model based on the functioning of human brain. Neural network models are developed by training the network to represent the relationships and processes that are inherent within the data. Being essentially non-linear regression models, they perform an input–output mapping using a set of interconnected simple processing nodes or neurons. Each neuron takes in inputs either externally or from other neurons and passes it through an activation or transfer function such as a logistic or sigmoid curve. Data enter the network through the input units arranged in what is called an input layer. These data are then fed forward through successive layers including the hidden layer in the middle to emerge from the output layer on the right. The inputs can be any combination of variables that are thought to be important for predicting the output; therefore, some knowledge of the hydrological system is important. The hidden layer is the essential component that allows the neural network to learn the relationships in the data. Multilayer perceptron (MLP) is one of the most commonly used neural network. The backpropagation algorithm is a variation of a gradient descent optimisation algorithm that minimises the error between the predicted and actual output values. The weighted connections between neurons are adjusted after each training cycle until the error in the validation data set begins to rise. The validation data set is a second data set that is given to the network to evaluate during training. Once the networks are trained to satisfaction, it can be put to operation when the new input data are passed through the trained network in its non-training mode to produce the desired model outputs. In order to validate the performance of the trained network before it is put into real operation, however, the operation mode is usually imitated by using the test data set.

3.2. Data Division for Model Generalization

The main objective of modeling the stage-discharge relationship (for any modeling) is the ability of the model to predict beyond the data set used for training, which is referred to as the ability of the model to generalize. There are several approaches by which the available data can be split into separate sets for training and testing. The hold out method is best for proper utilisation of the available data [2] . In this method, small subsets of the data are selected and withheld in turn. Each withheld subset is kept for independent testing. The remaining data are used for training of the model. The main advantage of this method is maximizing the utilization of available data, avoiding undesirable reduction of sample size.

Statistical properties of these 12 events are summarized as follows. The observed peaks of these runoff events included one peak larger than 500m^3/s, four peaks between 350 m^3/s and 450 m^3/s, and seven peaks between 200 m^3/s and 350 m^3/s. Each of the 12 significant runoff events is withheld one at a time to independently test the model accuracy. The remaining data are used for model training and to compute and update the network weights. However, during training, and depending on the network complexity, running ANN for too many epochs iterations may cause the network to memorize the training sample without learning to generalize to new data patterns and relationships. Therefore, the remaining data are divided into a training set and a validation set. This approach is known as cross validation and is used to avoid the problem of overfitting in ANN models. During the training process, the errors of both the training and the validation sets are computed and monitored. In the initial phases of training, errors of both the training and validations sets will continue to decrease with the increase of training iterations. When the network starts to over-fit the training subset of data, the validation error will start to increase. At this stage, the training process is stopped and the current set of network weights and thresholds is retained and assumed to be the optimal parameters of the model. This process is repeated for each of the 12 events and the performance of the developed ANN is assessed statistically.

3.3. ANN Structure and Selection of Input Variables

Before setting up an ANN, it is necessary to determine the number of hidden layers in the network and number of nodes in each hidden layer. By trial and error, it was found that two hidden layers were adequate for modeling the stage discharge relationship in this case. The number of nodes in the hidden layers were also selected in a similar manner. The number of nodes used in the hidden layers was kept between five and seven, depending on the number of input variables. In the present study, the input variables are selected using a combination of physical reasoning and a heuristic approach, which is discussed in [1]. In this approach, the number of input variables is increased gradually to assess their effect on the model performance. The minimum number of inputs were chosen as three, as the models with one or two inputs were not sufficient to model the complex stage-discharge relationship in a tidal river.

3.4. Performance Evaluation

Various statistical parameters are used as a measure of performance evaluation [1]. Statistical measures used in this study to assess the model performance, are the root mean square error RMSE, the coefficient of efficiency E, and the adjusted coefficient of efficiency E1, defined as follows:

$$\mathrm{RMSE}(\mathrm{m}^3/\mathrm{s}) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Q_{pi} - Q_{0i})^2} \quad (1)$$

$$\mathrm{RMSE}(\%) = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(Q_{pi}-Q_{0i})^2}}{\overline{Q_0}} \mathrm{X}100 \quad (2)$$

$$E(\%) = 1 - \frac{\sum_{i=1}^{n}(Q_{pi}-Q_{0i})^2}{\sum_{i=1}^{n}(Q_{pi}-\overline{Q_{0i}})^2} \quad (3)$$

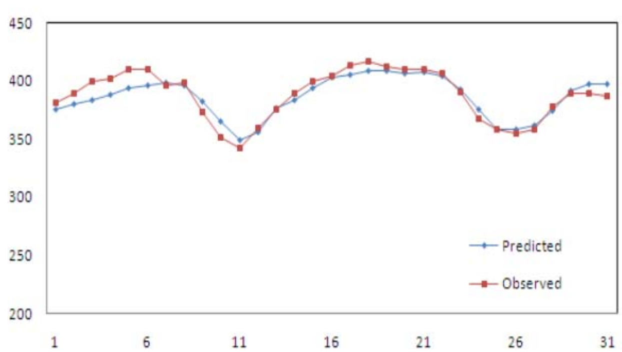
$$r = \frac{\sum_{i=1}^{N-1} t_i p_i}{\sqrt{\sum_{i=1}^{N-1} t_i^2}\sqrt{\sum_{i=1}^{N-1} p_i^2}} \quad (4)$$

