Modeling of stage-discharge relationship for Gharraf River, southern Iraq by using data driven techniques: A case study

A. M. Atiaa
Department of Geology, College of Sciences, University of Basra, Basra, Iraq
e-mail: alaaatiaa@gmail.com

Abstract: The potential of using three different data driven techniques namely multilayer perceptron with backpropagation neural network (MLP), M5P decision tree model, and Takagi-Sugeno (TS) inference system for mimic stage-discharge relationship at Gharraf River system, southern Iraq have been investigated and discussed in this study. The study used the available stage and discharge data for predicting discharge using different combinations of stage, antecedent stages, and antecedent discharge values. The models’ results were compared using root mean squared error (RMSE) and coefficient of determination ($R^2$) error statistics. The results of the comparison in testing stage reveal that M5 and Takagi-Sugeno techniques have certain advantages for setting up stage-discharge than multilayer perceptron artificial neural network. Although the performance of TS inference system was very close to that for M5 model in terms of $R^2$, the M5 method has the lowest RMSE (8.10 m$^3$/s). The study implies that both M5 and TS inference systems are promising tool for identifying stage-discharge relationship in the study area.

Key words: stage-discharge relationship, M5 model, artificial neural network, Gharraf River, Iraq

1. INTRODUCTION

The reliable estimation of river flow rate (discharge) is a prerequisite and crucial component for hydrological applications and analyses. Because of the dynamic nature of hydrological system, direct measurements of discharge are typically time consuming, costly and even impossible, especially during flood. Therefore, most discharge records are derived from converting the measured water levels (stages) to discharges by a functional relationship that is expressed as a rating curve. A calibrated stage-discharge rating offers an easy, cheap, and fast technique to estimate discharge (WMO 1980; Kennedy 1984; Herschy 1999). Stage-discharge rating is generally treated as the following power curve (Herschy 1999):

$$Q = b(a + H)^\alpha$$  \hfill (1)

where $Q$ is the discharge, $H$ is the stage, $\alpha$ is an index exponent, $a$ and $b$ are constants (depending on the study area).

Unfortunately, the functional relationship between stage and discharge is complex, time-varying, and cannot always captured by simple rating curve, even with the help of traditional modeling techniques such as polynomial regression or auto regressive integrated moving average ARIMA technique (Bhattacharya and Solomatine 2000). Many researchers attempt to establish this relation via data driven techniques such as artificial neural networks ANNs (Tawfik et al. 1977; Bhattacharya and Solomatine 2000; Sudheer et al. 2003; Bisht et al. 2010; Goel 2011), decision trees (Bhattacharya and Solomatine 2003; Ghimire and Reddy 2010; Ajmera and Goyal 2012), support vector machine (Aggarwal et al. 2012), wavelet-regression model (Kişi 2011), Takagi-Sugeno fuzzy inference system (Lohani et al. 2006), evolutionary based data driven models (Ghimire and Reddy 2012; Azamathulla et al. 2011). The results approve that these techniques are very efficient and reliable.
The aim of this study is to investigate the potential of the different data-driven models (artificial neural networks, fuzzy inference system, and M5P decision trees) to emulate stage-discharge rating curve of the Gharraf River at Hay, south of Iraq. Daily records of the stage and discharge are available for this river at Hay station for the period from April 2005 to May 2006. The performance of these techniques was compared and the best one with smaller estimation error was selected for future estimation of discharge from available data of previous discharge and stage values.

2. MATERIAL AND METHODS

2.1 Artificial neural networks

Artificial neural networks (ANNs) are massively parallel systems composed of many processing elements connected by links of variable weights. Given sufficient data and complexity, ANNs can be trained to model any relationship between a series of independent and dependent variables. For this reason, ANNs are considered to be universal approximates and have been successfully applied to a wide variety of problems that are difficult to understand, define and quantify. There are many different types of ANNs based on topology. One of the many ANN paradigms, the Multilayer Perceptron (MLP) network, is by far the most popular (Lippman 1987). The MLP is a layered feed-forward network which is typically trained with static Backpropagation (BP) algorithm. MLP is capable of approximating any measurable function from one finite-dimensional space to another within a desired degree of accuracy (Hornik et al. 1989). The MLP network consists of layers of parallel processing nodes. Each layer is fully connected to the preceding layer by interconnection strength, or weights, w. Figure 1 presents a three-layer MLP neural network consisting of layers i, j, and k, with interconnection weights wij and wjk between layers of neurons. Each neuron in a layer receives and processes weighted input from a previous layer and transmits its output to nodes in the following layer through links. The connection between i\textsuperscript{th} and j\textsuperscript{th} neuron is characterized by the weight coefficient wij and the i\textsuperscript{th} neuron by the threshold coefficient \( i\vartheta \). The weight coefficient reflects the degree of importance of the given connection in the network. The output value of the ith neuron \( x_i \) is computed as follows (Haykin, 1998):

\[
x_i = f(\xi_i)
\]

with \( \xi_i = \vartheta_i + \sum_{j} w_{ij} x_j \)  

where \( f(\xi_i) \) is the activation function. The threshold coefficient can be understood as a weight coefficient of the connection. With formally added neuron j, where \( x_j = 1 \), Sigmoid shape activation functions are normally defined as:

\[
f(\xi_i) = \frac{1}{1 + e^{-\xi_i}}
\]

The Backpropagation algorithm works by computing the error between the network output and the corresponding target value and propagating this backward through the network to update the weights. The weight updates are calculated based on:

\[
\Delta w_{ij}(t) = -\eta \frac{\partial E}{\partial w_{ij}} + \mu \Delta w_{ij}(t-1)
\]
where \( \eta \) and \( \mu \) are the learning and momentum rates, respectively. \( E \) is the error, or objective function, and \( \Delta w_i(t) \) and \( \Delta w_j(t-1) \) are the weight increments between nodes \( i \) and \( j \) for iterations \( t \) and \( t-1 \). A detail description of this algorithm can be found in Fausett (1994) and Haykin (1998).

![Multilayer perceptron with one hidden layer architecture](image)

**Figure 1. Multilayer perceptron with one hidden layer architecture**

### 2.2 M5P decision tree

A decision tree is a logical model represented as a binary (two-way split) tree that shows how the values of a target (dependent) variable can be predicted by using the values of a set of predictor (independent) variables. There are basically two types of decision trees: (1) classification trees are the most common and used to predict a symbolic attribute (class) (2) regression trees which are used to predict the value of a numeric attribute (Witten and Frank 2005). If each leaf in the tree contains a linear regression model, which is used to predict the target variable at that leaf, then it called a model tree.

The M5P model tree algorithm was originally developed by Quinlan (1992). Detail description of this technique is beyond of this paper and can be found in Witten and Frank (2005). A short description of this technique follows. The M5P algorithm constructs a regression tree by recursively splitting the instance space using tests on a single attributes that maximally reduce variance in the target variable. Figure 2 illustrates this concept. The formula to compute the standard deviation reduction (SDR) is (Quinlan 1992):

\[
SDR = sd(T) - \sum \frac{|P|}{|T|} sd(T_i)
\]

where \( T \) represents a set of example that reaches the node; \( T_i \) represents the subset of examples that have the \( i^{th} \) outcome of the potential set; and \( sd \) represents the standard deviation.