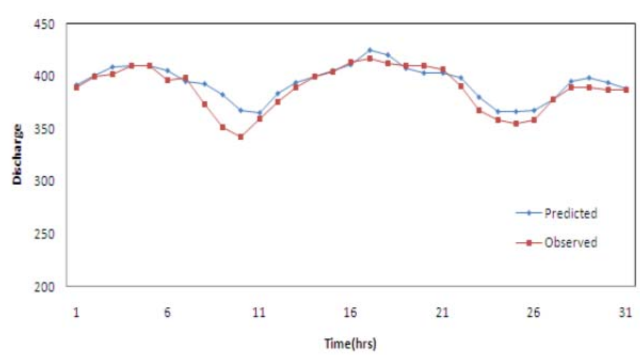
Where, Q_{0i} and Q_{pi} are observed and predicted discharges of the i th number of sample, respectively;

and n is the size of the stage–discharge sample. $\overline{Q_{0i}}$ denote the mean of Q_{0i} . The RMSE describes the average difference between model results and observations in units of the discharge variable and Eq.1 can be normalized to provide a relative measure with respect to the mean observed discharge as in Eq.2. Physically, the coefficient of efficiency, E, measures the differences between the observations and predictions relative to the variability in the observed data itself. According to Eq.3, E may range from -∞ to 1.0, where E=1.0 indicates a perfect model, E=0.0 indicates that the observed mean is as good a predictor as the model, and E < 0 indicates that the model is worse than using the observed mean as a predictor, or in other words the residual variance (described by the numerator in the expression above), is larger than the data variance (described by the denominator). The combined use of r, RMSE, and E, helps to assess each model's performance and compare the accuracy of any two modeling approaches. Coefficient of correlation r has been defined in Eq. 4. With increase in the number of inputs the performance of the models improved. Statistical measures of some of the tested cases are presented in Table 1.

Table 1: Statistical Analysis of ANN Prediction Accuracy.

Event	Parameters	Notation	r	RMSE(m3/s)	RMSE(%)	E
C	ht,ht-1,Qt-1	III	0.9363	7.76	2.00	0.868
C	ht-1,ht-2,Qt-1	IV	0.9327	10.14	2.61	0.758
C	ht,Qt-1,Qt-2	V	0.947	6.75	1.74	0.894
C	ht-1,ht-2,Qt-2	VI	0.867	10.51	2.71	0.747
C	ht-1, ht-2,ht-3, ht-4	IX	0.838	12.31	3.18	0.651
J	ht,ht-1,Qt-1	III	0.979	8.14	3.68	0.942
J	ht-1,ht-2,Qt-1	IV	0.846	21.11	9.5	0.619
J	ht,Qt-1,Qt-2	V	0.957	10.26	4.62	0.91
J	ht-1,ht-2,Qt-2	VI	0.844	20.37	9.17	0.644
J	ht,ht-1,ht-2, ht-3	VIII	0.844	19.87	8.98	0.387
J	ht-1, ht-2,ht-3, ht-4	IX	0.795	24.01	10.88	0.545





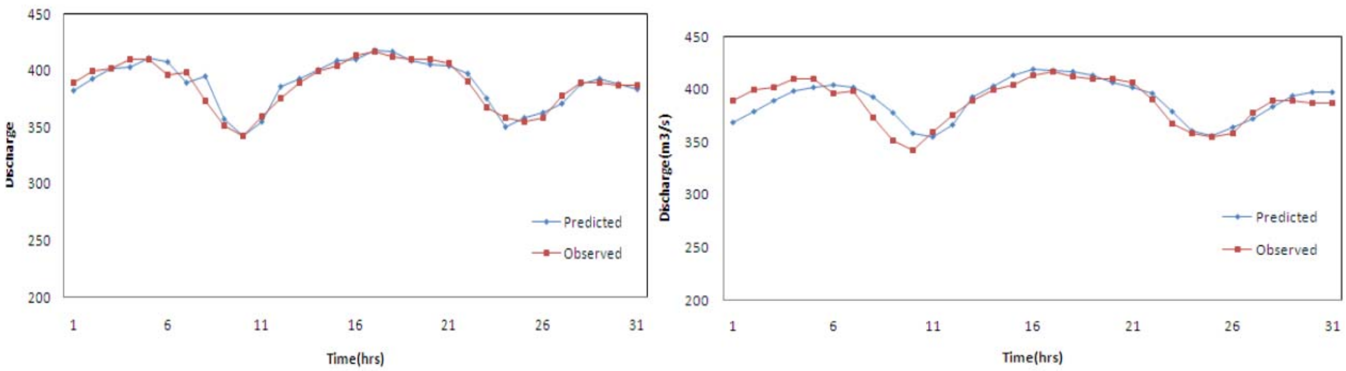


Fig. 1: Predicted and Observed tidal data for set C (i) Case III (ii)Case IV(iii) Case V (iv)Case VI