After the tree has been grown, a linear multiple regression is built for every inner node using the data associated with that node and all the attributes that participate for tests in the subtree to that node. After that, every subtree is considered for pruning process to overcome the overfitting problem. Pruning occurs when the estimated error for the linear model at the root of a subtree is smaller or equal to the expected error for the subtree. Finally, the smoothing process is employed to compensate for the sharp discontinuities between adjacent linear models at the leaves of the pruned tree.
2.3 Fuzzy logic

The term “fuzzy logic” has in fact two distinct meanings. In a narrow sense, it is viewed as a generalization of classical multi valued logics (Demicco and Klir 2004). In a broad sense, fuzzy logic is viewed as a system of concepts, principles, and methods for dealing with modes of reasoning that are approximate rather than exact. The fuzzy logic system is a cognitive artificial intelligence scientific technique developed in 1965 by Professor Lotfi Zadeh of the Department of Computer Science, University of California, Berkeley, USA. It provides a means of representing uncertainties and vagueness that characterize human perception, judgmental reasoning, and decision (Emamiet al. 1998). The generation of a fuzzy model is based on expert knowledge and historical data. Fuzzy inference is the process of formulating the mapping from a given input to an output equation using fuzzy logic, and then the mapping provides a basis from which decisions can be made or discerned. The fuzzy inference system (FLS) consists of four main interconnected components (Figure 3): rules, fuzzifier, inference engine, and output processor. Once the rules have been established, a fuzzy logic system can be viewed as a map from inputs to outputs. The rules are the heart of a FLS and can be expressed as a collection of IF-THEN statements. The IF part of a
rule is antecedent and the THEN part is consequent. Depending on the particular structure of the consequent proposition, three main types of fuzzy models are distinguished as:

(1) Linguistic (Mamdani Type) fuzzy model (Zadeh, 1973; Mamdani, 1977),
(2) Fuzzy relational model (Pedrycz, 1984; Yi and Chung, 1993),
(3) Takagi–Sugeno (TS) fuzzy model (Takagi and Sugeno, 1985).

In this paper, the TS fuzzy model is employed to emulate stage-discharge rating curve, so, a brief description of this method is outlined below.

\[ R_i = IF \text{x is } A_i, \text{THEN } y_i = a_i^T x + b_i \quad i = 1, 2, \ldots, M \]  

(7)

where \( x \in \mathbb{R}^p \) is the antecedent and \( y_i \in \mathbb{R} \) is the consequent of the \( i^{th} \) rule. In the consequent, \( a_i \) is the parameter vector and \( b_i \) is the scalar offset. The number of rules is denoted by \( M \) and \( A_i \) is the (multivariate) antecedent fuzzy set of the \( i^{th} \) rule defined by the membership function:

\[ \mu_i(x): \mathbb{R}^p \rightarrow [0,1] \]  

(8)

The fuzzy antecedent in the TS model is defined as an and-conjunction by means of the products operator (Wolfs and Willems 2013):

\[ \mu_i(x) = \prod_{j=1}^p \mu_{ij}(x_j) \]  

(9)

where \( x_j \) is the \( j^{th} \) input variable in the \( p \) dimension of input data space, and \( \mu_{ij} \) the membership degree of \( x_j \) to the fuzzy set describing the \( j^{th} \) premise part of the \( i^{th} \) rule, \( \mu_i(x) \) is the overall truth value of the \( i^{th} \) rule.
\[ y = \sum_{i=1}^{M} u_i(x) \cdot y_i \]  

(10)

where \( u_i \) is the normalized degree of fulfillment of the antecedent clause of rule \( R_i \) (Setnes 2000):

\[ u_i = \frac{\mu_i(x)}{\mu_i(x)} \sum_{j} \mu_j(x) \]  

(11)

The \( y_i \)'s are called consequent functions of the M rules and defined by:

\[ y_i = W_{i0} + W_{i1}x_1 + W_{i2}x_2 + \ldots + W_{ip}x_p \]  

(12)

where \( W_{ij} \) are the linear weights for the \( i \)th rule consequent function.

3. STUDY AREA AND DATA DESCRIPTION

The Gharraf River system is located in Mesopotamian plain, southern Iraq (Figure 4). The drainage area of this system is \( 435052 \times 10^6 \) m\(^2\). The river begins in the Kut Barrage and runs south between the great Euphrates and Tigris Rivers, and ends in Al-Hammer marsh land in Nassyria city. The main length of the river is approximately 230 km. The Gharraf area is characterized by hot and dry summer and cold and wet winter. The climate of the area is classified as semi-arid one. The course of the Shatt Al Gharraf can be subdivided according to the conditions that governed its development as follows (Iraqi Ministries of Environment, Water resources, Municipalities and Public works 2006):

1. The Hai Delta, which ends at Kalaat Sukkar in which expansion can take place,
2. The Rafai gully extends to about 10 km upstream of Bada’a in which flow is restricted, no lateral expansion being possible,
3. The Bada’a Delta is the most recent region of expansion on the left bank towards the Hor Abu Ijul, HorH’weynah and HorGhamukah depressions, and
4. The Shattrah and Kasser-Ibrahim Deltas are the regions of expansion at the end of the Rafai gully.

The daily averages of stage and discharge data for the Hay station on the Gharraf River were used in this study. The observed data covers the period from April 2005 to May 2006. In Iraq, it is difficult to obtain enough time series to build data driven models; hence, approximately one year was used. The available data are arbitrary divided into two parts sets; 66% for training and 34% for testing for all models developed in this study. The statistical parameters of the used data were given in Table 1. In this table, \( N, \text{Min}., \text{Max.}, \bar{x}, \text{Me}, s, C_v, \text{and } K_s \) refer to total number of data, minimum, maximum, arithmetic average, standard deviation, coefficient of variation, and coefficient of skewness, respectively. From Table 1, one could conclude that variation of discharge values is higher than that for stage. The maximum values of stage in testing set are higher than that for training set, this may cause difficulty to estimate discharge at extreme values. One the other hand, the maximum and minimum values of discharge in testing set fall within the range in training test. This may overcome the problem of estimation extreme discharge values which previously mention.
Figure 4. Map of the Gharraf River system, southern Iraq.

Table 1. Summary of statistical parameters of the used data

<table>
<thead>
<tr>
<th>Data set</th>
<th>N</th>
<th>Min.</th>
<th>Max.</th>
<th>$\bar{x}$</th>
<th>Me</th>
<th>s</th>
<th>$C_v$</th>
<th>$K_s$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall</td>
<td>H</td>
<td>331</td>
<td>8.75</td>
<td>10.95</td>
<td>9.7</td>
<td>0.23</td>
<td>2.39</td>
<td>2.31</td>
</tr>
<tr>
<td></td>
<td>Q</td>
<td>331</td>
<td>75</td>
<td>175</td>
<td>145.48</td>
<td>150</td>
<td>20.54</td>
<td>14.12</td>
</tr>
<tr>
<td>Training</td>
<td>H</td>
<td>218</td>
<td>8.75</td>
<td>10.95</td>
<td>9.71</td>
<td>9.7</td>
<td>0.27</td>
<td>2.76</td>
</tr>
<tr>
<td></td>
<td>Q</td>
<td>218</td>
<td>75</td>
<td>175</td>
<td>146.21</td>
<td>150</td>
<td>19.19</td>
<td>13.12</td>
</tr>
<tr>
<td>Testing</td>
<td>H</td>
<td>113</td>
<td>9.2</td>
<td>10.2</td>
<td>9.71</td>
<td>9.7</td>
<td>0.138</td>
<td>1.43</td>
</tr>
<tr>
<td></td>
<td>Q</td>
<td>113</td>
<td>75</td>
<td>173</td>
<td>144.7</td>
<td>150</td>
<td>22.07</td>
<td>15.25</td>
</tr>
</tbody>
</table>

4. PERFORMANCE CRITERIA FOR THE DEVELOPED MODELS

The performance of the various data driven models were evaluated by means of errors statistics criteria such as root mean squared error (RMSE) and coefficient of determination ($R^2$). The mathematical formulations of these criteria are outlined below:

(a) Root Mean Square Error (RMSE)
where $Q_i$ is the measured discharge and $\hat{Q}$ is the simulated discharge, $n$ is the number of observations (instants). As the value of this criterion approaches zero, the better fit between observed and modeled data is obtained.