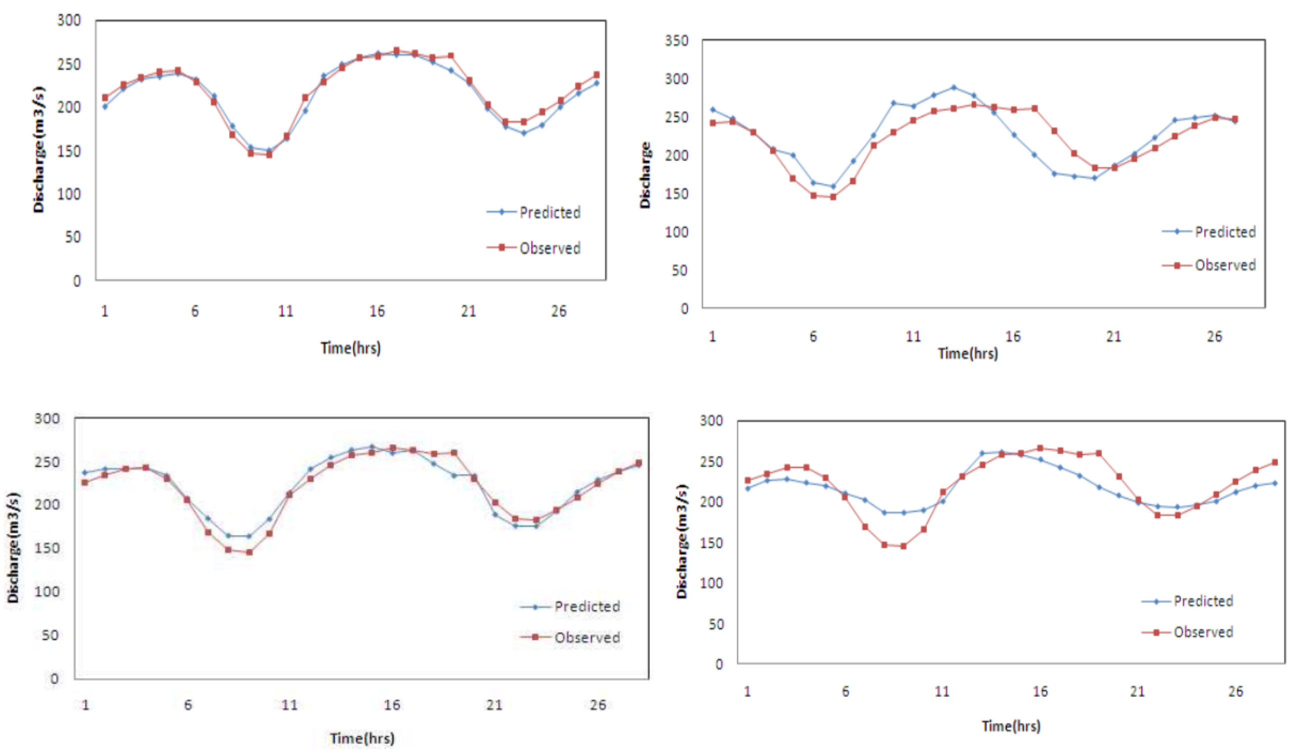


Fig. 2: Predicted and Observed tidal data for set J (i) Case III (ii) Case IX(iii) Case V(iv) Case VI

4. RESULTS AND DISCUSSION

Out of the 12 events (subsets of data) simulated, two events, Event C having a high range of discharge and Event J with comparatively low ranges of discharge are presented. Results of the following ANN models are presented. Model III, IV, V, VI, VIII and IX are presented. Fig.1 shows the plots of the observed and predicted series for event C and Fig. 2 shows the plots of the observed and predicted series for event J. Model III and V each having three numbers of input variables give the best result for both events C and J (in terms of the statistical parameters and plots). For high discharges Model VI gives good results. This shows that just increasing the number of input variables may not always improve the results. The variables used and the situations to be modeled are also to be considered.

5. CONCLUSION

The ANN Models has been developed for estimation of discharges, and can be used as an alternative of flow rating curves. A three-layered network is used. The number of nodes used in the hidden layers was kept between five and seven, depending on the number of input variables. In the present study, the input variables are selected using a combination of physical reasoning and a heuristic approach.

REFERENCES

[1] Habib E.H., Meselhe E.A. "Stage–Discharge Relations for Low-Gradient Tidal Streams Using Data-Driven Models," *Journal of Hydraulic Engineering, ASCE,* 482-492, May 2006

[2] Maier, H.R., Dandy, G.C. "Neural networks for the prediction and forecasting of water resources variables: a review of modeling issues and applications," *Environmental Modeling and Software* 15 (1), 101–124, 2000

[3] Rantz S.E. "An empirical method of determining momentary discharge of tide-affected streams*" Hydrology of tidal streams, Geological survey water-supply paper*, 1586-D, 1963