(b) Coefficient of Determination

$$R^2 = 1 - \frac{SSE}{SSy}$$

where

$$SSE = \sum_{i=1}^{n} (Q_i - \hat{Q}_i)^2$$

$$SSy = \sum_{i=1}^{n} (Q_i - \bar{Q})^2$$

where $\bar{Q}$ is the arithmetic mean of the observed $Q$. The better the fit, the closer $R^2$ is to ±1.

5. APPLICATIONS OF THE TECHNIQUES

In this study, feed forward neural network (MLP) with backpropagation algorithm was employed for developing ANN models. The popularity of MLP in hydrological application (Zhang and Govindaraju 2003; Leahy et al. 2008) is the main reason for selecting this network. Although, the architecture of MLP can have many hidden layers, works by Cybenco (1989) and Coulibaly et al. (1999) have been shown that a single hidden layer is sufficient for the MLP to approximate any complex non-linear function. For all the developed models, the Levenberg–Marquard algorithm was applied to train the networks. The logistic sigmoid transfer function is used in the hidden layer and a linear one in the output layer for all the developed networks. The early stopping method was selected to overcome over fitting problem. Demo version of Alyuda NeuroIntelligent commercial software was used in this study to build different neural networks. NeuroIntelligence is a neural network software for experts. It is used to apply neural networks to solve real-world forecasting, classification and function approximation problems. It is full-packed with proven techniques for neural network design and optimization. To ensure that each variable is treated equally in the models, all the input and output data were automatically normalized into the range [-1, 1]. The default values of learning rate (0.1) and momentum rate (0.1) were used for build network models. The number of nodes in the hidden layer for each developed models were determined by trial and error procedure considering the need to drive reasonable results.

The study examined various combinations of river stage $H_t$ with specified lag times $H_{t-1}$ and $H_{t-2}$ and the antecedent discharges $Q_{t-1}$ and $Q_{t-2}$ as inputs to the ANN models in order to evaluate the degree of effect of each of these variables on output variable $Q_t$. The input combinations evaluated in the present study are shown in Table 2. The same variable input combinations were also used for M5P and TS fuzzy inference system techniques. Also, to reduce network error, different numbers of iterations for the best network were examined. These test conduct in order to verify whether an increase iteration numbers could reduce error rate or not.
Table 2. Input combinations for the developed models

<table>
<thead>
<tr>
<th>Model</th>
<th>Input combinations</th>
<th>Output variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>( H_t )</td>
<td>( Q_t )</td>
</tr>
<tr>
<td>2</td>
<td>( H_t, Q_{t-1} )</td>
<td>( Q_t )</td>
</tr>
<tr>
<td>3</td>
<td>( H_{t-1}, H_t, Q_{t-1} )</td>
<td>( Q_t )</td>
</tr>
<tr>
<td>4</td>
<td>( H_{t-1}, H_t, Q_{t-2}, Q_{t-1} )</td>
<td>( Q_t )</td>
</tr>
<tr>
<td>5</td>
<td>( H_{t-2}, H_{t-1}, H_t, Q_{t-2}, Q_{t-1} )</td>
<td>( Q_t )</td>
</tr>
</tbody>
</table>

For building M5P models, Weka3.6 software was used. Weka is open-source machine learning/data mining software written in Java (Witten and Frank 2005). The software contains a comprehensive set of pre-processing tools, learning algorithms and evaluation methods. For this study, the parameters of M5P algorithm were set to their default values; pruning factor 4.0 and smoothing option. The software was available on http://www.cs.waikato.ac.nz/~ml.

A fuzzy toolbox in Matlab 2012a software was used for building fuzzy models. Membership functions were extracted via subtractive clustering method. Subtractive clustering method (Chiu 1994) is an extension of the mountain clustering method, where data points (not grid points) are considered as the potential candidates for cluster centers. It uses the positions of the data points to calculate the density function, thus reducing the number of calculations significantly. Since each data point is a candidate for cluster centers, a density measure at data point \( x_i \) is defined as (Chiu 1994):

\[
D_i = \sum_j \exp\left(-\frac{\|x_i - x_j\|^2}{(r_a/2)^2}\right)
\]

where \( r_a \) is a positive constant representing a neighborhood radius. Therefore, a data point will have a high-density value if it has many neighboring data points. A trial and error procedure was used to assigning a suitable value of calculus radius. After many trials the best result was 0.2. Three Gaussian membership functions were extracted for each model which were labeled as low, medium, and high. The same labels were used for \( Q_t \). Default values of the TS inference system were used in this study.

6. RESULTS AND DISCUSSIONS

The RMSE and \( R^2 \) statistics of each ANN model in testing period were given in Table 3. The ANN model whose inputs were \( H_{t-1}, H_t, Q_{t-2}, \) and \( Q_{t-1} \) (input combination 4) with [4 15 1] architecture has the smallest RMSE (9.91 m³/s) and the highest \( R^2 \) (0.82). As shown in Table 3, using only the stage \( H_t \) (input combination 1) gives poor estimate with the RMSE (21.99) and \( R^2 \) (0.05). Among the ANN models, whose inputs were the antecedent discharges (input combinations 2, 3, 4, and 5), the ANN model with \( Q_{t-1} \) has the biggest RMSE (12.06 m³/s) and the lowest \( R^2 \) (0.67). This emphasizes that the \( Q_t \) is mostly depended on the antecedent discharge values. Among the ANN models, whose inputs were the antecedent stages (input combinations 3, 4, and 5), the ANN model with inputs \( H_{t-2}, H_{t-1}, \) and \( H_t \) has the biggest RMSE (12.03 m³/s) and the lowest \( R^2 \) (0.72). In general, all the developed ANN models except ANN-1 and ANN-2 with [2 3 1] have good capabilities to emulate stage – discharge relationship because they have reasonable RMSE and \( R^2 \). Table 3 also shows that the increasing of hidden nodes brought slightly better performance for the developed models. The \( Q_t \) estimates of the best performance models were also represented graphically in Figure 5. It is obviously seen from these figures that measured and estimated discharge was reasonably good. All the figures show that the estimated discharge \( Q_t \) for all the developed models were underestimated especially with the lowest values of discharge.
Table 3. Statistical performance criteria for one hidden layer ANN's models

<table>
<thead>
<tr>
<th>Model</th>
<th>Input combinations</th>
<th>ANN architecture</th>
<th>Testing set</th>
<th>RMSE</th>
<th>R²</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANN-1</td>
<td>Ht</td>
<td>[ 1 3 1 ]</td>
<td>21.99</td>
<td>0.05</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>[ 2 3 1 ]</td>
<td>12.06</td>
<td>0.67</td>
<td></td>
</tr>
<tr>
<td>ANN-2</td>
<td>Ht, Qt</td>
<td>[ 2 5 1 ]</td>
<td>10.67</td>
<td>0.75</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>[ 2 10 1 ]</td>
<td>10.97</td>
<td>0.73</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>[ 3 3 1 ]</td>
<td>10.58</td>
<td>0.79</td>
<td></td>
</tr>
<tr>
<td>ANN-3</td>
<td>Ht, Ht, Qt</td>
<td>[ 3 8 1 ]</td>
<td>10.47</td>
<td>0.76</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>[ 3 15 1 ]</td>
<td>10.58</td>
<td>0.82</td>
<td></td>
</tr>
<tr>
<td>ANN-4</td>
<td>Ht, Ht, Qt, Qt</td>
<td>[ 4 5 1 ]</td>
<td>10.16</td>
<td>0.79</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>[ 4 15 1 ]</td>
<td>9.91</td>
<td>0.82</td>
<td></td>
</tr>
<tr>
<td>ANN-5</td>
<td>Ht, Ht, Ht, Ht, Qt</td>
<td>[ 5 5 1 ]</td>
<td>12.3</td>
<td>0.72</td>
<td></td>
</tr>
</tbody>
</table>

Table 4 presents the statistical performances of M5 decision tree models in which the model whose inputs were Ht and Qt-2 (input combination 2) was the best model among all other developed models with lowest RMSE and R², 8.10 and 0.88, respectively. The other models also perform best except the MT1 with single input H value. Figure 6 shows a graphical comparison between measured and estimated discharges. It is obvious from Figure 5 that the MT2-5 four models have very good agreement between measured and estimated discharges for both low and high values. For the MT5 model, the following rule was extracted form M5 algorithm:

\[
Q_{t-1} \leq 136.5 : LM1 (61/39.078\%)
\]

\[
Q_{t-1} > 136.5 :
| Q_{t-1} \leq 151.5 : LM2 (102/31.145\%)
| Q_{t-1} > 151.5 : LM3 (54/64.323\%)
\]

LM num: 1

\[
Q_t = -17.2053 * H_{t-2} + 2.2944 * H_{t-1} + 0.2323 * Q_{t-2} + 0.1694 * Q_{t-1} + 2.992 * H + 237.7584
\]

LM num: 2

\[
Q_t = -0.4746 * H_{t-2} - 2.5982 * H_{t-1} - 0.0163 * Q_{t-2} + 0.6489 * Q_{t-1} + 3.1867 * H + 53.6424
\]

LM num: 3

\[
Q_t = -0.4746 * H_{t-2} - 14.9457 * H_{t-1} - 0.0276 * Q_{t-2} + 0.2338 * Q_{t-1} + 19.5879 * H + 85.3705
\]

For the MT2 with minimal input data (input combinations 2), the following tree was extracted:

LM num: 1

\[
Q_t = 0.8555 * Q_{t-1} + 21.2475
\]

The statistical performance of TS inference system was shown in Table 5. Results of this data driven model were similar to that of M5 model. The TS2 with two inputs parameter, i.e. H and Qt-1 was the best among the other models with lowest RMSE (8.17) and R² (0.88). The worst model was the model whose input was stage only. The other three models (TS3-5) also perform very well where both high and low values were reasonable predicted (Figure 7). The TS2 was selected in this study a candidate for comparison with other data driven models because it has minimal input data and perform the best for all other developed models as mention previously. The membership editor
and fuzzy rules for this model were shown in Figures 8 and 9, respectively. Three simple fuzzy rules were generated for this model. These are:

\[
\begin{align*}
IF \ Q_{t-1} \text{ is low and } H \text{ is low} & \quad \text{THEN } Q \text{ is low} \\
IF \ Q_{t-1} \text{ is medium and } H \text{ is medium} & \quad \text{THEN } Q \text{ is medium} \\
IF \ Q_{t-1} \text{ is high and } H \text{ is high} & \quad \text{THEN } Q \text{ is high}
\end{align*}
\]

The TS inference system for this model was illustrated in Figure 10.

The comparison between the three data driven best models were presented in Table 6. The best data driven model for estimating \( Q_t \) was M5 model tree. Although, the performance of TS inference system was very close to that for M5 model in terms of \( R^2 \), the M5 method has the lowest RMSE (8.10 m\(^3\)/s). Results also reveal that the M5 model performed better than the ANN model for both low and high discharge predictions. The complex structure of ANN and the many parameters which must be assigned for successful training make the ANN a second priority when it compare with the simple structure and very fast training M5 algorithm. The generated tree structure with linear models on the leaves bears another benefit for this technique; it was very easy to understand even from those people who are unfamiliar with hydrology. The same results obtained by Ajmera and Goyal (2012) when they compared between ANN and M5 techniques for mimic flow rating curve. The result of this study agree with Ajmer and Goyal (2012) and added another comparison, i.e. between TS inference system and M5 which enhance the capability of M5 model for emulating stage-discharge relationship. The results also indicated that TS and MT models that used only two variables (\( Q_{t-1} \) and \( H \)) were very good for prediction \( Q_t \) for the study area.

<table>
<thead>
<tr>
<th>Model</th>
<th>Input combinations</th>
<th>Testing set</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>RMSE</td>
</tr>
<tr>
<td>DT1</td>
<td>( H_t )</td>
<td>16.44</td>
</tr>
<tr>
<td>DT2</td>
<td>( H_t, Q_{t-1} )</td>
<td>8.10</td>
</tr>
<tr>
<td>DT3</td>
<td>( H_{t-1}, H_t, Q_{t-1} )</td>
<td>8.32</td>
</tr>
<tr>
<td>DT4</td>
<td>( H_{t+1}, H_t, Q_{t+1}, Q_{t-1} )</td>
<td>8.32</td>
</tr>
<tr>
<td>DT5</td>
<td>( H_t, H_{t+2}, H_t, Q_{t+2, 1} )</td>
<td>8.32</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model</th>
<th>Input combinations</th>
<th>Testing set</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>RMSE</td>
</tr>
<tr>
<td>TS1</td>
<td>( H_t )</td>
<td>22.22</td>
</tr>
<tr>
<td>TS2</td>
<td>( H_t, Q_{t-1} )</td>
<td>8.17</td>
</tr>
<tr>
<td>TS3</td>
<td>( H_{t-1}, H_t, Q_{t-1} )</td>
<td>8.31</td>
</tr>
<tr>
<td>TS4</td>
<td>( H_{t+1}, H_t, Q_{t+1}, Q_{t+1} )</td>
<td>8.46</td>
</tr>
<tr>
<td>TS5</td>
<td>( H_{t+2}, H_{t+1}, H_t, Q_{t+2, 1} )</td>
<td>8.44</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model</th>
<th>Testing set</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RMSE</td>
</tr>
<tr>
<td>ANN</td>
<td>9.91</td>
</tr>
<tr>
<td>DT2</td>
<td>8.10</td>
</tr>
<tr>
<td>TS2</td>
<td>8.17</td>
</tr>
</tbody>
</table>
Figure 5. Comparison between measured and estimated discharge and best fit lines for best Performance ANN's models
(a) ANN-3 [3 3 1]  (b) ANN-3 [3 15 1]  (c) ANN-4 [4 5 1]  (d) ANN-4 [4 15 1]  and (e) ANN-6 [3 8 8 1].
Figure 6. Comparison between measured and estimated discharges and best fit lines for the best performance of M5 models.
7. CONCLUSIONS

The abilities of the artificial neural networks, M5 decision trees, and Takagi and Sugeno fuzzy inference techniques for emulating stage-discharge relationship for Gharraf River system, southern Iraq have investigated and discussed in this study. The study demonstrated that modeling of this relationship is possible through using these techniques. The M5 decision tree technique models with minimal data ie, current stage and one antecedent discharge perform better than that ANN models and TS inference engine. The root mean squared error and correlation of determination for best M5 model were (8.17 m$^3$/s) and (0.88), respectively. The best M5 and TS model were able to prediction discharge on both high and low values. Most of the developed ANN models were slightly capable to predict the discharge but most predictions were underestimating. All the developed models with stage as a single input failed to mimic stage-discharge relationship. This implies that antecedent discharges were needed for better relationship at this area. The study used data from one station and further studies using more data may enhance the results getting by this study.

Figure 7. Comparison between measured and estimated discharge and best fit lines for the best Performance of TS inference engines.

Figure 8. Membership editor for TS2 inference engine.
REFERENCES


